

https://dspace.org/

Unfolding mechanism of human glutathione transferase M1a-1a

Wiid, Kimberly Jade

2018

https://hdl.handle.net/10539/25749

Wiid, Kimberley Jade (2018) Unfolding mechanism of human glutathione transferase M1a-1a, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/25749>

Downloaded from a DSpace repository. https://dspace.org/



Unfolding mechanism of human glutathione transferase M1a-1a

Kimberley Jade Wiid
360483

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg in fulfilment of the requirements for the degree of Doctor of Philosophy

May 2018

Declaration

I, Kimberley Jade Wiid (student number: 360483), am a student registered for the degree of Doctor of Philosophy (PhD) in the academic year 2018.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this, or any other University.
- The information used in the Thesis has not been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature:  Date: 28 May 2018

Abstract

Proteins exist in equilibrium between the native (N) and the denatured (D) states. In order to form the biologically active native state, the amino acid sequence has to fold to form a stable three-dimensional structure. The large scientific community of biochemists and biophysicists has not yet been able to gain a complete understanding of this process. In this study, the unfolding of the homodimeric detoxification enzyme hGST M1a-1a (WT dimer) was investigated. Additionally, an F56S/R81A double-mutant (mutant monomer) was engineered to create a monomeric form of the protein. The mutant monomer was used to gain a better understanding of the unfolding events occurring at the subunit level, in the absence of quaternary interactions. Data from various techniques indicate the mutant monomer to closely resemble the tertiary structure of the subunits in the WT homodimer, making it a suitable model to study the unfolding mechanism of hGST M1a in the absence of quaternary interactions. A four-state equilibrium unfolding mechanism, involving two stable intermediate species, is proposed. HDX-MS studies indicate that disruption of the conserved lock-and-key motif, as well as the structures surrounding the mu loop, results in a destabilisation of domain 1. However, dimer dissociation cannot occur until the mixed charge cluster at the dimer interface has been destabilised. The destabilisation of domain 1 results in destabilisation of α4 and α5 in domain 2, because the domains unfold in a concerted manner. hGST M1a-1a dissociates to form monomeric intermediate (M), with weak interdomain interactions and compromised short-range contacts. The unstable M intermediate self-associates to form an oligomeric intermediate (I). The destabilisation of α6 and α7 in the hydrophobic core of domain 2 drives the formation of the partially structured denatured state. Further investigation will need to be pursued to determine whether hGST M1a-1a unfolds via transient intermediate states; however, the elucidation of the equilibrium unfolding pathway of a complex homodimeric protein is a valuable addition to the ever-growing knowledge base of protein folding.

This work is dedicated to those I love most:

My Mooks, Mirella Wiid, all that I am or ever hope to be is because of your beautiful light

My Nonna, Mara Molino, who has always shown me the value of knowledge

My brother, Kyle Wiid, who truly is the wind beneath my wings

My Zon, who is my best friend, my biggest fan, my partner in science and in life

"Around here, however, we don't look backwards for very long. We keep moving forward, opening up new doors and doing new things, because we're curious...and curiosity keeps leading us down new paths."

-Walt Disney

Acknowledgements

Thank you to my supervisor, Professor Heini Dirr, for your support, guidance and unwavering belief in my abilities throughout my postgraduate studies. Your passion for science is contagious. It has been a privilege and an honour to learn from you.

Thank you to my co-supervisor, Dr Derryn Legg-E'Silva, for your encouragement, advice and friendship. I truly appreciate your tireless assistance throughout the duration of my postgraduate studies.

Thank you to both Dr Stoyan Stoychev and Dr Previn Naicker at the CSIR for their invaluable intellectual and technical contribution to the HDX-MS experimental aspect of this research.

Thank you to Dr Terence Van Zyl and Jake Zondagh for writing the Python script that helped me with the initial data visualisation of the HDX-MS component of this work. I appreciate all of your time and effort.

Thank you to Professor Osman Bilsel at the University of Massachusetts, Boston, for his assistance with the Savuka fitting software. Your kind assistance is truly appreciated.

Thank you to my friends and colleagues in the Protein Structure-Function Research Unit. You have all taught me so much. A special thank you to Professor Yasien Sayed and Dr Sylvia Fanucchi for the many invaluable intellectual conversations – you both helped me more than you know.

Thank you to the National Research Foundation and the University of the Witwatersrand for the generous financial support.

And lastly, thank you to my family and close friends. The patience, kindness and love you have all shown me throughout this journey have really kept me going. Your enthusiasm and support has been inspirational.

Conference outputs

1. Molecular Biosciences Research Thrust, Johannesburg, 2017. *Oral presentation*. "**Human glutathione transferase M1a-1a unfolds/refolds via intermediates.**" Kimberley Wiid, Derryn Legg-E'Silva, Stoyan Stoychev, Previn Naicker and Heinrich Dirr.

2. University of the Witwatersrand Cross Faculty Postgraduate Symposium, Johannesburg, 2017. *Oral presentation*: "**Equilibrium unfolding/refolding of human glutathione transferase M1a-1a.**" Kimberley Wiid, Derryn Legg-E'Silva, Stoyan Stoychev, Previn Naicker and Heinrich Dirr.

3. European Molecular Biology Organisation at the University of Kent, UK; June 2015; **Application of transient kinetic methods to biological macromolecules** (Participant).

Table of contents

Declaration i

Abstract ii

Acknowledgements iv

Conference outputs v

List of abbreviations ix

List of figures xii

Chapter 1: Introduction 1

1.1. Multidomain protein assemblies 1

1.1.1. Protein-protein interface interactions 2

1.1.1.1. Geometric complementarity 2

1.1.1.2. Physiochemical complementarity 3

1.1.2. Charge clusters 4

1.1.3. Domain interface interactions 5

1.2. Protein folding 7

1.2.1. Folding models 7

1.2.2. Levinthal, energy landscapes and foldons 9

1.2.3. Protein folding pathway 11

1.2.3.1. The native state 11

1.2.3.2. The intermediate state 12

1.2.3.3. The unfolded state 15

1.3. Protein unfolding 16

1.3.1. The push-and-pull hypothesis 16

1.3.2. Equilibrium unfolding 17

1.3.3. Unfolding kinetics 19

1.4. Human glutathione transferase M1a-1a 21

1.4.1. Structure 21

1.4.2. Function 23

1.4.3. Intersubunit interactions 24

1.4.4. Conformational stability and folding 26
1.4.5. A modular approach to a complex investigation 29
1.5. Aim and objectives 30
Chapter 2: Experimental Procedures 31
2.1. Materials 31
2.2. Plasmid preparation 31
2.2.1. Wild-type 31
2.2.2. F56S/R81A mutant 32
2.3. Protein overexpression and purification 33
2.3.1. Protein concentration determination 34
2.3.2. SDS-PAGE 34
2.4. Structural characterisation of the WT and F56S/R81A mutant 34
2.4.1. Size exclusion high-performance liquid chromatography (SE-HPLC) 34
2.4.2. Far-UV circular dichroism 35
2.4.3. Intrinsic fluorescence 35
2.5. Functional characterisation of the WT and F56S/R81A mutant 36
2.5.1. Specific activity 36
2.5.2. ANS binding 36
2.6. Urea-induced equilibrium unfolding 37
2.6.1. Reversibility of unfolding 37
2.6.2. Urea-induced equilibrium unfolding spectroscopic probes 38
2.6.3. Data fitting 39
2.7. Pulsed labelling hydrogen-deuterium exchange mass spectrometry (HDX-MS) 39
2.7.1. Data analysis 40
Chapter 3: Results 42
3.1. Plasmid sequence verification 42
3.2. WT and F56S/R81A mutant protein purification 42
3.3. Structural characterisation of WT and F56S/R81A mutant 42
3.3.1. SDS-PAGE 42

3.3.2. SE-HPLC 46
3.3.3. Far-UV circular dichroism 49
3.3.4. Intrinsic fluorescence 49
3.4. Functional characterisation of WT dimer and mutant monomer 52
3.4.1. Specific activity 52
3.4.2. ANS binding 52
3.5. Urea-induced equilibrium unfolding 54
3.5.1. Recovery and reversibility of WT dimer and mutant monomer unfolding 54
3.5.2. Unfolding monitored by tryptophan fluorescence and far-UV CD 55
3.5.3. Protein-concentration dependent unfolding 59
3.5.4. Unfolding monitored by ANS binding 61
3.5.5. Unfolding monitored by SE-HPLC 61
3.5.6. Unfolding monitored by DLS 64
3.5.7. Structure of the intermediates 67
3.5.8. Data fitting of urea-induced equilibrium unfolding 70
3.6. Pulsed-labelling HDX-MS 74
Chapter 4: Discussion 88
4.1. The mutant monomer is a suitable model of the WT dimer subunit 88
4.2. Unfolding proceeds via two stable equilibrium intermediates 92
4.3. Proposed equilibrium unfolding pathway of hGST M1a-1a 99
4.4. Conclusions 101
References 102
APPENDIX 130

List of abbreviations

Å	Angstrom
ANS	8-anilino-1-naphthalene sulfonate
ΔASA	change in solvent accessible surface area
°C	degrees Celsius
C_{m1}	urea-induced midpoint of unfolding event one
C_{m2}	urea-induced midpoint of unfolding event two
C_{mtotal}	urea-induced midpoint of unfolding
CDNB	1-chloro-2,4-dinitrobenzene
CD	circular dichroism
D_2O	deuterium oxide
Da	dalton
D	denatured protein
DMG	dry molten globule
D_{Mutant}	denatured mutant monomer
D_{WT}	denatured wild-type dimer
DLS	dynamic light scattering
ε	molar extinction coefficient
EDTA	ethylenediaminetetraacetic acid
ELT	energy landscape theory
F56S/R81A	a double mutant in which residue F56 is substituted with S, and residue R81 substituted with A
FD	fully deuterated
$\Delta G(H_2O)_1$	change in Gibbs free energy in the absence of denaturant for equilibrium unfolding event one
$\Delta G(H_2O)_2$	change in Gibbs free energy in the absence of denaturant for equilibrium unfolding event two
$\Delta G(H_2O)_{total}$	change in Gibbs free energy in the absence of denaturant for equilibrium unfolding
G-site	glutathione-binding site
GSH	reduced glutathione
GSSG	oxidised glutathione
GST	glutathione *S*-transferase
HDX-MS	hydrogen-deuterium exchange mass spectrometry

His_6	hexahistidine
H-site	hydrophobic electrophile-binding site
hGST A1-1	human class Alpha glutathione *S*-transferase with two type one subunits
hGST M1a-1a	human class Mu glutathione *S*-transferase with two type one subunits
hGST P1-1	human class Pi glutathione *S*-transferase with two type one subunits
M_{Mutant}	native mutant monomer
M_{WT}	monomeric equilibrium intermediate formed by wild-type dimer hGST M1a-1a
I_{WT}	oligomeric equilibrium intermediate formed by wild-type dimer hGST M1a-1a
I_{Mutant}	oligomeric equilibrium intermediate formed by F56S/R81A mutant hGST M1a
IPTG	isopropyl β-D-1-thiogalactopyranoside
K	kelvin
LB	Luria-Bertani
LMW	low molecular weight
m_1-value	susceptibility of protein to urea in unfolding event one
m_2-value	susceptibility of protein to urea in unfolding event two
m_{total}-value	susceptibility of protein to urea denaturation
MG	molten globule
N	native protein
N_{WT}	native wild-type dimer
nm	nanometer
M	molar
OD	optical density
PDB	Protein Data Bank
p*I*	isoelectric point
rmp	revolutions per minute
ppm	parts per million
rGST M1-1	rat class Mu glutathione *S*-transferase with two type one subunits
s	seconds
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE-HPLC	size exclusion high-performance liquid chromatography
T	temperature
U	unfolded

UD	undeuterated
UV	ultraviolet
WMG	wet molten globule
WT dimer	dimeric human class Mu glutathione *S*-transferase with two type one subunits

Amino acids are named according to the IUPAC single and three letter codes

List of figures

Figure 1: Visual representation of protein folding concepts 8
Figure 2: Structural features of hGST M1a-1a 22
Figure 3: Structure-based sequence alignment of related GSTs 27
Figure 4: Sequence identity of WT and F56S/R81A mutant 43
Figure 5: Elution profiles of the WT protein 44
Figure 6: Elution profiles of the F56S/R81A mutant protein 45
Figure 7: SDS-PAGE analysis of WT and F56S/R81A mutant 47
Figure 8: SE-HPLC elution profiles of WT and F56S/R81A mutant 48
Figure 9: Far-UV CD spectra for WT dimer and mutant monomer 50
Figure 10: Fluorescence emission spectra for WT dimer and mutant monomer 51
Figure 11: Functional characterisation of WT dimer and mutant monomer 53
Figure 12: Recovery of the tertiary and secondary structure of WT dimer and mutant monomer 56
Figure 13: Reversibility of WT dimer and mutant monomer unfolding 57
Figure 14: Urea-induced equilibrium unfolding of WT dimer and mutant monomer monitored by intrinsic tryptophan fluorescence and ellipticity at 222 nm 58
Figure 15: Protein concentration-dependence of WT dimer and mutant monomer unfolding 60
Figure 16: Urea-induced equilibrium unfolding of WT dimer and mutant monomer monitored by ANS binding 62
Figure 17: SE-HPLC elution profiles of WT dimer and mutant monomer incubated in 0 M to 8 M urea 63
Figure 18: SE-HPLC retention time of each peak eluted for the WT dimer and mutant monomer when unfolded in 0 M - 8 M urea 65
Figure 19: Urea-induced equilibrium unfolding of WT dimer and mutant monomer monitored by DLS 66
Figure 20: Secondary structure of WT dimer and mutant monomer incubated in various concentrations of urea 68
Figure 21: Tertiary structure of WT dimer and mutant monomer incubated in various concentrations of urea 69
Figure 22: Global fitting of the WT dimer unfolding events 71
Figure 23: Global fitting of the mutant monomer unfolding events 72
Figure 24: Structural location and mass spectra of peptic fragment number 10 (residues 54-79) 76

Figure 25: Structural location and mass spectra of peptic fragment number 14 (residues 82-115) 77
Figure 26: Structural location and mass spectra of peptic fragment number 18 (residues 116-139) 78
Figure 27: Structural location and mass spectra of peptic fragment number 22 (residues 167-185) 79
Figure 28: The percentage deuteration for each peptide at varying urea concentrations for WT dimer and mutant monomer 82
Figure 29: Deuterium incorporation for the 30 common peptides obtained under urea-induced equilibrium unfolding conditions 84
Figure 30: Population of unfolded WT dimer and mutant monomer in the three unfolding groups 87
Figure 31: Conformational stability of native WT dimer and native mutant monomer 90

List of tables

Table 1: Thermodynamic parameters obtained from urea-induced equilibrium unfolding of WT dimer and mutant monomer.....73
Table 2: The Cm values for the 30 common peptides obtained for WT dimer and mutant monomer unfolding.....85

Chapter 1: Introduction

The chemical reactions that form the basis of life are governed by proteins, the most diverse and versatile molecules found within living systems. Proteins are also used in numerous biomedical, agricultural, industrial and environmental applications. However, the vast majority of systems that utilise proteins rely on the three-dimensional folded, functional state of the protein.

Major uncertainties exist in understanding the complex structural rearrangement of the primary amino acid sequence that results in the formations of secondary, tertiary and quaternary structural elements, to form well-defined and functional structures. Additionally, the majority of protein folding studies have been conducted on small monomeric proteins, and not on multidomain proteins. Although the study of simpler monomeric proteins has been of great practical value, 80% of the eukaryotic proteome is composed of multidomain proteins. Elucidation of the processes that govern multidomain protein folding will not only enable researchers to gain an understanding of the vast majority of biological processes, but will also provide insights into many disease states (Chiti and Dobson, 2006; de Oliveira and Silva, 2017; Ferguson and Fersht, 2003; Hartl, 2010).

1.1. Multidomain protein assemblies

Multidomain proteins control the regulation of several highly specific biochemical reactions within the cells of all living organisms (Kumar *et al.*, 2017; McCammon, 1998; Vogel *et al.*, 2004). Complex multidomain structures are formed through the extensive rearrangement and duplication of individual domains (Han *et al.*, 2007). Oligomerisation of these evolved multidomain structures not only enhances catalytic properties through the evolution of novel functions, but also provides an additional mechanism for stabilisation by shielding unfavourable hydrophobic surfaces from solvent (Ali and Imperiali, 2005; Dautant *et al.*, 2017; Janin *et al.*, 1988).

Multidomain dimeric proteins may be classified either as homodimeric or heterodimeric, with either heterologous or isologous subunit interface organisation (Monod *et al.*, 1965; Nooren and Thornton, 2003). Heterologous interfaces consist of interacting subunits that display non-identical surfaces, whereas isologous interfaces consist of two identical interacting subunits and can, therefore, only occur in dimeric complexes (Mosca *et al.*, 2014; Nooren and Thornton, 2003). Furthermore, oligomeric assemblies may be obligate complexes. Or non-obligate complexes (Nooren and Thornton, 2003). Obligate complexes, like the interleukin 5

cytokine dimer (Acuner Ozbabacan *et al.,* 2011), are permanent associations that require the association of subunits to function (Nooren and Thornton, 2003). On the other hand, non-obligate complexes are composed of stable, independently functioning interacting partners with either permanent (e.g. an enzyme-inhibitor complex) or transient associations (e.g. heterotrimeric G protein) (Jones and Thornton, 1996; Nooren and Thornton, 2003).

Oligomeric assemblies show an increase in the stability of individual subunit tertiary structures, as well as novel structural and functional elements. The evolution of novel elements occurs as a result of protein-protein interactions that are present both between and within oligomers. Protein interactions are precise and highly evolved networks that contribute towards the complexity of protein systems (Goncearenco *et al.*, 2014), while still maintaining a critical role in the correct folding and functioning of these complexes (George and Heringa, 2002; Jones and Thornton, 1995; Misra, 2017)

1.1.1. Protein-protein interface interactions

Protein-protein interfaces can occur between domains to form the tertiary structure and/or between subunits to form the quaternary structure (Han *et al.*, 2007). Characterisation of both the intra-molecular interactions (which govern the manner in which the polypeptide chain associates with itself), and the inter-molecular interactions (which govern the manner in which the polypeptide chain associates with domains and subunits), is essential in understanding protein dimerisation (Nooren and Thornton, 2003). A high degree of specificity is required for the formation of the correct molecular associations, and for correct dimerisation (Brinda *et al.,* 2002; Griffin and Gerrard, 2012; Janin *et al.*, 1988). These interactions are crucial in the construction and folding of structural elements, and the formation of these interactions requires the interacting surfaces to have complementary shape and charge (Jaenicke, 1991; Jones and Thornton, 1996; Larsen *et al.*, 1998).

1.1.1.1. Geometric complementarity

The geometric interlocking, charge reciprocation and amino acid side chain packing at the subunit-subunit interface is referred to as molecular complementarity (Jones and Thornton, 1996). Protein subunits that exhibit complementary molecular surfaces reduce the solvent-exposed surface area between the subunits, allowing an increase in intramolecular hydrophobic effects. Intramolecular hydrophobic interactions are crucial in the correct orientation of individual subunits for the formation of complementary van der Waals interactions, hydrogen bonds and other electrostatic interactions during oligomerisation (Kuroda and Gray, 2016; Meyer *et al.*, 1996; Novak, 2014).

Hydrogen bonds and salt bridges confer specificity to the association between subunits; however, geometric interface complementarity is required for complete specificity and association (Fersht *et al.*, 1985; Larsen *et al.*, 1998). The size of the buried surface area upon association, the packing density at the interface, the paucity of buried water molecules as well as the contour correlation between subunits are all defining features of geometric complementarity (Chothia and Janin, 1975; Janin *et al.*, 1988; Kuroda and Gray, 2016).

1.1.1.2. Physiochemical complementarity

Hydrophilic interactions, hydrophobic interactions, amino acid composition and electrostatic interactions are the fundamental molecular interactions that govern the subunit interface (Bahadur *et al.*, 2003; Janin *et al.*, 1988; Tsai *et al.*, 1997). Understanding their role in subunit-subunit association and dissociation is crucial, because these interactions are the determinants of proper folding, and hence the functioning, of proteins.

The hydrophobic effect

The hydrophobic effect is a result of the absence of hydrogen bonding between water and other non-polar groups dispersed amongst the hydrophobic patches of a protein (Dill, 1990). The contribution of the hydrophobic effect is more dominant in protein folding, but it is also essential in the stabilisation because the burial of non-polar amino acids via protein-protein associations is energetically-favourable (Balaji, 2015; Tsai *et al.*, 1997)

Protein interfaces may be composed of a central hydrophobic core surrounded by a hydrophilic charged region (Janin *et al.*, 1988), or proteins may have numerous small hydrophobic patches with hydrogen bonds across the subunit interface (Larsen *et al.*, 1998; Zhu and Karlin, 1996). The burial of non-polar amino acids is critical in oligomerisation, and the subunit-subunit interfaces of homodimers have been found to contain twice the buried surface area compared to other protein complexes (Bahadur *et al.*, 2003).

Amino acid composition

The amino acid residues located along interacting surfaces within a protein are fundamental in determining the efficacy of the interactions occurring between two subunits (Cunningham *et al.*, 1991; Fersht, 1984). Intersubunit interfaces are largely composed of charged residues, with 33% of the interface area consisting of the non-polar residues leucine, isoleucine, valine, cysteine, methionine and phenylalanine (Bahadur *et al.*, 2003; Janin *et al.*, 1988) whereas charged residues such as aspartic acid and glutamic acid contribute 14% to the interface area (Tsai *et al.*, 1997). Despite the apolar nature of interfaces, arginine and leucine are the most

abundant interface-residues. Arginine is typically involved in numerous salt bridges across interfaces and also contributes the majority of hydrogen bonds across the subunit interface (Janin *et al.*, 1988; Tsai *et al.*, 1997). Amino acids can also confer stability because it has been shown that amino acid residues that protrude from one subunit into the surrounding area of an interacting subunit are implicated in securing the subunit-subunit association due to the electrostatic interactions occurring between these amino acids (Cunningham *et al.*, 1991; Heringa and Argos, 1991).

Electrostatic complementarity

Electrostatic interactions such as salt bridges and hydrogen bonds are crucial in the formation of specific interactions (Fersht, 1984; Fersht *et al.*, 1985). The relative positioning of charged side-chain residues at each interacting surface contributes towards complementarity (Jones and Thornton, 1996). Interactions between the N-terminal amide and C-terminal carboxyl group as well as interactions between charged chemical groups of ionisable residues such as arginine, histidine, lysine, aspartic acid and glutamic acid are responsible for the electrostatic interactions that take place at the subunit-subunit interface (Jones and Thornton, 1995; Wodak *et al.*, 2015).

Electrostatic interactions may be repulsive or attractive, and they may serve to either stabilise or destabilise protein-protein interactions (Fersht, 1984). Destabilising effects are a result of repulsion between groups with similar charges, whereas stabilising effects are as a result of the formation of salt bridges (ion pairs) between groups that are oppositely charged and within 4 Å of one another (Barlow and Thornton, 1983). Electrostatic interactions may also occur between residues that are further apart from one another on the polypeptide chain, and between residues of two different polypeptide chains via a charge cluster, as is the case with subunit-subunit electrostatic interactions (Barlow and Thornton, 1983; Brinda *et al.*, 2002).

1.1.2. Charge clusters

A charge cluster describes the increased distribution of positively charged (lysine and arginine) or negatively charged (glutamic acid and aspartic acid) residues in a protein structure (Zhu and Karlin, 1996). Positive, negative and mixed charge clusters make more numerous contacts with residues within the cluster as opposed to residues in surrounding areas (Heringa and Argos, 1991) and these clusters confer highly specific functions to proteins.

Charged clusters are often found at the surface of tertiary or quaternary structures, and function to stabilise the protein conformation (Kharrat *et al.*, 2016; Kumar and Nussinov, 2002). Mixed-charge clusters are involved in the formation of multidomain complexes and have been suggested as potential mediators of highly specific protein-protein interactions (Zhu and Karlin, 1996). Additionally, these clusters and have been shown to stabilise multidomain proteins through the formation of quaternary structure (Thompson *et al.*, 2006). The residues that form charged clusters have been identified as key residues involved in dimerisation (Brinda *et al.*, 2002) because they have been shown to force secondary structural elements of a protein to fold either by inward collapse, or through the development of a central interaction core (Heringa and Argos, 1991). Clusters responsible for dimerisation are scarce; however, charge clusters have been found near the centre of subunit-subunit interfaces, and the amino acid residues within these clusters are inaccessible to solvent (Zhu and Karlin, 1996).

Amino acid residues in charge clusters facilitate long-range interactions through side-chain clustering (Heringa and Argos, 1991). Clusters are more commonly composed of larger residues such as arginine, tyrosine, histidine, tryptophan, glutamic acid, phenylalanine and glutamine because a greater degree of contact is possible amongst these groups (Heringa and Argos, 1991). Polar residues are more commonly found in charge clusters than hydrophobic residues. Additionally, amino acids containing larger numbers of methyl groups are highly favoured to facilitate maximum surface contact for recognition and interaction (Heringa and Argos, 1991).

1.1.3. Domain interface interactions

Domains can be described as compact, local, semi-independent structural units within proteins (Richardson, 1981). Many proteins are composed of domains, and multidomain proteins often display domain repeats, particularly in eukaryotic proteomes (Bhaskara *et al.*, 2013; Schüler and Bornberg-Bauer, 2016; Vogel *et al.*, 2004). It was previously thought that 13 % of proteins contained domain-repeats (Marcotte *et al.*, 1999), but more recently it has been shown that 25 % of proteins are repeat-containing (Pellegrini *et al.*, 2012).

The domain interface includes the buried surface area created during the association of domains and the interactions formed between domains (McCammon, 1998). The degree of interdomain interactions is dependent on the surface area of the domain interface (Jones *et al.*, 2000). Understanding domain interface interactions is important because domains are highly conserved and they fold and associate to confer biological function (Bhaskara and

Srinivasan, 2011; Janin *et al.*, 1988). The rearrangement and association of conserved domains assists the ordered assembly of numerous protein systems (Marsh *et al.*, 2013).

Hydrophobic cores within domains enable ordered assembly through the cooperative folding of numerous sites along the polypeptide chain, thereby enhancing the folding rate of a protein (Jaenicke, 1999; Marsh *et al.*, 2013). Domains act as cooperative thermodynamic units during protein folding, and some domains even exhibit autonomous folding trajectories with distinct equilibrium unfolding transitions during equilibrium unfolding (Privalov, 1979; Rowe and Tanford, 1973). Cooperativity between domains is dependent on the structure, and the topology of a protein, and cooperative folding of multidomain proteins is reliant on dynamic coupling and communication between stabilising interactions (Itoh and Sasai, 2008; Malhotra and Udgaonkar, 2016a; Seelig, 2018). Non-cooperative domains show three to four states at equilibrium, whereas cooperative domains show only two states (Jaenicke, 1999), but cooperativity is not necessary for all proteins to fold correctly (Bhutani and Udgaonkar, 2003; Lakshmikanth *et al.*, 2001; Malhotra and Udgaonkar, 2016b). It is possible for an overall reaction to show reduced cooperativity even though the events that constitute the reaction are fully cooperative as shown by the reduced cooperativity of multi-state transitions with discrete intermediates (Malhotra and Udgaonkar, 2016a).

Determining the folding trajectory of a multidomain protein is complex because determining contact boundaries for individual domains remains challenging (Jaenicke, 1999). However, residues at the domain interface are conserved between protein families and these residues are responsible for the majority of domain interface interactions during protein folding (Valdar and Thornton, 2001). Studies conducted on numerous protein classes within a superfamily can, therefore, provide valuable information on the nature of both interdomain and intersubunit contacts, which will ultimately contribute to the knowledge base of how protein-protein interactions impact on the stability and folding of proteins.

1.2. Protein folding

The exact mechanisms behind protein folding are not fully understood; however, extensive research has elucidated the driving forces that are responsible for protein folding. A large entropic penalty is involved when disordered chains form the highly defined structure, and a large enthalpic penalty exists as a result of solvent interaction disruption (Anfinsen, 1973). Intramolecular contacts such as hydrogen bonding and electrostatic interactions stabilise protein structures, but hydrophobic interactions assist the folding process by counteracting the unfavourable entropic and enthalpic penalties (see section 1.1.3.1.).

The amino acid sequence determines the driving hydrophobic force, as well as the stabilising electrostatic forces and hydrogen bonds within a protein. The importance of the primary structure has been thoroughly considered. Forty-five years ago, research conducted by Anfinsen proved that the information required for bovine pancreatic nuclease (RNase) to rearrange from a disordered state to a folded and fully functional state is contained within the primary amino acid sequence (Anfinsen, 1973). It was then concluded that the linear chains of amino acids found within all proteins assume a folded and functional three-dimensional structure through a process known as protein folding (Anfinsen, 1973). However, Levinthal postulated that it would take longer than the lifetime of the universe for a polypeptide chain to attain the native conformation (Levinthal, 1968) even though *in vivo* most proteins fold within a few seconds. The work conducted by these pioneers has resulted in one of the most intensely pursued research questions of our time: how do proteins fold?

1.2.1. Folding models

The two most generally accepted explanations for the principle of protein folding are the thermodynamic hypothesis proposed by Anfinsen (1973) and the kinetic hypothesis based on the calculations of Levinthal (1968). Several protein folding models have been proposed, but it is unlikely that a single mechanism describing protein folding exists. Therefore, protein folding models can only serve as an explanation of folding events occurring in a specified time frame.

A visual representation of four models of protein folding is shown in figure 1A. The nucleation-condensation model proposes that native-like interactions are formed between several key residues to construct folding nuclei, which propagate the formation of secondary and then tertiary structures in a step-wise manner (Fersht, 1997). In contrast, the framework model (Figure 1A) proposes that hydrogen-bonded secondary structural elements are able to form prior to and independently from the tertiary structure (Ptitsyn and Rashin, 1975).

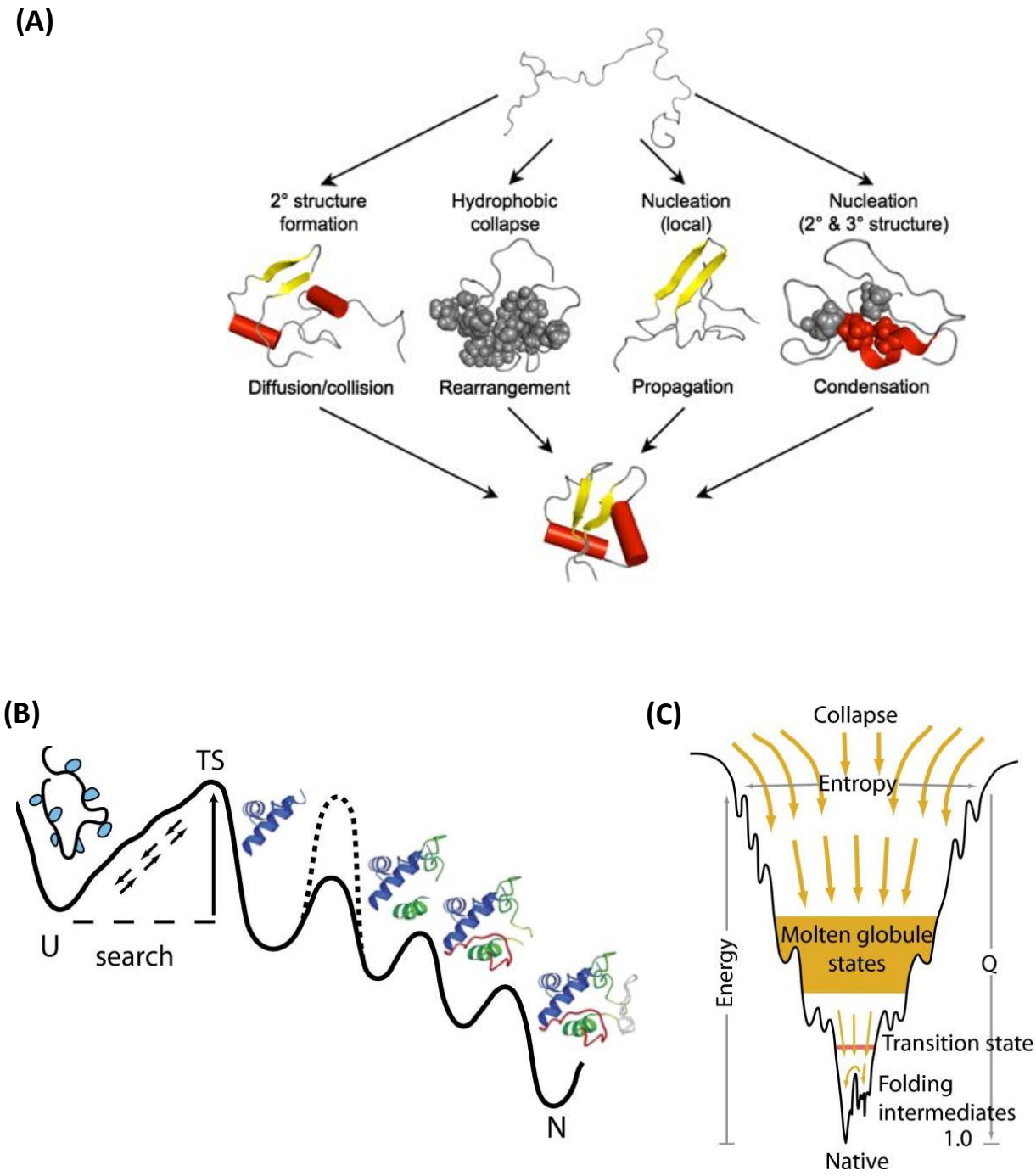


Figure 1: Visual representation of protein folding concepts

(A) Visual representation of the diffusion-collision, hydrophobic collapse, nucleation-growth and nucleation-condensation models, respectively (Nickson and Clarke, 2010). (B) The classical specific pathway concept with distinct intermediates, as proposed by Levinthal. The presence of an error-dependent kinetic barrier (- - -) can affect certain populations, thereby mimicking multi-pathway folding (Englander and Mayne, 2014). (C) Schematic representation of the "new view" energy landscape. Entropy decreases as the protein approaches the native state. Folding may involve transient intermediates in local energy minima, or misfolded states in significant kinetic traps (Englander and Mayne, 2014).

The diffusion-collision model (Figure 1A) expands upon the concept that secondary structure forms prior to tertiary structure because this model proposes that folding occurs by the diffusional collisions of structured microdomains (Karplus and Weaver, 1994). The successful collision of microdomains results in the cooperative folding of a protein to form a functional three-dimensional structure. The nucleation-growth model (Figure 1A) postulates that secondary structural elements are formed by numerous neighbouring residues, and that the structures formed act as propagation sites for step-wise folding (Wetlaufer, 1973). The hydrophobic collapse model (Figure 1A) proposes that folding is driven by the collapse of hydrophobic residues (Dill, 1985, 1990). Secondary structural elements, therefore, form as a result of restricted conformational space, with a bias towards the native structure (Dill, 1990). The unifying concept between the proposed models is that each model implies the existence of partially folded structures along the pathway. Additionally, researchers are certain that global folding occurs subsequent to all other folding events because correct folding of elementary processes and intermediate species must be complete in order for a protein to fold into a three-dimensional structure (Eaton *et al.*, 2000).

Numerous protein folding studies have been conducted resulting in the publication of several excellent review articles. However, protein folding researchers have not been able to reach a consensus for a unified mechanism for protein folding, and the protein folding problem has not yet been solved (Arai, 2018; Baldwin, 2017; Englander, 2000; Englander and Mayne, 2014, 2017b; Finkelstein *et al.*, 2017; Krishna and Englander, 2007).

1.2.2. Levinthal, energy landscapes and foldons

The classical specific pathway concept (Figure 1B) indicates that the folding process would take an enormously long time if a protein were to attain the correctly folded configuration by randomly sampling all the possible conformations (Levinthal, 1968). However, the short timescales of protein folding prove that proteins do not attain their native structure by a random process. Levinthal's classical specific pathway concept was widely accepted for decades, until the "new view" of protein folding was proposed (Figure 1C) (Baldwin, 1994; Sali *et al.*, 1994). The "new view" postulates that an ensemble of conformations follow parallel folding events to attain the native, folded state, forming the basis of the funnel-shaped energy landscape model (Baldwin, 1994; Chan and Dill, 1998; Dill and Chan, 1997).

The energy landscape theory (ELT) proposes that folding is a conformational search on a conceptual folding funnel that is biased towards the native state (Baldwin, 1994; Plotkin and

Onuchic, 2002a). The top of the funnel is composed of random protein states that have high enthalpy and entropy compared with the states further along the funnel (Figure 1C). These random protein states move down the protein funnel so as to minimise enthalpic forces as the native state is attained (Dill and Chan, 1997; Dinner *et al.*, 2000). Local minima in the funnel represent the presence of both stable intermediates and other proteins that exist along the folding pathway whereas a lack of intermediates is represented by a smooth funnel landscape (Figure 1C) (Baldwin, 1994; Chan and Dill, 1998; Plotkin and Onuchic, 2002b).

The complex nature of protein folding, and the diversity of protein molecules, has resulted in an as yet unresolved difference in opinion regarding the folding mechanism employed by proteins. The modern many-pathway funnelled energy landscape model has been challenged, and a case for a defined folding pathway has been proposed (Bai *et al.*, 1995; Englander and Mayne, 2017b). Englander and Mayne argue that the ELT is too generalised and that the system can be applied to the folding of any protein, RNA or polymer. A detailed review of the studies that have elucidated and explored protein foldons (Gianni *et al.*, 2003; Hu *et al.*, 2016, 2013; Xu *et al.*, 1998) concludes that a foldon consists of approximately 25 residues that fold cooperatively, and that each foldon strip folds on a single pathway (Englander and Mayne, 2017b). The foldon-dependent protein folding theory has received support from a number of researchers (Baldwin, 2017; Bédard *et al.*, 2008; Hu *et al.*, 2016; Krishna and Englander, 2007), but it has also been contested that the claims made by Englander, Mayne and co-workers pertaining to the existence of foldons are false (De *et al.*, 2017; Eaton and Wolynes, 2017; Yu *et al.*, 2015). Eaton and Wolynes have argued that theory, experimental research and simulations provide an irrefutable case that proteins fold by multiple pathways and not by a single pathway (Eaton and Wolynes, 2017; Kim *et al.*, 2014). Englander and Mayne conclude the argument stating that recent progress suggests that while the ELT has benefitted the field of protein folding greatly; it is “time to move on” towards a different vision (Englander and Mayne, 2017a).

It is clear that the current conflict needs to be resolved before the nature of protein folding can be understood (Baldwin, 2017). Novel folding studies are conducted on a regular basis (Arai, 2018; Dautant *et al.*, 2017; Finkelstein *et al.*, 2017; Malhotra and Udgaonkar, 2016b; Moulick and Udgaonkar, 2017; Neupane *et al.*, 2016; Puri and Chaudhuri, 2017) in an attempt to understand the various states that are available to a protein along the folding pathway. A deeper understanding of the various states in which a protein can exist will provide further evidence and will enable researchers to move towards a more unified understanding of the protein folding problem.

1.2.3. Protein folding pathway

The folding pathway of a protein cannot be understood unless all of the states that are present along the pathway are characterised. Therefore, the folded, unfolded and stable intermediate states, as well as the transition states that link each stable state, need to be identified and characterised.

1.2.3.1. The native state

The native state determines the biological function of a protein and the contacts formed by the native state are pivotal. The three dimensional folded state of a protein is a highly ordered structure that is determined by the rotations occurring around single bonds in the backbone and sidechain atoms. The existence of the native state depends on the delicate equilibrium achieved between the destabilising loss of entropy (as a result of folding) and the stabilising hydrophobic interactions, hydrogen bonds and electrophilic and interactions.

Hydrophobic interactions have been shown to be the dominant force driving protein folding (Dill, 1990; Heringa and Argos, 1991). Crystal structures of folded proteins have shown that the non-polar hydrophobic residues cluster together in the interior of the protein to form the stable core whereas polar and charged residues interact with solvent (Chothia and Janin, 1975; Hubbard and Argos, 1994; Janin *et al.*, 1988; Kauzmann, 1959). Water molecules surrounding the apolar amino acids assumes a conformation resulting in low entropy (to maximise positive hydrogen bonding between water molecules), whereas water molecules surrounding polar amino acids form a hydration shell that enables hydrogen bonding to the solution (Baldwin, 2014; Dill, 1990; Dill *et al.*, 1995). Therefore, the stability of a protein is a balance between hydrophobic driving forces and entropic opposing forces (Dill, 1990; Pace, 1990; Privalov, 1979).

Long- and short-range contacts are important non-covalent interactions that assist in maintaining the stability of proteins. Electrostatic forces are long-range interactions that occur between the charged groups of ionisable sidechains. The stabilising effect of electrostatic interactions occurs when two oppositely charged groups are within 4 Å of one another (salt-bridge). However, a destabilising effect can also occur as a result of non-specific repulsive forces when identically charged groups are within 4 Å of one another (Barlow and Thornton, 1983; Neet and Timm, 1994; Pace, 1986). Hydrogen bonding is a weak short-range contact that occurs between an electronegative atom covalently bound to a hydrogen atom and another electronegative atom. Hydrogen bonding occurs between the carbonyl oxygen and the amide nitrogen of peptide bonds in protein structures, and these

bonds stabilise secondary structural elements such as helix-coil transitions (Dill, 1990; Grantcharova *et al.*, 2001; Hubbard and Argos, 1994). Hydrogen bonding plays a role in the stabilisation of the folded, functional protein, but these interactions are not the dominant driving force in protein folding (Dill, 1990). Additionally, numerous interactions, such as disulfide bond formation, serve to stabilise the native protein. Disulfide bond formation is an important co- and post-translational modification that involves the formation of covalent interactions between intramolecular cysteines (Oka and Bulleid, 2013; Shimizu and Hendershot, 2009). Disulfide bond formation occurs in the endoplasmic reticulum, where electrons from reduced cysteines in the unfolded state are transferred to oxidoreductases of the protein disulfide isomerise family (Bader *et al.*, 1999; Oka and Bulleid, 2013). The formation of disulfide bonds is an important step in the oxidative folding of numerous secretory proteins, because the correct formation of native disulfide bonds ensures the stabilisation and assembly of multidomain protein complexes (Shimizu and Hendershot, 2009; Tu and Weissman, 2004).

1.2.3.2. The intermediate state

Folding has been shown to be a highly cooperative two-state process with no detectable intermediates for a number of smaller proteins (Jackson and Fersht, 1991; Kaplan *et al.*, 1997; Kiefhaber, 1995). However, examination of the folding mechanisms of larger multidomain proteins has elucidated the presence of intermediate and intermediate-like species as the protein moves from the denatured state towards a stable, native structure (and vice versa) (Baldwin, 1996; Creighton, 1990; Ptitsyn *et al.*, 1990). Researchers have exploited numerous methodologies in an attempt to define and characterise intermediate forms (Fink *et al.*, 1998). Stable thermodynamic intermediates are more easily examined, and detailed information regarding the structure of these intermediates can be obtained (Bhutani and Udgaonkar, 2003; Chamberlain *et al.*, 1996; Englander, 2000; Fink *et al.*, 1998). In contrast, kinetic folding intermediates have brief lifetimes. There are techniques that can be used to follow the folding process as it is occurring, these techniques provide good kinetic data, but limited structural information (Dasgupta and Udgaonkar, 2012; Kiefhaber, 1995; Krishna *et al.*, 2004; Wallace *et al.*, 1998a).

Several types of intermediates have been observed in the equilibrium pathways of large multidomain proteins. Intermediates include dimeric, monomeric and molten globule states. The molten globule (MG) state has been identified as a productive intermediate between the native and denatured states for a number of globular proteins (Englander, 2000; Khan *et al.*, 2016b; Pande and Rokhsar, 1998). Two variations of the MG state have been characterised:

the ‘wet’ molten globule (WMG) and the dry molten globule (DMG). The WMG is a highly dynamic protein state that exhibits native-like secondary structure and a loosely- packed solvent exposed core (Fink *et al.*, 1998; Ptitsyn *et al.*, 1990). Conversely, the DMG is an expanded structure that lacks close-packing interactions, with a native-like structure that does not permit solvent penetration of the core (Shakhnovich, 1997; Shakhnovich and Finkelstein, 1989). The role of molten globules in protein folding has been reviewed and discussed in great detail (Arai and Kuwajima, 2000; Baldwin *et al.*, 2010; Baldwin and Rose, 2013; Englander and Mayne, 2014) and studies have shown that equilibrium molten globules may even represent kinetic intermediate analogs (Dabora *et al.*, 1996; Jennings and Wright, 1993; Lim and Marqusee, 2017; Oliveberg, 1998; Raschke and Marqusee, 1997).

Kinetics studies have detected transiently formed intermediates, even though these metastable states are only present in low stoichiometric amounts (Englander, 2000; Kiefhaber, 1995). Characterisation of intermediates detected in kinetic studies remains challenging, but advances in hydrogen exchange methods have made it possible to obtain informative structural information about transient intermediate states (Englander, 2000; Englander and Mayne, 2017b; Krishna *et al.*, 2004). The presence of a kinetic intermediate indicates the occurrence of an event, such as *cis-trans* prolyl peptide bond isomerisation (Reimer *et al.*, 1998). The Xaa-proline peptide bond exists in either the *cis* or the *trans* conformation; with the *trans* conformation of non-proline peptide bonds displaying increased stability relative to the *cis* conformation. In native protein conformations, one of the isomers will be favoured as a result of structural constraints of the native conformation (Kiefhaber *et al.*, 1992; Reimer *et al.*, 1998). In an unfolded protein, a specific Xaa-proline peptide bond will favour the *trans* conformation because of steric effects and as a result of $n \rightarrow \pi^*$ interaction between the oxygen of the peptide bond and the consequent carbonyl carbon in the polypeptide (Hinderaker and Raines, 2003). Proteins that display the *cis* conformation in the native protein display isomerisation of 70-90% of the species during folding (Brandts *et al.*, 1977) and identification of peptide bonds that undergo isomerisation during unfolding is key in determining the folding pathway of that protein (Eyles and Gierasch, 2000). Isomerisation does not necessarily result in the formation of rate-limiting intermediate species, but the rate-limiting significance of *cis-trans* isomerisation during folding must be considered in the determination of the folding pathway of a protein (Cook *et al.*, 1979; Kiefhaber *et al.*, 1992; Segawa and Sugihara, 1984).

It has been suggested that intermediates direct the folding process (Kim and Baldwin, 1990; Walkenhorst *et al.*, 1997); however, it has also been argued that intermediates could represent trapped species in the ELT, retarding the folding process (Jackson and Fersht, 1991; Silow and Oliveberg, 1997). The two differing viewpoints have resulted in a critical debate that discusses whether intermediate states are obligate on-pathway intermediates or whether they are trapped species that occur off the folding pathway (Baldwin, 1996; Jackson, 1998; Miranker *et al.*, 1993). Despite much debate surrounding intermediates, identification and characterisation of both stable equilibrium intermediates and unstable transiently formed kinetic intermediates is critical. The characterisation of these states elucidates the (un)folding pathway of a protein and also provides useful insight into misfolding and aggregation.

Misfolding and aggregation

Protein folding most often results in a fully folded, functional, native structure *in vivo*. Consideration of the deep complexity of protein folding highlights that the process will not necessarily always occur without error. Regions that comprise the hydrophobic core of the folding structure are exposed to solvent when proteins are synthesised in the cell, resulting in unfavourable interactions with the other molecules (Ellis, 2001). Systems have evolved to prevent unfavourable interactions from resulting in misfolding, but particular conditions do result in failure of a protein to fold correctly or to remain correctly folded (Dobson, 2003; Englander *et al.*, 2007; Moulick and Udgaonkar, 2017; Yu *et al.*, 2015). Misfolded proteins are responsible for loss of function diseases such as cystic fibrosis as well as certain cancers (Dobson, 2003). Additionally, protein misfolding intermediates are also able to form precipitates known as protein aggregates (Dobson, 2003; Fink, 1998). Protein aggregates form insoluble fibrillar structures that accumulate in tissues like the brain and heart and result in diseases such as Alzheimer's, Parkinson's and diabetes (type II) (Chiti and Dobson, 2006; Dobson, 2003; Ross and Poirier, 2004).

Characterisation of a misfolded intermediate from a small folding domain (Gianni *et al.*, 2010) indicates that it is possible to characterise not only stable and transient intermediates, but also misfolded intermediates. It is essential that we apply our knowledge of protein folding pathways to elucidate structural and mechanistic details of misfolded and aggregation-prone intermediates because these states are the potential drug targets of the future (Aguzzi and O'Connor, 2010; Borgia *et al.*, 2013; de Oliveira and Silva, 2017; Khan *et al.*, 2016a; Perchiacca *et al.*, 2012; Woods *et al.*, 2012).

1.2.3.3. The unfolded state

The folded state is diverse between proteins, and this state has received much attention because of its biological relevance. Misfolding and aggregation have also been thoroughly pursued because these states are responsible for a wide variety of disease states. However, researchers have been less intrigued by the unfolded state of a protein because it was assumed that the unfolded state is a generic ensemble of highly disordered species that rearrange rapidly in the search for an energetically favourable configuration (Dill and Shortle, 1991; Griko, 1999). In tissues and cells, the unfolded state of a protein is in equilibrium with the native state of the protein and protein stability is defined as the difference in free energy between these two opposing states, indicating that the unfolded state is just as significant as the native state, when assessing the folding pathway (Dill and Shortle, 1991).

The theoretical ideal of the unfolded state is a random coil, in which all conformations have comparable free energies, whereas the native state is well defined by a set of coordinates (Creighton, 1990). However, the unfolded state is less random than initially thought, as unfolded states show inter-residue interactions as well as varying degrees of residual structure (Dill and Shortle, 1991; Hammarström and Jonsson, 2005; Plaxco and Gross, 2001; Shortle, 2001). Additionally, native-like topology has been shown to exist when proteins are unfolded under strongly denaturing conditions (Shortle, 2001). It is important to distinguish between a denatured and a fully unfolded protein state because the residual structure of a denatured state has significant implications in the refolding pathway and stability estimations of a protein (Myers *et al.*, 1995; Soulages, 1998).

The importance of the denatured state cannot be overlooked because these states are not random. The defined nature of the denatured state indicates that a pre-configuration of native-like bond angles may exist for a number of proteins, which would have significant implications for Levinthal's paradox (Plaxco and Gross, 2001). Increasing the knowledge-base pertaining to both denatured states and unfolded ensembles is, therefore, an important endeavour.

1.3. Protein unfolding

In order to understand how a protein folds, it is also important to assess the unfolding mechanism of the protein. Elucidation of the unfolding mechanism of a protein can provide details about the stability, cooperativity, intermediate states, and the mechanism of action of the protein. Protein unfolding is an essential physiological process *in vivo* (see review Matouschek, 2003). Protein unfolding assists in protein degradation (Baumeister *et al.*, 1998), membrane insertion (Gatzeva-Topalova *et al.*, 2010; Stoychev *et al.*, 2009), and even functionality (Dunker *et al.*, 2001), but has also been implicated in a number of misfolding diseases (see section 1.1.3.2.).

The unfolded state is neither generic, nor random, and as discussed above (section 1.1.3.3.), and not all proteins reach a fully unfolded, disordered conformation. Several proteins unfold via intermediates to a more defined denatured state, with a high degree of structural order (Hammarström and Jonsson, 2005; Plaxco and Gross, 2001). The unfolding of proteins, and how the denatured state is obtained, is of great importance because there are implications for the (re)folding pathway both *in vivo* and *in vitro* (Lapidus, 2017).

1.3.1. The push-and-pull hypothesis

Protein unfolding *in vitro* is a standard technique used to perturb the native state, and to promote the formation of intermediate states. Various methods are available to unfold proteins, but different methods produce unfolded and denatured states that are distinct from one another (Lapidus, 2017). Force (Ritchie and Woodside, 2015; Yu *et al.*, 2015), temperature (Phillips *et al.*, 1995; Privalov and Potekhin, 1986) and pH (Rami and Udgaonkar, 2001; Yeh *et al.*, 2017) are all used as denaturing agents; however, the two most commonly used methods, which form the basis of the push-and-pull hypothesis, are high-pressure denaturation and chemical denaturation (de Oliveira and Silva, 2015).

The molecular mechanism of pressure-induced denaturation is well-defined, and can be proven both experimentally and computationally (Lapidus, 2017). High pressures (~2 kbar) initiate the hydration of hydrophobic amino acids located within the core of the protein (Font *et al.*, 2006; Zhang and Smith, 1993). Hydration of the hydrophobic amino acids results in the formation of unstable cavities within the protein, which in turn causes the protein to unfold (Schuler, 2007).

In contrast, the molecular mechanism of denaturation by chemical denaturants, such as urea and guanidine hydrochloride, is not yet fully understood. One possible mechanism is an indirect mechanism that involves the alteration of the structure of water to reduce the magnitude of the hydrophobic effect (Tanford, 1970). A second theory is a direct mechanism that suggests interactions between the chemical denaturant and the protein results in the disruption of native contacts (Khan *et al.*, 2016b; Povarova *et al.*, 2010; Privalov, 1979). Simulations using the same protein indicate that urea disrupts native conformation by forming hydrogen bonds with the peptide backbone (Baldwin, 2014; Kauzmann, 1959) whereas guanidine hydrochloride disrupts hydrophobic interactions between the side chains of amino acid residues (Bedouelle, 2016; de Oliveira and Silva, 2015). A recent study conducted by de Oliveira and Silva (2015) shows that the direct mechanism is the driving force of urea-induced unfolding, a conclusion supported by several computational studies (Khan *et al.*, 2016b; Seelig, 2018).

Pressure-induced unfolding differs from urea-induced unfolding in that cooperative units are often formed as a result of the non-homogenous nature of pressure-induced unfolding (de Oliveira and Silva, 2015). The compressibility effects of pressure on a protein causes hydrogen bond shortening within the cooperative units, favouring hydration to push water molecules against atoms within the protein (push effect) unfolding (de Oliveira and Silva, 2015). Conversely, the direct preferential binding mechanism of urea results in the homogenous pulling of water molecules away from the protein structure (pull effect) to unfold the protein unfolding (de Oliveira and Silva, 2015). The dissimilar effects of urea and force on the unfolding of a protein are this explained by the push-and-pull hypothesis, which provides a molecular interpretation for the varying effects on unfolding (de Oliveira and Silva, 2015, 2017). Additionally, the hypothesis provides possible explanations for the nature of cooperativity, which is critical in the evaluation of unfolding mechanisms observed via both equilibrium unfolding and kinetic unfolding studies.

1.3.2. Equilibrium unfolding

Equilibrium unfolding studies are useful in providing detailed information of the structure, stability, cooperativity and folding of the native, denatured and stable intermediate states (Neet and Timm, 1994). The conformational stability of a protein can be determined by disrupting the equilibrium that exists between the native and denatured state through the use of a denaturant (temperature, chemical, force, pressure, pH). A denaturant will shift the equilibrium towards the denatured state and re-establish equilibrium. The choice of denaturant is critical because, to obtain thermodynamic parameters, denaturation needs to be

reversible (Pace, 1986). Unfolding at varying urea concentrations is monitored by a number of spectroscopic probes (far-UV circular dichroism, intrinsic fluorescence, ligand binding, size exclusion high-performance liquid chromatography (see section 3.5), to produce an unfolding curve.

Equilibrium unfolding curves are analysed by fitting the data to either a two-state model, or to a multistate model, to obtain the free energy change in the absence of denaturant ($\Delta G(H_2O)$) and the *m*-value (Myers *et al.*, 1995; Pace, 1986). The two-state model suggests that only the native and denatured states exist along the unfolding pathway, and that these forms interconvert in a highly cooperative manner, whereas a multistate model suggests the presence of at least one stable intermediate along the unfolding pathway (Neet and Timm, 1994; Pace, 1986; Soulages, 1998).

The cooperative two-state model (N↔U) has been applied to many small monomeric proteins (Kim and Baldwin, 1990), and to several dimeric proteins (Gloss and Matthews, 1997; Main *et al.*, 1999; Rumfeldt *et al.*, 2008; Wallace *et al.*, 1998a). However, despite the numerous examples of two-state unfolding, the basis for cooperativity is not yet fully understood (Malhotra and Udgaonkar, 2016a). Cooperativity is determined by the stability of the native state, and the interactions that stabilise the native structure are cooperative because disruption of one or more of these interactions results in the destabilisation of other nearby interactions (Matouschek and Fersht, 1993; Privalov and Potekhin, 1986). Cooperativity in folding pathways is important because partially folded states show unfavourable interactions that are not present in the native or denatured states. Recently it has been concluded that stabilisation of the denatured state, induced by an increase in denaturant, can result in a two-state cooperative global unfolding transition over a large energy barrier (Malhotra and Udgaonkar, 2015). Therefore, cooperativity might have evolved as a mechanism to decrease the propensity of a protein to aggregate (Malhotra and Udgaonkar, 2015, 2016a).

If the unfolding data is not a simple two-state sigmoidal curve, or if the transition region obtained from multiple probes are not superimposable, the unfolding is multistate (Pace, 1986). A general three-state model (N ↔ I ↔ U) would involve a native dimer, an unfolded monomeric intermediate and the unfolded/denatured state. A stable equilibrium intermediate can be detected because the species is significantly populated under equilibrium conditions, which enables the characterisation of the intermediate state. It is important to note that the deceptively simple two-state model does not necessarily exclude the possibility of high-energy intermediates (Kim and Baldwin, 1990). High-energy intermediates can still populate

the unfolding pathway, without any observed effect on the thermodynamic unfolding curve, if the energy of the intermediate is sufficiently high (Englander, 2000; Zaidi *et al.*, 1997). Therefore, to understand the complete unfolding pathway of a protein, the equilibrium and kinetic unfolding pathways must be assessed.

1.3.3. Unfolding kinetics

Equilibrium unfolding studies are fundamental in the design of kinetic experiments, and for the correct interpretation of the results (Matthews, 1987). Kinetic unfolding studies are useful in providing information about the transition states that populate the folding and unfolding mechanisms of proteins, and the mechanisms that govern transition state formation (Jaenicke and Rudolph, 1989). Kinetic unfolding studies are generally conducted by mixing protein with denaturant, and monitoring the progress of the unfolding reaction. Details on how the protein unfolds are obtained through the use of numerous probes, such as fluorescence, circular dichroism, ligand binding and mass spectrometry (Kiefhaber and Baldwin, 1995; Walters *et al.*, 2009).

Rapid mixing methods, such as stopped-flow studies, are commonly used to monitor the unfolding kinetics of proteins on a millisecond to microsecond timescale (Tonomura *et al.*, 1978). The unfolding reaction is initiated by rapidly mixing denaturants, such as urea, with native protein. The subsequent unfolding reaction can then be monitored using a suitable probe (Hoa, 1973; Roder *et al.*, 2008). If only two species are present throughout the unfolding reaction and unfolding proceeds via a two-state mechanism, only one single kinetic phase will be detected (Kim and Baldwin, 1990; Utiyama and Baldwin, 1986). However, if the monophasic unfolding kinetic traces detected using multiple probes show variation, or if a burst phase is present, an intermediate species may be present (Zaidi *et al.*, 1997). A burst phase indicates that unfolding events are occurring within the mixing time (dead-time) of the apparatus, and that there is a rapid accumulation of intermediate(s) in the early stages of protein unfolding (Lim and Marqusee, 2017; Roder *et al.*, 2008). Protein unfolding intermediates may also be present if the unfolding traces are not monophasic. Multiphase kinetic unfolding traces can occur if multiple states of the native protein exist or if there is coupling between the unfolding reaction and peptide bond isomerisation (Kiefhaber *et al.*, 1995). Additionally, the formation of a structured transient intermediate will also result in multiphase kinetic traces (Galani *et al.*, 2002; Shi *et al.*, 2017; Walkenhorst *et al.*, 1997; Žoldák *et al.*, 2017).

The presence of kinetic unfolding intermediates indicates a complex unfolding pathway, which can be defined as either sequential or parallel. In a sequential unfolding pathway, kinetic events occur in a sequential manner, and intermediates may form either on- or off-pathway (Deng and Smith, 1999; Wallace and Matthews, 2002). Conversely, in a parallel unfolding pathway, events occur simultaneously (parallel) (Kiefhaber, 1995; Wallace and Matthews, 2002; Wright *et al.*, 2004). It is also important to characterise whether an intermediate species is on-pathway or off-pathway. On-pathway unfolding intermediates are obligate intermediates that are required for the formation of the unfolded state, whereas off-pathway intermediates are non-productive and have been shown to slow down the unfolding process in a lag phase (for a detailed review see Baldwin, 1996).

Complex unfolding reactions are difficult to interpret, and in order to determine whether a pathway is sequential or parallel a number of experimental tests should be employed (for a thorough and detailed review see Wallace and Matthews, 2002). However, the use of genetically modified proteins can also assist in the elucidation of complex unfolding (and refolding) pathways (Matthews, 1987; Vallée-Bélisle and Michnick, 2012).

1.4. Human glutathione transferase M1a-1a

The canonical cytosolic glutathione transferases (GSTs) (EC 2.5.1.18) are a supergene family of multifunctional detoxification enzymes (Combes and Stakelum, 1961). Soluble mammalian GSTs are divided into eight classes: Alpha, Mu, Pi, Omega, Kappa, Sigma, Theta and Zeta (Board *et al.*, 2000; Mannervik *et al.*, 1992; Meyer *et al.*, 1991; Pemble *et al.*, 1996). The Tau, Phi, Delta and Beta GST classes are found in plants, insects and bacteria, respectively (Hayes and McLellan, 1999).

Evolution of the GST classes is thought to have proceeded via divergent evolution from the two electron transfer proteins glutaredoxin and thioredoxin (Ladner *et al.*, 2004). The Kappa gene class encodes mitochondrial GST and has been proposed as the progenitor of the Theta gene class (Pemble *et al.*, 1996). The cDNA sequence of the Alpha, Mu and Pi classes indicate that these classes have evolved from the Theta gene class, with evidence that class Mu diverged from the Theta precursor prior to class Alpha and Pi (Allocati *et al.*, 2006; Ladner *et al.*, 2004; Pemble *et al.*, 1996; Sheehan *et al.*, 2001). Evolutionary distinctions between the various gene classes are also indicated by the three-dimensional structures of the enzymes. Differences in the catalytic site residues provide evidence that the class Theta enzyme is an evolutionary precursor because an essential serine residue is present at the catalytic site of class Theta, whereas the other gene classes contain a highly conserved tyrosine residue (Armstrong, 1997). There is also evidence for class Theta as a precursor protein at the dimer interface because classes Alpha, Mu and Pi differ from older GSTs in that the dimer interface of these proteins is curved rather than flat and a hydrophobic lock-and-key interaction exists between the subunits of these classes in place of hydrophilic contacts (Armstrong, 1997; Ladner *et al.*, 2004; Piromjitpong *et al.*, 2007).

1.4.1. Structure

Human glutathione transferase M1a-1a (hGST M1a-1a)[1] is composed of two ~28 kDa subunits, each consisting of 217 amino acids. In accordance with other GST structures, each subunit of the homodimeric hGST M1a-1a is composed of two domains: the smaller thioredoxin-like N-terminal domain (domain 1) and the all α-helical C-terminal domain (domain 2). The domains are separated by a short six residue interdomain linker region (Figure 2) (Hayes and Pulford, 1995; Ji *et al.*, 1992; Sheehan *et al.*, 2001).

[1]hGST M1a-1a refers to human GST from Mu class that is composed of two type 1a allelic subunits. The nomenclature used throughout this thesis is in accordance with (Mannervik *et al.*, 1992).

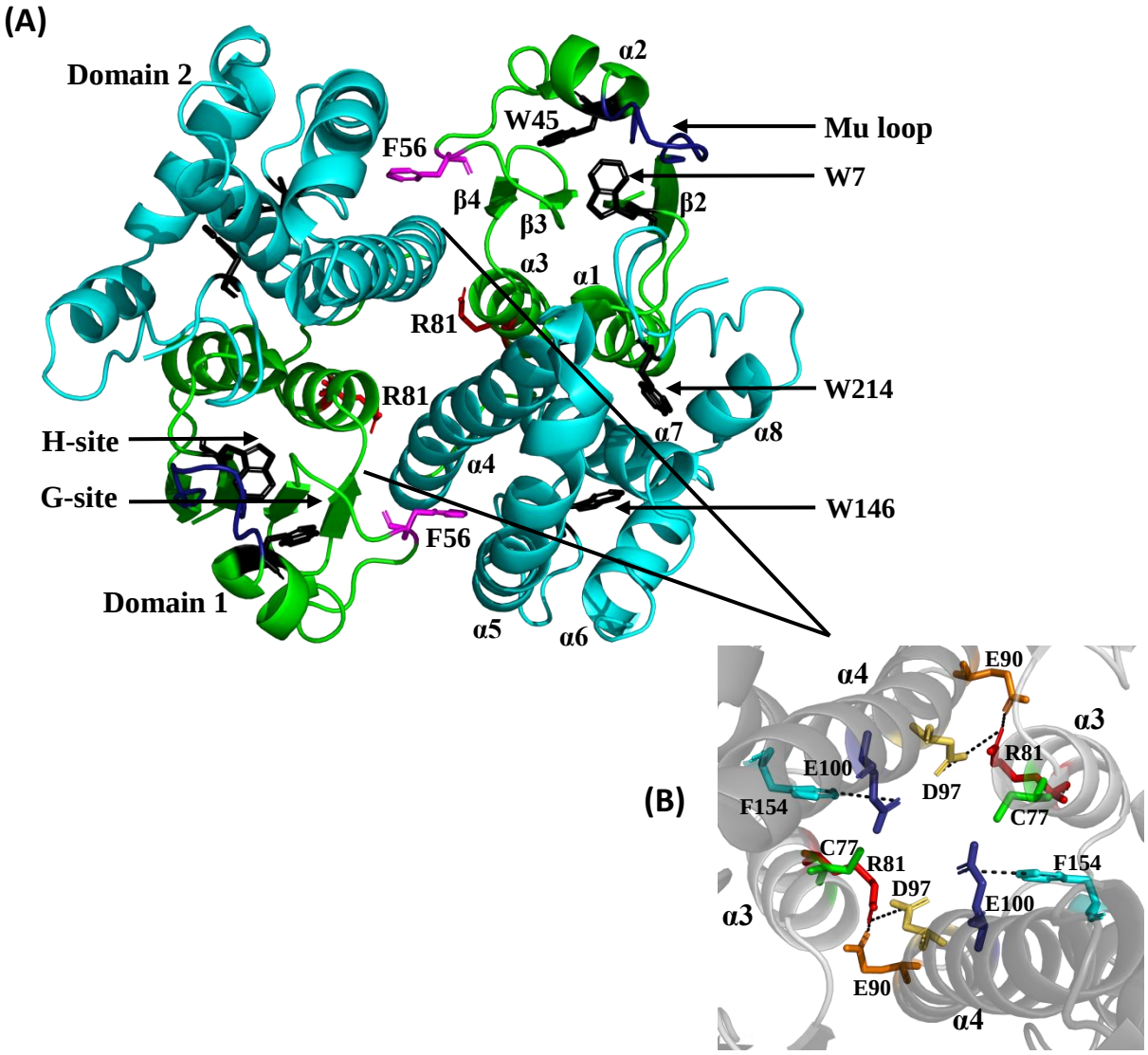


Figure 2: Structural features of hGST M1a-1a

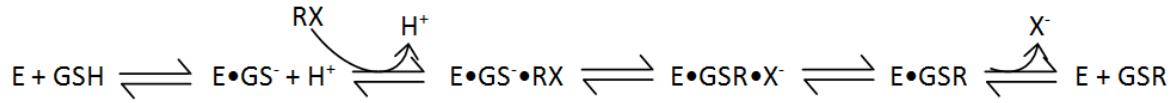
(A) The cartoon representative of hGST M1a-1a shown down a two-fold axis shows the smaller N-terminal thioredoxin-like domain 1 (green) and the larger all α-helical C-terminal domain (cyan). The Mu loop (dark blue) is located near the active site. The F56 key residue (magenta) and the conserved R81 (red) stabilise the quaternary structure of the protein. (B) The mixed charge cluster at the subunit interface is composed of residues C77 (green), R81 (red), E90 (orange), D97 (yellow), E100 (dark blue) and F154 (cyan). R81 interacts with E90 and D97 from the opposing subunit whereas E100 interacts with F154 within the same subunit. C77 is unable to interact electrostatically. PDB file 1GTU was used to generate the cartoon figures using the PyMOL Molecular Graphics System, V2.0 Schrödinger, LLC.

Domain 1 has been identified as the glutathione binding domain in thioredoxin (Kelley and Richards, 1987), glutathione peroxidase (Hayes and McLellan, 1999) and others (Martin, 1995) and the highly conserved domain consists of four antiparallel β-sheets sandwiched between three α-helices (β1-α1-β2-α2-β3- β-α3). Domain 2 is the larger all α-helical C-terminal domain, composed of five amphipathic α-helices to create a hydrophobic protein core (Figure 2) (Hayes and Pulford, 1995; Ji *et al.*, 1992; Sheehan *et al.*, 2001).

Each class displays unique structural features, such as the additional α-helix (α9) found in domain 2 of class Alpha GSTs (Sinning *et al.*, 1993; Wilce and Parker, 1994). Class Mu contains the unique insertion of an extended, mobile Mu loop region connecting the sequence between β2 and α2 (residues 33-43) in domain 1 (Figure 2) (Ji *et al.*, 1992). The Mu loop has been shown to contribute towards the substrate affinity of class Mu GSTs, but it is not essential for the maintenance of structure (Hearne and Colman, 2006). Despite the unique structural elements that are present in each class and the low sequence identity amongst classes, the quaternary structure and overall fold of GST enzymes is conserved. Dimerisation of the subunits is highly specific amongst GST gene classes and the interactions formed at the subunit interface govern the structure, stability and, therefore, the function of the enzyme.

1.4.2. Function

Homodimeric hGST M1a-1a shows a two-fold axis of symmetry, and each subunit contains a single active site at the subunit interface (Figure 2). The active site of each monomer is composed of a glutathione-binding site (G-site) and a hydrophobic binding site (H-site) for the binding of hydrophobic and electrophilic substrates (Figure 2) (Mannervik and Danielson, 1988). The GSTs function in the intracellular detoxification of carcinogens, mutagens and other toxic compounds via catalysis of a reduced glutathione (GSH) nucleophilic attack on non-polar compounds (Mannervik and Danielson, 1988; Sheehan *et al.*, 2001). The reaction serves to metabolise xenobiotic compounds that are then dispelled through the mercapturate pathway (Sheehan *et al.*, 2001). The following scheme represents the general reaction between GSH and an electrophilic substrate (R) catalysed by GST (E) (Armstrong, 1997):



The enzyme binds GSH at the G-site and removes the GSH sulfhydryl group proton to yield a glutathionate anion-enzyme complex (GS-E) (Armstrong, 1997). The proton is released and the electrophilic substrate binds to the H-site. The nucleophilic addition of a thiolate to the

electrophilic substrate is followed by the elimination of the leaving group (R) from the electrophilic substrate (Armstrong, 1997). The GSR conjugate is released with the free enzyme (E) in the final stage of the reaction (Armstrong, 1997).

The role of GSTs in toxicology is diverse and immense because of the multifunctional nature of these enzymes *in vivo* (Hayes and Pulford, 1995; Hearne and Colman, 2006; Henderson *et al.*, 1998; Hou *et al.*, 2007; Kinsley *et al.*, 2008; Pajaud *et al.*, 2012; Strange *et al.*, 2000). Recently, it was found that GSTs play a role in neurological disorder progression (Kumar *et al.*, 2017) and an increase in the levels of GST Mu in the cerebral cortex of patients with Alzheimer's disease has been reported (Sidell *et al.*, 2003). The different class Mu subunits possess distinct catalytic properties because the M1 subunits catalyse the conjugation of GSH to p-nitrobenzyl chloride and to trans-4-phenyl-3-buten-2-one whereas the M2 subunits catalyse the conjugation of GSH to 1,2-dichloro-4-nitrobenzene (Hussey and Hayes, 1993). Additionally, class Mu GSTs are of particular interests because they exhibit polymorphic expression of isoenzyme forms with varying function (Chenevix-Trench *et al.*, 1995; Mannervik *et al.*, 1992).

Polymorphisms in the GST M1-1 are thought to increase brain tumour susceptibility (Ezer *et al.*, 2002) as a result of altered catalytic activity. Additionally, hepatic GST M1a-1a has shown high activity towards 4-hydroxyalkenals, indicating that the protein may be involved in cellular defence against peroxidative damage (Hussey and Hayes, 1993; Mannervik *et al.*, 1988). The null phenotype of GST Mu has been associated with an increased risk of cancers because the enzyme is inactive, which eliminates the ability of the GST to detoxify carcinogens (Strange *et al.*, 2000). The role of GST M1 in anticancer drug resistance is amplified by its role in ASK1 regulation (McIlwain *et al.*, 2006). GST M1 regulates stress-induced apoptosis by binding to and inhibiting the activity of the MAP kinase kinase kinase ASK1 (Cho *et al.*, 2001; Ichijo *et al.*, 1997). Oligomerisation of GST M1 has been shown to release ASK1 and induce apoptosis in conditions such as oxidative stress and heat shock (Dorion *et al.*, 2002). These functions indicate that class Mu GST enzymes are vital for cell survival, making them important targets for studies that will improve our understanding of GST-related diseases.

1.4.3. Intersubunit interactions

Subunit-subunit interactions act not only as stabilising forces for the individual tertiary subunit structures, but also assist in the formation of quaternary structure (Alves *et al.*, 2006; Hornby *et al.*, 2002; Sayed *et al.*, 2000). Dimerisation is class specific and absolutely

essential for the functioning of the catalytic site, with no catalytically active GST monomers being present in solution (Abdalla *et al.*, 2002; Gildenhuys *et al.*, 2010; Thompson *et al.*, 2006). Dimerisation of ancestral GSTs, such as class Sigma and Theta, results in less burial of solvent accessible surface area upon dimerisation than the 14 % that has been reported for class Alpha, Mu and Pi GSTs (Board *et al.*, 2000; Dirr *et al.*, 1994). The differences observed are because of the more open dimer interface configuration observed in the ancestral GSTs (Board *et al.*, 2000), compared with the V-shaped cleft configuration formed by the hydrogen and ionic bond interactions that occur between α1 and α3 in domain 1 and between α4 and α6 in domain 2 of class Alpha, Mu and Pi GSTs (Dirr *et al.*, 1994; Wilce and Parker, 1994). The dimer interface of class Beta is very slightly packed, but lacks the V-shaped cleft formed in other classes, which opens the dimer interface to bulk solvent (Allocati *et al.*, 2006).

Crystal structure analyses of class Mu GST have identified two major interactions at the subunit-subunit interface. The first is the hydrophobic interactions of a lock-and-key mechanism and the second is a cluster of hydrogen and ionic bond interactions that occur between the two subunits as a result of a mixed charged cluster (Alves *et al.*, 2006; Dirr *et al.*, 1994; Hegazy *et al.*, 2004; Ji *et al.*, 1992; Reinemer *et al.*, 1992).

The lock-and-key motif is conserved amongst class Alpha, Mu and Pi GSTs, and the motif is responsible for a major hydrophobic intersubunit interaction (Reinemer *et al.*, 1992; Sinning *et al.*, 1993). A phenylalanine (F56) located between α2 and β3 (F52, Alpha; Y50, Pi) in domain 1 of one subunit is fixed into the hydrophobic cavity constructed by α4 and α5 of domain 2 of the interacting subunit (Figure 2) (Alves *et al.*, 2006; Hegazy *et al.*, 2006, 2004; Sayed *et al.*, 2000; Vargo *et al.*, 2004). Ancestral classes lack this interaction and are thought to compensate with increased electrostatic interactions in their predominantly hydrophilic interfaces (Meyer *et al.*, 1991; Rossjohn *et al.*, 1998; Stevens *et al.*, 1998). It has been shown that the lock-and-key motif plays a significant role in the stabilising the predominantly hydrophobic dimer interface of class Alpha, Mu and Pi GSTs, but that it is not solely responsible for dimerisation (Sayed *et al.*, 2000; Vargo *et al.*, 2004).

Charge clusters are a common and well-studied group of protein-protein interactions (see section 1.3.1.1.). A mixed charge cluster exists at the subunit interface of class Mu GSTS, even though the dimer interface is predominantly hydrophobic (Zhu and Karlin, 1996). Analysis of individual subunits did not reveal the presence of any charge clusters, indicating that the interdomain salt bridges that exist between interacting residues at the subunit-subunit interface are critical in the formation of the mixed charge cluster and the dimerisation of the

enzyme (Zhu and Karlin, 1996). The mixed charge cluster of hGST M1a-1a is composed of residues C77, R81, E90, D97, E100 and F154 (Figure 2B). R81 is conserved in class Pi and class Sigma and the charged residue forms salt bridges with E90 and D97 of the opposing subunit (Huang *et al.*, 2008; Zhu and Karlin, 1996). Additionally, hydrogen bonds connect F154 from chain A with E100 from chain A and F154 from chain B with E100 from chain B (Zhu and Karlin, 1996). The hydrogen bonds and ionic interactions formed within the mixed charge cluster are involved in the stabilisation of the quaternary structure of the enzyme (Hornby *et al.*, 2002; Thompson *et al.*, 2006; Zhu and Karlin, 1996).

A sequence alignment of closely related GST structural neighbours (class Alpha, Mu from both rat and human, Pi, Sigma and Theta) revealed interesting details about the subunit interface residues (Figure 3). The subunit interface of hGST M1a-1a has 13 charged residues whereas rGST M1-1 has 15. The two additional charged residues are R67 and R77 in rGST M1-1 whereas in hGST M1a-1a these residues are H67 and C77 (Figure 3). The difference between the mixed charge cluster residue at position 77 could result in differences in subunit stability between the two class Mu enzymes. These differences could be as a result of the formation of I interdomain salt bridges between R77 of domain 1 and D97 and E100 of domain 2, within each subunit (Ji *et al.*, 1992). Furthermore, the symmetrically equivalent R77 in each rGST M1-1 subunit is structurally conserved (R70, Pi; R68, Sigma) and the two guanidino groups of R77 in opposing subunits stack together near the dimer two-fold axis (Ji *et al.*, 1992). The hydrophobic C77 is unable to interact electrostatically, and the substitution of the residue in hGST M1a-1a could affect the properties of the subunit interface. Therefore, C77 in hGST M1a-1a may result in an altered dimerisation mechanism and conformational stability compared with rGST M1-1.

1.4.4. Conformational stability and folding

Dimer interface interactions affect the stability of proteins, and the obligate dimeric nature of GSTs suggests that subunit interactions are of great importance in the proper folding, and the maintenance of stability, of quaternary structure (Rumfeldt *et al.*, 2008). Furthermore, these interactions ensure that heterodimers can only occur naturally between members of the same class (Pettigrew and Colman, 2001; Stenberg *et al.*, 2000). The monomer-dimer transition of GSTs is relevant in the physiological functioning of these enzymes (Fabrini *et al.*, 2009). Monomeric forms of GSTs have been reported inside cells using antibody reactions. For example, hGST P1-1 appears to be bound to JNK in a monomeric form (Adler *et al.*, 1999; Kura *et al.*, 1996).

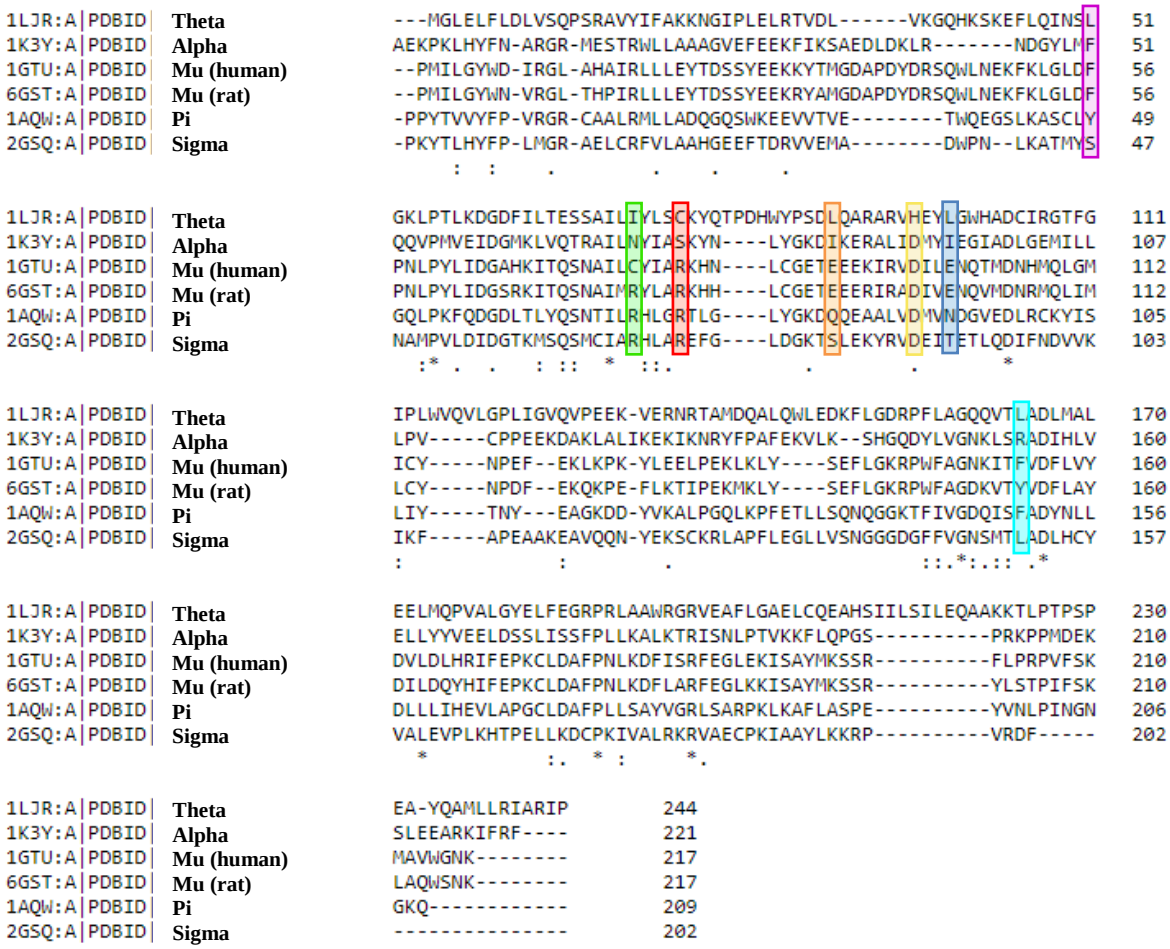


Figure 3: Structure-based sequence alignment of related GSTs

The structure-based sequence alignment shows class Theta (1LJR), Alpha (1K3Y), Mu from human (1GTU), Mu from rat (2GST), Pi (1AQW) and Sigma (2GSQ). The key residues from the lock and key motif (magenta) as well as the mixed charge cluster residues C77 (green), R81 (red), E90 (orange), D97 (yellow), E100 (blue) and F155 (cyan) are highlighted in boxes to highlight the similarities and differences between the classes. The sequence alignment was performed using the Clustal Omega tool (Sievers *et al.*, 2011). Clustal Omega is available for download at http://www.clustal.org/omega/.

The unfolding mechanisms of various GST classes are remarkably different, most likely because of the complex and varied dimer interface interactions that have been observed between the classes (Dirr, 2001). The GST isoenzymes from *Schistosoma japonicum* and the porcine class Pi enzyme (pGST P1-1) unfold via a highly cooperative two-state mechanism, with no intermediate species present in significant concentrations under equilibrium conditions (Dirr and Reinemer, 1991; Erhardt and Dirr, 1995). Human class Alpha GST (hGST A1-1) also unfolds via a two-state equilibrium unfolding mechanism; however, a native-like kinetic intermediate with partially dissociated domains also populates the unfolding pathway (Wallace *et al.*, 1998b).

The equilibrium unfolding model of hGST P1-1 is disputed amongst research groups: one research group supports the notion of structurally independent subunits and the presence of a stable monomeric intermediate (Aceto *et al.*, 1992), whereas another group found no evidence of a stable monomeric intermediate (Gildenhuys *et al.*, 2010). Aceto *et al.* (1992) show that hGST P1-1 denaturation follows a multistep process: the active dimeric structure dissociates into two structured, yet inactive monomers. Dissociation is thought to result in two inactive monomers via a bimolecular reaction rather than as a result of an unspecific effect caused by a chemical denaturant (Aceto *et al.*, 1992). In contrast, the unfolding of hGST P1-1 examined through the use of both equilibrium and kinetic unfolding experiments by Gildenhuys *et al.* (2010) revealed that α2 unfolds to form a dimeric intermediate. The dimeric intermediate proceeds via a two-state cooperative unfolding mechanism and dissociation of the dimer is tightly coupled to the complete global unfolding of the enzyme (Gildenhuys *et al.*, 2010).

The ability of GST M2 and GST P1 monomers to dimerise indicates a high degree of similarity between the subunit interface of class Mu and class Pi GSTs (Pettigrew and Colman, 2001). However, class Mu and class Pi GSTs do not unfold via the same mechanism. The unfolding mechanism for rat class Mu isoenzymes (rGST M1-1 and tGST M2-2) has been elucidated (Hornby *et al.*, 2000). The rat enzymes unfold in a three-state mechanism via an inactive monomeric intermediate, even though the rGST M1-1 and M2-2 isoenzymes display differences in stability (Hornby *et al.*, 2000; Luo *et al.*, 2002). The Sigma class GSTs show an even more complex unfolding mechanism, which includes thermodynamically stable dimeric and monomeric intermediates (Stevens *et al.*, 1998).

1.4.5. A modular approach to a complex investigation

Elucidation of the mechanisms that govern protein (un)folding is difficult for complex homo- and heterodimeric proteins like GSTs. However, the deconvolution of complex systems can be achieved through the use of comparative studies using simplified models. Therefore, a simplified monomeric variant of hGST M1a-1a was created to study the unfolding mechanism of the homodimer.

In order to create the monomeric mutant, the structural features of the subunit interface were considered (section 1.4.3.). Electrostatic interactions have been shown to govern dimerisation, and charged amino acids at the subunit interface of GSTs are responsible for these electrostatic interactions (Reinemer *et al.*, 1992). A study on hGST P1-1 identified the charged amino acids Arg-70, Arg-74, Asp-90 and Asp-94 as residues that are sufficiently close to be able to participate in the intra- and intermolecular interactions governing monomer-dimer equilibrium, and mutation of these residues produced a monomeric variant of hGST P1-1 (Huang *et al.*, 2008). Additionally, it has been shown that the simultaneous disruption of both the lock-and-key mechanism (F56S) and the mixed charge cluster (R81A) of rGST M1-1 results in the formation of a folded, monomeric variant protein.

The tertiary structures of the WT dimer and mutant monomer needed to be conformationally comparative to use the mutant monomer in a modular approach because major differences between the subunits of the dimeric hGST M1a-1a (WT dimer) and the simplified monomeric variant (mutant monomer) would result in discrepancies between the unfolding pathways. The globular structure of the F56S/R81A monomer generated from rGST M1-1 was shown to be comparable to a single subunit within the dimer through the use of hydrogen-deuterium exchange mass spectrometry (HDX-MS) fluorescence and far-UV circular dichroism (Thompson *et al.*, 2006). The differences in the subunit interfaces of hGST M1a-1a and rGST M1-1 (section 1.4.3.) do not affect residues F56 and R81, and the F56S/R81A monomeric variant is an ideal model to study the complex unfolding mechanism of dimeric hGST M1a-1a using a modular approach.

1.5. Aim and objectives

The numerous unfolding pathways utilised by various GST classes proves that determining the unfolding pathway of a multidomain oligomeric protein is highly complex. The use of different techniques and approaches may even elucidate different states that are populated throughout the unfolding mechanism. Differences between the primary structures within a class can result in the stabilisation of intermediates significantly enough that the unfolding mechanism is altered, as is the case for class Pi GST from pig (Erhardt and Dirr, 1995) and class Pi GST from the *Bufo bufo* embryo (Sacchetta *et al.*, 1999). Comparative studies will develop our understanding of topology, amino acid sequence, the balance between entropy and enthalpy and propensity to form secondary structures (Nickson and Clarke, 2010). Therefore, an investigation into the unfolding pathways of each class of GST enzymes, from various organisms, will allow insight into fundamental questions about protein folding mechanisms.

The main aim of this work is to propose an unfolding pathway for hGST M1a-1a.

The specific objectives of this study were to:

1. Perform site-directed mutagenesis to obtain a monomeric form hGST M1a-1a.
2. Over-express and purify wild-type and mutant hGST M1a-1a.
3. Assess the oligomeric status of wild-type and mutant hGST M1a-1a using size exclusion high-performance liquid chromatography.
4. Determine the catalytic activity of wild-type and mutant hGST M1a-1a using the GSH-CDNB conjugation assay.
5. Assess the secondary structure of wild-type and mutant hGST M1a-1a using far-UV circular dichroism.
6. Assess the tertiary structure of wild-type and mutant hGST M1a-1a using intrinsic fluorescence and hydrogen-deuterium exchange mass spectrometry.
7. Assess the stability of wild-type and monomeric mutant hGST M1a-1a using urea-induced equilibrium unfolding studies.
8. Assess the structural features and (un)folding intermediates of wild-type and mutant hGST M1a-1a using hydrogen-deuterium exchange mass spectrometry.

Chapter 2: Experimental Procedures

2.1. Materials

The *Escherichia coli* T7 competent cells and the Quikchange™ Site-Directed Mutagenesis kit were obtained from Stratagene (USA). E. Cloni EXPRESS competent cells were purchased from Promega (Madison, Wisconsin, USA). The GeneJET® Plasmid Miniprep Kit, dithiothreitol (DTT) and isopropyl-β-D-1-thiogalactopyranoside (IPTG) were obtained from Fermentas Life Sciences (St. Leon-Rot, Germany). Ampicillin was purchased from Roche Diagnostics (Mannheim, Germany). The compounds 1-chloro-2,4-dinitrobenzene (CDNB), 8-anilino-1-naphthalene sulfonate (ANS), reduced glutathione (GSH) and ethylenediaminetetraacetic acid (EDTA) were purchased from Sigma-Aldrich Co. (St. Louise, MO, USA). Ultrapure urea with a purity of 99.5% was purchased from Merck Chemicals (Darmstadt, Germany) and purification columns were purchased from GE Healthcare Life Sciences (Uppsala, Sweden). All other reagents used were of analytical grade.

2.2. Plasmid preparation

2.2.1. Wild-type

Dr Ikechekwu Achilonu (University of the Witwatersrand, South Africa) designed the codon harmonised nucleic acid sequence encoding wild-type (WT) hGST M1a-1a with an N-terminal his_6-tag and thrombin cleavage site. Additionally, a conservative and naturally occurring R205K mutation was incorporated to prevent thrombin cleavage occurring at this site within the protein sequence. The insert DNA sequence was synthesised and sub-cloned into the pET-11a plasmid (GenScript Corporation, USA). His_6-hGST M1a-1a plasmid DNA was extracted and purified from an overnight culture using the GeneJET™ Plasmid Miniprep Kit (Inqaba Biotech, Pretoria, South Africa). Plasmid DNA was sent to Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa) for sequencing to verify the insert and to confirm the absence any other mutations. The his_6-hGST M1a-1a plasmid DNA was transformed into *E. coli* T7 cells for overexpression.

2.2.2. F56S/R81A mutant

His_6-hGST M1a-1a dsDNA was used to design the oligonucleotide primers required to introduce the R81A mutation. His_6-R81A mutant dsDNA was then used to design the oligonucleotide primers required to introduce the F56S mutation. Mutagenesis was therefore performed twice, with two separate sets of oligonucleotide primers in order to generate the his_6-F56S/R81A double variant DNA sequence. Primers were synthesised by Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa).

R81A mutant:

R81A forward: 5'atcttgtgctacattgcc**gcc**aagcacaacctgtgtggg3'

R81A reverse: 5'cccacacaggttgtgctt**ggc**ggcaatgtagcacaagat3'

The nucleotides in bold show the codon change from arginine (cgc) to alanine (gcc).

F56S mutant:

F56S forward: 5'ttcaagctgggcctggac**tct**cccaatctgccctacttg3'

F56S reverse: 5' caagtagggcagattggg**aga**gtccaggcccagcttgaa3'

The nucleotides in bold show the codon change from phenylalanine (ttt) to serine (tct).

The Quikchange™ Site-Directed Mutagenesis kit (Stratagene, CA) was used to perform the site-directed mutagenesis according to the kit manual (Braman *et al.*, 1996). The plasmid encoding F56S/R81A mutant hGST M1a-1a was used to transform XL10-Gold supercompetent *E. coli* cells. Plasmid DNA was obtained from cell culture grown using a colony selected from a plate treated with kanamycin and streaked with transformed cells. The GeneJET™ Plasmid Miniprep Kit (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa) was used to extract plasmid DNA. Plasmid DNA was sent to Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa) for nucleotide sequencing. Nucleotide sequencing confirmed each mutation. The nucleotide sequences were aligned with the known mRNA sequence encoding WT hGST M1a-1a using BLASTN in the NCBI BLAST suite of (https://blast.ncbi.nlm.nih.gov/Blast.cgi) tools. The nucleotide sequence was translated using the translate tool on the ExPASy server (https://web.expasy.org/translate/) and the his_6-F56S/R81A peptide sequence was aligned with the known WT hGST M1a-1a peptide sequence using the peptide alignment tool LALIGN on the ExPASy server. The his_6-F56S/R81A plasmid DNA was transformed into *E. coli* T7 cells for overexpression.

2.3. Protein overexpression and purification

The WT his_6-hGST M1a-1a and the double mutant his_6-F56S/R81A proteins were overexpressed in *E. coli* T7 cells grown in 2xYT media. The cells were initially incubated at 37 °C until OD_{600nm} reached 0.5 AU. Overexpression of the WT was induced using a final IPTG concentration of 1 mM, and cells were incubated at 37 °C overnight. The mutant overexpression was induced using an IPTG concentration of 0.2 mM, and cells were incubated at 20°C overnight. The cultured cells were harvested by centrifugation and resuspended in 50 mM Tris-HCl buffer at pH 8.0, containing 500 mM NaCl, 2 mM phenylmethylsulfonyl fluoride (PMSF) and 0.02% (w/v) NaN_3 and stored at -20 °C.

Before purification of the WT and mutant GST M1a, cells were thawed at 30 °C and sonicated on ice for eight rounds of 30 second pulses at medium intensity. The soluble fraction was obtained by centrifugation at 24000 x g for 30 minutes at 4 °C. The supernatant was applied to a 5ml Ni^{2+} column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden), that was pre-equilibrated with 50 mM Tris-HCl buffer at pH 8.0, containing 500 mM NaCl, 30 mM imidazole and 0.02% (w/v) NaN_3. The column was first washed with equilibration buffer containing 30 mM imidazole to remove unbound protein. Subsequently, 4 ml fractions of the bound his_6-hGST M1a-1a proteins were eluted with a 30-300 mM imidazole gradient in 50 mM Tris-HCl buffer at pH 8.0, containing 150 mM NaCl and 0.02% (w/v) NaN_3. The absorbance at 280 nm was recorded and fractions were pooled. The pooled eluted protein was dialysed against 20 mM Tris-HCl buffer at pH 8.0, containing 200 mM NaCl, 5 mM CaCl and 0.02% (w/v) NaN_3 overnight at 4 °C. Thrombin cleavage to remove the His-tag from the GST M1a proteins was performed using a 1 U/ml thrombin stock solution. One microlitre thrombin stock was used per millilitre of purified protein for 8 hours at 20 °C. Thrombin was removed using a 5 ml HiTrap Benzamidine FF (High Sub) column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden), which binds the thrombin (Guimarães and Bicca de Alencastro, 2002). The free his_6-tag was removed from the solution of protein by binding it to a 5 ml Ni^{2+} column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) equilibrated with 50 mM Tris-HCl buffer at pH 8.0, containing 500 mM NaCl, 30 mM imidazole and 0.02% (w/v) NaN_3. The WT and mutant proteins were concentrated and dialysed into 20 mM sodium phosphate buffer at pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3 for storage. The 20 mM sodium phosphate buffer at pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3 buffer was optimised from the methods of Habig *et al.* (1974) to reduce protein aggregation by preventing metal-induced oxidation (Wingfield, 1995). Samples were flash-frozen in 500 µL aliquots and stored at -80 °C.

2.3.1. Protein concentration determination

The protein concentration of the WT and mutant were obtained using a dilution series. All protein concentrations of WT and mutant hGST M1a reported in this study refer to subunit/monomer concentrations, and were determined spectrophotometrically at 280 nm using the Beer Law

$$A = \varepsilon C l \quad (1)$$

where A is the absorbance at 280 nm, ε is the molar extinction coefficient, C is the concentration and *l* is the path length (1 cm).

The molar extinction coefficient was calculated according to Pace *et al.* (1995):

$$\varepsilon = 5550\Sigma Trp + 1340\Sigma Tyr + 150\Sigma Cys \quad (2)$$

where ΣTrp, ΣTyr and ΣCys are the total number of Trp, Tyr and Cys in the subunit/monomer sequence of the proteins, respectively, and the numbers shown are their respective extinction coefficients. The resultant molar extinction coefficient of 40130 $M^{-1}.cm^{-1}$ was used for both the WT and mutant.

2.3.2. SDS-PAGE

The purity, homogeneity and molecular mass of the WT and the mutant were assessed using a 4% stacking gel and 15% separating gel (Laemmli, 1970). Proteins were incubated with SDS-PAGE sample buffer (0.125 M Tris-HCl, 4 % (w/v) SDS, 20% (v/v) glycerol, 5% (v/v) β-mercaptoethanol and 0.02% (w/v) bromophenol blue, pH 6.8) for 5 minutes at 95 °C prior to loading. Samples were electrophoresed for 40 minutes at 180 V and gels were stained in 2% (w/v) Coomassie Blue R250 staining solution containing 13.5% (v/v) glacial acetic acid and 18.75% (v/v) ethanol. The gels were destained with 40% (v/v) ethanol and 10% (v/v) glacial acetic acid until the background was clear. The distances migrated by individual proteins were compared to a set of known standards (molecular weight marker SM0431)

2.4. Structural characterisation of the WT and F56S/R81A mutant

2.4.1. Size exclusion high-performance liquid chromatography (SE-HPLC)

The hydrodynamic volumes of the WT and mutant were characterised by high-performance liquid chromatography (SE-HPLC). A TSK Gel SuperSW2000 size-exclusion column with a resolution of 5–150 kDa (TOSOH Corporation, Tokyo, Japan) was used in conjunction with a TSK gel SWXL guard column (TOSOH Corporation, Tokyo, Japan). The column was

connected to a Shimadzu pump and equilibrated at 20 °C with 20 mM sodium phosphate buffer pH 7.0, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. The pump maintained an isocratic pressure of approximately 900 psi at a constant flow rate of 0.2 $ml.min^{-1}$. Triplicate samples of 20 µM WT and 20 µM mutant were prepared and injected onto the column. A sample containing a mixture of 20 µM WT and 20 µM mutant was also run. The low molecular weight (LMW) gel filtration calibration kit (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) was used to construct a standard curve.

2.4.2. Far-UV circular dichroism

Far-UV circular dichroism (CD) measurements were conducted with 1 µM and 10 µM of both WT and mutant proteins in 20 mM sodium phosphate buffer pH 7.45, containing 1 mM EDTA and 0.02% (w/v) NaN_3. The Jasco model 810 CD spectropolarimeter was connected to a Peltier temperature controller. All experiments were conducted in a 2 mm cuvette at 20 °C. A data pitch of 0.2 nm, a bandwidth of 1 nm and a response of 0.5 s were used. Readings were taken at 222 nm over a 30 s time period with an average of five accumulations per sample at various urea concentrations. The continuous scanning mode was used for spectra recorded between 190 nm and 250 nm, and seven accumulations were collected per sample. CD spectra were recorded in triplicate in millidegree ellipticity, and buffer contributions were subtracted for all data collected. Triplicate readings were averaged and converted to mean residue ellipticity, [θ] ($deg.cm^2.dmol^{-1}$), using the following equation (Woody, 1995):

$$[\theta] = (100.\theta)/ Cnl \qquad (3)$$

where θ (mdeg) is the measured ellipticity signal at the respective wavelength, *C* is the protein concentration (mM), *n* is the total number of residues, and *l* is the path length (cm).

2.4.3. Intrinsic fluorescence

Fluorescence emission spectra were collected for 1 µM and 10 µM concentrations of both the WT and mutant proteins in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. All spectra were collected on a Jasco FP-6300 fluorescence spectrophotometer connected to a Peltier temperature controller set at 20 °C. A quartz cuvette with a 10 mm path length was used for all experiments. The excitation and emission bandwidths were set at 5 nm for 1 µM protein samples and at 2.5 nm for 10 µM protein samples, respectively. The emission spectra were collected using a scanning speed of 200 nm/min. Intrinsic tryptophan fluorescence was monitored using an excitation wavelength of 280 nm and 295 nm. Emission was monitored between 290-500 nm. The spectra obtained

are an average of three accumulations per triplicate sample, and all spectra were corrected for buffer contributions. Denatured proteins were prepared in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3 with 8 M urea.

A Bradford test was conducted to ensure that the same concentration of protein was present in all samples. Three-hundred microlitres of Bradford reagent was added to the protein samples in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. The absorbance at 595 nm was measured in triplicate.

2.5. Functional characterisation of the WT and F56S/R81A mutant

2.5.1. Specific activity

The enzyme-catalysed conjugation of CDNB to GSH at 20 °C was monitored spectroscopically by measuring the formation of 1-(*S*-glutathionyl)-2,4-dinitrobenzene at 340 nm (Habig *et al.*, 1974). The specific activity of the WT and mutant was assessed using the standard GSH-CDNB conjugation assay on a Jasco V-600 spectrometer. Samples of 1–10 nM protein were prepared in 100 mM sodium phosphate buffer at pH 6.5, containing 1 mM EDTA and 0.02% sodium azide in the presence of 1 mM GSH. The conjugation reaction was initiated by the addition of a 30 mM stock solution of CDNB solubilised in ethanol to create a final assay CDNB concentration of 1 mM and a final assay volume of 3 ml. The reaction was followed as linear progress curves by measuring the absorbance at 340 nm for 1 min at 20 °C. All reactions were corrected for the non-enzymatic controls. The specific activity of the enzyme was calculated using the extinction coefficient of the chromophore at 340 nm (ε_{340}=9600 $M^{-1}.cm^{-1}$) (Habig *et al.*, 1974). The specific activity ($\mu mol.min^{-1}.mg^{-1}$) was determined by linear regression of a plot between the initial velocity of complex formation ($\mu mol.min^{-1}$) versus protein amount (mg). All measurements were performed in triplicate and corrected for the non-enzymatic reaction rate.

2.5.2. ANS binding

A 20 mM ANS stock solution was prepared in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. The concentration of the ANS solution was determined using equation 1 (ε_{350} = 4950 $M^{-1}.cm^{-1}$) (Weber and Young, 1964). A concentration of 200 μM ANS was added to 1 μM of both the WT and mutant protein solutions prepared in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. The mixture was incubated at room temperature for 1 hour in the dark to promote the complete binding of ANS to the protein. The excitation

and emission bandwidths were set at 5 nm, and the emission spectra were collected using a scanning speed of 200 nm/min. Triplicate samples were excited at 390 nm, and emission spectra were recorded from 390-600 nm using a quartz cuvette with a 10 mm path length. The spectra obtained are an average of three accumulations per sample. All spectra were subsequently corrected for the fluorescence contribution from the unbound ANS.

2.6. Urea-induced equilibrium unfolding

Equilibrium unfolding studies were performed in the presence of urea so as to establish the degree of stability, (un)folding and cooperativity of the WT and mutant proteins. All unfolding experiments were performed at 20 °C in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. Proteins were denatured in urea concentrations ranging from 0 M to 8.5 M. A 10 M urea stock, prepared as described previously (Pace, 1986), was used to prepare all urea samples The pH of the urea stock solution was adjusted to pH 7.45, the solution was filtered and the urea concentration was determined using a refractometer (Pace, 1986). All protein solutions were prepared in triplicate and allowed to equilibrate for 1 hour before the structure of the protein was probed using CD, intrinsic tryptophan fluorescence, ANS binding, SE-HPLC and dynamic light scattering (DLS). Aggregation throughout all equilibrium unfolding studies was monitored by Rayleigh scattering (Brahma *et al*., 2009).

2.6.1. Reversibility of unfolding

Triplicate samples of 1 µM WT and 1 µM mutant were incubated in 8.5 M urea for 1 hour at 20 °C so as to denature the proteins. Refolding was induced by a six-fold dilution of the unfolded protein with 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. The reaction was allowed to proceed for 1 hour at 20 °C. The recovery of the secondary and tertiary structure of both the WT and mutant was determined. Far-UV CD spectra (see section 2.4.3.) of the refolded proteins were recorded from 190-250 nm so as to determine the extent of recovery of the secondary structure of both the WT and mutant. These spectra were compared with spectra recorded for 1 µM native and 1 µM denatured proteins to determine the percentage of secondary structure recovery. Similarly, fluorescence emission spectra recorded from 300-500 nm, using an excitation wavelength of 295 nm, were recorded for 1 µM refolded WT and 1 µM mutant proteins (see section 2.4.4.). These spectra were compared with native and denatured reference samples to calculate the percentage recovery of the tertiary structure.

In order to determine reversibility, triplicate samples of denatured WT and denatured mutant were diluted with 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3 to final urea concentrations ranging between 0-8.5 M urea. Each reaction was left for 1 hour at 20 °C. Refolding was monitored by CD measurements obtained at 222 nm (see section 2.4.3.) and by tryptophan fluorescence emission (see section 2.4.4.) at both 340 nm and 350 nm.

2.6.2. Urea-induced equilibrium unfolding spectroscopic probes

Secondary-structural changes were monitored by CD measurements obtained at 222 nm on a Jasco model 810 CD spectropolarimeter (see section 2.4.3.). Local changes in tertiary structure were monitored by tryptophan fluorescence on a Jasco FP-6300 fluorescence spectrophotometer as described (see section 2.4.4.). An excitation wavelength of 295 nm was used, and emission was monitored at 340 nm for folded protein and 350 nm for unfolded protein. CD measurements and intrinsic tryptophan measurements for unfolding experiments were carried out on both 1 μM and 10 μM protein samples.

ANS binding to protein was measured using a final concentration of 200 μM ANS added to 1 μM of WT and mutant and incubated at room temperature for 1 hour in the dark (see section 2.5.2). The concentration of the ANS stock solution was determined using the Beer-Lambert equation (ε_{350} = 4950 $M^{-1}.cm^{-1}$) (Weber and Young, 1964). ANS binding was monitored using an excitation wavelength of 390 nm, and emission was recorded from 400-700 nm on a Jasco FP-6300 fluorescence spectrophotometer connected to a Peltier temperature controller set at 20 °C. The spectra obtained are an average of three accumulations per triplicate sample, and all spectra were corrected for unbound ANS in buffer. The relative ANS binding was calculated according to the methods of Chaudhuri *et al*., 1993.

The hydrodynamic diameter of the various states of the WT and mutant in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3 was probed using DLS. A Zetasizer Nano-S light scattering device (Malvern Instruments Ltd, UK) was used with the laser set at 523 nm. All solutions were filtered through a 0.1 μM filter to remove dust particles. Measurements are an average of five readings carried out on 5 μM protein.

SE-HPLC was used to probe the oligomeric status and hydrodynamic volumes of the WT and mutant in 20 mM sodium phosphate buffer pH 7.0, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3 (see section 2.4.2.). The column was equilibrated with 20 mM sodium

phosphate buffer pH 7.0, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3 containing the appropriate concentration of urea. The concentration of urea in the flow-through was determined using a refractometer to ensure that the column was equilibrated before applying the corresponding unfolded protein to ensure that no refolding could take place.

2.6.3. Data fitting

Savuka, a global analysis package developed by Professor Osman Bilsel, was used to fit the WT dimer and mutant monomer data obtained from the far-UV CD and intrinsic tryptophan fluorescence probes. The Savuka package is freely available online, with detailed download and installation instructions (http://www.osmanbilsel.net/software/savuka).

The WT dimer data were fit using Savuka function 190, which is a custom model incorporated by Professor Bilsel to fit the data obtained in this study. Function 190 is a three-state model that allows for a dimerisation (or dimer dissociation) event of both the native state and an intermediate state. The mutant monomer data were fit using Savuka function 108, which was already available in Savuka. Function 108 is a three-state unfolding model that allows for the dimerisation of the intermediate only. The thermodynamic parameters $\Delta G(H_2O)$ and m were obtained from the fits for both WT dimer and mutant monomer.

2.7. Pulsed labelling hydrogen-deuterium exchange mass spectrometry (HDX-MS)

The equilibrium unfolding of the WT and mutant was monitored by pulsed labelling HDX-MS in an automated manner using the LEAP PAL HDX system (Leap Technologies, USA). The LC-MS/MS analysis was performed on an Agilent 1100 HPLC system coupled to an AB Sciex 6600 TripleTOF. Protein samples were diluted to yield a final concentration of 2 mg/ml protein. Protein was incubated in urea from 0 to 9 M in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. Protein samples were allowed to unfold for 1 hour to reach equilibrium. On-exchange reactions were initiated by transferring 4 µl of protein to a vial containing 16 µl D_2O containing the same urea concentration as in the protein solution to ensure that refolding did not occur during labelling. Stock urea/D_2O was prepared as previously described (Bai *et al.*, 1993) and the urea concentration of each labelling solution was determined by refractometry. Following labelling, the solution was transferred to a vial containing 30 µL of 2 M guanidinium

hydrochloride, 100 mM TCEP at 0 °C to quench the labelling reaction and to minimise the back-exchange of incorporated deuterium.

Quenched samples were injected onto a Poroszyme immobilised pepsin column 2.1 × 5 mm (Life Technologies) at a flow rate of 100 μl/min using 0.1 % (v/v) formic acid to fragment the proteins. The resulting peptides were desalted on an Acclaim PepMap trap column (0.3 x 5 mm) for 2 min using 0.1 % (v/v) formic acid and separated at 200 μl/min using a linear 10 min gradient of 10-40 % of 80 % acetonitrile/ 0.1% (v/v) formic acid on a Kinetex C18 column (2.1 x 5 mm). Proteolysis, desalting and peptide separation were all performed in a column incubated at 4 °C. The fully deuterated control was prepared by conducting deuteration overnight in a buffer supplemented with 0.02% (v/v) formic acid. The non-deuterated control was prepared using MilliQ water in place of D_2O. The 6600 TipleTOF mass spectrometer was operated in Data Dependent Acquisition (DDA) mode for peptide identification, and a precursor scan was collected for deuterium labelling of each sample. In DDA mode, precursor scans were acquired from *m/z* 360-1500 using an accumulation time of 250 ms followed by 30 product scans which were acquired from *m/z* 100-1800 at 100 ms each, for a total scan time of 3.3 sec. Charged ions ranging from 1^+ to 5^+ which fall in the mass range 360 -1500 *m/z* were automatically fragmented in Q2 collision cells using nitrogen as the collision gas. Collision energies were chosen automatically as a function of *m/z* and charge.

2.7.1. Data analysis

The peptide pool for both proteins was sequenced using PEAKS 6 (Bioinformatics Solutions Inc.). Deuterium exchange data for all peptides that are in common between the two proteins were processed using HD Examiner 1.3 (Sierra Analytics, LLC, Modesto,CA). Each peptide was manually filtered using the HD Examiner 1.3 interface, and 30 high-quality common peptides were selected. These 30 peptides covered 98.6 % of the sequence.

The level of deuterium incorporated for each peptide at each unfolding condition was calculated as described previously (Zhang and Smith, 1993) through the use of HD Examiner 1.3 software.

The total number of amide peptide hydrogens in each peptide was calculated:

$$N = T_N - 2 - T_{Pro} \quad (4)$$

where N is the total number of amide peptide hydrogens in the peptide, T_N is the number of residues that can be deuterated within the peptide, and T_{Pro} is the number of proline residues

contained within the peptide. The value of 2 in the equation refers to the first two amino acid residues of a peptide, because they are unable to retain a deuterium at the peptide bond between them.

Back-exchange of deuterium after labelling was corrected for by using the following formula (Zhang and Smith, 1993):

$$D_O = \frac{(m - m_{0\%})}{(m_{100\%} - m_{0\%})} \times N \qquad (5)$$

where D_O is the average number of deuterons per peptide, m is the average isotopic centroid of the partially deuterated peptide, $m_{0\%}$ is the average isotopic centroid of the non-deuterated peptide, $m_{100\%}$ is the average isotopic centroid of the fully-deuterated peptide and N is the total number of amide peptide hydrogens in the peptide.

Peptide mass spectra that show both unimodal and bimodal distributions were analysed to identify the species present as the proteins unfold at urea concentrations ranging from 0-9 M. A number of isotope envelopes could not be clearly resolved, and the data was therefore not fitted to Gaussian areas in order to quantify the populations of the various states in each peptide. The population of folded and unfolded molecules was determined using the deuterium levels found in each peptide and comparing these levels to the deuterium levels found in the folded reference and in the denatured reference, respectively (Yang and Smith, 1997).

The percentage protection indicates what percentage of the population is folded in each peptide:

$$\text{Percentage protection} = \frac{(\%D - \%D_{\text{denatured reference}})}{(\%D_{\text{folded reference}} - \%D_{\text{denatured reference}})} \qquad (6)$$

The percentage exposed indicates what percentage of the population is denatured:

$$\text{Percentage exposed} = \frac{(\%D - \%D_{\text{folded reference}})}{(\%D_{\text{denatured reference}} - \%D_{\text{folded reference}})} \qquad (7)$$

where % protection is the percentage of the population that is folded, % exposed is the percentage of the population that is denatured, %D is the percentage deuteration of the peptide in a particular concentration of urea, $\%D_{\text{folded reference}}$ is the percentage deuteration in the folded reference of a particular peptide and $\%D_{\text{denatured reference}}$ is the percentage deuteration in the denatured reference of a particular peptide.

Chapter 3: Results

3.1. Plasmid sequence verification

Sequencing of the plasmid DNA in the region coding for his_6-hGST M1a-1a WT fusion protein indicated that the codon harmonised nucleic acid sequence encoding the WT protein, with an N-terminal his_6-tag and thrombin cleavage site, contained the K205R mutation that was incorporated to prevent thrombin cleavage from occurring at this coding region (Figure 4A). Sequencing results obtained for plasmid DNA purified from the plasmid encoding his_6-F56S/R81A mutant hGST M1a confirmed that the desired double mutation had been incorporated, and that no additional mutations were incorporated during the thermal cycling reactions (Figure 4A and 4B).

3.2. WT and F56S/R81A mutant protein purification

The GST proteins were purified from the supernatants of sonicated *E. coli* cells, as described in section 2.3. The WT his_6-hGST M1a-1a was eluted from the Ni^{2+} affinity column using a 0-100 % gradient of 30-300 mM imidazole, using 50 mM Tris-HCl buffer at pH 8.0, containing 150 mM NaCl, 300 mM imidazole and 0.02% (w/v) NaN_3 (Figure 5A). The hexahistidine tag was then removed by thrombin cleavage to yield pure WT protein (Figure 5B).

His_6-F56S/R81A mutant was also eluted using a 0-100 % gradient of 30-300 mM imidazole, using 50 mM Tris-HCl buffer at pH 8.0, containing 150 mM NaCl, 300 mM imidazole and 0.02% (w/v) NaN_3 (Figure 6A) and the hexahistidine tag was also removed to yield pure F56S/R81A mutant (Figure 6B).

3.3. Structural characterisation of WT and F56S/R81A mutant

The structural features of WT hGST M1a-1a and F56S/R81A mutant were assessed using SDS-PAGE, SE-HPLC, far-UV circular dichroism and intrinsic fluorescence.

3.3.1. SDS-PAGE

SDS-PAGE was used to determine the size and the purity of the WT and the F56S/R81A mutant. The size of each protein was determined by constructing a standard curve using a set of protein molecular weight standards.

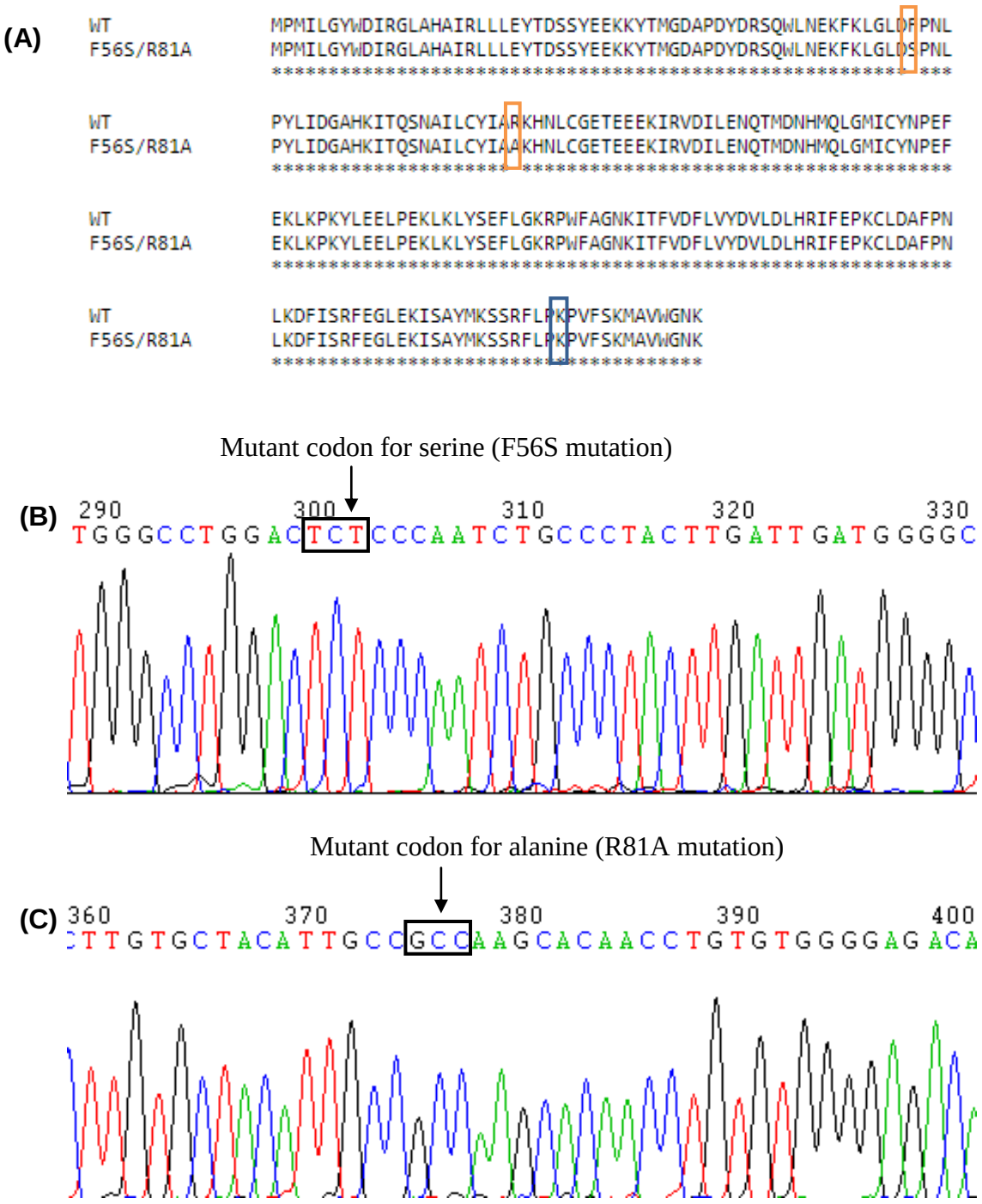


Figure 4: Sequence identity of WT and F56S/R81A mutant

(A) Sequence alignment showing F56S/R81A double mutation (orange box). The R205K mutation has also been incorporated to prevent cleavage by thrombin (blue box). WT and F56S/R81A mutant sequences were aligned using the Multiple Sequence Alignment Tool by CLUSTALW (http://www.genome.jp/tools-bin/clustalw). (B) Chromatogram showing truncated segment of F56S mutant sequence. The WT TTT codon for phenylalanine was replaced with the TCT codon for serine. (C) Chromatogram showing truncated segment of the R81A mutant sequence. The WT CGC codon for arginine was replaced with the GCC codon for alanine. The sequences were viewed using Chromas version 1.45 (32 bit) (http://www.technelysium.com.au/chromas.html; Technelysium Pty. Ltd., Helensvale, Australia).

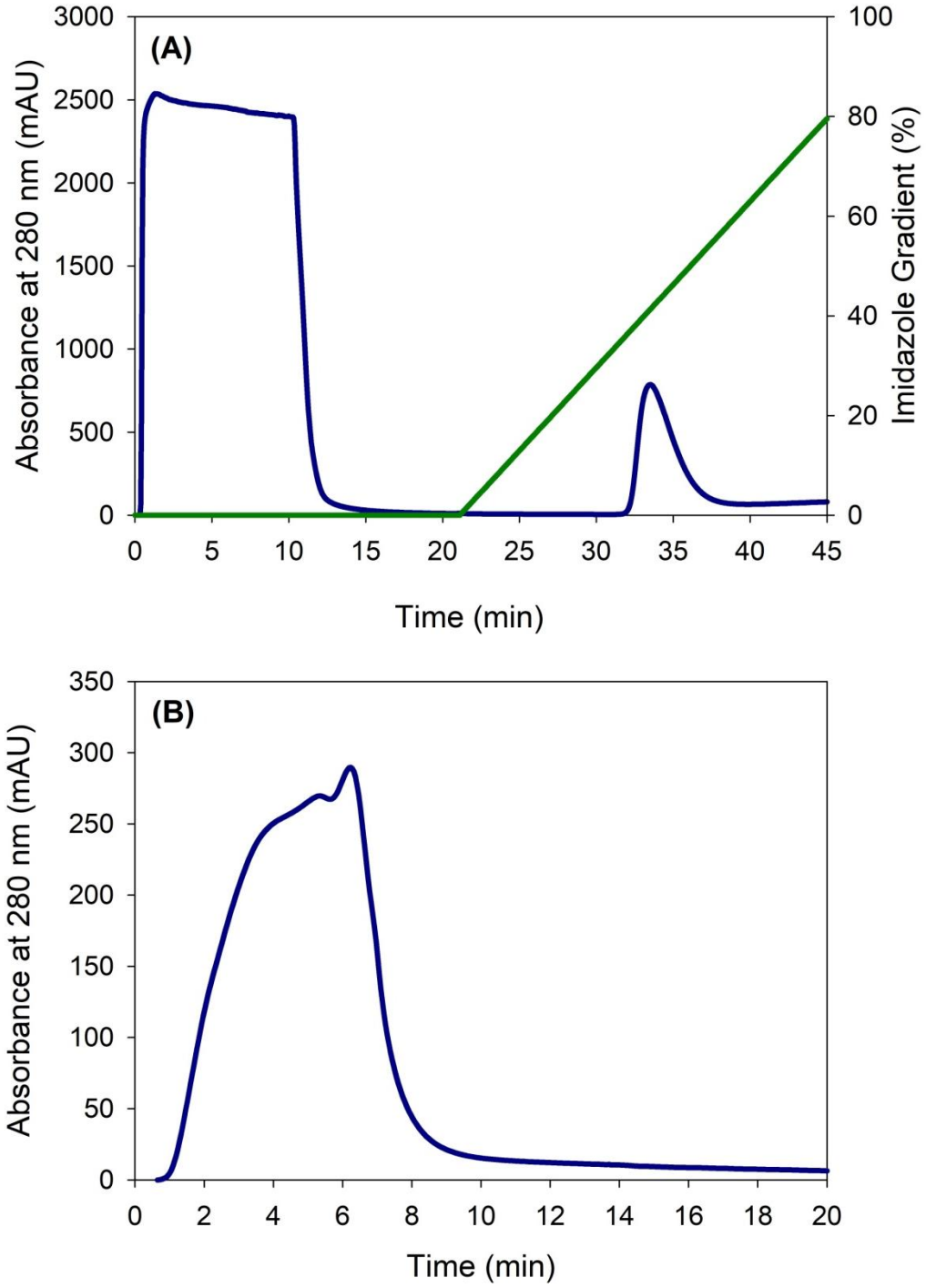


Figure 5: Elution profiles of the WT protein

(A) Elution profile of WT his_6-hGST M1a-1a shows the absorbance at 280 nm of the effluent (blue) and the 30-300 mM imidazole gradient, using 50 mM Tris-HCl buffer at pH 8.0, containing 150 mM NaCl, 300 mM imidazole and 0.02% (w/v) NaN_3 (green). Effluent was collected in 4 ml fractions. (B) Elution profile of WT protein cleaved overnight, and eluted by means of a Ni^{2+} affinity column coupled with a HiTrap Benzamidine FF (High Sub) column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden), using 50 mM Tris-HCl buffer at pH 8.0, containing 150 mM NaCl, 30 mM imidazole and 0.02% (w/v) NaN_3.

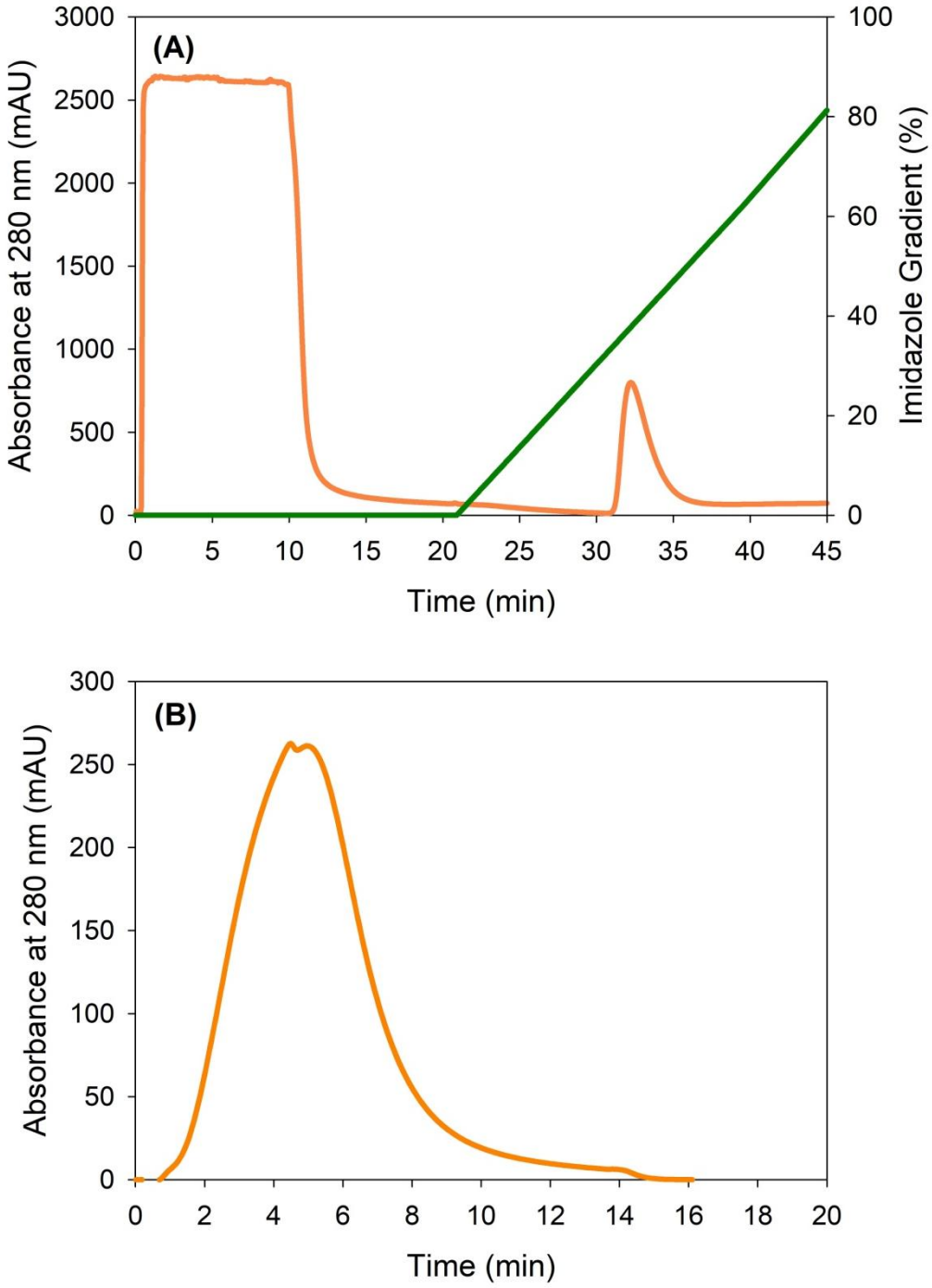


Figure 6: Elution profiles of the F56S/R81A mutant protein

(A) Elution of his_6-F56S/R81A mutant shows the absorbance at 280 nm of the effluent (orange) and the 30-300 mM imidazole gradient using 50 mM Tris-HCl buffer at pH 8.0, containing 150 mM NaCl, 300 mM imidazole and 0.02% (w/v) NaN_3 (green). Effluent was collected in 4 ml fractions. (B) Elution profile of F56S/R81A mutant cleaved overnight, and eluted by means of a Ni^{2+} affinity column coupled with a HiTrap Benzamidine FF (High Sub) column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden), using 50 mM Tris-HCl buffer at pH 8.0, containing 150 mM NaCl, 30 mM imidazole and 0.02% (w/v) NaN_3.

The proteins have approximately the same molecular mass under denaturing and reducing conditions (Figure 7). The size of the WT is ~29 kDa, which is consistent with results obtained for rGST M1a-1a (Hornby *et al.*, 2000; Luo *et al.*, 2002; Thompson *et al.*, 2006). The size of the F56S/R81A mutant is ~28 kDa. Both proteins are electrophoretically pure, as seen by the single bands (Figure 7 inset).

3.3.2. SE-HPLC

The hydrodynamic volume of a protein gives an indication of the oligomeric state of the protein. SE-HPLC was used to assess the hydrodynamic volumes, and therefore the oligomeric state, of the WT and F56S/R81A mutant.

The single, symmetrical peaks in figure 8A indicate that WT (blue) and F56S/R81A mutant (orange) proteins are pure. WT elutes at 16.8 minutes and F56S/R81A mutant elutes at 18.1 minutes, providing evidence that the F56S/R81A mutant has a lower molecular mass compared to the WT protein. A calibration curve was constructed and linear regression analysis was used to calculate the molecular masses of both proteins (Figure 8B). The mass of WT (measured using 20 µM protein) is 37 kDa and the mass of the F56S/R81A mutant (measured using 20 µM protein) is 25 kDa. The sample containing a mixture of 20 µM WT and 20 µM F56S/R81A mutant shows that WT and F56S/R81A mutant elute separately (Figure 8C). The symmetrical peaks of the SE-HPLC elution profiles of mutant monomer spiked with WT dimer (Figure 8C) indicate that the mutant monomer is a stable, folded species that does not associate with itself or with the WT dimer subunit. This result proves that the hydrodynamic volume, and therefore the molecular mass of the WT are greater than that of the F56S/R81A mutant. The smaller than expected size of the WT dimer is because of the anomalous behaviour of globular GST proteins when they pass through size-exclusion columns (Alves *et al.*, 2006; Chang *et al.*, 1999; Hornby *et al.*, 2000). SE-HPLC data is not of a high enough resolution to determine the exact mass of proteins that are not globular, like GSTs. However, the data is of a high enough resolution to conclude that the WT protein is dimeric, and the F56S/R81A mutant is monomeric.

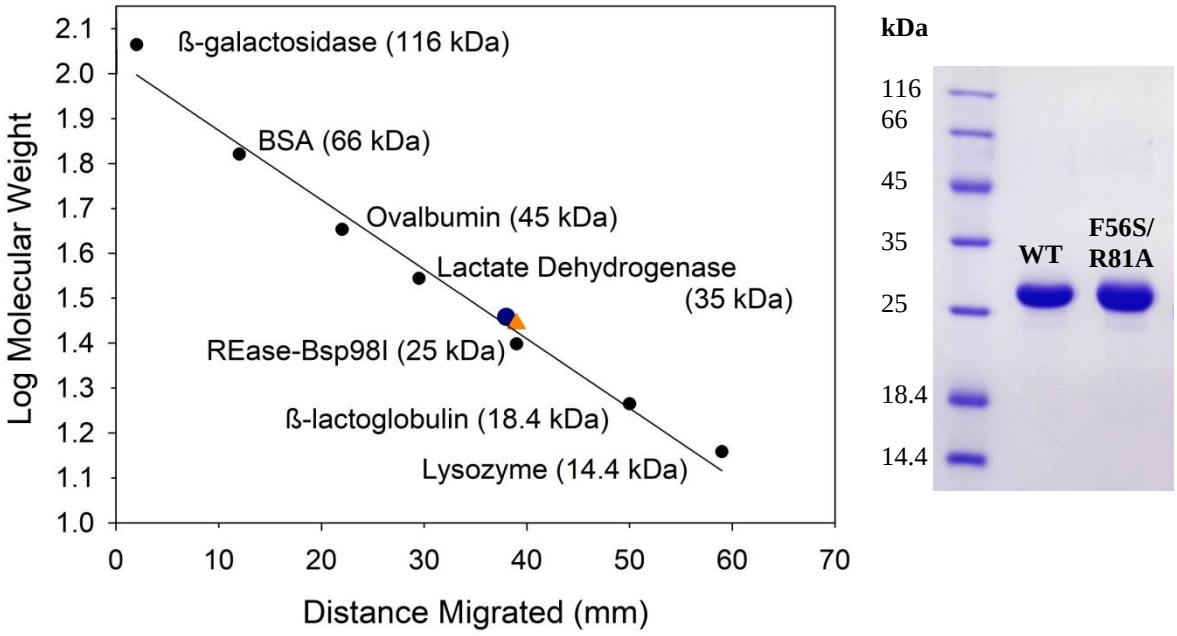


Figure 7: SDS-PAGE analysis of WT and F56S/R81A mutant

The calibration curve of known standards shows that WT (●) is ~29 kDa and F56S/R81A mutant (▲) is ~28 kDa. The standard curve was fitted to a straight line (R^2 = 0.98). The equation of the fitted line is y = -0.015x + 2.2028. The SDS-PAGE gel (inset) that was used to construct the standard curve shows that both proteins are pure.

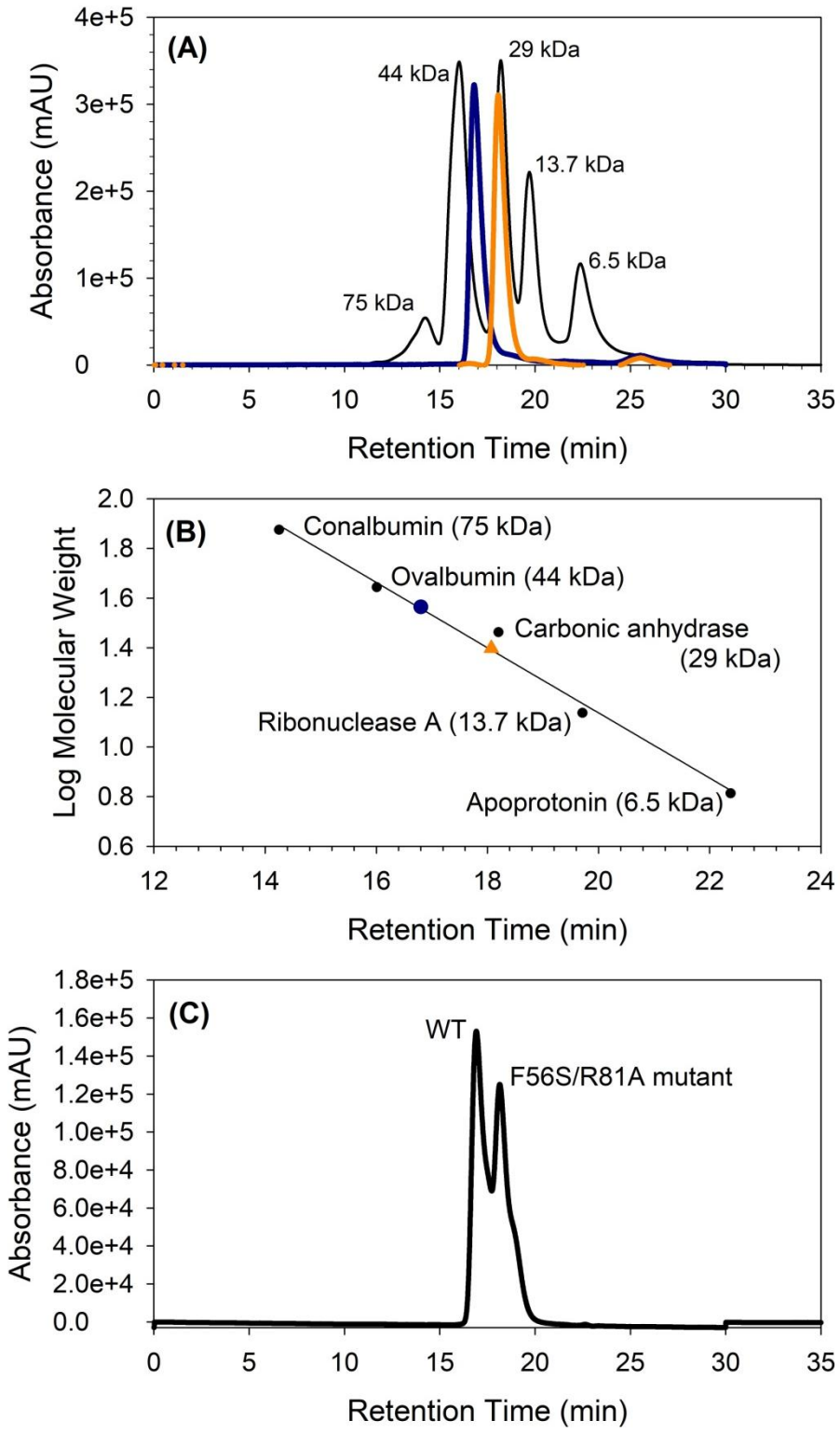


Figure 8: SE-HPLC elution profiles of WT and F56S/R81A mutant

(A) Elution profiles of WT (—) and F56S/R81A mutant (—) in contrast to the LMW gel filtration calibration kit standards (—). (B) Calibration curve of gel filtration kit standards fitted using linear regression. The equation of the straight line is y = -0.131x + 3.765 (R^2 = 0.99). WT (●) and F56S/R81A mutant (▲) are also indicated on the curve. (C) Elution profile of 20 µM WT spiked with 20 µM F56S/R81A mutant.

3.3.3. Far-UV circular dichroism

The far-UV region ranges from 180-250 nm, which corresponds to the wavelength at which the amide bonds that link amino acids are able to absorb circularly polarised light (Woody, 1995). The far-UV CD spectra of proteins exhibit characteristic features depending on the secondary structural elements of the polypeptide backbone of proteins (Woody, 1995). Far-UV CD is therefore an excellent probe to use when analysing the secondary structure of a protein.

Far-UV CD spectra were obtained for WT dimer (1 μM monomer) and mutant monomer (1 μM) (Figure 9). The predominantly α-helical WT dimer and mutant monomer spectra both show a peak at 190 nm and minima at 208 nm and at 222 nm. These characteristics are in accordance with far-UV CD spectra for predominantly α-helical proteins (Woody, 1995). The native-like spectrum of the mutant monomer differs slightly from that of WT dimer because mutant monomer exhibits a reduced intensity at 190 nm and at 222 nm.

3.3.4. Intrinsic fluorescence

Protein molecules have three dominant fluorescent amino acids: phenylalanine, tyrosine and tryptophan. However, only tryptophan absorbs light maximally at wavelengths greater than 295 nm, and can therefore be selectively excited (Lakowicz, 2006; Lakowicz and Masters, 2008).

The WT dimer and mutant monomer contain four tryptophan residues per subunit: W7, W45, W146 and W214 (Figure 2). Tyrosine and tryptophan were excited at 280 nm (Figure 10A), and tryptophan was selectively excited at 295 nm (Figure 10B). Fluorescence emission maxima values depend on the local environment of the tryptophan fluorophore. Figure 10 shows that the emission maximum for both the WT dimer and mutant monomer is at 340 nm. However, the fluorescence emission intensity for the mutant monomer is reduced compared to that of WT dimer. The difference in intensity seen in figure 10 is not a result of different protein concentrations between the two samples, as they were confirmed to be the same (see section 2.4.4).

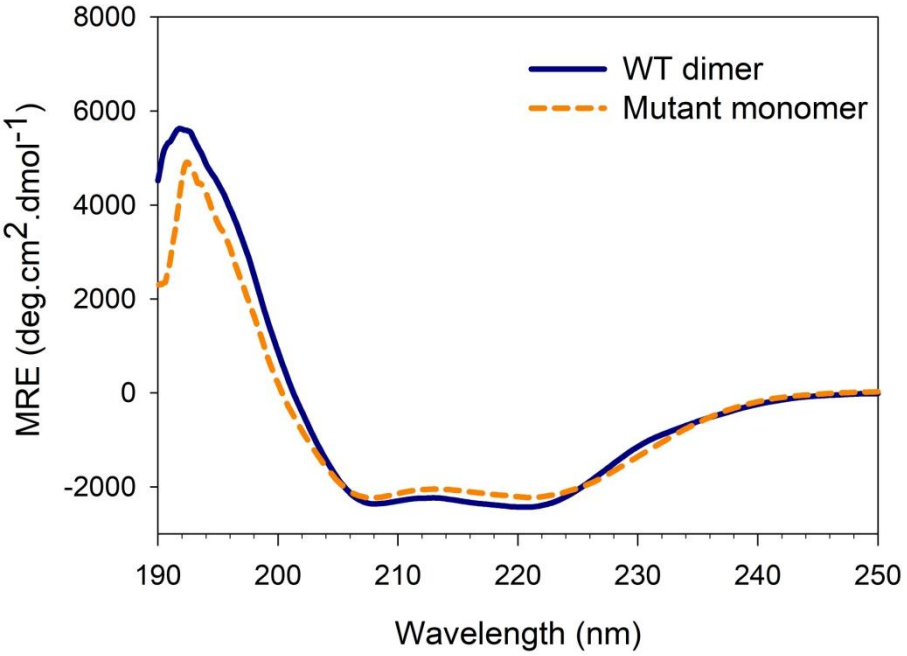


Figure 9: Far-UV CD spectra for WT dimer and mutant monomer

The resultant spectra for the WT dimer (1 µM monomer) (——) and mutant monomer (1 µM) (- - -**)** in 20 mM sodium phosphate buffer pH 7.45, containing 1 mM EDTA and 0.02% (w/v) NaN_3.are shown. Each spectrum represents the average obtained from three samples at 20 °C. All spectra were corrected for baseline signals.

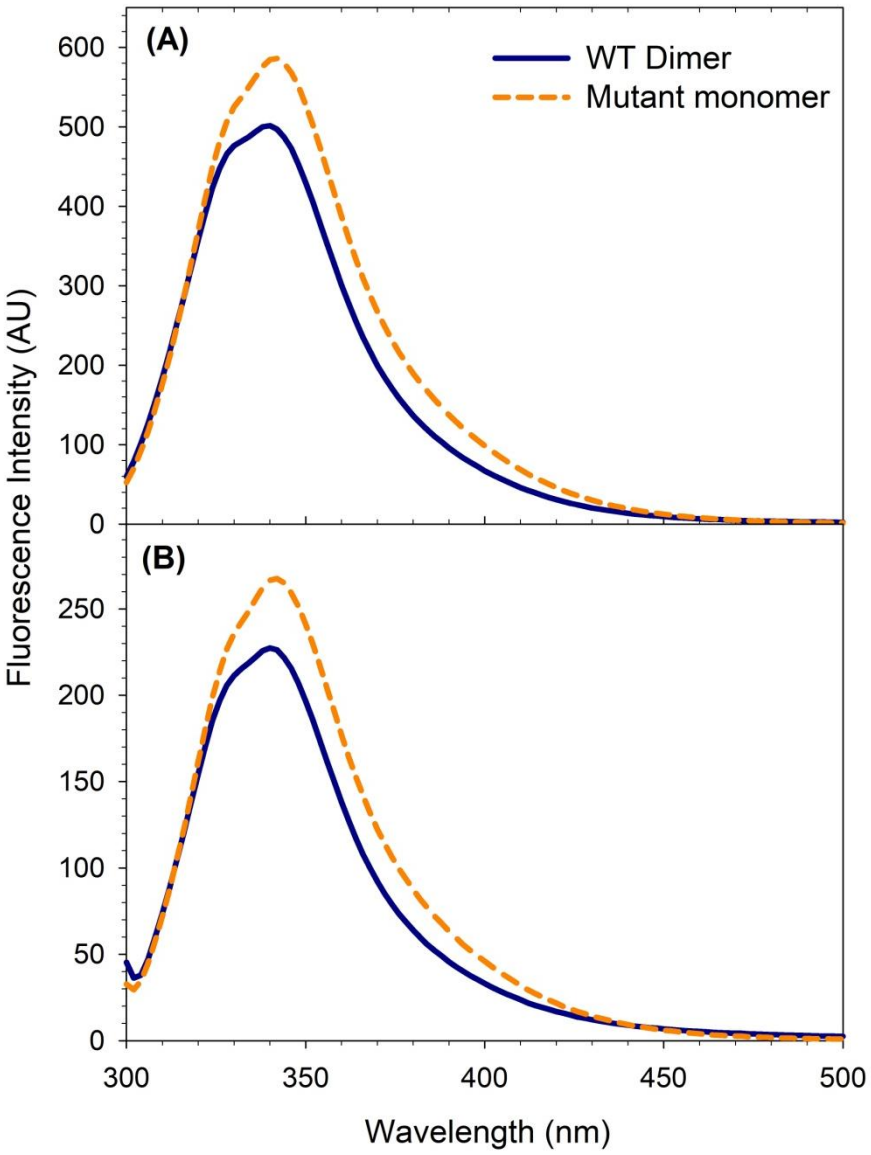


Figure 10: Fluorescence emission spectra for WT dimer and mutant monomer

(A) Tyrosine and tryptophan were excited at 280 nm. (B) Tryptophan was selectively excited at 295 nm. The emission spectra for WT dimer (1 µM monomer) (——) and mutant monomer (1 µM) (– – –) indicate that both proteins emit maximally at 340 nm when excited at 280 nm and at 295 nm in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. Each spectrum is the average of three separate protein samples at 20 °C. All spectra have been corrected for buffer contributions.

3.4. Functional characterisation of WT dimer and mutant monomer

The functional characteristics of mutant monomer were assessed in comparison with the functional characteristics of WT dimer. The standard GSH-CDNB conjugation assay was used to monitor specific activity, and ANS binding was used to monitor the non-substrate ligand binding properties of the protein.

3.4.1. Specific activity

The specific activity of both WT dimer and mutant monomer was determined using the standard GSH-CDNB conjugation assay (see section 2.5.1) (Habig *et al.*, 1974). Linear regression analysis gives the slope of the linear progress curves, which corresponds to the specific activity of that protein (Figure 11A). The specific activity of the WT dimer is 34.9±1.78 $\mu mol.min^{-1}.mg^{-1}$ and the specific activity of the mutant monomer is 0.001±0.005 $\mu mol.min^{-1}.mg^{-1}$.

A systematic error is observed in the plots in figure 11A because the plots do not intercept the origin at zero protein concentration. This is due to compensating for the non-enzymatic reaction. However, this does not impact the slopes (i.e., specific activity values) of the plots.

3.4.2. ANS binding

ANS has been shown to bind native GSTs, with little impact on the conformational stability and structure of the proteins (Erhardt and Dirr, 1995; Sluis-Cremer *et al.*, 1996; Stevens *et al.*, 1998). Free ANS in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3 shows an emission maximum at 520 nm. ANS bound to WT dimer displays a spectral blue shift from 520 nm to 500 nm, and ANS bound to mutant monomer displays a shift from 520 nm to 490 nm (Figure 11B) because the mutant monomer ANS binding site is less polar than the WT dimer ANS binding site. These spectral blue shifts are in accordance with those reported in other studies performed on rGST M1a-1a (Hornby *et al.*, 2000; Thompson *et al.*, 2006). A 3-fold increase in fluorescence intensity is observed for ANS bound to mutant monomer, when compared with that for ANS bound to the WT dimer (Figure 11B).

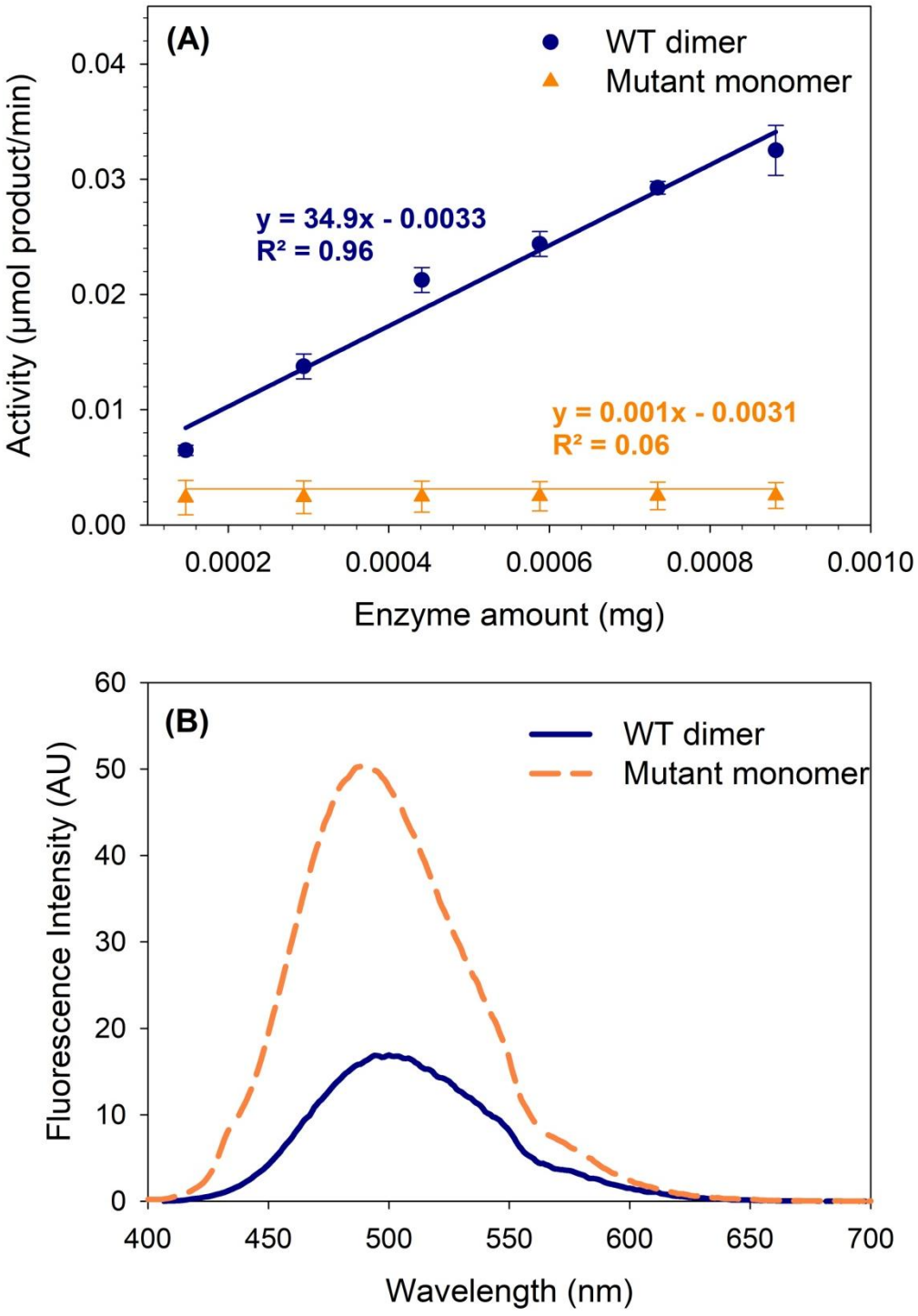


Figure 11: Functional characterisation of WT dimer and mutant monomer

(A) The specific activity for WT dimer (●) and mutant monomer (▲) was determined from the slope of the straight line fit. The experiment was performed in triplicate at 20 °C and the error bars show the standard deviation. (B) Fluorescence emission spectra of ANS bound to WT dimer (1 µM monomer) (——) and mutant monomer (1 µM) (- - -). ANS was selectively excited at 390 nm and emission spectra were recorded from 400-700 nm. A free ANS control was used to correct each spectrum for free ANS in solution, and each spectrum is the average of three separate protein samples at 20 °C.

3.5. Urea-induced equilibrium unfolding

A variety of spectroscopic probes can be used to monitor the unfolding and refolding transitions of proteins. Multiple probes should be used to gain an in-depth understanding of unfolding transitions under equilibrium conditions, because different probes report on changes that occur in different regions of the protein structure.

The equilibrium unfolding transitions of GSTs have been studied using far-UV CD to monitor secondary structural changes, tryptophan fluorescence to monitor both global and local tertiary structural changes, and ANS binding to monitor the formation of hydrophobic patches and intermediates (Dirr and Reinemer, 1991; Erhardt and Dirr, 1995; Fabrini *et al.*, 2009; Gildenhuys *et al.*, 2010, 2008; Hornby *et al.*, 2000; Kaplan *et al.*, 1997; Stevens *et al.*, 1998; Wallace *et al.*, 1998b). Furthermore, SE-HPLC can be used to determine the number and sizes of structural species present throughout the unfolding reaction, and DLS is useful in determining the hydrodynamic radius of various structural states as well as to ascertain whether aggregates have formed along the unfolding pathway.

3.5.1. Recovery and reversibility of WT dimer and mutant monomer unfolding

GST equilibrium unfolding studies show that GSTs unfold reversibly when using urea as a denaturant (Abdalla *et al.*, 2002; Aceto *et al.*, 1992; Gildenhuys *et al.*, 2010, 2008; Hornby *et al.*, 2000; Parbhoo *et al.*, 2011; Stevens *et al.*, 1998; Thompson *et al.*, 2006; Wallace *et al.*, 1998b). The thermodynamic parameters of an unfolding transition can only be determined if the reversibility of the unfolding pathway is established (Pace, 1986, 1990). Therefore, one must prove that the native state of the protein can be recovered once the protein has been denatured in urea. WT dimer and mutant monomer denatured in 8.5 M urea were diluted with 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. The recovery of the native secondary structure and tertiary structure was assessed using far-UV CD (see section 3.3.3) and intrinsic tryptophan fluorescence (see section 3.3.4), respectively.

Figures 12A and 12B show the spectroscopic data for refolded WT dimer and refolded mutant monomer in comparison with those for the native and denatured proteins. The spectra for both native and refolded WT dimer are similar and exhibit an emission maximum at 340 nm when excited at 295 nm (Figure 12A). The percentage recovery of native signal calculated from these spectra is ~100 %. A similar recovery was also calculated from the far-UV CD spectra (Figure 12B). The recoveries of native signals calculated from the spectroscopic data for the mutant monomer were also ~100% (Figures 12C and 12D). Table A in the appendix shows the exact values used to calculate the percentage recoveries. The reversibility of the unfolding-refolding processes for both proteins, monitored by far-UV CD and fluorescence, is demonstrated by the overlapping unfolding and refolding curves shown in figure 13, with is no evidence of unfolding/refolding hysteresis.

3.5.2. Unfolding monitored by tryptophan fluorescence and far-UV CD

Intrinsic fluorescence and far-UV CD were used as probes to monitor the equilibrium unfolding of 1 µM WT dimer and 1 µM mutant monomer (see section 2.6.2). The proteins were unfolded in 0 - 8.5 M urea and allowed to reach equilibrium for 1 hour at 20 °C. Tertiary structural changes during unfolding were monitored using fluorescence emission intensity values at 340 nm and 350 nm when excited at 295 nm, and secondary structural changes during unfolding were monitored using ellipticity values at 222 nm (Figure 14). The data indicate that unfolding is multi-state for both the WT dimer and mutant monomer (Figure 14). There are two distinct transitions (unfolding events): the first unfolding event takes place between 0 M to 4.5 M urea and the second unfolding event begins at approximately 5 M urea and continues until the proteins are denatured.

The change in the local environments of the four tryptophan residues (W7 and W45 are located in domain 1 and W146 and W214 are located in domain 2) is a good measure of the changes in the tertiary structure of the proteins as they unfold (Figure 14A and 14B). The WT dimer and mutant monomer unfolding curves in figure 14A are not superimposable during the first transition. However, the data do overlay during the second transition, indicating that the unfolding event(s) occurring during the second transition are equally cooperative and in all likelihood involve the same structural events. Figure 14B shows F350/F340 monitored by tryptophan fluorescence. At urea concentrations ranging between 2 M to 4.5 M the unfolding curves are not superimposable, whereas for all other urea concentrations the curves overlay.

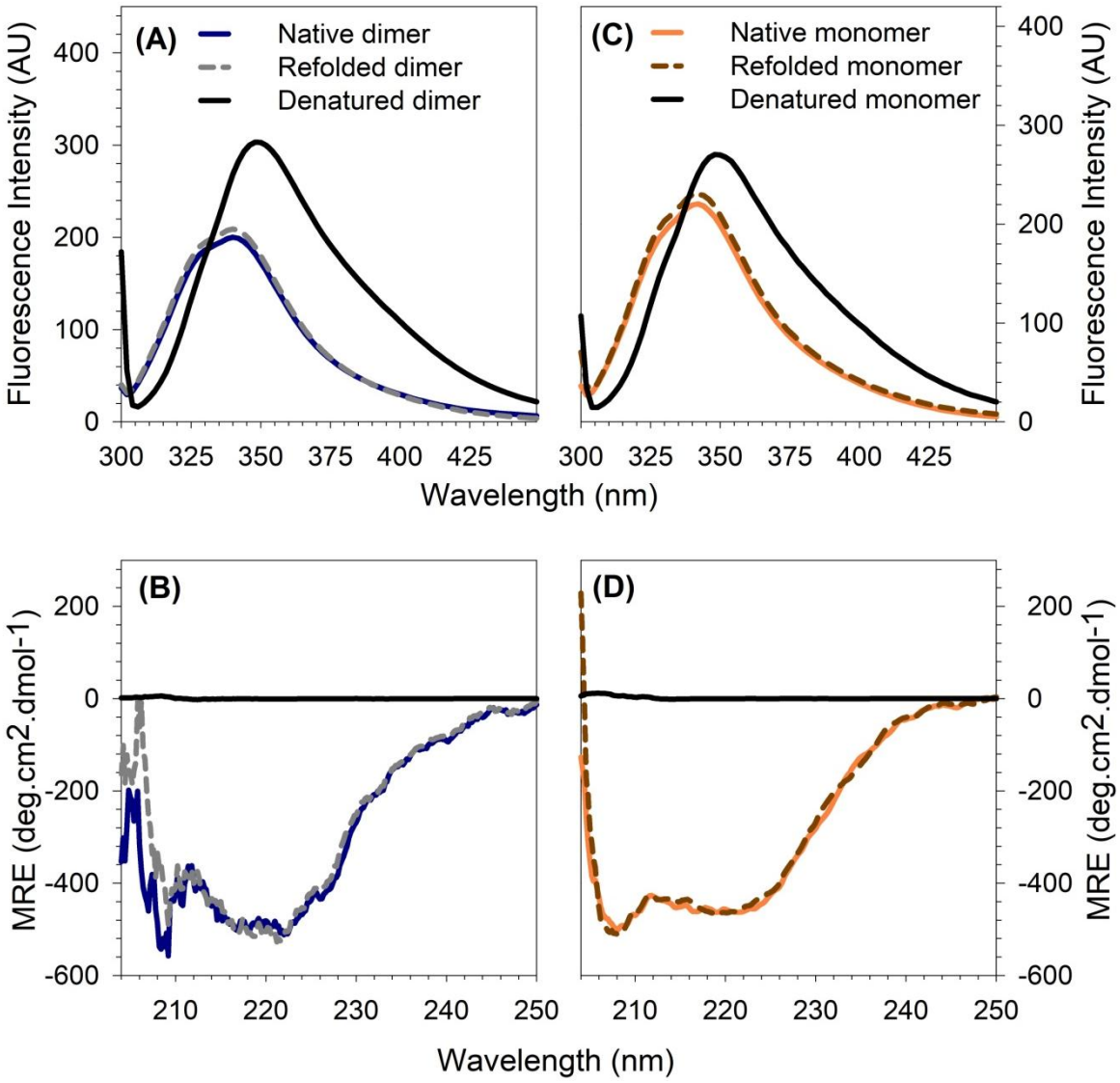


Figure 12: Recovery of the tertiary and secondary structure of WT dimer and mutant monomer

(A) The fluorescence spectrum for native WT dimer (——) excited at 295 nm is almost identical to that of refolded WT dimer (- - -). Both spectra show emission maxima at 340 nm, indicative of native protein because the emission maximum of denatured WT dimer (——) is 350 nm. (B) The far-UV CD spectrum for both native WT dimer (——) and refolded WT dimer (- - -) overlay fairly well. The two spectra are identical at 222 nm. (C) Fluorescence spectra for native mutant monomer (——) and refolded mutant monomer (- - -) overlay and both exhibit an emission maximum at 340 nm when excited at 295 nm. (D) Native mutant monomer (——) and refolded mutant monomer (- - -) far-UV CD spectra are almost identical. Both spectra exhibit identical troughs at 208 nm and at 222 nm. The final concentration for native, refolded and denatured protein was 1 µM (monomer) for WT dimer and mutant monomer. Each spectrum is the average of three samples at 20 °C. The percentage recovery for A, B, C and D can be obtained from table A in the appendix.

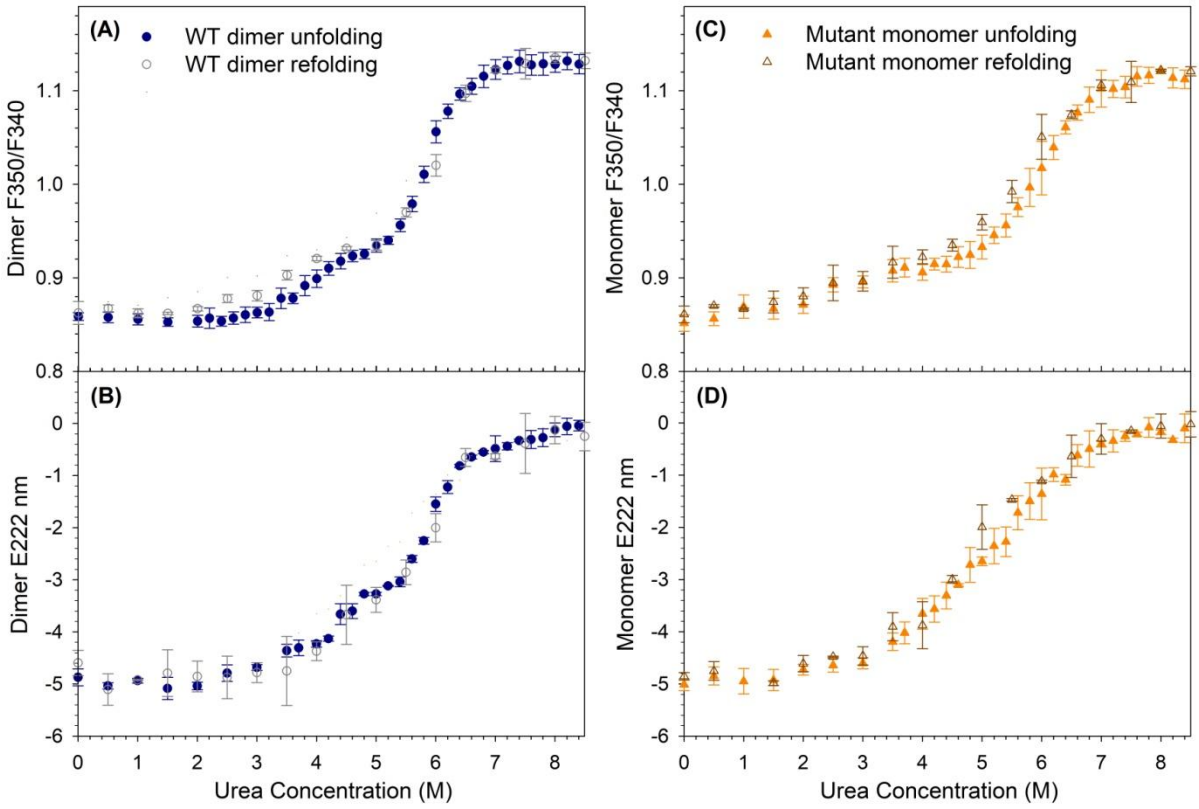


Figure 13: Reversibility of WT dimer and mutant monomer unfolding

The refolding curve of the WT dimer (1 µM monomer) (○) is super-imposable with the unfolding curve of the WT dimer (1 µM monomer) (●) when monitored using (A) intrinsic tryptophan fluorescence (F350/F340) and (B) far-UV CD ellipticity at 222 nm. The refolding curve of the mutant monomer (1 µM) (Δ) is super-imposable with the unfolding curve of the mutant monomer (1 µM) (▲) when monitored using (C) intrinsic tryptophan fluorescence (F350/F340) and (D) far-UV CD ellipticity at 222 nm. An excitation wavelength of 295 nm was used for fluorescence measurements. Refolding was allowed to proceed for 1 hour at 20 °C. All experiments were performed in triplicate in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3.

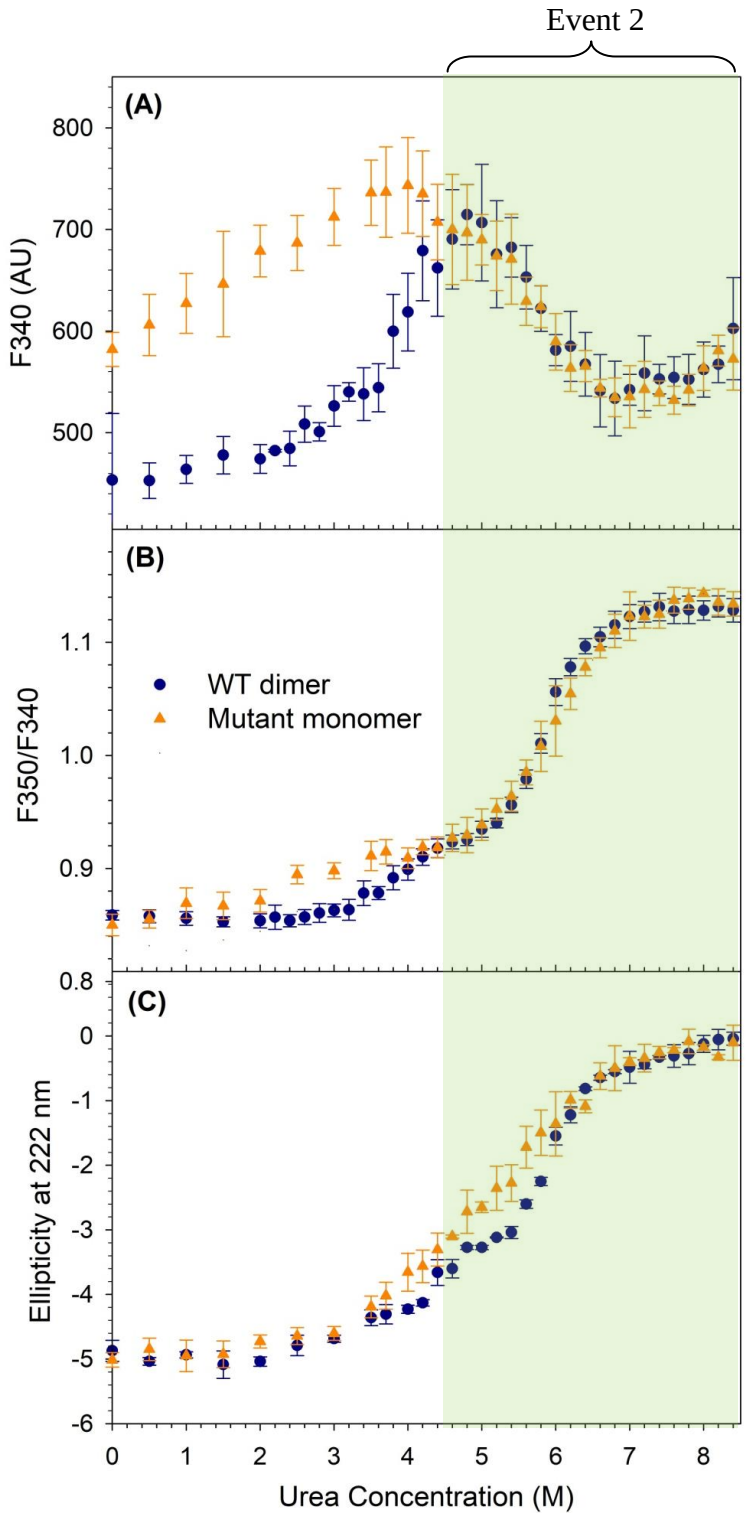


Figure 14: Urea-induced equilibrium unfolding of WT dimer and mutant monomer monitored by intrinsic tryptophan fluorescence and ellipticity at 222 nm

Unfolding transitions for the WT dimer (1 µM monomer) (●) and mutant monomer (1 µM) (▲) protein were monitored by (A) fluorescence emission at 340 nm, (B) ratio of intrinsic fluorescence intensities at 350 nm and 340 nm and (C) far-UV CD ellipticity at 222 nm. All unfolding experiments were performed in triplicate at 20 °C in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3.

Unfolding monitored by ellipticity at 222 nm (E222 nm) is evidently multi-state for the WT dimer, but appears to be two-state for the mutant monomer (Figure 14B). However, the intrinsic fluorescence probe has shown that unfolding is not a simple two-state process for both WT dimer and mutant monomer (Figure 14 A and 14B). The unfolding curves are non-coincident from 4 M to 6 M urea, indicating that the secondary structural changes that occur during the unfolding of WT dimer and mutant monomer differ in both unfolding event 1 and unfolding event 2.

The multi-state unfolding transitions indicates that stable intermediate state(s) are formed along the unfolding pathways of the WT dimer and mutant monomer, and before the data could be fitted to obtain thermodynamic parameters further investigation was required.

3.5.3. Protein-concentration dependent unfolding

A protein concentration-dependent unfolding study was conducted to understand at what urea concentration dimer dissociation occurs, and whether the intermediate(s) formed are monomeric or dimeric for both the WT dimer and mutant monomer. Urea-induced equilibrium unfolding transitions for the WT dimer and mutant monomer were monitored at 1 µM and at 10 µM protein concentrations. Intrinsic tryptophan fluorescence (F350/F340) and far-UV CD (ellipticity at 222 nm) were used as tertiary and secondary structural probes, respectively.

The first event of WT dimer unfolding (0 M to 4.5 M urea) is dependent on protein concentration because the unfolding curves show a shift to a higher urea concentration when protein concentration is increased (Figure 15A and 15B). There is a clear shift to a higher urea concentration when the protein concentration is increased for WT dimer unfolding event 2 when monitored using fluorescence (Figure 15A). However, the shift to a higher urea concentration is less evident when monitored by ellipticity at 222 nm (Figure 15B). Figure 15C and 15D show that mutant monomer unfolding is protein concentration-dependent. Unfolding monitored using intrinsic tryptophan fluorescence shows a shift to a higher urea concentration for 10 µM protein compared with 1 µM protein at urea concentrations from 3.5 M to 6.5 M (Figure 15C). Ellipticity at 222 nm was used to monitor changes in the secondary structure of mutant monomer and this probe elucidated a protein concentration-dependent unfolding transition from 4.5 M to 6.5 M urea. The protein-concentration dependence of WT dimer unfolding is more apparent when monitored by intrinsic tryptophan fluorescence because the dataset is more populated than the dataset obtained using far-UV CD as a probe (Figure 15B).

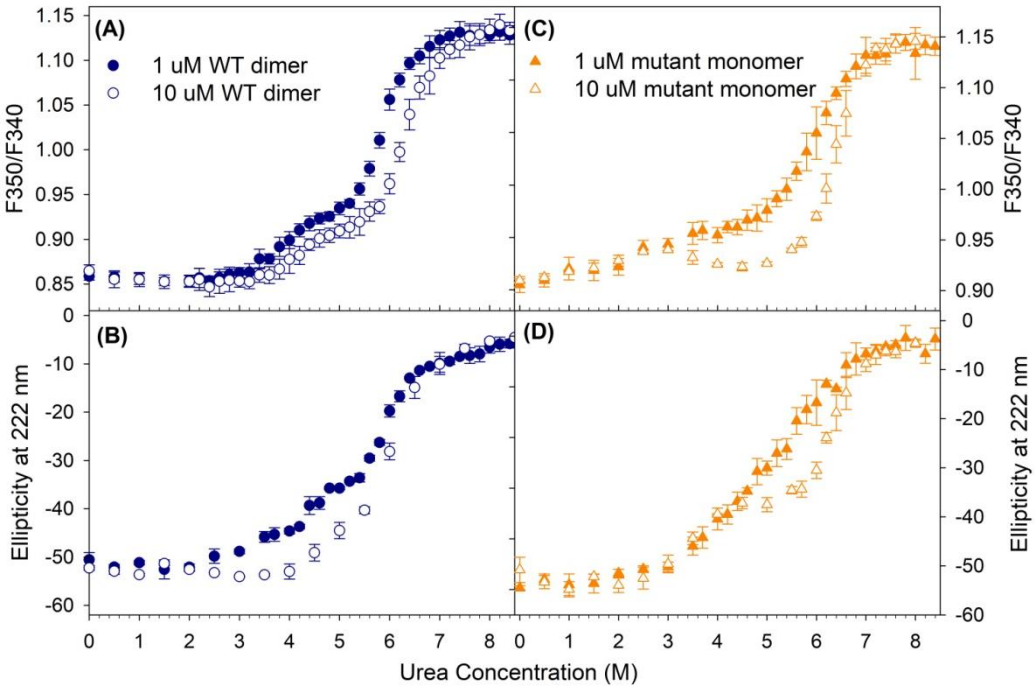


Figure 15: Protein concentration-dependence of WT dimer and mutant monomer unfolding

WT dimer unfolding was probed by (A) intrinsic tryptophan fluorescence (F350/F340) and (B) far-UV CD ellipticity at 222 nm using 1 µM (monomeric concentration) (●) and at 10 µM (monomeric concentration) (○). Mutant monomer unfolding was also probed by (C) intrinsic tryptophan fluorescence (F350/F340) and (D) far-UV CD ellipticity at 222 nm using 1 µM (▲) and at 10 µM (△). An excitation wavelength of 295 nm was used for fluorescence measurements. All unfolding experiments were performed in triplicate at 20 °C in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3.

3.5.4. Unfolding monitored by ANS binding

ANS bound to native WT dimer displays a spectral blue shift from 520 nm to 500 nm and ANS bound to native mutant monomer shifts from 520 nm to 490 nm, indicating that the dye has bound the hydrophobic regions in both proteins (Figure 16A and 16B). Figure 16C shows that ANS binding steadily increases between 1 M and 4.5 M urea during WT dimer unfolding, whereas mutant monomer shows no increase in ANS binding with an increase in urea concentration in this range. The unfolding curves are coincident for both proteins between 5 M and 8 M urea and ANS binding decreases for both proteins until only the denatured protein species exists from 7 M to 8.5 M urea (Figure 16C). The Rayleigh scatter at 390 nm shows no significant increase as the protein unfolds in various urea concentrations, indicating the absence of protein aggregation during unfolding (Figure 16D).

3.5.5. Unfolding monitored by SE-HPLC

Unfolding monitored by fluorescence emission when excited at 295 nm, ellipticity at 222 nm and ANS binding suggest that hGST M1a-1a unfolding might involve more than one intermediate. In order to identify all of the intermediates formed during unfolding, SE-HPLC was used to determine the hydrodynamic size of all protein species present at urea concentrations between 0 M and 8 M.

Figure 17A-D shows that the native WT dimer (N_{WT}) dissociates to form a monomeric intermediate (M_{WT}) at 4.5 M urea (Figure 17E and 17F). An additional intermediate, which has a greater hydrodynamic volume than the native protein, is also formed at 4.5 M urea for WT dimer (Figure 17E and 17F). The large oligomeric intermediate (I_{WT}) is only present when the monomeric intermediate (M_{WT}) is present, suggesting that I forms as a result of the non-native association of M (Figure 17E and 17F). Figures 17J and 17K show that native mutant monomer (M_{Mutant}) becomes destabilised in 3 M urea (Figure 17L), resulting in the formation of the oligomeric intermediate (I_{Mutant}) (Figure 17 L-O). At 6 M urea oligomeric I is present for both WT dimer (Figure 17G) and mutant monomer (Figure 17P) unfolding, along with a fraction of denatured (D) protein. At urea concentrations above 7 M only denatured protein forms exist for the WT dimer (Figure 17H-I) and for mutant monomer (Figure 17Q-R). These data provide evidence that the oligomeric I is an additional intermediate present on the unfolding pathway of the WT dimer hGST M1a-1a and mutant monomer hGST M1a.

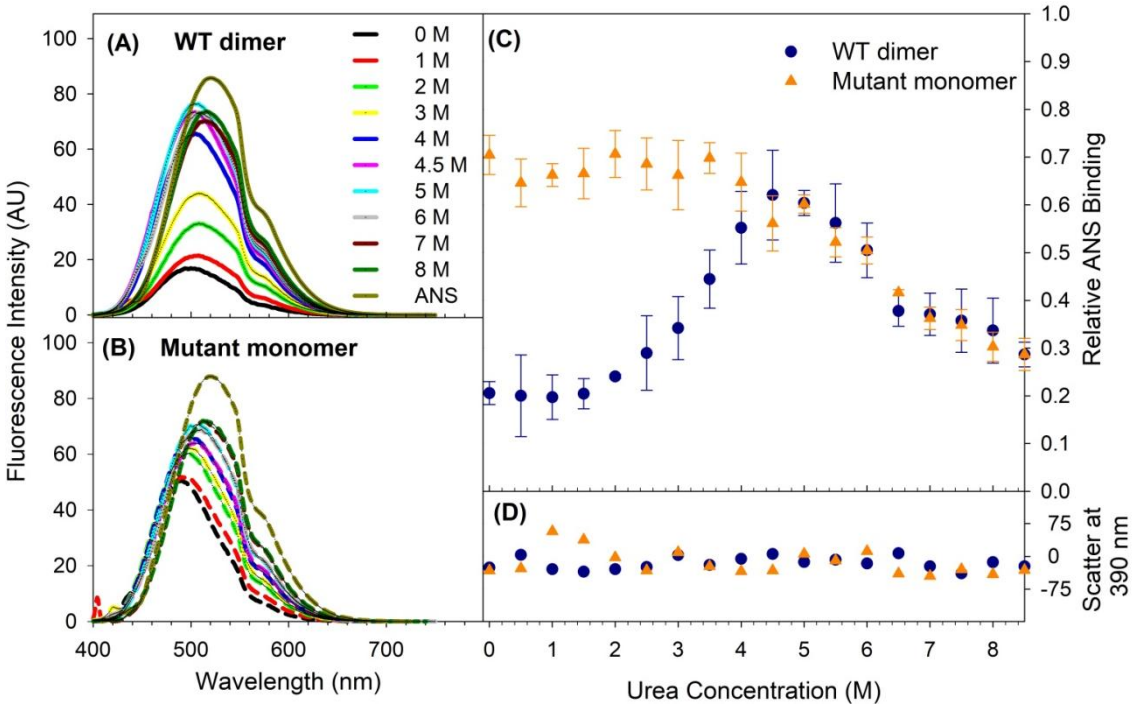


Figure 16: Urea-induced equilibrium unfolding of WT dimer and mutant monomer monitored by ANS binding

Fluorescence emission spectra of ANS bound to (A) WT dimer (1 µM monomer) and (B) mutant monomer (1 µM) in urea concentrations ranging from 0M to 8 M. ANS was selectively excited at 390 nm and emission spectra were recorded from 400-700 nm. The free ANS control (—) was used to correct each spectrum for free ANS in solution. Each spectrum is the average of three samples at 20 °C. (C) Unfolding transitions for 1 µM WT dimer (●) and 1 µM mutant monomer (▲) was monitored by ANS binding. All unfolding experiments were performed in triplicate at 20 °C in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3.

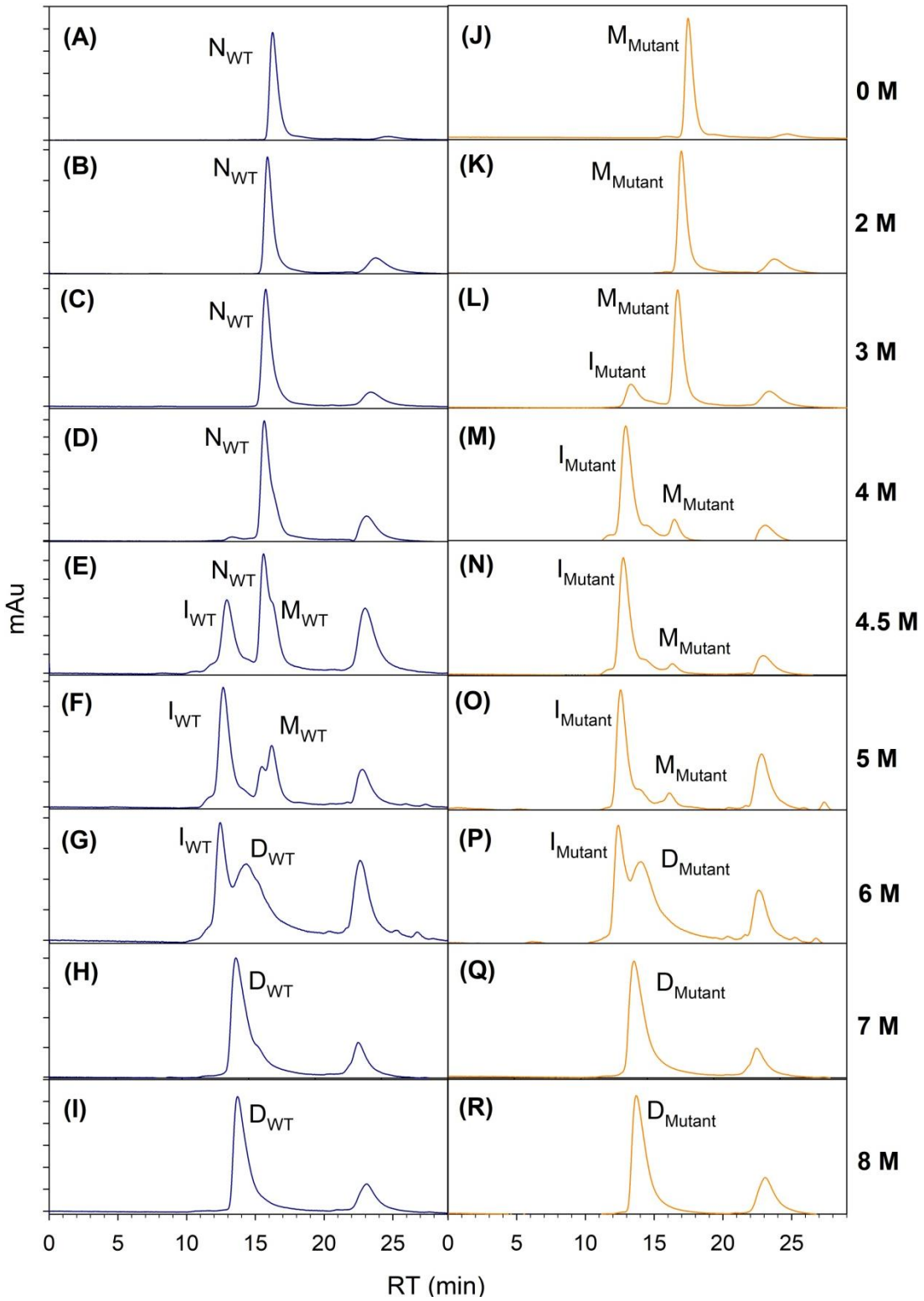


Figure 17: SE-HPLC elution profiles of WT dimer and mutant monomer incubated in 0 M to 8 M urea

(A-I) WT unfolding (5 µM monomer) (—) and (J-R) mutant monomer unfolding (1 µM) (—) at 20 °C, in 20 mM sodium phosphate buffer pH 7, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. The column was pre-equilibrated at the same urea concentration as the sample. Native proteins are labelled N, intermediate states are labelled I and denatured states are labelled D. The elution peak seen at ~23 minutes (A-R) is an artefact caused by the column.

It is interesting to note that when the retention time of each peak was analysed, monomeric M_{WT} at 5 M urea elutes at the same time as M_{Mutant} destabilised in 5 M urea (Figure 18). The oligomeric intermediates formed by both WT dimer (I_{WT}) and mutant monomer (I_{Mutant}) are the same size in urea concentrations from 4.5 M to 6 M urea and the denatured states are also equal in size (Figure 18). I_{WT} and I_{Mutant} will be referred to as I from this point forward. The data in figure 17 are in accordance with unfolding curves for all probes being coincident throughout the second unfolding event.

3.5.6. Unfolding monitored by DLS

DLS was used to assess the hydrodynamic diameter of the predominant species at varying urea concentrations. Monodisperse samples of both WT dimer and mutant monomer indicate the WT dimer has a greater diameter than the mutant monomer. Native WT dimer has a diameter of 8.7 nm, in accordance with previously reported results (Lipin *et al.*, 2008) (Figure 19A). Native mutant monomer has a diameter of 5.6 nm (Figure 19B).

The SE-HPLC data (see section 3.5.5) shows that at 3 M urea the native mutant monomer forms M_{Mutant} and I_{Mutant} (Figure 17L). The polydisperse sample at 3 M urea indicates that the mutant monomer has been denatured to such an extent that the hydrodynamic diameter of the protein is 8.7 nm (Figure 19B). Finally, from 5 M to 8 M urea the species present along the unfolding pathway are the same diameter for both WT dimer and mutant monomer (Figure 19A and 19B), indicating that the same species are present in the polydisperse WT dimer and mutant monomer samples. .

Absorbance spectra were recorded for each of the 5 µM protein samples that were used for the DLS measurements. The absorbance at 340 nm was recorded for each triplicate sample to confirm that the protein does not form insoluble aggregates during unfolding (Figure 19C). The absorbance at 340 nm remained low for each sample, indicating that no insoluble aggregates exist in each sample.

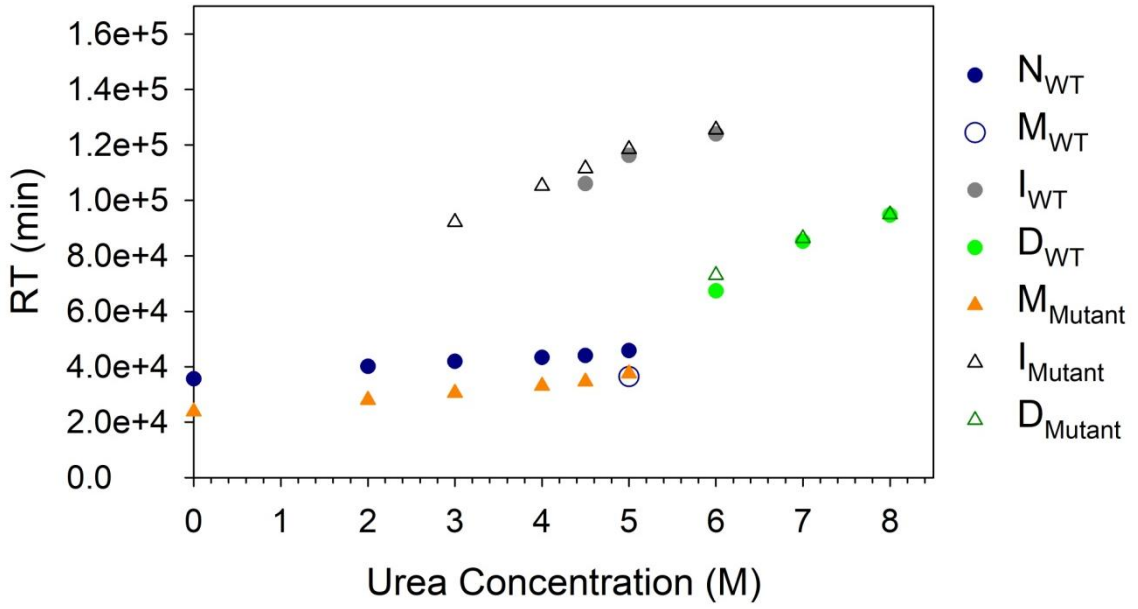


Figure 18: SE-HPLC retention time of each peak eluted for the WT dimer and mutant monomer when unfolded in 0 M - 8 M urea

The various species present at different urea concentrations are shown for both WT dimer and mutant monomer. The native WT (N_{WT}) elutes at a later time and therefore increases in size as the urea concentration is increased (●) until the dimer dissociates to form the monomeric intermediate M_{WT} (○) and the oligomeric intermediate I_{WT} (●). Denatured WT protein (D_{WT}) is the least globular structure (●). The native mutant monomer (M_{Mutant}) elutes at the same time from 0 M to 3 M urea, indicating that no major structural changes have occurred (▲). M_{WT} (○) and M_{Mutant} (▲) are the same size because they elute at the same retention time at 5 M urea. M_{Mutant} (▲) becomes destabilised and associates to form I_{Muant} (Δ). Oligomeric I formed by WT dimer (●) and mutant monomer (Δ) have similar retention times from 4.5 M to 6 M urea. Denatured mutant monomer (D_{Mutant}) forms at 6 M urea and is the only species present from 7 M to 8 M urea (Δ). D_{WT} (●) and D_{Mutant} (Δ) elute at the same time, indicating that the structures of these species are very similar.

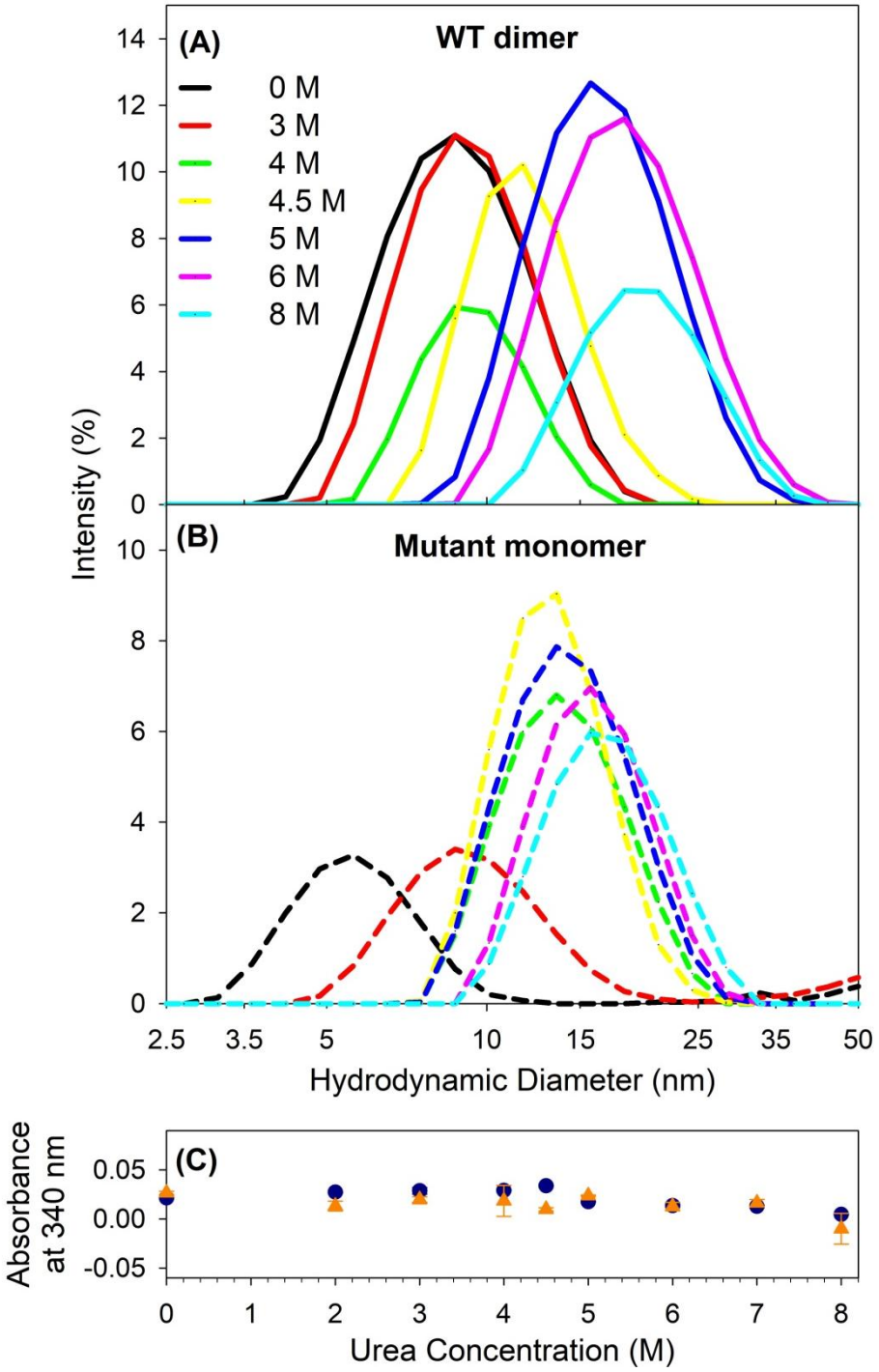


Figure 19: Urea-induced equilibrium unfolding of WT dimer and mutant monomer monitored by DLS

(A) DLS measurements showing the hydrodynamic diameter of (A) WT dimer (5 µM monomer) and (B) mutant monomer (5 µM) incubated in urea concentrations ranging from 0 M to 8 M. (C) Absorbance values at 340 nm of WT dimer (5 µM monomer) and mutant monomer (5 µM) incubated in urea concentrations ranging from 0 M to 8 M. All readings were performed in triplicate and spectra are an average of the triplicate readings. All protein samples were incubated in urea for 1 hour at 20 °C in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3.

3.5.7. Structure of the intermediates

The secondary and tertiary structure of the WT dimer and mutant monomer was assessed at various urea concentrations in order to determine structural elements of each intermediate. However, this was challenging because a mixture of species exists at the various urea concentrations and it was not possible to isolate each of these species.

Far-UV CD spectra of 1 µM WT dimer and 1 µM mutant monomer were compared under different urea conditions at 20 °C. The WT dimer maintains secondary structure in 0 M to 4 M urea (Figure 20A). At 4.5 M urea there is a loss in the secondary structure of the protein (Figure 20A). M_{WT} and I_{WT} are both present at 4.5 M urea and the spectrum at 4.5 M urea shows that both intermediates have a reduced secondary structure compared to that of the native WT. However, it is not possible to determine which intermediate contributes most to the signal. Secondary structure is gradually lost until the protein is denatured in 8 M urea. Mutant monomer shows a loss of secondary structure at 4 M urea (Figure 20B) even though all other unfolding data show that M_{Mutant} and I_{Mutant} are formed at 3 M urea. Again, this is because the far-UV CD spectra show an average of the signals from all species present. The mutant monomer shows a decrease in secondary structure until the protein is denatured in 8 M urea.

Intrinsic tryptophan fluorescence emission spectra of 1 µM WT dimer and 1 µM mutant monomer were compared under different urea conditions at 20 °C. The WT dimer shows no decrease in tertiary structure at 2 M urea (Figure 21A). However, at 3 M urea there is a loss in tertiary structure; this loss gradually increases until 5 M urea (Figure 21A). This differs from the far-UV CD data in that at 3 M urea the WT dimer still exhibits native-like secondary structure (Figure 21A). At 6 M urea the predominant species within the sample is denatured because the maximum emission shifts from 340 nm to 350 nm (Figure 21A) and at 8 M urea the protein is denatured. The mutant monomer shows a loss of tertiary structural contacts in 2 M urea (Figure 21B). The tertiary structure of the species present at 4 M, 4.5 M and 5 M urea is the same because these spectra overlay very well (Figure 21B). Mutant monomer also shows that at 6 M the predominant species is denatured because the emission maximum shifts from 340 nm to 350 nm (Figure 21B), until the protein is fully denatured at 8 M urea.

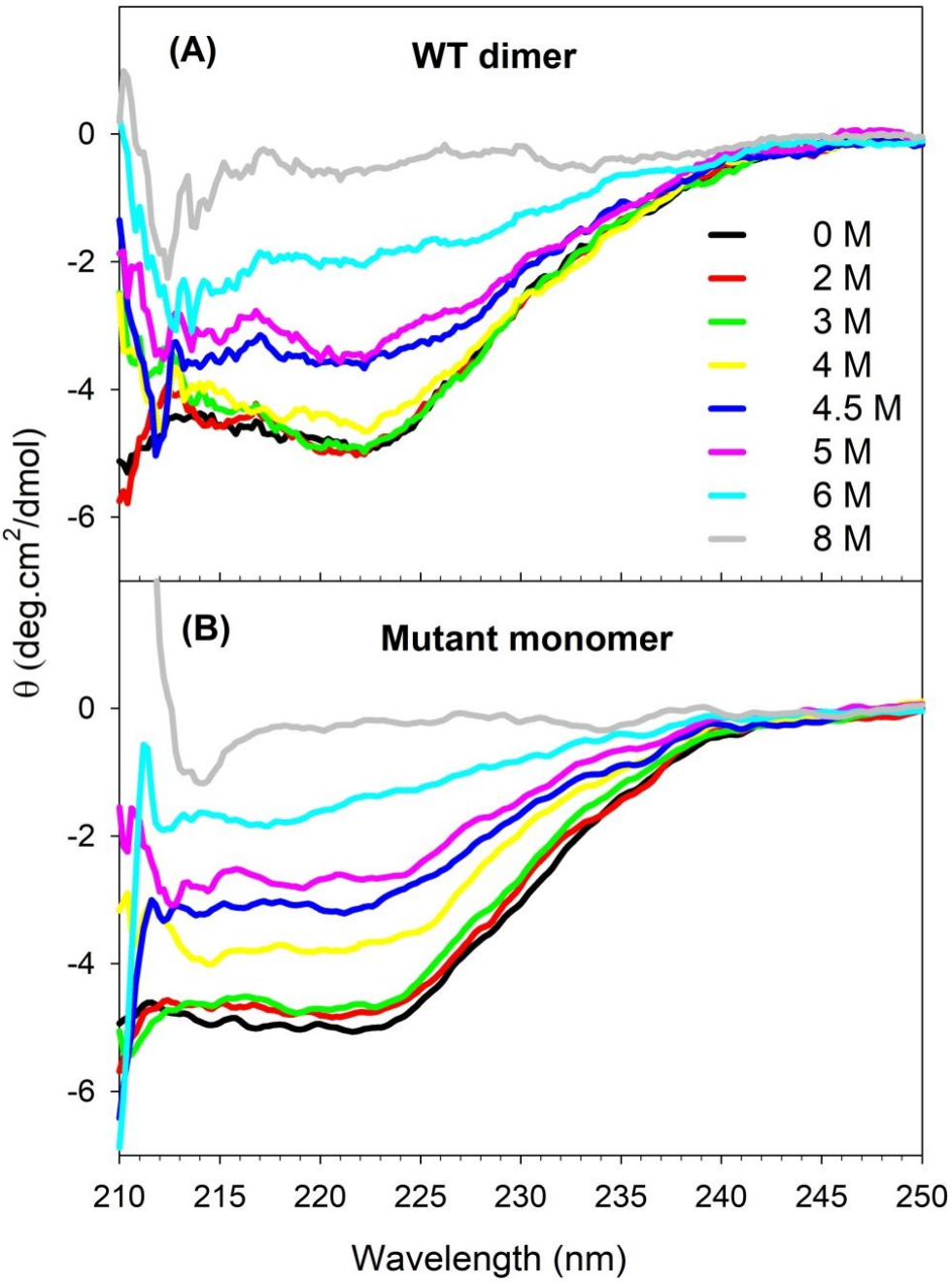


Figure 20: Secondary structure of WT dimer and mutant monomer incubated in various concentrations of urea

Far-UV CD spectra for (A) WT dimer (1 µM monomer) and (B) mutant monomer (1 µM) incubated in 0 M to 8 M urea for 1 hour at 20 °C in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. Each spectrum is the average of three samples and all spectra have been corrected for buffer contributions.

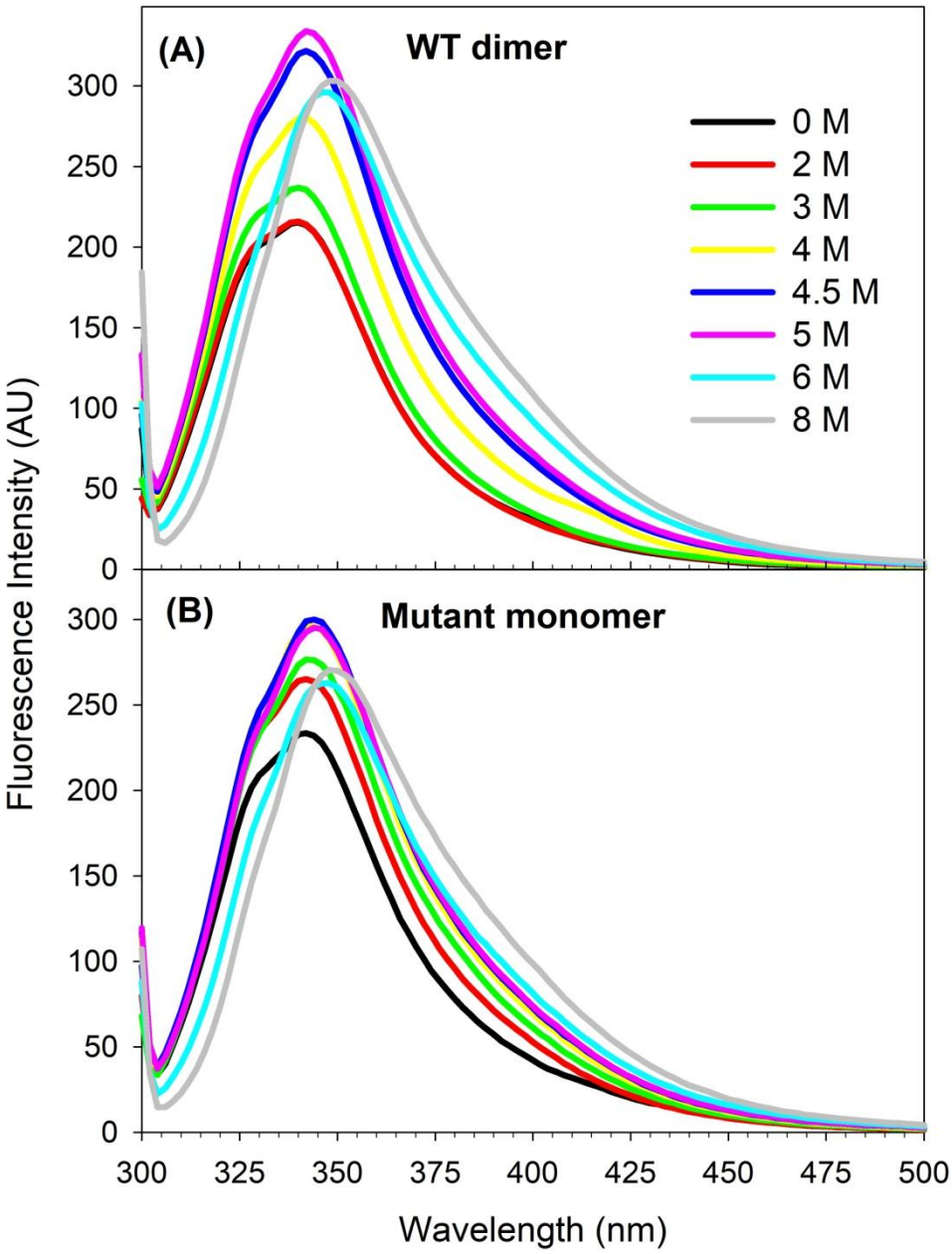


Figure 21: Tertiary structure of WT dimer and mutant monomer incubated in various concentrations of urea

Tryptophan emission fluorescence spectra of (A) WT dimer (1 µM monomer) and (B) mutant monomer (1 µM) incubated in 0 M to 8 M urea for 1 hour at 20 °C in 20 mM sodium phosphate buffer pH 7.45, containing 150 mM NaCl, 1 mM EDTA and 0.02% (w/v) NaN_3. Each spectrum is the average of three samples and all spectra have been corrected for buffer contributions. Protein was excited at 295 nm and emission was monitored between 300-500 nm.

3.5.8. Data fitting of urea-induced equilibrium unfolding

Appropriate thermodynamic models could only be selected to fit the complex multi-state unfolding curves once the intermediate states had been detected using probes such as ANS binding (section 3.5.4), SE-HPLC (section 3.5.5) and DLS (section 3.5.6). The WT dimer and mutant monomer unfolding data obtained from the intrinsic fluorescence and far-UV CD probes were globally fit using Savuka (see Section 2.6.3).

The number of chains in the oligomeric intermediate (I), formed through the non-native association of the monomeric intermediate (M), is not well determined. However, it is possible that I is dimeric, with a loosely packed configuration (Figure 17). Therefore, the WT dimer and mutant monomer data were fit to functions that incorporate parameters for the dimerisation of M to form I. WT dimer unfolding data were fit using Savuka function 190, which is a three-state model in which both the native state (N) and the M state can associate to form dimers (Figure 22). Similarly, mutant monomer unfolding data were fit to Savuka function 108, which is a three-state unfolding model with only an intermediate (M) that dimerises (Figure 23).

The thermodynamic parameters obtained from the fits are shown in Table 1. The parameters obtained from the fit were added to estimate the total $\Delta G(H_2O)_{total}$ and the total m_{total}-value. The $\Delta G(H_2O)_{total}$ is 13.5±0.9 kcal.mol^{-1} for WT dimer unfolding and 11.3±0.7 kcal.mol^{-1} for mutant monomer unfolding. The m_{total}-values are 2.6±0.2 kcal.mol^{-1}.M^{-1} for WT dimer unfolding and 2.1±0.2 kcal.mol^{-1}.M^{-1} for mutant monomer unfolding. Additionally, the C_{m1} for WT dimer is 0.8 M urea greater than the C_{m1} for the mutant monomer, whereas the C_{m2} values are the same (Table 1).

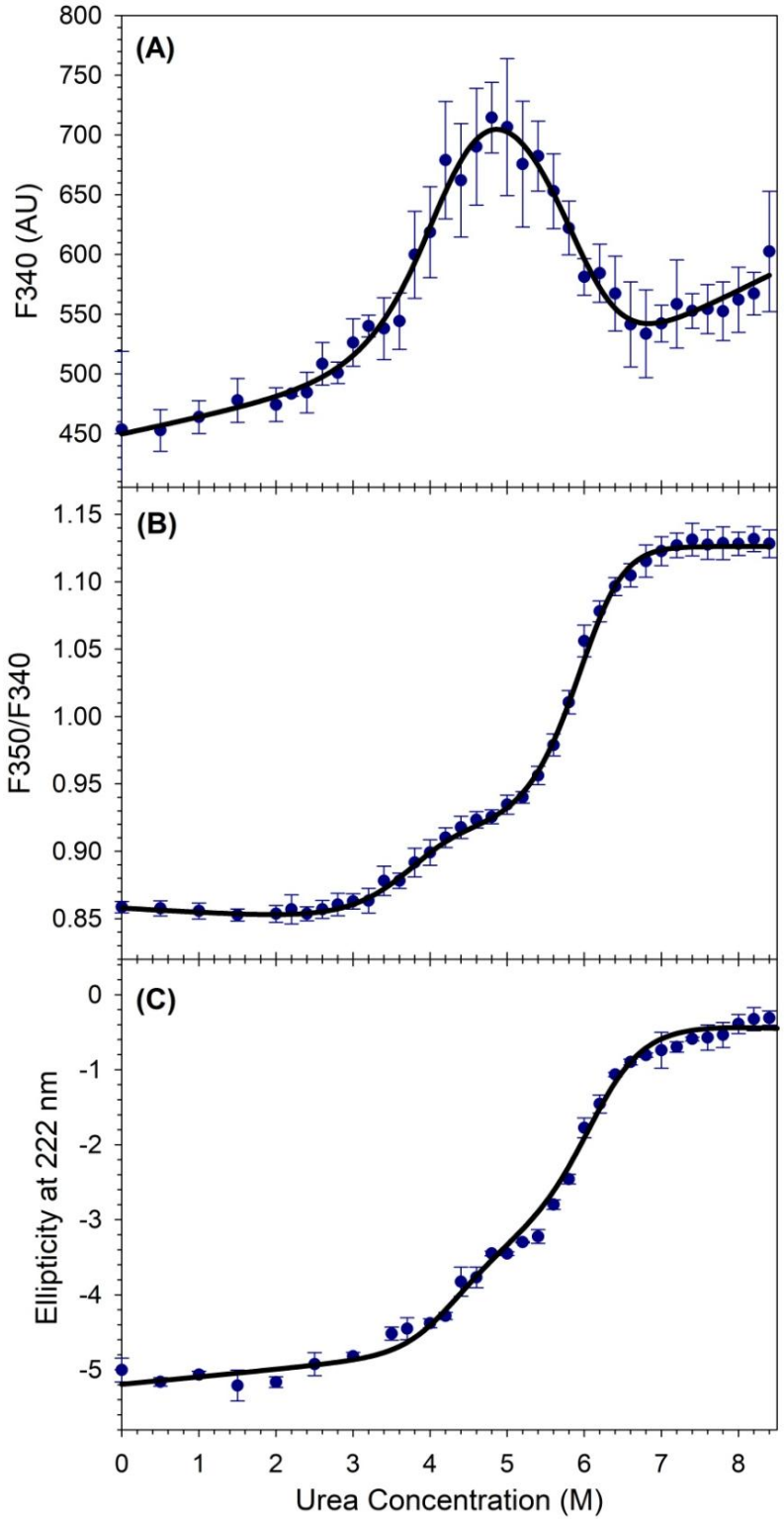


Figure 22: Global fitting of the WT dimer unfolding events

Solid lines (—) show the fit of WT dimer unfolding curves (●) monitored by (A) fluorescence using an excitation wavelength of 280 nm (B) tryptophan fluorescence (F350/F340) using an excitation wavelength of 295 nm and (C) far-UV CD (ellipticity at 222 nm). Global fitting was performed on Savuka v. 1 (developed by Professor Osman Bilsel http://www.osmanbilsel.net/software/savuka). The fits were obtained using data obtained from the far-UV CD and intrinsic tryptophan fluorescence probes.

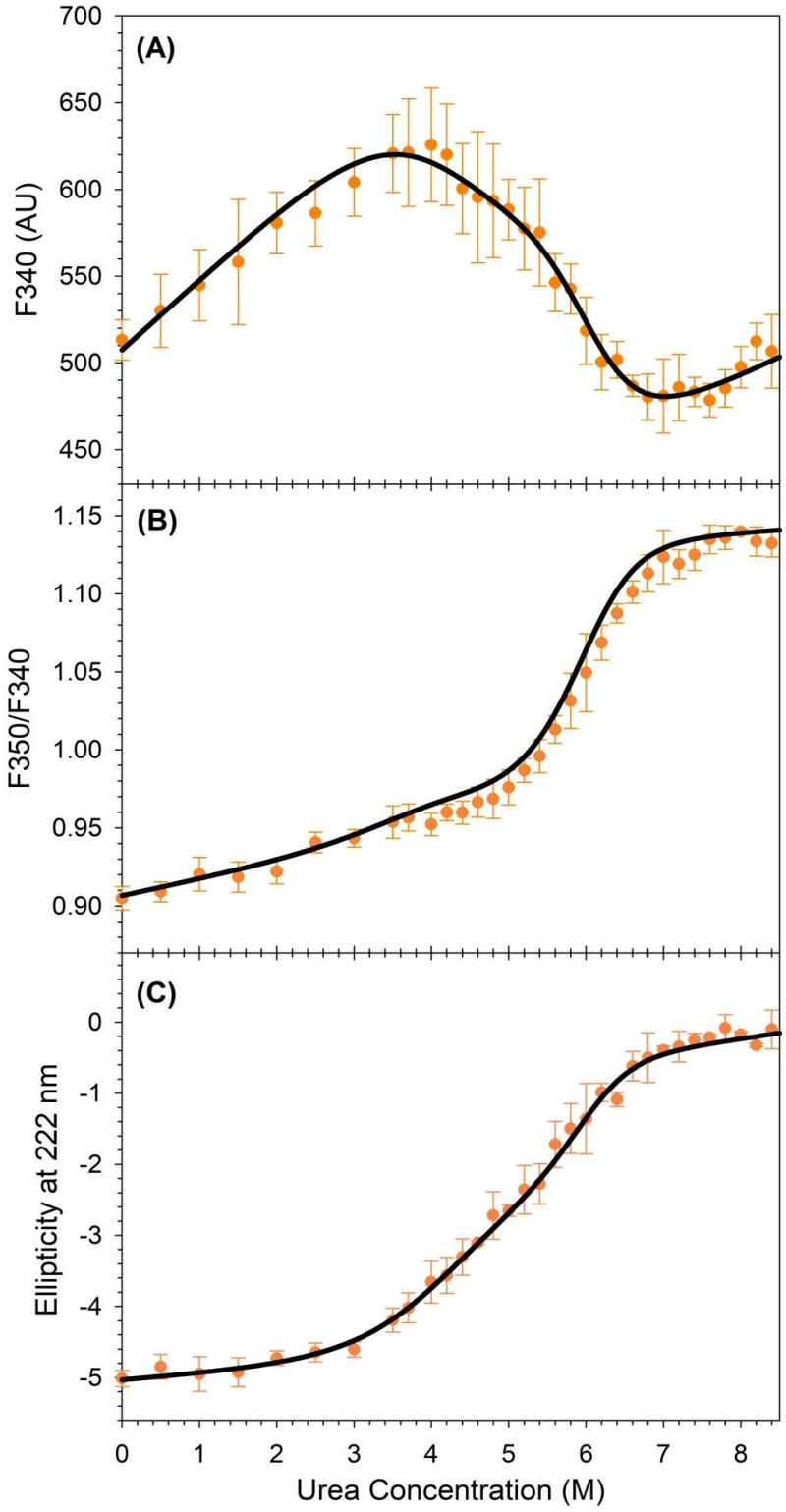


Figure 23: Global fitting of the mutant monomer unfolding events

Solid lines (—) show the fit of mutant monomer unfolding curves (▲) monitored by (A) fluorescence using an excitation wavelength of 280 nm (B) tryptophan fluorescence (F350/F340) using an excitation wavelength of 295 nm and (C) far-UV CD (ellipticity at 222 nm). Global fitting was performed on Savuka v. 1 (developed by Professor Osman Bilsel http://www.osmanbilsel.net/software/savuka). The fits were obtained using data obtained from the far-UV CD and intrinsic tryptophan fluorescence probes.

Table 1: Thermodynamic parameters obtained from urea-induced equilibrium unfolding of WT dimer and mutant monomer

	$\Delta G(H_2O)_1$ (kcal.mol^{-1})	$\Delta G(H_2O)_2$ (kcal.mol^{-1})	m_1-value (kcal.mol^{-1}.M^{-1})	m_2-value (kcal.mol^{-1}.M^{-1})	C_{m1} (M)	C_{m2} (M)
WT dimer	5.4 ± 0.6	8.1 ± 0.3	1.2 ± 0.1	1.4 ± 0.1	3.8	6.0
Mutant monomer	3.9 ± 0.4	7.4 ± 0.3	0.8 ± 0.1	1.3 ± 0.1	3.0	6.0

3.6. Pulsed-labelling HDX-MS

Pulsed-labelling HDX-MS is a powerful technique used in determining the stability of various regions within a protein. The technique relies on the fact that the deuterium levels observed in each peptide show unfolding events that occur in the backbone of the intact protein. Specific structural events, such as intermediate formation, can be detected using this technique because unfolded regions are expected to be completely deuterated whereas folded regions will be undeuterated or minimally deuterated (Deng and Smith, 1998, 1999). Two stable intermediates were detected during equilibrium unfolding using numerous probes (see section 3.5). In order to identify the native folded regions from regions that are denatured and less structured, pulse-labelling HDX-MS was performed on WT dimer and mutant monomer under conditions that minimise the loss of deuterium (Deng *et al.*, 1999; Krishna *et al.*, 2004; Pan and Smith, 2004), as described in section 2.7.

Labelled intact protein was digested with pepsin and over 200 unique peptides were identified for both the WT dimer and mutant monomer. The peptides were analysed and the mass spectra of 30 common peptides obtained for both proteins, were used for further analysis. The 30 common peptides cover 98.6 % of the sequence and were of good quality (Appendix Figure A and Figure B). Each peptide obtained after incubation in various concentrations of urea was compared with an undeuterated reference (UD), a deuterated native folded reference (0 M urea), an unfolded reference (8 M urea), and a fully deuterated reference (FD). Comparison of each peptide at each urea concentration with the reference samples, showed the extent of denaturation of each peptide at each condition.

Mass spectra for each peptide at each condition were used to calculate the deuterium levels and to assess the intermolecular distributions of the various species present within each peptide throughout unfolding. Intermolecular distributions and deuterium levels were obtained using the mass spectra obtained for each WT dimer and mutant monomer peptide. If the rate of D_2O exchange with the solvent is faster than the rate of refolding, the exchange occurs in the EX1 limit. Therefore, the addition of a denaturant such as urea will shift the reaction to the EX1 regime (Fang *et al.*, 2011; Krishna *et al.*, 2004; Liang *et al.*, 2004; Miranker *et al.*, 1996; Weis *et al.*, 2006). Unimodal mass spectra could indicate regions that are fully folded or fully denatured depending on whether the isotope pattern is similar to that of the folded reference (0 M urea) or similar to that of the denatured reference (8 M urea). Unimodal and bimodal mass spectra were observed across the 30 common peptides obtained from WT dimer and mutant monomer denatured in various concentrations of urea.

The electrospray ionisation mass spectra of four representative peptides are shown to highlight differences in the modality of the mass spectra produced, when the WT dimer and the mutant monomer are exposed to various concentrations of urea. Peptide 10: 54-79 (Figure 24A), peptide 14: 82-115 (Figure 52A) and peptide 18: 116-139 (Figure 26A) are all located at the domain interface and these three peptides all show broadly similar mass spectra envelopes signatures (Figure 24B-26B).

Peptide 10: 54-79 (Figure 24) is located in domain 1, and this peptide contains the F56 residue that was mutated to serine in the mutant monomer. The peptide forms β3 and β4 and part of α3 secondary structures (Figure 24A). Figure 24B shows the mass spectra of the WT dimer and mutant monomer at varying urea concentrations. The native (0 M) samples for both WT dimer and mutant monomer show an increased mass compared with the respective UD samples. At 4 M urea the WT dimer is still native, as shown by the unimodal peak corresponding to the native reference. At 4 M the mutant monomer shows a broadened unimodal peak. At 5 M and 6 M urea the mutant monomer shows a unimodal envelope that does not correspond with the 8 M denatured reference. The species that are shown in the monomeric unimodal mass spectra at 5 M and 6 M urea are more deuterated than the 8 M reference sample. The more complex WT dimer shows bimodal mass spectra at 5 M and 6 M urea. The peak intensity for the heavy envelope increases at higher urea concentrations, but does not correspond with the unimodal WT dimer peak at 8 M urea. The WT dimer and mutant monomer are not fully deuterated at 8 M because the peaks do not correspond with the FD reference peak.

Peptide 14: 82-115 (Figure 25) is located at the dimer interface, and this peptide contains the R81 residue that was mutated to alanine in the mutant monomer. The peptide forms part of α3 in domain 1 and the entirety of α4 in domain 2 (Figure 25A). The WT dimer and mutant monomer mass spectra of peptide 14 (Figure 25B) follow the same trend as the mass spectra of peptide 10 (Figure 24B). The WT dimer displays a unimodal peak at 0 M and 4 M, which correspond to the native protein envelope. At 5 M and 6 M urea the WT dimer displays bimodal isotope peaks with the lighter mass envelope corresponding to the native (0 M) reference and the heavy mass envelope representing an intermediate state (Figure 24B and 25B). The mutant monomer displays unimodal isotope peaks from 4 M to 6 M urea. At 4 M urea the unimodal peak shows broadening towards the lighter mass range (Figure 24B and 25B). The 5 M and 6 M urea peaks are more deuterated than the 8 M denatured reference peak (Figure 24B and 25 B).

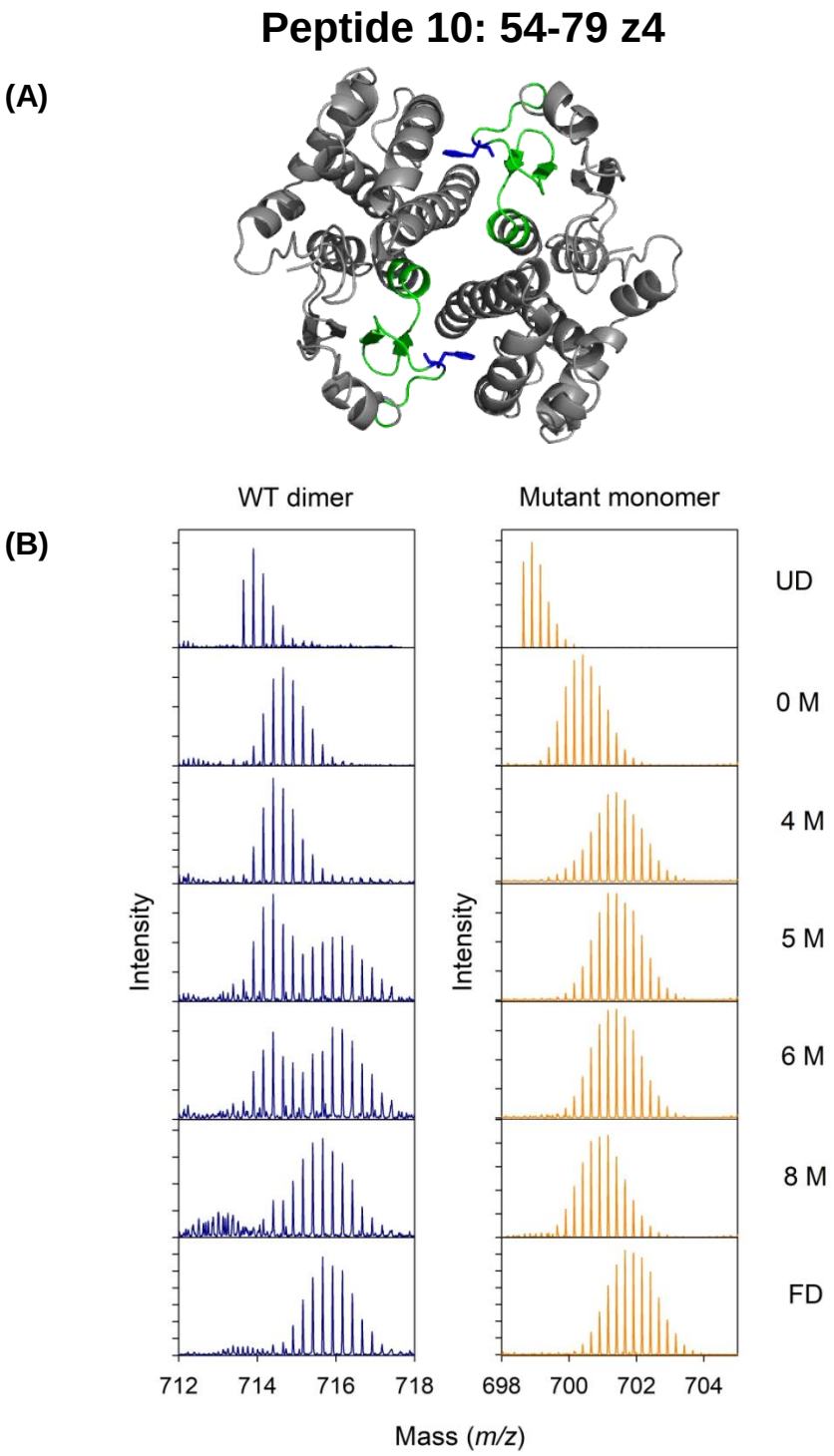


Figure 24: Structural location and mass spectra of peptic fragment number 10 (residues 54-79)

(A) Peptide 10:54-79 (green) is located in domain 1and contains the F56 residue (blue sticks). PDB: 1XW6 was used to generate the image using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (B) Mass spectra of WT dimer (─) and mutant monomer (─) at various urea concentrations indicates the unfolding of the peptide in each protein. Undeuterated (UD) indicates native protein that was not exposed to deuterium and fully deuterated (FD) indicates the mass of a peptide that is cleaved from native protein that was exposed to deuterium for 18 hours. The *m*/z x-axis differs between the WT dimer and mutant monomer because the F56S mutation is present in this peptide.

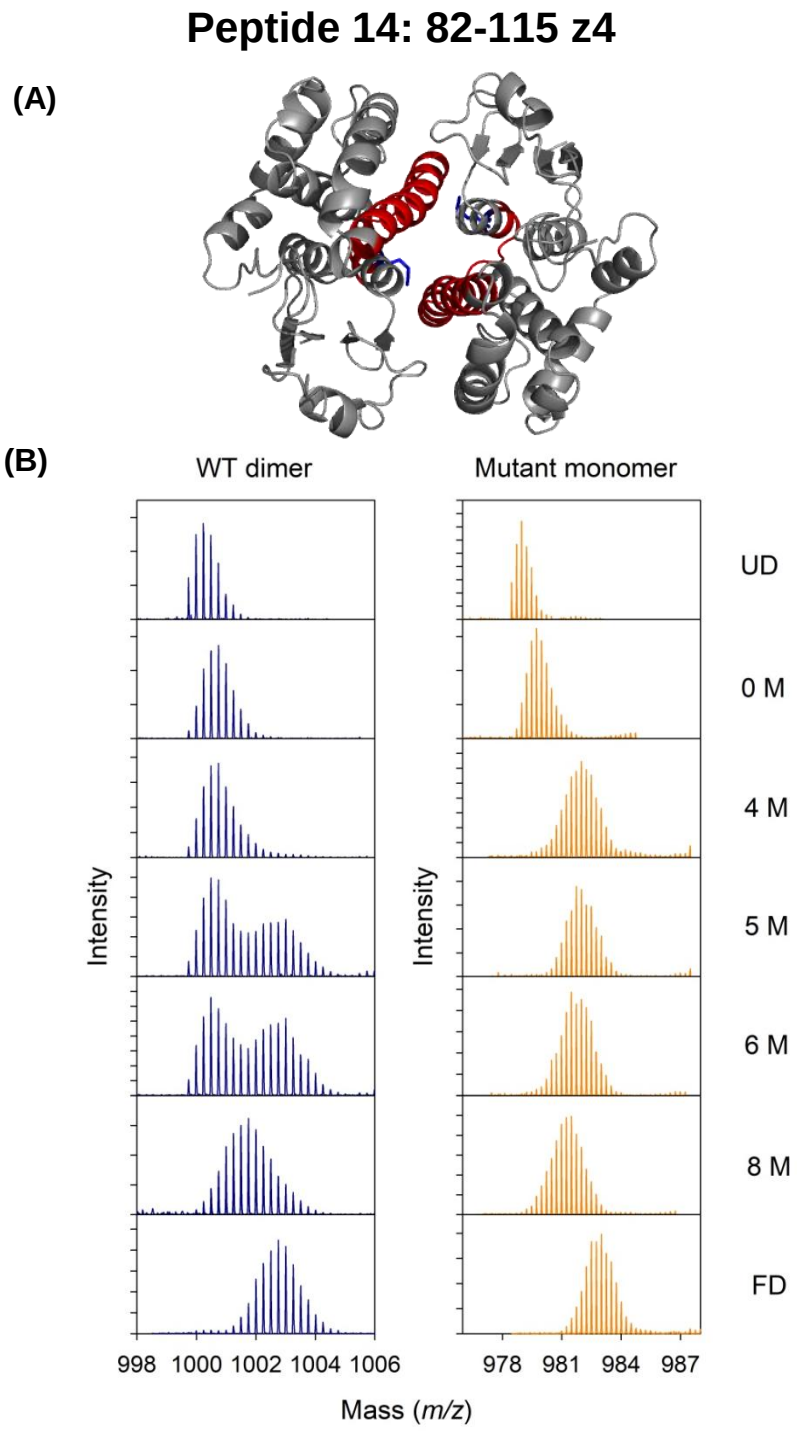


Figure 25: Structural location and mass spectra of peptic fragment number 14 (residues 82-115)

(A) Peptide 14:82-115 (red) contains the R81 residue (blue sticks). PDB: 1XW6 was used to generate the image using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (B) Mass spectra of WT dimer (─) and mutant monomer (─) at various urea concentrations indicates the unfolding of the peptide in each protein. Undeuterated (UD) indicates native protein that was not exposed to deuterium and fully deuterated (FD) indicates the mass of a peptide that is cleaved from native protein that was exposed to deuterium for 18 hours. The *m*/z x-axis differs between the WT dimer and mutant monomer because the R81A mutation is present in this peptide.

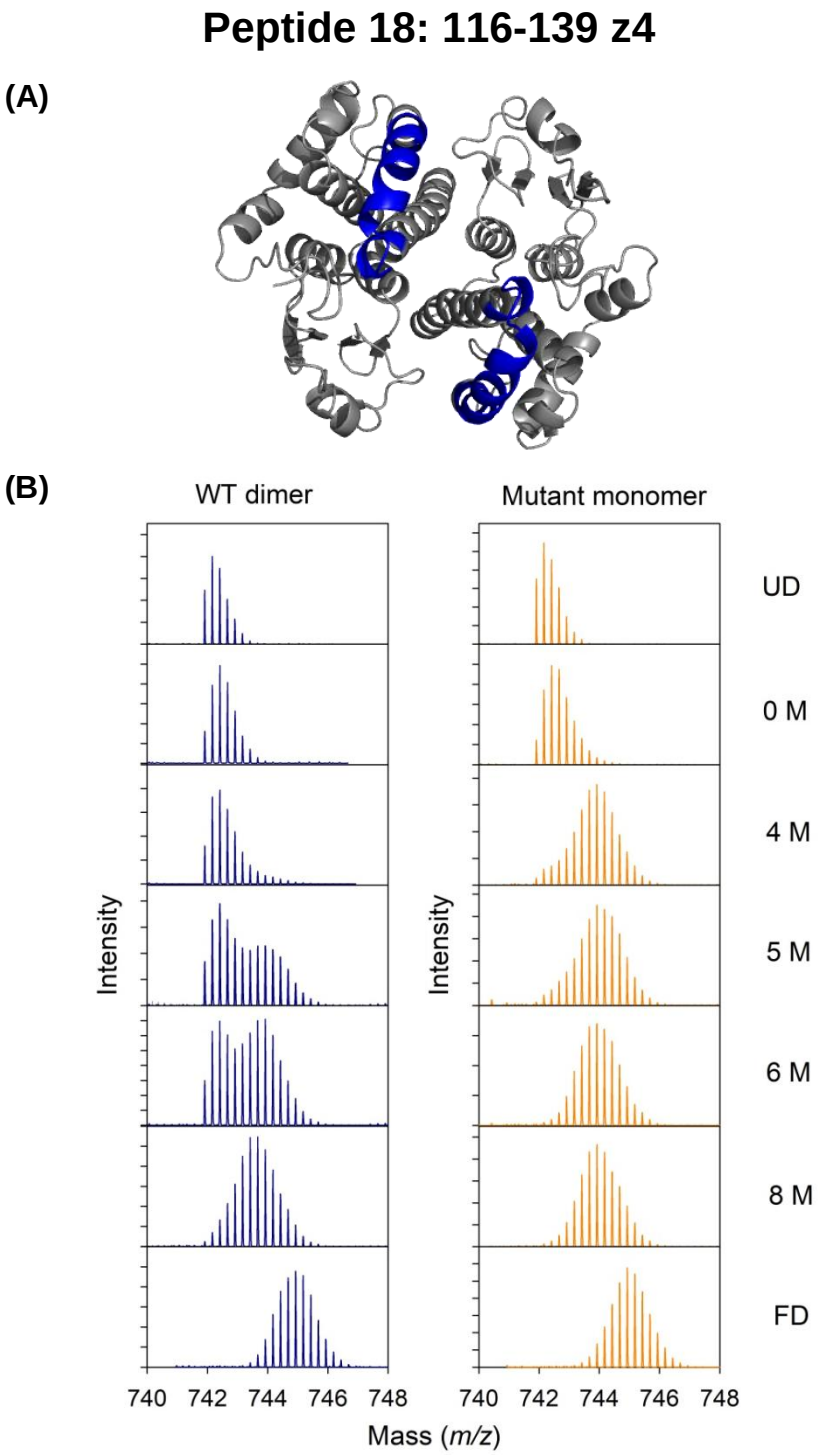


Figure 26: Structural location and mass spectra of peptic fragment number 18 (residues 116-139)

Peptide 3:22-51 (blue) is located in domain 1. PDB: 1XW6 was used to generate the image using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (B) Mass spectra of WT dimer (─) and mutant monomer (─) at various urea concentrations indicates the unfolding of the peptide in each protein. Undeuterated (UD) indicates native protein that was not exposed to deuterium and fully deuterated (FD) indicates the mass of a peptide that is cleaved from native protein that was exposed to deuterium for 18 hours.

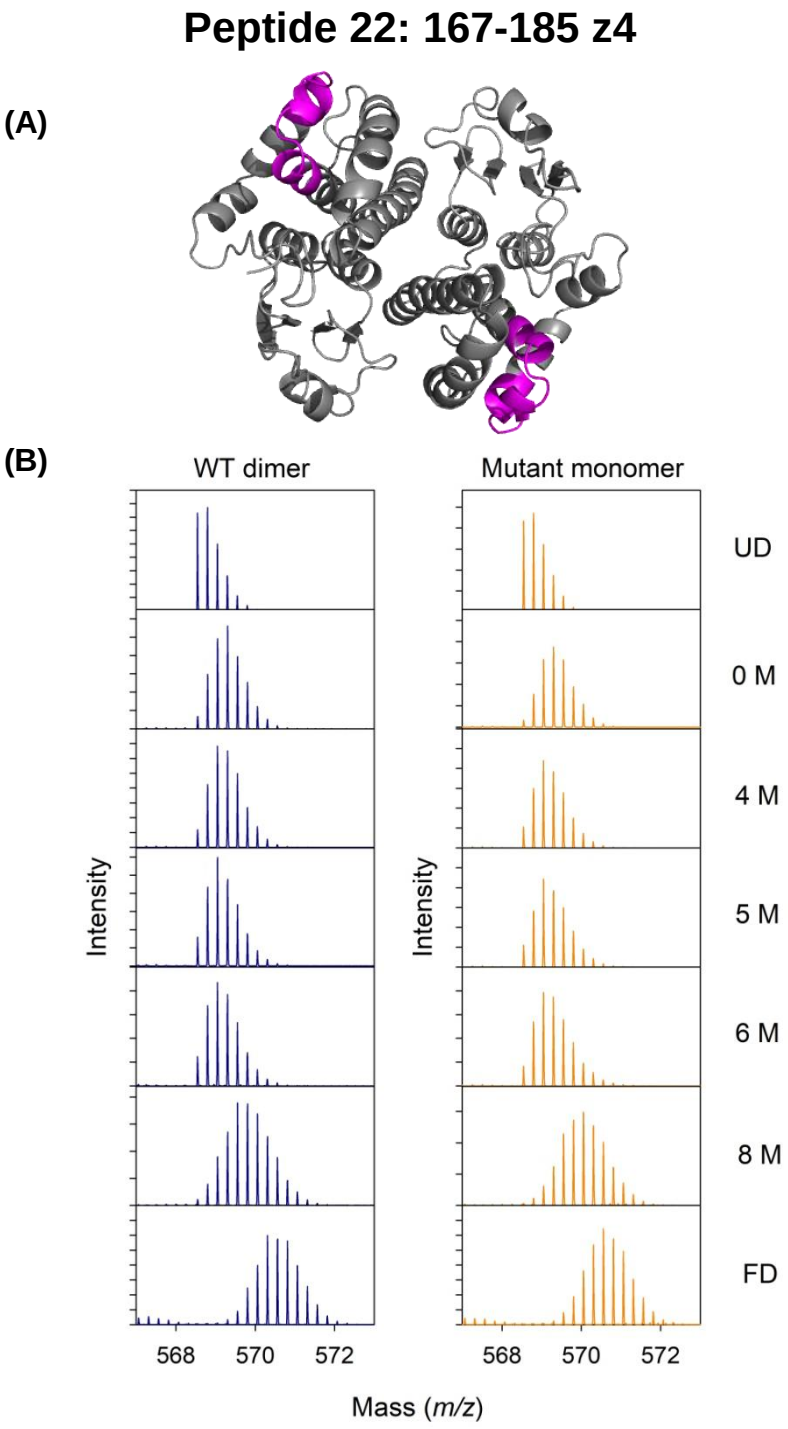


Figure 27: Structural location and mass spectra of peptic fragment number 22 (residues 167-185)

(A) Peptide 22:167-185 (magenta) is located in domain 2. PDB: 1XW6 was used to generate the image using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (B) Mass spectra of WT dimer (─) and mutant monomer (─) at various urea concentrations indicates the unfolding of the peptide in each protein. Undeuterated (UD) indicates native protein that was not exposed to deuterium and fully deuterated (FD) indicates the mass of a peptide that is cleaved from native protein that was exposed to deuterium for 18 hours.

Peptide 18: 116-139 (Figure 26) is located in domain 2, and forms part of α4 and the whole of α5 (Figure 26A). The WT dimer and mutant monomer amino acid composition are identical for this peptide. The 0 M native references for both WT dimer and mutant monomer have unimodal peaks (Figure 26B). At 4 M urea both WT dimer and mutant monomer have bimodal mass envelopes for peptide 18. However, the WT dimer shows a greater intensity at the lower mass envelope whereas the mutant monomer shows a greater intensity at the higher mass envelope**.** At 5 M urea the WT dimer and mutant monomer show an increase in the intensity of the heavy mass envelope, and the heavy mass envelope corresponds with the denatured 8 M reference sample (Figure 26B). At 6 M urea the WT dimer still shows a bimodal mass distribution but the mutant monomer only shows a unimodal mass spectrum that corresponds with the denatured 8 M sample.

Peptide 22: 167-185 (Figure 27) is located in domain 2 of the protein, and forms part of both α6 and α7 (Figure 27A). The mass spectra of this peptide differs from the three aforementioned peptides in that this peptide displays only unimodal mass spectra that correspond with the native spectra for both WT dimer and mutant monomer up until 8 M urea (Figure 27B). At 8 M urea there is a shift to a heavier mass**,** indicating that peptide 22 represents a highly, and equally, stable region in both proteins and proves that different regions within WT dimer and mutant monomer differ in stability.

The deuterium recovery for the 4 representative peptides was calculated by comparing the values of the denatured 8 M reference with the values of the deuterium levels obtained for the FD samples. The masses of WT dimer peptides 10, 14, 18 and 22 obtained from the denatured reference sample show that the peptides have an average of 6.3, 8.1, 6.6 and 4.4 deuterons, respectively. Deuterium levels of the FD reference samples for peptide 10, 12, 18 and 22 are 7.1, 9.4, 10.3 and 6.8. Therefore, the deuterium recoveries for the WT dimer peptic fragments were 89 %, 86 %, 64 % and 65 %, respectively. The denatured reference samples for mutant monomer showed an average of 7.8, 8.9, 6.7 and 4.8 deuterons for peptide 10, 14, 18 and 22. The FD reference samples for mutant monomer had an excess of 11.4, 15.1, 10.7 and 7.2 for peptides 10, 14, 18 and 22. Therefore, the deuterium recoveries for peptide 10, 14, 18 and 22 obtained from mutant monomer were 68 %, 59 %, 63 % and 67%, respectively. It has been reported that uncertainties between 10-500 ppm in mass measurements can be expected (Pan and Smith, 2004). The deuterium recovery results indicate that there is a greater percentage recovery for WT dimer than for mutant monomer, suggesting that the experiment is more reliable for WT dimer than for mutant monomer. Considering that the percentage recovery is calculated by comparing the values of the

denatured 8 M reference with the corresponding FD sample, it is more likely that the lower percentage recoveries indicate that the mutant monomer protein retains more structure in 8 M urea than the WT dimer protein, and not that the WT dimer data is more reliable.

The average molecular mass of a peptide can be determined from the centroid values of all the isotope peaks and this gives the deuterium levels found in each fragment (Pan and Smith, 2004; Yang and Smith, 1997). The deuterium levels found in each peptide were converted to the percentage deuteration. The percentage deuteration for each WT dimer and mutant monomer peptide were plotted for each urea condition (Figure 28).

Peptides 3-10 differ in the percentage deuteration between the WT dimer and mutant monomer at 0 M (Figure 28A) and at 2 M urea (Figure 28B). The native mutant monomer is highly deuterated in this region whereas the native WT dimer is not. At 5 M urea (Figure 28D) peptides 3-10 of the WT dimer become more deuterated and at 6 M (Figure 28E) and 8 M urea (Figure 28F) the WT dimer and the mutant monomer are equally deuterated in this region. Peptides 11-19 contain the R81 residue, which is mutated to alanine in the mutant monomer. This region is equally deuterated at 0 M (Figure 28A) and at 2 M urea (Figure 28B) for both proteins, indicating that both native proteins show no differences in structure in this region (amino acids 80-140). At 4 M urea the mutant monomer becomes more deuterated in this region, whereas the WT dimer remains minimally deuterated (Figure 28C). Only at 5 M urea does the WT dimer become more deuterated in this region (Figure 28D). At 6 M (Figure 28E) and 8 M (Figure 28F) urea, the WT dimer and the mutant monomer are equally deuterated in this region. Peptides 20-26 are equally, and minimally, deuterated for the WT dimer and mutant monomer from 0 M to 6 M urea (Figure 28A, 28B, 28C, 28D and 28E). However, at 8 M urea there is an increase in the deuteration for both the WT dimer and mutant monomer (Figure 28E).

The peptides were broadly grouped into three main groups based on the data in figure 28. The grouping was verified, and subsequently expanded, by determining the stability of each peptide within each group. The stability of each peptide was determined by an uptake plot for each corresponding WT dimer and mutant monomer peptide. An uptake plot is a graphical summary of the deuterium incorporation by each peptide during the unfolding of protein (Figure 29). The local unfolding midpoint (C_m), determined from the peptide uptake plot, is an indication of the stability of the protein region in which the peptide was originally located (Table 2) (Miranker *et al.*, 1996; Pan *et al.*, 2004; Pan and Smith, 2004; Rojsajjakul *et al.*, 2004).

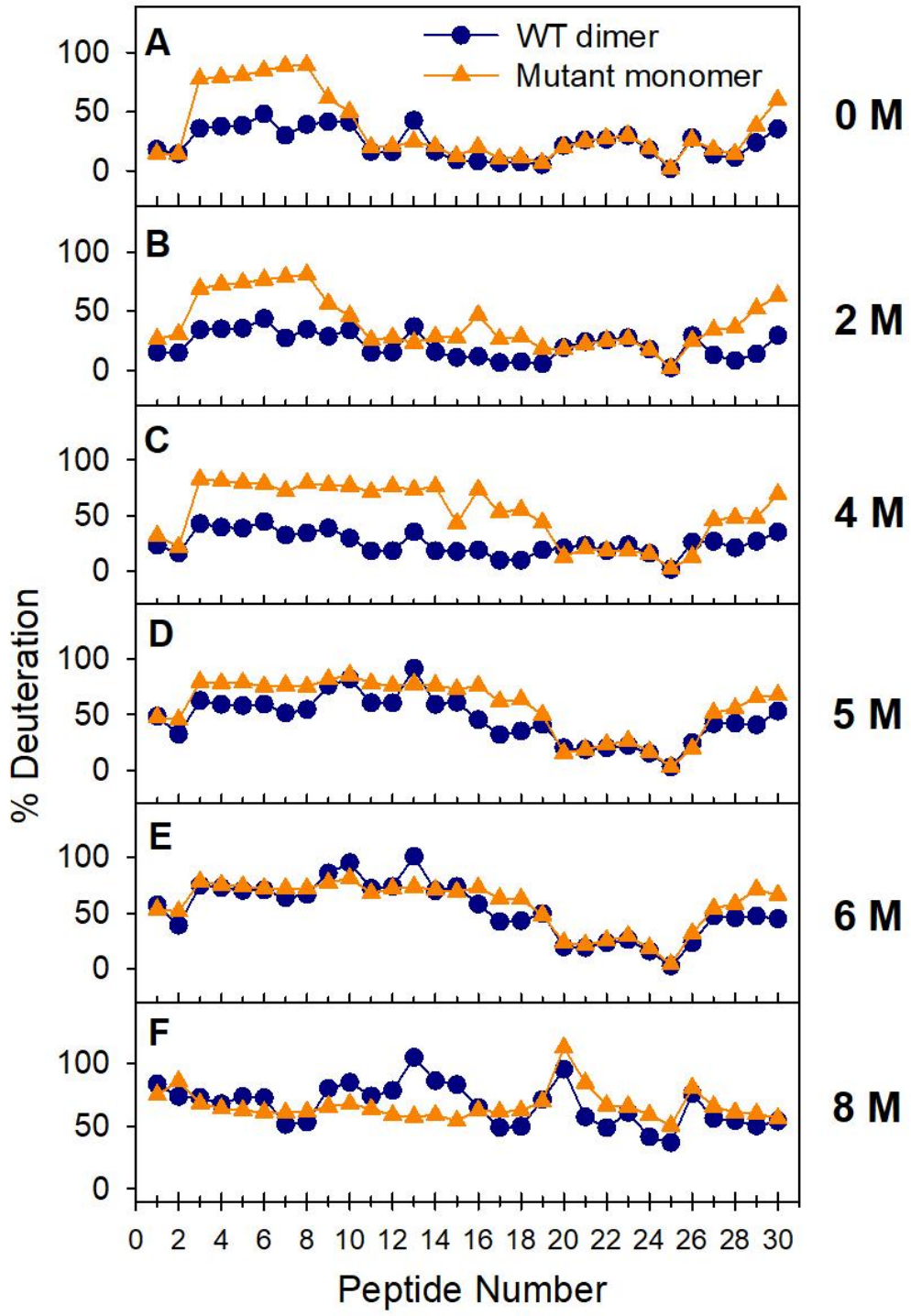
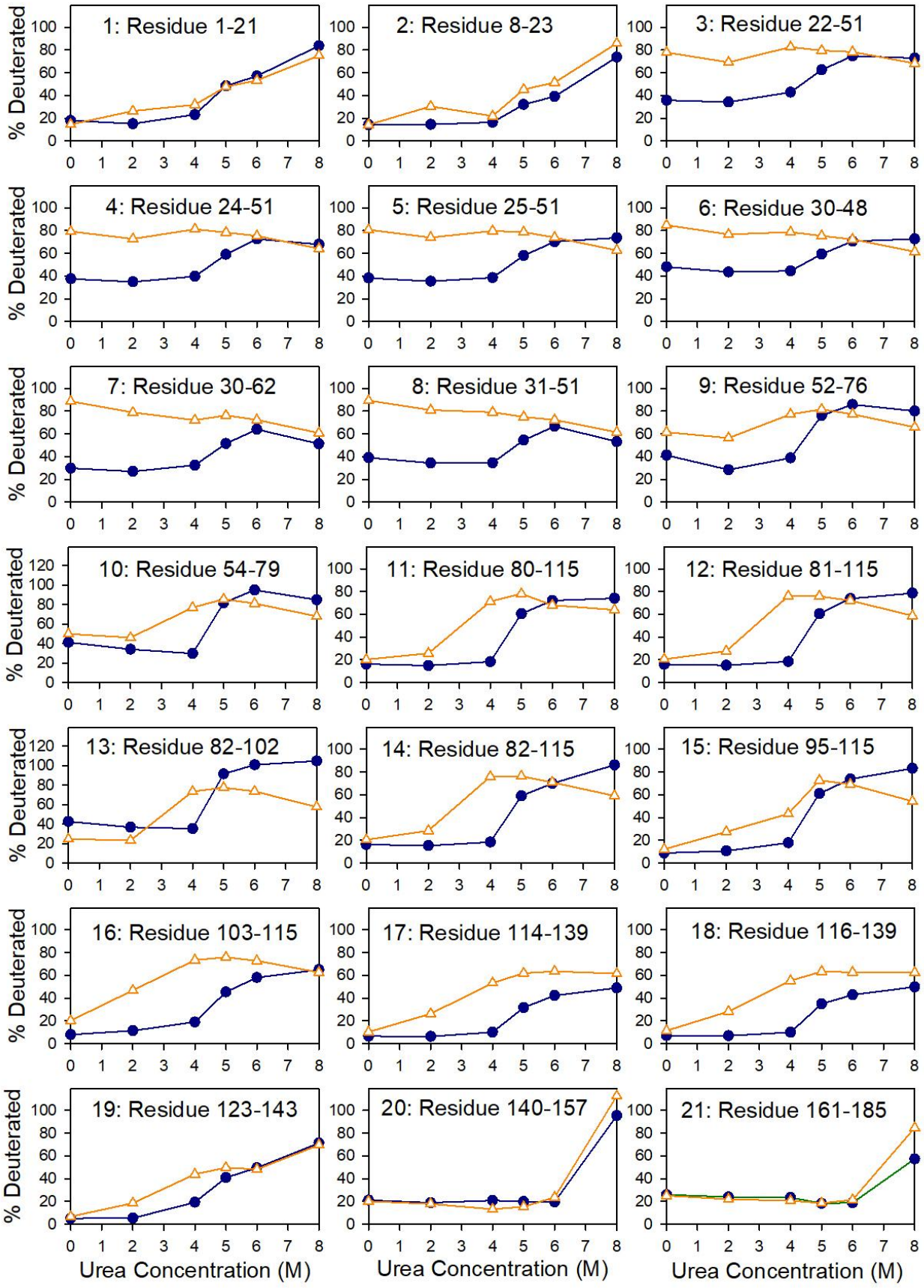


Figure 28: The percentage deuteration for each peptide at varying urea concentrations for WT dimer and mutant monomer

WT dimer (●) and mutant monomer (▲) were exposed to (A) 0 M urea (B) 2 M urea (C) 4 M urea (D) 5 M urea (E) 6 M urea and (F) 8 M urea. Each peptide was cleaved from intact protein that had been allowed to unfold for 1 hour. The protein was then pulse-labelled with urea/D_2O for 10 seconds prior to acid quenching and pepsin digestion


1: Residue 1-21
2: Residue 8-23
3: Residue 22-51
4: Residue 24-51
5: Residue 25-51
6: Residue 30-48
7: Residue 30-62
8: Residue 31-51
9: Residue 52-76
10: Residue 54-79
11: Residue 80-115
12: Residue 81-115
13: Residue 82-102
14: Residue 82-115
15: Residue 95-115
16: Residue 103-115
17: Residue 114-139
18: Residue 116-139
19: Residue 123-143
20: Residue 140-157
21: Residue 161-185
% Deuterated
Urea Concentration (M)

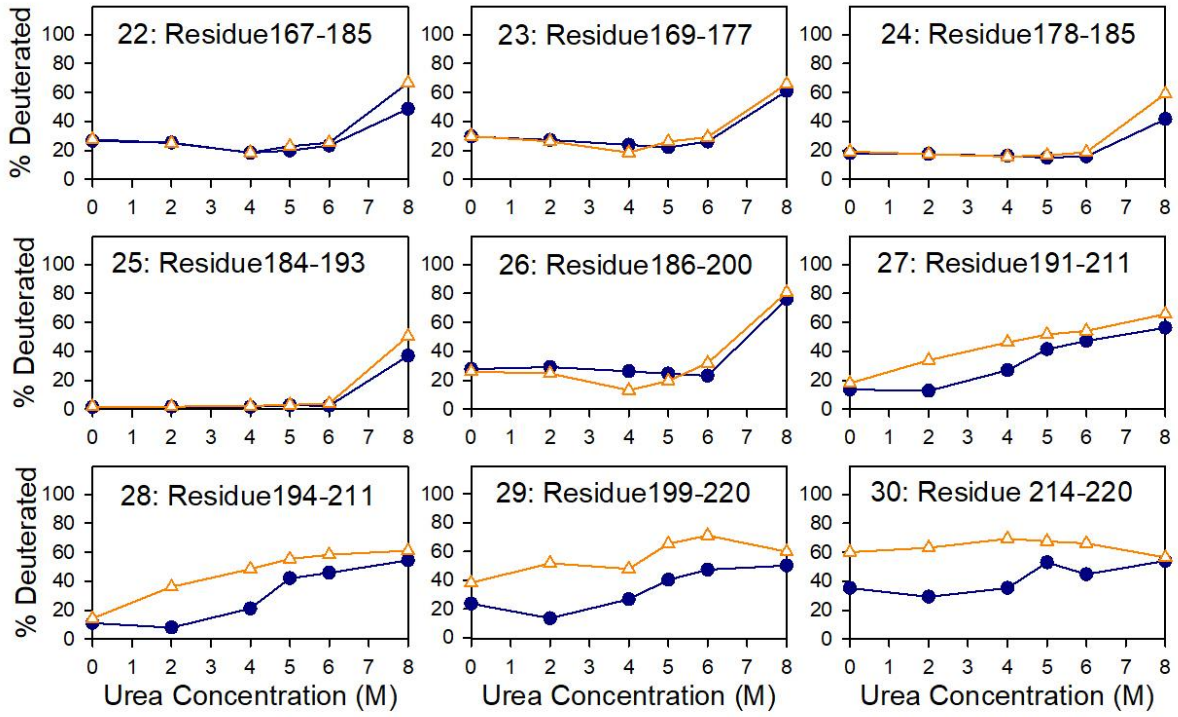


Figure 29: Deuterium incorporation for the 30 common peptides obtained under urea-induced equilibrium unfolding conditions

The percentage deuterium uptake (% deuterated) when WT dimer (●) and mutant monomer (△) are unfolded in 0 M to 8 M urea was calculated using the maximum exchangeable amide protons for each peptide. Each peptide was cleaved from intact protein that had been allowed to unfold for 1 hour. The protein was then pulse-labelled with urea/D_2O for 10 seconds prior to acid quenching and pepsin digestion

Table 2: The Cm values for the 30 common peptides obtained for WT dimer and mutant monomer unfolding. Unfolded protein samples were pulse-labelled with urea/D_2O for 10 seconds prior to acid quenching and pepsin digestion

Least stable Key group[1]			**Medium stable Charged cluster group[2]**			**Highly stable Core group[3]**		
Peptide	WT dimer C_m (M)	Mutant monomer C_m (M)	Peptide	WT dimer C_m (M)	Mutant monomer C_m (M)	Peptide	WT dimer C_m (M)	Mutant monomer C_m (M)
3: 22-51	4.9	0	1: 1-21	5.8	4.9	20: 140-157	7.3	7.2
4: 24-51	4.8	0	2: 8-23	5.7	5.0	21: 161-185	7.3	7.1
5: 25-51	5.0	0	11: 80-115	5.5	3.4	22: 167-185	7.4	7.2
6: 30-48	5.1	0	12: 81-115	5.5	3.3	23: 169-177	7.3	7.2
7: 30-62	5.0	0	13: 82-102	5.6	3.2	24: 178-185	7.2	7.1
8: 31-51	5.1	0	14: 82-115	5.8	3.2	25: 184-193	7.4	7.2
9: 52-76	4.8	0	15: 95-115	5.7	3.0	26: 186-200	7.4	7.0
10: 54-79	4.9	0	16: 103-115	5.6	2.6			
27: 191-211	4.7	0	17: 114-139	5.8	2.8			
28: 194-211	4.8	0	18: 116-139	5.9	3.0			
29: 199-220	4.7	0	19: 123-143	5.7	3.1			
30: 214-220	4.9	0						

[1]: The group is called the key group because it contains the F56 key residue in the WT dimer, which forms part of the conserved lock-and-key mechanism (see section 1.4). F56S/R81A is mutated to S56 in the mutant monomer to form the F56S mutation.
[2]: The group is called the charged cluster group because it contains the conserved R81 residue in the WT dimer, which forms part of the mixed charged cluster (see section 1.4). R81 is mutated to A81 in the mutant monomer to form the R81A mutation.
[3]: The core group consists of the peptides that form the highly stable core of the protein.

The peptides representing the three unfolding groups are plotted onto the structure of hGST M1a-1a (PDB code 1XW6) (Figure 30A). The least stable key group (orange) contains peptides 3-10 and 27-30. This group contains the F56 key residue in the WT dimer and the S56 residue in the mutant monomer (F56S mutation). The C_m values for WT dimer range from 4.7 M to 5.1 M urea (Table 2). The mutant monomer C_m values could not be determined because these peptides were maximally deuterated throughout the experiment (Table 2). The medium stable charged cluster group is shown in blue on figure 30A. This group contains peptides 1, 2 and 11-19 (Table 2). The conserved R81 residue is located within the WT dimer peptides of this group (R81A in mutant monomer). The C_m values range from 5.5 M to 5.9 M urea for WT dimer and from 2.6 M to 5.0 M urea for the less stable mutant monomer (Table 2). The highly stable core group consists of peptides 20-26 and is shown in green on figure 30A. The C_m values of this group range from 7.1 M to 7.4 M for both the WT dimer and mutant monomer (Table 2).

The population of denatured protein can be obtained if the bimodal peptides are compared with the low mass UD envelope and the high mass FD envelope. The envelopes were not always clearly resolved (monomer data at 4 M in Figure 26B), so all spectra could not be fitted to Gaussian distributions to obtain the areas under each envelope. Therefore, calculated deuterium levels were used to determine the populations of denatured molecules in each peptide at varying concentrations of urea, in accordance with literature (Pan *et al.*, 2004; Pan and Smith, 2004; Rojsajjakul *et al.*, 2004; Yang and Smith, 1997). The population of denatured protein in each unfolding stability group was calculated using equation 7 (section 2.7.1) and plotted as a function of urea concentration (Figure 27B and 27C). The least stable key group (peptides 3-10 and 27-30) (Figure 30A) shows a transition midpoint at approximately 5 M urea for the WT dimer (Figure 30B), which is in accordance with the C_m values obtained for this group (Table 2). In contrast, the corresponding least stable key group transition from folded to unfolded mutant monomer is non-cooperative, and a transition midpoint cannot be determined (Figure 30C). The medium stable charged cluster group (peptides 1,2 and 11-19) show a transition midpoint of 5.5 M urea for the WT dimer and 3.5 M for the mutant monomer (Figure 30), in accordance with the C_m values for this group (Table 2). The most stable core group (peptides 20-26) unfolds in the same manner for both WT dimer (Figure 30B) and mutant monomer (Figure 30C) and both proteins have unfolding midpoints at approximately 7 M urea and an average C_m of 7.3 M and 7.1 M, respectively.

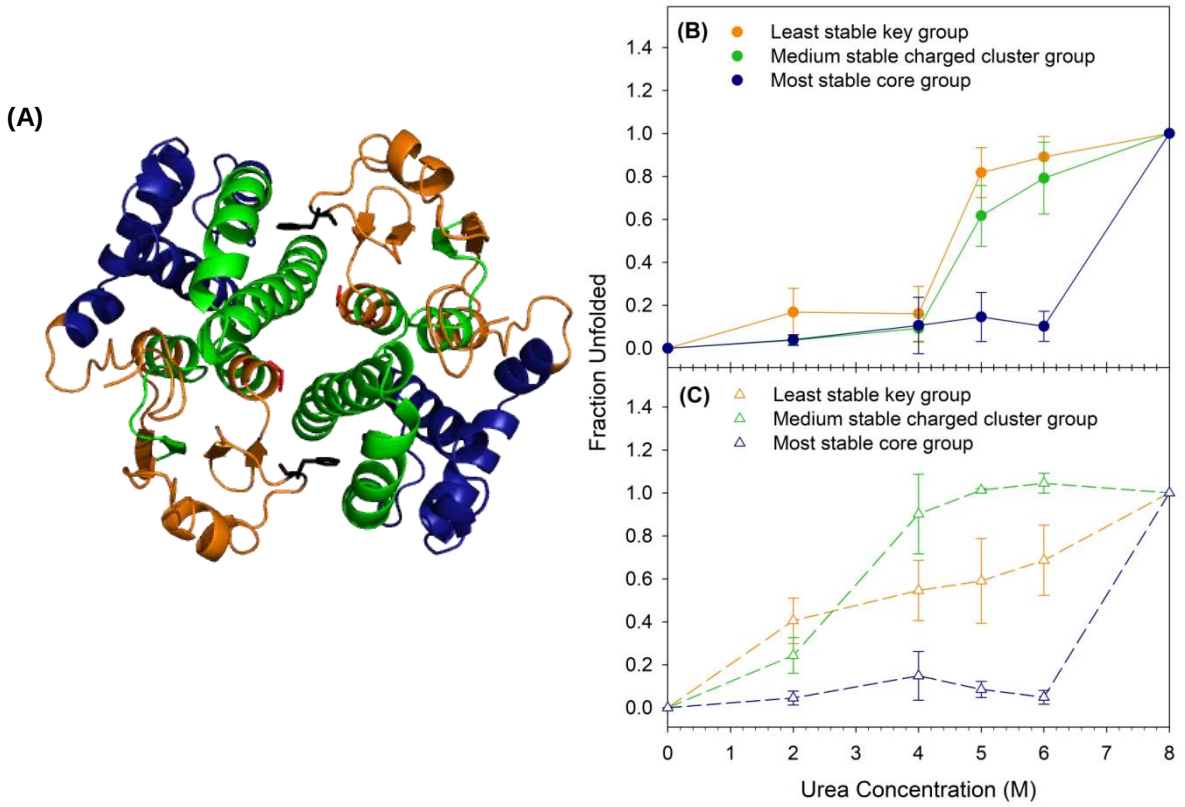


Figure 30: Population of unfolded WT dimer and mutant monomer in the three unfolding groups

(A) Unfolding groups are highlighted on the WT dimer structure. The least stable key group (peptides 3-10 and 27-30) is shown in orange, the medium stable charged cluster group (peptides 1,2 and 11-19) is shown in blue and the most stable core group (peptides 20-26) is shown in green. F56 is shown in black and R81 is shown in red. PDB: 1XW6 was used to generate the image using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (B) The fraction unfolded for WT dimer showing the least stable key group (peptides 3-10 and 27-30) (--●--), the medium stable charged cluster group (peptides 1,2 and 11-19) (--●--) and the most stable core group (peptides 20-26) (--●--) (C) The fraction unfolded for mutant monomer showing the least stable key group (peptides 3-10 and 27-30) (--Δ--), the medium stable charged cluster group (peptides 1,2 and 11-19) (--Δ--) and the most stable core group (peptides 20-26) (--Δ--). Data points are the average of all the peptides contained within the specific unfolding group.

Chapter 4: Discussion

In-depth analyses have been conducted on the (un)folding pathways of numerous monomeric proteins, due to their relative structural simplicity. The fundamental principles learnt from this research can now be extended to larger and more complex multi-domain and oligomeric proteins. However, obtaining the (un)folding pathway for a multi-domain protein is challenging because of the numerous domain- and subunit-interface interactions, and the autonomous folding of individual domains.

The dimeric structure of GST proteins is stabilised by multiple intersubunit interactions. The most prominent of these interactions are the conserved lock-and-key motif and the mixed charge cluster at the dimer interface (see section 1.4.3). The bulky F56 residue acts as the stabilising key residue in class Mu GSTs (Figure 2), but substitution of this residue alone is not sufficient to disrupt the dimeric structure (Hegazy *et al.*, 2006; Hornby *et al.*, 2002; Sayed *et al.*, 2000). The simultaneous disruption of both the F56 key residue and the R81 residue in the mixed charge cluster is required for the formation of a monomeric GST, as shown in this study (Figure 8), and in a study conducted on rGST M1-1 (Thompson *et al.*, 2006).

The F56S/R81A monomeric variant of hGST M1a-1a was produced so that the complete unfolding pathway of hGST M1a-1a could be assessed at the subunit level, in the absence of the processes involved in dimer dissociation. The comparative study of the complex WT dimer and the less complex mutant monomer allowed for a more in-depth understanding of the unfolding pathway of hGST M1a-1a.

4.1. The mutant monomer is a suitable model of the WT dimer subunit

The success of this study relied on the fact that the mutant monomer is structurally comparable to the subunit in the WT dimer, and that the mutations have not resulted in any major tertiary structural alterations (see section 1.4.5). The 25 kDa mass obtained for the native mutant monomer using SE-HPLC (Figure 8B) corresponds with the 26 kDa molecular mass reported for the folded GST class Mu subunit (Armstrong, 1997). Additionally, the mass of the mutant monomer is similar to the masses obtained for the F56S/R81A monomeric mutant of rGST M1-1 (26 kDa) (Thompson *et al.*, 2006), and other monomeric variants of GSTs (Abdalla *et al.*, 2002; Huang *et al.*, 2008).

The insertion of the F56 functional group into the hydrophobic cavity constructed by α4 and α5 on the opposing subunit anchors the region surrounding the mu loop in domain 1 (Figure 2), thereby stabilising the active site of hGST M1a-1a (Alves *et al.*, 2006; Sayed *et al.*, 2000). The F56S mutation increases the dynamics of the mu loop (Figure 28A and Figure 31B), which disrupts Y6 in the active site, a residue that is essential for efficient catalysis (Thompson *et al.*, 2006). The stabilising quaternary structure of the dimer is also required for the formation of a complete active site in each subunit, because of the proximity of the subunit interface to the active site. If the subunit interface is disrupted, Q71, S72 and D105 cannot form hydrogen bonds with GSH from the opposite subunit (Ji *et al.*, 1992; Wilce and Parker, 1994), and the protein cannot catalyse the nucleophilic attack on non-polar xenobiotic compounds (see section 1.4.2). Therefore, the loss in specific activity of the mutant monomer (Figure 11) is due to the destabilisation of the active site, in accordance with studies conducted on other GST subunit interface variants (Alves *et al.*, 2006; Hornby *et al.*, 2002; Sayed *et al.*, 2000; Thompson *et al.*, 2006).

The loss of quaternary structure in a protein may result in structure modifications, such as a reduction in the α-helical content of a protein (Hennessey *et al.*, 1982; Woody, 1995). However, spectroscopic studies reveal that there are no major differences in the secondary (Figure 9) and tertiary (Figure 10) structures of the WT dimer and the mutant monomer. In fact, pulsed-labelling HDX-MS data show that the subunit of the native WT dimer (Figure 31A) is conformationally comparable to the native structure of the mutant monomer (Figure 31B). The difference in intensity of the fluorescence spectra obtained for WT dimer and mutant monomer could be a result of an increase in the dynamics of domain 1. The W7 and W45 residues in domain 1 are located at the active site (nearby the mu loop), whereas W146 and W214 remain more buried in the hydrophobic core of domain 2 (Figure 2A and 2B). The increased dynamics of the regions surrounding the mu loop in the absence of quaternary structure (Figure 31B), therefore, explain the increase in the intensity of the mutant monomer fluorescence spectra (Figure 10), because a more mobile environment can change the local quenching chemical environment and result in a higher quantum yield (Lakowicz, 2006).

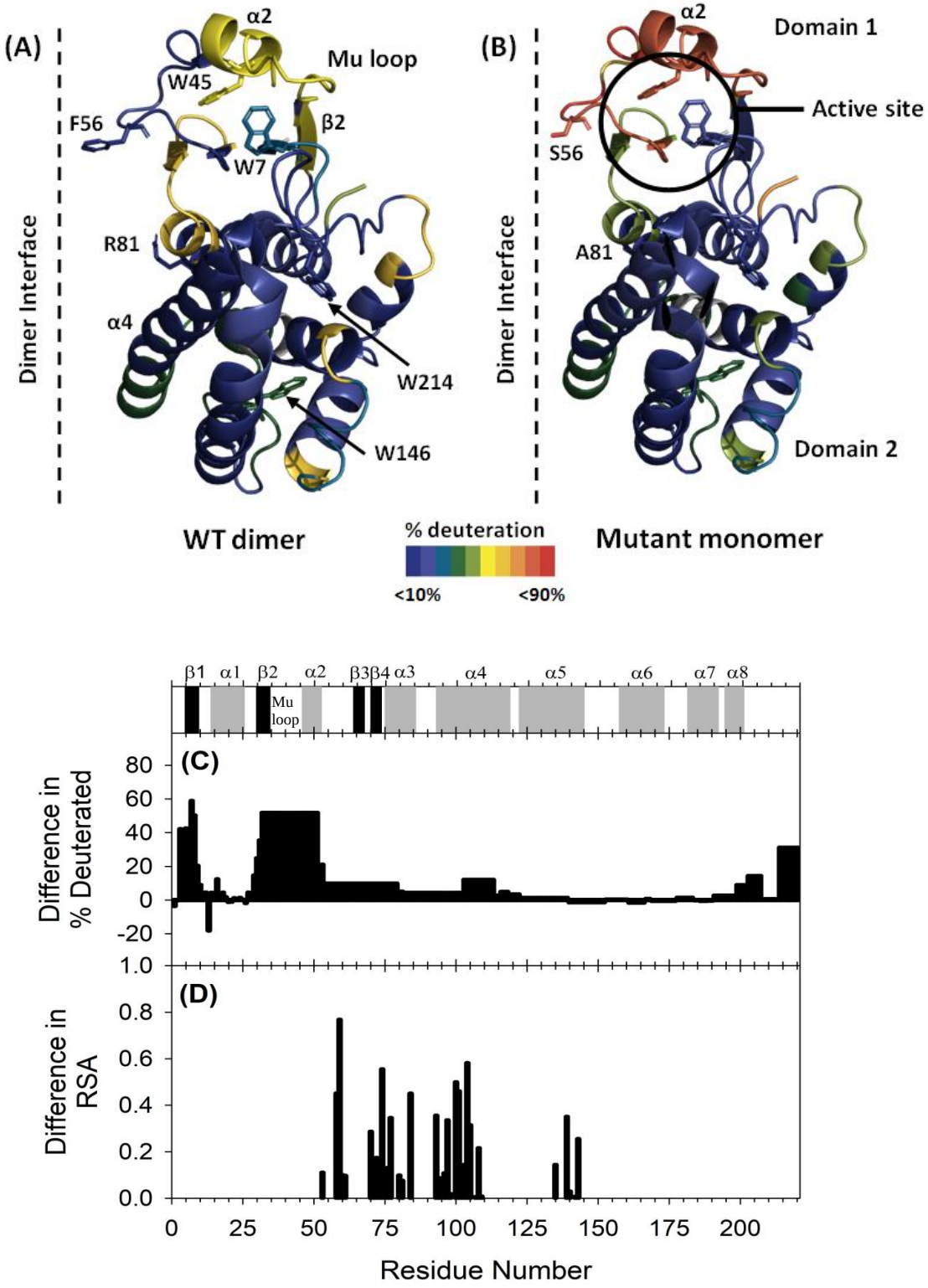


Figure 31: Conformational stability of native WT dimer and native mutant monomer

The percentage deuteration of the tertiary structures of (A) native WT dimer and (B) native mutant monomer indicates that α2 becomes highly dynamic in the absence of stabilising quaternary structure (PDB file 1GTU). A comparison between the differences in percentage deuteration (C) and relative solvent accessibility (RSA) between WT dimer and mutant monomer at 0 M urea indicate that even though the dimer interface has been disrupted in the mutant monomer, the tertiary and secondary structures have not been altered significantly.

The percentage deuteration of a peptide gives an indication of how dynamic the region is, because regions that are highly dynamic will show a high percentage deuteration (Bai *et al.*, 1993; Deng *et al.*, 1999; Loftus *et al.*, 1986; Zhang *et al.*, 1996). Continuous labelling HDX-MS experiments performed on the native F56S/R81A variant of rGST M1-1 indicate that β2, α2, β3 and β4 in domain 1 are highly dynamic (Thompson *et al.*, 2006). The most apparent difference in the levels of deuteration between native WT dimer and mutant monomer occurs in domain 1 (residues 22-79), indicating that like the rGST M1-1 monomeric variant, the human mutant monomer shows an increase in the dynamics of domain 1 structures (Figure 28A). The highly deuterated mutant monomer residues 22-79 include the F56 key residue and the mu loop, located between β2 and α2 (Figure 31A, 31B and 31C), confirming previous findings that in the absence of quaternary structure the regions surrounding the mu loop become highly dynamic (Hearne and Colman, 2006; Thompson *et al.*, 2006). The R81 residue forms part of the mixed charge cluster at the dimer interface (Figure 2B). Disruption of the mixed charge cluster increases the relative solvent accessibility (RSA) at the subunit interface (Figure 31D), and results in an increase in the percentage deuteration of β3, β4, α3 and α4 in the mutant monomer (Figure 31C). The dynamics of domain 1, and the solvent accessible surface of the dimer interface, are altered in the monomeric mutant; however, domain 2 remains relatively unchanged (Figure 31).

Therefore, in the absence of stabilising quaternary structure, the regions surrounding the mu loop become more dynamic (Figure 28A and Figure 31) and the dimer interface becomes more solvent-exposed. However, the global secondary (Figure 9) and tertiary (Figure 10) structure of the mutant monomer remain relatively unchanged. Therefore, the mutant monomer is a suitable model of the WT dimer subunit.

4.2. Unfolding proceeds via two stable equilibrium intermediates

The unfolding pathway of WT dimer hGST M1a-1a must begin with the native, folded dimer (N_2) and end with two denatured monomers (2D), and most likely proceeds via either a monomeric (M), or a dimeric (I_2) intermediate. The equilibrium unfolding of oligomeric proteins is protein-concentration dependent because of the law of mass action (Jaenicke, 1991; Neet and Timm, 1994; Ragone, 2000). The protein-concentration dependence of WT dimer unfolding during event 1 indicates that the dimer dissociates to form two monomeric intermediates (M_{WT}) during this event (Figure 15A and 15B). WT dimer unfolding, using ANS binding as a probe, provides further evidence for the formation of M_{WT} during unfolding event 1 (Figure 16A and 16C). ANS is an anionic dye that has been shown to bind to the hydrophobic H-site of numerous native GSTs (Erhardt and Dirr, 1995; Sayed *et al.*, 2000; Sluis-Cremer *et al.*, 1996; Stevens *et al.*, 1998), including class Mu GSTs (Hornby *et al.*, 2000; Kinsley *et al.*, 2008). The increase in ANS binding during event 1 indicates that there is an increase in the exposure of hydrophobic surface that was previously buried at the subunit interface (Figure 16C), which is consistent with the formation of a monomeric intermediate. Additionally, SE-HPLC data show that an intermediate, which has a smaller mass than native dimer, appears at approximately 4 M urea (Figure 17D and 17E). Therefore, it can be concluded that during unfolding event 1, at approximately 4 M urea, the inter-subunit interactions that govern dimer stabilisation are disrupted, and WT dimer dissociates to form $M_{WT.}$ The finding is in accordance with equilibrium unfolding studies conducted on rGST M1-1, which indicate that a ~20 kDa monomeric intermediate is present in the three-state unfolding of rGST M1-1 (Hornby *et al.*, 2000; Luo *et al.*, 2002; Thompson *et al.*, 2006). It is challenging to characterise M_{WT}, formed during unfolding event 1, because the intermediate cannot be separated from the other equilibrium unfolding species (Figure 17E and 17F).

Structural characterisation of the mutant monomer revealed that the tertiary structure of the native mutant monomer is comparable to the WT dimer subunit (see section 4.1). The unfolding pathway of the mutant monomer must begin with a native, folded monomer (M_{Mutant}) and end with denatured monomer (D). Therefore, elucidation of the equilibrium unfolding pathway of mutant monomer provides insight into the characteristics and unfolding trajectory of M_{WT}, and provides insight into the unfolding pathway of hGST M1a-1a. Previously, the F56S/R81A monomeric variant of rGST M1-1 was shown to lose secondary and tertiary structure simultaneously via a two-state unfolding process (Thompson *et al.*, 2006). The folding of many small dimeric and monomeric proteins, including GSTs and

GST-like proteins, has been shown to be a two-state process (Erhardt and Dirr, 1995; Gildenhuys *et al.*, 2008; Gloss and Matthews, 1998; Kaplan *et al.*, 1997; Wallace *et al.*, 1998b).

Mutant monomer unfolding appears to be two-state when monitored using far-UV CD (Figure 14C). However, secondary structural changes can occur in a diffuse manner during unfolding (Malhotra and Udgaonkar, 2016b), and it is well known that two-state unfolding can be more complex than it appears, depending on the probe that is used (Pace, 1990; Zaidi *et al.*, 1997). HDX-MS studies show that peptides 11-19 (residues 80-140) become deuterated in 4 M urea in the absence of quaternary structure (Figure 28C), indicating that α3, α4 and α5 do not unfold in urea concentrations below 4 M (event 1). Additionally, peptides 20-26 (residues 141-200) only become deuterated in 8 M urea (Figure 28F), indicating that α6 and α7 are not altered during unfolding event 1 from 0 M to 4.5 M urea. Therefore, unfolding monitored using far-UV CD appears two-state for the mutant monomer because ellipticity at 222 nm monitors the α-helical content of a protein, and the majority of the α-helices only begin to unfold in urea concentrations above 4 M in a diffuse manner. Mutant monomer unfolding was also assessed using intrinsic fluorescence, and this probe indicates that unfolding does not occur via a two-state mechanism (Figure 14A and 14B). Residue W214 is located in the C-terminal loop (Figure 31A and 31B), and it is not a good probe to monitor structural events because the highly dynamic region is predominantly solvent exposed, even in the native state (peptide 30 Figure 28). However, the other three tryptophan found within the protein subunit do serve as good probes for structural events. W7 and W45 are located at the active site, near the mu loop and monitor unfolding events in domain 1 (Figure 31A and 31B), whereas W146 is located in the loop region between α5 and α6, and inserts in to the region between α6 and α7, thereby monitoring unfolding events occurring in domain 2 (Figure 31A and 31B). The gradual and non-cooperative transition observed for mutant monomer in unfolding event 1 (Figure 14A) is a result of the minor conformational changes that occur in the regions surrounding the dynamic mu loop in the absence of quaternary structure, as detected by W7 and W45. Therefore, the minor conformational changes that occur in the mutant monomer at low concentrations of urea show that the mutant monomer loses structure gradually during unfolding event 1 (Figure 14A) to form a native-like M_{Mutant} intermediate (Figure 17L-O).

WT dimer dissociates to form the monomeric M_{WT} during unfolding event 1. WT dimer has a higher $\Delta G(H_2O)_1$ than mutant monomer (Table 1), which indicates that the WT dimer is more conformationally stable. However, a greater difference between the $\Delta G(H_2O)_1$ was expected because of the stabilising quaternary structure of the WT dimer. Dimer dissociation is highly cooperative and results in a large exposure of surface area, as indicated by the greater m_1-value obtained for WT dimer unfolding when compared with the m_1-value of mutant monomer (Table 1). The lower m_1-value obtained for mutant monomer unfolding suggests that only minor structural changes occur during unfolding event 1, and that the native monomer becomes more loosely packed at urea concentrations between 0 M and 4 M. The mechanisms employed, and the energy barriers that must be overcome in unfolding event 1 differ between the two proteins; however, the structural characteristics of M_{WT} and mutant monomer in approximately 4 M to 5 M urea are similar.

Hydrophobicity was assessed by monitoring unfolding using ANS binding as a probe. At 5 M urea, both proteins bind ANS to produce spectra with the same characteristics (Figure 16A and 16B), and M_{WT} and M_{Mutant} show equal relative ANS binding (Figure 16C), proving that M_{WT} and M_{Mutant} are equally hydrophobic (Figure 16A and 16B). M_{WT} and M_{Mutant} also have the same hydrodynamic size (Figure 18) and the same hydrodynamic diameter, as measured by DLS (Figure 19A and 19B), at 5 M urea. Finally, HDX-MS data show that WT dimer and mutant monomer are equally deuterated at 5 M urea, indicating that the equilibrium states that are populated at 5 M urea are conformationally similar (Figure 28D). Therefore, M_{Mutant} can be used to deconvolute the unfolding of hGST M1a-1a because M_{WT} and M_{Mutant} (referred to as M from this point forward) are structurally similar.

It is not unusual for stable monomeric intermediates to populate the unfolding pathway of GST proteins. A structured, stable intermediate has been observed for a GST from *Proteus mirabilius* (Sacchetta *et al.*, 1993), and bbGST P1-1 from *Bufo bufo* embryos unfolds via partially structured monomeric intermediates (Sacchetta *et al.*, 1999). Additionally, rGST M1-1 unfolds via a stable, monomeric intermediate (Hornby *et al.*, 2000). However, GST proteins have also been shown to unfold via more than one stable intermediate (Stevens *et al.*, 1998). The equilibrium unfolding data for WT dimer and mutant monomer provide evidence for a second equilibrium intermediate, in addition to the monomeric M intermediate formed on the urea-induced equilibrium unfolding pathway of hGST M1a-1a. The WT dimer displays protein-concentration dependence for both unfolding event 1 and 2 (Figure 15A and 15B). Monomeric proteins show no dependence upon protein concentration during two-state unfolding because these molecules are unimolecular and do not undergo bimolecular events

(Ragone, 2000); however, the mutant monomer shows protein-concentration dependence for unfolding event 2 (Figure 15C and 15D). Therefore, a protein-concentration dependent event occurs subsequent to the formation of M, indicating that a second multi-molecular intermediate populates the unfolding pathway of hGST M1a-1a. The oligomeric intermediate (I) formed by WT dimer and mutant monomer have the same hydrodynamic volume, a volume that is even greater than that of the denatured proteins (Figure 18). SE-HPLC data also show that oligomeric I is present at the same urea concentrations as M, the monomeric intermediate: at 4.5 M and at 5 M urea for WT dimer (Figure 17E and 17F) and from 3 M to 5 M urea for mutant monomer (Figure 17L-O). However, unlike dimer dissociation (event 1), the unfolding curves of WT dimer and mutant monomer are superimposable for unfolding event 2 (Figure 14 and Figure 16C). The $\Delta G(H_2O)_2$ of WT dimer is only 0.7 kcal.mol^{-1} greater than the $\Delta G(H_2O)_2$ of mutant monomer. The small difference indicates that oligomeric I is equally stable for both WT dimer and mutant monomer. Event 2 is also equally cooperative for both WT dimer and mutant monomer because the m_2-values are approximately equal (Table 1). Therefore, the event(s) occurring from 5 M to 8 M urea are the same for WT dimer and mutant monomer.

It is possible that I forms as a result of insoluble aggregates; however, this is unlikely because unfolding is reversible (Figure 12), and refolding curves for both the WT dimer and mutant monomer are superimposable with the unfolding curves of each respective protein (Figure 13). Additionally, the Rayleigh scattering at 390 nm shows no significant increase in scattering (Figure 16D), and absorbance spectra of protein samples unfolded in various concentrations of urea also show no increase in absorbance at 340 nm (Figure 19C). Therefore, it is more likely that oligomeric I forms as a result of the association of monomeric M, and that I is a stable intermediate and not an insoluble aggregate.

The formation of two stable equilibrium intermediates may result in a difference between the experimental and theoretical *m*-values, because when the *m*-value is obtained through analysis of a complex process, an underestimation of the *m*-value can be observed (Soulages, 1998). The *m*-value has been shown to correlate with the change in solvent accessible surface area (ΔASA), making it possible to calculate a theoretical *m*-value using the ΔASA (Myers *et al.*, 1995). The ΔASA for WT dimer is 40013 Å^2 and the theoretical *m*-value using urea as a denaturant is 4.8 kcal.mol^{-1}.M^{-1}. The theoretical *m*-value is nearly double the experimental *m*-value of 2.6 ± 0.2 kcal.mol^{-1}.M^{-1} (Table 1). The difference between the experimental and theoretical *m*-values is also observed for rGST M1-1, which has an experimental *m*-value of 2.4 kcal.mol^{-1}.M^{-1} and a theoretical *m*-value of 4.7 kcal.mol^{-1}.M^{-1} (Hornby *et al.*, 2000). The

ΔASA for mutant monomer is 19553 Å^2 and the theoretical *m*-value is 2.5 kcal.mol^{-1}.M^{-1}, and the mutant monomer m_{total}-value is 2.1 ± 0.2 kcal.mol^{-1}.M^{-1}. The m_{total}-value is in accordance with both the theoretical *m*-value and the 2.0 kcal.mol^{-1}.M^{-1} value reported for the F56S/R81A monomeric double mutant (Thompson *et al.*, 2006). Therefore, it is possible that WT dimer m_1-value (Table 1) is underestimated as a result of the formation of the monomeric intermediate (dimer dissociation), and not as a result of the formation of the oligomeric intermediate.

The 13.5 ± 0.9 kcal.mol^{-1} $\Delta G(H_2O)_{total}$ value for WT dimer falls below the 19.8–26.8 kcal.mol^{-1} range obtained for other dimeric GSTs (Erhardt and Dirr, 1995; Kaplan *et al.*, 1997; Wallace *et al.*, 1998b) and below the 19.8 kcal.mol^{-1} obtained for rGST M1-1 (Hornby *et al.*, 2000). The reduced $\Delta G(H_2O)_{total}$ could be a result of an inappropriate fitting model, or as a result of overparameterisation, which results in an underestimated $\Delta G(H_2O)_1$ for the dimer dissociation event. Conversely, the 11.3 ± 0.7 kcal.mol^{-1} $\Delta G(H_2O)_{total}$ value obtained for mutant monomer is within the 6-14 kcal.mol^{-1} range stated for monomeric proteins (Neet and Timm, 1994), and compares well with the 10.1 kcal.mol^{-1} obtained for the monomeric F56S/R81A rGST M1-1 mutant (Thompson *et al.*, 2006).

The change in Gibbs free energy in the absence of urea ($\Delta G(H_2O)_{total}$) will be affected by intermediate formation because the $\Delta G(H_2O)_{total}$ gives an indication of the conformational stability of a protein (Pace, 1986). Therefore, the underestimation of $\Delta G(H_2O)_1$ could be a result of the formation of M. Regardless of the underestimation of $\Delta G(H_2O)_1$, the $\Delta G(H_2O)_{total}$ value of the mutant monomer is lower than that of the WT dimer. The reduced $\Delta G(H_2O)_{total}$ value of the mutant monomer indicates that the native state of the mutant monomer is less stable than the native state of the WT dimer. Additionally, the reduced $\Delta G(H_2O)_{total}$ value is a result of the reduced $\Delta G(H_2O)_1$ value, and not as a result of the $\Delta G(H_2O)_2$ value. The conclusion is supported by the fact that the C_m value obtained for unfolding event 1 (C_{m1}) is lower for mutant monomer (3 M) than for WT dimer (3.8 M) whereas the C_m value obtained for unfolding event 2 (C_{m2}) is the same (6 M) for both proteins (Table 1).

Analysis of the C_m values obtained for each peptide from the HDX-MS data reveal that unfolding occurs cooperatively, based on the stability of different regions in hGST M1a-1a (Figure 30B and 30C). A lower C_m value indicates that the peptide is unfolding at a lower urea concentration, and is less susceptible to denaturation by urea. A higher C_m value indicates that the peptide unfolds at a higher urea concentration, and is more stable (Pan *et*

al., 2004). Figure 30 C indicates that the least stable key group peptides have a smaller fraction unfolded than the medium stable key group between 4 M and 6 M urea. The unexpected result could be due to aggregation of the mutant monomer at the high initial concentrations of protein required to conduct successful HDX-MS studies. Additionally, the mutant monomer begins to destabilise in 2 M urea (Figure 17 K) and at 3 M urea the oligomeric intermediate has already formed (Figure 17 L). Therefore, at urea concentrations between 4 M and 6 M, the monomeric intermediate (M) forms the oligomeric intermediate (I). The formation of I will result in a decrease in the percent deuteration, which will result in a decrease in the fraction of unfolded protein. The fraction of unfolded protein in each group reveals that the lock-and-key motif, as well as the structural elements surrounding the mu loop (domain1), are destabilised before the mixed-charge cluster at the subunit interface (domain 2) (Figure 30). It can be concluded that dimer dissociation requires the destabilisation of domain 1 interface structures (β2, α2, β3, β4 and part of α3) but that it is the destabilisation of α3 and domain 2 interface structures (α4 and α5) that results in complete dimer dissociation and the progression of the unfolding reaction.

In the EX1 limit, amide sites that are involved in a cooperative unfolding event will result in bimodal distributions in the mass spectra for that peptide at a specific urea concentration (Malhotra and Udgaonkar, 2015). These bimodal distributions will likely show two distinct peaks whose masses will reflect the undeuterated and the fully deuterated sites (Deng and Smith, 1999; Pan and Smith, 2004). Peptides within the highly unstable key group (Figure 24) and the medium stable charged cluster group (Figure 25 and Figure 26) display bimodal distributions in urea concentrations between 4 M and 8 M, which suggests that unfolding in these regions is cooperative. Additionally, analysis of the bimodality of the mass spectra indicates that the heavy peak at 5 M and 6 M are more deuterated than the denatured 8 M peak for both WT dimer and mutant monomer (Figure 24-26). The more heavily deuterated peak could indicate that subunit interface residues are involved in the formation of the exposed hydrophobic monomeric intermediates with numerous exposed residues, because an oligomeric intermediate would be shielded and show a lower mass to charge ratio (*m*/z) than a monomeric or denatured sample. The peptide data indicate that there is a high degree of interdomain cooperativity between domain 1 and domain 2, and that the two domains unfold in a concerted manner rather than independently, in accordance with studies performed on other GSTs (Gildenhuys *et al.,* 2010; Thompson *et al.,* 2006; Wallace *et al.,* 1998b).

The unfolding of hGST M1a-1a results in denatured protein that maintains partial residual structures. Residual structure can refer to native-like structural elements that guide the protein to the native state during refolding or simply to the non-native protection of solvent accessible surface area (Daggett and Fersht, 2003). The far-UV CD (Figure 20) and intrinsic tryptophan data (Figure 21) suggest that the denatured state at 8 M urea shows no residual tertiary or secondary structure. However, hGST M1a-1a is still able to bind ANS in 8 M urea, which indicates that residual structures are present (Figure 16). The residual structures of denatured states are often contained within the hydrophobic core of the protein (Hammarström and Jonsson, 2005). Domain 2 is defined as the hydrophobic core of GST class Mu (Ji *et al.*, 1992); however, the HDX-MS data presented in this study suggests that the highly stable core group is composed of α6 and α7, and that these helices are predominantly responsible for maintaining the stable core of the protein (Figure 28F and Figure 30). At 8 M urea peptides 10 (Figure 24), 14 (Figure 25) and 18 (Figure 26) are not fully deuterated, indicating that the denatured state is more shielded than the heavy peak at 5 M and 6 M. It is important to note that the residual structure of the denatured state could be as minimal as the protection of solvent accessible surface area in α6 and α7. However, the possibility that at 8 M both WT dimer and mutant monomer are not fully unfolded, as a result of the formation of oligomeric I (Figure 24B and Figure 25B), cannot be ignored.

The stability of the core structures of hGST M1a-1a, even in the absence of stabilising subunit interactions (Figure 28 and Figure 30), indicates that intersubunit interactions have evolved more recently and that ancestral GSTs were likely monomeric, which is in accordance with the hypothesis that canonical GSTs have evolved from monomeric enzymes like glutaredoxin 2 (Ladner *et al.*, 2004; Sheehan *et al.*, 2001). It is probable that the ancestral GSTs evolved to incorporate slightly varied intersubunit interactions between classes so that the biological function of GSTs could be increase by expanding the substrate specificity between various GSTs (see section 1.4).

4.3. Proposed equilibrium unfolding pathway of hGST M1a-1a

The data obtained from the in-depth comparative study on the equilibrium unfolding of WT dimer and the F56S/R81A mutant monomer has elucidated following proposed unfolding pathway for monomeric hGST M1a:

$$M \leftrightarrow I \leftrightarrow D \qquad \text{(Scheme 1)}$$

where M is the native F56S/R81A mutant monomer, I is an oligomeric equilibrium intermediate and D is the denatured state.

Additionally, the following equilibrium unfolding pathway for WT dimer hGST M1a-1a has been proposed:

$$N_2 \leftrightarrow 2M \leftrightarrow 2I \leftrightarrow 2D \qquad \text{(Scheme 2)}$$

where N_2 is the native dimer, M is a monomeric equilibrium intermediate, I is an oligomeric equilibrium intermediate and D is the denatured state.

The four-state equilibrium unfolding mechanism of hGST M1a-1a (Scheme 2) is in contrast with the two-state mechanism reported for class Alpha and class Pi GSTs (Erhardt and Dirr, 1995; Gildenhuys *et al.*, 2010; Kaplan *et al.*, 1997; Wallace *et al.*, 1998b) and the three-state mechanism reported for dimeric rGST M1-1 (Hornby *et al.*, 2000), bbGST P1-1 (Sacchetta *et al.*, 1999) and *Physa acuta* GST_3 (Abdalla and Hamed, 2006). The Sigma class also unfolds via a four-state mechanism, but unlike hGST M1a-1a, the oligomeric (dimeric) intermediate forms before the monomeric intermediate (Stevens *et al.*, 1998). A monomeric intermediate is also formed along the equilibrium unfolding pathway of rGST M1-1 and rGST M2-2, but no other intermediate states are observed for class Mu GST isolated from rat (Hornby *et al.*, 2000).

The differences observed in the equilibrium unfolding mechanisms of rGST M1-1 and hGST M1a-1a indicates that even minor differences in the primary structure of a protein can have an impact on the stability of the structural elements of the protein (Figure 3). It is not unheard of for minor differences in the primary structure within a class to cause significant differences in the equilibrium unfolding mechanism. An intermediate species that is stabilised in the equilibrium unfolding pathway of the GST class Pi isoenzyme isolated from the *Bufo bufo* embryo (bbGST P1-1) (Sacchetta *et al.*, 1999) is not detected in the unfolding pathway of class Pi GST isolated from pig (Erhardt and Dirr, 1995) or human (Gildenhuys *et al.*, 2010).

Differences in the conformational stability between isoenzymes within the rat GST Mu have also been reported (Hornby *et al.*, 2000). The rGST M1-1 and rGST M2-2 isoenzymes have a 78 % sequence similarity, but M1-1 was shown to have a weaker quaternary structure and a more stable monomeric intermediate than M2-2 (Hornby *et al.*, 2000; Hussey and Hayes, 1993; Luo *et al.*, 2002). Similarly, there is a 79 % similarity in sequence identity between hGST M1a-1a and rGST M1-1, with the majority of the differences occurring at the subunit interface. These differences could result in the altered properties of the monomeric intermediates formed by hGST M1a-1, when compared with the rGST M1-1 monomeric intermediate.

HDX-MS studies conducted on rGST M1-1 also show that α4 and α5 form a structural motif of stability in domain 2, which is only disrupted once domain 1 has become destabilised (Thompson *et al.*, 2006). However, the monomeric intermediate of rGST M1-1 does not associate to form an additional oligomeric intermediate. The 21 % difference in sequence similarity could result in the formation of a monomeric intermediate that is more stable than the M formed when hGST M1a-1a dissociates. The difference could arise from the fact that once the dimer has dissociated, the stability and conformation of α1 in domain 1 of rGST M1-1 remains relatively unchanged (Thompson *et al.*, 2006), whereas α1 of hGST M1a-1a becomes highly dynamic after dimer dissociation. Additionally, in hGST M1a-1a a C77 is present at the subunit interface, whereas in rGST M1-1 the residue is a conserved R77 (Figure 3) (Section 1.4.3). The R77 in domain 1 of rGST M1-1 forms an interdomain salt-bridge with D97 and E100 of domain 2 (Ji *et al.*, 1992), whereas the C77 in domain 1 of hGST M1a-1a is not able to form stabilising interdomain interaction (Figure 2B). Therefore, the monomeric intermediate (M) formed when hGST M1a-1a dissociates lacks the stabilising interdomain interaction that is present in the monomeric intermediate formed by rGST M1-1, which could account for the non-native association of M to form oligomeric I.

4.4. Conclusions

The equilibrium unfolding of hGST M1a-1a is four-state. The dissociation of N_2 to form monomeric M requires the destabilisation of structural elements in both domain 1 and domain 2. Destabilisation of the lock-and-key motif, as well as the structures surrounding the mu loop, results in a destabilisation of domain 1. The destabilisation of domain 1 results in destabilisation of α4 and α5 in domain 2, because the domains unfold in a concerted manner. Dimer dissociation results in the formation of a loosely-packed and hydrophobic monomeric equilibrium intermediate, with compromised short-range contacts (M). The compromised M self-associates to form a bulky, shielded oligomeric equilibrium intermediate, I. Finally, the destabilisation of α6 and α7 in the hydrophobic core of domain 2, results in further denaturation of hGST M1a-1a. The denatured state (D) is not fully unfolded, and minor residual structural elements are suspected to persist.

The equilibrium unfolding pathway proposed at this stage describes the stable equilibrium intermediates formed along the urea denaturation axis satisfactorily. In order to attain more reliable thermodynamic parameters for the WT dimer, complex and rigorous fitting models, which are able to fit the complete 4-state WT dimer unfolding, will have to be developed. Additionally, unfolding kinetics studies will be undertaken to obtain a more comprehensive understanding of the unfolding mechanism. An initial conditions test and double-jump unfolding experiments will clarify whether the unfolding pathway of hGST M1a-1a is sequential or parallel, and whether denaturation of I is the rate-limiting step. Chemical unfolding is a dynamic, multistate process with the coexistence of numerous conformations (Seelig, 2018). It will, therefore, also be useful to study the unfolding pathway of hGST M1a-1a using other denaturants, such as guanidinium chloride, pH and pressure to provide further details about cooperativity and the energy barriers to unfolding. Investigation into these characteristics will provide a more complete understanding of the structural basis of hGST M1a-1a unfolding, and provide novel insights into the mechanisms of protein (un)folding.

The field of protein (un)folding is constantly being critically re-examined and the established views of protein (un)folding are thoroughly questioned, tested and challenged (Arai, 2018; Lapidus, 2017; Žoldák *et al.*, 2017). Researchers are constantly working at defining additional parameters that can be used to describe protein (un)folding (Hall *et al.*, 2018), and it is important that we continue to add to the knowledge base of one of most pertinent cumulative scientific investigations of the last century.

References

Abdalla, A. M., Bruns, C. M., Tainer, J., Mannervik, B., and Stenberg, G. (2002). Design of a monomeric human glutathione transferase GSTP1, a structurally stable but catalytically inactive protein. *Protein Engineering*, **15**(10), 827–34.

Abdalla, A. M., and Hamed, R. R. (2006). Multiple unfolding states of glutathione transferase from Physa acuta (Gastropada: Physidae). *Biochemical and Biophysical Research Communications*, **340**(2), 625–632.

Aceto, A., Caccuri, A. M., Sacchetta, P., Bucciarelli, T., Dragani, B., Rosato, N., Federici, G., and Di Ilio, C. (1992). Dissociation and unfolding of Pi-class glutathione transferase. Evidence for a monomeric inactive intermediate. *The Biochemical Journal*, **285 (Pt 1**, 241–5.

Acuner Ozbabacan, S. E., Engin, H. B., Gursoy, A., and Keskin, O. (2011). Transient protein-protein interactions. *Protein Engineering, Design & Selection : PEDS*, **24**(9), 635–48.

Adler, V., Yin, Z., Fuchs, S. Y., Benezra, M., Rosario, L., Tew, K. D., Pincus, M. R., Sardana, M., Henderson, C. J., Wolf, C. R., Davis, R. J., and Ronai, Z. (1999). Regulation of JNK signaling by GSTp. *The EMBO Journal*, **18**(5), 1321–34.

Aguzzi, A., and O'Connor, T. (2010). Protein aggregation diseases: Pathogenicity and therapeutic perspectives. *Nature Reviews Drug Discovery*, **9**(3), 237–248.

Ali, M. H., and Imperiali, B. (2005). Protein oligomerization: how and why. *Bioorganic & Medicinal Chemistry*, **13**(17), 5013–20.

Allocati, N., Masulli, M., Pietracupa, M., Federici, L., and Di Ilio, C. (2006). Evolutionarily conserved structural motifs in bacterial GST (glutathione S-transferase) are involved in protein folding and stability. *The Biochemical Journal*, **394**(Pt 1), 11–7.

Alves, C. S., Kuhnert, D. C., Sayed, Y., and Dirr, H. W. (2006). The intersubunit lock-and-key motif in human glutathione transferase A1-1: role of the key residues Met51 and Phe52 in function and dimer stability. *The Biochemical Journal*, **393**(Pt 2), 523–8.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science (New York, N.Y.)*, **181**(4096), 223–30.

Arai, M. (2018). Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophysical Reviews*. doi:10.1007/s12551-017-0346-7

Arai, M., and Kuwajima, K. (2000). Role of the molten globule state in protein folding. In *Advances in protein chemistry*, Vol. 53, pp. 209–282.

Armstrong, R. N. (1997). Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chemical Research in Toxicology*, **10**(1), 2–18.

Bader, M., Muse, W., Ballou, D. P., Gassner, C., and Bardwell, J. C. A. (1999). Oxidative protein folding is driven by the electron transport system. *Cell*, **98**(2), 217–227.

Bahadur, R. P., Chakrabarti, P., Rodier, F., and Janin, J. (2003). Dissecting subunit interfaces in homodimeric proteins. *Proteins*, **53**(3), 708–19.

Bai, Y., Milne, J. S., Mayne, L. C., and Englander, W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins*, **17**(1), 75–86.

Bai, Y., Sosnick, T. R., Mayne, L. C., and Englander, W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science (New York, N.Y.)*, **269**(5221), 192–7.

Balaji, S. (2015). Internal symmetry in protein structures: Prevalence, functional relevance and evolution. *Current Opinion in Structural Biology*, **32**, 156–166.

Baldwin, R. L. (1994). Matching speed and stability. *Nature*, **369**(6477), 183–184.

Baldwin, R. L. (1996). On-pathway versus off-pathway folding intermediates. *Folding & Design*, **1**(1), R1–R8.

Baldwin, R. L. (2014). Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. *Proceedings of the National Academy of Sciences*, **111**(36), 13052–13056.

Baldwin, R. L. (2017). Clash between energy landscape theory and foldon-dependent protein folding. *Proceedings of the National Academy of Sciences*, 201709133.

Baldwin, R. L., Frieden, C., and Rose, G. D. (2010). Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins: Structure, Function, and Bioinformatics*, **78**(13), 2725–2737.

Baldwin, R. L., and Rose, G. D. (2013). Molten globules, entropy-driven conformational change and protein folding. *Current Opinion in Structural Biology*, **23**(1), 4–10.

Barlow, D. J., and Thornton, J. M. (1983). Ion-pairs in proteins. *Journal of Molecular Biology*, **168**(4), 867–885.

Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998). The proteasome: Paradigm of a self-compartmentalizing protease. *Cell*, **92**(3), 367–380.

Bédard, S., Mayne, L. C., Peterson, R. W., Wand, J., and Englander, W. (2008). The foldon substructure of staphylococcal nuclease. *Journal of Molecular Biology*, **376**(4), 1142–54.

Bedouelle, H. (2016). Principles and equations for measuring and interpreting protein stability: From monomer to tetramer. *Biochimie*, **121**, 29–37.

Bhaskara, R. M., de Brevern, A. G., and Srinivasan, N. (2013). Understanding the role of domain-domain linkers in the spatial orientation of domains in multi-domain proteins. *Journal of Biomolecular Structure & Dynamics*, **31**(12), 1467–80.

Bhaskara, R. M., and Srinivasan, N. (2011). Stability of domain structures in multi-domain proteins. *Scientific Reports*, **1**, 40.

Bhutani, N., and Udgaonkar, J. B. (2003). Folding subdomains of thioredoxin characterized by native-state hydrogen exchange. *Protein Science : A Publication of the Protein Society*, **12**(8), 1719–31.

Board, P. G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermiin, L. S., Schulte, G. K., Danley, D. E., Hoth, L. R., Griffor, M. C., Kamath, A. V., Rosner, M. H., Chrunyk, B. A., Perregaux, D. E., Gabel, C. A., Geoghegan, K. F., and Pandit, J. (2000). Identification, characterization, and crystal structure of the omega class glutathione transferases. *Journal of Biological Chemistry*, **275**(32), 24798–24806.

Borgia, M. B., Nickson, A. A., Clarke, J., and Hounslow, M. J. (2013). A mechanistic model for amorphous protein aggregation of immunoglobulin-like domains. *Journal of the American Chemical Society*, **135**(17), 6456–6464.

Braman, J., Papworth, C., and Greener, A. (1996). Site-directed mutagenesis using double-stranded plasmid DNA templates. *Methods in Molecular Biology*, **57**, 31–44.

Brandts, J. F., Brennan, M., and Lung-Nan Lin. (1977). Unfolding and refolding occur much faster for a proline-free proteins than for most proline-containing proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **74**(10), 4178–81.

Brinda, K. V, Kannan, N., and Vishveshwara, S. (2002). Analysis of homodimeric protein interfaces by graph-spectral methods. *Protein Engineering*, **15**(4), 265–277.

Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nature Structural Biology*, **3**(9), 782–7.

Chan, H. S., and Dill, K. A. (1998). Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins*, **30**(1), 2–33.

Chang, M., Bolton, J. L., and Blond, S. Y. (1999). Expression and Purification of Hexahistidine-Tagged Human Glutathione S-Transferase P1-1 in Escherichia coli. *Protein Expression and Purification*, **17**(3), 443–448.

Chenevix-Trench, G., Young, J., Coggan, M., and Board, P. G. (1995). Glutathione S-transferase M1 and T1 polymorphisms: susceptibility to colon cancer and age of onset. *Carcinogenesis*, **16**(7), 1655–7.

Chiti, F., and Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, **75**(1), 333–66.

Cho, S. G., Lee, Y. H., Park, H. S., Ryoo, K., Kang, K. W., Park, J., Eom, S. J., Kim, M. J., Chang, T. S., Choi, S. Y., Shim, J., Kim, Y., Dong, M. S., Lee, M. J., Kim, S. G., Ichijo, H., and Choi, E. J. (2001). Glutathione S-Transferase Mu Modulates the Stress-activated Signals by Suppressing Apoptosis Signal-regulating Kinase 1. *Journal of Biological Chemistry*, **276**(16), 12749–12755.

Chothia, C., and Janin, J. (1975). Principles of protein–protein recognition. *Nature*, **256**(5520), 705–708.

Combes, B., and Stakelum, G. S. (1961). A LIVER ENZYME THAT CONJUGATES SULFOBROMOPHTHALEIN SODIUM WITH GLUTATHIONE*. *Journal of Clinical Investigation*, **40**(6), 981–988.

Cook, K. H., Schmid, F. X., and Baldwin, R. L. (1979). Role of proline isomerization in folding of ribonuclease A at low temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, **76**(12), 6157–6161.

Creighton, T. E. (1990). Protein folding. *The Biochemical Journal*, **270**(1), 1–16.

Cunningham, B., Mulkerrin, M., and Wells, J. (1991). Dimerization of human growth hormone by zinc. *Science*, **253**(5019), 545–548.

Dabora, J. M., Pelton, J. G., and Marqusee, S. (1996). Structure of the acid state of Escherichia coli ribonuclease HI. *Biochemistry*, **35**(37), 11951–11958.

Daggett, V., and Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends in Biochemical Sciences*, **28**(1), 18–25.

Dasgupta, A., and Udgaonkar, J. B. (2012). Four-state folding of a SH3 domain: Salt-induced modulation of the stabilities of the intermediates and native state. *Biochemistry*, **51**(23), 4723–4734.

Dautant, A., Meyer, P., and Georgescauld, F. (2017). Hydrogen/Deuterium Exchange Mass Spectrometry Reveals Mechanistic Details of Activation of Nucleoside Diphosphate Kinases by Oligomerization. *Biochemistry*, **56**(23), 2886–2896.

De, D., Singh, A., and Gupta, A. N. (2017). Locating transition path region in the free energy landscape of protein folding. Retrieved from http://arxiv.org/abs/1705.01246

de Oliveira, G. A. P., and Silva, J. L. (2015). A hypothesis to reconcile the physical and chemical unfolding of proteins. *Proceedings of the National Academy of Sciences*, **112**(21), E2775–E2784.

de Oliveira, G. A. P., and Silva, J. L. (2017). The push-and-pull hypothesis in protein unfolding, misfolding and aggregation. *Biophysical Chemistry*, **231**, 20–26.

Deng, Y., and Smith, D. L. (1998). Identification of Unfolding Domains in Large Proteins by Their Unfolding Rates †. *Biochemistry*, **37**(18), 6256–6262.

Deng, Y., and Smith, D. L. (1999). Hydrogen exchange demonstrates three domains in aldolase unfold sequentially. *Journal of Molecular Biology*, **294**(1), 247–58.

Deng, Y., Zhang, Z., and Smith, D. L. (1999). Comparison of continuous and pulsed labeling amide hydrogen exchange/mass spectrometry for studies of protein dynamics. *Journal of the American Society for Mass Spectrometry*, **10**(8), 675–684.

Dill, K. A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry*, **24**(6), 1501–1509.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**(31), 7133–7155.

Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. (1995). Principles of protein folding--a perspective from simple exact models. *Protein Science : A Publication of the Protein Society*, **4**(4), 561–602.

Dill, K. A., and Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural & Molecular Biology*, **4**(1), 10–19.

Dill, K. A., and Shortle, D. (1991). Denatured states of proteins. *Annual Review of Biochemistry*, **60**, 795–825.

Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M. (2000). Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences*, **25**(7), 331–9.

Dirr, H. W. (2001). Folding and assembly of glutathione transferases. *Chemico-Biological Interactions*, **133**, 24–27.

Dirr, H. W., and Reinemer, P. (1991). Equilibrium unfolding of class pi glutathione S-transferase. *Biochemical and Biophysical Research Communications*, **180**(1), 294–300.

Dirr, H. W., Reinemer, P., and Huber, R. (1994). X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *European Journal of Biochemistry / FEBS*, **220**(3), 645–61.

Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, **426**(6968), 884–890.

Dorion, S., Lambert, H., and Landry, J. (2002). Activation of the p38 signaling pathway by heat shock involves the dissociation of glutathione S-transferase Mu from Ask1. *Journal of Biological Chemistry*, **277**(34), 30792–30797.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., … Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, **19**(1), 26–59.

Eaton, W. A., Muñoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R., and Hofrichter, J. (2000). Fast kinetics and mechanisms in protein folding. *Annual Review of Biophysics and Biomolecular Structure*, **29**(1), 327–59.

Eaton, W. A., and Wolynes, P. G. (2017). Theory, simulations, and experiments show that proteins fold by multiple pathways. *Proceedings of the National Academy of Sciences*, **114**(46), E9759–E9760.

Ellis, R. J. (2001). Macromolecular crowding: An important but neglected aspect of the intracellular environment. *Current Opinion in Structural Biology*, **11**(1), 114–119.

Englander, W. (2000). Protein Folding Intermediates and Pathways Studied by Hydrogen Exchange. *Annual Review of Biophysics and Biomolecular Structure*, **29**(1), 213–238.

Englander, W., and Mayne, L. C. (2014). The nature of protein folding pathways. *Proceedings of the National Academy of Sciences*, **111**(45), 15873–15880.

Englander, W., and Mayne, L. C. (2017a). Reply to Eaton and Wolynes: How do proteins fold? *Proceedings of the National Academy of Sciences*, (1), 201716929.

Englander, W., and Mayne, L. C. (2017b). The case for defined protein folding pathways. *Proceedings of the National Academy of Sciences*, **114**(31), 8253–8258.

Englander, W., Mayne, L. C., and Krishna, M. M. G. (2007). Protein folding and misfolding: mechanism and principles. *Quarterly Reviews of Biophysics*, **40**(4), 287–326.

Erhardt, J., and Dirr, H. W. (1995). Native dimer stabilizes the subunit tertiary structure of porcine class pi glutathione S-transferase. *European Journal of Biochemistry*, **230**(2), 614–620.

Eyles, S. J., and Gierasch, L. M. (2000). Multiple roles of prolyl residues in structure and folding. *Journal of Molecular Biology*, **301**(3), 737–747.

Ezer, R., Alonso, M., Pereira, E., Kim, M., Allen, J. C., Miller, D. C., and Newcomb, E. W. (2002). Identification of glutathione S-transferase (GST) polymorphisms in brain tumors and association with susceptibility to pediatric astrocytomas. *Journal of Neuro-Oncology*, **59**(2), 123–134.

Fabrini, R., De Luca, A., Stella, L., Mei, G., Orioni, B., Ciccone, S., Federici, G., Lo Bello, M., and Ricci, G. (2009). Monomer-dimer equilibrium in glutathione transferases: A critical re-examination. *Biochemistry*, **48**(43), 10473–10482.

Fang, J., Rand, K. D., Beuning, P. J., and Engen, J. R. (2011). False EX1 signatures caused by sample carryover during HX MS analyses. *International Journal of Mass Spectrometry*, **302**(1–3), 19–25.

Ferguson, N., and Fersht, A. R. (2003). Early events in protein folding. *Current Opinion in Structural Biology*, **13**(1), 75–81.

Fersht, A. R. (1984). Basis of biological specificity. *Trends in Biochemical Sciences*, **9**(4), 145–147.

Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Current Opinion in Structural Biology*, **7**, 3–9.

Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., and Winter, G. (1985). Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, **314**(6008), 235–238.

Fink, A. L. (1998). Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design*, **3**(1), R9–R23.

Fink, A. L., Oberg, K. A., and Seshadri, S. (1998). Discrete intermediates versus molten globule models for protein folding: characterization of partially folded intermediates of apomyoglobin. *Folding and Design*, **3**(1), 19–25.

Finkelstein, A. V., Badretdin, A. J., Galzitskaya, O. V., Ivankov, D. N., Bogatyreva, N. S., and Garbuzynskiy, S. O. (2017). There and back again: Two views on the protein folding puzzle. *Physics of Life Reviews*, (April). doi:10.1016/j.plrev.2017.01.025

Font, J., Torrent, J., Ribó, M., Laurents, D. V, Balny, C., Vilanova, M., and Lange, R. (2006). Pressure-jump-induced kinetics reveals a hydration dependent folding/unfolding mechanism of ribonuclease A. *Biophysical Journal*, **91**(6), 2264–74.

Galani, D., Fersht, A. R., and Perrett, S. (2002). Folding of the yeast prion protein Ure2: kinetic evidence for folding and unfolding intermediates. *Journal of Molecular Biology*, **315**(2), 213–27.

Gatzeva-Topalova, P. Z., Warner, L. R., Pardi, A., and Sousa, M. C. (2010). Structure and Flexibility of the Complete Periplasmic Domain of BamA: The Protein Insertion Machine of the Outer Membrane. *Structure*, **18**(11), 1492–1501.

George, R. A., and Heringa, J. (2002). An analysis of protein domain linkers: their classification and role in protein folding. *Protein Engineering*, **15**(11), 871–9.

Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2003). Unifying features in protein-folding mechanisms. *Proceedings of the National Academy of Sciences*, **100**(23), 13286–13291.

Gianni, S., Ivarsson, Y., De Simone, A., Travaglini-Allocatelli, C., Brunori, M., and Vendruscolo, M. (2010). Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nature Structural and Molecular Biology*, **17**(12), 1431–1437.

Gildenhuys, S., Wallace, L., Burke, J. P., Balchin, D., Sayed, Y., and Dirr, H. W. (2010). Class Pi Glutathione Transferase Unfolds via a Dimeric and Not Monomeric Intermediate: Functional Implications for an Unstable Monomer. *Biochemistry*, **49**(24), 5074–5081.

Gildenhuys, S., Wallace, L., and Dirr, H. W. (2008). Stability and unfolding of reduced Escherichia coli glutaredoxin 2: A monomeric structural homologue of the glutathione transferase family. *Biochemistry*, **47**(40), 10801–10808.

Gloss, L. M., and Matthews, C. R. (1997). Urea and thermal equilibrium denaturation studies on the dimerization domain of Escherichia coli Trp repressor. *Biochemistry*, **36**(19), 5612–5623.

Gloss, L. M., and Matthews, C. R. (1998). The barriers in the bimolecular and unimolecular folding reactions of the dimeric core domain of Escherichia coli Trp repressor are dominated by enthalpic contributions. *Biochemistry*, **37**(45), 16000–16010.

Goncearenco, A., Shoemaker, B. A., Zhang, D., Sarychev, A., and Panchenko, A. R. (2014). Coverage of protein domain families with structural protein–protein interactions: Current progress and future trends. *Progress in Biophysics and Molecular Biology*, **116**(2–3), 187–193.

Grantcharova, V., Alm, E. J., Baker, D., and Horwich, A. L. (2001). Mechanisms of protein folding. *Curr Opin Struct Biol*, **11**(1), 70–82.

Griffin, M. D. W., and Gerrard, J. A. (2012). The relationship between oligomeric state and protein function. *Advances in Experimental Medicine and Biology*, **747**, 74–90.

Griko, Y. V. (1999). Denaturation versus unfolding: energetic aspects of residual structure in denatured alpha-lactalbumin. *Journal of Protein Chemistry*, **18**(3), 361–9.

Guimarães, C. R. W., and Bicca de Alencastro, R. (2002). Thrombin Inhibition by Novel Benzamidine Derivatives: A Free-Energy Perturbation Study. *Journal of Medicinal Chemistry*, **45**(23), 4995–5004.

Habig, W. H., Pabst, M. J., and Jakoby, W. B. (1974). Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. *The Journal of Biological Chemistry*, **249**(22), 7130–9.

Hall, D., Kinjo, A. R., and Goto, Y. (2018). A new look at an old view of denaturant induced protein unfolding. *Analytical Biochemistry*, **542**, 40–57.

Hammarström, P., and Jonsson, B.-H. (2005). Protein Denaturation and the Denatured State. In *eLS*, Chichester, UK: John Wiley & Sons, Ltd, pp. 1–7.

Han, J.-H., Batey, S., Nickson, A. A., Teichmann, S. A., and Clarke, J. (2007). The folding and evolution of multidomain proteins. *Nature Reviews Molecular Cell Biology*, **8**(4), 319–330.

Hartl, U. (2010). Protein folding : mechanisms and role in disease.

Hayes, J. D., and McLellan, L. I. (1999). Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress. *Free Radical Research*, **31**(4), 273–300.

Hayes, J. D., and Pulford, D. J. (1995). The Glut athione S-Transferase Supergene Family: Regulation of GST and the Contribution of the lsoenzymes to Cancer Chemoprotection and Drug Resistance Part I. *Critical Reviews in Biochemistry and Molecular Biology*, **30**(6), 445–520.

Hearne, J. L., and Colman, R. F. (2006). Contribution of the mu loop to the structure and function of rat glutathione transferase M1-1. *Protein Science*, **15**(6), 1277–1289.

Hegazy, U. M., Hellman, U., and Mannervik, B. (2006). Replacement Surgery with Unnatural Amino Acids in the Lock-and-Key Joint of Glutathione Transferase Subunits. *Chemistry and Biology*, **13**(9), 929–936.

Hegazy, U. M., Mannervik, B., and Stenberg, G. (2004). Functional Role of the Lock and Key Motif at the Subunit Interface of Glutathione Transferase P1-1. *Journal of Biological Chemistry*, **279**(10), 9586–9596.

Henderson, C. J., McLaren, A. W., Moffat, G. J., Bacon, E. J., and Wolf, C. R. (1998). Pi-class glutathione S-transferase: regulation and function. *Chemico-Biological Interactions*, **111**–**112**, 69–82.

Hennessey, J. P., Johnson, W. C., Bahler, C., and Wood, H. G. (1982). Subunit interactions of transcarboxylase as studied by circular dichroism. *Biochemistry*, **21**(4), 642–646.

Heringa, J., and Argos, P. (1991). Side-chain clusters in protein structures and their role in protein folding. *Journal of Molecular Biology*, **220**(1), 151–171.

Hinderaker, M. P., and Raines, R. T. (2003). An electronic effect on protein structure. *Protein Science : A Publication of the Protein Society*, **12**(6), 1188–94.

Hoa, B. (1973). Stopped Flow, pp. 127–136.

Hornby, J. A. T., Codreanu, S. G., Armstrong, R. N., and Dirr, H. W. (2002). Molecular Recognition at the Dimer Interface of a Class Mu Glutathione Transferase: Role of a Hydrophobic Interaction Motif in Dimer Stability and Protein Function †. *Biochemistry*, **41**(48), 14238–14247.

Hornby, J. A. T., Luo, J.-K., Stevens, J. M., Wallace, L., Kaplan, W., Armstrong, R. N., and Dirr, H. W. (2000). Equilibrium Folding of Dimeric Class μ Glutathione Transferases Involves a Stable Monomeric Intermediate †. *Biochemistry*, **39**(40), 12336–12344.

Hou, L., Honaker, M. T., Shireman, L. M., Balogh, L. M., Roberts, A. G., Ng, K. C., Nath, A., and Atkins, W. M. (2007). Functional promiscuity correlates with conformational heterogeneity in A-class glutathione S-transferases. *Journal of Biological Chemistry*, **282**(32), 23264–23274.

Hu, W., Kan, Z.-Y., Mayne, L. C., and Englander, W. (2016). Cytochrome c folds through foldon-dependent native-like intermediates in an ordered pathway. *Proceedings of the National Academy of Sciences*, **113**(14), 3809–3814.

Hu, W., Walters, B. T., Kan, Z., Mayne, L. C., Rosen, L. E., Marqusee, S., and Englander, W. (2013). Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proceedings of the National Academy of Sciences*, **110**(19), 7684–7689.

Huang, Y., Misquitta, S., Blond, S. Y., Adams, E., and Colman, R. F. (2008). Catalytically Active Monomer of Glutathione S -Transferase π and Key Residues Involved in the Electrostatic Interaction between Subunits. *Journal of Biological Chemistry*, **283**(47), 32880–32888.

Hubbard, S. J., and Argos, P. (1994). Cavities and packing at protein interfaces. *Protein Science : A Publication of the Protein Society*, **3**(12), 2194–206.

Hussey, A. J., and Hayes, J. D. (1993). Human Mu-class glutathione S-transferases present in liver, skeletal muscle and testicular tissue. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, **1203**(1), 131–141.

Ichijo, H., Nishida, E., Irie, K., Ten Dijke, P., Saitoh, M., Moriguchi, T., Takagi, M., Matsumoto, K., Miyazono, K., and Gotoh, Y. (1997). Induction of apoptosis by ASK1, a mammalian MAPKKK that activates SAPK/JNK and p38 signaling pathways. *Science*, **275**(5296), 90–94.

Itoh, K., and Sasai, M. (2008). Cooperativity, connectivity, and folding pathways of multidomain proteins. *Proceedings of the National Academy of Sciences*, **105**(37), 13865–13870.

Jackson, S. E. (1998). How do small single-domain proteins fold? *Folding and Design*, **3**(4), R81–R91.

Jackson, S. E., and Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry*, **30**(43), 10436–10443.

Jaenicke, R. (1991). Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry*, **30**(13), 3147–3161.

Jaenicke, R. (1999). Stability and folding of domain proteins. *Progress in Biophysics and Molecular Biology*, **71**(2), 155–241.

Jaenicke, R., and Rudolph, R. (1989). Folding proteins. In T. E. Creighton, ed., *Protein Structure a Practical Approach*, Oxford University Press, pp. 191–223.

Janin, J., Miller, S., and Chothia, C. (1988). Surface, subunit interfaces and interior of oligomeric proteins. *Journal of Molecular Biology*, **204**(1), 155–64.

Jennings, P. A., and Wright, P. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*, **262**(5135), 892–896.

Ji, X., Zhang, P., Armstrong, R. N., and Gilliland, G. L. (1992). The three-dimensional structure of a glutathione S-transferase from the mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-A resolution. *Biochemistry*, **31**(42), 10169–84.

Jones, S., Marin, A., and Thornton, J. M. (2000). Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Engineering, Design and Selection*, **13**(2), 77–82.

Jones, S., and Thornton, J. M. (1995). Protein-protein interactions: A review of protein dimer structures. *Progress in Biophysics and Molecular Biology*, **63**(1), 31–65.

Jones, S., and Thornton, J. M. (1996). Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences*, **93**(1), 13–20.

Kaplan, W., Hüsler, P., Klump, H., Erhardt, J., Sluis-Cremer, N., and Dirr, H. W. (1997). Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag. *Protein Science : A Publication of the Protein Society*, **6**(2), 399–406.

Karplus, M., and Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Science : A Publication of the Protein Society*, **3**(4), 650–68.

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

Kelley, R. F., and Richards, F. M. (1987). Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry*, **26**(21), 6765–74.

Khan, M. S., Bhat, S. A., Tabrez, S., Alama, M. N., Alsenaidy, M. A., and Al-Senaidy, A. M. (2016a). Denaturation induced aggregation in α-crystallin: differential action of chaotropes. *Journal of Molecular Recognition*, **29**(11), 536–543.

Khan, P., Prakash, A., Haque, M. A., Islam, A., Hassan, M. I., and Ahmad, F. (2016b). Structural basis of urea-induced unfolding: Unraveling the folding pathway of hemochromatosis factor E. *International Journal of Biological Macromolecules*, **91**, 1051–1061.

Kharrat, N., Belmabrouk, S., Abdelhedi, R., Benmarzoug, R., Assidi, M., Al Qahtani, M. H., and Rebai, A. (2016). Screening for clusters of charge in human virus proteomes. *BMC Genomics*, **17**(Suppl 9). doi:10.1186/s12864-016-3086-3

Kiefhaber, T. (1995). Kinetic traps in lysozyme folding. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(20), 9029–33.

Kiefhaber, T., and Baldwin, R. L. (1995). Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. *Proceedings of the National Academy of Sciences*, **92**(7), 2657–2661.

Kiefhaber, T., Labhardt, A., M., and Baldwin, R. L. (1995). Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature*, **375**(6531), 513–515.

Kiefhaber, T., Schmid, F. X., Willaert, K., Engelborghs, Y., and Chaffotte, A. (1992). Structure of a rapidly formed intermediate in ribonuclease T1 folding. *Protein Science : A Publication of the Protein Society*, **1**, 1162–1172.

Kim, B. L., Schafer, N. P., and Wolynes, P. G. (2014). Predictive energy landscapes for folding {alpha}-helical transmembrane proteins. *Pnas*, **243101**, 1410529111-.

Kim, P. S., and Baldwin, R. L. (1990). Intermediates In The Folding Reactions Of Small Proteins. *Annual Review of Biochemistry*, **59**(1), 631–660.

Kinsley, N., Sayed, Y., Mosebi, S., Armstrong, R. N., and Dirr, H. W. (2008). Characterization of the binding of 8-anilinonaphthalene sulfonate to rat class Mu GST M1-1. *Biophysical Chemistry*, **137**(2–3), 100–4.

Krishna, M. M. G., and Englander, W. (2007). A unified mechanism for protein folding: Predetermined pathways with optional errors. *Protein Science*, **16**(3), 449–464.

Krishna, M. M. G., Hoang, L., Lin, Y., and Englander, W. (2004). Hydrogen exchange methods to study protein folding. *Methods (San Diego, Calif.)*, **34**(1), 51–64.

Kumar, A., Dhull, D. K., Gupta, V., Channana, P., Singh, A., Bhardwaj, M., Ruhal, P., and Mittal, R. (2017). Role of Glutathione-S-transferases in neurological problems. *Expert Opinion on Therapeutic Patents*, **27**(3), 299–309.

Kumar, S., and Nussinov, R. (2002). Relationship between Ion Pair Geometries and Electrostatic Strengths in Proteins. *Biophysical Journal*, **83**(3), 1595–1612.

Kura, T., Takahashi, Y., Takayama, T., Ban, N., Saito, T., Kuga, T., and Niitsu, Y. (1996). Glutathione S-transferase-π is secreted as a monomer into human plasma by platelets and tumor cells. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, **1292**(2), 317–323.

Kuroda, D., and Gray, J. J. (2016). Shape complementarity and hydrogen bond preferences in protein-protein interfaces: Implications for antibody modeling and protein-protein docking. *Bioinformatics*, **32**(16), 2451–2456.

Ladner, J. E., Parsons, J. F., Rife, C. L., Gilliland, G. L., and Armstrong, R. N. (2004). Parallel evolutionary pathways for glutathione transferases: structure and mechanism of the mitochondrial class kappa enzyme rGSTK1-1. *Biochemistry*, **43**(2), 352–61.

Lakowicz, J. R. (2006). *Principles of fluorescence spectroscopy*. *Principles of Fluorescence Spectroscopy*. doi:10.1007/978-0-387-46312-4

Lakowicz, J. R., and Masters, B. R. (2008). Principles of Fluorescence Spectroscopy, Third Edition. *Journal of Biomedical Optics*, **13**(2), 29901.

Lakshmikanth, G. S., Sridevi, K., Krishnamoorthy, G., and Udgaonkar, J. B. (2001). Structure is lost incrementally during the unfolding of barstar. *Nature Structural Biology*, **8**(9), 799–804.

Lapidus, L. J. (2017). Protein unfolding mechanisms and their effects on folding experiments. *F1000Research*, **6**(0), 1723.

Larsen, T. A., Olson, A. J., and Goodsell, D. S. (1998). Morphology of protein–protein interfaces. *Structure*, **6**(4), 421–427.

Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*, **65**, 44–45.

Liang, Z. X., Tsigos, I., Lee, T., Bouriotis, V., Resing, K. A., Ahn, N. G., and Klinman, J. P. (2004). Evidence for increased local flexibility in psychrophilic alcohol dehydrogenase relative to its thermophilic homologue. *Biochemistry*, **43**(46), 14676–14683.

Lim, S. A., and Marqusee, S. (2017). The burst-phase folding intermediate of ribonuclease H changes conformation over evolutionary history. *Biopolymers*, (August), e23086.

Lipin, D. I., Lua, L. H. L., and Middelberg, A. P. J. (2008). Quaternary size distribution of soluble aggregates of glutathione-S-transferase-purified viral protein as determined by asymmetrical flow field flow fractionation and dynamic light scattering. *Journal of Chromatography. A*, **1190**(1–2), 204–14.

Loftus, D., Gbenle, G. O., Kim, P. S., and Baldwin, R. L. (1986). Effects of denaturants on amide proton exchange rates: a test for structure in protein fragments and folding intermediates. *Biochemistry*, **25**(6), 1428–36.

Luo, J., Hornby, J. A. T., Wallace, L., Chen, J., Armstrong, R. N., and Dirr, H. W. (2002). Impact of domain interchange on conformational stability and equilibrium folding of chimeric class micro glutathione transferases. *Protein Science : A Publication of the Protein Society*, **11**(9), 2208–17.

Main, E. R., Fulton, K. F., and Jackson, S. E. (1999). Folding pathway of FKBP12 and characterisation of the transition state. *J. Mol. Biol.*, **291**(2), 429–444.

Malhotra, P., and Udgaonkar, J. B. (2015). Tuning Cooperativity on the Free Energy Landscape of Protein Folding. *Biochemistry*, **54**(22), 3431–3441.

Malhotra, P., and Udgaonkar, J. B. (2016a). How cooperative are protein folding and unfolding transitions? *Protein Science*, **25**(11), 1924–1941.

Malhotra, P., and Udgaonkar, J. B. (2016b). Secondary Structural Change Can Occur Diffusely and Not Modularly during Protein Folding and Unfolding Reactions. *Journal of the American Chemical Society*, **138**(18), 5866–5878.

Mannervik, B., Awasthi, Y. C., Board, P. G., Hayes, J. D., Di Ilio, C., Ketterer, B., Listowsky, I., Morgenstern, R., Muramatsu, M., Pearson, W. R., Pickett, C. B., Sato, K., Widersten, M., and Wolf, C. R. (1992). Nomenclature for human glutathione transferases. *Biochemical Journal*, **282**(1), 305–306.

Mannervik, B., and Danielson, U. H. (1988). Glutathione transferases--structure and catalytic activity. *CRC Critical Reviews in Biochemistry*, **23**(3), 283–337.

Mannervik, B., Danielson, U. H., and Ketterer, B. (1988). Glutathione Transferases - Structure and Catalytic Activity. *Critical Reviews in Biochemistry and Molecular Biology*, **23**(3), 283–337.

Marcotte, E. M., Pellegrini, M., Yeates, T. O., and Eisenberg, D. (1999). A census of protein repeats. *Journal of Molecular Biology*, **293**(1), 151–160.

Marsh, J. A., Hernández, H., Hall, Z., Ahnert, S. E., Perica, T., Robinson, C. V., and Teichmann, S. A. (2013). Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell*, **153**(2), 461–470.

Martin, J. L. (1995). Thioredoxin -a fold for all reasons. *Structure*, **3**(3), 245–250.

Matouschek, A. (2003). Protein unfolding — an important process in vivo? *Current Opinion in Structural Biology*, **13**(1), 98–109.

Matouschek, A., and Fersht, A. R. (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. *Proceedings of the National Academy of Sciences*, **90**(16), 7814–7818.

Matthews, C. R. (1987). Effect of Point Mutations of the Folding of Globular Proteins. *Methods in Enzymology*, **154**(C), 498–511.

McCammon, J. A. (1998). Theory of biomolecular recognition. *Current Opinion in Structural Biology*, **8**(2), 245–249.

McIlwain, C. C., Townsend, D. M., and Tew, K. D. (2006). Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene*, **25**(11), 1639–1648.

Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M., and Ketterer, B. (1991). Theta, a new class of glutathione transferases purified from rat and man. *The Biochemical Journal*, **274**, 409–14.

Meyer, M., Wilson, P., and Schomburg, D. (1996). Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking. *Journal of Molecular Biology*, **264**(1), 199–210.

Miranker, A. D., Robinson, C. V., Radford, S. E., Aplin, R. T., and Dobson, C. M. (1993). Detection of transient protein folding populations by mass spectrometry. *Science (New York, N.Y.)*, **262**(5135), 896–900.

Miranker, A. D., Robinson, C. V., Radford, S. E., and Dobson, C. M. (1996). Investigation of protein folding by mass spectrometry. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, **10**(1), 93–101.

Misra, G. (2017). *Introduction to Biomolecular Structure and Biophysics*. (G. Misra, Ed.)*Introduction to Biomolecular Structure and Biophysics: Basics of Biophysics*, Singapore: Springer Singapore. doi:10.1007/978-981-10-4968-2

Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, **12**(1), 88–118.

Mosca, R., Céol, A., Stein, A., Olivella, R., and Aloy, P. (2014). 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Research*, **42**(Database issue), D374-9.

Moulick, R., and Udgaonkar, J. B. (2017). Identification and Structural Characterization of the Precursor Conformation of the Prion Protein which Directly Initiates Misfolding and Oligomerization. *Journal of Molecular Biology*, **429**(6), 886–899.

Myers, J. K. K., Pace, C. N., and Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science*, **4**(10), 2138–48.

Neet, K. E., and Timm, D. E. (1994). Conformational stability of dimeric proteins: Quantitative studies by equilibrium denaturation. *Protein Science*, **3**(12), 2167–2174.

Neupane, K., Foster, D. A. N., Dee, D. R., Yu, H., Wang, F., and Woodside, M. T. (2016). Direct observation of transition paths during the folding of proteins and nucleic acids. *Science*, **352**(6282), 239–242.

Nickson, A. A., and Clarke, J. (2010). What lessons can be learned from studying the folding of homologous proteins? *Methods*, **52**(1), 38–50.

Nooren, I. M. A., and Thornton, J. M. (2003). Diversity of protein-protein interactions. *The EMBO Journal*, **22**(14), 3486–92.

Novak, W. R. P. (2014). Tertiary Structure Domains, Folds, and Motifs. In E. Bell, ed., *Molecular Life Sciences*, New York, NY: Springer New York, pp. 1–5.

Oka, O. B. V, and Bulleid, N. J. (2013). Forming disulfides in the endoplasmic reticulum. *Biochimica et Biophysica Acta - Molecular Cell Research*, **1833**(11), 2425–2429.

Oliveberg, M. (1998). Alternative Explanations for "Multistate" Kinetics in Protein Folding: Transient Aggregation and Changing Transition-State Ensembles †. *Accounts of Chemical Research*, **31**(11), 765–772.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in Enzymology*, **131**(July 1985), 266–80.

Pace, C. N. (1990). Conformational stability of globular proteins. *Trends in Biochemical Sciences*, **15**(1), 14–17.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Science : A Publication of the Protein Society*, **4**(11), 2411–23.

Pajaud, J., Kumar, S., Rauch, C., Morel, F., and Aninat, C. (2012). Regulation of Signal Transduction by Glutathione Transferases. *International Journal of Hepatology*, **2012**, 1–11.

Pan, H., Raza, A. S., and Smith, D. L. (2004). Equilibrium and Kinetic Folding of Rabbit Muscle Triosephosphate Isomerase by Hydrogen Exchange Mass Spectrometry. *Journal of Molecular Biology*, **336**(5), 1251–1263.

Pan, H., and Smith, D. L. (2004). Amide hydrogen exchange/mass spectrometry applied to cooperative protein folding: equilibrium unfolding of Staphylococcus aureus aldolase. *Methods in Enzymology*, **380**(1966), 285–308.

Pande, V. S., and Rokhsar, D. S. (1998). Is the molten globule a third phase of proteins? *Proceedings of the National Academy of Sciences of the United States of America*, **95**(4), 1490–4.

Parbhoo, N., Stoychev, S. H., Fanucchi, S., Achilonu, I., Adamson, R. J., Fernandes, M., Gildenhuys, S., and Dirr, H. W. (2011). A conserved interdomain interaction is a determinant of folding cooperativity in the GST fold. *Biochemistry*, **50**(32), 7067–7075.

Pellegrini, M., Renda, M. E., and Vecchio, A. (2012). Ab initio detection of fuzzy amino acid tandem repeats in protein sequences. *BMC Bioinformatics*, **13**(SUPPL.3). doi:10.1186/1471-2105-13-S3-S8

Pemble, S. E., Wardle, A. F., and Taylor, J. B. (1996). Glutathione S-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue. *The Biochemical Journal*, **319 (Pt 3**, 749–754.

Perchiacca, J. M., Ladiwala, A. R. A., Bhattacharya, M., and Tessier, P. M. (2012). Structure-based design of conformation- and sequence-specific antibodies against amyloid . *Proceedings of the National Academy of Sciences*, **109**(1), 84–89.

Pettigrew, N. E., and Colman, R. F. (2001). Heterodimers of glutathione S-transferase can form between isoenzyme classes pi and mu. *Archives of Biochemistry and Biophysics*, **396**(2), 225–230.

Phillips, C. M., Mizutani, Y., and Hochstrasser, R. M. (1995). Ultrafast thermally induced unfolding of RNase A. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(16), 7292–6.

Piromjitpong, J., Wongsantichon, J., and Ketterman, A. J. (2007). Differences in the subunit interface residues of alternatively spliced glutathione transferases affects catalytic and structural functions. *Biochemical Journal*, **401**(3), 635–644.

Plaxco, K. W., and Gross, M. (2001). Unfolded, yes, but random? Never! *Nature Structural Biology*, **8**(8), 659–60.

Plotkin, S. S., and Onuchic, J. N. (2002a). Understanding protein folding with energy landscape theory Part I: Basic concepts. *Quarterly Reviews of Biophysics*, **35**(2), 111–167.

Plotkin, S. S., and Onuchic, J. N. (2002b). Understanding protein folding with energy landscape theory Part II: Quantitative aspects. *Quarterly Reviews of Biophysics*, **35**(3), 111–167.

Povarova, O. I., Kuznetsova, I. M., and Turoverov, K. K. (2010). Differences in the Pathways of Proteins Unfolding Induced by Urea and Guanidine Hydrochloride: Molten Globule State and Aggregates. *PLoS ONE*, **5**(11), e15035.

Privalov, P. L. (1979). Stability of Proteins Small Globular Proteins. In *Advances in Protein Chemistry*, Vol. 33, pp. 167–241.

Privalov, P. L., and Potekhin, S. A. (1986). Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods in Enzymology*, **131**(2), 4–51.

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I. (1990). Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Letters*, **262**(1), 20–24.

Ptitsyn, O. B., and Rashin, A. A. (1975). A model of myoglobin self-organization. *Biophysical Chemistry*, **3**(1), 1–20.

Puri, S., and Chaudhuri, T. K. (2017). Folding and unfolding pathway of chaperonin GroEL monomer and elucidation of thermodynamic parameters. *International Journal of Biological Macromolecules*, **96**, 713–726.

Ragone, R. (2000). How the protein concentration affects unfolding curves of oligomers. *Biopolymers*, **53**(3), 221–5.

Rami, B. R., and Udgaonkar, J. B. (2001). pH-jump-induced folding and unfolding studies of barstar: Evidence for multiple folding and unfolding pathways. *Biochemistry*, **40**(50), 15267–15279.

Raschke, T. M., and Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Structural Biology*, **4**(4), 298–304.

Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998). Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *Journal of Molecular Biology*, **279**(2), 449–460.

Reinemer, P., Dirr, H. W., Lo Bello, M., Rome, T. V., Vincent, S., and Parker, W. (1992). Three-dimensional Structure of Class n : Glutathione S-Transferase from Human Placetta in Complex wit at 243 A Resolution. *Cycle*, 214–226.

Richardson, J. S. (1981). The Anatomy and Taxonomy of Protein Structure. In *Structure*, pp. 167–339.

Ritchie, D. B., and Woodside, M. T. (2015). Probing the structural dynamics of proteins and nucleic acids with optical tweezers. *Current Opinion in Structural Biology*, **34**, 43–51.

Roder, H., Maki, K., Latypov, R. F., Cheng, H., and Shastry, M. C. R. (2008). Early Events in Protein Folding Explored by Rapid Mixing Methods. *Protein Folding Handbook*, **1**, 491–535.

Rojsajjakul, T., Wintrode, P. L., Vadrevu, R., Robert Matthews, C., and Smith, D. L. (2004). Multi-state Unfolding of the Alpha Subunit of Tryptophan Synthase, a TIM Barrel Protein: Insights into the Secondary Structure of the Stable Equilibrium Intermediates by Hydrogen Exchange Mass Spectrometry. *Journal of Molecular Biology*, **341**(1), 241–253.

Rose, G. D., Fleming, P. J., Banavar, J. R., and Maritan, A. (2006). A backbone-based theory of protein folding. *Proceedings of the National Academy of Sciences*, **103**(45), 16623–16633.

Ross, C. A., and Poirier, M. A. (2004). Protein aggregation and neurodegenerative disease. *Nature Medicine*, **10 Suppl**(1078–8956 (Print)), S10-7.

Rossjohn, J., McKinstry, W. J., Oakley, a J., Verger, D., Flanagan, J., Chelvanayagam, G., Tan, K. L., Board, P. G., and Parker, M. W. (1998). Human theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site. *Structure (London, England : 1993)*, **6**, 309–322.

Rowe, E. S., and Tanford, C. (1973). Equilibrium and kinetics of the denaturation of a homogeneous human immunoglobulin light chain. *Biochemistry*, **12**(24), 4822–7.

Rumfeldt, J. A. O. O., Galvagnion, C., Vassall, K. A., and Meiering, E. M. (2008). Conformational stability and folding mechanisms of dimeric proteins. *Progress in Biophysics and Molecular Biology*, **98**(1), 61–84.

Sacchetta, P., Aceto, A., Bucciarelli, T., Dragani, B., Santarone, S., Allocati, N., and Di Ilio, C. (1993). Multiphasic denaturation of glutathione transferase B1-1 by guanidinium chloride. Role of the dimeric structure on the flexibility of the active site. *European Journal of Biochemistry*, **215**(3), 741–5.

Sacchetta, P., Pennelli, A., Bucciarelli, T., Cornelio, L., Amicarelli, F., Miranda, M., and Di Ilio, C. (1999). Multiple Unfolded States of Glutathione Transferase bbGSTP1-1 by Guanidinium Chloride. *Archives of Biochemistry and Biophysics*, **369**(1), 100–106.

Sali, A., Shakhnovich, E. I., and Karplus, M. (1994). How does a protein fold? *Nature*, **369**(6477), 248–251.

Sayed, Y., Wallace, L., and Dirr, H. W. (2000). The hydrophobic lock-and-key intersubunit motif of glutathione transferase A1-1: Implications for catalysis, ligandin function and stability. *FEBS Letters*, **465**(2–3), 169–172.

Schüler, A., and Bornberg-Bauer, E. (2016). Evolution of protein domain repeats in metazoa. *Molecular Biology and Evolution*, **33**(12), 3170–3182.

Schuler, B. (2007). Application of single molecule Förster resonance energy transfer to protein folding. *Methods in Molecular Biology (Clifton, N.J.)*, **350**, 115–38.

Seelig, J. (2018). Cooperative protein unfolding. A statistical-mechanical model for the action of denaturants. *Biophysical Chemistry*, **233**(October 2017), 19–25.

Segawa, S., and Sugihara, M. (1984). Characterization of the transition state of lysozyme unfolding. I. Effect of protein-solvent interactions on the transition state. *Biopolymers*, **23**(11 Pt 2), 2473–88.

Shakhnovich, E. I. (1997). Theoretical studies of protein-folding thermodynamics and kinetics. *Current Opinion in Structural Biology*, **7**(1), 29–40.

Shakhnovich, E. I., and Finkelstein, A. V. (1989). Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers*, **28**(10), 1667–80.

Sheehan, D., Meade, G., Foley, V. M., and Dowd, C. A. (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal*, **360**(Pt 1), 1–16.

Shi, J., Nobrega, R. P., Schwantes, C., Kathuria, S. V., Bilsel, O., Matthews, C. R., Lane, T. J., and Pande, V. S. (2017). Atomistic structural ensemble refinement reveals non-native structure stabilizes a sub-millisecond folding intermediate of CheY. *Scientific Reports*, **7**(May), 44116.

Shimizu, Y., and Hendershot, L. M. (2009). Oxidative Folding: Cellular Strategies for Dealing with the Resultant Equimolar Production of Reactive Oxygen Species. *Antioxidants & Redox Signaling*, **11**(9), 2317–2331.

Shortle, D. (2001). Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea. *Science*, **293**(5529), 487–489.

Sidell, K. R., Montine, K. S., Picklo, M. J., Olsen, S. J., Amarnath, V., and Montine, T. J. (2003). Mercapturate metabolism of 4-hydroxy-2-nonenal in rat and human cerebrum. *Journal of Neuropathology and Experimental Neurology*, **62**(2), 146–153.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**(539). doi:10.1038/msb.2011.75

Silow, M., and Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proceedings of the National Academy of Sciences of the United States of America*, **94**(12), 6084–6.

Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., Board, P. G., Olin, B., Mannervik, B., and Jones, T. A. (1993). Structure Determination and Refinement of Human Alpha Class Glutathione Transferase A1-1, and a Comparison with the Mu and Pi Class Enzymes. *Journal of Molecular Biology*, **232**(1), 192–212.

Sluis-Cremer, N., Naidoo, N., and Dirr, H. W. (1996). Class-pi glutathione S-transferase is unable to regain its native conformation after oxidative inactivation by hydrogen peroxide. *European Journal of Biochemistry / FEBS*, **242**(2), 301–7.

Soulages, J. L. (1998). Chemical denaturation: potential impact of undetected intermediates in the free energy of unfolding and m-values obtained from a two-state assumption. *Biophysical Journal*, **75**(1), 484–92.

Stenberg, G., Abdalla, A. M., and Mannervik, B. (2000). Tyrosine 50 at the subunit interface of dimeric human glutathione transferase P1-1 is a structural key residue for modulating protein stability and catalytic function. *Biochemical and Biophysical Research Communications*, **271**(1), 59–63.

Stevens, J. M., Hornby, J. A. T., Armstrong, R. N., and Dirr, H. W. (1998). Class Sigma Glutathione Transferase Unfolds via a Dimeric and a Monomeric Intermediate: Impact of Subunit Interface on Conformational Stability in the Superfamily †. *Biochemistry*, **37**(44), 15534–15541.

Stoychev, S. H., Nathaniel, C., Fanucchi, S., Brock, M., Li, S., Asmus, K., Woods, V. L., and Dirr, H. W. (2009). Structural dynamics of soluble chloride intracellular channel protein CLIC1 examined by amide hydrogen-deuterium exchange mass spectrometry. *Biochemistry*, **48**(35), 8413–8421.

Strange, R. C., Jones, P. W., and Fryer, A. A. (2000). Glutathione S-transferase: genetics and role in toxicology. *Toxicology Letters*, **112**–**113**, 357–63.

Tanford, C. (1970). Protein denaturation: Part C. Theoretical models for the mechanism of denaturation. *Advances in Protein Chemistry*, **24**(C), 1–95.

Thompson, L. C., Walters, J., Burke, J. P., Parsons, J. F., Armstrong, R. N., and Dirr, H. W. (2006). Double Mutation at the Subunit Interface of Glutathione Transferase rGSTM1-1 Results in a Stable, Folded Monomer †. *Biochemistry*, **45**(7), 2267–2273.

Tonomura, B., Nakatani, H., Ohnishi, M., Yamaguchi-Ito, J., and Hiromi, K. (1978). Test reactions for a stopped-flow apparatus. *Analytical Biochemistry*, **84**(2), 370–383.

Tsai, C.-J., Xu, D., and Nussinov, R. (1997). Structural motifs at protein-protein interfaces: Protein cores versus two-state and three-state model complexes. *Protein Science*, **6**(9), 1793–1805.

Tu, B. P., and Weissman, J. S. (2004). Oxidative protein folding in eukaryotes: mechanisms and consequences. *The Journal of Cell Biology*, **164**(3), 341–6.

Utiyama, H., and Baldwin, R. L. (1986). Kinetic mechanisms of protein folding. *Methods in Enzymology*, **131**(October), 51–70.

Valdar, W. S., and Thornton, J. M. (2001). Protein-protein interfaces: analysis of amino acid conservation in homodimers. *Proteins*, **42**(1), 108–24.

Vallée-Bélisle, A., and Michnick, S. W. (2012). Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis. *Nature Structural & Molecular Biology*, **19**(7), 731–6.

Vargo, M. A., Nguyen, L., and Colman, R. F. (2004). Subunit Interface Residues of Glutathione S -Transferase A1-1 that Are Important in the Monomer−Dimer Equilibrium †. *Biochemistry*, **43**(12), 3327–3335.

Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., and Teichmann, S. A. (2004). Structure, function and evolution of multidomain proteins. *Current Opinion in Structural Biology*, **14**(2), 208–16.

Walkenhorst, W. F., Green, S. M., and Roder, H. (1997). Kinetic Evidence for Folding and Unfolding Intermediates in Staphylococcal Nuclease †. *Biochemistry*, **36**(19), 5795–5805.

Wallace, L., Blatch, G. L., and Dirr, H. W. (1998a). A topologically conserved aliphatic residue in alpha-helix 6 stabilizes the hydrophobic core in domain II of glutathione transferases and is a structural determinant for the unfolding pathway. *The Biochemical Journal*, **336 (Pt 2**, 413–8.

Wallace, L., and Matthews, C. R. (2002). Sequential vs. parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophysical Chemistry*, **101**–**102**, 113–131.

Wallace, L., Sluis-Cremer, N., and Dirr, H. W. (1998b). Equilibrium and kinetic unfolding properties of dimeric human glutathione transferase A1-1. *Biochemistry*, **37**(15), 5320–5328.

Walters, J., Milam, S. L., and Clark, A. C. (2009). Practical approaches to protein folding and assembly. *Methods in Enzymology*, **455**(8), 1–39.

Weber, G., and Young, L. (1964). Fragmentation of Bovine Serum Albumin by Pepsin. I. the Origin of the Acid Expansion of the Albumin Molecule. *The Journal of Biological Chemistry*, **239**(5), 1415–23.

Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M., and Eyck, L. F. (2006). Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *Journal of the American Society for Mass Spectrometry*, **17**(11), 1498–1509.

Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **70**(3), 697–701.

Wilce, M. C. J., and Parker, M. W. (1994). Structure and function of glutathione S-transferases. *Biochimica et Biophysica Acta*, **1205**(1), 1–18.

Wingfield, P. T. (1995). Use of Protein Folding Reagents. In *Current Protocols in Protein Science*, Vol. Appendix 3, Hoboken, NJ, USA: John Wiley & Sons, Inc., p. A.3A.1-A.3A.4.

Wodak, S. J., Malevanets, A., and MacKinnon, S. S. (2015). The Landscape of Intertwined Associations in Homooligomeric Proteins. *Biophysical Journal*, **109**(6), 1087–1100.

Woods, L. A., Platt, G. W., Hellewell, A. L., Hewitt, E. W., Homans, S. W., Ashcroft, A. E., and Radford, S. E. (2012). UKPMC Funders Group amyloid-forming protein, **7**(10), 730–739.

Woody, R. W. (1995). Circular dichroism. *Methods in Enzymology*, **246**, 34–71.

Wright, C. F., Steward, A., and Clarke, J. (2004). Thermodynamic characterisation of two transition states along parallel protein folding pathways. *Journal of Molecular Biology*, **338**(3), 445–451.

Xu, Y., Mayne, L. C., and Englander, W. (1998). Evidence for an unfolding and refolding pathway in cytochrome c. *Nature Structural Biology*, **5**(9), 774–778.

Yang, H., and Smith, D. L. (1997). Kinetics of cytochrome c folding examined by hydrogen exchange and mass spectrometry. *Biochemistry*, **36**(48), 14992–9.

Yeh, Y.-Q., Liao, K.-F., Shih, O., Shiu, Y.-J., Wu, W.-R., Su, C.-J., Lin, P.-C., and Jeng, U.-S. (2017). Probing the Acid-Induced Packing Structure Changes of the Molten Globule Domains of a Protein near Equilibrium Unfolding. *The Journal of Physical Chemistry Letters*, **8**(2), 470–477.

Yu, H., Dee, D. R., Liu, X., Brigley, A. M., Sosova, I., and Woodside, M. T. (2015). Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(27), 8308–13.

Zaidi, F. N., Nath, U., and Udgaonkar, J. B. (1997). Multiple intermediates and transition states during protein unfolding. *Nature Structural Biology*, **4**(12), 1016–24.

Zhang, Z., Post, C. B., and Smith, D. L. (1996). Amide hydrogen exchange determined by mass spectrometry: application to rabbit muscle aldolase. *Biochemistry*, **35**(3), 779–91.

Zhang, Z., and Smith, D. L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Science : A Publication of the Protein Society*, **2**(4), 522–31.

Zhu, Z. Y., and Karlin, S. (1996). Clusters of charged residues in protein three-dimensional structures. *Proceedings of the National Academy of Sciences*, **93**(16), 8350–8355.

Žoldák, G., Jancura, D., and Sedlák, E. (2017). The fluorescence intensities ratio is not a reliable parameter for evaluation of protein unfolding transitions. *Protein Science*, **26**(6), 1236–1239.

APPENDIX

List of figures

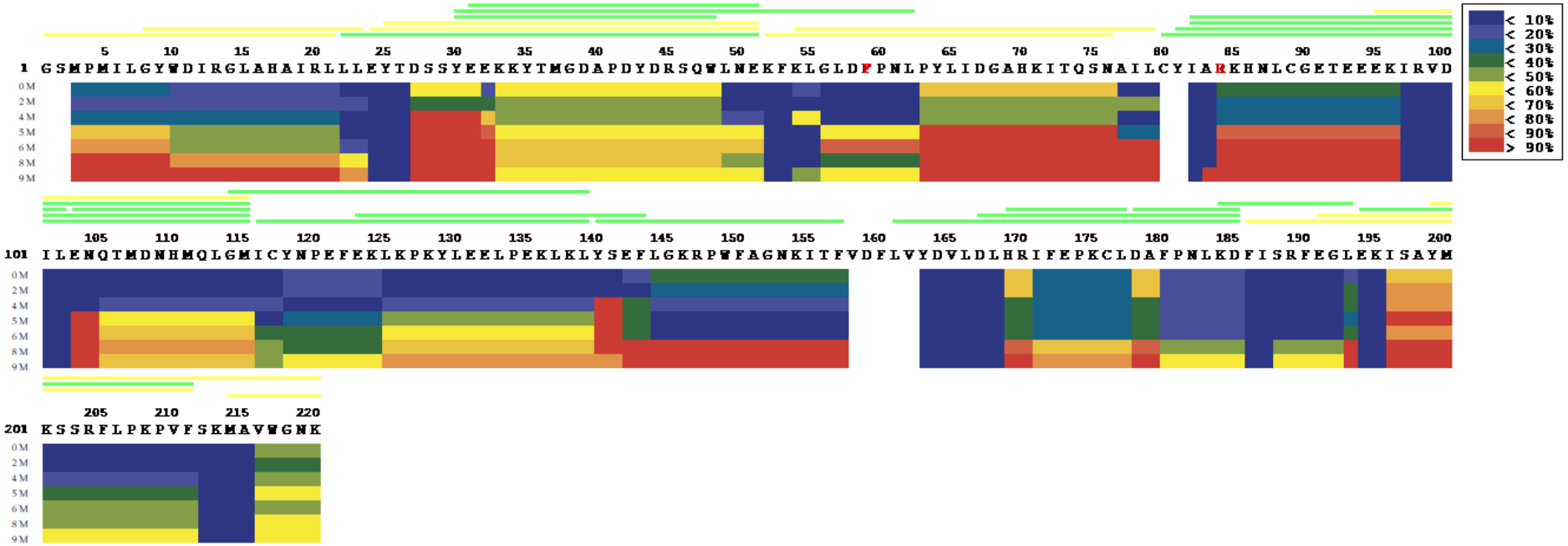
APPENDIX Figure A: Peptide heat map for the WT dimer.....132
APPENDIX Figure B: Peptide heat map for the mutant monomer.....133

List of tables

APPENDIX Table A: Percentage recovery of the WT dimer and F56S/R81A mutant monomer.....131

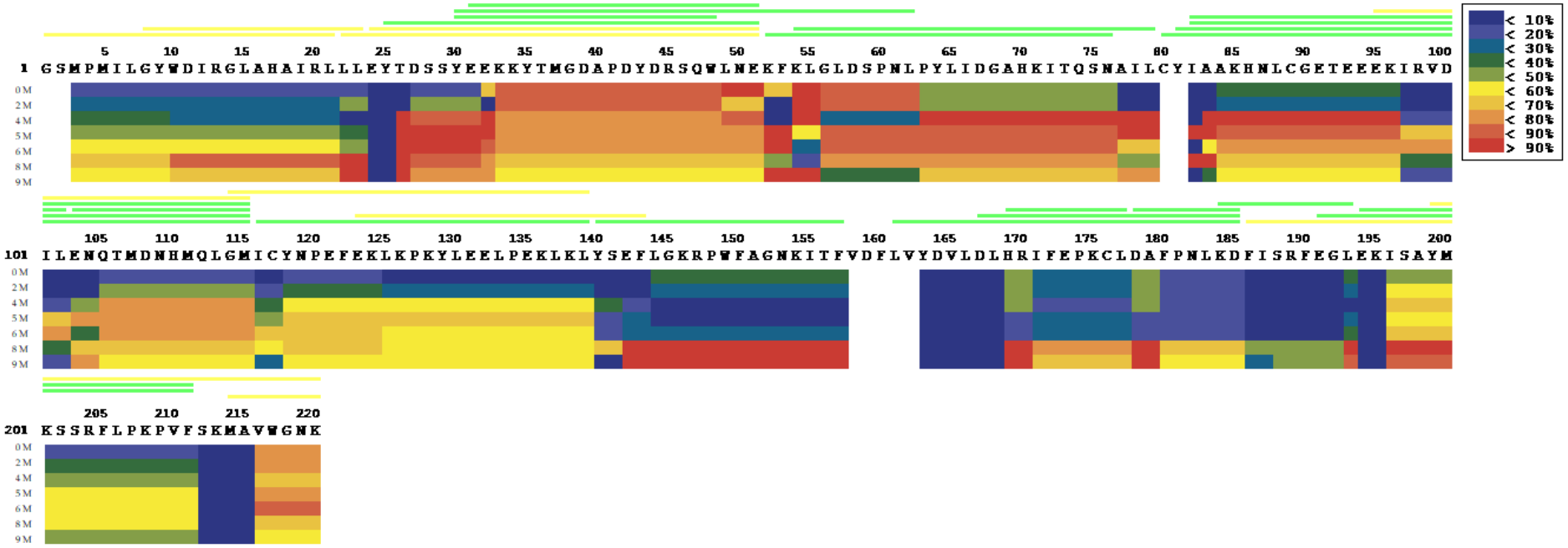
APPENDIX Table A: Percentage recovery of the WT dimer and F56S/R81A mutant monomer

Ellipticity at 222 nm			
	Native protein	Refolded protein	% Recovery
WT dimer	- 511.7	- 510.6	99.8
Mutant monomer	- 461.3	- 453.3	98.3
Fluorescence intensity at 340 nm			
	Native protein	Refolded protein	% Recovery
WT dimer	200.2	209.1	104.4
Mutant monomer	219.5	230.1	104.8



APPENDIX Figure A: Peptide heat map for the WT dimer

The peptides (green and yellow) were analysed using HD Examiner 1.3 (Sierra Analytics, LLC, Modesto,CA). Each peptide is mapped to the corresponding amino acid sequence that it covers. The percentage deuteration at various urea concentrations is coloured according to the key (insert). The thrombin cleavage site results in the addition of two amino acids (glycine and serine) at the N-terminus of the protein. These two amino acids have been accounted for in the sequence and numbering. The F56 residue and the R81 residue are highlighted in red.



APPENDIX Figure B: Peptide heat map for the mutant monomer

The peptides (green and yellow) were analysed using HD Examiner 1.3 (Sierra Analytics, LLC, Modesto,CA). Each peptide is mapped to the corresponding amino acid sequence that it covers. The percentage deuteration at various urea concentrations is coloured according to the key (insert). The thrombin cleavage site results in the addition of two amino acids (glycine and serine) at the N-terminus of the protein. These two amino acids have been accounted for in the sequence and numbering.