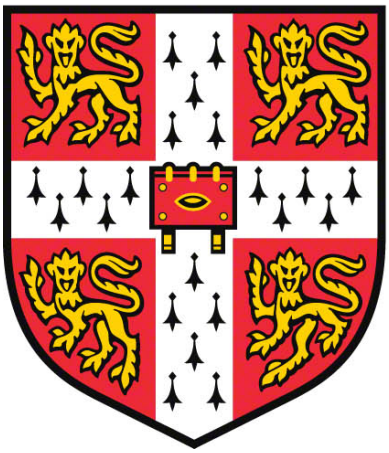
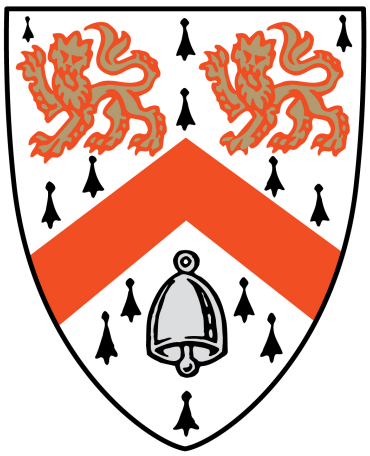
From disorder to order:
the importance of context in protein folding and binding mechanisms





Carolina Altomar Testa Furtado de Mendonça

Wolfson College

University of Cambridge

September 2019

This dissertation is submitted for the degree of *Doctor of Philosophy*

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the Preface or specified in the text. It is not substantially the same as any that I have submitted for a degree or diploma or other qualification at the University of Cambridge or any other University, and no part has already been or is concurrently being submitted for any degree, diploma or other qualification. It does not exceed the 60,000-word limit, as imposed by the Physics and Chemistry Degree Committee.

Carolina A. T. F. de Mendonça

September 2019

Summary

Anfinsen's seminal work has shown that the information required for a protein to fold into its specific three-dimensional structure is encoded into its amino acid sequence. The protein structure was believed to determine its activity, meaning that a protein needs to fold in order to function. More recently, intrinsically disordered proteins (IDPs) have been shown to represent a significant portion of the proteome. Despite the lack of a predefined structure, they still play important roles in cellular function, challenging the structure-function paradigm. Proteins are largely studied in isolated conditions, but in a cellular environment they are part of a vastly more complex system. The work presented here aims to shed light on how context can influence folding and binding mechanisms.

First, we used SasG – a bacterial protein that defies the disorder prediction with its unique sequence composition and unusual structure – as a template to investigate co-translational folding, and how the presence of the ribosome can affect its folding mechanism. SasG *in vitro* translation was investigated utilising force profile experiments. We showed that both the $G5^2$ and E-$G5^2$ constructs can fold very early, when still inside the vestibule of the ribosome. Moreover, our results suggest that non-native interactions can also provide sufficient force to release the stall sequence.

Next we employed protein members of the BCL-2 family – involved in controlling the cell death mechanism – to understand what encodes a coupled folding and binding reaction. Although displaying a variety of conformations, some IDPs can fold upon binding to a partner protein. Promiscuous binding is a great advantage of disordered molecules, as multiple IDPs are able to bind and fold to the same partner protein. This raises the question of what orchestrates a coupled folding and binding reaction: the IDP or the partner protein? Using phi-value analysis we studied four IDP-partner protein complexes, composed of alternative pairs of BCL-2 family members. In the bimolecular context, the disordered protein dictates the transition state interactions. Therefore, analogous to Anfinsen's postulate, the folding pathway is encoded by the protein that folds, even when binding to another macromolecule is required.

Finally, studies of the BCL-2 member BID on its full-length context showed that it cannot interact with its partner A1 unless it is cleaved and unfolded (tBID). These results provide insights on the role of tBID as a player during programmed cell death and hence why the pathway of cleaving BID with caspase is energetically favourable.

Acknowledgements

When I first arrived in Cambridge in September 2012, I had many dreams and aspirations, but I could have never imagined how impactful this experience would be in my life.

A few months into adapting to my new life (while still learning English), I started a job as a lab technician at the Chemistry department. The lab 290 was shared by several groups, but one was particularly welcoming: the Clarke group. Intrigued by where all the media I was preparing was going to, I started to read the papers and have informal chats with students and post-docs about their work. That was when, on a typical morning in the lab, I had a conversation with Jane. After hinting my intentions to continue my academic career and my actual interest in knowing more about the research done by her group, she warmly considered my words. And that changed everything. A few weeks later, I started my first project in the lab, in a part-time/volunteer basis. That project evolved to a formal MPhil and provided the principles for the PhD thesis I now submit.

To Jane, my wholeheartedly gratitude. Thank you for believing in me. Thank you for giving me an opportunity to pursue my dreams when I thought I would never be able to. Thank you for being the best mentor and supervisor that I could have asked for. Your encouragement and guidance really pushed me to always wanting to do better.

The opportunity to join the Clarke group would not have been accomplished without Dominika. She supervised and trained me, aware of my limited experience in biochemistry. I still remember our very first meeting, where she patiently explained the fundamentals of molecular biology and protein biophysics. Certainly, I could not have had anyone more suitable for this role. Her rigorous scientific thinking and experimental skills provided lessons that I will carry with me forever. Thanks Dominika, for being such a fantastic supervisor and great friend. I will always miss sharing my bench and desk with you.

I am also grateful for the time spent with all Clarke members when I started: Sarah, Katie, Joe, Chi, Alex, Jeff, Andrew and Adrian. In the following years, new people have joined our group: Michael, Liza, Basile, Sophie, Annette, Tristan, Pavel, Quenton and Lee. I would like to thank Jeff for always be willing to help, especially for the training he provided and for designing the constructs for the ribosome project. Thanks to Lee for cheering me up in so many mornings and for the assistance in the challenging molecular biology involved in the ribosome project. I am

also very grateful to Annette. Her solid experience was impactful and brought me confidence in many occasions. I would like to thank her for doing repeats of my force profile experiments and for proofreading one of my chapters.

My experience would not have been so fulfilling without my buddies: Liza, Basile, Michael, Quenton and Tristan. I was close to Liza from the start. Her determination and talent to do a good science were very inspiring. She was always available to hear me — to discuss my exciting or frustrating results, as well as to share my personal feelings. Her friendship was essential during this PhD and is a jewel that I will always cherish. Basile was our daily boost of laugh and happiness. Always very busy but also happily available to help, he was fundamental in our group. His passion for science was continuously pushing and cheering us all. Michael would always do everything to encourage his peers. His analytical thinking and organised schedule would always motivate us do more. I am particularly grateful for the time we shared and for the support given by him during our joint project, which also hooked me to the "IDP side" of the group. The project was also shared with Quenton. Although brief, his time in the lab was very fruitful and filled with exciting scientific (and non-scientific) discussions. Tristan assisted us all in a very efficient and organised way. I am particularly grateful for all the protein production and purification that were essential to my work.

During my time in the lab, I had the chance to supervise Pavel, an Erasmus student. Despite the limited time, he made a huge effort to make the most out of his experience. I am grateful for the opportunity to work with such a determined and focused young scientist.

I would like to thank everyone that provided technical support and assistance in the department, particularly, Anita, Jo, Helen, Kevin, Gary, Ryan and Nathan. Your jobs contributed massively for everything to run smoothly and pleasantly. Thanks also to Peter Sharaat (Biochemistry department), for all the amino acid analysis and Asha, for the mass spec analysis, both fundamental for the knowledge and accuracy of protein concentrations and identity.

Although far away from my country and family, Wolfson college provided me with a great home-like environment. I am grateful to all the staff members and for the opportunity to live in college for all these years. The college was also home for Ali. Always willing to help and listen, her friendship and support were very important during my PhD journey. I am particularly grateful for all the time she dedicated taking care of Sophia, granting me precious working/writing hours.

I would also like to thank all my friends that made this journey more cheerful and who always made sure I felt loved and supported. Big thanks to Juan, for always being present, even when our physical distance and time difference were enormous. Thanks to Rafa, for all the countless WhatsApp messages exchanged last year. Being able to share my challenges, concerns and worries were priceless and you were fundamental on my supportive network. I am very grateful to the Brazilian friends I made in the UK, for being my big family away from home: Vanessa & Ildo (best godparents ever!) and all the other *Só a nata* members.

My sincerest gratitude to all my family. Thanks for all the encouragement and for understanding my absence. A special thanks to Tia Lucia, who cheered and supported me throughout my entire life. Thanks for always being there for me. Thanks to Soninha, Zé Paulo and to my brothers- and sisters-in-law. Life is much better having you all with me.

I would like to thank Philipe for sharing this adventure with me. The years have passed, the challenges have changed, and our bond keeps getting stronger. Finally, I would like to thank Sophia, *minha Coquinho*, for being the best writing partner I could have ever asked for. You are the last and finest chapter of this PhD journey.

Aos meus avós, Cleuza e Ulisses

Ao meu pai, Fabinho

Aos meus amores, Philipe e Sophia

Abbreviations

AMP	Ampicillin
BCL-2	B-Cell Lymphoma 2
D	Denatured state
ddH_2O	Ultrapure deionised water
[den]	Denaturant concentration
$[den]_{50\%}$	Denaturant concentration at which 50% of proteins are denatured
EDTA	Ethylenediaminetetraacetic acid
E. coli	*Escherichia coli*
FPLC	Fast protein liquid chromatography
Gly-to-Ala	Glycine to alanine mutation
IDP	Intrinsically disordered protein
IPTG	Isoprophyl β-D-thiogalactoside
IS	Ionic strength
KAN	Kanamycin
K_d	Equilibrium dissociation constant
k_{obs}	Observed association rate constant
k_{on}	Association rate constant
k_{off}	Dissociation rate constant
LB	Lysogeny broth
LB-KAN	LB- kanamycin media
LB-AMP	LB- ampicillin media
MCL-1	Induced Myleloid Leukemia Cell Differentiation Protein

$m_{D\text{-}N}$	Equilibrium *m*-value
N	Native state
NMR	Nuclear magnetic resonance
PAGE	Poly-acrylamide gel electrophoresis
PCR	Polymerase chain reaction
PUMA	p53 Upregulated Modulator of Apoptosis
PBS	Phosphate buffered saline
Pro-to-Ala	Proline to alanine mutation
PTC	Peptidyl transferase centre
RMSD	Root mean square deviation
S. aureus	*Staphylococcus aureus*
SASA	Solvent accessible surface area
SasG	*S. aureus* surface protein G
SDM	Site direct mutagenesis
SDS	Sodium dodecyl sulphate
Tris	Tris[hydroxymethyl]aminomethane hydrocloride
TS	Transition state
Tyr-to-Trp	Tyrosin to tryptophan mutation
WT	Wild-type protein

Amino acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartate
E	Glu	Glutamate
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Try	Tyrosine

Contents

List of figures

List of tables

1 Introduction

1.1 From DNA to protein

The discovery of the DNA double helix structure in 1953 (Watson & Crick, 1953) is considered a milestone in the history of science as it opened new perspectives in the way biology is perceived. The subsequent interest in how genes control the cellular chemical processes gained vast attention in the scientific community and gave rise to very powerful techniques (*e.g.* recombinant DNA tools, gene sequencing and engineering). Moreover, major advances such as mapping the human genome (Venter *et al.*, 2001; Lander *et al.*, 2001) and the promise of gene therapy, all have their origins in the unveiling of the DNA structure.

The two DNA strands are known as polynucleotides, since the monomeric units are called nucleotides. DNA nucleotides are comprised of a phosphate, a pentose sugar (deoxyribose) and four nitrogenous bases (adenosine, A; cytosine, C; guanine, G and thymine, T; Figure 1.1).

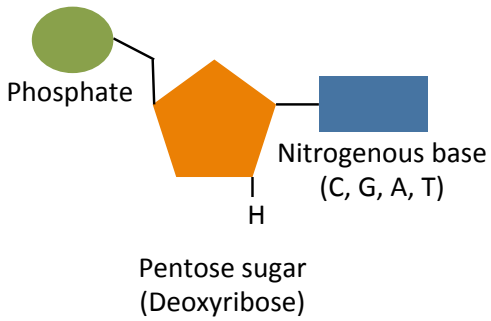


Figure 1.1:Schematic of DNA nucleotides.

Each group of three nucleotides, named codons, contain the coding information for an amino acid (*e.g.* GGG = glycine). The great potential of DNA is to encode genes from codons. Genes are a sequence of codons that encode multiple amino acids. These covalently linked amino acid polymers, classically known as proteins, are key players in cellular function.

1.2 Protein building blocks

Considering that in a codon the 4 different DNA nucleotide bases should be arranged in a group of 3 to encode an amino acid, one could expect that 64 different amino acids could be produced (4^3). However, only 20 naturally occurring amino acids exist, with different codons being able to encode the same amino acid. This biological strategy reduces the chances of a single nucleotide mutation resulting in a different amino acid and potentially altering the protein sequence, structure and function. Amino acids are classically divided into four different groups, based on their chemical properties: charged, aromatic, non-polar and polar (Figure 1.2).

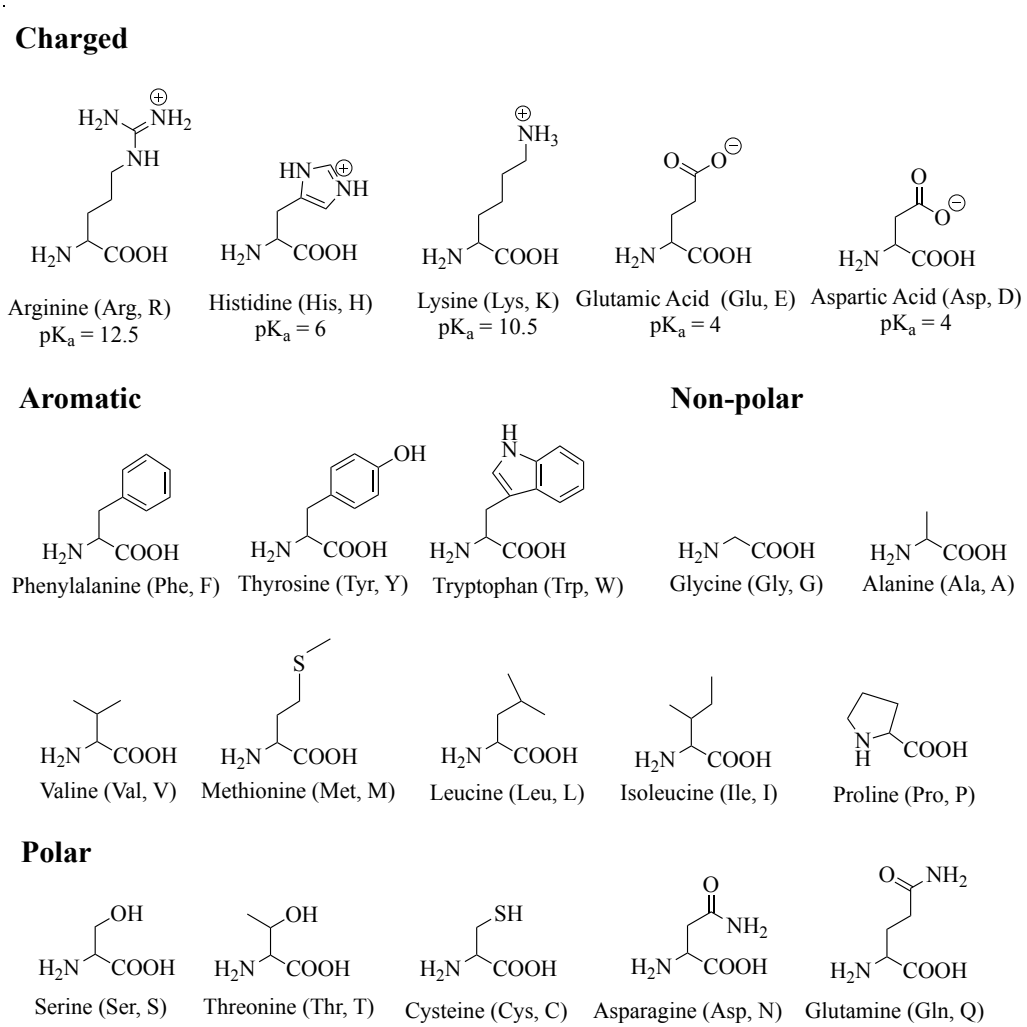


Figure 1.2: Naturally occurring amino acids represented by their structures, full names, three letters and single letter codes. pKa values at pH 7 are reported for charged residues. Structures were constructed using ChemDraw 18.2.

Amino acids are the protein building blocks, as they form a linear polymer, known as a polypeptide chain, where the monomeric units are linked through the peptide bonds (ω). The peptide bond is planar and can be in two conformations: trans (ω =180˚) or cis (ω =0˚) (Figure 1.3). Due to steric clashes of the polypeptide side chain, the vast majority of peptide bonds are found in trans conformation (Stewart *et. al*., 1990). Polypeptide backbones provide two extra rotational bonds, known as psi (ψ) and phi (ϕ) torsion angles (Figure 1.3).

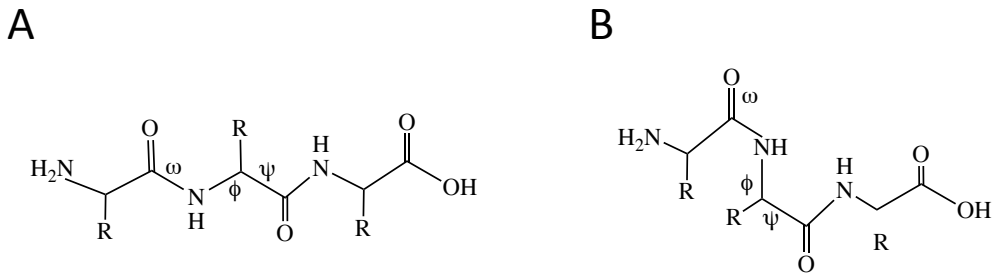


Figure 1.3: Peptide bond conformations. Tripeptide structures showing the first peptide bond (ω) in (A) trans and (B) cis conformations. Psi (ψ) and (ϕ) angles are also represented. R indicates the amino acid side chains. Structures were constructed using ChemDraw 18.2.

1.3 From a polypeptide chain to a folded protein

In vivo, proteins are normally found in solution. Interactions between the amino acids and the solvent can influence the psi and phi angles of a polypeptide chain. However, intra-molecular interactions can also occur and allow the extended chain of residues to fold into a more compact, well-packed three-dimensional structure (Figure 1.4). When the folded state is lower in energy compared to the denatured state, folding becomes a spontaneous event. Despite the specificity of the inter-residue interactions, folded proteins are only marginally stable, with the Gibbs free energy for folding ($\Delta G_{folding}$) of about -2 to -10 kcal.mol^{-1} (Savage et al. 1993; Taverna and Goldstein 2002). A single hydrogen bond, for instance, is estimated to contribute with 1-5 kcal.mol^{-1} (Fersht et al. 1985), suggesting that all interactions can significantly affect protein stability.

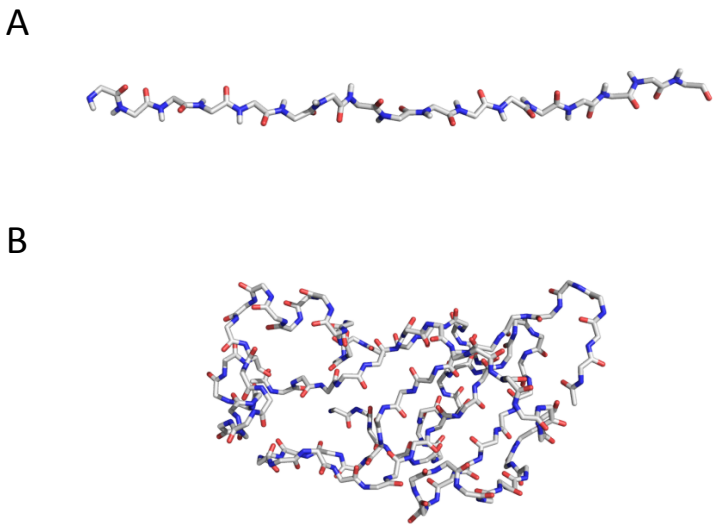


Figure 1.4: A polypeptide chain can fold to form a more compact 3-dimensional structure. (A) an extended, 1-dimentional structure of a polypeptide chain. (B) interactions of residues in the polypeptide chain can drive the folding of the molecule, forming a more compact, structured conformation. For simplicity, only the backbone structure is represented, with carbons (gray), nitrogen (blue) and oxygen (red) atoms illustrated. Structures were constructed using PyMOL (version 1.5.0.4, Schrödinger) and the PDB code 1A2P.

1.3.1 Hydrogen bonds

A hydrogen bond is essentially an electrostatic attraction between a hydrogen covalently bound to a highly electronegative atom – the donor (*e.g.* oxygen, nitrogen, fluorine) and another electronegative atom containing an unshared pair of electrons – the acceptor. In a polypeptide chain, hydrogen bonds can be formed between the amide (donor) and the carbonyl groups (acceptor) present in the amino acids backbones. Although the formation of intra-molecular hydrogen bonds seems to be sufficient and very favourable, the solvent also needs to be considered (Dill 1990; Hunter 2004). Water can act as a hydrogen bond donor or acceptor. Since hydrogen bonds can be formed between residue and solvent, intramolecular hydrogen bonds do not confer a significant contribution to protein stability. Nevertheless, specific structures can be constructed with the hydrogen bonds between the amino acids, which makes this type of interaction very important to the ability of proteins to adopt a folded conformation.

1.3.2 Van der Waals

Van der Waals is a distance-dependant interaction based on attraction and repulsion forces. It differs from covalent and ionic interactions, where attractions result in the formation of a chemical electronic bond. In comparison, van der Waals force is weaker and is strongly influenced by the distance between the interacting molecules (Roth *et al*, 1996). Atoms of any kind have fluctuating electron distribution relative to their nuclei, resulting in transient dipoles. When two atoms are close in proximity, the dipole of one can induce a dipole on the other, leading to an attraction force. However, the decrease of the distance between the two atoms can cause the two electron orbits to overlay, causing a repulsion effect. The optimal distance between the two atoms required for an effective van der Waals interaction is dependent on the nature of the atom. When a protein folds into its three-dimensional conformation, the atoms are packed in a distance that favours dipole interaction, van der Waals force play significant role in protein structure formation and stability (Pace, 2014).

1.3.3 Effect of electrostatics

Considering the 20 naturally occurring amino acids, at different pH environment, a subset of them can be ionised, resulting in either loss or acceptance of a proton (Figure 1.2). Electrostatic interactions can be formed by the attraction and repulsion of charges. Although hydrogen bonds and van der Waals force are also considered to be electrostatic in nature, charge-charge interactions can be effective at much longer distances (e.g. 5-10 Å) (Zhou and Pang 2018). Therefore, opposite charged interactions can be expected to substantially drive protein folding. However, the aqueous nature of the solvent needs to be considered: dipoles present in water molecules allow not only the hydrogen bond formation but also charge-charge interactions to occur between the water and the charged residues. Consequently, when polypeptide chains fold, most polar side chains seek the exterior, where they can be solvated by water molecules in the solvent (Perutz 1978). Nonetheless, charged interactions can contribute to the formation of the correct structure upon folding. An unpaired charged residue is likely to be located in the solvent accessible surface of the protein, where it can interact with water. For that reason, electrostatics interactions can also play a role in promoting protein solubility.

1.3.4 Hydrophobic effect

The hydrophobic effect is the propensity of non-polar substances to attract each other in an aqueous environment, in order to exclude the water molecules (Chandler 2005). Considering the protein building blocks, non-polar amino acids would repel water molecules, forming a hydrophobic core. Note that the attractive forces between non-polar residues are not driven by the amino acids *per se*, but originate from the water molecules. There are two hypothesis that try to explain this effect: 1) Water molecules have a decrease in entropy in the presence of hydrophobic molecules. The water molecules undergo a special rearrangement around the hydrophobic residues in order to maximise their ability to form hydrogen bonds (Ball 2011; Camilloni et al. 2016). 2) Hydrophobic residues create an enthalpic penalty by disrupting the water hydrogen bond network (Lum, 1999). Either way, a solvent exposed hydrophobic group in a protein structure would imply an energetic disadvantage. Therefore, the hydrophobic packing of a polypeptide chain is associated with a substantial gain in free energy. For that reason, the hydrophobic effect is considered the driving force of protein folding (Dill, 1990; Pace 2014).

1.3.5 Disulphide bonds

Typically called S-S bonds, disulphide bonds are covalent interactions between two thiol groups that are located in close proximity in space. The thiol group present in the cysteine residue can form S-S bonds when two of these residues are in an oxidising environment and close to each other. Hence, it is more likely to form these interactions in a folded state, where the residues are in a much closer proximity, compared to the unfolded state. Although, the formation of a disulphide bond after the protein is folded can provide an enthalpic gain and contribute to the stability of the folded state, this is not always true (Creighton 1988). Formation of disulphide bonds can increase the strain of the native state structure (Clarke, 1995a; Clarke, 1995b) as well as stabilising the unfolded state by the presence of the intramolecular interactions (Clarke et al. 2000). Although some of the proteins studied in this thesis contain cysteine residues, none of them have native disulphide bonds present in their folded structure.

1.4 Protein structure

Proteins can adopt different conformational structures, with the simplest being the uni-dimensional covalent linkage of the amino acids forming a polypeptide chain, known as primary structure. In order to adopt more complex structures, amino acids psi and phi angles can experience torsions that allow a structure rearrangement. Although the possibilities might be vast, many of the psi (ψ) and phi (ϕ) angle combinations, and the associated structure conformations, are not favourable due to steric clashes (Ramachandran, 1963). The Ramachandran plot (Figure 1.5) describes the torsional psi and phi angles of the amino acids in a peptide.

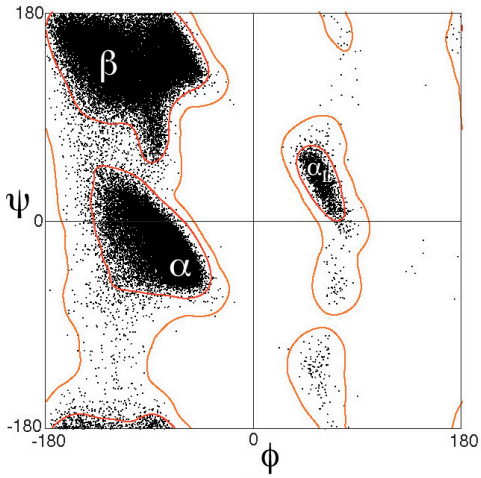


Figure 1.5: General Ramachandran plot. psi (ψ) and phi (ϕ) angle combinations of a typical amino acid residue is restricted to the outlined regions. Amino acids (black points) usually fall within favored regions (interior lines). The regions favor distinct secondary structures: α-helices (α), left-handed α-helices (α_L) and β-sheets (β). Figure was adapted from Lovell et al., 2003 and shows the favourable psi and phi angle combinations for all amino acids, except glycine and proline.

As previously discussed, all amino acid residues with the exception of prolines, are capable of forming hydrogen bonds with each other through both amide and carbonyl groups. Considering the unfavourable psi and phi angle torsions, intramolecular hydrogen bonds can occur in two manners: 1) the carbonyl group of one residue (i) can interact with the amide group of another residue located 3 to 4 residues ahead (i+4) (Figure 1.6, A). These interactions originate a helical structure, most commonly found with turns every 3.6 residues, known as α-helices. 2) Interactions can also occur between the carbonyl and amide groups of amino acids that are much further apart (*e.g.* when a turn is introduced in a polypeptide chain). These can originate much more extended structures, known as β-sheets

(Figure 1.6, B). The torsion angles in α-helices and β-sheets confers a much more ordered assembly in comparison with the flexible primary structure of a polypeptide chain, and together they represent the secondary structures of proteins.

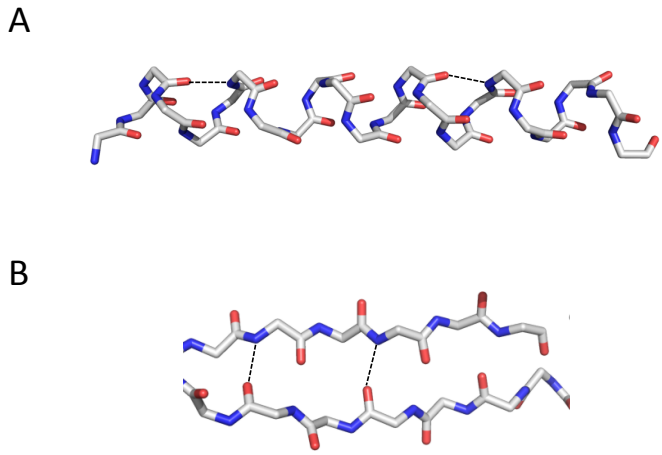


Figure 1.6: Intramolecular hydrogen bonding generates protein secondary structures. (A) In an α-helix, the carbonyl of one residue can hydrogen bond with the amide group of another residue when they are 4 residues apart. (B) In a β-sheet, residues located at much further distances can come closer when a turn is made in the primary structure allowing the residues to hydrogen bond. Some hydrogen bonds are represented by dashed lines. For simplicity, only the backbone structure is represented, with carbons (gray), nitrogen (blue) and oxygen (red) atoms illustrated. Structures were constructed using PyMOL (version 1.5.0.4, Schrödinger).

Elements of secondary structure can then interact and build the most complex level of unimolecular protein structure: the tertiary (Figure 1.7). These interactions are driven by the amino acid side chains and the forces previously discussed in section 1.3.

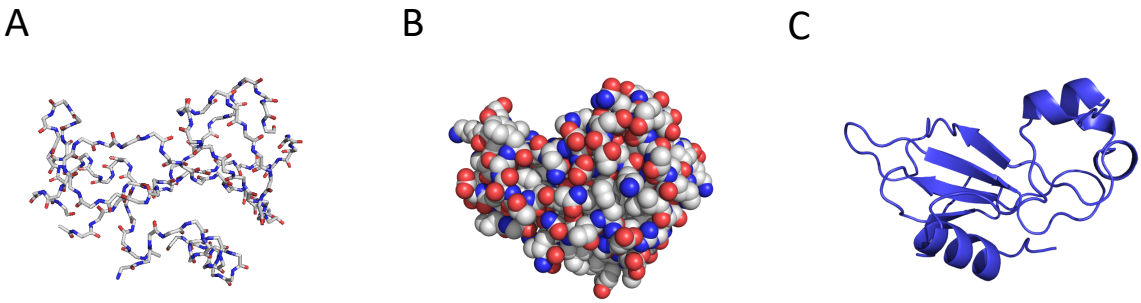


Figure 1.7: Different representations of a protein tertiary structure. (A) Barnase represented by the secondary structure elements that come together to form the protein structure. (B) The tight packing of the protein residues can be easily visualised when we represent all atoms. (C) Cartoon representation only shows the secondary structure elements. The three structures were constructed using PyMOL (version 1.5.0.4, Schrödinger) and the PDB code: 1A2P.

1.5 Why study protein folding

Proteins differ from one another in their sequence of amino acids and most fold to form highly specific, three-dimensional structures. *In vivo* proteins are not isolated, and many other factors will influence folding, such as metal ions, salts, ligands, temperature, post-translational modifications and chaperones.

The protein folding field started to gain importance and began to develop in the late 50's, after the first protein structure was determined (Kendrew *et al.*, 1958). Since then, there has been a continuous and increasing interest in understanding how protein structures are formed and the mechanisms by which they fold.

Further investigation and experiments showed that all the information required for a protein to fold into its native conformation is carried in the amino acid sequence (Anfinsen *et al.*, 1961). However, the mechanisms for folding and the relationship of folding mechanisms from primary sequence to tertiary structure are still not completely understood.

The protein folding timescale can vary from hours to milliseconds, but what drives this difference in time is not well defined. In 1968, Levinthal demonstrated, using simple calculations, that because of the large number of degrees of freedom in the psi/phi angles in a polypeptide chain, a random search of all possible pathways for folding would be on a geological timescale. Hence, it would not be possible for a protein to fold into its native conformation in a short timescale if it had to sequentially sample all the possible conformations. He proposed the idea of the folding being driven by a rapid formation of a local nucleus that determines the following folding of the polypeptide chain (Levinthal 1968, 1969). This conclusion supported the idea of folding pathways, which are pre-determined routes that allow the protein to fold from its denatured to its native state in a biologically relevant timescale.

There are a number of principles that drive the protein folding field. The reason why a protein folds is due to the native state being lower in energy than its denatured state. However, proteins are very easily destabilized, and even small changes in amino acid sequence can be enough to make the native state higher in energy compared to the denatured, thus preventing the protein from folding (Fersht, 1999).

Protein folding studies are still a significant challenge in science. Tackling the most fundamental questions on how a protein folds, is stabilized, interacts with ligands and functions, have a profound impact on how we may perceive metabolic regulation, cellular function and possibly better understand underlying processes in disease states. Ultimately, such studies could open possibilities for new treatments in order to ameliorate a patient's quality of life (Hammarström et al. 2003).

1.5.1 Models of protein folding

Over the past decades, different models have been described in order to explain how proteins fold. The first models, also known as the "classical view" of protein folding, were based on the idea of folding pathways. The *framework* model (Ptitsyn 1973), for instance, suggests that the folding starts with the formation of secondary structure elements that then interact to form subsequent intermediates. The final structure is accomplished with the packing of the side chains. Another model, typically referred to as the *nucleation condensation*, states that the folding limiting step is the formation of a nucleus, followed by the rapid propagation of the protein structure (Fersht 1997). The nucleus is an expanded version of the protein native structure, which is further stabilised by long range interactions within the protein sequence. Thus, the formation of the nucleus is coupled with the condensation of the whole structure (Itzhaki, 1995).

Subsequent studies have proposed what is commonly referred to as the "new view" of protein folding model: the energy landscape theory applied to protein folding (Bryngelson *et al.*, 1995; Onuchic *et al*., 1997). The landscape describes the dependence of free energy on all protein conformations. High and low energy conformations are represented by a hill and a valley, respectively. Hence, the energy surface is typically referred to as the energy funnel. The energy landscape theory suggests that a protein does not follow a single folding pathway but different routes down the folding funnel are possible. The downhill nature of folding allows it to proceed one residue at a time, randomly. Populated intermediates are interpreted as regions of kinetic traps in the landscape.

A lot has been discussed in the protein folding community regarding folding mechanisms, with the majority of the work based on globular proteins. In this thesis we want to investigate

non-traditional protein structures: SasG, a bacterial protein with an usual sequence composition and structure, as well as, a subset of proteins that are disordered in isolation and only fold upon binding to a partner protein.

1.6 Staphylococcus aureus surface protein G (SasG)

1.6.1 Biological background

Bacterial cells enclosed in a polymeric matrix can adhere to a living or inert surface to form functional communities known as bacterial biofilms (Costerton, Stewart, and Greenberg 1999). The formation of a biofilm involves an initial attachment of these communities onto a surface and a subsequent maturation phase that is physiologically specific to each organism. The maturation step consists of an intracellular aggregation and the formation of a 3-dimensional structure, characteristic of a mature biofilm. Once the maturation step is complete, a final detachment phase disperses the bacterial cells to new infection sites in the human body (Otto 2008). Biofilms are very difficult to eradicate as they are resistant to the host immune system and to antimicrobial agents (Dunne 2002; Lewis 2001; Steward and Costdrton 2001).

The most frequent bacteria found on human skin and mucous surfaces are Staphylococci, which are a common cause of infections caused by biofilms. *S. aureus* and *S. epidermis* can form biofilms on implanted medical devices causing infections that often require device removal (Harris and Richards 2006). *Staphylococcus aureus* surface protein G (SasG) promotes cell-to-cell accumulation during biofilm formation (Geoghegan *et al.*, 2010). SasG is a protein from the MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) group that contains an LPXTG motif essential for covalent linkage to the cell peptidoglycan (Corrigan et al. 2007). SasG has an A domain at its N-terminus, which is responsible for promoting adhesion to epithelial cells (Roche, Meehan, and Foster 2003) followed by a B region, that promotes accumulation during biofilm formation (Geoghegan *et al.*, 2010).

The SasG B region is composed of nine repeats of two domains (Figure 1.4, A): the G5 domain which is approximately 80 residues (PFAM accession number PF07501; Bateman *et al.*, 2004; Bateman, Holden and Yeats, 2005) and the E domain which is approximately 50 residues (D. T. Gruszka et al. 2012). The G5 domain was named after its five conserved glycine residues, also present in the E domain. The sequence identity between G5 domains (except for the first and last G5s) and between E domains is >97% (D. T. Gruszka et al.

2012). On the PONDR-FIT (Xue et al. 2010) and other disorder predictor analysis, G5 and E domains are expected to be disordered as they have an intrinsically disordered protein (IDP) sequence composition (Figure 1.8).

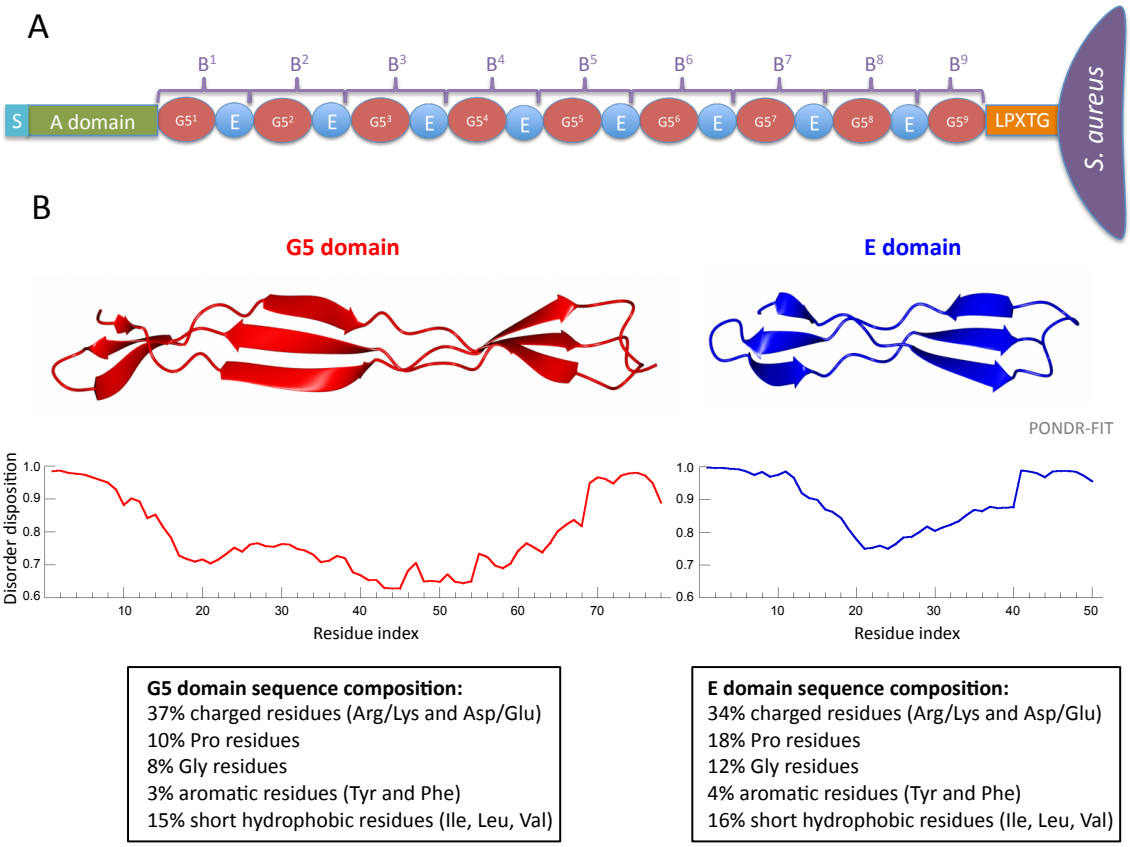


Figure 1.8: SasG G5 and E domains disorder prediction. (A) Schematic representation of the domains arrangement within SasG. (B) G5 and E domains PONDR-FIT analysis show that both domains are predicted to be disordered. The sequence composition of G5 and E have also been shown.

1.6.2 Biophysical background

During my MPhil, I investigated the role of the glycine residues in the SasG protein. The project involved stability and kinetics studies, which provided evidence of the importance of the interface between the E and the $G5^2$ domains and solid information for the characterization of the folding pathway of $G5^2$ and E-$G5^2$ constructs. All the following data and analysis were previously presented and discussed in my MPhil thesis.

The role of the E-G5^2 interface in stability

Equilibrium studies of the G5^2 domain in isolation showed that mutations of the most conserved glycines G584A and G587A, located in the triple helical region, were so significantly destabilizing to the domain structure, that the G5^2 domain was mostly unfolded in 0 M urea (G587A~70% unfolded and G584A~100% unfolded). In the E-G5^2 construct, the same mutations are also significantly destabilizing but both E and G5^2 are folded in that context (Figure 1.9).

The data show that the addition of the intrinsically disordered E domain to an unfolded G5^2 rescues the stability of the system. This result shows the significant free energy contribution that comes from the inter-domain interface. As has been shown previously, the E-G5^2 interface is more stable than the domains themselves (estimated to be > 6 kcal.mol^{-1}; (Gruszka *et al.*, 2015).

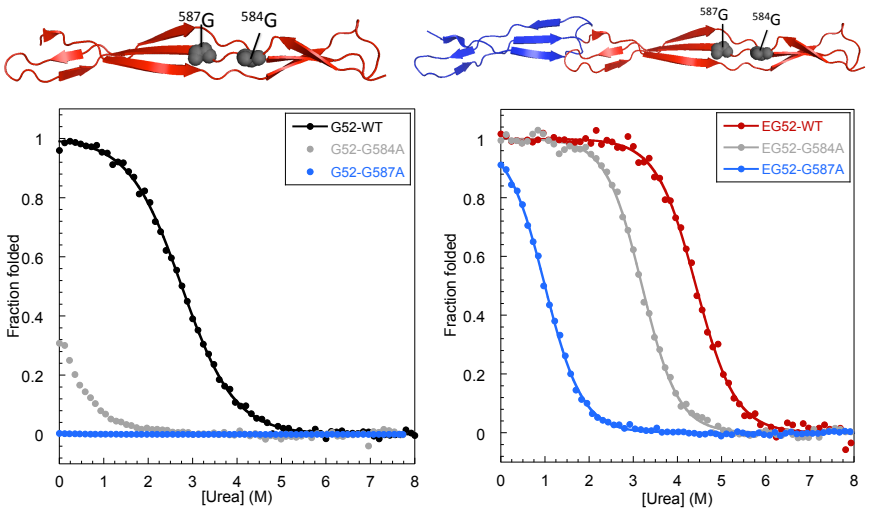


Figure 1.9: Equilibrium curves comparison of the mutants G584A and G587A in G5^2 in isolation and E-G5^2 construct. (A) G5^2 domain crystal structure with mutants highlighted and equilibrium fraction folded data for G5^2-WT (black), and its mutants G584A (grey) and G587A (blue). (B) E- G5^2 domain crystal structure with mutants highlighted and equilibrium fraction folded data for E-G5^2-WT (red), and its mutants G584A (grey) and G587A (blue).

Folding mechanism of SasG

In the context of the G5^2 in isolation, all the mutations in the C-terminal region of the domain show a Φ-value of about 1, whereas mutations in its N-terminal region result in a Φ-value of approximately 0 (Figure 1.10). The data clearly suggest that in the transition state (TS)

the G5^2 domain has structure in its C-terminal region that is almost completely native-like. This is where the domain starts to fold.

In the E-G5^2 construct, mutations in the N-terminal region of G5^2 and all mutants located in the E domain show a Φ-value of 0 suggesting that this whole region of the protein is unstructured in the TS. Dr. Gruszka's Φ-value analysis for Pro-to-Ala and Tyr-to-Trp mutations are in agreement with all the glycine to alanine data. Pro-to-Ala mutations located in the E domain and in the N-terminal region of G5^2 domain also have a Φ-value of 0 whereas in the C-terminal end of the G5^2, Φ-values of 1 are also observed (Figure 1.10). The Φ-value analysis show that the folding of G5^2 domain starts via its C-terminal end, which is also the rate-limiting step for the folding of the entire E-G5^2 construct.

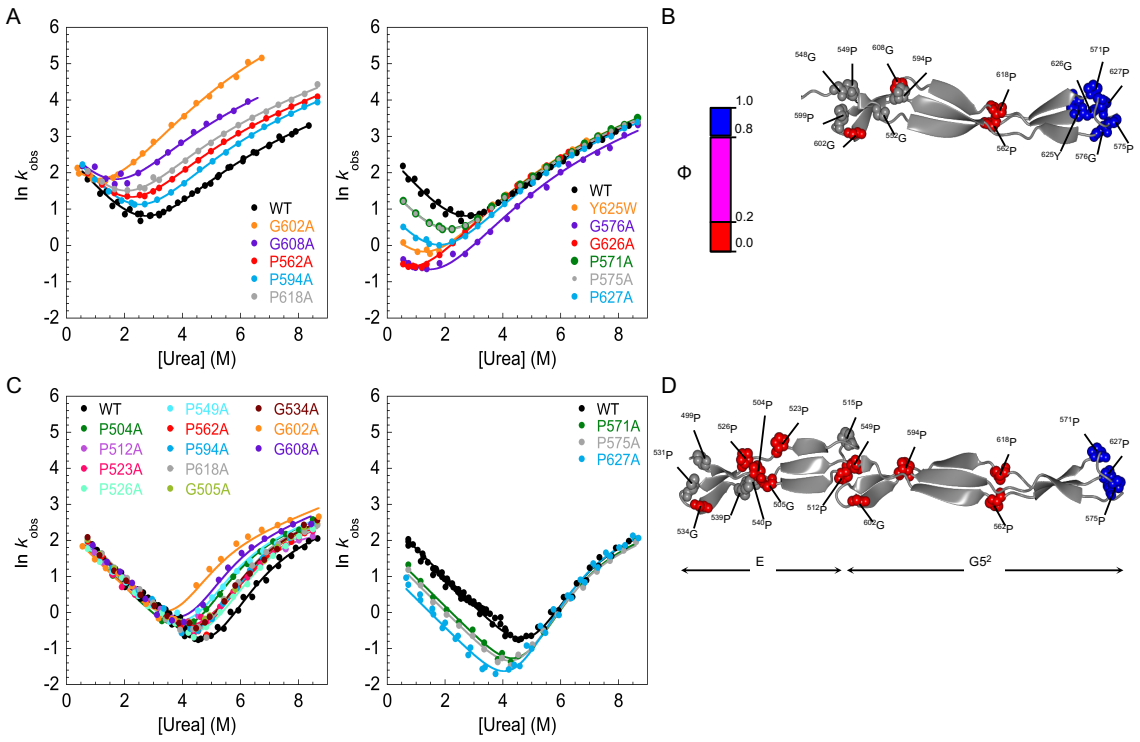


Figure 1.10: Mapping the structure of the WT folding pathway for G5^2 and E-G5^2 constructs (Proline data and global fitting provided by Dr. Dominika Gruszka). (A) Chevron plots for G5^2-WT (black) and its mutants (colour-coded). Left panel shows mutants that unfold faster than G5^2-WT and fold at approximately the same rate (Φ=0). Right panel shows mutants that fold slower than G5^2-WT but unfold at the same rate (Φ=1). (B) Structure of the G5^2 TS mapped into the domain crystal structure. (C) Chevron plots for E-G5^2-WT (black) and its mutants (colour-coded). Left panel shows mutants that unfold faster than E-G5^2-WT and fold at approximately the same rate (Φ=0). Right panel shows mutants that fold slower than E-G5^2-WT but unfold at the same rate (Φ=1). (D) Structure of the E-G5^2 TS mapped into the construct crystal structure. (Figure taken from Gruszka et al., 2016).

Alternative folding pathway of E-G5^2

In the context of the E-G5^2, Gly-to-Ala mutations located in the C-terminal region (G576A and G626A) and in the triple helical region (G584A and G587A) of the G5^2 domain, show a change in the dependence of the rate constant for folding on denaturant concentration (m_{kf}). The data suggest an alternative pathway for folding once the C-terminal region of G5^2 is destabilized. Dr. Gruszka observed the same change in the folding *m*-value for the mutant E-G5^2-Y625W (also in the C-terminal region of the G5^2).

As the change in the folding *m*-value is only detected in the E-G5^2 construct and it is observed for the proteins with the triple helical region or C-terminal end destabilized, we hypothesised that the alternative pathway for folding could start via the inter-domain interface.

A pseudo-WT analysis, in the context of E-G5^2-Y626, two mutations (P540A and P599A) revealed results that contributed to the characterization of the alternative pathway for folding.

1) The E domain is partially folded in the transition state of the alternative pathway: The mutant P540A resulted in an intermediate phi ($0.3< \Phi <0.7$) in the context of the E-G5^2-Y625W, implying that the triple helical region of the E domain is partially structured in the TS (Figure 1.11).

2) If the interface between the two domains is destabilized, then the E-G5^2 reverts to the original folding pathway: the chevron plot of the mutant E-G5^2-Y625W-P599A has the same m_{kf} as E-G5^2-P599A and E-G5^2-WT, indicating the wild-type-like folding pathway. This supports the hypothesis that in the alternative folding pathway the interface between the two domains is folded (Figure 1.11).

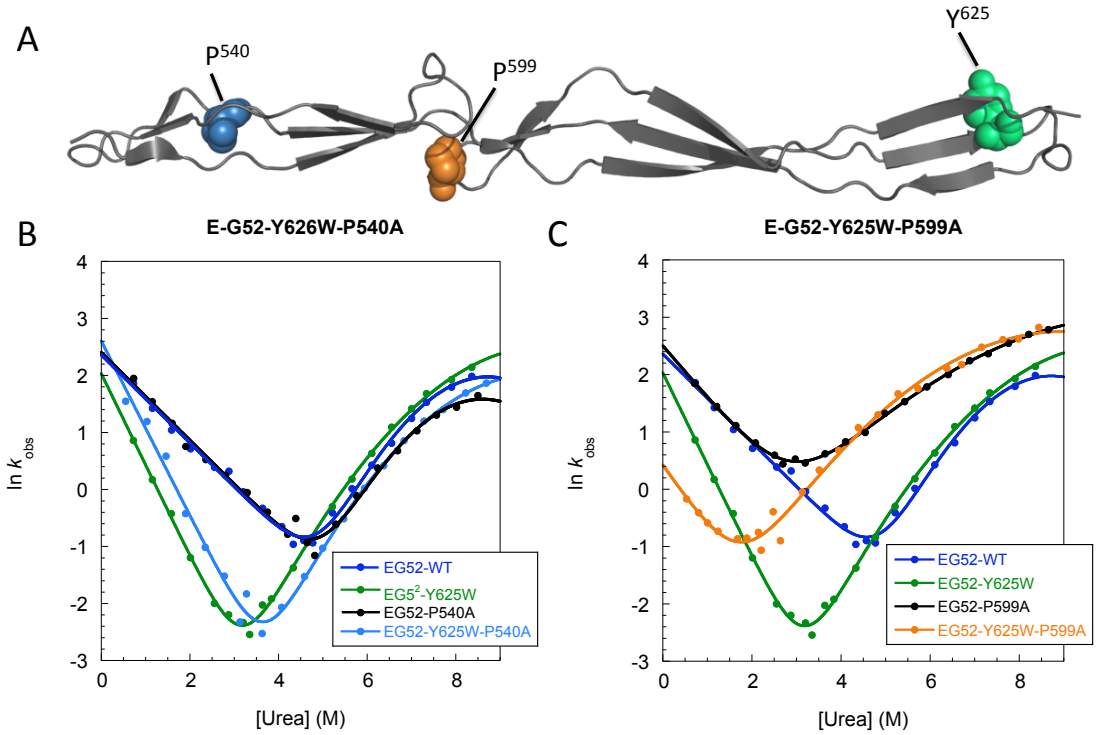


Figure 1.11: Kinetic data of the Pro-to-Ala mutants in the context of E-G5²-Y625W pseudo-WT. (A) Proline and tyrosine residues in E-G5² structure. (B) Chevron plots for E-G5²-WT (blue), E-G5²-Y625W (green), E-G5²-P540A (black) and E-G5²-Y625W-P540A (cyan). (C) Chevron plots for E-G5²-WT (blue), E-G5²-Y625W (green) E-G5²-P599A (black) and E-G5²-Y625W-P599A (orange).

1.7 Intrinsically disordered proteins

Since the first protein structure was determined 6 decades ago (Kendrew *et al.*, 1958,1960), proteins were typically associated with the classical structure-function paradigm, where the protein biological function was defined by its specific three-dimensional structure. More recently, the discovery of proteins that lack a pre-defined structure has intriguided the scientific community. The carriers of such "structure-free" conformations were named intrinsically disordered proteins (IDPs) or protein hybrids comprised by ordered domains and disordered regions (IDRs). Following the idea that the information required for a protein to fold is encoded in its amino acid sequence (Anfinsen 1973), the lack of strucutre must also be encoded in the IDP sequences. In comparison to folded proteins, disordered proteins typically comprise a higher number of prolines, glycines and charged residues (Romero et al. 2001; Theillet et al. 2013; Uversky, Gillespie, and Fink 2000), whereas hydrophobic residues are underrepresented. Together, IDPs and IDRs can often be found in nature, with noticeble amounts of both being present in all life kindoms and viral proteomes (Dunker *et al.*, 2000; Xue, 2012; Peng *et al.*, 2014).

The growing interest of this intriguing class of proteins starts precisely by the challenge that they represent in the structure-function paradigm. Despite being very flexible and lacking a well defined strucutre, IDPs are biologically active and predominantly involved in transcription and signaling regulation (Ward, 2004). It has been suggested that the lack of a globular structure might represent an advantage for IDPs, as the strucuture flexibility might play a role in allowing them to interact with different targets in the cell (Dunker et al. 2001; Wright and Dyson 1999). Furthermore, the disorder-to-order transition of IDPs during binding processes in the cell can be a fine mechanism for cellular regulation.

Although evolution can provide positive changes leading to better adaptation, mutations in proteins can be unfavourable. In folded proteins, for instance, mutations can largely affect protein stability (Tokuriki et al. 2007), which explains the evolutionary pressure to maintain the position and nature of the residues that are crutial to the protein thermodynamic stability (Mirny and Shakhnovich 1999). Due to the lack of structure, IDPs are more susceptible to mutations and evolutionary changes (Chen *et al.*, 2006). Nevertheless, the evolution persistance of disorder confirms the relevance of such class of proteins and indicates that the functional benefits of disorder surpass the ability to tolerate mutations. The lack of

structure also provides more acessibility to amino acid side chains, making them more prone to post translation modifications (PTMs) (Bah and Forman-Kay 2016). PTMs can significantly affect the energy landscapes of IDPs causing structural changes and directly impacting on protein function. Phosphorylation and acetylation modifications, for instance, allow changes in the IDP conformation possibilities as well as can allow them to act as regulatory switches (Kouzarides 2007; Nash et al. 2001). Another feature of IDPs is their ability to bind weakly and promiscuously to different targets (Chen 2012; Zhou 2012). The kinase inhibitory domain of the cyclin-dependant kinase (CDK), for instance, can bind to a diverse family of cyclin-CDK complexes (Kriwacki et al. 1996). Likewise, the GTPase-domain of the Wiskott–Aldrich syndrome protein (WASP) can bind to its own CTA-domain, resulting in its inhibition, whereas when in a different conformation, it can bind to the GTP-ase Cdc42, resulting in WASP activation (Kim et al. 2000).

Due to their unusual strucuture and biological relevance in cellular processes, the presence of IDPs in the cell requires a very specific and tight regulatory mechanism (Babu et al. 2012). Studies have shown that human transcripts that encode IDPs have higher mRNA decay rates as well as an increased proportion of miRNA, suggesting lower gene expression levels of these proteins (Chen, 2008; Edwards *et al.*, 2009). Moreover, an increased proportion of ubiquitination sites are found in IDPs compared with folded proteins (Edwards et al. 2009; Radivojac et al. 2010), targeting those molecules to fast degradation (Tompa et al. 2008). Together, these are suggestions that most IDPs are predicted to have shorter half-lifes comapared to ordered proteins (Van Der Lee et al. 2014). In some cases, once this cellular regulation is somehow compromised, increased expression levels and persistance of IDPs in the cell, can lead to multiple diseases, including cancer and different types of neurodegenerative disorders (Babu 2016; Uversky, Oldfield, and Dunker 2008). Together, all these intriguing features and functional relevance makes this class of proteins very interesting, which can explain the increasing number of studies involving IDPs since their discovery.

1.7.1 Coupled folding and binding of IDPs

In the cell, transcription and signalling regulation are typically achieved by protein interactions. Although lacking a pre-defined structure, a subset of IDPs can fold upon

binding to a partner macromolecule, known as coupled folding and binding reaction (Dyson and Wright 2002). In these cases, the folded partner provides the necessary free energy for the folding of the IDP. IDPs coupled folding and binding reactions are typically involved in signalling systems (Wright and Dyson 2015). A question regarding the mechanism can emerge: what comes first, folding or binding? The route by which the folded complex is formed may indeed affect function. Two extreme mechanisms are proposed: in the *induced fit* mechanism, all members of the conformational ensemble of the IDPs are binding-competent, resulting in folding after binding to the partner macromolecule. In the *conformational selection* mechanism, only correctly folded members of the ensemble are binding-competent, resulting in folding before binding (Figure 1.12).

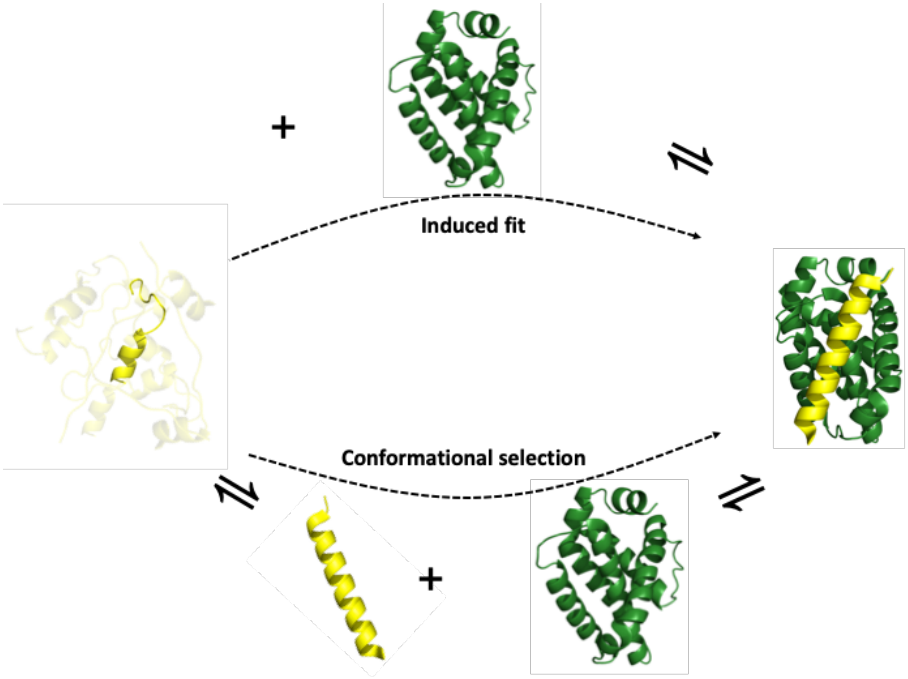


Figure 1.12: Schematic of IDPs coupled folding and binding mechanisms. A subset of IDPs can fold upon binding to a folded partner. Two extreme mechanisms have been proposed. In the induced fit mechanism (top), any member of the structural ensemble of IDPs are binding-competent and can fold subsequently to binding. In the conformational selection (bottom), only the correctly folded structures are able to bind, hence folding precedes binding. Structures were constructed using PyMOL (version 1.5.0.4, Schrödinger) and the PDB code: 2VOI and 1DDB.

It is important to emphasise that coupled folding and binding reactions with elements of both mechanisms, are possible (Hammes *et al.* 2009; Greives and Zhou, 2014). Although thermodynamic arguments can often be made in order to explain coupled folding and binding mechanisms, only kinetic studies of these complexes can properly shed light on mechanisms (Gianni, Dogan, and Jemth 2016; Kiefhaber, Bachmann, and Jensen 2012; Shammas et al. 2016).

1.8 Aims of this thesis

Throughout the decades that protein folding has been investigated, studies are typically done with proteins in isolation. Understanding proteins single domains and how certain properties (*e.g.* thermodynamic stability, folding and unfolding mechanisms) are conserved through families of proteins, is a very powerful way to understand the fundamentals of protein science. However, *in vivo,* proteins are immersed in a crowded cellular environment, where many other factors can alter the way they fold and function: temperature, pH, chaperones, interaction with other macromolecules These are just a few examples of how context can affect the properties of proteins. Moreover, a vast proportion of the proteome is known to be intrinsically disordered, with a subset of these molecules being able to fold only upon binding to another protein. The aim of the work designed in this thesis was to shed light on how context can influence protein dynamics and protein-protein interactions.

After a previous study of the bacterial protein SasG, where the folding mechanisms were extensively characterised and described (Gruszka et al. 2016), the next intriguing question was on the charged nature of this protein. What is the role of the numerous charged residues in SasG stability and ability to fold? Then, to explore protein folding in a more complex context, co-translational folding was also targeted: how the elongated beta-sheet structure of SasG can fold upon translation on the ribosome? How that compares with the folding of the domains in isolation?

Intrigued by the disorder-order fine line drawn by SasG, the next step of the thesis targets intrinsically disordered proteins and protein-protein interactions. How coupled folding and binding reactions of IDPs compare to the folding of a single protein? What encodes the folding information: the IDP or the partner protein? Finally, as IDPs are largely studied as peptides, how the coupled folding and binding of an IDP peptide compares with the protein in the full-length context?

2 Thermodynamics and kinetics analysis

2.1 Protein folding

2.1.1 Equilibrium studies of proteins

The use of denaturants, such as urea and guanidine hydrochloride, is one of the primary ways of measuring the conformation stability of proteins and comparing this with the stability of the mutated proteins. The use of denaturants is widespread (Pace 1986) even though their mechanism of action is not yet completely understood. It is known that the interactions of denaturant with protein constituent groups are more favourable in comparison with the interaction between these groups and water (Tanford 1970) and alter the equilibrium between native and denatured states of the protein.

Different probes can be used in order to investigate to what extent a protein is folded (CD, fluorescence, etc.). Although many proteins denature irreversibly (Kauzmann 1959), many of them regain their folded conformation once the denaturant is removed. Thermal unfolding (another method of denaturation) of chymotrypsinogen (Brandts 1964) characterized the transition between folded and denatured state as being very abrupt (Figure 2.1), a phenomenon that was also observed for many other systems and led to the idea of a "two-state" process representing the equilibrium between native and denatured states. In a "two-state" process, the equilibrium reactions (D⇌N) are cooperative, occurring in a single step with no populated intermediates.

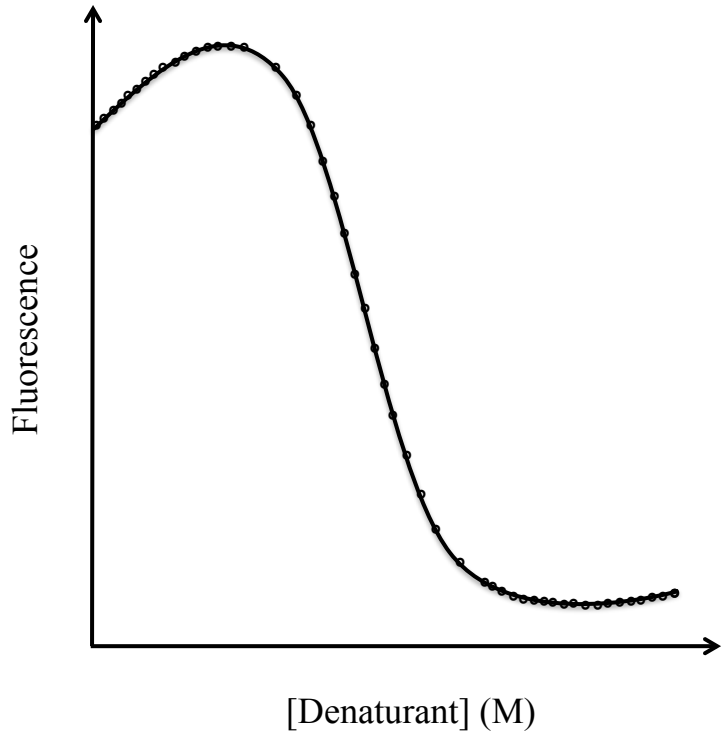


Figure 2.1: Representation of a fluorescence denaturation curve under equilibrium conditions.

The free energy of transfer of a protein from pure water to a denaturant solution shows a linear dependence (Figure 2.2) in denaturant concentration (Pace 1986) and is expressed by the equation (Greene and Pace 1974):

$$\Delta G_{\mathrm{D-N}} = \Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}} - m_{\mathrm{D-N}}\,[\mathrm{den}] \qquad \textbf{Equation 2.1}$$

Where,

$\Delta G_{\mathrm{D-N}}$ is the free energy of unfolding at a specific denaturant concentration.

$\Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}}$ is the free energy of unfolding in pure water.

$m_{\mathrm{D-N}}$ is the equilibrium *m*-value (see details bellow).

[den] is the concentration of denaturant.

The equilibrium *m*-value is the dependence (slope) of the free energy of unfolding on denaturant concentration and it is related to the change in the solvent accessible area (SASA) as the protein unfolds (Myers, Pace, and Scholtz 1995).

There is a point in the denaturant concentration ($[den]_{50\%}$) at which 50% of proteins are denatured. When [den] = $[den]_{50\%}$, both states D and N are equally populated ([D] = [N]) and the energy difference between them (ΔG_{D-N}) is zero, so:

$$\Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}} = m_{\mathrm{D-N}}\,[\mathrm{den}]_{50\%} \qquad \textbf{Equation 2.2}$$

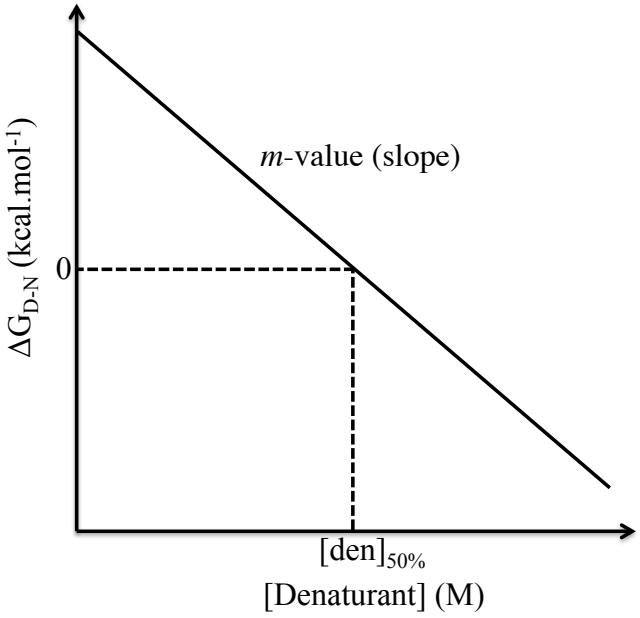


Figure 2.2: Free energy of unfolding has a linear dependence with the concentration of denaturant.

The values $\Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}}$, $m_{D\text{-}N}$ and $[den]_{50\%}$ can be extracted from an equilibrium denaturation curve using the following equation (Clarke and Fersht 1993):

$$F = \frac{[(\alpha_{\mathrm{N}} + \beta_{\mathrm{N}}\,[\mathrm{den}]) + (\alpha_{\mathrm{D}} + \beta_{\mathrm{D}}[\mathrm{den}])\exp\{m_{\mathrm{D-N}}([\mathrm{den}] - [\mathrm{den}]_{50\%}/RT\}]}{1 + \exp\{m_{\mathrm{D-N}}\,([\mathrm{den}] - [\mathrm{den}]_{50\%})/RT\}} \qquad \textbf{Equation 2.3}$$

Where,

F is the fluorescence reading at given [den].

α_N is the fluorescence of the native protein in pure water.

β_N is the gradient of the fluorescence baseline for native protein.

α_D is the fluorescence of the denatured protein in pure water.

β_D is the gradient of the fluorescence baseline for denatured protein.

T is the temperature at which the experiment is performed (K).

R is the universal gas constant (1.987 $cal.mol^{-1}.K^{-1}$).

The terms α_N, β_N, α_D and β_D are shown graphically in Figure 2.3:

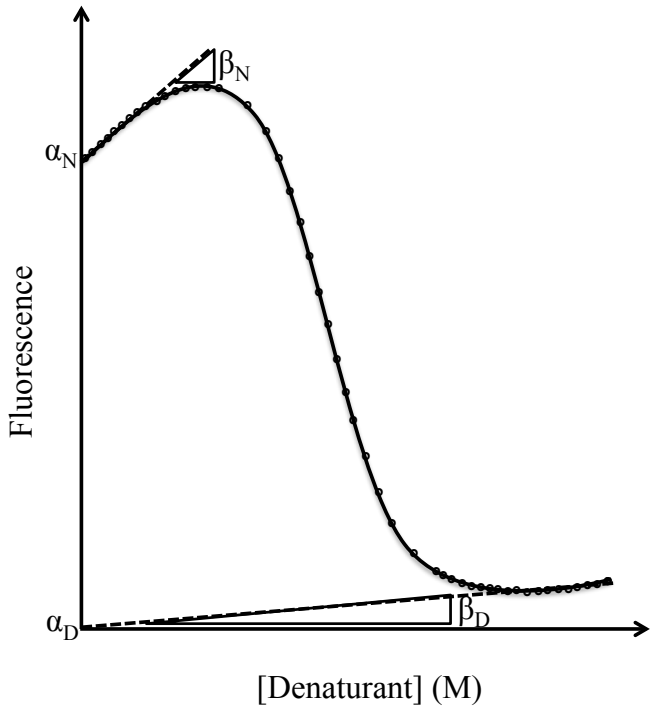


Figure 2.3: Diagram to explain visually the meaning of the terms from Equation 2.3.

To allow comparison between different equilibrium curves, the fluorescence reading (*F*) is converted to fraction folded reading, which varies between 0 and 1 to indicate the proportion of the protein found in the native state at given denaturant concentration:

$$\text{Fraction folded} = \frac{F-(\alpha_D+\beta_D[\text{den}])}{\alpha_N+\beta_N[\text{den}])-(\alpha_D+\beta_D[\text{den}])} \qquad \textbf{Equation 2.4}$$

2.1.2 Kinetics studies of proteins

Protein folding was initially considered to be a process involving a series of intermediates so that a random search for the correct conformation was not feasible (Levinthal paradox). However, the study of the chymotrypsin inhibitor 2 showed a two-state kinetic behaviour with no *populated* intermediates (Jackson and Fersht 1991). In this case, proteins would fold apparently directly from the denatured to the native state via a transition state (TS), which is the highest energy point on the reaction pathway (Figure 2.4).

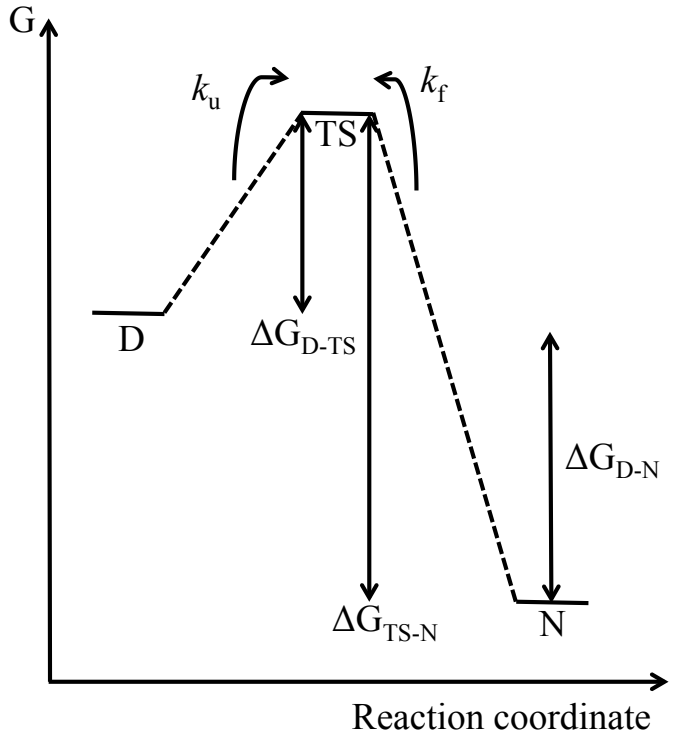


Figure 2.4: Free energy diagram showing the denatured, transition and native states.

The free energies of transfer of the native, transition and denatured states from water into a denaturant solution are linearly proportional to [den]. As well as in the Equation 2.1, the following equations can be stated:

$$\Delta G_{\mathrm{D-TS}} = \Delta G_{\mathrm{D-TS}}^{\mathrm{H_2O}} - m_{\mathrm{D-TS}}\,[\mathrm{den}] \qquad \textbf{Equation 2.5}$$

$$\Delta G_{\mathrm{TS-N}} = \Delta G_{\mathrm{TS-N}}^{\mathrm{H_2O}} - m_{\mathrm{TS-N}}\,[\mathrm{den}] \qquad \textbf{Equation 2.6}$$

Using the relationship $k = A \exp(-\Delta G/RT)$:

$$\Delta G_{\mathrm{D-TS}} = RT \ln k_{\mathrm{f}} - RTln\ A \qquad \textbf{Equation 2.7}$$

$$\Delta G_{\mathrm{TS-N}} = -RT \ln k_{u} - RTln\ A \qquad \textbf{Equation 2.8}$$

$$\ln k_{u} = \ln k_{u}^{\mathrm{H_2O}} + m_{ku}\,[\mathrm{den}] \qquad \textbf{Equation 2.9}$$

$$\ln k_{f} = \ln k_{\mathrm{f}}^{\mathrm{H_2O}} + m_{\mathrm{kf}}\,[\mathrm{den}] \qquad \textbf{Equation 2.10}$$

Assuming two-state kinetics, at concentrations below $[\mathrm{den}]_{50\%}$ the natural log of the folding rate against [den] (chevron plot) will be a straight line with gradient m_{kf}. Similarly, above $[\mathrm{den}]_{50\%}$ the natural log of the folding rate against [den] will be a straight line with gradient m_{ku} (Figure 2.5). Kinetic parameters can be extracted from the observed chevron plot by using the equation:

$$\ln k_{\mathrm{obs}} = \ln[k_{f}^{\mathrm{H_2O}} \exp(-m_{kf}\,[\mathrm{den}]) + [k_{u}^{\mathrm{H_2O}} \exp(m_{ku}\,[\mathrm{den}])] \qquad \textbf{Equation 2.11}$$

k_{obs} is the observed rate constant at a particular [den].

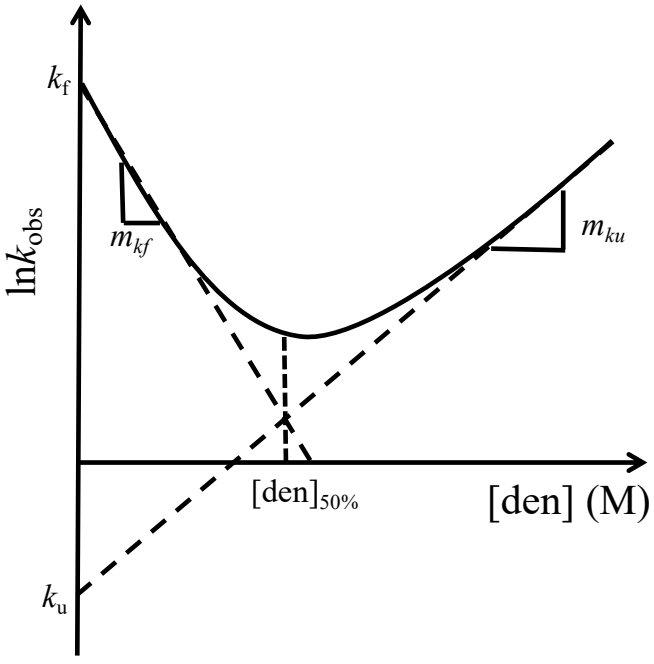


Figure 2.5: Chevron plot and its kinetic parameters.

The kinetic free energy of unfolding in water, $\Delta G^{\mathrm{H_2O}}_{\mathrm{D-N\ kin}}$ can be calculated from:

$$\Delta G^{\mathrm{H_2O}}_{\mathrm{D-N\ kin}} = -RTln(k_{\mathrm{u}}^{\mathrm{H_2O}}/k_{\mathrm{f}}^{\mathrm{H_2O}}) \qquad \textbf{Equation 2.12}$$

And since $m_{\mathrm{D-TS}} = RTm_{kf}$ and $m_{\mathrm{TS-N}} = RTm_{ku}$:

$$m_{\mathrm{D-N}\,kin} = m_{\mathrm{D-TS}} + m_{\mathrm{TS-N}} = RT(m_{ku} + m_{kf}) \qquad \textbf{Equation 2.13}$$

So, if kinetic and equilibrium $m_{\mathrm{D-N}}$ are the same and if both equilibrium and kinetic $\Delta G^{\mathrm{H_2O}}_{\mathrm{D-N}}$ are also the same, folding can be described as two-state, where there are no populated intermediates in the folding pathway.

2.1.3 Φ-value analysis

A standard experimental procedure to investigate the transition state in protein folding consists of mutating a single residue in the primary sequence of the polypeptide and seeing the effect of the mutation on the equilibrium and kinetics studies. It is known that mutations can affect the free energy of the native and transition states and consequently cause a difference in their equilibrium stabilities ($\Delta\Delta G_{\text{D-N}}$ and $\Delta\Delta G_{\text{D-TS}}$).

$$\Delta\Delta G_{\mathrm{D-N}} = \Delta G^{\mathrm{WT}}_{\mathrm{D-N}} - \Delta G^{\mathrm{MUT}}_{\mathrm{D-N}} \qquad \textbf{Equation 2.14}$$

$$\Delta\Delta G_{\mathrm{D-TS}} = \Delta G^{\mathrm{WT}}_{\mathrm{D-TS}} - \Delta G^{\mathrm{MUT}}_{\mathrm{D-TS}} \qquad \textbf{Equation 2.15}$$

If a particular residue is located in a region of the protein with a native conformation in the transition state, then $\Delta\Delta G_{\mathrm{D-TS}}$ will be the same as $\Delta\Delta G_{\mathrm{D-N}}$ (Figure 2.6, A).

$$\Phi = \frac{\Delta\Delta G_{D-TS}}{\Delta\Delta G_{D-N}} = 1 \qquad \textbf{Equation 2.16}$$

However, if a residue is located in a region of the protein with a denatured conformation in the transition state (no native contacts), then the $\Delta\Delta G_{\mathrm{D-TS}}$ will be 0 (Figure 2.6, B).

$$\Phi = \frac{\Delta\Delta G_{D-TS}}{\Delta\Delta G_{D-N}} = 0 \qquad \textbf{Equation 2.17}$$

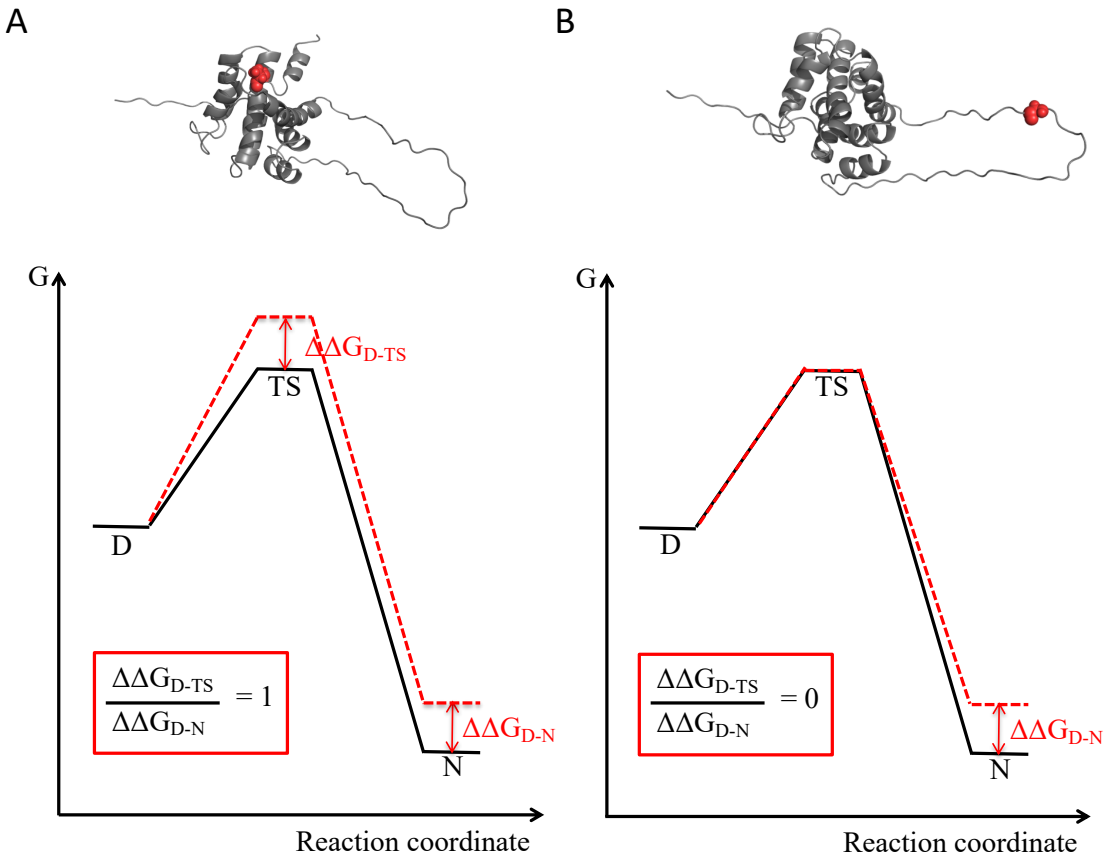


Figure 2.6: Φ-value analysis. (A) Mutation destabilizes the native and the transition states to the same extent, giving a Φ-value of 1. (B) Mutation destabilizes the native state but does not affect the transition state giving a Φ-value of 0.

2.2 Coupled folding and binding reactions

2.2.1 Reaction rates

As previously stated, in a reversible reaction, the rate constant k describes the flux of the forward and reverse reactions. For protein folding events, the rates are termed k_u (rate constant for unfolding) and k_f (rate constant for folding). Analogous to that, in a two-state bimolecular reaction, the rates are termed k_{on} (association rate constant) and k_{off} (dissociation rate constant).

For a two-state protein, the change in concentration of the denatured protein is described by Equation 2.18:

$$\frac{\mathrm{d}[D]}{\mathrm{d}t} = k_{\mathrm{u}}[N] - k_{\mathrm{f}}[D] \qquad \textbf{Equation 2.18}$$

Where [N] and [D] are the concentrations of native and denatured protein, respectively, and k_{u} and k_{f} are the rate constant of unfolding and folding, respectively.

When equilibrium is achieved, the change in concentration of native and denatured protein is equal. Therefore, the equilibrium constant K for a two-state protein is described by Equation 2.20:

$$0 = k_{\mathrm{u}}[N] - k_{\mathrm{f}}[D] \qquad \textbf{Equation 2.19}$$

$$K = \frac{[D]}{[N]} = \frac{k_{\mathrm{u}}}{k_{\mathrm{f}}} \qquad \textbf{Equation 2.20}$$

In a reaction involving two proteins, where proteins A and B interact to form a complex C, like in IDP coupled folding and binding reactions, the rate by which the complex is formed, will depend on the concentration of both proteins. Hence, the change in concentration of protein A is described by Equation 2.21:

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k_{\mathrm{off}}[C] - k_{\mathrm{on}}[A].[B] \qquad \textbf{Equation 2.21}$$

Where [A], [B] and [C] are the concentrations of proteins A, B and complex C, respectively, and k_{on} and k_{off} are the association and dissociation rates constants, respectively.

Analogous to the uni-molecular process, once equilibrium is achieved, the change in concentration of protein A is zero and the equilibrium constant K_d is described by Equation 2.23:

$$0 = k_{\text{off}}[C] - k_{\text{on}}[A][B] \qquad \textbf{Equation 2.22}$$

$$K_d = \frac{[A][B]}{[C]} = \frac{k_{\text{off}}}{k_{on}} \qquad \textbf{Equation 2.23}$$

As previously discussed, the unfolding of a protein upon rapidly mixing with denaturant can be followed by a change in the fluorescence. The fluorescence change can be described by a single exponential function, that is derived from the rate law. Starting from a two-state reaction (Equation 2.18), the concentration of native protein can be replaced by the change in concentration of denatured protein:

$$\frac{\mathrm{d}[D]}{\mathrm{d}t} = k_{\text{u}}([D]_0 - [D]) - k_{\text{f}}[D] \qquad \textbf{Equation 2.24}$$

$$\frac{\mathrm{d}[D]}{\mathrm{d}t} + (k_{\text{u}} + k_{\text{f}})[D] = k_{\text{u}}[D]_0 \qquad \textbf{Equation 2.25}$$

Equation 2.25 can be then integrated:

$$[D] = \frac{k_u[D]_0}{k_{u+}k_f} + C\mathrm{e}^{(-k_{u+}k_f)t} \qquad \textbf{Equation 2.26}$$

Which is equivalent to a single exponential function, described by equation 2.27:

$$[D] = F_{\text{final}} + \Delta F\mathrm{e}^{(-k_{\text{obs}})t} \qquad \textbf{Equation 2.27}$$

Where F_{final} is the fluorescence reading at the end of the reaction, ΔF in the fluorescence change along the reaction, k_{obs} is the observed rate constant ($k_u + k_f$) and t is time.

For a reaction involving two proteins in near equimolar concentrations, the concentrations of both proteins are relevant and must be considered. The model used by Shammas *et al.* (2013) contemplates this case and the integration of the rate law is more complex than previously demonstrated. When one of the proteins in the reaction has a significantly higher concentration in comparison to the other (pseudo-first order condition), the model is again simplified. For instance, if the concentration of protein B is 10 times higher than the concentration of protein A, it can be assumed that only 10% of protein B will associate and go into complex. In that case, the concentration of protein B will not change upon the reaction and therefore can be incorporated in the association rate constant:

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k_{\mathrm{off}}[C] - k'_{\mathrm{on}}[A] \qquad \textbf{Equation 2.28}$$

where k'_{on} is the new association rate constant and is equal to k_{on} multiplied by the concentration of B (10 times in excess).

3 Materials and methods

Throughout the entire course of the biophysical experiments for this thesis, all the buffers and reagents were prepared at room temperature, volumetrically, with deionised water (ddH_2O) purified at 18 mΩ cm^{-1} of resistance by an ELGA maxima system. The pH of the buffers and solutions were checked using Radiometer MeterLab PHM210 instrument.

In this chapter, suppliers will be referenced as a superscript:

(a) MP Biomedicals Inc., (b) Life Technologies, (c) Sigma-Aldrich Company, (d) Fisher Scientific International Inc., (e) STARLAB group (f) Biotium, (g) Thermo Scientific, (h) Acros organics, (i) BDH laboratory supplies, (j) GE Healthcare Life Sciences, (k) Agarose Bead Technologies, (l) New England Biolabs, (m) Eurogentec (n) Merck.

3.1 Molecular Biology

LB

LB media was prepared from commercially available capsules[a]. 10 g tryptone, 5 g yeast extract and 10 g NaCl were diluted in 1 litre ddH_2O and autoclaved. When stored, the solutions were kept 4 °C.

LB agar plates

The plates were prepared from commercially available capsules[a]. 10 g tryptone, 5 g yeast extract, 10 g NaCl and 15 g agar were diluted in 1 litre ddH_2O and autoclaved. After cooling down to about 50 °C, kanamycin or ampicillin was added (60 μg.ml^{-1} and 100 μg.ml^{-1} final concentration, respectively) and the plates were stored at 4 °C.

Kanamycin (KAN)

3 g of kanamycin[b] was diluted in 50 ml ddH_2O to prepare a [1000x] stock (60 mg.ml^{-1} final concentration). The solution was syringe filtered through a 0.2 µm cellulose acetate membrane and stored at 4 °C.

Ampicillin (AMP)

5 g of ampicillin[d] was diluted in 50 ml ddH_2O (100 mg.ml^{-1} final concentration). The solution was syringe filtered through a 0.2 µm cellulose acetate membrane and stored in 1 mL aliquots at -20 °C.

LB - Kanamycin and LB- Ampicillin (LB- KAN and LB- AMP)

After the LB media was prepared using the standard procedure described before, kanamycin or ampicillin was added to a final concentration of 60 µg.ml^{-1} or 100 µg.ml^{-1}, respectively.

Tris- Acetate-EDTA buffer (TAE)

A [1x] buffer composition consists in 40 mM Tris (pH 8.0), 20 mM acetic acid, and 1mM EDTA. A [50x] stock was prepared using 242 g Tris base[c], 100 ml of 0.5 M EDTA[g] (pH 8.0) and 57.1 ml glacial acetic acid[d], diluted in 1 litre of ddH_2O.

Agarose gel

0.5 g agarose[e] was added to 50 ml [1x] TAE buffer and heated up (> 100 °C) until it was dissolved. Then, 5 µl of the DNA stain GelRed[f] was added to produce a 1% (*w/v*) agarose gel.

Phusion Hot Start II DNA polymerase[g]

It is the DNA polymerase used in all site direct mutagenesis (SDM) and PCR protocols. It is commercially available and it is supplied with its appropriate buffer. It was stored at -20 °C.

Deoxynucleotide mix (dNTP)[g]

Commercially available solution containing dATP, dCTP, dGTP and dTTP, each at final concentration of 10 mM. 10 µl aliquots stored at -20°C.

*Dpn*I [g]

The restriction enzyme used to digest methylated DNA. It is supplied in appropriate buffer and stored at -20°C.

Gene ruler 1kb plus [g]

Commercially available ladder used for quantification and size determination of double-stranded DNA on agarose gels. It covers a wide range (75-2000 bp) of DNA lengths and it is supplied with its appropriate buffer. Stored at 4 °C.

3.2 Buffers and reagents

3.2.1 Protein production and purification

SasG

Binding buffer 1: 20 mM Tris (pH 8.0), 100 mM NaCl, 20 mM imidazole.

1.76 g Trizma acid[c], 1.05 g Trizma base[c], 5.84 g NaCl[c], 1.36 g imidazole[h] were diluted in 1 litre of ddH_2O and vacuum filtered through a 0.2 µm cellulose acetate membrane.

Gradient buffer 1: 20 mM Tris (pH 8.0), 100 mM NaCl, 500 mM imidazole.

1.76 g Trizma acid[c], 1.05g Trizma base[c], 5.84g NaCl[c], 34.04g imidazole[h] were diluted in 1 litre of ddH_2O and vacuum filtered through a 0.2 µm cellulose acetate membrane.

Phosphate buffered saline (PBS) [20x]

A [1x] solution consists in 10 mM phosphate (pH 7.4), 136.5 mM NaCl, 2.7 mM KCl (164 mM ionic strength). A [20x] stock solution was prepared with 28.83 g Na_2HPO_4 (dihydrate)[d], 5.93 g NaH_2PO_4 (dihydrate) [d], 159.54 g NaCl[c], 4.03 g KCl[i] diluted in 1 litre of ddH_2O.

Binding buffer 2: 10 mM phosphate (pH 7.4), 136.5 mM NaCl, 2.7 mM KCl, 20 mM imidazole. A [1x] solution was prepared from 50 ml [20x] PBS stock and 1.36 g imidazole[h] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

Gradient buffer 2: 10 mM phosphate (pH 7.4), 136.5 mM NaCl, 2.7 mM KCl, 500 mM imidazole.

A [1x] solution was made with 50 ml [20x] PBS stock and 34.04 g imidazole[h] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

Coomassie brilliant blue (CBB) stain: 0.2% (*w/v*) CBB R-250, 50% (*v/v*) EtOH, 10% (*v/v*) CH_3COOH.

A 500 ml stock solution was prepared dissolving 1 g of CBB R-250[g] in 250 ml ethanol[c], 50 ml glacial acetic acid[d] and 200 ml of ddH_2O.

Coomassie brilliant blue (CBB) destain: 10% (*v/v*) EtOH, 10% (*v/v*) CH_3COOH.

A 1 litre stock solution was prepared by adding 100 ml of ethanol[c], 100 ml glacial acetic acid[d] in 800 ml ddH_2O.

Isopropyl β-D-1-thiogalactopyranoside (IPTG)[g]

A 1 M stock was prepared diluting 2.38 g of IPTG in 10 mL dd H_2O. The solution was filtered through a 0.22 μm sterile membrane and 400 μl aliquots stored at -20 °C.

Mark 12 unstained standard[b]

Commercially available ladder used for size determination of proteins on a NuPAGE gels. It covers a range 2.5-200 kDa and it is supplied ready to use. It was stored at 4 °C.

HRV 3C protease[g]

Commercially available protease supplied as a 10,000 units with a highly specific recognition sequence for Leu-Glu-Val-Leu-Phe-Gln-↓-Gly-Pro, cleaving after the glutamine residue. It was used to cleave histidine tags in the SasG purification process.

A1 and BID full-length

Tobacco etch virus (TEV) protease:

TEV protease was expressed and purified in the lab by Tristan Kwan following Blommel and Fox protocol (Blommel and Fox 2007). It was stored in 20 mM TRIS (pH 7.5), 125 mM NaCl, 2 mM DTT, 2 mM EDTA and 50% (v/v) glycerol at -20 °C.

TEV cleavage buffer:

10 mM TRIS pH 8.0, 150 mM NaCl. It was prepared with 0.866 g Trizma acid[c], 0.533 g Trizma base[c] and 8.519g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 μm cellulose acetate membrane.

Instant blue[c]:

Commercially available coomassie dye-based protein stain. It is a solution formulated for fast and sensitive protein detection. The bands were visible in less than 15 minutes without the need to destain.

Biophysical buffer:

50 mM Sodium Phosphate, 0.05% Tween, pH 7. It was prepared with 5.29 g Na_2HPO_4 (dihydrate)[d], 3.17 g NaH_2PO_4 (dihydrate) [d] and 500 μL of Tween 20[d] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 μm cellulose acetate membrane.

Ion-exchange buffer:

20 mM Tris, 0 and 1 M NaCl, pH 7. A [10x] stock was prepared with 28.8 g Trizma acid[c], 1.94 g Trizma base[c] and 0 or 58.44 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 μm cellulose acetate membrane.

A [1x] solution was made by dilution of the [10x] stock.

Page ruler unstained protein ladder[g]:

Commercially available ladder used for protein size determination on SDS gels. It covers a range from 5 kDa to 250 kDa and it is supplied ready to be used. It was stored at -20 °C.

Glutathione Sepharose 4B GST-tagged protein purification resin[j]:

Commercially available affinity chromatography resin for batch purification of GST- tagged proteins.

Dithiothreitol (DTT)[b]:

A 1 M stock was prepared with 1.54 g DTT diluted in 10 mL ddH_2O or in an appropriate buffer. 400 µL aliquots were stored at -20 °C.

Truncated BID (t-BID):

Caspase 8[n]:

Commercially available protein (Uniprot Q14790) employed as a protease to cleave full-length BID generating t-BID. Caspase was resuspended in 50 µL of biophysical buffer with 15% glycerol.

PUMA

Nickel resin[k]:

Commercially available nickel resin that recognizes exposed histidine residues and retain a protein containing a His-tag. Used in PUMA purification process.

Factor Xa[l]:

Commercially available protease with a recognition sequence of Ile-Glu/Asp-Gly-Arg, cleaving after the Arg residue. It was used to cleave PUMA from GB1.

Factor Xa cleavage buffer:

20 mM Tris pH 8.0, 50 mM NaCl, 5 mM $CaCl_2$. Prepared by the dilution of the [10x] 20 mM Tris, pH 8 with the addition of 2.92 g of NaCl and 0.74 g $CaCl_2$ in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane. A [10x] stock of 20 mM Tris was prepared with 28.8 g Trizma acid[c], 1.94 g Trizma base[c].

PBS 25 mM imidazole buffer:

Prepared by the dilution of the [20x] PBS and the addition of 1.70 g of Imidazole diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

3.2.2 Ionic strength studies

MOPS 5 mM ionic strength buffer:

10 mM MOPS (pH 7.4), 5 mM ionic strength due to the buffer. It was prepared with 0.88 g of MOPS acid[c], 1.32 g of MOPS base[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

MOPS 1 M ionic strength buffer:

10 mM MOPS (pH 7.4), 1 M ionic strength. It was prepared with 0.75 g of MOPS acid[c], 1.46 g of MOPS base[c] and 58.18 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

3.2.3 pH dependence studies

Phosphate buffer, pH 2

10 mM phosphate, 8 mM ionic strength, pH 2. It was prepared with 0.63 g Na_2HPO_4 (dihydrate)[d], 0.67 g NaH_2PO_4 (dihydrate)[d], 0.212 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

Phosphate buffer, pH 3

10 mM phosphate, 8 mM ionic strength, pH 3. It was prepared with 0.123 g Na_2HPO_4 (dihydrate)[d], 1.37 g NaH_2PO_4 (dihydrate)[d] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

Acetate buffer, pH 4

10 mM acetate, 8 mM ionic strength, pH 4. It was prepared with 0.50 g acetic acid, 0.218 g sodium acetate, 0.374 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

Acetate buffer, pH 4.5

10 mM acetate, 8 mM ionic strength, pH 4.5. It was prepared with 0.372 g acetic acid, 0.503 g sodium acetate, 0.247 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

Acetate buffer, pH 5

10 mM acetate, 8 mM ionic strength, pH 5. It was prepared with 0.204 g acetic acid, 0.880 g sodium acetate, 0.084 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

MES buffer, pH 6

10 mM MES, 8 mM ionic strength, pH 6. It was prepared with 1.15 g of MES acid[c], 0.160 g of MES base[c] and 0.231 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 µm cellulose acetate membrane.

MOPS buffer, pH 7

10 mM MOPS, 8 mM ionic strength, pH 7. It was prepared with 1.339 g of MOPS acid[c], 0.809 g of MOPS base[c] and 0.263 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 μm cellulose acetate membrane.

TRIS buffer, pH 8

10 mM TRIS, 8 mM ionic strength, pH 8. It was prepared with 0.866 g Trizma acid[c], 0.533 g Trizma base[c], 0.141 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 μm cellulose acetate membrane.

TRIS buffer, pH 9

10 mM TRIS, 8 mM ionic strength, pH 9. It was prepared with 0.173 g Trizma acid[c], 1.066 g Trizma base[c], 0.402 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 μm cellulose acetate membrane.

CAPS buffer, pH 10

10 mM CAPS, 8 mM ionic strength, pH 10. It was prepared with 1.63 g acid component[c], 0.10 g basic component[c], 0.319 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 μm cellulose acetate membrane.

CAPS buffer, pH 11

10 mM CAPS, 8 mM ionic strength, pH 11. It was prepared with 0.487 g acid component[c], 0.30 g basic component[c], 0.016 g NaCl[c] diluted in 1 litre of ddH_2O. The solution was vacuum filtered through a 0.2 μm cellulose acetate membrane.

3.2.4 Urea solutions

A 9.5 M urea[j] stock solution was made in the appropriate buffer (PBS or biophysical buffer). For equilibrium experiments, a series of dilutions were made in order to get a range of different concentrations. For proteins containing cysteines, DTT (10 mM final concentration) was also added to the urea solution in order to prevent disulphide bond formation. The urea concentration was measured using an Atago 1T refractometer (at 25°C ± 0.5°C) and the equation:

$$[\text{Urea}] = 117.66x + 29.753x^2 + 185.56x^3 \quad \textbf{Equation 3.1}$$

Where x is the difference in reading between the urea solution and the pure buffer.

3.3 Experimental protocols

3.3.1 Preparation of chemically competent cells

E. coli strains (DH5α, BL21- DE3 and C41-pLysS) were used to prepare competent cells following the protocol:

In a sterile Erlenmeyer, 200 µL of bacteria cell was added to 50 mL of LB media and incubated in a shaker at 37 °C for approximately 5 hours (OD_{600} 0.3-0.6). The solution was then pelleted by centrifugation (Heraeus Multifugue X3R[g], 3000 RPM, 10 minutes, 4 °C) and the supernatant was discarded. The cell pellet was resuspended in 10 ml 100 mM $CaCl_2$, 15% (v/v) glycerol solution and incubated on ice for 20 minutes. Once again, the solution was pelleted by centrifugation and the supernatant discarded. The cell pellet was resuspended in 2 ml 100 mM $CaCl_2$, 15% (v/v) glycerol solution, aliquoted in 50 µl samples, flash-frozen in liquid nitrogen and stored at -80°C.

3.3.2 Heat-shock transformation

1 µl of a plasmid stock of interest was added into 50 µl of chemically competent *E. coli* cells and incubated on ice for approximately 30 minutes. The cells were heated up in a water bath at 42 °C for 45 seconds and then transferred back into ice for 5 more minutes. Then, 850 µl of LB media was added and the cells were incubated in a shaker at 37 °C for 45 minutes. The cells were pelleted by centrifugation (2400 ×*g*, 3 minutes, 25 °C) and resuspended in approximately 50 µl of its supernatant to be spread onto either a LB-KAN or LB-AMP agar plate. The plate was incubated overnight at 37 °C.

3.3.3 Plasmid production and purification

On the next day following the transformation, a single colony was taken from the agar plate, added to 5 mL of either LB-KAN or LB-AMP media and incubated overnight at 37 °C. The cells were then pelleted by centrifugation (17000 ×*g*, 5 minutes, 25 °C) and the plasmid was purified using a QIAprep spin Miniprep kit (QIAGEN) following the manufacturers protocol and stored at - 20 °C.

SasG, pH and ionic strength dependence studies: Dr. Dominika Gruszka kindly provided me with SasG genes (E-G5^2 and G5^2) that were inserted into a pSKB2 plasmid that contained a N-terminal histidine tag and a HRV 3C protease recognition cleavage site. The plasmid contains a T7 promoter and a kanamycin resistant gene (KanR).

SasG, folding on the ribosome studies: A modified version of the commercially available pRSETa vector was used in all the folding on the ribosome experiments. The plasmid contains a T7 promoter and an ampicillin resistant gene (AmpR). This modified plasmid also encodes an N-terminal polyhistidine tag (6 x His) and a Thrombin cleavage site.

SasG, effect of the charges on the ribosome studies: To study the effect of the charges on the co-translational folding of SasG, three versions of G5^2 and E-G5^2 were designed: one

with all the negative charges replaced, one with all the positive charges replaced and one with all the charges replaced. The 3 genes for E-G5^2 were purchased from GenScript in pUC57 vector. The 3 genes for G5^2 were cut from the purchased E-G5^2 ones. Then, a molecular biology step was done to insert all genes into the same modified pRSETa, the same used for all the other ribosome studies (see above).

A1: Bonsu Ku (Korea Research Institute of Bioscience and Biotechnology) donated the A1 gene inserted into a slighted modified pGEX-4T-3 plasmid, which contained a TEV cleavage site. As a result of TEV cleavage, an additional GS was left at the N-terminus of A1.

PUMA: Expressed as a GB1 fusion protein. PUMA gene was inserted into a modified version of pRSETa plasmid, engineered by Dr Joseph Rogers, which contained an N-terminal histidine tag preceding the GB1 and a Factor Xa cleavage site between GB1 and PUMA.

Full-length BID and t-BID: The protein encoding genes were designed and purchased from GenScript in pUC57 vector. In the lab, I designed a molecular biology strategy to insert both genes into the pGEX-4T-3 plasmid, the same used for A1. The proteins were expressed and purified using a very similar protocol to A1.

3.3.4 Calculating plasmid concentration

With the purified plasmid DNA, its concentration was calculated using A_{260} nm absorption readings of a 10x diluted sample, measured on a CARY 50 BIO UV-Vis spectrometer. The sample was prepared using 10 µL of plasmid DNA diluted into 90 µL ddH_2O and measured against a blank sample consisted of 10 µL buffer EB (elution buffer, QIAGEN) diluted in 90 µL ddH_2O.

Considering that a solution containing 50 μg.ml^{-1} of double-stranded DNA has an absorbance at 260 nm of 1 and taking into account the dilution factor of the sample (DF = 10), the plasmid concentration was calculated following the equation:

$$[\text{Plasmid}] = A_{260} \cdot DF \cdot 50 \text{ ug.ml}^{-1} \qquad \textbf{Equation 3.2}$$

The quality of the DNA was determined by the ratio between A_{260} and A_{280} nm readings, with values in the range of 1.8-2.0 taken to mean good quality DNA.

3.3.5 Polymerase chain reaction (PCR)

The PCR protocol used in this thesis followed the Phusion Hot Start II DNA Polymerase specifications and the samples prepared in order to have a 50 μL total volume per reaction tube.

Table 3.1: Phusion Hot Start II DNA Polymerase recommended PCR sample preparation protocol for optimal performance.

Reaction Mix	Final concentration
H_2O	-
HF buffer	1x
dNTPs	200 μM each
Primer forward	0.5 μM
Primer reverse	0.5 μM
Template DNA (plasmid)	10 ng. μl^{-1}
Phusion Hot Start II DNA Polymerase	0.02 U/μl

The reaction was performed in a PTC-100 Programmable Thermal Controller (Thermo Fisher Scientific).

Table 3.2: Cycling program used for PCR

Step	Temperature (°C)	Time
Initial denaturation	95	2 min
	35 cycles	
Denaturation	95	30 sec
Annealing	55	30 sec
Extension	72	4 min
	End of cycles	
Final extension	72	20 min
Hold	10	∞

3.3.6 Site directed mutagenesis (SDM)

In order to edit a gene by adding, removing or mutating its DNA sequence, SDM was performed. The desired plasmid DNA was used as a template and complementary primers (30-40 nucleotides in length) were designed containing the desired change in its sequence.

The SDM protocol followed the Phusion Hot Start II DNA Polymerase specifications in order to get an optimal condition for the enzyme. The volumes for the sample preparation were calculated in order to have a 50 µL total volume per reaction tube.

Table 3.3: Phusion Hot Start II DNA Polymerase recommended SDM sample preparation protocol for optimal performance.

Reaction Mix	Final concentration
H_2O	-
HF buffer	1x
dNTPs	200 µM each
Primer forward	0.5 µM
Primer reverse	0.5 µM
Template DNA (plasmid)	20-50 ng. μl^{-1}
Phusion Hot Start II DNA Polymerase	0.02 U/µl

The SDM was performed in a PTC-100 Programmable Thermal Controller (Thermo Fisher Scientific).

Table 3.4: Cycling program used for SDM

Step	Temperature (°C)	Time
Initial denaturation	98	3 sec
	25 cycles	
Denaturation	98	30 sec
Annealing	65	30 sec
Extension	72	3 min
	End of cycles	
Final extension	72	10 min
Hold	4	∞

To digest the remaining methylated template DNA, 3 µL of *Dpn*I (10 U/ µL) was then added to each reaction tube and incubated overnight at 37 °C. Then, 10 µL of the sample was transformed into 100 µL chemically competent cells (DH5α) and the purified plasmid DNA was stored at -20 °C. In order to confirm the SDM success, the sample was sent for sequencing (Genewiz Inc.).

3.3.7 Ligation

Ligation reaction was employed in order to join two nucleic acid fragments and to create recombinant DNA molecules, such as when a foreign DNA fragment was inserted into a plasmid. The ligation reaction is described in Table 3.5.

Table 3.5: Sample preparation for ligation.

Reaction Mix	Volumes
H_2O	6 µl
[10x] ligase buffer T4	1 µl
T4 ligase (5 U/µL)	1 µl
Plasmid DNA (~100 ng/µL)	1 µl
DNA insert (~100 ng/µL)	1 µl

The sample was incubated at 4 °C overnight in order to guarantee an optimal efficiency of the reaction. In the following day, the sample was transformed into DH5α cells (10 µl of the ligation sample in 100 µl cells).

The plasmid was then purified using the QIAprep spin Miniprep kit following the manufacturer protocol (QIAGEN) and stored at -20 °C. The purified plasmid was sent for sequencing to confirm the reaction's success.

3.3.8 Protein production

1 µl of the desired plasmid DNA (~100 ng.µL^{-1}) was transformed into 50 µL BL21-DE3 (SasG and BID full-length) or C41-pLysS (PUMA and A1) chemically competent cells. On the following day, a single colony was added to 50 mL of LB-KAN or LB-AMP media and the solution incubated overnight at 37 °C.

Then, 10 mL of the pre-culture was used to inoculate 1 litre of LB-KAN or LB-AMP media. The flask was incubated in a shaker at 37 °C until OD_{600} was 0.6-0.8 when the cells have reached the exponential growth phase. To induce protein expression, 400 µl of IPTG (1 M stock) was added to each flask. The temperature was reduced (25 °C for SasG, 24 °C for A1 and 18 °C for PUMA) and the flask incubated overnight, shaking at 180 rpm. For BID full-length, after induction, the temperature was kept at 37 °C for 6 hours.

3.3.9 Protein purification

SasG

***S. aureus*, Uniprot Q2G2B2, residues 502-629**

EG5^2

GPETIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVE
KEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKD
PINELTEYGPET

G5^2

YGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIIS
KGESKEEITKDPINELTEYGPET

Figure 3.1: Amino acid sequence of SasG E-G5^2 and G5^2.

After the protein expression step, the cells were centrifuged (8000 x g, 15 minutes, 8° C) and the pellet was resuspended in 35 ml binding buffer 1 on ice. Cells were lysed by sonication (3 minutes total time, 3 seconds on, 7 seconds off) and the cell debris was pelleted by centrifugation (38000 ×*g*, 45 minutes, 8° C). The supernatant was filtered (0.22 µm) and purified by FPLC (AKTA, GE Healthcare). The sample was initially loaded onto a 5 ml HisTrap column equilibrated with Binding buffer 1. The protein was eluted from the HisTrap column using a linear Gradient buffer 1, (20-500 mM Imidazole gradient). The fractions containing the protein were collected and cleaved overnight with HRV 3C protease in order to remove the N-terminal His-tag. The sample was dialysed (3.5 KDa membrane) against 4 litres of [1x] PBS buffer overnight at 4 °C.

The protein sample in PBS buffer was submitted to a second step of purification using a HisTrap column. This step was to separate the cleaved His-tag and the remaining impurities from the protein. The sample was loaded into the column and eluted from the HisTrap with the Binding buffer 2. As a final step, Gradient buffer 2 was used to wash the HisTrap column. The purified protein was stored at -20 °C.

The purity of the protein was verified by SDS-PAGE (NuPAGE Bis Tris Mini Gels, Novex by Life Technologies) and the correct molecular weight confirmed by mass spectroscopy.

A1

Mouse, Uniprot Q07440, residues 1-152 containing P104K and C113S mutations

GSMAESELMHIHSLAEHYLQYVLQVPAFESAPSQACRVLQRVAFSVQKEVEKNLKSYLDD
FHVESIDTARIIFNQVMEKEFEDGIINWGRIVTIFAFGGVLLKKLKQEQIALDVSAYKQV
SSFVAEFIMNNTGEWIRQNGGWEDGFIKKFEPKS

Figure 3.2: Amino acid sequence of A1. The plasmid containing the P104K and C113S mutations was kindly donated by Bonsu Ku. The additional N-terminus GS was left as a result of a TEV cleavage.

The A1 sequence containing the mutations P104K and C113S was the same used to solve the structures of A1: PUMA and A1: BID (Smits et al. 2008). Cells were pelleted by centrifugation (8000 x g, 15 minutes, 8° C), resuspended in 35 mL of PBS and sonicated (3 minutes total time, 15 seconds on, 45 seconds off). The cell debris was pelleted by centrifugation (38000 ×*g*, 45 minutes, 8° C) and the supernatant was incubated with glutathione sepharose 4B resin (approximately 10 mL) for four hours at 4° C. Then, the resin containing the bound protein was washed once with PBS and twice with TEV cleavage buffer. Each washing step consisted in topping up the falcon tube with buffer, mixing it with the resin by inverting the tube and spinning it down by centrifugation (500 ×*g*, 5 minutes, 4° C). The supernatant was discarded after each wash. The resin was resuspended in TEV cleavage buffer and a final concentration of 2 mM EDTA and 10 mM DTT were added to the solution. In order to cleave the protein off the resin, 5 µM of TEV protease was added and the protein was incubated overnight at room temperature. The purification of the cleaved protein consisted of two FPLC steps: an ion-exchange followed by a size exclusion. Using an AKTA FPLC instrument, cleaved A1 was loaded onto an ion-exchange column (HiTrapTM SP HP, 5 mL) using the buffers 20 mM TRIS pH 7.0 with 0 and 1 M NaCl. A gradient of 0 – 23% NaCl over 60 mL was used to elute A1. Fractions containing the almost pure A1 (assessed by SDS-PAGE) were combined. As a final step of purification, the protein was loaded into a Superdex G75 gel filtration column used also to exchange the

buffer into the 50 mM sodium phosphate biophysical buffer. The purity of the protein was verified by SDS-PAGE and the correct molecular weight confirmed by mass spectroscopy. The protein was stored at 4° C.

PUMA peptide

Mouse, Uniprot Q99ML1, residues 127-161, containing the mutation M144A

RVEEEEWAREIGAQLRRAADDLNAQYERRRQEEQH

Figure 3.3: Amino acid sequence of PUMA peptide. The peptide was produced containing the M144A mutation to prevent oligomerization.

The mutation M144A was introduced to make the peptide less prone to self-association (Joseph M. Rogers, Wong, and Clarke 2014). Cells were pelleted by centrifugation (8000 x g, 15 minutes, 8° C), resuspended in 35 mL of PBS, 25 mM imidazole buffer and cells were lysed by sonication (3 minutes total time, 15 seconds on, 45 seconds off). Sonicate supernatant was incubated with Ni^{+2} agarose resin[k] (1.25 mL per litre of culture) for 1 hour at 4° C. The resin was washed twice with PBS, 25 mM imidazole and once with Factor Xa cleavage buffer. Washing steps consisted in filling up the falcon tube with buffer, mixing it by inversion and decanting the resin by centrifugation (500 ×*g*, 5 minutes, 4°C). The supernatant was discarded after each wash. The resin was resuspended in Factor Xa cleavage buffer and incubated with 20 ug of Factor Xa[l] overnight at room temperature. The purification of PUMA consisted of two steps: an ion exchange followed by size exclusion. Using an AKTA FPLC instrument, the resin supernatant was loaded onto an ion-exchange column (HiTrapTM SP HP, 5 mL), using the buffers 20 mM TRIS pH 7.0 with 0 and 1 M NaCl. PUMA was eluted from the column with a gradient of 0 – 250 mM NaCl over 90 mL. Fractions containing PUMA (assessed by SDS-PAGE) were pooled and loaded onto a size exclusion column (HiLoadTM 26/600 Superdex 30 pg). This final step was used also to exchange the buffer into the 50 mM sodium phosphate biophysical buffer. The purity of the protein was verified by SDS-PAGE and the correct molecular weight confirmed by mass spectroscopy. Purified PUMA was aliquoted, flash frozen in liquid nitrogen and stored at - 80 °C.

BID-TAMRA peptides

Mouse, Uniprot P70444, residues 76- 110

Mouse BID peptides, wild-type and mutants, were purchased from Biomatik (Canada). The peptides were N-terminally labelled with TAMRA fluorescent dye. The peptides were synthetized as trifluroacetate salts and were reconstituted in biophysical buffer. The solutions were filtered, aliquoted, flash frozen in liquid nitrogen and stored at - 80°C.

WT	SESQEEIIHNIARHLAQIGDEMDHNIQPTLVRQLA
E81G	SESQEGIIHNIARHLAQIGDEMDHNIQPTLVRQLA
E81A	SESQEAIIHNIARHLAQIGDEMDHNIQPTLVRQLA
I82A	SESQEEAIHNIARHLAQIGDEMDHNIQPTLVRQLA
I83A	SESQEEIAHNIARHLAQIGDEMDHNIQPTLVRQLA
N85G	SESQEEIIHGIARHLAQIGDEMDHNIQPTLVRQLA
N85A	SESQEEIIHAIARHLAQIGDEMDHNIQPTLVRQLA
I86A	SESQEEIIHNAARHLAQIGDEMDHNIQPTLVRQLA
R88G	SESQEEIIHNIAGHLAQIGDEMDHNIQPTLVRQLA
R88A	SESQEEIIHNIAAHLAQIGDEMDHNIQPTLVRQLA
L90A	SESQEEIIHNIARHAAQIGDEMDHNIQPTLVRQLA
Q92G	SESQEEIIHNIARHLAGIGDEMDHNIQPTLVRQLA
Q92A	SESQEEIIHNIARHLAAIGDEMDHNIQPTLVRQLA
I93A	SESQEEIIHNIARHLAQAGDEMDHNIQPTLVRQLA
E96G	SESQEEIIHNIARHLAQIGDGMDHNIQPTLVRQLA
E96A	SESQEEIIHNIARHLAQIGDAMDHNIQPTLVRQLA
M97A	SESQEEIIHNIARHLAQIGDEADHNIQPTLVRQLA
H99G	SESQEEIIHNIARHLAQIGDEMDGNIQPTLVRQLA
H99A	SESQEEIIHNIARHLAQIGDEMDANIQPTLVRQLA
I101A	SESQEEIIHNIARHLAQIGDEMDHNAQPTLVRQLA
I86A-M97A	SESQEEIIHNAARHLAQIGDEADHNIQPTLVRQLA

Figure 3.4: Amino acid sequences of synthetized BID wild-type and mutants. The peptides were synthesized with an N-terminal TAMRA fluorescent dye.

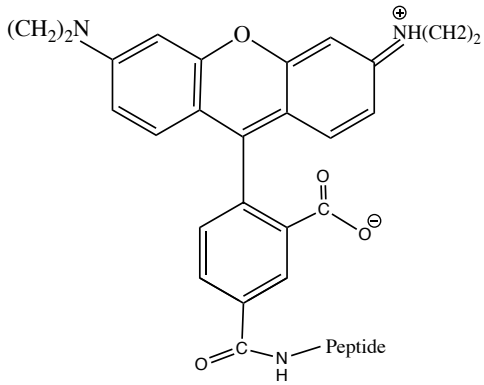


Figure 3.5: Chemical structure of TAMRA fluorescent dye. 5 Carboxytetramethylrhodamine (TAMRA) dye was attached to the N-terminus of all synthesized BID peptides. The presence of TAMRA did not change the net charge of the peptides. TAMRA chemical structure was constructed using ChemDraw Professional (version 17.0).

BID full-length

Mouse, Uniprot P70444, residues 1- 195

GSMDSEVSNGSGLGAEHITDLLVFGFLQSSGCTRQELEVLGRELPVQAYWEADLEDELQT
DGSQASRSFNQGRIEPDSESQEEIIHNIARHLAQIGDEMDHNIQPTLVRQLAAQFMNGSL
SEEDKRNCLAKALDEVKTAFPRDMENDKAMLIMTMLLAKKVASHAPSLLRDVFHTTVNFI
NQNLFSYVRNLVRNEMD

Figure 3.6: Amino acid sequence of BID full-length. The gene that encodes the protein was purchased from GenScript in the pUC57 vector. The additional GS at the N-terminus of the sequence was left as a result of a TEV cleavage.

Cells were pelleted by centrifugation (8000 x g, 15 minutes, 8° C), resuspended in 25 mL of PBS and the cells were lysed by sonication (3 minutes total time, 15 seconds on, 45 seconds off). The sonicate supernatant was incubated with glutathione sepharose 4B resin (approximately 10 mL) for four hours at 4° C. Resin was washed once with PBS and twice with TEV cleavage buffer. The beads were resuspended in TEV cleavage buffer and EDTA and DTT (2 and 10 mM final concentrations, respectively) were added to the solution. Then, 5 µM final concentration of TEV protease was added and the protein was incubated at room

temperature overnight. Differently than what was done for A1 purification, a HisTrap step was used for BID instead of an ion-exchange. BID full- length has an overall charge of -15 (32 positively charged and 17 negatively charged residues) and adding an ion-exchange step would be particularly challenging, as a very high salt concentration would be required. Since TEV has a His-tag, a His-Trap step was employed to separate cleaved TEV from the protein. Firstly, 20 mM imidazole was added to the protein sample to prevent BID binding to the nickel resin. Then, the solution was passed through a HisTrap column (2 mL/min flow rate). As a final purification step, the protein was loaded onto a Superdex G75 gel filtration column used also to exchange the buffer into the 50 mM sodium phosphate biophysical buffer. The purity of the protein was verified by SDS-PAGE and the correct molecular weight confirmed by mass spectroscopy. The protein was stored at 4 °C.

t-BID

Mouse, Uniprot P70444, residues 61- 195

SQASRSFNQGRIEPDSESQEEIIHNIARHLAQIGDEMDHNIQPTLVRQLAAQFMNGSLSE
EDKRNCLAKALDEVKTAFPRDMENDKAMLIMTMLLAKKVASHAPSLLRDVFHTTVNFINQ
NLFSYVRNLVRNEMD

Figure 3.7: Amino acid sequence of t-BID. The gene that encodes t-BID was purchased from GenScript in the pUC57 vector but the truncated protein could not be produced directly. It required the production of BID full-length and its truncation was produced using caspase 8.

Purified BID full-length was cleaved by caspase 8 in order to produce t-BID. The 25 U of the purchased caspase was resuspended in 50 µL of biophysical buffer with 15% glycerol. The protease was added to 10 mL of BID full-length (approximately 70 µM) with 5% glycerol, 10 mM DTT and 10 mM EDTA. The reaction was carried out at room temperature overnight. After the reaction, SDS-PAGE shows three components present: t-BID, BID fragment (60 residues cleaved from BID) and a small amount of uncleaved BID full-length. To unfold t-BID and BID fragment that remain folded after cleavage, the protein sample was incubated with 8 M urea for 2 hours at room temperature, before a size exclusion gel

filtration step. Using an AKTA FPLC instrument, the protein sample was loaded onto a G75 gel filtration column, previously equilibrated with 8 M urea in the biophysical buffer. Fractions containing separated t-BID, BID fragment and BID full-length were collected and the proteins were stored at 4 °C.

3.3.10 Calculating protein concentration

Protein concentration was calculated following the Beer-Lambert equation:

$$A = \varepsilon.c.l \qquad \textbf{Equation 3.3}$$

Where,

A is the absorbance at a specific wavelength.

ε is the extinction coefficient of the solution ($M^{-1}.cm^{-1}$).

c is the solution concentration (M).

l is the path length of the cuvette in which the sample is contained (cm).

The absorbance readings were taken in a CARY 50 BIO UV-Vis spectrometer. A sample of the corresponding buffer was used as a blank and the total absorbance calculated by the A_{280} values from the protein sample. The extinction coefficient for SasG (ε= 2980 $M^{-1}.cm^{-1}$) was obtained from the ProtParam tool (ExPASy). For a more accurate extinction coefficient and concentration calculation, multiple samples of A1, BID peptides and BID full-length were sent for amino acid analysis (AAA) performed at the Biochemistry department. The experimental concentration values from AAA were then plotted against the A_{280} (A_{555} for TAMRA labeled peptides) values and the gradients of the straight-line fits gave the extinction coefficients. (A1: ε= 24200 $M^{-1}.cm^{-1}$, BID-TAMRA peptides ε= 85000 $M^{-1}.cm^{-1}$ and BID full-length ε= 8700 $M^{-1}.cm^{-1}$).

3.3.11 Equilibrium folding stability: data acquisition and data analysis

Two solutions were prepared volumetrically to perform equilibrium experiments: 9 M and 0 M urea, both in 1x PBS. A Hamilton Microlab® was used to dispense the solutions in order to create a set of 68 tubes (800 µl samples) with urea concentration ranging from 0 to 9 M. A 100 µl sample of 45 µM protein stock solution in 1x PBS was added to each tube in order

to give a final protein concentration of 5 µM. The set of tubes were left to equilibrate at 25 °C for 30 minutes. The data were collected using a LS-50 luminescence spectrophotometer (PerkinElmer). Fluorescence readings were obtained from each solution equilibrated at 25 °C. The excitation wavelength was set to 276 nm and the scan taken from 290-390 nm (300 nm.min^{-1}).

The acquired data were exported to a Kaleidagraph spreadsheet (Synergy Software) plotted and fitted to Equation 2.3 (chapter 2), from which thermodynamic parameters ($m_{\mathrm{D-N}}$, $[\mathrm{den}]_{50\%}$ and $\Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}}$; see chapter 2 for details) were obtained. To allow comparison the fluorescence data were converted into fraction folded using the Equation 2.4 (chapter 2).

3.3.12 Folding on the ribosome: the arrest peptide assay

The arrest peptide assay with the incorporation of ^{35}S methionine[l] into nascent protein chains was performed using the PURExpress[m] *in vitro* translation kit. Up to 24 reactions could be run simultaneously.

A master mix reaction containing the solutions with the volumes recommended on the kit protocol and the ^{35}S methionine was prepared. 9 µl of the master mix was added to 1 µl of the template DNA within a 20 seconds interval between each sample. The final 10 µl reaction (Table 3.6) was incubated in a thermo mixer at 37 °C for 15 minutes and 500 rpm.

Table 3.6: Reaction mix for the incorporation of ^{35}S methionine into nascent protein chains.

Reaction	Volume
Solution I	5 µl
Solution II	0.5 µl
Solution III	0.5 µl
Template DNA	1.0 µl
^{35}S methionine[l]	0.5 µl
ddH_2O	2.5 µl

To stop the reaction after 15 minutes, 10 µl of ice-cold 10 % (*v/v*) trichloroacetic acid (TCA) in water was added to each tube within a 20 seconds interval between each sample. They were incubated on ice for 30 minutes.

Precipitated samples were pelleted by centrifugation for 5 minutes at 4°C and 14000 rpm. The supernatants were removed and discharged via the radioactivity waste sink. 10 µl of [1x] protein loading buffer was added to each tube and the pellets were resuspended by agitation (15 minutes, 37°C and 1400 rpm). Then 1.5 µl of RNase A (4 mg/ml) was added to each tube and samples were incubated for 15 minutes at 37° C with shaking at 700 rpm.

The samples were loaded (11.5 µl) onto a standard NuPage gel and run for 50 minutes at 200 V. After running, the gel was fixed in 50 ml Instant Blue solution for 45 minutes and then dried in a gel drier for 2 hours at 80 °C.

The dried gel was transferred to a phosphorimaging plate and left for two days. After that, the phosphorimaging plate was read on the Typhoon trio[i] and the image saved and analysed using the program ImageJ (Schneider, Rasband, and Eliceiri 2012); Figure 2.1).

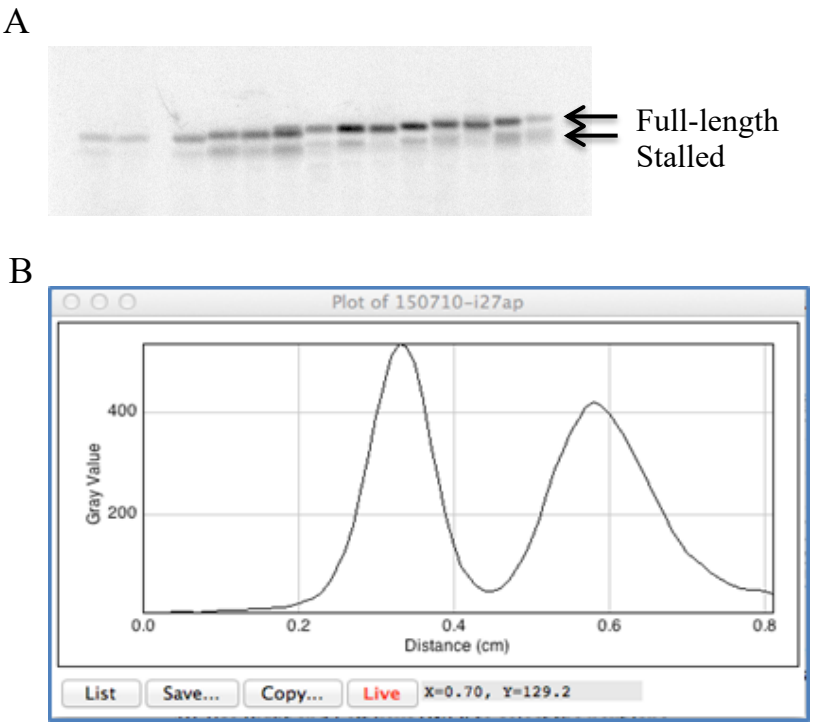


Figure 3.8: Arrest peptide assay. (A) A generic protein imaged gel. (B) The intensity of the two bands is quantified to generate a profile of intensity vs. position using the ImageJ program.

The data can be fitted to an equation involving the two Guassians and a flat baseline.

$$Y = A + Ce^{\frac{-(X-D)^2}{2E^2}} + Fe^{\frac{-(X-G)^2}{2H^2}} \qquad \textbf{Equation 3.4}$$

From equation 3.4, the area of each band can be quantified using:

$$\text{Area 1} = CE\sqrt{2\pi} \qquad \textbf{Equation 3.5}$$

$$\text{Area 2} = FH\sqrt{2\pi} \qquad \textbf{Equation 3.6}$$

And finally, each band can be quantified as a percentage of the total area or, in other words, we can find out the fraction full-length of the protein in each linker length sample and then get the force profile (fraction full-length vs. linker length).

3.3.13 Circular Dichroism (CD)

CD was employed in order to investigate the proteins and peptides secondary structures. The experiments were performed using an Applied Photophysics Chirascan spectrometer. Samples were prepared by weight to ensure an accurate concentration. In each experiment, the buffer in use was scanned and subtracted from each protein spectrum. The CD signal (millidegrees (mdeg)), obtained from far-UV CD is dependent on protein secondary structure as well as on the number of peptide bonds. Therefore, to allow comparison between different proteins, the CD signal was converted to mean residual ellipticity (MRE), which accounts for protein concentration, number of amino acids (aa) and cuvette path length (Equation 3.7):

$$\text{MRE} = \frac{\text{CD signal (mdeg)}}{\text{10 x concentration (M) x number of aa x path lenght (cm)}} \qquad \textbf{Equation 3.7}$$

Path length needs to be taken into account as the interactions between the polarised light and the protein molecules will vary not only with the protein concentration and its number of residues, but with the length that the light will pass through as well. A longer path length will result in a greater mdeg signal. However, increasing the path length also increases the background absorbance, which can be minimised by adjusting the detector voltage. The manufacturer's recommendation for the Applied Photophysics instrument was a maximum of 800 V. An optimization of signal to noise ratio was obtained carefully balancing the protein concentration and the path length for each experiment.

A1

Due to the presence of a single cysteine residue in A1 sequence, DTT (10 mM final concentration) was added to the protein sample to avoid disulphide-bond formation. DTT absorbs in the far-UV region, which increases the background absorbance. To optimize this condition, a 0.2 mm cuvette was chosen to allow enough light to pass through and reach the detector. Experimental settings were 1 nm bandwidth, 5 s adaptive time-per-point sampling at 25 °C. Different A1 concentrations were used and compared to check for oligomerization.

BID TAMRA peptides

To check for oligomerization propensity of TAMRA labelled BID peptides, two different concentrations (5 and 10 μM) were scanned and the spectra of each peptide compared. The experimental settings were 2 mm path length, 1 nm bandwidth, 5 s adaptive time-per-point sampling at 25 °C. All peptide samples were prepared and scanned on the same day. Each sample was sent for amino acid analysis (Department of Biochemistry) for an accurate concentration determination. Peptide helicity was calculated from the highest concentration sample.

Coupled folding and binding

CD was used to probe coupled folding and binding between IDP and partner protein, based on the assumption that a change in the structures would occur. Spectra of equimolar concentrations of each individual component were collected and converted to MRE. Then, the mean of the two individual sample spectra was calculated and the MRE would represent the spectrum expected if no structure change would occur upon binding. Next, a mixture of the two samples (1:1 volume ratio) was incubated at 25 °C for two hours and then scanned. The difference between the expected spectrum calculated from the individual scans and the one from the mixture, is related to the structure change upon the formation of the complex. To allow this comparison, the difference spectra were multiplied by the fraction of protein in complex. The concentration of the complex ([AB]) was estimated by the equation 3.8.

$$[\mathrm{AB}] = \frac{([\mathrm{A}]+[\mathrm{B}]+K_\mathrm{d}-\sqrt{(([\mathrm{A}]+[\mathrm{B}]+K_\mathrm{d})-4\,[\mathrm{A}].[\mathrm{B}])})}{2} \qquad \textbf{Equation 3.8}$$

Where, [A] and [B] are the total concentration of each protein and K_d is the dissociation constant for the complex. To calculate the fraction bound, the concentration of the complex ([AB]) was divided by the total amount of either protein A or B.

Helical content estimation

The helical content estimation in this thesis was calculated from the MRE value at 222 nm following Muñoz and Serrano method (Muñoz and Serrano 1995).

The method describes that the expected MRE of a helical protein at 0 °C is given by:

$$\mathrm{MRE}_{\mathrm{helix\ (0°C)}} = -39500\left(1-\frac{2.57}{\text{number of amino acids}}\right) \qquad \textbf{Equation 3.9}$$

For a random coil at a given temperature, the MRE is calculated by:

$$\mathrm{MRE_{coil}} = 400 - 45\mathrm{t} \qquad \textbf{Equation 3.10}$$

Where 400 is the MRE value at 222 nm for a random coil at 0 °C and t is the temperature in Celsius.

The MRE for a helix at a given temperature is calculated by:

$$\mathrm{MRE_{helix}} = \mathrm{MRE_{helix\ (0°C)}} + 100\mathrm{t} \qquad \textbf{Equation 3.11}$$

Where $\mathrm{MRE_{helix\ (0°C)}}$ can be calculated from the Equation 3.9 and t is the temperature in Celsius.

Then, the helical content can be estimated by:

$$\%\ \mathrm{Helix} = \frac{100}{(1+\left(\frac{(\mathrm{MRE_{222}} - \mathrm{MRE_{helix}})}{(\mathrm{MRE_{coil}} - \mathrm{MRE_{222}})}\right)} \qquad \textbf{Equation 3.12}$$

Where $\mathrm{MRE_{222}}$ is the measured MRE at 222 nm and $\mathrm{MRE_{coil}}$ and $\mathrm{MRE_{helix}}$ are calculated from the equations 3.10 and 3.11.

3.3.14 Association kinetics

Association kinetics experiments were performed by rapidly mixing protein solutions in a stopped-flow spectrometer (Applied Photophysics) and following the TAMRA dye fluorescence. The excitation wavelength was 555 nm and the emission was recorded above 570 nm, with a slit width of 1 mm. The temperature was maintained at 25 °C. A cut-off filter (570 nm) was utilized to reduce the detection of scattered light. A minimum of 15 traces were collected and averaged before analysis. Data collected before the instrument's mixing dead-time (1 ms) were removed.

The experiments were performed under pseudo-first order conditions, with the concentration of one protein at least 10-fold higher than the other. To minimize the signal to noise problem, the protein with the lower fluorescence was used in excess. The averaged trace of each concentration of protein in excess was fit to a single exponential decay function (Equation 2.26, chapter 2) and the observed rate constant (k_{obs}) was calculated by:

$$k_{obs} = k_{on}\,[A] + k_{off} \qquad \textbf{Equation 3.13}$$

Where [A] is the concentration of the protein in excess, and k_{on} and k_{off} are the association and dissociation rate constants, respectively. Thus, k_{on} was determined by the gradient from the fit of k_{obs} versus [A].

3.3.15 Dissociation kinetics

Dissociation kinetics experiments were performed using either a stopped-flow spectrometer (Applied Photophysics) (for $k_{obs} > 0.03$ s^{-1}) or a fluorescence spectrophotometer (Cary Eclipse, Varian) (for $k_{obs} < 0.03$ s^{-1}). Samples comprised of peptide/protein and its partner were mixed with various concentrations of an excess of competitor peptide. The idea being that the higher affinity peptide competitor would form a complex with the partner protein, releasing the peptide of interest, reaching a new equilibrium. The fluorescence change upon formation of the new complex was used to extract the k_{obs}. Traces were averaged and the mean was fit to a single exponential decay function (Equation 2.26, chapter 2). Excess of the competitor was used to guarantee that k_{obs} represents k_{off}, whereas at a low competitor/complex ratio, k_{obs} would represent the k_{off} plus a component of k_{on}, from the peptide/partner complex being reformed.

Experimental settings were 340 nm excitation and 360 nm emission wavelengths, when following intrinsic tryptophan fluorescence, or 555/575 nm for TAMRA labelled peptides. The temperature was maintained at 25 °C and slit widths were 2.5 and 10 nm on the fluorimeter or 1 nm on the stopped-flow. For TAMRA labelled BID peptides, the fluorescence change was small. Therefore, fluorescence anisotropy (vertical/vertical channel) was used to follow the dissociation kinetics.

Pre-formed complexes of 5 μM BID-TAMRA peptides/partner protein were mixed with different concentrations of PUMA, in a 1:9 or 1:10 ratio for the fluorimeter and stopped-flow, respectively. Samples were manually mixed before starting to be recorded in the fluorimenter, which caused the start of the reaction to be missed. This lag time was taken into account by recording the time between the addition of the competitor and the return of the cuvette to the fluorimeter. That information was then used to adjust the reaction time scale during the data analysis. Three data points of different concentrations of competitor were recorded and the observed rate constants (k_{obs}) were averaged. As the concentrations of competitor were sufficiently high, k_{obs} was equal k_{off}.

3.3.16 Equilibrium binding affinities

Equilibrium dissociation constants (K_d) were measured following fluorescence anisotropy, using a spectrophotometer (Cary Eclipse, Varian) at 25 °C. The TAMRA dye fluorophore was excited at 555 nm and its emission was recorded at 575 nm. The emission/excitation slit widths were 2.5 and 10, respectively. Vertical and horizontal channels were used to calculate the anisotropy.

Thus, a correction factor (G) was calculated to take into account the instrument capacity to detect the fluorescence in the different channels. G factor was calculated by:

$$G = \frac{I_{HV}}{I_{HH}} \qquad \textbf{Equation 3.14}$$

Where I_{HH} is the fluorescence intensity when using the horizontal excitation/emission channel and $I_{HV,}$ the horizontal excitation and vertical emission channels.

The anisotropy (R) can then be calculated by:

$$R = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}} \qquad \textbf{Equation 3.15}$$

Where I_{VV} is the fluorescence intensity when using the vertical excitation/emission channel and $I_{VH,}$ the vertical excitation and horizontal emission channels.

To account for the changes in fluorescence intensity upon the binding reaction, the anisotropy can be adjusted (R_{adj}) using the equation from Dandliker *et al* (Dandliker et al. 1981):

$$R_{adj} = \frac{\frac{x}{y} \times \frac{Q_f}{Q_b} \times R_b + R_f}{1 + \frac{x}{y} \times \frac{Q_f}{Q_b}} \qquad \textbf{Equation 3.16}$$

Where Q is the measured fluorescence, x is the difference between the measured and the free peptide anisotropy, y is the difference between the anisotropy of the bound peptide and the measured anisotropy and R is the measured anisotropy.

Total fluorescence intensity of the free (Q_f) and bound (Q_b) peptide was calculated by:

$$Q = I_{VV} + 2I_{VH} \qquad \textbf{Equation 3.17}$$

After the anisotropy calculation (Equation 3.16), it is then possible to determine the equilibrium binding dissociation constant (K_d):

$$R_{adj} = R_f + \Delta R \left(\frac{K_d + [A] + [B] - \sqrt{((K_d + [A] + [B])^2 - 4[A][B])}}{2[A]} \right) \qquad \textbf{Equation 3.18}$$

where $[A]$ and $[B]$ are the concentrations of the IDP peptide and partner protein, respectively, R_f is the anisotropy of the free peptide and ΔR is the difference between the free and bound peptide anisotropies.

3.3.17 Data Analysis

Throughout the thesis, biophysics data were fit using Kaleidagraph (version 4.1, Synergy). Quantification of the protein gels bands for the ribosome experiments was done using ImageJ (version 1.52a, NIH).

4 Role of charged residues in the stability of SasG

4.1 Introduction

4.1.1 SasG biological background

Certain bacterial species are able to produce a polymeric matrix that adheres to living or inert surface to form functional communities, known as bacterial biofilms (Costerton et al. 1999). Biofilms are very difficult to eradicate as they are resistant to the host immune system and to antimicrobial agents (Lewis 2001; Dunne 2002; Steward & Costdrton 2001), and present a challenge when treating chronic infections – as usually seen with inserted medical devices (Harris and Richards 2006). Staphylococci are a common source of infections mediated by biofilms, as they populate the human skin and mucous surfaces.

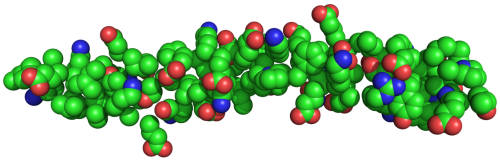
Staphylococcus aureus surface protein G (SasG) promotes cell-to-cell accumulation during biofilm formation (Geoghegan et al. 2010b). Non-pathogenic *S. aureus* normally lives in anterior nares, where the pH is just below neutral, around 6 to 6.5 (Washington et al. 2000). When infection starts and *S. aureus* turns into a pathogen somewhere in the body, the pH is likely to be more acidic. It is known that during immune responses, neutrophils and other immune cells release their content in acidic microenvironments (Cao et al. 2015), where the pH is likely to drop below 5. In fact, there are some suggestions in the literature that release of zinc by neutrophils (and associated drop in pH) might be the trigger for biofilm formation (Conrady, Wilson, and Herr 2013). In this context, it is expected that during biofilm formation, SasG would experience a variety of different conditions in the body, including changes in pH.

4.1.2 SasG biophysical background

SasG amino acid composition is very intriguing: it is comprised mainly of charged residues, prolines and glycines. Thus, it has a sequence composition of an IDP, but yet it folds. Its G5 domains are folded in isolation and in multidomain constructs whereas the E domains only fold when preceding a G5 domain or in multidomain constructs (Dominika T Gruszka et al. 2012).

The second G5 domain ($G5^2$) has 38% of its sequence composed by charged amino acids (Arg, Lys, Asp, Glu) and lacks a hydrophobic core as the β-sheets are exposed to solvent on both sides. The E domain is also highly charged, which makes the E-$G5^2$ construct analogous, with a total of 36% of charged residues (Figure 4.1).

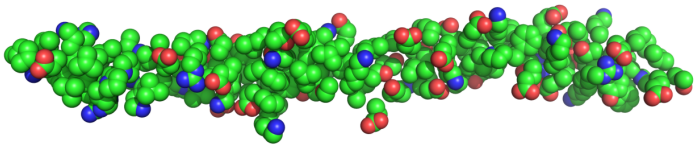
A



GPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITT
PTLKNPLTGEIISKGESKEEITKDPINELTEYGPET

38% charged (31/82): Arg (2), Lys (11), Asp (3), Glu (15)

B



GPEMIAPGHRDEFDPKLPTGEKEEVPGKPGIKNPETGDVVRPPVD
SVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKG
EKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPET

36% charged (48/132): Arg (4), Lys (16), Asp (7), Glu (21)

Figure 4.1: Charged residues in SasG $G5^2$ and E-$G5^2$ constructs. Side chains of (A) $G5^2$ and (B) E-$G5^2$, showing exposed charged amino acids and its sequences with charged residues highlighted and their percentage within the domain amino acid composition.

4.1.3 Kappa value

Highly charged proteins can be extended or collapsed, depending on their patterning. The patterning can be measured using a kappa value (κ), a parameter that can vary from 0-1 (Das and Pappu 2013). κ is low for well-mixed sequences, where the intra-chain interactions are well balanced. Segregated opposite charges generates a high κ (Figure 4.2).

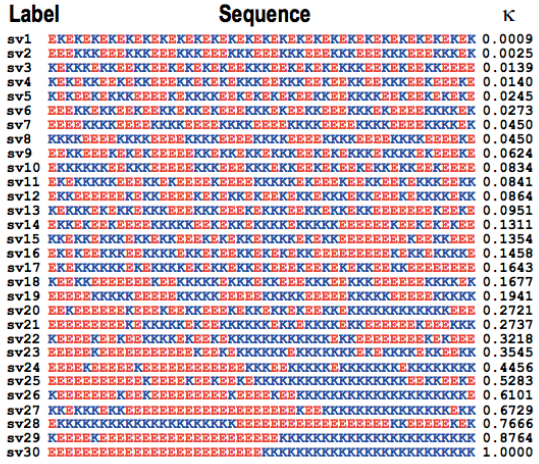


Figure 4.2: Kappa value example of thirty sequence variants composed by Glu-Lys residues. Column 1 shows the label of each sequence variant. Column 2 shows the actual sequence, with Glu residues in red and Lys residues in blue. Column 3 shows the κ-values (Figure taken from Das and Pappu 2013).

The pattern of charged residues in a protein sequence can be related to its ability to fold. Low κ-values are often encountered in random coil ensembles, whereas high κ-values are observed in hairpin-like conformations where long-range electrostatic interactions occur. The κ-value for SasG G5^2 domain was calculated using CIDER (Holehouse et al. 2017), a tool developed by the Pappu laboratory that uses various algorithms to analyse protein sequences.

κ-values for G5^2 domain and E-G5^2 are 0.072 and 0.080, respectively (Figure 4.3). According to the charged residues pattern distribution, both constructs were likely to be expanded. This is in agreement with the PONDR-FIT (Xue et al. 2010) disorder predictor analysis (Introduction, chapter 1), but not in agreement with the structural data. One might infer that the charges are involved in salt bridges formation, as is the case for the highly patterned sequences.

```
Protein Name = SasG.G52

Amino Acid Sequence =
GPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKTITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPET

Length = 82
Molecular Weight (ProtParam) = 9069.02 da
pI (ProtParam) = 4.82
Kappa Value (Assuming all Charged) = 0.072

Amino Acid Content =
{'A': 1, 'C': 0, 'E': 15, 'D': 3, 'G': 8, 'F': 2, 'I': 7, 'H': 0, 'K': 11, 'M': 0, 'L': 4,
'N': 3, 'Q': 1, 'P': 8, 'S': 3, 'R': 2, 'T': 10, 'W': 0, 'V': 3, 'Y': 1}


Protein Name = SasG.E-G52

Amino Acid Sequence =
GPEMIAPGHRDEFDPKLPTGEKEçEVPGKPGIKNPETGDVVRPPVDSVTKYGPVKGDSIVEKEEIPFEKERKFNPDLAPGTEKVTREGQKGEKT
ITTPTLKNPLTGEIISKGESKEEITKDPINELTEYGPET

Length = 132
Molecular Weight (ProtParam) = 14463.01 da
pI (ProtParam) = 4.84
Kappa Value (Assuming all Charged) = 0.080

Amino Acid Content =
{'A': 2, 'C': 0, 'E': 21, 'D': 7, 'G': 14, 'F': 3, 'I': 9, 'H': 1, 'K': 16, 'M': 1, 'L': 5,
'N': 4, 'Q': 1, 'P': 17, 'S': 4, 'R': 4, 'T': 13, 'W': 0, 'V': 8, 'Y': 2}
```

Figure 4.3: Kappa values results for SasG G5^{2}and E-G5^2 constructs. Values for both G5^2 and E-G5^2 are low, 0.072 and 0.080 respectively.

It has been shown that proline and glycine residues are very important for SasG stability at physiological pH (Gruszka et al. 2016). Additionally, it is known that the cooperativity between SasG domains could be important for sensing mechanical stress (Gruszka et al. 2015), allowing the protein to fold and unfold in response to environmental conditions.

Considering that the protein also naturally experiences environments with different pH and given the unusually charged nature of the protein, we were interested to investigate whether protein stability was also sensitive to this environmental input.

4.1.4 Chapter aims

In this chapter, bioinformatics analysis and biophysical experiments were performed in order to investigate the role of charged amino acids in the stability of SasG. It will include results from the peptide charge calculator tool, kappa-value analysis, salt bridges predictor, ionic strength dependence (for both G5^2 and E-G5^2 constructs) and pH dependence studies of G5^2. The data provide insights on the importance of the charged residues for SasG stability.

4.2 Results

4.2.1 Peptide charge calculator

The "Peptide charge calculator" (Gale Rhodes, University of Southern Maine), is a script that calculates the sum of each charged species of a protein sequence using the Henderson-Hasselbach equation at each pH. Using this tool, it was possible to generate a plot with the profile of the net charge against increasing pH for SasG $G5^2$ and E-$G5^2$ constructs. (Figure 4.4).

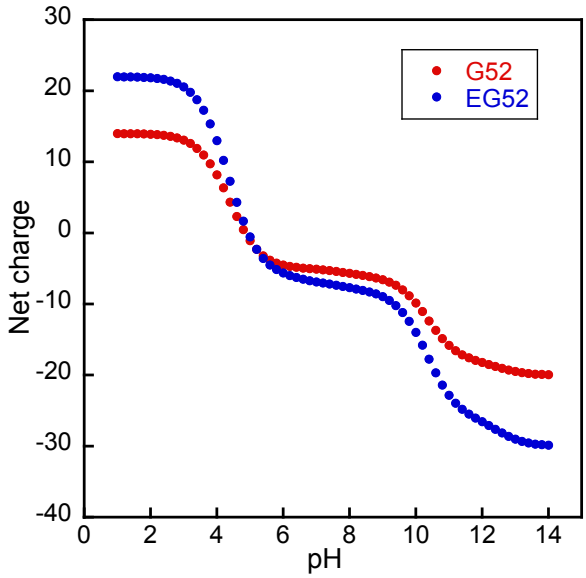


Figure 4.4: Peptide charge calculator results for $G5^2$ and E-$G5^2$constructs. The script computes net charge on peptide from composition of ionisable residues.

The plot shows the expected net charge across the pH range for $G5^2$ and E-$G5^2$. Any residues involved in salt bridges may experience altered pKa values (Anderson, Becktel, and Dahlquist 1990), stabilising the ionised state. Protein stability is altered by changes in total net charge (Shaw et al. 2001). Therefore, based on this net charge calculator, we might expect the protein stability to remain relatively unaltered in the 6-9 region, with changes in stability occurring in the range 3-5 and 10-12. On the other hand, if the residues are involved in salt bridge formation, we might expect the protein stability to remain unchanged over a wider range of pH.

4.2.2 SasG salt bridges

The ESBRI tool (Evaluating the salt bridges in proteins online tool; Costantini et al. 2008), was used in order to check if charged residues were involved in salt bridges along SasG structure. A cut off of 4 Å was used as a maximum distance between two opposite charged residues (Kumar and Nussinov 2002). The results show that despite having 31 charged amino acids, $G5^2$ has only six salt bridges (Figure 4.5, A). E-$G5^2$ has 48 charged amino acids but only 8 salt bridges (Figure 4.5, B). Perhaps this should not be a surprise since all residues are solvent exposed and able to interact with water.

A

G52-WT

Residue 1	Residue 2	Distance
LYS 565	GLU 621	3.84
ARG 567	GLU 578	3.09
LYS 586	GLU 560	3.23
LYS 589	GLU 612	3.46
LYS 597	ASP 553	3.90
LYS 611	GLU 613	3.96

B

EG52-WT

Residue 1	Residue 2	Distance
HIS 506	ASP 621	2.92
HIS 506	ASP 578	3.27
LYS 565	GLU 560	3.84
ARG 567	GLU 612	3.09
LYS 586	GLU 553	3.23
LYS 589	GLU 613	3.46
LYS 597	ASP 613	3.90
LYS 611	GLU 613	3.96

Figure 4.5: Salt bridges results of SasG contructs. Salt bridges predictions for (A) $G5^2$ and (B) E-$G5^2$, using ESBRI tool (Costantini et al. 2008) assuming a 4 Å cut-off.

4.2.3 Ionic strength effect on SasG contructs: $G5^2$ and E-$G5^2$

Both constructs $G5^2$-WT and E-$G5^2$-WT were successfully expressed in *E. coli* and purified by nickel affinity chromatography. A yield of approximately 60 mg of protein per litre of cell culture was obtained. The purity of each protein was verified by SDS-PAGE and confirmed by mass spectrometry.

Neither $G5^2$ or E domains have a tryptophan residue in their sequence composition, so all equilibrium studies were performed following tyrosine fluorescence signal. There is one tyrosine located in the C-terminal end of each domain; one in E and one in G5 (figure 4.6).

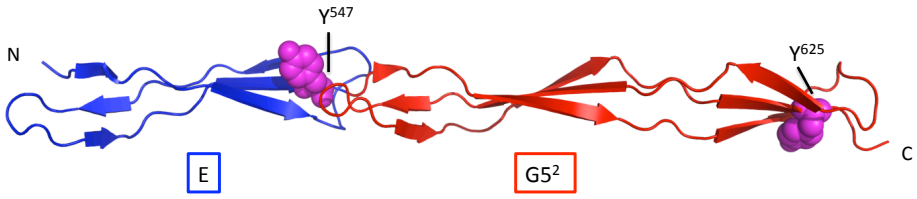
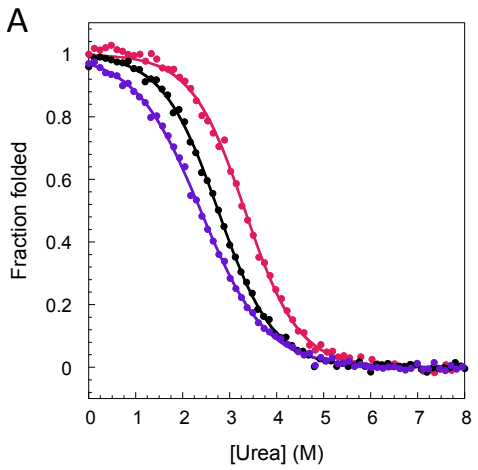
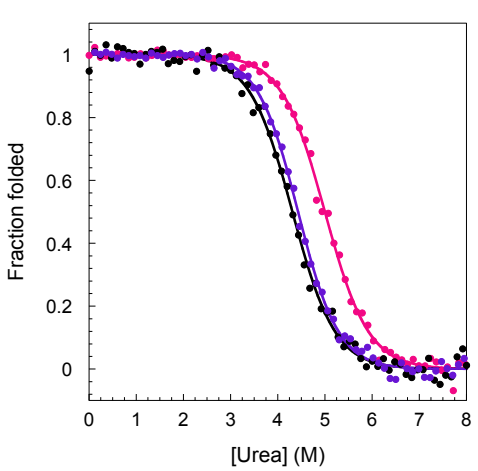


Figure 4.6: Tyrosine residues in E-$G5^2$ construct.

Previous equilibrium and kinetics studies showed that the observed fluorescence signal in SasG E$G5^2$ construct comes mainly from the tyrosine located in the $G5^2$ domain (Y^{625}) (Gruszka et al. 2016).

In order to do a screening of the charge-charge interactions of SasG and gain some insights on the protein stability, we investigated the effect of the ionic strength. Equilibrium studies were performed in both constructs $G5^2$-WT and E-$G5^2$-WT. Urea-induced equilibrium unfolding curves (pH 7.4) were recorded for each construct (one in MOPS, 5 mM ionic strength and the other in MOPS, 1 M ionic strength) and then compared with the curves recorded in PBS (164 mM ionic strength) in each case (Figure 4.7).





Ionic strength (IS)	*m*-value (kcal.mol^{-1}.M^{-1})	[den]$_{50\%}$ (M)	ΔG (kcal.mol^{-1})
5 mM	1.03 ± 0.01	3.33 ± 0.01	3.43 ± 0.03
164 mM	0.99 ± 0.01	2.76 ± 0.01	2.73 ± 0.03
1 M	0.85 ± 0.01	2.38 ± 0.01	2.02 ± 0.02

Ionic strength (IS)	*m*-value (kcal.mol^{-1}.M^{-1})	[den]$_{50\%}$ (M)	ΔG (kcal.mol^{-1})
5 mM	1.29 ± 0.03	5.00 ± 0.04	6.45 ± 0.16
164 mM	1.31 ± 0.04	4.31 ± 0.01	5.65 ± 0.17
1 M	1.38 ± 0.02	4.43 ± 0.02	6.11 ± 0.09

Figure 4.7: Effect of ionic strength on the stability of SasG. Urea-induced equilibrium denaturation studies at pH 7.4 and varying the ionic strength. (A) G5^2-WT equilibrium curves and its parameters in different buffers: PBS with 164 mM (black), MOPS with 5 mM IS (pink) and MOPS with 1 M IS (purple). (B) E-G5^2-WT equilibrium curves and its parameters in different buffers: PBS with 164 mM (black), MOPS with 5 mM IS (pink) and MOPS with 1 M IS (purple).

In the cases of G5^2-WT the stability decreased slightly whereas for E-G5^2-WT the stability remained the same with an increasing concentration of salt. The data was considered the first sign suggesting that charge-charge interactions may not be key for the stability of both G5^2-WT and E-G5^2-WT in PBS.

4.2.4 pH dependence of the SasG stability

When a change in a folded protein environment is made, the charged amino acids, especially arginine, lysine, aspartic acid and glutamic acid may have an altered pKa. The pKa is the pH value at which 50% of the molecules will be protonated. As SasG is highly charged, the initial question was to investigate the role of the charged amino acids on the stability of G5^2 domain (Table 4.1).

Table 4.1: Charged amino acid composition of the $G5^2$ domain.

Amino acid	Number	pKa
Lysine (K)	11	10.4
Glutamic acid (E)	15	4.5
Aspartic acid (D)	3	4.0
Arginine (R)	2	12.5
Tyrosine (Y)	1	9.6

In addition to the ionic strength studies a series of urea-induced equilibrium experiments were performed on the $G5^2$ construct, with a range of buffers at different pH values (Table 4.2), in order to investigate possible chemical shifts on the pKa values of the charged amino acids. This study would help to provide insights into the nature of the folded $G5^2$ domain, despite the lack of a hydrophobic core and the high percentage of charged residues.

To discount the effect of salts on protein stability, all buffers were prepared with the lowest possible ionic strength, which was kept the same for the preparation of all of them (8 mM ionic strength).

Table 4.2 Buffers used for the pH dependence studies of SasG $G5^2$-WT domain.

pH	Buffer	pKa
2.0	Phosphate	2.15
3.0	Phosphate	2.15
4.0	Acetate	4.76
4.5	Acetate	4.76
5.0	Acetate	4.76
6.0	MES	6.21
7.0	MOPS	7.31
8.0	Tris	8.06
9.0	Tris	8.06
10.0	CAPS	10.51
11	CAPS	10.51

In the urea-induced equilibrium denaturation studies, the pH transpired to have an effect on stability in the range between pH 2 and pH 5 and then in the range between pH 10 and pH 11. Moreover, no effect on stability was verified for pH values between 6 and 9, as previously suggested by peptide charge calculator (Figure 4.8, A and fluorescence data on Appendix 1).

This effect is very clear when we plot the variation in stability (ΔG_{D-N}) against pH value where we can see a significant change in stability from pH 2 to pH 5 and pH 10 to pH 11 whereas in the pH range between 6 and 9, the stability is relatively constant (Figure 4.8, B).

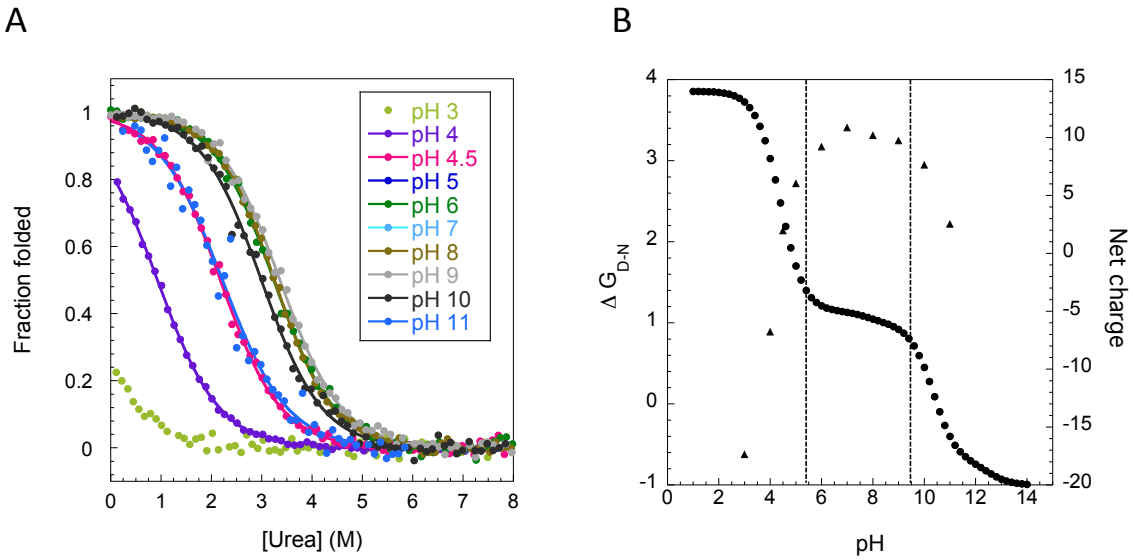


Figure 4.8: Equilibrium studies and effect of pH on the stability of the G5² domain. (A) Equilibrium denaturation curves of G5²-WT in different pH buffers (pH ranging from 3 to 11). Due to a very strong destabilising effect, we were not able to calculate fraction full-length at pH 2 (see fluorescence data on Appendix 1). (B) Variation of stability (ΔG_{D-N}) plotted against pH values (triangles). The peptide charge calculator results for G5² were also plotted for comparison (data previously shown in figure 4.4).

Due to the topology and structural similarity between G5² and E-G5² and because they both show a similar prediction using the peptide charge calculator, we decided not to perform the same pH dependence experiments for E-G5².

4.3 Conclusion

The pH dependence studies of the $G5^2$ domain are in agreement with results observed from ionic strength studies, as it does not show a dramatic effect on the domain stability over the range of pH values. Another interesting tool, was the salt bridges predictor (ESBRI-evaluating the salt bridges in proteins online tool; Costantini et al. 2008), which shows that $G5^2$ has six salt bridges and $EG5^2$ has 8. These do not appear to have either a significant pKa shift and/or are not particularly stabilising to the fold of the protein. All the remaining charged residues (19 in $G5^2$ and 32 in E-$G5^2$) are apart from each other with distances greater than 4 Å.

The acquired data when compared with the prediction from the net charge calculator, suggest that no significant effect on the pKa values of the amino acids is observed. From the results, it is clear that the unusual charge composition of SasG does not significantly contribute to the protein capacity to stay folded at different pH values during biofilm formation in the cell.

Kappa-values of both $G5^2$ and E-$G5^2$ are low, which would be an indicative of a random coil. The result is in agreement with the PONDR-FIT (Xue et al. 2010) disorder predictor analysis, that suggest that both constructs should be disordered. Yet, the protein is folded in solution. The bioinformatics analysis and biophysical results of this chapter are in agreement, showing that SasG stability does not come greatly from the presence or the pattern distribution of its charged residues. Clearly this protein goes against current predictive methods of protein stability. Since this protein has few hydrophobic and aromatic residues, the reason for the stability of SasG is unclear. However, our previous studies have demonstrated that the interface between the two domains can contribute to the formation of the E-$G5^2$ construct, even when both domains were destabilised and are unfolded in isolation (Gruszka et al. 2016).

5 Co-translational Folding of SasG

I would like to thank Dr. Jeffrey Hollins for building the constructs containing the different liker lengths and Dr. Lee Kwa for the assistance with the molecular biology of the charged constructs. I am also very grateful to Annette Steward for doing some repeats of the force profile experiments.

5.1 Introduction

In vivo, proteins fold and function under a variety of conditions, including temperature, pH and mechanical stress, all of which can modulate the folding energy landscapes (Sosnick & Barrick 2011, Guinn *et al*. 2015, Baldwin 2006). The ribosome, for instance, can influence protein folding: since folding is often faster than the rate of translation, molecules can attempt to fold while being translated (Cabrita, Dobson, and Christodoulou 2010). Although the idea that the ribosome can affect the folding process is well accepted, the mechanism by which it occurs is not fully understood (Kaiser et al. 2011).

The extent to which studies *in vitro* compare with the folding of proteins *in vivo* is also not entirely comprehended. In contrast with *in vitro* conditions, aspects such as i) the unidirectional vectorial orientation that the peptide chain arises from the ribosome; ii) the conformational possibilities allowed to the emerging sequence before the entire protein is translated; and iii) the possible interactions between the nascent chain and the highly charged surface of the ribosome are examples of new possibilities of conditions which might influence folding *in vivo* (Knight et al. 2013; Nilsson et al. 2015; Tian et al. 2018).

Our previous studies regarding the folding of SasG showed that both $G5^2$ and E-$G5^2$ fold via the same, highly polarised transition state, with formation of the C-terminal region of $G5^2$ domain being the limiting step for the folding of the entire structure. Once the main pathway is destabilised, it was also possible to observe and characterise an alternative folding pathway that starts via the interface between the two domains (Gruszka et al. 2016). After this extensive study to characterise SasG folding pathway *in vitro*, we were interested in using SasG as model to investigate co-translational folding and get some insights on how comparable the *in vitro* and *in vivo* results are.

5.1.1 The ribosome

The ribosome is a macromolecular machine – present in all living cells – that decodes genetic information and converts it into amino acid sequence. It contains three large ribosomal RNA (rRNA) molecules, which represent two-thirds of its mass, and over fifty proteins (Kurland 1960). One of the first attempts to reveal the ribosome structure dates a bit more than 40 years, when a model of an *E. coli* ribosome was obtained by negative electron microscopy (Lake 1976). Although rudimentary compared to the most recent data, the study was able to show the two ribosomal subunits, referred to as small and large. Later on, with the development of new techniques, much more detailed structures of the ribosome allowed us to gain more knowledge on its architecture and function. Cryogenic electron microscopy (cryo-EM) for instance, was recently able to reveal the atomic structure of a mammalian ribosome-Sec61 complex, determined to 3.4 Å resolution (Voorhees et al. 2014).

Ribosome small and large subunits are named 30S and 50S in bacteria and 40S and 60S in eukaryotes, respectively. The small subunit is involved in decoding the messenger RNA (mRNA) information whereas the large one contains the peptidyl transferase centre (PTC), where the formation of peptide bonds occurs. The ribosome tunnel connects the PTC to the site where the nascent chain emerges (Frank et al. 1995). During the translation step, the nascent polypeptide chain travels through the ribosome tunnel in order to emerge into the cell cytoplasm (Milligan and Unwin 1986). The average distance from the PTC to the end of the tunnel is approximately 100 Å and the diameter of the tunnel is 15 Å. At 80 Å from the PTC, the tunnel widens significantly to form the vestibule (Figure 5.1) (Nissen et al. 2000).

Early results demonstrated that the ribosome tunnel fits approximately 30-40 amino acids of a typical nascent chain (Malkin and Rich 1967), suggesting that translated polypeptides transit the ribosome tunnel in an extended conformation of approximately 3.5 Å per residue. For consistency to the other studies in our lab, we are going to assume that it takes about 35 residues of an extended nascent chain to span the approximate 100 Å from the PTC to exit the ribosome tunnel (Tian et al. 2018).

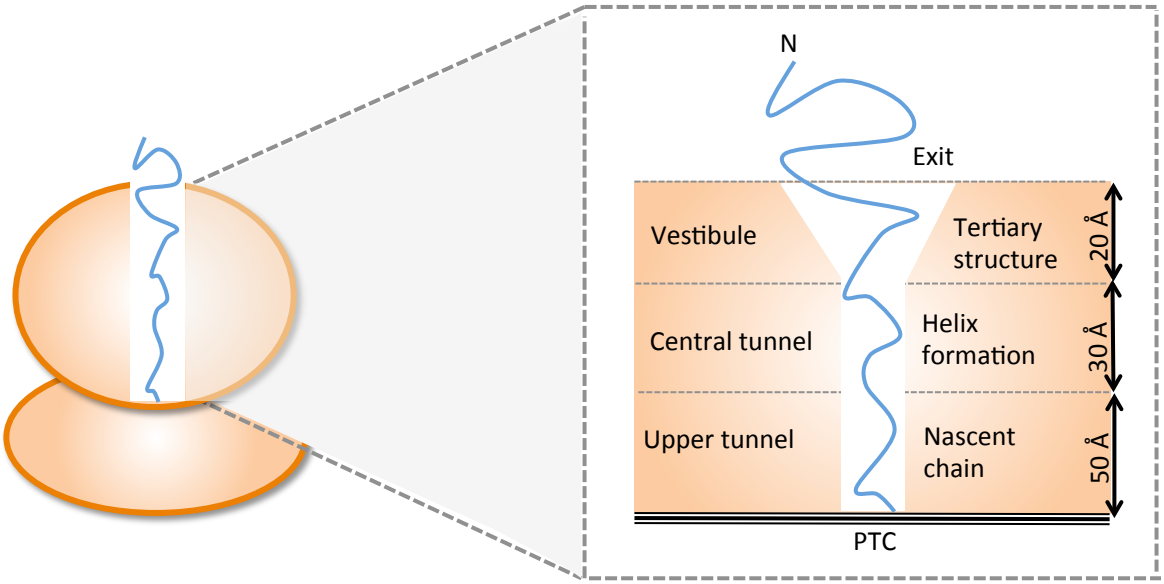


Figure 5.1: A schematic representation of the ribosome tunnel cross-section. The ribosome tunnel connects the peptidyl transferase centre (PTC) to the exit site, where the polypeptide chain emerges. The tunnel total length is approximately 100 Å and can protect approximately 30-40 of a nascent chain. Studies showed helices being formed in the lower tunnel and formation of tertiary structure was encountered in the vestibule region

5.1.2 Co-translational folding

The relationship between the protein folding *in vitro* and co-translational folding has gained a lot of relevance in recent years and a variety of methods have been employed for that purpose (Thommen et al. 2017, Komar 2018). Elements of secondary and tertiary structures (Kudva et al. 2018), as well as small peptides, have been found to be formed in the ribosome tunnel (Tu & Deutsch 2017, Su et al. 2017, Mercier & Rodnina 2018; Figure 5.1). Moreover, entire domains like spectrin R16 and the Ig I27 have been shown to be able to fold while still in the ribosome vestibule, before exiting the tunnel (Nilsson et al. 2017; Tian et al. 2018).

In comparison to domains folding in isolation, the presence of the ribosome has been reported to be able to affect the folding process. For instance, in the co-translational folding of the Ig-like domain FLN5, the ribosome was found to modulate the folding process, as the domain could only fold well beyond the tunnel whereas it can fold spontaneously when in isolation (Cabrita et al. 2016). For T4 lysozyme, the ribosome was shown to slow down the

folding to the native structure compared to *in vitro* results. Moreover, incomplete T4 lysozyme polypeptides are able to fold near the ribosome surface whereas misfolding and aggregation is observed when they are free in solution (Kaiser et al. 2011).

Quantitative pulse-proteolysis experiments were able to identify a destabilising effect of the ribosome on the nascent polypeptide chain (Samelson et al. 2016) and electrostatics was also shown to play a role in the nascent chain. Due to the highly charged ribosome surface, the net charge of the polypeptide chain can define its mobility as it arises from the ribosome tunnel: negatively charged chains are more dynamic whereas positively charged ones are more static, as the latter can interact with the ribosome surface (Knight et al. 2013).

Although some proteins are shown to fold via the same pathway on and off the ribosome (Tian et al. 2018), others are reported to differ (Evans et al. 2005, Nilsson et al. 2017). Understanding the co-translational folding has recently demonstrated that the ribosome can also play a role providing a further level of protein regulation (Gloge et al. 2014, Kirchner et al. 2017, Chaney et al. 2017).

5.1.3 The arrest peptide assay

The experimental procedures used in my project are based in the set up used in Gunnar von Heijne's lab (Nilsson et al. 2015). The gene encoding the DNA sequence of the protein of interest is cloned into a series of different constructs containing a SecM stall sequence and a linker with variable length. Cryo-EM results were able to unravel the SecM stalling mechanisms upon translation (Zhang et al. 2015). The arrest sequence of SecM can interact extensively with the ribosome tunnel and can alter the conformations of the ribosome A-site (where the tRNA first binds) and P-site (where the amino acid is incorporated). The change in conformation can cause two major consequences for translation: the tRNA binding in the A-site is compromised and its passage to the P-site is slowed down. Together, both effects contribute to the SecM stalling ability (Zhang et al. 2015). The interaction between the stall sequence and the ribosome can be released when a tension is applied. Therefore, the effectiveness of the stall is proportional to the force applied by the folding of the nascent polypeptide (Goldman *et al*. 2015) (Figure 5.2).

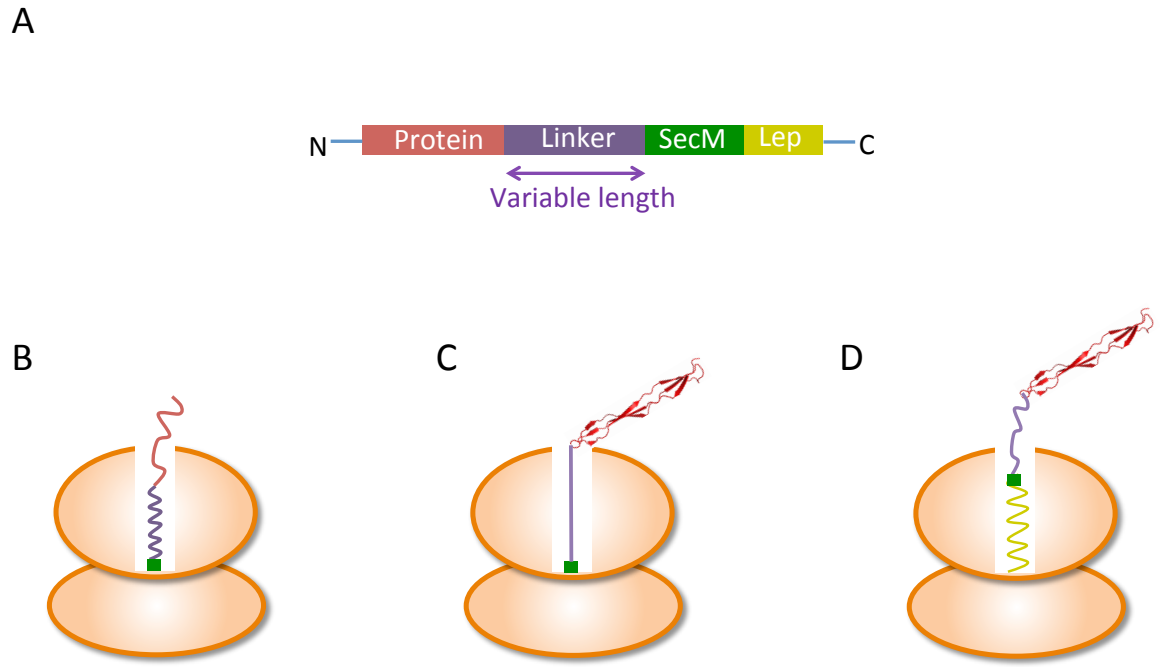


Figure 5.2: Schematic of the arrest peptide assay based on force measurements. (A) Schematic representation of SasG constructs. (B) Translated sequence containing the G5² domain (red), followed by a variable linker (purple) and the stall sequence (green). (C) Once a tension is applied by the folding of the G5² domain, (D) the SecM is released and the following sequence is translated (Lep domain, yellow).

The efficiency of the arrest at each linker length can act as an indirect measurement of the tension applied in the stall sequence caused by the folding of the translated nascent chain. If the linker is too long the polypeptide chain will fold without applying a tension and will not release the stall. If the linker is too short, the nascent chain will not be able to fold due to a spatial constraint inside the ribosome tunnel. The length of the linker and the tension applied by the folding of the polypeptide chain needs to be optimal in order to release the stall sequence. The stall sequence used in this work is the relatively weak *E. coli* SecM arrest peptide, also used in previous co-translational force profile assays (Nilsson *et al.* 2015, Nilsson et al. 2017; Tian et al. 2018; Marsden et al. 2018). The proline residue located at the end of the SecM sequence is the key residue for stalling and mutating it to an alanine completely disrupts the SecM translational stalling capacity (Yap and Bernstein 2009).

Stall sequence: F S T P V W I S Q A Q G I R A G P

5.1.4 SasG as a model to study on the ribosome

As previously described in the Introduction and in the previous chapter (chapters 1 and 4), SasG is an unusual protein, composed of tandem repeats of two domains, E and G5. Both domains are predicted to be disordered as their sequences are composed mainly of prolines, glycines and charged residues. Although the E domain is unfolded on its own, it folds when preceding a G5 domain. Both G5 in isolation and E-G5 can naturally fold to an extended three-dimensional β-sheet structure, exposed to the solvent on both sides as it lacks a hydrophobic core (Dominika T Gruszka et al. 2012).

This intriguing structure folds via a very polarised transition state, with the formation of the far C-terminal end of the G5 domain being the limiting-step for folding the entire structure, in both G5 and E-G5 cases. The addition of the disordered E domain contributes to an increase in stability for E-G5 (6.3 $kcal.mol^{-1}$) when compared to G5 in isolation (2.8 $kcal.mol^{-1}$). Furthermore, with the presence of E, an alternative pathway for folding is identified for E-G5 when the C-terminus of G5 is destabilised. In this context, folding of E-G5 starts via the interface between the two domains. The interdomain interface was also shown to be crucial to maintain the long-range cooperativity and the cooperative folding of SasG (Gruszka et al. 2016).

After characterising SasG folding pathways *in vitro* and the importance of the interface between E and G5 domains, we decided to use SasG as a model to study folding on the ribosome. Its unusual sequence composition, long-range interactions, extended structure and alternative folding pathway makes SasG a very interesting candidate to study co-translational folding.

5.2 Chapter aims

After an extensive study to characterise and understand how SasG is able to fold and form its structure *in vitro*, we wanted to expand this knowledge and investigate its co-translational folding. With the previous information regarding the role of SasG domains interface and the presence of an alternative pathway for folding, we performed co-translational folding experiments to determine the effect that the additional E domain (and the formation of the inter-domain interface) has on the folding of the $G5^2$ domain on the ribosome.

This project aimed to help us to shed some light into how *in vitro* and *in vivo* studies can be compared and to understand to what extent the presence of the ribosome and the vectoral emergence of a nascent chain can influence the co-translational folding of a protein.

5.3 Results

5.3.1 Molecular biology strategy

The project started with a series of molecular biology steps in order to insert SasG DNA sequence in the different constructs with all linker lengths (L = 21 to 61 residues). SasG DNA construct, originally in the pSKB2 vector, was subcloned into the pRSETa vector for consistency with all the work done in the lab. The restriction sites before and after the gene of interest were mutated to *Bam*HI and *Eco*RI sites, respectively, by site directed mutageneis (SDM). The gene was extracted by double digestion and was ligated into the pRSETa containing the longest linker length (L61). Following up, another step of SDM was performed to remove the *Eco*RI restriction site and the stop codon in between the protein gene and the linker.

In order to create all the other twenty constructs with different linker lengths, a series of SDM reactions were performed using the gene containing the L61 linker as template DNA. A series of primers were designed by Dr. Jeff Hollins based on the idea that linkers would be truncated from its N-terminal end, next to the protein sequence. Previous work had shown that truncation from the C-terminal end of the linker can alter the strength of the stall (Ismail, N., *et al*., 2012). For consistency and to keep the immediate sequence after the protein constant, all linker sequences started with Ser-Gly-Ser-Gly. Following the SGSG, the middle of the linker sequences are composed by a fraction of the Lep domain known to be disordered and of a variable length, followed by the SecM stall sequence. An extra fraction of the Lep domain comes after the stall sequence in order to differentiate by length the translated protein that was stalled from the one released from the ribosome (Figure 5.3).


```
L21 SGSG                                          FSTPVW
L23 SGSG                                       SS FSTPVW
L25 SGSG                                    WMSS FSTPVW
L27 SGSG                                 GDWMSS FSTPVW
L29 SGSG                              MMGDWMSS FSTPVW
L31 SGSG                           YFMMGDWMSS FSTPVW
L33 SGSG                        GQYFMMGDWMSS FSTPVW
L35 SGSG                     PPGQYFMMGDWMSS FSTPVW
L37 SGSG                  IVPPGQYFMMGDWMSS FSTPVW
L39 SGSG               TWIVPPGQYFMMGDWMSS FSTPVW
L41 SGSG            NATWIVPPGQYFMMGDWMSS FSTPVW
L43 SGSG         QQNATWIVPPGQYFMMGDWMSS FSTPVW
L45 SGSG      PGQQNATWIVPPGQYFMMGDWMSS FSTPVW
L47 SGSG   LVPGQQNATWIVPPGQYFMMGDWMSS FSTPVW
L49 SGSG KTLVPGQQNATWIVPPGQYFMMGDWMSS FSTPVW
L51 SGSG YQKTLVPGQQNATWIVPPGQYFMMGDWMSS FSTPVW
L53 SGSG PIYQKTLVPGQQNATWIVPPGQYFMMGDWMSS FSTPVW
L55 SGSG KDPIYQKTLVPGQQNATWIVPPGQYFMMGDWMSS FSTPVW
L57 SGSG GIKDPIYQKTLVPGQQNATWIVPPGQYFMMGDWMSS FSTPVW
L59 SGSG AYGIKDPIYQKTLVPGQQNATWIVPPGQYFMMGDWMSS FSTPVW
L61 SGSG KFAYGIKDPIYQKTLVPGQQNATWIVPPGQYFMMGDWMSS FSTPVW
```

Figure 5.3: Protein sequences of all primers designed to create the different linker length constructs for the force profile assays. Using the gene containing the protein, the longest linker length (L61) and the SecM stall sequence, a series of SDM were applied to truncate the linker and generate all constructs with variable liker lengths. Figure shows that the linker was truncated from its N-terminus, just after the protein sequence. The primer sequences start from SGSG (green), followed by a sequence extracted from the Lep domain that is known to be disordered (black), and by the SecM stall sequence (first six residues showed in red).

In order to create controls, the key proline residue responsible for the stall within the SecM sequence was mutated to an alanine in order to generate a non-stop control. In addition, the proline was also mutated to a stop codon (TAA) to create a control that, once translated, would generate a shorter sequence as translation would stop at the end of the linker and not at the subsequent Lep domain (Figure 5.4). At each linker length, these two controls would allow us to associate the bands on the gel to its correct protein: either the stalled or the full-length translated polypeptide nascent chain.

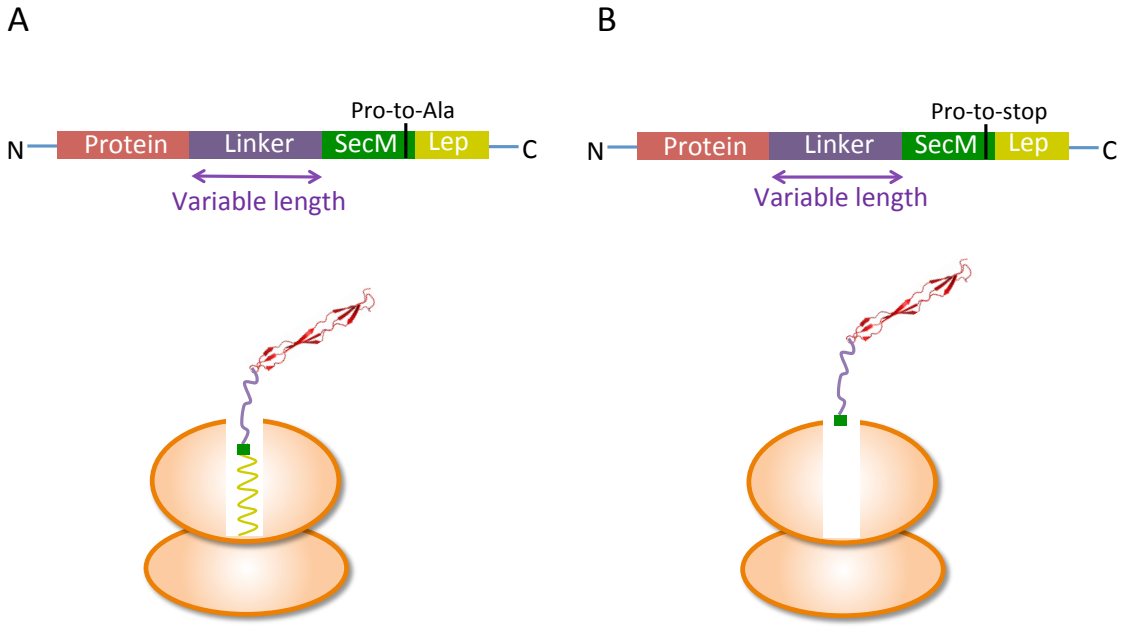


Figure 5.4: Controls involved in the folding on the ribosome experiments. (A) Proline to alanine mutation in the stall sequence prevents the stalling on the ribosome and the subsequent Lep domain sequence is translated. (B) Proline to stop-codon mutation in the stall sequence prevents the stalling on the ribosome and stops translation after the linker.

5.3.2 Folding on the ribosome of the $G5^2$ domain

Once the constructs containing the different linker lengths were created, folding on the ribosome studies were performed in order to generate the force profile of the $G5^2$ domain. Pro-Ala and one Pro-stop codon controls were also generated for some linker lengths. The experiments were performed using the *in vitro* translation kit (see Materials and Methods, chapter 3) and the force profile results for $G5^2$ are plotted in figure 5.5.

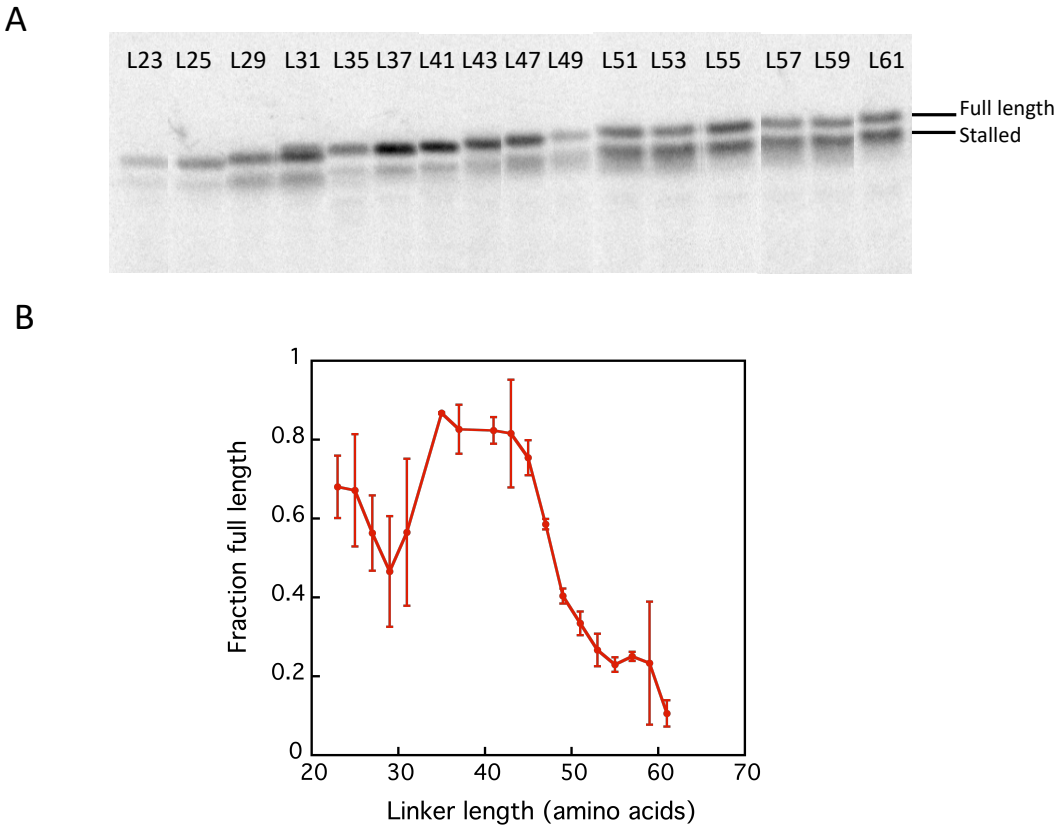


Figure 5.5: Cotranslational force profile for the G5[2] domain. (A) Example of a SDS-PAGE gel that was used to quantify the amount of stalled relative to full-length translated protein at different linker lengths. (B) Force profile of SasG G5[2] domain. Fraction full-length protein is plotted against the number of residues in the linker (linker length). The results are the average of two experiments for each linker length and the error bars represent the standard error.

From the force profile experiments for the G5[2] domain we can see that the highest force is applied at the linker length L= 35 which is the position at which the protein is emerging from the vestibule. Apparently SasG starts to fold at linker length L31, in other words, before it is fully emerged from the tunnel. Interestingly, we can also observe a high force for the linker lengths L23 and L25, when the polypeptide chain is still inside the ribosome tunnel. At longer linker lengths, no significant force was observed, as the folding does not generate enough tension to release the stall. These results are the average of at least three experiments for each linker length. Pro-to-Ala and Pro-to-stop controls were also quantified by the SDS-PAGE gels, and some examples are shown in Figure 5.6.

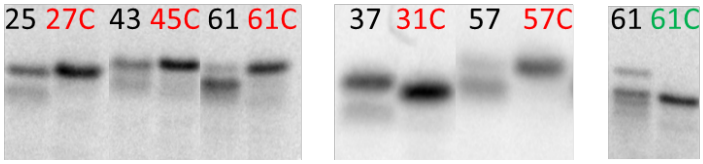


Figure 5.6: SDS-PAGE gels used to quantify the G5² domain bands at different linker lengths and its controls. The gel shows some examples of G5² bands at different linker lengths (black) and some Pro-to-Ala (red) and Pro-to-stop (green) controls.

The controls allowed the association of each band on the gel to its respective result: Pro-Ala controls correspond to full-length protein, whereas Pro-stop correspond to the stalled protein. We chose not to do a control for every linker length as the gels were very clear, with just two bands appearing for each construct.

From biophysical studies during my MPhil, I was able to characterise one Gly-to-Ala mutation, G587A, in G5² that was completely disruptive to the domain structure (see Introduction, Chapter 1). For the folding on the ribosome studies, I used G5²-G587A as a non-folding control. The mutation was introduced to different linker length constructs via SDM. Force profile results are shown in figure 5.7.

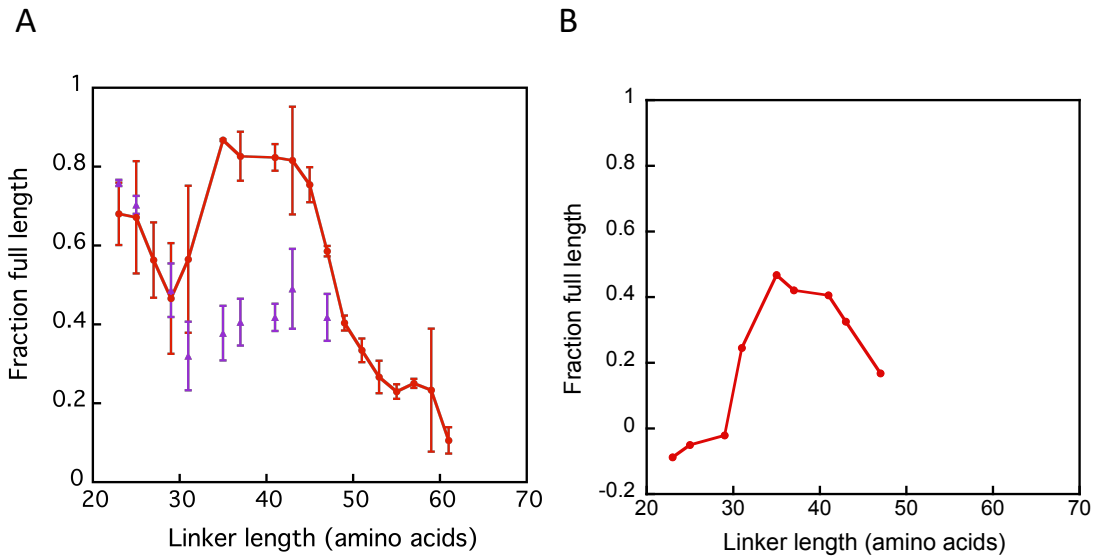


Figure 5.7: Force profile results for the G5² construct and its non-folding control G5²-G587A. (A) Force profiles of G5² (red) and its non-folding control G5-G587A (purple triangles). Although the non-folding control was expected to not exert any force, the nascent chain was able to release the stall at different linker lengths. (B) Normalised data for G5² taking into account the contribution of extra effects apart from folding that promote the stall release for G5²-G587A. The results are the average of at least three experiments for each linker length. The results are the average of at least two experiments for each linker length and the error bars represent the standard error.

Although the non-folding control differs from the WT results, their fraction full-length is significantly higher from what we expected. At lower linker lengths (L23 and L25), as well as we see for $G5^2$-WT, a great proportion of the translated molecules can release the stall. This means that the force applied while the protein is still inside the ribosome tunnel, does not come from a folding event. In the peak region of the force profile of $G5^2$-WT, we see a much lower fraction full-length for $G5^2$-G587A. This suggests that between linkers L31-L47 is where the folding of $G5^2$ happens during translation. This is represented by normalised data (Figure 5.7, B) that takes into account the contribution on fraction full-length that comes from the non-folding control.

The results show that the $G5^2$ domain can fold while still in the ribosome vestibule. Additionally, it also suggests that the domain is able to apply force and release the stall from additional sources that are not folding. High fraction full-length readings at lower linker lengths and for the non-folding control could possibly be explained by the formation of non-native interactions or structures being formed inside the ribosome tunnel or even that the translated polypeptide chain could be interacting with the ribosome.

5.3.3 Folding on the ribosome of E-G5^2

In order to gain more insights, E-G5^2 construct was also studied for comparison. The additional folding pathway characterised *in vitro*, where folding starts with the formation of the interface between the two domains, could possibly add some new possibilities to fold co-translationally. In the main folding pathway, formation of the far end of the G5^2 domain is the limiting step that drives the folding of the entire structure. Because of the vectoral emergence of the nascent chain on the ribosome, the E-G5^2 interface would be available earlier than the C-terminus of the G5^2 domain. We would predict that E-G5^2 could start to form its structure earlier than G5^2.

Results of E-G5^2 on the ribosome showed otherwise. As well as for G5^2, the stall sequence is released at lower linker lengths (<L30) and the main peak of the E-G5^2 force profile overlays with the one from G5^2 (Figure 5.8), resulting in an almost identical force profiles for both constructs.

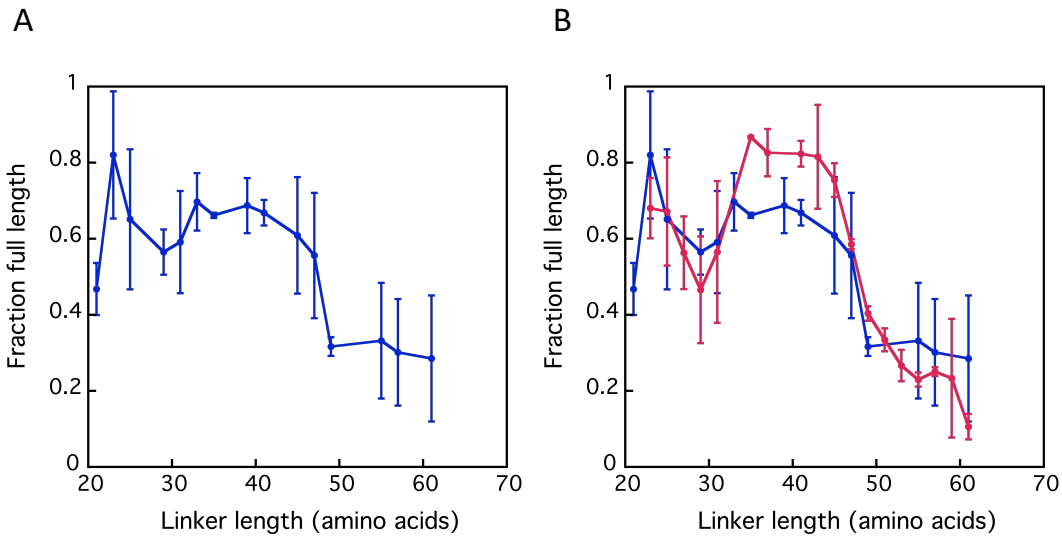


Figure 5.8: Co-translational folding force profile for E-G5^2. (A) Force profile of SasG E-G5^2 construct. (B) Comparative results of G5^2 and E-G5^2 force profiles. Force profiles of G5^2 (red) and E-G5^2 (blue). Fraction full-length protein is plotted against the linker length. The results are the average of at least two experiments for each linker length and the error bars represent the standard error.

In order to get a non-folding control for E-G5^2, we identified a mutation that disrupts the construct stability and prevent its formation. It consists of three Ala-Gly mutations: G587A, G602A and G626A. Those three mutations were previously shown to be the most destabilising ones for E-G5^2 (Gruszka et al. 2016). The mutations were introduced to the E-G5^2 gene, one at the time, by SDM. The protein containing the triple mutant was expressed and purified and equilibrium studies confirmed its unfolded state (Figure 5.9).

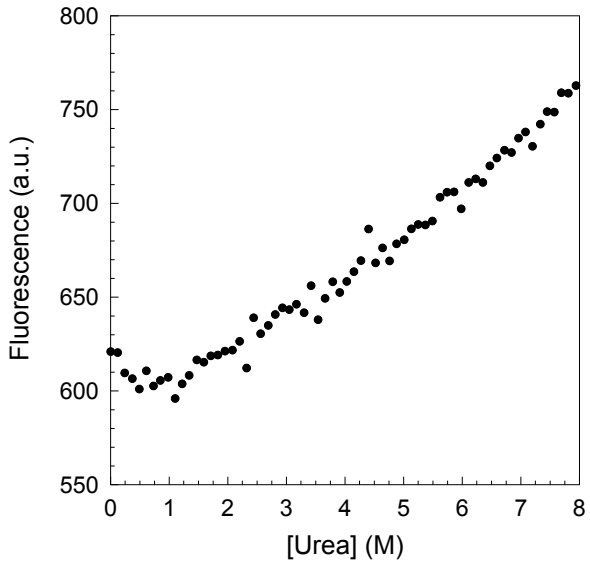


Figure 5.9: Equilibrium denaturation curve of purified E-G5^2 non-folding control. Fluorescence data at 305 nm of the triple mutant E-G5^2-G587A-G602A-G626A shows its unfolded nature.

Force profile experiments were also performed for E-G5^2 non-folding control. A scattered profile was observed throughout the entire plot. Fraction full-length values were much lower for all points of the non-folding control, including the ones of shorter linker lengths (Figure 5.10). The results for E-G5^2 non-folding control correspond to what would be expected for a polypeptide nascent chain that is unable to form a structure and release the stall, which is not what was observed for G5^2-non folding control.

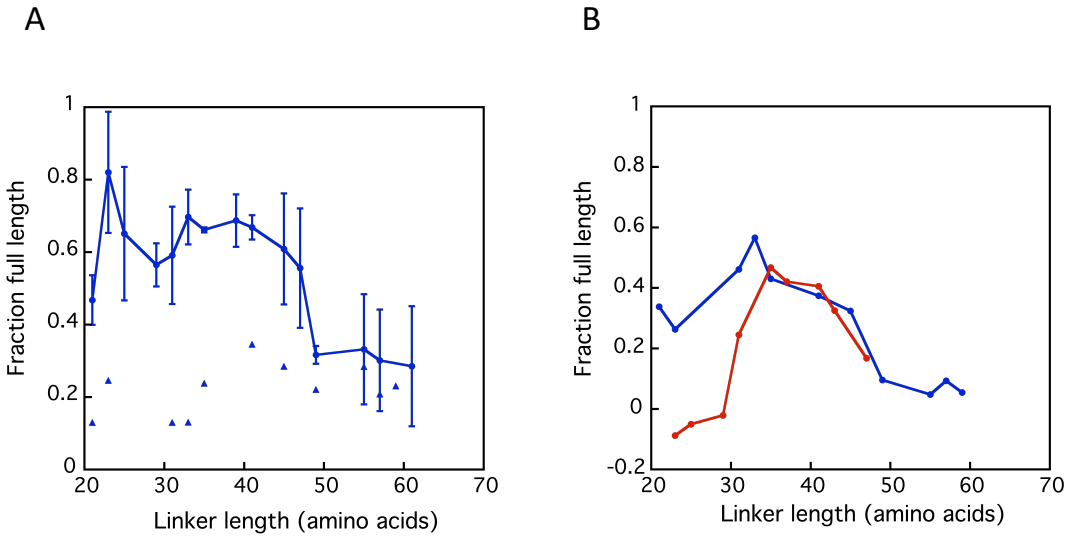


Figure 5.10: Force profile results for E-G5^2 and its non-folding control E-G5^2-G587A-G602A-G626A. (A) Force profiles of E-G5^2 (dots) and its non-folding control G5-G587A-G602A-G626A (triangles). The results are the average of at least two experiments for each linker length. (B) Comparison of the normalised data for E-G5^2 (blue) and G5^2 (red) taking into account the non-folding controls of each construct. The results are the average of at least two experiments for each linker length and the error bars represent the standard error.

We also studied on the ribosome E-G5^2 with a single Gly-Ala mutation, the one that disrupts the G5^2 domain structure in isolation (used as G5^2 non-folding control) but does not have the same effect on E-G5^2. It was previously showed that E-G5^2-G587A folds via the alternative pathway that starts by the formation of the E-G5^2 interface (Gruszka et al. 2016). Thus, the force profile of E-G5^2-G587A could shed some light in the role of the alternative folding pathway on the co-translational folding of SasG.

To our surprise, the force profile results of E-G5^2-G587A are very similar to the E-G5^2 non-folding control, the triple mutant G587A-G602A-G626A (Figure 5.11), suggesting that this single mutation in the E-G5^2 construct is unable to fold on the ribosome.

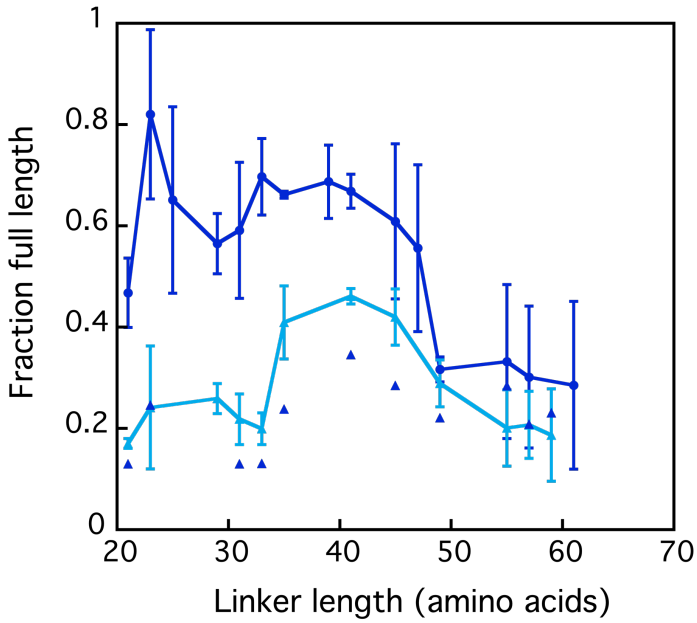


Figure 5.11: Comparative results of E-G5², E-G5²-G587A and E-G5² non-folding control. Force profiles of E-G5² (blue), EG5²-G587A mutant (cyan) and EG5²-G587A-G602A-G626A non-folding control (triangles). The results are the average of at least two experiments for each linker length and the error bars represent the standard error.

These results suggest that the alternative pathway for folding might not be an available option for E-G5² to fold on the ribosome. This is in agreement with what we see for G5² and E-G5², where no difference between the two force profiles was identified. The initial hypothesis that E-G5² would fold earlier, once the interface between the two domains was translated, did not prove to be correct. It has been shown that the ribosome can have destabilising effect on polypeptide nascent chains (Liu et al. 2017; Samelson et al. 2016), which could maybe explain the inability of E-G5² construct to fold via its alternative pathway for folding.

5.3.4 Role of charged residues on co-translational folding of SasG

The unexpectedly high values of fraction full-length for $G5^2$ non-folding control (G587A) were very intriguing. The ribosome has a highly charged surface, mainly negatively charged and electrostatics is known to play an important role on the nascent chain dynamics. It has been shown that the amplitude of local motion of a highly negatively charged nascent chain that is in close proximity with the ribosome is higher when compared with disordered polypeptides under physiological conditions. Moreover, highly positively charged polypeptides are shown to be less dynamic, as they can interact with the ribosome surface (Knight et al. 2013, Figure 5.12).

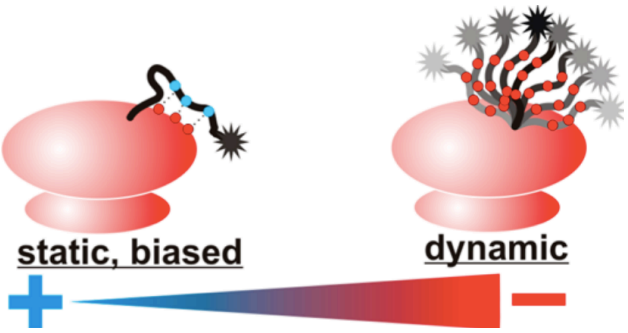


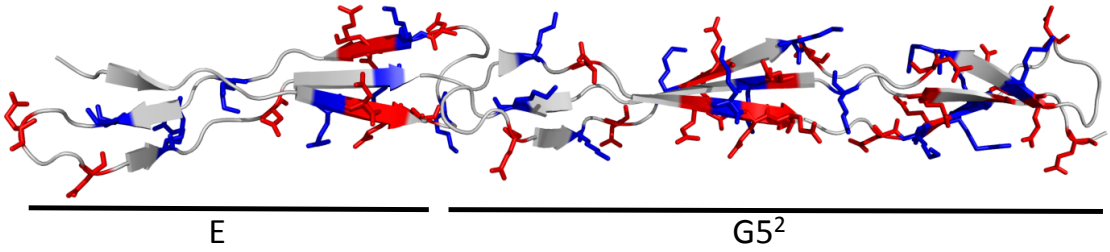
Figure 5.12: Schematics of the role of electrostatics on ribosome-nascent chain interaction. Due to the overall negatively charged nature of the ribosome surface, the more negatively charged the nascent chain, the more dynamic it will behave. Positively charged nascent chains will tent to interact with the ribosome surface and stay static (Figure taken from Knight et al. 2013).

The unusual charged content of SasG raised the hypothesis that the folding on the ribosome experiments could also be reporting some electrostatics contributions. In order get some insights on that, we designed three SasG constructs for both $G5^2$ and E-$G5^2$: one where all the positively charged residues were removed (only negative left), one with all negatively charged residues removed (only positive left) and one with all the charges removed (no charges left).

Positively charged lysine and arginine were mutated to methionine whereas negatively charged residues aspartic acid and glutamic acid were mutated to asparagine and glutamine,

respectively (all sequences are shown in Appendix 2). The mutations were chosen to minimise their effect as much as possible, as we wanted to test for the effect of the charges only. Charged residues in SasG are distributed all over the domain E and $G5^2$ domains (Figure 5.12).

A



B

	Positively charged	Negatively charged
E domain	7	10
$G5^2$ domain	13	18
E-$G5^2$	20	28

Figure 5.13: Charges content and distribution in SasG. (A) Cartoon representation of SasG E-$G5^2$ structure showing the charged residues distributions represented as sticks (positively charged in blue and negatively charged in red)). (B) Number of charged amino acids in E-$G5^2$ and E-$G5^2$ constructs.

The genes were purchased from GenScript in the standard pUC57 vector. The same molecular biology strategy previously used was employed to insert the genes into the pRSETa vector containing all linker lengths for the folding on the ribosome studies. Furthermore, the genes were also inserted into the pSKB2 plasmid for protein expression.

The six constructs, $G5^2$ and E-$G5^2$ only positive, only negative and no charges, were expressed. All proteins but E-$G5^2$ only positive and only negative charges were insoluble and went to the cell pellet. To check for the protein stabilities, equilibrium denaturation studies were performed for both E-$G5^2$ only positive and only negative charges. The results showed that the two proteins were unfolded (Figure 5.13).

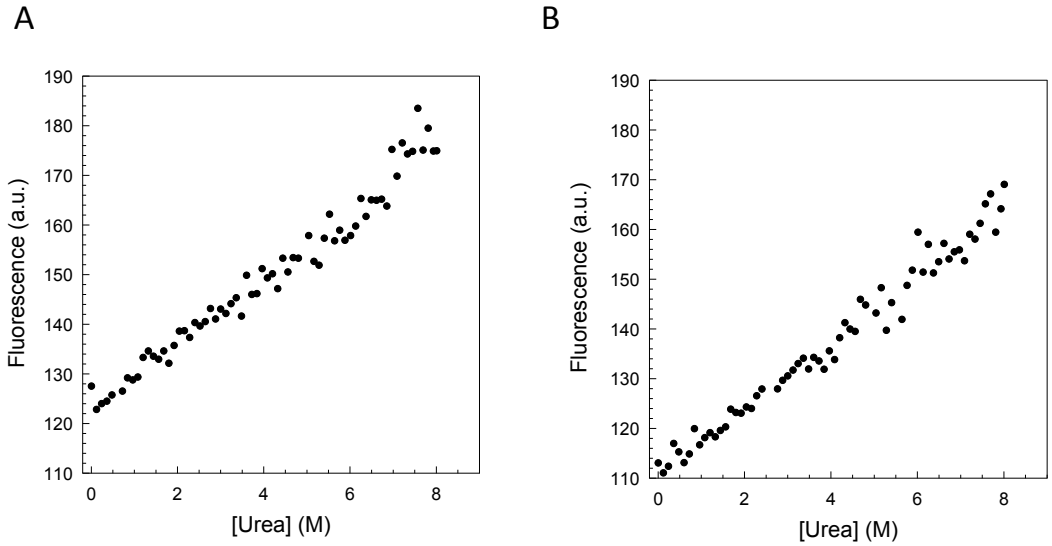


Figure 5.14: Equilibrium denaturation curves of purified proteins: E-G5^2 only negative and only positive charges. Fluorescence data at 305 nm of (A) E-G5^2 only negative and (B) E-G5^2 only positive charges show their unfolded nature.

With the preliminary knowledge that all designed proteins cannot fold, force profile experiments were performed for all six constructs. Results of both G5^2 and E-G5^2 were similar, showing that in the absence of negatively charged residues (only positive constructs), force profiles showed the lower fraction full-length readings. In contrast, the presence of negatively charged residues (only negative charges constructs) showed the higher read through. Constructs with no charges showed intermediate fraction full-length values almost through the entire force profiles (Figure 5.14).

The data suggest that the charge composition of the nascent chain can play a significant role in the force profile experiments, as the stall sequence can be released even by disordered polypeptides. The results show that force profile experiments can report not only on folding events but also other sources of non-native interactions intra- or/and inter-molecular, possibly between nascent chain and ribosome. They also reveal that the interactions largely occur while some parts of the chain are still sequestered in the tunnel (i.e. at linker lengths <35).

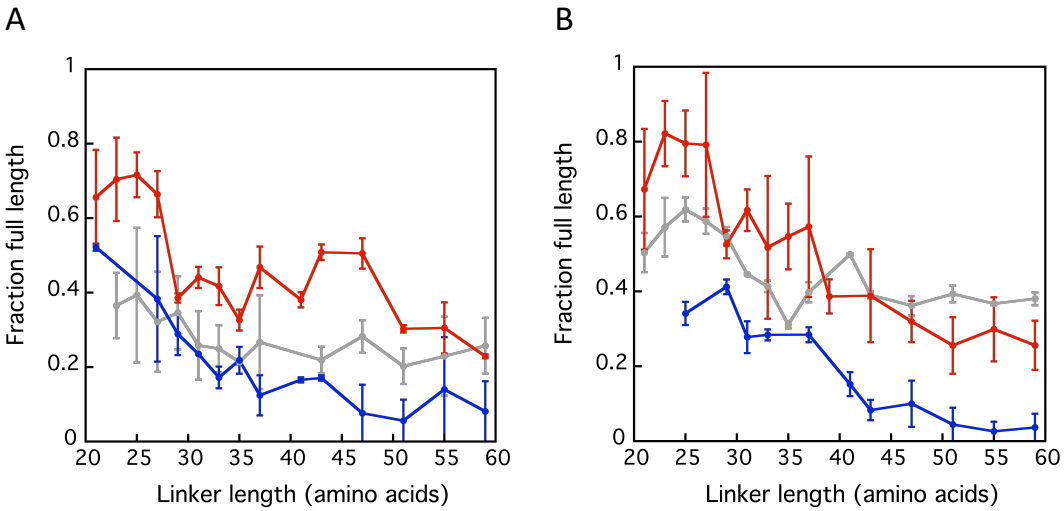


Figure 5.15: Effect of charged residues in SasG folding on the ribosome. (A) Comparative force profile results of G5[2]: only positive charges (blue), only negative charges (red) and no charge (grey). (B) Comparative force profile results of G5[2]: only positive charges (blue), only negative charges (red) and no charge (grey). The results are the average of at least two experiments for each linker length and the error bars represent the standard error.

5.4 Discussion

After an extensive study of SasG *in vitro*, it was possible not only to understand the folding pathway of both $G5^2$ domain in isolation and E-$G5^2$, but we have also identified an alternative pathway for folding of the multidomain construct E-$G5^2$ (Gruszka et al. 2016). The results showed very clearly that the interface between the two domains is key for this intriguing protein to maintain its stability. The lack of a hydrophobic core, the highly charged sequence composition and with the preliminary knowledge of the ability of E-$G5^2$ to fold via the interface between the two domains, SasG became an interesting system to study on the ribosome.

Force profile experiments results showed an early folding of $G5^2$, while the domain is still inside the ribosome vestibule (linker length L31). Surprisingly, high fraction full-length values were also observed for $G5^2$-G587A, a Gly-Ala mutant that is known to disrupt the domain structure. The result was an indication that force profile experiments can report not only the folding of a nascent chain, but other factors could also contribute to the release of the stall sequence.

It was interesting to notice high read through of fraction full-length for shorter linker lengths (<L30). It has been previously reported that at very short linker lengths (L18-L20), the ribosome tunnel available space (10-15 Å wide) can accommodate alpha helices formation (Wilson et al. 2011). It has also been shown that small β-sheet proteins (<4.2 KDa) can apply a force and release the stall at short linker lengths, folding while still inside the ribosome tunnel (Marino, Von Heijne, and Beckmann 2016). But SasG results also showed high fraction full-length values for $G5^2$ non-folding control ($G5^2$-G587A), which is an indication that the stall release does not come from the folding of the domain in this case.

In comparison to the $G5^2$ domain, the E-$G5^2$ construct shows an almost identical force profile. They are both able to fold at similar positions when being translated, while still inside the ribosome vestibule. Our initial prediction was that E-$G5^2$ could fold earlier than $G5^2$ on the ribosome, as it has an alternative pathway that starts by the formation of the inter-domain interface. The results did not support this hypothesis. Moreover, a single mutant G587A, that is known to disrupt the $G5^2$ domain in isolation but does not prevent

the E-G5^2 to fold *in vitro*, showed a similar force profile result when compared to the non-folding control (E-G5^2-G587A-G602A-G626A). *In vitro*, this Gly-Ala mutation on E-G5^2 disrupts the main folding pathway (via the C-terminal end of G5^2) but the structure is formed via its alternative one (via the domains interface). All together, these data suggest that the alternative folding pathway might not be available on the co-translational folding of SasG E-G5^2, maybe due to the additional destabilising effect that the ribosome can confer to the nascent chain. Despite the unexpected results, co-translational studies of SasG show the first example of the formation of large all β-sheet structures (≈10 kDa for G5^2 and ≈15 kDa for E-G5^2) inside the ribosome.

Based on the highly charged nature of SasG and to try to gain some more insights about other factors that could contribute to the stall release on the force profiles, electrostatic effects were tested. Constructs containing only positive, only negative and no charged residues were designed for both G5^2 and E-G5^2. The results show that nascent chains containing only positive charges were mostly unable to release the stall whereas the ones with only negatively charged residues showed the higher fraction full-length values throughout their force profiles. In the absence of charged residues, the constructs showed intermediate fraction full-length values for a great proportion of linker lengths, including the ribosome vestibule region, where both G5^2 and E-G5^2 were shown to fold co-translationally (≈L30 – L40).

It has been shown that nascent chain dynamics varies depending on its charged nature when in close proximity with the ribosome (Knight et al. 2013). Highly negatively charged polypeptides were found to be very dynamic, with higher amplitude of local motion in comparison to disordered polypetides. Highly positively charged nascent chains, on the other extreme, were shown to be static as they can interact with the negatively charged ribosome surface. In light of those findings, we could hypothesise that SasG only negative charges constructs are the most dynamic of the designed constructs, what could explain the release of the stall at most linker lengths. The only positive charges constructs can interact with the ribosome surface, thus they stay attached to the surface and are not able to release the stall. No charged SasG variants show an intermediate behaviour. Maybe the lack of charged residues causes a hydrophobic collapse of the structures that can also apply a tension on the stall and allow its release.

The results also show high fraction full-length readings at shorter linker lengths for nearly all constructs, suggesting that electrostatics on its own is not enough to explain that behaviour. Despite that, the study suggests that force profile experiments can report both on intramolecular folding of a protein when being translated, as well as, intermolecular interactions between the nascent protein chain and the ribosome.

6 Encoding IDP coupled folding and binding

This chapter is the result of a project done in collaboration with other two students in my group, Dr. Michael Crabtree and Quenton Bubb. Each of us worked with one system composed by an IDP and a partner protein. The biophysical results of the three systems were then combined with the work done by Dr. Joseph Rogers. Together, the four systems allowed us to gain some knowledge on the encoding of coupled folding and binding reactions of IDPs. I would like to thank Tristan Kwan for the expression and purification of PUMA-35 used in my dissociation experiments as a competitor peptide. The results of this chapter were published in Crabtree et al. 2018.

6.1 Introduction

6.1.1 What does it mean for a protein to be disordered?

Natively disordered proteins comprise an important and well represented group of proteins. Their biological relevance is well established in nature (e.g. cell signalling and regulatory processes) as we can see their persistence throughout evolution. But what are these proteins?

To be considered intrinsically disordered, a protein should be an extended, flexible polypeptide with very little or no secondary structure present under physiological conditions, although transient secondary structures is known to occur in such disordered ensembles. Over the past two decades, an increasing amount of those proteins have been reported and characterised (Uversky 2002).

Although disordered in isolation, a variety of those IDPs can gain structure and become relatively rigid upon binding to natural partners. This phenomenon is known as coupled folding and binding (Figure 6.1).

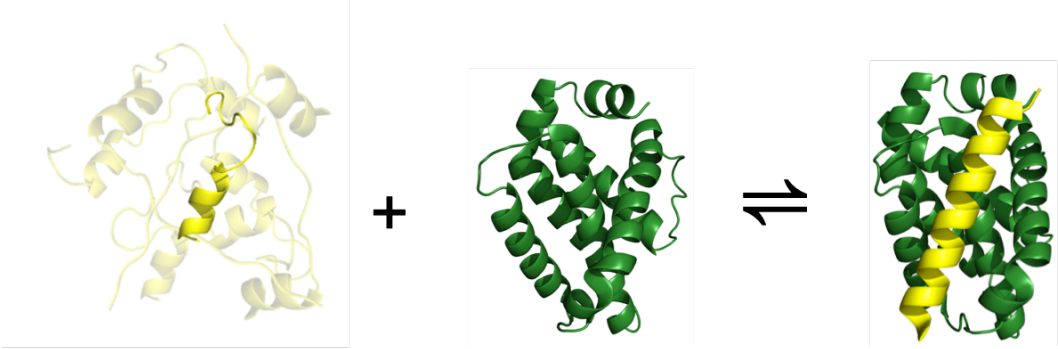


Figure 6.1: Schematic of a coupled folding and binding reaction. Adapted from Rogers et al. 2014.

6.1.2 Encoding protein folding and function

Since the discovery of the first protein structure reported in 1958, the protein structure-function paradigm has been simply described as: in order for a polypeptide to function, it needs to fold. Adopting a three-dimensional structure was believed to be the key requirement for a protein to be able to function (Alberts 2008). With the discovery of IDPs, natively unstructured proteins that are able to play important roles *in vivo,* our attention was brought back to the fundamental definition of what encodes protein folding and function.

Early experiments performed by Anfinsen demonstrated that the information required for a polypeptide chain to fold into its unique three-dimensional structure is built in the protein sequence composition (Anfinsen *et al.*, 1961; Anfinsen, 1973). However, IDPs are natively unstructured, so we might say that they have disorder encoded in their sequence (Romero et al. 2001). Nonetheless, some IDPs become folded upon binding to a partner protein, which raises the question: what encodes a coupled and folding and binding reaction – IDPs or partner proteins?

6.1.3 The BCL-2 family network

B-cell lymphoma 2 (BCL-2) family is well known and described as the cell apoptosis regulator. The family is composed by a number of proteins that share the evolutionary conserved BCL-2 homology (BH) domains (Reed, 1998; Korsmeyer, 1999). Protein members of this family act as pro- or anti-apoptotic in the cell by controlling the permeabilisation of the mitochondrial outer membrane (Kvansakul et al. 2008). Pro-survival and pro-apoptotic proteins share four BH domains (BH1-BH4, Figure 6.2, A) and fold to similar globular structures: seven amphipathic α-helices surrounding a central hydrophobic helix (α5) (Muchmore *et al.*, 1996; Figure 6.2, B). BH3-only proteins are composed only by the BH3 amphipathic helix (Figure 6.2, A), which mediates their interaction with the hydrophobic groove (formed by helices 2-5) on the surface of the globular protein members of the family. Unlike the other family members, BH3-only proteins are intrinsically disordered in isolation but become helical upon binding to the partner proteins (Hinds et al. 2007), which makes the BCL-2 family an interesting network to study coupled folding and binding.

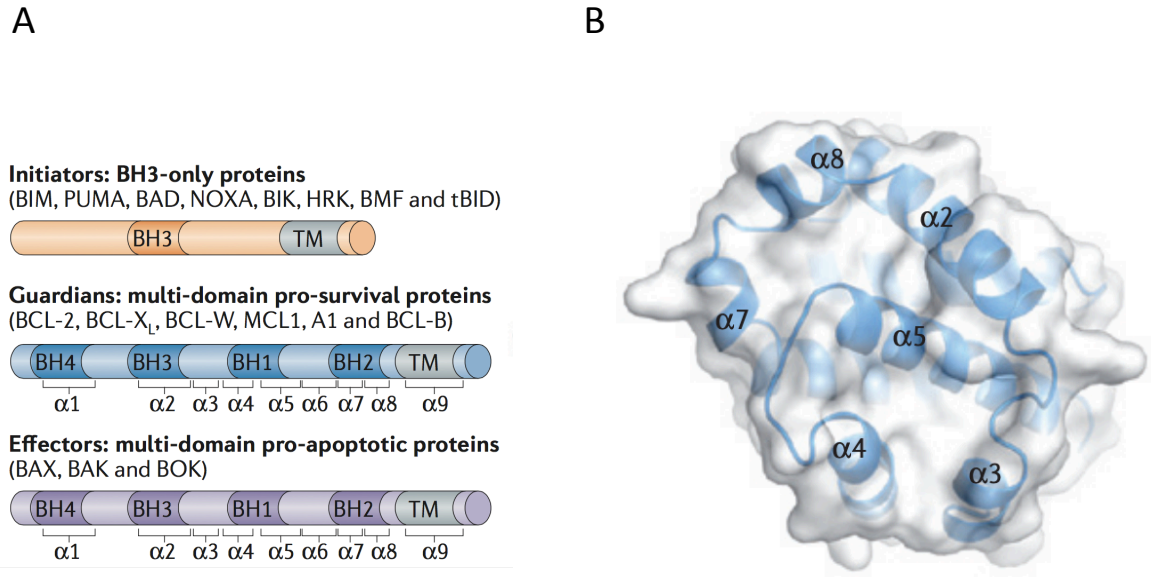


Figure 6.2: The role of the BCL-2 family network in controlling cell death. (A) Schematic of the BCL-2 family protein members divided in three categories: initiators, pro-survival and pro-apoptotic. Initiator proteins contain only the BH3 helix, whereas pro-survival and pro-apoptotic proteins have four BCL-2 homology regions (BH1-BH4) (B) Example of a BCL-2 family member fold, composed by seven amphipathic α-helices and a central hydrophobic helix (α5). Figure adapted from Czabotar et al. 2014.

6.1.4 The BCL-2 family as a model to study encoding of coupled folding and binding reactions using Φ-value analysis

Previous work carried out by Dr. Joseph Rogers was able to characterise the transition state of the coupled folding and binding reaction between the disordered PUMA peptide and MCL-1. Φ-value analysis revealed that at the transition state, PUMA is partially folded towards its N-terminus whereas the rest of the peptide remains disordered (Rogers et al. 2014; Figure 6.3).

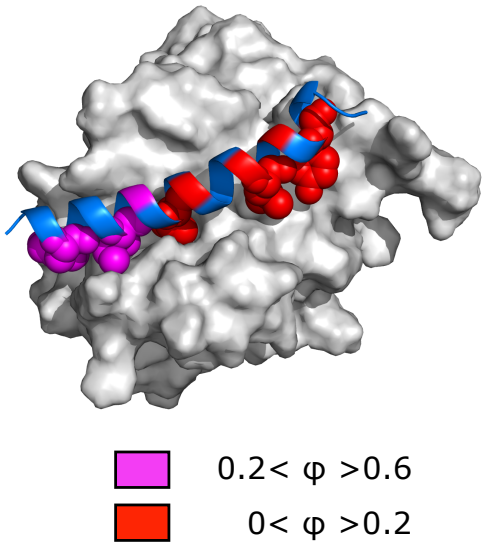


Figure 6.3: Φ-value analysis of PUMA peptide binding to MCL-1. The results show that formation of PUMA structure starts via the N-terminus (intermediate Φ-values) whereas the rest of the molecule is still disordered at the transition state (low Φ-values). Data and figure from Rogers et al. 2014.

Using the previous results from PUMA:MCL-1, we decided to use the BCL-2 family as a model to unravel the encoding of IDPs coupled folding and binding. We chose to study two IDPs (PUMA and BID) and two partner proteins (MCL-1 and A1), which have low sequence identity (≈20%) but very high structure identity between them (Figure 6.4 and 6.5). Comparative Φ-value analysis of the four systems (BID:A1, BID:MCL-1, PUMA:A1 and PUMA:MCL-1) would help us to shed some light and understand the encoding of coupled folding and binding reactions.

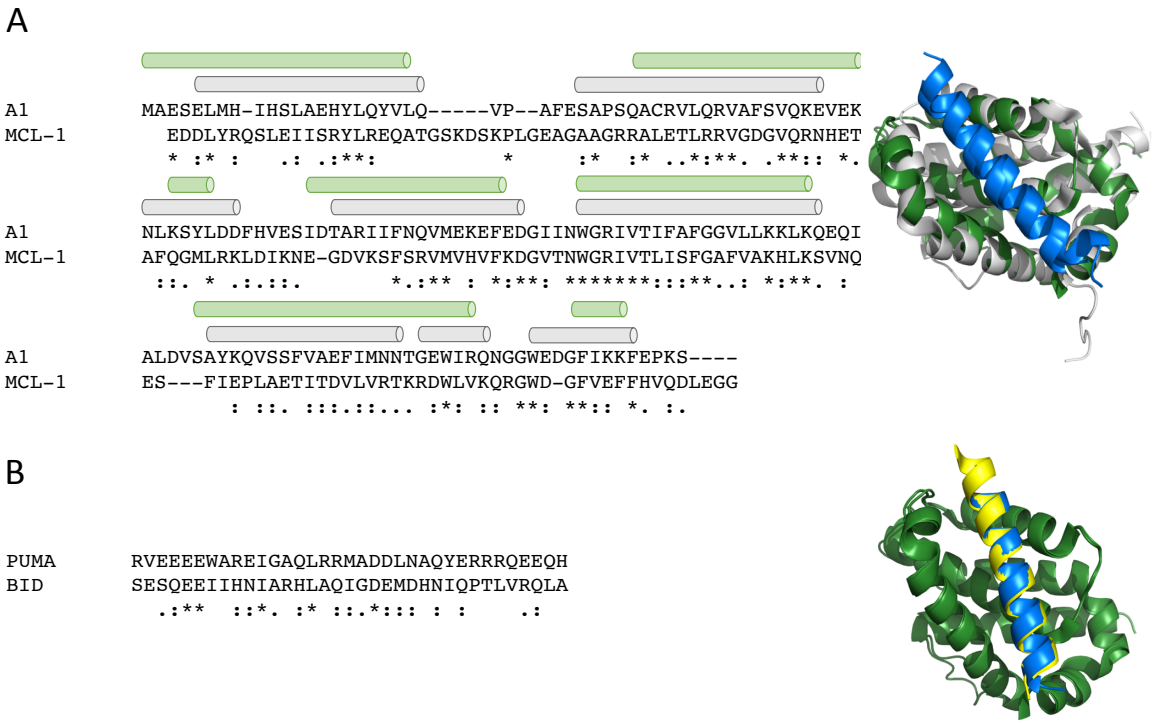


Figure 6.4: Sequence and structural alignment of A1, MCL-1, PUMA and BID constructs. (A) Left: sequence alignment for A1 and MCL-1 with helical regions represented by green and grey cylinders, respectively. Right: structural alignment of A1 (green) and MCL-1 (grey). The proteins are in complex with PUMA peptide, shown in blue. (B) Left: sequence alignment for PUMA and BID peptides. Right: structural alignment of PUMA (blue) and BID (yellow) peptides. The peptides are in complex with A1, shown in green. Alignments were produced using Clustal Omega (Goujon et al. 2010, Sievers et al. 2011): conserved residues are indicated with an asterisk, residues with strongly similar properties with a colon, and residues with weakly similar properties with a period. Note that in addition to the sequences shown above, both MCL-1 and A1 contained a GS at the N-terminus, which was a remnant from protease cleavage during purification. Figure adapted from Crabtree et al. 2018.

6.2 Chapter aims

The protein structure-function paradigm has been assumed for decades. With the discovery of IDPs and their biological relevance, we had to rethink the assumption that function is strictly related to folding. However, a subset of IDPs is able to fold upon binding to a partner protein. Since the early studies performed by Anfinsen, we know that the information required for a protein to fold is encoded in its own sequence, but what about IDPs? Is the folding of an IDP in a coupled folding and binding reaction encoded by the disordered polypeptide or is it templated by the folded partner? In this chapter we will use Φ-value analysis to investigate different systems and get some insights in what drives the folding of proteins that have disorder encoded in their sequences.

To do so, we are going to use four systems composed by two BH3-only IDPs (PUMA and BID) and two partner proteins (A1 and MCL-1), all members of the BCL-2 family (Figure 6.5).

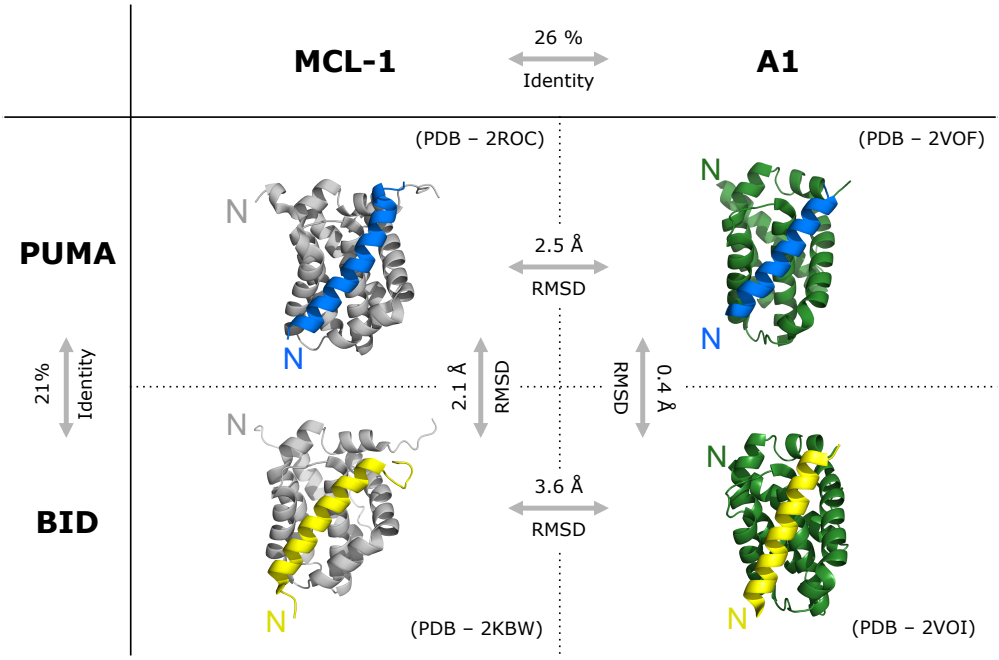


Figure 6.5: Structural homology and sequence identity of the complexes formed by the BH3-only BID and PUMA with the pro-survival MCL-1 and A1 proteins. The BH3-only IDPs PUMA (blue) and BID (yellow) can fold to an alpha helix structure upon binding to partner proteins like MCl-1 (grey) and A1 (green). Root mean square deviation (RMSD) were obtained by structural aligments using PyMOL and sequence identities were calculated from the sequence alignments of the mouse protein constructs. Figure taken from Crabtree et al. 2018.

6.3 Results

6.3.1 Molecular biology strategy

Mouse A1 gene was kindly donated from Bonsu Ku's lab in the pGEX vector as a GST fusion protein, containing a TEV cleavage site. Unlabeled PUMA-35 peptide used in this chapter as a competitor of BID in the dissociation experiments was expressed and purified in the lab by Tristan Quan. TAMRA labeled BID-WT and mutants peptides were purchased from Biomatik (Canada).

6.3.2 Purification of A1

A1 was expressed as a GST fusion protein, with a TEV cleavage site in between. The expressed protein was bound to the glutathione sepharose (GS) resin and cleaved using TEV. A1 was purified by ion exchange chomatography. An additional step in Superdex 75 gel filtration (Figure 6.6, A) was employed to buffer exchange the protein. SDS-PAGE confirms the purity of the protein although the sample runs further than it would be expected by its mass (Figure 6.6, B).

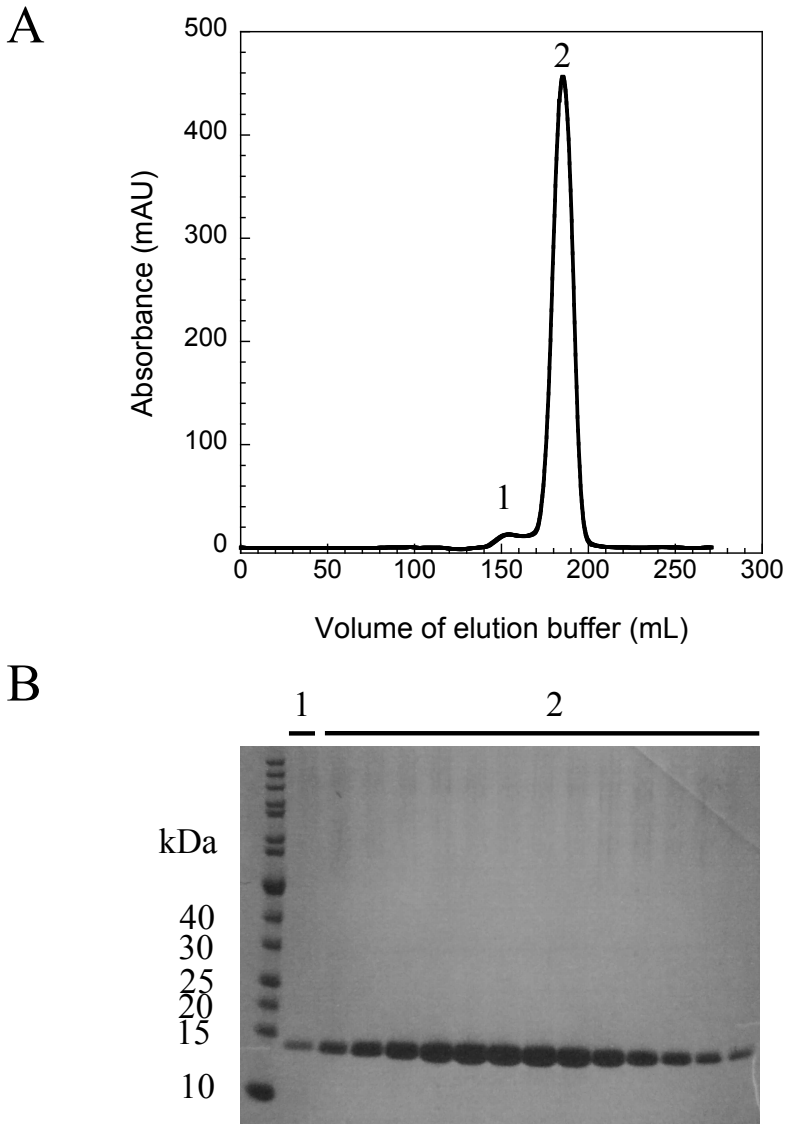


Figure 6.6: Purification of A1. (A) Chromatogram image of A1 purification after G75 size exclusion run showing peaks 1 and 2. (B) SDS-PAGE of purified A1 after a G75 run showing protein fractions and NuPage unstained protein ladder for comparison. The gel starts from marker and show one fraction of peak 1 followed by peak 2. Pure A1 is present in all fractions.

Purified A1 was sent for mass spectroscopy analysis at the Department of Chemistry. Results confirmed A1 identity and purity with the confirmation of its expected mass. Figure 6.7 shows a chromatogram as an example (sample sent by Dr. Michael Crabtree).

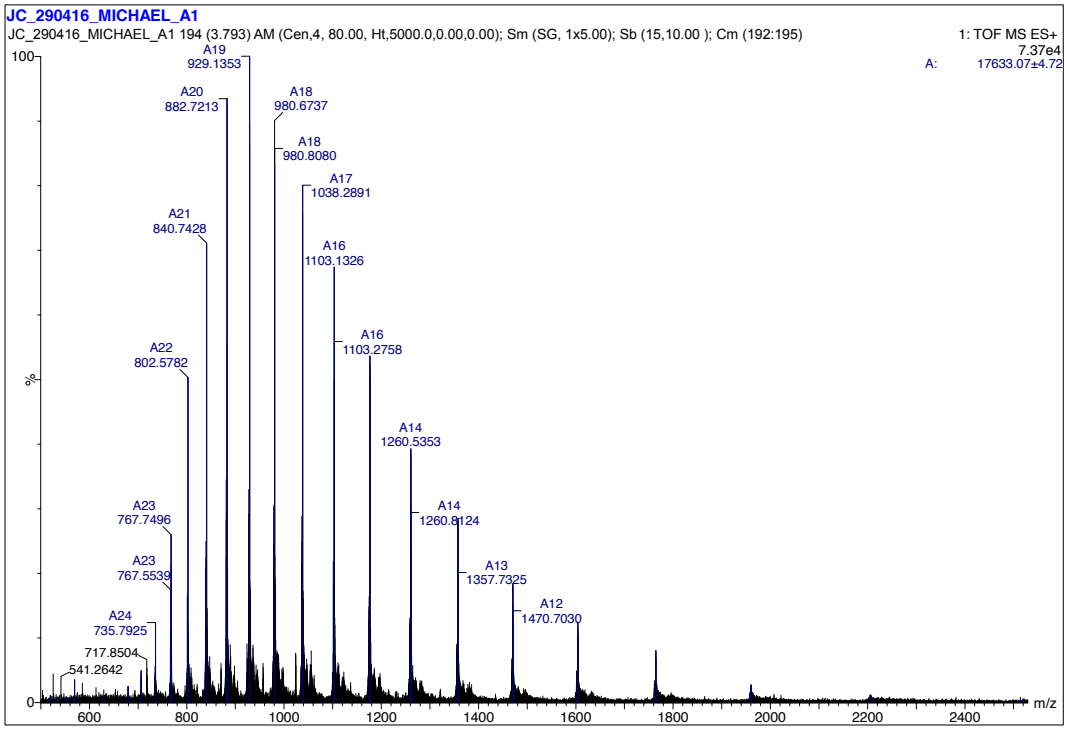


Figure 6.7: Mass spectroscopy results of A1 after purification. The obtained mass of A1 was 17633 ± 5 Da and the expected mass was 17635 Da. Dr. Michael Crabtree submitted the sample to analysis.

6.3.3 TAMRA labelled peptides

Mouse BID peptides (WT and mutants) with an N-terminal TAMRA fluorescent dye attached were designed and purchased from Biomatik. The mutations were aimed to allow the investigation of coupled folding and binding of BID and a partner protein using Φ-value analysis.

Similarly to folding, Φ-value analysis of a coupled folding and binding reaction is performed by comparing the difference in free energy between the transition state and the

bound state of the two proteins (complex formation). To do so, mutations are designed to disrupt interactions between the proteins allowing a change in free energy between the WT and mutant. This difference in free energy ($\Delta\Delta G$) is required to be greater than 0.6 kcal.mol^{-1} to allow us to have confidence in the calculated Φ-value. (Fersht and Sato 2004).

Two sets of single mutations were designed:

- **Hydrophobic to alanine mutants: designed to disrupt the interactions between BID and its partner protein (e.g. A1, MCL-1).** Hydrophobic residues of BID are found buried in the partner protein's hydrophobic groove upon binding. Mutating them to alanine reduces the side chain of the residues resulting in a loss in affinity of the complex.
- **Alanine to glycine mutants: designed to probe the peptide helicity within the BID:partner complex transition state.** In a coupled folding and binding reaction, the helicity of the peptide can be investigated by mutating residues that are not involved in intermolecular interactions. Charged residues in the peptide that are solvent exposed and do not make contact with the partner protein in the bound state are the target for this mutations. Firstly they are mutated to alanine, which will be our pseudo-WT. Next, they are mutated to glycine and the difference in free energy between the glycine and our pseudo-WT allow us to calculate the Φ-values.

Importantly, when designing the mutations for Φ-value analysis, we want to guarantee that they are not too disruptive or they can affect the transition state otherwise. This could change the reaction pathway by involving different interactions in the transition state when compared to WT. To test for that hypothesis, a double mutant comprised of two mutations that have been previously characterised individually is employed: if we take BID-I86A-M97A double mutant for instance, the Φ-value of M97A can be calculated from the difference in free energy between the double mutant and I86A mutant. If the same Φ-value is obtained, we can assume that the I86A mutation does not affect the interaction of M97A in the transition state.

6.3.4 Extinction coefficient

In order to get an accurate concentration measurement, experimental extinction coefficients were calculated for both A1 and the TAMRA labelled BID peptides. A precise volume of each sample was sent for amino acid analysis (Biochemistry department) and the accurate concentration was plotted against the absorption at 280 nm for A1 and 555 nm for TAMRA. The results allowed us to calculate the experimental extinction coefficient (ε_{280}) (Figure 6.8). The values were 24,200 ± 200 $M^{-1}.cm^{-1}$ and 85,000 ± 1,000 $M^{-1}.cm^{-1}$, for A1 and BID peptides respectively. Throughout this chapter, A1 and BID peptides concentrations were calculated using the experimental extinction coefficient values.

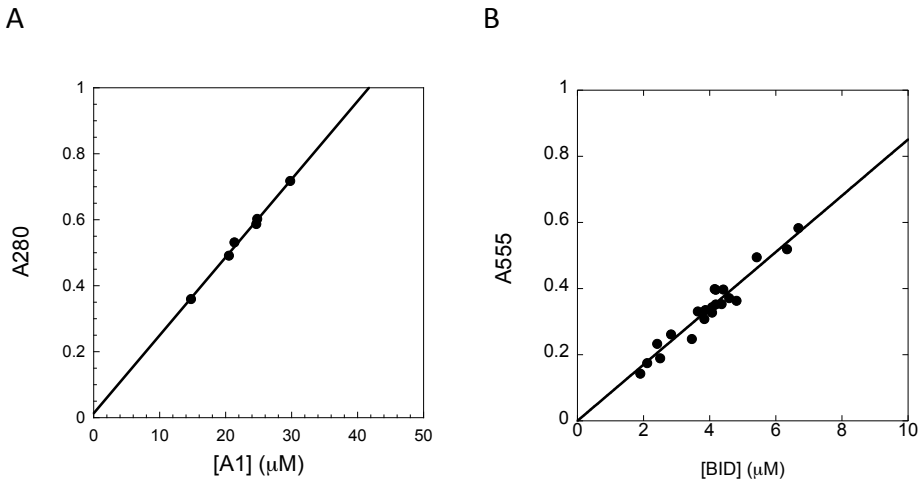


Figure 6.8: Extinction coefficient for A1 and TAMRA labelled BID. Extinction coefficient values were calculated from amino acid analysis (AAA). Concentration results for different A1 samples and TAMRA BID peptides were plotted against its absorbance readings at (A) 280 nm for A1 and (B) 550 nm for BID. The experimental extinction coefficients obtained were 24,200 ± 200 $M^{-1}.cm^{-1}$ and 85,000 ± 1,000 $M^{-1}.cm^{-1}$ for A1 and BID, respectively. The experimental values were used for all A1 and BID concentration calculations in this chapter.

6.3.5 Effect of mutations in TAMRA-BID peptides

To investigate the oligomerisation propensity of BID peptides, CD data were collected for all designed peptides at two concentrations: 5 and 10 μM. Results are shown in figures 6.9 and 6.10. For all peptides, no significant change was observed with the increase on concentration, meaning that the residual helicity was independent of the peptide concentration. The results gave us enough confidence that the CD spectrum represents the peptide monomer.

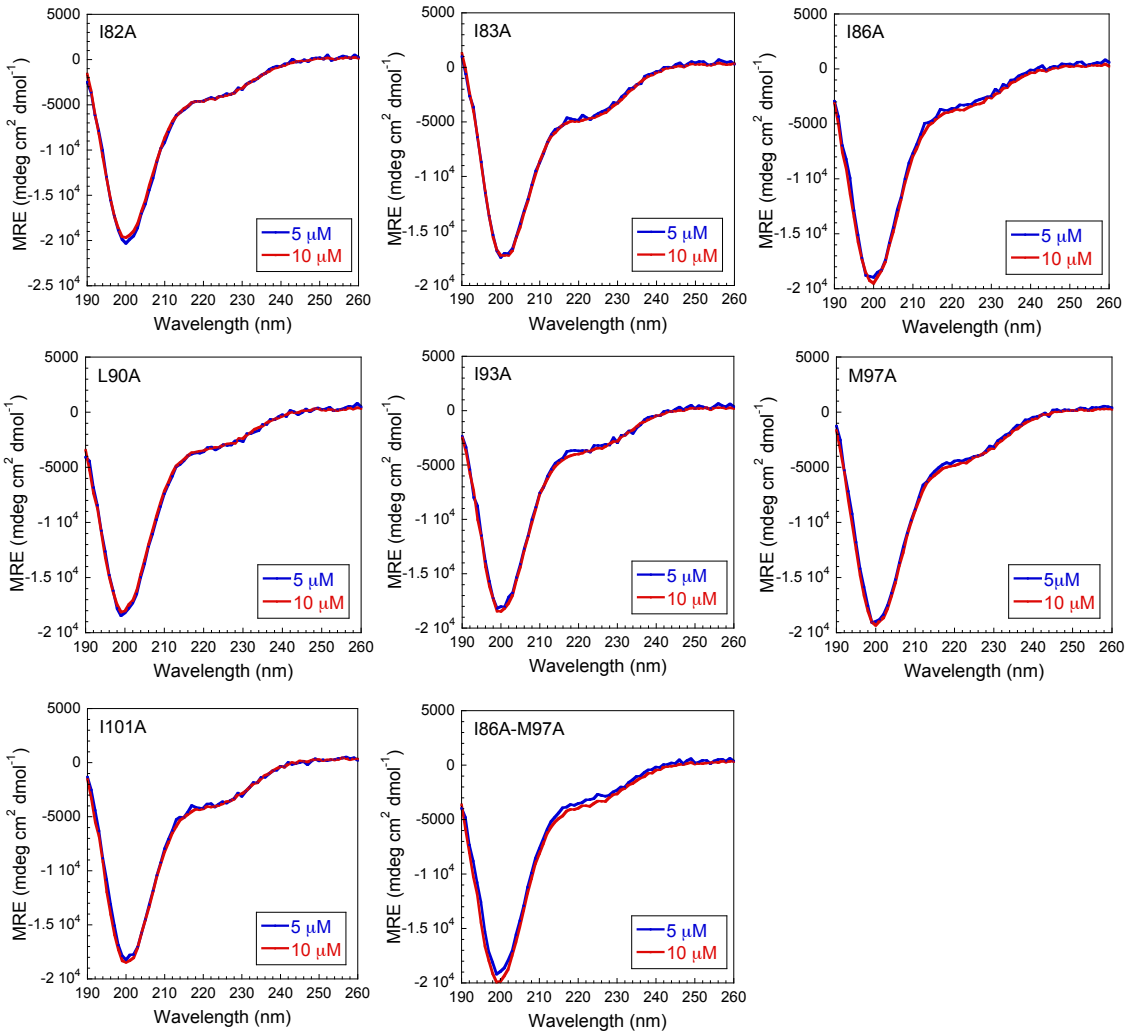


Figure 6.9: Concentration dependence CD spectra of hydrophobic to alanine mutants of BID. CD spectra of all BID peptides containing a hydrophobic to alanine mutation at two different concentrations (5 and 10 μM) show that the residual structure is maintained for all peptides with an increase in concentration.

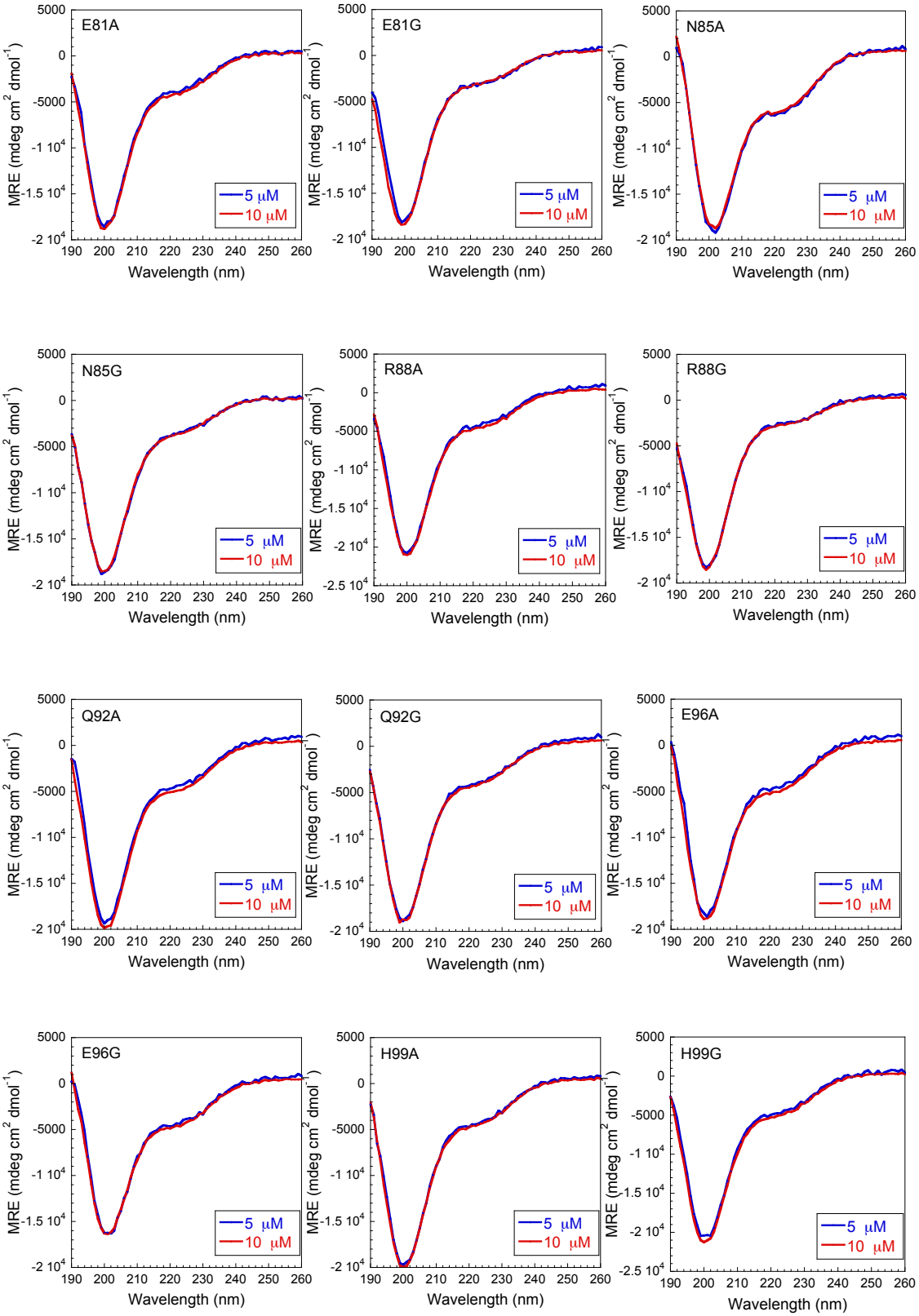


Figure 6.10: Concentration dependence CD spectra of alanine to glycine mutants of BID. CD spectra of BID peptides containing a charged to alanine/glycine mutation at two different concentrations (5 and 10 μM). The data show that the residual structure is maintained for all peptides with an increase in concentration.

Residual helicity was calculated for all BID peptides using the mean residual ellipticity (MRE) at 222 nm and the Muñoz and Serrano method (1995) (Figure 6.11). The mutation N85A showed an evident increase in helicity compared to WT whereas all other peptides displayed a very similar or reduced result. Alanine to glycine mutations decreased the residual helicity of the peptides as expected (Serrano et al. 1992).

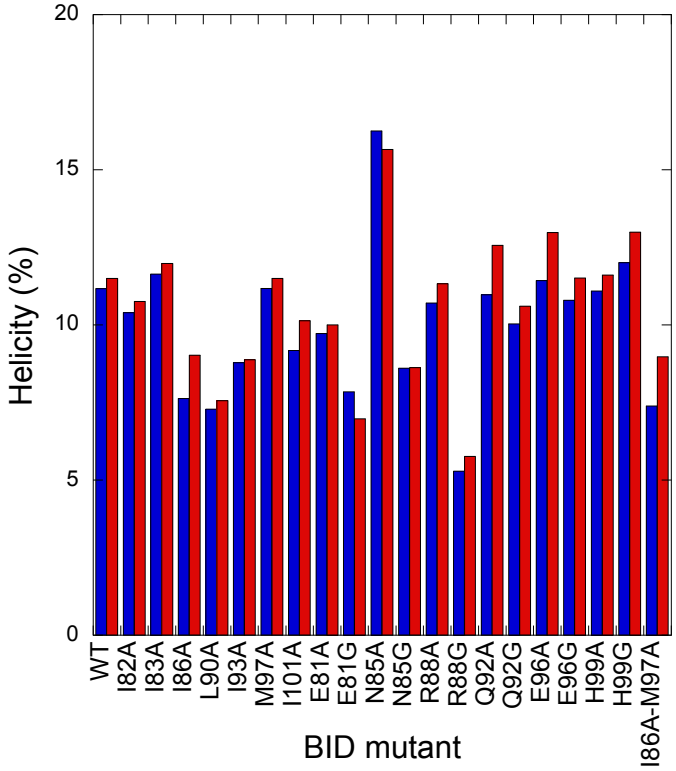


Figure 6.11: Comparative helicity of BID mutants. BID peptides helicities at 5 μM (blue) and 10 μM (red). Residual helicities were calculated from the MRE at 222 nm.

6.3.6 Φ-value analysis

In order to characterise the transition state of BID:A1 complex and to gain some knowledge on how that would compare with the other three systems (BID:MCL-1, PUMA:A1 and PUMA:MCL-1), Φ-value analysis was employed. Association and dissociation kinetics of A1 and TAMRA labelled BID peptides (WT and mutants) were investigated. Fluorescence

intensity change or anisotropy over time were monitored. Association experiments were performed by rapidly mixing BID and different concentrations of A1 under pseudo-first order conditions using a stopped flow instrument. A series of individual traces were averaged for each concentration and the data fitted to a single exponential (Figure 6.12, A, B and C). Dissociation kinetics were performed by competing a pre-formed complex of BID:A1 with an excess of unlabelled WT-PUMA peptide at different concentrations (Figure 6.12, D, E and F).

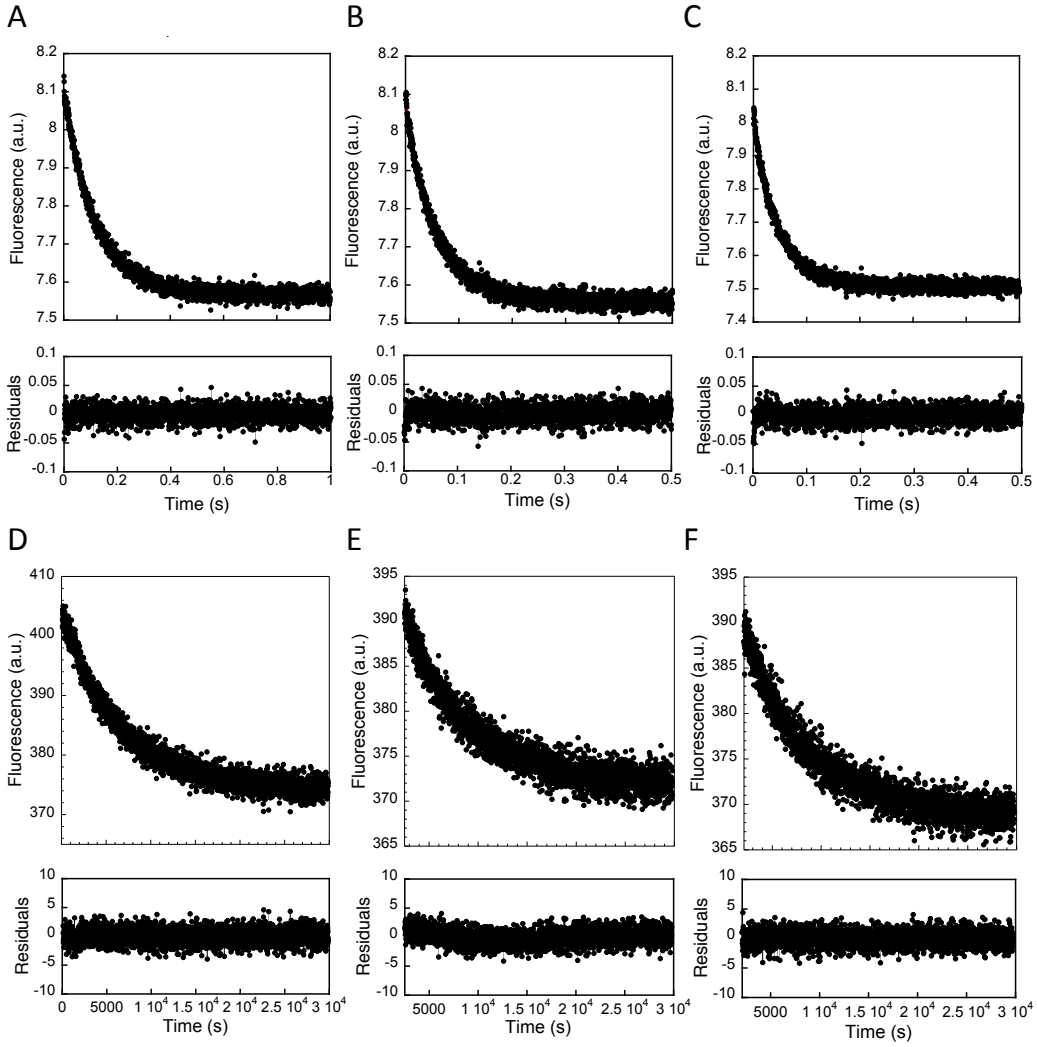


Figure 6.12: Association and dissociation kinetics raw data for WT-BID:A1. Fluorescence intensity change of the association kinetics were obtained by rapidly mixing BID and A1 under a pseudo-first order condition: 0.05 μM BID with (A) 1 μM A1, (B) 2 μM A1 and (C) 2.5 μM A1. Data represent the average of 20-30 traces at each concentration. The results were fit to a single exponential function. Fluorescence intensity change of the dissociation kinetics were obtained by mixing a pre-formed complex of BID:A1 with PUMA in excess: 4 μM BID:A1 competing with (D) 80 μM PUMA, (E) 120 μM PUMA and (F) 160 μM PUMA. The data were fit to a single exponential plus drift function.

For both association and dissociation experiments, the fit of the signal change at each concentration gave us an observed rate constant (k_{obs}). k_{obs} was plotted against the concentration of the partner protein and the association rate constant (k_{on}) was obtained from the gradient of the straight line (Figure 6.13, A). For dissociation experiments, k_{obs} was plotted against the ratio between the concentrations of competitor over the complex and the dissociation rate constant (k_{off}) was the means of the observed rates (Figure 6.13, B).

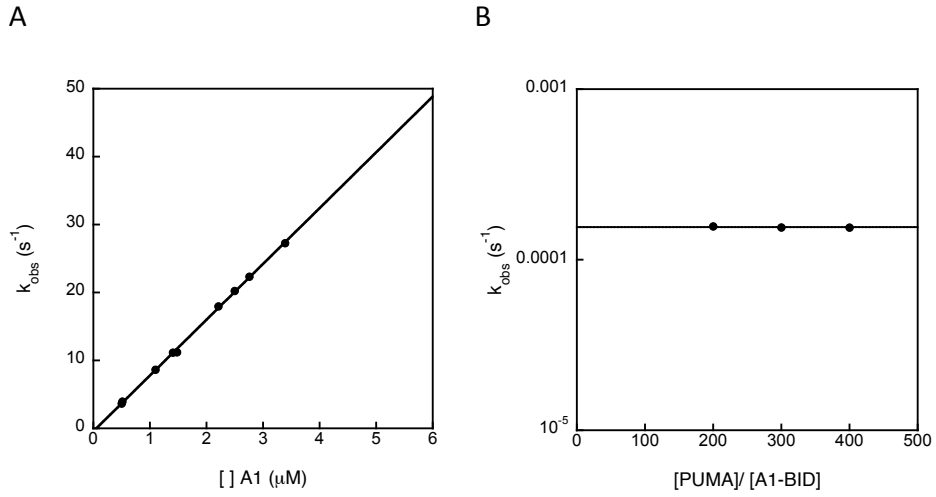


Figure 6.13: Kinetics data for WT-BID:A1. (A) Association kinetics of BID interacting with A1. The experiments were performed in a pseudo-first order condition with two different A1 stocks. The gradient of the straight line gave an association rate constant of 8.2 ± 0.2 μM^{-1} s^{-1}. (B) Dissociation kinetics was done using a pre-formed complex of BID:A1 mixed with unlabelled PUMA peptide. The mean of the observed rate constants gave a dissociation rate constant of 0.155 ± 0.005 x10-3 s-1. Errors in kon represent the error from the fit whereas in koff it comes from the standard error of the mean.

After characterising WT-BID, the same experiments were performed for all BID mutants. Figures 6.14 and 6.15 show the association and dissociation data for all hydrophobic to alanine mutants in comparison with WT-BID. Alanine to glycine mutants data are shown in figures 6.16 and 6.17.

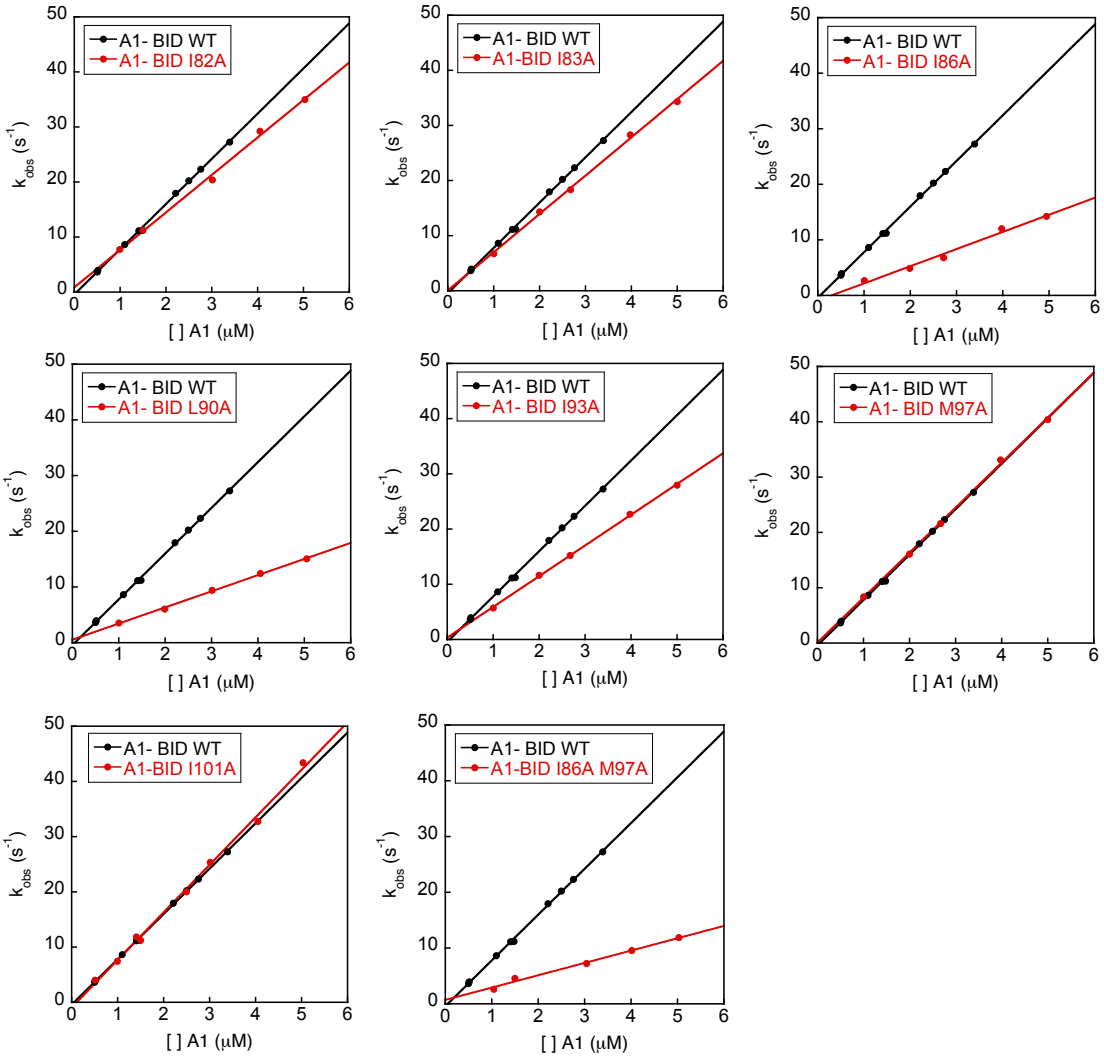


Figure 6.14: Association kinetics of BID hydrophobic to alanine mutants interacting with A1. BID peptides were rapidly mixed (stopped flow) with an excess of A1 at different concentrations. The observed rate constant (k_{obs}) was obtained by fitting the change in fluorescence at each concentration to a single exponential. Association rate constant (k_{on}) was obtained from the gradient of the straight line fit of k_{obs} against the concentration of A1.

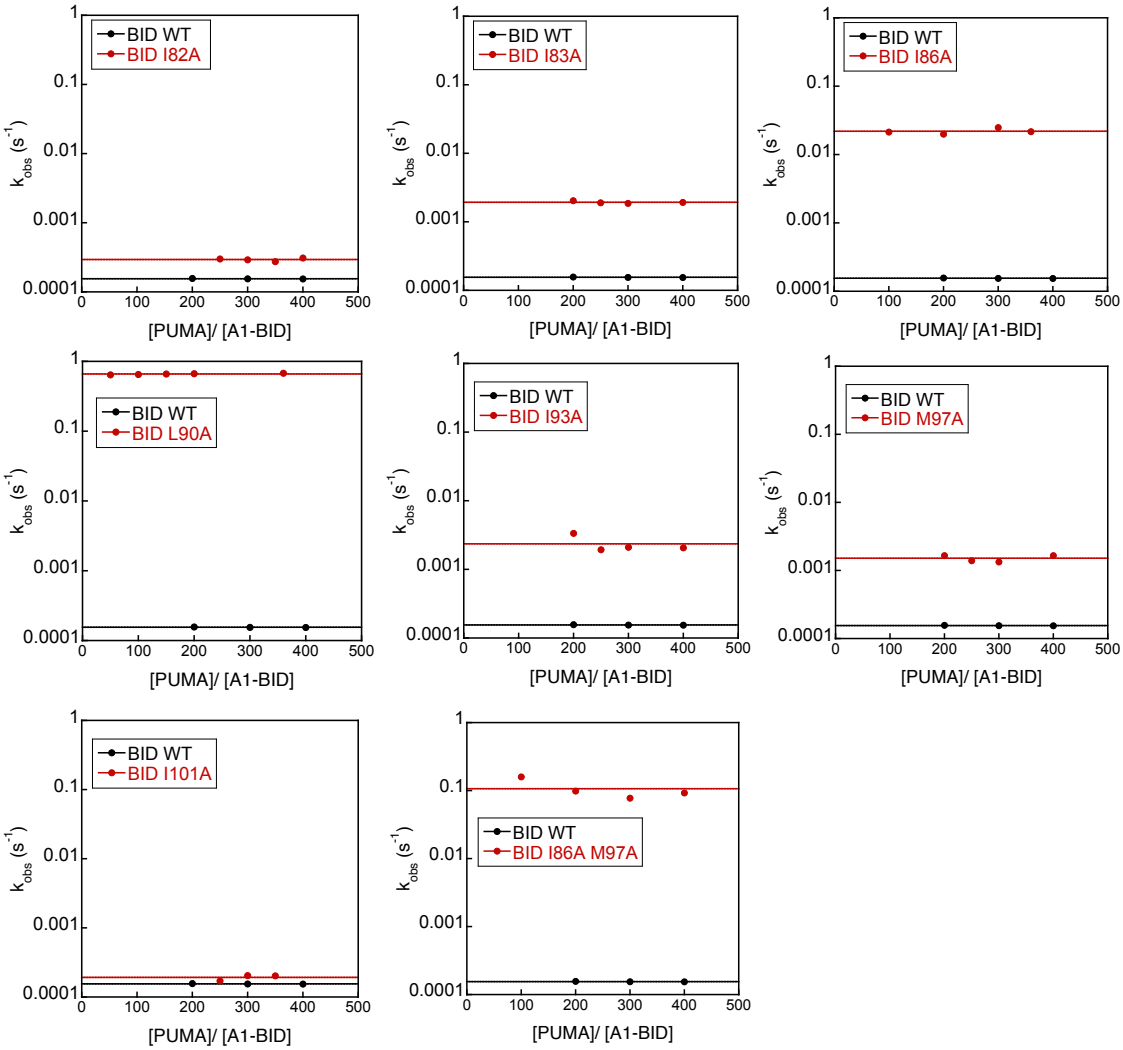


Figure 6.15: Dissociation kinetics of BID hydrophobic to alanine mutants interacting with A1. A pre-formed complex composed by BID and A1 was mixed with unlabelled PUMA peptide at different concentrations. The observed rate constant (k_{obs}) was obtained by fitting the change in fluorescence at each concentration to a single exponential. Dissociation rate constant (k_{off}) was obtained from the mean of the observed rate constants.

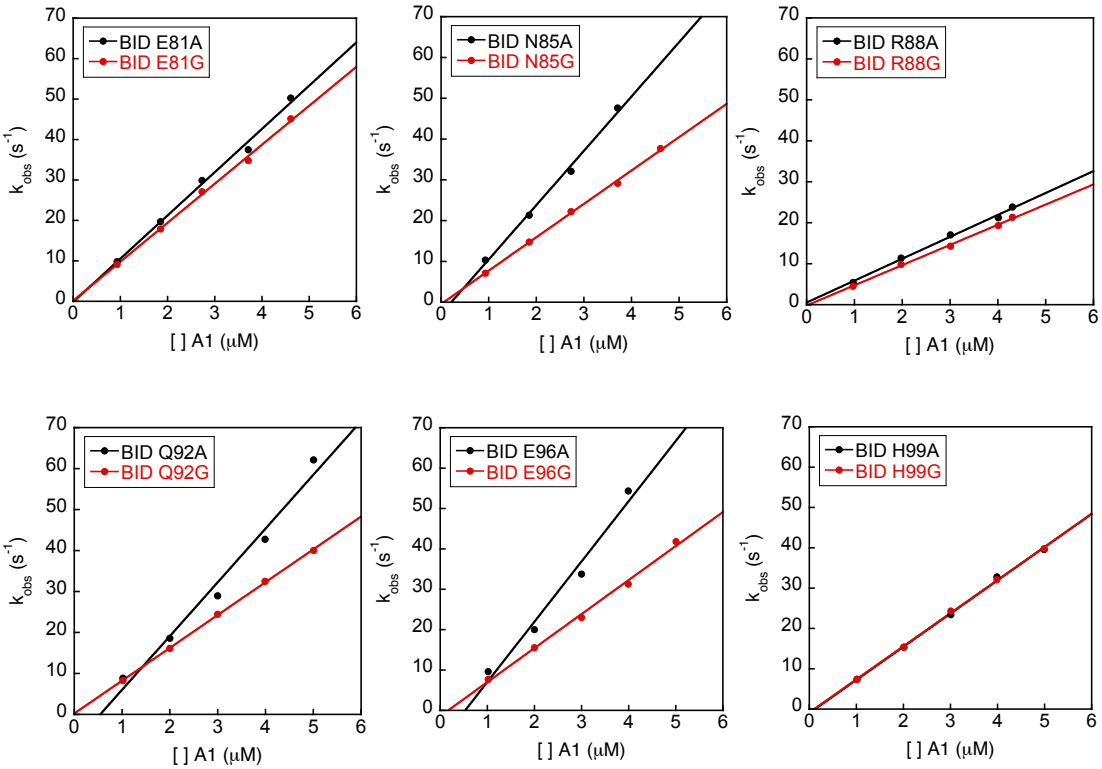


Figure 6.16: Association kinetics of BID alanine to glycine mutants interacting with A1. BID peptides were rapidly mixed (stopped flow) with an excess of A1 at different concentrations. The observed rate constant (k_{obs}) was obtained by fitting the change in fluorescence at each concentration to a single exponential. Association rate constant (k_{on}) was obtained from the gradient of the straight line fit of k_{obs} against the concentration of A1.

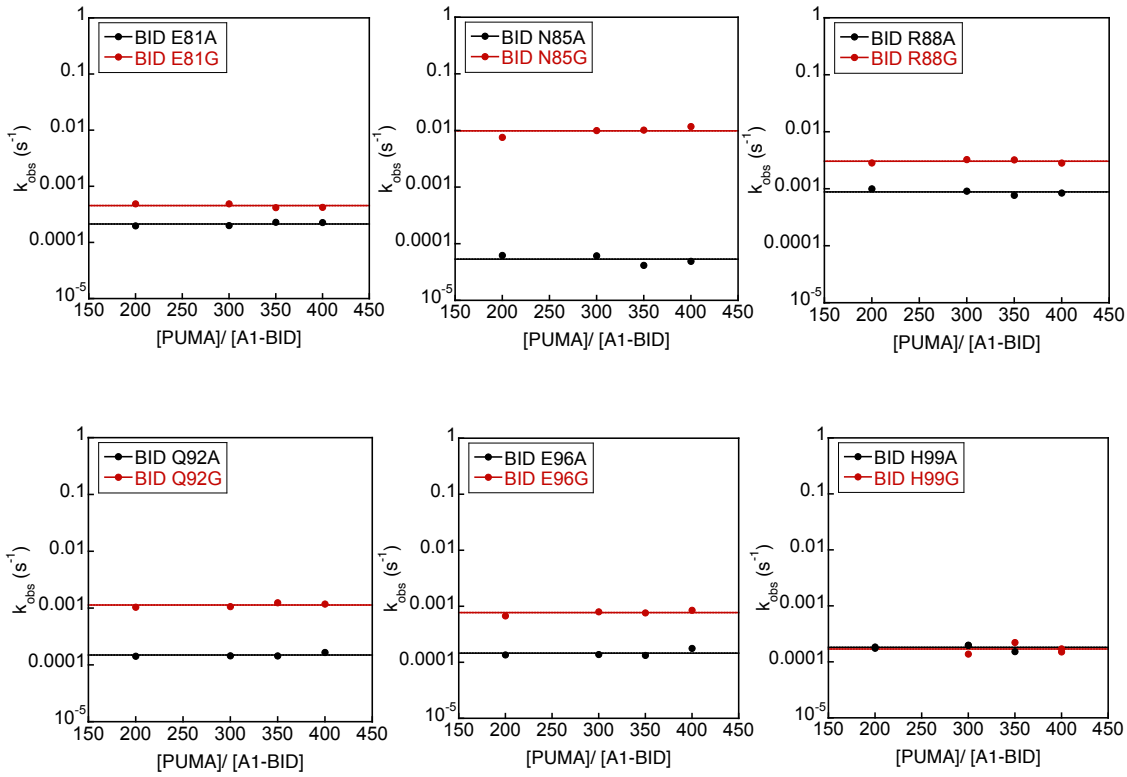


Figure 6.17: Dissociation kinetics of BID alanine to glycine mutants interacting with A1. A pre-formed complex composed by BID and A1 was mixed with unlabelled PUMA peptide at different concentrations. The observed rate constant (k_{obs}) was obtained by fitting the change in fluorescence at each concentration to a single exponential. Dissociation rate constant (k_{off}) was obtained from the mean of the observed rate constants.

In a two-state reaction, equilibrium and kinetics dissociation constants ($K_{d\ (eq)}$ and $K_{d\ (kin)}$) should be the same. To check for that hypothesis, an equilibrium binding experiment was performed for the most destabilising mutant L90A. The reason for choosing that mutation is simply because complexes of higher affinity would require very low concentrations pushing the experimental limits and increasing the errors. An equilibrium dissociation constant of 360 ± 40 nM was obtained (Figure 6.18). $K_{d\ (eq)}$ showed reasonable agreement with the kinetic dissociation constant of 210 ± 14 nM. For that reason, the Φ-values for all mutants were then calculated using the kinetic data.

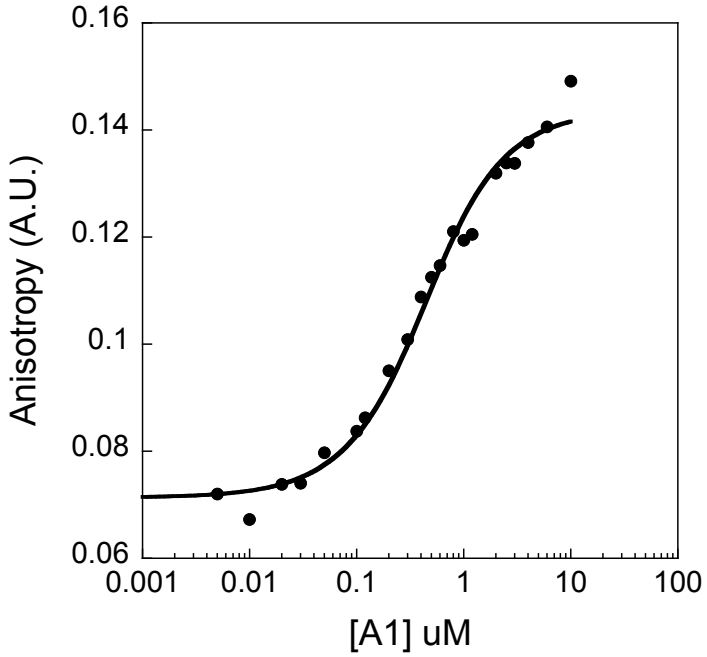


Figure 6.18: Equilibrium binding curve for BID-L90A binding A1. BID-L90A was incubated with different concentrations of A1. Data were fit to equation 3.18 (chapter 3) to obtain equilibrium K_d. Equilibrium and kinetic comparison of K_d show a two-state coupled folding binding reaction for BID-A1.

Assuming a two-state reaction, K_d can be calculated from the ratio of k_{off} over k_{on} (Equation 2.22, chapter 2). Φ-values were calculated using Equation 6.1:

$$\Phi = \frac{\ln\left(\frac{k_{on}^{WT}}{k_{on}^{Mut}}\right)}{\ln\left(\frac{K_d^{Mut}}{K_d^{WT}}\right)} \qquad \textbf{Equation } 6.1$$

Table 6.1 summarises all the biophysical parameters that characterise the coupled folding and binding of BID peptide and A1. Φ-values were very low (0 - 0.2) throughout the entire BID peptide, suggesting that the peptide is still mainly disordered at the transition state.

Table 6.1. Coupled folding and binding of BID–A1: biophysical parameters.

BID	**Helicity (CD) %**	k_{on} **(μM^{-1}s^{-1})**	k_{off} **(s^{-1}) x10^{-3}**	**n**	K_d **(k_{off}/k_{on}) (nM)**	**ΔΔ*G* (kcal.mol^{-1})**	**Φ**
WT	11.5	8.2 ± 0.2	0.155 ± 0.005	3	0.019 ± 0.001	-	-
I82A	10.7	6.8 ± 0.2	0.300 ± 0.004	4	0.004 ± 0.001	0.49 ± 0.03	0.23 ± 0.04
I83A	12.0	7.0 ± 0.2	2.0 ± 0.5	4	0.29 ± 0.01	1.47 ± 0.02	0.06 ± 0.01
I86A	9.0	2.9 ± 0.1	22.0 ± 0.5	4	7.6 ± 0.2	3.55 ± 0.03	0.17 ± 0.01
L90A	7.5	3.1 ± 0.2	660 ± 3	5	210 ± 14	5.52 ± 0.04	0.10 ± 0.01
I93A	8.9	5.6 ± 0.1	2.5 ± 0.2	4	0.44 ± 0.03	1.87 ± 0.04	0.12 ± 0.01
M97A	11.5	8.1 ± 0.1	1.50 ± 0.04	4	0.190 ± 0.006	1.35 ± 0.03	0.004 ± 0.015
I101A	10.1	8.6 ± 0.3	0.20 ± 0.01	3	0.022 ± 0.001	0.10 ± 0.03	-0.3 ± 0.3
I86A- M97A	9.0	2.2 ± 0.1	110 ± 36	4	50 ± 16	4.650 ± 0.002	-
E81A	10.0	10.7 ± 0.5	0.213 ± 0.004	4	0.020 ± 0.001	-	-
E81G	7.0	9.6 ± 0.1	0.454 ± 0.001	4	0.047 ± 0.001	0.51 ± 0.03	0.12 ± 0.05
R88A	11.3	5.3 ± 0.2	0.88 ± 0.02	4	0.16 ± 0.08	-	-
R88G	5.8	4.9 ± 0.6	3.03 ± 0.06	4	0.6 ± 0.1	0.8 ± 0.3	0.06 ± 0.09
E96A	13.0	10 ± 2	0.146 ± 0.005	4	0.015 ± 0.002	-	-
E96G	11.5	7.8 ± 0.3	0.77 ± 0.02	4	0.098 ± 0.005	1.1 ± 0.1	0.13 ± 0.04
H99A	11.6	8.2 ± 0.2	0.182 ± 0.004	5	0.002 ± 0.001	-	-
H99G	13.0	8.2 ± 0.1	0.17 ± 0.01	4	0.002 ± 0.001	-0.04 ± 0.04	-0.03 ± 0.01

Helicity of all peptides were calculated using the MRE at 222 nm and the method of Muñoz & Serrano 1995. k_{on} and k_{off} errors were obtained from error of the curve fit and from the standard error of the mean, respectively. n is number of repeats. K_d was calculated from the ratio between k_{off} and k_{on}, with the respective errors propagated using standard methods. Φ-values were calculated using the kinetic rate constants and errors represent the propagated error. For Ala-Gly mutants ΔΔG and Φ are the comparison of the Gly and the pseudo-WT Ala mutants. Table adapted from Crabtree et al. 2018.

6.4 Discussion

Intrinsically disordered proteins and intrinsically disordered regions are well represented in nature, comprising over 30% of the proteome (Ward *et al.*, 2004). They are commonly found in cell signalling mechanisms, as is the case of our family of interest, the BCL-2. IDPs biological function together with their lack of a predefined structure challenges the idea that in order to function a protein needs to fold.

Although disordered in isolation, some IDPs can fold to a stable three-dimensional structure upon binding to a partner protein. Our main interest in this project was to identify what encodes a coupled folding and binding reaction: the IDP or is it templated by the folded partner? Using the BCL-2 family as a model to study coupled folding and binding, we have compared four systems composed by two IDPs (BID and PUMA) and two folded proteins (A1 and MCL-1) using Φ-value analysis.

Φ-value analysis has been widely employed in protein folding studies to reveal the pathway by which a protein goes from unfolded to its folded state. A few IDP coupled folding and binding studies also uses Φ-value analysis in order to unravel the mechanism by which a disordered peptide folds and binds to a folded partner (Dahal et al. 2017; Dogan et al. 2013; Gianni et al. 2014; Giri et al. 2013; Hill et al. 2014; J. M. Rogers et al. 2014; Toto et al. 2014). Most studies have shown that few interactions are usually formed at the transition state of the coupled folding and binding reactions, which is in agreement with the results we got for disordered BH3-only proteins binding to folded BCL-2–like proteins.

My results on BID:A1 using Φ-value analysis revealed that BID peptide remains unstructured at the transition state upon binding to A1, with low Φ-values observed along the entire peptide ($0< \Phi <0.2$). Considering that some IDPs can fold to different structures upon binding to different partner proteins (Avalos et al. 2002; Lowe et al. 2002; Richard R. Rustandi, Baldisseri, and Weber 2000), it would be reasonable to hypothesise that coupled folding and binding reactions would be templated by the folded partner. If this hypothesis is correct, using a different partner protein should affect the transition state interactions. To challenge this hypothesis, we did a comparative Φ-value analysis using the same BID peptide binding to a different partner, MCL-1 (Figure 1.19, A).

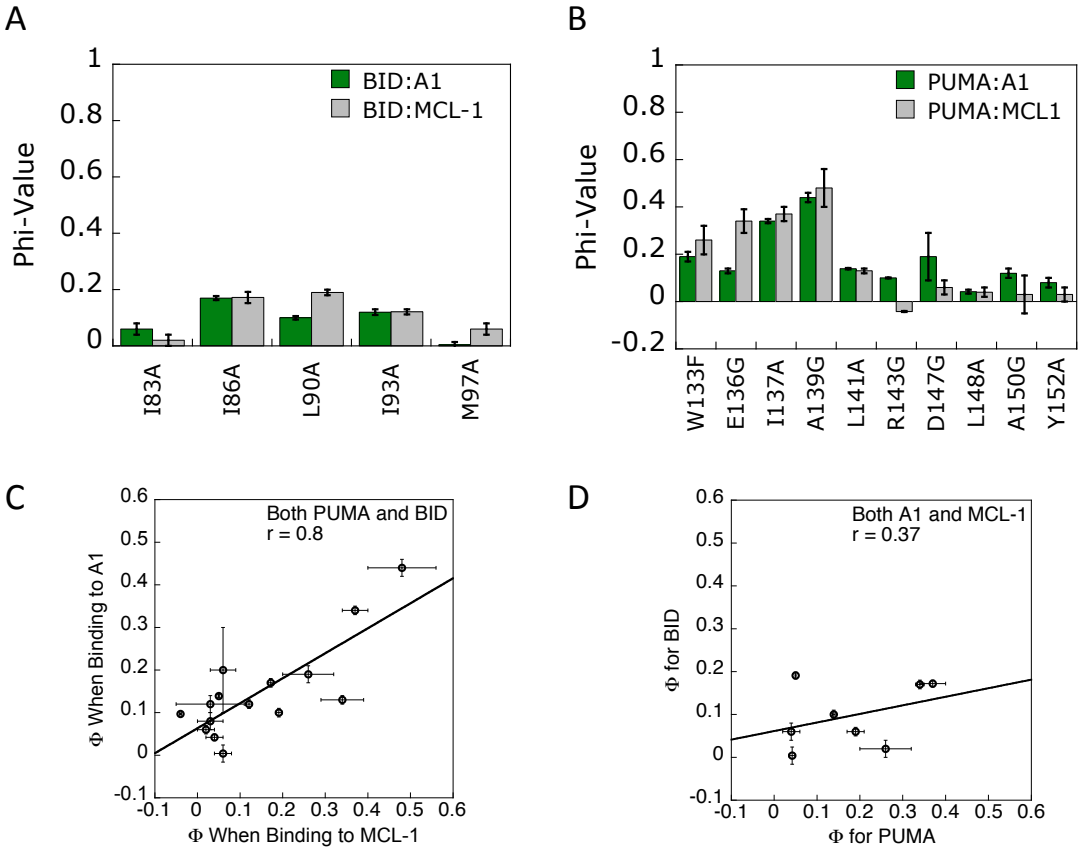


Figure 6.19: Comparative Φ-values of PUMA and BID binding to MCL-1 and A1. (A) Φ-value analysis data of BID peptides binding to the partners MCL-1 and A1. BID:MCL-1 data was taken from Quenton Bubb. (B) Φ-value analysis data of PUMA peptides binding to the partners MCL-1 and A1. Data were taken from Rogers et al. 2014 and Dr. Michael Crabtree. (C) Correlation plot of Φ-values obtained for PUMA and BID peptides when binding to MCL-1 and A1. Together with the hydrophobic mutants, Ala-Gly mutations in PUMA that destabilises both MCL-1 and A1 by >0.6 kcal.mol-1 are included in the plot. (D) Correlation plot of Φ-values obtained for MCL-1 and A1 when binding to PUMA and BID peptides. Only hydrophobic mutants are represented, as Ala-Gly mutations did not destabilise BID complexes enough to allow Φ-value calculations. Error bars represent the propagated errors. Figure adapted from Crabtree et al. 2018.

Surprisingly, the results showed a very good agreement for all mutants. In both cases, BID remains disordered at the transition state upon binding to the partner protein, showing low Φ-values throughout the entire peptide sequence. These results indicated that the disordered peptide BID is responsible for encoding the transition state upon binding to a partner protein.

Interestingly, the results for PUMA peptide binding to A1 (work done by Dr. Michael Crabtree) show a similar Φ-value pattern to PUMA binding to MCL-1 (work done by Dr. Joseph Rogers), with intermediate Φ-values ($0.2< \Phi <0.6$) observed at the N-terminal of the PUMA peptide whereas low Φ-values ($0< \Phi <0.2$) are observed along rest of the peptide

sequence (Figure 1.19, B). Most importantly, the results show that the pattern of Φ-values is maintained for the peptides PUMA and BID when binding to MCL-1 and A1. A strong correlation of Φ-values is observed for both peptides when binding to MCL-1 and A1 (Figure 6.19, C). In contrast, a correlation is not observed when we plot Φ-values obtained for MCL-1 and A1 when binding to PUMA and BID peptides. (Figure 6.19, D). The results support the conclusion that the IDPs encode the interactions of the transition states.

Simulations of the unbound PUMA suggest that although mainly disordered, the peptide displays a higher degree of helicity at its N-terminus (Harmon et al. 2016), whereas CD data show that BID peptide is largely disordered in the unbound state. Both results are in agreement with the transition states structures findings: PUMA is partially structured with higher Φ-values at the N-terminal region whereas BID remains disordered. That might indicate that the residual structure of the IDP in the unbound state can determine the transition state interactions upon binding to a protein partner. IDPs encode their level of residual structure, and our study shows that they have also the potential to encode the transition state structures in a coupled folding and binding reaction. The ability of controlling residual structure and encoding transition state interactions provides a fine way of how evolution can modulate the affinity of the IDP and partner protein complexes, which could be one good reason for the strong conservation of disorder on the proteome.

7 Coupled folding and binding in a full-length IDP context

7.1 Introduction

7.1.1 The role of IDP sequence context on its ability to bind and interact

Natively unfolded or intrinsically disordered proteins (IDPs) lack a defined structure but nonetheless represent a significant proportion of the proteome (Ward *et al.*, 2004). They play significant roles in transcription regulation, DNA binding, intracellular transportation etc, where they are commonly involved in networks comprised by multiple partners (Ward *et al.*, 2004; Dunker *et al.*, 2005).

Although the discovery of IDPs is relatively new, a significant number have been reported over the past 20 years (Uversky 2002). A great proportion of experimental studies investigate IDPs as peptides or shorter versions of the whole protein sequence, usually the binding motif. However, these motifs are often part of larger proteins.

Even though studying IDPs as peptides is a much simpler way to work and can offer significant insights on these intriguing molecules, it has already been shown that the sequence context of the binding motif could affect IDPs properties. For instance, mutations on conserved proline residues located in flanking regions of IDPs were shown to modulate their binding affinities for partner proteins by controlling the lifetime of the IDP/protein complex and its *in vivo* function (Borcherds et al. 2014; Crabtree et al. 2017). Furthermore, the coupled folding and binding of cMyb and KIX is altered when cMyb is extended by the addition of the transactivation domain TAD (Shammas et al. 2014, Arai et al. 2015).

On the previous chapter, we demonstrated that IDPs encode coupled folding and binding reactions with its partner proteins, as their sequences contain the necessary information for complex formation: same IDP binds to different partner proteins following the same

pathway (Crabtree et al. 2018). The research was performed using the binding motif (35 residues) of the IDPs but they all exist in a longer context. The study opened new ideas on how to explore the same IDP-partner complex using the IDP in its full-length form. That would allow us to shed some light onto how the full-length context compares to what we know for the peptides.

7.1.2 BID as a model of an IDP in its full-length context

The BCL-2 family is well known to be involved in cell apoptosis comprising a complex network between pro-apoptotic, pro-survival and "initiator" proteins. The precise mechanism by which apoptosis occurs is still not fully understood (L.Omonosova and C.Hinnadurai 2008). That makes the BCL-2 a very intriguing and interesting family of proteins to be explored. Protein members of the BCL-2 family all have in common the presence of a BH3 (BCL-2 homology 3) motif. Pro-apoptotic (e.g BAK, BAX, BOK) and pro-survival (e.g MCL-1, A1, BCL-XL) are multi domain proteins, whereas the stimuli activators or "initiators" (e.g. PUMA, tBID, BIM, BAD) are BH3-only proteins and mostly disordered (Czabotar et al. 2014).

A potential candidate to be studied in a full-length context is the previously studied BID. In contrast to all other "initiators" proteins of the BCL-2 family, full-length BID (BID_{FL}) has a defined globular structure with eight helices. Helix 3 contains the BH3 motif and it is connected to helices 1 and 2 by a long flexible loop, which contains a caspase cleavage site (Figure 7.1). Despite the low sequence identity, the structure of BID_{FL} is very similar to the other BCL-2 family proteins, like BAX and MCL-1 (Petros et al. 2004, Fesik 2000).

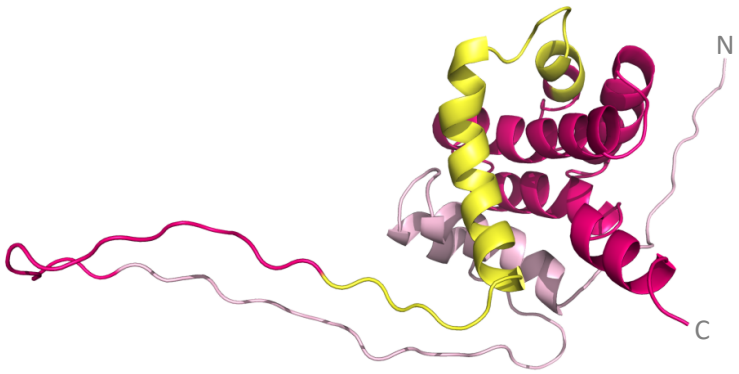


Figure 7.1: BID_{FL} structure (PDB 1DDB). BID_{FL} structure showing N- and C-terminus. The structure shows the first 60 residues that are cleaved in tBID in light pink, the 35 residues that comprise the BH3 motif are shown in yellow and the rest of the molecule is shown in dark pink. The 35 amino acids long fragment was previously used as a peptide to study coupled folding and binding (Chapter 6).

In the cell, BID is cleaved by caspase 8, where 60 N-terminal residues are removed generating the truncated BID (tBID) (Figure 7.2). tBID plays a key role in the cell death pathway, amplifying the apoptotic response which is required for an effective apoptosis (Czabotar et al. 2014). Although BID_{FL} is located in the cytosol, tBID translocates to the mitochondria where it signals apoptosis (Li et al. 1998).

```
BIDfl       MDSEVSNGSGLGAEHITDLLVFGFLQSSGCTRQELEVLGRELPVQAYWEADLEDELQTDG 60
tBID        ------------------------------------------------------------ 0

BIDfl       SQASRSFNQGRIEPDSESQEEIIHNIARHLAQIGDEMDHNIQPTLVRQLAAQFMNGSLSE 120
tBID        SQASRSFNQGRIEPDSESQEEIIHNIARHLAQIGDEMDHNIQPTLVRQLAAQFMNGSLSE 60
            ************************************************************

BIDfl       EDKRNCLAKALDEVKTAFPRDMENDKAMLIMTMLLAKKVASHAPSLLRDVFHTTVNFINQ 180
tBID        EDKRNCLAKALDEVKTAFPRDMENDKAMLIMTMLLAKKVASHAPSLLRDVFHTTVNFINQ 120
            ************************************************************

BIDfl       NLFSYVRNLVRNEMD   195
tBID        NLFSYVRNLVRNEMD   135
            ***************
```

Figure 7.2: Sequence alignment of BID_{FL} and tBID. BID_{FL} is N-terminally truncated by caspase 8, where 60 residues are removed generating tBID. Alignment was produced using Clustal Omega (Goujon et al. 2010, Sievers et al. 2011).

Even though tBID preserves the secondary helical structure of BID_{FL}, it adopts an α-helical but dynamically disordered conformation in solution (Gong et al. 2004, Wang & Tjandra

2013, Yao et al. 2009). Perhaps, that might facilitate the interaction between BID and the pro-apoptotic and pro-survival proteins, as its BH3 motif will be exposed after cleavage. Studying BID_{FL} and tBID would possibly give us insights on why nature chooses this pathway for the apoptosis network trigger and function, as well as gain insights into the contribution of the protein context to the interaction profile of the binding motif.

7.1.3 Fluorescent residues in BID_{FL} and tBID

BID_{FL} contains two tryptophan and one tyrosine residues (Figure 7.3). After truncation by caspase and the removal of $BID_{1\text{-}60}$ fragment, one tryptophan and one tyrosine located in the flexible loop are deleted from BID, leaving tBID with one tyrosine only. Biophysical experiments of BID_{FL} are performed following the tryptophan fluorescence, whereas tyrosine was used for tBID.

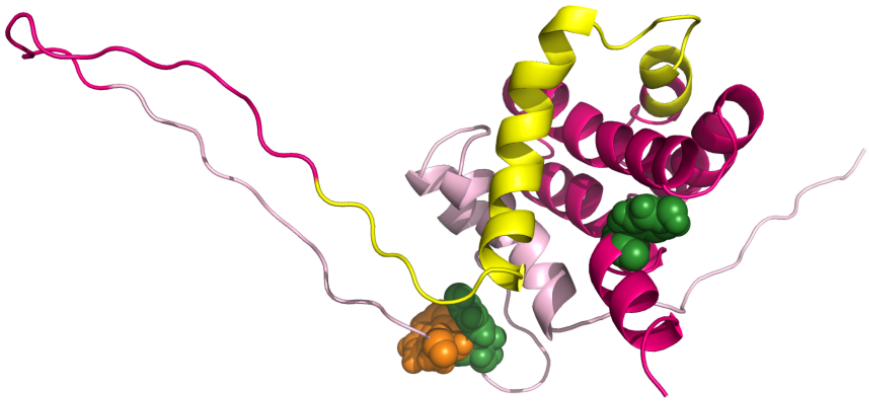


Figure 7.3: BID_{FL} structure (PDB 1DDB) showing aromatic residues. The structure shows the position of the aromatic residues represented as spheres: one tryptophan (orange) and two tyrosines (green). The tryptophan and one tyrosine are located in the flexible loop, that gets truncated on tBID (light pink), whereas the other tyrosine is positioned on helix 8.

7.1.4 Chapter aims

The majority of the studies on IDPs are performed using peptides, which are the binding regions of a larger protein. This chapter aims to investigate an IDP in its full-length context to shed some light on how that affects its ability to form complexes with partner proteins and how it modulates the complex affinity. For that reason, I have chosen to study BID, which is active as a BH3-only protein tBID but is a member of the BCL-2 family that is structured in a full-length context.

7.2 Results

7.2.1 Design and production of BID and tBID

Mouse BID_{FL} and tBID genes were purchased from GenScript in the standard pUC57 vector. A molecular biology strategy was designed to insert both genes into pGEX-4T-3 plasmid. This plasmid was previously used for A1, which made the purification protocol very similar to the one I was already using. Primers were designed in order to extract BID_{FL} and tBID genes from their original pUC57 vector and to remove A1 gene from pGEX-4T-3. The primers for BID/tBID contained extra base pairs in both ends that were complementary with the desired pGEX-4T-3 plasmid. A ligation was performed as a last step in order to insert the protein genes into the vector. The choice to work with mouse over human BID was to keep the consistency with the previous work (chapter 6) and allow comparison between the results.

Once the molecular biology steps were finished and we had both BID and tBID in the pGEX-4T-3 plasmid, we wanted to develop the protocol for the proteins expression and purification. In order to find the optimal condition for BID_{FL} and tBID expressions, different parameters were tested and varied: three different OD_{600} upon induction (0.4, 0.6 and 0.8), two different IPTG concentrations (0.1 and 1 mM) and two temperatures after IPTG induction (25° and 37° C). Results for the different conditions for the two proteins are shown in Figure 8.4. From the expression trials, it became quite clear that tBID expression was really poor. It was thought that the unstructured protein is likely to be insoluble so we decided to express only BID_{FL} and use it to obtain tBID with the caspase-8 cleavage later on. For BID_{FL}, we found as an optimal condition to induce expression with 0.1 mM IPTG at OD_{600} 0.6-0.8 and incubate it at 37° C for 6 hours after induction (Figure 7.3).

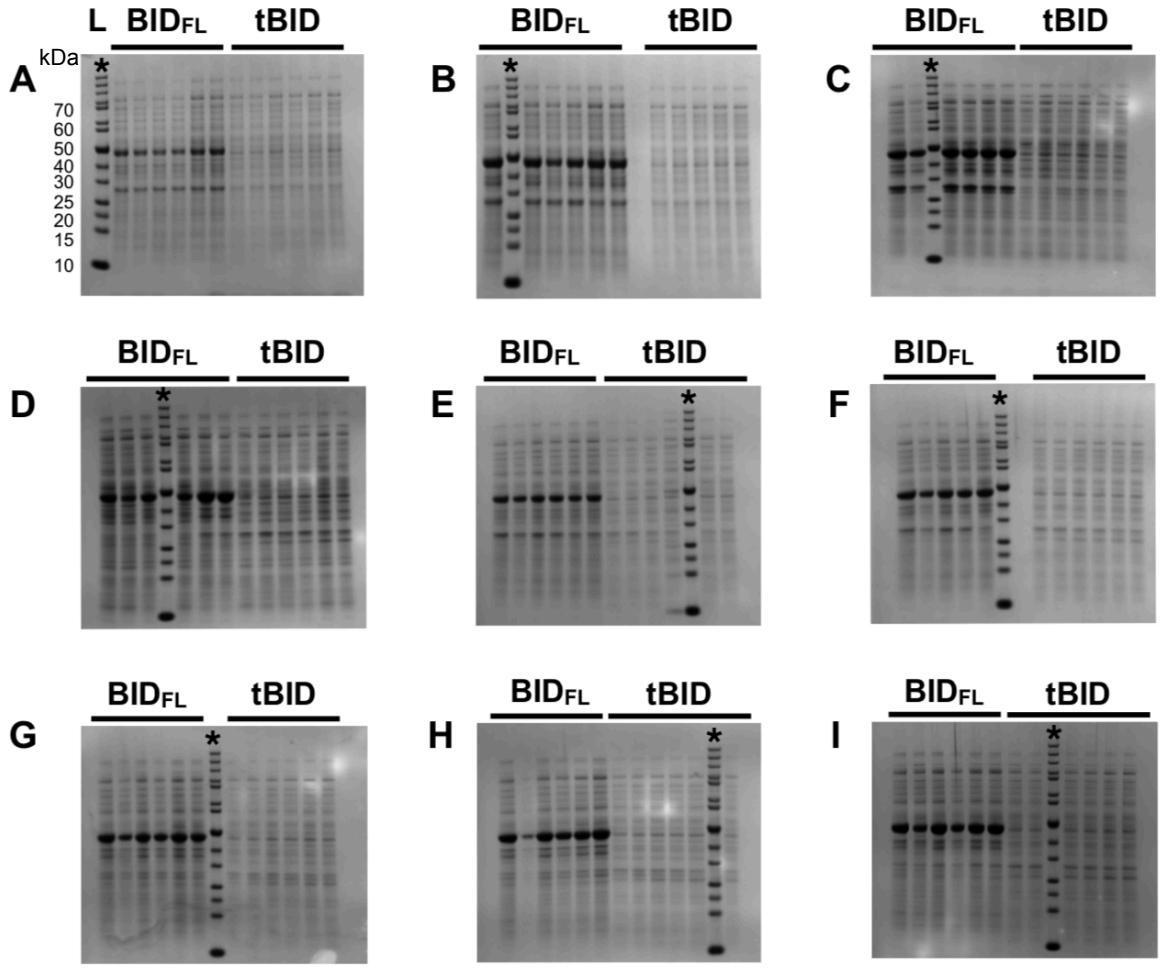


Figure 7.4: Expression trials of BID_{FL} and tBID. Different parameters were tested in order to find the optimal expression protocol for the two protein constructs: three different OD_{600} (0.4, 0.6 and 0.8), two different IPTG concentrations (0.1 and 1 mM) and two temperatures after induction (25° and 37° C). Each gel contain six samples of BID_{FL} and tBID in the different conditions: 1) OD_{600} = 0.4, IPTG = 0.1 mM, 2) OD_{600} = 0.4, IPTG = 1 mM, 3) OD_{600} = 0.6, IPTG = 0.1 mM, 4) OD_{600} = 0.6, IPTG = 1 mM, 5) OD_{600} = 0.8, IPTG = 0.1 mM, 6) OD_{600} = 0.8, IPTG = 1 mM. Gels represent the different incubation temperature and times after IPTG induction: (A) 25° C, 4 hours; (B) 25° C, 6 hours; (C) 25° C, overnight; (D) 37° C for 6 hours, then 25 ° C overnight; (E) 37° C, 2 hours; (F) 37° C, 3 hours; (G) 37° C, 4 hours; (H) 37° C, 5 hours (I) 37° C, 6 hours. PageRuler unstained standard ladder is shown in (A) for comparison (in all gels). The trials show a very poor expression of tBID for all the conditions. The optimal condition for BID_{FL} expression was to induce with 0.1 mM IPTG at OD_{600} 0.6-0.8 and to incubate it at 37° C for 6 hours after induction (gel I).*

BID_{FL} was expressed as a GST fusion protein and the detailed purification protocol is described in Chapter 3 (Materials and Methods). In the final step of purification, using a Superdex 75 gel filtration column, we noticed that extra peaks of higher molecular weight appeared before the main BID_{FL} one (Figure 7.5, A). SDS gel of the fractions confirm the expected BID_{FL} mass and purity not only for fractions on the main peak, but also the ones of higher mass (Figure 7.5, B). These results suggest that that BID_{FL} is capable of homooligomerisation.

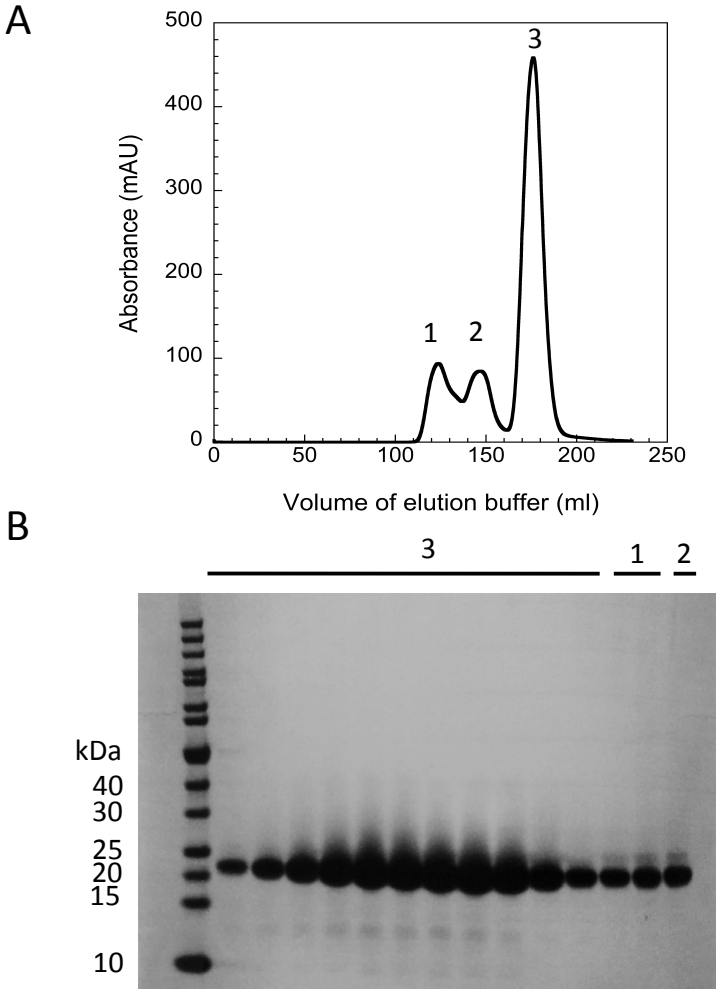


Figure 7.5: Purification of BID_{FL}. (A) Chromatogram image after BID_{FL} G75 size exclusion run showing the peaks 1, 2 and 3. (B) SDS gel of BID_{FL} fractions after G75 purifications and NuPage unstained protein ladder for comparison. Gel starts from marker and show fractions from peak 3, followed by peaks 1 and 2. Pure BID_{FL} can be identified in all fractions, suggesting that the larger species (first and second peaks on the chromatogram) are also formed by BID_{FL}.

Purified BID_{FL} samples were sent for mass spectroscopy analysis at the Chemistry department to confirm its identity and purity. All three samples results confirmed the expected mass for BID_{FL}. Figure 7.6 show one chromatogram as an example.

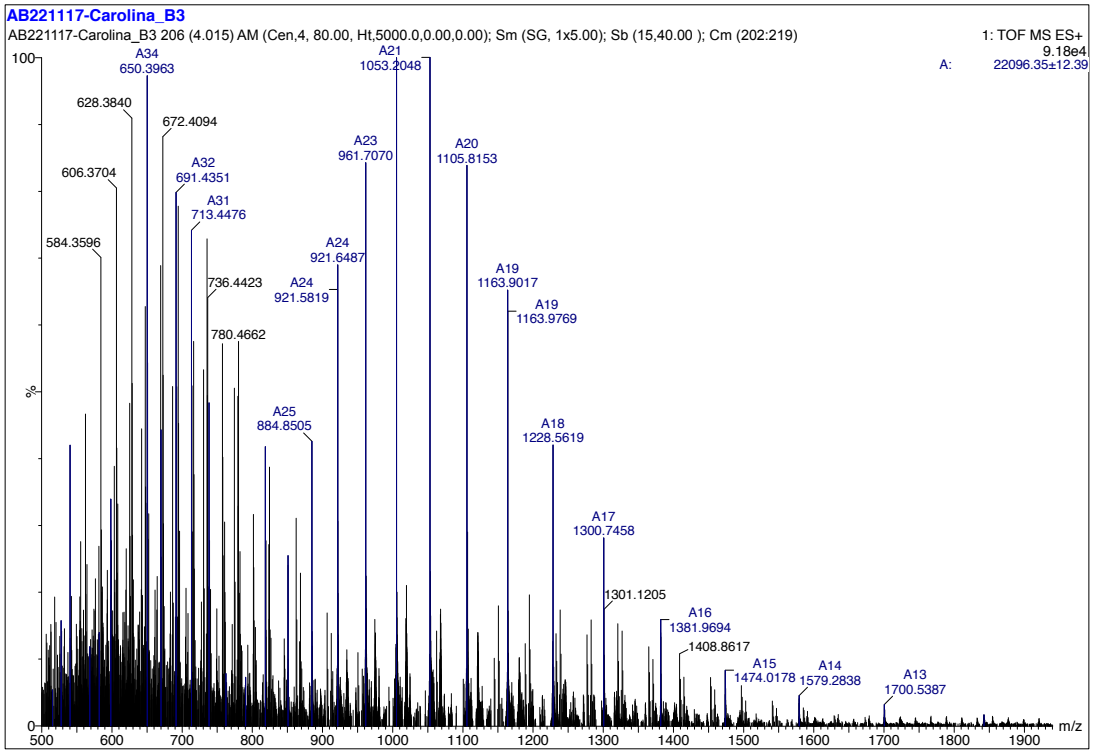


Figure 7.6: Mass spectroscopy results for BID_{FL}. Mass spectroscopy analysis of the final BID_{FL} sample in 50 mM sodium phosphate, pH 7.0 (0.05% Tween-20) provided a mass of 22,096±12.39 Da, which is within error of the expected mass of 22,095 Da.

For accurate concentration calculations, samples of purified BID_{FL} were sent for Amino Acid Analysis (AAA) at the Biochemistry department, performed by Peter Sharratt. The results allowed us to calculate the experimental extinction coefficient (ε_{280}) (Figure 7.7). The experimental and the ProtParam theoretical values were slightly different, 8700 and 8480 $M^{-1}.cm^{-1}$ respectively. Throughout this chapter, BID_{FL} concentrations were calculated using the experimental extinction coefficient value (ε_{280} = 8700 $M^{1}.cm^{-1}$).

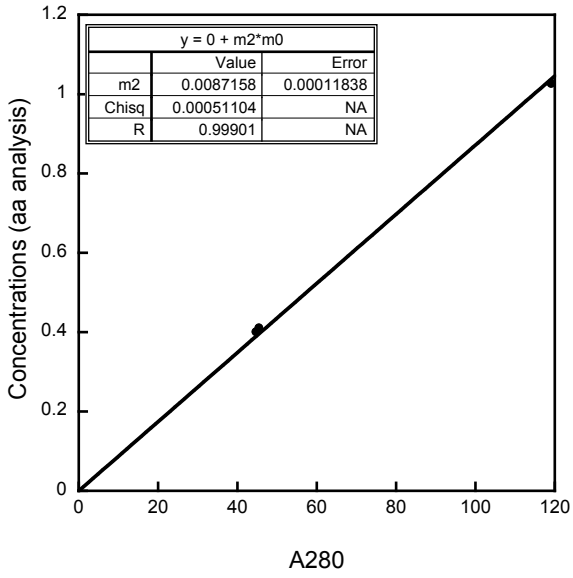


Figure 7.7: Extinction coefficient for BID_{FL}. Extinction coefficient value calculated from amino acid analysis (AAA). Concentration results for three different samples of BID_{FL} plotted against its absorbance readings at 280 nm. ProtParam expected value was ε=8480 $M^{-1}.cm^{-1}$ whereas the experimental one was ε=8700 $M^{-1}.cm^{-1}$. The experimental value was used for all BID_{FL} concentration calculations in this chapter.

Production of BID_{FL} was somehow straightforward whereas tBID was not. Therefore, the latter was obtained by cleaving BID_{FL} with caspase-8 after it was produced. Even though the cleavage was successful (as confirmed by SDS-PAGE), the N-terminal 60-residue fragment (BID_{1-60}) was still bound to the rest of the molecule (as observed by SEC). Hence, an extra step was added to the purification protocol in order to separate tBID from BID_{1-60}. After cleavage with caspase-8, the protein was incubated in 8 M urea and unfolding of the protein allowed tBID and BID_{1-60} to be separated by size exclusion chromatography (Figure 7.8). Pure tBID was dialysed back into buffer to remove denaturant. That is an unexpected and interesting result demonstrating that even though caspase-8 is able cleave BID_{FL}, this event does not sufficiently destabilise the protein to result in the separation of tBID from the fragment.

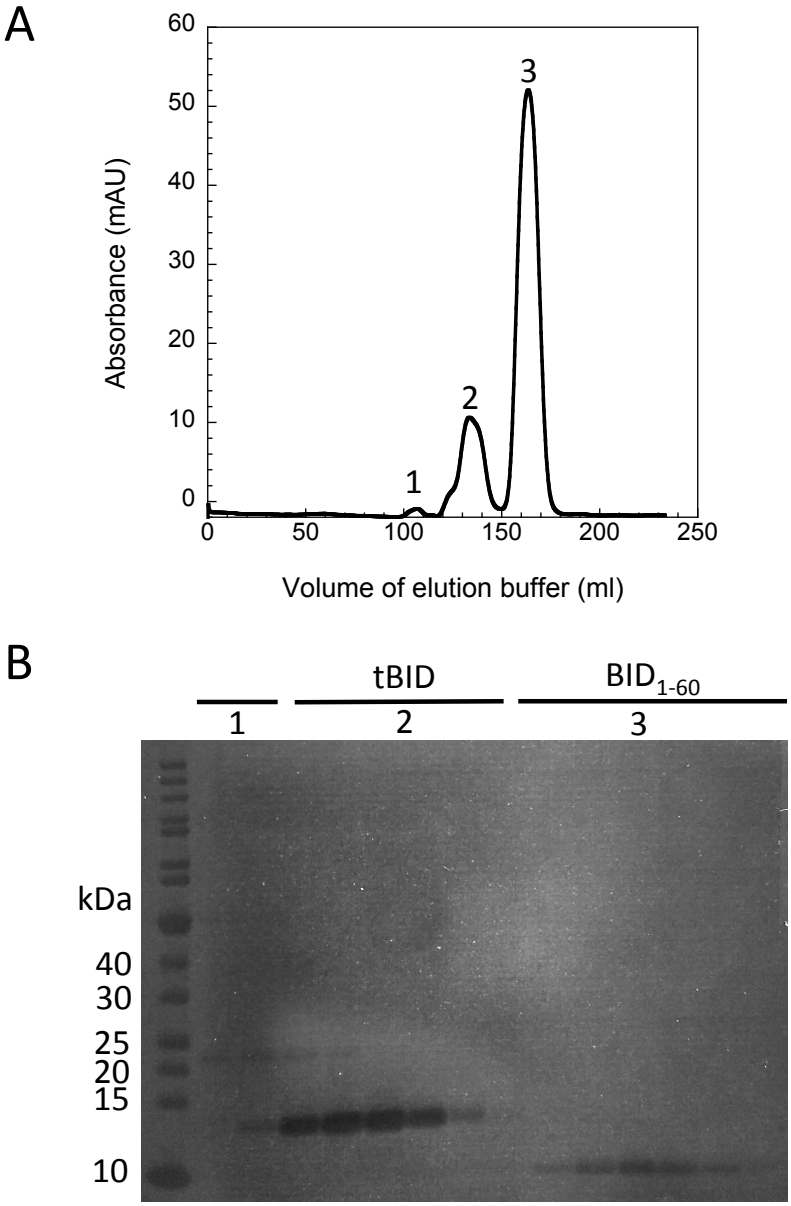


Figure 7.8: Purification of tBID after unfolding in denaturant. (A) Chromatogram image after G75 size exclusion run in 8 M urea showing peaks 1, 2 and 3. (B) SDS gel of tBID fractions after G75 purification in urea and NuPage unstained protein ladder for comparison. Gel starts from marker and shows fractions from peak 1 (uncleaved BID_{FL}), 2 (tBID) and 3 (BID_{1-60} fragment).

7.2.2 Stability of BID_{FL}

The stability of BID_{FL} was measured by urea-induced equilibrium denaturation. The experiment was performed in the fluorimeter with BID_{FL} in 50 mM sodium phosphate, pH 7.0 (0.05% Tween-20) and 10 mM DTT, at 25 °C. Tryptophan fluorescence was followed with the excitation wavelength of 280 nm and emission recorded from 300 to 400 nm. Data were analysed calculating the fluorescence at 356 nm and the results were fitted to a two-state model (Figure 7.9). Although BID_{FL} has a flexible unfolded loop on its structure, its stability (12.9 kcal.mol^{-1}) is very similar to other BCL-2 family proteins, MCL-1 and A1 (≈ 12 kcal.mol^{-1}) for instance.

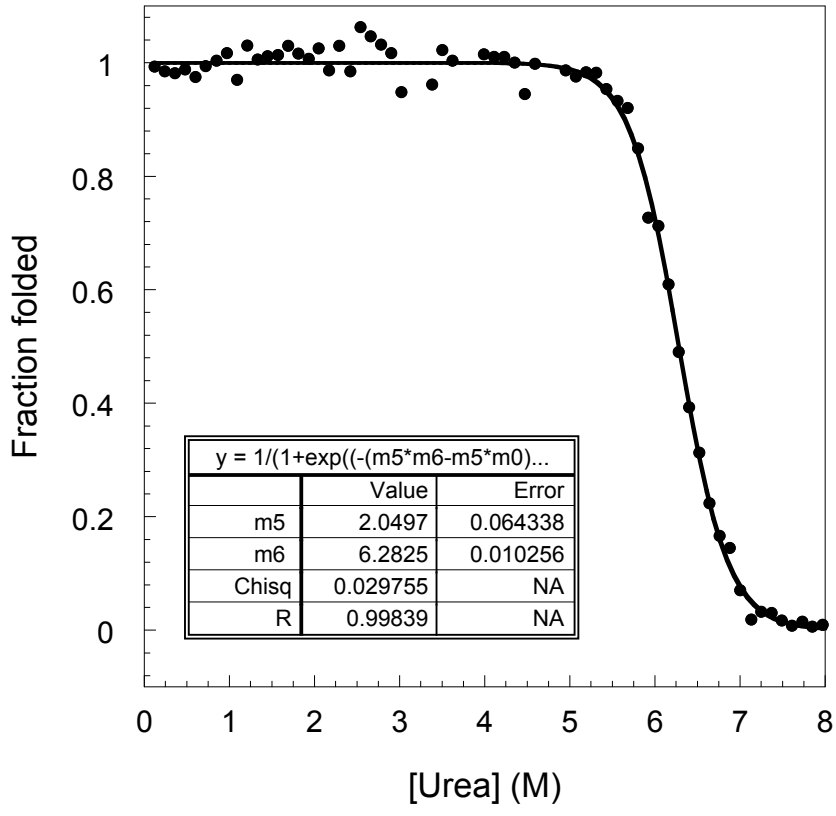


Figure 7.9: Equilibrium curve of BID_{FL}. Urea-induced equilibrium denaturation curve of BID_{FL} in 50 mM sodium phosphate, pH 7.0 (0.05% Tween-20) with 10 mM DTT.

7.2.3 CD data of BID and tBID

CD data was collected during the purification steps of tBID as well as for pure BID_{FL}. After incubation of BID_{FL} with caspase-8 and the first run of size exclusion chromatography (SEC), CD spectra overlay with the one from BID_{FL} (Figure 7.10, A) suggesting that the protein has the same structure before and after cleavage. Together with the SEC and SDS PAGE results, CD result agrees that the fragment $BID_{1\text{-}60}$ is still present with tBID after BID_{FL} is cleaved by caspase.

The following stage was to unfold the cleaved BID, so it was incubated in denaturant and then another SEC run separated tBID and fragment. Although tBID is mostly disordered, CD spectra show a recovery in residual structure between tBID in urea and after dialysis, in buffer (Figure 7.10, B). That result demonstrates the tBID ability to recover residual structure upon denaturant removal. Figure 7.10, C shows a comparison between BID_{FL} and tBID CD spectra. Although BID_{FL} contains a long flexible loop on its structure, it has similar CD spectra when compared with other previously studied BCL-2 family proteins like A1 (Chapter 6). In contrast, tBID is largely disordered, as are the other BH3-only proteins such as PUMA.

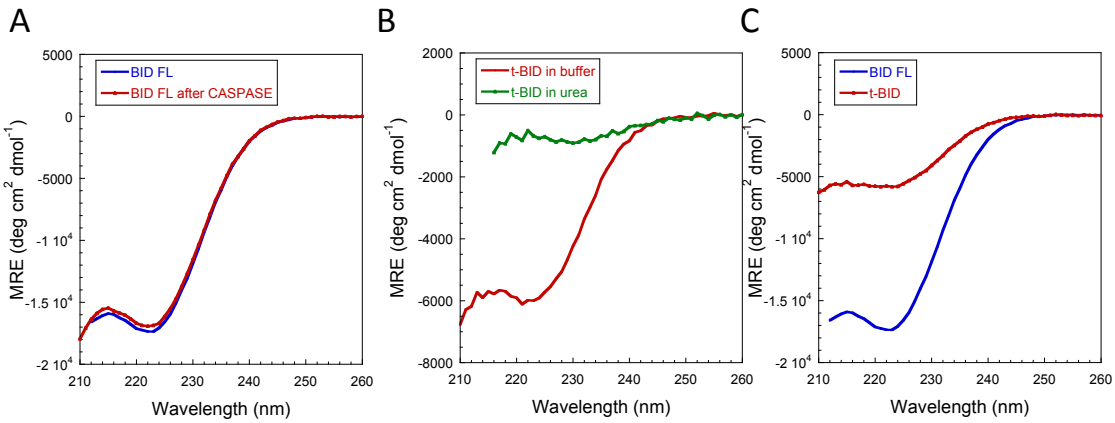


Figure 7.10: CD results during the purifications steps of tBID. (A) CD spectra show the same residual structure of purified BID_{FL} before and after cleavage with caspase-8. (B) tBID after the final purification step using urea shows a recovery in residual structure after dialysis to remove the denaturant. (C) Comparison between purified BID_{Fl} and tBID.

7.2.4 Oligomerisation of BID_{FL}

BCL-2 family oligomerisation propensity

The BCL-2 family proteins mechanism of cell apoptosis is being extensively discussed and studied over the past years. From the biological point of view, it is known that pro-apoptotic family members like BAK and BAX are the cell death executioners (Dewson & Kluck 2009, Peña-Blanco & García-Sáez 2018). Once the signal for apoptosis is triggered, these proteins are able to form pores in the mitochondrial outer-membrane and release its inter-membrane contents, activating caspases resulting in cell death. Thus, control of pro-apoptotic proteins oligomerisation is believed to be key in apoptosis regulation by the BCL-2 family (Czabotar et al. 2014, (Kale, Osterlund, and Andrews 2018). Although the relevance of BAK/BAX oligomerisation is appreciated, its structures and mechanism are not yet fully understood.

The fact that many PhD students' projects in the lab over the last few years involved the BCL-2 family, allowed us to gain significant insights and gave us an unique chance to compare results between different protein members of the family as they came out of the instruments and were analysed. Dr. Basile Wicky worked with pro-apoptotic BAK and BAX and found out that neither of them forms oligomers under buffer conditions: not spontaneously, nor in the presence of BH3-only peptides. Interestingly, using detergents to provide a membrane-like environment, BAK and BAX oligomerise *in vitro* in a spontaneous manner, without requiring the presence of any other BCL-2 protein (Wicky, 2018). With the focus on the oligomerisation properties of BCL-2 proteins, Dr. Michael Crabtree spotted some interesting and intriguing results: equilibrium denaturation curves of pro-survival BCL-XL and A1 proteins, showed two transitions. Further investigation showed that this biphasic behaviour under denaturing conditions could be explained by oligomers formation, with the first transition indicating monomer to oligomer formation and the second representing dissociation of the oligomer into unfolded monomer. This behaviour was not reproduced by MCL-1, for example, another pro-survival protein. When I first purified BID_{FL} I could not ignore the intriguing result from its purification: together with the protein's main peak, other peaks of higher molecular weight could be seen in the chromatogram. SDS gel revealed that those fractions of higher order species were only

composed by BID_{FL}, again a strong indication of oligomers formation. Michael and Basile also observed a similar behaviour when purifying BCL-XL. The main oligomer peak was in fact a dimer, confirmed by analytical ultracentrifugation (AUC).

BID_{FL} sequence contains two cysteines (Figure 7.11) that could be a sensible reason for multimerisation under oxidising conditions. In the absence of a reducing agent, although not huge, oxidation over time would be expected. During the BID_{FL} purification steps, as well as in all experiments, DTT was added (10 mM final concentration) to avoid oligomers formation. Thus, the higher species observed could not be explained by disulphide bond formation.

Figure 7.11: BID_{FL} structure (PDB 1DDB) showing cysteine residues. The structure shows the position of the two cysteine residues as blue spheres: one located in the loop between helices 1 and 2 and the other on helix 5.

Cross-linking of BID_{FL}

Although the oligomerisation was not the focus of this project, I could not ignore this interesting result. As well as pro-apoptotic BAK and BAX and pro-survival BCL-XL and A1, BID in a full-length context was also able to form higher order species. The oligomeric state was investigated using chemical cross-linking.

The experiment was carried out by using two crosslinkers: EDC (1-ethyl-3-[3-dimethyllaminopropyl]carbodiimide hydrochloride) and BS^3 (bis[sulfosuccinimidyl) suberate). This dual composition was chosen to allow short as well as longer length cross species formation. EDC is known as a "zero length" crosslinker as its reaction forms new bonds between carboxyls and amides without incorporation of any EDC's atoms. It reacts with a carboxyl group forming an O-acylisourea and this intermediate rapidly reacts with an amino group. This reaction results in the formation of a new amide bond, or in other words, the proteins cross-linking. Although EDC allows cross-linking without adding a spacer group, it requires the two amino acids to be in close proximity, and that is why BS^3 came into use. BS^3 is a N-hydroxysuccinimide (NHS) ester with an 8-atom spacer arm (11.4 Å). This reagent is specific for primary amines on the side chain of lysine residues and O-acylisoureas (formed by EDC reaction). The use of EDC and BS^3 together allows cross-linking of lysine with glutamate and aspartate residues that are located far from each other.

Crossliking gel results showed formation of dimers, trimers and tetramers of BID_{FL}, confirming the preliminary indication from SEC peaks (Figure 7.12). It is important to highlight that chemical cross-linking is not an equilibrium reaction, which makes the bands distribution on the gel not a direct association with the oligomers present in solution. Therefore the gel result does not necessarily represent the distribution observed by SEC, but it shows the capability of BID to form up to tetrameric oligomers.

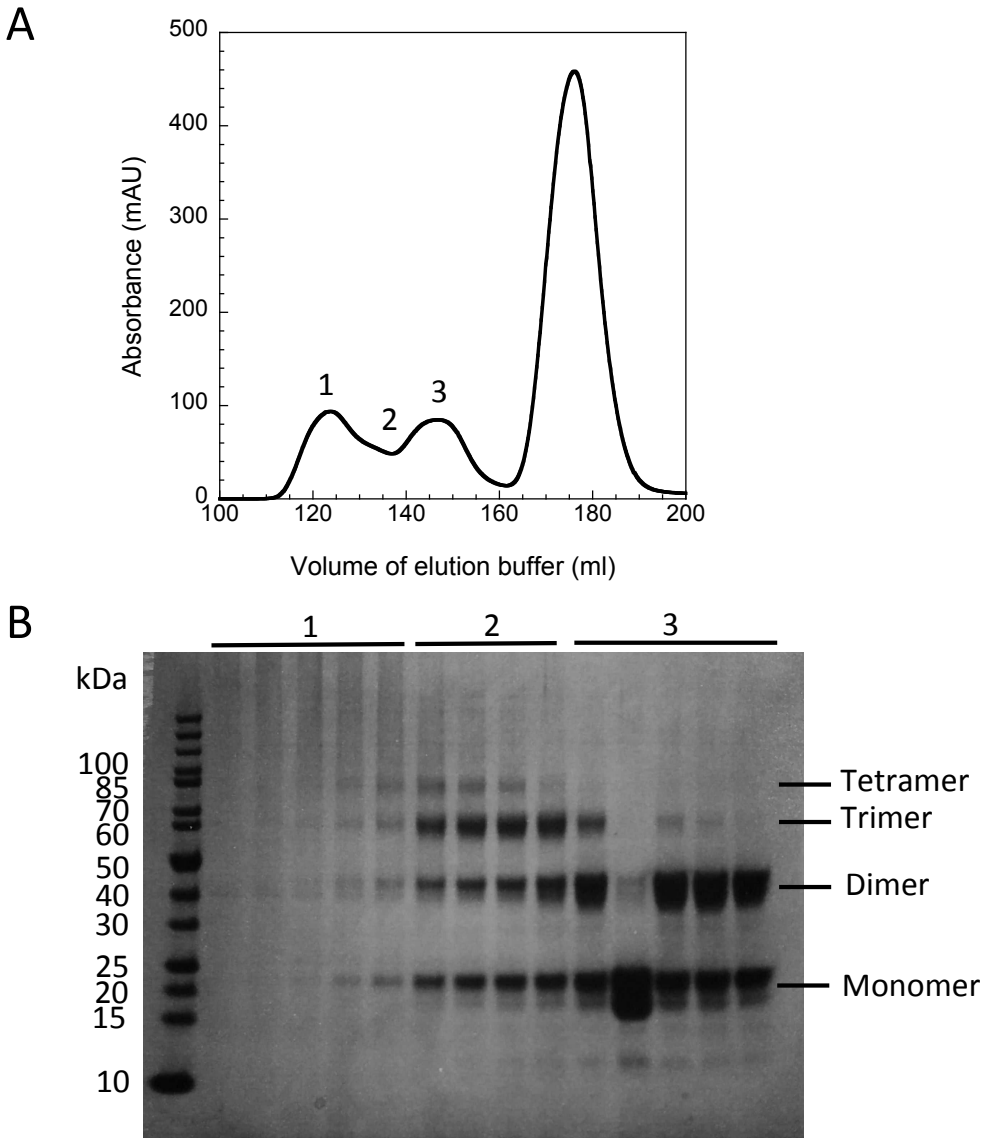


Figure 7.12: Crosslinking results of BID_{FL}. (A) Chromatogram image after G75 size exclusion run showing the peak of purified BID_{FL} as well as the extra peaks (regions 1, 2 and 3). (B) SDS gel of BID_{FL} fractions after a crosslinking reaction and NuPage ruler unstained protein ladder for comparison. Gel starts from marker and show fractions from regions 1, 2, and 3 from the chromatogram image. From the gel image, we can observe the presence of BID_{FL} monomer and its oligomers: dimers, trimers and tetramers.

7.2.5 Association kinetics of BID_{FL} and tBID with the pro-survival protein A1

Association kinetics experiments of both BID_{FL} and tBID were performed against A1 protein. Although any other folded BCL-2 family protein could have been used, the choice to work with A1 was to be consistent and comparable with my previous studies of A1 and BID-BH3 peptide (Chapter 6).

Experiments for BID_{FL}:A1 were performed under pseudo-first order conditions with A1 10x in excess. Results show a very slow association rate (Figure 7.13). Fitted traces represent an average of 25-35 measurements.

Although the current result is not sufficient to calculate the concentration dependent association rate constant (k_{on}), the significantly low apparent rate constant observed ($\approx 0.02\ s^{-1}$) can give us insights into this system. Two reasonable hypotheses can emerge:

1) The two folded proteins do not interact. The folded structure of BID_{FL} might prevent it to bind and interact with A1 as its BH3 region is found enclosed in the structure.

2) Unfolding of BID_{FL} is required in order to expose its BH3 region and allow complex formation. In the latter case, the observed rate could represent the unfolding rate of BID and be the rate-limiting step for BID_{FL} and A1 to interact.

A quick test where the light source was temporary blocked revealed that a photobleaching effect plays a role on the signal, but it is not sufficient to explain the observed fluorescence change.

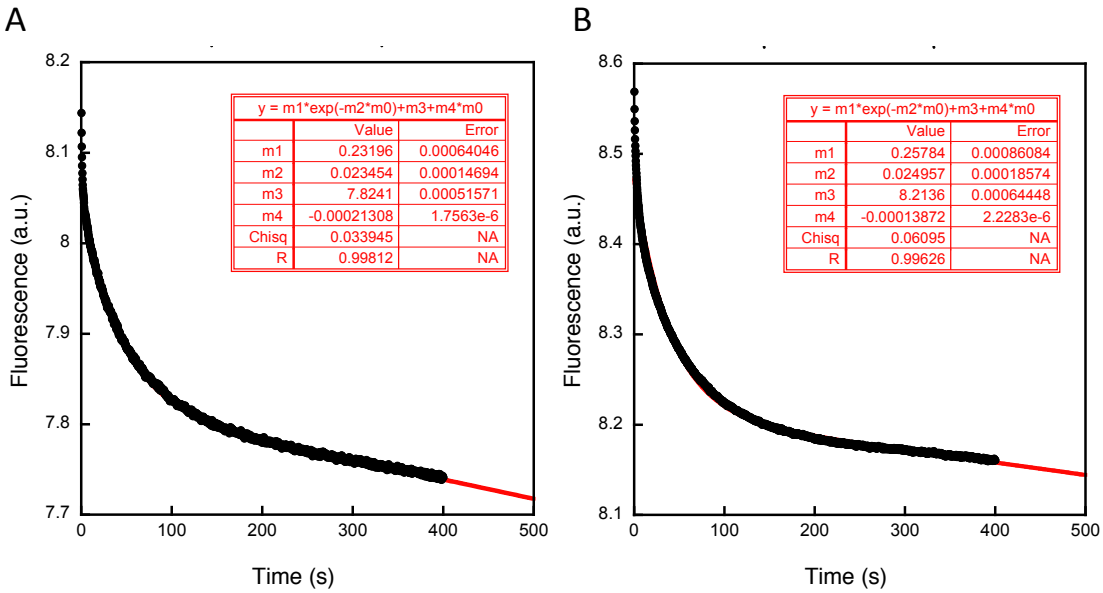


Figure 7.13: Association kinetics of BID_{FL} and A1. BID_{FL} and A1 were rapidly mixed in a stopped flow instrument on a 1:1 volume ratio under pseudo-first order conditions. The final concentrations in the cell were (A) 1 and 10 µM and (B) 1.5 and 15 µM, of BID_{FL} and A1 respectively. A 320 nm cut-off filter was employed, and each trace is a result of 35 measurements averaged.

Unfolding kinetics of BID_{FL} revealed that the protein unfolds in a rate of ≈0.01 s^{-1} (Figure 7.14, A), a very similar rate to the one observed for BID_{FL}:A1 association. The results support the idea that the unfolding of BID_{FL} is the rate limiting step for the association of BID_{FL} and A1. Surprisingly, a very similar unfolding rate (≈0.02 s^{-1}) is observed for the cleaved BID_{FL}, after incubation with caspase but before separation from the $BID_{1\text{-}60}$ fragment (Figure 7.14, B). The results show that the cleavage process on its own, is not sufficient to speed up the unfolding of the BID_{FL} molecule.

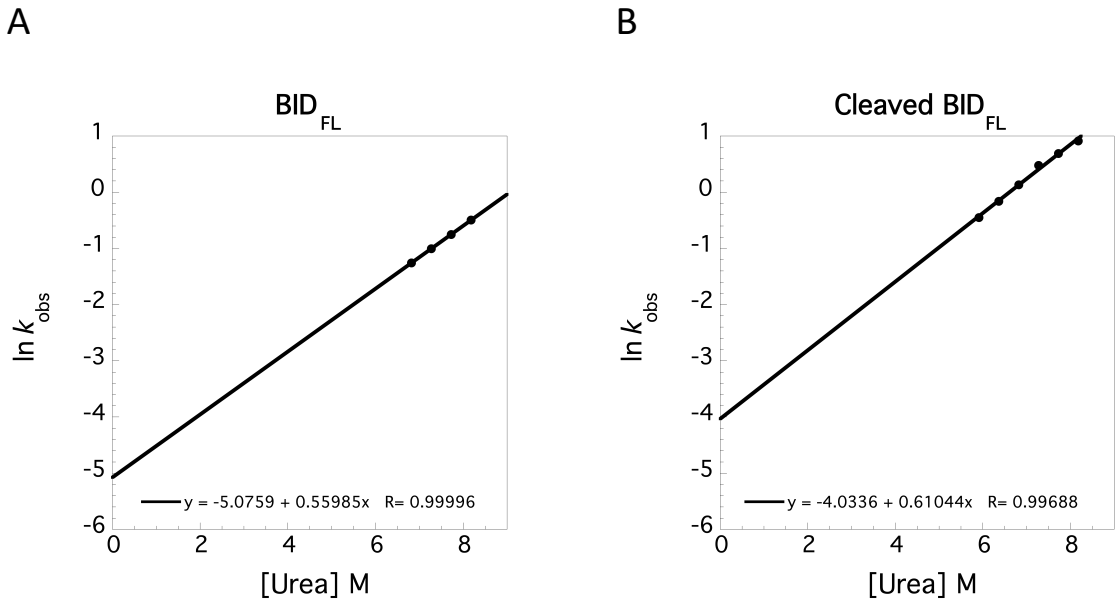


Figure 7.14: Unfolding kinetics of BID_{FL} before and after cleavage with caspase. (A) BID_{FL} and (B) BID_{FL} upon cleavage with caspase unfolding kinetics. The experiments were performed in a stopped-flow instrument, in a 1:10 volume ratio. Each point represents an average of 20-25 traces.

As tBID does not have a tryptophan residue, some preliminary tests were required in order to find the optimised settings to do association experiments with A1. That was to decide which excitation wavelength to use, as well as if a cut-off filter would be employed. Stopped flow experiments were performed in a 1:1 volume ratio, with both proteins at the same concentration (2 μM final). The results show that a 276 nm excitation wavelength with no cut-off filter is a good condition. Figure 7.15, A show an example of a trace using the chosen settings.

Later on, association kinetics were done under pseudo-first order conditions using a 1:1 volume ratio. A signal change was observed, and the data were fitted to a single exponential plus drift. The observed association rate for that concentration was ≈1.5 s^{-1} (Figure 7.15, B), which is two orders of magnitude faster when compared with the previous one from BID_{FL} and A1. Although preliminary, this result is more consistent with fast association as is observed for BID-BH3 peptide binding to A1.

Additional kinetics experiments (association and dissociation) would have allowed us to compare BID_{FL}, tBID and BID peptide data and get insights on how the BID-BH3 motif sequence context can modulate the binding affinities for BID and A1. Unfortunately, these final experiments were challenged by time limitations.

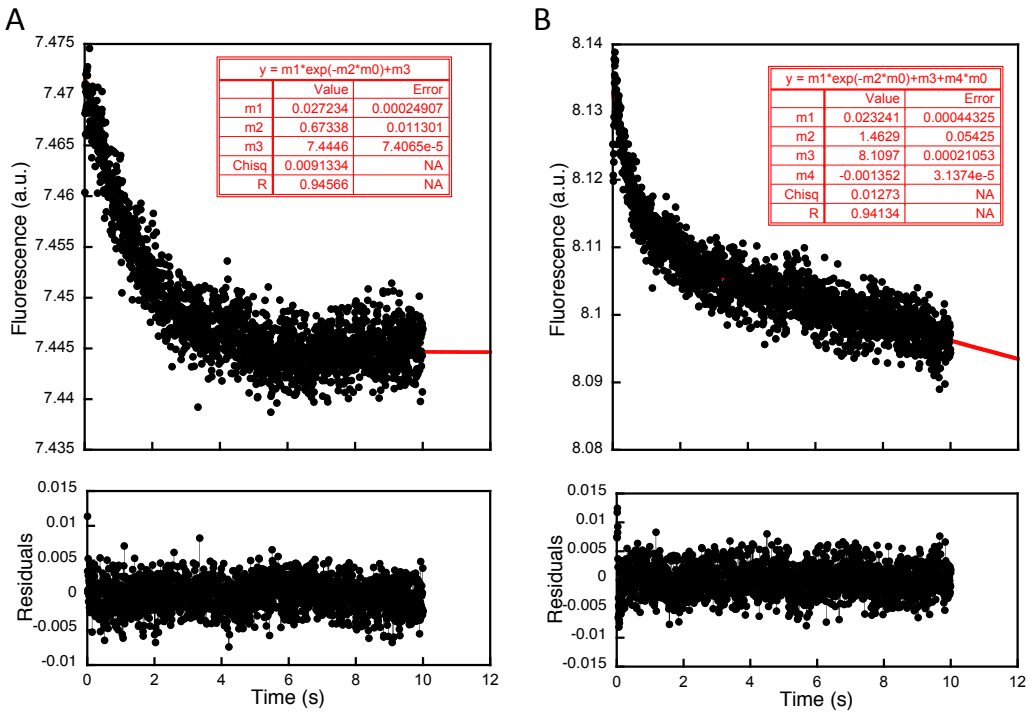


Figure 7.15: Association kinetics of tBID and A1. The two proteins were rapidly mixed in a stopped flow instrument on a 1:1 volume ratio. (A) Experiment done on a 1:1 concentration ratio: 2 μM final concentration of each protein. The trace is a result of 31 measurements averaged. (B) Experiment done under pseudo-first order conditions with A1 10x in excess: 1 and 10 μM of tBID and A1, respectively. The trace is a result of 15 measurements averaged.

7.3 Discussion

Starting from the gene design and expression and purification protocols developed, BID_{FL} and tBID were successfully produced in the lab. Even under different conditions, expression of tBID transpired to be very poor, which is the reason why we chose to produce the protein via the cleavage of BID_{FL} by caspase-8. Unexpectedly, after cleavage, tBID and $BID_{1\text{-}60}$ fragment were not separated as we first expected. Thus, unfolding of the protein was necessary in order to get free tBID in solution. This fact showed that although caspase-8 is necessary for BID_{FL} cleavage, it is not sufficient to release it from its N-terminal 60 residues part.

CD data also confirms the very similar structure of BID_{FL} and BID after caspase-8 cleavage. After the final purification step, although very disordered, tBID is able to recover some residual structure in buffer when compared with the its previous condition in denaturant. The CD spectra of both BID_{FL} and tBID are in agreement with expected structural knowledge from the literature: BID_{FL} is a folded alpha-helical protein whereas tBID is largely unstructured.

Purification of BID_{FL} showed an intriguing result: extra peaks were observed on its chromatogram. Although the peaks would correspond to higher molecular weight species, SDS PAGE revealed that they are only composed by BID_{FL}. Considering that the cysteine potential of multimerisation is abrogated by the reducing agent, presence of higher order species comes from propensity of the protein sequence to form oligomers. That was also previously observed for other BCL-2 family members, both pro and anti-apoptotic proteins. BCL-XL also shows extra peaks upon purification. Similarly, A1 and BCL-XL can oligomerise once their folded structure is destabilized by a reducing agent. Moreover, pro-apoptotic BAK and BAX oligomerise in a membrane-like environment, as when in contact with detergents. In the case of $BID_{FL,}$ cross-linking experiments identified the size of the oligomers, with dimers, trimers and tetramers observed. Although only pro-apoptotic BAK and BAX have oligomerisation as an *in vivo* function (pore forming proteins), a spectrum of oligomerising ability is observed throughout the BCL-2 family members. That instigating evidence could shed some light in the functional evolution of these proteins. Evolutionary vestiges of the BCL-2 family ancestry could possibly explain self-assembly propensities of the proteins.

Biophysical characterisation of BID_{FL} showed that despite its long flexible loop, it has the stability similar to other BCL-2 proteins like A1 and MCL-1. To gain some insights into the BID_{FL} and tBID abilities to form complexes when compared to BID-BH3, some initial kinetics experiments were conducted. Association kinetics of BID_{FL} and A1 show a very slow rate ($\approx 0.02\ s^{-1}$). Interestingly, unfolding kinetics of BID_{FL} show a very similar rate ($\approx 0.01\ s^{-1}$), suggesting that the unfolding and consequent exposure of the BH3 motif is the rate-limiting step for the complex formation. Although one might think that the cleavage step with caspase would increase the unfolding rate of the BID molecule, no significant change was observed upon incubation with caspase ($\approx 0.02\ s^{-1}$). Thus, in order to activate BID to give it pro-apoptotic activity, BID has to not only be cleaved by caspase but also unfold.

Preliminary association kinetics of tBID and A1 seemed promising. Further experiments would be required to show how the rates for association, dissociation and the modulation of complex formation of the protein in its physiological relevant length compares with the BH3 peptide, or in other words, how relevant and informative is the use of peptides as a model for studying IDPs.

8 Conclusions and future perspectives

8.1 Conclusions from this thesis

The work designed and performed in this PhD thesis aimed to deepen our understanding of how context can influence protein folding dynamics and protein-protein interactions. SasG is a protein that challenges the predictions of folding and stability. The first step of the thesis aimed to explore the potential role of SasG unusual charged composition in the protein stability. Next, SasG was used as a system to study co-translational folding. Force profile experiments were used to shed light on how these elongated beta sheet structure would fold on the ribosome and how that compares with the *in vitro* results of the protein in isolation.

Intrigued by protein disorder and seeking to gain more information of protein-protein interactions, the subsequent project was designed to investigate what encodes IDPs coupled folding and binding reactions. For that purpose, comparative ϕ-value analysis was employed utilising a family of proteins involved in cell-death regulation – the BCL-2 family. Finally, considering that IDPs are largely studied as peptides, the last chapter aimed to investigate an IDP as a full-length protein. The BCL-2 family member BID, a BH3-only protein that is structured in a full-length context, was used to shed some light on the role of this protein in the apoptosis mechanism in the cell.

8.1.1 SasG stability cannot be explained exclusively by highly charged amino acids composition and distribution

SasG is a bacterial multidomain protein, predicted to be disordered, but yet it folds. With an unusual sequence composition (65% of Glycines, Prolines and charged residues), it has an elongated beta sheet structure, exposed to the solvent in both sides. The lack of a hydrophobic core and the abundance of charged residues makes SasG structure very intriguing. Seeking to understand what is the contribution of the charged residues on SasG stability, we hypothesised that electrostatics could be key for the protein stability and that the pattern of charged residues could contribute for the structure formation.

Surprisingly, equilibrium studies of SasG at different pH and ionic strength values did not support the initial hypothesis, as they show no dramatic effect in the $G5^2$ domain stability. Moreover, bioinformatics analysis confirms that the majority of the charged residues in SasG $G5^2$ and E-$G5^2$ constructs are not involved in salt bridges.

SasG clearly defies current prediction of protein disorder and its stability cannot be simply explained by the presence or the distribution of charged residues across the molecule. Nevertheless, our previous studies were able to demonstrate the role of the inter-domain interface in the protein stability. Formation of SasG interface was shown to drive the folding of E-$G5^2$ even when both domains were destabilised and unfolded in isolation (Gruszka et al. 2016).

8.1.2 Force profile experiments can report both intramolecular folding and intermolecular interactions between the nascent protein chain and the ribosome

A fundamental question of how *in vitro* and co-translational protein folding studies compare has gained more attention in recent years. As translation happens slowly, it can allow nascent chains to fold while still being translated, tethered to the ribosome, even before the entire sequence emerges from the ribosome tunnel.

After previously characterising SasG folding mechanism *in vitro*, the follow up question comprised how this intriguing protein would fold upon translation on the ribosome. Force profile experiments were employed to characterise the co-translational folding of $G5^2$, revealing that the domain is able to fold while is still inside the ribosome vestibule (linker length L31). Remarkably, a control experiment involving a non-folding variant of the $G5^2$ domain was also able to release the stall. Furthermore, a high read through was also observed at very short linker lengths (<L30), where the protein is still largely sequestered inside the ribosome tunnel. Together, these results suggest that other factors rather than folding can contribute to the stall release and subsequent high fraction full-length results.

With the prior knowledge that E-$G5^2$ has an additional pathway for folding that starts with the formation of the inter-domain interface (Gruszka et al. 2016), we hypothesised that E-

$G5^2$ would fold earlier than $G5^2$ when being translated, as soon as the interface is available. Surprisingly, force profile results showed that $G5^2$ and E-$G5^2$ have almost identical results, with both constructs folding at similar positions upon translation. That was the first indication that the alternative folding pathway for E-$G5^2$ might not be available during translation. Moreover, force profile results for E$G5^2$-G587A, a mutant that is shown *in vitro* to fold via the domains interface, basically overlays with the non-folding variant of E-$G5^2$. The results support the idea that the alternative folding pathway, well characterised *in vitro,* is not accessible until the peptide chain emerges from the ribosome.

SasG cotranslational folding results showed that the presence of the ribosome can affect the folding process in comparison with the protein in isolation, including preventing alternative folding mechanism known to be accessible in the absence of the ribosome. In agreement with recent findings, our results contribute to the idea that the ribosome can play a role in protein regulation.

Interestingly, the highly charged composition of SasG allowed us to explore the role of electrostatics in the co-translational folding. Non-folding variants of SasG with different charge composition were used to shed light on how the presence of charged residues can contribute for the stall sequence release. Overall, although disordered, highly positively charged constructs were shown to release the stall whereas highly negatively charged nascent chains remained stalled. Although force profile experiments can be a powerful technique to study co-translational folding, non-folding controls transpired to be essential in order to attribute the stall release to folding events.

8.1.3 Intrinsically disordered proteins are able to encode coupled folding and binding reactions

Intrinsically disordered proteins (IDPs) are gaining more attention as their biological relevance is increasingly revealed. Although lacking a structure in isolation, a subset of IDPs is able to fold upon binding to a partner protein. Coupled folding and binding reactions can be very promiscuous, with different IDPs being able to bind to the same partner protein. Moreover, as the stability of the IDP-partner complex is dependent on the folded molecule,

an immediate hypothesis would be that the folded partner templates the coupled folding and binding reaction.

Comparative ϕ-value analysis of BCL-2 family proteins revealed that the folding pathway is essentially encoded by the IDP, not templated by the stabilising partner protein as the initial hypothesis suggested. The results open space for new insights regarding the role of disorder. The ability of IDPs to encode both their residual structure and transition states, provide them the capacity to develop specific kinetic profiles. *In vivo*, disordered proteins are conserved and typically present in cell signalling processes, where responses to stimuli may need to be irreversible (e.g. apoptosis stimulation) or occur significantly fast (e.g. cell receptor activation). Altering the residual structure or the encoded transition state provide a fine manner for evolution to modulate the lifetime of these complexes, which could be a good explanation for disorder conservation.

8.1.4 Insights of an IDP in its full-length context

IDPs are typically studied as peptides, the binding motifs of a bigger protein. Investigation of the BCL-2 family member BID on its full-length context (BID_{FL}) could shed some light on the biological mechanism by which this molecule plays a role in the apoptosis network. It is well known that BID_{FL} gets truncated by caspase in the cell. Association and unfolding kinetics revealed that BID cannot bind its partner protein A1 in a full-length context. Cleavage with caspase followed by unfolding is required for the BID molecule to open and expose its binding motif, the BH3 region.

Interestingly, an oligomerisation propensity, previously identified for other BCL-2 family members (BAK, BAX, A1, BLC-XL), was also observed for BID_{FL}. Although only the pore forming pro-apoptotic BAX and BAK have oligomerisation as an *in vivo* function, the potential to form oligomers is observed across the BCL-2 family members. That *per se* can lead to a better understanding of the functional evolution of these proteins.

8.2 Future directions

In terms of the work described in this thesis, it would be very interesting to compare how the association, dissociation and lifetime of the complexes are affected between BID peptide, cleaved BID_{FL} and tBID binding to A1. That would shed some light on how the use of IDP binding motifs only (peptides) can report on what actually happens in the cellular context.

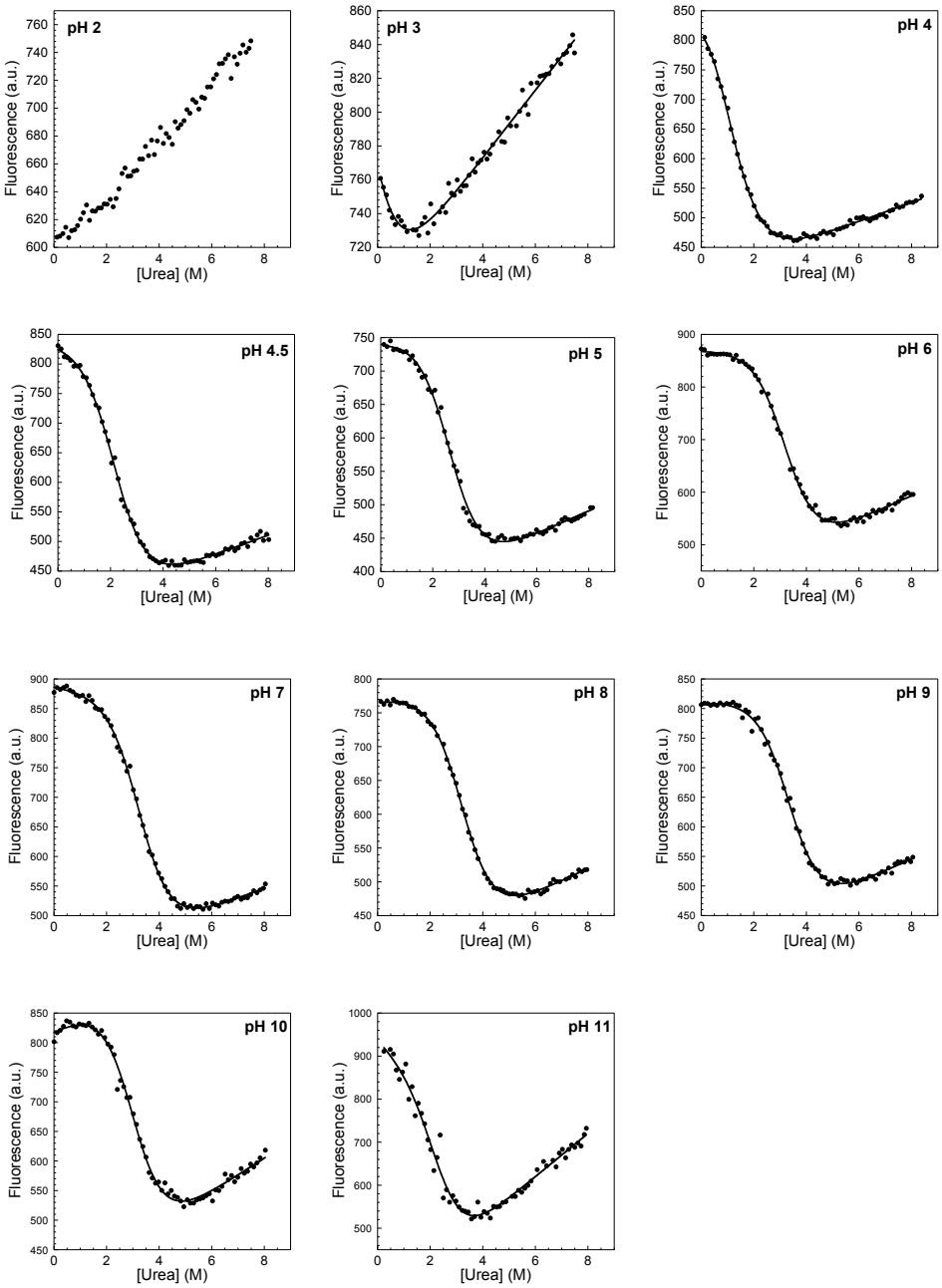
Another very interesting project to pursue is the oligomerisation propensity of the BCL-2 family proteins. An evolutionary reconstruction could unravel whether the ancestor of the current family members was more, or less, susceptible to self-assembly.

It would be also fascinating to study an IDP or IDR that can adopt different conformations upon binding to a partner molecule. The disordered C-terminal domain of p53, for instance, can bind as a helix (Rustandi, Baldisseri and Weber, 2000), as a coil (Lowe et al. 2002; Mujtaba et al. 2004) or as a strand (Avalos et al. 2002). Understanding how the transition state is encoded when multiple binding structures are possible would be very significant.

9 Apendix

Appendix 1-

Equilibrium curves of $G5^2$ at different pH values: fluorescence at 305 nm.



Appendix 2-

SasG folding on the ribosome DNA and protein sequences

$G5^2$ protein sequence:

Y G P V K G D S I V E K E E I P F E K E R K F N P D L A P G T
E K V T R E G Q K G E K T I T T P T L K N P L T G E I I S K G
E S K E E I T K D P I N E L T E Y G P E T

$G5^2$ DNA sequence:

TACGGTCCGGTTAAAGGTGACTCTATCGTTGAAAAAGAAGAAATCCCGTTCGAAAAAGAA
CGTAAATTCAACCCGGACCTGGCGCCGGGTACCGAAAAAGTTACCCGTGAAGGTCAGAAA
GGTGAAAAAACCATCACCACCCCGACCCTGAAAAACCCGCTGACCGGTGAAATCATCTCT
AAAGGTGAATCTAAAGAAGAAATCACCAAAGACCCGATCAACGAACTGACCGAATACGGT
CCGGAAACC

E-$G5^2$ protein sequence:

G P E T I A P G H R D E F D P K L P T G E K E E V P G K P G I
K N P E T G D V V R P P V D S V T K Y G P V K G D S I V E K E
E I P F E K E R K F N P D L A P G T E K V T R E G Q K G E K T
I T T P T L K N P L T G E I I S K G E S K E E I T K D P I N E
L T E Y G P E T

E-$G5^2$ DNA sequence:

GGGCCCGAAACGATCGCGCCGGGTCACCGTGACGAATTTGACCCGAAACTGCCGACCGGT
GAAAAAGAAGAAGTTCCGGGTAAACCGGGTATCAAAAACCCGGAAACCGGTGACGTTGTT
CGTCCGCCGGTTGACTCTGTTACCAAATACGGTCCGGTTAAAGGTGACTCTATCGTTGAA
AAAGAAGAAATCCCGTTCGAAAAAGAACGTAAATTCAACCCGGACCTGGCGCCGGGTACC
GAAAAAGTTACCCGTGAAGGTCAGAAAGGTGAAAAAACCATCACCACCCCGACCCTGAAA
AACCCGCTGACCGGTGAAATCATCTCTAAAGGTGAATCTAAAGAAGAAATCACCAAAGAC
CCGATCAACGAACTGACCGAATACGGTCCGGAAACC

G5[2] only negative charges protein sequence:

Y G P V M G D S I V E M E E I P F E M E M M F N P D L A P G T
E M V T M E G Q M G E M T I T T P T L M N P L T G E I I S M G
E S M E E I T M D P I N E L T E Y G P E T

G5[2] only negative charges DNA sequence:

TACGGTCCGGTTATGGGTGACTCTATCGTTGAAATGGAAGAAATCCCGTTCGAAATGGAA
ATGATGTTCAACCCGGACCTGGCGCCGGGTACCGAAATGGTTACCATGGAAGGTCAGATG
GGTGAAATGACCATCACCACCCCGACCCTGATGAACCCGCTGACCGGTGAAATCATCTCT
ATGGGTGAATCTATGGAAGAAATCACCATGGACCCGATCAACGAACTGACCGAATACGGT
CCGGAAACC

E-G5[2] only negative charges protein sequence:

G P E T I A P G H M D E F D P M L P T G E M E E V P G M P G I
M N P E T G D V V M P P V D S V T M Y G P V M G D S I V E M E
E I P F E M E M M F N P D L A P G T E M V T M E G Q M G E M T
I T T P T L M N P L T G E I I S M G E S M E E I T M D P I N E
L T E Y G P E T

E-G5[2] only negative charges DNA sequence:

GGGCCCGAAACGATCGCGCCGGGTCACATGGACGAATTTGACCCGATGCTGCCGACCGGT
GAAATGGAAGAAGTTCCGGGTATGCCGGGTATCATGAACCCGGAAACCGGTGACGTTGTT
ATGCCGCCGGTTGACTCTGTTACCATGTACGGTCCGGTTATGGGTGACTCTATCGTTGAA
ATGGAAGAAATCCCGTTCGAAATGGAAATGATGTTCAACCCGGACCTGGCGCCGGGTACC
GAAATGGTTACCATGGAAGGTCAGATGGGTGAAATGACCATCACCACCCCGACCCTGATG
AACCCGCTGACCGGTGAAATCATCTCTATGGGTGAATCTATGGAAGAAATCACCATGGAC
CCGATCAACGAACTGACCGAATACGGTCCGGAAACC

G5[2] only positive charges protein sequence:

Y G P V K G N S I V Q K Q Q I P F Q K Q R K F N P N L A P G T
Q K V T R Q G Q K G Q K T I T T P T L K N P L T G Q I I S K G
Q S K Q Q I T K N P I N Q L T Q Y G P Q T

G5[2] only positive charges DNA sequence:

TACGGTCCGGTTAAAGGTAATTCTATCGTTCAAAAACAACAAATCCCGTTCCAAAAACAA
CGTAAATTCAACCCGAATCTGGCGCCGGGTACCCAAAAAGTTACCCGTCAAGGTCAGAAA
GGTCAAAAAACCATCACCACCCCGACCCTGAAAAACCCGCTGACCGGTCAAATCATCTCT
AAAGGTCAATCTAAACAACAAATCACCAAAAATCCGATCAACCAACTGACCCAATACGGT
CCGCAAACC

E-G5[2] only positive charges protein sequence:

G P Q T I A P G H R N Q F N P K L P T G Q K Q Q V P G K P G I
K N P Q T G N V V R P P V N S V T K Y G P V K G N S I V Q K Q
Q I P F Q K Q R K F N P N L A P G T Q K V T R Q G Q K G Q K T
I T T P T L K N P L T G Q I I S K G Q S K Q Q I T K N P I N Q
L T Q Y G P Q T

E-G5[2] only positive charges DNA sequence:

GGGCCCCAAACGATCGCGCCGGGTCACCGTAATCAATTTAATCCGAAACTGCCGACCGGT
CAAAAACAACAAGTTCCGGGTAAACCGGGTATCAAAAACCCGCAAACCGGTAATGTTGTT
CGTCCGCCGGTTAATTCTGTTACCAAATACGGTCCGGTTAAAGGTAATTCTATCGTTCAA
AAACAACAAATCCCGTTCCAAAAACAACGTAAATTCAACCCGAATCTGGCGCCGGGTACC
CAAAAAGTTACCCGTCAAGGTCAGAAAGGTCAAAAAACCATCACCACCCCGACCCTGAAA
AACCCGCTGACCGGTCAAATCATCTCTAAAGGTCAATCTAAACAACAAATCACCAAAAAT
CCGATCAACCAACTGACCCAATACGGTCCGCAAACC

G5[2] no charges protein sequence:

Y G P V M G N S I V Q M Q Q I P F Q M Q M M F N P N L A P G T
Q M V T M Q G Q M G Q M T I T T P T L M N P L T G Q I I S M G
Q S M Q Q I T M N P I N Q L T Q Y G P Q T

$G5^2$ no charges DNA sequence:

TACGGTCCGGTTATGGGTAATTCTATCGTTCAAATGCAACAAATCCCGTTCCAAATGCAA
ATGATGTTCAACCCGAATCTGGCGCCGGGTACCCAAATGGTTACCATGCAAGGTCAGATG
GGTCAAATGACCATCACCACCCCGACCCTGATGAACCCGCTGACCGGTCAAATCATCTCT
ATGGGTCAATCTATGCAACAAATCACCATGAATCCGATCAACCAACTGACCCAATACGGT
CCGCAAACC

E-$G5^2$ no charges protein sequence:

G P Q T I A P G H M N Q F N P M L P T G Q M Q Q V P G M P G I
M N P Q T G N V V M P P V N S V T M Y G P V M G N S I V Q M Q
Q I P F Q M Q M M F N P N L A P G T Q M V T M Q G Q M G Q M T
I T T P T L M N P L T G Q I I S M G Q S M Q Q I T M N P I N Q
L T Q Y G P Q T

E-$G5^2$ no charges DNA sequence:

GGGCCCCAAACGATCGCGCCGGGTCACATGAATCAATTTAATCCGATGCTGCCGACCGGT
CAAATGCAACAAGTTCCGGGTATGCCGGGTATCATGAACCCGCAAACCGGTAATGTTGTT
ATGCCGCCGGTTAATTCTGTTACCATGTACGGTCCGGTTATGGGTAATTCTATCGTTCAA
ATGCAACAAATCCCGTTCCAAATGCAAATGATGTTCAACCCGAATCTGGCGCCGGGTACC
CAAATGGTTACCATGCAAGGTCAGATGGGTCAAATGACCATCACCACCCCGACCCTGATG
AACCCGCTGACCGGTCAAATCATCTCTATGGGTCAATCTATGCAACAAATCACCATGAAT
CCGATCAACCAACTGACCCAATACGGTCCGCAAACC

Appendix 3 - Table 1. Biophysical parameters for the coupled folding and binding of PUMA–A1.

PUMA	Helicity %	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1}) x10^{-3}	n	K_d (k_{off}/k_{on}) (nM)	$\Delta\Delta G$ (kcal.mol^{-1})	Φ
WT	20.1	5.9 ± 0.2	0.50 ± 0.02	4	0.085 ± 0.004	-	-
W133F	17.0	4.27 ± 0.04	2.03 ± 0.03	3	0.475 ± 0.008	1.02 ± 0.03	0.19 ± 0.02
I137A	17.2	1.64 ± 0.05	6.0 ± 0.2	4	3.7 ± 0.2	2.23 ± 0.04	0.34 ± 0.01
L141A	19.8	1.67 ± 0.01	1227 ± 8	4	735 ± 7	5.37 ± 0.03	0.139 ± 0.003
L148A	20.1	4.90 ± 0.05	34.8 ± 0.3	3	7.10 ± 0.09	2.62 ± 0.03	0.042 ±0.008
Y152A	28.6	5.1 ± 0.1	2.51 ± 0.03	3	0.49 ± 0.01	1.04 ± 0.03	0.08 ± 0.02
E132A	21.0	8.8 ± 0.5	0.77 ± 0.01	4	0.088 ± 0.005	0.02 ± 0.05	-
E132G	14.3	8.4 ± 0.2	0.86 ± 0.01	4	0.102 ± 0.003	0.09 ± 0.03	0.30 ± 0.02
E136A	21.4	10.4 ± 0.2	0.48 ± 0.04	4	0.046 ± 0.004	-0.36 ± 0.06	-
E136G	15.3	8.2 ± 0.2	2.38 ± 0.09	4	0.29 ± 0.01	1.09 ± 0.04	0.13 ± 0.01
A139G	13.2	3.26 ± 0.05	1.08 ± 0.04	4	0.33 ± 0.01	0.81 ± 0.04	0.44 ± 0.02
R143A	22.2	3.67 ± 0.04	0.88 ± 0.03	4	0.240 ± 0.008	0.62 ± 0.04	-
R143G	15.8	3.03 ± 0.01	5.2 ± 0.2	3	1.72 ± 0.07	1.17 ± 0.04	0.097 ± 0.002
D147A	30.2	9.4 ± 0.3	0.48 ± 0.02	4	0.051 ± 0.003	-0.92 ± 0.04	-
D147G	15.9	6.76 ± 0.08	1.95 ± 0.05	4	0.288 ± 0.008	1.03 ± 0.04	0.2 ± 0.1
A150G	13.3	4.84 ± 0.08	2.15 ± 0.02	4	0.444 ± 0.008	0.98 ± 0.03	0.12 ± 0.02
R154A	20.0	4.86 ± 0.08	0.54 ± 0.02	4	0.111 ± 0.005	0.16 ± 0.04	-
R154G	13.5	4.79 ± 0.03	0.44 ± 0.03	3	0.092 ± 0.006	-0.11 ± 0.05	-0.076 ± 0.002
A139G (A150G)	9.7	2.78 ± 0.02	3.9 ± 0.1	4	1.40 ± 0.04	0.68 ± 0.03	0.48 ± 0.01
A150G (A139G)	9.7	2.78 ± 0.02	3.9 ± 0.1	4	1.40 ± 0.04	0.82 ± 0.03	0.11 ± 0.01

Percentage of helicity was calculated using the MRE at 222 nm and the method of Munoz and Serrano (1995). k_{on} and k_{off} were experimentally determined, and the associated errors represent the curve fit error and the standard error of the mean (n= number of repeats used to determine k_{off}), respectively. K_d was calculated by taking the ratio between the dissociation and association rate constants, and the respective errors were propagated using standard methods. Φ was calculated using the kinetic rate constants and errors represent the propagated error. Table taken from Crabtree et al. 2018.

Table 2. Biophysical parameters for the coupled folding and binding of BID–MCL-1.

BID	Helicity (CD) %	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1}) x10^{-3}	n	K_d (k_{off}/k_{on}) (nM)	$\Delta\Delta G$ (kcal.mol^{-1})	Φ
WT	11.5	8.9 ± 0.1	3.0 ± 0.1	4	0.33 ± 0.01	-	-
I82A	10.7	9.5 ± 0.9	6.20 ± 0.07	2	0.65 ± 0.06	0.40 ± 0.06	0 ± 300
I83A	12.0	8.6 ± 0.1	14 ± 4	4	1.7 ± 0.4	0.9 ± 0.1	0.02 ± 0.02
I86A	9.0	4.0 ± 0.1	140 ± 14	3	35 ± 4	2.78 ± 0.08	0.172 ± 0.008
L90A	7.5	2.1 ± 0.1	1300 ± 150	4	610 ± 80	4.50 ± 0.09	0.191 ± 0.008
I93A	8.9	4.5 ± 0.2	430 ± 26	4	96 ± 7	3.39 ± 0.06	0.121 ± 0.008
M97A	11.5	7.8 ± 0.3	30 ± 4	3	3.8± 0.5	1.47 ± 0.09	0.06 ± 0.02
I101A	10.1	8.2 ± 0.1	2.5 ± 0.2	4	0.30 ± 0.03	-0.06 ± 0.07	0 ± 1
I86A-M97A	9.0	3.5 ± 0.2	1360 ± 70	4	390 ± 30	4.23 ± 0.06	0.133 ± 0.007
E81A	10.0	6.5 ± 0.3	3.6 ± 0.3	4	0.55 ± 0.05	-	-
E81G	7.0	4.8 ± 0.1	8 ± 2	3	1.6 ± 0.3	0.6 ± 0.1	0.27 ± 0.06
R88A	11.3	10.0 ± 0.3	6.3 ± 0.7	3	0.6 ± 0.08	-	-
R88G	5.8	8.3 ± 0.5	15 ± 1	3	1.8 ± 0.2	0.6 ± 0.1	0.17 ± 0.06
E96A	13.0	9.9 ± 0.8	3.8 ± 0.6	4	0.38 ± 0.07	-	-
E96G	11.5	10.9 ± 0.4	10 ± 1	4	0.9 ± 0.1	0.5 ± 0.1	-0.1 ± 0.1
H99A	11.6	9.2 ± 0.4	1.0 ± 0.3	4	0.12 ± 0.03	-	-
H99G	13.0	9.0 ± 0.2	4.0 ± 0.1	4	0.4 ± 0.1	0.8 ± 0.2	0.01 ± 0.04

Percentage of helicity was calculated using the MRE at 222 nm and the method of Munoz and Serrano (1995). k_{on} and k_{off} were experimentally determined, and the associated errors represent the curve fit error and the standard error of the mean (n= number of repeats used to determine k_{off}), respectively. K_d was calculated by taking the ratio between the dissociation and association rate constants, and the respective errors were propagated using standard methods. Φ was calculated using the kinetic rate constants and errors represent the propagated error. Table taken from Crabtree et al. 2018.

10 References

Alberts, Bruce. 2008. “The Shape and Structure of Proteins.” in *Molecular Biology of the Cell*.

Anderson, D. Eric, Wayne J. Becktel, and F. W. Dahlquist. 1990. “PH-Induced Denaturation of Proteins: A Single Salt Bridge Contributes 3-5 Kcal/Mol to the Free Energy of Folding of T4 Lysozyme.” *Biochemistry* 29(9):2403–8.

Anfinsen, C. B., E. Haber, M. Sela, and F. H. White. 1961. “The Kinetics of Formation of Native Ribonuclease during Oxidation of the Reduced Polypeptide Chain.” *Proceedings of the National Academy of Sciences of the United States of America* 47:1309–14.

Anfinsen, Christian B. 1973. “Principles That Govern the Folding of Protein Chains.” *Science*.

Arai, Munehito, Kenji Sugase, H. Jane Dyson, and Peter E. Wright. 2015. “Conformational Propensities of Intrinsically Disordered Proteins Influence the Mechanism of Binding and Folding.” *Proceedings of the National Academy of Sciences* 112(31):9614–19.

Avalos, Jose L., Ivana Celic, Shabazz Muhammad, Michael S. Cosgrove, Jef D. Boeke, and Cynthia Wolberger. 2002. “Structure of a Sir2 Enzyme Bound to an Acetylated P53 Peptide.” *Molecular Cell*.

Babu, M. Madan. 2016. “The Contribution of Intrinsically Disordered Regions to Protein Function, Cellular Complexity, and Human Disease.” *Biochemical Society Transactions*.

Babu, M. Madan, Robin van der Lee, Natalia Sanchez de Groot, and Jörg Gsponer. 2012. “Intrinsically Disordered Proteins: Regulation and Disease.” in *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map*.

Bah, Alaji and Julie D. Forman-Kay. 2016. “Modulation of Intrinsically Disordered Protein Function by Post-Translational Modifications.” *Journal of Biological Chemistry*.

Baldwin, R. L. 2006. “Temperature Dependence of the Hydrophobic Interaction in Protein Folding.” *Proceedings of the National Academy of Sciences* 83(21):8069–72.

Ball, Philip. 2011. “More than a Bystander.” *Nature*.

Bateman, Alex, Lachlan Coin, Richard Durbin, Robert D. Finn, Volker Hollich, Sam Griffiths-Jones, Ajay Khanna, Mhairi Marshall, Simon Moxon, Erik L. L. Sonnhammer, David J. Studholme, Corin Yeats, and Sean R. Eddy. 2004. “@Pfam@The Pfam Protein Families Database.” *Nucleic Acids Research* 32:D138-41.

Bateman, Alex, Matthew T. G. Holden, and Corin Yeats. 2005. “The G5 Domain: A Potential N-Acetylglucosamine Recognition Domain Involved in Biofilm Formation.” *Bioinformatics* 21:1301–3.

Blommel, Paul G. and Brian G. Fox. 2007. “A Combined Approach to Improving Large-Scale Production of Tobacco Etch Virus Protease.” *Protein Expression and Purification* 55(1):53–68.

Borcherds, Wade, François Xavier Theillet, Andrea Katzer, Ana Finzel, Katie M. Mishall, Anne T. Powell, Hongwei Wu, Wanda Manieri, Christoph Dieterich, Philipp Selenko, Alexander Loewer, and Gary W. Daughdrill. 2014. “Disorder and Residual Helicity Alter P53-Mdm2 Binding Affinity and Signaling in Cells.” *Nature Chemical Biology* 10(12):1000–1002.

Brandts, John F. 1964. “The Thermodynamics of Protein Denaturation. II. A Model of Reversible Denaturation and Interpretations Regarding the Stability of Chymotrypsinogen.” *Journal of the American Chemical Society* 86(20):4302–14.

Bryngelson, Joseph D., José Nelson Onuchic, Nicholas D. Socci, and Peter G. Wolynes. 1995. “Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis.” *Proteins: Structure, Function, and Bioinformatics*.

Cabrita, Lisa D., Anaïs M. E. Cassaignau, Hélène M. M. Launay, Christopher A. Waudby, Tomasz Wlodarski, Carlo Camilloni, Maria Evangelia Karyadi, Amy L. Robertson, Xiaolin Wang, Anne S. Wentink, Luke S. Goodsell, Cheryl A. Woolhead, Michele Vendruscolo, Christopher M. Dobson, and John Christodoulou. 2016. “A Structural Ensemble of a Ribosome-Nascent Chain Complex during Cotranslational Protein Folding.” *Nature Structural and Molecular Biology* 23(4):278–85.

Cabrita, Lisa D., Christopher M. Dobson, and John Christodoulou. 2010. “Protein Folding on

the Ribosome." *Current Opinion in Structural Biology* 20(1):33–45.

Camilloni, Carlo, Daniela Bonetti, Angela Morrone, Rajanish Giri, Christopher M. Dobson, Maurizio Brunori, Stefano Gianni, and Michele Vendruscolo. 2016. "Towards a Structural Biology of the Hydrophobic Effect in Protein Folding." *Scientific Reports*.

Cao, Shannan, Peng Liu, Haiyan Zhu, Haiyan Gong, Jianfeng Yao, Yawei Sun, Guangfeng Geng, Tong Wang, Sizhou Feng, Mingzhe Han, Jiaxi Zhou, and Yuanfu Xu. 2015. "Extracellular Acidification Acts as a Key Modulator of Neutrophil Apoptosis and Functions." *PLoS ONE* 10(9).

Chandler, David. 2005. "Interfaces and the Driving Force of Hydrophobic Assembly." *Nature*.

Chaney, Julie L., Aaron Steele, Rory Carmichael, Anabel Rodriguez, Alicia T. Specht, Kim Ngo, Jun Li, Scott Emrich, and Patricia L. Clark. 2017. "Widespread Position-Specific Conservation of Synonymous Rare Codons within Coding Sequences." *PLoS Computational Biology* 13(5).

Chen, Jessica Walton, Pedro Romero, Vladimir N. Uversky, and A. Keith Dunker. 2006. "Conservation of Intrinsic Disorder in Protein Domains and Families: II. Functions of Conserved Disorder." *Journal of Proteome Research*.

Chen, Jianhan. 2012. "Towards the Physical Basis of How Intrinsic Disorder Mediates Protein Function." *Archives of Biochemistry and Biophysics*.

Chen, Jianping, Han Liang, and Ariel Fernández. 2008. "Protein Structure Protection Commits Gene Expression Patterns." *Genome Biology*.

Clarke, J. and A. R. Fersht. 1993. "Engineered Disulfide Bonds as Probes of the Folding Pathway of Barnase: Increasing the Stability of Proteins against the Rate of Denaturation." *Biochemistry* 32:4322–29.

Clarke, Jane, Kim Henrick, and Alan R. Fersht. 1995. "Disulfide Mutants of Barnase I: Changes in Stability and Structure Assessed by Biophysical Methods and X-Ray Crystallography." *Journal of Molecular Biology*.

Clarke, Jane, Andrea M. Hounslow, and Alan R. Fersht. 1995. "Disulfide Mutants of Barnase II: Changes in Structure and Local Stability Identified by Hydrogen Exchange." *Journal*

of Molecular Biology.

Clarke, Jane, Andrea M. Hounslow, Alan R. Fersht, Chris J. Bond, and Valerie Daggett. 2000. "The Effects of Disulfide Bonds on the Denatured State of Barnase." *Protein Science*.

Conrady, Deborah G., Jeffrey J. Wilson, and Andrew B. Herr. 2013. "Structural Basis for Zn²⁺-Dependent Intercellular Adhesion in Staphylococcal Biofilms." *Proceedings of the National Academy of Sciences* 110(3):E202 LP-E211.

Corrigan, Rebecca M., David Rigby, Pauline Handley, and Timothy J. Foster. 2007. "The Role of Staphylococcus Aureus Surface Protein SasG in Adherence and Biofilm Formation." *Microbiology* 153(8):2435–46.

Costantini, Susan, Giovanni Colonna, and Angelo M. Facchiano. 2008. "ESBRI: A Web Server for Evaluating Salt Bridges in Proteins." *Bioinformation* 3(3):137–38.

Costerton, J. W., P. S. Stewart, and E. P. Greenberg. 1999. "Bacterial Biofilms: A Common Cause of Persistent Infections." *Science (New York, N.Y.)* 284:1318–22.

Crabtree, Michael D., Carolina A. T. F. Mendonça, Quenton R. Bubb, and Jane Clarke. 2018. "Folding and Binding Pathways of BH3-Only Proteins Are Encoded within Their Intrinsically Disordered Sequence, Not Templated by Partner Proteins." *Journal of Biological Chemistry* 293(25):9718–23.

Crabtree, Michael David, Wade Borcherds, Anusha Poosapati, Sarah L. Shammas, Gary W. Daughdrill, and Jane Clarke. 2017. "Conserved Helix-Flanking Prolines Modulate IDP:Target Affinity by Altering the Lifetime of the Bound Complex." *Biochemistry* acs.biochem.7b00179.

Craig Venter, J., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M.

Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Yuan Wang, A. Wang, X. Wang, J. Wang, M. H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. Lai Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Yu H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. Ni Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Deslattes Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. 2001. “The Sequence of the Human Genome.” *Science*.

Creighton, Thomas E. 1988. “Disulphide Bonds and Protein Stability.” *BioEssays*.

Czabotar, Peter E., Guillaume Lessene, Andreas Strasser, and Jerry M. Adams. 2014. “Control

of Apoptosis by the BCL-2 Protein Family: Implications for Physiology and Therapy." *Nature Reviews Molecular Cell Biology* 15(1):49–63.

Dahal, Liza, Tristan O. C. Kwan, Sarah L. Shammas, and Jane Clarke. 2017. "PKID Binds to KIX via an Unstructured Transition State with Nonnative Interactions." *Biophysical Journal*.

Dandliker, Walter B., Mao Lin Hsu, Jacques Levin, and B. Ramanath Rao. 1981. "Equilibrium and Kinetic Inhibition Assays Based upon Fluorescence Polarization." *Methods in Enzymology* 74(C):3–28.

Das, R. K. and R. V. Pappu. 2013. "Conformations of Intrinsically Disordered Proteins Are Influenced by Linear Sequence Distributions of Oppositely Charged Residues." *Proceedings of the National Academy of Sciences* 110(33):13392–97.

Dewson, G. and R. M. Kluck. 2009. "Mechanisms by Which Bak and Bax Permeabilise Mitochondria during Apoptosis." *Journal of Cell Science* 122(16):2801–8.

Dill, Ken A. 1990. "Dominant Forces in Protein Folding." *Biochemistry*.

Dogan, Jakob, Xin Mu, Ake Engstrom, and Per Jemth. 2013. "The Transition State Structure for Coupled Binding and Folding of Disordered Protein Domains." *Scientific Reports*.

Dunker, A. K., Z. Obradovic, P. Romero, E. C. Garner, and C. J. Brown. 2000. "Intrinsic Protein Disorder in Complete Genomes." *Genome Informatics. Workshop on Genome Informatics*.

Dunker, A. Keith, J. David Lawson, Celeste J. Brown, Ryan M. Williams, Pedro Romero, Jeong S. Oh, Christopher J. Oldfield, Andrew M. Campen, Catherine M. Ratliff, Kerry W. Hipps, Juan Ausio, Mark S. Nissen, Raymond Reeves, Chulhee Kang, Charles R. Kissinger, Robert W. Bailey, Michael D. Griswold, Wah Chiu, Ethan C. Garner, and Zoran Obradovic. 2001. "Intrinsically Disordered Protein."

Dunker, Keith, Marc Cortese, Pedro Romero, Lilia Iakoucheva, and Vladimir Uversky. 2005. "Flexible Nets." *FEBS Journal* 272(20):5129–48.

Dunne, W. Michael. 2002. "Bacterial Adhesion: Seen Any Good Biofilms Lately?" *Clinical Microbiology Reviews* 15:155–66.

Dyson, H. Jane and Peter E. Wright. 2002. “Coupling of Folding and Binding for Unstructured Proteins.” *Current Opinion in Structural Biology*.

Edwards, Yvonne J. K. JK, Anna E. Lobley, Melissa M. Pentony, and David T. Jones. 2009. “Insights into the Regulation of Intrinsically Disordered Proteins in the Human Proteome by Analyzing Sequence and Gene Expression Data.” *Genome Biology* 10(5):R50.

Evans, Michael S., Thomas F. Clarke, and Patricia L. Clark. 2005. “Conformations of Co-Translational Folding Intermediates.” *Protein and Peptide Letters* 12(2):189–95.

Fersht, Alan R. 1997. “Nucleation Mechanisms in Protein Folding.” *Current Opinion in Structural Biology*.

Fersht, Alan R. 1999. “Structure and Mechanism in Protein Science, a Guide to Enzyme Catalysis and Protein Folding.” *Lavoisier.Fr*.

Fersht, Alan R. and Satoshi Sato. 2004. “Phi-Value Analysis and the Nature of Protein-Folding Transition States.” *Proceedings of the National Academy of Sciences of the United States of America* 101:7976–81.

Fersht, Alan R., Jian Ping Shi, Jack Knill-Jones, Denise M. Lowe, Anthony J. Wilkinson, David M. Blow, Peter Brick, Paul Carter, Mary M. Y. Waye, and Greg Winter. 1985. “Hydrogen Bonding and Biological Specificity Analysed by Protein Engineering.” *Nature*.

Fesik, Stephen W. 2000. “Insights into Programmed Cell through Structural Biology.” *Cell* 103(2):273–82.

Frank, Joachim, Jun Zhu, Pawel Penczek, Yanhong Li, Suman Srivastava, Adriana Verschoor, Michael Radermacher, Robert Grassucci, Ramani K. Lata, and Rajendra K. Agrawal. 1995. “A Model of Protein Synthesis Based on Cryo-Electron Microscopy of the E. Coli Ribosome.” *Nature* 376(6539):441–44.

Geoghegan, Joan A., Rebecca M. Corrigan, Dominika T. Gruszka, Pietro Speziale, James P. O’Gara, Jennifer R. Potts, and Timothy J. Foster. 2010a. “Role of Surface Protein SasG in Biofilm Formation by Staphylococcus Aureus.” *Journal of Bacteriology* 192(21):5663–73.

Geoghegan, Joan A., Rebecca M. Corrigan, Dominika T. Gruszka, Pietro Speziale, James P. O'Gara, Jennifer R. Potts, and Timothy J. Foster. 2010b. "Role of Surface Protein SasG in Biofilm Formation by Staphylococcus Aureus." *Journal of Bacteriology* 192:5663–73.

Gianni, Stefano, Carlo Camilloni, Rajanish Giri, Angelo Toto, Daniela Bonetti, Angela Morrone, Pietro Sormanni, Maurizio Brunori, and Michele Vendruscolo. 2014. "Understanding the Frustration Arising from the Competition between Function, Misfolding, and Aggregation in a Globular Protein." *Proceedings of the National Academy of Sciences of the United States of America*.

Gianni, Stefano, Jakob Dogan, and Per Jemth. 2016. "Coupled Binding and Folding of Intrinsically Disordered Proteins: What Can We Learn from Kinetics?" *Current Opinion in Structural Biology*.

Giri, R., A. Morrone, A. Toto, M. Brunori, and S. Gianni. 2013. "Structure of the Transition State for the Binding of C-Myb and KIX Highlights an Unexpected Order for a Disordered System." *Proceedings of the National Academy of Sciences*.

Gloge, Felix, Annemarie H. Becker, Günter Kramer, and Bernd Bukau. 2014. "Co-Translational Mechanisms of Protein Maturation." *Current Opinion in Structural Biology* 24(1):24–33.

Goldman, Daniel H., Christian M. Kaiser, Anthony Milin, Maurizio Righini, Ignacio Tinoco, and Carlos Bustamante. 2015. "Mechanical Force Releases Nascent Chain-Mediated Ribosome Arrest in Vitro and in Vivo." *Science* 348(6233):457–60.

Gong, Xiao Min, Jungyuen Choi, Carla M. Franzin, Dayong Zhai, John C. Reed, and Francesca M. Marassi. 2004. "Conformation of Membrane-Associated Proapoptotic TBid." *Journal of Biological Chemistry* 279(28):28954–60.

Goujon, Mickael, Hamish McWilliam, Weizhong Li, Franck Valentin, Silvano Squizzato, Juri Paern, and Rodrigo Lopez. 2010. "A New Bioinformatics Analysis Tools Framework at EMBL-EBI." *Nucleic Acids Research* 38(SUPPL. 2).

Greene, R. F. and C. N. Pace. 1974. "Urea and Guanidine Hydrochloride Denaturation of Ribonuclease, Lysozyme, ?? Chymotrypsin, and ?? Lactoglobulin." *Journal of Biological Chemistry* 249:5388–93.

Greives, Nicholas and Huan Xiang Zhou. 2014. “Both Protein Dynamics and Ligand Concentration Can Shift the Binding Mechanism between Conformational Selection and Induced Fit.” *Proceedings of the National Academy of Sciences of the United States of America*.

Gruszka, D. T., J. A. Wojdyla, R. J. Bingham, J. P. Turkenburg, I. W. Manfield, A. Steward, A. P. Leech, J. A. Geoghegan, T. J. Foster, J. Clarke, and J. R. Potts. 2012. “PNAS Plus: Staphylococcal Biofilm-Forming Protein Has a Contiguous Rod-like Structure.” *Proceedings of the National Academy of Sciences* 109:E1011–18.

Gruszka, Dominika T., Carolina A. T. F. Mendonça, Emanuele Paci, Fiona Whelan, Judith Hawkhead, Jennifer R. Potts, and Jane Clarke. 2016. “Disorder Drives Cooperative Folding in a Multidomain Protein.” *Proceedings of the National Academy of Sciences* 113(42):11841–46.

Gruszka, Dominika T., Fiona Whelan, Oliver E. Farrance, Herman K. H. Fung, Emanuele Paci, Cy M. Jeffries, Dmitri I. Svergun, Clair Baldock, Christoph G. Baumann, David J. Brockwell, Jennifer R. Potts, and Jane Clarke. 2015. “Cooperative Folding of Intrinsically Disordered Domains Drives Assembly of a Strong Elongated Protein.” *Nature Communications* 6:7271.

Gruszka, Dominika T, Justyna a Wojdyla, Richard J. Bingham, Johan P. Turkenburg, Iain W. Manfield, Annette Steward, Andrew P. Leech, Joan a Geoghegan, Timothy J. Foster, Jane Clarke, and Jennifer R. Potts. 2012. “Staphylococcal Biofilm-Forming Protein Has a Contiguous Rod-like Structure.” *Proceedings of the National Academy of Sciences of the United States of America* 109(17):E1011-8.

Guinn, Emily J., Bharat Jagannathan, and Susan Marqusee. 2015. “Single-Molecule Chemo-Mechanical Unfolding Reveals Multiple Transition State Barriers in a Small Single-Domain Protein.” *Nature Communications* 6.

Hammarström, Per, R. Luke Wiseman, Evan T. Powers, and Jeffery W. Kelly. 2003. “Prevention of Transthyretin Amyloid Disease by Changing Protein Misfolding Energetics.” *Science (New York, N.Y.)* 299:713–16.

Hammes, Gordon G., Yu Chu Chang, and Terrence G. Oas. 2009. “Conformational Selection

or Induced Fit: A Flux Description of Reaction Mechanism." *Proceedings of the National Academy of Sciences of the United States of America*.

Harmon, Tyler S., Michael D. Crabtree, Sarah L. Shammas, Ammon E. Posey, Jane Clarke, and Rohit V. Pappu. 2016. "GADIS: Algorithm for Designing Sequences to Achieve Target Secondary Structure Profiles of Intrinsically Disordered Proteins." *Protein Engineering, Design and Selection*.

Harris, Llinos G. and R. Geoff Richards. 2006. "Staphylococci and Implant Surfaces: A Review." *Injury* 37.

Hill, Stephanie A., Lee Gyan Kwa, Sarah L. Shammas, Jennifer C. Lee, and Jane Clarke. 2014. "Mechanism of Assembly of the Non-Covalent Spectrin Tetramerization Domain from Intrinsically Disordered Partners." *Journal of Molecular Biology*.

Hinds, M. G., C. Smits, R. Fredericks-Short, J. M. Risk, M. Bailey, D. C. S. Huang, and C. L. Day. 2007. "Bim, Bad and Bmf: Intrinsically Unstructured BH3-Only Proteins That Undergo a Localized Conformational Change upon Binding to Prosurvival Bcl-2 Targets." *Cell Death and Differentiation*.

Holehouse, Alex S., Rahul K. Das, James N. Ahad, Mary O. G. Richardson, and Rohit V. Pappu. 2017. "CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins." *Biophysical Journal* 112(1):16–21.

Hunter, Christopher A. 2004. "Quantifying Intermolecular Interactions: Guidelines for the Molecular Recognition Toolbox." *Angewandte Chemie - International Edition*.

Ismail, Nurzian, Rickard Hedman, Nina Schiller, and Gunnar Von Heijne. 2012. "A Biphasic Pulling Force Acts on Transmembrane Helices during Translocon-Mediated Membrane Integration." *Nature Structural and Molecular Biology* 19(10):1018–23.

Itzhaki, Laura S., Daniel E. Otzen, and Alan R. Fersht. 1995. "The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analysed by Protein Engineering Methods: Evidence for a Nucleation-Condensation Mechanism for Protein Folding." *Journal of Molecular Biology*.

Jackson, S. E. and A. R. Fersht. 1991. "Folding of Chymotrypsin Inhibitor 2. 1. Evidence for

a Two-State Transition." *Biochemistry* 30:10428–35.

Kaiser, Christian M., Daniel H. Goldman, John D. Chodera, Ignacio Tinoco, and Carlos Bustamante. 2011. "The Ribosome Modulates Nascent Protein Folding." *Science* 334(6063):1723–27.

Kale, Justin, Elizabeth J. Osterlund, and David W. Andrews. 2018. "BCL-2 Family Proteins: Changing Partners in the Dance towards Death." *Cell Death and Differentiation* 25(1):65–80.

Kauzmann, W. 1959. "Some Factors in the Interpretation of Protein Denaturation." *Advances in Protein Chemistry* 14:1–63.

KENDREW, J. C., G. BODO, H. M. DINTZIS, R. G. PARRISH, H. WYCKOFF, and D. C. PHILLIPS. 1958. "A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis." *Nature* 181:662–66.

Kendrew, J. C., R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore. 1960. "Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 . Resolution." *Nature*.

Kiefhaber, Thomas, Annett Bachmann, and Kristine Steen Jensen. 2012. "Dynamics and Mechanisms of Coupled Protein Folding and Binding Reactions." *Current Opinion in Structural Biology*.

Kim, Annette S., Lazaros T. Kakalis, Norzehan Abdul-Manan, Grace A. Liu, and Michael K. Rosen. 2000. "Autoinhibition and Activation Mechanisms of the Wiskott-Aldrich Syndrome Protein." *Nature*.

Kirchner, Sebastian, Zhiwei Cai, Robert Rauscher, Nicolai Kastelic, Melanie Anding, Andreas Czech, Bertrand Kleizen, Lynda S. Ostedgaard, Ineke Braakman, David N. Sheppard, and Zoya Ignatova. 2017. "Alteration of Protein Function by a Silent Polymorphism Linked to TRNA Abundance." *PLoS Biology* 15(5).

Knight, Anders M., Peter H. Culviner, Neşe Kurt-Yilmaz, Taisong Zou, S. Banu Ozkan, and Silvia Cavagnero. 2013. "Electrostatic Effect of the Ribosomal Surface on Nascent Polypeptide Dynamics." *ACS Chemical Biology* 8(6):1195–1204.

Komar, Anton A. 2018. “Unraveling Co-Translational Protein Folding: Concepts and Methods.” *Methods* 137:71–81.

Korsmeyer, Stanley J. 1999. “BCL-2 Gene Family and the Regulation of Programmed Cell Death.” in *Cancer Research*.

Kouzarides, Tony. 2007. “Chromatin Modifications and Their Function.” *Cell*.

Kriwacki, Richard W., Ludger Hengst, Linda Tennant, Steven I. Reed, and Peter E. Wright. 1996. “Structural Studies of P21Waf1/Cip1/Sdi1 in the Free and Cdk2-Bound State: Conformational Disorder Mediates Binding Diversity.” *Proceedings of the National Academy of Sciences of the United States of America*.

Kudva, Renuka, Pengfei Tian, Fátima Pardo-Avila, Marta Carroni, Robert Best, Harris D. Bernstein, and Gunnar von Heijne. 2018. “The Shape of the Ribosome Exit Tunnel Affects Cotranslational Protein Folding.” *ELife*.

Kumar, Sandeep and Ruth Nussinov. 2002. “Close-Range Electrostatic Interactions in Proteins.” *ChemBioChem* 3(7):604–17.

Kurland, C. G. 1960. “Molecular Characterization of Ribonucleic Acid from Escherichia Coli Ribosomes: I. Isolation and Molecular Weights.” *Journal of Molecular Biology* 2(2):83–91.

Kvansakul, M., H. Yang, W. D. Fairlie, P. E. Czabotar, S. F. Fischer, M. A. Perugini, D. C. S. Huang, and P. M. Colman. 2008. “Vaccinia Virus Anti-Apoptotic F1L Is a Novel Bcl-2-like Domain-Swapped Dimer That Binds a Highly Selective Subset of BH3-Containing Death Ligands.” *Cell Death and Differentiation*.

L.Omonosova, E. and G. C.Hinnadurai. 2008. “BH3-Only Proteins in Apoptosis and beyond: An Overview.” *Oncogene* 27:S2–19.

Lake, James A. 1976. “Ribosome Structure Determined by Electron Microscopy of Escherichia Coli Small Subunits, Large Subunits and Monomeric Ribosomes.” *Journal of Molecular Biology* 105(1):131–59.

Lander, Eric S., Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William Fitzhugh, Roel Funke, Diane

Gage, Katrina Harris, Andrew Heaford, John Howland, Lisa Kann, Jessica Lehoczky, Rosie Levine, Paul McEwan, Kevin McKernan, James Meldrim, Jill P. Mesirov, Cher Miranda, William Morris, Jerome Naylor, Christina Raymond, Mark Rosetti, Ralph Santos, Andrew Sheridan, Carrie Sougnez, Nicole Stange-Thomann, Nikola Stojanovic, Aravind Subramanian, Dudley Wyman, Jane Rogers, John Sulston, Rachael Ainscough, Stephan Beck, David Bentley, John Burton, Christopher Clee, Nigel Carter, Alan Coulson, Rebecca Deadman, Panos Deloukas, Andrew Dunham, Ian Dunham, Richard Durbin, Lisa French, Darren Grafham, Simon Gregory, Tim Hubbard, Sean Humphray, Adrienne Hunt, Matthew Jones, Christine Lloyd, Amanda McMurray, Lucy Matthews, Simon Mercer, Sarah Milne, James C. Mullikin, Andrew Mungall, Robert Plumb, Mark Ross, Ratna Shownkeen, Sarah Sims, Robert H. Waterston, Richard K. Wilson, Ladeana W. Hillier, John D. McPherson, Marco A. Marra, Elaine R. Mardis, Lucinda A. Fulton, Asif T. Chinwalla, Kymberlie H. Pepin, Warren R. Gish, Stephanie L. Chissoe, Michael C. Wendl, Kim D. Delehaunty, Tracie L. Miner, Andrew Delehaunty, Jason B. Kramer, Lisa L. Cook, Robert S. Fulton, Douglas L. Johnson, Patrick J. Minx, Sandra W. Clifton, Trevor Hawkins, Elbert Branscomb, Paul Predki, Paul Richardson, Sarah Wenning, Tom Slezak, Norman Doggett, Jan Fang Cheng, Anne Olsen, Susan Lucas, Christopher Elkin, Edward Uberbacher, Marvin Frazier, Richard A. Gibbs, Donna M. Muzny, Steven E. Scherer, John B. Bouck, Erica J. Sodergren, Kim C. Worley, Catherine M. Rives, James H. Gorrell, Michael L. Metzker, Susan L. Naylor, Raju S. Kucherlapati, David L. Nelson, George M. Weinstock, Yoshiyuki Sakaki, Asao Fujiyama, Masahira Hattori, Tetsushi Yada, Atsushi Toyoda, Takehiko Itoh, Chiharu Kawagoe, Hidemi Watanabe, Yasushi Totoki, Todd Taylor, Jean Weissenbach, Roland Heilig, William Saurin, Francois Artiguenave, Philippe Brottier, Thomas Bruls, Eric Pelletier, Catherine Robert, Patrick Wincker, André Rosenthal, Matthias Platzer, Gerald Nyakatura, Stefan Taudien, Andreas Rump, Douglas R. Smith, Lynn Doucette-Stamm, Marc Rubenfield, Keith Weinstock, Mei Lee Hong, Joann Dubois, Huanming Yang, Jun Yu, Jian Wang, Guyang Huang, Jun Gu, Leroy Hood, Lee Rowen, Anup Madan, Shizen Qin, Ronald W. Davis, Nancy A. Federspiel, A. Pia Abola, Michael J. Proctor, Bruce A. Roe, Feng Chen, Huaqin Pan, Juliane Ramser, Hans Lehrach, Richard Reinhardt, W. Richard McCombie, Melissa De La Bastide, Neilay Dedhia, Helmut Blöcker, Klaus Hornischer, Gabriele Nordsiek, Richa Agarwala, L. Aravind, Jeffrey A. Bailey, Alex Bateman, Serafim Batzoglou, Ewan Birney, Peer Bork, Daniel G. Brown, Christopher B. Burge, Lorenzo Cerutti, Hsiu Chuan

Chen, Deanna Church, Michele Clamp, Richard R. Copley, Tobias Doerks, Sean R. Eddy, Evan E. Eichler, Terrence S. Furey, James Galagan, James G. R. Gilbert, Cyrus Harmon, Yoshihide Hayashizaki, David Haussler, Henning Hermjakob, Karsten Hokamp, Wonhee Jang, L. Steven Johnson, Thomas A. Jones, Simon Kasif, Arek Kaspryzk, Scot Kennedy, W. James Kent, Paul Kitts, Eugene V. Koonin, Ian Korf, David Kulp, Doron Lancet, Todd M. Lowe, Aoife McLysaght, Tarjei Mikkelsen, John V. Moran, Nicola Mulder, Victor J. Pollara, Chris P. Ponting, Greg Schuler, Jörg Schultz, Guy Slater, Arian F. A. Smit, Elia Stupka, Joseph Szustakowki, Danielle Thierry-Mieg, Jean Thierry-Mieg, Lukas Wagner, John Wallis, Raymond Wheeler, Alan Williams, Yuri I. Wolf, Kenneth H. Wolfe, Shiaw Pyng Yang, Ru Fang Yeh, Francis Collins, Mark S. Guyer, Jane Peterson, Adam Felsenfeld, Kris A. Wetterstrand, Richard M. Myers, Jeremy Schmutz, Mark Dickson, Jane Grimwood, David R. Cox, Maynard V. Olson, Rajinder Kaul, Christopher Raymond, Nobuyoshi Shimizu, Kazuhiko Kawasaki, Shinsei Minoshima, Glen A. Evans, Maria Athanasiou, Roger Schultz, Aristides Patrinos, and Michael J. Morgan. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature*.

Van Der Lee, Robin, Marija Buljan, Benjamin Lang, Robert J. Weatheritt, Gary W. Daughdrill, A. Keith Dunker, Monika Fuxreiter, Julian Gough, Joerg Gsponer, David T. Jones, Philip M. Kim, Richard W. Kriwacki, Christopher J. Oldfield, Rohit V. Pappu, Peter Tompa, Vladimir N. Uversky, Peter E. Wright, and M. Madan Babu. 2014. “Classification of Intrinsically Disordered Regions and Proteins.” *Chemical Reviews*.

Levinthal, C. 1968. “Are There Pathways for Protein Folding?” *Journal de Chimie Physique et de Physico-Chimie Biologique* 65:44–45.

Levinthal, Cyrus. 1969. “How to Fold Graciously.” Pp. 22–24 in *Mössbaun Spectroscopy in Biological Systems Proceedings*. Vol. 24.

Lewis, K. 2001. “Riddle of Biofilm Resistance.” *Antimicrobial Agents and Chemotherapy* 45:999–1007.

Li, Honglin, Hong Zhu, Chi Jie Xu, and Junying Yuan. 1998. “Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis.” *Cell* 94(4):491–501.

Liu, Kaixian, Joseph E. Rehfus, Elliot Mattson, and Christian M. Kaiser. 2017. “The Ribosome

Destabilizes Native and Non-Native Structures in a Nascent Multidomain Protein." *Protein Science* 26(7):1439–51.

Lovell, Simon C., Ian W. Davis, W. B. Adrendall, P. I. W. de Bakker, J. Michael Word, Michael G. Prisant, J. S. Richardson, and David C. Richardson. 2003. "Structure Validation by C Alpha Geometry: Phi,Psi and C Beta Deviation." *Proteins-Structure Function and Genetics*.

Lowe, Edward D., Ivo Tews, Kin Yip Cheng, Nick R. Brown, Sheraz Gul, Martin E. M. Noble, Steven J. Gamblin, and Louise N. Johnson. 2002. "Specificity Determinants of Recruitment Peptides Bound to Phospho-CDK2/Cyclin A." *Biochemistry*.

Lum, Ka, David Chandler, and John D. Weeks. 1999. "Hydrophobicity at Small and Large Length Scales." *Journal of Physical Chemistry B*.

Malkin, Leonard I. and Alexander Rich. 1967. "Partial Resistance of Nascent Polypeptide Chains to Proteolytic Digestion Due to Ribosomal Shielding." *Journal of Molecular Biology* 26(2):329–46.

Marino, Jacopo, Gunnar Von Heijne, and Roland Beckmann. 2016. "Small Protein Domains Fold inside the Ribosome Exit Tunnel." *FEBS Letters* 590(5):655–60.

Marsden, Andrew P., Jeffrey J. Hollins, Charles O'Neill, Pavel Ryzhov, Sally Higson, Carolina A. T. F. Mendonça, Tristan O. Kwan, Lee Gyan Kwa, Annette Steward, and Jane Clarke. 2018. "Investigating the Effect of Chain Connectivity on the Folding of a Beta-Sheet Protein On and Off the Ribosome." *Journal of Molecular Biology* 430(24):5207–16.

Mercier, Evan and Marina V. Rodnina. 2018. "Co-Translational Folding Trajectory of the HemK Helical Domain." *Biochemistry* 57(25):3460–64.

Milligan, R. A. and P. N. T. Unwin. 1986. "Location of Exit Channel for Nascent Protein in 80S Ribosome." *Nature* 319(6055):693–95.

Mirny, L. a and E. I. Shakhnovich. 1999. "Universally Conserved Positions in Protein Folds: Reading Evolutionary Signals about Stability, Folding Kinetics and Function." *Journal of Molecular Biology* 291(1):177–96.

Muchmore, Steven W., Michael Sattler, Heng Liang, Robert P. Meadows, John E. Harlan, Ho Sup Yoon, David Nettesheim, Brian S. Chang, Craig B. Thompson, Sui Lam Wong, Shi Chung Ng, and Stephen W. Fesik. 1996. "X-Ray and NMR Structure of Human Bcl-XL, an Inhibitor of Programmed Cell Death." *Nature*.

Mujtaba, Shiraz, Yan He, Lei Zeng, Sherry Yan, Olga Plotnikova, Sachchidanand, Roberto Sanchez, Nancy J. Zeleznik-Le, Ze'ev Ronai, and Ming-Ming Zhou. 2004. "Structural Mechanism of the Bromodomain of the Coactivator CBP in P53 Transcriptional Activation." *Molecular Cell* 13(2):251–63.

Muñoz, V. and L. Serrano. 1995. "Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. III. Temperature and PH Dependence." *Journal of Molecular Biology* 245(3):297–308.

Myers, J. K., C. N. Pace, and J. M. Scholtz. 1995. "Denaturant m Values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding." *Protein Science : A Publication of the Protein Society* 4:2138–48.

Nash, Piers, Xiaojing Tang, Stephen Orlicky, Qinghua Chen, Frank B. Gertler, Michael D. Mendenhall, Frank Sicheri, Tony Pawson, and Mike Tyers. 2001. "Multisite Phosphorylation of a CDK Inhibitor Sets a Threshold for the Onset of DNA Replication." *Nature*.

Nick Pace, C., J. Martin Scholtz, and Gerald R. Grimsley. 2014. "Forces Stabilizing Proteins." *FEBS Letters*.

Nilsson, Ola B., Rickard Hedman, Jacopo Marino, Stephan Wickles, Lukas Bischoff, Magnus Johansson, Annika Müller-Lucks, Fabio Trovato, Joseph D. Puglisi, Edward P. O'Brien, Roland Beckmann, and Gunnar von Heijne. 2015. "Cotranslational Protein Folding inside the Ribosome Exit Tunnel." *Cell Reports* 12(10):1533–40.

Nilsson, Ola B., Adrian A. Nickson, Jeffrey J. Hollins, Stephan Wickles, Annette Steward, Roland Beckmann, Gunnar Von Heijne, and Jane Clarke. 2017. "Cotranslational Folding of Spectrin Domains via Partially Structured States." *Nature Structural and Molecular Biology* 24(3):221–25.

Nissen, P., J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz. 2000. "The Structural Basis of

Ribosome Activity in Peptide Bond Synthesis." *Science* 289(5481):920–30.

Onuchic, José Nelson, Zaida Luthey-Schulten, and Peter G. Wolynes. 1997. "THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective." *Annual Review of Physical Chemistry*.

Otto, M. 2008. "Staphylococcal Biofilms." *Current Topics in Microbiology and Immunology* 322:207–28.

Pace, C. N. 1986. "Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves." *Methods in Enzymology* 131:266–80.

Peña-Blanco, Aida and Ana J. García-Sáez. 2018. "Bax, Bak and beyond — Mitochondrial Performance in Apoptosis." *FEBS Journal* 285(3):416–31.

Peng, Zhenling, Jing Yan, Xiao Fan, Marcin J. Mizianty, Bin Xue, Kui Wang, Gang Hu, Vladimir N. Uversky, and Lukasz Kurgan. 2014. "Exceptionally Abundant Exceptions: Comprehensive Characterization of Intrinsic Disorder in All Domains of Life." *Cellular and Molecular Life Sciences*.

Perutz, M. F. 1978. "Electrostatic Effects in Proteins." *Science* 201(4362):1187–91.

Petros, Andrew M., Edward T. Olejniczak, and Stephen W. Fesik. 2004. "Structural Biology of the Bcl-2 Family of Proteins." *Biochimica et Biophysica Acta - Molecular Cell Research* 1644(2–3):83–94.

Ptitsyn, O. B. 1973. "Stages in the Mechanism of Self-Organization of Protein Molecules." *Doklady Akademii Nauk SSSR* 210(5):1213—1215.

Radivojac, Predrag, Vladimir Vacic, Chad Haynes, Ross R. Cocklin, Amrita Mohan, Joshua W. Heyen, Mark G. Goebl, and Lilia M. Iakoucheva. 2010. "Identification, Analysis, and Prediction of Protein Ubiquitination Sites." *Proteins: Structure, Function and Bioinformatics*.

Ramachandran, G. N., C. Ramakrishnan, and V. Sasisekharan. 1963. "Stereochemistry of Polypeptide Chain Configurations." *Journal of Molecular Biology*.

Reed, John C. 1998. "Bcl-2 Family Proteins." *Oncogene*.

Roche, Fiona M., Mary Meehan, and Timothy J. Foster. 2003. "The Staphylococcus Aureus Surface Protein SasG and Its Homologues Promote Bacterial Adherence to Human Desquamated Nasal Epithelial Cells." *Microbiology* 149:2759–67.

Rogers, J. M., V. Oleinikovas, S. L. Shammas, C. T. Wong, D. De Sancho, C. M. Baker, and J. Clarke. 2014. "Interplay between Partner and Ligand Facilitates the Folding and Binding of an Intrinsically Disordered Protein." *Proceedings of the National Academy of Sciences* 111(43):15420–25.

Rogers, Joseph M., Chi T. Wong, and Jane Clarke. 2014. "Coupled Folding and Binding of the Disordered Protein PUMA Does Not Require Particular Residual Structure." *Journal of the American Chemical Society* 136(14):5197–5200.

Romero, Pedro, Zoran Obradovic, Xiaohong Li, Ethan C. Garner, Celeste J. Brown, and A. Keith Dunker. 2001. "Sequence Complexity of Disordered Protein." *Proteins: Structure, Function and Genetics*.

Roth, Charles M., Brian L. Neal, and Abraham M. Lenhoff. 1996. "Van Der Waals Interactions Involving Proteins." *Biophysical Journal*.

Rustandi, R R, D. M. Baldisseri, and D. J. Weber. 2000. "Structure of the Negative Regulatory Domain of P53 Bound to S100B(Bβ)." *Nature Structural Biology* 7(7):570–74.

Rustandi, Richard R., Donna M. Baldisseri, and David J. Weber. 2000. "Structure of the Negative Regulatory Domain of P53 Bound to S100B(Bβ)." *Nature Structural Biology*.

Samelson, Avi J., Madeleine K. Jensen, Randy A. Soto, Jamie H. D. Cate, and Susan Marqusee. 2016. "Quantitative Determination of Ribosome Nascent Chain Stability." *Proceedings of the National Academy of Sciences* 113(47):201610272.

Savage, H. J., C. J. Elliott, C. M. Freeman, and J. L. Finney. 1993. "Lost Hydrogen Bonds and Buried Surface Area: Rationalising Stability in Globular Proteins." *Journal of the Chemical Society, Faraday Transactions*.

Schneider, Caroline A., Wayne S. Rasband, and Kevin W. Eliceiri. 2012. "NIH Image to ImageJ : 25 Years of Image Analysis HISTORICAL Commentary NIH Image to ImageJ : 25 Years of Image Analysis." *Nat. Methods* 9(7):671–75.

Serrano, Luis, Jose Luis Neira, Javier Sancho, and Alan R. Fersht. 1992. "Effect of Alanine versus Glycine in α-Helices on Protein Stability." *Nature*.

Shammas, S. L., A. J. Travis, and J. Clarke. 2014. "Allostery within a Transcription Coactivator Is Predominantly Mediated through Dissociation Rate Constants." *Proceedings of the National Academy of Sciences* 111(33):12055–60.

Shammas, Sarah L., Michael D. Crabtree, Liza Dahal, Basile I. M. Wicky, and Jane Clarke. 2016. "Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies." *Journal of Biological Chemistry*.

Shammas, Sarah L., Alexandra J. Travis, and Jane Clarke. 2013. "Remarkably Fast Coupled Folding and Binding of the Intrinsically Disordered Transactivation Domain of CMyb to CBP KIX." *Journal of Physical Chemistry B*.

Shaw, Kevin L., Gerald R. Grimsley, Gennady I. Yakovlev, Alexander A. Makarov, and C. Nick Pace. 2001. "The Effect of Net Charge on the Solubility, Activity, and Stability of Ribonuclease Sa." *Protein Science* 10(6):1206–15.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D. Thompson, and Desmond G. Higgins. 2011. "Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology* 7.

Smits, Callum, Peter E. Czabotar, Mark G. Hinds, and Catherine L. Day. 2008. "Structural Plasticity Underpins Promiscuous Binding of the Prosurvival Protein A1." *Structure* 16(5):818–29.

Sosnick, Tobin R. and Doug Barrick. 2011. "The Folding of Single Domain Proteins-Have We Reached a Consensus?" *Current Opinion in Structural Biology* 21(1):12–24.

Steward, Philip S. and J. William Costdrton. 2001. "Antibiotic Resistance of Baqteria in Biofilms." *The Lancet* 358(9276):135–38.

Stewart, David E., Atom Sarkar, and John E. Wampler. 1990. "Occurrence and Role Ofcis Peptide Bonds in Protein Structures." *Journal of Molecular Biology*.

Su, Ting, Jingdong Cheng, Daniel Sohmen, Rickard Hedman, Otto Berninghausen, Gunnar von Heijne, Daniel N. Wilson, and Roland Beckmann. 2017. "The Force-Sensing Peptide VemP Employs Extreme Compaction and Secondary Structure Formation to Induce Ribosomal Stalling." *ELife* 6.

Tanford, C. 1970. "Protein Denaturation." *Adv. Protein Chem.* 24:1–95.

Taverna, Darin M. and Richard A. Goldstein. 2002. "Why Are Proteins Marginally Stable?" *Proteins: Structure, Function and Genetics*.

Theillet, Francois-Xavier, Lajos Kalmar, Peter Tompa, Kyou-Hoon Han, Philipp Selenko, A. Keith Dunker, Gary W. Daughdrill, and Vladimir N. Uversky. 2013. "The Alphabet of Intrinsic Disorder: I. Act like a Pro: On the Abundance and Roles of Proline Residues in Intrinsically Disordered Proteins." *Intrinsically Disordered Proteins*.

Thommen, Michael, Wolf Holtkamp, and Marina V. Rodnina. 2017. "Co-Translational Protein Folding: Progress and Methods." *Current Opinion in Structural Biology* 42:83–89.

Tian, Pengfei, Annette Steward, Renuka Kudva, Ting Su, Patrick J. Shilling, Adrian A. Nickson, Jeffrey J. Hollins, Roland Beckmann, Gunnar von Heijne, Jane Clarke, and Robert B. Best. 2018. "Folding Pathway of an Ig Domain Is Conserved on and off the Ribosome." *Proceedings of the National Academy of Sciences* 115(48):E11284–93.

Tokuriki, Nobuhiko, Francois Stricher, Joost Schymkowitz, Luis Serrano, and Dan S. Tawfik. 2007. "The Stability Effects of Protein Mutations Appear to Be Universally Distributed." *Journal of Molecular Biology*.

Tompa, P., J. Prilusky, I. Silman, and J. L. Sussman. 2008. "Structural Disorder Serves as a Weak Signal for Intracellular Protein Degradation." *Proteins: Structure, Function and Genetics*.

Toto, Angelo, Rajanish Giri, Maurizio Brunori, and Stefano Gianni. 2014. "The Mechanism of Binding of the KIX Domain to the Mixed Lineage Leukemia Protein and Its Allosteric Role in the Recognition of C-Myb." *Protein Science*.

Tu, Li Wei and Carol Deutsch. 2017. "Determinants of Helix Formation for a Kv1.3 Transmembrane Segment inside the Ribosome Exit Tunnel." *Journal of Molecular*

Biology 429(11):1722–32.

Uversky, Vladimir N. 2002. "What Does It Mean to Be Natively Unfolded?" *European Journal of Biochemistry* 269(1):2–12.

Uversky, Vladimir N., Joel R. Gillespie, and Anthony L. Fink. 2000. "Why Are 'natively Unfolded' Proteins Unstructured under Physiologic Conditions?" *Proteins: Structure, Function and Genetics*.

Uversky, Vladimir N., Christopher J. Oldfield, and a Keith Dunker. 2008. "Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept." *Annual Review of Biophysics* 37(1):215–46.

Voorhees, Rebecca M., Israel S. Fernández, Sjors H. W. Scheres, and Ramanujan S. Hegde. 2014. "Structure of the Mammalian Ribosome-Sec61 Complex to 3.4 Å Resolution." *Cell* 157(7):1632–43.

Wang, Yu and Nico Tjandra. 2013. "Structural Insights of TBid, the Caspase-8-Activated Bid, and Its BH3domain." *Journal of Biological Chemistry* 288(50):35840–51.

Ward, JJ, JS Sodhi, and LJ McGuffin. 2004. "Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life." *Journal of Molecular Biology* 337(3):635–45.

Washington, N., R. J. C. Steele, S. J. Jackson, D. Bush, J. Mason, D. A. Gill, K. Pitt, and D. A. Rawlins. 2000. "Determination of Baseline Human Nasal PH and the Effect of Intranasally Administered Buffers." *International Journal of Pharmaceutics* 198(2):139–46.

WATSON, J. D. and F. H. C. CRICK. 1953. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." *Nature* 171(4356):737–38.

Wilson, Daniel N., Shashi Bhushan, Thomas Becker, and Roland Beckmann. 2011. "Nascent Polypeptide Chains within the Ribosomal Tunnel Analyzed by Cryo-EM." Pp. 393–404 in *Ribosomes*.

Wright, Peter E. and H. Jane Dyson. 1999. "Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm." *Journal of Molecular Biology*.

Wright, Peter E. and H. Jane Dyson. 2015. “Intrinsically Disordered Proteins in Cellular Signalling and Regulation.” *Nature Reviews Molecular Cell Biology*.

Xue, B., R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky. 2010. “PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids.” *Biochimica et Biophysica Acta* 1804(4):996–1010.

Xue, Bin, A. Keith Dunker, and Vladimir N. Uversky. 2012. “Orderly Order in Protein Intrinsic Disorder Distribution: Disorder in 3500 Proteomes from Viruses and the Three Domains of Life.” *Journal of Biomolecular Structure and Dynamics*.

Yao, Yong, Andrey A. Bobkov, Leigh A. Plesniak, and Francesca M. Marassi. 2009. “Mapping the Interaction of Pro-Apoptotic TBID with pro-Survival BCL-XL.” *Biochemistry* 48(36):8704–11.

Yap, Mee Ngan and Harris D. Bernstein. 2009. “The Plasticity of a Translation Arrest Motif Yields Insights into Nascent Polypeptide Recognition inside the Ribosome Tunnel.” *Molecular Cell* 34(2):201–11.

Zhang, Jun, Xijiang Pan, Kaige Yan, Shan Sun, Ning Gao, and Sen Fang Sui. 2015. “Mechanisms of Ribosome Stalling by SecM at Multiple Elongation Steps.” *ELife*.

Zhou, Huan Xiang. 2012. “Intrinsic Disorder: Signaling via Highly Specific but Short-Lived Association.” *Trends in Biochemical Sciences*.

Zhou, Huan Xiang and Xiaodong Pang. 2018. “Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation.” *Chemical Reviews*.