

# **Structural, Ligand Binding and Redox Studies of hPDI fragments using NMR Spectroscopy**

Kirsty Louise Richards

PhD Biochemistry 2015

A thesis submitted to the University of Kent for the degree of PhD in Biochemistry at  
the School of Biosciences, Faculty of Science, Technology and Medicine

University of  
**Kent**

## **Declaration**

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or institute of learning.

Kirsty Louise Richards

September 2015

## Acknowledgements

I would like to begin by thanking my supervisors **Dr. Richard Williamson** and **Dr. Mark Howard**, for all of their help and guidance throughout my PhD and thesis writing. My thanks also go to **Dr. Michelle Rowe** for help with NMR data acquisition and analysis, and **Mr. Kevin Howland** of the Protein Research Facility for help with mass spectrometry and RP-HPLC analysis. I would also like to extend my gratitude to **Mr. Paul Hudson** for his preliminary  $^{19}\text{F}$  peptide work that I have continued to develop upon within this thesis.

My special thanks also go to the past members of our **NMR group**, especially **Dr. Jane Wagstaff**, for demonstrating such a high standard of practice within the lab, to **Drs. Denisa Doko** and **Sam Lodge**, **Mr. Nathan Cross** and **Miss. Robyn Tucker**, for our productive discussions and seemingly endless cups of tea and coffee. Lastly **Miss. Rose Curtis-Marof**, whom I am very lucky to call my friend- Rose, you and James have helped me more than you will ever know!

I would also like to thank my mad, amazing *E. colins*. You lot really are the best (and not just because of all the biscuits and cake).

Finally, I would like to express the warmest thanks and gratitude to my family and friends, especially to my wonderful parents **Karen & Dan**, my little brother **Craig**, my best friend **Claire** and my dear **Rob**, for their continued and unwavering love and support.

Thank you, I couldn't have done it without you all,

Kirsty



“No one likes to fall. And few people would ever choose to drown. But in struggling through the ocean of this life, sometimes it’s so hard not to let the world in. Sometimes the ocean does enter us. It shatters the boat, and you sink down into the depths of the sea.

On the ocean floor you may feel as though you are at your lowest point. You may feel broken, surrounded by darkness. But that dark place isn’t the end. Sometimes the ocean floor is only a stop in the journey, and it is when you are at this lowest point that you are faced with a choice. You can stay there at the bottom, until you drown. Or you can gather pearls and rise back up—stronger from the swim, and richer from the jewels.”

**Yasmin Mogahed**

## Contents

A. Acknowledgements .....	i
B. Contents .....	iii
C. Abbreviations .....	x
D. Abstract .....	xiii

## Chapter 1: Introduction

1.1 Protein Folding in the ER .....	1
1.2 An Overview of hPDI .....	3
1.2.1 <i>hPDI Structure</i> .....	4
1.2.2 <i>Ligand Binding</i> .....	9
1.2.3 <i>Catalytic Activity</i> .....	12
1.2.4 <i>Re-oxidation of hPDI by Ero1<math>\alpha</math></i> .....	15
1.2.5 <i>Chaperone Activity</i> .....	17
1.3 PDI Family Members .....	20
1.3.1 <i>PDIp (PDIA2)</i> .....	21
1.3.2 <i>ERp57 (PDIA3)</i> .....	22
1.3.3 <i>ERp27 (PDIA8)</i> .....	23
1.3.4 <i>PDILT (PDIA7)</i> .....	24
1.3.5 <i>ERp44 (PDIA10)</i> .....	25
1.4 Implication of hPDI in Disease States .....	25
1.5 An Introduction to Biomolecular NMR .....	26
1.6 Project Aims .....	28
1.7 Project Impact .....	29

## Chapter 2: Expression and Purification of hPDI Binding Peptide $\Delta$ -som

2.1 Introduction .....	31
2.1.1 <i>Recombinant Peptide Fusion Strategies</i> .....	32

2.1.1.1 <i>Insoluble Fusion Method (KSI)</i> .....	32
2.1.1.2 <i>Solubility Enhancing Fusion Method (GB1)</i> .....	34
<b>2.2 Materials and Methods</b> .....	<b>35</b>
2.2.1 <i>Materials</i> .....	36
2.2.2 <i>E. coli strains and Vectors</i> .....	36
2.2.3 <i>Bacterial Culture Media</i> .....	36
2.2.4 <i>Competent Cells</i> .....	36
2.2.5 <i>Transformations</i> .....	37
2.2.6 <i>Fusion Protein Expression</i> .....	37
2.2.7 <i>SDS-PAGE Analysis</i> .....	38
2.2.8 <i>Cell Lysis</i> .....	41
2.2.9 <i>KSI Fusion Protein Isolation</i> .....	41
2.2.10 <i>KSI Fusion Protein Cleavage and Peptide Purification</i> .....	41
2.2.11 <i>GB1 Fusion Protein IMAC Purification and Cleavage</i> .....	43
2.2.12 <i>Peptide Analysis by Mass Spectrometry</i> .....	43
<b>2.3 Results</b> .....	<b>44</b>
2.3.1 <i>Expression and Purification of <math>\Delta</math>-som from KSI Fusions</i> .....	44
2.3.2 <i>Expression and Purification of <math>\Delta</math>-som from GB1 Fusions</i> .....	51
<b>2.4 Discussion</b> .....	<b>57</b>
2.4.1 <i>KSI Peptide Fusions</i> .....	57
2.4.2 <i>GB1 Peptide Fusions</i> .....	58
2.4.3 <i>Overall Summary</i> .....	59
 <b>Chapter 3: Expression and Purification of hPDI b'x and W111F abb'x</b>	
<b>3.1 Introduction</b> .....	<b>61</b>
3.1.1 <i>E. coli strains and Vectors</i> .....	62
3.1.2 <i>Protein Expression and Purification</i> .....	63

<b>3.2 Materials and Methods</b> .....	63
3.2.1 <i>Glycerol Stocks</i> .....	63
3.2.2 <i>Bacterial Culture Media</i> .....	64
3.2.3 <i>Recombinant Protein Expression</i> .....	65
3.2.4 <i>Cell Lysis</i> .....	65
3.2.5 <i>IMAC Chromatography</i> .....	66
3.2.6 <i>Dialysis</i> .....	66
3.2.7 <i>Anion Exchange Chromatography</i> .....	67
3.2.8 <i>Size Exclusion Chromatography</i> .....	67
3.2.9 <i>Gels: Reducing SDS-PAGE and Native-PAGE</i> .....	67
3.2.10 <i>Concentration Determination</i> .....	68
3.2.11 <i>Mass Determination and Isotope Incorporation</i> .....	68
<b>3.3 Results</b> .....	68
3.3.1 <i>Expression of <b>b</b>'x and W111F <b>abb</b>'x</i> .....	68
3.3.2 <i>IMAC Chromatography</i> .....	68
3.3.3 <i>Anion Exchange Chromatography</i> .....	70
3.3.4 <i>Size Exclusion Chromatography</i> .....	71
3.3.5 <i>Mass Determination and Isotopic Incorporation by Mass Spectrometry</i> ....	75
3.3.6 <i>Final Yields</i> .....	76
<b>3.4 Discussion</b> .....	78

## **Chapter 4: Assignment of hPDI *b*'x and W111F *abb*'x by NMR Spectroscopy**

<b>4.1 Introduction</b> .....	81
4.1.1 <sup>15</sup> N, <sup>1</sup> H-HSQC Experiments.....	82
4.1.2 <sup>13</sup> C, <sup>1</sup> H-HSQC Experiments.....	82
4.1.3 <i>Chemical Shift Perturbation Mapping</i> .....	83
4.1.4 <i>Sequential Backbone Assignment of <b>b</b>'x and W111F <b>abb</b>'x by Triple Resonance</i> .....	84

4.1.5	<i>Secondary Structure Prediction</i>	86
4.1.6	<i>Sequential Side-Chain Assignment: HCCH TOCSY</i>	87
4.2	<b>Materials and Methods</b>	88
4.2.1	<i>NMR Sample Preparation</i>	88
4.2.2	<i>NMR Data Acquisition and Processing</i>	89
4.2.2.1	<i><math>^{15}\text{N}, ^1\text{H}</math>-HSQC Experiments of <b>b'x</b> and W111F <b>abb'x</b></i>	89
4.2.2.2	<i><math>^{13}\text{C}, ^1\text{H}</math>-HSQC Experiments of <b>b'x</b></i>	90
4.2.2.3	<i>Sequential Backbone Assignment of <b>b'x</b> and W111F <b>abb'x</b></i>	90
4.2.2.4	<i>Sequential Side-Chain Assignment of <b>b'x</b></i>	90
4.2.3	<i>NMR Data Analysis</i>	90
4.2.3.1	<i>Triple Resonance Backbone and Side-Chain Assignments</i>	90
4.2.3.2	<i>Minimal Chemical Shift Perturbation Mapping</i>	91
4.2.3.3	<i>Secondary Structure Predictions of <b>b'x</b> and W111F <b>abb'x</b></i>	91
4.3	<b>Results</b>	92
4.3.1	<i>Effect of Redox State on W111F <b>abb'x</b></i>	92
4.3.2	<i>Sequential Backbone Assignment of Reduced W111F <b>abb'x</b></i>	93
4.3.3	<i>Effect of Temperature on W111F <b>abb'x</b></i>	96
4.3.4	<i>Sequential Backbone Assignment of <b>b'x</b></i>	98
4.3.5	<i>Side-Chain Assignment of <b>b'x</b></i>	101
4.3.6	<i>Secondary Structure Predictions of <b>b'x</b> and W111F <b>abb'x</b></i>	103
4.4	<b>Discussion</b>	106
4.4.1	<i>Effect of Redox State on W111F <b>abb'x</b></i>	106
4.4.2	<i>Sequential Backbone Assignment of Reduced W111F <b>abb'x</b></i>	108
4.4.3	<i>Effect of Temperature on W111F <b>abb'x</b></i>	110
4.4.4	<i>Sequential Backbone and Methyl Side-Chain Assignment of <b>b'x</b></i>	111
4.4.5	<i>Secondary Structure Predictions of <b>b'x</b> and W111F <b>abb'x</b> by DANGLE</i>	113

## Chapter 5: Ligand Binding Studies of *b'x* and W111F *abb'x* by NMR Spectroscopy

<b>5.1</b>	<b>Introduction</b>	115
5.1.1	<i>Ligand Binding by <math>^{15}\text{N}, ^1\text{H}</math>-HSQC spectra</i>	116
5.1.2	<i>Ligand Binding by <math>^{13}\text{C}, ^1\text{H}</math>-HSQC spectra</i>	118
5.1.3	<i>3D <math>^{15}\text{N}</math> NOESY-HSQC of <i>b'x</i> and W111F <i>abb'x</i></i>	118
5.1.4	<i>Synthetic Peptide Fluorination</i>	119
5.1.5	<i>Ligand Binding by <math>^{19}\text{F}</math> 1D spectra</i>	120
<b>5.2</b>	<b>Materials and Methods</b>	121
5.2.1	<i>NMR Sample Preparation</i>	121
5.2.2	<i>NMR Data Acquisition and Processing</i>	123
5.2.2.1	<i><math>^{15}\text{N}, ^1\text{H}</math>-HSQC Experiments of <i>b'x</i> and W111F <i>abb'x</i></i>	123
5.2.2.2	<i><math>^{19}\text{F}</math> 1D Experiments of <i>b'x</i> and W111F <i>abb'x</i></i>	123
5.2.2.3	<i><math>^{13}\text{C}, ^1\text{H}</math>-HSQC Experiments of <i>b'x</i></i>	123
5.2.2.4	<i><math>^{15}\text{N}</math> NOESY-HSQC Experiments of <i>b'x</i> and W111F <i>abb'x</i></i>	123
5.2.3	<i>NMR Data Analysis</i>	124
5.2.3.1	<i><math>^{19}\text{F}</math> Peptide Assignment</i>	124
5.2.3.2	<i>Chemical Shift Perturbation Mapping</i>	124
<b>5.3</b>	<b>Results</b>	125
5.3.1	<i>Effect of Neighbouring Domains on Peptide Ligand Binding</i>	126
5.3.2	<i>Effect of Protein Redox State on Peptide Ligand Binding</i>	133
5.3.3	<i>Further Evidence of Capping</i>	135
5.3.4	<i><math>^{15}\text{N}</math> NOESY-HSQC for Evidence of Binding</i>	137
5.3.5	<i>Ligand Binding by <math>^{13}\text{C}, ^1\text{H}</math>-HSQC spectra</i>	138
5.3.6	<i>Use of Fluorinated Ligands to Monitor Peptide Ligand Binding</i>	140
5.3.7	<i>Effect of Alanine Substitutions on Peptide Ligand Binding</i>	148
<b>5.4</b>	<b>Discussion</b>	153
5.4.1	<i>Peptide Ligand Binding Specificity of hPDI</i>	153
5.4.2	<i>Proof of Binding by NOESY-HSQC</i>	155

5.4.3 Ligand Binding by $^{13}\text{C}$ Methyl Shifts .....	156
<b>Chapter 6: Redox Studies of W111F <i>abb</i>'x by NMR Spectroscopy</b>	
<b>6.1 Introduction .....</b>	<b>157</b>
6.1.1 Redox Potential by NMR Spectroscopy .....	158
6.1.2 Alternative Methods of Measuring hPDI Redox Potential .....	159
<b>6.2 Materials and Methods .....</b>	<b>160</b>
6.2.1 NMR Sample Preparation .....	160
6.2.2 NMR Data Acquisition and Processing .....	160
6.2.2.1 $^{15}\text{N},^1\text{H}$ -HSQC Experiments of W111F <i>abb</i> 'x .....	161
6.2.3 NMR Data Analysis .....	161
6.2.3.1 Redox Potential of W111F <i>abb</i> 'x.....	161
6.2.3.2 Graphical Fits .....	162
<b>6.3 Results .....</b>	<b>162</b>
6.3.1 Equating Redox Potential of W111F <i>abb</i> 'x by $^{15}\text{N},^1\text{H}$ -HSQC .....	162
<b>6.4 Discussion .....</b>	<b>166</b>
<b>Chapter 7: Recombinant Fluoroindole Incorporation of W111F <i>abb</i>'x</b>	
<b>7.1 Introduction .....</b>	<b>169</b>
7.1.1 Uses of Fluorine in Biochemistry .....	170
7.1.2 Recombinant Protein Fluorination using <i>E. coli</i> .....	171
7.1.3 Fluorination of W111F <i>abb</i> 'x .....	173
<b>7.2 Materials and Methods .....</b>	<b>174</b>
7.2.1 Recombinant Protein Expression.....	174
7.2.2 Mass Determination and Isotope Incorporation.....	174
7.2.3 NMR Sample Preparation, Data Acquisition and Processing.....	174
7.2.4 NMR Data Analysis.....	174
7.2.4.1 $^{19}\text{F}$ Protein Assignment.....	175

7.2.4.2 Chemical Shift Perturbation Mapping .....	175
7.2.4.3 Redox Potential of 6-F-Trp <sup>15</sup> N W111F <b>abb'x</b> .....	175
7.2.4.4 Graphical Fits.....	175
<b>7.3 Results</b> .....	<b>175</b>
7.3.1 Expression of 6-F-Trp <sup>15</sup> N W111F <b>abb'x</b> .....	175
7.3.2 Mass Determination and Isotopic Incorporation .....	177
7.3.3 Effects of <sup>19</sup> F Incorporation Observed by <sup>15</sup> N, <sup>1</sup> H-HSQC .....	180
7.3.4 Ligand Binding of 6-F-Trp <sup>15</sup> N W111F <b>abb'x</b> .....	181
7.3.5 Redox Potential of 6-F-Trp <sup>15</sup> N W111F <b>abb'x</b> .....	183
<b>7.4 Discussion</b> .....	<b>188</b>
7.4.1 Expression of 6-F-Trp <sup>15</sup> N W111F <b>abb'x</b> .....	188
7.4.2 Protein-Observed Ligand Binding by <sup>19</sup> F NMR .....	190
7.4.3 Redox Potential by <sup>19</sup> F NMR .....	191
7.4.4 Overall Summary.....	196
<b>Chapter 8: General Summary</b> .....	<b>197</b>
<b>References</b> .....	<b>204</b>
<b>Appendix Contents</b> .....	<b>223</b>

## Abbreviations

$^1\text{H}$	proton (magnetic nuclei: 99.99% natural abundance)
$^{13}\text{C}$	13-carbon isotope (magnetic nuclei: 1.11% natural abundance)
$^{15}\text{N}$	15-nitrogen isotope (magnetic nuclei: 0.36% natural abundance)
$^{19}\text{F}$	19-fluorine (magnetic nuclei: 100% natural abundance)
1D	one dimensional
2D	two dimensional
3D	three dimensional
4-F-indole	4-fluoroindole
4-F-Phe	4-fluorophenylalanine
5-F-indole	5-fluoroindole
6-F-indole	6-fluoroindole
6-F-Trp	6-fluorotryptophan
7-F-indole	7-fluoroindole
$\Delta$ -som	$\Delta$ -somatostatin
$\phi$	phi angle
$\psi$	psi angle
$\sigma$	standard deviation
CCPN	Collaborative Computing Project for NMR
CNBr	cyanogen bromide
CV	column volume

D <sub>2</sub> O	deuterium oxide
Da	Dalton
DANGLE	Dihedral ANgles from Global Likelihood Estimates
dH <sub>2</sub> O	distilled water
DNA	deoxyribonucleic acid
DTT <sup>ox</sup>	dithiothreitol (oxidised form)
DTT <sup>red</sup>	dithiothreitol (reduced form)
EDTA	ethylenediaminetetraacetic acid
Ero1 $\alpha$	Endoplasmic Reticulum oxidoreductin 1 alpha
Ero1 $\beta$	Endoplasmic Reticulum oxidoreductin 1 beta
FXa	Factor Xa
g	gravity
GB1	protein G B1 domain
GdnHCl	guanidine hydrochloride
GSH	glutathione (reduced form)
GSSG	glutathione (oxidised form)
hPDI	human Protein Disulphide Isomerase
HSQC	Heteronuclear Single Quantum Coherence
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
KSI	ketosteroid isomerase
LB	Lysogeny Broth
LC-MS	Liquid Chromatography- Mass Spectrometry

MilliQ	ultra-pure distilled water
MM	minimal media
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PDB	RCSB Protein Data Bank
PDI	Protein Disulphide Isomerase
PDIp	pancreas-specific Protein Disulphide Isomerase
ppm	parts per million
RP-HPLC	Reverse Phase High Performance Liquid Chromatography
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
WT	wild type
yPDI	yeast Protein Disulphide Isomerase

## Abstract

Human Protein Disulphide Isomerase (hPDI) is an important protein folding catalyst involved in the reduction, oxidation and isomerisation of disulphide bonds, a major rate-limiting step in the formation of many proteins. hPDI also exhibits molecular chaperone activity, selectively binding small, unfolded peptides and a large range of mis-folded protein intermediates. hPDI is the founding member of a family of 21 structurally related proteins, and consists of domain subunits **a** and **a'** (which are catalytically active) and **b** and **b'** (which are not). **b** domain is thought to confer structural stability to the protein, and **b'** contains the primary ligand binding site. There is also an **x-linker** region, and a C-terminal acidic tail **c**, in the complete domain structure **abb'xa'c**.

Despite a recently solved crystal structure of hPDI **abb'xa'** (4EKZ.pdb) and over 50 years of research, we still have limited knowledge of how hPDI works, and more specifically how it interacts with ligands. Obtaining more detailed information is of significant importance; not only is hPDI associated with several human disease states, but it shows great potential as a therapeutic treatment, and is also used industrially to improve recombinant protein yields in mammalian and bacterial cell expression systems. Nonetheless, physiological ligand specificity of hPDI is still poorly understood.

This study focuses on characterising the ligand binding ability and specificity of hPDI using a model binding peptide,  $\Delta$ -somatostatin ( $\Delta$ -som), and two recombinantly expressed fragments of hPDI: **b'x**, the smallest monomeric fragment of hPDI capable of ligand binding, and **abb'x**, a larger fragment previously uncharacterised by NMR. Backbone and side-chain resonance assignments were carried out, and allowed hPDI binding affinity to be quantified using both protein-observed and ligand-observed NMR methods. Peptide affinity was in agreement with the previously estimated 0.1-1 mM affinity, and was found to be independent of neighbouring domains and protein redox state. Fluorination of the peptide at key aromatic residues and subsequent Ala substitutions showed that the three Phe residues of the peptide all contributed in some part to hPDI binding, with substitution of the C-terminal Phe having the largest negative effect. Substitution of the three Phe residues to Ala abolished binding to hPDI, suggesting that unlike the family member PDIp (pancreas-specific Protein Disulphide Isomerase), hPDI ligand specificity may be mediated by Phe residues.

Investigation of **abb'x** redox potential by NMR supported recently published values of full-length protein as determined by mass spectrometry, but also highlighted how the redox state of **a** domain becomes progressively more reducing upon addition of subsequent domains to the protein, further confirming cross-talk among the four domains of hPDI.

Recombinant fluorindole incorporation into **abb'x** showed the incorporation rate was not as extensive as previously reported for **b'x** fragment. Preliminary work showed the fluorine probe did not significantly alter ligand binding affinity, but in contrast reported a markedly more reducing redox potential, suggesting care must be taken to consider the potential impact protein fluorine modification can infer.

# CHAPTER 1

## Introduction

### **1.1 Protein Folding in the ER**

The endoplasmic reticulum (ER) is a cellular compartment present in all eukaryotes, where a large number of membrane and secretory proteins obtain their correct three-dimensional structure by post-translational modification (Sitia and Braakman, 2003). One of these modifications is the formation of disulphide bonds, which are made by the oxidation of two side chain cysteine thiols to form a covalent sulphur-sulphur bond.

The ER is a naturally oxidising environment that promotes the formation of disulphides. These bonds can help stabilise the native tertiary structure of a protein, and it is estimated that 16% of all human proteins contain disulphide bonds (Wang et al., 2014). Many proteins contain several cysteine residues that need to form the correct intramolecular disulphide bond pairs before the protein can reach its native or active state. Often, molecular chaperones and protein folding catalysts play a role in aiding protein folding by increasing the speed at which a protein folds into its correct native state, so that the event happens on a feasible time scale for secretion.

Molecular chaperones are a class of proteins present in most cellular compartments of the eukaryotic cell, and as such are ranked as some of the most abundant cellular proteins (Saibil, 2008).

Defining roles common for all molecular chaperones are as follows:-

- The chaperone will exhibit little substrate specificity
- It will recognise some common elements present in unfolded substrates (e.g. an exposed hydrophobic surface)
- The chaperone will only bind to a mis-folded substrate (and not native protein).

The most abundant ER chaperone is binding immunoglobulin protein (BiP), which uses energy generated from the hydrolysis of ATP to prevent aggregation of proteins in the ER (Hendershot et al., 1988). This thesis is involved in the study of human Protein Disulphide Isomerase (hPDI), another highly abundant ER resident protein and a multifunctional enzyme that possesses the ability to act as both a protein folding catalyst and molecular chaperone. Protein folding catalysts are enzymes that work in a similar manner to molecular chaperones, but are able to catalyse specific reactions. In the case of hPDI this is the reduction, oxidation and isomerisation of disulphide bonds, a crucial step in the correct folding of many proteins.

Despite a large suite of ER resident protein folding catalysts and molecular chaperones, it is still quite common for unfolded proteins to build up in the ER. These newly synthesised (unfolded) or mis-folded proteins are prone to aggregation, owing to their exposed hydrophobic regions and high concentrations. The role of a molecular chaperone is to aid in the prevention of this aggregation. If the ER lumen comes under oxidative stress due to an accumulation of mis-folded protein, the unfolded protein response (UPR) is activated to try and counteract the cellular stress (Walter and Ron, 2011). The UPR attempts to restore normal cellular activity by:-

- Targeting unfolded or mis-folded proteins for proteasomal degradation
- Up-regulating signalling pathways that lead to increased expression of foldases and molecular chaperones
- Reducing cellular translation rates

Prolonged activation of the UPR will ultimately lead to apoptosis of the cell, and chronic activation of the UPR is implicated in many different disease states including metabolic diseases (atherosclerosis, obesity, insulin resistance, fatty liver), renal disease, cancer and also many neurodegenerative diseases (Diaz-Villanueva et al., 2015).

## 1.2 An Overview of hPDI

Protein Disulphide Isomerase (PDI) was the first protein folding catalyst to be discovered some 50 years ago, isolated and characterised by two groups independently whilst they were both performing studies on RNase. Christian Anfinsen's group purified the enzyme from microsomes of rat liver homogenate samples (Goldberger et al., 1963), and Bruno Straub's group collected extracts containing PDI protein from pigeon and chicken pancreas (Venetianer and Straub, 1963).

The protein was located primarily within the ER of cells, where it was found to play a crucial role: PDI is capable of facilitating the formation of correctly folded and functional proteins for secretion out of the cell, by acting as a catalyst for the oxidation, reduction and isomerisation of native disulphide bonds, a rate limiting step in the formation of many proteins. PDI also exhibits molecular chaperone activity, selectively binding small, unfolded peptides and a large range of mis-folded protein intermediates (Hatahet and Ruddock, 2007). As the first and founding member, PDI now lends its name to a family of over 20 structurally related proteins.

Human PDI (hPDI) is a 491 amino acid, 57 kDa multi-domain protein found within the lumen of the ER at concentrations estimated to be ~0.8% of total cell protein, making it the most abundant member of its family (Freedman et al., 1994). In addition to its roles as an oxidoreductase and chaperone, hPDI forms the  $\beta$  subunit of prolyl-4-hydroxylase, an enzyme involved in the synthesis of collagens (Pihlajaniemi, Helaakoski et al. 1987), and is also a subunit of microsomal triglyceride transfer protein, an enzyme involved in the assembly of lipoproteins (Wetterau et al., 1990). hPDI also binds and stabilises the major histocompatibility complex class I peptide loading complex (Lee et al., 2009). In addition to the ER, hPDI is also found at the cell surface, where it is thought to have a mediatory role of the adhesion, secretion and aggregation of platelets (Essex and Li, 1999), and facilitates fusion of the HIV virus with the cell membrane (Barbouche et al., 2003).

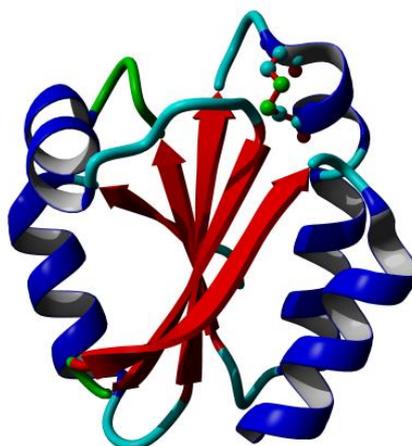
In humans, PDI is up-regulated in several disease states that occur due to protein aggregation or incorrect folding, most notably the neurodegenerative diseases Alzheimer's, Parkinson's, Huntington's and Motor Neurone Disease, where the accumulation of non-functional or incorrectly folded proteins leads to a loss of normal brain function (Andreu et al., 2012). There is also increasing evidence that PDI works to support the survival and

progression of several types of cancer (Xu et al., 2014), and that other members of the PDI family play a direct role in the resistance of certain cancers to chemotherapeutics (Tufo et al., 2014).

### 1.2.1 hPDI Structure

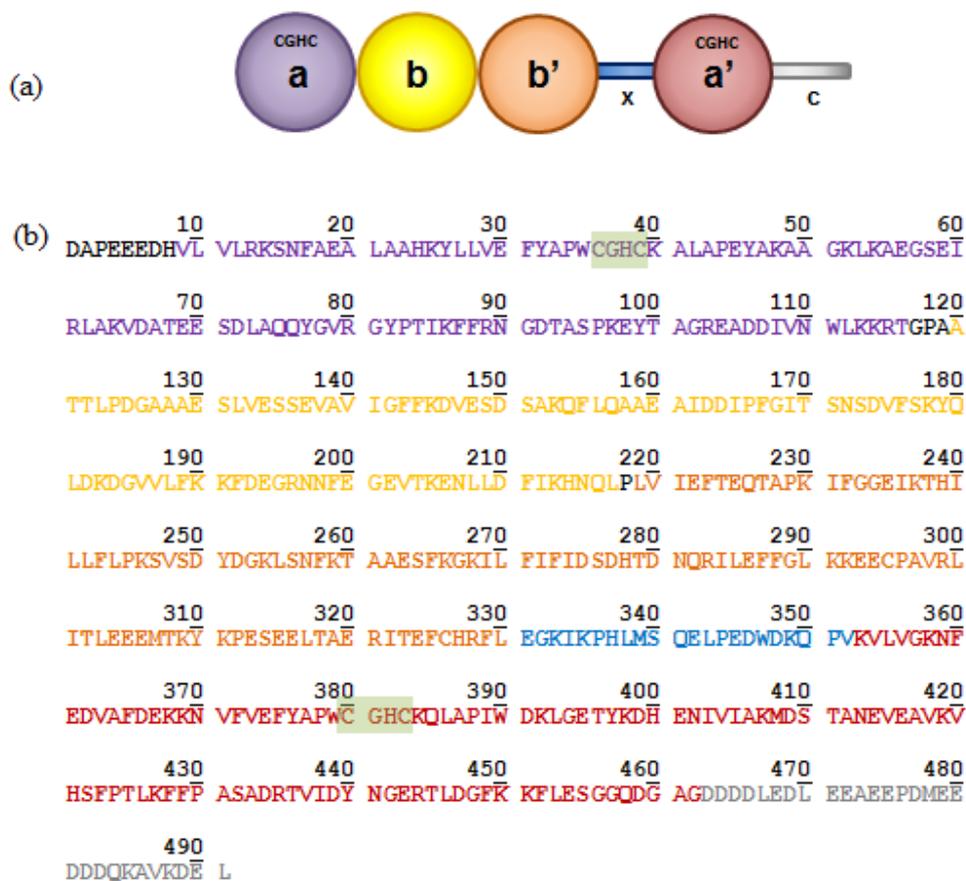
PDI is the archetypal member of a family of 21 structurally related proteins that are part of the thioredoxin superfamily, as they all contain at least one thioredoxin-like domain. Thioredoxins are a class of redox proteins present in all organisms, that contain a distinctive architectural fold named after the reductase Thioredoxin A (Martin, 1995). The fold in hPDI is characterised by 5 central  $\beta$ -sheet strands flanked by 4  $\alpha$ -helices, in a  $\beta\alpha\beta\alpha\beta\alpha$  formation (see Figure 1.1) (Kemnick et al., 1997).

PDI family members are thought to have all originally derived from a two thioredoxin-like-domain ancestor that has since evolved into the current family members by means of multiple gene duplication or deletion events. This theory is supported by homology of the ligand binding site of one domain of hPDI to the two catalytically active redox sites of other domains (Kanai et al., 1998).



**Figure 1.1: Crystal structure of oxidised thioredoxin (1ERU.pdb)** Alpha helices are shown in blue, and beta sheets in red. The two active site cysteines are shown as ball and stick figures and the sulphur atoms are shown in green.

Full length hPDI consists of four thioredoxin-like domains, defined by Alanen et al. in 2003 (Alanen et al., 2003). The two 'a' domains, **a** and **a'**, are catalytically active and contain a conserved CXXC active site motif, whilst the two 'b' domains, **b** and **b'**, are non-catalytic, and therefore lack the conserved cysteine residues. There is also a short 19 amino acid **x-linker** region between the **b'** and **a'** domains, and a C-terminal acidic tail region **c**, arranged in a complete domain structure of **abb'xa'c** (see Figure 1.2).



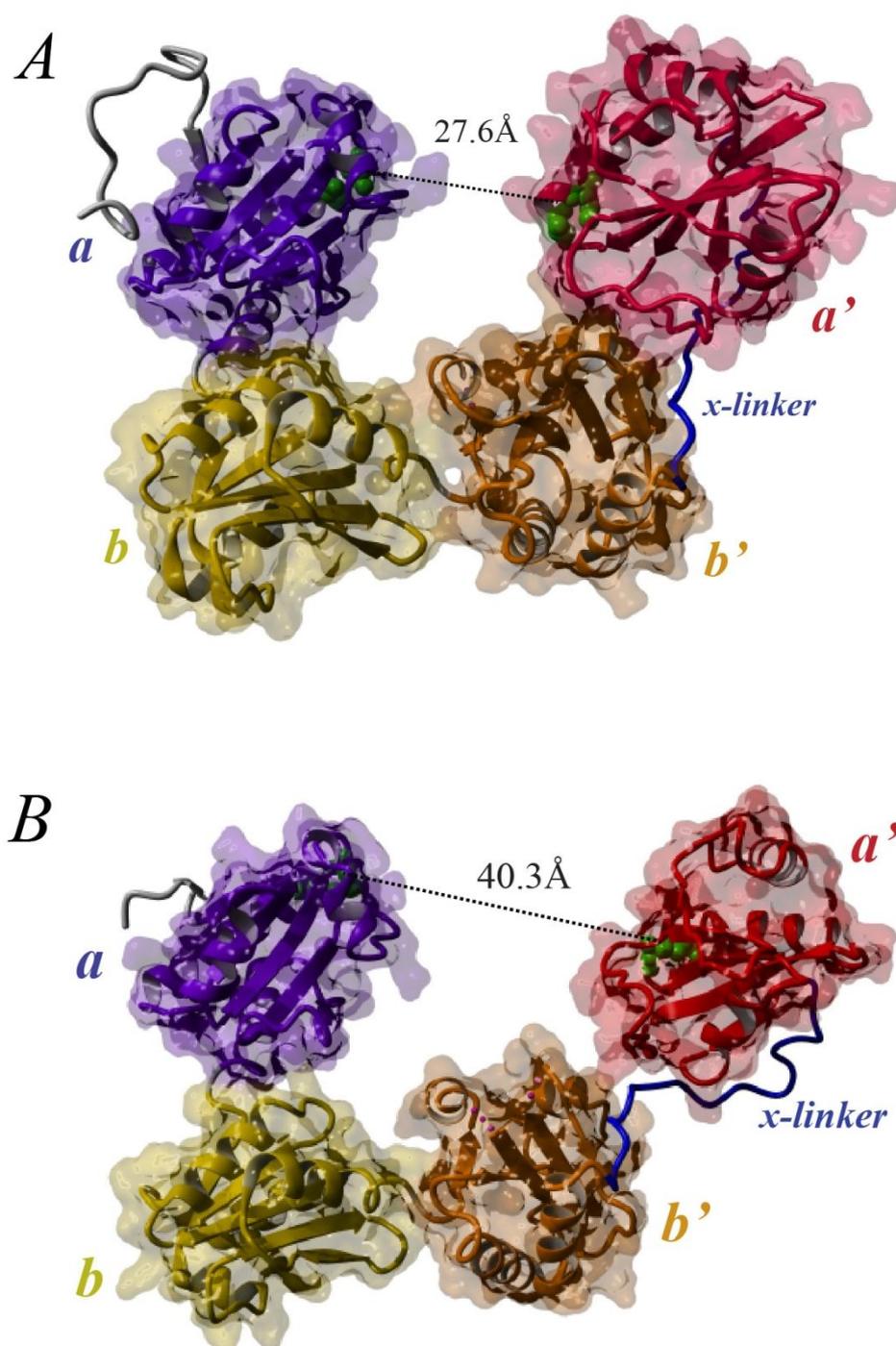
**Figure 1.2: Domain architecture of hPDI (a) and amino acid sequence of hPDI (b).** The domains are coloured purple (a), yellow (b), orange (b') and red (a'). The **x-linker** is coloured blue, the **c-terminal tail** is coloured grey, and the two redox active sites are highlighted in the sequence in green.

The catalytic **a** and **a'** domains are responsible for hPDI's action as an oxidase or reductase, whilst the **b'** domain is reported to be the primary substrate binding site for peptide ligands and mis-folded proteins (Klappa et al., 1998). Whilst isolated catalytic domains of hPDI can work as red/ox catalysts, the binding site of the **b'** domain is essential for the protein to carry out isomerase activity, with the minimum isomerase fragment identified as **b'xa'** (Wang et al., 2012). The **b** domain has additional hydrophobic patches that increase the size of the substrate binding cleft, and is thought to play a structural role as a physical 'spacer' between the two catalytic domains (Kozlov et al., 2010). In addition, the **b** domain is also thought to help aid in protein solubility, on account of its overall hydrophilic nature (Kober et al., 2013).

Whilst individual domain structures of hPDI have been solved over the last 20 years by a combination of x-ray crystallography and NMR spectroscopy, it was only during the period of this thesis that the first crystal structures of near full-length reduced and oxidised hPDI **abb'xa'** were published by the Wang group (see Figure 1.3), with the same group also publishing a crystal structure of the **bb'xa'** hPDI fragment the year prior (Wang et al., 2012), (Wang et al., 2013).

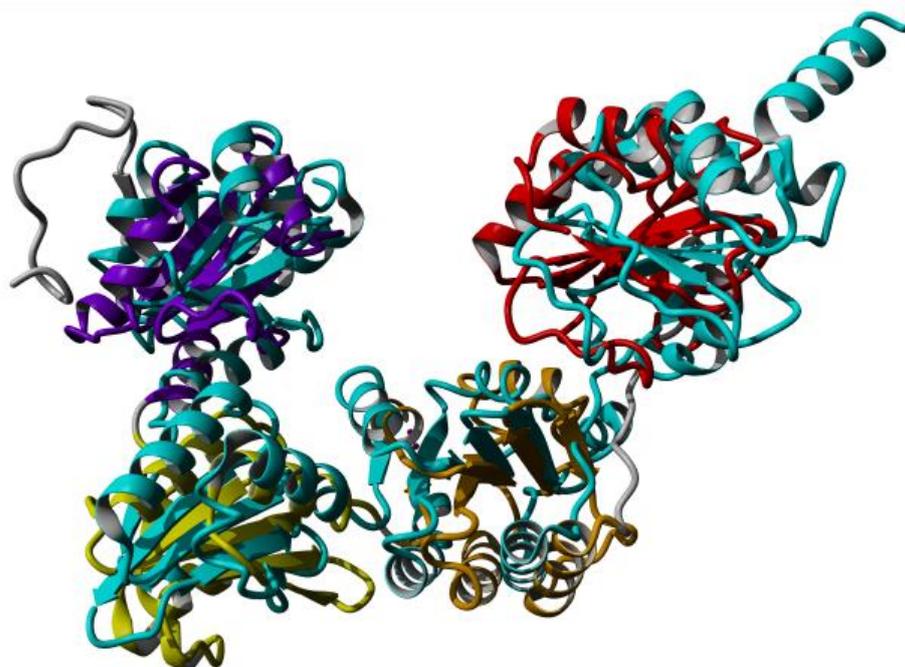
Whilst these structures lack the **c**-terminal tail extension of hPDI (a region high in acidic residues that encodes the protein's KDEL ER-retention motif), other functional significances of the tail have not been reported, and its removal helped crystallisation of the protein (Wang et al., 2012). NMR spectra of full length hPDI protein also benefit from removal of the **c**-terminal tail, which appears to be unstructured and adds significantly to the hydrodynamic volume of the protein in solution (unpublished observations).

In the oxidised state of hPDI, the four domains all appear within the same plane, where as in the reduced structure the **a'** domain rotates out of plane  $\sim 45^\circ$ , around the **x-linker** and away from the substrate binding site of **b'** domain. The distance between active sites increases from 27.6Å in the reduced state to 40.3Å in the oxidised state, with the cleft between the domains doubling in size (see Figure 1.3) (Wang et al., 2013). Thus, chaperone activity of hPDI is controlled in a redox-dependent manner: as hPDI becomes oxidised, it adopts a more open conformation and is able to accommodate larger substrates than the reduced form, by exposing its substrate binding area (Wang et al., 2012), (Wang et al., 2013).



**Figure 1.3: Crystal structures of (A) reduced (4EKZ.pdb) and (B) oxidised (4EL1.pdb) hPDI abb'xa' as published by Wang et al. (2013). The domains are coloured purple (a), yellow (b), orange (b') and red (a'). The x-linker is coloured blue and the redox active cysteines are shown in green. The distances between the two active sites is also labelled for both reduced and oxidised structures.**

Both reduced and oxidised structures of hPDI **abb'xa'** show the four domains to be arranged in a horseshoe shape, where the **bb'** fragment forms the base and the catalytic domains form the two ends, a similar conformation as observed for yeast PDI (yPDI). Figure 1.4 shows an overlay of hPDI and yPDI structures for comparison.



**Figure 1.4: Overlaid crystal structures of yPDI (2B5E.pdb) and hPDI **abb'xa'** (4EKZ.pdb).** The domains of hPDI are coloured as Figures 1.2 and 1.3, with yPDI coloured cyan. Note that the **c** region in the yPDI structure forms a helical conformation- this region is absent from the hPDI structure, but is thought to be unstructured.

Historically, crystallization of hPDI protein had been hindered by the proteins intrinsic inter-domain flexibility (Hatahet and Ruddock, 2009). Until the Wang group published **abb'xa'** structures, the two structures of yPDI were routinely used to model hPDI, despite hPDI and yPDI only sharing 31% sequence identity (Wang et al., 2015). In yPDI, the **bb'** base is fairly rigid, with the catalytic **a** domains being much more flexible. Superimposition of the hPDI structures onto yPDI show a domain rotation of  $\sim 20^\circ$  of **b'** relative to **b** domain, suggesting that **bb'** is more flexible than yPDI, behaving more like the hPDI family member ERp27 (composing a **bb'** structure) (Amin et al., 2013). In addition, whilst the yPDI structure shows greater flexibility around its **a** domain (Tian et

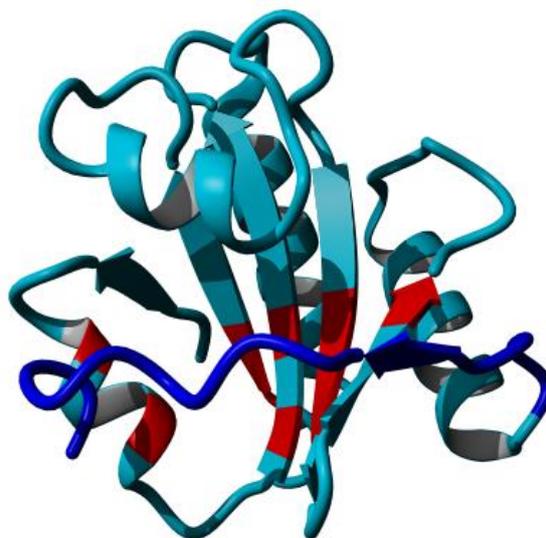
al., 2008), it is the **a'** domain of hPDI that is more flexible. This is thought to be due to the plasticity of the **x-linker** region between the **b'** and **a'** domains (Wang et al., 2010).

More recently, the oxidised and reduced hPDI crystal structures have been used to simulate how the protein may behave in solution. It is imagined that the crystal structures of hPDI were more indicative of ligand bound structures than that of free protein, as both the oxidised and reduced resolved structures show the **b'** substrate binding site partly occupied by another hPDI molecule (Wang et al., 2013). Molecular dynamic simulations showed the protein adopts a much more compact structure, stabilised by additional salt bridges between **a** and **b** domains, **a** and **b'** (which was an unexpected observation as the **a** and **b'** domains are  $\sim 45\text{\AA}$  away from each other in the crystal structures) and **b'** and the **x-linker**. Many of these interactions were confirmed *in vitro* by site-directed mutagenesis studies (Yang et al., 2014).

### 1.2.2 Ligand Binding

In 1998 the **b'** domain of hPDI was identified as the domain responsible for primary binding of ligand peptides and non-native proteins (Klappa et al., 1998; Pirneskoski et al., 2004). A crystal structure of (I272A) **b'**x was published by Nguyen et al. in 2008 (Nguyen et al., 2008), and the substrate binding site was subsequently mapped using NMR spectroscopy (Byrne et al., 2009).

Unlike the catalytic domains, **b'** has no specific conserved binding motif. The **b'** domain does contain two cysteine residues, but they are hidden deep within the protein, too far apart for a disulphide bond to form and inaccessible to thiol-modification reagents such as iodoacetamide (unpublished observation). Instead, the domain contains a large multivalent hydrophobic binding pocket that binds ligands primarily via hydrophobic interactions (Byrne et al., 2009; Denisov et al., 2009). The site was found to consist mainly of residues from the core  $\beta$ -sheet and  $\alpha$ -helices 1 and 3 (see Figure 1.5), including residues <sup>228</sup>Ala, <sup>232</sup>Phe, <sup>238</sup>Thr, <sup>240</sup>Ile, <sup>270</sup>Leu, <sup>271</sup>Phe, <sup>300</sup>Leu-<sup>302</sup>Thr and <sup>310</sup>Tyr.



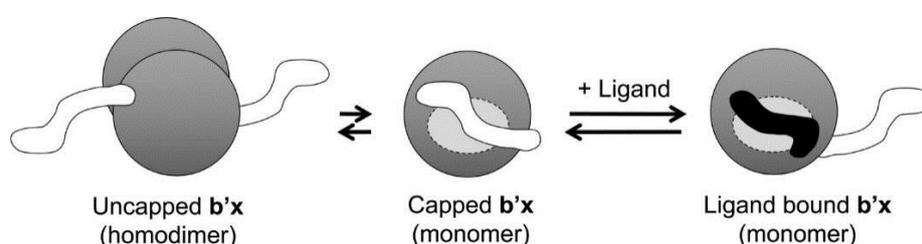
**Figure 1.5:** Crystal structure of the hPDI ligand binding domain *b'*x I272A mutant (3BJ5.pdb). I272A *b'*x is a 'capping' mutant, where the *x*-linker (in blue) caps the hydrophobic ligand binding site of the *b'* domain (in cyan). Residues mapped by Byrne et al. (2009) as composing the ligand binding site are coloured red.

As an individual domain, *b'* is able to bind a range of different peptides, including somatostatin and mastoparan. This binding is driven by reversible hydrophobic interactions, and was shown to be independent of the presence of disulphide bonds or cysteine residues (Klappa et al., 1997). Hence, the historic gold-standard peptide used to study hPDI ligand binding is  $\Delta$ -somatostatin ( $\Delta$ -som), a double Cys-to-Ser form of the peptide hormone somatostatin (AG(C $\rightarrow$ S)KNFFWKTFST(C $\rightarrow$ S)), which lacks the WT  $^3$ Cys- $^{14}$ Cys disulphide bond. This 14-mer is employed extensively within this thesis to study the ligand binding specificity of hPDI.

In 2000, work was published that identified mutations made in the *a'* domain of full length hPDI protein that indirectly affected peptide binding of the *b'* domain, but it was suspected the mutations caused destabilisation of the *a'* domain, such that it behaved like a disordered structure and bound the ligand binding site in *b'* (Klappa et al., 2000). It is anticipated that the peptide ligand binding affinity of *b'* domain is independent of neighbouring domains, however other domains of PDI have been shown to contribute to the substrate binding of larger peptides and proteins, with the *a'* domain having been shown essential for binding of mis-folded BPTI and scrambled RNase (Klappa et al.,

1997). The binding site in **b'** is also crucial for re-oxidation of the protein to occur by Ero1 $\alpha$  (see 1.2.4). More recently, small hydrophobic patches on the surface of domains **a**, **b** and **a'** have been identified as regions likely assisting the hydrophobic binding of larger mis-folded proteins by hPDI (Wang et al., 2013a), (Wang et al., 2015). Having large, low-affinity binding site(s) like this allows PDI to weakly bind to unstructured or mis-folded protein intermediates, and then release the correctly folded protein once the correct conformation is reached. An unfolded protein will display several hydrophobic residues on its surface, facilitating binding to the hydrophobic sites on PDI. As the protein adopts its correct conformation, fewer hydrophobic patches will be exposed, so PDI will lose its affinity to the now correctly folded substrate.

Studies of the **b'x** fragment of hPDI by NMR spectroscopy and intrinsic fluorescence have shown the **x-linker** is in conformational exchange, existing as either a 'capped' or 'uncapped' conformation (Wallis et al., 2009). In the 'capped' form, the **x-linker** binds across the hydrophobic ligand binding site of the **b'** domain, whereas in the uncapped form the **x-linker** is free in solution (Nguyen et al., 2008). Monomeric **b'x** can be isolated from dimeric species by gel filtration, but the protein slowly uncaps and dimerises upon prolonged storage (Curtis-Marof et al., 2014). Capping helps to explain why purified hPDI **b'** and **bb'** fragments exist in exclusively dimeric states; removal of the **x-linker** region promotes dimerization of the protein at its hydrophobic binding site (Wallis et al., 2009). These different states are shown in **b'x** by Figure 1.6.



**Figure 1.6: Alternative conformations of the *x-linker* region in *b'x* (taken from Byrne et al. 2009).** When **b'x** is in solution it can exist in a 'capped' monomeric/ 'uncapped' homo-dimeric equilibrium, with both states stabilising the protein by occlusion of the hydrophobic binding site. In the presence of ligand, the protein adopts an 'uncapped' monomeric form, as the ligand occupies the substrate binding site, and the **x-linker** is free in solution.

The observation of capping lead to the theory that the **x-linker** region may work to modulate the ligand binding site of hPDI, with exposure of the hydrophobic site allowing the protein to bind ligands (Byrne et al., 2009). A similar capping event had been observed in the PDI family member ERp44, which is capped by its C-terminal tail (Wang et al., 2008). Whilst this phenomena is observed in fragments of PDI that end immediately after the **x-linker**, ‘capping’ has not been conclusively proven in full length hPDI.

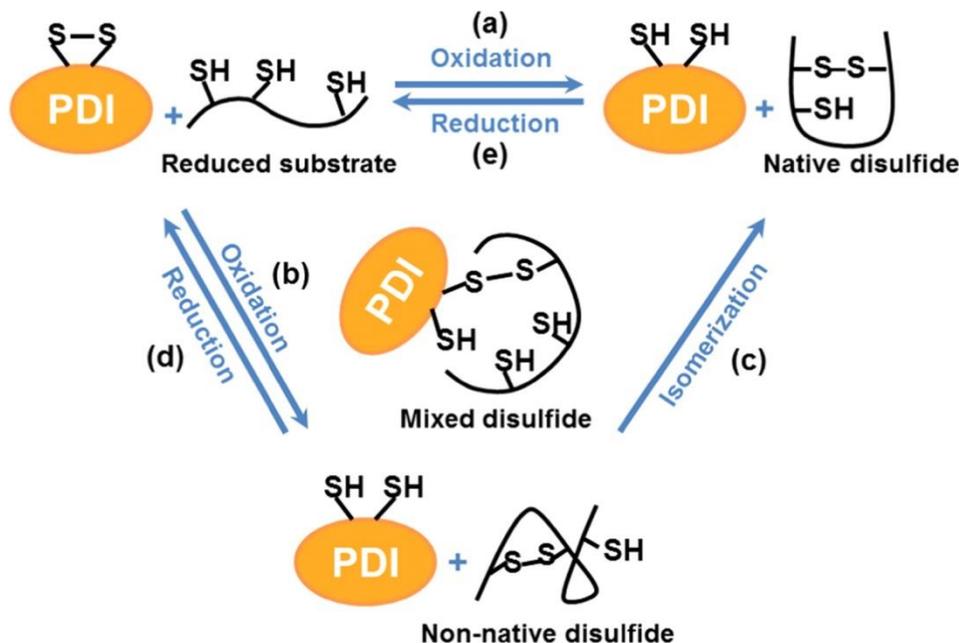
The peptide binding specificity of hPDI has not been extensively studied, and there is scant recent work published on this functional role of hPDI. Early work showed that other peptides can compete with  $\Delta$ -som to bind hPDI, but these peptides have few residues in common with  $\Delta$ -som (for example the sequence of mastoparan is INLKALAALAKKIL), suggesting there is no specific binding motif recognition (Ruddock et al., 2000). More detailed work has been carried out to ascertain the specificity of PDIp (a pancreatic specific PDI family member), where it has been shown that a single Tyr or Trp residue is sufficient for binding recognition. Mutation of these residues to Phe abolishes recognition by PDIp (Ruddock et al., 2000), and subsequently the binding motif of ligands was identified as an hydroxyaryl group (Klappa et al., 2001). This simple recognition motif opened up a wide range of potential non-physiological ligands for PDIp, including xenobiotics and oestrogens, making PDIp a much more ‘druggable’ target than hPDI. No such recognition motif has been identified for hPDI, and it is anticipated that substrate recognition by hPDI may be much more complex.

### 1.2.3 Catalytic Activity

Depending on its redox state, hPDI is able to catalyse the reduction, oxidation or isomerisation of disulphide bonds, on account of its two CGHC motif redox active sites (the CXXC motif is conserved among most catalytic domains of the thioredoxin and PDI families). During substrate oxidation, the disulphide of a PDI active site is reduced to the dithiol state, after which PDI is subsequently re-oxidised (see 1.2.4). Similarly, disulphide reduction is catalysed by the transfer of reducing equivalents from PDI to the oxidised substrate (Wilkinson and Gilbert, 2004).

There are two potential methods of disulphide ‘isomerisation’: either several cycles of reduction and oxidation will occur until the correct native conformation is reached, or true isomerisation will occur, where disulphide bonds are re-arranged to form the correct

structure, with no net change to the catalysts redox state (unlike reduction and oxidation reactions) (Hatahet and Ruddock, 2009). This occurs in hPDI because the N-terminal active site cysteine of PDI is maintained primarily in a thiolate  $S^-$  state by virtue of its low  $pK_a$  ( $pK_a$  of N-terminal Cys in the **a** domain is  $4.8 \pm 0.1$ , compared to a typical cysteine residue  $pK_a$  of  $\sim 8.3$ ) (Karala et al., 2010). The abnormally low  $pK_a$  is brought about by the residues structural location (at the N-terminus of an  $\alpha$ -helix, which has a permanent dipole associated with it) and stabilisation of the ionised state by the histidine imidazole group within the CGHC motif (Darby and Creighton, 1995), (Kortemme and Creighton, 1995). In contrast, the C-terminal cysteines of the hPDI catalytic sites have somewhat limited solvent exposure, and exist with a higher than average  $pK_a$  ( $pK_a$  of C-terminal Cys in the **a** domain is  $\sim 8.6$  (Lappi et al., 2004), (Karala et al., 2010)). This high  $pK_a$  would ordinarily limit the re-oxidation of hPDI *in vivo*, however an arginine residue conserved in **a** domain of hPDI and other PDI family members is able to move in and out of the active site, sufficiently lowering the  $pK_a$  of the C-terminal cysteine, enabling re-oxidation of hPDI (Lappi et al., 2004), (Karala et al., 2010). Disulphide reduction, oxidation and isomerisation by hPDI is summarised in Figure 1.7.



**Figure 1.7: PDI-catalysed disulphide bond formation (taken from Wang et al. 2015).** Oxidised PDI can oxidise reduced substrates to form native (a) or non-native (b) disulphide-bonded protein, which leaves PDI in its reduced form. Reduced PDI can subsequently be re-oxidised by reduction of native disulphides (e), mis-folded substrates (d) or a variety of different oxidising agents (not shown). Non-native disulphides can be remedied either by cycles of reduction and subsequent re-oxidation, or by isomerisation with PDI by direct intra-molecular disulphide shuffling (c). Note only one redox catalytic site has been shown for clarity.

Whilst PDI research is commonly carried out on different fragment constructs of PDI, it is important to remember that the functional ability of the construct is dependent on the composition of domains. Domains can work independently *in vitro* (for example, a simple oxidation reaction requires only a single catalytic domain (Pirneskoski et al., 2004), but for isomerisation reactions the minimum redox fragment is **b'xa'** (Wang et al., 2012). Maximum isomerisation efficiency, or more complex isomerisation reactions that involve substantial changes in structure to the substrate, require all domains of hPDI (Darby et al., 1998), findings that indicate hPDI domains work synergistically (Freedman et al., 2002).

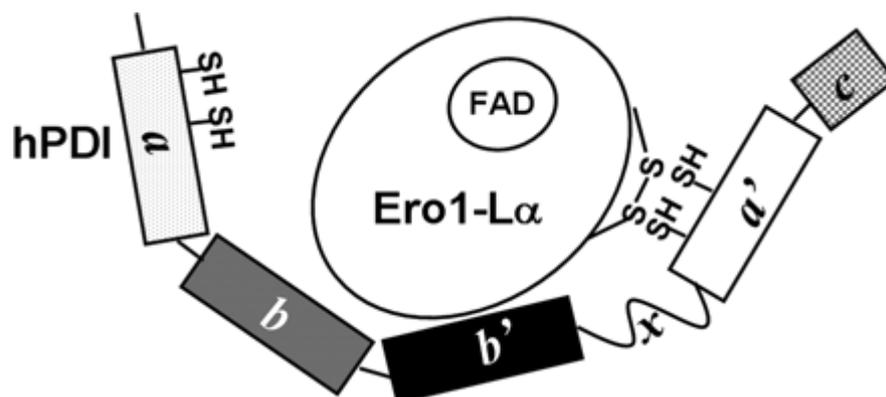
In full length protein, the **a** domain of hPDI is thought to primarily act as an isomerase and the **a'** domain primarily as an oxidase, due to the preferential oxidation of **a'** domain. This is despite the two active sites sharing highly similar reduction potentials (Chambers et al., 2010) (detailed further in Chapter 6).

The recent molecular modelling simulations by Wang et al. also show the **a** and **a'** redox catalytic sites may come within 5.4Å of one another- close enough to exchange disulphides. This work was supported by mutagenesis, and the group were able to identify a <sup>36</sup>Cys-<sup>380</sup>Cys disulphide-bonded form of their C38A/C383A mutant, suggesting disulphide exchange can occur among the active sites of full length hPDI (Yang et al., 2014).

#### 1.2.4 Re-oxidation of PDI by Ero1 $\alpha$

After oxidation of a PDI substrate, PDI is left either in a mixed disulphide state or as a fully reduced protein. *In vitro* PDI can be selectively reduced or oxidised by either dithiothreitol (DTT) or oxidised glutathione (GSSG). It can also be re-oxidised by GSSG *in vivo*, or additionally by hydrogen peroxide or other ER oxidoreductins, peroxiredoxins or peroxidases (Ruddock, 2012). However, the largest contributor to PDI re-oxidation *in vivo* is the ER oxidoreductin 1 family member Ero1 $\alpha$ , a protein expressed constitutively by most cells (Tavender and Bulleid, 2010). The family member Ero1 $\beta$  can also re-oxidise hPDI, but this protein is expressed by a smaller subset of cells (Pagani et al., 2000).

PDI is the preferred substrate of the flavin dependent Ero1 $\alpha$ , and this interaction has been shown to be highly specific to PDI, for instance, ERp57, a PDI family member which shares the same domain structure as PDI is not able to reduce Ero1 $\alpha$  (Inaba et al., 2010). The transfer of disulphides occurs through contacts between the PDI redox active sites and an 'outer' active site of Ero1 $\alpha$ . This idea is demonstrated by Figure 1.8.



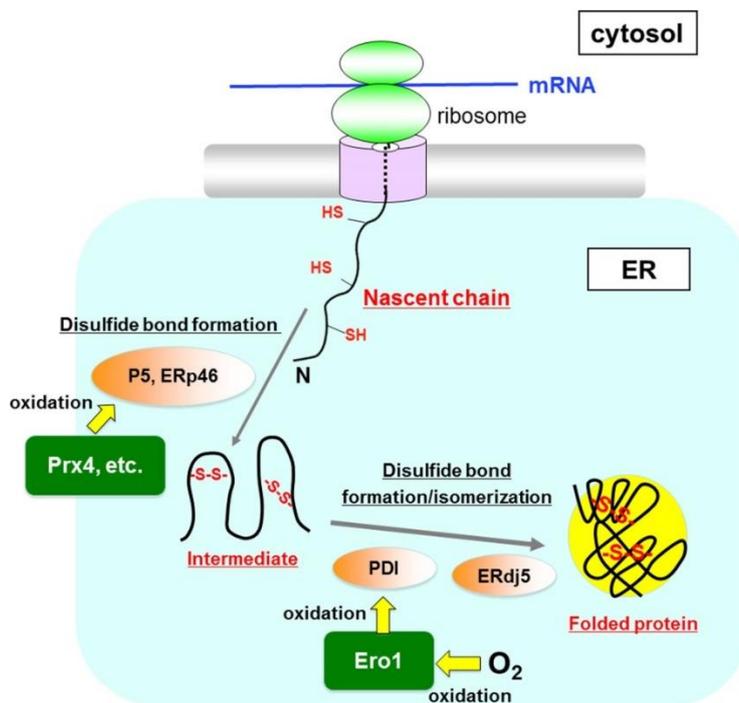
**Figure 1.8: Proposed model of hPDI re-oxidation by Ero1 $\alpha$**  (taken from Wang et al. 2009). This figure demonstrates that the minimal binding fragment for re-oxidation by Ero1 $\alpha$  is **b'xa'**. Due to Ero1 $\alpha$  binding at the **b'** substrate binding site, the <sup>94</sup>Cys-<sup>99</sup>Cys redox pair of Ero1 $\alpha$  face the redox active site of the **a'** domain of hPDI.

A  $\beta$ -hairpin loop of Ero1 $\alpha$  binds to hPDI at the **b'** substrate binding site, which due to the overall binding orientation means its electron shuttle loop is within close proximity of the **a'** active site of hPDI; thus Ero1 $\alpha$  preferentially oxidises the **a'** domain. This preferential oxidation was proven by the domain swap mutant **a'bb'xac**, which was oxidised at its **a** domain (Wang et al., 2009).

The Ero1 $\alpha$  active site is then re-oxidised by disulphide exchange with its 'inner' active site disulphides (Tavender and Bulleid 2010) and the electrons are ultimately passed to molecular oxygen to form hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>).

The ability of several enzymes to re-oxidise PDI suggests a complex network of disulphide bond formation in mammalian cells, and recently a new model of disulphide bond formation in the ER has been proposed. The Inaba group have shown that the peroxiredoxin Prx4, (which first reduces H<sub>2</sub>O<sub>2</sub> to form water), preferentially oxidises P5 and ERp46 (two other PDI family members), which go on to mediate rapid, albeit error-prone, disulphide bond formation of newly synthesised proteins, in a reaction independent of molecular oxygen (Sato et al., 2013). It is then proposed that PDI (oxidised by Ero1 $\alpha$ ) follows with a slow but highly precise isomerisation of the non-natively folded proteins, in

an oxygen dependent manner (Ero1 $\alpha$  is oxidised by molecular O<sub>2</sub>). The idea is presented in Figure 1.9.



**Figure 1.9: Proposed model of mammalian oxidative protein folding (taken from Okumura et al. 2015).** This figure suggests that P5 and Erp46 rapidly form error-prone disulphides in nascent proteins, with PDI playing a later role in slower but more efficient disulphide isomerisation.

This theory is supported by the observation that the presence of these components *in vitro* accelerates the overall rate of both reactions (thus showing the two systems could work synergistically (Sato et al., 2013),(Kojima et al., 2014)), and the proposal is also supported by observations made *in vivo* (Koritzinsky et al., 2013). This idea may in part also help to explain the seeming redundancy of so many PDI family members, if certain family members are found to carry out distinct roles in the cell (Okumura et al., 2015).

### 1.2.5 Chaperone Activity

In addition to its activity as a sulfhydryl protein folding catalyst, hPDI has been shown to exhibit activity as a molecular chaperone, by aiding in protein folding and preventing aggregation in a way that is independent of cysteine residues. It was theorised that PDI

must display some chaperone activity before it is able to function as an isomerase, promoting peptide chain folding by bringing together distant cysteines prior to native bond formation. The idea was first proposed in 1993 by Wang and Tsou (Wang and Tsou 1993), and after strong support from experimental data the idea is now widely accepted.

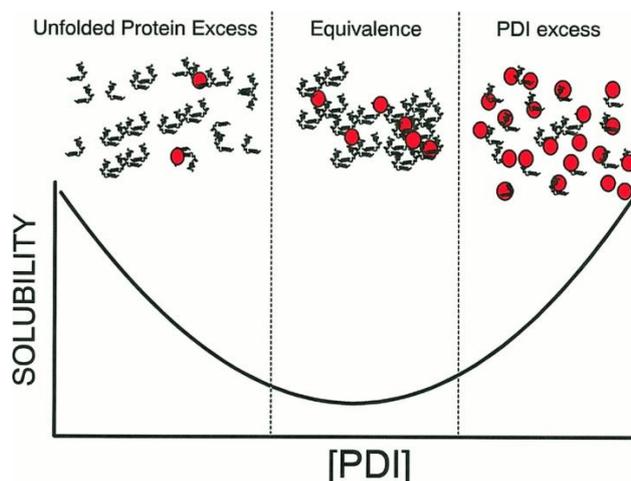
Instead of interaction with the redox active domains, PDI can bind non-cysteine containing unfolded substrates primarily at its **b'** domain. PDI chaperone activity can therefore be abolished by the binding of a small peptide ligand at the **b'** substrate binding site (Quan et al., 1995). Binding of larger mis-folded substrates also occurs throughout the entire hydrophobic cleft of hPDI, and it has been shown that the oxidised (open) conformation of hPDI exhibits greater chaperone activity than its reduced (closed) protein form (Wang et al., 2012), (Yang et al., 2014). This seems logical- a more open conformation would allow the protein to bind larger misfolded protein intermediates.

Measuring the molecular chaperone activity specifically of PDI is challenging, as many of PDI's native substrates also need to form disulphide bonds- hence it can be hard to differentiate between PDI functioning as a molecular chaperone, or as an oxidase or isomerase. PDI chaperone activity is traditionally assayed using D-glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) (Cai et al., 1994), as GAPDH does not form any disulphide bonds, but hPDI can also prevent aggregation of lactate dehydrogenase, citrate synthase, rhodanese, or alcohol dehydrogenase (Watanabe et al., 2014), in an ATP independent manner (Cai et al., 1994).

The ability of PDI to behave as a molecular chaperone *in vitro* is often measured by activity of the refolded substrate protein. This result may not be indicative of the role PDI plays *in vivo*- the substrate protein may not always be folded by PDI to completion, or PDI may be just one of many molecular chaperones working to assist the correct refolding of a substrate (Cai et al., 1994). Indeed GAPDH is a cytoplasmic enzyme, and Cai et al. acknowledge the ER-resident hPDI may not be involved in GAPDH biogenesis *in vivo* (Cai et al., 1994).

Somewhat paradoxically, under certain conditions PDI has also been shown to behave as an anti-chaperone *in vitro*, forming large, insoluble complexes with lysozyme (Puig and Gilbert, 1994). This is thought to occur when the concentration of PDI present in a solution

of denatured lysozyme (that already has a tendency to self associate into aggregates) is lower than to be in excess (5-10 unfolded protein: 1 PDI), (see Figure 1.10).



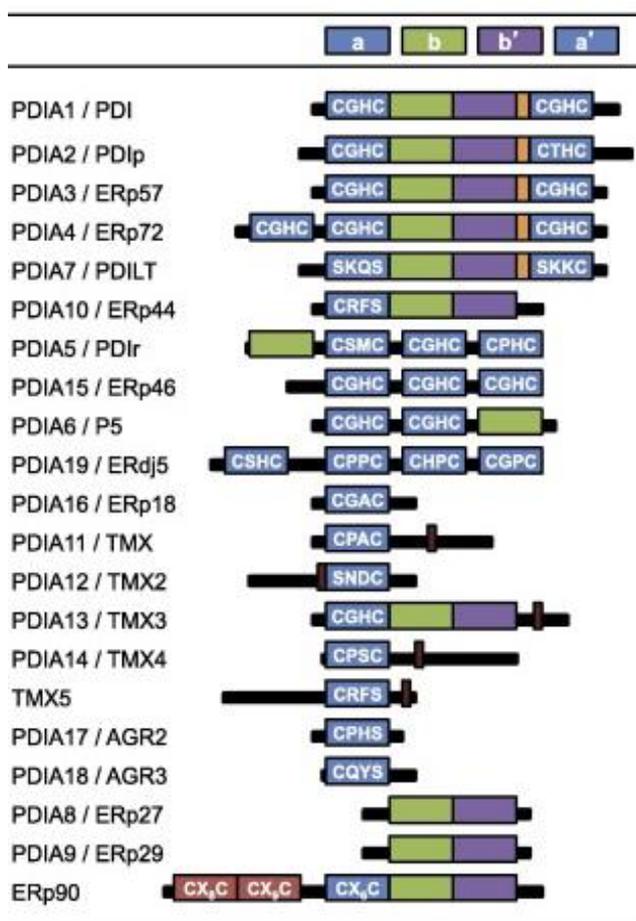
**Figure 1.10: The chaperone/anti-chaperone behavior of PDI (taken from Gilbert, 1997).** Increasing the concentration of PDI from low concentrations (left region), can first lead to a decrease in solubility of the substrate, by providing extra ways for the unfolded protein to cross-link into larger aggregates (anti-chaperone behavior, shown by middle region). At higher concentrations this self-associated cross-linking is inhibited by PDI (chaperone behavior, right region).

In this case, multivalent interactions with PDI can inadvertently work to promote larger aggregate formation (Gilbert, 1997). *In vitro* this type of aggregation is often irreversible, but *in vivo*, DTT-induced ER chaperone associated aggregates of misfolded proteins are correctly folded and released from the ER upon removal of DTT (Tatu et al. 1993). It is thought this could be a crucial preservation mechanism to retain unfolded proteins in the ER, and to protect these proteins from degradation when the when the cell is under stress, or when the stoichiometric chaperone capacity of PDI is exceeded (Gilbert, 1997). Finally, PDI has also been implicated in the degradation of some terminally misfolded proteins (Gillece et al., 1999), and more recently it has been proposed that this degradation is mediated by an interaction with signal peptide peptidase (SPP), where PDI bound to SPP works to unfold the misfolded protein into a dislocation competent state, for transport back into the cytosol for degradation (Lee et al., 2010). In this example, hPDI was shown to be

required for the degradation of Major Histocompatibility Complex (MHC) Class I heavy chains (Lee et al., 2010).

### 1.3 PDI Family Members

PDI is the founding member of a group of 21 structurally related proteins, grouped on account of their thioredoxin-like domain structure. Like PDI, family members contain thiol-reactive sites that display red/ox/isomerase activity, and/or thiol-non-reactive thioredoxin-like domains that display chaperone-like activities, to facilitate ER protein folding (Kozlov et al., 2010). Figure 1.11 shows the current hPDI family members.



**Figure 1.11: Current human PDI family members (taken from Andreu et al. 2012).** Catalytic thioredoxin-like domains **a** and **a'** are shown in blue, non-catalytic **b** domains in green, and **b'** substrate binding domains are shown in purple. The **x-linker** region is shown in orange, and transmembrane domains are shown in brown. The specific CXXC sequence for each member is shown in the **a** and **a'** domains.

It is still not fully understood why there is such a large number of functionally similar enzymes present in the ER lumen. Whilst it has been shown in mammalian cell lines that knockout of PDI can be compensated for by ERp57 (Rutkevich et al., 2010), knockouts of PDI or ERp57 are both lethal at an embryonic stage (Ni and Lee, 2007), and mice deficient in ERp57 suffer from severe bone growth complications during puberty (Linz et al., 2015). Researchers suggest that each enzyme has its own substrate specificity, or that *in vivo* each enzyme catalyses a distinct redox reaction (Ellgaard and Ruddock, 2005), (Jessop et al., 2009), (Rutkevich et al., 2010). Whilst most catalytic PDI family members have been extensively studied, the functions of many of the redox-inactive family members are somewhat less well defined, with several enzymes still classed as having unknown functions.

An overview of several PDI family members is now given, with particular focus given to family members that are homologous to PDI, or those that are also capable of binding peptide ligands.

### 1.3.1 PDIp (PDIA2)

PDIp is a homologue of PDI primarily secreted by acinar cells of the pancreas, with expression also observed in the human brain, bladder and stomach (Hatahet and Ruddock, 2009), and in rat liver, kidney, lung, GI tract and reproductive organs (Conn et al., 2004), (Klappa et al., 1998). The protein has the same **abb'xa'c** domain architecture as PDI and shares 45% sequence identity (Walker et al., 2013). There is no full length structure of PDIp currently available, but the protein is anticipated to share high structural similarity to PDI.

PDIp displays both disulphide isomerase and chaperone activity; PDIp is able to bind scrambled RNase but not native protein, as well as a range of peptides derived from pancreatic digestive enzymes, and  $\Delta$ -som (Klappa et al., 1998)

Unfortunately there is little published on the redox activity of PDIp. The protein is known to deviate from the conserved CGHC motif observed in PDI and has a CTHC redox active motif in its **a'** domain, showing ~50% activity compared to PDI in a peptide oxidase assay (Alanen et al., 2006a).

Early ligand binding studies of PDIp suggested that a single Trp or Tyr residue in the peptide was sufficient to trigger peptide binding (Ruddock et al., 2000), with subsequent work showing the protein interacts specifically with a hydroxyaryl group (Klappa et al., 2001). PDIp was also shown to bind the hormone oestradiol in a hydrophobic pocket between the **b** and **b'** domains with an approximate  $K_d$  of 150 nM (Fu et al., 2011), (Fu et al., 2012). Proposed physiological roles of PDIp include the modulation of oestrogenic hormones (Fu et al., 2009) or vice versa (Klappa et al., 2001), and the isomerisation of incorrectly folded secretory proteins (for example zymogens) (Klappa et al., 1998). A loss of PDIp expression is also observed in pancreatic cancer, suggesting that PDIp could be used as an important biological marker.

### 1.3.2 ERp 57 (PDIA3)

ERp57 is a homologue of PDI, which shares the same domain structure but lacks a specific hydrophobic substrate binding site in its **b'** domain. Instead, the corresponding patch of ERp57 is negatively charged, and the protein is therefore unable to bind PDI model peptides or other larger substrates (Zapun et al., 1998). ERp57 interacts almost exclusively at this modified site with calnexin or calreticulin, two ER resident proteins (calnexin is found anchored into the ER membrane and calreticulin is resident in the ER lumen), that themselves interact specifically with glycoproteins (Oliver et al., 1997).

Unlike other PDI family members, instead of binding its substrates directly ERp57 relies on the association of calnexin or calreticulin with glycoproteins to bring its substrates into close proximity of its **a** and **a'** active sites (Kozlov et al., 2006). Hence the main functional role of ERp57 is thought to be the isomerisation of non-native disulphide bonds in a specific subset of glycoproteins (Jessop et al. 2009).

#### **The calnexin/calreticulin cycle**

Calnexin and calreticulin are components of the protein quality control system in the ER, and work either to promote the correct folding of a secretory pathway protein, or alternatively to target misfolded proteins for degradation, and the interaction of a glycoprotein with either calnexin or calreticulin commits them to processing by the calnexin/calreticulin cycle.

Protein glycosylation is an extremely common post-translational modification that involves the covalent addition of sugar moieties, in this case a core oligosaccharide containing three glucose molecules, to particular residues in a nascent polypeptide chain. Glycosylation is used within the ER as a quality control mechanism, and to directly monitor the status of protein folding (for detailed reviews see (Ellgard and Helenius, 2003), (Caramelo and Parodi, 2008)). Successive hydrolysis of two of the three glucose molecules of the oligosaccharide by ER glucosidase I and II leaves a monoglucosylated glycoprotein that is able to interact with calnexin or reticulon- which in turn enables the folding glycoprotein to interact with ERp57 and undergo disulphide bond isomerisation. Upon removal of the final glucose molecule by ER glucosidase II, the protein is able to dissociate from the bound lectin. Correctly folded proteins are then directed to the Golgi- but not before it is ‘checked’ by glycosyltransferase enzyme. If the protein displays a non-native structure or the enzyme identifies exposed hydrophobic patches, the protein is re-glycosylated with a single glucose molecule, and consequently re-enters the calnexin/calreticulin cycle until it adopts a correct structure.

ERp57 also functions as a component of the MHC Class I loading complex, (where it forms a disulphide-linked complex with tapasin, and is thought to facilitate the correct formation of Class I molecules) (Radcliffe et al., 2002), and is also involved in the correct folding of influenza protein HA (Daniels et al., 2003). More recently, ERp57 has been shown crucial for the ‘uncoating’ of Simian Virus 40 by isomerisation of inter-chain disulphides of major capsid protein VPI (Schelhaas et al., 2007), a reaction that happens independently of calnexin or calreticulin.

### 1.3.3 ERp27 (PDIA8)

ERp27 is another ER resident protein expressed by the pancreas, kidney, spleen, bone marrow and lungs of vertebrates (Lash et al., 2000). It is homologous to the non-catalytic **bb'** domains of hPDI, and whilst currently classed as a protein with unknown function, it is known to be up-regulated in times of ER oxidative stress and the UPR (Kober et al., 2013). It consists of two thioredoxin folds, although neither contain a CXXC redox motif (Amin, Freedman, 2013), so the protein is not able to function as an oxido-reductase. The **b** domain is thought to aid protein solubility *in vivo* (Kober et al., 2013). ERp27 is able to

bind  $\Delta$ -som at a hydrophobic site in its second domain, that is homologous to the **b'** substrate ligand binding site of hPDI, but in contrast to PDI, ERp27 does not contain a flexible **x-linker** after its homologous **b'** domain.

In addition to disordered peptide binding, the protein has also been shown to interact with a model protein substrate exclusively in an unfolded state (Alanen et al., 2006b), (Kober et al., 2013) suggesting the protein acts as a molecular chaperone. ERp27 also binds ERp57 at the same site ERp57 binds lectins (the C-terminus of its **b'** domain), a site distinct from its ligand binding site. With recent observations of an ERp27-ERp57 interaction, Kober et al. have proposed that in times of oxidative stress, the up-regulation of ERp27 allows the chaperone to recruit ERp57 to the global oxidative protein folding effort; binding of ERp27 presents ERp57 with a greatly increased pool of mis-folded substrates (ERp57 would otherwise only interact with certain lectin bound glycoproteins) (Kober et al., 2013).

#### 1.3.4 PDILT (PDIA7)

PDILT is another tissue specific (testis) ER resident glycoprotein that is homologous to PDI, however it is the only PDI family member under developmental control. It is induced during puberty, and plays a role in the quality control of the sperm membrane protein ADAM3. It has been implicated in cases of male mouse infertility (Tokuhiro et al. 2012), and further work aims to see if this relationship is also inferred in some cases of male human infertility.

The protein lacks the catalytic cysteine residues of PDI, and therefore does not exhibit oxidoreductase activity *in vitro*. Instead of the CXXC motifs of PDI, PDILT displays a SKQS motif in **a** domain, and the single cysteine motif SKKC in **a'** domain. This motif is highly conserved in PDILT of other species, and also among bacterial thioredoxin proteins (van Lith et al., 2007). This lack of cysteines suggests the protein functions primarily as a molecular chaperone, aiding with the correct folding of spermatid-specific proteins.

The recent crystal structure of PDILT **b'** domain shows the hydrophobic substrate binding pocket occupies a larger area that observed in PDI (Bastos-Aristizabal et al., 2014). The crystal structure shows the binding pocket to be occupied by Trp and Tyr side chain residues of another PDILT molecule, suggesting that **b'** preferentially binds exposed aromatic residues in its substrates; indeed PDILT also binds  $\Delta$ -som (van Lith et al., 2007).

### 1.3.5 ERp44 (PDIA10)

ERp44 is another PDI family member discovered due to its interaction with Ero1 $\alpha$  and Ero1 $\beta$ , with which it forms mixed disulphides. It has a similar clover-shape arranged domain structure as PDI, with domain sequence **abb'**, and a CRFS catalytic motif in its **a** domain (Wang et al., 2008). ERp44 is upregulated in times of oxidative cellular stress, although the protein shows little activity as an oxidoreductase or isomerase, thought to be due in part to the protein's flexible **c**-terminal tail. In solution the tail partly occludes the hydrophobic substrate binding pocket of **b'** domain, as well as an additional hydrophobic patch around the active site motif of **a** (Anelli et al., 2002). The **c**-tail is thought to regulate ERp44 activity *in vivo*, as removal of it improves protein activity five-fold (Wang et al., 2008).

### 1.4 Implication of hPDI in Disease States

Whilst initially PDI and its family members were thought to play a protective role in the cell, owing to their up-regulation as part of the UPR (to reduce oxidative stress levels of the cell), more recent PDI publications describe how it may actually have a causal role in some disease states. Recently, Erp72 and ERp5 have been implicated in mediating drug resistance of metastatic lung cancer to cisplatin (Tufo et al., 2014), and there is an increasing number of publications linking post-translationally modified hPDI to the umbrella of neurodegenerative diseases (Jeon et al., 2014). In addition, PDI is implicated in the progression of other cancers (Xu et al., 2014), diabetes (Grek and Townsend, 2014) and cardiovascular disease (Ali Kahn et al., 2014), although the exact role PDI plays in these diseases is still to be determined.

There is evidence suggesting that some viruses are able to take advantage of PDI's ability to carry out thiol-disulphide exchange, to either enter the cell, or a specific cellular compartment. Newcastle virus is able to gain entry into its host cell by large conformational changes of its F fusion protein (mediated by disulphide re-arrangement carried out by PDI) (Jain et al., 2008), and the toxic A1 subunit of a holotoxin secreted by *Vibrio cholerae* is displaced from the secretory pathway into the cytoplasm by PDI (Tsai et al., 2001).

There is increasing evidence to suggest that hPDI can even escape the ER, and mediate thiol-disulphide exchange at the surface of the cell, to play a role in haemostasis (Jordan and Gibbins, 2006). While it has been difficult to prove that it is PDI itself, and not another PDI family member being recruited to the surface, recent work suggests that PDI is able to activate tissue factor, which itself controls the production of fibrin- a crucial factor in blood plasma that allows clotting to occur (Reinhardt et al., 2008). Cell surface PDI has also been shown to increase infectivity of Th2 (T helper 2) cells to HIV virus, by binding to and enhancing the activity of galectin-9, a secretory lectin (Bi et al., 2011).

Detrimental roles of PDI may also be down to a loss of its normal protective function, or a toxic gain of function- for example, a common post translational modification of proteins is s-nitrosylation, a normal cellular mechanism analogous to phosphorylation, that controls a range of protein functions by the addition of a nitroso (NO) group to a reactive thiol group (Stamler et al., 2001). However, such a modification to hPDI is an aberrant, irreversible reaction that inhibits the normal protein activity, reducing both chaperone and isomerase activity (Uehara et al., 2006). This topic has been particularly well reviewed recently by Parakh et al., (2015), and makes clear that these newly emerging deleterious roles of hPDI require further investigation. Either way, PDI shows promise as a biomarker for diagnosis or prognosis of many disease states, as well a potential therapeutic treatment and drug target (Andreu et al., 2012).

### **1.5 An Introduction to Biomolecular NMR**

Since publication of the first solution structure solved by NMR in 1985 (that of BPTI by (Chazin et al., 1985)), biomolecular NMR has proved an invaluable tool in the field of protein and protein-ligand complex structure determination. NMR has many advantages over other techniques like x-ray crystallography or cryo-EM, particularly for the study of a conformationally active enzyme like PDI. First, proteins do not always readily crystalize, or may not form the 'correct' type of crystal for adequate diffraction (structure determination of PDI has historically been hindered by poor crystal formation). Although the field of cryo-EM has rapidly advanced to a stage where data can now be of a comparative resolution to these two methods, it is still better suited to the study of larger proteins (Bai et al., 2015). An advantage of NMR is that the collected spectra can provide

‘dynamic’ information of the protein, whereas x-ray and cryo-EM only supply ‘snapshot’ images of the protein in particular conformations, that may not be indicative of how the protein behaves in solution. NMR is particularly well-suited to study dynamic events such as ligand binding or a change in redox state, and as such is employed throughout this thesis.

An NMR spectrometer is very powerful in detecting the subtle changes of the local chemical environment a given NMR-active nuclei may experience. It is a technique made possible by a quantum property that is intrinsic to some nuclei, known as nuclear spin angular momentum, or simply ‘nuclear spin’. Certain isotopes of elements have a non-zero spin, a property that means the placement of such nuclei into a static magnetic field will cause the nuclei to interact with the field to induce a magnetic moment. The magnetic moment aligns either with or against the magnetic field, and resonates at a frequency related to the field strength of the applied magnet. For example, within a 14.1 Tesla magnet,  $^1\text{H}$  nuclei resonate with a frequency of approximately 600 MHz, and  $^{15}\text{N}$  nuclei with a frequency of 60.9 MHz. These resonance frequencies are usually quoted as a ‘chemical shift’ in parts per million (ppm), a unit that was derived to be independent of field strength, allowing the direct comparison of data collected using different spectrometers.

Nuclei that are typically studied by NMR are either naturally abundant isotopes (e.g.  $^1\text{H}$  and  $^{19}\text{F}$  have natural abundances of 99.99% and 100% respectively) or proteins can be artificially enriched with stable NMR-active nuclei by supplementing the growth media with  $^{13}\text{C}$  and/or  $^{15}\text{N}$  sources (natural abundance of these isotopes is otherwise only 1.11% and 0.36% respectively). The unique chemical shifts of  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^{19}\text{F}$  nuclei have been used within this thesis for both the assignment and subsequent ligand binding/redox studies of both hPDI protein fragments and the  $\Delta$ -som peptide ligand. Further theory detailing each particular experiment will be described in its relevant Chapter.

## 1.6 Project Aims

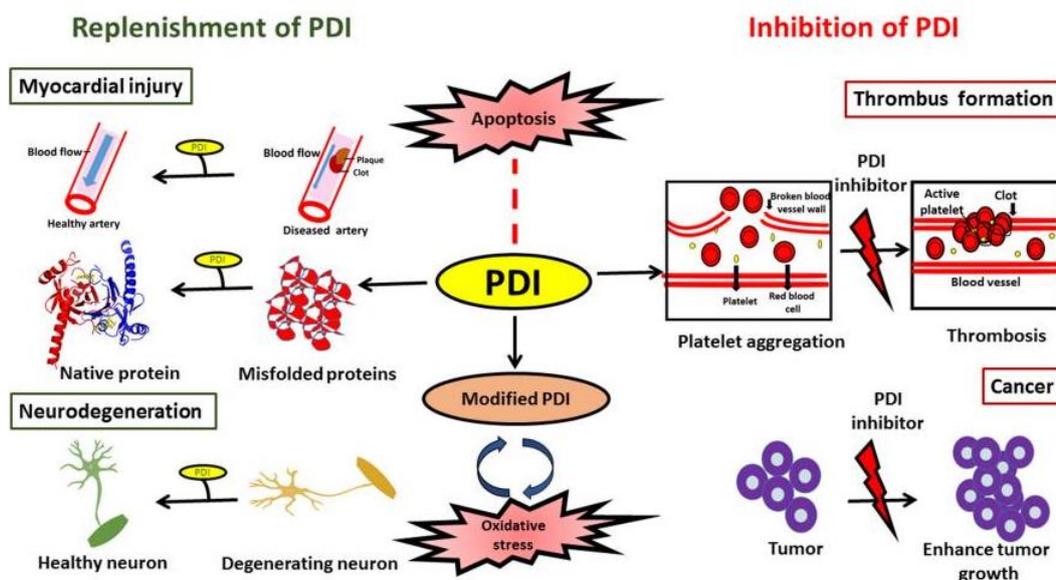
This work aims to further characterise the ligand binding ability and specificity of hPDI to its model binding peptide,  $\Delta$ -som. Early binding work carried out by Klappa et al. suggested the binding event relied upon hydrophobic interactions, and recently published work located the primary binding site of hPDI to a hydrophobic patch of the **b'** domain (Byrne et al., 2009). Since publication of Klappa et al.'s original work (Klappa et al., 1997), little has been reported of hPDI's peptide specificity. Obtaining more detailed information is of significant importance; hPDI is associated with several human disease states, but its physiological ligand specificity is still poorly understood.

The aims of this thesis were:

1. Recombinant expression, purification and isotopic labelling of the model binding peptide  $\Delta$ -somatostatin, and the hPDI fragments **b'x** and W111F **abb'x**.
2. Characterisation of hPDI **b'x** protein and the previously unassigned W111F **abb'x** fragment by NMR spectroscopy, to include: investigation of 'capping' by the **x-linker** region (a published observation seen in **b'x**), full backbone assignments by triple resonance experiments, and additional methyl side-chain assignment.
3. Determination of accurate and quantitative binding affinities of  $\Delta$ -som to hPDI by  $^{15}\text{N}/^1\text{H}$ ,  $^{13}\text{C}/^1\text{H}$ , and  $^{19}\text{F}$  NMR spectroscopy, to investigate the effects of temperature, neighbouring domain and protein redox state on binding affinity.
4. Elucidation of  $\Delta$ -som peptide ligand binding specificity to hPDI, using residue substitutions and specific peptide labelling techniques.
5. Investigation of hPDI redox potential and ligand binding affinity by  $^{19}\text{F}$  NMR spectroscopy (specific site incorporation of the  $^{19}\text{F}$  nucleus into the protein enables a simplified view of the protein redox state or binding event). In addition,  $^{19}\text{F}$  incorporation into the peptide allows a unique ligand-observed perspective of the binding event- this is explained further in Chapter 5.

## 1.7 Project Impact

It is anticipated that furthering our overall knowledge of hPDI will aid in a more general understanding of the structure/function based mechanism of hPDI, a protein that has been extensively studied for over 50 years, but until recently had eluded crystallisation. In addition, more detailed information on binding specificity may help in the rational design of hPDI targeted drugs. There are a multitude of potential therapeutic applications of hPDI, owing to its broad involvement in various fundamental cellular processes. Whilst many of these applications involve the up-regulation or restoration of normal hPDI cellular levels, the reduction or partial inhibition/inactivation of hPDI may be a more appropriate treatment for some disease states. Possible therapeutic applications have been suggested by Parakh et al. (see Figure 1.12).



**Figure 1.12: Possible therapeutic applications of hPDI modulation (taken from Parakh et al. 2015).** Restoring the function of modified hPDI or increasing the levels of hPDI may work to reduce the toxic load of misfolded proteins present in neurodegenerative disease states. hPDI treatment may also be valuable in cases of vascular injury, where the enzyme could help support initial platelet adhesion. Conversely, in cases of cancer or cardiovascular disease, it may be more appropriate to inhibit hPDI activity, where the enzyme inadvertently works to promote tumour growth and thrombus formation respectively.

Small molecular inhibitors of PDI are being researched for various different roles including neuroprotection (Kaplan and Stockwell, 2015), cancer treatments (Xu et al., 2012), (Eirich et al., 2014) prevention of thrombi (Khodier et al., 2010), as antivirals (Khan, Simizu et al., 2011), as well as general PDI inhibition (Ge et al., 2013). However, current small molecule inhibitors of PDI are mostly directed to the active sites of hPDI, and are plagued by cytotoxic effects and broad specificity. For example bacitracin, a peptide antibiotic used widely in research as an inhibitor of hPDI reduction, is not suitable for clinical use on account of its low specificity, nephrotoxicity and poor membrane permeability. In addition, commercially available bacitracin is supplied as no less than 22 structurally related peptides, with  $IC_{50}$  values of each peptide ranging between 20-1050  $\mu$ M (Dickerhof et al., 2011). It may be that if the peptide specificity of hPDI could be better understood a suite of peptide ligands could be designed, working to block the chaperone and isomerase activities of hPDI, to deliver a much more target specific, clinically practical treatment.

In addition to its diverse therapeutic applications, PDI is also used industrially to increase the yield of recombinantly produced protein therapeutics, many of which require formation of disulphide bonds for correct folding and biological activity (Bulaj, 2005), (Matos et al., 2014). A better knowledge of the substrate specificity of PDI might allow future engineering to increase PDI activity, or alternatively, if we had a better understanding of how the activity and specificity differ among the PDI family members, a more suitable PDI family member could be used instead.

## CHAPTER 2

# Expression and Purification of hPDI Binding Peptide $\Delta$ -som

### 2.1 Introduction

The production of small peptides for biological study is commonly carried out via solid-phase synthesis, a method developed in 1963 that works by progressively coupling chemically protected amino acids (residues that have their reactive side chains ‘blocked’ by other chemical groups to stop non-specific reactions or branching of the peptide) to the N-terminus of a growing peptide chain (Merrifield, 1963). The technique allows for the incorporation of either standard or non-natural amino acids, and whilst initial methods produced low yields and were labour intensive, solid-phase synthesis is now a common and cost-effective method of producing peptides <50 residues. The technique is utilised later within this thesis for the production of fluorine-containing peptides (Chapter 5).

For detailed study by NMR, it is preferable to have uniform isotopic enrichment of the sample (usually  $^{15}\text{N}$  and/or  $^{13}\text{C}$  labelling). Uniform labelling strategies allow multidimensional NMR and nucleus-selective editing techniques to be carried out. However, isotopic enrichment of synthetic peptides significantly increases their cost,

owing to the requirement of large excesses of each individually enriched and chemically protected amino acid. In addition to this cost, the purity of commercially available isotope-labelled amino acids is much lower than their unlabelled counterparts, making the purification of the isotope-labelled peptides extremely difficult from a crude mixture. On a small scale, it is not commercially viable to produce uniform labelled peptides in this manner (personal communication with Peptide Synthetics, [www.peptidesynthetics.co.uk](http://www.peptidesynthetics.co.uk)).

By comparison, recombinant *in vivo* isotopic labelling methods are much more cost effective. The *E. coli* expression system is widely exploited for protein expression, and equally the system lends itself well to stable isotope labelling (e.g.  $^{13}\text{C}$  and/or  $^{15}\text{N}$ ) by growth in a minimal media using a sole  $^{13}\text{C}$  carbon and/or  $^{15}\text{N}$  nitrogen source. However, owing to their small size and lack of secondary structure, peptides are readily degraded by proteases in the *E. coli* cytoplasm (Shen, 1984). This is a problem that can be circumvented by expression of the peptide fused to a suitable protein partner. A variety of different fusions can be employed, including soluble (for example GB1, GST and HSA binding domain of protein G) (Koenig et al., 2003), (Williamson et al., 2000), (Jonasson et al., 1998) or insoluble fusion partners (for example KSI) (Kuliopulos and Walsh, 1994), (also see (Hwang et al., 2014) for a good review).

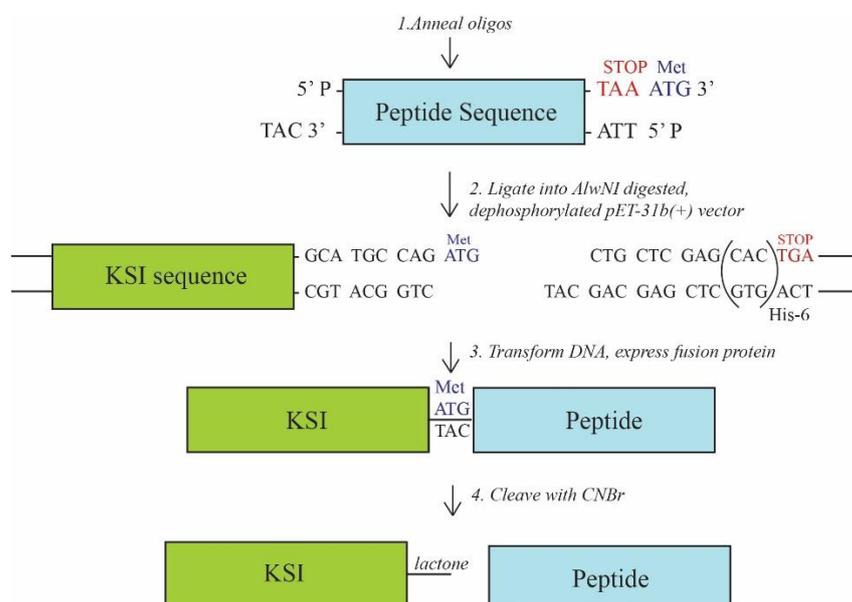
This chapter demonstrates attempts to express and purify  $\Delta$ -som (AGSKNFFWKTFSTSS), the model ligand binding peptide of hPDI, using both soluble and insoluble protein fusion partners. Expression of the fusion protein was successful using both systems, however, cleavage and purification of the peptide from an insoluble recombinant product was complicated by various factors including poor cleavage rates, residue formylation, peptide loss and peptide instability upon cleavage.

## **2.1.1 Recombinant Peptide Fusion Expression Strategies**

### **2.1.1.1 Insoluble Fusion Method (KSI)**

There are several insoluble proteins suitable for use as fusion partners, with the Ketosteroid Isomerase (KSI) expression system available commercially as the pET-31b(+) plasmid vector (a pBR322 derived plasmid), supplied by Novagen (Kuliopulos and Walsh 1994). The vector has a sequence encoding the insoluble bacterial protein Ketosteroid Isomerase (125 a.a), followed by an AlwNI restriction site and a C-terminal His<sub>6</sub>-tag (please see

Appendix 2.1 for pET-31b(+) vector map). The vector allows for the insertion of multiple tandem repeats of a peptide sequence (spaced by cyanogen bromide cleavable methionine residues) to increase overall yields. However, previous work within the laboratory has shown that tandem expression results in only a modest increase in peptide yield (unpublished observation) and the chemical cleavage by cyanogen bromide (CNBr) at the C-terminal side of methionine residues results in their modification to homoserine lactone, an undesired modification that would occur to all but the last tandem peptide. Therefore a stop codon was engineered immediately after the peptide coding sequence, to allow translation of a KSI-Met-peptide fusion, that when cleaved by CNBr would leave a KSI-lactone protein, and an unmodified peptide (see Figure 2.1).



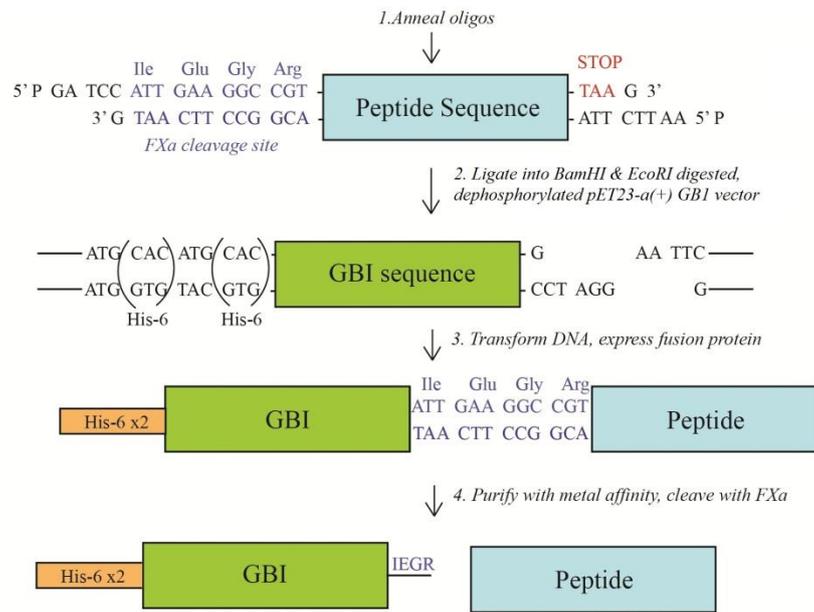
**Figure 2.1: Scheme showing pET-31b(+) vector KSI fusion expression system.** The annealed oligos are ligated into AlwNI digested, dephosphorylated vector to yield a KSI-Met-peptide-STOP construct. The methionine residues (cleavage site of CNBr) are highlighted in blue. The stop codons are highlighted in red. Figure adapted from (Kuliopulos and Walsh 1994).

There are many advantages to using an insoluble fusion method- insoluble proteins form inclusion bodies within the cytoplasm of the *E. coli*, which can greatly improve protein stability and hence expression yields as the insoluble nature of the aggregates protect the peptide from proteolysis (Kuliopulos and Walsh 1994). Using an inclusion body

purification method developed within the laboratory (Williamson et al., 1996), (Williamson et al., 2000), (Williamson, 2001), the inclusion bodies can be easily isolated from the majority of other soluble and contaminating membrane proteins to >80% purity, negating the requirement of an affinity tag for purification.

#### **2.1.1.2 Solubility Enhancing Fusion Method (GB1):**

A soluble GB1 fusion vector was designed within the laboratory, engineered using a pET-23a(+) vector kindly gifted by Prof. C. C. Wang, Beijing, China. The vector encodes for the soluble B1 domain of G protein (6.2 kDa), with an N-terminal His<sub>6</sub>-tag sequence (please note the vector incidentally codes for two sequential His<sub>6</sub>-tags; this was not found to affect expression or purification so was not altered). The GB1 sequence is followed by BamHI and EcoRI restriction sites (please see Appendix 2.2 and 2.3 for vector maps). Peptide coding sequences were designed for insertion between the two cloning sites and engineered with a proteolytic cleavage site (either Factor Xa (Ile-Glu/Asp-Gly-Arg-↓-X) or enterokinase (Asp-Asp-Asp-Asp-Lys-↓-X) and a stop codon, to allow translation of a His<sub>6</sub>-tag-GB1-protease cleavage site-peptide fusion. The expressed protein was purified using metal affinity chromatography and subsequently cleaved with the appropriate protease (see Figure 2.2).



**Figure 2.2: Scheme showing GB1 fusion expression system with Factor Xa cleavage site.** The annealed oligos are ligated into a BamHI and EcoRI digested, dephosphorylated vector to yield a His<sub>6</sub>-tag-GBI-peptide-STOP construct. The IEGR cleavage recognition site of FXa is labelled and the stop codon is highlighted in red.

Both Factor Xa and enterokinase were chosen due to their highly specific recognition sequences and apparent lack of secondary protease activity. More importantly, both enzymes cleave at the C-terminus of their respective recognition sequences, yielding an un-modified, cleaved peptide (Jenny et al., 2003), (Light and Janska, 1989).

## 2.2 Materials and Methods

**Please note:** The KSIΔ-som construct was made by an undergraduate student prior to the start of this thesis. Transformation of the vector, expression and purification of the KSIΔ-som fusion were carried out by myself. The GB1Δ-som constructs were made by undergraduate students under the supervision of Dr. Richard Williamson and myself (2014). Transformations, initial expression and lysis steps were carried out by myself, after which time I began my thesis write up. Larger scale expression, cleavage and characterisation of the GB1 fusions were carried out by further undergraduate students during this time (2015).

### 2.2.1 Materials

Unless stated otherwise all laboratory reagents were supplied by Sigma Aldrich (Gillingham, UK).

### 2.2.2 *E. coli* Strains and Vectors

Both modified pET vectors are under the control of a T7 promoter. The DH5 $\alpha$  *E. coli* strain was used for all DNA manipulation steps, and *E. coli* BL21 (DE3) pLysS cells were used for subsequent protein expression. These cells are deficient in the proteases *lon* and *ompT*, and carry the prophage  $\lambda$ DE3, which contains the T7 bacteriophage gene I. This codes for the T7 RNA polymerase enzyme under control of the isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) inducible *lac* UV5 promoter. T7 RNA polymerase binds specifically to the T7 promoter to initiate transcription. The pLysS plasmid carries the gene encoding T7 lysozyme- it confers chloramphenicol resistance to the cell, lowers 'leaky' background expression of the target protein before induction (T7 lysozyme is a natural inhibitor of T7 RNA polymerase) and has the additional benefit of aiding cell lysis after protein expression.

### 2.2.3 Bacterial Culture Media

Glycerol stocks and all transformed cells were plated out onto fresh LB agar plates (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl and 15 g/L agar) with antibiotic selection appropriate for each plasmid (ampicillin at 100  $\mu$ g/mL for pET vectors, chloramphenicol at 34  $\mu$ g/mL for pLysS). All *E. coli* cultures were grown using sterile LB media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl in MilliQ water) unless stated otherwise, again with the appropriate antibiotic selection in Erlenmeyer flasks 5 times the media volume.

### 2.2.4 Competent Cells

Competent DH5 $\alpha$  and BL21(DE3)pLysS *E. coli* cells were made for DNA transformation using a method adapted from Sambrooks and Russell (2001). Master glycerol stocks of each strain were streaked out onto LB agar plates with antibiotic selection if appropriate (chloramphenicol at 34 $\mu$ g/mL for BL21(DE3)pLysS strains), and incubated overnight at 37°C. A single colony was selected and used to inoculate 50 mL of sterile SOB media (20 g/L tryptone, 5 g/L yeast extract, 2 g/L NaCl, 10 mL/L of 250 mM KCl and 5 mL/L of 2M

MgCl<sub>2</sub>, pH 7). Cultures were incubated at 37°C with 200 rpm shaking until an A<sub>600</sub> of 0.5 was reached, at which point the cultures were transferred to sterile 50 mL falcon tubes and centrifuged at 1,811 g for 15 mins at 4°C. Each cell pellet was re-suspended in 25 mL of sterile, ice-cold 50 mM CaCl<sub>2</sub> and incubated on ice for 30 min. Cells were then centrifuged as before and each pellet re-suspended in 2 mL ice-cold 50 mM CaCl<sub>2</sub>. Sterile 50% glycerol was added to each suspension to a final concentration of 10% glycerol, and the suspension portioned into sterile Eppendorfs in 200  $\mu$ L aliquots, flash frozen using liquid nitrogen or an ethanol/dry ice bath and stored at -80°C until required.

### **2.2.5 Transformations**

Vector DNA was transformed into competent DH5 $\alpha$  and BL21(DE3)pLysS *E. coli* cells prepared as stated in 2.2.4. Competent cells stored at -80°C were defrosted on ice and split into sterile, pre-chilled Eppendorf tubes (60  $\mu$ L aliquot per transformation). Purified miniprep DNA concentrations were calculated using a NanoDrop 2000c UV-Vis Spectrophotometer (Thermo Scientific); competent DH5 $\alpha$  cells were transformed using 50 ng DNA, and competent BL21(DE3)pLysS cells transformed with 150 ng DNA (BL21 cells have reduced competency when compared to DH5 $\alpha$ ). After the addition of vector DNA, cells were incubated on ice for 10 mins, followed by a heat-shock at 42°C for 45 secs. After a second incubation on ice for 5 mins, 150  $\mu$ L of sterile LB media was added to each tube and incubated for 1 hr at 37°C with 200 rpm shaking. 10, 25, 50 and 100  $\mu$ L aliquots of transformed cells were then spread onto LB agar plates containing appropriate antibiotics and incubated overnight at 37°C. In addition, a negative control transformation (using 5  $\mu$ L of sterile MilliQ water), and a positive control (vector known to transform with high efficiency) were transformed each time to check overall transformation efficiency.

### **2.2.6 Fusion Protein Expression**

For main grows, starter cultures of 100 mL LB media with ampicillin (100  $\mu$ g/mL) and chloramphenicol (34  $\mu$ g/mL) were inoculated using a single colony from LB agar plates. The starter culture was grown overnight at 37°C with 200 rpm shaking, to an A<sub>600</sub> of approximately 2.0. The cells were then centrifuged in sterile 50 mL falcons at 3,220 g for 15 mins at 4°C, and the cell pellets re-suspended in 1-5 mL of fresh, sterile LB media. Each large main culture (up to 3x 400 mL cultures from 1x 100 mL starter culture) was then inoculated from the starter culture re-suspension to give a starting A<sub>600</sub> of 0.1.

The main cultures were incubated at 37°C with shaking at 200 rpm and monitored until they reached an  $A_{600}$  of 0.7. They were then induced using 1 mM IPTG (1 M stock). After induction at 37°C for 4 hours, the cells were cooled on ice for 5 mins and then harvested by centrifugation at 10,000 g for 10 mins at 4°C. Each 400 mL cell pellet was then re-suspended in 10 mL peptide lysis buffer (KSI fusions: 50 mM Tris-HCl, 2 mM EDTA, pH 8.0, GB1 fusions: 20 mM sodium phosphate, 50 mM NaCl, pH 7.3) and stored at -20°C until required.

### 2.2.7 SDS-PAGE Analysis

Reducing SDS-PAGE gels were used at all stages to characterise expression, lysis and subsequent purification steps of the fusion proteins.

**Reducing SDS-PAGE:** Protein gels were prepared using the Laemmli Tris-Glycine buffer system (Laemmli, 1970), using a 12.5 or 15% w/v resolving gel (pH 8.8) and a 5% w/v stacking gel (pH 6.8) (see Table 2.1 for gel composition).

**Table 2.1: Components of reducing SDS-PAGE resolving and stacking gels.**

Component	15% Resolving Gel (/20 mL)	12.5% Resolving Gel (/20 mL)	5% Stacking Gel (/10 mL)
<b>40% Acrylamide</b> (37.5:1 A:B)	7.5 mL	6.25 mL	1.25 mL
<b>1.5 M Tris-HCl, pH 8.8</b>	5 mL	5 mL	-
<b>0.5 M Tris-HCl, pH 6.8</b>	-	-	1.25 mL
<b>10% w/v SDS</b>	200 $\mu$ L	200 $\mu$ L	100 $\mu$ L
<b>MilliQ</b>	7.3 mL	8.55 mL	7.4 mL
<b>10% AMPS</b>	50 $\mu$ L	50 $\mu$ L	50 $\mu$ L
<b>TEMED</b>	14 $\mu$ L	14 $\mu$ L	10 $\mu$ L

A- acrylamide, B- bisacrylamide, SDS- sodium dodecyl sulphate, AMPS- ammonium persulphate and TEMED- tetramethylethylenediamine.

Gels were polymerised using ammonium persulphate (AMPS) and tetramethylethylenediamine (TEMED) (both Bio-Rad, Hertfordshire, UK). Twenty millilitres of resolving gel was sufficient to pour 2-3 gels in 1.0 mm cassettes (Novex, Thermo Fisher Scientific). The gels were left for 3 hours at RT to polymerise. Gel samples were prepared for analysis using 4x gel loading buffer (0.33 M Tris-HCl, 8% w/v SDS, 0.4% Bromophenol blue, glycerol to a final concentration of 40%- Sambrook & Russell, 2001) made reducing with 200 mM dithiothreitol (DTT<sup>red</sup>) and incubated at 100°C for 3 mins. Ten microliter sample volumes were loaded, with all sample loads equivalent to material derived from 50  $\mu$ L of original culture volume (unless stated otherwise). Each gel was run with 10  $\mu$ L Precision Plus protein standard (Bio-Rad) in the first lane and empty wells were run with 10  $\mu$ L 1x reducing sample buffer. Gels were run using 500 mL 1x TGS gel running buffer (5x stock of 0.95 M glycine, 0.125 M Tris and 0.5% w/v SDS, pH 8.3) at 100 V through the stacking gel (15-20 mins) and 125 V through the resolving gel (approximate run time 1.5 hrs). Gels were typically stained overnight in Coomassie stain (40% methanol, 10% acetic acid and 1 mg/mL Brilliant Blue G) and then de-stained using a solution of 10% methanol, 10% acetic acid. All steps were carried out under gentle agitation at RT. Once de-stained, gels were stored in de-stain solution or equilibrated in 30% methanol, 3% glycerol for 1 hour and dried between cellophane sheets.

**Reducing ‘Peptide Gel’ SDS-PAGE:** As  $\Delta$ -som peptide is significantly small (1.6 kDa) it was not observed by standard SDS-PAGE. Therefore peptide gels were prepared, using a modified version of the original Tris-Tricine buffer system (Schagger and von Jagow, 1987), adapted for the resolution of small peptides (<10 kDa) by inclusion of 6 M urea in the resolving gel (Schagger, 2006). Gels were made using a 16% w/v resolving gel, a 1 cm 10% w/v ‘spacer’ gel and a 4% w/v stacking gel (see Table 2.2 for gel composition).

**Table 2.2: Components of reducing peptide Tris-Tricine SDS-PAGE.**

<b>Component</b>	<b>16% Resolving Gel (/10 mL)</b>	<b>10% Spacer Gel (/10 mL)</b>	<b>4% Stacking Gel (/6 mL)</b>
<b>AB-6 Polyacrylamide</b> (49.5:6 A:B)	3.86 mL	-	-
<b>AB-3 Polyacrylamide</b> (49.5:3 A:B)	-	2.49 mL	0.6 mL
<b>3x Gel Buffer</b> (3M Tris, 1M HCl, 0.3% SDS, pH 8.45)	3.33 mL	3.33 mL	1.5 mL
<b>Urea</b>	3.6 g	-	-
<b>MilliQ</b>	2.81 mL	4.18 mL	3.9 mL
<b>10% AMPS</b>	33.3 $\mu$ L	50 $\mu$ L	45 $\mu$ L
<b>TEMED</b>	3.33 $\mu$ L	10 $\mu$ L	4.5 $\mu$ L

A- acrylamide, B- bisacrylamide, SDS- sodium dodecyl sulphate, AMPS- ammonium persulphate and TEMED- tetramethylethylenediamine.

Gel samples were prepared for analysis using 4 x peptide gel loading buffer (3% SDS (w/v), 6%  $\beta$ -mercaptoethanol, 30% glycerol (w/v), 0.05% Bromophenol blue, 150 mM Tris-HCl, pH 7.0). Samples and buffer were diluted to 1 x with MilliQ and incubated at 37°C for 60 mins. Gels were run using 400 mL 1 x anode buffer (10 x stock of 1 M Tris and 225 mM HCl, pH 8.9) and 200 mL 1 x cathode buffer (10 x stock of 1 M Tris, 1 M Tricine, 1% SDS, pH 8.25), run at 30 V through the stacking gel, and 125 V through the spacer and resolving gels (approximate run time 2 hrs). The gels were fixed for 30 mins (100 mM ammonium acetate in a 50:40:10 solution of methanol: H<sub>2</sub>O: acetic acid), then stained for 60 mins using Coomassie stain (0.025% Brilliant Blue G in 10% acetic acid), and de-stained in 2 x 60 min steps using a 10% acetic acid solution. All steps were carried out under gentle agitation at RT. Once de-stained, gels were stored in de-stain solution or equilibrated in 30% methanol, 3% glycerol for 1 hr and dried between cellophane sheets.

### **2.2.8 Cell Lysis**

Previously frozen cells were thawed at RT, and each 10 mL re-suspended pellet (equivalent to 400 mL original culture volume) was treated with 200  $\mu$ L of Triton® X-100 (5% v/v in lysis buffer stock), gently mixed and left to incubate for 20 mins. This was followed by treatment with 100  $\mu$ L DNase I (Bovine, 2 mg/mL stock) and 50  $\mu$ L 2 M MgCl<sub>2</sub> for a second incubation at RT for 20 mins. The lysate was then sonicated on icy water (3 min cycle of 10 sec pulse, 10 sec rest) and the insoluble material pelleted by centrifugation for 10 mins, 12,000 g at 4°C. Samples of the total cell lysate, supernatant and re-suspended pellet were taken for reducing SDS-PAGE analysis. Either the insoluble (KSI fusion protein) or soluble (GB1 fusion protein) fractions were taken forward for purification.

### **2.2.9 KSI Fusion Protein Isolation**

The insoluble material was re-suspended in 2.5 mL wash buffer (50 mM Tris-HCl, 10 mM EDTA and 0.5% Triton X-100 (w/v), pH 8.0) and homogenised for 5 mins using a hand-held glass homogeniser. The insoluble material was recovered by centrifugation (10 mins, 12,000 g at 4°C), and the homogenisation step repeated twice more using 2.5 mL wash buffer, followed by two steps using 2.5 mL MilliQ water. A sample of washed inclusion bodies was taken for SDS-PAGE analysis and the remaining material stored as a wet pellet at -20°C.

### **2.2.10 KSI Fusion Protein Cleavage and Peptide Purification**

Washed inclusion bodies from up to 1.2 L original culture volume were re-suspended in 0.5 mL MilliQ water and dissolved by addition of either 5 mL formic acid or TFA. The solubilised fusion protein mix was poured into a falcon tube containing 0.2 g CNBr, mixed by inversion and left to incubate at RT in the dark fume hood for 18-24 hours.

The cleaved protein solution was diluted into 20 mL MilliQ water, flash frozen on liquid nitrogen and lyophilised to remove CNBr. The lyophilised material was then re-suspended in 2.5 mL phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, pH 7.5), and corrected to pH 7.5 if necessary. The suspensions were left to stir overnight in the dark at 4°C, and free peptide was recovered by centrifugation (15,700 g for 5 mins at RT).

Additionally, recovery of  $\Delta$ -som peptide was trialled by serial dilution from denaturing phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, 6 M guanidine hydrochloride (GdnHCl), pH 7.5) into non-denaturing phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, pH 7.5). Lyophilised material was initially dissolved into 1 mL denaturing buffer, and 100  $\mu$ L aliquots were then each made up to 1 mL with varying amounts of each buffer to yield final GdnHCl concentrations of 6, 5, 4, 3.4, 3, 2.7, 2.4, 1.2 and 0.6 M. Samples were monitored using absorbance at 280 nm and KSI pellets were subsequently analysed by reducing SDS-PAGE. The samples were centrifuged at 15,700 g for 5 mins at RT and the pellets washed with 1 mL fresh non-reducing buffer to remove residual denaturant (GdnHCl precipitates SDS). Re-suspended KSI pellets were then centrifuged as before and taken up in 1 x sample buffer.

Recovery of  $\Delta$ -som peptide was also trialled by size exclusion chromatography under denaturing conditions using a 50 mL CV of Sephadex G15 medium, a matrix of cross-linked dextran (GE Healthcare). The column was stored at RT in de-gassed 20% ethanol and equilibrated using 1.5 CV de-gassed running buffer (50 mM sodium phosphate, 100 mM NaCl, either 6 M or 2.4 M GdnHCl, pH 7.5). The lyophilised material was re-suspended in 1 mL running buffer and loaded onto the column using a three way tap. The peptide was eluted using 1.5 CV of running buffer at a flow rate of 2 mL/min, collected in 5 mL elution fractions and monitored using absorbance at 280 nm by an ÄKTA FPLC analyser. After use the column was washed using 3 CV of de-gassed 20% ethanol and stored at RT for future use.

Unlabelled peptide was purified by RP-HPLC (Waters 600/486 series HPLC with a preparative Vydac Resolve C18 column). Peptide was loaded and eluted using a gradient of acetonitrile (HPLC-grade) in HPLC-grade water (+ 0.05% TFA). Peak fractions were collected in 15 mL falcon tubes when the absorbance at 200 nm  $>0.1$  AU, lyophilised and re-suspended in phosphate buffer. Peptide absorbance was measured at this stage using a wavelength of 280 nm and a phosphate buffer blank. The  $\Delta$ -som peptide amino acid sequence was submitted to ExPASy ProtParam (<http://web.expasy.org/protparam/>) to calculate extinction coefficients, and molar protein concentration was calculated using the Beer-Lambert Law. The extinction coefficient of the peptide was 5500  $M^{-1}cm^{-1}$ . Additionally, samples of purified peptide were analysed by SDS-PAGE gel for comparison with known peptide standards.

### 2.2.11 GB1 Fusion Protein IMAC Purification and Cleavage

An equivalent of up to 1.2 L original cell culture lysate was loaded per 5 mL packed column of Chelating Sepharose Fast Flow (GE Healthcare), using multiple columns when scaling up. The re-usable columns were stored at 4°C in 20% ethanol, and immediately prior to use were washed using 10 CV of MilliQ water. All following steps were loaded and eluted using gravity flow, and fractions from each elution were collected for analysis by reducing SDS-PAGE. All wash and elution buffers were used ice-cold to limit the proteolysis observed after initially using buffers at RT.

The resin was first charged using 1 CV of 0.2 M NiSO<sub>4</sub> and washed using 5 CV sodium acetate buffer (20 mM sodium acetate, 0.5 M NaCl, pH 4). The column was then equilibrated using 10 CV of binding buffer (20 mM sodium phosphate, 0.5 M NaCl, 10 mM imidazole, pH 7.3). The supernatant was filtered using a 0.2  $\mu$ m Minisart single use syringe filter and loaded onto the column. A second 10 CV of binding buffer was used to wash the column followed by 5 CV of wash buffer (20 mM sodium phosphate, 0.5 M NaCl, 50 mM imidazole, pH 7.3), used to remove any weakly bound impurities. The nickel and His<sub>6</sub>-tag fusion protein were then eluted using 10 CV of strip buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM EDTA, pH 7.3) and collected in 2 mL fractions. The absorbance of each fraction was measured at A<sub>280</sub> nm using a wash buffer blank and a cut-off of 0.1 AU to select fractions for reducing SDS-PAGE analysis and subsequent steps. After use the column was washed using 10 CV of MilliQ water and again stored in 20% ethanol at 4°C for future use.

Strip buffer elutions containing His<sub>6</sub>-tagged protein were treated according to their respective cleavage site, with the addition of either 0.05 U/ mL enterokinase, incubated at 37°C for 2 hours, or 5  $\mu$ g/ mL Factor Xa, incubated at 25°C for 16 hours.

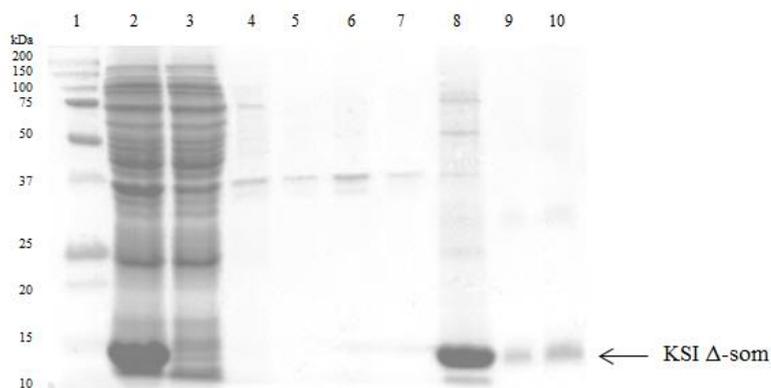
### 2.2.12 Peptide Analysis by Mass Spectrometry

To confirm the correct peptide had been produced, portions of each recovered peptide were analysed using Electrospray Mass Spectrometry (coupled to analytical HPLC), kindly performed by Kevin Howland, Biomolecular Science Facility Manager, University of Kent.

## 2.3 Results

### 2.3.1 Expression and Purification of $\Delta$ -som from KSI Fusions

pET 31b(+) KSI $\Delta$ -som DNA was previously generated by an undergraduate project student prior to the start of this thesis, and the sequence confirmed by DNA sequencing (see Appendix 2.4.1). This DNA was transformed into DH5 $\alpha$  cells to maintain glycerol stocks, and the plasmid DNA from an overnight culture of these cells was purified by mini-prep (QIAGEN), then transformed into BL21(DE3)pLysS cells for protein expression. Growth of *E. coli* cells with the insert produced white, uniform colonies on LB agar plates, and showed normal doubling times in both LB and minimal media. SDS-PAGE analysis of the induced cell lysate showed a prominent band of the overexpressed fusion protein, which as anticipated was only present in the insoluble fraction (see Figure 2.3). On comparison of the fusion protein with the Precision Plus protein standards, the fusion appears to migrate between the 10-15 kDa standards. KSI has a molecular weight of 13.6 kDa and the  $\Delta$ -som insert is 1.6 kDa (giving a total molecular weight of 15.2 kDa)- so the fusion protein migrates faster than would be expected for its size. This phenomena was also observed with the GB1 $\Delta$ -som fusion (see Figure 2.11), suggesting fusion of  $\Delta$ -som to either protein may cause faster migration through the gel.

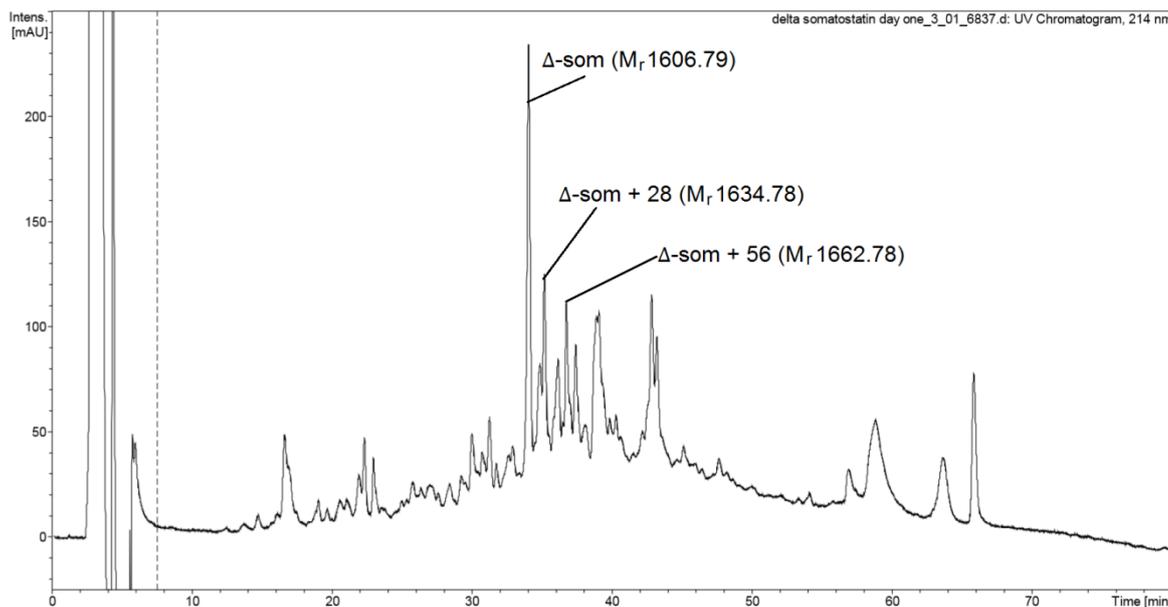


**Figure 2.3: 15% Reducing SDS-PAGE analysis of KSI $\Delta$ -som cell lysis and inclusion body wash steps.** Lane 1: Precision Plus protein marker, lane 2: total cell lysate, lane 3: lysis supernatant, lanes 4-7: supernatants from inclusion body wash steps 1-4, lane 8: Washed inclusion bodies, lane 9 and 10: 1  $\mu$ g and 2  $\mu$ g N-TIMP-2 protein standard (14.1 kDa).

The vast majority of contaminating soluble proteins were removed by centrifugation of the lysate to pellet the insoluble inclusion bodies (see Figure 2.3 lane 3, lysis supernatant). The inclusion body wash steps were successful in removing other contaminating proteins, and after the four wash steps KSI $\Delta$ -som was estimated to be >80% pure. The major contaminating band removed in the wash step elutions (see lanes 4-7 of Figure 2.3) is likely  $\beta$ -lactamase, a 40 kDa protein expressed by plasmid-containing cells (the enzyme confers ampicillin resistance). This band was especially prominent as a contaminant in washed inclusion bodies from overnight expressions, therefore induction was carried out at 37°C for 4 hrs (where comparable fusion protein expression was observed as for overnight induction, but with lower  $\beta$ -lactamase expression).

From the SDS-PAGE gel in Figure 2.3 we can estimate an excess of 4  $\mu$ g washed fusion protein per 50  $\mu$ L of original culture, giving a yield of >80 mg/L. As  $\Delta$ -som peptide constitutes 10.53% of the total fusion product by mass, the maximum theoretical yield of recoverable peptide is >8 mg/L.

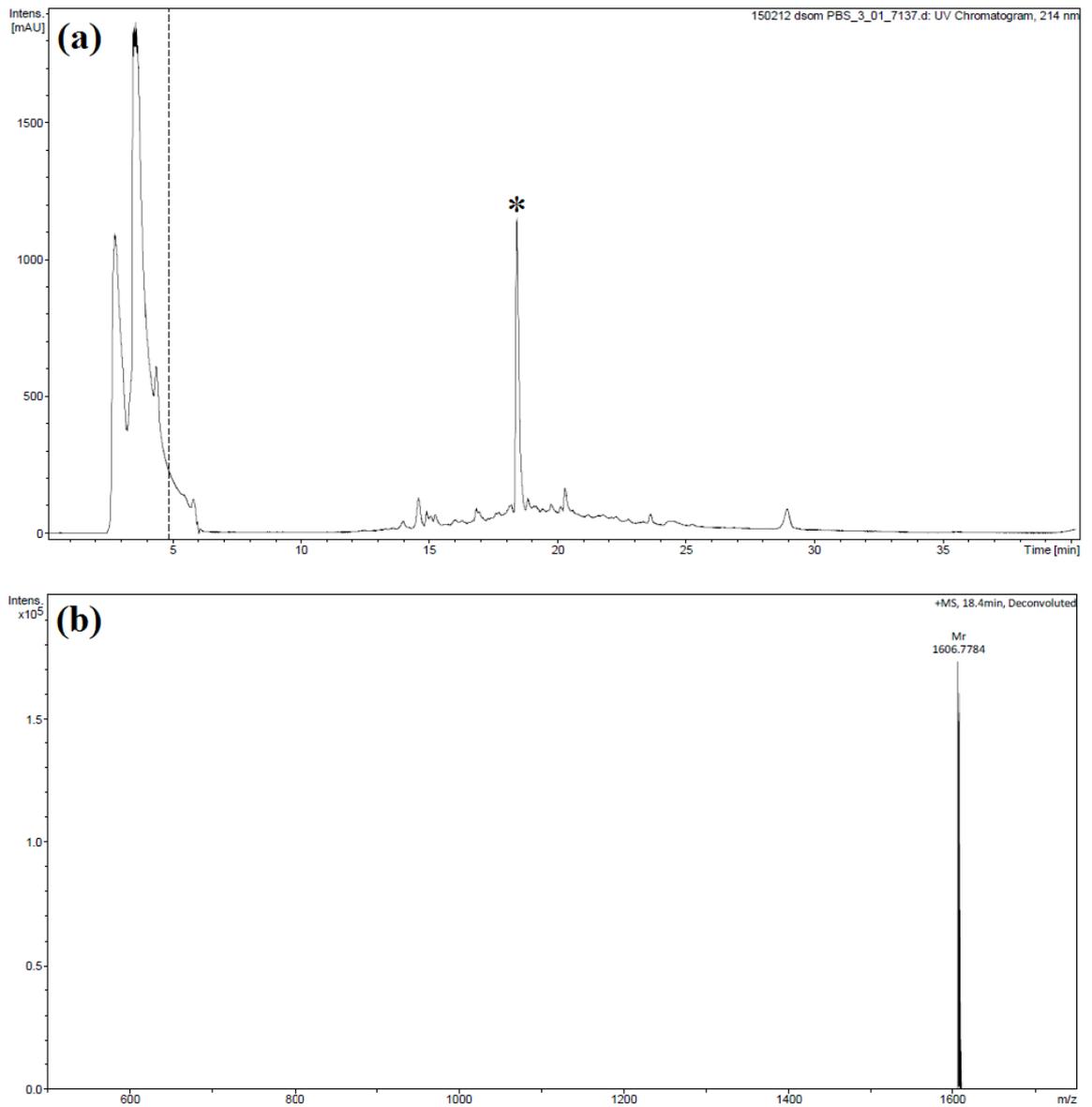
Cleavage of  $\Delta$ -som peptide from KSI was first carried out as described in 2.2.10 using CNBr in formic acid. After lyophilisation, the material was re-suspended in phosphate buffer and centrifuged to remove the insoluble KSI. Analytical RP-HPLC analysis of the soluble fraction and corresponding masses (as identified by mass spectrometry) are shown in Figure 2.4.



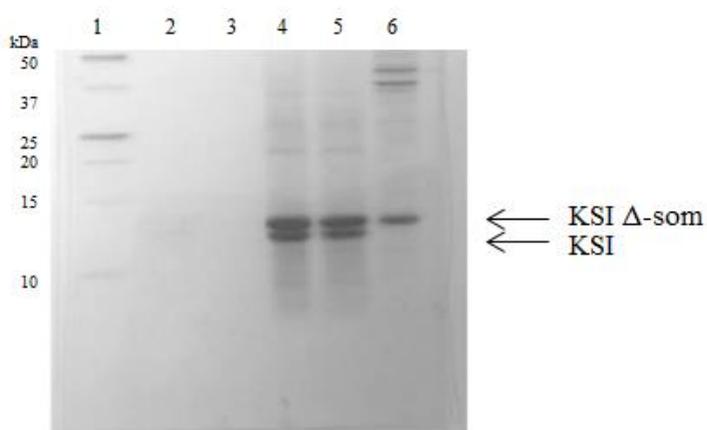
**Figure 2.4:** HPLC trace of  $\Delta$ -som after cleavage from KSI using CNBr in formic acid. Corresponding masses of each relevant peak are labelled.

Whilst the major peak observed was identifiable as  $\Delta$ -som peptide (1606.8 Da), additional peaks were also seen in the analysis that eluted soon after the peptide in the HPLC trace. Mass spectrometry analysis revealed masses of 1634.78 and 1662.78 Da, corresponding to  $\Delta$ -som peptide +28 or +56 Da. Probing of these masses using Delta Mass ([www.abrf.org](http://www.abrf.org)) showed them to be a common post-translational modification observed when using formic acid (Goodlett et al., 1990); an N,N-di-methylation that can occur to either lysine or arginine residues. Since there are two lysine residues present in the peptide, it is likely that the observed masses correspond to single di-methylation (+28 Da) and double di-methylation (+56 Da).

As a result of this, shorter cleavage times using formic acid were trialled (2, 4, 8 and 16 hrs) but these times also showed formylation of the peptide. Substitution of the formic acid for TFA prevented formylation from occurring (see Figure 2.5), however overall cleavage rates were much poorer (estimated at 50% cleavage, as shown by Figure 2.6). Extended cleavage times in the TFA solution (>24 hrs) were trialled to improve this cleavage rate, but resulted in degradation of the peptide, as observed by mass spectrometry.



**Figure 2.5:** HPLC trace of  $\Delta$ -som after cleavage from KSI using CNBr in TFA acid (a) and corresponding mass identification of major peak \* in HPLC trace (b).  $\Delta$ -som peptide was identified as the major species with a mass of 1606.8 Da (expected 1607.7 Da).



**Figure 2.6: 16% Reducing peptide SDS-PAGE analysis of KSI $\Delta$ -som cleavage in TFA.** Lane 1: Precision Plus protein marker, lanes 2 and 3: supernatant of re-suspended, cleaved material (should contain  $\Delta$ -som peptide but peptide is extremely dilute), lanes 4 and 5: re-suspended pellet of cleaved material (cleaved and un-cleaved fusion protein present), lane 6: washed inclusion bodies (un-cleaved KSI $\Delta$ -som). A cleavage rate of 50% in TFA conditions was estimated from the two KSI bands (cleaved and un-cleaved) in lanes 4 and 5. All loads are equivalent to the material derived from 200  $\mu$ L of original culture.

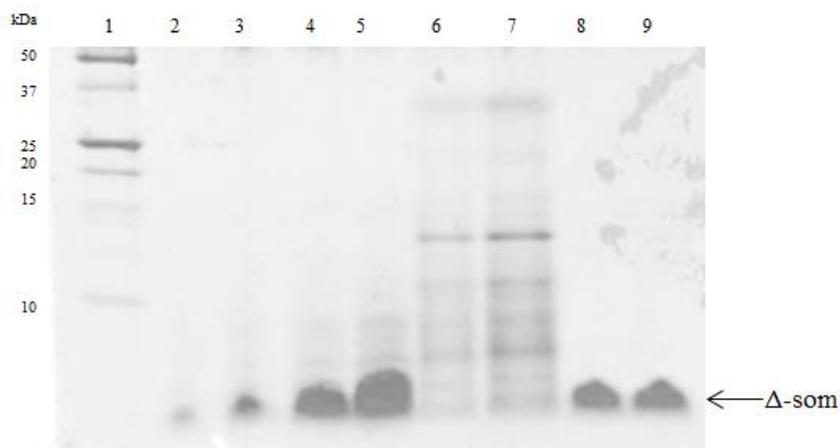
Despite large losses at this cleavage step, isotopic enrichment was attempted for the expression of  $^{15}\text{N}$ ,  $^{13}\text{C}$  labelled fusion protein. Expression in defined minimal media (MM) (see Table 3.1 for components) typically results in lower protein expression, and expression of  $^{15}\text{N}$ ,  $^{13}\text{C}$  labelled KSI $\Delta$ -som was estimated at 20 mg/L (for comparison expression in LB was estimated at 80 mg/L). Typically a drop of 30-40% expression is observed in the laboratory upon change to MM for  $^{15}\text{N}$  labelling, however the drop in expression reported here is much lower. This is likely due to the addition of  $^{13}\text{C}$  labelling, where the  $^{13}\text{C}$  source is added to the MM at only half the quantity standard  $^{12}\text{C}$  source is added (this is due to cost;  $^{13}\text{C}$  D-glucose is around £150/g, and is required at 2 g/L).

In addition to this drop in expression (if 50% cleavage is assumed as per reported for LB preparations, MM would yield  $\sim$ 1 mg/L cleaved peptide), peptide recovery from the overnight re-solubilisation step from MM expressions was very poor and not reproducible. It was found once cleaved the peptide adhered to the glassware and consumable plastic used, and cleaved peptide was also identified in the centrifuged KSI pellet. Analytical HPLC of an acetonitrile wash of a falcon tube used in the overnight re-solubilisation step

estimated an almost equal amount of peptide was ‘sticking’ to the inside of the falcon tube as was recovered from the supernatant fraction.

It is likely that peptide expressed in LB, once cleaved, behaved similarly.  $A_{280}$  nm readings of RP-HPLC purified peptide from a 400 mL LB culture using the standard overnight re-solubilisation technique gave an estimated peptide yield of 1.93 mg/L, a recovery rate of just 24% from total peptide expressed as fusion protein. For comparison, this method was used previously within the laboratory to produce the peptide A20fmdv2 (a 20-mer, 2.2 kDa in size) from LB cultures with an average yield of 16 mg/L (Wagstaff et al., 2010).

RP-HPLC purified  $\Delta$ -som peptide from LB cultures along with attempts to purify isotopically enriched  $\Delta$ -som from MM cultures is visualised by reducing peptide SDS-PAGE in Figure 2.7, with synthetically produced peptide run for comparison.



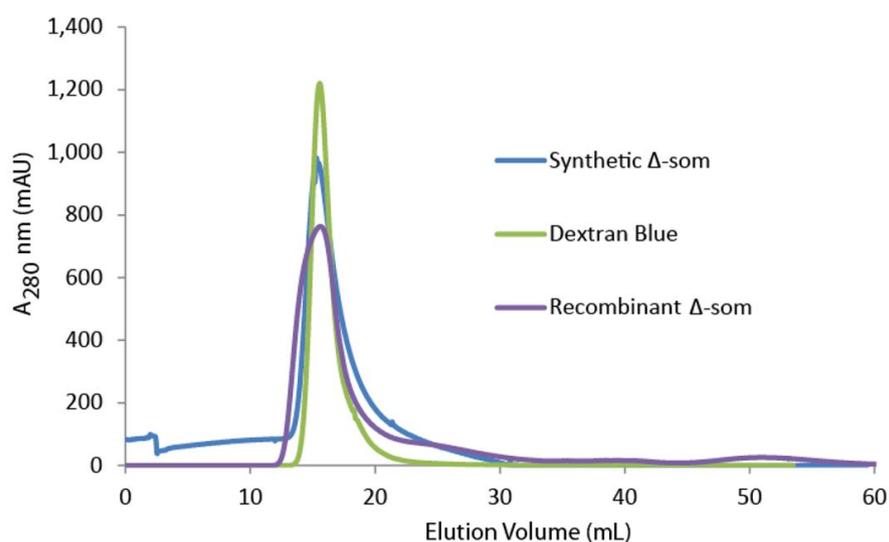
**Figure 2.7: 16% Reducing peptide SDS-PAGE analysis of  $\Delta$ -som purified from LB and MM cultures.**

Lane 1: Precision Plus protein marker, lanes 2-5: 1, 5, 10 and 20  $\mu$ g RP-HPLC purified recombinant  $\Delta$ -som (from a 400 mL LB culture), lanes 6 and 7:  $\sim$ 10 and  $\sim$ 20  $\mu$ g equivalent loads of soluble protein from overnight re-solubilisation step after cleavage (from a 400 mL MM culture- no identifiable  $\Delta$ -som peptide present). Lanes 8 and 9: 5 and 10  $\mu$ g synthetic  $\Delta$ -som peptide for comparison.

Attempts were therefore directed to improving peptide separation from KSI after cleavage. Instead of resuspension in PBS, the lyophilised, cleaved material was re-suspended in a

denaturing 6 M GdnHCl buffer, to solubilise both the insoluble KSI protein and  $\Delta$ -som peptide. This was followed by serial dilution of the material from 6 M down to a 0.6 M GdnHCl buffer. KSI is highly insoluble in the absence of denaturant, and began dropping out of solution in samples containing <4 M GdnHCl.  $A_{280}$  nm values of the samples after centrifugation (to remove insoluble KSI) reported similar absorbances from samples containing <2.4 M GdnHCl, indicating that KSI was largely insoluble at these concentrations of denaturant, and composed the centrifuged pellet.

Size exclusion chromatography under denaturing conditions was performed, using either 2.4 M or 6 M GdnHCl buffers, in an attempt to resolve denatured peptide from the supernatant alone (2.4 M), or resolution of both peptide and solubilised KSI from the total re-suspension (6 M). Unfortunately the peptide did not appear to interact with the column matrix, and the peptide eluted in the column void volume along with KSI (shown in Figure 2.8). This was confirmed using Dextran Blue (2 MDA) to determine the void volume of the column and 1 mg synthetic  $\Delta$ -som peptide as an elution standard.



**Figure 2.8: Gel Filtration elution profile of cleaved KSI and  $\Delta$ -som using a 50 mL G-15 Superdex column under denaturing conditions (2.4 M GdnHCl). 1 mg synthetically produced  $\Delta$ -som peptide (blue), 1 mg Dextran Blue (green) and 200 mL equivalent load of cleaved  $\Delta$ -som peptide (purple).**

Mass spectrometry collected using the soluble fraction of a 0.6 M GdnHCl buffer re-suspension identified intact peptide, but the solution also contained a greater proportion of solubilised contaminants than was observed in samples re-suspended in PBS, indicating a size exclusion based separation step would still be preferable prior to RP-HPLC.

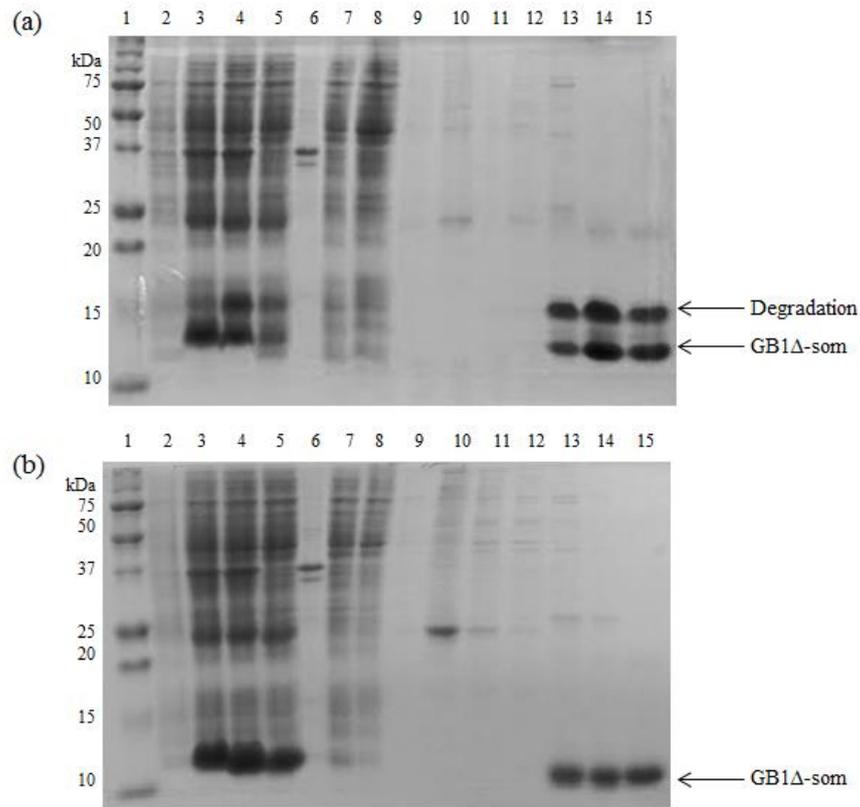
After discussions with Mr. Kevin Howland (Biomolecular Science Facility Manager, University of Kent) it became evident that a dedicated RP-HPLC column would be required to purify peptides under denaturing conditions, as denaturants can alter the column retention time. In addition, Mr. Howland did not suggest purifying samples of the peptide on a RP-HPLC column if they had any possibility of containing solubilised KSI or KSI fragments, which could become insoluble upon removal of the denaturant in the mobile phase, thus blocking the column.

This method was deemed impractical and an uneconomic use of stable isotopes to produce the quantity of isotopically enriched peptide as required for NMR. Consequently, attempts to produce recombinant peptide using an alternative fusion partner were made.

### **2.3.2 Expression & Purification of $\Delta$ -som from GB1 Fusions**

pET-23a(+) GB1 $\Delta$ -som DNA with either enterokinase or Factor Xa cleavage sites were generated by undergraduate project students in Summer 2014 (please see Appendix 2.4 for sequencing). This DNA was transformed as outlined in 2.3.1 for KSI $\Delta$ -som. Growth of *E. coli* cells with either insert produced white, uniform colonies on LB agar plates, and showed normal doubling times in LB. SDS-PAGE analysis of the induced cell lysate showed prominent bands for both overexpressed fusion proteins, which as anticipated were present in the soluble fraction (see Figure 2.9 for example expression and IMAC of GB1-FXa- $\Delta$ -som).

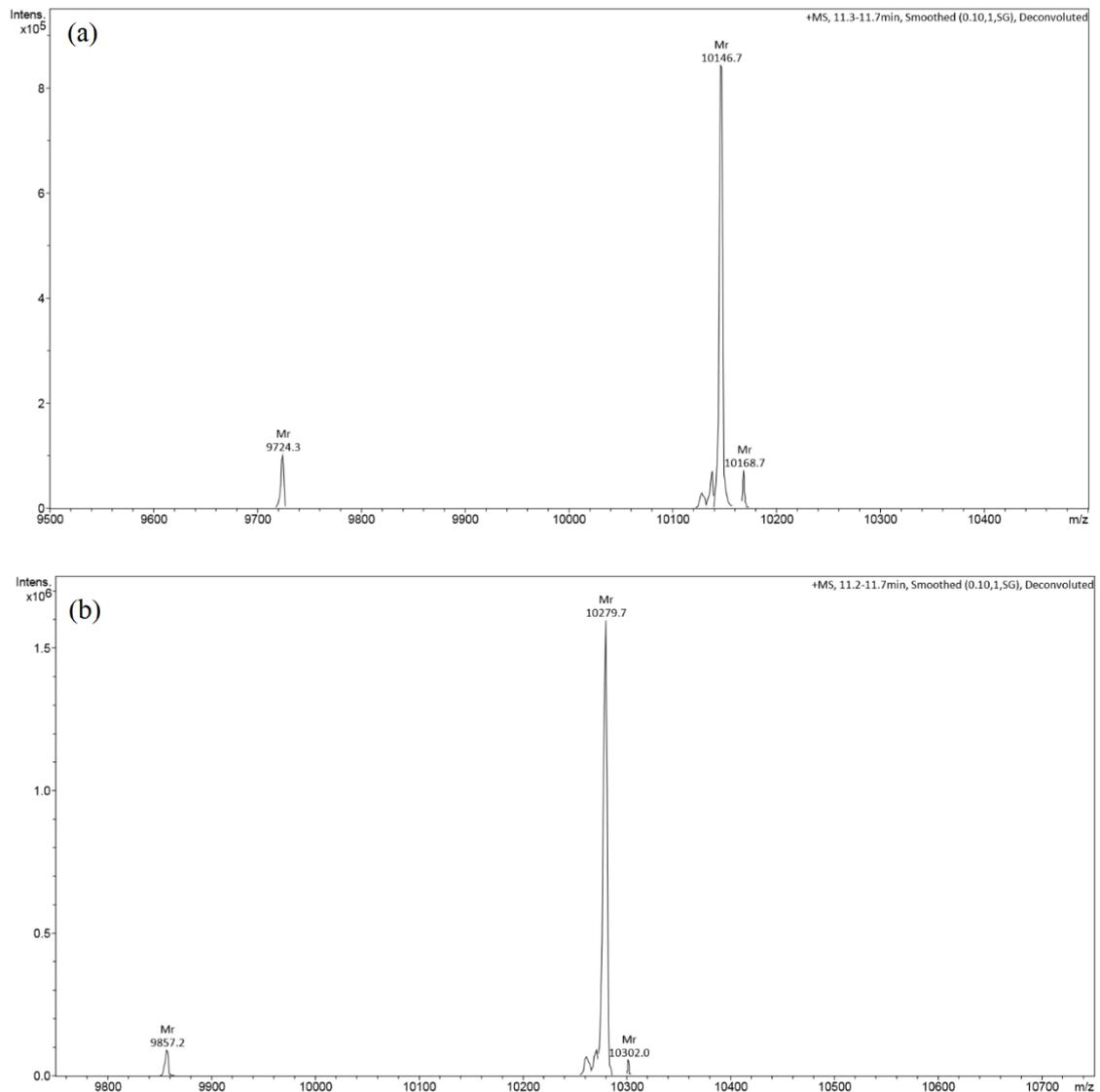
Overexpression of the fusion protein in the post-induction pellet was observed as a single protein band, however an additional band appeared in the gel samples upon lysis at RT. These extra bands co-purified by IMAC chromatography (see Figure 2.9 (a) lanes 13-15), suggesting the new band could be a degraded form of the fusion protein, albeit with an intact His<sub>6</sub>-tag. It was imagined the fusion protein was targeted by proteases released upon cell lysis, an action confirmed by the addition of protease inhibitors to the lysis buffer.



**Figure 2.9: 15% Reducing SDS-PAGE analysis of GB1-FXa- $\Delta$ -som cell lysis and metal affinity chromatography at RT (a) and using protease inhibitors (b).** Lane 1: Precision Plus protein marker, lane 2: pre-induction cell pellet, lane 3: post-induction cell pellet, lane 4: total cell lysate, lane 5: lysis supernatant, lane 6: lysis pellet, lane 7: supernatant load flow through, lane 8: binding buffer flow through, lane 9: wash buffer flow through, lanes 10-15: fractions from strip buffer flow through. Proteolytic digest of the fusion protein is shown in (a), which appears to run at a larger MW (observed in lane 4 onwards).

This addition was successful in inhibiting degradation observed by SDS-PAGE, and a single overexpressed band followed through lysis steps and IMAC chromatography for both fusion proteins (see Figure 2.9 (b)). It is unusual that the degraded fusion protein should run at a seemingly higher MW, however this phenomena was observed upon cleavage of  $\Delta$ -som from KSI fusions, and additionally upon cleavage of  $\Delta$ -som from intact GB1 fusions (see Figure 2.11). Whilst the use of protease inhibitors were shown to limit proteolysis, their addition would also complicate purification of the peptide downstream, as the inhibitors would need removing prior to cleavage of the fusion with the enzymes Factor Xa or enterokinase.

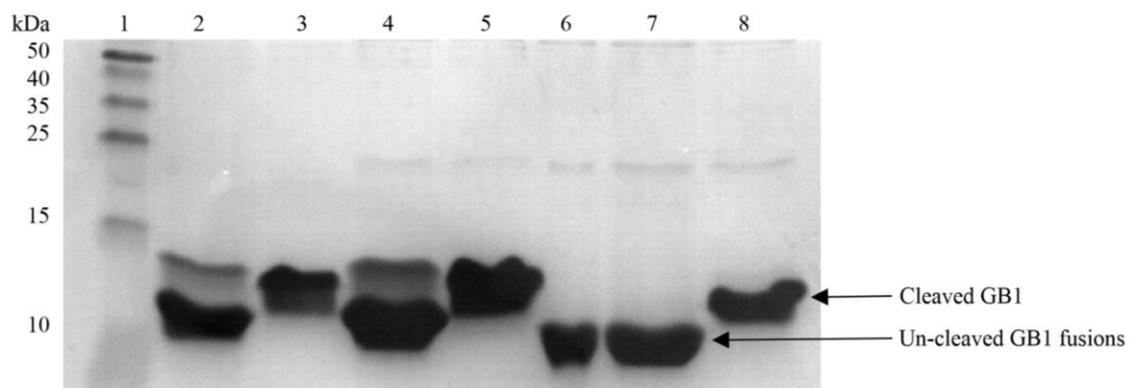
It was at this stage undergraduate project students took on the project. The students instead purified the fusion proteins by carrying out the lysis steps on ice, and using ice-cold buffers during IMAC chromatography to minimise any potential proteolytic activity, removing the requirement for protease inhibitors. Mass spectrometry of the purified GB1 fusions are shown in Figure 2.10.



**Figure 2.10: Mass spectrometry analysis of GB1-FXa- $\Delta$ -som (a) and GB1-enterokinase- $\Delta$ -som (b) fusion proteins. Major species in (a) 10146.7 Da (expected 10147.0 Da) and in (b) 10279.7 Da (expected 10280.0 Da).**

In both cases the major peak in each spectra identified the intact GB1 fusions at the correct MW: GB1-FXa- $\Delta$ -som was identified as 10146.7 Da (expected MW 10147.0 Da) and GB1-Enterokinase- $\Delta$ -som was identified as 10279.7 Da (expected MW 10280.0 Da). In addition, each spectra contained a +22 Da species, a sodium adduct commonly observed by LC-MS. Each spectra also identified a minor species of expected MW -422.4 Da, which corresponds to GB1 $\Delta$ -som -FTSS (last four residues at C-terminus). This observation suggests that a minor degree of proteolysis is still occurring.

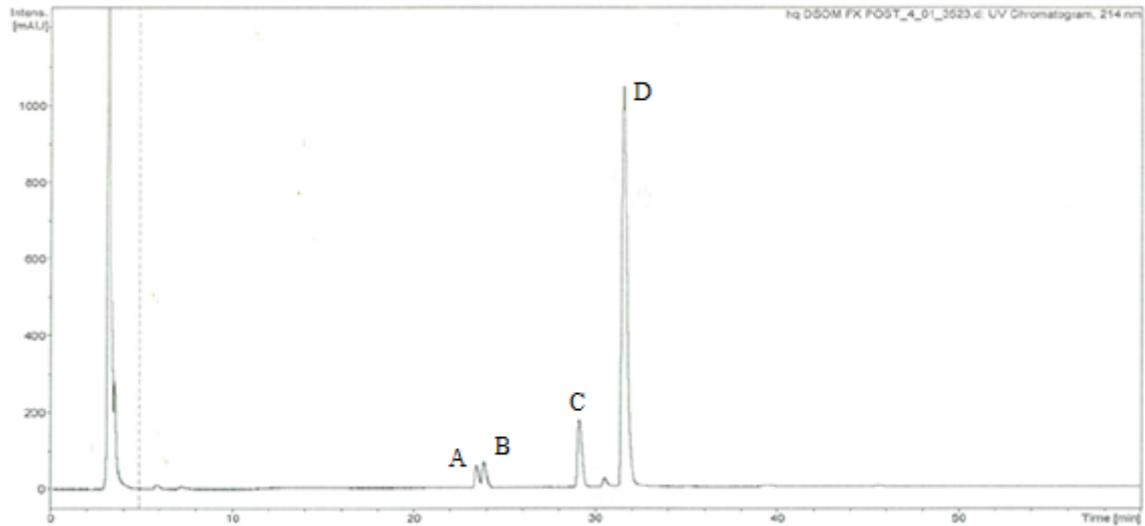
Cleavage of the fusion proteins was carried out by the students using varying concentrations of protease. It was determined that either 5  $\mu$ g/mL Factor Xa or 0.05 U/mL enterokinase resulted in efficient cleavage rates of the peptide from the GB1 fusion partner (>95% as determined by SDS-PAGE, Figure 2.11).



**Figure 2.11: 17% Reducing SDS-PAGE analysis of GB1 $\Delta$ -som cleavage by Factor Xa or enterokinase.** Lane 1: protein ladder, lanes 2 + 4: GB1-enterokinase- $\Delta$ -som pre-cleavage, lanes 3 + 5: GB1-enterokinase- $\Delta$ -som post-cleavage, lanes 6 + 7: GB1-FXa- $\Delta$ -som pre-cleavage, lane 8: GB1-FXa- $\Delta$ -som post-cleavage. Note: a small proportion of degraded GB1-enterokinase- $\Delta$ -som is clearly visible prior to cleavage with enterokinase. As also observed in the expression of KSI $\Delta$ -som, the GB1 $\Delta$ -som protein migrates at a lower MW, whilst cleaved GB1 protein migrates at a higher MW. Gel kindly reproduced from an undergraduate student report (Miss. Hannah Quinn).

Analytical HPLC and mass spectrometry analysis of the cleaved mixtures was carried out and summarised results are shown in Figures 2.12 and 2.13. Please see Appendix 2.5 and 2.6 for the corresponding mass spectrometry spectra.

(a)



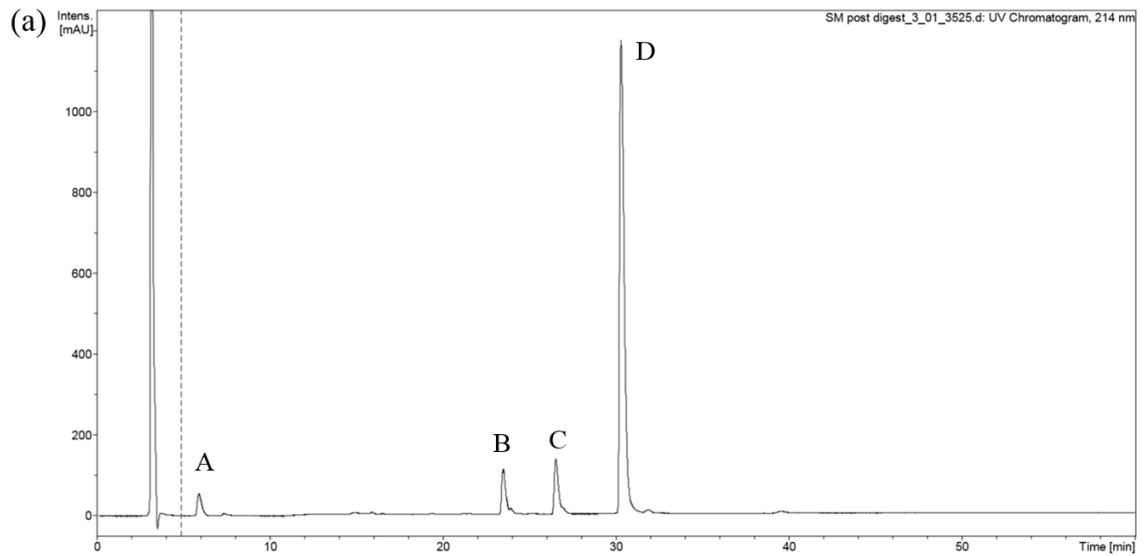
(b)

Peak	MW (Da)	Sequence Identity	Relative Intensity
A	1083.56	Peptide (AGSKNFFWK-)	3.34%
B	1184.61	Peptide (AGSKNFFWKT-)	4.17%
C	1606.8	Intact $\Delta$ -som	11.56%
D	8556.21	Intact GB1	79.53%

(c)



**Figure 2.12: HPLC trace of GB1 $\Delta$ -som as cleaved by Factor Xa (a). Corresponding MW and relative intensities of each labelled peak are shown in (b). Specific and non-specific cleavage sites are denoted by black and red arrows respectively in (c). See Appendix 2.5 for supporting mass spectrometry of each labelled peak. All peptides identified using ExPASy Find Pept software (<http://web.expasy.org/findpept/>).**



(b)

Peak	MW (Da)	Sequence Identity	Relative Intensity
A	541.24	Peptide (-TFTSS)	3.45%
B	1083.56	Peptide (AGSKNFFWK-)	6.27%
C	740.37	Peptide (-NFFWK-)	8.13%
D	8689.11	Intact GB1	82.16%

(c)



Figure 2.13: HPLC trace of GB1 $\Delta$ -som as cleaved by enterokinase (a). Corresponding MW and relative intensities of each labelled peak are shown in (b). Specific and non-specific cleavage sites are denoted by black and red arrows respectively in (c). See Appendix 2.6 for supporting mass spectrometry of each labelled peak. All peptides identified using ExPASy Find Pept software (<http://web.expasy.org/findpept/>).

LC-MS of the material cleaved with Factor Xa showed 80% composed of GB1 protein (peak D), whilst ~12% was identified as intact  $\Delta$ -som peptide (peak C). These values agree closely with theoretical yields ( $\Delta$ -som constitutes 16% of the total fusion and GB1 constitutes 84%). Two smaller peaks (A and B), together composing 7%, were also identified as  $\Delta$ -som peptide fragments that had been correctly cleaved from GB1, but also cleaved after Lys and Thr residues (cleavage at this same Thr residue was observed in Figure 2.10- intact fusion protein before cleavage). The two peptide fragments therefore suggest that whilst cleavage by Factor Xa is highly specific, a small amount of proteolysis may still be occurring during the lysis and/or IMAC steps of GB1 fusion protein purification.

LC-MS analysis of the material cleaved using enterokinase showed 82% of the mixture composed of GB1 protein (peak D), as expected. The peptide however was recovered as three partial fragments: ~8% NFFWK (peak C), ~6% AGSKNFFWK (peak B), and ~3% TFTSS (peak A). Therefore whilst successful cleavage of the fusion was achieved using enterokinase, the incubation appears to have led to cleavage at sites in the fusion other than the engineered cleavage site. This appears to have happened at lysine residues in the peptide sequence (see Figure 2.13 (c)). Whilst an unexpected result, a search of the literature showed it to be a previously published observation (Light and Janska, 1989), (Liew et al., 2005). It is also of note that the fusion protein pre-cleavage ran as multiple species by SDS-PAGE (see Figure 2.11), suggesting minor protein degradation may have additionally occurred before the enterokinase cleavage step.

To summarise, Factor Xa cleavage has proven to be most suitable for cleavage of  $\Delta$ -som peptide from GB1 fusions, and GB1-FXa- $\Delta$ -som fusion protein was successfully cleaved to yield 2.6 mg/L intact  $\Delta$ -som peptide.

## **2.4 Discussion**

### **2.4.1 KSI Peptide Fusions**

The pET31b(+) KSI fusion system was initially used as the method for recombinant peptide production; the system has been used successfully within the laboratory to produce milligram quantities of uniform isotopically labelled peptides for NMR spectroscopy.

Expression of the KSI $\Delta$ -som fusion protein gave yields >80 mg/L culture in LB media, and >20 mg/L in minimal media, as estimated by SDS-PAGE analysis.

Cleavage of the fusion protein by CNBr in formic acid cleaved with >90% efficiency, but mass spectrometry of the peptide showed the use of formic acid resulted in formylation of the peptide. Cleavage in an alternative TFA solution was less efficient, with around 50% efficiency after 24 hours (with longer incubations leading to degradation of the peptide). Attempts to purify the peptide after cleavage were met with difficulty; the cleaved peptide was found to adhere to KSI and/or the plastic and glassware, a problem that seemed magnified with material expressed in MM owing to the overall lower protein expression.

Alternative attempts to purify the adherent peptide were made by denaturation of the cleaved material using GdnHCl, but upon purification the  $\Delta$ -som peptide (synthetic and recombinant forms) was not found to interact with the size exclusion media. Depending upon supplier, there is a discrepancy as to the resolving limits of G-15 Sephadex media: GE (the original manufacturers and suppliers of this media) state on their website the dextran is capable of resolving peptides >1500 Da in size ( $\Delta$ -som is 1.6 kDa and should therefore have been resolved). Contrastingly, Sigma state their G-15 Sephadex media is only suitable for resolving peptides that are 500 – 1500 Da in size. A response from GE regarding this matter was not received, but it is possible the GE website should read <1500 Da. It was not possible to resolve  $\Delta$ -som peptide using this media.

The adherent property of the peptide may be a result of its high hydrophobicity; synthetically synthesised peptide was also poorly soluble and prone to aggregation in aqueous buffers at concentrations >10 mM. To circumvent this problem synthetic  $\Delta$ -som peptides were dissolved into 100% DMSO. It is interesting to note that previous success of this method was achieved within the laboratory for the expression of more hydrophilic peptides.

Ultimately, this recombinant method was not deemed viable for the production of isotopically enriched  $\Delta$ -som peptide, on account of peptide losses at the cleavage and recovery stages.

### **2.4.2 GB1 Peptide Fusions**

The GB1 fusion system was trialled as a method for recombinant peptide production, in an attempt to circumvent the various problems that arose from using an insoluble fusion partner coupled with a chemical cleavage site. Expression of both the GB1 $\Delta$ -som fusion

proteins gave yields >80 mg/L culture in LB media, with the fusion found exclusively within the soluble fraction (as determined by SDS-PAGE analysis). A combination of lysis steps carried out on ice and IMAC purification using ice-cold buffers somewhat limited the proteolysis observed when these steps were carried out at RT, and intact GB1 fusion protein was identified by mass spectrometry as composing >90% pooled IMAC fractions. Cleavage of the fusion proteins using varying concentrations of either enterokinase or FXa was trialled by the undergraduate students, and use of either 5  $\mu$ g/mL Factor Xa or 0.05 U/mL enterokinase resulted in efficient cleavage rates of the peptide from the GB1 fusion (>95% as determined by SDS-PAGE).

LC-MS analysis of the material cleaved using enterokinase showed the peptide was efficiently cleaved but was additionally cleaved at lysine residues within the peptide, to be recovered only as three small peptides. In contrast, material cleaved by Factor Xa showed intact peptide comprised 12% of the total cleaved sample.

A small amount of proteolysis was still identified by SDS-PAGE and mass spectrometry prior to cleavage by Factor Xa. Protease inhibitors could be added to the lysis buffer to limit this proteolysis, however an additional purification step would be required to remove these inhibitors before the enzymatic cleavage, a step which conversely could lead to lower peptide yields.

Nevertheless, assuming 100% cleavage by Factor Xa this is a peptide recovery rate of >70%, giving a final yield of 2.6 mg/L intact  $\Delta$ -som peptide.

### **2.4.3 Overall Summary**

Initial attempts to purify  $\Delta$ -som peptide from an insoluble KSI fusion were met with limited success, despite several attempts at optimisation. The production of  $\Delta$ -som peptide was deemed crucial to achieve the aims of this thesis, as it is the model binding peptide of hPDI. This thesis instead employs the use of unlabelled and specifically fluorinated synthetic peptides, the latter of which can be produced for a moderately increased cost compared to unlabelled synthetic peptides.

Meanwhile, work conducted by undergraduate students showed reasonable quantities of unlabelled  $\Delta$ -som peptide could be purified from soluble GB1 fusions encoding a Factor Xa cleavage site. The purification of this fusion protein was shown to be successful in large LB culture scale-ups, and the peptide was easily isolated from the cleaved mixture to yield a modest 2.6 mg/L intact  $\Delta$ -som peptide (compared to 1.9 mg/L from KSI fusions). This yield is still relatively low compared to the original GB1 literature, where yields >10

mg/L are quoted (Koenig et al., 2003). Unfortunately whilst there was not time to develop and optimise this method further in LB or trial expression and cleavage using a defined minimal media, it is now hoped that a system is in place for future researchers to express and purify milligram quantities of isotopically enriched  $\Delta$ -som peptide.

## CHAPTER 3

# Expression and Purification of hPDI **b'x** and W111F **abb'x**

### 3.1 Introduction

Biological studies of a protein by Nuclear Magnetic Resonance (NMR) typically require a highly pure, isotopically enriched sample of micromolar to millimolar concentration. Recombinant protein expression is a technique developed in the early 1970s, and today is widely used by scientists as a means of selectively overexpressing their protein of interest. A variety of different *in vivo* host strains have been developed, and examples of common cell hosts include bacteria (*E. coli*), yeast (*S. cerevisiae*, *P. pastoris*), insect (*S. frugiperda*) or mammalian cell lines (CHO (Chinese hamster ovary), HEK (human embryonal kidney)). It is also possible to selectively express proteins *in vitro*, using translation compatible whole-cell extracts (Sitaraman and Chatterjee, 2009) or by chemical synthesis (Nilsson et al., 2005). The cell host or expression system chosen is largely dependent on the type of protein to be expressed; for example, in the production of antibodies, expression in a mammalian cell line can prove vital for production of a correctly folded and post-translationally modified protein (Wurm, 2004). However, mammalian cell culture is both costly and labour intensive with relatively low yield, so is not always an appropriate

or economic choice for protein expression. Chemical protein synthesis is better suited to small proteins and peptides, and whilst cell-free methods can allow for the stable isotopic protein labelling required, it is the *E. coli* bacterial expression system that is most widely exploited for expression of proteins for study by NMR. The bacteria grow rapidly in either rich or minimal media to a high density, generally with good protein yield and recovery (Baneyx, 1999). This expression system lends itself well to stable isotope labelling (e.g.  $^{13}\text{C}$  or  $^{15}\text{N}$ ) by growth in a minimal media using a sole  $^{13}\text{C}$  carbon and/or  $^{15}\text{N}$  nitrogen source. Recombinant proteins expressed this way can easily be engineered with small terminal extensions (such as a His<sub>6</sub>-tag) that can aid in affinity purification steps after cell lysis.

### 3.1.1 *E. coli* Strains and Vectors

For this project *E. coli* BL21 (DE3) pLysS cells were used for protein expression. BL21 cells allow expression of genes under control of a T7 promoter, and are deficient in the proteases *lon* and *ompT*, making them suitable for stable protein expression (see 2.2.2 for further details). A modified form of the Novagen pET-23a(+) vector containing the coding sequence for a His<sub>6</sub>-tag upstream of the *NdeI* site was used for the expression of hPDI **b'x** and W111F **abb'x**. This construct was made and generously gifted by Prof. L Ruddock (University of Oulu, Finland) (see Appendix 2.2 and 3.1 for vector maps).

WT hPDI **abb'x** contains three Trp residues:  $^{35}\text{Trp}$  (heavily quenched by the adjacent active site in **a** domain),  $^{111}\text{Trp}$  (also in **a** domain), and  $^{347}\text{Trp}$  in the **x-linker** region (see Figures 1.2 and 7.2). The W111F mutant was initially designed as a probe for fluorescence that would report exclusively on the **x-linker** region of the protein. This mutant has been shown to be structurally unaffected by the substitution, and behaves in a similar manner to WT protein (Wang et al., 2010). With this in mind it was decided not to mutate W111F **abb'x** back to WT protein. This thesis utilises selective fluorination of the Trp residues of **abb'x**, and  $^{111}\text{Trp}$  is buried within the protein, forming part of the hydrophobic core. Previous studies within the laboratory (Curtis-Marof et al., 2014) suggested that fluorination of this hidden Trp residue may impact upon protein folding (discussed further in Chapter 7).

### 3.1.2 Protein Expression and Purification

All proteins requiring isotopic enrichment for NMR were grown in a minimal medium containing  $^{15}\text{N}$  ammonium sulphate and/or  $^{13}\text{C}$  D-glucose as sole nitrogen and glucose sources respectively. Initial characterisation of the two proteins was carried out using protein expressed in LB media, and whilst overall protein yields were high the purified protein showed variable protein-dependent dimer formation, undesirable for NMR. Similar dimer formation also presented upon expression in a rich autoinduction media. Expression of the fragments in an unlabelled minimal media (MM) resulted in less dimer formation for both **b'x** and W111F **abb'x**. Any unlabelled protein for NMR was therefore expressed in MM using laboratory grade  $^{14}\text{N}$  ammonium sulphate and  $^{12}\text{C}$  D-glucose as nitrogen and glucose sources.

After expression, the cells were lysed by an overnight freeze-thaw and treated with detergent and DNase. After lysis, the proteins were purified from the cell lysate supernatant using metal affinity chromatography. The proteins were desalted by overnight dialysis, and residual impurities were removed using anion exchange chromatography (theoretical pI of **b'x**: 5.98 and W111F **abb'x**: 5.23). At this stage the proteins were of high purity, but existed as mixed monomer: dimer populations. They were separated using size exclusion chromatography, yielding highly pure (>95%) monomeric species suitable for NMR. In addition, analytical gel filtration was used to calculate each proteins hydrodynamic volume, and electrospray mass spectrometry was collected to accurately confirm the molecular weight and isotopic incorporation of W111F **abb'x** (the **b'x** construct has been previously characterised (Byrne et al., 2009)).

## 3.2 Materials and Methods

*Unless stated otherwise all laboratory reagents were supplied by Sigma Aldrich (Gillingham, UK).*

### 3.2.1 Glycerol Stocks

The *E.coli* strain BL21 (DE3) pLysS containing a pET-23a(+) vector with the coding sequences for the human PDI fragments **b'x** or W111F **abb'x** were used. Both sequences

are expressed with an N-terminal His<sub>6</sub>-tag extension and residues <sup>213</sup>Lys-<sup>351</sup>Pro (**b'x**) and <sup>1</sup>Asp-<sup>351</sup>Pro (**abb'x**) of hPDI. The amino acid sequences of the expected recombinant products and relevant ExPASy analysis are shown in Appendix 3.2.

### 3.2.2 Bacterial Culture Media

Glycerol stocks were streaked out onto fresh LB agar plates (10 g/L Tryptone, 5 g/L Yeast Extract, 10 g/L NaCl and 15 g/L Agar) with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) and incubated overnight at 37°C. For the production of isotopically enriched hPDI proteins, a minimal medium was used (see Table 3.1).

**Table 3.1: Components of minimal media.**

Component (stock concentrations)	Stock Solution	Volume/ L	Method of sterilisation	Final working concentration	
1. (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> (x50)	( <sup>15</sup> N)	30 g/L	20 mL	Filter	4.51 mM
	or ( <sup>14</sup> N)	50 g/L			7.57 mM
2. PO <sub>4</sub> /NaCl (x10)	Na <sub>2</sub> HPO <sub>4</sub>	68 g/L	100 mL	Autoclave	47.90 mM
	+KH <sub>2</sub> PO <sub>4</sub>	30 g/L			22.04 mM
	+NaCl	5 g/L			8.56 mM
3. Na <sub>2</sub> SO <sub>4</sub> (x1000)		42.6 g/L	1 mL	Autoclave	0.30 mM
4. EDTA Trace	EDTA	10 g/L	10 mL	Autoclave	0.34 mM
Elements (x100)	MnCl <sub>2</sub>	3.2 g/L			0.25 mM
	FeCl <sub>3</sub>	1 g/L			6.17 µM
	CuCl <sub>2</sub>	20 mg/L			0.15 µM
	CoCl <sub>2</sub>	20 mg/L			0.15 µM
	H <sub>3</sub> BO <sub>3</sub>	20 mg/L			0.32 µM
	ZnCl <sub>2</sub>	20 mg/L			0.15 µM
5. MgSO <sub>4</sub> (x1000)		246 g/L	1 mL	Filter	2.04 mM
6. CaCl <sub>2</sub> (x1000)		44.1 g/L	1 mL	Filter	0.40 mM
7. Biotin (x1000)		1 g/L	1 mL	Filter	4.09 µM
8. Thiamine (x1000)		1 g/L	1 mL	Filter	3.77 µM
9. D-Glucose (x50)	( <sup>13</sup> C)	100 g/L	20 mL	Filter	10.74 mM
	or ( <sup>12</sup> C)	200 g/L			22.20 mM
10. MilliQ water			845 mL	Autoclave	

Parts 2, 3 and 4 were added to an Erlenmeyer flask with the correct volume of MilliQ water (part 10) and autoclaved. All other components were filter sterilised, stored at 4°C and added immediately prior to inoculation.

PO<sub>4</sub>/NaCl, Na<sub>2</sub>SO<sub>4</sub> and EDTA trace elements were added to the correct volume of MilliQ water in Erlenmeyer flasks 5 times the final media volume, sterilised by autoclaving, and stored at room temperature until required. All other minimal media components were made using MilliQ water (with the exception of Biotin, which was made in 50% ethanol solution), and filter sterilised using a 0.2 µm Minisart single use syringe filter, stored at 4°C and added aseptically to the medium immediately prior to inoculation. MM promoted least dimer formation when compared to LB and autoinduction media, so was also used for unlabelled protein preparations, where (<sup>14</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and <sup>12</sup>C<sub>6</sub>H<sub>12</sub>O<sub>6</sub> were used in place of their isotopic counterparts.

### 3.2.3 Recombinant Protein Expression

For main grows, starter cultures of 100 mL unlabelled MM with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) were inoculated using a single colony from the overnight plates. The starter was grown for 24 hrs at 37°C with 200 rpm shaking, to an A<sub>600</sub> of approximately 2.0 AU. The cells were then centrifuged in 50 mL falcon tubes at 3,220 g for 15 mins at 4°C and the cell pellets re-suspended in 1-5 mL of fresh, sterile media.

Each large main culture (up to 3 x 400 mL cultures from 1x 100 mL starter culture) was then inoculated from the starter culture re-suspension to give a starting A<sub>600</sub> of 0.1 AU.

With a doubling time of approximately 60 min, the cells were grown at 37°C and monitored until they reached an A<sub>600</sub> of 0.7 AU. They were then induced by the addition of 1 mM IPTG (1 M stock). Optimised induction times that promoted greatest monomer formation were 2.5 hrs at 37°C for **b'x**, and overnight induction at 27°C for W111F **abb'x**. After induction, the cells were cooled on ice for 5 mins and then harvested by centrifugation at 10,000 g for 10 mins at 4°C. The cell pellets were then re-suspended in protein lysis buffer (20 mM sodium phosphate, 50 mM sodium chloride, pH 7.3) and stored at -20°C until required (10 mL protein lysis buffer per each 400 mL of original culture pellet).

### 3.2.4 Cell Lysis

Previously frozen cells were thawed at RT. Each 10 mL re-suspended pellet (equivalent to 400 mL original culture volume) was treated with 200 µL of Triton® X-100 (from a stock of 5% v/v in lysis buffer), gently mixed and left to incubate for 20 mins. This was followed

by treatment with 100  $\mu$ L DNase I (Bovine, 2 mg/mL stock) and 50  $\mu$ L 2 M  $MgCl_2$  for a second incubation at RT for 20 mins. The lysate was then centrifuged at 10,000 g for 15 mins at 4°C and the supernatant carefully separated from the insoluble cell debris. Samples of the total cell lysate and supernatant were taken for reducing SDS-PAGE analysis.

### **3.2.5 IMAC Chromatography**

This affinity purification step was carried out using Chelating Sepharose Fast Flow (GE Healthcare). The column was charged with nickel ions and the His<sub>6</sub>-tag proteins were eluted using EDTA, as outlined previously in 2.2.11. An equivalent of up to 800 mL original cell culture lysate was loaded per column, using multiple columns when scaling up.

### **3.2.6 Dialysis**

Strip buffer fractions containing  $A_{280\text{ nm}} > 0.1$  AU of hPDI protein were pooled, pipetted into dialysis tubing (12-14 kDa MWCO, Medicell International Ltd) and dialysed overnight using 2 L phosphate buffer (20 mM sodium phosphate, pH 7.3) at 4°C with constant stirring.

### **3.2.7 Anion Exchange Chromatography**

Anion exchange chromatography was carried out on the desalted proteins using a Source 30Q column (5 mL CV, 30  $\mu$ m size beads, GE Healthcare) connected to an ÄKTA FPLC analyser (GE Healthcare). The column was stored at RT in de-gassed 20% (v/v) ethanol, and equilibrated for use with a cycle of 10 CV buffer A (20 mM sodium phosphate, pH 7.3, de-gassed), 10 CV buffer B (20 mM sodium phosphate, 0.5 M NaCl, pH 7.3, de-gassed), followed by a second 10 CV of buffer A. A total protein amount of up to 800 mL equivalent original culture was filtered using a 0.2  $\mu$ m Minisart single use syringe filter and loaded onto the column using a 50 mL injection superloop. The protein was eluted from the column with an increasing salt gradient using buffer B from 0-100% in 20 CV at a flow rate of 5 mL/min, collected in 2mL fractions and monitored using absorbance at 280 nm. A cut-off of higher than 0.1 AU at  $A_{280\text{ nm}}$  was again used to select fractions for reducing SDS and native-PAGE analysis and were pooled for size exclusion chromatography. After use the anion exchange column was washed using 10 CV of de-gassed 20% (v/v) ethanol and stored at RT for future use.

### 3.2.8 Size Exclusion Chromatography

Both preparative and analytical size exclusion chromatography were carried out on pooled proteins from anion exchange using a 300 mL CV of Superdex 200 medium, a matrix of cross-linked dextran and agarose (GE Healthcare). The column was stored at RT in de-gassed 20% ethanol, and equilibrated overnight using 1.5 CV de-gassed running buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.3).

For preparative size exclusion chromatography, pooled protein from a single anion exchange run was concentrated to a volume of approximately 1 mL using a Vivaspin 20 Centrifugal Filter (10 kDa MWCO, Sartorius), and loaded onto the column under gravity flow using a three way tap. The protein was eluted using 1.5 CV of running buffer at a flow rate of 2 mL/min, collected in 5 mL fractions and monitored using absorbance at 280 nm. After use the column was washed using 1.5 CV of de-gassed 20% (v/v) ethanol and stored at RT for future use.

For analytical size exclusion chromatography, the column was calibrated using 1 mg each of Cytochrome C (12.4 kDa), Carbonic Anhydrase (29 kDa), Bovine Serum Albumin (66 kDa), Alcohol Dehydrogenase (150 kDa) and  $\beta$  Amylase (200 kDa), (MW Gel Filtration Marker Kit 12,000-200,000 Da). An additional 1 mg of Apoferritin (443 kDa) was also used, and 1 mg Dextran Blue (2 MDa) to determine the void volume of the column.

### 3.2.9 Gels: Reducing SDS-PAGE and Native-PAGE

Reducing SDS-PAGE and native-PAGE gels were used at all stages to assess and confirm expression, lysis and subsequent purification steps of the proteins.

**Reducing SDS-PAGE** gels were prepared, run and stained as previously outlined in 2.2.7.

**Native-PAGE** gels were also prepared using the Tris-Glycine buffer system, with a 12.5% w/v resolving gel (pH 8.3) and no stacking gel. The gel recipe follows those in Table 2.1, omitting the sodium dodecyl sulphate (SDS) and substituting with 200  $\mu$ L MilliQ water. The gel samples were prepared using 4x non-reducing sample buffer (as previously described for SDS-PAGE but without SDS and DTT, and no heat incubation). Each lane was loaded with 10  $\mu$ L sample, and gels were run negative to positive (theoretical pI of both proteins is lower than pH 8.3) in 1x TG gel running buffer (TGS buffer with SDS

omitted), at 12 mA for 4-6 hrs. Native gels were stained, de-stained and dried as for SDS-PAGE gels.

### 3.2.10 Concentration Determination

Protein absorbance was measured throughout the purification process using an Ultrospec 2000 UV/Visible Spectrophotometer (Pharmacia Biotech) and a suitable buffer blank at a wavelength of 280 nm. Protein amino acid sequences were submitted to ExPASy ProtParam (<http://web.expasy.org/protparam/>) to calculate extinction coefficients, and molar protein concentrations were calculated using the Beer-Lambert Law. The extinction coefficient of **b'x** was  $8,605 \text{ M}^{-1}\text{cm}^{-1}$  and for W111F **abb'x** was  $24,660 \text{ M}^{-1}\text{cm}^{-1}$ .

### 3.2.11 Mass Determination and Isotope Incorporation

The molecular weight of unlabelled and isotopically enriched monomeric W111F **abb'x** protein was determined using Electrospray Mass Spectrometry, performed by Kevin Howland, Biomolecular Science Facility Manager, University of Kent.

## 3.3 Results

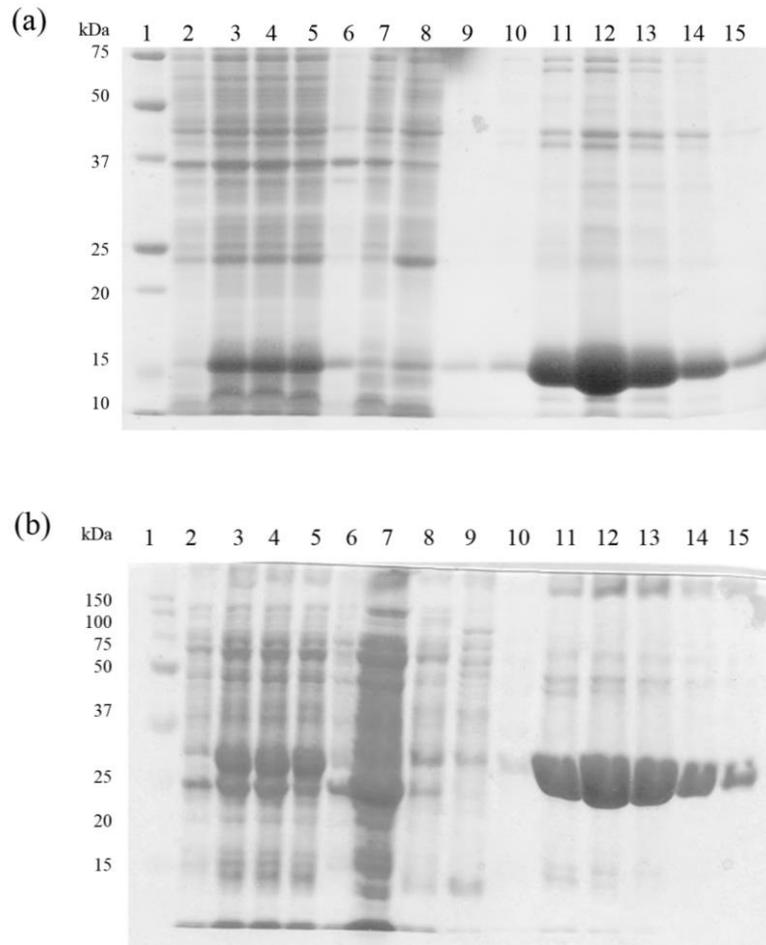
### 3.3.1 Expression of *b'x* and W111F *abb'x*

Expression of *E. coli* BL21 (DE3) pLysS with hPDI construct inserts produced white uniform colonies on LB agar plates and showed doubling times of approximately 30 min in LB medium and 60 min in minimal medium. After induction both fragments were processed using the same lysis and purification protocol, and the following results show expression and purification of  $^{15}\text{N}$  labelled **b'x** and W111F **abb'x** protein.

### 3.3.2 IMAC Chromatography

Both hPDI fragments were expressed as soluble proteins (as shown in Figure 3.1). By comparison with the Precision Plus protein standards, the 17 kDa **b'x** protein runs with a molecular weight around 15 kDa, and W111F **abb'x** also runs smaller than expected with an apparent molecular weight between 25 and 37 kDa (actually 41 kDa). Purification by His<sub>6</sub>-tag affinity successfully removed the majority of contaminating proteins from the

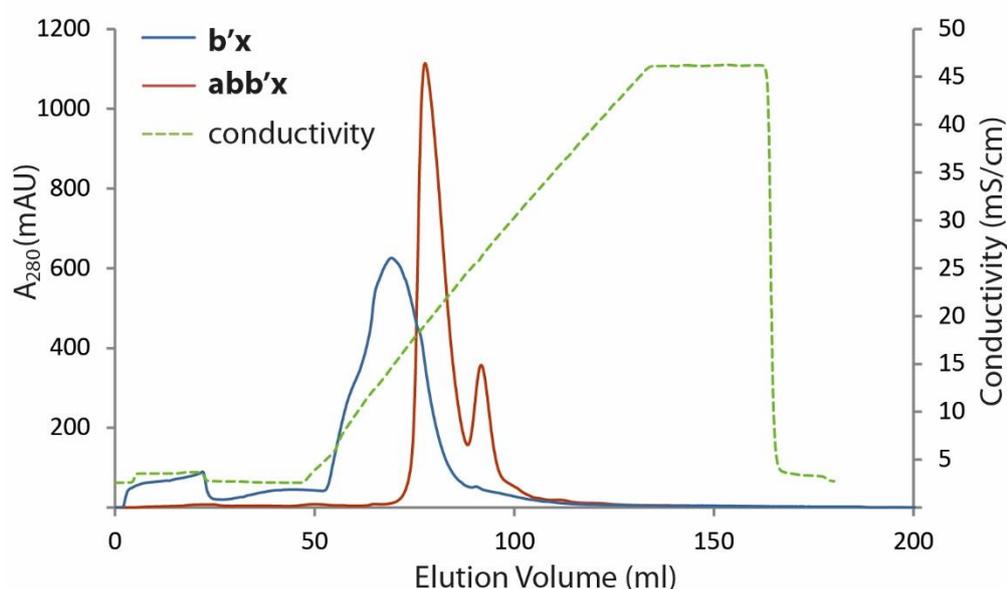
lysate supernatant, but some higher and lower molecular weight impurities were still visible in the purified strip buffer fractions of both **b'**x and W111F **abb'**x.



**Figure 3.1: 12.5% reducing SDS-PAGE analysis of *b'*x (a) and W111F *abb'*x (b): expression, lysis and IMAC chromatography.** Lane 1: Precision Plus protein marker, lane 2: pre-induction cell pellet, lane 3: post-induction cell pellet, lane 4: total cell lysate, lane 5: lysis supernatant, lane 6: lysis pellet, lane 7: supernatant load flow through, lane 8: binding buffer flow through, lane 9: wash buffer flow through, lanes 10-15: fractions from strip buffer flow through.

### 3.3.3 Anion Exchange Chromatography

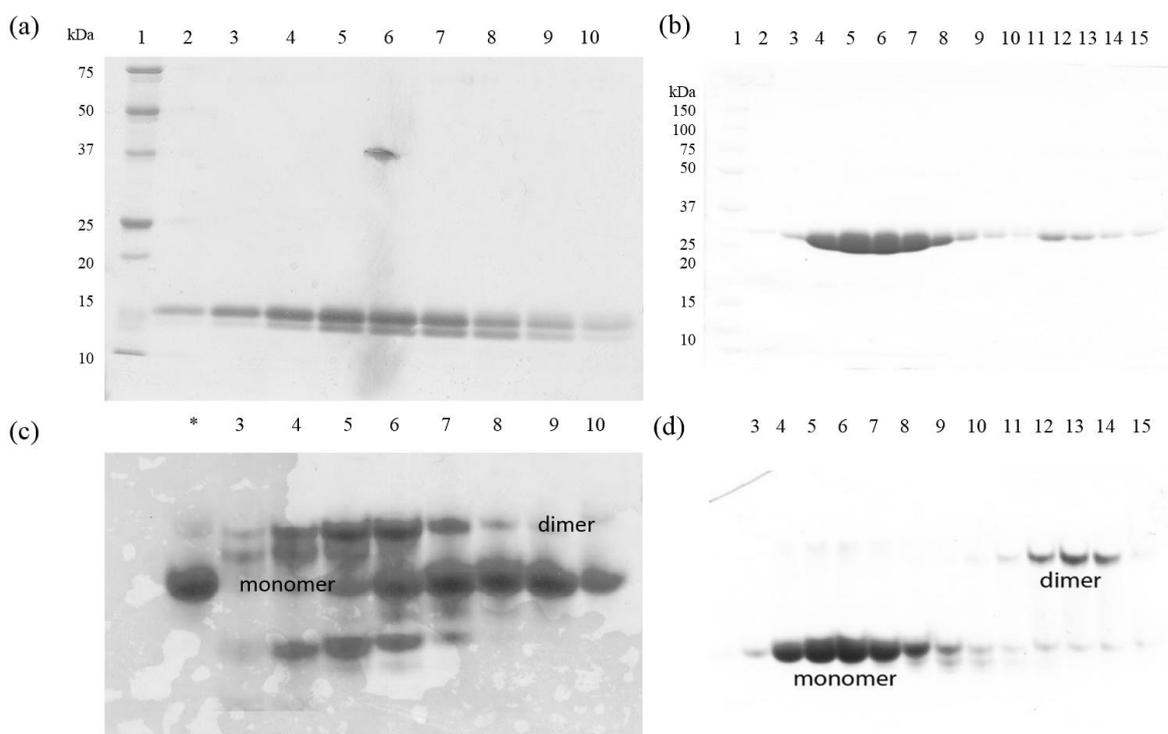
Anion exchange chromatography was successful in removing a large number of the remaining impurities from the protein samples. **b'**x began eluting upon 8% Buffer B, and W111F **abb'**x slightly later at 27% Buffer B (see Figure 3.2). The chromatogram shows a single but non symmetrical peak for **b'**x and multiple elution peaks for W111F **abb'**x, both suggestive of the presence of multiple purified species. Fractions across the peaks from both proteins were run on reducing SDS and native-PAGE gels (Figure 3.3) for analysis.



**Figure 3.2: Anion exchange profile of 800 mL equivalent culture volume *b'*x and W111F *abb'*x protein using a 5 mL Source 30Q Column.** The UV absorbance at 280 nm is coloured blue (**b'**x) or red (**abb'**x) and the conductivity is shown in green.

Reducing SDS-PAGE analysis (Figure 3.3 (a) and (b)) of the peaks showed a single species for both proteins; however native-PAGE of each protein yielded two distinct species (Figure 3.3 (c) and (d)). It is well documented that hPDI fragments containing the **b'** domain form homodimers (Wallis et al., 2009), therefore it is likely that the peaks correspond to monomer and dimer species.

In the case of W111F **abb**'x, native-PAGE shows the monomer elutes first before the dimer (Figure 3.3 (d)), and anion exchange provided adequate resolution to pool largely monomer fractions. The **b**'x peak however is an unresolved mixture of monomer and dimer protein (with a small proportion of another species) (Figure 3.3 (c)), so the entire peak was pooled for size-based separation of the two species by size exclusion chromatography.

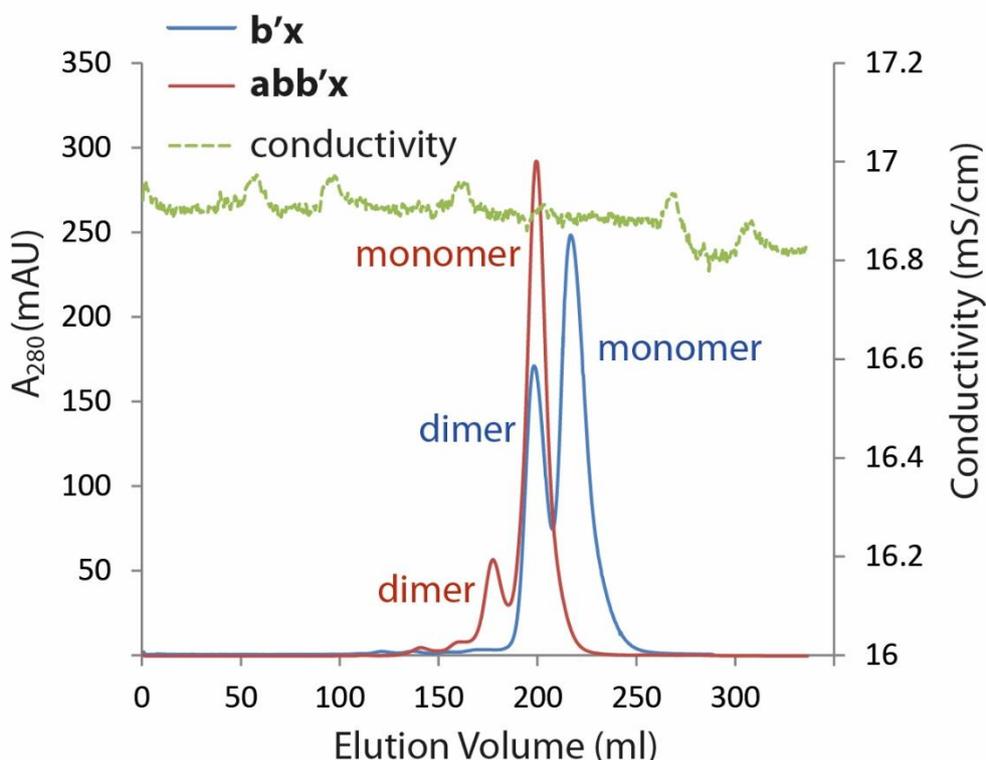


**Figure 3.3: 12.5% Reducing SDS-PAGE analysis of anion exchange fractions of *b*'x (a) and W111F *abb*'x (b). Native-PAGE analysis of corresponding samples of *b*'x (c) and W111F *abb*'x (d). Lane 1: Precision Plus protein marker (gels (a) and (b) only), lanes 2-10 in (a) are fractions across the anion exchange peak (50-80 mL), and lanes 2-15 in (b) are fractions across the two anion exchange peaks (70-100 mL), as shown in Figure 3.2. Numbered lanes in (c) and (d) correspond to the samples run in lanes in (a) and (b). Lane \* in (c) is purified monomeric *b*'x protein for comparison.**

### 3.3.4 Size Exclusion Chromatography

**Preparative Size Exclusion-** The anion exchange peak fractions containing monomer protein were pooled, concentrated by ultrafiltration and loaded in 1 mL volumes onto the

Superdex 200 column. Both proteins eluted in two distinct peaks. A typical elution profile of **b'x** and W111F **abb'x** is shown in Figure 3.4. W111F **abb'x** monomer was simplest to isolate from its dimer species upon anion exchange, thus the protein elutes largely as monomer on size exclusion, with a minor dimer peak. Elution of **b'x** varied depending on preparation; it was found that higher protein expression yields generated a larger fraction of dimer upon purification.

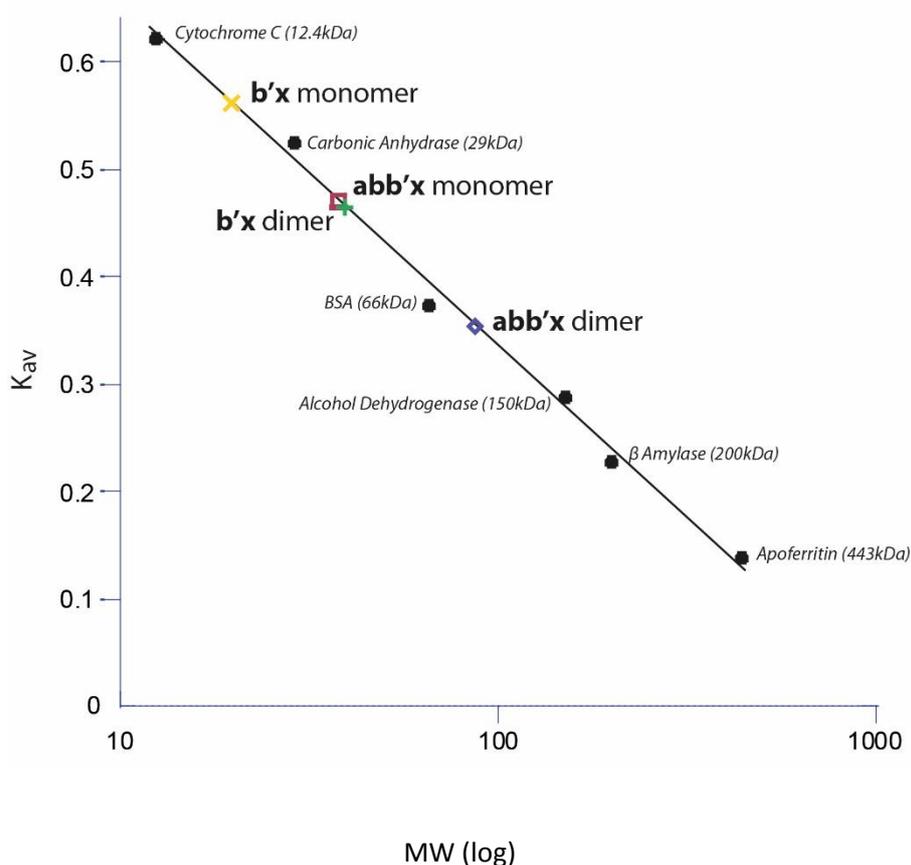


**Figure 3.4:** Size exclusion chromatography profile of 800 mL equivalent culture volume **b'x** and W111F **abb'x** protein using a 300 mL Superdex 200 Column. Monomer peak fractions were pooled for NMR studies. The UV absorbance at 280 nm is coloured blue (**b'x**) or red (**abb'x**) and the conductivity is shown in green.

**Analytical Size Exclusion-** To determine the molecular weight of each protein species the size exclusion column was calibrated using a set of protein standards to generate a calibration curve, plotted as  $K_{av}$ , the gel phase distribution coefficient, versus  $\log MW$ . The  $K_{av}$  of each protein was determined using equation 3.1:

$$3.1 \quad K_{av} = \frac{V_e - V_0}{V_c - V_0}$$

Where  $V_e$  is the peak elution volume,  $V_0$  is the void volume (as determined by Dextran Blue), and  $V_c$  is the geometric column volume. The standard calibration curve was then used to determine the molecular weight of each proteins monomer and dimer species (see Figure 3.5). Molecular weight calculations from the calibrated curve show that  $^{15}\text{N}$  **b'x** monomer behaves as a protein with a slightly larger (12.4%) hydrodynamic volume than expected, and  $^{15}\text{N}$  W111F **abb'x** behaves slightly smaller (7.2%) (see Table 3.2).



**Figure 3.5: Size exclusion column calibration curve.** The standard proteins used are represented by black circles.  $^{15}\text{N}$  labelled **b'x** is a yellow (monomer) or green (dimer) cross, and  $^{15}\text{N}$  W111F **abb'x** is a red (monomer) or blue (dimer) square.

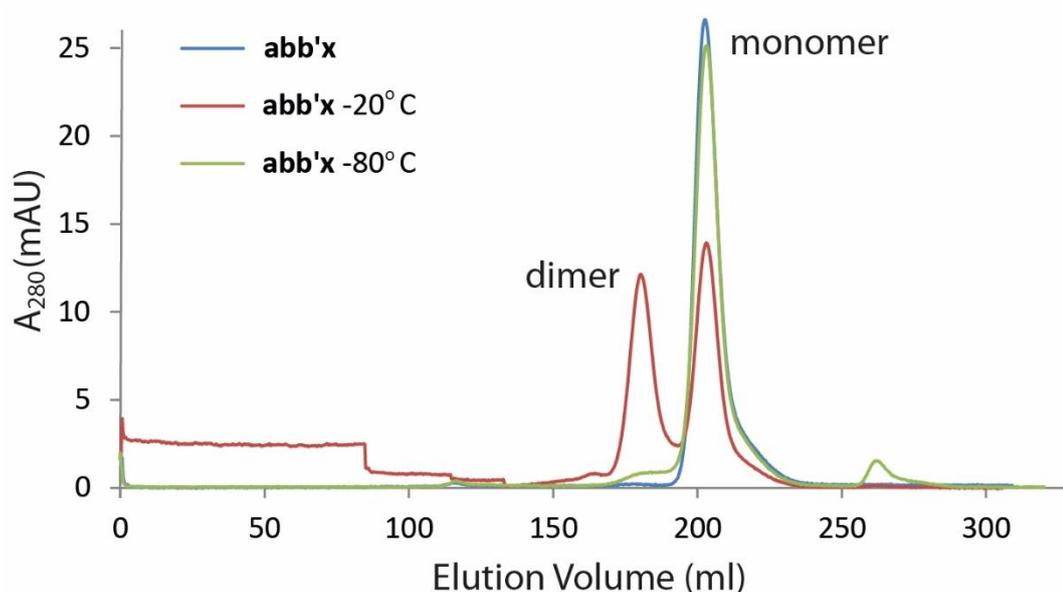
**Table 3.2: Expected and measured molecular weights of <sup>15</sup>N **b'**x and <sup>15</sup>N W111F **abb'**x.**

hPDI Fragment	Oligomeric State	Expected MW (Da)	Measured MW (Da)
<sup>15</sup> N <b>b'</b> x	<i>Monomer</i>	17486	19650
	<i>Dimer</i>	34971	39344
<sup>15</sup> N W111F <b>abb'</b> x	<i>Monomer</i>	41005	38049
	<i>Dimer</i>	82009	86671

The expected molecular weights were calculated based on amino acid sequence and assuming 98.9% <sup>15</sup>N incorporation (see section 3.3.5). The measured molecular weights were obtained using the size exclusion column calibration curve in Fig 3.5.

The largest molecular weight peaks visualised on size exclusion chromatography are approximately double the molecular weight of their corresponding monomer species, strongly suggesting dimer forms of the two proteins. However, both protein dimers ran larger than expected- **b'**x 12.5% larger and the W111F **abb'**x dimer 5.7% larger.

**Sample Stability on Storage-** The rate of conversion of monomer to dimer was studied for each protein at different storage temperatures (4°C, -20°C and -80°C). 1 mL aliquots of purified monomeric protein at 1 mg/mL in NMR buffer (20 mM sodium phosphate, 50 mM NaCl, pH 7.0) were either flash frozen on liquid nitrogen and stored at -80°C, or transferred to a -20°C freezer. After 1 week of storage **b'**x was present largely in a dimeric form, regardless of the storage temperature. W111F **abb'**x also formed a large proportion of dimer after storage for 1 week at -20°C, but remained predominantly monomeric after storage for the same length of time at -80°C. Only a small amount of dimer was observed, although a new smaller molecular weight peak was also present, suggesting possible degradation of the protein (see Figure 3.6). W111F **abb'**x was stable in its monomeric form after weekly storage at 4°C at concentrations of <1 mM, and **b'**x at concentrations <0.4 mM (higher concentrations promoted dimer formation as determined by size exclusion). As a result of these findings, protein was expressed and purified immediately prior to each NMR experiment, with minimal storage time at 4°C only, for optimal protein spectra.



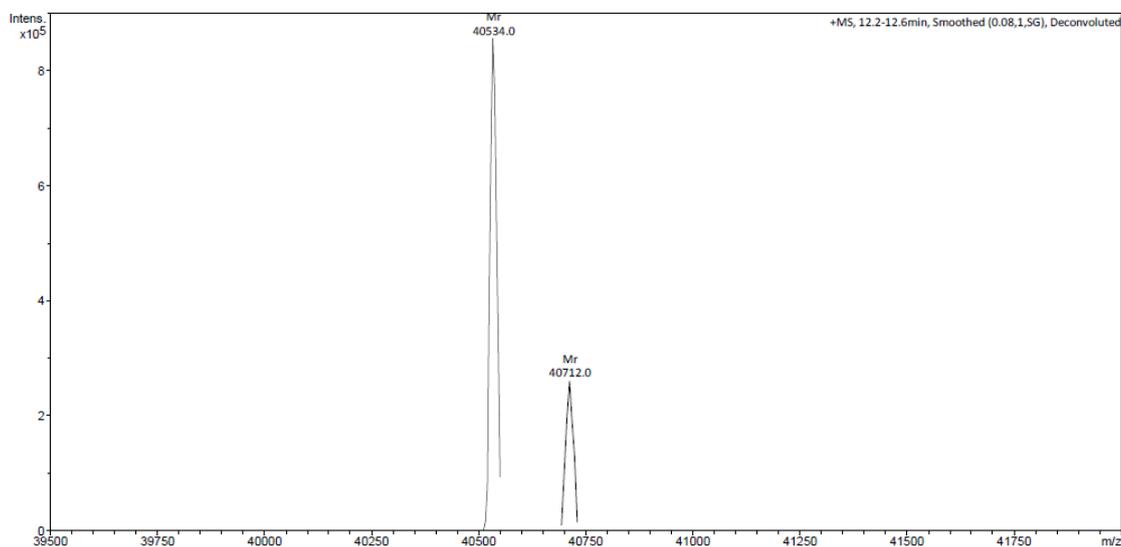
**Figure 3.6:** Size exclusion profile of 1 mg W111F **abb**'x protein using a 300 mL Superdex 200 Column. The blue trace was freshly purified monomeric W111F **abb**'x, the red trace W111F **abb**'x stored at -20°C for 1 week, and the green trace W111F **abb**'x stored at -80°C for 1 week.

### 3.3.5 Mass Determination and Isotopic Incorporation by Mass Spectrometry

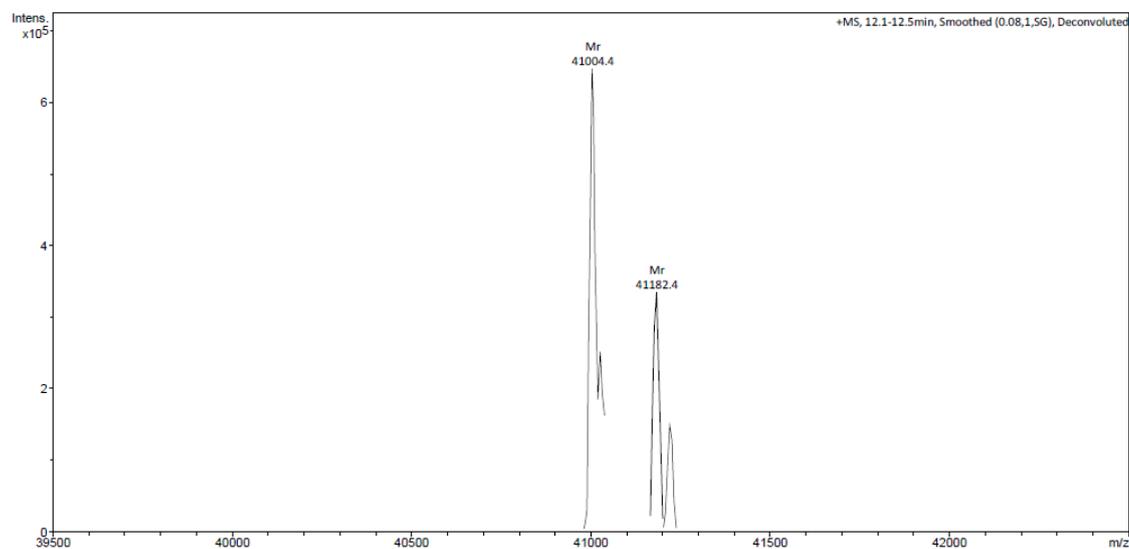
The **b**'x construct has been worked on extensively within the laboratory, and the monomeric protein mass previously determined (Byrne et al., 2009). Therefore in this thesis only mass spectrometry analysis of W111F **abb**'x was carried out.

The predicted molecular weight of W111F **abb**'x + His<sub>6</sub>-tag (as determined from its amino acid sequence) is 40534.7 Da. Mass spectrometry analysis confirmed the protein has a molecular weight of 40534.0 Da; the 0.7 Da difference is well within the 0.01% error range of analysis ( $\pm 4.05$  Da). There was also an additional species present in the sample, corresponding to a mass increase of +178 Da (see Figure 3.7). This is consistent with an  $\alpha$ -N-6-phosphogluconoylation of the His<sub>6</sub>-tag, a modification documented in recombinant proteins that should not affect the recombinant protein (Geoghegan et al., 1999). The modification is present in both the unlabelled and <sup>15</sup>N labelled protein samples (see Figure 3.8).

The major species present in  $^{15}\text{N}$  labelled W111F **abb**'x protein has a molecular weight of 41004.4 Da. The mass change for 100%  $^{15}\text{N}$  incorporation would be +475 Da; therefore the mass spectrometry analysis reports an incorporation extent of 98.9%.



**Figure 3.7:** Mass spectrometry analysis of unlabelled W111F **abb**'x.



**Figure 3.8:** Mass spectrometry analysis of  $^{15}\text{N}$  labelled W111F **abb**'x.

### 3.3.6 Final Yields

Accurate monomeric protein yields were determined using  $A_{280}$  values of pooled size exclusion chromatography fractions against a buffer blank. Total yields of monomeric **b**'x

and W111F **abb'x** protein are summarised in tables 3.3 and 3.4. An overnight induction of **b'x** at 27°C was trialled, as was growth of both proteins in an auto-induction media. Both resulted in a marked increase in dimer formation where the dimer was the major species seen on size exclusion. Overnight expression of W111F **abb'x** at 27°C, and expression of **b'x** for 2.5 hrs at 37°C were found to be optimal expression conditions; whilst this resulted in slightly lower overall protein expression levels, subsequent purification showed the highest ratio of monomer to dimer protein, simplifying purification and decreasing the risk of purifying a heterogeneous protein sample for NMR.

A drop in expression is always expected when switching from a rich (LB) to a defined (minimal) media, similarly a drop is seen between <sup>15</sup>N labelling and <sup>15</sup>N/<sup>13</sup>C labelling, on account of only using 1 g/L <sup>13</sup>C source in place of a 2 g/L <sup>12</sup>C source.

**Table 3.3: Average total yields of monomeric *b'x* as purified by size exclusion chromatography.**

<b>2.5 hour induction at 37°C</b>	
<b>b'x</b>	Yield (mg/L)
<b>LB</b>	30 ( <i>significant dimer present</i> )
<b>Unlabelled Minimal</b>	19
<b><sup>15</sup>N Minimal</b>	21
<b><sup>15</sup>N/<sup>13</sup>C Minimal</b>	11

**Table 3.4: Average total yields of monomeric W111F *abb'x* from varying expression conditions as purified by size exclusion chromatography.**

<b>abb'x</b>	<b>3 hour induction at 37°C</b>	<b>12 hour induction at 27°C</b>
	Yield (mg/L)	Yield (mg/L)
<b>LB</b>	54 ( <i>significant dimer present</i> )	63 ( <i>significant dimer present</i> )
<b>Unlabelled Minimal</b>	-	42
<b><sup>15</sup>N Minimal</b>	18	38
<b><sup>15</sup>N/<sup>13</sup>C Minimal</b>	-	10

### 3.4 Discussion

Both **b'**x and W111F **abb'**x were expressed as soluble proteins in rich and minimal media. The His<sub>6</sub>-tagged proteins were readily purified by metal affinity chromatography from the majority of other contaminating soluble *E.coli* proteins. Second-step purification using anion exchange chromatography generated a broad, single asymmetrical peak for **b'**x and two separate peaks for W111F **abb'**x. Both elution profiles were suggestive of multiple conformational forms. It has been previously published that hPDI fragments containing the **b'** domain form non-covalent homodimers (Wallis et al., 2009), (Byrne et al., 2009) so this result was not unexpected. As such, samples from across both elution profiles ran as single species on reducing SDS-PAGE, however by native-PAGE two distinct species were observed. Whilst some separation of W111F **abb'**x monomer/dimer was achieved at this stage, a further purification step of size exclusion chromatography was necessary to sufficiently separate the monomer and dimer forms of both proteins.

The molecular weights of the monomer and dimer species as determined by analytical size exclusion chromatography generally agreed with the expected molecular weights. Both **b'**x monomer and dimer ran on average 12.5% larger than expected. Full length hPDI is also published as having an anomalously large elution profile on gel filtration (Li et al., 2006), on account of its open and elliptical conformation. W111F **abb'**x monomer and dimer molecular weights agreed more closely with the expected values, with the monomer behaving 7% smaller and the dimer 5% larger. This suggests that W111F **abb'**x monomer may exist in a more compact state compared to full-length hPDI, before adopting a more similar open conformation upon dimerization.

Mass spectrometry of the W111F **abb'**x monomer protein confirmed a correct mass (40,534 Da) and mass spectrometry of <sup>15</sup>N labelled protein showed a good incorporation rate of 98.9%. Both unlabelled and <sup>15</sup>N labelled protein showed a minor species with a +178 Da mass increase. This modification is consistent with an  $\alpha$ -N-6-phosphogluconoylation of the His<sub>6</sub>-tag (Geoghegan et al., 1999). It wasn't shown to affect metal affinity chromatography purification of the protein, and the minor modified species was not resolvable from unmodified protein using size exclusion chromatography, on account of its comparatively small mass increase. Analysis of the modification was not

pursued further. It is a well-documented modification, and like the His<sub>6</sub>-tag itself is not likely to adversely impact upon protein structure or activity.

Final yields of monomer from both proteins were dependent on both growth medium and expression conditions. Both proteins gave lowest overall protein yields using minimal media; conversely this media also gave the highest monomer proportions. Induction for 2.5 hours at 37°C for **b'x** and overnight expression at 27°C for W111F **abb'x** gave the highest purified monomer yields. It may be that upon over expression in rich media (where expression rates are faster), the recombinant protein builds up to an overall higher concentration, and is therefore more likely to form stable dimers. It is likely that these dimers are present from the point of cell lysis.

It is interesting to note that, with exception of <sup>15</sup>N/<sup>13</sup>C labelled protein, W111F **abb'x** monomer yields were approximately double that of monomeric **b'x**, thought to be partly due to a greater monomer yield of W111F **abb'x** at the onset of protein purification. W111F **abb'x** was also more easily resolved from its dimeric form, both by anion exchange chromatography and size exclusion chromatography (for size exclusion this is likely due to the larger MW difference between monomer and dimer protein). A more suitable narrow-range size exclusion media could be sourced to increase separation of the monomer and dimer species of **b'x**, and so increase the overall monomeric yield.

*In vivo*, hPDI exists in a monomer:dimer equilibrium, a ratio that is thought to be influenced by the number of PDI substrates present within the ER at a given time. Dimerisation of hPDI renders the two hydrophobic ligand binding sites inaccessible to misfolded ligands, thus reducing PDI activity (for a recent review and structure of the dimer interface see (Bastos-Aristizabal et al., 2014a). *In vitro* hPDI forms dimers at high, but still physiological concentrations- as is reported in this Chapter. Unfortunately purified dimer is of little use in a laboratory setting, as it cannot be returned to its monomeric form. It is highly preferable that only homogenous, monomeric samples be taken forward for analysis using NMR spectroscopy, particularly if the dimeric protein shows reduced activity.

Both overall monomeric protein yields and the storage and stability experiments suggest that of the two hPDI fragments, W111F **abb'x** is more stable in a monomeric form at higher concentrations (dimerisation was promoted by storage for 1 week at 4°C at concentrations >1mM W111F **abb'x** and >0.4mM **b'x**). The differing results show that

when handling hPDI fragments their propensity to form dimers is something to be mindful of, particularly when running multi-dimensional NMR experiments, which can span time periods of several days.

## CHAPTER 4

# Assignment of hPDI **b'x** and W111F **abb'x** by NMR Spectroscopy

### 4.1 Introduction

Protein assignment by NMR spectroscopy allows identification and correlation of each active NMR nuclei within a protein. Specific  $^{15}\text{N}$  and  $^{13}\text{C}$  labelling techniques are exploited to allow sequential assignment of a protein's backbone and/or side-chain (Wüthrich, 1986). The approach relies upon each residue of the protein displaying a distinctive spin system—first each spin system is identified, then correlations between these systems are matched by use of a set of through-bond triple resonance experiments, to assign adjacent spin systems that correspond to amino acids sequential in primary sequence. The method is well used and reviewed, and allows the assignment of even large proteins in a relatively short period of time (Frueh, 2014; Göbl et al., 2014). NMR backbone and side-chain assignments can be used to accurately map changes in the protein in response to temperature, pH, protein redox state, or upon addition of other components to the system such as a binding ligand.

This thesis focuses on two fragments of hPDI: **b'x** and **abb'x**. **b'x** is the smallest monomeric fragment of hPDI required for peptide ligand binding, and the W111F **abb'x** fragment whilst also capable of peptide ligand binding also contains a redox active

catalytic site in its **a** domain. The W111F **abb'**x fragment has been used previously in fluorescence and limited proteolysis studies (Wang et al., 2010a), but NMR assignment of the protein fragment was not carried out as part of this work. Characterisation of the **b'**x fragment has previously been undertaken (Byrne et al., 2009; Nguyen et al., 2008).

Work in this thesis showed each protein gave optimal spectra at 37°C, and due to significant differences in the spectra of the **b'**x protein to previously assigned spectra (resonances were collected under differing temperature and buffer conditions), triple resonance backbone assignments of both protein fragments was carried out. In addition to this, methyl side-chain assignments were also made for the **b'**x fragment. These assignments enabled the investigation of peptide ligand binding in the presence or absence of neighbouring domains to the **b'** domain, or under different protein redox states, as discussed in subsequent Chapters.

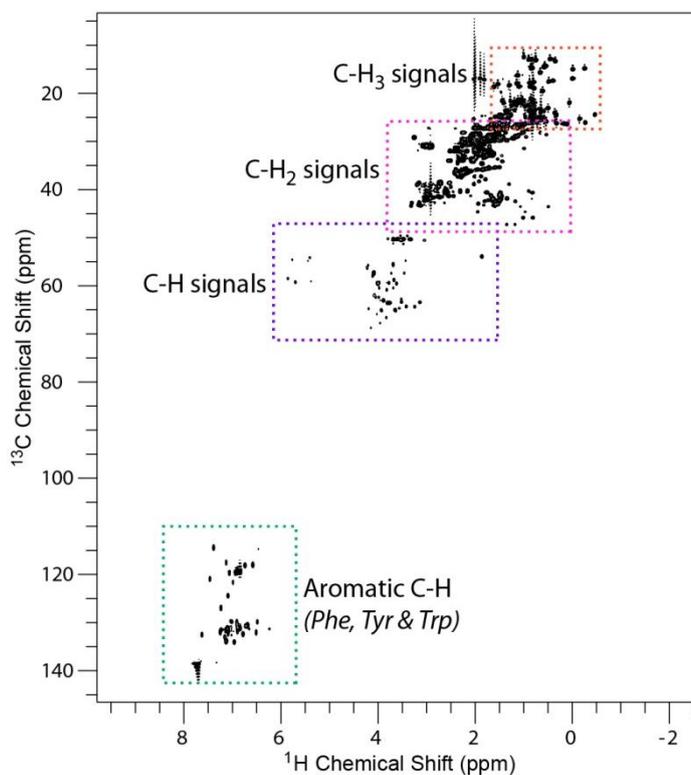
#### 4.1.1 $^{15}\text{N}, ^1\text{H}$ -HSQC Experiments

The  $^{15}\text{N}, ^1\text{H}$ -Heteronuclear Single Quantum Correlation (HSQC) is often the first two-dimensional NMR experiment performed on a biological protein. It is a 2D experiment that is optimised to show correlations between  $^1\text{H}$  directly bonded to NMR active  $^{15}\text{N}$  nuclei. The large majority of peaks visible will be the backbone amide groups, so one peak will be visible for each amino acid residue, with the omission of Pro residues (which do not have an amide proton for the correlation to be detected). Peaks can also be seen for Trp indole groups, and under acidic conditions the side chain groups of Asn and Gln. A  $^{15}\text{N}, ^1\text{H}$ -HSQC can be used to give general information about the protein quality- for example, if the protein is folded, and if the temperature and pH parameters are optimal. A  $^{15}\text{N}, ^1\text{H}$ -HSQC assigned using triple resonance experiments can inform on changes to the chemical environment that particular residues sense upon redox change or peptide ligand binding.

#### 4.1.2 $^{13}\text{C}, ^1\text{H}$ -HSQC Experiments

The  $^{13}\text{C}, ^1\text{H}$ -Heteronuclear Single Quantum Correlation (HSQC) experiment is the carbon equivalent of the  $^{15}\text{N}, ^1\text{H}$ -HSQC. It is another 2D experiment that is optimised to show all C-H correlations, between  $^1\text{H}$  directly bonded to NMR active  $^{13}\text{C}$  nuclei. Each carbon chemical environment can be observed in a specific area of the spectrum (see Figure 4.1)

and the  $^{13}\text{C}, ^1\text{H}$ -HSQC can be used much like a  $^{15}\text{N}, ^1\text{H}$ -HSQC, to report on protein status or upon changes the protein experiences, as sensed by the protein side-chain resonances.



**Figure 4.1:**  $^{13}\text{C}, ^1\text{H}$ -HSQC spectra of 0.25mM  $^{13}\text{C}$  **b**'x. Specific C-H chemical groups are highlighted. Tetramethylsilane ( $\text{Si}(\text{CH}_3)_4$ ) is used as a reference (0 ppm), as each of its protons are equivalent and highly shielded. Protons in  $\text{CH}_3$ ,  $\text{CH}_2$  and  $\text{CH}$  groups become progressively more de-shielded, becoming more exposed to the magnetic field and hence display a chemical shift  $>0$  ppm. Aromatic CH protons typically show the largest chemical shifts as the delocalised electrons in the benzene ring cause the nuclei to experience a stronger magnetic field.

### 4.1.3 Chemical Shift Perturbation Mapping

Chemical shift perturbation mapping is a method routinely used to measure sensitive changes in a protein; these perturbations report on the extent of shielding or de-shielding of the observed nuclei. They are measured using the 'chemical shift', measured in parts per million (ppm). 'ppm' is a value proportional to the peak position in Hz, that also factors in

the spectrometer frequency, allowing shifts to be directly compared independently of the spectrometer field strength. In this thesis they are used to show changes that occur from a change in temperature, redox state or upon addition of a binding ligand/protein.

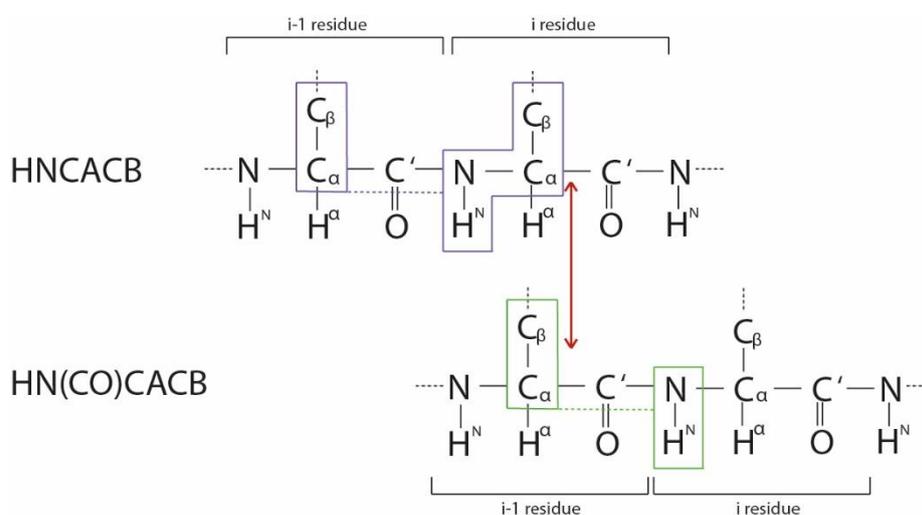
$^{15}\text{N}$ ,  $^1\text{H}$ -HSQCs are perhaps the most widely used spectra in biological NMR to track changes in a protein via chemical shift perturbation; each chemical shift reports on the unique chemical environment that a given backbone amide experiences in the protein. These chemical shifts can be mapped onto a protein sequence or structure to highlight specific residues involved in the event. For  $^1\text{H}$  and  $^{15}\text{N}$  nuclei, perturbations can be influenced by both through bond or through space interactions. It is also important to emphasise that chemical shift perturbations are exclusively reporting on a change in chemical environment. This means that in the case of ligand binding, the perturbations can either report directly upon the ligand binding event itself, or if the protein undergoes a structural re-arrangement upon binding, then alternatively the induced conformational change the protein experiences as well.

$^{13}\text{C}$  chemical shift perturbations are also used within this thesis to study ligand binding in more detail. Carbon chemical shifts are increasingly being used to monitor ligand binding, because  $^{13}\text{C}$  nuclei are less sensitive to structural changes in the protein, and should therefore be more sensitive reporters of direct ligand-binding events (Williamson, 2013). The study in this thesis focuses on the methyl region of the  $^{13}\text{C}$ ,  $^1\text{H}$ -HSQC, as these chemical groups form the hydrophobic side-chains of the aliphatic amino acids alanine, threonine, valine, leucine and isoleucine. Aliphatic residues are scattered throughout the **b'x** protein core, but also compose a large proportion of the hydrophobic binding site. These methyl 'reporters' of ligand binding are used to give more specific information about residues involved in the binding event.

#### 4.1.4 Sequential Backbone Assignment of **b'x** and W111F **abb'x** by Triple Resonance

Triple resonance experiments typically allow a large percentage of the  $^1\text{H}_\text{N}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  protein resonances to be unambiguously assigned. Assignment by HNCACB/HN(CO)CACB or HNCA/HN(CO)CA experiments were carried out using  $^{15}\text{N}/^{13}\text{C}$  isotopically enriched samples of **b'x** or W111F **abb'x** to aid in backbone assignment of the protein fragments. A backbone assignment allows each N-H peak in a  $^{15}\text{N}$ ,  $^1\text{H}$ - HSQC to be assigned to a specific amino acid residue. Triple resonance

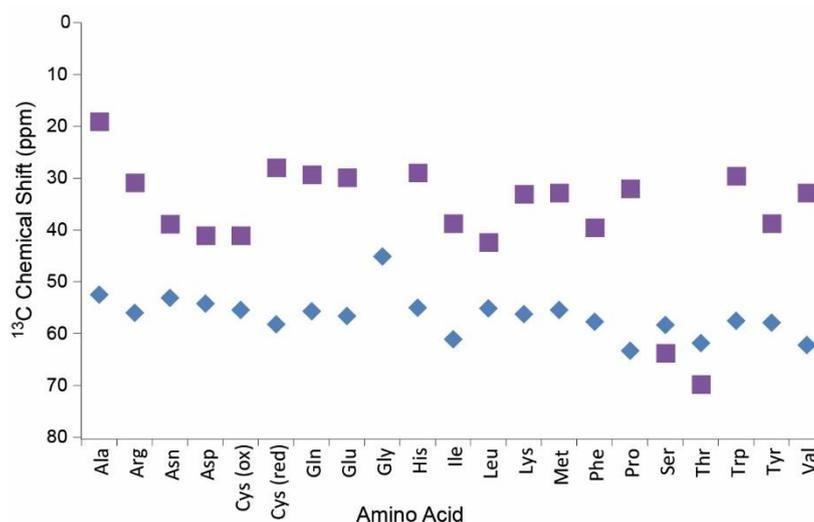
experiments generate  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  shifts for each residue, which with suitable software can be viewed in such a way that any given amide  $^1\text{H}$  and  $^{15}\text{N}$  correlation contains either the  $^{13}\text{C}$  resonances of the correlated residue and its sequentially preceding residue, denoted as 'i' and 'i-1' (HNCACB and HNCA), or just the  $^{13}\text{C}$  resonance of the preceding residue 'i-1' (HN(CO)CA/HN(CO)CACB). The experiments are named for the type of correlations they detect, and Figure 4.2 demonstrates how these correlations are linked for the HNCACB/HN(CO)CACB pair used to assign **b**'x (Grzesiek and Bax, 1992).



**Figure 4.2: Dipeptide assignment of a protein backbone showing correlations obtained in HNCACB and HN(CO)CACB experiments.** HNCACB correlates each NH strip with the  $\text{C}_\alpha$  and  $\text{C}_\beta$  shifts of its own residue (i) and that of the preceding (i-1) residue (shown in purple). HN(CO)CACB correlates each NH strip with just the preceding (i-1) residues  $\text{C}_\alpha$  and  $\text{C}_\beta$  shifts (shown in green). HNCA and HN(CO)CA experiments correlate shifts in the same manner, with omission of the  $\text{C}_\beta$  shifts.

Large proteins tend to give poorer NMR spectra as a result of slower molecular tumbling. HNCA (Kay et al., 1990) and HN(CO)CA (Bax and Ikura, 1991) experiments were used to assign the W111F **abb**'x backbone; these experiments are typically used for proteins >150 residues, and give the same information as the HNCACB and HN(CO)CACB, with the omission of  $\text{C}_\beta$  resonances.

The data from each triple resonance experiment was analysed using CcpNmr Analysis software, which ranks each data strip in order of best fit. The analysis relies upon characteristic carbon chemical shifts of each amino acid, which are assigned based on the random coil chemical shifts of amino acids (Wishart et al., 1995b). There are generally several matching 'strips', so the fits can first be manually assigned using the characteristic chemical shift patterns of alanine, glycine, serine and threonine  $C_{\alpha}$  and/or  $C_{\beta}$  resonances (see Figure 4.3), to assign small sequential sequences of the amino acid chain. After initial unambiguous assignment of small lengths of the protein, the segments can be lengthened and pieced together after other potential matches have been eliminated. Assignments can then be mapped onto the  $^{15}\text{N}, ^1\text{H}$ -HSQC to generate an assigned protein spectra.

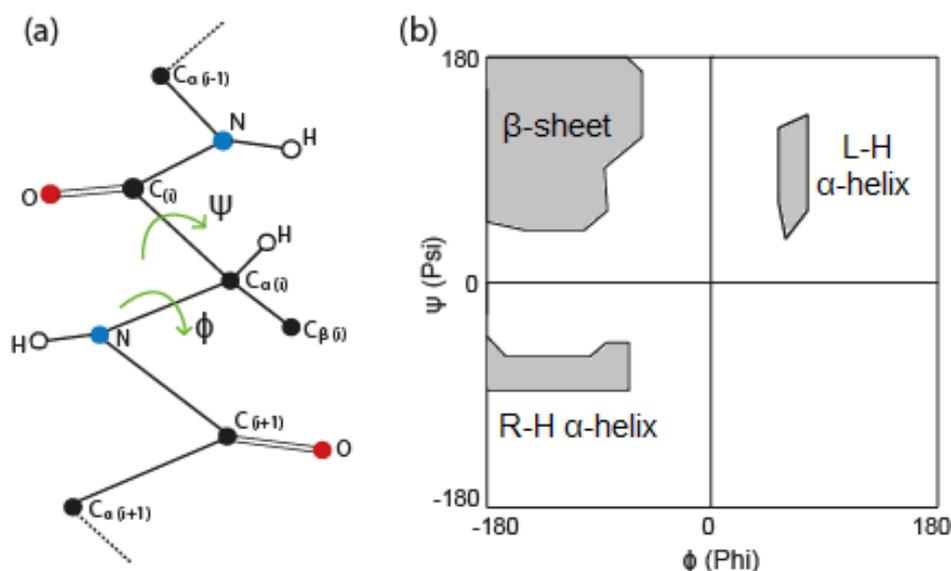


**Figure 4.3:** Random coil chemical shifts for  $C_{\alpha}$  (♦) and  $C_{\beta}$  (■) resonances. Adapted from data published by Wishart and Sykes (1994).

#### 4.1.5 Secondary Structure Prediction

The CcpNMR Analysis software contains an inbuilt programme called DANGLE (Dihedral Angles from Global Likelihood Estimates), that can be used for secondary structure predictions of proteins from their assigned resonances (Cheung et al., 2010). It uses an algorithm that can predict the most likely secondary structure of a protein, by

factoring in the amino acid type and its corresponding chemical shift in relation to its neighbouring residues. Protein backbones tend to adopt regular structural arrangements that are defined by dihedral angles, specific angles of rotation about the N-C $\alpha$  bonds ( $\phi$  Phi) and C-C $\alpha$  bonds ( $\psi$  Psi). Backbone secondary structures and their angles are typically displayed using Ramachandran plots as shown in Figure 4.4.

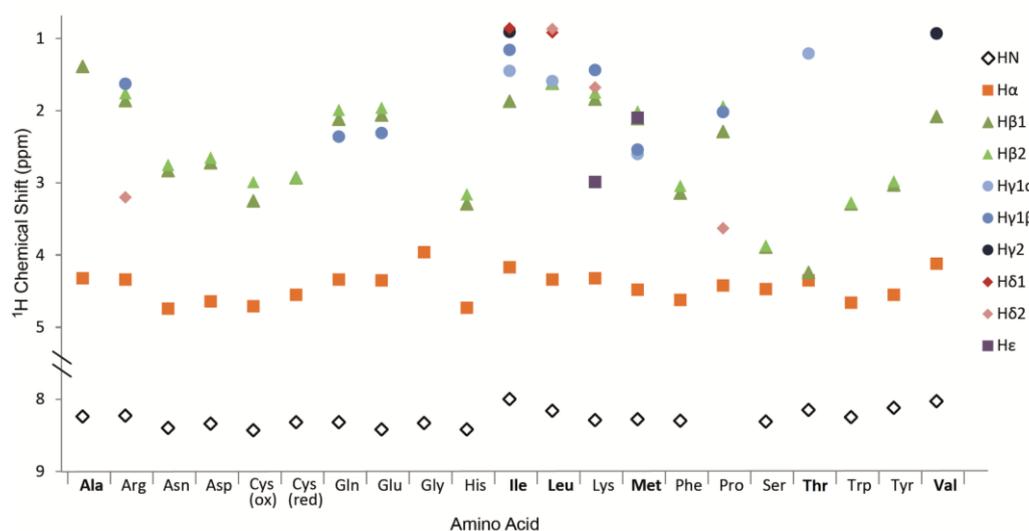


**Figure 4.4: A Ramachandran plot representation.** The illustration (a) shows the Phi angles ( $\phi$ ) between N-C $\alpha$  bonds and the Psi angles ( $\psi$ ) between C-C $\alpha$  bonds. Typical angles of each secondary structure type are highlighted on the graph in grey (b). L-H (left-handed) and R-H (right-handed).

#### 4.1.6 Sequential Side-Chain Assignment: HCCH-TOCSY

The HCCH-TOCSY experiment, along with backbone  $C_{\alpha}$  and  $C_{\beta}$  assignments, allows the side-chain resonances of a protein to be assigned. It uses  $^{13}\text{C}$  enriched protein and works by transferring magnetisation from side-chain  $^1\text{H}$  nuclei to the attached  $^{13}\text{C}$  nuclei, and then back to  $^1\text{H}$  for detection. In each CH strip, all of the side chain  $^1\text{H}$  resonances of that residue are potentially visible. Using the  $C_{\alpha}$  and  $C_{\beta}$  shifts of the backbone assignment, there will be a point that lies on the diagonal of the spectrum that has the same proton chemical shift and a distinctive strip of shifts. By navigating to these shifts, assignment of the  $\text{H}_{\alpha}$  and  $\text{H}_{\beta}$  resonances is possible. Using reference random coil proton chemical shifts

as a guide (see Figure 4.5), the assignment of the  $H_\gamma$ ,  $H_\delta$  and  $H_\epsilon$  resonances of the strip is also possible, and can be subsequently mapped onto the  $^{13}\text{C}, ^1\text{H}$ -HSQC. The HCCH-TOCSY was used to unambiguously assign the  $^{13}\text{C}, ^1\text{H}$ -HSQC methyl region of **b**'x in the absence and presence of ligand, to highlight hydrophobic side-chains that may be involved in the binding event.



**Figure 4.5: Proton random coil conformation chemical shifts for HN, Ha, Hβ, Hy, Hδ and Hε resonances.** Residues along the x-axis in bold contain CH<sub>3</sub> groups to be assigned. Shifts adapted from (Wishart et al., 1995a).

## 4.2 Materials and Methods

### 4.2.1 NMR Sample Preparation

Monomeric protein isotopically labelled and purified as outlined in Chapter 3 was buffer exchanged into NMR buffer (20 mM sodium phosphate, 50 mM NaCl, pH 7.0) using a Vivaspin 20 Centrifugal Filter (10 kDa MWCO, Sartorius) and then concentrated down into stocks to use as required (concentration determined by A<sub>280</sub> nm absorbance). Typical samples contained 0.1-1 mM protein, 5% D<sub>2</sub>O and 0.05% sodium azide, and for W111F **abb**'x an excess of reducing/oxidising agent. Deuterated samples were prepared in the

same manner by using a deuterated NMR buffer (20 mM sodium phosphate and 50 mM NaCl in D<sub>2</sub>O, pH 7.0). Samples were made up to a total volume of 330  $\mu$ L using NMR buffer in a sterile eppendorf, before being transferred to a 5mm Shigemi NMR tube (BMS-005V, Sigma).

#### 4.2.2 NMR Data Acquisition and Processing

NMR data was acquired with the assistance of Drs. Mark Howard and Michelle Rowe at the University of Kent using a 4-channel, 5-amplifier Bruker Avance III 14.1 T (600MHz <sup>1</sup>H) NMR spectrometer equipped with a 5mm QCI-F cryoprobe, except where stated. All NMR experiments were solvent suppressed using either WATERGATE suppression by GrAdient-Tailored Excitation (Piotto et al., 1992) or excitation sculpting (Hwang and Shaka, 1995) to reduce the water signal intensity. All experiments were carried out at 37°C, pH 7.0 unless stated otherwise. <sup>1</sup>H chemical shift referencing was carried out using the position of the <sup>1</sup>H<sub>2</sub>O resonance at the given temperature. <sup>15</sup>N and <sup>13</sup>C chemical shifts were referenced using spectrometer internal referencing but also confirmed or modified using the gyromagnetic ratio relationship (Wishart and Sykes, 1994).

All 1D data was processed using Bruker Topspin 3.2 software, and 2D and 3D NMR data was processed using NMRPipe software (Delaglio et al., 1995).

##### 4.2.2.1 <sup>15</sup>N,<sup>1</sup>H-HSQC Experiments of hPDI **b**'x and W111F **abb**'x

<sup>15</sup>N,<sup>1</sup>H-HSQC spectra were collected of each protein fragment over 45 mins with 2048 points in the F2 (<sup>1</sup>H) dimension and 256 points in the F1 dimension (<sup>15</sup>N). Assigned <sup>15</sup>N,<sup>1</sup>H-HSQC spectra are used throughout this thesis to observe the effect of temperature or redox state, and in subsequent chapters to monitor peptide ligand binding. Reduced or oxidised spectra of W111F **abb**'x were acquired using a 10x molar excess of either DTT or GSSG respectively. Temperature experiments were carried out using samples of reduced W111F **abb**'x and spectra collected at 25, 27.5, 30, 32.5, 35 and 37°C. Temperature titrations were also used to enable the assignment of W111F **abb**'x spectra at 25°C, to allow its comparison to other assigned hPDI spectra. Temperature experiments were carried out using samples of reduced **b**'x and spectra collected at 10, 15, 20, 25, 30, 35 and 37°C.

#### 4.2.2.2 $^{13}\text{C}$ , $^1\text{H}$ -HSQC Experiments of hPDI **b'x**

$^{13}\text{C}$ ,  $^1\text{H}$ -HSQC spectra of 0.4 mM **b'x** were acquired over 90 mins with 2048 points in the F2 dimension ( $^1\text{H}$ ) and 512 points in the F1 dimension ( $^{13}\text{C}$ ) in deuterated NMR buffer. Assigned  $^{13}\text{C}$ ,  $^1\text{H}$ -HSQC spectra were used in subsequent chapters to monitor peptide ligand binding.

#### 4.2.2.3 Sequential Backbone Assignment of **b'x** and W111F **abb'x**

**b'x Triple Resonance:** Backbone resonance assignments were collected using a 0.4 mM  $^{13}\text{C}/^{15}\text{N}$  labelled sample of **b'x**. Triple resonance datasets HNCACB and HN(CO)CACB were acquired with 2048 points in the F3 dimension ( $^1\text{H}$ ), 32 points in the F2 dimension ( $^{15}\text{N}$ ) and 256 points in the F1 dimension ( $^{13}\text{C}$ ).

**W111F *abb'x* Triple Resonance:** Backbone resonance assignments were collected using a 0.4 mM  $^{13}\text{C}/^{15}\text{N}$  labelled sample of W111F **abb'x**. Triple resonance datasets HNCA and HN(CO)CA were acquired with 2048 points in the F3 dimension ( $^1\text{H}$ ), 40 points in the F2 dimension ( $^{15}\text{N}$ ) and 128 points in the F1 dimension ( $^{13}\text{C}$ ). Carrier frequencies for experiments were set to 4.586 ppm for  $^1\text{H}$ , 41.521 ppm for  $^{13}\text{C}$  and 117.868 ppm for  $^{15}\text{N}$ .

#### 4.2.2.4 Sequential Side-Chain Assignment of **b'x**

Side-chain resonance assignments were collected using a 0.4 mM  $^{13}\text{C}/^{15}\text{N}$  labelled sample of **b'x**. HCCH-TOCSY spectra were acquired with 1024 points in the F3 dimension ( $^1\text{H}$ ), 60 points in the F2 dimension ( $^{15}\text{N}$ ) and 256 points in the F1 dimension ( $^{13}\text{C}$ ) in deuterated NMR buffer. Carrier frequencies for experiments were set to 4.586 ppm for  $^1\text{H}$ , 45.919 ppm for  $^{13}\text{C}$  and 119.454 ppm for  $^{15}\text{N}$ .

### 4.2.3 NMR Data Analysis

Analysis of 2D and 3D spectra was carried out using CcpNmr Analysis software Version 2 (Vranken et al., 2005).

#### 4.2.3.1 Triple Resonance Backbone and Side-Chain Assignments

Sequential backbone assignment of **b'x** and W111F **abb'x** was achieved using triple resonance datasets, assigned as outlined in the introduction of this chapter, using random

coil chemical shifts (starting with the characteristic resonances of alanine, glycine, serine and threonine) as determined by Wishart et al. (Wishart et al., 1995b).

#### 4.2.3.2 Minimal Chemical Shift Perturbation Mapping

Chemical shift perturbation mapping was used to identify changes to specific residues of the protein upon a change in chemical environment (in this instance changes in reduced or oxidised protein). Minimal chemical shift perturbations are used to compare peaks from a fully assigned spectrum with the nearest peaks in a related but unassigned spectrum. This is useful when spectra are sufficiently different that unassigned peaks from a spectrum cannot be unambiguously identified from nearest neighbouring peaks of an assigned spectrum. One such example was the significant shift or line broadening changes observed as a result of oxidising W111F **abb'**x and comparing this to the assigned spectrum from the reduced protein. Minimal chemical shifts can sometimes be an underestimation of true shifts occurring, but require assignment of only one spectrum.

Minimal chemical shift perturbations were calculated using equation 4.1:-

$$4.1 \text{ Chemical Shift Perturbation} = \sqrt{(\Delta {}^1H_N)^2 + \left(\frac{1}{6}\Delta {}^{15}N\right)^2}$$

Where  $\Delta {}^1H_N$  and  $\Delta {}^{15}N$  are the chemical shift perturbations (ppm) of a peak in the  ${}^1H_N$  and  ${}^{15}N$  dimensions respectively. The minimal change in chemical shift perturbation was plotted graphically against residue number.  ${}^{15}N$  shifts are weighted by 1/6 their actual size, to compensate for the large chemical shift range they report over when compared to proton chemical shifts; the chemical shift range of  ${}^{15}N$  is  $\sim 30$  ppm compared to  ${}^1H_N \sim 5$  ppm.

#### 4.2.3.3 Secondary Structure Predictions of **b'**x and W111F **abb'**x

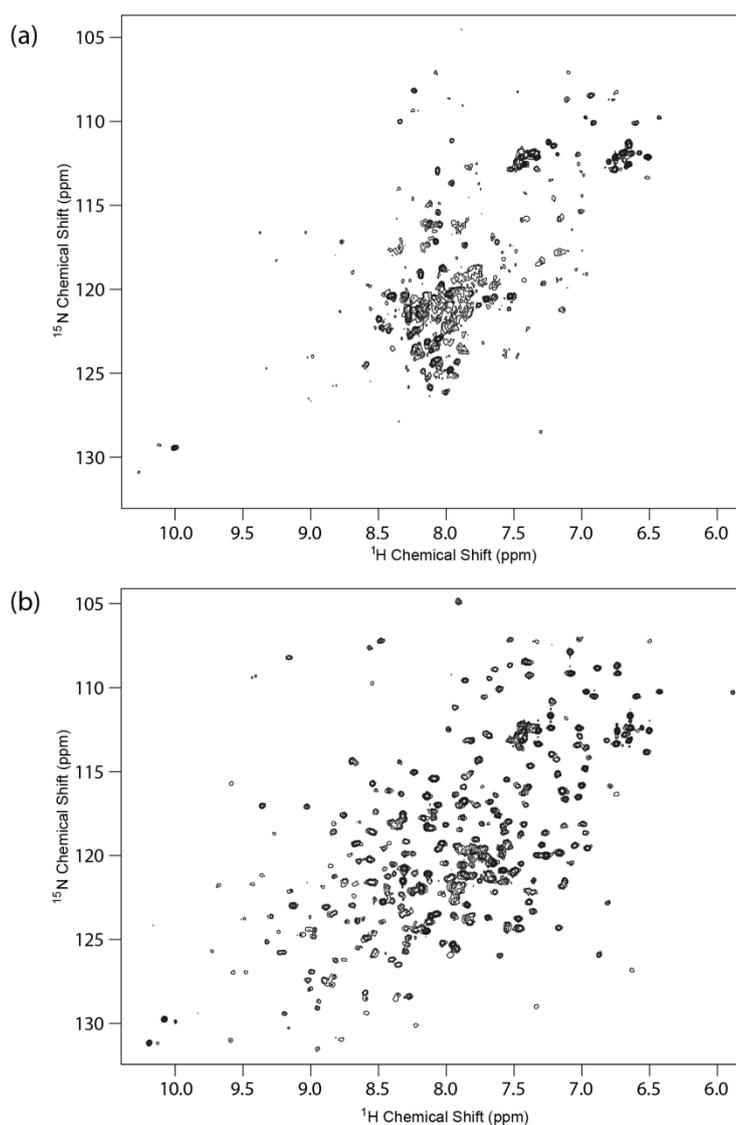
Using the  ${}^1H_N$ ,  ${}^{15}N$ , and  ${}^{13}C_\alpha$  chemical shifts obtained from triple resonance experiments of **b'**x and W111F **abb'**x, Ramachandran plots of each residue were generated using CcpNMR Analysis software's inbuilt DANGLE programme (Dihedral Angles from Global Likelihood Estimates). This software takes chemical shift assignments and compares them against a database of published protein structures and their chemical shifts, to generate a likely prediction of secondary structure. The software analyses the population distribution

expected for each specific residue (for example glycine has much greater conformational freedom ( $\phi$  and  $\psi$ ) than most residues, so can appear over a larger range of regions than shown by Figure 4.4). Prediction limits were set such that any residue with several predicted  $\phi/\psi$  angle regions was excluded from the secondary prediction. Each residue was assigned a likely secondary structure that was either random coil,  $\beta$ -strand or  $\alpha$ -helix.

## 4.3 Results

### 4.3.1 Effect of Redox State on W111F *abb*'x

Preliminary NMR studies of W111F **abb**'x showed that the protein is folded, stable and gives fair spectra at 37°C. Spectra were collected for oxidised and reduced protein at pH 7.0 using a 10x molar excess of GSSG or DTT<sup>red</sup> respectively (see Figure 4.6). The reduced form of the protein (Figure 4.6 (b)) gave a greater number of well-resolved peaks when compared to the oxidised spectra (Figure 4.6 (a)), therefore triple resonance experiments were carried out using the reduced form of W111F **abb**'x.

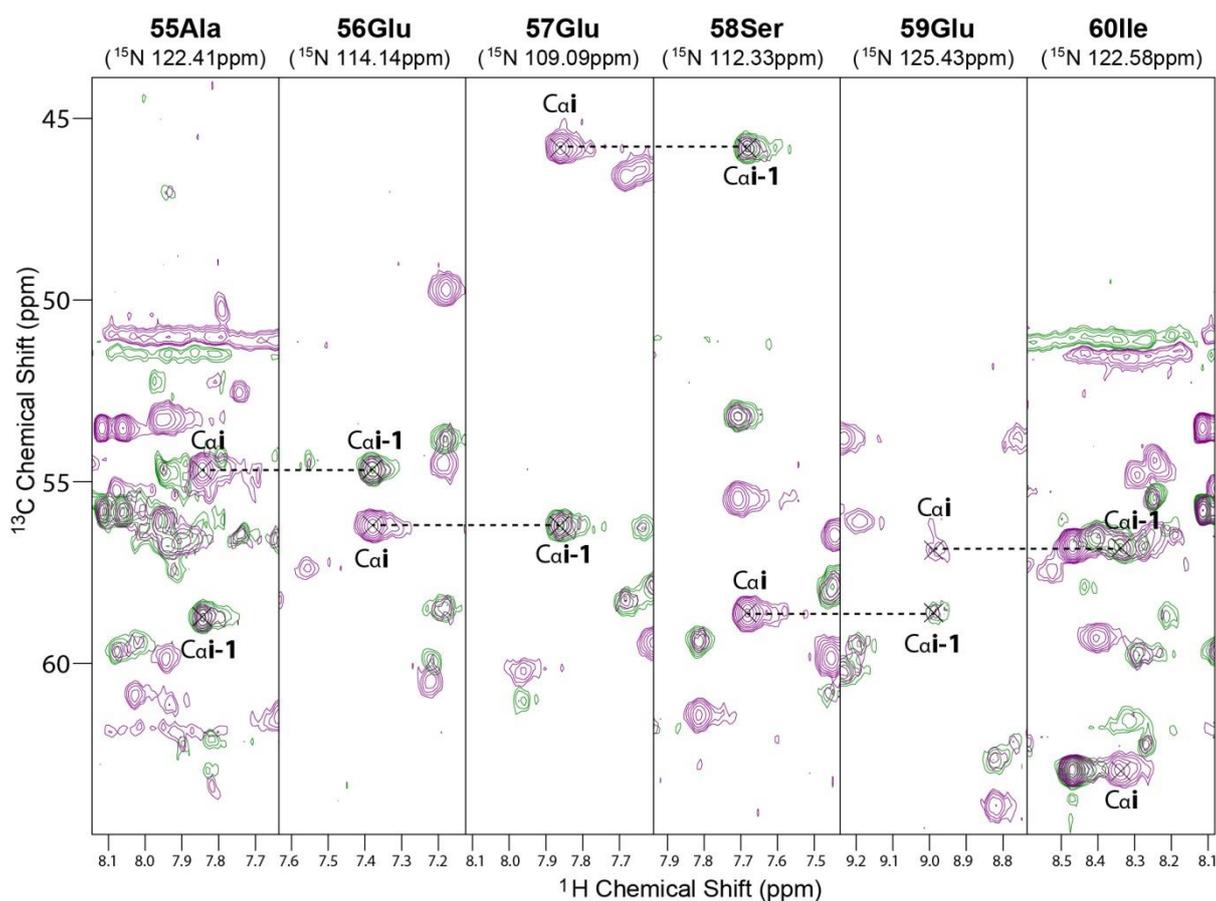


**Figure 4.6:**  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of 0.25mM W111F **abb'**x + 5 mM GSSG (a) or + 5 mM DTT<sup>red</sup> (b). Resolution and number of peaks improves when the protein is in a reduced form.

### 4.3.2 Sequential Backbone Assignment of Reduced W111F **abb'**x

Sequential backbone assignments of the  $^1\text{H}_\text{N}$ ,  $^{15}\text{N}_\text{H}$ , and  $^{13}\text{C}_\alpha$  nuclei of W111F **abb'**x were made using the triple resonance experiments HNCA and HN(CO)CA. Sequential assignments were made by using the characteristic shifts of residues such as glycine, serine or threonine, then ‘walking’ through to the next residue. A typical example of the sequential assignment process is shown in Figure 4.7 for A55-I60. Ignoring the His<sub>6</sub>-tag

and unassignable Pro residues, backbone assignments were made for 68% of the amide  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}_\alpha$  resonances (see Appendix 4.1 for assignment tables). Triple resonance assignments were transferred onto a  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC of W111F **abb**'x (as shown in Figure 4.8), for use in future experiments.

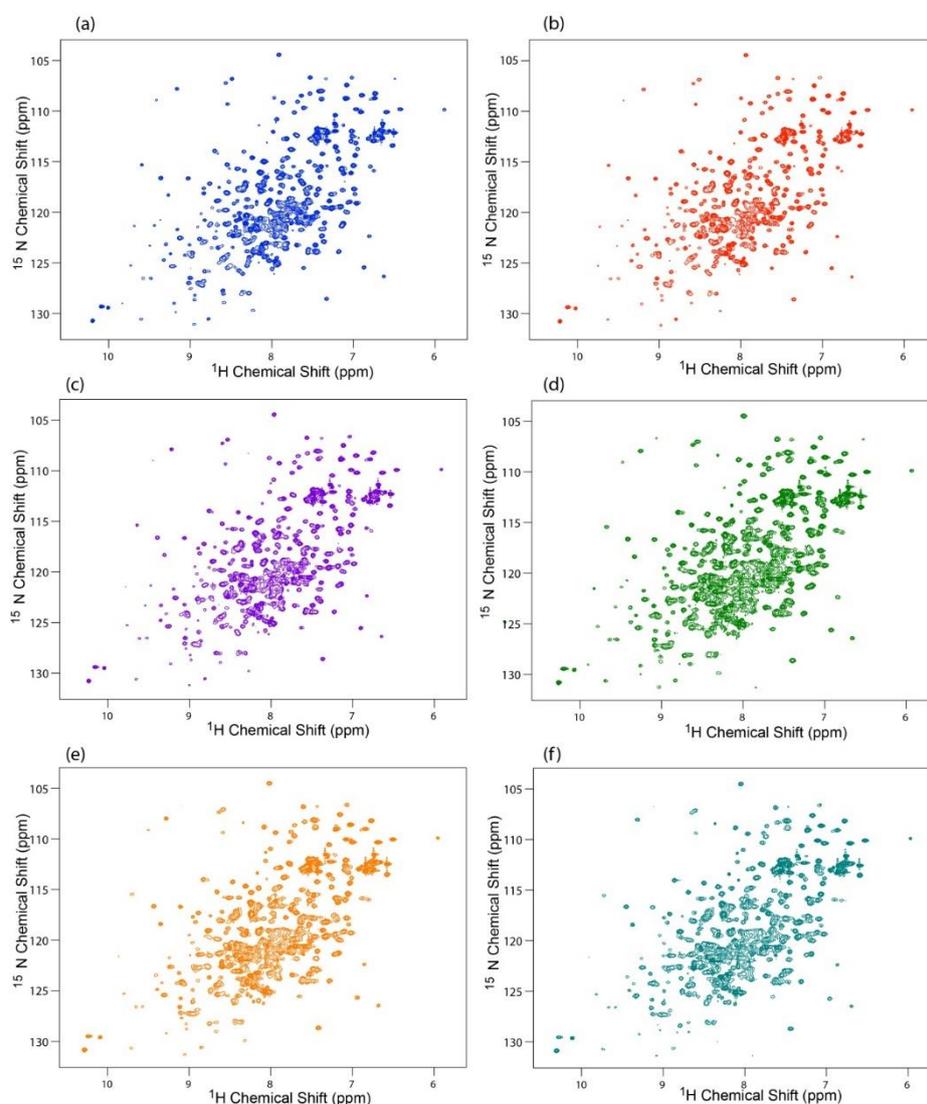


**Figure 4.7: Protein backbone triple resonance sequential assignment example of W111F *abb*'x.** HNCA (purple) and HN(CO)CA (green) strips demonstrating the assignment stretch for residues A55 to I60. Sequential matches between the strips are shown by dotted lines. The corresponding  $^{15}\text{N}$  chemical shift of each plane is shown above each strip (ppm).



### 4.3.3 Effect of Temperature on W111F *abb*'x

Within the laboratory larger multi-domain fragments of hPDI (>30kDa) are typically studied at 37°C, as this temperature yields spectra with greatest peak resolution. Smaller individual domains have typically been assigned at 25°C. To enable comparison of the W111F **abb**'x fragment of hPDI with other individual domain assignments,  $^{15}\text{N}, ^1\text{H}$ -HSQC spectra were collected of the protein at 37°C, 35°C, 32.5°C, 30°C, 27.5°C and 25°C (spectra shown in Figure 4.9). Temperature dependent shifts could then be tracked throughout the titration series to assign the spectra at 25°C (see Figure 4.10) to 63% of the backbone amide  $^1\text{H}$  and  $^{15}\text{N}$  resonances.

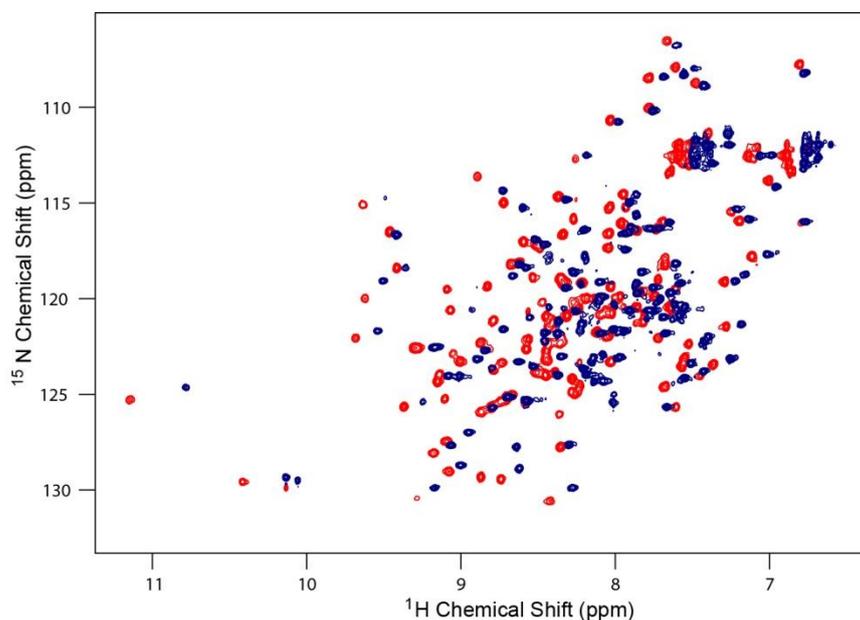


**Figure 4.9: Temperature titration  $^{15}\text{N}, ^1\text{H}$ -HSQC spectra of W111F *abb*'x.** Spectra collected at 37°C (a), 35°C (b), 32.5°C (c), 30°C (d), 27.5°C (e) and 25°C (f) allowed assignments of W111F **abb**'x at 37°C to be mapped onto spectra at 25°C, for comparison with individual domain spectra.



#### 4.3.4 Sequential Backbone Assignment of *b'*x

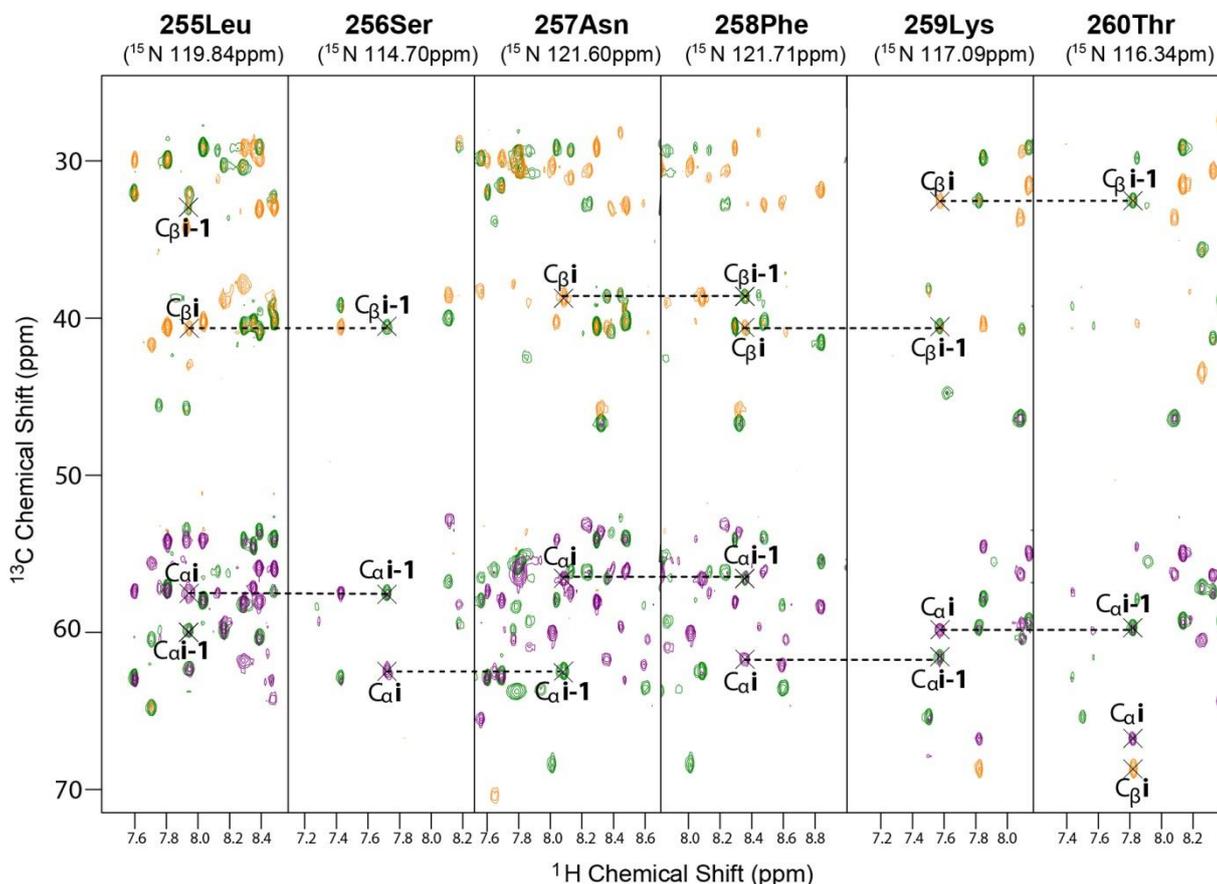
The **b'**x fragment of hPDI has previously been assigned at pH 6.5 and 25°C. As shown previously within the laboratory, observed protein chemical shifts of W111F **abb'**x fragment are highly sensitive to temperature, and a direct transfer of previously assigned resonances onto protein at 37°C by titration was not possible (see Figure 4.11 for overlaid spectra).



**Figure 4.11** Overlaid  $^{15}\text{N}, ^1\text{H}$ -HSQC spectra of *b'*x, pH 7.0 and 37°C (blue), and *b'*x at pH 6.5 and 25°C (red). (Referencing was based on  $^1\text{H}$  carrier positions of 4.72 ppm at 25°C and 4.586 ppm at 37°C).

To allow a direct comparison of the W111F **abb'**x and **b'**x fragments at 37°C, sequential backbone assignments of the  $^1\text{H}_\text{N}$ ,  $^{15}\text{N}_\text{H}$ ,  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  nuclei were made using the triple resonance experiments HNCACB and HN(CO)CACB. Ignoring the His<sub>6</sub>-tag and unassignable Pro residues, backbone assignments were made for 85% of the amide  $^1\text{H}$  and  $^{15}\text{N}$ , and 76% of the  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  resonances (see Appendix 4.2 for assignment tables). Sequential assignments were made by using the characteristic shifts of residues such as serine, threonine, alanine and glycine, then ‘walking’ through to the next residue. A typical

example of the sequential assignment process is shown in Figure 4.12. Triple resonance assignments were then transferred onto a  $^{15}\text{N}, ^1\text{H}$ -HSQC of **b'**x as shown in Figure 4.13.

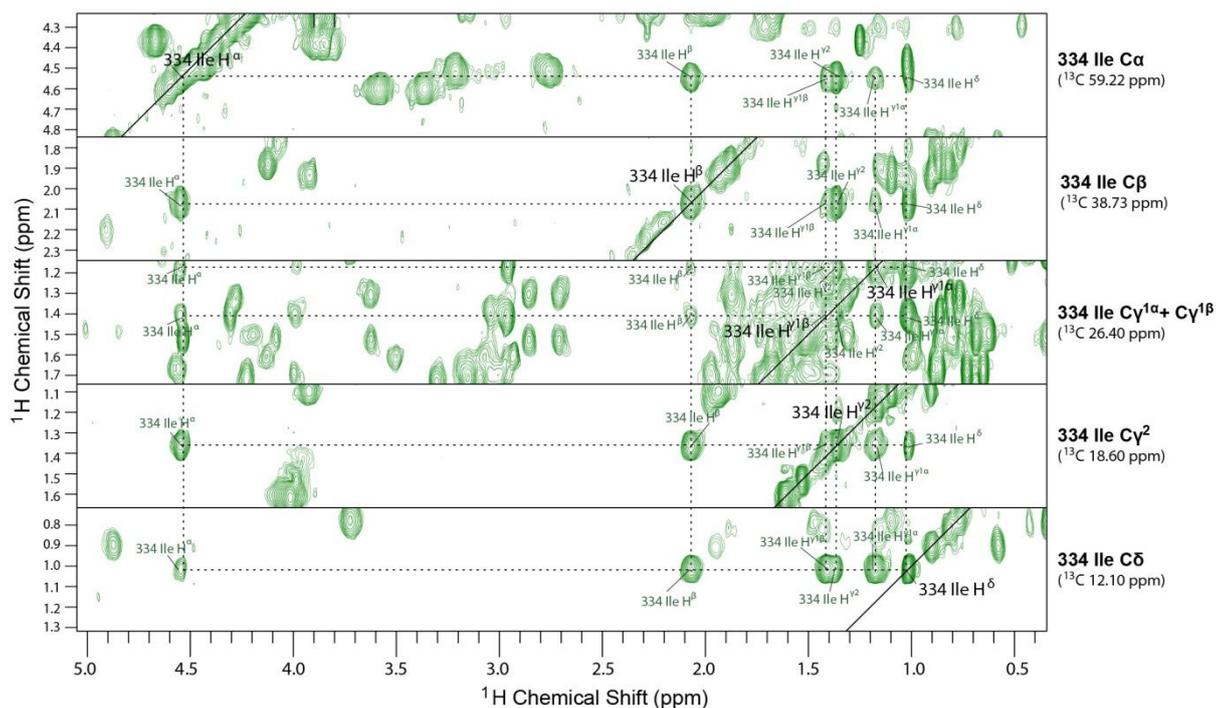


**Figure 4.12: Protein backbone triple resonance sequential assignment example of *b'*x.** HNCACB and HN(CO)CACB strips demonstrating the assignment stretch for residues L255 to T260. The HNCACB shows  $C_\alpha$  (purple) and  $C_\beta$  (yellow) resonances of the 'i' and 'i-1' residues, and the HN(CO)CACB spectra shows the resonances of just the  $C_\alpha$  and  $C_\beta$  'i-1' residues (green). Sequential matches between strips are shown by dashed lines. The corresponding  $^{15}\text{N}$  chemical shift of each plane is shown above each strip. Please note that S256 is missing its  $C_\beta$ , so in this instance could only be followed by its  $C_\alpha$  resonance (all Gly residues had already been assigned).

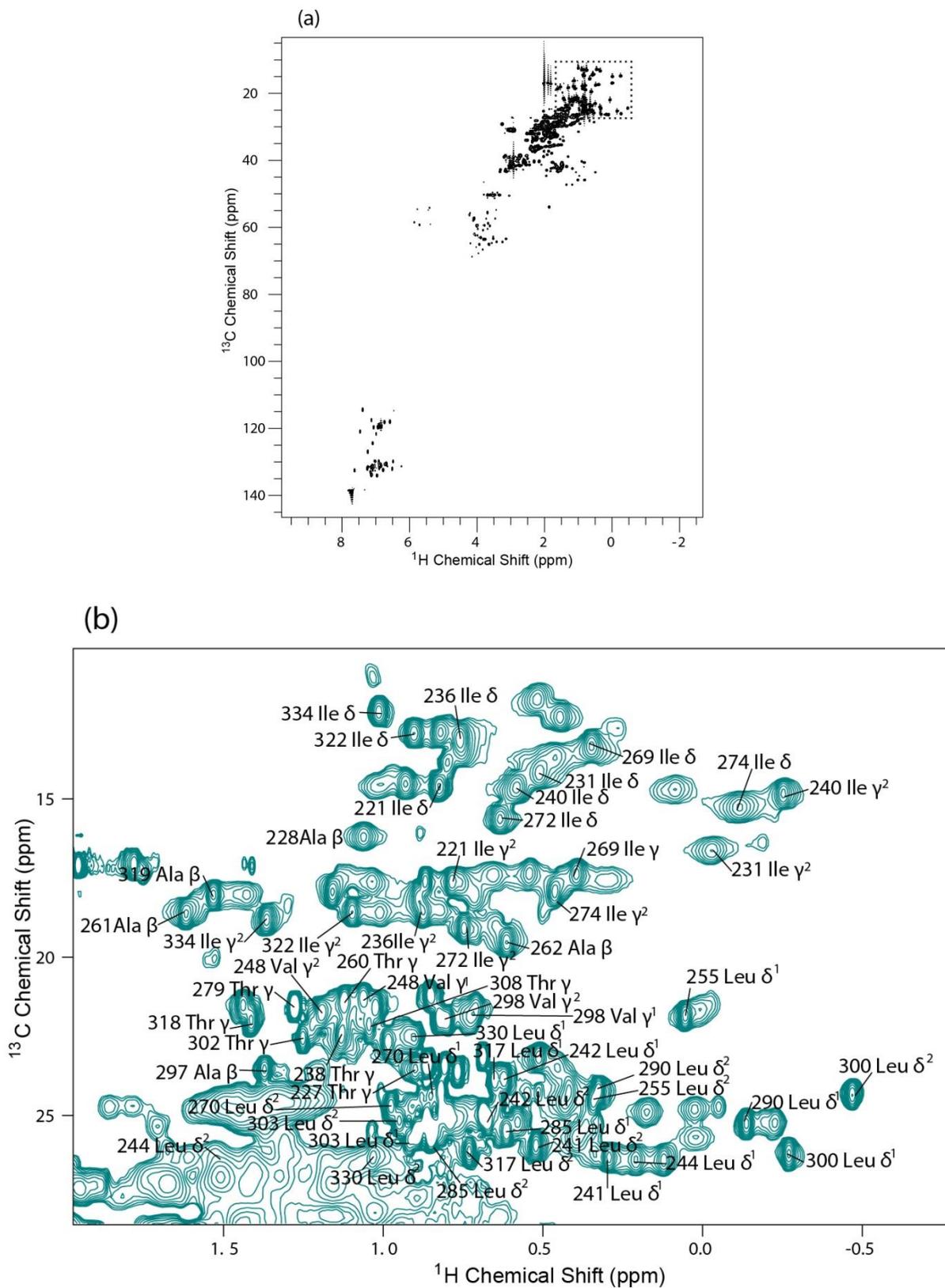


### 4.3.5 Side-Chain Assignment of *b'*x

Methyl  $^{13}\text{C}$  side-chain assignments of the alanine, threonine, valine, leucine and isoleucine resonances of **b'**x were made using the HCCH-TOCSY experiment, in addition to the triple resonance backbone assignments.  $^{13}\text{C}$   $\text{C}\alpha$  and  $\text{C}\beta$  assignments could be linked to the HCCH-TOCSY to assign the  $\text{H}\alpha$  and  $\text{H}\beta$  resonances, and subsequently the  $\text{H}\gamma$ ,  $\text{H}\delta$  and  $\text{H}\epsilon$  resonances, using the random proton chemical shifts as illustrated in Figure 4.5. A typical example of strips used in the assignment process is shown in Figure 4.14 for residue I334. Side-chain assignments were made for 89% of the methyl  $^1\text{H}$  and  $^{13}\text{C}$  resonances (see Appendix 4.3 for assignment tables). Figure 4.15 shows the assigned  $^{13}\text{C}$ ,  $^1\text{H}$ -HSQC methyl region (b) as shown by the dotted box in (a).



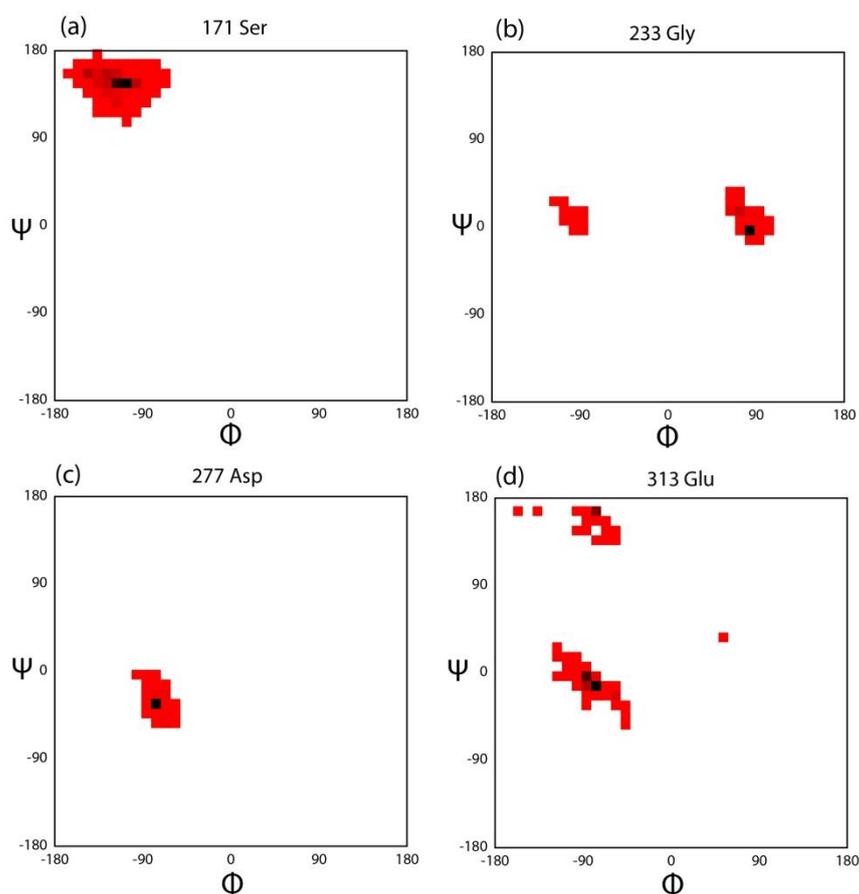
**Figure 4.14:** Example C-H 'strips' showing correlations obtained in the HCCH-TOCSY experiment of  $^{334}\text{Ile}$ . HCCH-TOCSY correlates all proton resonances bound to a particular  $^{13}\text{C}$  nucleus. The corresponding  $^{13}\text{C}$  chemical shift of each plane is shown to the right of each strip.



**Figure 4.15:**  $^{13}\text{C}, ^1\text{H}$ -HSQC spectra of **b'**x (a) and methyl region (b). Assignable methyl resonances (as highlighted by the dashed box in (a)) are labelled on the methyl region in (b).

### 4.3.6 Secondary Structure Predictions of *b*'x and W111F *abb*'x

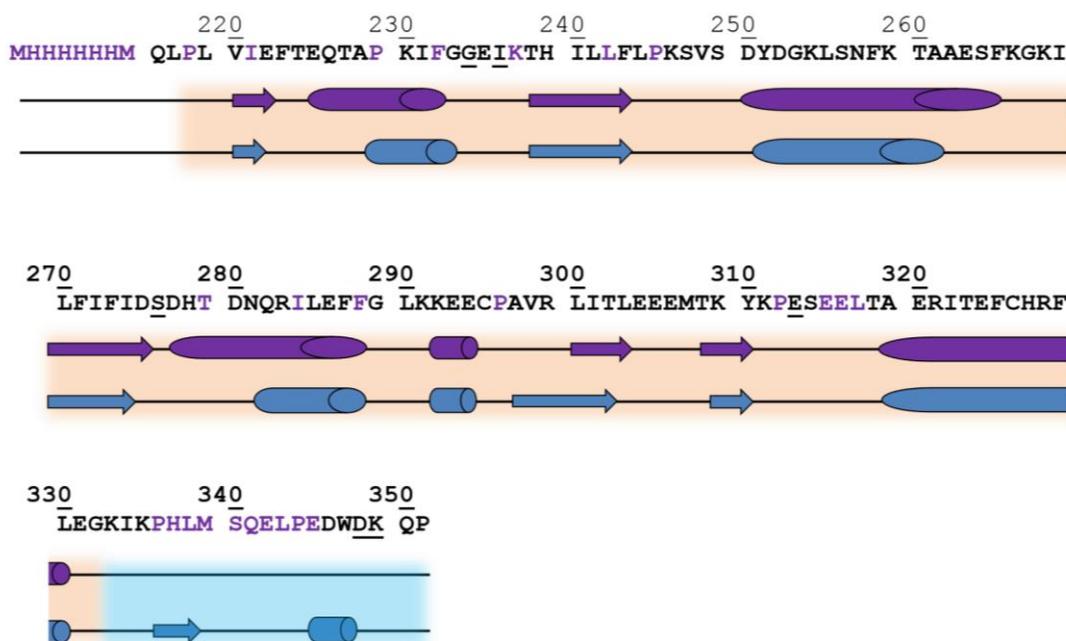
The  $^1\text{H}_\text{N}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}_\alpha$  and/or  $^{13}\text{C}_\beta$  assignments of **b**'x and W111F **abb**'x were used for secondary structure prediction by DANGLE. The programme predicted the Psi ( $\psi$ ) and Phi ( $\phi$ ) angles along with the most likely secondary structure of each amino acid, generating three possible outcomes: random coil, strand or helix. Typical Ramachandran plot examples of W111F **abb**'x are shown in Figure 4.16 for residues S171, G233, D277 and E313. Angles predicted for a single location (Figure 4.16 (a) and (c)) were assigned as  $\beta$ -strand and  $\alpha$ -helix respectively. When two islands of angles were present (Figure 4.16 (b)) a random coil was assigned. Residues that generated more than two islands of angles (Figure 4.16 (d)) were not included in the secondary structure prediction.



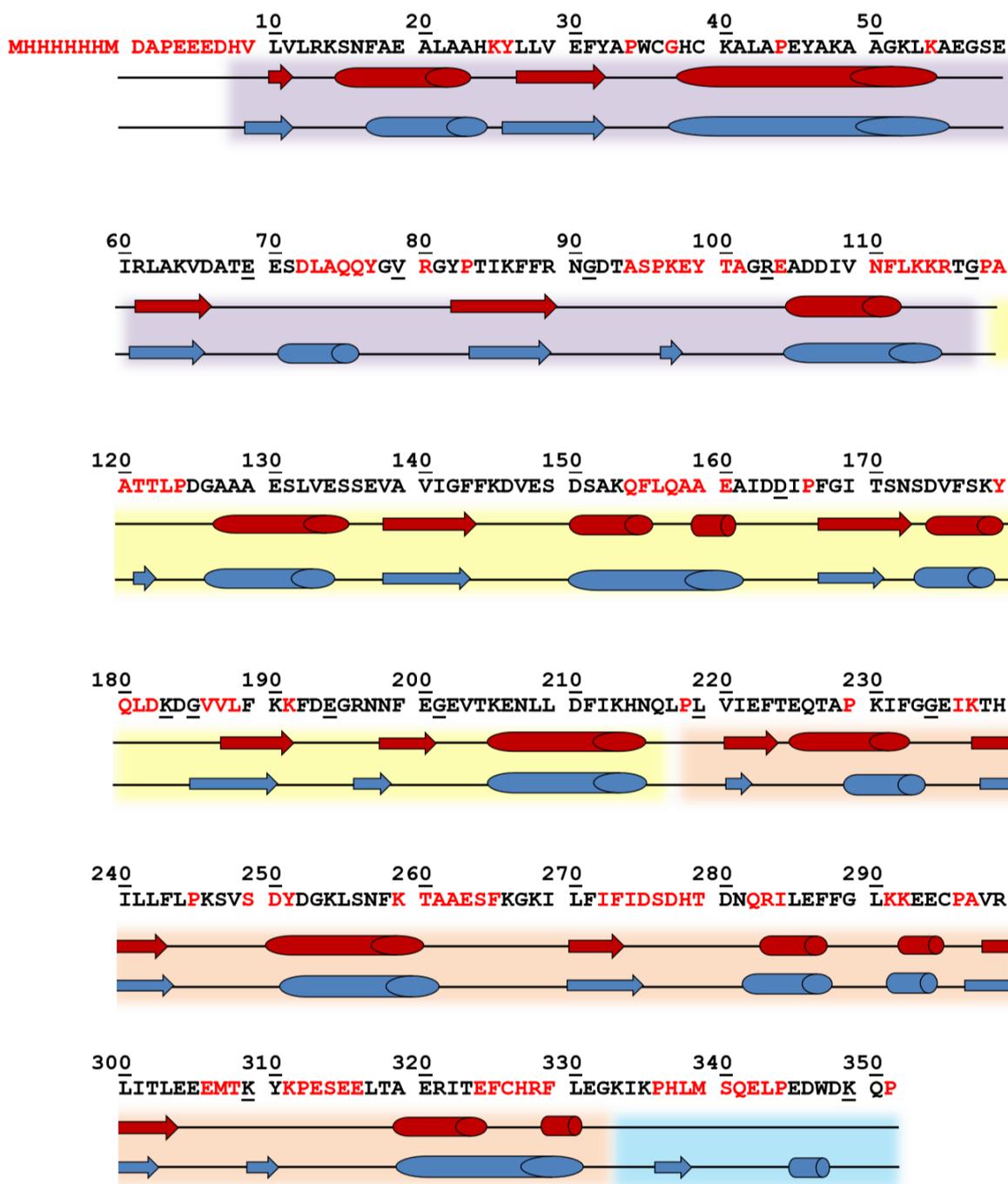
**Figure 4.16:** Example of secondary structure prediction of W111F *abb*'x as generated by DANGLE. S171 predicted as  $\beta$ -strand (a), G233 predicted as coil (b), D277 predicted as  $\alpha$ -helix (c) and E313 excluded from secondary structure calculation as multiple island locations were present (d).

The secondary structure predictions for **b'**x and W111F **abb'**x, with the corresponding region of the reduced hPDI **abb'**xa' crystal structure (4EKZ.pdb) for comparison, are shown in Figure 4.17 and Figure 4.18 respectively. DANGLE was successful in predicting secondary structure of both **b'**x and W111F **abb'**x, with the predicted structures in overall agreement with the published crystal structure of near full-length protein.

Small differences were seen in the W111F **abb'**x structure predictions and the crystal structure; an additional  $\alpha$ -helix and  $\beta$ -strand in the **a** domain, and small  $\beta$ -strands at the start of the **b** domain and end of the **b'** domain, which were present only in the crystal structure. These secondary structure elements were not predicted from the triple resonance data due to missing assignments of these stretches by NMR. Additionally, no secondary structure of the **x-linker** was predicted for either fragment, thought to be due in part to the linker regions low assignment (56%), caused by its conformational activity in solution.



**Figure 4.17: Backbone assignments and predicted secondary structure of *b'*x.** The *b'*x predicted secondary structure sequence is shown in purple, with the corresponding sequence of reduced *abb'*xa' crystal structure (4EKZ.pdb) for comparison shown in blue.  $\beta$ -strand structure is denoted by arrows and  $\alpha$ -helices by cylinders. Domain boundaries are **b'** (P218-G332) in orange and **x-linker** (K333-P351) in blue. Residues that could not be assigned as a result of ambiguous or missing triple resonance data are highlighted in the primary sequence in purple. Residues not included in the secondary structure predictions due to their  $\phi$  and  $\psi$  angles mapping to more than two clusters are shown underlined.



**Figure 4.18: Backbone assignments and predicted secondary structure of W111F **abb'**x.** The W111F **abb'**x predicted secondary structure sequence is shown in red, with the corresponding sequence of reduced **abb'**xa' crystal structure (4EKZ.pdb) for comparison in blue.  $\beta$ -strand structure is denoted by arrows and  $\alpha$ -helices by cylinders. Domain boundaries are **a** (D7-G117) in purple, **b** (A119-Q216) in yellow, **b'** (P218-G332) in orange and **x-linker** (K333-P351) in blue. Residues that could not be assigned as a result of ambiguous or missing triple resonance data are highlighted in the primary sequence in red. Residues not included in the secondary structure predictions due to their  $\phi$  and  $\psi$  angles mapping to more than two clusters are shown underlined.

## 4.4 Discussion

### 4.4.1 Effect of Redox State on W111F *abb'x*

$^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of either fully reduced or oxidised W111F **abb'x** were collected as outlined in 4.2.2. Spectrum differences were pronounced, with the oxidised spectra displaying far fewer peaks with an overall poorer resolution. The majority of peaks visible in oxidised protein spectra were located within the narrow random coil chemical shift range (8.0-8.5 ppm), with dispersed resonances that are indicative of structure in the reduced protein appearing at a much lower intensity in the oxidised spectra. The two spectra shown in Figure 4.6 suggest that in its oxidised state W111F **abb'x** is subject to significant line broadening, such that many residues do not appear in the spectrum. Much like other hPDI fragments, the reduced form of the protein gives rise to well-resolved peaks and suggests more conformational stability under the NMR conditions used. Overall these spectra confirm that the reduced form of W111F **abb'x** protein is better suited for assignment by NMR triple resonance experiments, with the protein showing a greater resolution and peak number when compared to its oxidised counterpart.

It is published that hPDI **abb'xa'** undergoes significant structural change upon reduction/oxidation, with the reduced form of protein adopting a more compact conformation (Wang et al., 2013). Wang et al. determined that it is the **b'xa'** region of **abb'xa'** that exhibits greatest structural change upon reduction/oxidation. This is in direct contrast to yPDI, where the **bb'** region is rigid, and the **a** domain is most flexible (Tian et al., 2008). Figure 4.19 shows minimal chemical shift perturbations of reduced and oxidised W111F **abb'x** protein, and highlights shifting residues of varying magnitude throughout the entire protein. This work indicates that in solution all domains of W111F **abb'x** are perturbed by changing redox state, however the majority of shifting residues comprise the redox active **a** domain. Minimal chemical shift perturbation mapping can underestimate the true extent that resonances move, as it maps shifts of assigned residues to the next nearest resonance. Hence large shifts can be underestimated if the true chemical shift change is of considerable magnitude or in a crowded region of the spectra.

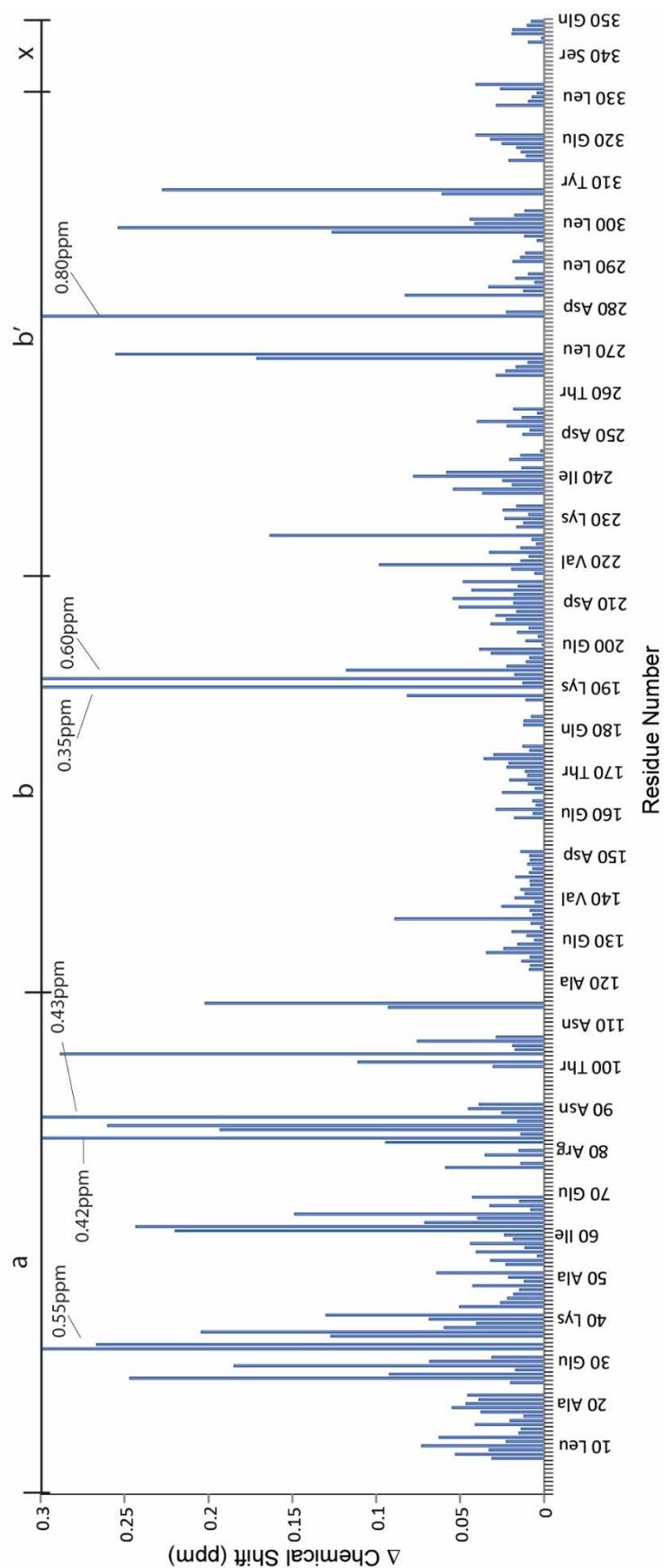


Figure 4.19: Minimal chemical shift perturbation map of reduced and oxidised W111F **abb'***x* protein

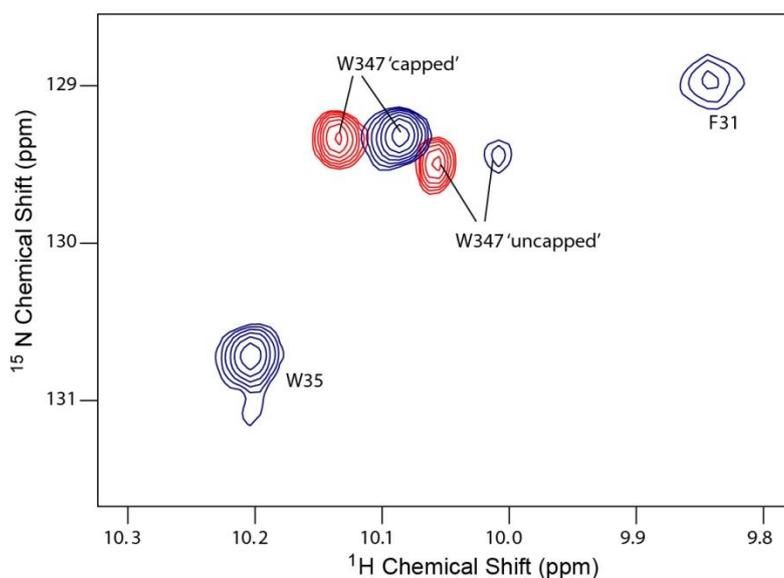
In agreement with Wang et al. findings are the NMR solution assignments discussed next in section 4.4.2. The **b'**x portion of W111F **abb'**x presented the greatest number of unassigned resonances- 48% of its resonances were unassignable (48% of **b'** and 44% of the **x-linker**), compared to an average of 23% of **ab** resonances (23% in **a** and 22% in **b** domain).

NMR spectroscopy shows an 'average' of all possible conformations a protein adopts, meaning that a residue can exist as a single peak, or if in slow exchange can appear as multiple peaks, that may broaden out or disappear as the exchange rate increases, preventing their correct assignment in solution. It is entirely possible that the unassigned resonances of the **b'** domain and **x-linker** region exhibit the highest shifts upon change in redox state.

#### 4.4.2 Sequential Backbone Assignment of Reduced W111F **abb'**x

NMR backbone assignments of W111F **abb'**x were completed to assign the  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra to 68%, as shown in section 4.3.2. HNCA and HN(CO)CA triple resonance experiments were used to correlate  $^1\text{H}$  and  $^{15}\text{N}$  with the  $\text{C}_\alpha$  nuclei of the *i* and *i*-1 residues. The assignments correspond to 77% of **a** domain, 78% of **b** domain, 52% of **b'** domain and 56% of the **x-linker** region (see Appendix 4.1 for assignment tables). Given the size of W111F **abb'**x protein (40.5 kDa), predictably there were several cases of ambiguity in the triple resonance data, leading to several unassignable resonances.

Furthermore, triple resonance assignment of the backbone also confirms that W347 indole is present in two distinct chemical environments, as seen in spectra of the **b'**x fragment (Byrne et al., 2009). This suggests the **x-linker** of W111F **abb'**x also exhibits a capped and uncapped conformation (see Figure 4.20).



**Figure 4.20: Overlaid  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of *b*'x (red) and W111F *abb*'x (blue) tryptophan indole NH region.**

Figure 4.18 shows the primary amino acid sequence of W111F **abb**'x, with the residues that could not be assigned as either a result of ambiguous or missing triple resonance data highlighted in red. Additional assignments could have been made from the ambiguous data, if additional  $\text{C}_\beta$  triple resonances had been collected, to remove such ambiguity. This data was not collected however, due to the size of W111F **abb**'x; as protein increase in molecular weight, their associated correlation times in solution become longer, which has the effect of increasing the nuclear transverse relaxation rate. This rate is proportional to the NMR resonance line width and so this series of proportionalities explains increasing line broadening with molecular weight.  $\text{C}_\beta$  resonances would likely broaden out to such an extent that many would be absent from the triple resonance data collected. This limitation can be resolved by running triple resonance experiments using a triple-labelled ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$ -labelled) protein sample. Deuteration using  $\text{D}_2\text{O}$  in place of  $\text{H}_2\text{O}$  removes around 70-80% of sidechain  $^1\text{H}$ , improving the relaxation rate and thus the spectral quality (LeMaster, 1990).

It was apparent from the triple resonance data that there are several regions of W111F **abb**'x that were unassignable because of missing resonances in the HNCA/HN(CO)CA

experiment pair, most notably in the **b'**x portion of the protein. Missing stretches included K259-F265 and I272-T279 of **b'** domain, and P336-P344 of the **x-linker**. Absence of these residues are likely to be due to the conformational exchange exhibited when in solution, as previously documented at Kent when assigning the **b'**x fragment (Byrne et al., 2009b).

Due to the high cost of generating a deuterated sample, together with the conformational exchange seen of the **b'**x fragment, it was decided not to proceed with deuteration, and that the achieved backbone assignments would be sufficient for planned NMR studies.

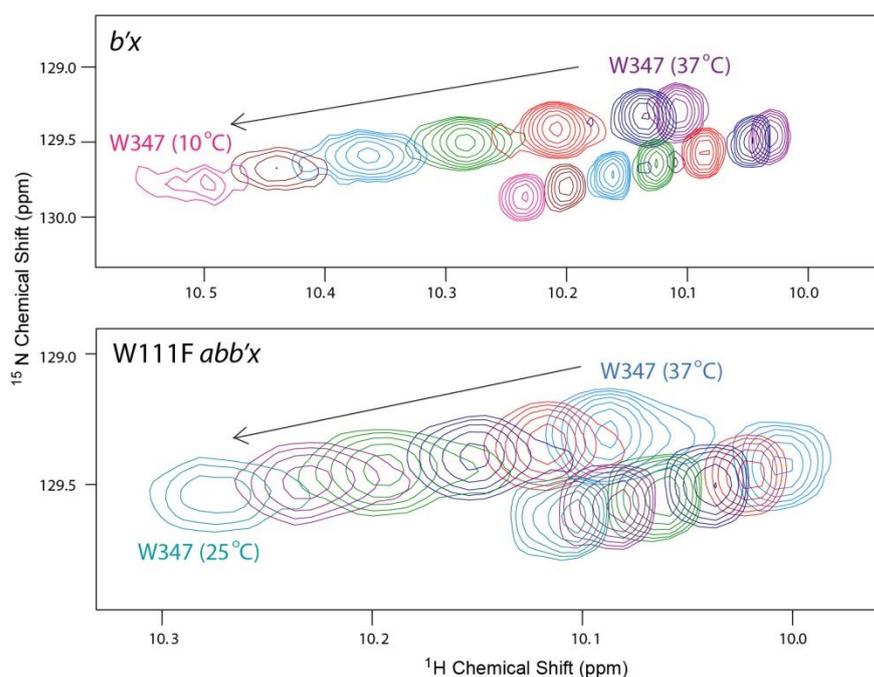
#### 4.4.3 Effect of Temperature on W111F *abb'*x

Previous studies of **b'**x and other hPDI domains within our laboratory show small hPDI fragments give fair spectra at 25°C, with the **b'**x fragment previously assigned at this temperature (Byrne et al., 2009b). Spectra of W111F **abb'**x collected at 25°C were poor, with considerable peak broadening and reduced signal intensity, due to both the size of the fragment and the increased viscosity of water at lower temperatures. In light of this, all subsequent spectra and triple resonance data were instead collected at 37°C. Running experiments at an increased temperature can narrow peak lines and boost signal, whilst also more closely mimicking physiological conditions. Additionally, studies by Wang et al. on WT **abb'**x showed the fragment displays lower chaperone activity at 20°C compared to 40°C (Wang et al., 2010), something not shown by full-length hPDI. They suggest this is because at low temperatures **abb'**x exists as a predominantly capped form, with ligand having to compete more strongly with the **x-linker** to bind- an event that is thought to be limited by the presence of **a'** domain in full-length hPDI. This is interesting to note, because spectra of W111F **abb'**x collected for this thesis at 25°C still show defined 'capped' and 'uncapped' resonances of the **x-linker** residue W347, as did spectra of **b'**x collected as low as 10°C (shown in Figure 4.21).

Nevertheless, higher temperatures can shift the NMR timescale to show molecules in fast exchange, meaning conformational exchange rates will be increased. This is particularly significant for **b'**x and W111F **abb'**x, as it may promote the rate of capping and uncapping of the **x-linker** region to **b'** domain, to facilitate ligand binding.

The assignments for W111F **abb'**x carried out at 37°C were mapped onto the spectrum of the protein collected at 25°C for comparison with previous single domain assignments

collected at the lower temperature. 92% of the backbone assignments made at 37°C could be mapped onto the spectrum at 25°C, however this is only 63% of all possible assignable resonances. This limited the comparison of W111F **abb**'x at 25°C with previously collected individual domain data, where percentage assignment is much higher (typically >80%).



**Figure 4.21:** Overlaid  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC indole region **b**'x collected at 37, 35, 30, 25, 20, 15 and 10°C (**b**'x) and W111F **abb**'x collected at 37, 35, 32.5, 30, 27.5 and 25°C. The figure highlights the presence of capped and uncapped W347 forms at all temperatures.

#### 4.4.4 Sequential Backbone and Methyl Side-Chain Assignment of **b**'x

**Backbone Assignment:** As mentioned in 4.4.3, W111F **abb**'x gives best spectra at 37°C. Spectra of **b**'x collected at 37°C were sufficiently different to the previously assigned spectra within the laboratory (collected at pH 6.5 and 25°C), highlighting sensitive pH and temperature-mediated perturbations of several resonances. Therefore NMR backbone assignments of **b**'x were also completed to assign the  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra.

HNCACB and HN(CO)CACB triple resonance experiments were used to successfully assign the  $^{15}\text{N}, ^1\text{H}$ -HSQC spectra to 85%, by correlating  $^1\text{H}$  and  $^{15}\text{N}$  resonances with the  $\text{C}_\alpha$  and  $\text{C}_\beta$  nuclei of the  $i$  and  $i-1$  residues, as shown in section 4.3.2 (see Appendix 4.2 for assignment tables). This additional information improved assignment of the **b'** domain to 89% (compared to just 52% assignment of **b'** in W111F **abb'**x). In comparison to W111F **abb'**x, there were fewer ambiguous assignments, in part due to the collection of the additional  $\text{C}_\beta$  resonances, but also the smaller size of **b'**x meant fewer overlapping resonances. Figure 4.17 shows the amino acid sequence of **b'**x with the residues that could not be assigned, as either a result of ambiguous or missing triple resonance data, highlighted in purple.

A 56% assignment of the **x-linker** region of **b'**x directly supports W111F **abb'**x assignment of the **x-linker** (also to 56%), and previous published data of **b'**x (Byrne et al., 2009) with the largest unassigned stretch of the protein again being P336-E345 of the **x-linker** region. This was due to missing resonances in the triple resonance data, thought to be due to the high conformational activity the **x-linker** region displays in solution.

**Side-chain Assignment:**  $^{13}\text{C}$  methyl side-chains were assigned to more accurately probe direct ligand-binding events of the protein, by allowing detailed assignment of the hydrophobic ligand binding site. As mentioned previously, the  $^{15}\text{N}$  amide backbone assignments whilst being easier to collect are more limited to reporting on changes of the protein backbone, and are also more sensitive to induced structural changes the protein experiences than sidechain  $^{13}\text{C}$  nuclei are. This work was carried out on the smaller **b'**x fragment of hPDI; it is only 18.5 kDa in size, giving good resolution of individual peaks in a  $^{13}\text{C}, ^1\text{H}$ -HSQC spectra. In addition to this, both  $\text{C}_\alpha$  and  $\text{C}_\beta$  data were collected on **b'**x, and the backbone was assigned to a higher degree, allowing a more complete assignment of **b'**x sidechains than would have been achievable using the larger W111F **abb'**x fragment.

Of the 66 assignable methyl sidechains of all alanine, threonine, valine, leucine and isoleucine residues previously 'backbone' assigned in the  $^{15}\text{N}, ^1\text{H}$ -HSQC, 89% were assigned using the HCCH-TOCSY, by correlating all  $^1\text{H}$  resonances attached to each  $^{13}\text{C}$  nuclei, as explained in section 4.1.6 (see Appendix 4.3 for assignment tables). The majority of the 11% unassigned methyl resonances comprised the published ligand binding site.

#### 4.4.5 Secondary Structure Predictions of *b'x* and W111F *abb'x* by DANGLE

The DANGLE secondary structure predictions of **b'x** and W111F **abb'x** both agree with one another, and also closely match the secondary structure seen in the crystal structure of **abb'xa'** published by Wang et al. (Wang et al., 2013).

Prediction for the W111F **abb'x** fragment, where overall backbone assignment was lower, yielded several stretches of unassigned residues, leaving DANGLE unable to accurately predict these regions (see Figure 4.18). There were several small differences between the structure predictions of this fragment and the **abb'xa'** crystal structure. These include the addition of three small  $\beta$ -strands and an  $\alpha$ -helix that are only present in the crystal structure: S71-Q76 and K97-Y99 in **a** domain, T121-L123 at the start of **b** domain and K309-L310 in **b'** domain. These predictions were simply not possible using the triple resonance data, as stretches D72-Y77, A94-T100 and P118-P124 of W111F **abb'x** were unassignable. Residues K309-L310 were flanked on either side by unassigned residues, therefore it is likely that DANGLE was unable to correctly predict secondary structure of these residues. Also in contrast to the **abb'xa'** crystal structure was the **x-linker** region of both **b'x** and W111F **abb'x**, which were not assigned any secondary structure (see Figures 4.17 and 4.18). In the **abb'xa'** crystal structure the **x-linker** is shown to contain a small  $\beta$ -strand composed of residues P336-L338, and an  $\alpha$ -helix of residues E345-W347. The I272A **b'x** 'capping mutant' crystal structure also reports the same secondary structure of **x-linker** (Nguyen et al., 2008). A large region of **x-linker** resonances were unassignable by triple resonance due to the conformational activity **x-linker** exhibits in solution, so it is also possible there was not enough information for DANGLE to correctly predict secondary structure for this region.

In full-length hPDI, it is proposed the **x-linker** adopts a secondary conformation, with recent structural work of the **bb'xa'** fragment of hPDI demonstrating that in reduced crystal structures the **x-linker** is found tightly packed into the **a'** domain (Wang et al., 2012). This packing could also explain the differences seen, as the **a'** domain is not present in either the **b'x** or W111F **abb'x** fragments for this phenomena to occur. It has also been shown that presence of the **a'** domain in **abb'xa'** hPDI somewhat limits the ability of the **x-linker** to move between a capped and uncapped conformation (Wang et al., 2013)- this could also help to explain the differences seen in the crystal and solution predicted

structures. It is interesting to note however that the same paper shows the **x-linker** region undergoes a 45° rotation in reduced/oxidised **abb'xa'**, demonstrating that it remains highly flexible even in the presence of the **a'** domain.

## CHAPTER 5

# Ligand Binding Studies of **b'**x and W111F **abb'**x by NMR Spectroscopy

### 5.1 Introduction

Given PDI's role as a critical ER-resident protein folding catalyst, it is hardly surprising the majority of PDI research tends to focus on the protein's oxidoreductase or isomerase activity. Yet PDI is also capable of binding small peptide ligands, at a primary binding pocket in its **b'** domain (Klappa et al., 1998). This pocket is also crucial for the binding of larger, unfolded or mis-folded proteins, in addition to Ero1 $\alpha$ , the greatest contributor to hPDI re-oxidation *in vivo* (Tavender and Bulleid, 2010). Hence this binding site makes for an attractive target for which to design therapeutics, as any peptide that is able to compete with physiological ligands to 'block' the **b'** binding pocket could reduce both the chaperone and isomerase activities of hPDI.

The binding site of hPDI was mapped using NMR spectroscopy in 2009 (Byrne et al., 2009), yet ligand specificity of hPDI is still poorly understood and to date no structural models exist for the protein bound to a target peptide ligand. Specificity of the PDI family member PDIp has been studied more comprehensively, and it was shown either a Trp or Tyr residue in the peptide sequence is sufficient for binding recognition (Ruddock et al., 2000). Substrate recognition by hPDI is anticipated to be somewhat more complex.

This Chapter introduces the use of NMR spectroscopy to first quantitate the binding affinity of hPDI to its historic model peptide  $\Delta$ -som, followed by detailed analysis using fluorinated residues and specific residue substitutions to probe the binding specificity. There are several advantages of using NMR to study ligand binding over other biomolecular techniques like isothermal titration calorimetry, surface plasmon resonance or Förster resonance energy transfer. Principally NMR allows the binding interaction to be observed at an atomic level, meaning that as well as direct ligand binding affinities, highly detailed structural information about the binding interaction can also be generated. With the collection of various other NMR spectra, this information can even be used to generate structural models of the protein-ligand complex.

### 5.1.1 Ligand Binding by $^{15}\text{N}$ , $^1\text{H}$ -HSQC spectra

A well-documented method of measuring ligand binding by NMR spectroscopy is via chemical shift perturbation, a technique comprehensively reviewed by both Fielding and Williamson (Fielding, 2003; Williamson, 2013). The approach requires the collection of a set of spectra of the protein sample with varying concentrations of ligand in each spectra. The overlaid spectra can inform on the binding event and highlight residues that are involved in the binding interaction.

Ligand binding by NMR spectroscopy is traditionally measured using chemical shift perturbation of resonances in overlaid  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra. Results presented in Chapter 4 have already highlighted how sensitive the protein backbone is to changes in redox state, and shifts of a similar magnitude can also be observed in the protein upon addition of a binding ligand.

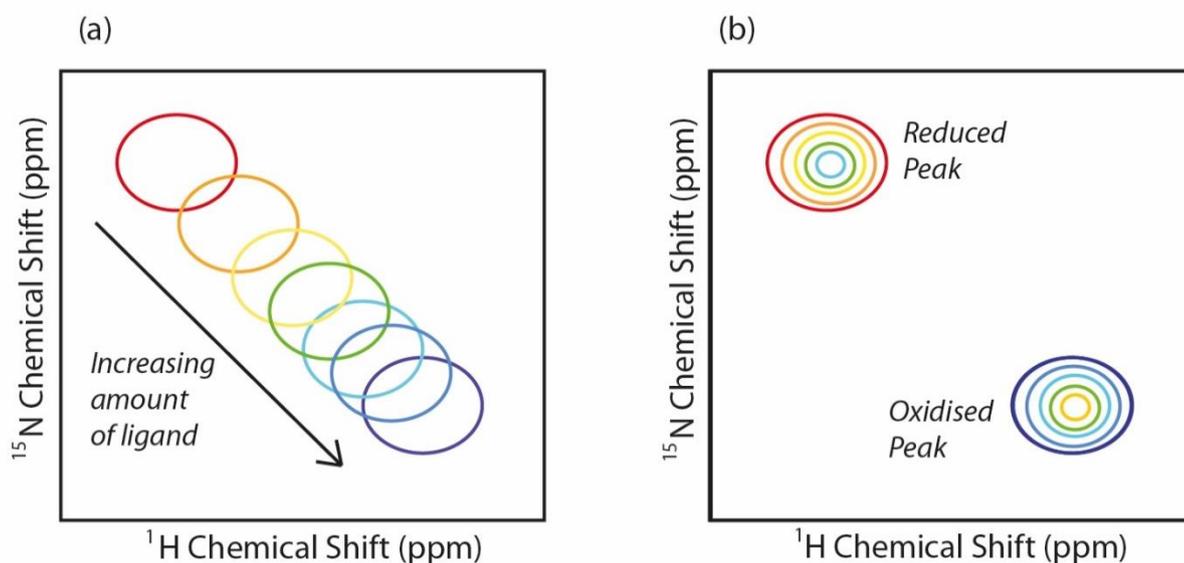
This particular method uses the controlled titration of a ligand into a constant amount of protein, to observe how the protein shifts and changes upon binding. The method is traditionally protein-observed, therefore only the protein is isotopically enriched, so the resonances in the spectra are not complicated by the addition of a ligand. The perturbation of each protein resonance can be measured throughout the titration series, weighted as previously outlined in Equation 4.1, and plot graphically against ligand concentration to equate a dissociation constant ( $K_d$ ). Equation 5.1 defines the relationship between protein and ligand in free and bound complex states (of a single binding site model):-

$$5.1 \text{ Dissociation Constant } (K_d) = \frac{[P][L]}{[PL]}$$

Where [P] is concentration of free protein, [L] is concentration of free ligand, and [PL] concentration of the protein-ligand complex.

$K_d$  is a concentration of ligand at which half the protein ligand binding sites become occupied, so the smaller the value, the tighter the binding affinity. Determination of  $K_d$  by NMR typically works best at the 'weak' end of biological affinity (mM- $\mu$ M range) (Vaynberg and Qin, 2006), and is therefore well-suited to probe ligand affinity of hPDI, where it is estimated the binding affinity to  $\Delta$ -som is 0.1-1mM (Byrne et al., 2009b).

Exchange rates in NMR can be in fast or slow timescales. Ligand binding is easiest to measure when the exchange rate between free and ligand bound states is fast, as is the case for hPDI binding of  $\Delta$ -som. Signals in the overlaid spectra will gradually move from their signal in the free spectra, to their signal in the bound spectra, and intermediate spectra will report a 'weighted average' trajectory of bound and free shifts (as shown by Figure 5.1 (a)).



**Figure 5.1:**  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC NMR peak shape with fast (a) and slow (b) exchange rates. In fast exchange (e.g. ligand binding), the peaks move smoothly from free to bound states upon addition of increasing amounts of ligand. Figure adapted from (Williamson, 2013). In slow exchange (e.g. redox change), the reduced peak decreases in intensity as the oxidised peak increases in intensity, upon addition of an increased oxidising:reducing agent ratio.

During the titration series the total concentrations of [L] and [P] are known, each the sum of free [L] or [P] and bound complex [PL] forms. If the binding event is in fast exchange we know that the observed chemical shift perturbation is an average of all the shifts in free and complex bound states, and we assume that the fractions of free and bound are equal to 1. Together these assumptions are used to calculate a dissociation constant ( $K_d$ ), by fitting experimental data to Equation 5.2, taken from (Morton et al., 1996) :-

$$5.2 \quad \Delta = \Delta_0 \frac{(K_d + [L] + [P]) - \sqrt{((K_d + [L] + [P])^2 - 4 [P][L])}}{2[P]}$$

Where  $\Delta$  is the observed chemical shift perturbation in Hz,  $\Delta_0$  the maximum chemical shift perturbation in Hz, [P] and [L] the protein and ligand concentrations respectively. Williamson has also explained the mathematical generation of this equation, and the derivation can be found in Appendix 5.1 (Williamson, 2013).

### 5.1.2 Ligand Binding by $^{13}\text{C}, ^1\text{H}$ -HSQC spectra

Ligand binding was also probed using overlaid methyl regions of assigned  $^{13}\text{C}, ^1\text{H}$ -HSQCs. It was carried out with the intention that any observed shifts might report more sensitively upon the ligand-binding event (as  $^{13}\text{C}$  nuclei are thought to be less influenced by structural change upon binding, and  $^{15}\text{N}$  shifts can report on both direct ligand binding or induced conformational change upon binding). Few methyl resonances of the ligand binding site could be assigned, but surprisingly the data still reported accurately throughout the protein on what is likely to be a ligand-induced conformational change. Hence,  $^{13}\text{C}, ^1\text{H}$ -HSQCs were not a suitable method to use to observe ligand binding, due to limited assignable reporters within the binding site.

### 5.1.3 3D $^{15}\text{N}$ NOESY-HSQC of *b'x* and W111F *abb'x*

The  $^{15}\text{N}$  NOESY-HSQC is a through-space experiment that reports inter and/or intramolecular interactions in 3D, collected by combining a  $^{15}\text{N}$ -NOESY with a  $^{15}\text{N}$ -HSQC step (Marion et al., 1989; Zwieterweg and Fesik, 1989). Magnetisation is transferred between all hydrogens using the Nuclear Overhauser Effect (NOE), to neighbouring  $^{15}\text{N}$  nuclei, and then back again for detection. Using an assigned  $^{15}\text{N}, ^1\text{H}$ -HSQC of the protein,

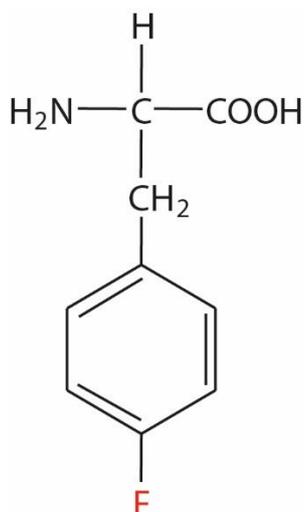
it is possible to look at the NOE resonances of each NH group from all other nearby hydrogen atoms, typically within a 5-6Å distance, where intensity of the peak will be dependent on the distance between each atom.

Spectra were run in anticipation of observing extra NOE peaks (intermolecular peaks) in the protein spectra upon addition of the binding ligand  $\Delta$ -som. If extra peaks are observed, then isotope edited NOESY spectra can be used to distinguish between the intra- and intermolecular contacts (Fesik et al., 1988) to generate structural restraints, that can be used with docking software to generate structural ligand-bound models (Constantine, Davis et al. 2006). Data was collected using **b'x** and W111F **abb'x** at several temperatures and field strengths, but insufficient intermolecular NOEs were observed between protein and binding ligand to allow further structure determination.

#### 5.1.4 Synthetic Peptide Fluorination

Due to the various difficulties encountered in attempts to produce recombinant isotopically labelled  $\Delta$ -som (see Chapter 2), synthetically fluorinated forms of the peptide were ordered from Peptide Synthetics (Fareham). Synthetic fluorination allowed us to circumvent recombinant peptide problems whilst utilising Kent's highly sensitive  $^{19}\text{F}$  cryoprobe, to view binding to hPDI from a unique ligand-observed perspective.  $^{19}\text{F}$  fluorine is close to 100% natural abundance, and is also active by NMR spectroscopy (Gerig, 1994). Additionally it has the benefit of having extremely low uptake in nature, meaning there will be no contaminating protein background signal of NMR samples (David and B. Harper, 1999).

It is well documented that other molecular chaperones rely upon hydrophobic residues for ligand recognition (Hammond and Helenius, 1995), and it was expected hPDI recognises its ligands by a similar mechanism, as previous work has shown ligand binding is a hydrophobically driven event (Klappa et al., 1997). Therefore we have targeted the three hydrophobic phenylalanine residues of  $\Delta$ -som for fluorination. This chapter reports the incorporation of 4-fluoro-phenylalanine (4-F-Phe) (see Figure 5.2) into  $\Delta$ -som peptide, as an additional method to monitor peptide ligand binding to hPDI.



**Figure 5.2: Chemical structure of 4-Fluoro Phenylalanine.** The fluorinated amino acid is made by substituting the C-4 hydrogen for a fluorine atom (highlighted in red).

The introduction of fluorine into phenylalanine residues should have limited effect when compared to substitution of an entire residue, but successive fluorination, whilst initially stabilising hydrophobic interactions, can also destabilise them (Salwiczek et al., 2012). Therefore, careful attempts were made to observe any effects fluorination conferred to the peptides, and a comprehensive study of fluorinated and non-fluorinated peptides was undertaken. Results inferred by fluorinated peptides were used to create non-fluorinated alanine substitutions to validate these findings.

### 5.1.5 Ligand Binding by $^{19}\text{F}$ 1D spectra

Selective fluorination of the peptide allows ligand binding of hPDI to be viewed from the direct ligand-observed perspective. Data was collected by controlled titration of  $^{15}\text{N}$  **b'x** protein into a constant amount of fluorinated  $\Delta$ -som peptide, to observe how the peptide experiences the binding event.  $^{19}\text{F}$  data was treated in a similar manner to  $^{15}\text{N}, ^1\text{H}$ -HSQC shifts, but instead uses peak maxima shift in Hz instead of 2D weighted chemical shifts. These shifts were plotted against protein concentration and Equation 5.2 was used to generate a dissociation constant. This method was exclusively ligand-observed, but the same experiments were repeated by  $^{15}\text{N}, ^1\text{H}$ -HSQC spectra (as outlined in 5.1.1), to gain confidence in the binding affinities generated by  $^{19}\text{F}$  1D spectra.

## 5.2 Materials and Methods

### 5.2.1 NMR Sample Preparation

Monomeric protein isotopically labelled and purified as outlined in Chapter 3 was buffer exchanged using a Vivaspin 20 Centrifugal Filter (10 kDa MWCO, Sartorius) into NMR buffer (20 mM sodium phosphate, 50 mM NaCl, pH 7.0) and then concentrated down into stocks to use as required (determined by  $A_{280}$  nm absorbance).

**Ligand Binding:** Protein stocks were made to 0.3 mM or 0.45 mM in NMR buffer, and peptide stocks were made to 20 mM or 10 mM in 100% DMSO- $d_6$  (several peptides showed poor solubility in aqueous buffers). Table 5.1 shows the peptides used for ligand binding experiments.

**Table 5.1:** List of substituted peptides used to probe peptide ligand binding of hPDI.

Peptide	Sequence
$\Delta$ -som	AGSKNFFWKTFTSS
<i>F1</i> $\Delta$ -som	AGSKNF(4-F-Phe)FWKTFTSS
<i>F2</i> $\Delta$ -som	AGSKNFF(4-F-Phe)WKTFTSS
<i>F3</i> $\Delta$ -som	AGSKNFFWKT(4-F)TSS
<i>F1,2,3</i> $\Delta$ -som	AGSKNF(4-F-Phe)F(4-F-Phe)WKT(4-F-Phe)TSS
<b>Phe 6</b> →A $\Delta$ -som	AGSKN <b>A</b> FWKTFTSS
<b>Phe 7</b> →A $\Delta$ -som	AGSKNF <b>A</b> WKTFTSS
<b>Phe 11</b> →A $\Delta$ -som	AGSKNFFWKT <b>A</b> TSS
<b>Phe 6,7</b> →A $\Delta$ -som	AGSKN <b>AA</b> WKTFTSS
<b>Phe 6,11</b> →A $\Delta$ -som	AGSKN <b>A</b> FWKT <b>A</b> TSS
<b>Phe 7,11</b> →A $\Delta$ -som	AGSKNF <b>A</b> WKT <b>A</b> TSS
<b>Phe 6,7,11</b> →A $\Delta$ -som	AGSKN <b>AA</b> WKT <b>A</b> TSS

Fluorinated Phe residues are shown by 4-F-Phe (4-fluorophenylalanine). Phe-Ala substitutions are highlighted in red.

**b'x** + ligand ‘protein-observed’ samples were made using a constant 0.25 mM protein, varying concentrations of peptide, 5% total DMSO-d<sub>6</sub>, 10% D<sub>2</sub>O and 0.05% sodium azide, made to a total volume of 350 μL using NMR buffer (see Table 5.2 for typical titration series sample preparations). Samples of 1:0, 1:0.25, 1:0.5 1:1, 1:1.5, 1:2 and 1:2.5 protein: ligand ratios were collected.

**Table 5.2:** Typical ‘protein-observed’ titration series using 0.25 mM <sup>15</sup>N *b'x* and varying concentrations of Δ-som peptide.

Δ-som [mM]	0	0.0625	0.125	0.25	0.375	0.5	0.625
<b>Protein (0.3 mM stock)</b>	291.7	291.7	291.7	291.7	291.7	291.7	291.7
<b>Peptide (20 mM stock)</b>	-	1.09	2.19	4.38	6.56	8.75	10.94
<b>DMSO-d<sub>6</sub></b>	17.5	16.41	15.31	13.12	10.94	8.75	6.56
<b>D<sub>2</sub>O</b>	35	35	35	35	35	35	35
<b>Sodium azide (5% stock)</b>	3.5	3.5	3.5	3.5	3.5	3.5	3.5
<b>NMR Buffer (de-gassed)</b>	2.3	2.3	2.3	2.3	2.3	2.3	2.3
<b>TOTAL:</b>	350	350	350	350	350	350	350

All volumes are in μL.

W111F **abb'x** ‘protein-observed’ samples were made as for **b'x**, with the addition of a 5x molar concentration of DTT<sup>red</sup> or GSSG to reduce or oxidise the protein respectively. Due to the instability of oxidised protein at high concentrations, oxidised spectra were collected using 0.2 mM protein, and molar ratios were kept constant (so the titration series finished at a 2.5 molar excess peptide).

**b'x** + ligand ‘ligand-observed’ samples were made using a constant 0.15 mM peptide, varying concentrations of **b'x** protein, a 5% total DMSO-d<sub>6</sub>, 10% D<sub>2</sub>O and 0.05% sodium

azide, made to a total volume of 350  $\mu$ L using NMR buffer. Samples of 1:0, 1:0.25, 1:0.5 1:1, 1:1.5, 1:2 and 1:2.5 ligand:protein ratio were collected.

### 5.2.2 NMR Data Acquisition and Processing

NMR data was acquired at the University of Kent using a 4-channel, 5-amplifier Bruker Avance III 14.1T (600MHz  $^1$ H) NMR spectrometer equipped with a 5 mm QCI-F cryoprobe, unless otherwise stated. All experiments were carried out at 37°C, pH 7.0, unless otherwise stated.  $^1$ H chemical shift referencing was carried out using the position of the  $^1$ H<sub>2</sub>O resonance at the given temperature.  $^{15}$ N and  $^{13}$ C chemical shifts were referenced using a spectrometer-based macro, and  $^{19}$ F chemical shifts were referenced using the position of TFA (-76.55 ppm).

All 1D data was processed using Bruker Topspin software, and 2D or 3D NMR data was processed using NMRPipe software (Delaglio et al., 1995).

#### 5.2.2.1 $^{15}$ N, $^1$ H-HSQC Experiments of *b'x* and W111F *abb'x*

$^{15}$ N, $^1$ H-HSQC spectra were acquired of each protein fragment with varying amounts of peptide ligand or oxidising: reducing agent as previously outlined in 4.2.2.1.

#### 5.2.2.2 $^{19}$ F 1D Experiments of *b'x* and W111F *abb'x*

$^{19}$ F 1D spectra were acquired of each peptide ligand with varying amounts of **b'x** over 60 minutes with 256 points in the F1 ( $^{19}$ F) dimension. The italicised peptides in Table 5.1 show the peptides that  $^{19}$ F data was collected for.

#### 5.2.2.3 $^{13}$ C, $^1$ H-HSQC Experiments of *b'x*

$^{13}$ C, $^1$ H-HSQC spectra were acquired of 0.3 mM **b'x** with and without 0.75 mM  $\Delta$ -som as previously outlined in 4.2.2.2.

#### 5.2.2.4 $^{15}$ N NOESY-HSQC Experiments of *b'x* and W111F *abb'x*

$^{15}$ N NOESY-HSQC spectra of 0.4 mM  $^{15}$ N **b'x** with and without 2.5x molar excess  $\Delta$ -som were collected at 37°C at Mill Hill using an 800 MHz spectrometer. Additional spectra were collected at 10°C at the University of Kent. Data at Kent was acquired with 2048 points in the F3 dimension ( $^1$ H), 220 points in the F2 dimension ( $^1$ H) and 32 points in the F1 dimension ( $^{15}$ N).

Additionally,  $^{15}\text{N}$  NOESY-HSQC spectra of 0.4 mM W111F  $^{15}\text{N}$  **abb'x** with and without 2.5x molar excess  $\Delta$ -som were collected at 10°C at Mill Hill using a 950 MHz spectrometer.

### 5.2.3 NMR Data Analysis

All subsequent data analysis of 1D spectra was carried out using Bruker Topspin software, and all 2D and 3D spectra analysis was carried out using CcpNmr Analysis software Version 2 (Vranken et al., 2005).

#### 5.2.3.1 $^{19}\text{F}$ Peptide Assignment

Singly fluorinated peptides each yielded a single peak that tracked throughout the  $^{19}\text{F}$  spectra upon addition of **b'x** protein. It was hoped that the peak trajectories would overlay well with their equivalent peaks in spectra of F1,2,3  $\Delta$ -som, but this was not the case.  $^{13}\text{C}$ ,  $^{19}\text{F}$  2D spectra collected at 37°C were blank, meaning unambiguous assignment of F1,2,3  $\Delta$ -som could not be carried out. Instead, the tight  $^{19}\text{F}$  calculated  $K_d$  fits of the single fluorinated peptides were used to identify the fluorine peaks in the F1,2,3  $\Delta$ -som spectra (explained further in 5.3.6).

#### 5.2.3.2 Chemical Shift Perturbation Mapping

The backbone assignments of **b'x** and W111F **abb'x** described in Chapter 4 were used to assign all subsequent  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQCs. Peak assignments were automatically copied across onto overlaid spectra by the software upon increasing addition of peptide ligand. Peaks were then manually moved to their new peak maxima. Whilst time consuming, manual fits were found to give better graphical fits than the automatic peak picking software available in Analysis.

Chemical shift perturbations were then weighted if required for further analysis:-

**Using  $^{15}\text{N}$  data:** Chemical shift perturbations measured in  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQCs were calculated using equation 4.1, to weight observed  $^{15}\text{N}$  shifts relative to their  $^1\text{H}$  shifts.

**Using  $^{19}\text{F}$  data:** Chemical shifts monitored via  $^{19}\text{F}$  spectra were collected in 1D, therefore did not need weighting. Each fluorine peak maxima was measured and directly plotted against protein concentration.

In  $^{19}\text{F}$  spectra of F1,2,3  $\Delta$ -som, the F1 and F2 peaks are reporting a sufficiently similar chemical environment that they appear as a single peak. The peak resolves into two separate peaks upon a 2:1 addition of **b'x** protein. Attempts to integrate the peak into two separate peaks were not successful, therefore data points prior to this concentration were measured by using the peak maxima of the overlapped peak. This does not appear to have adversely affected graphical fits, so it is likely that the peaks are closely on top of one another at lower concentrations of protein.

**Using  $^{13}\text{C}$  data:** Upon attempting to weight  $^{13}\text{C}$  methyl shifts relative to their respective  $^1\text{H}$  shifts, nothing could be found published within the literature for methyl shifts specifically. Therefore chemical shift perturbations were calculated similarly to  $^{15}\text{N},^1\text{H}$ -HSQC shifts (Williamson et al., 1997), to allow for the difference in spectral width of  $^{13}\text{C}$  shifts relative to  $^1\text{H}$  shifts, using equation 5.3:-

$$5.3 \text{ } ^{13}\text{C methyl chemical shift perturbation} = \sqrt{(\Delta^1\text{H})^2 + \left(\frac{1}{8.1}\Delta^{13}\text{C}\right)^2}$$

Where  $\Delta^1\text{H}$  and  $\Delta^{13}\text{C}$  are the chemical shift perturbations (ppm) of a peak in the  $^1\text{H}$  and  $^{13}\text{C}$  dimensions respectively.  $^{13}\text{C}$  methyl shifts were weighted by 1/8.1 their actual size, to compensate for the chemical shift range they report over when compared to proton chemical shifts (observed chemical shift range of  $^{13}\text{C}$  was 17 ppm compared to observed  $^1\text{H}$  chemical shift range of 2.1 ppm).

## 5.3 Results

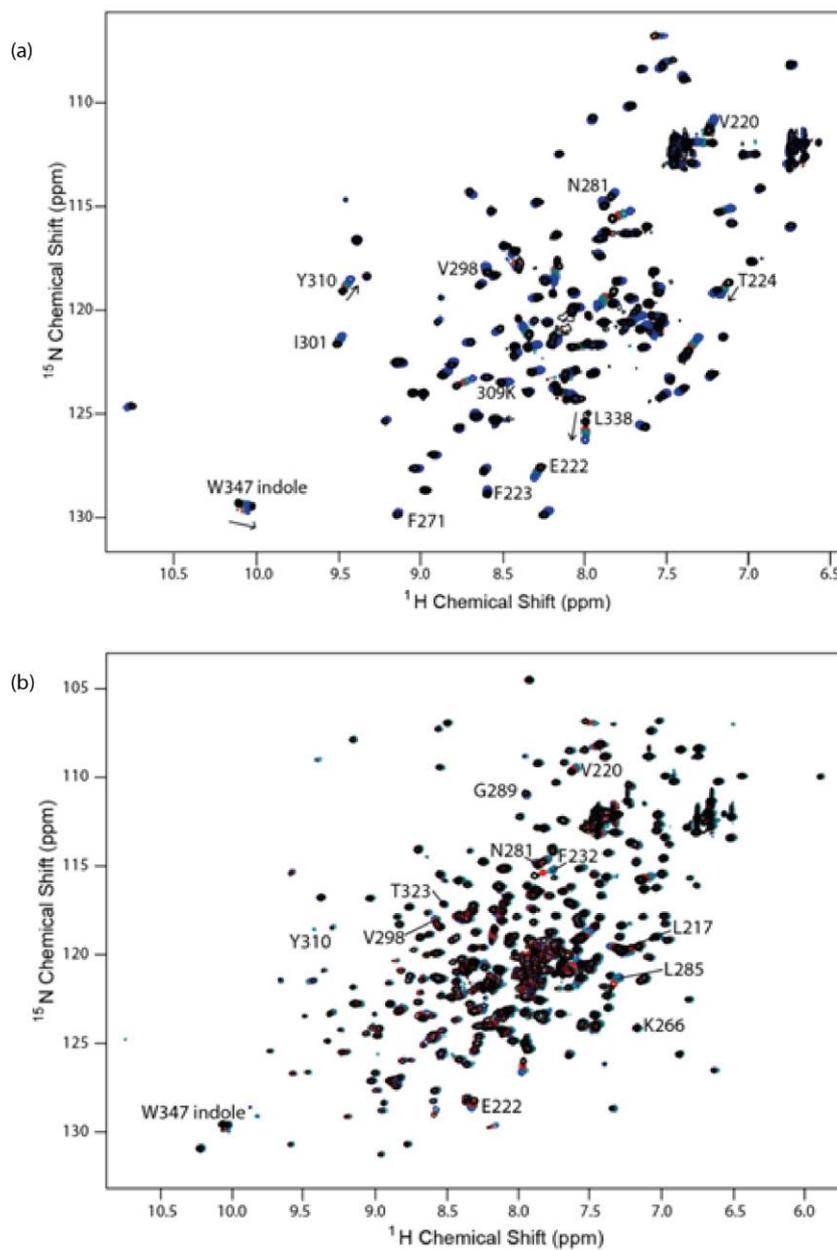
*Upon addition of ligand there were typically 44 shifting residues from the **b'x**  $^{15}\text{N},^1\text{H}$ -HSQCs and 33 shifting residues from the reduced W111F **abb'x**  $^{15}\text{N},^1\text{H}$ -HSQCs that fit graphically to give the stated dissociation constants of each section. For each results section a representative sample of graphs are shown- additional raw data can be found in the Appendix as stated in each section.*

### 5.3.1 Effect of Neighbouring Domains on Peptide Ligand Binding

Dissociation constants were calculated from practical data by careful titration of increasing concentrations of  $\Delta$ -som ligand into either 0.25 mM **b'x** or W111F **abb'x** protein. 7 data points were collected for each protein, corresponding to a titration endpoint of 2.5:1 molar ratio of ligand:protein. The overlaid spectra of **b'x** shown in Figure 5.3 (a) and of W111F **abb'x** in (b) demonstrate the observed shifts upon addition of increasing  $\Delta$ -som ligand, and in each case show the W347 indole resonance shifting from a 'capped' to 'uncapped' conformation; ligand is competing to displace the **x-linker** region from the **b'** ligand binding site.

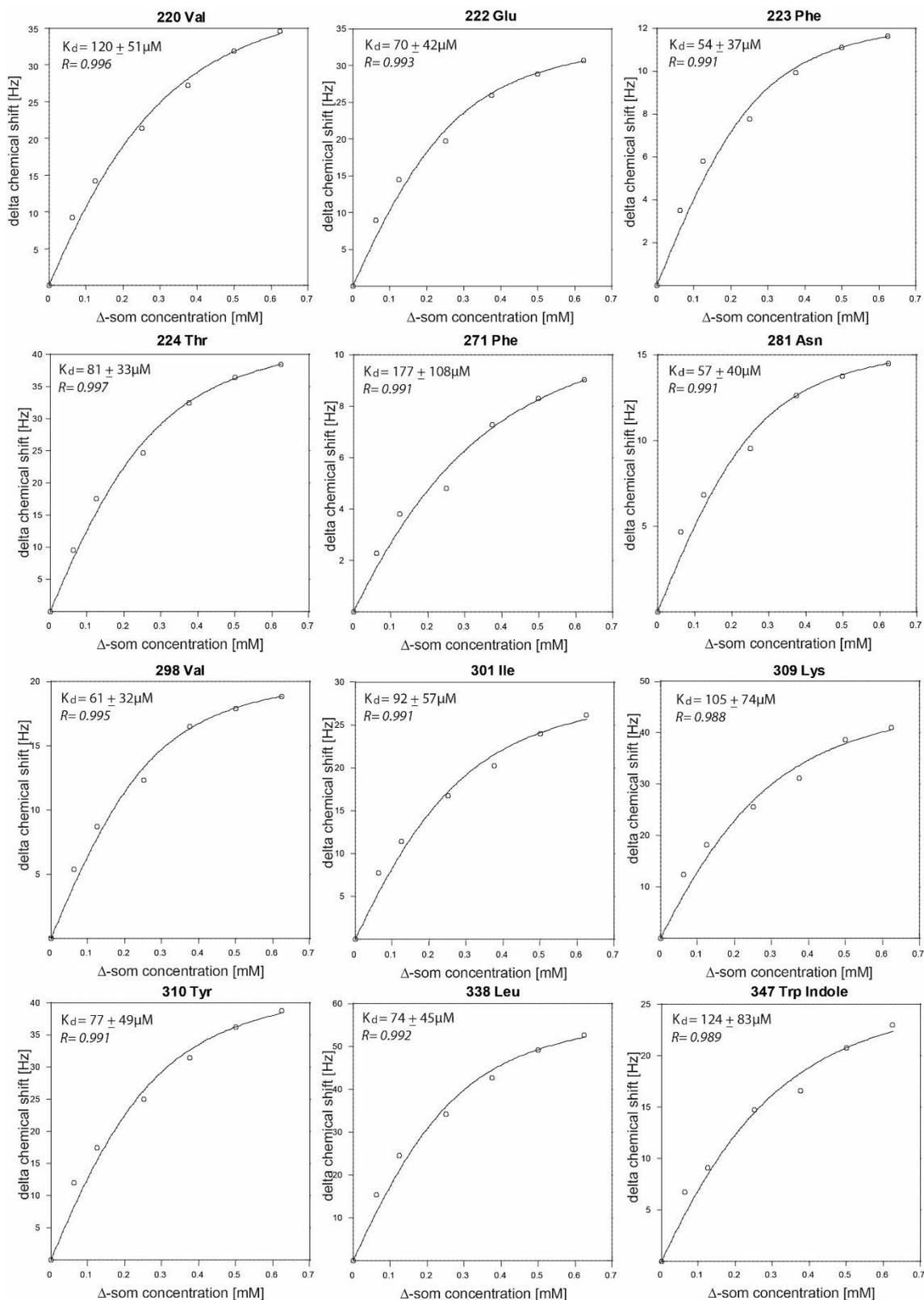
The 44 residues of **b'x** that fitted the curve gave an average dissociation constant of  $103 \pm 47 \mu\text{M}$  (see Figure 5.4 for sample fits). Similarly, the 33 residues of reduced W111F **abb'x** that fitted the curve gave an average dissociation constant of  $73 \pm 31 \mu\text{M}$ . (see Figure 5.5 for sample fits). These values are both in agreement with the estimated binding affinity of  $\Delta$ -som of 0.1-1 mM (Byrne et al., 2009), suggesting that binding affinity is at the tighter end of this estimated region.

The data points collected fit well to the dissociation curves but gave large errors, likely due to the limitations of the curve fits not reaching a full occupancy plateau- a large error can be associated with extrapolating the curve fits to reach  $\Delta_0$  (Fielding, 2003). This is a particular problem when fitting weak affinities, as is the case here, but likewise the fits were also limited practically, by both protein and peptide concentrations: preliminary work carried out with the  $\Delta$ -som peptides showed several precipitate at concentrations higher than 0.75 mM when in aqueous buffer solutions (so 0.625 mM was used as the maximum ligand concentration). Equally, a minimum of 0.25 mM protein in each sample was required for collection of high quality spectra within a reasonable time period.



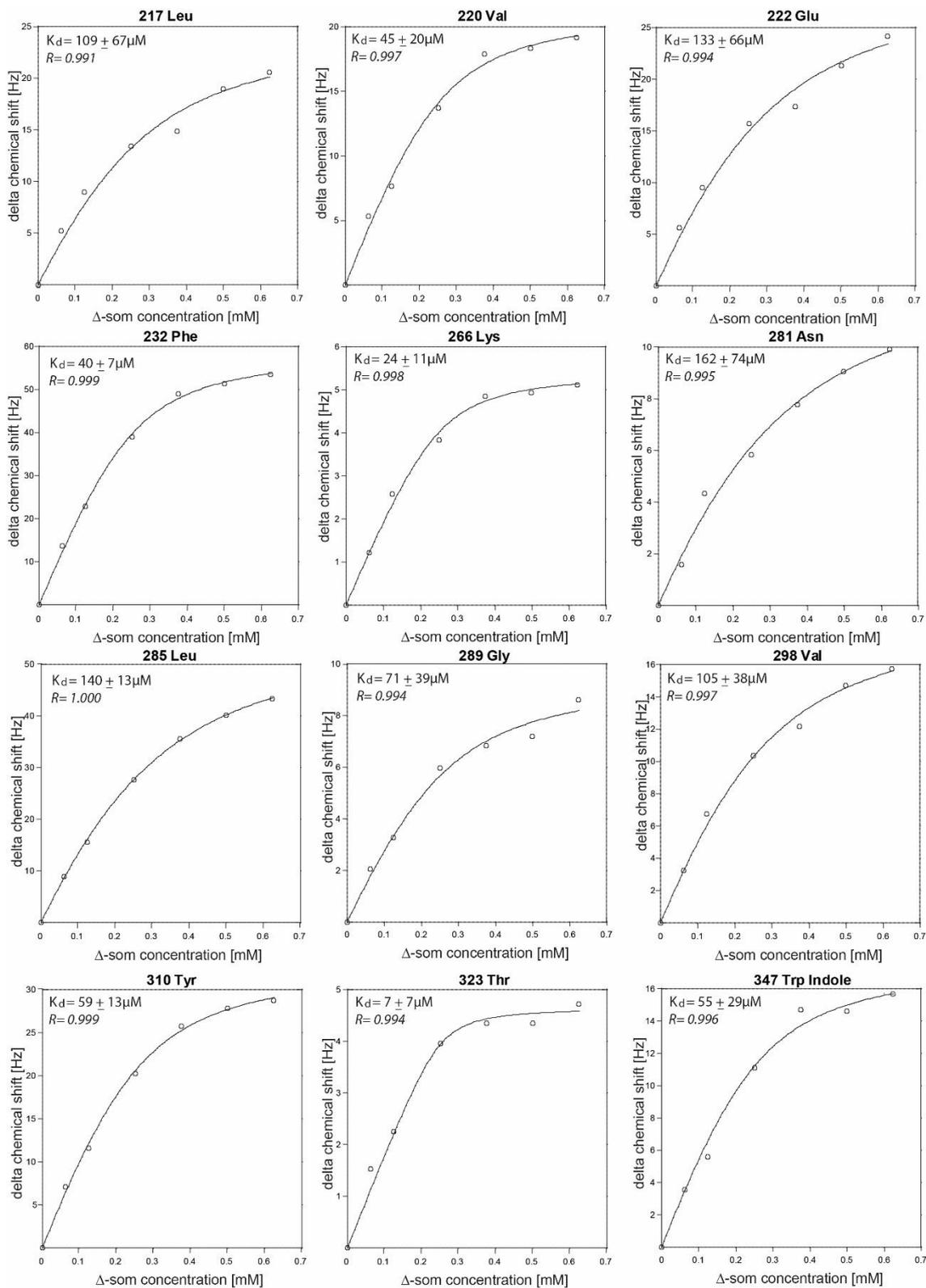
**Figure 5.3: Overlaid  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of  $0.25\text{ mM } ^{15}\text{N } b'x$  (a) and W111F  $abb'x$  (b) + increasing amounts of  $\Delta$ -som. Spectra: +0 mM  $\Delta$ -som (black), +0.125 mM  $\Delta$ -som (red), +0.375 mM  $\Delta$ -som (cyan), +0.625 mM  $\Delta$ -som (blue). Arrows in (a) indicate the direction example peaks track throughout the titration series. Sample graph fits seen in Figures 5.4 and 5.5 are highlighted on each spectra.**

# Ligand Binding Studies of $b'x$ and W111F $abb'x$ by NMR Spectroscopy



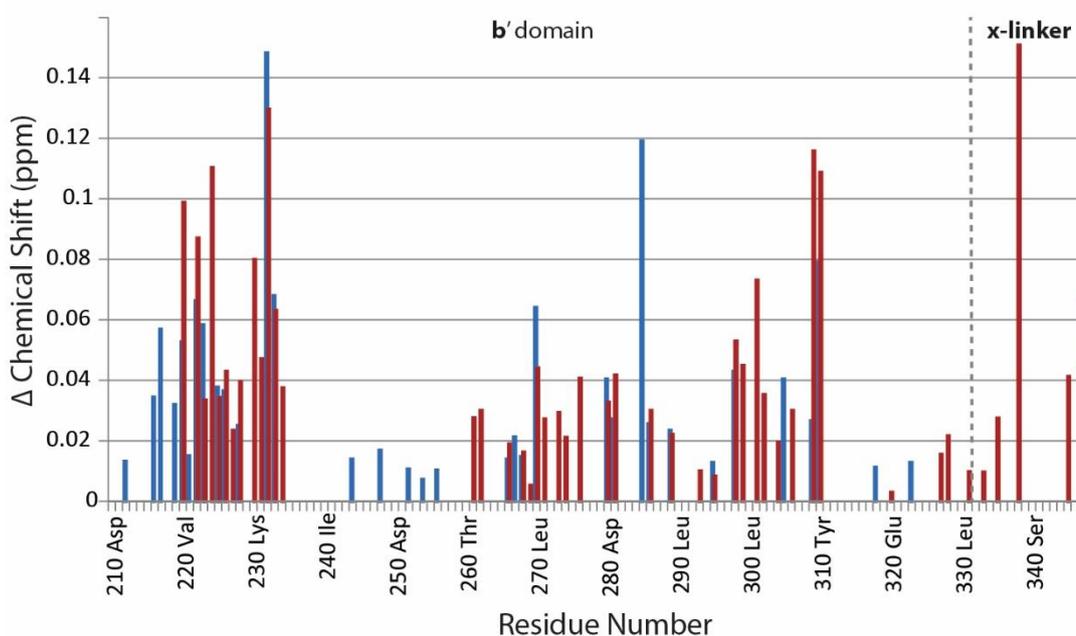
**Figure 5.4: Example dissociation constant titration curves using overlaid  $^{15}\text{N}, ^1\text{H}$ -HSQC data of 0.25 mM  $^{15}\text{N}$   $b'x$  + increasing amounts of  $\Delta$ -som. The curves give an average dissociation constant of  $103 \pm 47 \mu\text{M}$ .**

# Ligand Binding Studies of $\mathbf{b}'\mathbf{x}$ and W111F $\mathbf{abb}'\mathbf{x}$ by NMR Spectroscopy



**Figure 5.5:** Example dissociation constant titration curves using  $^{15}\text{N}, ^1\text{H}$ -HSQC data of 0.25 mM  $^{15}\text{N}$  reduced W111F  $\mathbf{abb}'\mathbf{x}$  + increasing amounts of  $\Delta$ -som. The curves give an average dissociation constant of  $73 \pm 31 \mu\text{M}$ .

Chemical shift perturbations of both protein fragments reveal that it is largely the same areas of the **b'** domain that are perturbed upon ligand binding as published previously (Byrne et al., 2009b): this includes a chain of residues at the start of the **b'** domain that form an  $\alpha$ -helix bordering the c-terminal end of the binding site (T224-G233), and residues L270, L300 and Y310, which all form the  $\beta$ -sheet 'floor' of the binding site. In addition, L338 and W347 indole resonance of the **x-linker** region were also perturbed upon binding, showing displacement of the **x-linker** (see Figure 5.6 for chemical shift perturbations of **b'**x (in red) and W111F **abb'**x (in blue)).



**Figure 5.6:** Chemical shift perturbation map of the **b'**x portion of either 0.25 mM **b'**x (red) or 0.25 mM reduced W111F **abb'**x (blue), upon addition of 0.625 mM  $\Delta$ -som. The 'ab' portion of W111F **abb'**x experiences very minor shifts (average shifts of 0.011 ppm) that did not fit graphically, and so were not included in the shift map.

The overlaid perturbation maps show that there are more trackable resonances in **b'**x spectra than in W111F **abb'**x (35 in W111F **abb'**x as opposed to 46 in **b'**x), particularly those of **x-linker** and the ligand binding site. Due to its large size, spectra of W111F **abb'**x were subject to residue overlap- a significant number of trajectories could not be accurately

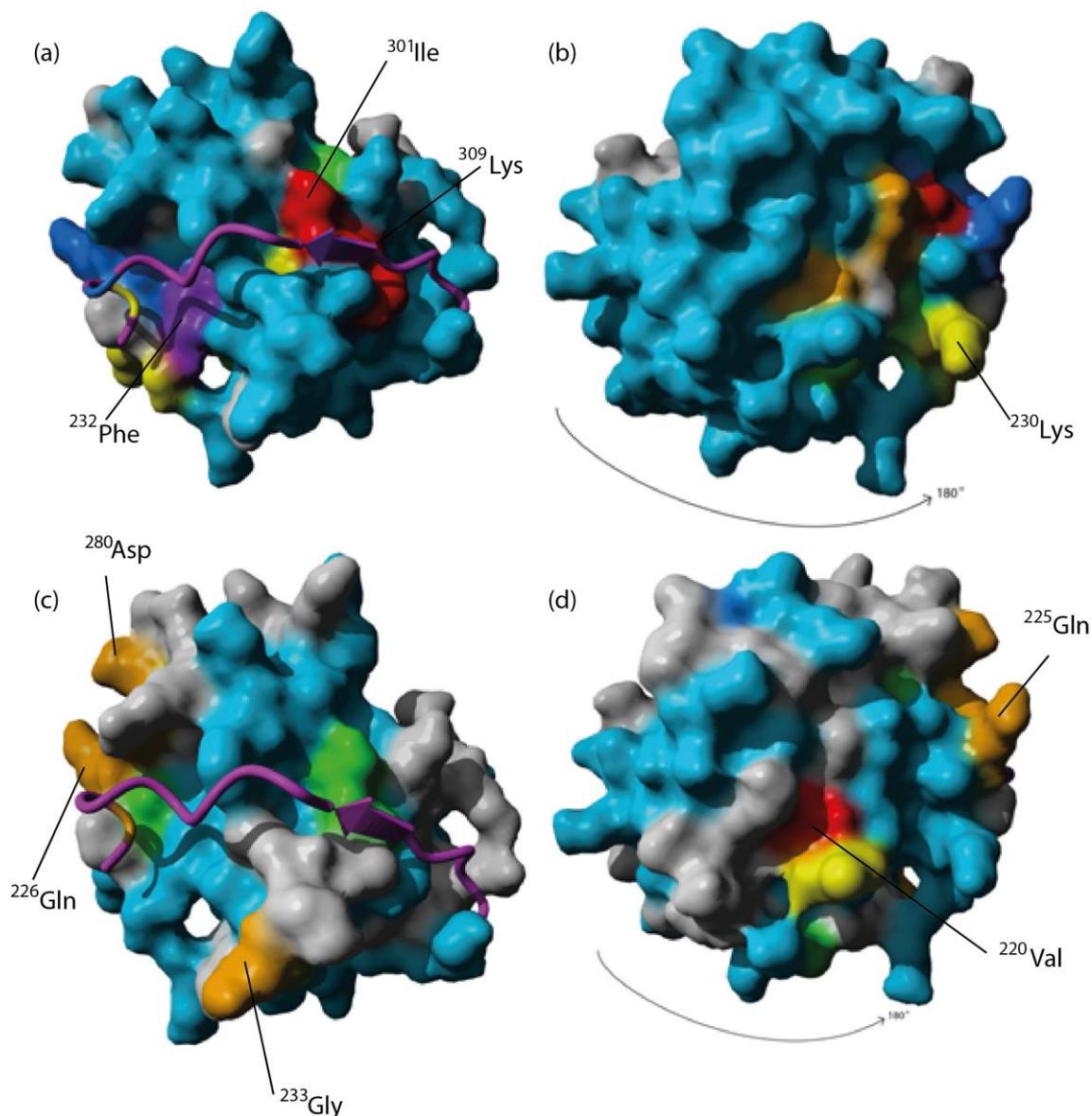
tracked throughout the titration series, and therefore could not be included in the shift maps or further analysis. Shifts observed in the '**ab**' portion of W111F **abb'x** fragment are not included in the maps; there were no observable resonances of **a** domain that shifted significantly upon binding, and the 19 shifts observed in **b** domain were very small in comparison to shifts witnessed in **b'x** (average 0.011 ppm). It appears only **b'** domain and the **x-linker** region are significantly perturbed upon peptide ligand binding.

Significantly shifting residues for surface maps were determined using the method described by (Schumann et al., 2007). This method involves rounds of standard deviation ( $\sigma$ ) calculations, removing residues shifting greater than the mean +  $3\sigma$  from the calculations, thereby limiting skew of the data (by a few residues that may shift a disproportionately large amount). Shifts are coloured in Figure 5.7 as  $> \text{mean} + 0.5\sigma$  (blue),  $> \text{mean} + 1\sigma$  (green),  $> \text{mean} + 2\sigma$  (yellow),  $> \text{mean} + 3\sigma$  (orange),  $> \text{mean} + 4\sigma$  (red), and  $> \text{mean} + 5\sigma$  (purple). The 3BJ5.pdb **b'x** structure was used to map shifts for both **b'x** and W111F **abb'x** fragments, as no significant shifts were observable in the '**ab**' portion of W111F **abb'x** to map. In addition, the 3BJ5.pdb structure is the I272A 'capped' mutant of **b'x**, where the **x-linker** occludes the binding site, allowing for easy identification of the binding site.

Surface maps of the **b'x** fragment of both **b'x** and W111F **abb'x** show two major regions of significantly shifting residues- the first being residues reporting throughout the binding site (identified by the region 'capped' by the **x-linker** in Figure 5.7 (a) and (b)), and a region of shifting residues on the opposite side of the protein (that correspond to the  $\alpha$ -helix bordering the c-terminal end of the binding site, T224-G233). It is likely that these residues are reporting upon a binding induced conformational change to the protein, as they are not directly involved in the binding event itself.

Each protein reports a similar surface and magnitude of chemical shifts, but W111F **abb'x** reports a somewhat less well-defined binding surface due to the limitations discussed above. The similar  $K_d$  values generated, along with the small shifts witnessed in **b** domain and lack of observable shifts in **a** domain upon ligand binding, suggest these domains are not involved in the binding event. Work carried out within the laboratory on the reduced **abb'xa'** fragment of hPDI showed the average dissociation constant to be  $108 \pm 84 \mu\text{M}$  at  $37^\circ\text{C}$ . This figure is also in close agreement with the values determined by this thesis.

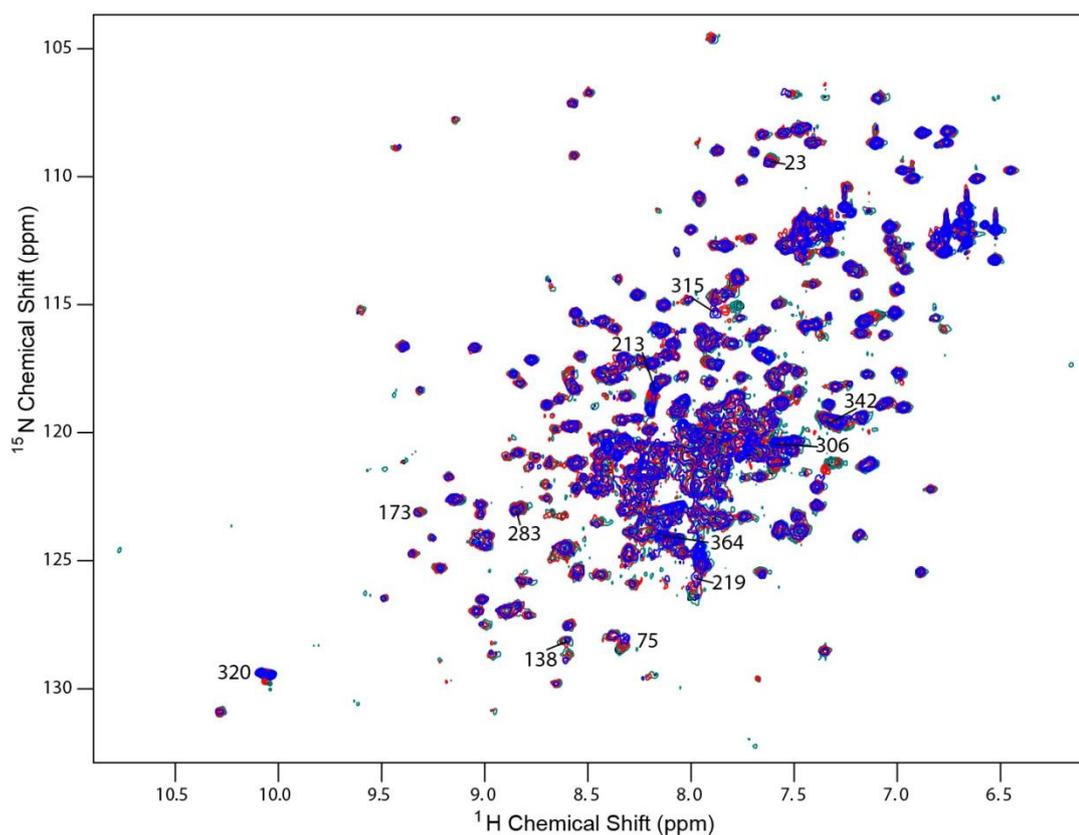
Together this suggests that the peptide ligand binding affinity of **b'** domain is not influenced by the presence of other neighbouring domains.



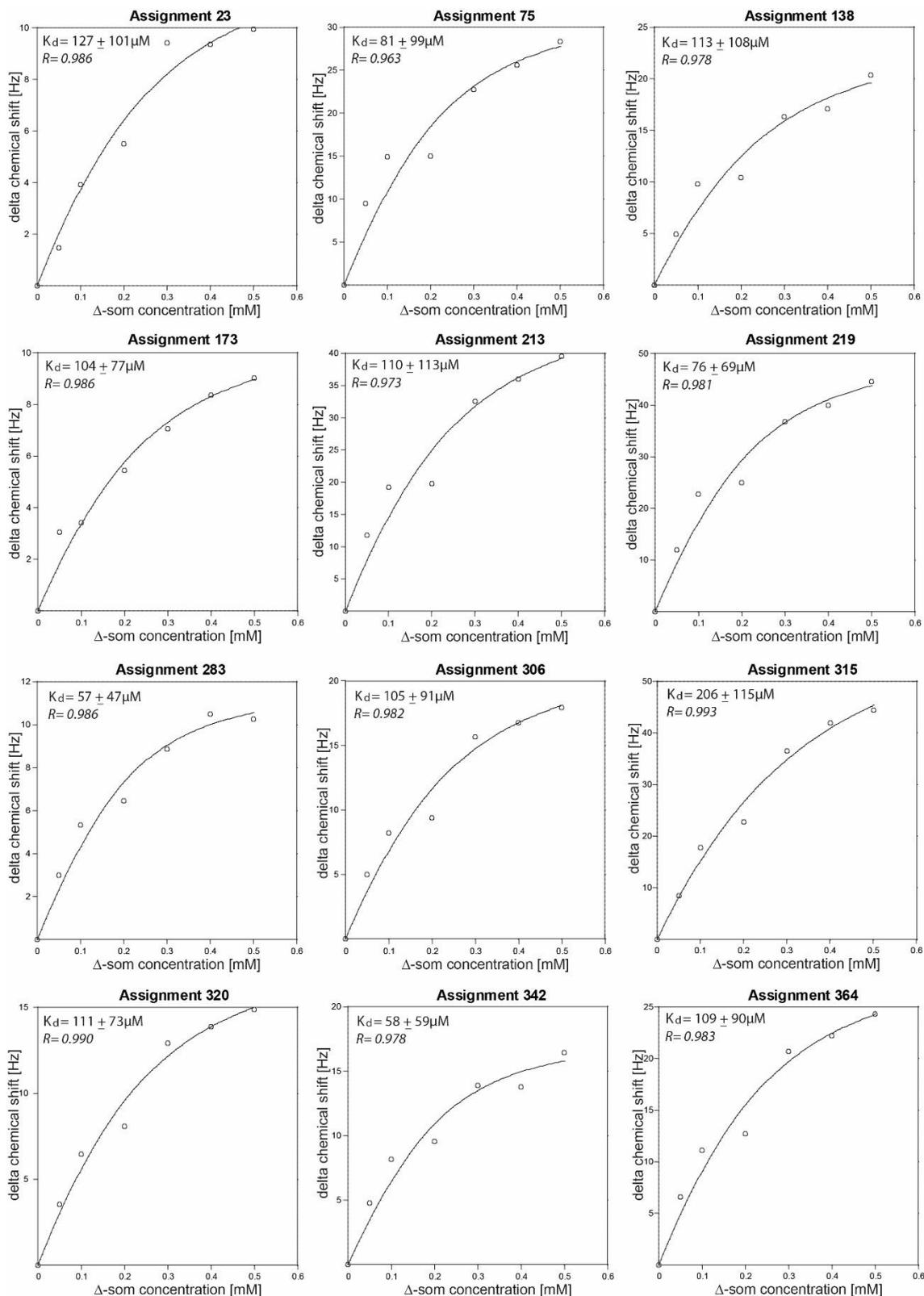
**Figure 5.7: Models of hPDI *b'*'x (3BJ5.pdb) showing significantly shifting backbone resonances of *b'*'x protein (a) and W111F *abb'*'x protein (b) upon 2.5 molar excess of binding ligand  $\Delta$ -som. Amide resonances shifting  $> \text{mean} + 0.5\sigma$  (blue),  $> \text{mean} + 1\sigma$  (green),  $> \text{mean} + 2\sigma$  (yellow),  $> \text{mean} + 3\sigma$  (orange),  $> \text{mean} + 4\sigma$  (red), and  $> \text{mean} + 5\sigma$  (purple). To aid site identification the **x-linker** is coloured magenta, and all other assigned residues are coloured teal. Unassigned residues are coloured grey. Molecular graphics created with YASARA ([www.yasara.org](http://www.yasara.org)).**

### 5.3.2 Effect of Protein Redox State on Peptide Ligand Binding

Whilst dissociation constants of the **b**'x and reduced W111F **abb**'x fragments of hPDI reported in 5.3.1 are in agreement with one another, a ligand titration of oxidised W111F **abb**'x was also carried out to determine if active site oxidation in **a** domain had any effect upon ligand binding affinity. Spectra collected of the oxidised titration are shown in Figure 5.8, and example dissociation constant titration curves are shown by Figure 5.9. Note that as oxidised spectra of W111F **abb**'x is sufficiently different to its reduced spectra (as shown previously by Figure 4.6) the titration series was carried out using unassigned resonances, but the graphical fits correspond to the numbered resonances present on the overlaid spectra.



**Figure 5.8:** Overlaid  $^{15}\text{N},^1\text{H}$ -HSQC spectra of 0.2 mM oxidised W111F **abb**'x + increasing amounts of  $\Delta$ -som. Spectra: +0 mM  $\Delta$ -som (blue), +0.1 mM  $\Delta$ -som (red), +0.3 mM  $\Delta$ -som (cyan), +0.5 mM  $\Delta$ -som (black). Unassigned sample graph peak fits for each protein are labelled.



**Figure 5.9:** Example dissociation constant titration curves using overlaid  $^{15}\text{N}, ^1\text{H}$ -HSQC data of 0.2 mM  $^{15}\text{N}$  oxidised **abb'x** + increasing amounts of  $\Delta$ -som. The curves give an average dissociation constant of  $100 \pm 79 \mu\text{M}$ .

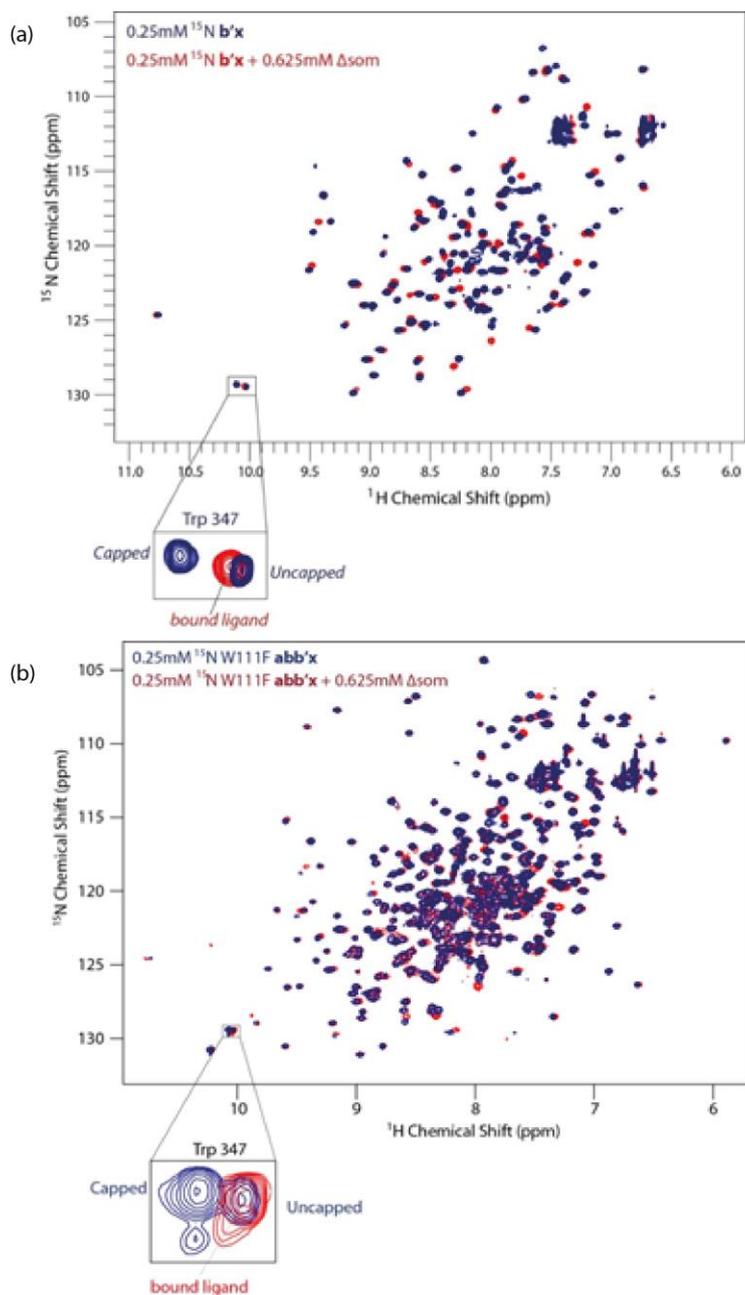
Some 21 residues could be tracked throughout the titration series, to give an average dissociation constant of  $100 \pm 79 \mu\text{M}$ . This is again in close agreement with the value of both the **b'x** and reduced W111F **abb'x** fragments.

From these findings the decision was made to probe ligand binding further using the **b'x** fragment; peptide ligand binding does not appear to be influenced by protein redox state or neighbouring domains. Due to its smaller size, the **b'x** fragment gives more simplified NMR spectra with a greater percentage of assignments, whilst still reporting on ligand binding representative to full-length protein.

### 5.3.3 Further Evidence of Capping

Capping is a published event that occurs in fragments of hPDI that are terminated at the end of the **x-linker** region (Nguyen et al., 2008). Further evidence of the capping event was shown in Chapter 4, of both the **b'x** and W111F **abb'x** fragments, where triple resonance assigned  $^{15}\text{N}, ^1\text{H}$ -HSQC spectra of each fragment show two resonances for the  $^{347}\text{Trp}$  indole (see Figure 4.21). The two resonances correspond to 'capped' and 'uncapped' forms of the **x-linker**. Upon addition of a binding ligand to each fragment (that competes with the **x-linker** to bind the peptide ligand binding site), only a single peak for  $^{347}\text{Trp}$  is visible (see Figure 5.10). This peak tracks throughout the titration from the 'capped' resonance towards the 'uncapped' resonance upon increasing concentrations of ligand, suggesting that, unlike capping, the ligand binding event is in fast exchange. As the concentration of ligand is increased, the protein reports a shift towards a more 'uncapped' form (as the binding site becomes increasingly more occupied by ligand than **x-linker**). This observation is further evidence that the **x-linker** is able to bind to and modulate the hydrophobic peptide ligand binding site of the **b'** domain in **b'x** and W111F **abb'x** hPDI fragments.

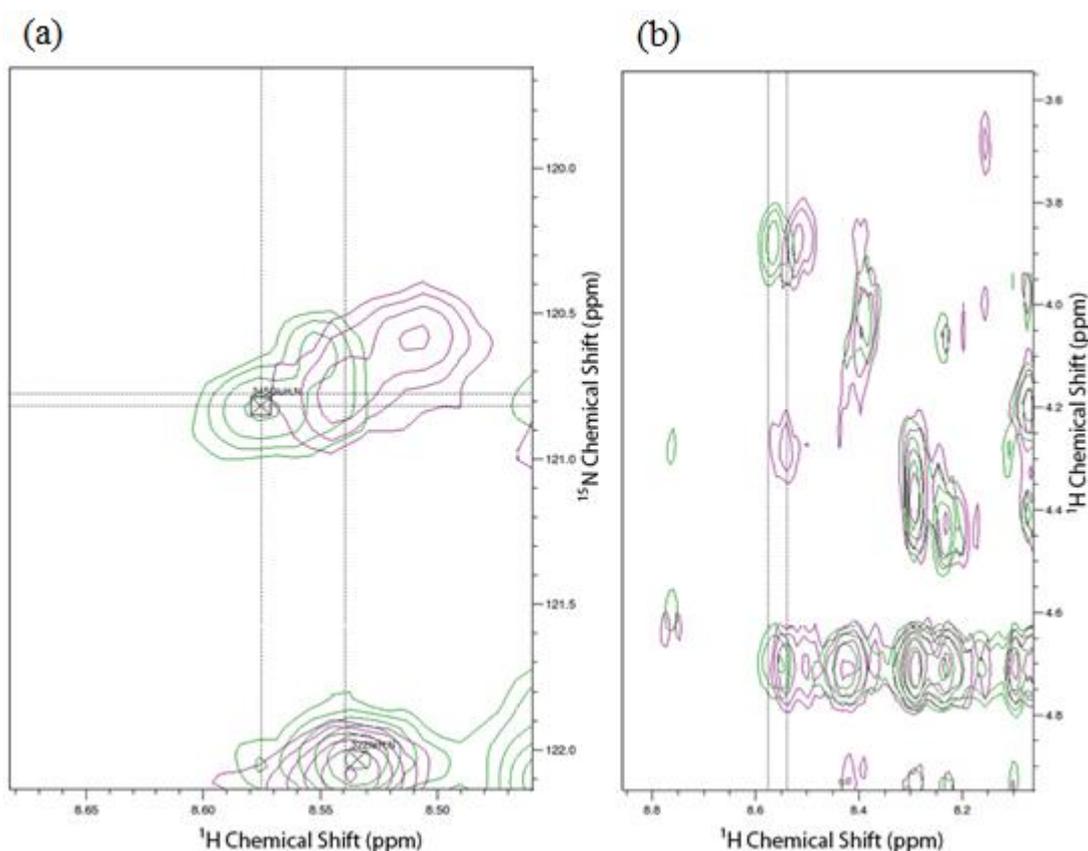
Interestingly, at the end point of the **b'x** ligand titration, the bound ligand  $^{347}\text{Trp}$  indole peak does not overlay with the uncapped  $^{347}\text{Trp}$  indole resonance (see Figure 5.10) to suggest a fully uncapped form of protein. Equally the graphical fits do not reach a plateau, suggesting 100% occupancy of the binding sites has not been reached (as discussed previously in 5.3.1).



**Figure 5.10:** Overlaid  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of  $0.25 \text{ mM } ^{15}\text{N}$  protein (blue) and  $0.25 \text{ mM } ^{15}\text{N}$  protein +  $0.625 \text{ mM } \Delta\text{-som}$  (red), for **b'x** in (a) and W111F **abb'x** in (b). The indole  $^{347}\text{Trp}$  resonances are expanded for each fragment, showing two peaks in the absence of ligand (x-linker in 'capped' and 'uncapped' conformations) and a single peak for the same resonance in the presence of excess ligand.

### 5.3.4 $^{15}\text{N}$ NOESY-HSQC for Evidence of Binding

NOESY-HSQC experiments were carried out using both **b'x** and W111F **abb'x** fragments, in an attempt to gain structural restraints and through-space evidence of  $\Delta$ -som binding, with the intention of creating a ligand bound solution structure of the fragments. Figure 5.11 shows an example plane of NOESY peaks collected at 37°C for residue  $^{345}\text{Glu}$ , one of only five residues that reported extra NOE peaks upon binding.



**Figure 5.11: Overlaid NOESY strips of 0.4 mM  $b'x$  (green) and 0.4 mM  $b'x$  + 1 mM  $\Delta$ -som (purple).** Window (a) shows resonance  $^{345}\text{Glu}$  shifting upon binding  $\Delta$ -som, and the corresponding  $^1\text{H}$ - $^1\text{H}$  strip in (b) shows an additional NOE peak at approximately 4.3 ppm in the ligand bound (purple) spectra.

Preliminary spectra have suggested  $\Delta$ -som is very conformationally active at 37°C (with spectra of the peptide collected at this temperature completely blank), so experiments were

repeated at 10°C to see if a lower running temperature could increase the number of extra NOE peaks observed (run using both **b'x** and W111F **abb'x** fragments). Unfortunately a lower running temperature was detrimental to the spectral quality. Each experiment was checked for NOESY peaks with and without addition of 2.5 molar excess of  $\Delta$ -som, but in this instance NOESY experiments did not yield a sufficient number of extra NOE peaks to perform structural restraint calculations.

### 5.3.5 Ligand Binding by $^{13}\text{C},^1\text{H}$ -HSQC spectra

As reported in Chapter 4, 89% of the possible **b'x** methyl  $^1\text{H}$  and  $^{13}\text{C}$  resonances were made (see Figure 4.15) using the HCCH-TOCSY experiments. A HCCH-TOCSY with ligand was collected, and a  $^{13}\text{C},^1\text{H}$ -HSQC of 0.3 mM  $^{13}\text{C}$  **b'x** and 0.75 mM unlabelled  $\Delta$ -som was assigned, with the intention of gaining more site-specific information of the binding event.

Observed chemical shift perturbations upon addition of ligand were weighted similarly to  $^{15}\text{N},^1\text{H}$  shifts, by weighting the  $^{13}\text{C}$  shifts relative to their  $^1\text{H}$  shifts. Figure 5.12 demonstrates the shifts witnessed throughout **b'x** protein, with the average methyl shift being 0.031 ppm (raw data can be found in Appendix 5.2).

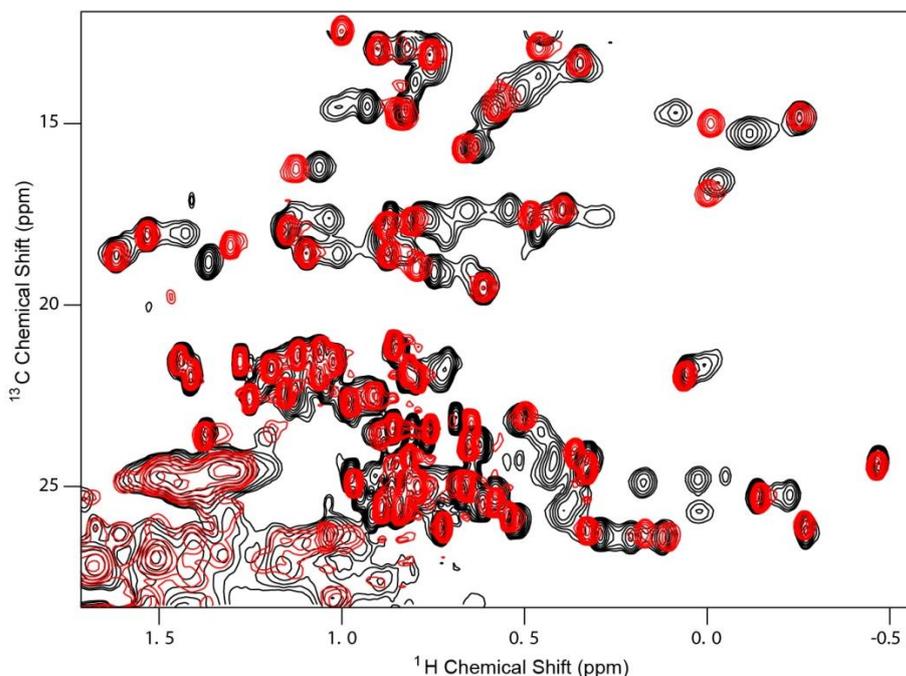
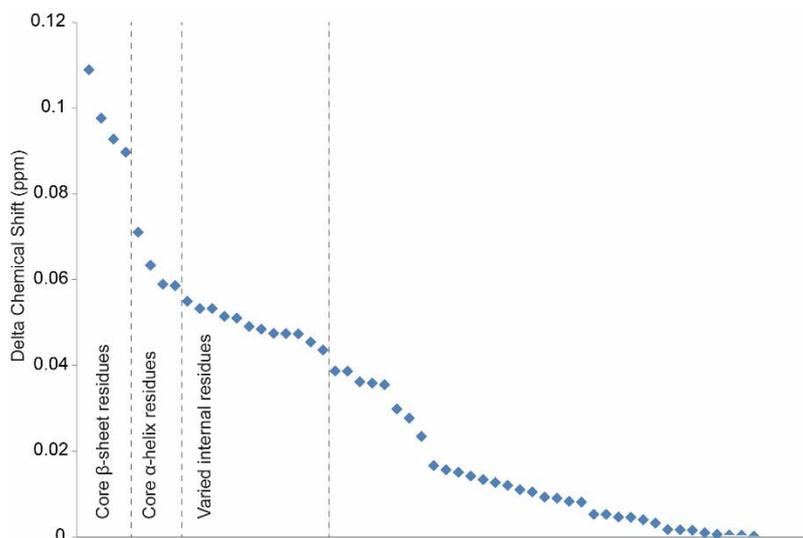


Figure 5.12: Overlaid  $^{13}\text{C},^1\text{H}$ -HSQC methyl region spectra of 0.3 mM  $^{13}\text{C}$  **b'x** (black) and 0.3 mM  $^{13}\text{C}$  **b'x** + 0.75 mM  $\Delta$ -som (red).

It was apparent that the  $^{13}\text{C}$  data could not be treated similarly to the  $^{15}\text{N}$  data; whilst of a comparable magnitude, no shifts were  $3\sigma$  greater than the mean. Shifts were instead ranked by size, as shown in Figure 5.13.



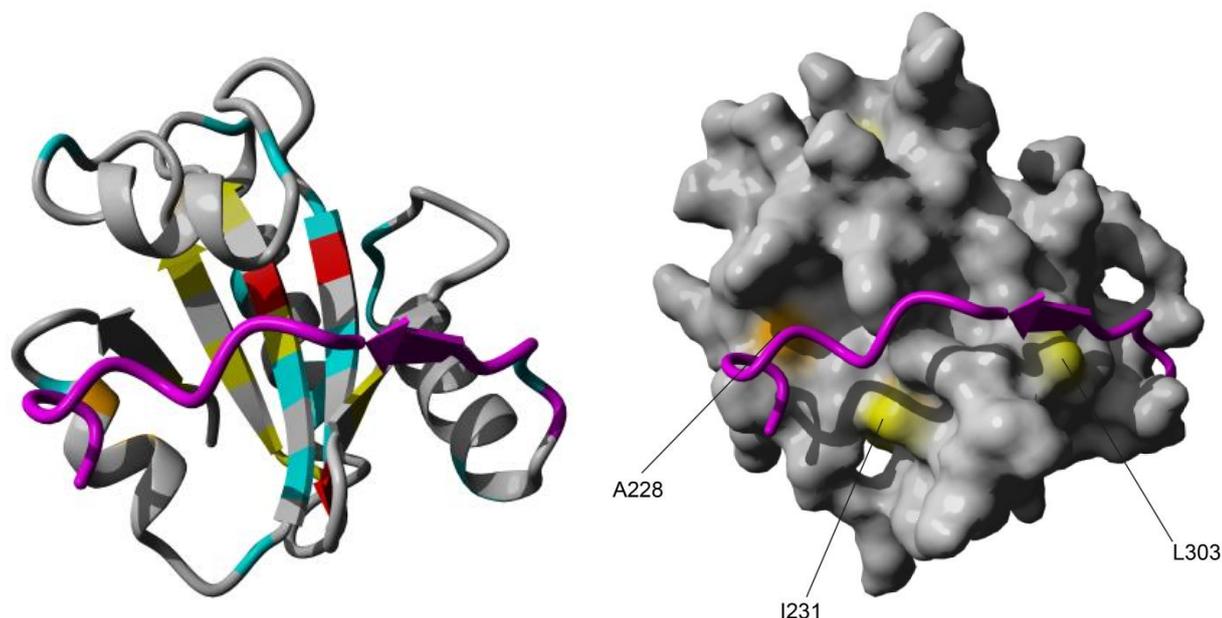
**Figure 5.13:**  $^{13}\text{C}$  methyl chemical shift perturbations upon addition of 0.75 mM  $\Delta$ -som to 0.3 mM **b'x**, ranked by size. Shifting resonances  $>0.09$  ppm were found to be  $\beta$ -strand, 0.055-0.09 ppm were  $\alpha$ -helical, and 0.04-0.055 ppm were additional internal residues.

A threshold shift of 0.04 ppm was chosen from the graph, and all shifts greater than this were mapped onto the **b'x** structure (see Figure 5.14). Shifts were witnessed throughout the entire protein core, but the surface map shows how few of the methyl 'reporters' were actually surface exposed- just three of the assignable methyl resonances were present within the binding site and surface exposed:  $^{228}\text{Ala}$ ,  $^{231}\text{Ile}$  and  $^{303}\text{Leu}$  (labelled).

It is interesting to note that the largest shifts ( $> 0.09$  ppm, coloured red) corresponded to core beta strand residues, and shifts 0.055-0.09 ppm (coloured orange) corresponded only to residues in alpha helices. Shifts between 0.04-0.055 ppm (coloured yellow) also all corresponded to internal residues.

Whilst in this instance collection of  $^{13}\text{C}$  chemical shifts did not yield more detailed binding-site specific data as anticipated, the shifts still help to confirm a conformational

change the protein structure undergoes upon ligand binding, as also observed by  $^{15}\text{N},^1\text{H}$ -HSQC in 5.3.1.

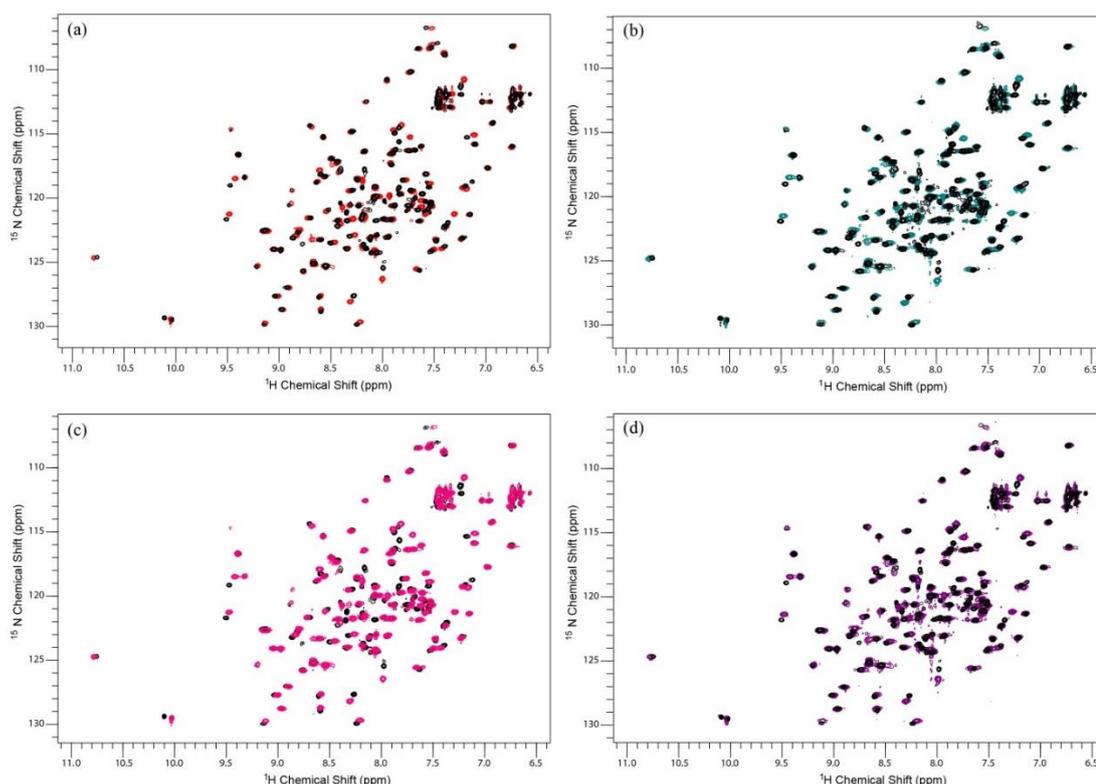


**Figure 5.14: Model of hPDI *b*'x (3BJ5.pdb) showing significantly shifting methyl resonances upon 2.5 molar excess of binding ligand  $\Delta$ -som, and surface map showing exposed methyls.** Methyl resonances shifting  $>0.04\text{ppm}$  are coloured yellow,  $>0.055\text{ppm}$  are coloured orange, and  $>0.09\text{ppm}$  are coloured red. To aid site identification the **x-linker** is coloured magenta, and the rest of the molecule coloured grey. Molecular graphics created using Yasara ([www.yasara.org](http://www.yasara.org))

### 5.3.6 Use of Fluorinated Ligands to Monitor Peptide Ligand Binding

Ligand binding studies were carried out using fluorinated analogues of the model binding peptide  $\Delta$ -som, in an attempt to gain more detailed information of the binding specificity of hPDI. The three Phe residues of the peptide were each fluorinated in turn (F1, F2 and F3  $\Delta$ -som) along with a peptide where all three residues were fluorinated (F1,2,3  $\Delta$ -som). As large residues with hydrophobic, aromatic side chains, their presence in the peptide sequence was thought to play a key role in the chaperone-like ligand recognition of hPDI to  $\Delta$ -som.

Dissociation constants of **b'x** and each fluorinated peptide were calculated from practical data, first by the titration of increasing concentrations of **b'x** protein into 0.15 mM of each fluorinated ligand. These titrations were monitored by  $^{19}\text{F}$  1D spectra, to view ligand-observed chemical shift perturbations. The titration series was then repeated, using constant protein and varying ligand, and monitored by  $^{15}\text{N},^1\text{H}$ -HSQC to track traditional protein-observed chemical shift perturbations (as outlined in 5.2.1). The  $^{15}\text{N},^1\text{H}$ -HSQC tracked spectra of protein-observed titrations (start and finish overlays) are shown below in Figure 5.15. Overlaid  $^{19}\text{F}$  1D ligand-observed spectra of single fluorinated peptides (F1, F2, and F3  $\Delta$ -som), and triple fluorinated F1,2,3  $\Delta$ -som peptide titrations are shown in Figures 5.16-5.19.



**Figure 5.15:** Overlaid  $^{15}\text{N},^1\text{H}$ -HSQC spectra 0.25 mM **b'x** (black) and +0.625 mM F1  $\Delta$ -som (red)(a), +0.625 mM F2  $\Delta$ -som (cyan)(b), +0.625 mM F3  $\Delta$ -som (pink)(c) and +0.625 mM F1,2,3  $\Delta$ -som (purple)(d).

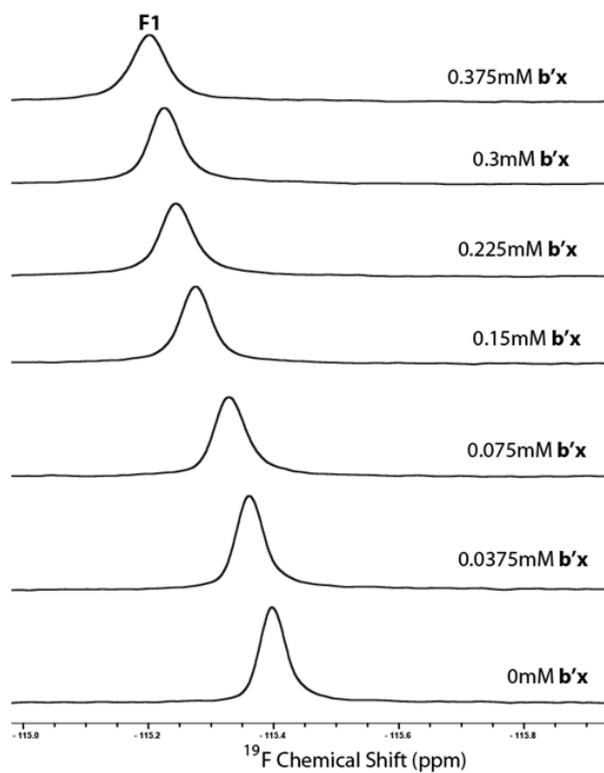


Figure 5.16: Overlaid  $^{19}\text{F}$  spectra 0.15 mM F1  $\Delta$ -som+ increasing  $^{15}\text{N}$  **b'x** titration series

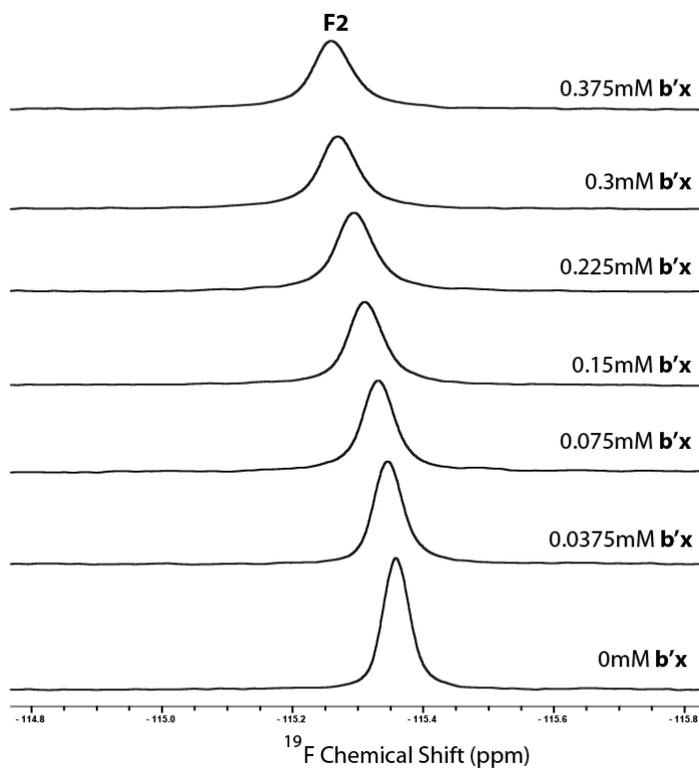


Figure 5.17: Overlaid  $^{19}\text{F}$  spectra 0.15 mM F2  $\Delta$ -som+ increasing  $^{15}\text{N}$  **b'x** titration series

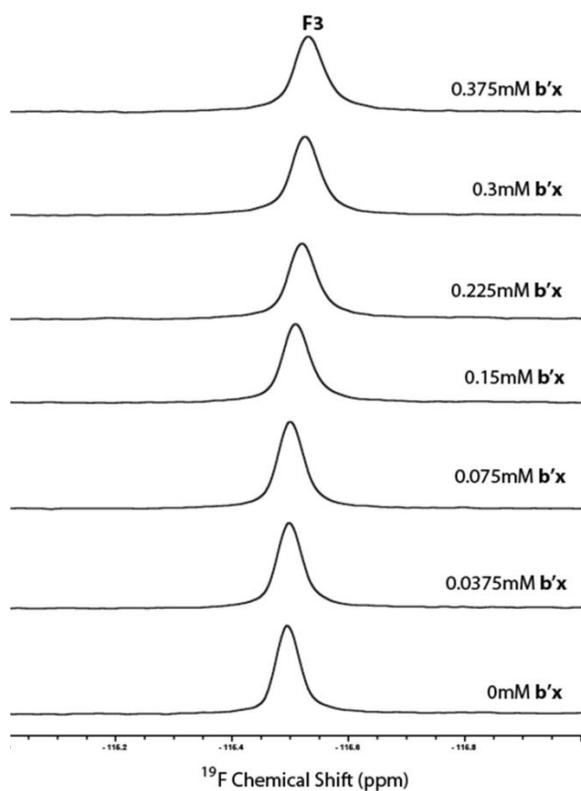


Figure 5.18: Overlaid  $^{19}\text{F}$  spectra 0.15 mM F3  $\Delta$ -som + increasing  $^{15}\text{N}$   $b'x$  titration series

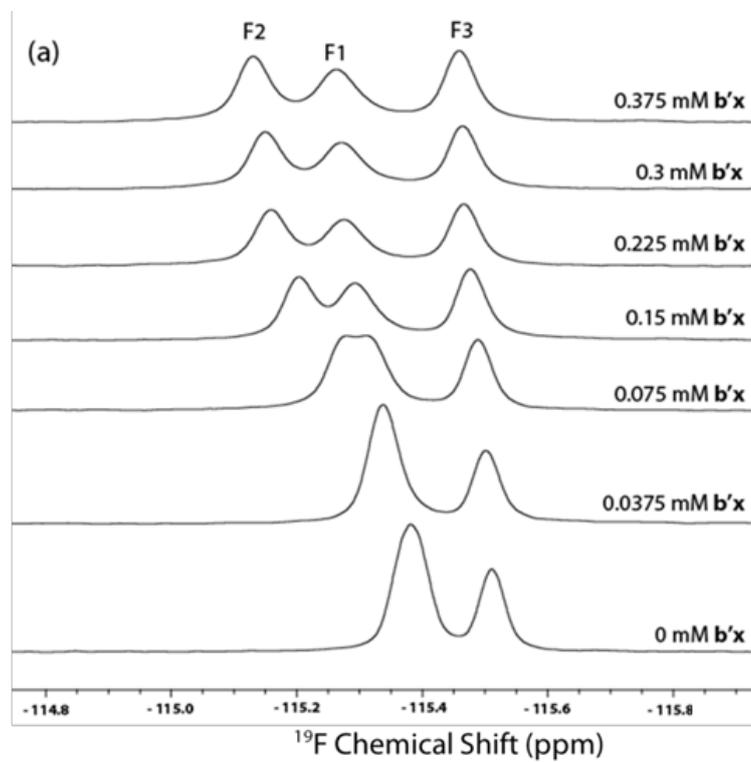
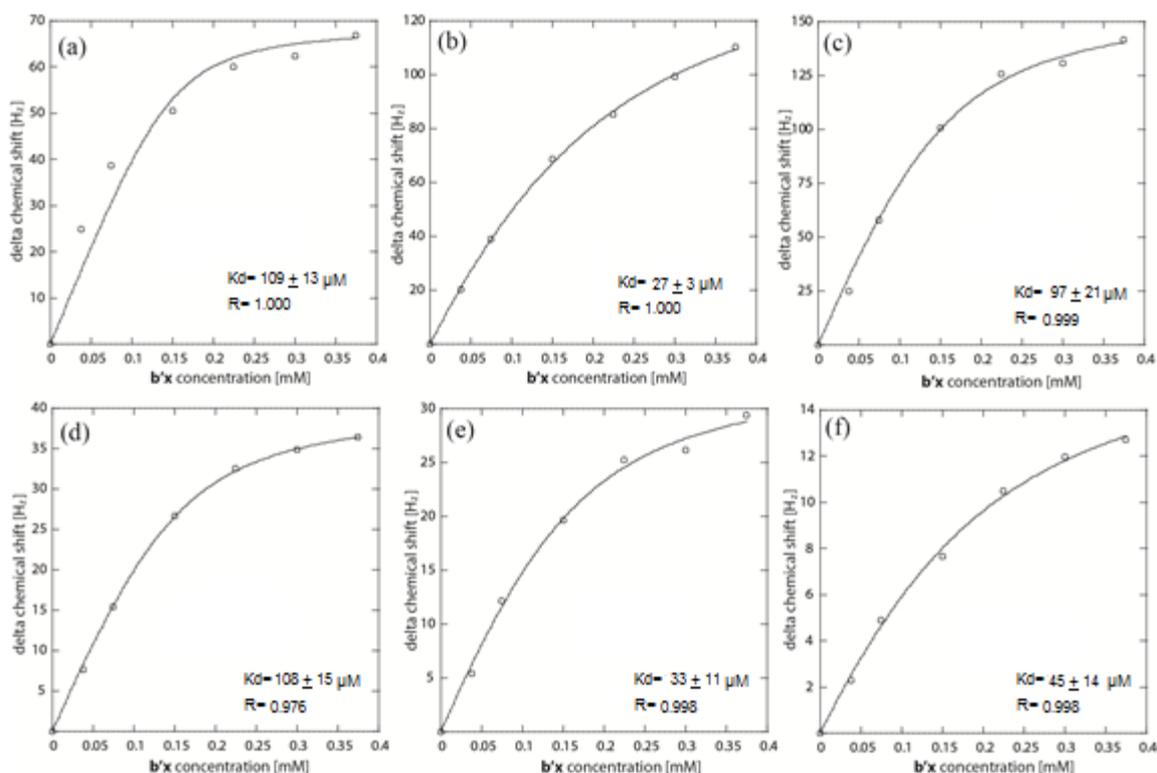


Figure 5.19: Overlaid  $^{19}\text{F}$  spectra 0.15 mM F1,2,3  $\Delta$ -som + increasing  $^{15}\text{N}$   $b'x$  titration series

Please note, only  $^{19}\text{F}$ -observed graphical fits are shown below in Figure 5.20, but all fluorine experiments were repeated using constant protein in  $^{15}\text{N}, ^1\text{H}$ -HSQC protein-observed titrations. Graphical fits of this data for each peptide can be found in Appendix 5.3.



**Figure 5.20: Dissociation constant titration curves of F1,2,3  $\Delta$ -som peptide: F1 peak (a), F2 peak (b) and F3 peak (c). Dissociation constant titration curves using single fluorinated  $\Delta$ -som peptide: F1  $\Delta$ -som (d), F2  $\Delta$ -som (e) and F3  $\Delta$ -som (f).**

The results of each single and triple fluorinated peptide as observed by  $^{15}\text{N}, ^1\text{H}$ -HSQC, and the individual fits of all  $^{19}\text{F}$  1D ligand-observed peak fits are summarised in Tables 5.3 and 5.4.

**Protein observed binding:** The 44 residues of **b'x** with triple fluorinated F1,2,3  $\Delta$ -som peptide that fitted to the curve gave an average dissociation constant of  $48 \pm 35 \mu\text{M}$ . This value was slightly tighter than the affinity with unfluorinated  $\Delta$ -som, however a modest

increase was anticipated as a potential result of increased hydrophobicity fluorine could confer to the peptide, working to stabilise a hydrophobically driven binding interaction. However, the calculated value is still well within the error range of unlabelled  $\Delta$ -som ( $103 \pm 47 \mu\text{M}$ ). The generated value is also in agreement with other work carried out within the laboratory, using fluorinated **b'x** protein and unlabelled ligand (Curtis-Marof et al., 2014).

**Table 5.3:  $K_d$  dissociation constants of all single fluorinated  $\Delta$ -som peptides with  $^{15}\text{N}$  *b'x*, as determined by either  $^{15}\text{N},^1\text{H}$ -HSQC or  $^{19}\text{F}$  1D spectra.**

Peptide Sequence		Protein $^{15}\text{N}$ Tracked $\mu\text{M}$	Ligand $^{19}\text{F}$ Tracked (Single-F) $\mu\text{M}$
<b>AGSKNFFWKTFTSS</b>	$\Delta$ -som	$103 \pm 47$	N/A
<b>AGSKN<u>F</u>FWKTFTSS</b>	F1 $\Delta$ -som	$61 \pm 47$	$109 \pm 13$
<b>AGSKNFF<u>W</u>KTFTSS</b>	F2 $\Delta$ -som	$51 \pm 30$	$27 \pm 3$
<b>AGSKNFFWKT<u>F</u>TSS</b>	F3 $\Delta$ -som	$84 \pm 53$	$97 \pm 21$

The observed fluorinated Phe residue for each experiment is underlined in bold.

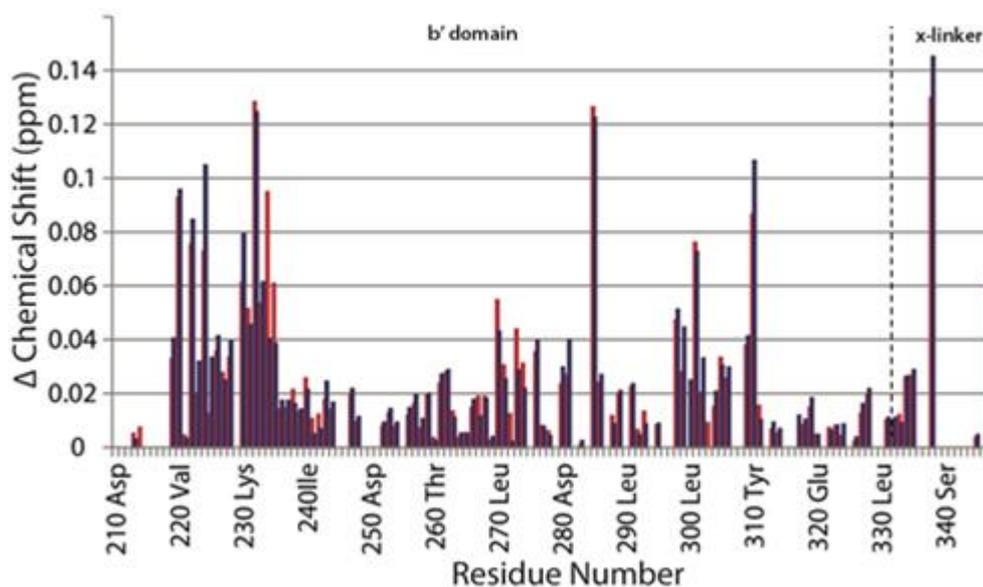
**Table 5.4:  $K_d$  dissociation constants of F1,2,3  $\Delta$ -som peptide with  $^{15}\text{N}$  *b'x*, as determined by either  $^{15}\text{N},^1\text{H}$ -HSQC or  $^{19}\text{F}$  1D spectra.**

Peptide Sequence		Protein $^{15}\text{N}$ Tracked $\mu\text{M}$	Ligand $^{19}\text{F}$ Tracked (Triple-F) $\mu\text{M}$
<b>AGSKNFFWKTFTSS</b>	$\Delta$ -som	$103 \pm 47$	N/A
<b>AGSKN<u>F</u>FWKTFTSS</b>	F1,2,3 $\Delta$ -som	N/A	$108 \pm 15$
<b>AGSKNFF<u>W</u>KTFTSS</b>	F1,2,3 $\Delta$ -som	N/A	$33 \pm 11$
<b>AGSKNFFWKT<u>F</u>TSS</b>	F1,2,3 $\Delta$ -som	N/A	$45 \pm 14$
<b>AGSKNFFWKT<u>F</u>TSS</b>	F1,2,3 $\Delta$ -som	$48 \pm 35^*$	$62 \pm 13^*$

The observed fluorinated Phe residue(s) for each experiment are underlined in bold. \*These  $K_d$  values are calculated as the mean of all individual  $K_d$  results obtained.

Further analysis of the  $^{15}\text{N},^1\text{H}$ -HSQC protein spectra showed the peptide appears to bind to the protein in the same manner as un-fluorinated peptide; residues track throughout the titration series in the same direction, and with comparable magnitude.

The chemical shift perturbation map in Figure 5.21 shows how the protein shifts upon addition of 2.5 molar excess of either un-fluorinated or triple fluorinated peptide.



**Figure 5.21:** Chemical shift perturbation map of 0.25mM  $^{15}\text{N}$  *b'x* and 0.625mM  $\Delta$ -som (blue) or 0.625mM F1,2,3  $\Delta$ -som (red).

It is clear that shifts are highly comparable in both magnitude and residue perturbed for both peptides, and similar results were obtained when using each singly fluorinated peptide. Overall the  $^{15}\text{N},^1\text{H}$ -HSQC protein-observed data showed that the fluorinated peptides behave in a similar manner to un-fluorinated  $\Delta$ -som, as highlighted by their highly similar chemical shift perturbations and similar calculated  $K_d$  values.

**Ligand observed binding:** Each single fluorinated peptide yielded a single peak in the spectra that tracked smoothly throughout the titration upon addition of increasing amounts of **b'x** protein. The calculated  $K_d$  values given in Table 5.3 demonstrate that  $^{19}\text{F}$  data is

concurrent with the  $^{15}\text{N}, ^1\text{H}$ -HSQC data; all single fluorinated  $^{19}\text{F}$  ligand-observed  $K_d$  fits are in agreement with their protein-observed  $K_d$  equivalents. A comparison of the three single fluorinated peptides shows that it is F2  $\Delta$ -som that reports a moderately tighter binding affinity when compared to the other two Phe residues, as reported by both  $^{15}\text{N}, ^1\text{H}$ -HSQC and  $^{19}\text{F}$  1D spectra. This suggests that the F2 Phe position of  $\Delta$ -som peptide may be crucial for the binding interaction to occur.

It was hoped that the single fluorine chemical shift trajectories would overlay well enough with the triple fluorinated peptide spectra to be able to assign each resonance- however it was soon apparent that peptide assignments could not simply be mapped over. In spectra of F1,2,3  $\Delta$ -som, the F1 and F2 resonances report such similar chemical shift perturbations that they appear as a single peak until the addition of 0.075 mM **b'x** protein (see Figure 5.19). The two residues are adjacent in the peptide sequence, and it is likely they experience highly similar chemical environments (and hence report similar chemical shift perturbations). Deconvolution attempts of the combined peak were unsuccessful, so the first two data points of each titration were fitted using the overlapped peak maxima (which may explain the slightly poorer fit of F1 peak of F1,2,3  $\Delta$ -som in Figure 5.20(a)). Sensitivity of chemical shift perturbations to more widespread fluorination was most apparent for the F3 peak, which tracked to the right as a single fluorinated peptide, and to the left as the triple fluorinated F1,2,3  $\Delta$ -som peptide (see Figure 5.18 for single F3 peak, and Figure 5.19 for F3 peak as part of F1,2,3  $\Delta$ -som). Several 2D  $^{19}\text{F}, ^{13}\text{C}$  NMR experiments using  $^{13}\text{C}$  natural abundance were carried out in an attempt to unambiguously assign each peak in F1,2,3  $\Delta$ -som, but as previously observed for spectra of  $\Delta$ -som collected at 37°C they were blank. Instead, the tight  $^{19}\text{F}$  calculated  $K_d$  fits of the single fluorinated peptides were used to assign the fluorine peaks in the F1,2,3  $\Delta$ -som spectra.

Overall the  $^{19}\text{F}$  observed data gave extremely good fits, with smaller errors. It appears that fluorinated peptides behave in a similar manner to unfluorinated  $\Delta$ -som, as observed by both protein and ligand viewpoints. The most prominent differences observed were not between fluorinated and unfluorinated peptides as anticipated, but between the individual Phe resonances of the fluorinated peptides.  $K_d$  of individual  $^{19}\text{F}$  peaks of F1,2,3  $\Delta$ -som peptide were in agreement with the values calculated from the F1 and F2  $\Delta$ -som peptides (see  $^{19}\text{F}$ -calculated values in Tables 5.3 and 5.4), with F1 values quoted within 1  $\mu\text{M}$  and F2 values within 6  $\mu\text{M}$ . However, F3  $\Delta$ -som reported a notably different  $K_d$  value in the

single and triple fluorinated peptides, reporting a two-fold tighter affinity as the triple fluorinated peptide. This finding, coupled with its trajectory change, suggests the F3 Phe residue is perhaps more sensitive to widespread fluorination across the peptide than F1 or F2 Phe residues are, or alternatively the residue may play a more prominent role in the binding event, that is stabilised by the widespread fluorination.

### **5.3.7 Effect of Alanine Substitutions on Peptide Ligand Binding**

Titration carried out using selectively fluorinated peptides hinted that particular residues may contribute more to the binding interaction than others. To investigate this further, alanine substitutions were introduced at each of the three Phe residues that were previously fluorinated. Single, double and triple Phe-Ala substitutions were made, resulting in 7 new peptides that were used to calculate dissociation constants, by protein observed titrations.  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of 0.25 mM  $^{15}\text{N}$  **b'x** + 0.625 mM  $\Delta$ -som, overlaid with 0.25 mM  $^{15}\text{N}$  **b'x** + 0.625 mM each alanine substituted peptide are given for comparison in Figure 5.22, along with the  $K_d$  values of each peptide with **b'x** protein in Table 5.5. All  $K_d$  values were generated from an average of 44 shifting residues in the protein as reported previously. Graphical fits of this data for each peptide can be found in Appendix 5.4.

## Ligand Binding Studies of $b'x$ and W111F $abb'x$ by NMR Spectroscopy

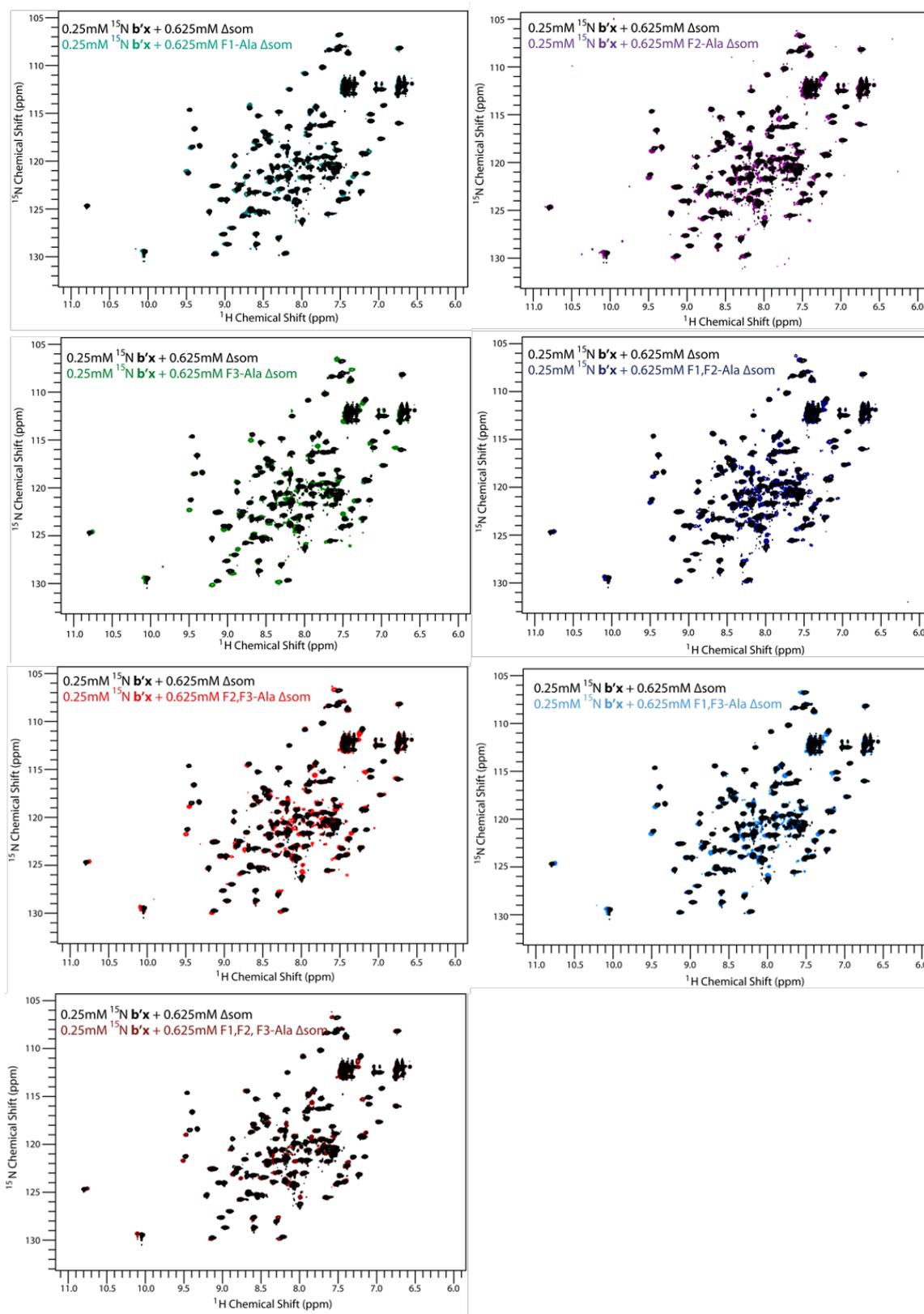


Figure 5.22:  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC overlaid spectra of  $0.25 \text{ mM } b'x + 0.625 \text{ mM } \Delta\text{-som}$  (black) overlaid with each alanine substituted peptide ligand ( $0.625 \text{ mM}$ ).

**Table 5.5:  $K_d$  dissociation constants of single, double and triple alanine substituted  $\Delta$ -som peptides as calculated by  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC.**

Peptide Sequence		$^{15}\text{N}$ Tracked / $\mu\text{M}$
<b>AGSKNFFWKTFTSS</b>	$\Delta$ -som	$103 \pm 47$
<b>AGSKN<u>A</u>FWKTFTSS</b>	Phe6 $\rightarrow$ A $\Delta$ -som	$199 \pm 100$
<b>AGSKNF<u>A</u>WKTFTSS</b>	Phe7 $\rightarrow$ A $\Delta$ -som	$123 \pm 96$
<b>AGSKNFFWKT<u>A</u>TSS</b>	Phe11 $\rightarrow$ A $\Delta$ -som	$261 \pm 101$
<b>AGSKN<u>AA</u>WKTFTSS</b>	Phe6,7 $\rightarrow$ A $\Delta$ -som	$242 \pm 177$
<b>AGSKN<u>A</u>FWKT<u>A</u>TSS</b>	Phe6,11 $\rightarrow$ A $\Delta$ -som	$587 \pm 174$
<b>AGSKNF<u>A</u>WKT<u>A</u>TSS</b>	Phe7,11 $\rightarrow$ A $\Delta$ -som	$412 \pm 219$
<b>AGSKN<u>AA</u>WKT<u>A</u>TSS</b>	Phe6,7,11 $\rightarrow$ A $\Delta$ -som	<i>Doesn't bind</i>

The alanine substitutions for each peptide are underlined in bold.

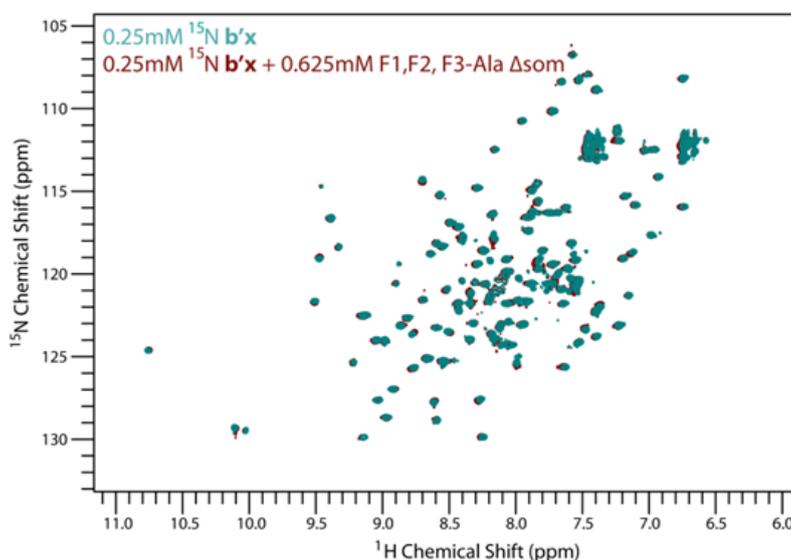
Progressive Phe $\rightarrow$ Ala mutation peptides showed weaker binding affinities to **b'x**, with around a two-fold weaker affinity. These findings demonstrate that whilst no single Phe residue is crucial for the binding interaction to occur, all three Phe residues must contribute to binding in part, as removal of any single Phe residue resulted in a weaker binding affinity. Interestingly, it was substitution of the F3 Phe (Phe 11) that resulted in the greatest negative change to binding affinity ( $261 \pm 101 \mu\text{M}$ ). The data collected using the fluorinated peptides suggested that the F2 (Phe7) residue may be most pivotal, but interestingly the Phe7 to alanine substitution is the tightest binding of the single mutant peptides (ie. it is least affected by the alanine substitution).

The double mutants all displayed weaker binding affinities than the single mutants, with peptides including the Phe11 $\rightarrow$ A substitution reporting the largest negative effect on the protein's ability to bind the peptide (some 4 or 5-fold reduced binding). The Phe 6,7 $\rightarrow$ A substitution resulted in a similar  $K_d$  to the single mutant peptides, whereas double substitutions involving Phe11 displayed weaker binding than both the Phe 6,7 $\rightarrow$ A and all single substituted peptides. This is an interesting observation, because it indicates primary recognition of  $\Delta$ -som by hPDI may involve the C-terminal region of the peptide. This

suggests an important role for Phe11 in  $\Delta$ -som recognition by **b'x**, which was also hinted by the single and triple fluorinated peptide data.

The shifts observed using the triple Phe to Ala substituted peptide were too small at the end of the titration to weight and fit accurately (the vast majority were smaller than 0.01 ppm). The lack of observable shifts in the spectra is highlighted in Figure 5.23, and Figure 5.24 shows the magnitude of shifts observed when compared to an equivalent concentration of  $\Delta$ -som. The small shifts observed appeared to be randomly scattered upon the protein's surface, and are similar in distribution to shifts seen in the protein when using a non-binding control peptide; helix-28 from human serum albumin (see Figures 5.25 and 5.26).

The loss of recognition of **b'x** to this peptide demonstrates that all three phenylalanine residues together are crucial for the binding interaction to occur. Substitution of any one Phe residue weakens binding affinity, showing they all in some part contribute to binding. The data reported here also suggests that Phe 11 contributes most to the binding, as its substitution to alanine has the greatest negative effect on binding in all observed peptides. In addition to this, Phe 11 was also the residue shown to be most sensitive to fluorination.



**Figure 5.23:**  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of 0.25 mM  $^{15}\text{N}$  **b'x** (blue) and 0.25 mM  $^{15}\text{N}$  **b'x** + 0.625 mM F1,2,3-Ala  $\Delta$ -som (brown).

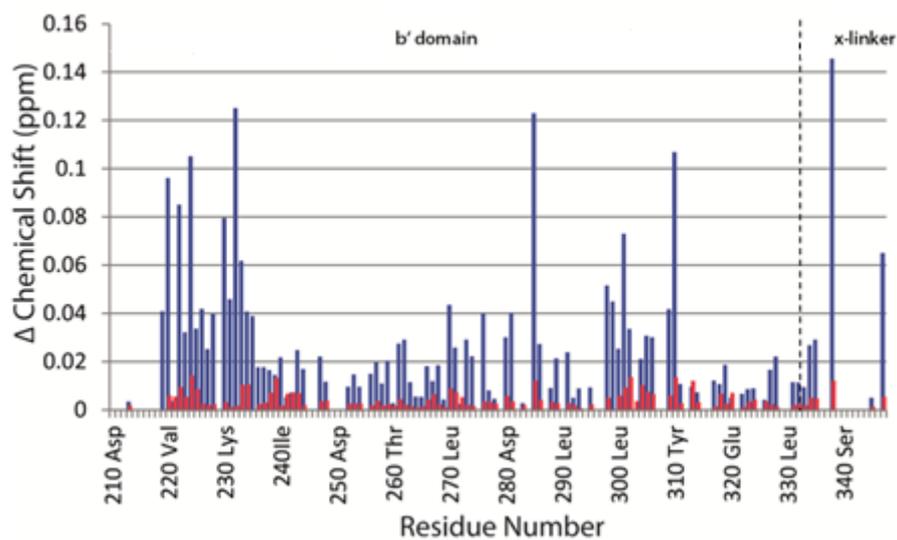


Figure 5.24: Chemical shift perturbation map of 0.25 mM  $b'x$  + 0.625 mM  $\Delta$ -som (blue) and 0.25 mM  $b'x$  + 0.625 mM F1,F2,F3-Ala  $\Delta$ -som (red).

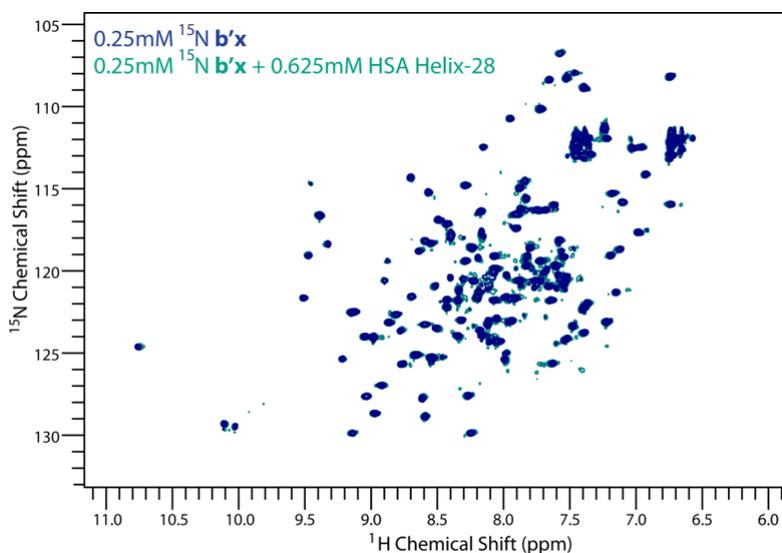


Figure 5.25:  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra of 0.25 mM  $^{15}\text{N}$   $b'x$  (blue) and 0.25 mM  $^{15}\text{N}$   $b'x$  + 0.625 mM HSA helix-28 (teal).

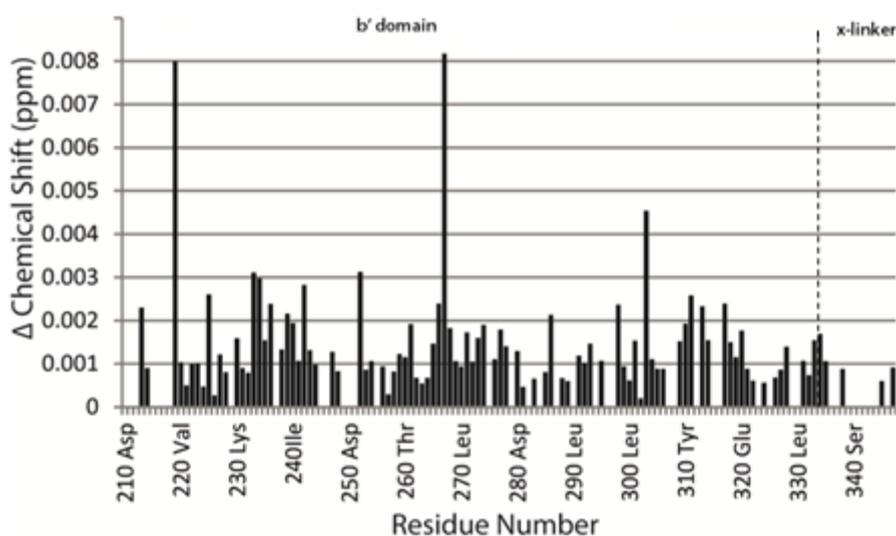


Figure 5.26: Chemical shift perturbation map of 0.25 mM *b'*x upon addition of 0.625 mM non-binding control peptide HSA helix-28.

## 5.4 Discussion

### 5.4.1 Peptide Ligand Binding Specificity of hPDI

Work in this Chapter has shown that the binding affinity of hPDI to its peptide ligand  $\Delta$ -som is in close agreement with the previously estimated affinity of 0.1-1mM (**b'**x fragment  $K_d$ :  $103 \pm 47 \mu\text{M}$ , W111F **abb'**x fragment  $K_d$ :  $73 \pm 31 \mu\text{M}$ ). It also appears binding affinity is independent of both active site redox state or proximity of other neighbouring domains.

In addition, incorporation of fluorine into the peptide has not only allowed a unique view of the binding event from a 'ligand-observed' perspective, but has also highlighted the important role the Phe residues play in this binding event. Binding affinities calculated from either the peptide or ligand viewpoint also gave highly comparable values, showing this method of selective fluorination is well suited to study ligand binding- particularly in this instance, where more traditional isotopic labelling techniques of the peptide have been met with limited success. Alanine substitution of the three Phe residues of  $\Delta$ -som abolished binding of the peptide to **b'**x, which further demonstrates that like other molecular chaperones, hydrophobic, aromatic amino acids must play a key role in hPDI's ligand recognition. It is also of note that fluorine is the only channel with which  $\Delta$ -som peptide

has been visible by our NMR spectrometer at 37°C, as all <sup>1</sup>H or <sup>13</sup>C natural abundance spectra collected using synthetic or recombinant peptide at 37°C were blank.

**<sup>19</sup>F Peptides:** Overall the <sup>19</sup>F ligand observed data gave extremely good correlation with the <sup>15</sup>N,<sup>1</sup>H-HSQC protein observed data, but with improved precision. This result may be due to the lack of background noise and thus greatly simplified spectra, or simply the increased resolution that is available when using fluorine as an NMR probe. On average the <sup>19</sup>F nucleus is surrounded by 9 electrons, and hence will report more sensitively on its given chemical environment than say a <sup>1</sup>H nucleus, which is surrounded by a single electron (Gerig, 1994).

The K<sub>d</sub> fits generated were slightly higher in affinity than for un-fluorinated peptide (non-fluorinated Δ-som K<sub>d</sub>: 103 ± 47 μM, F1,2,3 Δ-som K<sub>d</sub>: 48 ± 35 μM), as observed from the protein (**b'x**), but the values quoted are still well within the error range of un-fluorinated peptide. A comparison of protein chemical shift perturbations confirmed that fluorinated peptide appears to bind and interact with **b'x** protein in a similar manner to un-fluorinated peptide, and the direction and extent of observed chemical shift perturbations are highly comparable.

The <sup>19</sup>F ligand-observed method has several advantages over traditional uniform labelling techniques. Firstly, fluorinated peptides are easily synthesized and cost little more than non-fluorinated peptides (Peptide Synthetics quote ~£20 extra per 5 mg peptide). Additionally, this approach does not require full assignment of the peptide.

The single Phe peptides each produced unambiguous spectra with a single peak. The spectra of F1,2,3 Δ-som were perhaps more complicated, as the F1 and F2 peaks initially overlap. Attempts were made to assign each of the <sup>19</sup>F peaks of F1,2,3 Δ-som, but again, many of the <sup>13</sup>C spectra collected were blank. Spectra of double fluorinated Δ-som peptides may have been helpful to aid in the peak assignment using overlaid <sup>19</sup>F spectra, but there was no guarantee that the observed peaks would overlay with the triple fluorinated peptide (as was apparent with the single Phe peptides). Ultimately, this was not deemed necessary, as the assignments did not significantly affect the data collection or analysis. It was deemed that fluorination did not have any adverse effects on the binding interaction observed, with fluorinated peptides and protein reporting on the binding event in a similar manner.

**Alanine Substituted Peptides:** The seven alanine substituted peptides show that all three phenylalanine residues contribute to the recognition of  $\Delta$ -som by **b'x**. Whilst substitution of all three residues prevented any binding interaction from occurring, all single Phe substitutions were shown to weaken the binding affinity (albeit to different extents), with the two Phe-Ala substitutions generally weakening binding affinity further still.

The alanine substitutions also validated the results obtained by  $^{19}\text{F}$  observed spectra, by further suggesting that the c-terminal region of the peptide may exert a greater contribution to the binding event than the n-terminal region. The alanine substituted peptides indicate that of the three phenylalanine residues, it is Phe11 (F3) that has the greatest contribution to binding, as it has the greatest negative effect on binding when substituted. The  $^{19}\text{F}$  observed data also showed that this residue was most sensitive to more widespread fluorination of the peptide.

Together this work suggests that recognition of  $\Delta$ -som peptide may be mediated more by Phe residues, as opposed to Trp or Tyr residues, which is the case for PDIp ligand recognition (Ruddock et al., 2000a). This may go some way to explain how a synthetic peptide used within our laboratory, Pep-20 (YKLDKESYPVFYLFRR), binds to hPDI moderately more tightly than  $\Delta$ -som (unpublished data). It contains two Phe residues in its c-terminal region, and no tryptophan residues. The work in this Chapter could easily be extended to include titrations using a fluoro-Trp or fluoro-Tyr  $\Delta$ -som peptide, or fluoro-Phe incorporated Pep-20 peptides, to investigate this idea further.

#### 5.4.2 Proof of Binding by NOESY-HSQC

NOESY-HSQC spectra were run at a variety of field strengths and temperatures, using either **b'x** and W111F **abb'x** fragments with and without  $\Delta$ -som peptide, in attempts to identify intramolecular NOEs, for evidence of binding. This data would have been useful as it could have enabled structural restraints to be calculated for eventual generation of a ligand bound structure. Only 5 new NOE resonances were visible in **b'x** spectra with peptide at 37°C, which whilst confirming binding, was deemed a scant number of NOEs with which to proceed any further. Different running temperatures and field strengths did not improve the number of NOE resonances collected, which may in part be due to incompatible running temperatures of protein and peptide. Natural abundance spectra of  $\Delta$ -som could only be collected at temperatures lower than 10°C- any higher than this and the peptide is so conformationally active that all peaks broaden out and the spectra appear

blank. Chapter 4 showed that running protein spectra at temperatures lower than 37°C (at 25°C) lead to much broader peaks and a significant loss of resolution. Thus, NOESY spectra collected at 10°C were noticeably poorer than spectra collected at 37°C, and still lacked any extra NOEs. A ligand-bound structure would help to greatly expand our knowledge of hPDI ligand binding, but in this case it appears that NMR techniques are somewhat limited by running compatibility of ligand and protein.

### 5.4.3 Ligand Binding by $^{13}\text{C}$ Methyl Shifts

Ligand binding was also probed using overlaid methyl region  $^{13}\text{C},^1\text{H}$ -HSQCs of **b'x** protein, in the presence and absence of a binding ligand. Whilst data collected using  $^{15}\text{N},^1\text{H}$ -HSQCs could be used to accurately quantify binding affinities, each titration shows a large number of resonances that shift upon binding, of which not all are directly involved in the binding event itself (i.e. some are reporting on induced structural changes). The carbon nuclei is thought to be less influenced by these structural changes, and we hoped using  $^{13}\text{C},^1\text{H}$ -HSQCs would be a more direct reporter of ligand binding.  $^{13}\text{C},^1\text{H}$ -HSQC spectra can be collected in similar time frame to  $^{15}\text{N},^1\text{H}$ -HSQC spectra, and  $^{13}\text{C}$  labelling is becoming increasingly more affordable. This method of studying ligand binding is also gaining increased popularity for its application in studies of larger proteins, and as a preliminary screening tool (Hajduk et al., 2000), (Huth et al., 2007). Signals remain sharp and well-resolved even when using low concentrations of protein, and with recent developments in NMR sensitivity, spectra can even be collected on unlabelled proteins in a timely manner (Quintern et al., 2012). Unfortunately, **b'x** protein lacked a functional number of assignable methyl reporters in its binding site (just three surface exposed methyl residues were assigned), which limited the detail this method could provide. The data did however, provide further evidence of a ligand-induced conformational change upon binding, and methyl resonances throughout the protein core reported perturbation upon addition of a ligand, with the greatest shifts observed in core beta sheets and alpha helices.

## CHAPTER 6

# Redox Studies of W111F **abb'**x by NMR Spectroscopy

### 6.1 Introduction

The ability of hPDI to carry out its catalytic roles as an oxidoreductase and isomerase is dependent on both the redox status of the ER cellular compartment in which PDI is localised, as well as the proteins own redox potential. A redox potential, also referred to as a reduction potential ( $E'_o$ ), is a measure of a substances tendency to either gain or donate electrons, thereby becoming reduced or oxidised. In solution, a substance with a higher redox potential will have a tendency to gain electrons from another substance with a lower (less positive or more negative) redox potential. Thus, the substance with the lower redox potential will become oxidised, as it loses its electrons to the substance with a higher redox potential, which in turn will become reduced.

Redox potential is a measurement that helps to define a proteins role *in vivo*; for example thioredoxin is a strong reductant with a calculated redox potential of -230 mV (Watson et al., 2003). hPDI active site reduction potentials are somewhat less reducing by comparison, and are reported as  $E'_o$  -162 mV for **a** domain and  $E'_o$  -169 mV for **a'** domain (Chambers et

al., 2010). Whilst the ER is known to be an oxidising environment to facilitate disulphide bond formation, in this compartment hPDI is known to act as an oxidase, reductase and isomerase. This *in vivo* activity is defined not only by the proteins redox potential, but also its biochemical and kinetic ability to interact with other redox active substrates or smaller thiol containing compounds, and their relative concentrations (Sevier and Kaiser, 2002).

*In vitro*, the ability of PDI to reduce or oxidise model substrates can be controlled by adjustment of the redox buffer (Freedman, 1995). The Kent NMR group have recently shown how  $^{15}\text{N},^1\text{H}$ -HSQCs can be used to probe the redox state of different proteins in solution (Taylor et al., 2013). The NMR based method was used to calculate the reduction potential of hPDI **a** ( $E'_o -145.3 \pm 1.0$  mV) and hPDI **ab** ( $E'_o -158 \pm 2.4$  mV). This Chapter reports calculation of the redox potential of the W111F **abb'**x fragment, for comparison with the previously published data.

### 6.1.1 Redox Potential by NMR Spectroscopy

Redox potential by NMR is achieved by collecting a set of  $^{15}\text{N},^1\text{H}$ -HSQC spectra of the protein sample in redox equilibrium with a compound of known redox potential. Relative proportions of reduced and oxidised protein are calculated from the spectra and fit graphically to determine a redox potential. The published work uses a glutathione redox pair (GSH:GSSG) (Taylor et al., 2013), but the method is equally amenable to other redox pairs and this work also uses dithiothreitol ( $\text{DTT}^{\text{red}}:\text{DTT}^{\text{ox}}$ ).

The exchange rate observed by NMR between reduced and oxidised hPDI is within the slow exchange rate timescale. This means that as the ratio of oxidising agent is increased, the 'reduced' peaks in the overlaid spectra will decrease, and the 'oxidised' peaks will increase in intensity (as shown in Figure 5.1 (b)). Hence relative proportions of reduced and oxidised protein can be recorded from the NMR spectra using peak height or volume from either reduced or oxidised peaks (Taylor et al., 2013). The line broadening of resonances observed upon oxidation would complicate the dataset if both sets of peaks were monitored simultaneously, so in this instance only the assigned, reduced peaks were followed.

The peak intensities collected from each spectra are made relative to their corresponding peak in a fully reduced spectra ( $F_{\text{red}}$ ), and the relative value is used to determine an

equilibrium constant ( $K_{eq}$ ) by fitting to the Hill equation, as shown by Equations 6.1 (a) for glutathione or (b) for dithiothreitol:-

$$6.1 \quad (a) \quad F_{red} = ([GSH]^2/[GSSG]) / (K_{eq} + [GSH]^2/[GSSG])$$

$$(b) \quad F_{red} = ([DTT^{red}] / [DTT^{ox}]) / (K_{eq} + [DTT^{red}] / [DTT^{ox}])$$

Note that the equation used to generate  $K_{eq}$  is dependent upon which redox pair is used; oxidised glutathione (GSSG) yields two molecules of reduced glutathione (GSH), whereas oxidised dithiothreitol ( $DTT^{ox}$ ) yields one molecule of reduced dithiothreitol ( $DTT^{red}$ ).

The reduction potential of the protein can then be determined using the Nernst Equation (Equation 6.2):-

$$6.2 \quad \text{Nernst Equation: } E'_0 = E'_{0(\text{Reducing Agent})} - (RT/nF) \ln(K_{eq})$$

Where  $E'_0$  is reduction potential, R is the molar gas constant, T is the temperature in Kelvin,  $n$  is the number of reactive electrons in the red/ox process (in this case 2), and F is the Faraday constant.

This method of redox potential calculation is particularly useful when studying proteins with multiple redox active sites, as is the case in a multi-domain protein such as hPDI. If the  $^{15}N, ^1H$ -HSQC spectra is assigned, resonances for each redox active domain can be analysed separately to determine individual domain redox potentials.

### 6.1.2 Alternative Methods of Measuring hPDI Redox Potential

hPDI redox potentials have previously been calculated by radiolabelling and HPLC (Lundstrom and Holmgren, 1993), (Darby and Creighton, 1995). These methods 'freeze' the redox status of hPDI by acid quenching, followed by alkylation of the free thiols under denaturing conditions. Relative amounts of reduced and oxidised protein can then be determined, but if multiple redox active sites are present in the protein only an average redox potential of all sites can be calculated.

Recent work has been carried out using a differential alkylation method, whereby the free thiols trapped by acid quenching are first alkylated using iodoacetate. The alkylated sample is run reduced on SDS-PAGE, and the band gel excised. The free thiols (that would have originally been disulphide bonded upon quenching, but are now in a reduced form) are

then alkylated using 1,2  $^{13}\text{C}$  bromoacetate, and the protein is analysed by mass spectrometry (Chambers et al., 2010). Like our NMR spectroscopy method, this technique also allows individual redox potentials to be calculated for each catalytic domain of PDI.

## 6.2 Materials and Methods

### 6.2.1 NMR Sample Preparation

Monomeric protein isotopically labelled and purified as outlined in Chapter 3 was buffer exchanged using a Vivaspin 20 Centrifugal Filter (10 kDa MWCO, Sartorius) into NMR buffer (20 mM sodium phosphate, 50 mM NaCl, pH 7.0) and then concentrated down (determined by  $A_{280}$  nm absorbance) into stocks to use as required.

W111F **abb**'x protein stocks were made to 0.45 mM in NMR buffer, and reducing agents (DTT<sup>red</sup> and GSH) or oxidising agents (DTT<sup>ox</sup> and GSSG) were made to 100 mM, 10 mM and 1 mM stock concentrations. Samples were prepared using 0.25 mM W111F **abb**'x, varying ratios of oxidising: reducing agents (5 mM total concentration), 10% D<sub>2</sub>O and 0.05% sodium azide, made to a total volume of 350  $\mu\text{L}$  using NMR buffer (see Appendix 6.1 for a table of prepared samples). Prior to use, NMR buffer was degassed by sparging with nitrogen gas for two hours. All samples were made up in sterile Eppendorfs immediately prior to use and transferred to 5 mm Shigemi NMR tubes (BMS-005V, Sigma).

### 6.2.2 NMR Data Acquisition and Processing

NMR data was acquired at the University of Kent using a 4-channel, 5-amplifier Bruker Avance III 14.1T (600MHz  $^1\text{H}$ ) NMR spectrometer equipped with a 5mm QCI-F cryoprobe, unless otherwise stated. All experiments were carried out at 37°C, pH 7.0, unless otherwise stated.  $^1\text{H}$  chemical shift referencing was carried out using the position of the  $^1\text{H}_2\text{O}$  resonance at the given temperature.  $^{15}\text{N}$  chemical shifts were referenced using a spectrometer-based macro.

All 2D NMR data was processed using NMRPipe software (Delaglio et al., 1995).

### 6.2.2.1 $^{15}\text{N}$ , $^1\text{H}$ -HSQC Experiments of W111F **abb**'x

$^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra were acquired of each protein fragment as previously outlined in 4.2.2.1 using varying ratios of oxidising: reducing agent.

### 6.2.3 NMR Data Analysis

All subsequent data analysis of 2D spectra was carried out using CcpNmr Analysis software Version 2 (Vranken et al., 2005).

#### 6.2.3.1 Redox Potential of W111F **abb**'x

To measure protein redox potentials, peak volume intensities were collected from assigned peaks in spectra of W111F **abb**'x, and made relative to their corresponding peaks in the fully reduced spectra.

Data was initially collected using the GSH:GSSG redox pair, and fitted to the Hill equation defined by Equation 6.1 (a) to determine an equilibrium constant.  $F_{\text{red}}$  was then plotted against  $K_{\text{eq}}$  to determine a redox potential. Inspection of the graphs showed that when peak intensities collected using the higher concentrations of GSH ( $> 4.5$  mM) were made relative to a fully reduced spectra (collected using 5 mM  $\text{DTT}^{\text{red}}$ ), peaks appeared to 'plateau' at a  $F_{\text{red}} \sim 0.6$ . These graphs highlighted that the glutathione system had reached its reducing ability at  $\sim 60\%$  reduced, and hence could not fully reduce W111F **abb**'x. This result had not been observed in the laboratory when using hPDI fragments **a** and **ab**. Consequently an additional redox pair were chosen to collect data of the protein in a more reduced state. Since DTT has already been shown to be capable of fully reducing hPDI, and its oxidised form is readily available, it was chosen as a suitable alternative. It has the added benefit of being only moderately more reducing than glutathione, such that the equilibria of both redox pairs can be shifted to give buffers of the same redox potential (see table in Appendix 6.1).

By using the Nernst Equation (Equation 6.2) to calculate individual redox potentials of each sample, data points were fitted using reduction potential along the x-axis instead of  $K_{\text{eq}}$ . This allowed for data collected using different redox pairs (either dithiothreitol or glutathione) to be directly compared on the same graph.

### 6.2.3.2 Graphical Fits

All graphical fits were made using Kaleidagraph software version 4.1 (Synergy Software). The curve fit equation can be found in Appendix 6.2.

## 6.3 Results

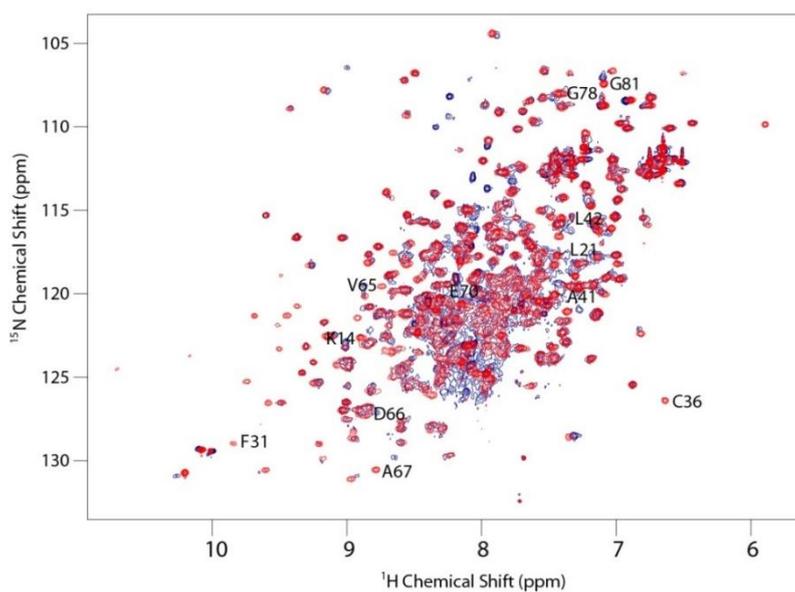
*There were 19 residues of W111F **abb**'x that could be tracked accurately throughout the  $^{15}\text{N}, ^1\text{H}$ -HSQC series to report on protein redox state. A representative sample of graphs are shown in this section, and additional raw data can be found in Appendix 6.2.*

### 6.3.1 Equating Redox Potential of W111F **abb**'x by $^{15}\text{N}, ^1\text{H}$ -HSQC

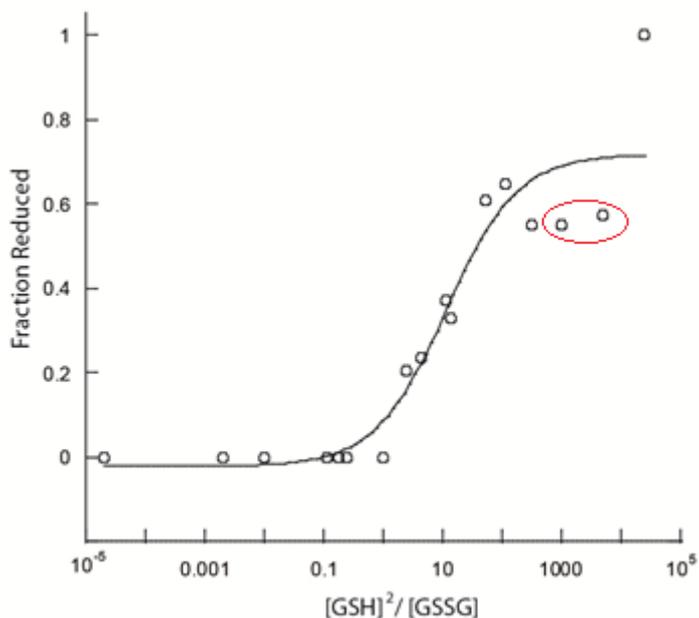
The redox potential of W111F **abb**'x was first attempted from practical data using the careful titration of differing ratios of the redox pair GSH:GSSG ( $E'_0$  -240 mV). Control spectra using 5 mM GSSG (fully oxidised) and 5 mM DTT<sup>red</sup> (fully reduced) were collected. There were 19 residues that could be tracked accurately throughout the redox series, and 12 example tracked resonances are highlighted in Figure 6.1.

All reduced peak volume intensities collected from each spectra were made relative to the corresponding resonance in the fully reduced spectra. In addition to the two control spectra, 15 data points were collected using differing ratios of GSSG:GSH, that totalled 5 mM red/ox agent.

An overlay of protein spectra collected using 5 mM GSH and 5mM DTT<sup>red</sup> showed that the protein appeared in a mixed redox state when 'reducing' using glutathione, with various peaks only present in oxidised spectra still being present in the 5 mM GSH spectra. This was not an expected result, as our laboratory have previously shown that the hPDI fragments **a** (at 0.065 mM) and **ab** (at 0.3 mM) are fully reduced by 5 mM GSH (Taylor et al., 2013). It has been suggested in the literature that there is cross-talk between the individual domains of hPDI- and **ab** fragment was shown by our laboratory to be moderately more reducing than **a** is as a single domain. It appears that W111F **abb**'x is more reducing still, such that GSH is no longer able to fully reduce the active site of **a** domain.



**Figure 6.1:** Overlaid  $^{15}\text{N}, ^1\text{H}$ -HSQC spectra of 0.25 mM  $^{15}\text{N}$  W111F **abb**'x reduced using 5 mM  $\text{DTT}^{\text{red}}$  (red) and oxidised using 5 mM GSSG (blue). Sample graph fits seen in Figure 6.3 are highlighted on the spectra.

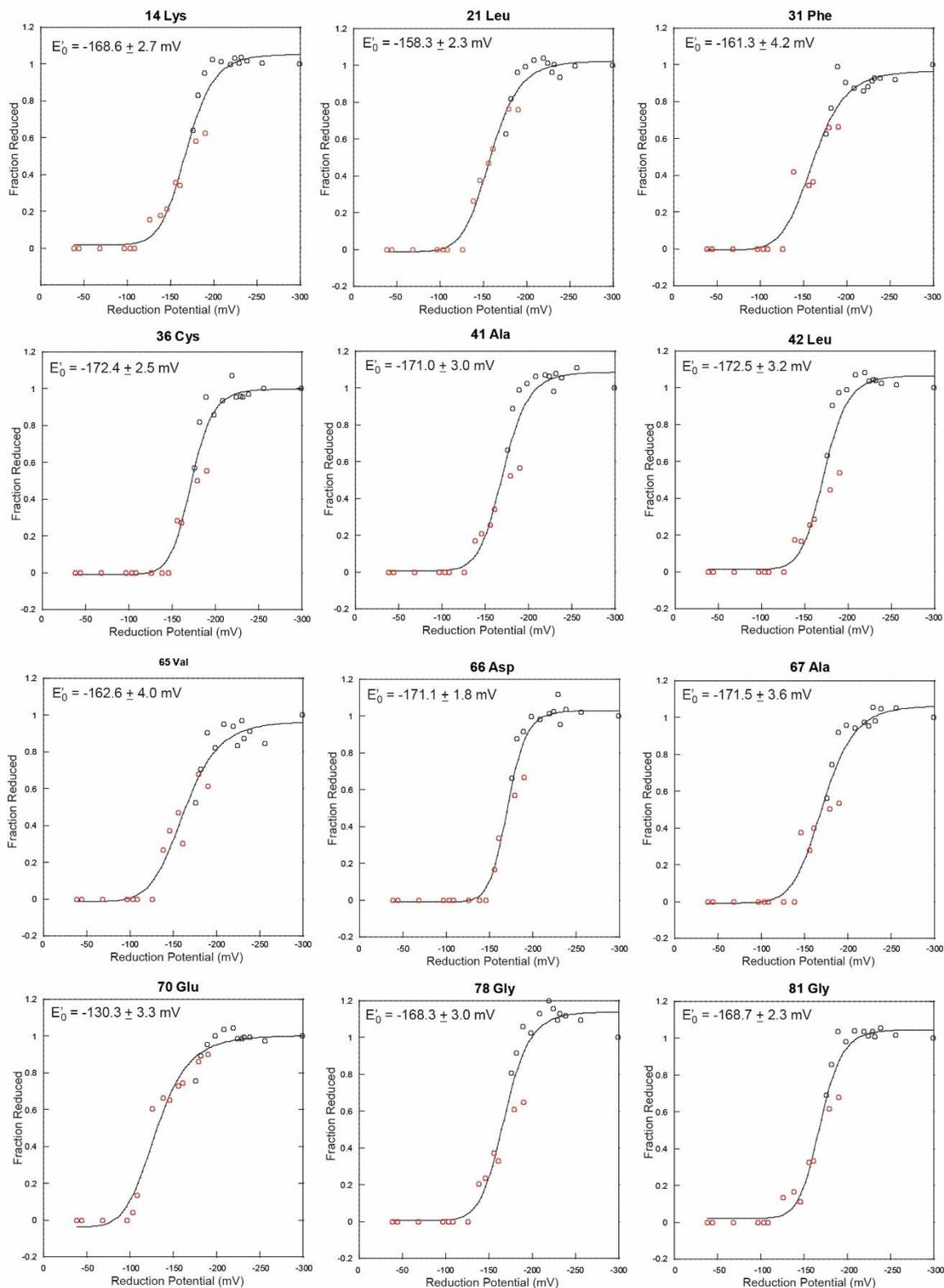


**Figure 6.2:** Fraction reduced plot of  $^{78}\text{Gly}$ , fit using GSH:GSSG redox pair relative to spectra reduced by DTT. This Hill plot shows that 5 mM GSH is only able to reduce around 60% of W111F **abb**'x, relative to fully reduced spectra collected using 5mM  $\text{DTT}^{\text{red}}$ . The  $\text{DTT}^{\text{red}}$  peak is plotted as 4.999 mM GSH: 0.001 mM GSSG for comparison with GSH: GSSG data points. Data points where it was deemed the reducing ability of GSH had been reached are circled in red.

Attempts to fit the data to Hill plots when resonances were made relative to the 5 mM DTT<sup>red</sup> spectra showed that resonances in samples using >4.95 mM GSH seemed to plateau around 60% fraction reduced (see Figure 6.2). Higher total concentrations of reducing: oxidising agents were trialled (10 mM, 20 mM and 30 mM total concentrations), but the protein was not stable in such a great excess of redox agent. Instead, DTT was picked as an alternative red/ox agent, as it has a moderately stronger reducing ability ( $E'_o$  - 330 mV), and is also capable of fully reducing hPDI. Hence additional spectra were collected using differing ratios of the DTT<sup>red</sup>:DTT<sup>ox</sup> redox pair (5 mM total concentration).

Dithiothreitol is a moderately stronger reducing agent, and it was not possible to fully oxidise W111F **abb'**x using DTT<sup>ox</sup>. To enable data collected using both the glutathione and DTT data, the reduction potential of each individual experiment was calculated using the  $K_{eq}$  of each experiment (see Equations 6.1 (a) and (b)) and the Nernst equation (Equation 6.2). Any glutathione collected data points that showed the reducing ability of GSH had been reached were removed from the fits (see two data points shown in red in Figure 6.2).

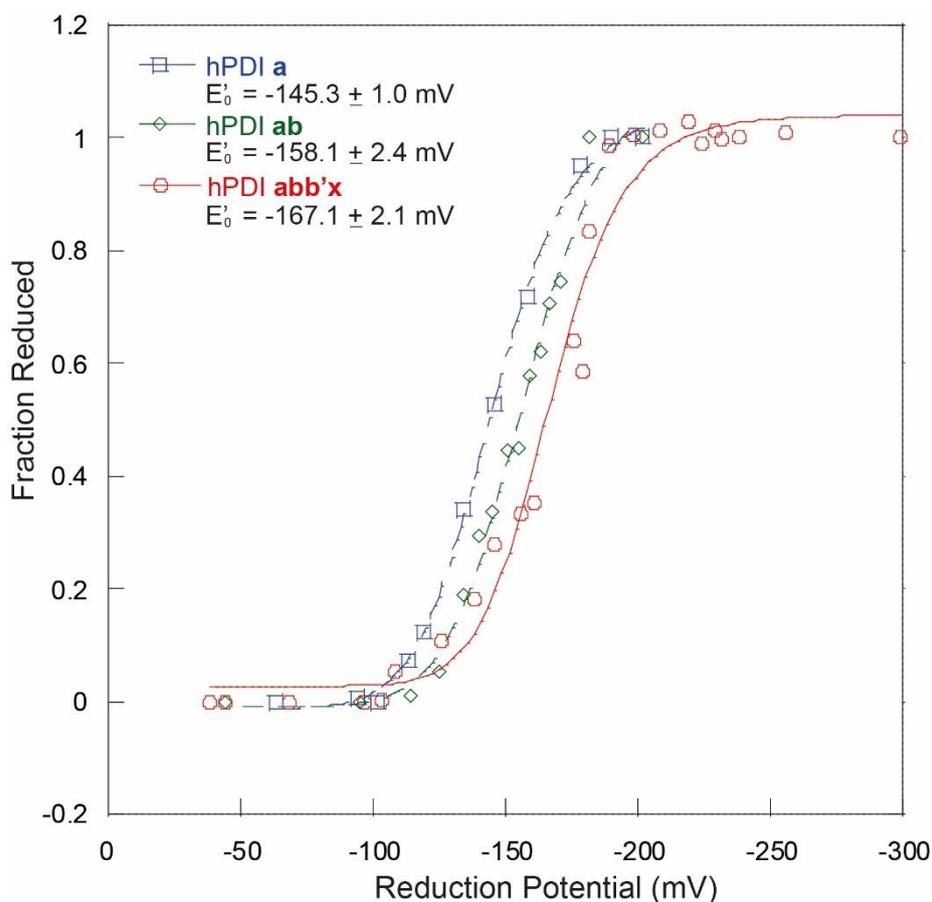
An additional 11 data points were collected using the DTT<sup>red</sup>:DTT<sup>ox</sup> pair, and example Hill plot fits of the data are shown in Figure 6.3. An average redox potential of  $E'_o$  -167.1  $\pm$  2.1 mV was calculated for W111F **abb'**x. This value is moderately more reducing than the values calculated for both **a** domain ( $E'_o$  -145.3  $\pm$  1.0 mV) and **ab** fragment ( $E'_o$  -158  $\pm$  2.4 mV), collected at 25°C using the same method (Taylor et al., 2013). The value is however highly comparable to the value published by the Bulleid group for the **a** domain of full-length hPDI ( $E'_o$  -162.7  $\pm$  2.9 mV), that was determined using a chemical modification and mass spectrometry method as outlined in 6.1.2 (Chambers, Tavender et al. 2010).



**Figure 6.3: Redox potential graphs of W111F *abb*'x as monitored by  $^{15}\text{N},^1\text{H}$ -HSQC, fit using the Hill Equation.** Data points in black were collected using the DTT<sup>ox</sup>:DTT<sup>red</sup> redox pair and data points in red were collected using the GSSG:GSH redox pair. The curves fit to give an average redox potential of  $-167.1 \pm 2.1$  mV.

## 6.4 Discussion

Overall this work shows the  $^{15}\text{N}, ^1\text{H}$ -HSQC method of determining redox potential is both reproducible and equally valid when different red/ox pairs are used. The redox potential of **a** domain as calculated here using W111F **abb**'x also closely agrees with the value of **a** domain as published by the Bulleid group for full-length hPDI ( $E'_0$   $-162.7 \pm 2.9$  mV). An interesting observation however is how the redox potential quoted here differs to values generated for other hPDI fragments containing **a** domain, as calculated within our laboratory using the same method. The raw data from fits of hPDI **a** and **ab** fragments (as published by Taylor et al. 2013) have been adapted to fit the same axis as the W111F **abb**'x data given here, and an overlay of all averaged datapoints is shown in Figure 6.4.



**Figure 6.4:** Redox potential graph of hPDI *a* (blue), hPDI *ab* (green) and hPDI W111F *abb*'x (red) as monitored by  $^{15}\text{N}, ^1\text{H}$ -HSQC and fit using the Hill Equation. hPDI *a* and *ab* data was generated using the same method as published by Taylor et. al (2013).

It appears that the redox potential of **a** domain becomes progressively more reducing upon addition of the extra domains. Note that the shape and gradient of each curve does not change for each of the hPDI fragments, but the curves merely shift towards a more reducing redox potential (read from the graph as  $0.5 F_{red}$ ). This suggests that additional domains are not destabilising the reduced or oxidised forms of the **a** domain, but the **b** domain and **b'** ligand binding domain are clearly having an effect on the redox potential.

It was previously thought that like yPDI, the two active sites of hPDI must have differing redox potentials. In yPDI, the Ero1 $\alpha$  yeast homolog, Ero1p, preferentially oxidises the more reducing **a** domain of yPDI (yPDI active site of **a**:  $E'_o$  -194 mV, active site of **a'**:  $E'_o$  -164 mV), allowing the protein to exhibit oxidase activity with its **a** domain and isomerase activity with its **a'** domain (Vitu et al., 2010). This phenomena is reversed in humans, with Ero1 $\alpha$  preferentially oxidising the **a'** domain over the active site in **a** (Baker et al., 2008),(Wang et al., 2009). One might therefore infer that it is the **a'** domain of hPDI that has the more reducing redox potential, as it is preferentially targeted by Ero1 $\alpha$ . However, the work by the Bulleid group suggests that the two active sites of hPDI are actually very similar in their reducing abilities (hPDI **a** active site :  $E'_o$  -162.7  $\pm$  2.9 mV, hPDI **a'** active site :  $E'_o$  -169.4  $\pm$  2.3 mV), and recent work by the Wang group also supports this. They instead believe the different abilities of the active sites are a result of asymmetric recognition of Ero1 $\alpha$  to the **b'xa'** fragment of hPDI, which causes specific oxidation of **a'** over the **a** domain active site (thereby allowing the **a** domain to primarily carry out isomerase activities, and the **a'** domain to carry out oxidase activity) (Zhang, Niu et al. 2014). They speculate that yPDI differs to hPDI in this manner, as yeast cells express far fewer disulphide-bonded proteins, thus regulated isomerase activity of yPDI is not as critical as it is for hPDI. However, this model of hPDI reduction/oxidation cannot explain why the redox potential of the active site in **a** domain might change upon addition of the **b** or **b'** domain, as has been published previously and reported in this thesis.

In addition to the work shown here, preliminary studies carried out within our laboratory using hPDI **abb'xa'** show that the active sites of **a** and **a'** have significantly different redox potentials (active site of **a**:  $E'_o$  -180.7  $\pm$  12mV, active site of **a'**:  $E'_o$  -219  $\pm$  8 mV, unpublished data). This could mean the yPDI and hPDI methods of redox control are much more analogous than is currently believed. Clearly, further work should be carried out to more accurately determine the redox potentials of the **a** and **a'** domains of full length hPDI,

ideally together with calculations of **a'** as part of the **b'xa'c** and **xa'c** fragments (individual **a'** domain would also be advantageous, but as a single domain gives exceedingly poor spectra). It is with regret that this work was not possible within the timescale of this thesis.

## CHAPTER 7

# Recombinant Fluoroindole Incorporation of W111F **abb'**x

### 7.1 Introduction

The work in Chapter 5 showed synthetic peptide fluorination is a technique well suited to study peptide ligand binding of hPDI fragments. The fluorination resulted in only modestly tighter  $K_d$  affinities, which were collected within a similar timescale to  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC data, but with much faster analysis and greater data precision. The  $K_d$  values equated also had a greater degree of accuracy, so as a laboratory we were keen to further develop our use of fluorine probes to study hPDI.

We recently published work that shows **b'**x protein is amenable to fluorination (Curtis-Marof et al., 2014), using a simple and inexpensive recombinant labelling technique recently developed by Crowley et al. (Crowley et al., 2012). Recombinant fluorine labelling allowed the ligand binding event to be observed using 'protein-observed' 1D  $^{19}\text{F}$  spectra and generated a  $K_d$  of  $23 \pm 4 \mu\text{M}$  (calculated at  $25^\circ\text{C}$ ).

This chapter expands upon the published work, to further study the W111F **abb'**x fragment of hPDI. It demonstrates the recombinant fluorine labelling of the protein fragment at its

two key tryptophan residues;  $^{35}\text{Trp}$ , that resides immediately adjacent to the active site of **a** domain, a reporter residue of protein redox state, and  $^{347}\text{Trp}$ , that is the 'capping' and ligand binding reporter of the **x-linker** region. Small preparative grows using 60 mg/L of each commercially available fluoroindole (4-F, 5-F, 6-F and 7-F) show that 6-F-indole gives greatest expression of W111F **abb'**x. This was in contrast to our work carried out using the **b'**x fragment of hPDI, where 5-F-indole gave greatest protein yields (and use of 6-F-indole lead to some 10-fold poorer expression), suggesting that fluoroindole incorporation must be employed on a protein by protein basis.

This chapter also demonstrates the use of  $^{15}\text{N}$  6-F-Trp W111F **abb'**x to further probe ligand binding and redox state, by both  $^{15}\text{N},^1\text{H}$ -HSQC and  $^{19}\text{F}$  1D NMR spectra. In addition, a recent paper by Wang et al. suggested the active site of **a** domain may come much closer to the ligand binding site (when the protein is modelled in simulations) than any of their crystal structures suggest (Yang et al., 2014), and attempts to corroborate this idea by 1D and experimental  $^{19}\text{F}$ - $^{19}\text{F}$  NMR are also reported.

### 7.1.1 Uses of Fluorine in Biochemistry

**Drug Development:** Fluorine is extensively exploited by medicinal chemists for the unique properties and improved bioavailability or affinity it can confer to a compound. Despite its incredibly low uptake in nature, fluorine is now routinely incorporated into drugs, with over 20% of current pharmaceuticals containing at least one fluorine atom (Yamazaki et al., 2009). Fluorine can be used to modulate the basicity or lipophilicity of a compound, increasing membrane permeability or absorption. More commonly, substituents of fluorine are specifically incorporated to 'block' metabolically labile sites of a drug, helping to increase its metabolic stability and half-life within the body (Bohm et al., 2004). Fluorine's high electronegativity means it also has a powerful effect on the acidity or basicity of nearby functional groups, promoting tighter binding of a compound to its intended target, as well as increasing the overall stability and reactivity of it. The most historic use of fluorine in drug development is perhaps the development of fluoroquinolones, where addition of fluorine at carbon-6 increases the antibiotics potency more than 10-fold, and the minimum inhibitory concentration is reduced 100-fold (Andersson and MacGowan, 2003).

**Structure/Function-based Protein Studies:** The incorporation of fluorinated amino acids into proteins to probe their structure and function is now heavily employed by scientists, with the topic well reviewed by Salwiczek et al. (Salwiczek et al., 2012). Fluorine makes for an attractive probe- it is not normally present in biological samples so has no competing background signal, and sterically the fluorine atom is a close mimic for hydrogen; fluorine is the next smallest atom after hydrogen, so a H-F or CH<sub>3</sub>- CH<sub>2</sub>F swap only occupies a 20% larger molecular volume (Yamazaki et al., 2009).

The incorporation of fluorine has proven particularly useful in the study of hydrophobic protein cores, where initially fluorination works to increase hydrophobic interactions and therefore stabilises secondary structure, with successive fluorination causing destabilisation of the native secondary structure, either by steric hindrance or over-packing of the hydrophobic core (Horng and Raleigh, 2003), (Salwiczek and Koksche, 2009).

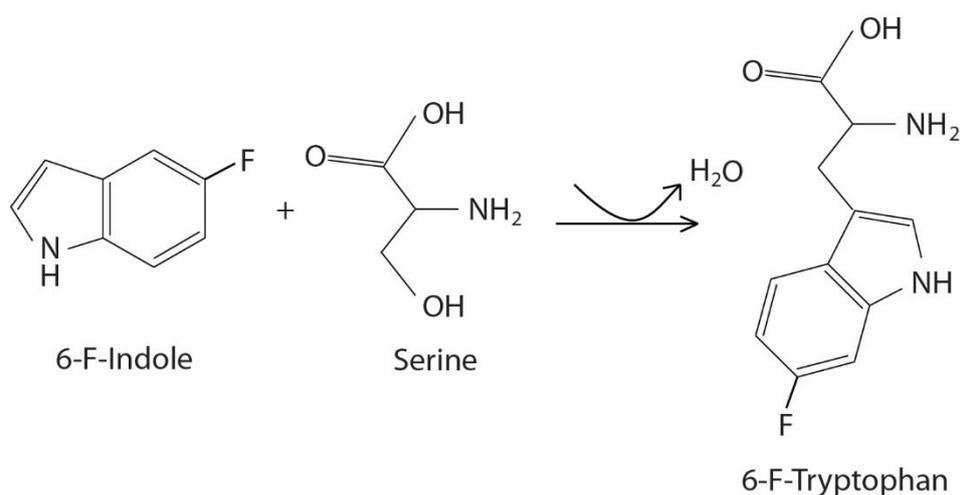
There are a large number of fluorinated amino acids available for incorporation into peptides by solid-state synthesis or semi-synthetically into larger polypeptides by native chemical ligation. Cell free expression systems can also be used to produce fluorinated protein, a method that works by exploiting the anticodon-codon recognition between mRNA and tRNA (coupling the corresponding 3'-terminus of a tRNA molecule to a fluorinated analogue results in the incorporation of fluorine into the polypeptide chain (Chapeville et al., 1962)). Site specific incorporation is also possible by use of orthogonal aminoacyl-tRNA synthetase-tRNA pairs that can incorporate non-canonical amino acids at amber 'stop' codons (Noren et al., 1989). This method of 'expanding the genetic code' has been intensively developed and it is now possible to incorporate non-canonical residues in bacterial (Neumann et al., 2008)), yeast (Hancock et al., 2010) and mammalian cell cultures (Mukai et al., 2008), and more recently has been made possible in multicellular organisms such as *C. elegans* (Greiss and Chin, 2011), (Parrish et al., 2012) and *D. melanogaster* (Bianco et al., 2012).

### **7.1.2 Recombinant Protein Fluorination using *E. coli***

Recombinant fluorine labelling of proteins is historically done in *E. coli* using auxotrophic strains that are not able to synthesise their own aromatic amino acids (Broos et al., 2003), or by using specific inhibitors of the aromatic synthesis pathway- indoleacrylic acid blocks tryptophan synthase (Matchett, 1972), and glyphosate blocks the Shikimate pathway

(Holländer and Amrhein, 1980). With these methods, supplementation with fluorinated aromatic amino acid analogues to the media consequently leads to their incorporation into the over-expressed protein. However, disruption of such an integral biosynthetic pathway can compromise the cells, leading to a build-up of metabolic intermediates, or more simply by causing secondary nutritional requirements (Whittaker, 2007). Equally, the fluorinated amino acid analogues and the inhibitors themselves can be costly to supplement.

A recent method published by Crowley et al. (Crowley et al., 2012) is able to circumvent these problems, and works by a straightforward incorporation of fluoroindole into the growth media shortly before induction. To synthesise tryptophan, *E. coli* adds indole to a serine residue, in a condensation reaction catalysed by tryptophan synthase. This enzyme is able to accommodate a wide range of indole 'mimics', including several fluorinated analogues (Ngo et al., 2007). Figure 7.1 demonstrates this reaction, showing the synthesis of 6-F-Trp from serine and 6-F-indole.



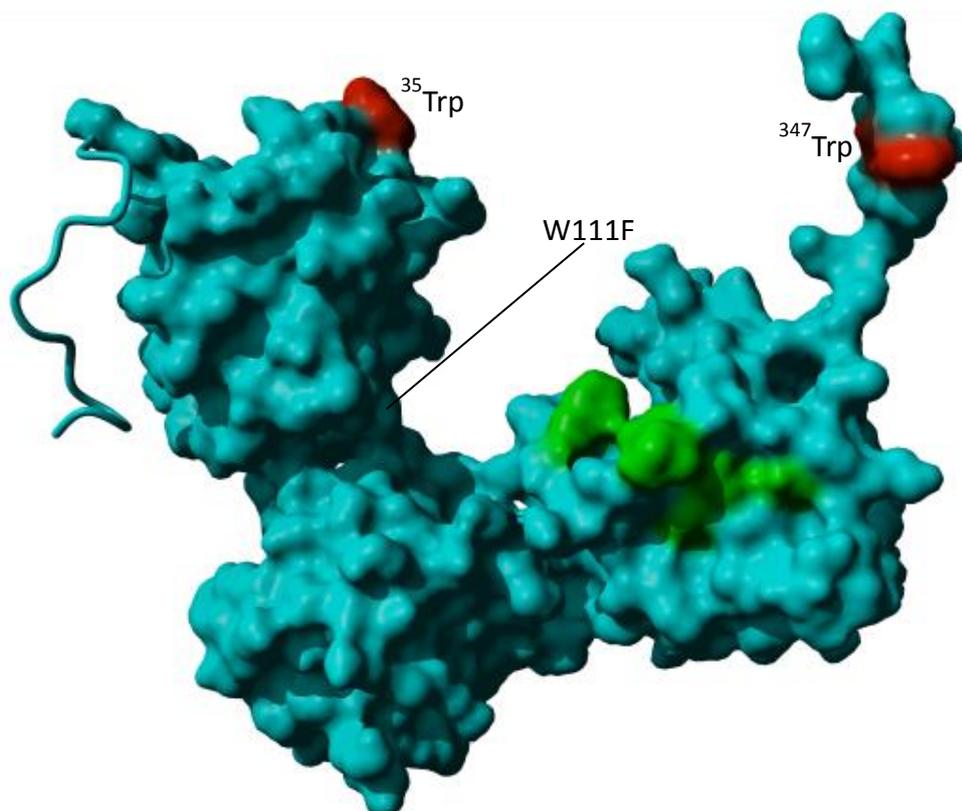
**Figure 7.1: 6-F-Tryptophan synthesis from serine and 6-F-indole.** The reaction is catalysed by tryptophan synthase in a condensation reaction.

The developed method does not require special auxotrophic cell strains or inhibitors, and has the added benefit of being much more cost-effective; gram for gram 5-fluoroindole is fifty times cheaper than 5-fluorotryptophan. Previous work by our laboratory has shown

incorporation of 5-fluoroindole into 5-F-Trp labelled **b'**x protein to be in excess of 80% (Curtis-Marof et al., 2014), proving the method as a simple and effective  $^{19}\text{F}$  labelling strategy for NMR studies.

### 7.1.3 Fluorination of W111F **abb'**x

There are two sites in W111F **abb'**x that should incorporate the fluoroindole into fluoro-Trp:  $^{35}\text{Trp}$ , a conserved residue of unknown function, that lies immediately adjacent to the active site in **a** domain, and  $^{347}\text{Trp}$ , the 'capping' reporter present in the **x-linker** region (see Figure 7.2).



**Figure 7.2:** The *abb'*x portion of the crystal model of reduced *abb'xa'* (4EKZ.pdb) highlighting the two Trp residues to be fluorinated (red), and the position of the W111F mutation. The ligand binding site is coloured green, and the rest of the molecule is coloured teal.

The W111F substitution is also highlighted in the Figure to demonstrate an additional advantage of using the W111F fluorescence mutant in place of WT **abb**'x. The  $^{111}\text{Trp}$  residue is buried within the protein core, and unpublished studies within the laboratory have shown that fluorination at this site may disrupt the stability of the protein (unpublished observation).

## **7.2 Materials and Methods**

### **7.2.1 Recombinant Protein Expression**

The method of recombinant fluorine incorporation as outlined by Crowley et al. requires 60 mg/L fluoroindole be added to the culture immediately prior to induction. Therefore, cells were grown initially as stated in 3.3.2 until an  $A_{600}$  of 0.7 AU was reached. Next, fluoroindole from a 200x stock (12 g/L stock, in 100% DMSO) was added to each flask to reach a final concentration of 60 mg/L, and incubated at 37°C for 15 minutes with 200 rpm shaking. Cells were then induced as normal using 1 mM IPTG and incubated overnight at 27°C with 200 rpm shaking (as per standard protocol). Pelleted cells were then lysed and purified as outlined in 3.2.4.

### **7.2.2 Mass Determination and Isotope Incorporation**

The molecular weight of  $^{19}\text{F}$ ,  $^{15}\text{N}$  isotopically enriched monomeric W111F **abb**'x protein was determined using Electrospray Mass Spectrometry, performed by Mr Kevin Howland, Biomolecular Science Facility Manager, University of Kent.

### **7.2.3 NMR Sample Preparation, Data Acquisition and Processing**

All  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC and  $^{19}\text{F}$  1D NMR data was acquired and processed at the University of Kent as previously outlined in Chapter 5.

### **7.2.4 NMR Data Analysis**

All subsequent data analysis of 1D spectra was carried out using Bruker Topspin software, and all 2D spectra analysis was carried out using CcpNmr Analysis software Version 2 (Vranken, Boucher et al. 2005).

#### 7.2.4.1 <sup>19</sup>F Protein Assignment

<sup>19</sup>F spectra of reduced 6-F-Trp **abb'x** protein yielded two distinct fluorine resonances that overlaid well with previous spectra collected within the laboratory of other fluorinated hPDI fragments with 5-F-Trp incorporation. These previously assigned spectra permitted the straightforward assignment of 6-F-Trp resonances.

#### 7.2.4.2 Chemical Shift Perturbation Mapping

All <sup>15</sup>N, <sup>1</sup>H-HSQC and <sup>19</sup>F 1D chemical shift perturbations were calculated as previously outlined in 5.2.3.2.

#### 7.2.4.3 Redox Potential of 6-F-Trp <sup>15</sup>N W111F **abb'x**

**Using <sup>15</sup>N data:** Reduction potential by <sup>15</sup>N was measured from peak volume intensities, and calculated as previously outlined in 6.2.3.1.

**Using <sup>19</sup>F data:** Reduction potential by <sup>19</sup>F was plotted using the <sup>35</sup>Trp integrated peak volume from each redox experiment, which was made relative to its corresponding peak in the fully reduced spectra and plot against the reduction potential as calculated by the Nernst Equation (Equation 6.2). Initial plots of the data from each redox pair (either glutathione or DTT) showed that in contrast to <sup>15</sup>N collected data of **abb'x**, the points did not readily fit to the same curve. Therefore, each redox pair was fitted to individual Hill Plots to calculate a reduction potential for each separate glutathione or DTT redox pair.

#### 7.2.4.4 Graphical Fits

All graphical fits were made as previously outlined in 6.2.3.2.

## 7.3 Results

### 7.3.1 Expression of 6-F-Trp <sup>15</sup>N W111F **abb'x**

As previously reported in Chapter 3, *E. coli* BL21 (DE3) pLysS cells with the W111F **abb'x** insert show normal growth times in minimal media, with a total cell increase upon standard overnight induction of around 170%. To assess the effect DMSO (the indole stock solvent), and each indole might have upon protein expression and cell growth, 100 mL test

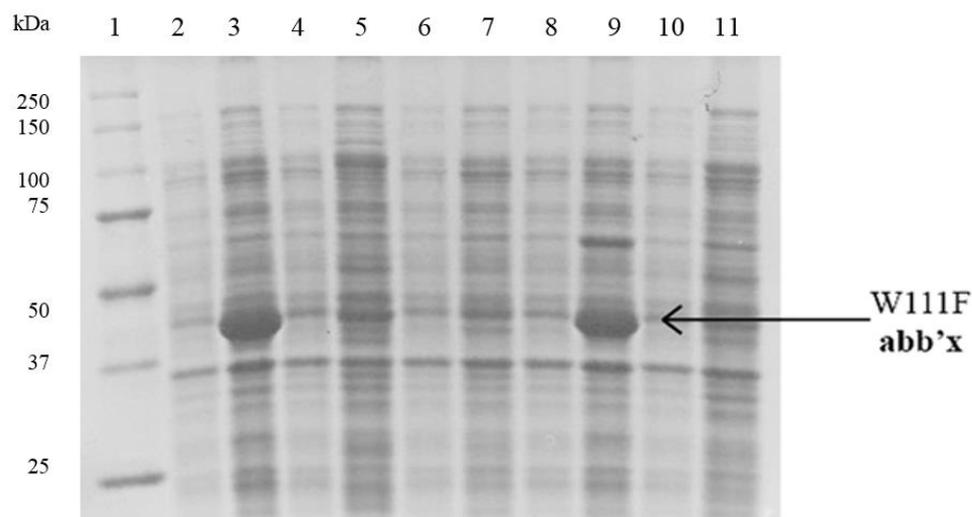
cultures were grown, with addition of the selected indole prior to induction. Pre and post induction  $A_{600}$  nm readings were recorded and pellets ran on reducing SDS-PAGE (see Table 7.1 and Figure 7.3).

**Table 7.1: Pre and post induction  $A_{600}$  nm readings of test indole cultures.**

Expression conditions	Pre-Induction $A_{600}$	Post-Induction $A_{600}$ (% cell increase)
$^{15}\text{N}$ labelling	0.686	1.862 (171%)
$^{15}\text{N}$ labelling + DMSO	0.695	1.831 (163%)
$^{15}\text{N}$ labelling + indole	0.720	1.783 (148%)
$^{15}\text{N}$ labelling + 4-F-indole	0.670	1.608 (140%)
$^{15}\text{N}$ labelling + 5-F-indole	<b>0.668</b>	<b>1.024 (53%)</b>
$^{15}\text{N}$ labelling + 6-F-indole	0.677	1.700 (151%)
$^{15}\text{N}$ labelling + 7-F-indole	0.686	1.646 (140%)

0.5% DMSO (the indole solvent) and 4, 6, or 7 F-indoles had minimal effect on % cell increase upon induction, whilst 5-F-indole appeared to have an inhibitory effect on cell growth after its addition, as shown by the markedly reduced % cell increase.

With the exception of 5-F-indole, the cultures showed only a minor drop in total percentage cell increase (reduction of 8% upon addition of DMSO, and reduction of 23% upon addition of indole), thus the addition of these indoles did not appear to adversely affect cell growth. In contrast, 5-F-indole appeared to have an inhibitory effect upon cell growth, with an observed overnight cell increase of only 53%, one third of the cell growth observed when un-fluorinated indole was added to the culture (see Table 7.1). This is an interesting observation, as previous work by the laboratory using the same expression system has shown that incorporation of 5-F-indole actually has no effect on cell growth, and of the four fluoroindoles gives greatest expression of hPDI **ab** and **b'**x fragments.



**Figure 7.3: 10% reducing SDS-PAGE analysis of  $^{19}\text{F}$  W111F **abb**'x expression in the presence of indole.** Lane 1: Precision Plus protein marker, lanes 2 & 3: pre and post induction pellets (+ unfluorinated indole), lanes 4 & 5: pre and post induction pellets (+ 4-F-indole), lanes 6 & 7: pre and post induction pellets (+ 5-F-indole), lanes 8 & 9: pre and post induction pellets (+ 6-F-indole), lanes 10 & 11: pre and post induction pellets (+ 7-F-indole).

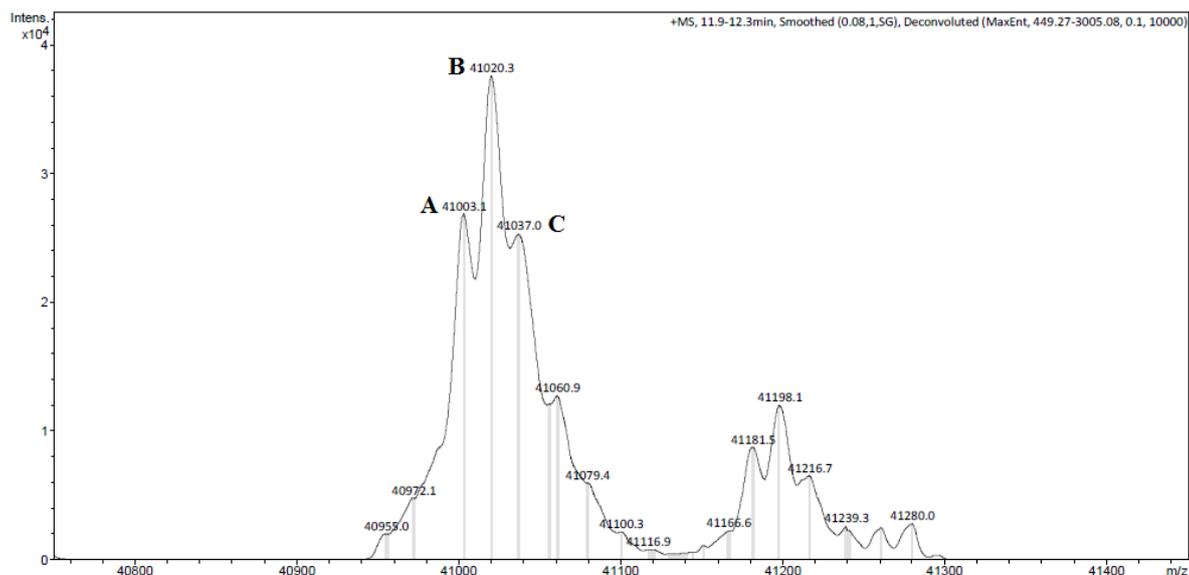
Figure 7.3 demonstrates the protein expression observed in pre and post induction pellets. It appears that only the 6-F-indole culture expressed an amount of W111F **abb**'x protein comparable to the un-fluorinated indole control culture. The over-expressed protein band was noticeably lower in intensity for the 4-F-indole, 5-F-indole and 7-F-indole test cultures. As a result, 6-F-indole was chosen for incorporation into W111F **abb**'x.

### 7.3.2 Mass Determination and Isotopic Incorporation

The molecular weight of  $^{15}\text{N}$  W111F **abb**'x as determined by mass spectrometry (see Figure 3.8) is 41004.4 Da (98.9%  $^{15}\text{N}$  incorporation). There are two potential sites of fluorine incorporation within the protein, and the hydrogen to fluorine swap will result in a mass increase of +18 Da at each site. Mass spectrometry analysis of 6-F-Trp  $^{15}\text{N}$  W111F **abb**'x is shown in Figure 7.4 (a). It is clear from the spectra that the protein purifies as a heterogenous mixture of fluorinated and un-fluorinated protein. Deconvolution of the major peaks identified three different protein species (see Figure 7.4 (b)), the largest of which (41,020.5 Da) corresponds to a singly fluorinated species (~40% relative intensity). Unfortunately it is not possible from mass spectrometry alone to state which of the two

sites has been fluorinated. The peak at 41,003.8 Da corresponds to un-fluorinated protein (~30% relative intensity), and the peak at 41,039.4 Da corresponds to double fluorinated protein (~30% relative intensity).

(a)



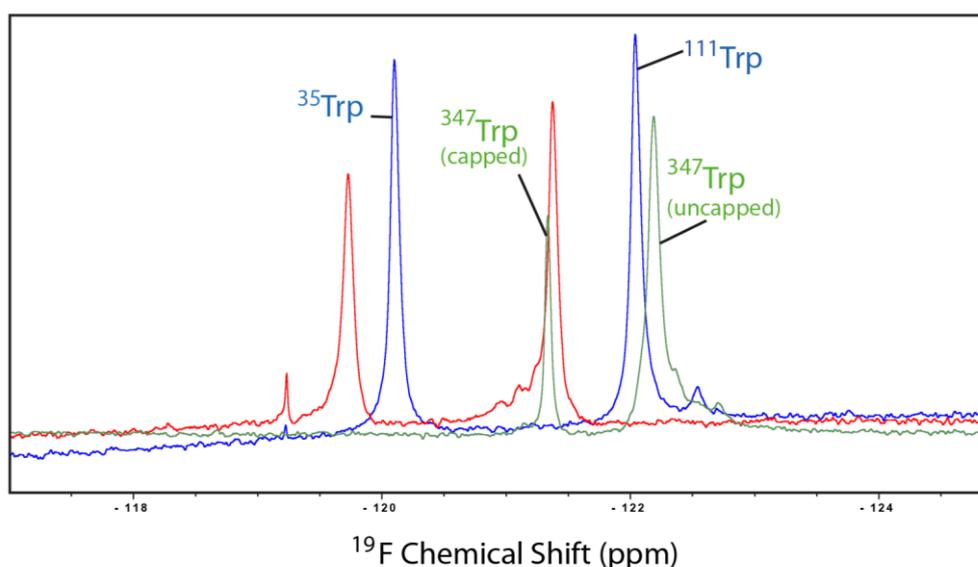
(b)

Peak	MW (Da)	Mass Increase (Da)	Sequence Identity	Relative Intensity
A	41003.1	N/A	<sup>15</sup> N W111F <b>abb</b> 'x (un-fluorinated)	30%
B	41020.3	+ 17.2	<sup>15</sup> N, <sup>19</sup> F W111F <b>abb</b> 'x (single fluorination)	40%
C	41037.0	+ 33.9	<sup>15</sup> N, <sup>19</sup> F W111F <b>abb</b> 'x (double fluorination)	30%

**Figure 7.4: Mass spectrometry analysis of 6-F-Trp <sup>15</sup>N W111F **abb**'x (a). Corresponding MW and relative intensities of each labelled peak (A, B and C from (a)) are shown in (b).**

To gain more information on the relative amounts of fluorination at each Trp residue, <sup>19</sup>F 1D spectra of the protein was also collected. The spectra was first overlaid with other spectra collected within the laboratory of two 5-F-Trp labelled hPDI fragments- **ab** and

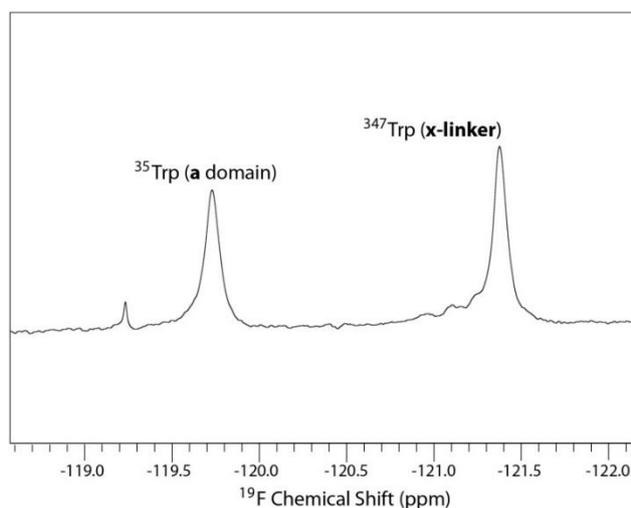
**b**'x, to attempt identification of each fluorine peak of 6-F-Trp  $^{15}\text{N}$  W111F **abb**'x (see Figure 7.5). The spectra of 5-F-Trp labelled fragments overlaid well with 6-F-Trp  $^{15}\text{N}$  W111F **abb**'x, and suggested the peak at -119.73 ppm corresponds to  $^{35}\text{Trp}$  of **a** domain, and the peak at -121.37 ppm corresponds to  $^{347}\text{Trp}$  of the **x-linker** region. The peak at -119.23 ppm is a minor fluorine contaminant found in our water supply, and is present in all subsequent  $^{19}\text{F}$  NMR spectra.



**Figure 7.5:** Overlaid  $^{19}\text{F}$  1D spectrum of 5-F-Trp **ab** (blue), 5-F-Trp **b**'x (green) and 6-F-Trp W111F **abb**'x (red). 5-F-Trp **ab** and **b**'x  $^{19}\text{F}$  spectra collected by other laboratory members were used to assign 6-F-Trp W111F **abb**'x. Note:  $^{35}\text{Trp}$  present in **a** domain,  $^{111}\text{Trp}$  present in **a** domain of WT **ab** only,  $^{347}\text{Trp}$  present in **x-linker** region.  $^{347}\text{Trp}$  as part of 5-F-Trp **b**'x is present in its 'capped' and 'uncapped' species.

Integration of the two fluorine peaks of 6-F-Trp  $^{15}\text{N}$  W111F **abb**'x shown in Figure 7.6 gives an estimated 40% fluorine incorporation at  $^{35}\text{Trp}$ , and 60% incorporation at  $^{347}\text{Trp}$ . Together with the mass spectrometry peak deconvolution, it can be estimated that their exists approximately 30% single 6-F- $^{35}\text{Trp}$  species and 10% single 6-F- $^{347}\text{Trp}$  species.

Overall this level of incorporation is much lower than was achieved using 5-F-indole within the laboratory (estimated 80% incorporation in **b**'x at  $^{347}\text{Trp}$ ) (Curtis-Marof et al., 2014).



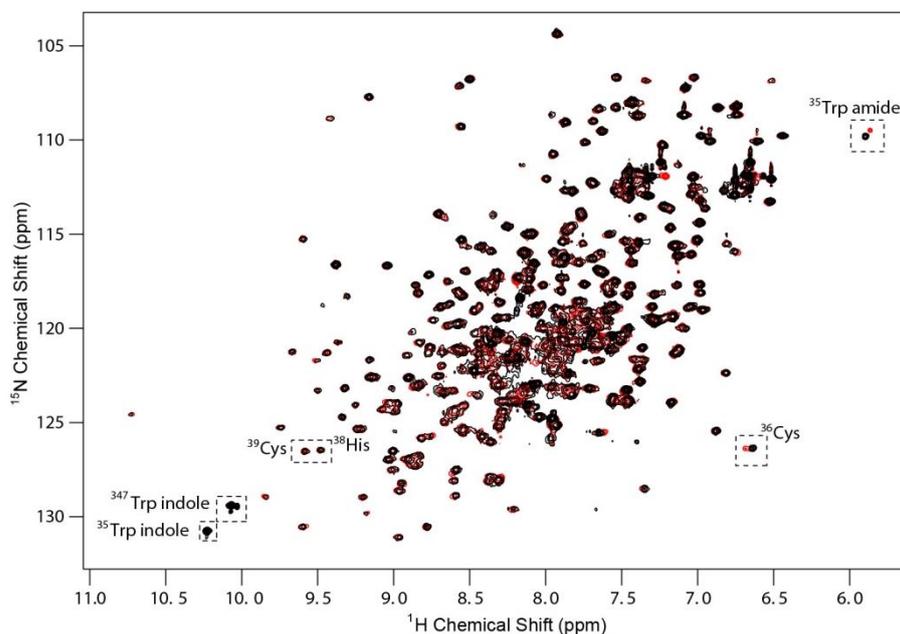
**Figure 7.6:**  $^{19}\text{F}$  1D spectrum of 0.25 mM 6-F-Trp  $^{15}\text{N}$  W111F **abb**'x. Integration of the two peaks gives an approximate fluorine incorporation of 40% at  $^{35}\text{Trp}$  (**a**) :60% at  $^{347}\text{Trp}$  (**x-linker**).

### 7.3.3 Effects of $^{19}\text{F}$ Incorporation Observed by $^{15}\text{N}$ , $^1\text{H}$ -HSQC

$^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra collected from fluorinated protein samples overlays well with spectra of non-fluorinated protein (see Figure 7.7). This was to be expected; the spectra is displaying resonances of both the double and single fluorinated, and un-fluorinated protein species in the sample. As such, chemical shift perturbation maps of overlaid spectra with un-fluorinated protein reported no obvious perturbations, as each resonance was simply mapping to the non-fluorinated protein resonances.

The  $^{35}\text{Trp}$  and  $^{347}\text{Trp}$  indole resonances are still prominent in the 'fluorinated' HSQC spectra; if fluorine is incorporated at these sites, they would not also be  $^{15}\text{N}$  labelled, and thus should disappear from the spectra. This observation is therefore consistent with a low rate of fluorine incorporation at the two sites. A closer inspection of the spectra showed there are extra peaks observable for both the  $^{35}\text{Trp}$  and  $^{36}\text{Cys}$  amide resonances, suggesting  $^{36}\text{Cys}$  also 'senses' a change upon fluorination of  $^{35}\text{Trp}$ . The new peaks are an estimated

two thirds the intensity of the non-fluorinated resonances, concurrent with the values estimated in 7.3.2 for fluorination at this Trp site (40%). Unfortunately  $^{347}\text{Trp}$  amide resonance is in a crowded region of the spectra, so any extra peak that may appear from the fluorinated protein is not readily visible. Unlike  $^{35}\text{Trp}$  and  $^{36}\text{Cys}$ , residues  $^{38}\text{His}$  and  $^{39}\text{Cys}$  (which also form part of the nearby active site), did not appear to be perturbed by the fluorination.



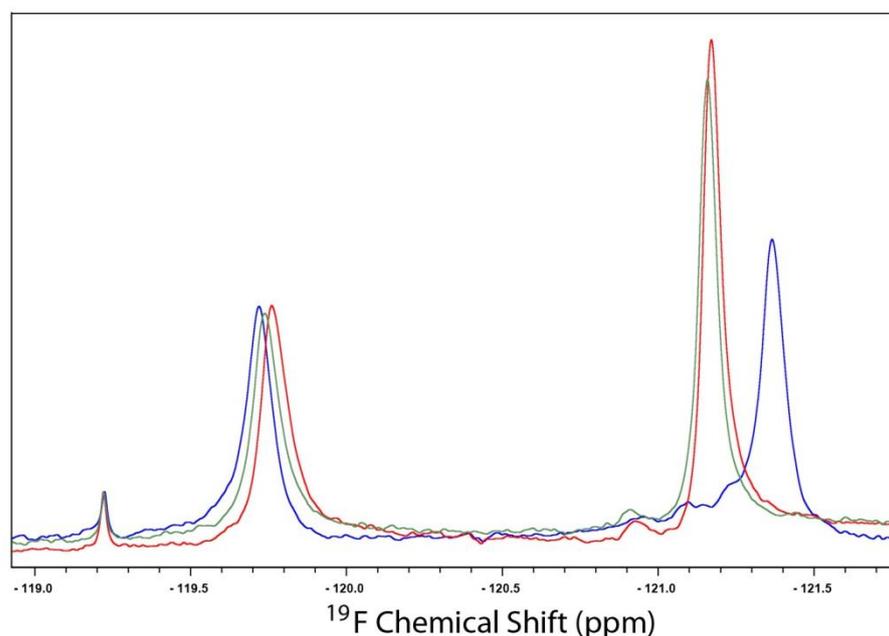
**Figure 7.7:**  $^{15}\text{N},^1\text{H}$ - HSQC spectrum of 0.25 mM  $^{15}\text{N}$  **abb**'x (black) overlaid with 0.25 mM 6-F-Trp  $^{15}\text{N}$  W111F **abb**'x (red). Resonances constituting the redox active site and the two fluorinated Trp residues are highlighted on the spectra by dashed boxes.

### 7.3.4 Ligand Binding of 6-F-Trp $^{15}\text{N}$ W111F **abb**'x

As demonstrated by mass spectrometry,  $^{15}\text{N},^1\text{H}$ -HSQC and  $^{19}\text{F}$  1D NMR, the total level of fluorine incorporation is relatively low when compared to uniform isotope labelling methods (98.9% incorporation of  $^{15}\text{N}$  was reported for W111F **abb**'x in 3.3.5). Practically, this level of incorporation limits the use of 'mixed fluorinated' protein for detailed ligand binding analysis, where a homogenous, uniform protein species is paramount for quantitative studies. Ligand binding titrations carried out by  $^{15}\text{N},^1\text{H}$ -HSQC generated an

average  $K_d$  of  $90 \pm 65 \mu\text{M}$ , but it must be appreciated this figure is reporting on a mixture of fluorinated and un-fluorinated protein within the sample (see Appendix 7.1 for graphical fits). As a result, the figure quoted may downplay the influence of fluorinated protein on peptide ligand binding.

Preliminary  $^{19}\text{F}$  1D spectra collected of exclusively fluorinated protein (whilst in a mixed fluorinated protein sample) showed the **x-linker**  $^{347}\text{Trp}$  residue is perturbed upon addition of either  $\Delta$ -som or F1,2,3  $\Delta$ -som peptides, with chemical shift perturbation of a similar magnitude and direction to that observed as published with 5-F-Trp **b'**x protein (Curtis-Marof et al., 2014). An interesting observation was the smaller perturbation of  $^{35}\text{Trp}$  resonance, that is found directly adjacent to active site in **a** domain. This resonance was seen to report shift changes upon addition of either fluorinated or un-fluorinated peptide (see Figure 7.8), suggesting that residues close to the catalytic active site in **a** are influenced by peptide ligand binding in **b'** domain. This phenomena had not previously been observed by traditional NMR methods.

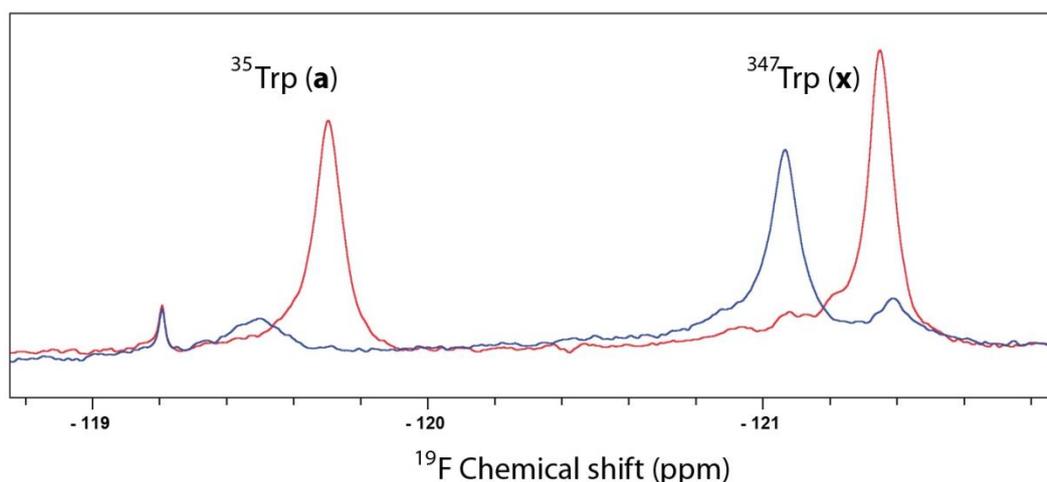


**Figure 7.8:** Overlaid  $^{19}\text{F}$  1D spectrum of 0.25 mM 6-F-Trp  $^{15}\text{N}$  W111F **abb'**x (blue) with addition of either 0.625 mM  $\Delta$ som (green) or 0.625 mM F1,2,3  $\Delta$ som peptide (red). The spectra show that the 6-F-Trp in the **x-linker** reports perturbation upon ligand binding with both fluorinated and un-fluorinated peptide. Interestingly, the 6-F-Trp in **a** domain is also perturbed by ligand binding of both peptides, with the greatest perturbation observed upon addition of F1,2,3  $\Delta$ -som peptide.

Close contact of the **a** domain active site and the ligand binding site have recently been suggested by the Wang group (Yang et al., 2014), so this was an exciting observation. The extent of perturbation was largest with the addition of fluorinated protein, therefore attempts were made by NOE difference and  $^{19}\text{F}$ - $^{19}\text{F}$  NOESY experiments (between fluorinated protein and fluorinated peptide) to confirm this occurrence, but unfortunately were not successful, resulting in blank spectra.

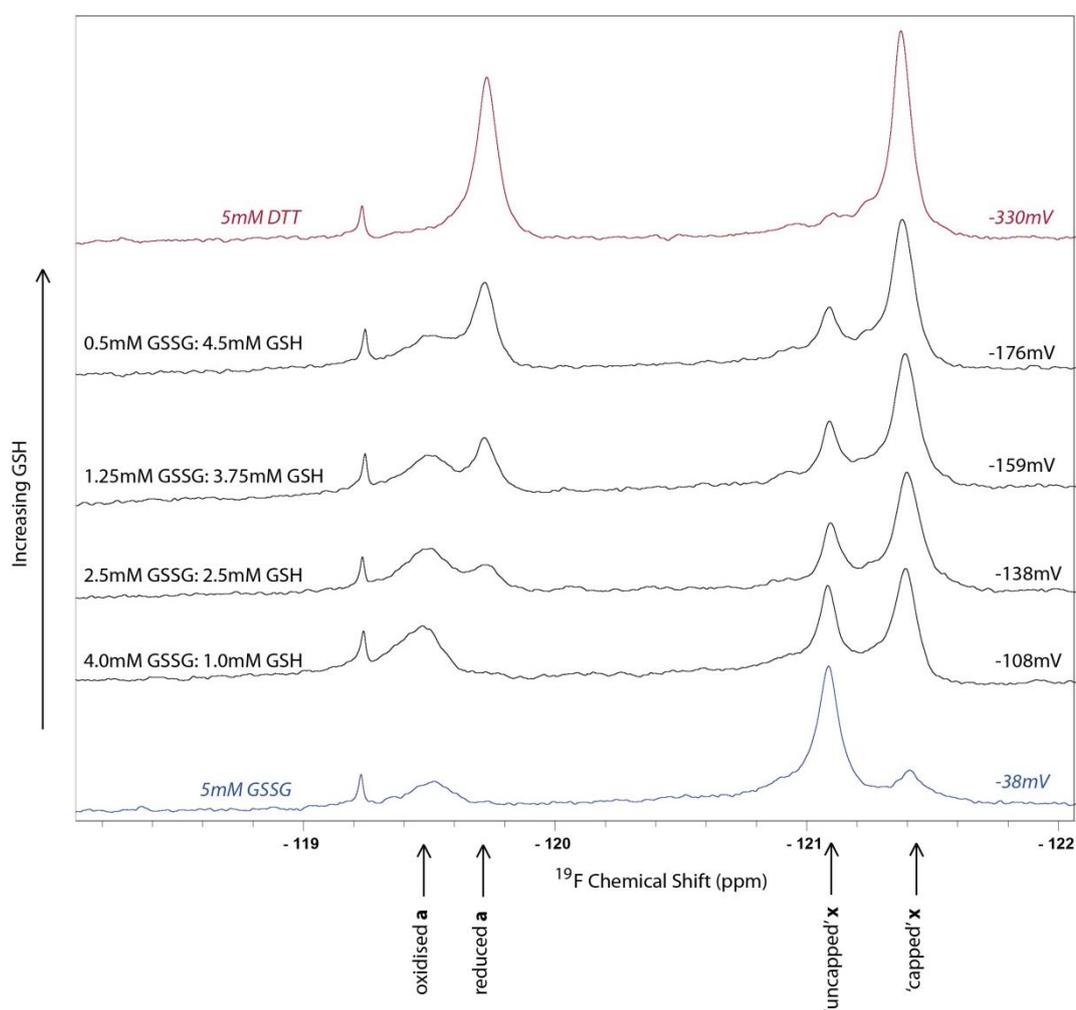
### 7.3.5 Redox Potential of 6-F-Trp $^{15}\text{N}$ W111F **abb**'x

$^{19}\text{F}$  1D spectra were collected of reduced and oxidised 6-F-Trp  $^{15}\text{N}$  W111F **abb**'x, and showed that both fluorine resonances reported a perturbation upon change in redox state (see Figure 7.9). A clear shift was observed for  $^{35}\text{Trp}$  resonance upon change in redox state, with the oxidised peak appearing much broader and lower in intensity, consistent with the previous data collected of  $^{15}\text{N}$  labelled protein by  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC (see 4.3.1). The  $^{347}\text{Trp}$  resonance exists as a mixture of 'capped' and 'uncapped' protein, with reduced protein adopting a mostly 'capped' form, and oxidised protein adopting a mostly 'uncapped' form. It was reasoned that as  $^{347}\text{Trp}$  reports this dynamic 'capping' and 'uncapping' event, and appears as multiple resonances, it would not be an accurate reporter of protein redox state. Efforts were therefore directed to observation of  $^{35}\text{Trp}$  resonance to solely monitor protein redox state.



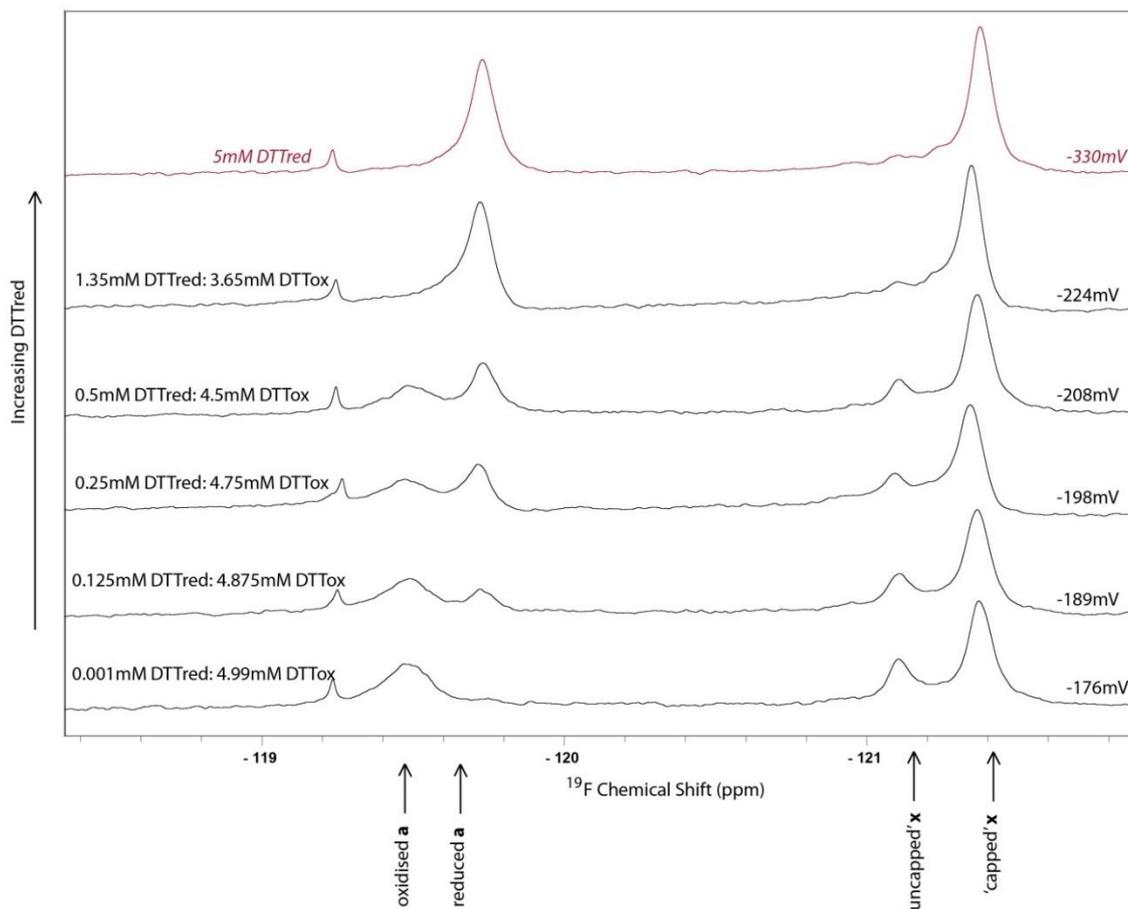
**Figure 7.9:** Overlaid  $^{19}\text{F}$  1D spectrum of 0.25 mM 6-F-Trp  $^{15}\text{N}$  W111F **abb**'x reduced (red) and oxidised (blue).

The redox potential of 6-F-Trp  $^{15}\text{N}$  W111F **abb'x** was attempted from practical data, using the careful titration of differing ratios of the redox pairs GSH:GSSG ( $E'_0$  -240 mV) and  $\text{DTT}^{\text{red}}:\text{DTT}^{\text{ox}}$  ( $E'_0$  -330 mV) to follow the  $^{35}\text{Trp}$  resonance. As per  $^{15}\text{N},^1\text{H}$ -HSQC redox potential data collection, control spectra using 5 mM GSSG (fully oxidised) and 5 mM  $\text{DTT}^{\text{red}}$  (fully reduced) were collected.  $^{35}\text{Trp}$  peak heights and volumes were recorded from each spectra and made relative to the fully reduced spectra. In addition to the two control spectra, 4 data points were initially collected using differing ratios of GSH:GSSG, as shown in Figure 7.10.



**Figure 7.10:**  $^{19}\text{F}$  1D spectra of 6-F-Trp W111F **abb'x** change in protein redox state, using the GSH:GSSG redox pair. The  $^{35}\text{Trp}$  resonance in **a** reports a gradual shift from a mixture of reduced/oxidised protein (at -176 mV), to a completely oxidised protein (at -108 mV). The  $^{347}\text{Trp}$  in **x-linker** also shows a gradual change from mostly 'capped' to mostly 'uncapped' forms upon oxidation.

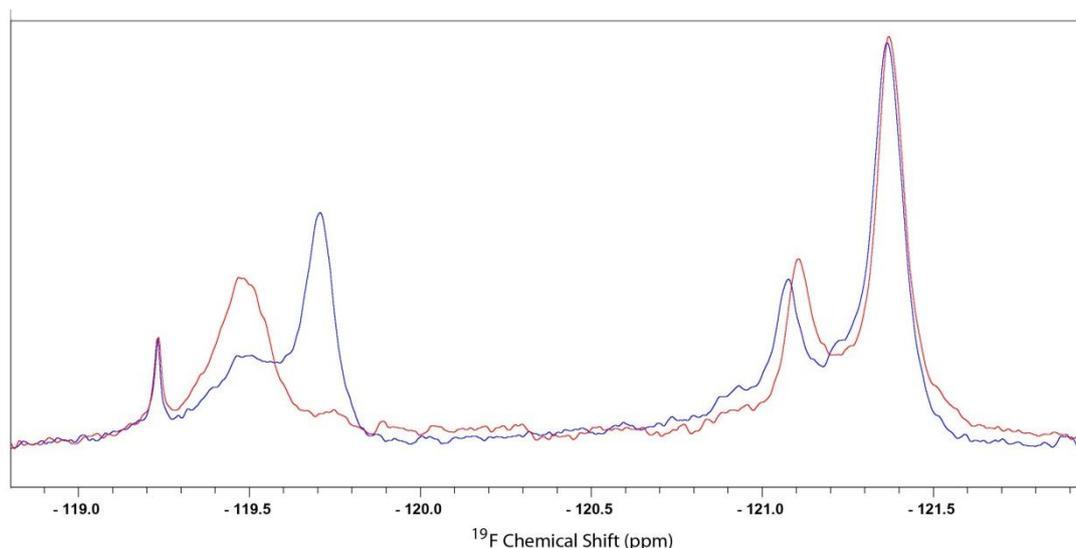
A similar threshold of the ability of GSH to reduce W111F **abb**'x was also observed by  $^{19}\text{F}$  1D, hence an additional 5 data points were collected using differing ratios of  $\text{DTT}^{\text{red}}:\text{DTT}^{\text{ox}}$ , to complete the reduced portion of the Hill plot. The spectra collected are shown in Figure 7.11.



**Figure 7.11:**  $^{19}\text{F}$  1D spectra of 6-F-Trp W111F **abb**'x change in redox state, using the  $\text{DTT}^{\text{red}}:\text{DTT}^{\text{ox}}$  redox pair. The  $^{35}\text{Trp}$  resonance in **a** reports a gradual change from reduced protein (at -330 mV) to a completely oxidised protein (at -176 mV). The  $^{347}\text{Trp}$  in **x-linker** shows a slight shift, from mostly 'capped' to a mixture of 'capped' and 'uncapped' forms upon change in oxidation state.

Similar experiments run using uniform  $^{15}\text{N}$  labelled W111F **abb**'x showed the DTT redox pair to be strongly reducing, capable of keeping the protein between 70-100% reduced (for a comparison see the black data points of Figure 6.3). In direct contrast, observation of the  $^{35}\text{Trp}$  resonance in  $^{19}\text{F}$  1D spectra here suggest the protein is more oxidised than

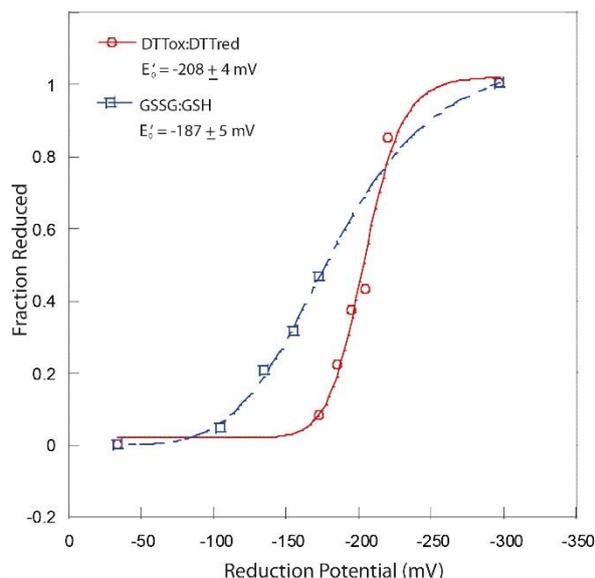
anticipated by the redox potential equilibria used. An overlay of two experiments, both at equilibria of -176 mV (achieved with either 4.99 mM DTT<sup>ox</sup>: 0.001 mM DTT<sup>red</sup> or 0.5 mM GSSG: 4.5 mM GSH) show <sup>35</sup>Trp reporting different protein redox states, depending on the redox pair used (see Figure 7.12).



**Figure 7.12: Overlaid <sup>19</sup>F 1D spectra of 4.99 mM DTT<sup>ox</sup>: 0.001 mM DTT<sup>red</sup> (red) and 0.5 mM GSSG:4.5 mM GSH (blue). Both experiments are conducted with the system at  $E'_0$ : -176 mV.**

It is interesting to note that whilst the <sup>35</sup>Trp appears in a mixed redox state with the glutathione redox pair, and subsequently an almost exclusively oxidised state with the DTT redox pair, that the <sup>347</sup>Trp resonances report similar proportions of 'capped' and 'uncapped' protein in the overlaid spectra. This suggests that the **x-linker** Trp does not 'sense' the change in redox state in the same way the Trp residue in **a** domain does.

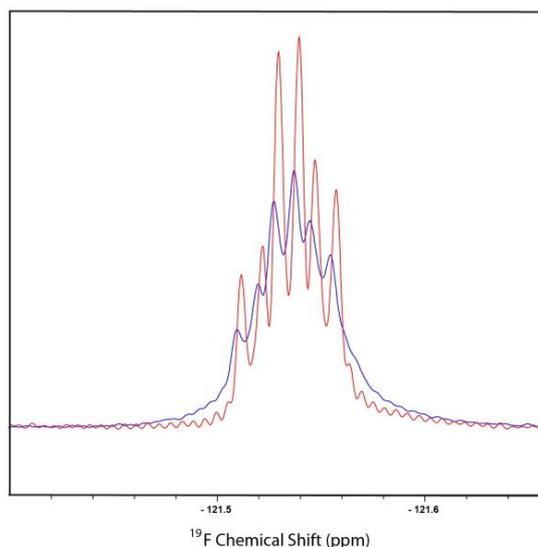
When the two datasets are fit independently to Hill plots (Figure 7.13), they give redox potentials of  $E'_0$   $-187 \pm 5$  mV for the glutathione pair, and  $E'_0$   $-208 \pm 4$  mV for the DTT pair. This stark difference was not observed when using <sup>15</sup>N uniform labelled protein, where all data points fit to a single curve to give a redox potential of  $E'_0$   $-167 \pm 2.1$  mV. This suggests that either the fluorine is altering the local environment to such an extent as to modify the redox potential of the **a** domain active site, or alternatively, the presence of fluorine is causing an unexpected interaction between the DTT and the catalytic site.



**Figure 7.13: Redox potential graph of 6-F-Trp W111F *abb*'x as monitored by <sup>19</sup>F 1D spectra, fit using the Hill Equation.** Data points in red were collected using the DTT<sup>ox</sup>:DTT<sup>red</sup> redox pair and fit to give a redox potential of  $E'_0 -208 \pm 4$  mV. Data points in blue were collected using the GSSG:GSH redox pair and fit to give a redox potential of  $E'_0 -187 \pm 5$  mV.

Indole is not thought to react with DTT, however spectra of 0.25 mM 6-F-indole in the presence of a 20 molar excess of either DTT<sup>red</sup> or DTT<sup>ox</sup> are markedly different, with fluoroindole peaks in the presence of DTT<sup>ox</sup> displaying peak broadening (see Figure 7.14), hinting there may be an interaction.

It is apparent more work needs to be carried out to ascertain what effect the fluorination is having on the hPDI redox system, which unfortunately was not within the timescale of this thesis.



**Figure 7.14:** Overlaid  $^{19}\text{F}$  1D spectra of 0.25 mM 6-F-indole with 5 mM  $\text{DTT}^{\text{red}}$  (red) or 5 mM  $\text{DTT}^{\text{ox}}$  (blue). Line broadening of the fluoroindole peaks suggests the  $\text{DTT}^{\text{ox}}$  could be interacting with 6-F-indole.

## 7.4 Discussion

### 7.4.1 Expression of 6-F-Trp $^{15}\text{N}$ W111F **abb**'x

Incorporation of 60 mg/L unfluorinated indole into cells grown using standard minimal media for expression of  $^{15}\text{N}$  W111F **abb**'x was successful, with the addition having no visible adverse effects to either cell growth or protein expression, and with comparable expression upon induction to standard minimal grows. Of the fluorinated indoles, addition of 4-F, 5-F or 7-F-indole to the growth medium were each found to have inhibitory effects—either by retarding cell growth (5-F), or with unaffected cell growth, but no visible W111F **abb**'x expression bands by SDS-PAGE (4-F and 7-F). Protein expression using 6-F-indole gave a prominent protein band comparable to those observed in standard minimal grows. The protein was readily purified using metal affinity, anion exchange and size exclusion chromatography, resolving as a predominantly monomeric protein, with a small portion of well resolved dimer (as previously reported for W111F **abb**'x, Chapter 3). Final monomeric protein yields were also highly comparable to those previously described for W111F **abb**'x.

Mass spectrometry of the monomeric protein indicated that several different forms of the monomer had been co-purified. Masses of 41,003.8 Da (corresponding to un-fluorinated protein: 30%), 41,020.5 Da (corresponding to singly fluorinated protein: 40%) and 41,039.4 Da (corresponding to doubly fluorinated protein: 30%) were all identified.  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC NMR spectra of the protein also demonstrated fluorine incorporation was not as high as had been anticipated- if fluoroindole is taken up by cells for synthesis of fluoro-tryptophan, then the indole of each Trp residue will not contain  $^{15}\text{N}$  labelling, (and should therefore disappear from the spectra). Both  $^{35}\text{Trp}$  and  $^{347}\text{Trp}$  indole resonances were still clearly visible in the spectra, suggesting a low proportion of protein was fluorinated at these sites.

$^{19}\text{F}$  1D NMR spectra of the protein showed two distinct peaks, that correlate well with corresponding 5-F-indole hPDI fragment spectra collected within the laboratory. Peak integration estimated 60% incorporation at residue  $^{347}\text{Trp}$ , and 40% at residue  $^{35}\text{Trp}$ . Together with the mass spectrometry analysis, this suggests 30% of protein was labelled at both Trp sites, 30% labelled at  $^{347}\text{Trp}$  (**x-linker**) only, 10% labelled at  $^{35}\text{Trp}$  (**a domain**) only, and the remaining 30% was expressed as un-fluorinated protein.

This analysis suggested fluorine incorporation into W111F **abb'x** by 6-F-indole was not as extensive or uniform as previously published for the **b'x** fragment (Curtis-Marof et al., 2014), where greater than 80% incorporation of **b'x** was achieved using 5-F-indole. Work carried out within this thesis showed 5-F-indole had an inhibitory effect on the growth of cells encoding W111F **abb'x**. 5-F-indole has long been reported as being non-specifically cytotoxic to cells (Taub, 1977); indeed the method developed by Crowley et al. also reported a toxic effect on cells when used at concentrations of 300 mg/L (Crowley et al., 2012). However, care was taken to use a final concentration of 60 mg/L 5-F-indole in the test expressions, so results from all fluoroindole grows could be directly compared.

It is possible the new stock of 5-F-indole used within this thesis could have been contaminated with something else the cells find toxic, perhaps introduced at a stage of purification or synthesis. Unfortunately there was not time to probe this idea any further. It may be helpful to note that some of the fluorinated indoles have only been recently synthesised by Sigma (they offer 'experimental' grades that do not come with analytical identification or purity), with 7-F-indole only becoming available in 2015. It is also

interesting to note that whilst 5-F-indole gave greatest protein expression for the **b'x** fragment, incorporation using 6-F-indole into **b'x** was also successful in the laboratory, with an estimated incorporation of 55%. This value is supportive of the incorporation observed here of 6-F-indole incorporation into W111F **abb'x**, where incorporation at  $^{347}\text{Trp}$  (the same Trp residue fluorinated as in the **b'x** fragment) was estimated at 60%.

In summary, the observed level of fluorine incorporation into W111F **abb'x** complicated further quantitative analysis of fluorinated protein.  $^{15}\text{N},^1\text{H}$  NMR spectra report on all  $^{15}\text{N}$  labelled protein in the sample, whether fluorinated or not. It is clear to see from the  $^{15}\text{N},^1\text{H}$ -HSQC in Figure 7.7 that some residues are shifting upon fluorination, but chemical shift perturbation maps will map to the nearest peak. In a mixed population this will likely be the unfluorinated peaks, consequently 'downplaying' the extent fluorination is having on the protein. Ideally, fluorinated protein needs to be isolated from unfluorinated protein prior to analysis by NMR. This was simply not possible by the purification techniques used within this thesis. Similarly, identification of either a +18 or +36 Da mass change to a 40.5 kDa protein was only possible by mass spectrometry using sophisticated deconvolution software.

It may be beneficial to trial fluoroindole expression using a fresh stock of 5-F-indole, to see if this fluoroindole can potentially be incorporated as successfully into W111F **abb'x** as was reported for the **b'x** fragment.

#### 7.4.2 Protein-Observed Ligand Binding by $^{19}\text{F}$ NMR

Detailed ligand binding analysis of fluorinated protein was made more complex by the presence of a mixed protein sample, hence the dissociation constant calculated by  $^{15}\text{N}-^1\text{H}$ -HSQC ( $K_d: 90 \pm 65 \mu\text{M}$ ) is reporting on both un-fluorinated and fluorinated protein in the sample. Whilst the measured  $K_d$  value suggests fluorination does not influence peptide binding affinity, it is difficult to conclusively say this is the case, due to mixed species being present. Nevertheless, the  $^{19}\text{F}$  peaks of 6-F-Trp W111F **abb'x** were well resolved, and were also distinct from fluorinated peptide peaks, illustrating that ligand binding could be tracked for both peptide and protein simultaneously by  $^{19}\text{F}$ , especially if higher rates of protein fluorine incorporation can be achieved. Quantitative studies were not undertaken using  $^{19}\text{F}$  1D spectra because work already carried out within the laboratory using the **b'x**

fragment of hPDI determined a modest increase in binding affinity when  $^{19}\text{F}$  was incorporated (Curtis-Marof et al., 2014).

It was an exciting observation that the fluorinated  $^{35}\text{Trp}$  resonance in **a** was perturbed upon ligand binding, suggesting that the active site of **a** domain is in a relatively close molecular proximity to the PDI binding site in **b'**, as speculated by Wang et al. recent paper (Yang et al., 2014). Unfortunately work carried out to attempt collection of  $^{19}\text{F}$ - $^{19}\text{F}$  NOEs (between  $^{19}\text{F}$  Trp in **a** domain,  $^{19}\text{F}$  Trp in **x-linker** and fluorinated F1,2,3  $\Delta$ -som peptide) was not successful, giving blank spectra. In addition, NOE difference spectra, that work by selectively saturating ligand peaks to view changes to the protein peaks (or vice versa), did not yield any difference spectra. Consequently, all attempts to validate this occurrence were inconclusive. However, work within the laboratory that was conducted after practical work for this thesis was concluded, has shown that at 14.1 tesla,  $^{19}\text{F}$  relaxation is severely influenced by chemical shift anisotropy. Although  $T_1$ -based NOE transfer is possible, transverse relaxation would drive exceptionally fast signal decay and would hamper the detection of  $^{19}\text{F}$ - $^{19}\text{F}$  NOEs (personal communication with M.J.Howard).

### 7.4.3 Redox Potential by $^{19}\text{F}$ NMR

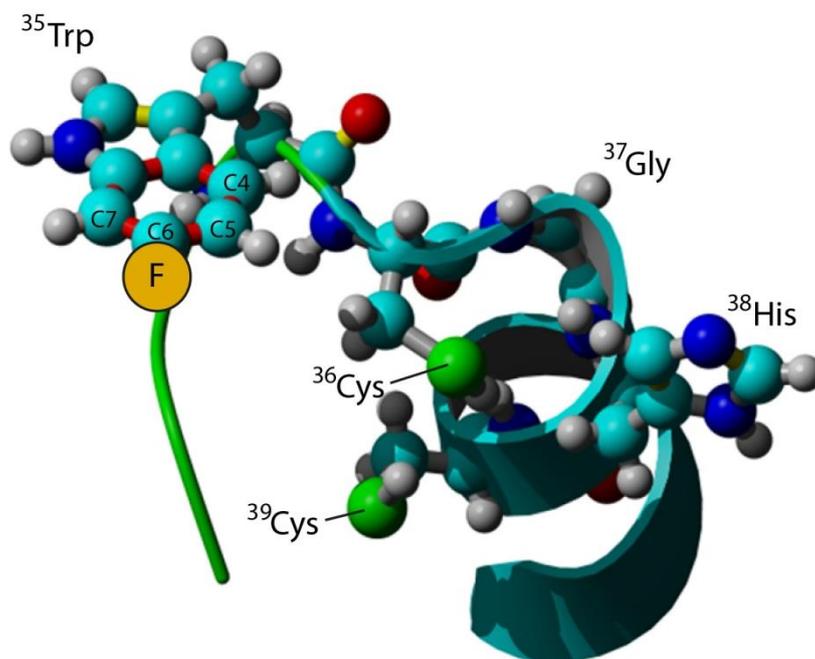
It was anticipated that the incorporation of fluorine probes into the W111F **abb'x** protein might allow  $^{19}\text{F}$  1D spectra to be used much like is reported in this thesis using traditional  $^{15}\text{N}$  labelling approaches- to follow the protein through a change in redox state and equate a redox potential.

Analysis of the redox state of fluorinated protein was complicated by several factors. Unfortunately, the  $^{347}\text{Trp}$  residue reports as a mixture of 'capped' and 'uncapped' protein, so whilst a general trend from 'capped' (in reduced spectra) to 'uncapped' (in oxidised spectra) was observed, neither peak could be followed accurately through the redox titration to fit this residue, leaving the single reporter of  $^{35}\text{Trp}$ . Data collected by  $^{19}\text{F}$  using a titration of the glutathione redox pair, fit  $^{35}\text{Trp}$  to give a redox potential of  $E'_o -187 \pm 5$  mV, suggesting the redox potential may be made moderately more reducing by the fluorination. As per the  $^{15}\text{N}$  labelling strategy, collection of additional data points using the DTT redox pair were made, to improve the reduced section of the Hill plot. With experiments carried out to report between -176 mV to -300 mV, it was expected the Trp residue would report a predominantly reduced state. Instead,  $^{35}\text{Trp}$  reported a more

oxidised state than expected; overlaid spectra at -176 mV (made using either 4.99 mM DTT<sup>ox</sup>: 0.001mM DTT<sup>red</sup> or 0.5 mM GSSG: 4.5 mM GSH), should have resulted in comparable proportions of reduced and oxidised protein, but delivered a spectrum with an almost exclusively oxidised peak for <sup>35</sup>Trp in the DTT spectra (see Figure 7.12). Interestingly, the 'capped' and 'uncapped' forms of <sup>347</sup>Trp overlay extremely well in both spectra, suggesting the **x-linker** does not sense this change in redox state. This data indicates that fluorination of the protein has an effect on the system; either, fluorotryptophan modifies the redox active site, or interacts with DTT (a finding which is also inferred by Figure 7.14).

It is possible that fluorine incorporation may change the local surface charge of the protein, to either repel reduced DTT or attract oxidised DTT. If either of these situations were to occur, then the local equilibria of DTT<sup>red</sup>:DTT<sup>ox</sup> might not accurately reflect the actual equilibria of the experiment, thus the protein would not report an accurate redox state. It is also reasonable to assume that addition of fluorine so close to the active site could also result in other changes to the local environment (such as steric effects). Whilst the redox potential fit achieved by <sup>19</sup>F using the glutathione pair reports only a modest increase in protein redox potential (see Figure 7.13), the curve fit is noticeably different to those generated using <sup>15</sup>N labelled protein, with a much shallower gradient, suggesting lower cooperativity (see curve fits of Figure 6.3).

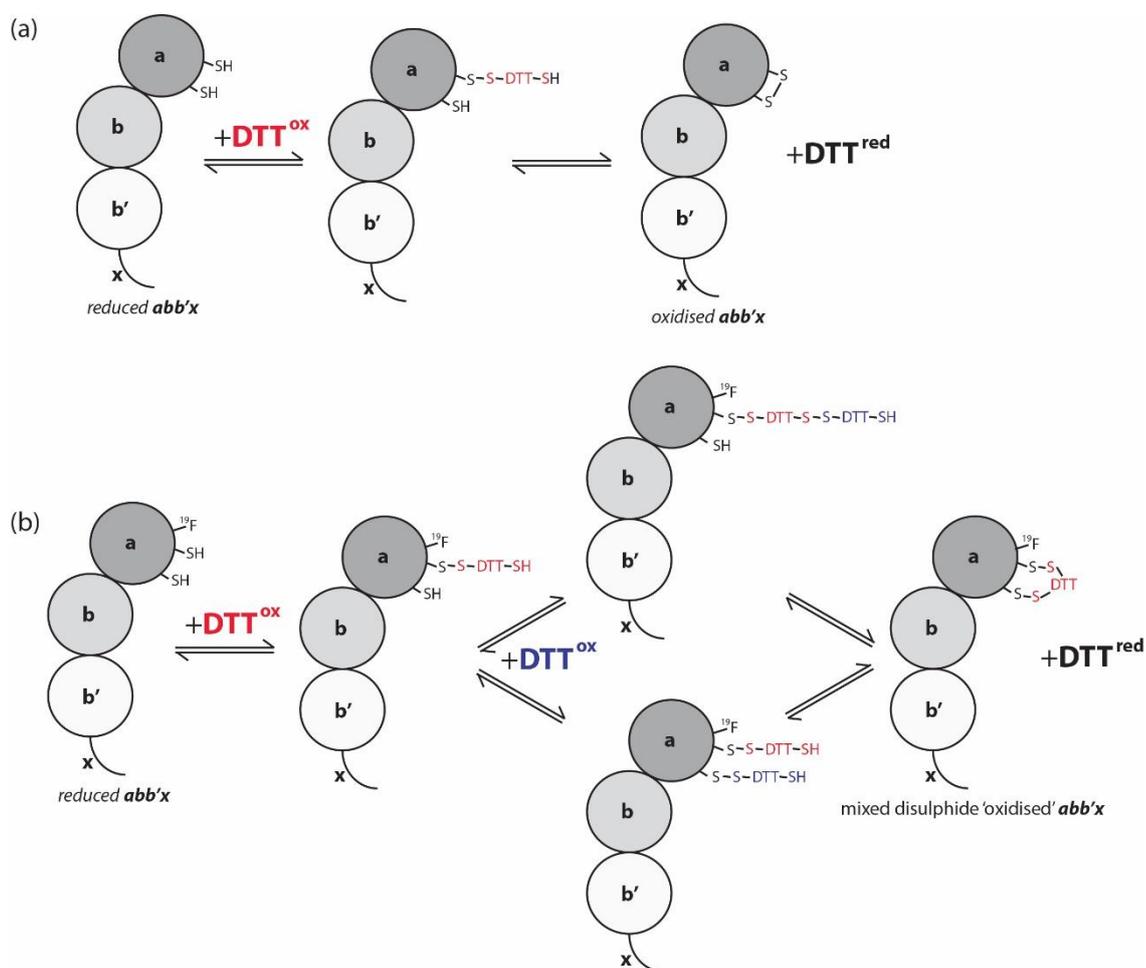
Figure 7.15 illustrates the fluorine modification made at the <sup>347</sup>Trp residue, and shows the fluorine atom is pointing directly towards the two sulphur atoms of <sup>36</sup>Cys (within 7.45 Å) and <sup>39</sup>Cys (within 8.73 Å). If this configuration is correct, a fluorine modification in close proximity to the active site sulphur atoms could impart some effect on the redox ability of the protein, or affect the stability of the protein.



**Figure 7.15: hPDI *a* domain active site (4EKZ.pdb) with 6-Fluoro-Trp incorporation: the fluorine atom of  $^{35}\text{Trp}$  points towards the two active site cysteines.** The active site motif (CGHC) and  $^{35}\text{Trp}$  are shown as ball and stick models, the sulphur atoms of the cysteine residues are coloured green, and the fluorine atom is coloured orange. The aromatic ring carbons of  $^{35}\text{Trp}$  are also numbered. Molecular graphics created using Yasara ([www.yasara.org](http://www.yasara.org)).

$^{35}\text{Trp}$  is a highly conserved residue among PDI in different species, but unfortunately there is little reported in the literature as to why it is conserved. It lies immediately adjacent to the active site in **a** domain, by which it is heavily quenched, and is therefore rarely targeted for mutation in protein fluorescence studies. A W35F/W111F/W379F/W390F mutant (all Trp residues of hPDI mutated to Phe) has been reported in the literature by Nguyen et al., with the mutations stated as having a destabilising effect on the protein, but it is not apparent which Trp residues cause this de-stabilisation (Nguyen et al., 2008). A recent paper by Gray et al. report how Tyr and Trp residues are present in close proximity to the redox active sites of a number of oxidoreductases. These residues may work to transport potentially damaging oxidising equivalents away from the active site of the protein (Gray and Winkler, 2015). It may be beneficial to explore this Trp conservation among PDI family members further, particularly if the residue plays a role in redox regulation.

Alternatively, whilst DTT is known as a strong reducing agent (it is believed not to form mixed disulphides, due to a rapid rate of recyclization), rare cases do occur where DTT adducts form. In a case published by Li et al. (Li et al., 1998), the two sulphur atoms of DTT form mixed disulphides with two different sulphur atoms of RNase A protein. A similar mode of action could be occurring for fluorinated W111F **abb'**x, and is demonstrated in Figure 7.16.



**Figure 7.16:** Scheme showing the oxidation of reduced W111F **abb'**x by DTT<sup>ox</sup> (a) and the proposed mixed disulphide oxidation of 6-F-Trp W111F **abb'**x by DTT<sup>ox</sup> (b). In both cases, the initial interaction of reduced **abb'**x with DTT<sup>ox</sup> leads to a mixed disulphide. The DTT should then rapidly cyclize to yield DTT<sup>red</sup> and oxidised protein. If instead the mixed disulphide state is somehow stabilised, a subsequent DTT<sup>ox</sup> could attack the free thiols of either the protein or the DTT molecule, resulting in an oxidised form of protein—where instead of an intramolecular disulphide bond as in (a), two intermolecular disulphides have formed instead with one DTT molecule.

It is possible that the addition of fluorine into  $^{35}\text{Trp}$  works to stabilize a mixed disulphide state of DTT and protein. If this happens, DTT is no longer able to cyclize (and become  $\text{DTT}^{\text{ox}}$ ), since it has no remaining free thiols. So, whilst each cysteine residue of the protein is principally in an 'oxidised' state, instead of being intra-molecularly bonded to one another, they are each bonded to a DTT molecule.

The case described in Figure 7.16 relies upon subsequent interactions of reduced protein with  $\text{DTT}^{\text{ox}}$ , which could explain why datapoints collected between -189 to -208 mV (where there is a 9 to 39 times excess of  $\text{DTT}^{\text{ox}}$  compared to  $\text{DTT}^{\text{red}}$ ) report a more 'oxidised' form of protein- the reaction conditions could work to promote formation of a 'mixed disulphide' species. The data point collected at -224 mV (with a much lower excess of  $\text{DTT}^{\text{ox}}$  to  $\text{DTT}^{\text{red}}$ ) actually fits much more closely with the glutathione generated Hill plot (see -224 mV DTT datapoint in Figure 7.13). Unfortunately there was not time within the project to explore this idea any further, but if a mixed disulphide form of protein/DTT is present and stable for a suitable time period, it may be possible to identify the species using mass spectrometry (visible as a + 158 Da mass increase).

It is also possible that  $\text{DTT}^{\text{ox}}$  may be interacting with the fluoroindole- the overlaid spectra of 6-F-indole with either  $\text{DTT}^{\text{red}}$  or  $\text{DTT}^{\text{ox}}$  (Figure 7.14) showed the  $^{19}\text{F}$  peaks of 6-F-indole broaden out upon addition of the  $\text{DTT}^{\text{ox}}$ . Saturation Transfer Difference NMR (STD-NMR) could be used to determine if an interaction is occurring. The technique is largely used as a qualitative tool, to screen for quick 'yes/no' answers to binding- in this type of experiment the protein is selectively saturated, and difference spectra (saturation on/off) will show perturbations of the ligand signals if binding occurs (Viegas et al., 2011). It has however been shown  $\text{DTT}^{\text{ox}}$  can interact with hPDI fragments at locations other than the catalytic active sites (the hPDI **bb'xa'** crystal structure was published with a molecule of  $\text{DTT}^{\text{ox}}$  packed tightly into the ligand binding site in **b'** domain (Wang et al., 2012)). This is something that has not previously been observable in HSQC spectra of hPDI, but highlights that STD-NMR data should in this case still be treated with caution.

It would be interesting to repeat the redox titration using another redox pair (for example using  $\beta$ -mercaptoethanol, which has a reducing ability intermediate of DTT and glutathione), to see how the protein reports its redox state in comparison to the data obtained. This data collection might also help to validate the idea of a mixed disulphide

forming with DTT<sup>ox</sup>, if the  $\beta$ -mercaptoethanol titration reports more similarly to the glutathione redox titration.

#### **7.4.4 Overall Summary**

Work within this Chapter has reported some of the potential problems fluorine incorporation may impart. In this case it is possible the fluorination of <sup>35</sup>Trp has modified the protein redox state in some way, or altered the biochemical properties local to the site of incorporation. Whilst the synthesis of fluorinated compounds in medicinal chemistry is now carried out on a routine basis, this work illustrates the continual requirement to explore the impact such a modification may make, and it is clear <sup>19</sup>F protein labelling should still be examined on a case by case basis.

# CHAPTER 8

## General Summary

Previously published research reports that hPDI is capable of binding a large range of unfolded or mis-folded proteins, as well as a variety of smaller unstructured peptide ligands. In addition, hPDI displays activity as a molecular chaperone, as incorrect disulphide bond formation or presence of cysteine residues are not prerequisite for binding.

The primary substrate binding site was mapped to the **b'** domain and shows a large multivalent hydrophobic binding pocket that spans the domain (Byrne et al., 2009). More recent work has highlighted small hydrophobic patches present in **a**, **b** and **a'** domains that likely contribute to the binding of larger mis-folded proteins (Wang et al., 2013).

hPDI is able to bind substrates at its primary **b'** binding site via reversible, hydrophobic interactions, but ultimately hPDI ligand specificity is still poorly understood. Early work showed hPDI bound two small peptides, somatostatin and mastoparan, but the peptides share no discernable binding motif. hPDI was also found to bind  $\Delta$ -som, a double Cys-Ser form of somatostatin that lacks the WT  $^3\text{Cys-}^{14}\text{Cys}$  disulphide bond, confirming binding of somatostatin was also not reliant on cysteine residues for recognition (Klappa et al., 1997). Work carried out at a similar time identified the binding specificity of the PDI family member PDIp to an hydroxyaryl group (Klappa et al., 2001). It was suggested that hPDI ligand specificity may work in a similar manner, in that ligands may be recognised by their exposed aromatic residues, similar to ligand recognition by other molecular chaperones.

However, as the binding could not be attributed to a specific motif like PDIp, recognition by hPDI was anticipated to be somewhat more complex.

Obtaining more detailed information of the substrate specificity of hPDI is of significant importance, as hPDI is implicated in many different disease states. Chronic up-regulation of hPDI can eventually lead to cell apoptosis, and additionally a major loss of physiological hPDI function can occur as a result of aberrant post-translational S-nitrosylation. hPDI is also industrially relevant, owing to its ability to fold recombinant therapeutics, many of which require correct disulphide bond formation to be biologically active.

To summarise, hPDI displays great potential as both a therapeutic treatment but also a druggable target *in vivo*. Unfortunately many of the small molecule inhibitors of hPDI are targeted to the catalytically active CGHC redox sites, and are plagued by cytotoxic effects and broad specificity. The main aims of this thesis were to characterise the **b'x** and previously unassigned **abb'x** fragments of hPDI, and use them to accurately determine the binding affinity of hPDI to its model ligand peptide  $\Delta$ -som. In addition, we aimed to use specific residue substitutions of  $\Delta$ -som to ascertain hPDI binding specificity.

Two fragments of hPDI were used: **b'x**, the smallest monomeric fragment capable of peptide ligand binding, and **abb'x**, a larger fragment of hPDI that also includes a redox active site. Studies of **b'xa'c** hPDI fragment within the laboratory show in solution the **a'** domain is on the verge of unfolding, even in the absence of denaturant (unpublished observations). As such, it gives exceedingly poor NMR spectra. It was anticipated the W111F **abb'x** fragment of hPDI would be more stable and hence give better NMR spectra. Studies of hPDI by NMR spectroscopy typically report the use of different protein fragments. Increased molecular weight introduces a greater number of resonances to the NMR spectra, complicating both resonance assignment and spectral analysis. In addition, larger molecules have slower tumbling rates in solution, which will also negatively impact spectral quality. The current size limit of protein NMR is ~35 kDa, so studies of the 40.5 kDa W111F **abb'x** fragment of hPDI had approached the 'useful' limit of standard NMR techniques. Larger proteins can be studied using deuteration, but this option is not always offset by the increased cost of raising a deuterated protein sample.

In Chapter 3, expression of **b'x** and **abb'x** in minimal media produced proteins that purified as mixed monomer: dimer species, as previously published for the **b'x** fragment. Dimer formation occurs at the **b'** substrate binding site, so **abb'x** dimer had also been anticipated. Average yields of 21 mg/L monomeric  $^{15}\text{N}$  **b'x** and 38 mg/L monomeric  $^{15}\text{N}$  W111F **abb'x** were purified by size exclusion chromatography for subsequent NMR studies.

Monomeric W111F **abb'x** displayed a hydrodynamic volume 7% smaller than expected for a protein of its size, which suggests the monomer may exist in a more compact state compared to full length hPDI, which has an anomalously large elution profile. This finding is also consistent with probable 'capping' of the **b'** binding site by the **x-linker** region, which if free in solution might be expected to add to the protein's hydrodynamic volume.

Capping was evident additionally in Chapter 4, by observation of two distinct resonances of  $^{347}\text{Trp}$  indole in  $^{15}\text{N},^1\text{H}$ -HSQC spectra of both **b'x** and W111F **abb'x**. Upon titration of a binding ligand into both proteins in Chapter 5, the Trp resonance corresponding to 'capped' protein shifts to overlay with the 'uncapped' resonance, indicative of an uncapped 'ligand bound' protein species. This event was also observed by  $^{19}\text{F}$  1D spectra of 6-F-Trp W111F **abb'x** in Chapter 7.

Despite the relatively large size of W111F **abb'x**, 68% of backbone resonances were assigned by NMR spectroscopy. In addition, secondary structure prediction by DANGLE closely agreed with the equivalent portion of the recently published **abb'xa'** crystal structure (Wang et al., 2013a).

Like many PDI fragments, W111F **abb'x** gave optimal spectra as a reduced protein, with a significant number of peaks in oxidised spectra appearing much broader or absent from the spectra. Minimal chemical shift maps of the protein show that all domains of W111F **abb'x** are perturbed upon a change in redox state. The redox potential of W111F **abb'x** was determined in Chapter 6 as  $E'_o$   $-167.1 \pm 2.1$  mV, a value in close agreement with the published **a** domain redox potential of full length hPDI (Chambers et al., 2010). Interestingly work carried out in this thesis, along with previously published work by the laboratory have shown that the redox potential of **a** domain becomes progressively more reducing upon addition of extra domains, to such an extent that GSH is no longer able to fully reduce the W111F **abb'x** fragment, when used in a 20x molar excess.

This observation of changing redox potential helps to demonstrate cross-talk among the four domains of hPDI. It is clear more work needs to be carried out to explore this notion, especially as preliminary work in our laboratory suggests **a** and **a'** as part of **abb'xa'** have markedly different redox potentials. Current literature states the two active sites of hPDI have redox potentials within 7 mV of one another, however we estimate much greater difference, similar to yPDI where a 30 mV difference is reported (Vitu et al., 2010).

The spectra of **b'x** and W111F **abb'x** assigned in Chapter 4 were used to confirm hPDI ligand binding of  $\Delta$ -som, with each fragment reporting  $K_d$  of  $103 \pm 47 \mu\text{M}$  and  $73 \pm 31 \mu\text{M}$  respectively, values that are moderately tighter than the estimated binding affinity of 0.1-1 mM (Klappa et al., 1997). Furthermore, the binding affinity was shown to be unaffected by protein redox state or the presence of neighbouring domains to the primary ligand binding site in **b'** domain, with the magnitude of shifts observed in the **ab** portion of W111F **abb'x** upon binding  $\Delta$ -som minimal in comparison to those observed in the **b'x** fragment. In light of this, subsequent peptide titrations were carried out using the smaller **b'x** fragment, expressed and purified immediately prior to NMR time to ensure only stable, monomeric protein was analysed.

Previously published work has demonstrated that the presence of a single Tyr or Trp residue in the peptide sequence is sufficient to promote binding of a peptide to the PDI family member PDIp. In Chapter 5 when these residues of  $\Delta$ -som were mutated to Phe, PDIp lost its ability to bind the peptide ligand, suggesting that PDIp ligand recognition was not mediated by Phe residues (Ruddock et al., 2000). Contrastingly, the three Phe residues of  $\Delta$ -som have proven crucial for binding recognition by hPDI in this study, with the substitution of a single Phe residue of  $\Delta$ -som weakening binding ~two-fold, the substitution of two Phe residues weakening binding still, and substitution of all three Phe residues resulting in a complete loss of binding affinity as measurable by NMR spectroscopy. This is despite there still being a Trp residue present in the peptide, proving that ligand specificity between hPDI and PDIp is clearly distinct.

An interesting observation was the C-terminal Phe residue ( $^{11}\text{Phe}$ ) of the peptide, which of the three Phe residues seemed most imperative to binding. Of the three Phe, it was the residue that reported the greatest difference in affinity between the single and triple

fluorinated peptides, and additionally its substitution to alanine had the greatest negative effect to binding of all the single substitutions. The single Phe 11→Ala substituted peptide reported a drop in affinity comparable to that observed for the double Phe 6,7→Ala peptide.

In summary, ligand titrations by NMR spectroscopy have allowed the binding affinity of hPDI to  $\Delta$ -som to be accurately quantified, and the  $^{19}\text{F}$  ligand-observed data closely supported the data collected by  $^{15}\text{N},^1\text{H}$ -HSQC protein-observed methods. Generally the curve fits published in this thesis do not reach 100% occupancy, due to a requirement of 0.25 mM minimum protein concentration for collection of adequate spectra. The titration end point was also limited to a 2.5x molar excess of peptide, as several  $\Delta$ -som peptides hydrolysed within hours of storage at  $>0.7$  mM in aqueous solutions. Hence, graphical fits were limited by practical working concentrations of protein and peptide, and it is not surprising the  $K_d$  fits quoted in this thesis have relatively large error ranges associated with them.

Overall the collection methods undertaken in this thesis were straightforward and reproducible, but additionally the  $^{19}\text{F}$  data collected in Chapters 5 and 7 also gave greater precision and was less labour intensive to analyse when compared to  $^{15}\text{N},^1\text{H}$ -HSQC spectra. In this instance, use of fluorinated peptides has provided a suitable method by which to probe ligand binding of hPDI, when traditional recombinant labelling techniques were met with limited success. The work described in this thesis could readily be extended to include substitution of the Trp residue of  $\Delta$ -som, or even a total alanine scan of the peptide. Indeed, if recombinant peptide production had been more straightforward it is likely several mutant peptides would have been expressed for study in NMR titrations.

The results obtained relating to the C-terminal Phe residue of  $\Delta$ -som may also explain observations made upon hPDI binding Pep20, a synthetic peptide used within our laboratory that binds to **b'x** with a marginally increased affinity. Interestingly Pep20 contains two Phe residues in its C-terminal region; further work could be carried out using fluoro-Phe incorporation into the peptide or alanine substitutions, to further develop the ideas proposed in this thesis.

If recombinant isotopic labelling of  $\Delta$ -som had been more straightforward, an additional aim of this thesis may have been to attempt a ligand bound structure of W111F **abb'x**.

Indeed, preliminary NOESY experiments were collected to gain structural restraint information upon binding, but collection of useful data was hindered by a running temperature incompatibility between peptide and protein. Owing to its large size, hPDI W111F **abb'x** gave best spectra at 37°C, however peptide was only visible by 2D NMR methods at temperatures lower than 10°C.

At the point of submission of this thesis, project students had demonstrated successful and straightforward purification of intact  $\Delta$ -som peptide, from GB1-FXa- $\Delta$ -som fusion protein expressed in LB media. The next step would be to try purification of isotopically labelled peptide from a minimal media growth, where overall expression (and therefore total fusion yields) will be lower. The final protein yield of intact peptide reported in this thesis is still lower than is routinely reported with GB1 fusions, however this could be due in part to the hydrophobic and unstable properties of the unstructured  $\Delta$ -som peptide.

Finally, owing to the relative ease with which  $^{19}\text{F}$  data was collected from fluorine containing peptides (Chapter 5), in Chapter 7 recombinant fluorination of the protein was trialled, to determine if ligand binding by  $^{19}\text{F}$  could also be collected from a protein-observed perspective, and additionally, if a fluorine probe at  $^{35}\text{Trp}$  of **a** domain could be used to calculate protein redox state, as was determined in this thesis using  $^{15}\text{N},^1\text{H}$ -HSQC titration (Chapter 6).

Fluorine incorporation into W111F **abb'x** was not as uniform as recently published for the **b'x** fragment (Curtis-Marof et al., 2014), and the presence of two potential fluorination sites complicated attempts to ascertain overall fluorine incorporation. A combination of mass spectrometry and  $^{19}\text{F}$  NMR were only able to give estimations: NMR could provide a relative ratio of fluorination at each site but could not report on un-fluorinated protein, whereas mass spectrometry could only report on relative amounts of non, single and double fluorinated protein. Together the data could estimate 30% un-fluorinated protein, 30% singly fluorinated at  $^{35}\text{Trp}$ , 10% singly fluorinated at  $^{347}\text{Trp}$ , and 30% double fluorinated protein. Such a heterogeneous population limited the application of 6-fluoro-Trp W111F **abb'x** in redox and ligand binding studies. In addition, analysis of the  $^{347}\text{Trp}$  fluorine peak in  $^{19}\text{F}$  1D spectra was complicated by the **x-linker's** ability to exist in solution in both capped and uncapped states.

Whilst a general trend from mostly ‘capped’ species in reduced spectra to a mostly ‘uncapped’ species in oxidised spectra could be observed, the **x-linker** could not be used to probe redox state or accurately report on ligand binding.

Interestingly, the  $^{35}\text{Trp}$  present in **a** domain reported perturbation upon ligand binding- this residue and both of the active site cysteines of W111F **abb’x** were all assignable by triple resonance, yet reported no significant perturbation upon ligand binding when observed by  $^{15}\text{N}, ^1\text{H}$ -HSQC. Hence, this observation by  $^{19}\text{F}$  is particularly relevant given the Wang groups recent publication of hPDI modelling simulations, which indicates the **a** domain of hPDI comes into much closer contact with **b’** in solution than is shown by their **abb’xa’** crystal structures (Yang et al., 2014).

Finally,  $^{19}\text{F}$  spectra revealed the  $^{35}\text{Trp}$  residue could not be used as a reporter of protein redox, as this resonance reported markedly different redox states depending on the redox buffer used. This was not something observed by  $^{15}\text{N}, ^1\text{H}$ -HSQC titration when using non-fluorinated protein, collected using both glutathione and dithiothreitol redox buffers.

It is not apparent if this observation was an effect of fluorine modifying the redox ability of W111F **abb’x**, or if  $^{19}\text{F}$  Trp is able to promote formation of a mixed disulphide state between DTT and the catalytic site of **a** domain. This work further suggests that the highly conserved  $^{35}\text{Trp}$  residue of PDI family members may be integral to protein redox control, and clearly warrants further investigation.

In closing, this thesis has confirmed peptide ligand binding of  $\Delta$ -som by hPDI occurs independently of protein redox state and neighbouring domains, unlike binding of larger mis-folded proteins, where additional hydrophobic patches in **a**, **b** and **b’** are thought to contribute to binding. Site specific fluorination and alanine substitutions of the  $\Delta$ -som peptide have indicated that the three Phe residues of  $\Delta$ -som are together crucial for binding recognition by hPDI, and suggest the C-terminal region of the peptide is most important for binding affinity. It is hoped future work will be carried out to more ascertain any potential binding motif of hPDI, with the ultimate goal of designing specific peptide inhibitors to hPDI for modulation of *in vitro* and *in vivo* hPDI activity.

## References

- Alanen, H.I., Salo, K.E., Pekkala, M., Siekkinen, H.M., Pirneskoski, A., Ruddock, L.W., 2003. Defining the domain boundaries of the human protein disulfide isomerases. *Antioxidants & Redox Signaling* 5, 367-374.
- Alanen, H.I., Salo, K.E., Pirneskoski, A., Ruddock, L.W., 2006a. pH dependence of the peptide thiol-disulfide oxidase activity of six members of the human protein disulfide isomerase family. *Antioxidants & Redox Signaling* 8, 283-291.
- Alanen, H.I., Williamson, R.A., Howard, M.J., Hatahet, F.S., Salo, K.E., Kauppila, A., Kellokumpu, S., Ruddock, L.W., 2006b. ERp27, a new non-catalytic endoplasmic reticulum-located human protein disulfide isomerase family member, interacts with ERp57. *The Journal of Biological Chemistry* 281, 33727-33738.
- Ali Khan, H., Mutus, B., 2014. Protein disulfide isomerase a multifunctional protein with multiple physiological roles. *Frontiers in Chemistry* 2, 70.
- Amin, N.T., Wallis, A.K., Wells, S.A., Rowe, M.L., Williamson, R.A., Howard, M.J., Freedman, R.B., 2013. High-resolution NMR studies of structure and dynamics of human ERp27 indicate extensive interdomain flexibility. *The Biochemical Journal* 450, 321-332.
- Andersson, M.I., MacGowan, A.P., 2003. Development of the quinolones. *The Journal of Antimicrobial Chemotherapy* 51 Suppl 1, 1-11.
- Andreu, C.I., Woehlbier, U., Torres, M., Hetz, C., 2012. Protein disulfide isomerases in neurodegeneration: From disease mechanisms to biomedical applications. *FEBS Letters* 586, 2826-2834.
- Anelli, T., Alessio, M., Mezghrani, A., Simmen, T., Talamo, F., Bachi, A., Sitia, R., 2002. ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. *The EMBO Journal* 21, 835-844.
- Bai, X.C., McMullan, G., Scheres, S.H., 2015. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences* 40, 49-57.
- Baker, K.M., Chakravarthi, S., Langton, K.P., Sheppard, A.M., Lu, H., Bulleid, N.J., 2008. Low reduction potential of Ero1alpha regulatory disulphides ensures tight control of substrate oxidation. *The EMBO Journal* 27, 2988-2997.
- Baneyx, F., 1999. Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* 10, 411-421.

- Barbouche, R., Miquelis, R., Jones, I.M., Fenouillet, E., 2003. Protein-disulfide isomerase-mediated reduction of two disulfide bonds of HIV envelope glycoprotein 120 occurs post-CXCR4 binding and is required for fusion. *The Journal of Biological Chemistry* 278, 3131-3136.
- Bastos-Aristizabal, S., Kozlov, G., Gehring, K., 2014. Structural insight into the dimerization of human protein disulfide isomerase. *Protein Science : a Publication of the Protein Society* 23, 618-626.
- Bax, A., Ikura, M., 1991. An efficient 3D NMR technique for correlating the proton and <sup>15</sup>N backbone amide resonances with the  $\alpha$ -carbon of the preceding residue in uniformly <sup>15</sup>N/<sup>13</sup>C enriched proteins. *Journal of Biomolecular NMR* 1, 99-104.
- Bi, S., Hong, P.W., Lee, B., Baum, L.G., 2011. Galectin-9 binding to cell surface protein disulfide isomerase regulates the redox environment to enhance T-cell migration and HIV entry. *Proceedings of the National Academy of Sciences of the USA* 108, 10650-10655.
- Bianco, A., Townsley, F.M., Greiss, S., Lang, K., Chin, J.W., 2012. Expanding the genetic code of *Drosophila melanogaster*. *Nature Chemical Biology* 8, 748-750.
- Bohm, H.J., Banner, D., Bendels, S., Kansy, M., Kuhn, B., Muller, K., Obst-Sander, U., Stahl, M., 2004. Fluorine in medicinal chemistry. *Chembiochem : a European Journal of Chemical Biology* 5, 637-643.
- Broos, J., Gabellieri, E., Biemans-Oldehinkel, E., Strambini, G.B., 2003. Efficient biosynthetic incorporation of tryptophan and indole analogs in an integral membrane protein. *Protein Science* 12, 1991-2000.
- Bulaj, G., 2005. Formation of disulfide bonds in proteins and peptides. *Biotechnology Advances* 23, 87-92.
- Byrne, L.J., Sidhu, A., Wallis, A.K., Ruddock, L.W., Freedman, R.B., Howard, M.J., Williamson, R.A., 2009. Mapping of the ligand-binding site on the b' domain of human PDI: interaction with peptide ligands and the x-linker region. *The Biochemical Journal* 423, 209-217.
- Cai, H., Wang, C.C., Tsou, C.L., 1994. Chaperone-like activity of protein disulfide isomerase in the refolding of a protein with no disulfide bonds. *The Journal of Biological Chemistry* 269, 24550-24552.
- Chambers, J.E., Tavender, T.J., Oka, O.B., Warwood, S., Knight, D., Bulleid, N.J., 2010. The reduction potential of the active site disulfides of human protein disulfide

- isomerase limits oxidation of the enzyme by Ero1 $\alpha$ . *The Journal of Biological Chemistry* 285, 29200-29207.
- Chapeville, F., Lipmann, F., Ehrenstein, G., Weisblum, B., Ray, W.J., Benzer, S., 1962. On the role of soluble ribonucleic acid in coding for amino acids. *Proceedings of the National Academy of Sciences of the United States of America* 48, 1086-1092.
- Chazin, W.J., Goldenberg, D.P., Creighton, T.E., Wuthrich, K., 1985. Comparative studies of conformation and internal mobility in native and circular basic pancreatic trypsin inhibitor by  $^1\text{H}$  nuclear magnetic resonance in solution. *European Journal of Biochemistry / FEBS* 152, 429-437.
- Cheung, M.-S., Maguire, M.L., Stevens, T.J., Broadhurst, R.W., 2010. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance* 202, 223-233.
- Conn, K.J., Gao, W., McKee, A., Lan, M.S., Ullman, M.D., Eisenhauer, P.B., Fine, R.E., Wells, J.M., 2004. Identification of the protein disulfide isomerase family member PDip in experimental Parkinson's disease and Lewy body pathology. *Brain Research* 1022, 164-172.
- Crowley, P.B., Kyne, C., Monteith, W.B., 2012. Simple and inexpensive incorporation of  $^{19}\text{F}$ -tryptophan for protein NMR spectroscopy. *Chemical Communications (Cambridge)* 48, 10681-10683.
- Curtis-Marof, R., Doko, D., Rowe, M.L., Richards, K.L., Williamson, R.A., Howard, M.J., 2014.  $^{19}\text{F}$  NMR spectroscopy monitors ligand binding to recombinantly fluorine-labelled b'x from human protein disulphide isomerase (hPDI). *Organic & Biomolecular Chemistry* 12, 3808-3812.
- Daniels, R., Kurowski, B., Johnson, A.E., Hebert, D.N., 2003. N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. *Molecular Cell* 11, 79-90.
- Darby, N.J., Creighton, T.E., 1995. Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* 34, 16770-16780.
- Darby, N.J., Penka, E., Vincentelli, R., 1998. The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency. *Journal of Molecular Biology* 276, 239-247.

- David, B. Harper, D., 1999. Fluorine-containing natural products. *Journal of Fluorine Chemistry* 100, 127-133.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A., 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* 6, 277-293.
- Denisov, A.Y., Maattanen, P., Dabrowski, C., Kozlov, G., Thomas, D.Y., Gehring, K., 2009. Solution structure of the bb' domains of human protein disulfide isomerase. *The FEBS Journal* 276, 1440-1449.
- Diaz-Villanueva, J.F., Diaz-Molina, R., Garcia-Gonzalez, V., 2015. Protein Folding and Mechanisms of Proteostasis. *International Journal of Molecular Sciences* 16, 17193-17230.
- Dickerhof, N., Kleffmann, T., Jack, R., McCormick, S., 2011. Bacitracin inhibits the reductive activity of protein disulfide isomerase by disulfide bond formation with free cysteines in the substrate-binding domain. *The FEBS Journal* 278, 2034-2043.
- Eirich, J., Braig, S., Schyschka, L., Servatius, P., Hoffmann, J., Hecht, S., Fulda, S., Zahler, S., Antes, I., Kazmaier, U., Sieber, S.A., Vollmar, A.M., 2014. A Small Molecule Inhibits Protein Disulfide Isomerase and Triggers the Chemosensitization of Cancer Cells. *Angewandte Chemie International Edition* 53, 12960-12965.
- Ellgaard, L., Ruddock, L.W., 2005. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Reports* 6, 28-32.
- Essex, D.W., Li, M., 1999. Protein disulphide isomerase mediates platelet aggregation and secretion. *British Journal of Haematology* 104, 448-454.
- Fesik, S.W., Luly, J.R., Erickson, J.W., Abad-Zapatero, C., 1988. Isotope-edited proton NMR study on the structure of a pepsin/inhibitor complex. *Biochemistry* 27, 8297-8301.
- Fielding, L., 2003. NMR methods for the determination of protein-ligand dissociation constants. *Current Topics in Medicinal Chemistry* 3, 39-53.
- Freedman, R.B., 1995. The formation of protein disulphide bonds. *Current Opinion in Structural Biology* 5, 85-91.
- Freedman, R.B., Hirst, T.R., Tuite, M.F., 1994. Protein disulphide isomerase: building bridges in protein folding. *Trends in Biochemical Sciences* 19, 331-336.

- Freedman, R.B., Klappa, P., Ruddock, L.W., 2002. Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. *EMBO Reports* 3, 136-140.
- Frueh, D.P., 2014. Practical aspects of NMR signal assignment in larger and challenging proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* 78, 47-75.
- Fu, X., Wang, P., Fukui, M., Long, C., Yin, L., Choi, H.J., Zhu, B.T., 2012. PDIp is a major intracellular oestrogen-storage protein that modulates tissue levels of oestrogen in the pancreas. *The Biochemical Journal* 447, 115-123.
- Fu, X.M., Dai, X., Ding, J., Zhu, B.T., 2009. Pancreas-specific protein disulfide isomerase has a cell type-specific expression in various mouse tissues and is absent in human pancreatic adenocarcinoma cells: implications for its functions. *Journal of Molecular Histology* 40, 189-199.
- Fu, X.M., Wang, P., Zhu, B.T., 2011. Characterization of the estradiol-binding site structure of human pancreas-specific protein disulfide isomerase: indispensable role of the hydrogen bond between His278 and the estradiol 3-hydroxyl group. *Biochemistry* 50, 106-115.
- Ge, J., Zhang, C.J., Li, L., Chong, L.M., Wu, X., Hao, P., Sze, S.K., Yao, S.Q., 2013. Small molecule probe suitable for in situ profiling and inhibition of protein disulfide isomerase. *ACS Chemical Biology* 8, 2577-2585.
- Geoghegan, K.F., Dixon, H.B., Rosner, P.J., Hoth, L.R., Lanzetti, A.J., Borzilleri, K.A., Marr, E.S., Pezzullo, L.H., Martin, L.B., LeMotte, P.K., McColl, A.S., Kamath, A.V., Stroh, J.G., 1999. Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in *Escherichia coli*: the cause of extra mass of 258 or 178 Da in fusion proteins. *Analytical Biochemistry* 267, 169-184.
- Gerig, J.T., 1994. Fluorine NMR of proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* 26, Part 4, 293-370.
- Gilbert, H.F., 1997. Protein disulfide isomerase and assisted protein folding. *Journal Biological Chemistry* 272, 29399-29402.
- Gillece, P., Luz, J.M., Lennarz, W.J., de la Cruz, F.J., Romisch, K., 1999. Export of a cysteine-free misfolded secretory protein from the endoplasmic reticulum for degradation requires interaction with protein disulphide isomerase. *Journal of Cell Biology* 147, 1443-1456.

- Göbl, C., Madl, T., Simon, B., Sattler, M., 2014. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy* 80, 26-63.
- Goldberger, R.F., Epstein, C.J., Anfinsen, C.B., 1963. Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *The Journal of Biological Chemistry* 238, 628-635.
- Goodlett, D.R., Armstrong, F.B., Creech, R.J., van Breemen, R.B., 1990. Formylated peptides from cyanogen bromide digests identified by fast atom bombardment mass spectrometry. *Analytical Biochemistry* 186, 116-120.
- Gray, H.B., Winkler, J.R., 2015. Hole hopping through tyrosine/tryptophan chains protects proteins from oxidative damage. *Proceedings of the National Academy of Sciences of the United States of America* 112, 10920-10925.
- Greiss, S., Chin, J.W., 2011. Expanding the genetic code of an animal. *Journal of the American Chemical Society* 133, 14196-14199.
- Grek, C., Townsend, D.M., 2014. Protein Disulfide Isomerase Superfamily in Disease and the Regulation of Apoptosis. *Endoplasmic Reticulum Stress in Diseases* 1, 4-17.
- Grzesiek, S., Bax, A., 1992. Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr. *Journal of the American Chemical Society* 114, 6291-6293.
- Hajduk, P.J., Augeri, D.J., Mack, J., Mendoza, R., Yang, J., Betz, S.F., Fesik, S.W., 2000. NMR-Based Screening of Proteins Containing <sup>13</sup>C-Labeled Methyl Groups. *Journal of the American Chemical Society* 122, 7898-7904.
- Hammond, C., Helenius, A., 1995. Quality control in the secretory pathway. *Current Opinion in Cell Biology* 7, 523-529.
- Hancock, S.M., Uprety, R., Deiters, A., Chin, J.W., 2010. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *Journal of the American Chemical Society* 132, 14819-14824.
- Hatahet, F., Ruddock, L.W., 2007. Substrate recognition by the protein disulfide isomerases. *The FEBS Journal* 274, 5223-5234.
- Hatahet, F., Ruddock, L.W., 2009. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. *Antioxidants & Redox Signaling* 11, 2807-2850.

- Hendershot, L.M., Ting, J., Lee, A.S., 1988. Identity of the immunoglobulin heavy-chain-binding protein with the 78,000-dalton glucose-regulated protein and the role of posttranslational modifications in its binding function. *Molecular and Cellular Biology* 8, 4250-4256.
- Holländer, H., Amrhein, N., 1980. The Site of the Inhibition of the Shikimate Pathway by Glyphosate: I. Inhibition by glyphosate of pheylpropanoid synthesis in buckwheat. *Plant Physiology* 66, 823-829.
- Horng, J.C., Raleigh, D.P., 2003. phi-Values beyond the ribosomally encoded amino acids: kinetic and thermodynamic consequences of incorporating trifluoromethyl amino acids in a globular protein. *Journal of the American Chemical Society* 125, 9286-9287.
- Huth, J.R., Park, C., Petros, A.M., Kunzer, A.R., Wendt, M.D., Wang, X., Lynch, C.L., Mack, J.C., Swift, K.M., Judge, R.A., Chen, J., Richardson, P.L., Jin, S., Tahir, S.K., Matayoshi, E.D., Dorwin, S.A., Lador, U.S., Severin, J.M., Walter, K.A., Bartley, D.M., Fesik, S.W., Elmore, S.W., Hajduk, P.J., 2007. Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies. *Chemical Biology & Drug Design* 70, 1-12.
- Hwang, P.M., Pan, J.S., Sykes, B.D., 2014. Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*. *FEBS Letters* 588, 247-252.
- Hwang, T.L., Shaka, A.J., 1995. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* 112, 275-279.
- Inaba, K., Masui, S., Iida, H., Vavassori, S., Sitia, R., Suzuki, M., 2010. Crystal structures of human Ero1 $\alpha$  reveal the mechanisms of regulated and targeted oxidation of PDI. *The EMBO Journal* 29, 3330-3343.
- Jain, S., McGinnes, L.W., Morrison, T.G., 2009. Overexpression of thiol/disulfide isomerases enhances membrane fusion directed by the Newcastle disease virus fusion protein. *Journal of Virology* 82, 12039-12048.
- Jenny, R.J., Mann, K.G., Lundblad, R.L., 2003. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expression and Purification* 31, 1-11.

- Jeon, G.S., Nakamura, T., Lee, J.S., Choi, W.J., Ahn, S.W., Lee, K.W., Sung, J.J., Lipton, S.A., 2014. Potential effect of S-nitrosylated protein disulfide isomerase on mutant SOD1 aggregation and neuronal cell death in amyotrophic lateral sclerosis. *Molecular Neurobiology* 49, 796-807.
- Jessop, C.E., Watkins, R.H., Simmons, J.J., Tasab, M., Bulleid, N.J., 2009. Protein disulphide isomerase family members show distinct substrate specificity: P5 is targeted to BiP client proteins. *Journal of Cell Science* 122, 4287-4295.
- Jonasson, P., Nygren, P.A., Johansson, B.L., Wahren, J., Uhlen, M., Stahl, S., 1998. Gene fragment polymerization gives increased yields of recombinant human proinsulin C-peptide. *Gene* 210, 203-210.
- Jordan, P.A., Gibbins, J.M., 2006. Extracellular disulfide exchange and the regulation of cellular function. *Antioxidant Redox Signalling* 8, 312-324.
- Kanai, S., Toh, H., Hayano, T., Kikuchi, M., 1998. Molecular evolution of the domain structures of protein disulfide isomerases. *Journal of Molecular Evolution* 47, 200-210.
- Kaplan, A., Stockwell, B.R., 2015. Structural Elucidation of a Small Molecule Inhibitor of Protein Disulfide Isomerase. *ACS Medicinal Chemistry Letters*.
- Karala, A.R., Lappi, A.K., Ruddock, L.W., 2010. Modulation of an active-site cysteine pKa allows PDI to act as a catalyst of both disulfide bond formation and isomerization. *Journal of Molecular Biology* 396, 883-892.
- Kay, L.E., Ikura, M., Tschudin, R., Bax, A., 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* 89, 496-514.
- Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M., Creighton, T.E., 1997. The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. *Current Biology : CB* 7, 239-245.
- Khodier, C., VerPlank, L., Nag, P.P., Pu, J., Wurst, J., Pilyugina, T., Dockendorff, C., Galinski, C.N., Scalise, A.A., Passam, F., van Hessem, L., Dilks, J., Kennedy, D.R., Flaumenhaft, R., Palmer, M.A.J., Dandapani, S., Munoz, B., Schrieber, S.L. 2010. Identification of ML359 as a Small Molecule Inhibitor of Protein Disulfide Isomerase, In: *Probe Reports from the NIH Molecular Libraries Program*. Bethesda MD.

- Klappa, P., Freedman, R.B., Langenbuch, M., Lan, M.S., Robinson, G.K., Ruddock, L.W., 2001. The pancreas-specific protein disulphide-isomerase PDIP interacts with a hydroxyaryl group in ligands. *The Biochemical Journal* 354, 553-559.
- Klappa, P., Hawkins, H.C., Freedman, R.B., 1997. Interactions between protein disulphide isomerase and peptides. *European Journal of Biochemistry / FEBS* 248, 37-42.
- Klappa, P., Koivunen, P., Pirneskoski, A., Karvonen, P., Ruddock, L.W., Kivirikko, K.I., Freedman, R.B., 2000. Mutations that destabilize the a' domain of human protein-disulfide isomerase indirectly affect peptide binding. *The Journal of Biological Chemistry* 275, 13213-13218.
- Klappa, P., Ruddock, L.W., Darby, N.J., Freedman, R.B., 1998a. The b' domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. *The EMBO Journal* 17, 927-935.
- Klappa, P., Stromer, T., Zimmermann, R., Ruddock, L.W., Freedman, R.B., 1998c. A pancreas-specific glycosylated protein disulphide-isomerase binds to misfolded proteins and peptides with an interaction inhibited by oestrogens. *European Journal of Biochemistry / FEBS* 254, 63-69.
- Kober, F.-X., Koelmel, W., Kuper, J., Drechsler, J., Mais, C., Hermanns, H.M., Schindelin, H., 2013. The Crystal Structure of the Protein-Disulfide Isomerase Family Member ERp27 Provides Insights into Its Substrate Binding Capabilities. *The Journal of Biological Chemistry* 288, 2029-2039.
- Koenig, B.W., Rogowski, M., Louis, J.M., 2003. A rapid method to attain isotope labeled small soluble peptides for NMR studies. *Journal of Biomolecular NMR* 26, 193-202.
- Kojima, R., Okumura, M., Masui, S., Kanemura, S., Inoue, M., Saiki, M., Yamaguchi, H., Hikima, T., Suzuki, M., Akiyama, S., Inaba, K., 2014. Radically different thioredoxin domain arrangement of ERp46, an efficient disulfide bond introducer of the mammalian PDI family. *Structure (London, England : 1993)* 22, 431-443.
- Koritzinsky, M., Levitin, F., van den Beucken, T., Rumantir, R.A., Harding, N.J., Chu, K.C., Boutros, P.C., Braakman, I., Wouters, B.G., 2013. Two phases of disulfide bond formation have differing requirements for oxygen. *The Journal of Cell Biology* 203, 615-627.

- Kortemme, T., Creighton, T.E., 1995. Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family. *Journal of Molecular Biology* 253, 799-812.
- Koslov, G., Maattanen, P., Schrag, J.D., Pollock, S., Cygler, M., Nagar, M., Thomas, D.Y., Gehring, K., 2006. Crystal structure of the bb' domains of the protein disulfide isomerase ERp57. *Structure* 14, 1331-1339.
- Kozlov, G., Maattanen, P., Thomas, D.Y., Gehring, K., 2010. A structural overview of the PDI family of proteins. *The FEBS Journal* 277, 3924-3936.
- Kuliopulos, A., Walsh, C.T., 1994. Production, Purification, and Cleavage of Tandem Repeats of Recombinant Peptides. *Journal of the American Chemical Society* 116, 4599-4607.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Lappi, A.K., Lensink, M.F., Alanen, H.I., Salo, K.E., Lobell, M., Juffer, A.H., Ruddock, L.W., 2004. A conserved arginine plays a role in the catalytic cycle of the protein disulphide isomerases. *Journal of Molecular Biology* 335, 283-295.
- Lash, A.E., Tolstoshev, C.M., Wagner, L., Schuler, G.D., Strausberg, R.L., Riggins, G.J., Altschul, S.F., 2000. SAGEmap: a public gene expression resource. *Genome Research* 10, 1051-1060.
- Lee, S., Park, B., Kang, K., Ahn, K., 2009. Redox-regulated Export of the Major Histocompatibility Complex Class I-Peptide Complexes from the Endoplasmic Reticulum. *Molecular Biology of the Cell* 20, 3285-3294.
- Lee, S.O., Cho, K., Cho, S., Kim, I., Oh, C., Ahn, K., 2010. Protein disulphide isomerase is required for signal peptide peptidase-mediated protein degradation. *The EMBO Journal* 29, 363-375.
- LeMaster, D.M., 1990. Deuterium labelling in NMR structural analysis of larger proteins. *Quarterly Reviews of Biophysics* 23, 133-174.
- Li, S.J., Hong, X.G., Shi, Y.Y., Li, H., Wang, C.C., 2006. Annular arrangement and collaborative actions of four domains of protein-disulfide isomerase: a small angle X-ray scattering study in solution. *The Journal of Biological Chemistry* 281, 6581-6588.

- Li, Y.-J., Rothwarf, D.M., Scheraga, H.A., 1998. An Unusual Adduct of Dithiothreitol with a Pair of Cysteine Residues of a Protein as a Stable Folding Intermediate. *Journal of the American Chemical Society* 120, 2668-2669.
- Liew, O.W., Ching Chong, J.P., Yandle, T.G., Brennan, S.O., 2005. Preparation of recombinant thioredoxin fused N-terminal proCNP: Analysis of enterokinase cleavage products reveals new enterokinase cleavage sites. *Protein Expression and Purification* 41, 332-340.
- Light, A., Janska, H., 1989. Enterokinase (enteropeptidase): comparative aspects. *Trends in Biochemical Sciences* 14, 110-112.
- Linz, A., Knieper, Y., Gronau, T., Hansen, U., Aszodi, A., Garbi, N., Hammerling, G.J., Pap, T., Bruckner, P., Dreier, R., 2015. ER Stress During the Pubertal Growth Spurt Results in Impaired Long-Bone Growth in Chondrocyte-Specific ERp57 Knockout Mice. *Journal of bone and mineral research : The Official Journal of the American Society for Bone and Mineral Research* 30, 1481-1493.
- Lundstrom, J., Holmgren, A., 1993. Determination of the reduction-oxidation potential of the thioredoxin-like domains of protein disulfide-isomerase from the equilibrium with glutathione and thioredoxin. *Biochemistry* 32, 6649-6655.
- Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., Bax, A., 1989. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *Journal of the American Chemical Society* 111, 1515-1517.
- Martin, J.L., 1995. Thioredoxin —a fold for all reasons. *Structure (London, England : 1993)* 3, 245-250.
- Matchett, W.H., 1972. Inhibition of Tryptophan Synthetase by Indoleacrylic Acid. *Journal of Bacteriology* 110, 146-154.
- Matos, C.F., Robinson, C., Alanen, H.I., Prus, P., Uchida, Y., Ruddock, L.W., Freedman, R.B., Keshavarz-Moore, E., 2014. Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. *Biotechnology Progress* 30, 281-290.
- Merrifield, R.B., 1963. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* 85, 2149-2154.
- Morton, C.J., Pugh, D.J., Brown, E.L., Kahmann, J.D., Renzoni, D.A., Campbell, I.D., 1996. Solution structure and peptide binding of the SH3 domain from human Fyn. *Structure* 4, 705-714.

- Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., Yokoyama, S., 2008. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochemical and Biophysical Research Communications* 371, 818-822.
- Neumann, H., Peak-Chew, S.Y., Chin, J.W., 2008. Genetically encoding N[epsilon]-acetyllysine in recombinant proteins. *Nature Chemical Biology* 4, 232-234.
- Ngo, H., Harris, R., Kimmich, N., Casino, P., Niks, D., Blumenstein, L., Barends, T.R., Kulik, V., Weyand, M., Schlichting, I., Dunn, M.F., 2007. Synthesis and Characterization of Allosteric Probes of Substrate Channeling in the Tryptophan Synthase Bienenzyme Complex. *Biochemistry* 46, 7713-7727.
- Nguyen, V.D., Wallis, K., Howard, M.J., Haapalainen, A.M., Salo, K.E., Saaranen, M.J., Sidhu, A., Wierenga, R.K., Freedman, R.B., Ruddock, L.W., Williamson, R.A., 2008. Alternative conformations of the x region of human protein disulphide-isomerase modulate exposure of the substrate binding b' domain. *Journal of Molecular Biology* 383, 1144-1155.
- Ni, M., Lee, A.S., 2007. ER chaperones in mammalian development and human diseases. *FEBS Letters* 581, 3641-3651.
- Nilsson, B.L., Soellner, M.B., Raines, R.T., 2005. Chemical synthesis of proteins. *Annual Review of Biophysics and Biomolecular Structure* 34, 91-118.
- Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., Schultz, P.G., 1989. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science (New York, N.Y.)* 244, 182-188.
- Okumura, M., Kadokura, H., Inaba, K., 2015. Structures and functions of protein disulfide isomerase family members involved in proteostasis in the endoplasmic reticulum. *Free Radical Biology and Medicine* 83, 314-322.
- Oliver, J.D., van der Wal, F.J., Bulleid, N.J., High, S., 1997. Interaction of the thiol-dependent reductase ERp57 with nascent glycoproteins. *Science* 275, 86-88.
- Oliver, J.D., Roderick, H.L., Llewellyn, D.H., High, S., 1999. ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Molecular Biology of the Cell* 10, 2573-2582.
- Pagani, M., Fabbri, M., Benedetti, C., Fassio, A., Pilati, S., Bulleid, N.J., Cabibbo, A., Sitia, R., 2000. Endoplasmic reticulum oxidoreductin 1-lbeta (ERO1-Lbeta), a

- human gene induced in the course of the unfolded protein response. *The Journal of Biological Chemistry* 275, 23685-23692.
- Parakh, S., Atkin, J.D., 2015. Novel roles for protein disulphide isomerase in disease states: a double edged sword? *Frontiers in Cell and Developmental Biology* 3, 30.
- Parrish, A.R., She, X., Xiang, Z., Coin, I., Shen, Z., Briggs, S.P., Dillin, A., Wang, L., 2012. Expanding the genetic code of *Caenorhabditis elegans* using bacterial aminoacyl-tRNA synthetase/tRNA pairs. *ACS Chemical Biology* 7, 1292-1302.
- Piotto, M., Saudek, V., Sklenar, V., 1992. Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions. *Journal of Biomolecular NMR* 2, 661-665.
- Pirneskoski, A., Klappa, P., Lobell, M., Williamson, R.A., Byrne, L., Alanen, H.I., Salo, K.E., Kivirikko, K.I., Freedman, R.B., Ruddock, L.W., 2004. Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase. *The Journal of Biological Chemistry* 279, 10374-10381.
- Puig, A., Lyles, M.M., Noiva, R., Gilbert, H.F., 1994. The role of the thiol/disulfide centers and peptide binding site in the chaperone and anti-chaperone activities of protein disulfide isomerase. *Journal of Biological Chemistry* 269, 19128-19135.
- Quan, H., Fan, G., Wang, C.C., 1995. Independence of the chaperone activity of protein disulfide isomerase from its thioredoxin-like active site. *The Journal of Biological Chemistry* 270, 17078-17080.
- Quinternet, M., Starck, J.P., Delsuc, M.A., Kieffer, B., 2012. Unraveling complex small-molecule binding mechanisms by using simple NMR spectroscopy. *Chemistry (Weinheim an der Bergstrasse, Germany)* 18, 3969-3974.
- Radcliffe, C.M., Diedrich, G., Harvey, D.J., Dwek, R.A., Cresswell, P., Rudd, P.M., 2002. Identification of specific glycoforms of major histocompatibility complex class I heavy chains suggests that class I peptide loading is an adaptation of the quality control pathway involving calreticulin and ERp57. *The Journal of Biological Chemistry* 277, 46415-46423.
- Reinhardt, C., von Bruhl, M.L., Manukyan, D., Grahl, L., Lorenze, M., Altmann, B., Dlugai, S., Hess, S., Konrad, I., Orschiedt, L., Mackman, N., Ruddock, L., Massberg, S., Engelmann, B., 2008. Protein disulfide isomerase acts as an injury

- response signal that enhances fibrin generation via tissue factor activation. *Journal of Clinical Investigation* 118, 1110-1122.
- Ruddock, L.W., 2012. Low-molecular-weight oxidants involved in disulfide bond formation. *Antioxidants & Redox Signaling* 16, 1129-1138.
- Ruddock, L.W., Freedman, R.B., Klappa, P., 2000. Specificity in substrate binding by protein folding catalysts: Tyrosine and tryptophan residues are the recognition motifs for the binding of peptides to the pancreas-specific protein disulfide isomerase PDI $\beta$ . *Protein Science* 9, 758-764.
- Rutkevich, L.A., Cohen-Doyle, M.F., Brockmeier, U., Williams, D.B., 2010. Functional relationship between protein disulfide isomerase family members during the oxidative folding of human secretory proteins. *Molecular Biology of the Cell* 21, 3093-3105.
- Saibil, H.R., 2008. Chaperone machines in action. *Current Opinion in Structural Biology* 18, 35-42.
- Salwiczek, M., Kokschi, B., 2009. Effects of fluorination on the folding kinetics of a heterodimeric coiled coil. *Chembiochem : a European Journal of Chemical Biology* 10, 2867-2870.
- Salwiczek, M., Nyakatura, E.K., Gerling, U.I., Ye, S., Kokschi, B., 2012. Fluorinated amino acids: compatibility with native protein structures and effects on protein-protein interactions. *Chemical Society Reviews* 41, 2135-2171.
- Sato, Y., Kojima, R., Okumura, M., Hagiwara, M., Masui, S., Maegawa, K., Saiki, M., Horibe, T., Suzuki, M., Inaba, K., 2013. Synergistic cooperation of PDI family members in peroxiredoxin 4-driven oxidative protein folding. *Scientific Reports* 3, 2456.
- Schagger, H., 2006. Tricine-SDS-PAGE. *Nature Protocols* 1, 16-22.
- Schagger, H., von Jagow, G., 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry* 166, 368-379.
- Schelhaas, M., Malmstrom, J., Pelkmans, L., Haugstetter, J., Ellgaard, L., Grunewald, K., Helenius, A., 2007. Simian Virus 40 depends on ER protein folding and quality control factors for entry into host cells. *Cell* 131, 516-529.

- Schumann, F.H., Riepl, H., Maurer, T., Gronwald, W., Neidig, K.P., Kalbitzer, H.R., 2007. Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *Journal of Biomolecular NMR* 39, 275-289.
- Sevier, C.S., Kaiser, C.A., 2002. Formation and transfer of disulphide bonds in living cells. *Nature Reviews Molecular Cell Biology* 3, 836-847.
- Shen, S.H., 1984. Multiple joined genes prevent product degradation in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 81, 4627-4631.
- Sitaraman, K., Chatterjee, D.K., 2009. High-throughput protein expression using cell-free system. *Methods in Molecular Biology (Clifton, N.J.)* 498, 229-244.
- Sitia, R., Braakman, I., 2003. Quality control in the endoplasmic reticulum protein factory. *Nature* 426, 891-894.
- Stamler, J.S., Lamas, S., Fang, F.C., 2001. Nitrosylation: The Prototypic Redox-Based Signaling Mechanism. *Cell* 106, 675-683.
- Tatu, U., Braakman, I., Helenius, A., 1993. Membrane glycoprotein folding, oligomerization and intracellular transport: effects of dithiothreitol in living cells. *EMBO Journal*; 12, 2151-2157.
- Taub, M., 1977. The mechanism of killing of mouse fibroblasts by the amino acid analogue 5-fluorotryptophan. *Journal of Cellular Physiology* 93, 189-195.
- Tavender, T.J., Bulleid, N.J., 2010. Molecular mechanisms regulating oxidative activity of the Ero1 family in the endoplasmic reticulum. *Antioxidants & Redox Signaling* 13, 1177-1187.
- Taylor, S.L., Crawley-Snowdon, H., Wagstaff, J.L., Rowe, M.L., Shepherd, M., Williamson, R.A., Howard, M.J., 2013. Measuring protein reduction potentials using <sup>15</sup>N HSQC NMR spectroscopy. *Chemical Communications (Cambridge)* 49, 1847-1849.
- Tian, G., Kober, F.X., Lewandrowski, U., Sickmann, A., Lennarz, W.J., Schindelin, H., 2008. The catalytic activity of protein-disulfide isomerase requires a conformationally flexible molecule. *The Journal of Biological Chemistry* 283, 33630-33640.
- Tokuhiro, K., Ikawa, M., Benham, A.M., Okabe, M., 2012. Protein disulfide isomerase homolog PDILT is required for quality control of sperm membrane protein

- ADAM3 and male fertility [corrected]. *Proceedings of the National Academy of Sciences of the United States of America* 109, 3850-3855.
- Tsai, B., Rodighiero, C., Lencer, W.I., Rapoport, T.A., 2001. Protein disulfide isomerase acts as a redox-dependent chaperone to unfold cholera toxin. *Cell* 104, 937-948.
- Tufo, G., Jones, A.W., Wang, Z., Hamelin, J., Tajeddine, N., Esposti, D.D., Martel, C., Boursier, C., Gallerne, C., Migdal, C., Lemaire, C., Szabadkai, G., Lemoine, A., Kroemer, G., Brenner, C., 2014. The protein disulfide isomerases PDIA4 and PDIA6 mediate resistance to cisplatin-induced cell death in lung adenocarcinoma. *Cell Death and Differentiation* 21, 685-695.
- Uehara, T., Nakamura, T., Yao, D., Shi, Z.Q., Gu, Z., Ma, Y., Masliah, E., Nomura, Y., Lipton, S.A., 2006. S-nitrosylated protein-disulphide isomerase links protein misfolding to neurodegeneration. *Nature* 441, 513-517.
- van Lith, M., Karala, A.R., Bown, D., Gatehouse, J.A., Ruddock, L.W., Saunders, P.T., Benham, A.M., 2007. A developmentally regulated chaperone complex for the endoplasmic reticulum of male haploid germ cells. *Molecular Biology of the Cell* 18, 2795-2804.
- Vaynberg, J., Qin, J., 2006. Weak protein-protein interactions as probed by NMR spectroscopy. *Trends in Biotechnology* 24, 22-27.
- Venetianer, P., Straub, F.B., 1963. The enzymic reactivation of reduced ribonuclease. *Biochimica et Biophysica Acta* 67, 166-168.
- Viegas, A., Manso, J., Nobrega, F.L., Cabrita, E.J., 2011. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *Journal of Chemical Education* 88, 990-994.
- Vitu, E., Kim, S., Sevier, C.S., Lutzky, O., Heldman, N., Bentzur, M., Unger, T., Yona, M., Kaiser, C.A., Fass, D., 2010. Oxidative activity of yeast Ero1p on protein disulfide isomerase and related oxidoreductases of the endoplasmic reticulum. *Journal of Biological Chemistry* 285, 18155-18165.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D., 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687-696.
- Wagstaff, J.L., Howard, M.J., Williamson, R.A., 2010. Production of recombinant isotopically labelled peptide by fusion to an insoluble partner protein: generation of

- integrin alphavbeta6 binding peptides for NMR. *Molecular bioSystems* 6, 2380-2385.
- Walker, A.K., Soo, K.Y., Levina, V., Talbo, G.H., Atkin, J.D., 2013. N-linked glycosylation modulates dimerization of protein disulfide isomerase family A member 2 (PDIA2). *The FEBS Journal* 280, 233-243.
- Wallis, A.K., Sidhu, A., Byrne, L.J., Howard, M.J., Ruddock, L.W., Williamson, R.A., Freedman, R.B., 2009. The ligand-binding b' domain of human protein disulphide-isomerase mediates homodimerization. *Protein Science : a publication of the Protein Society* 18, 2569-2577.
- Walter, P., Ron, D., 2011. The unfolded protein response: from stress pathway to homeostatic regulation. *Science (New York, N.Y.)* 334, 1081-1086.
- Wang, C., Chen, S., Wang, X., Wang, L., Wallis, A.K., Freedman, R.B., Wang, C.C., 2010. Plasticity of human protein disulfide isomerase: evidence for mobility around the X-linker region and its functional significance. *The Journal of Biological Chemistry* 285, 26788-26797.
- Wang, C., Li, W., Ren, J., Fang, J., Ke, H., Gong, W., Feng, W., Wang, C.C., 2013. Structural insights into the redox-regulated dynamic conformations of human protein disulfide isomerase. *Antioxidants & Redox Signaling* 19, 36-45.
- Wang, C., Yu, J., Huo, L., Wang, L., Feng, W., Wang, C.C., 2012. Human protein-disulfide isomerase is a redox-regulated chaperone activated by oxidation of domain a'. *The Journal of Biological Chemistry* 287, 1139-1149.
- Wang, C.C., Tsou, C.L., 1993. Protein disulfide isomerase is both an enzyme and a chaperone. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 7, 1515-1517.
- Wang, L., Li, S.J., Sidhu, A., Zhu, L., Liang, Y., Freedman, R.B., Wang, C.C., 2009. Reconstitution of human Ero1-Lalpha/protein-disulfide isomerase oxidative folding pathway in vitro. Position-dependent differences in role between the a and a' domains of protein-disulfide isomerase. *The Journal of Biological Chemistry* 284, 199-206.
- Wang, L., Vavassori, S., Li, S., Ke, H., Anelli, T., Degano, M., Ronzoni, R., Sitia, R., Sun, F., Wang, C.C., 2008. Crystal structure of human ERp44 shows a dynamic functional modulation by its carboxy-terminal tail. *EMBO Reports* 9, 642-647.

- Wang, L., Wang, X., Wang, C.C., 2015. Protein disulfide-isomerase, a folding catalyst and a redox-regulated chaperone. *Free Radical Biology & Medicine* 83, 305-313.
- Wang, L., Zhang, L., Niu, Y., Sitia, R., Wang, C.C., 2014. Glutathione peroxidase 7 utilizes hydrogen peroxide generated by Ero1alpha to promote oxidative protein folding. *Antioxidants & Redox Signaling* 20, 545-556.
- Watanabe, M.M., Laurindo, F.R., Fernandes, D.C., 2014. Methods of measuring protein disulfide isomerase activity: a critical overview. *Frontiers in Chemistry* 2, 73.
- Watson, W.H., Pohl, J., Montfort, W.R., Stuchlik, O., Reed, M.S., Powis, G., Jones, D.P., 2003. Redox potential of human thioredoxin 1 and identification of a second dithiol/disulfide motif. *The Journal of Biological Chemistry* 278, 33408-33415.
- Wetterau, J.R., Combs, K.A., Spinner, S.N., Joiner, B.J., 1990. Protein disulfide isomerase is a component of the microsomal triglyceride transfer protein complex. *The Journal of Biological Chemistry* 265, 9800-9807.
- Whittaker, J. 2007. Selective Isotopic Labeling of Recombinant Proteins Using Amino Acid Auxotroph Strains, In: Cregg, J. (Ed.) *Pichia Protocols*. Humana Press, 175-187.
- Wilkinson, B., Gilbert, H.F., 2004. Protein disulfide isomerase. *Biochimica et Biophysica Acta* 1699, 35-44.
- Williamson, M.P., 2013. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* 73, 1-16.
- Williamson, P.T., Roth, J.F., Haddingham, T., Watts, A., 2000. Expression and purification of recombinant neurotensin in *Escherichia coli*. *Protein Expression and Purification* 19, 271-275.
- Williamson, R.A., 2001. Refolding of TIMP-2 from *Escherichia coli* inclusion bodies. *Methods in Molecular Biology* (Clifton, N.J.) 151, 257-265.
- Williamson, R.A., Carr, M.D., Frenkiel, T.A., Feeney, J., Freedman, R.B., 1997. Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry* 36, 13882-13889.
- Williamson, R.A., Natalia, D., Gee, C.K., Murphy, G., Carr, M.D., Freedman, R.B., 1996. Chemically and conformationally authentic active domain of human tissue inhibitor of metalloproteinases-2 refolded from bacterial inclusion bodies. *European Journal of Biochemistry / FEBS* 241, 476-483.

- Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., Sykes, B.D., 1995a.  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest neighbour effects. *Journal of Biomolecular NMR* 5, 67-81.
- Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., Sykes, B.D., 1995b.  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift referencing in biomolecular NMR. *Journal of Biomolecular NMR* 6, 135-140.
- Wishart, D.S., Sykes, B.D., 1994. Chemical shifts as a tool for structure determination. *Methods in Enzymology* 239, 363-392.
- Wurm, F.M., 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology* 22, 1393-1398.
- Wüthrich, K., 1986. *NMR of Proteins and Nucleic Acids*. Wiley.
- Xu, S., Butkevich, A.N., Yamada, R., Zhou, Y., Debnath, B., Duncan, R., Zandi, E., Petasis, N.A., Neamati, N., 2012. Discovery of an orally active small-molecule irreversible inhibitor of protein disulfide isomerase for ovarian cancer treatment. *Proceedings of the National Academy of Sciences of the United States of America* 109, 16348-16353.
- Xu, S., Sankar, S., Neamati, N., 2014. Protein disulfide isomerase: a promising target for cancer therapy. *Drug Discovery Today* 19, 222-240.
- Yamazaki, T., Taguchi, T., Ojima, I. 2009. Unique Properties of Fluorine and their Relevance to Medicinal Chemistry and Chemical Biology, In: *Fluorine in Medicinal Chemistry and Chemical Biology*. John Wiley & Sons, Ltd, 1-46.
- Yang, S., Wang, X., Cui, L., Ding, X., Niu, L., Yang, F., Wang, C., Wang, C.C., Lou, J., 2014. Compact conformations of human protein disulfide isomerase. *PloS One* 9, e103472.
- Zapun, A., Darby, N.J., Tessier, D.C., Michalak, M., Bergeron, J.J., Thomas, D.Y., 1998. Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. *The Journal of Biological Chemistry* 273, 6009-6012.
- Zuiderweg, E.R.P., Fesik, S.W., 1989. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* 28, 2387-2391.

## Appendix Contents

### Chapter 2

<b>2.1 Vector Map of pET-31b(+)</b> .....	225
<b>2.2 Vector Map of pET-23a(+)</b> .....	226
<b>2.3 Expression Region of modified pET-23a(+)</b> expressing N-terminal His <sub>6</sub> -tag (x2) GB1 fusion proteins .....	227
<b>2.4 Translated Fusion Protein Sequences</b> .....	228
<b>2.5 Mass Spectrometry of each labelled peak from HPLC in Figure 2.13</b> .....	229
<b>2.6 Mass Spectrometry of each labelled peak from HPLC in Figure 2.14</b> .....	230

### Chapter 3

<b>3.1 Expression region of modified pET-23a(+)</b> expressing N-terminal His <sub>6</sub> -tag hPDI fragments .....	232
<b>3.2 Amino acid sequences and Expasy analysis of expressed hPDI fragments</b> .....	232

### Chapter 4

<b>4.1 W111F abb'x (37°C) Backbone Resonance Assignment Table</b> .....	234
<b>4.2 b'x (37°C) Backbone Resonance Assignment Tables</b> .....	239
<b>4.3 0.3 mM b'x (37°C) Methyl Sidechain Resonance Assignment Tables</b> .....	242

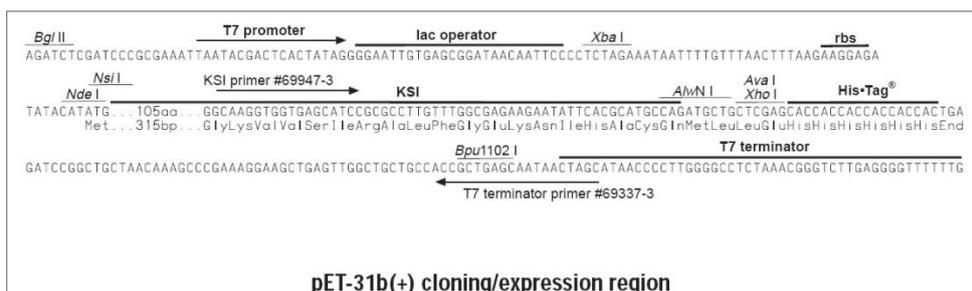
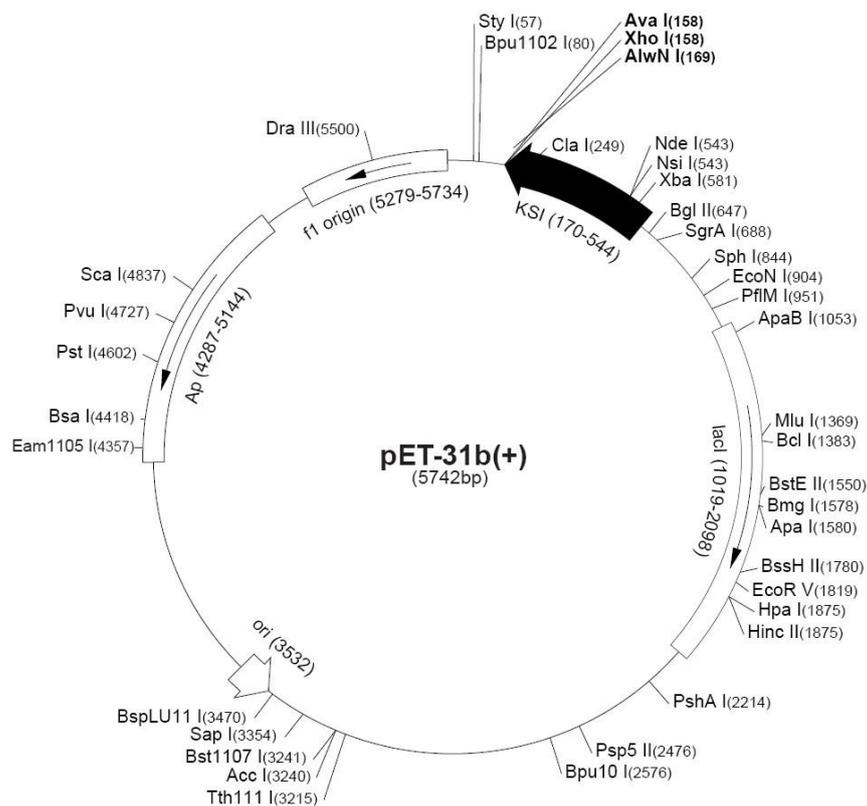
### Chapter 5

<b>5.1 Derivation of Dissociation Constant (K<sub>d</sub>)</b> .....	245
----------------------------------------------------------------------	-----

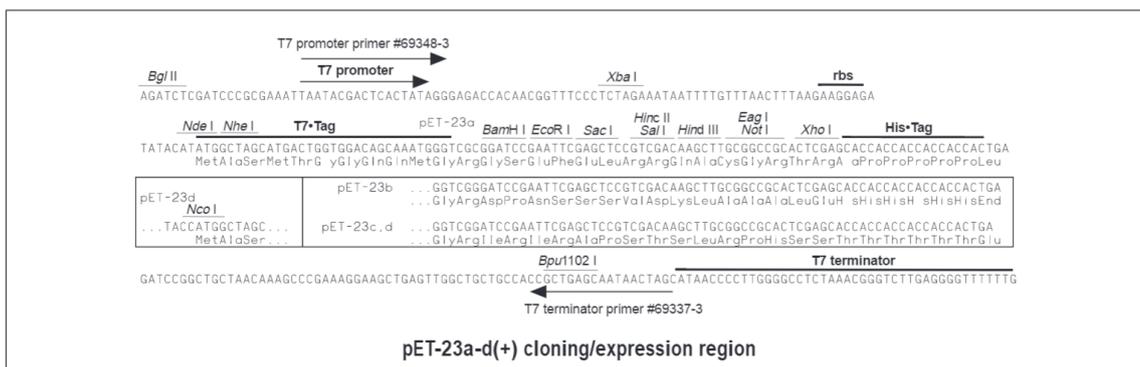
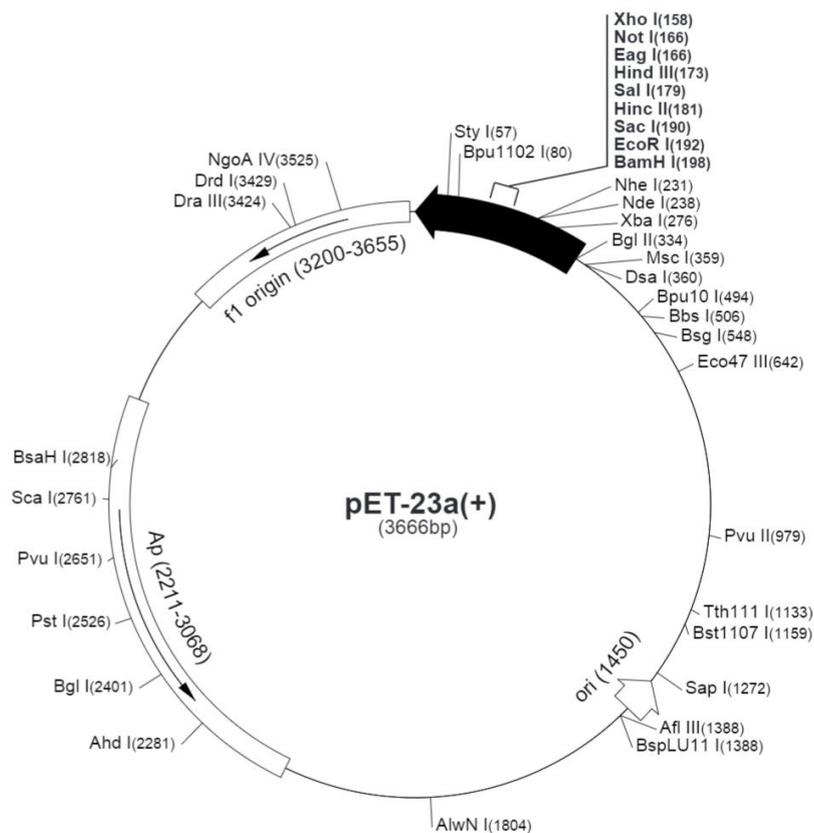
<b>5.2 0.3 mM <i>b'x</i> + 0.9 mM <math>\Delta</math>-som (37°C) Methyl Sidechain Resonance Assignment</b>	
<b>Tables</b> .....	246
<b>5.3 Representative fits of 0.25 mM <math>^{15}\text{N}</math> W111F <i>abb'x</i> (reduced) + <math>^{19}\text{F}</math> <math>\Delta</math>-som:</b>	
(a) + F1 $\Delta$ -som (protein observed) .....	249
(b) + F2 $\Delta$ -som (protein observed) .....	250
(c) + F3 $\Delta$ -som (protein observed) .....	251
(d) + F1,2,3 $\Delta$ -som (protein observed) .....	252
<b>5.4 Representative fits of 0.25 mM <math>^{15}\text{N}</math> W111F <i>abb'x</i> (reduced) + <math>\Delta</math>-som alanine substitutions:</b>	
(a) + Phe 6 $\rightarrow$ Ala $\Delta$ -som .....	253
(b) + Phe 7 $\rightarrow$ Ala $\Delta$ -som .....	254
(c) + Phe 11 $\rightarrow$ Ala $\Delta$ -som .....	255
(d) + Phe 6, 7 $\rightarrow$ Ala $\Delta$ -som .....	256
(e) + Phe 6, 11 $\rightarrow$ Ala $\Delta$ -som .....	257
(f) + Phe 7, 11 $\rightarrow$ Ala $\Delta$ -som .....	258
 <b>Chapter 6</b>	
<b>6.1 Table of Redox Potential Samples</b> .....	259
<b>6.2 <math>^{15}\text{N}</math> W111F <i>abb'x</i> Redox Potential Raw Data (<math>F_{\text{red}}</math>)</b> .....	260
 <b>Chapter 7</b>	
<b>7.1 Fits of 0.25 mM <math>^{15}\text{N}</math> 6-F-Trp W111F <i>abb'x</i> (reduced) + <math>\Delta</math>-som</b> .....	263

## APPENDIX

## 2.1 Vector Map of pET-31b(+)



## 2.2 Vector Map of pET-23a(+)



## 2.3 Expression Region of modified pET-23a(+) expressing N-terminal His<sub>6</sub>-tag (x2) GB1 fusion proteins

```

      rbs                               Nde1
actttaagaaggagatata (cat)ATGCATCACCATCACCACCATATGCACCATCATCATCATCACCAGTACAAG
      M H H H H H H M H H H H H H Q Y K

CTCGCTCTGAACGGTAAAACCTGAAAGGTGAAACCACCACCGAAGCTGTTGACGCT
L A L N G K T L K G E T T T E A V D A

GCTACCGCGGAAAAAGTTTTCAAACAGTACGCTAACGACAACGGTGTGACGGTGAA
A T A E K V F K Q Y A N D N G V D G E

TGGACCTACGACGACGCTACCAAAACCTTCACGGTAACCGAAggatccgaattcact
      BamH1  EcoR1
      W T Y D D A T K T F T V T E

      Sac1  Sal1  HindIII  Not1  Xho1  His-tag
agtgagctcgtcgacaagcttgcggccgcactcgagcaccaccaccaccaccactgagatccggctgctaacaagaagc

```

Black script is original pET-23a(+) vector

**Blue script** is sequence added to original pET-23a(+) vector

**Red script (cat)** is sequence removed from original pET-23a(+) vector

Uppercase script is translated

Lowercase script is not translated

**rbs** is in bold

restriction sites are highlighted

Translated polypeptide is shown in **ORANGE**

## 2.4 Translated Fusion Protein Sequences:

All fusion protein sequencing was carried out by Beckman Coulter Genomics, using their T7 promotor Primer (forward primer) and their T7 terminator primer (reverse primer).

Sequencing files were submitted to Expasy Translate (<http://web.expasy.org/translate>), and the translated region submitted to Expasy ProtParam (<http://web.expasy.org/protparam>) for calculation of protein parameters.

### 2.4.1 KSI- $\Delta$ -som

Met HTPEHITAVVQRFVAALNAGDLGIVALFADDATVEDP  
VGSEPRSGTAAIREFYANSLKLPLAVELTQEVRAVANEAA  
FAFTVSFEYQGRKTVVAPIDHFRFNGAGKVVSIRALFGEK  
NIHACQ Met **AGSKNFFWKTFTSS** Stop

Fusion MW: 15193.2 Da  
 Cleaved Peptide: 1607.7 Da

### 2.4.2 N-terminal His<sub>6</sub>-tag(x2)-GB1-enterokinase cleavage site- $\Delta$ -som

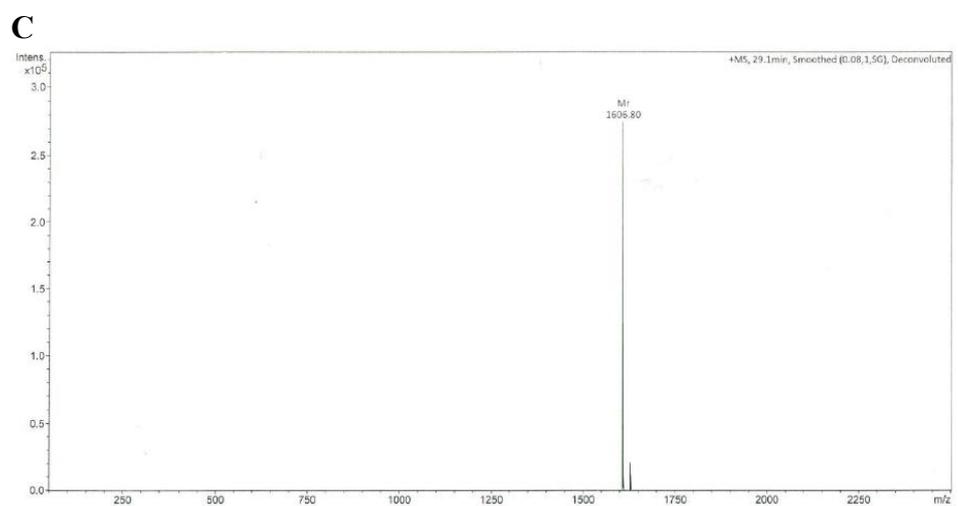
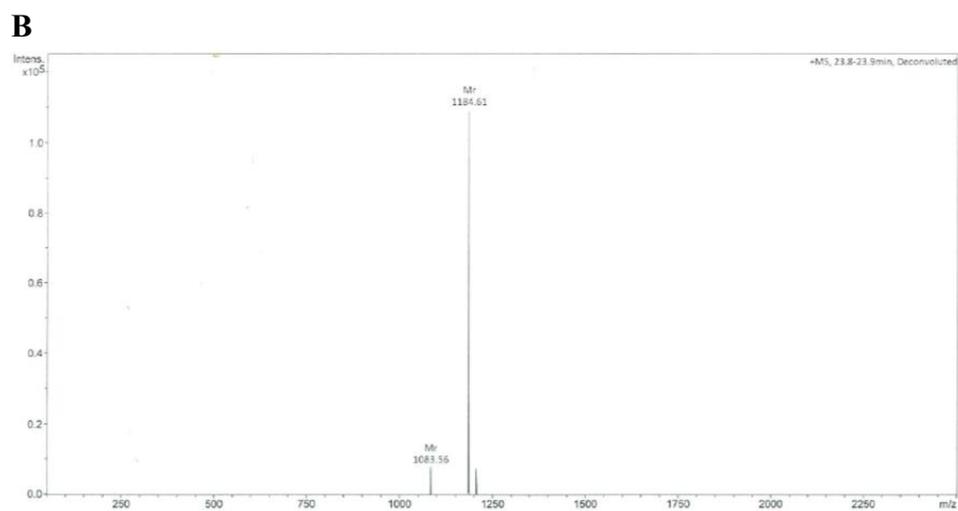
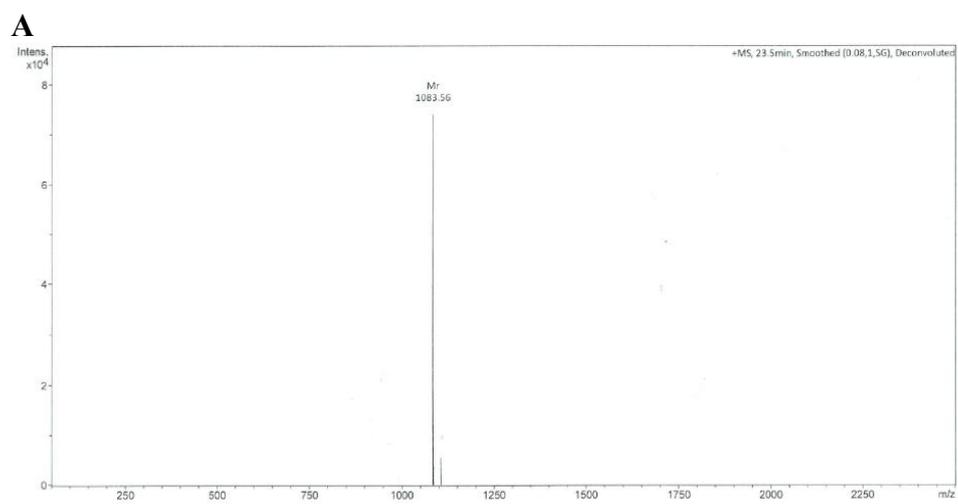
MHHHHHMHHHHHHHQYKLALNGKTLKGETTTEAVDAATAEKVFKQYANDNGV  
 DGEWTYDDATKTFTVTEGSDDDDKAGSKNFFWKTFTSS

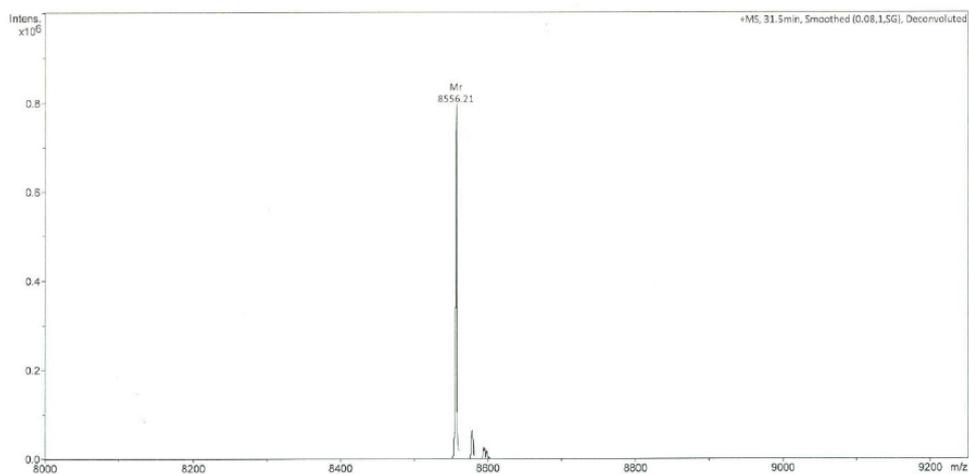
Fusion MW: 10280.0 Da  
 Cleaved Peptide: 1607.7 Da

### 2.4.3 N-terminal His<sub>6</sub>-tag(x2)-GB1-Factor Xa cleavage site- $\Delta$ -som

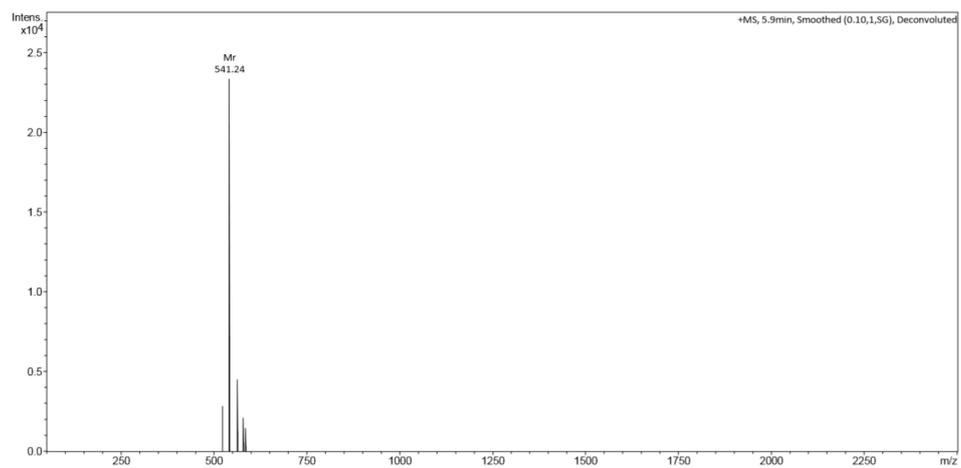
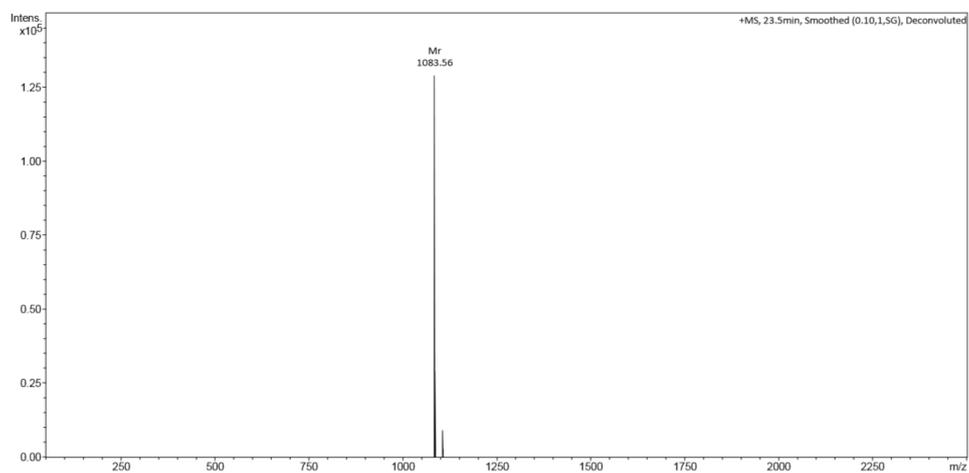
MHHHHHMHHHHHHHQYKLALNGKTLKGETTTEAVDAATAEKVFKQYANDNGV  
 DGEWTYDDATKTFTVTEGSIEGRAGSKNFFWKTFTSS

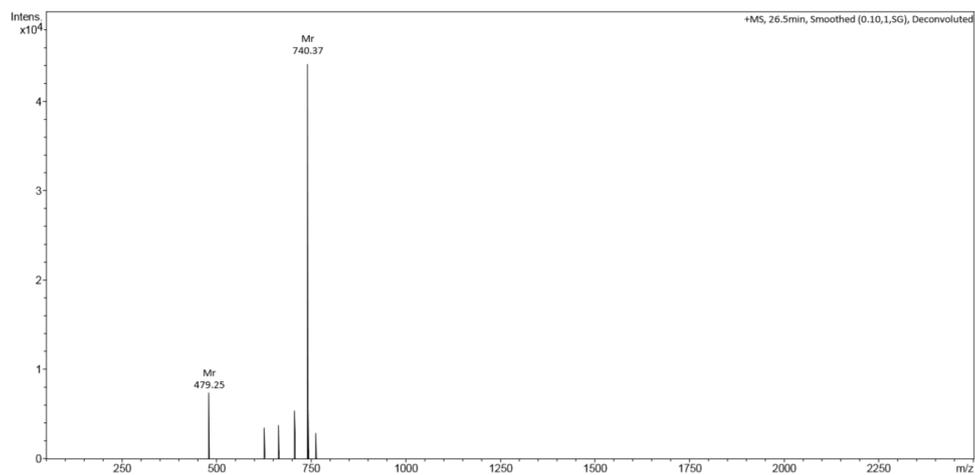
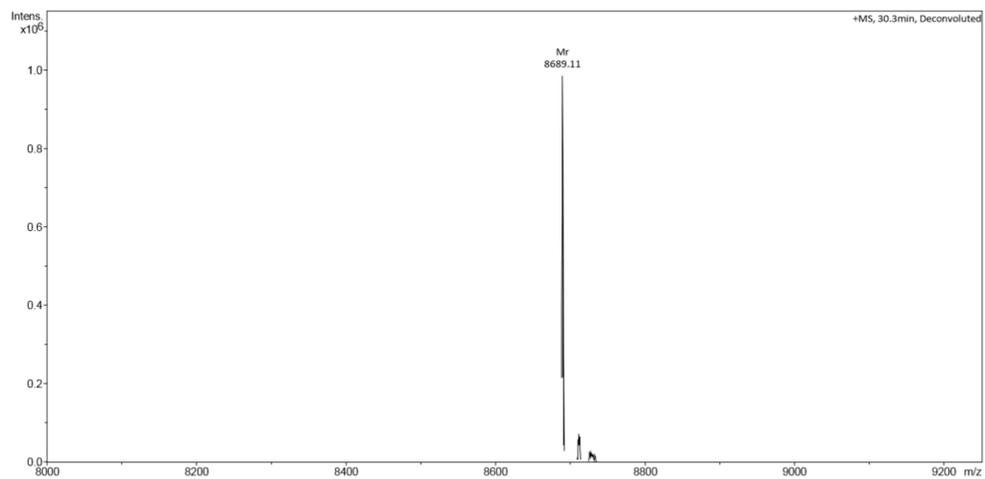
Fusion MW: 10147.0 Da  
 Cleaved Peptide: 1607.7 Da

**2.5 Mass Spectrometry of each labelled peak from HPLC in Figure 2.13**

**D**

## 2.6 Mass Spectrometry of each labelled peak from HPLC in Figure 2.14

**A****B**

**C****D**

### 3.1 Expression region of modified pET-23a(+) expressing N-terminal His<sub>6</sub>-tag hPDI fragments

```

          rbs                               Nde1                               BamH1  EcoR1
actttaagaaggagatatacatATGCATCACCATCACCACCATATG-Met-POI-Stop-ggatccgaattcact
          Sac1   Sall  HindIII  Not1    Xho1      His-tag
agtgagctccgtcgacaagcttggggcgggactcgagcaccaccaccaccaccactgagatccggctgctaacaaagc

```

Black script is original pET-23a(+) vector

Purple script is **Met-protein of interest-Stop** sequence added

Uppercase script is translated polypeptide

Lowercase script is not translated

rbs is in red

restriction sites are highlighted

Translated polypeptide is underlined

### 3.2 Amino acid sequences and Expsay analysis of expressed hPDI fragments

#### 3.2.1 His<sub>6</sub>-tag-b'x:

```

          10           20           30           40           50           60
MHHHHHMKH NQLPLVIEFT EQTAPKIFGG EIKTHILLEL PKSVSDYDGD LSNFKTAAES
          70           80           90           100          110          120
FKGKILFAFI DSDHTDNQRI LEFFGLKKEE CPAVRLITLE EEMTKYKPES EELTAERITE
          130          140
FCHRFLGKI KPHLMSQELP EDWDKQP

```

Number of amino acids: 147

Molecular weight: 17240.7 Da

Theoretical pI: 5.98

#### Atomic composition:

Carbon	C	781
Hydrogen	H	1203
<b>Nitrogen</b>	<b>N</b>	<b>205</b>
Oxygen	O	224
Sulfur	S	6

**Extinction coefficients:** Extinction coefficients are in units of M<sup>-1</sup> cm<sup>-1</sup>, at 280 nm measured in water.

**Ext. coefficient**      **8605**

### 3.2.2 His<sub>6</sub>-tag W111F abb'x:

```

      10      20      30      40      50      60
MHHHHHMDA PEEEDHVLVL RKSNFAEALA AHKYLLVEFY APWCGHCKAL APEYAKAAGK

      70      80      90      100     110     120
LKAEGSEIRL AKVDATEESD LAQQYGVRGY PTIKFFRNGD TASPKEYTAG READDIVNFL

      130     140     150     160     170     180
KKRTGPAATT LPDGAAAESL VESSEVAVIG FFKDVESDSA KQFLQAAEAI DDIPFGITSN

      190     200     210     220     230     240
SDVFSKYQLD KGVVLFKKF DEGRNFEGE VTKENLLDFI KHNQLPLVIE FTEQTAPKIF

      250     260     270     280     290     300
GGEIKTHILL FLPKSVSDYD GKLSNFKTAA ESFKGKILFI FIDSDHTDNQ RILEFFGLKK

      310     320     330     340     350
EECPAVRLIT LEEEMTKYKP ESEELTAERI TEFCHRFLEG KIKPHLMSQE LPEDWDKQP

```

*The W111F mutation is underlined (note the numbering is different owing to His<sub>6</sub>-tag)*

Number of amino acids: 359

Molecular weight: 40534.7 Da

Theoretical pI: 5.23

Atomic composition:

Carbon	C	1828
Hydrogen	H	2815
<b>Nitrogen</b>	<b>N</b>	<b>475</b>
Oxygen	O	552
Sulfur	S	8

**Extinction coefficients:** Extinction coefficients are in units of M<sup>-1</sup> cm<sup>-1</sup>, at 280 nm measured in water.

**Ext. coefficient**      **24660**

4.1 W111F *abb*'x (37°C) Backbone Resonance Assignment Table

Resonance not present in amino acid

- Resonances that could not be assigned

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
1 Asp	-	-	-
2 Ala	-	-	-
3 Pro			-
4 Glu	-	-	-
5 Glu	-	-	-
6 Glu	-	-	-
7 Asp	-	-	-
8 His	-	-	58.41
9 Val	7.82	123.43	61.54
10 Leu	8.94	128.18	55.72
11 Val	8.28	124.34	62.76
12 Leu	8.36	125.97	52.72
13 Arg	8.83	118.06	53.77
14 Lys	8.88	122.54	60.98
15 Ser	7.97	108.61	60.21
16 Asn	7.54	117.46	52.43
17 Phe	7.86	123.23	63.36
18 Ala	8.67	118.78	55.09
19 Glu	8.04	118.77	59.30
20 Ala	7.69	123.18	55.01
21 Leu	7.43	117.61	57.02
22 Ala	7.03	116.00	53.10
23 Ala	7.30	118.88	53.00
24 His	7.13	116.19	55.00
25 Lys	-	-	-
26 Tyr	-	-	55.06
27 Leu	8.40	125.67	53.46
28 Leu	8.96	131.01	53.64
29 Val	9.58	126.43	61.73
30 Glu	8.38	128.01	53.97
31 Phe	9.83	128.88	57.18
32 Tyr	8.91	121.35	54.80
33 Ala	7.08	119.94	47.62
34 Pro			64.57
35 Trp	5.89	109.79	53.81
36 Cys	6.63	126.31	60.04
37 Gly	-	-	47.61
38 His	9.46	127.72	59.21
39 Cys	9.64	127.79	64.40

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
40 Lys	8.66	123.39	60.09
41 Ala	7.45	119.96	54.28
42 Leu	7.42	116.49	55.99
43 Ala	7.36	120.98	57.18
44 Pro			65.61
45 Glu	6.98	118.13	58.18
46 Tyr	8.54	124.05	62.59
47 Ala	7.74	117.59	54.78
48 Lys	7.84	120.74	59.67
49 Ala	8.08	122.93	55.20
50 Ala	7.75	119.27	54.85
51 Gly	8.49	106.72	47.08
52 Lys	8.03	124.67	59.58
53 Leu	7.72	119.08	57.77
54 Lys	-	-	58.72
55 Ala	7.85	122.41	54.67
56 Glu	7.38	114.14	56.21
57 Gly	7.86	109.09	45.82
58 Ser	7.68	112.33	58.61
59 Glu	8.99	125.43	56.84
60 Ile	8.34	122.58	62.97
61 Arg	7.82	123.06	52.19
62 Leu	9.07	124.17	54.98
63 Ala	9.36	120.65	49.59
64 Lys	9.43	121.21	54.37
65 Val	8.73	119.46	60.25
66 Asp	8.85	127.14	52.42
67 Ala	8.78	130.46	53.22
68 Thr	8.54	109.23	63.24
69 Glu	6.81	122.30	57.19
70 Glu	8.31	120.19	52.72
71 Ser	7.40	115.41	61.31
72 Asp	-	-	-
73 Leu	-	-	-
74 Ala	-	-	-
75 Gln	-	-	-
76 Gln	-	-	-
77 Tyr	-	-	58.91
78 Gly	7.42	107.95	47.13

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
79 Val	7.74	119.53	64.09
80 Arg	-	-	55.23
81 Gly	7.09	107.39	43.70
82 Tyr	8.09	117.18	53.75
83 Pro			63.15
84 Thr	8.65	117.47	63.53
85 Ile	9.59	130.46	60.22
86 Lys	9.23	125.27	53.79
87 Phe	9.49	123.21	55.06
88 Phe	9.20	128.91	57.01
89 Arg	8.75	121.01	53.52
90 Asn	9.69	121.27	54.40
91 Gly	9.17	107.74	46.28
92 Asp	7.60	119.05	54.04
93 Thr	8.98	123.93	60.92
94 Ala	-	-	-
95 Ser	-	-	-
96 Pro			-
97 Lys	-	-	-
98 Glu	-	-	-
99 Tyr	-	-	-
100 Thr	-	-	-
101 Ala	-	-	50.79
102 Gly	7.91	104.37	45.34
103 Arg	8.42	118.78	55.67
104 Glu			55.23
105 Ala	9.73	125.17	56.39
106 Asp	8.70	113.86	57.61
107 Asp	7.27	119.49	57.39
108 Ile	7.40	121.58	65.03
109 Val	8.01	117.64	67.73
110 Asn	-	-	-
111 Phe	-	-	-
112 Leu	-	-	-
113 Lys	-	-	-
114 Lys	-	-	-
115 Arg	-	-	56.12
116 Thr	7.02	106.64	61.33
117 Gly	7.23	110.28	44.26
118 Pro			-
119 Ala	-	-	-
120 Ala	-	-	-
121 Thr	-	-	-
122 Thr	-	-	-
123 Leu	-	-	-

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
124 Pro			63.94
125 Asp	7.02	111.92	52.50
126 Gly	8.99	106.38	47.32
127 Ala	7.92	125.07	54.89
128 Ala	8.46	120.26	53.23
129 Ala	7.71	112.19	55.44
130 Glu	8.32	116.89	59.40
131 Ser	7.82	112.64	61.42
132 Leu	7.36	122.80	58.39
133 Val	7.87	119.26	66.81
134 Glu	8.10	115.01	58.03
135 Ser	7.46	112.68	59.88
136 Ser	7.00	115.30	57.24
137 Glu	8.57	124.53	59.93
138 Val	7.22	113.46	60.52
139 Ala	9.02	126.88	51.29
140 Val	7.88	120.03	60.79
141 Ile	8.98	124.34	60.11
142 Gly	8.51	115.63	44.43
143 Phe	8.36	127.82	54.98
144 Phe	8.10	116.88	55.64
145 Lys	9.33	124.63	58.99
146 Asp	8.76	117.09	51.87
147 Val	8.43	120.71	63.96
148 Glu	8.17	117.25	54.93
149 Ser	7.43	115.80	58.14
150 Asp	8.77	121.24	58.21
151 Ser	8.35	113.93	61.58
152 Ala	6.87	125.36	54.54
153 Lys	7.94	115.86	60.01
154 Gln	-	-	-
155 Phe	-	-	-
156 Leu	-	-	-
157 Gln	-	-	-
158 Ala	-	-	-
159 Ala	-	-	-
160 Glu	-	-	57.48
161 Ala	7.16	119.34	52.80
162 Ile	6.96	119.04	59.76
163 Asp	8.29	122.91	54.78
164 Asp	8.42	115.64	54.89
165 Ile	7.05	118.91	57.72
166 Pro			62.78
167 Phe	8.65	119.80	55.72
168 Gly	9.42	108.86	43.45

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
169 Ile	8.83	120.69	58.54
170 Thr	8.55	121.05	60.15
171 Ser	8.58	119.70	56.72
172 Asn	8.14	120.58	55.08
173 Ser	8.07	116.46	58.53
174 Asp	8.53	122.07	57.42
175 Val	7.92	122.10	65.87
176 Phe	7.88	120.24	61.81
177 Ser	7.99	111.97	61.65
178 Lys	7.56	123.83	-
179 Tyr	-	-	-
180 Gln	-	-	-
181 Leu	-	-	-
182 Asp	-	-	53.27
183 Lys	7.62	117.08	54.43
184 Asp	7.65	116.80	54.65
185 Gly	8.57	107.17	45.55
186 Val	-	-	-
187 Val	-	-	-
188 Leu	-	-	54.24
189 Phe	9.48	126.43	56.95
190 Lys	8.33	115.83	53.51
191 Lys	-	-	55.89
192 Phe	6.53	113.34	54.44
193 Asp	8.83	117.54	56.02
194 Glu	10.15	123.61	58.30
195 Gly	7.69	108.97	46.59
196 Arg	7.13	120.99	54.96
197 Asn	9.15	121.60	53.79
198 Asn	9.01	123.13	52.47
199 Phe	8.29	124.81	58.94
200 Glu	7.34	128.49	54.91
201 Gly	6.89	108.35	43.50
202 Glu	8.31	120.32	56.21
203 Val	9.00	126.39	64.45
204 Thr	7.13	117.73	58.06
205 Lys	9.24	123.99	60.74
206 Glu	9.03	116.55	60.88
207 Asn	7.90	117.92	55.34
208 Leu	8.44	121.38	57.61
209 Leu	8.39	120.00	58.98
210 Asp	7.56	118.79	57.83
211 Phe	7.79	120.64	60.41
212 Ile	8.66	121.14	65.51
213 Lys	8.10	117.91	59.39

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
214 His	7.77	113.93	58.47
215 Asn	7.19	113.69	54.46
216 Gln	7.56	114.94	57.41
217 Leu	7.33	119.47	51.90
218 Pro			61.75
219 Leu	8.24	120.53	57.88
220 Val	7.40	111.39	59.42
221 Ile	8.75	125.41	59.91
222 Glu	8.26	127.73	54.86
223 Phe	8.59	128.92	59.41
224 Thr	7.19	119.32	59.14
225 Glu	8.86	120.07	59.41
226 Gln	7.95	114.53	58.20
227 Thr	7.48	108.74	61.90
228 Ala	7.60	125.67	56.86
229 Pro			65.86
230 Lys	7.13	115.65	57.62
231 Ile	7.54	119.01	64.35
232 Phe	7.93	115.23	59.04
233 Gly	7.64	106.75	45.35
234 Gly	7.38	108.00	44.79
235 Glu	8.46	117.69	57.98
236 Ile	-	-	-
237 Lys	-	-	55.57
238 Thr	6.74	115.86	62.12
239 His	8.86	125.64	53.79
240 Ile	9.12	122.45	58.99
241 Leu	8.95	128.48	53.52
242 Leu	8.73	123.36	52.81
243 Phe	9.18	128.07	56.76
244 Leu	8.90	126.90	50.95
245 Pro			62.31
246 Lys	8.18	123.88	58.83
247 Ser	7.61	107.91	58.28
248 Val	7.55	123.19	62.06
249 Ser	-	-	-
250 Asp	-	-	-
251 Tyr	-	-	62.87
252 Asp	8.55	115.04	57.34
253 Gly	7.77	110.12	47.07
254 Lys	7.94	123.02	59.78
255 Leu	8.12	119.97	57.55
256 Ser	8.36	114.70	62.05
257 Asn	7.90	121.56	56.60
258 Phe	7.64	121.65	61.45

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
259 Lys	-	-	-
260 Thr	-	-	-
261 Ala	-	-	-
262 Ala	-	-	-
263 Glu	-	-	-
264 Ser	-	-	-
265 Phe	7.17	115.85	57.19
266 Lys	7.32	123.27	58.58
267 Gly	9.54	114.89	45.44
268 Lys	8.30	118.62	57.19
269 Ile	7.86	116.45	60.42
270 Leu	8.32	130.66	54.64
271 Phe	9.17	129.72	56.84
272 Ile	-	-	-
273 Phe	-	-	-
274 Ile	-	-	-
275 Asp	-	-	-
276 Ser	-	-	-
277 Asp	-	-	-
278 His	-	-	-
279 Thr	8.86	127.69	66.09
280 Asp	10.72	124.47	56.86
281 Asn	7.98	115.02	52.74
282 Gln	-	-	-
283 Arg	-	-	-
284 Ile	-	-	63.56
285 Leu	7.36	121.57	60.56
286 Glu	7.89	117.21	59.37
287 Phe	8.03	121.76	60.86
288 Phe	7.46	113.11	59.45
289 Gly	7.94	110.74	46.62
290 Leu	7.66	120.77	53.53
291 Lys	-	-	-
292 Lys	-	-	60.98
293 Glu	9.36	116.52	58.86
294 Glu	7.57	118.05	55.82
295 Cys	6.99	117.61	58.35
296 Pro	-	-	-
297 Ala	8.45	123.18	51.57
298 Val	8.56	117.88	59.38
299 Arg	8.85	122.98	53.06
300 Leu	8.76	125.58	54.33
301 Ile	9.64	122.39	58.46
302 Thr	8.66	114.19	59.13
303 Leu	8.52	125.19	54.49

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
304 Glu	7.47	123.16	56.52
305 Glu	8.41	122.09	59.30
306 Glu	-	-	-
307 Met	-	-	-
308 Thr	-	-	62.52
309 Lys	8.79	123.72	54.92
310 Tyr	9.62	120.00	57.66
311 Lys	-	-	-
312 Pro	-	-	-
313 Glu	-	-	-
314 Ser	-	-	-
315 Glu	-	-	-
316 Glu	-	-	57.40
317 Leu	8.54	125.32	53.14
318 Thr	6.74	108.18	59.46
319 Ala	9.20	125.22	56.06
320 Glu	9.27	118.19	60.32
321 Arg	7.59	119.44	57.94
322 Ile	8.52	121.75	65.52
323 Thr	8.57	116.99	67.93
324 Glu	8.12	121.75	-
325 Phe	-	-	-
326 Cys	-	-	-
327 His	-	-	-
328 Arg	-	-	-
329 Phe	-	-	60.67
330 Leu	8.14	123.44	57.48
331 Glu	7.72	116.37	56.29
332 Gly	7.73	108.43	46.34
333 Lys	7.99	116.56	56.18
334 Ile	7.98	120.03	59.25
335 Lys	8.91	128.04	54.29
336 Pro	-	-	-
337 His	-	-	-
338 Leu	-	-	-
339 Met	-	-	-
340 Ser	-	-	-
341 Gln	-	-	-
342 Glu	-	-	-
343 Leu	-	-	-
344 Pro	-	-	62.97
345 Glu	8.39	120.46	57.29
346 Asp	8.19	119.13	54.07
347 Trp	7.70	120.59	57.93
348 Asp	7.95	120.40	53.98

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>
349 Lys	7.51	120.51	55.83
350 Gln	8.06	122.93	53.53
351 Pro			-

4.2 *b'*x (37°C) Backbone Resonance Assignment Tables

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>	C <sub>β</sub>
213 Lys	-	-	-	-
214 His	-	-	-	-
215 Asn	-	-	-	-
216 Gln	-	-	-	-
217 Leu	-	-	-	-
218 Pro			-	-
219 Leu	-	-	57.09	42.92
220 Val	7.24	111.15	59.71	-
221 Ile	-	-	59.95	41.49
222 Glu	8.24	127.45	54.78	29.86
223 Phe	8.59	128.79	59.41	40.63
224 Thr	7.11	118.55	-	-
225 Glu	8.89	120.42	59.43	29.03
226 Gln	7.83	114.48	58.15	28.87
227 Thr	7.39	108.93	61.98	69.53
228 Ala	7.59	125.47	56.68	-
229 Pro			65.61	30.99
230 Lys	7.18	115.27	57.40	32.27
231 Ile	7.52	119.10	64.24	38.74
232 Phe	7.82	115.59	58.88	38.30
233 Gly	7.58	106.66	45.38	
234 Gly	7.41	107.81	44.79	
235 Glu	8.38	117.54	58.26	30.44
236 Ile	7.72	119.41	61.85	37.90
237 Lys	-	-	55.63	33.03
238 Thr	6.73	115.87	62.19	69.94
239 His	8.77	125.57	54.01	35.31
240 Ile	9.14	122.39	58.97	40.15
241 Leu	8.97	128.57	53.66	43.54
242 Leu	-	-	52.87	41.81
243 Phe	9.04	127.54	56.80	39.29
244 Leu	8.93	126.89	50.89	44.72
245 Pro			62.27	31.74
246 Lys	8.18	123.62	58.74	32.17
247 Ser	7.53	108.16	58.27	63.60
248 Val	7.41	122.31	62.02	32.64
249 Ser	8.30	120.50	59.75	63.25
250 Asp	8.88	122.43	52.90	35.62
251 Tyr	7.54	121.15	62.97	39.22
252 Asp	8.57	115.11	57.40	40.55
253 Gly	7.71	110.09	47.10	
254 Lys	7.94	122.98	59.99	32.97
255 Leu	8.06	119.84	57.56	40.60
256 Ser	8.28	114.70	62.50	-

<b>Residue</b>	<b>H<sub>N</sub></b>	<b>N<sup>H</sup></b>	<b>C<sub>α</sub></b>	<b>C<sub>β</sub></b>
<b>257 Asn</b>	7.92	121.60	56.57	38.66
<b>258 Phe</b>	7.64	121.71	61.63	40.59
<b>259 Lys</b>	8.43	117.10	59.79	32.52
<b>260 Thr</b>	8.18	116.34	66.70	68.62
<b>261 Ala</b>	7.53	124.08	54.87	18.66
<b>262 Ala</b>	7.20	118.99	53.98	19.52
<b>263 Glu</b>	6.93	114.04	58.82	29.71
<b>264 Ser</b>	7.39	111.93	60.71	63.51
<b>265 Phe</b>	7.11	115.85	57.28	41.16
<b>266 Lys</b>	7.22	123.03	58.50	31.85
<b>267 Gly</b>	9.47	114.75	45.60	
<b>268 Lys</b>	8.25	118.60	57.16	35.67
<b>269 Ile</b>	7.75	116.28	60.39	43.34
<b>270 Leu</b>	8.27	129.80	54.59	43.15
<b>271 Phe</b>	9.11	129.71	56.68	39.37
<b>272 Ile</b>	9.05	123.95	58.95	42.12
<b>273 Phe</b>	8.81	122.55	54.65	42.64
<b>274 Ile</b>	8.70	121.43	59.94	42.39
<b>275 Asp</b>	8.36	123.86	52.69	40.33
<b>276 Ser</b>	7.40	123.69	60.53	64.81
<b>277 Asp</b>	8.29	119.32	55.55	41.65
<b>278 His</b>	7.16	121.28	58.34	31.88
<b>279 Thr</b>	-	-	65.99	68.77
<b>280 Asp</b>	10.75	124.49	56.80	40.01
<b>281 Asn</b>	7.89	114.93	52.75	38.55
<b>282 Gln</b>	7.56	120.72	59.50	28.19
<b>283 Arg</b>	8.41	118.07	59.79	32.59
<b>284 Ile</b>	8.49	120.66	55.82	37.12
<b>285 Leu</b>	7.39	121.87	60.43	40.92
<b>286 Glu</b>	7.91	117.31	59.36	29.29
<b>287 Phe</b>	7.79	120.43	63.65	37.31
<b>288 Phe</b>	-	-	59.56	39.87
<b>289 Gly</b>	7.95	110.73	46.71	
<b>290 Leu</b>	7.68	120.93	53.48	45.80
<b>291 Lys</b>	8.08	119.04	54.12	34.19
<b>292 Lys</b>	8.99	123.95	60.95	32.11
<b>293 Glu</b>	9.38	116.54	58.83	28.33
<b>294 Glu</b>	7.58	118.09	55.87	30.67
<b>295 Cys</b>	7.00	117.60	58.46	27.28
<b>296 Pro</b>			61.51	-
<b>297 Ala</b>	8.34	122.97	51.58	23.50
<b>298 Val</b>	8.59	118.11	59.47	35.23
<b>299 Arg</b>	8.66	123.10	53.03	35.54
<b>300 Leu</b>	8.66	125.11	54.34	47.19

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>	C <sub>β</sub>
301 Ile	9.53	121.72	58.42	42.69
302 Thr	8.72	114.19	59.33	71.03
303 Leu	8.55	125.12	54.48	42.47
304 Glu	7.48	123.28	56.55	29.85
305 Glu	8.43	122.15	59.27	29.13
306 Glu	7.87	116.20	54.90	31.42
307 Met	8.83	123.87	56.10	33.90
308 Thr	8.35	121.55	62.45	70.37
309 Lys	8.55	123.31	54.87	35.92
310 Tyr	9.50	119.26	57.63	41.59
311 Lys	9.14	122.39	52.74	-
312 Pro			62.05	31.41
313 Glu	8.89	119.35	57.65	30.21
314 Ser	7.45	111.72	56.48	65.22
315 Glu	-	-	-	-
316 Glu	-	-	-	-
317 Leu	-	-	53.10	41.25
318 Thr	6.75	108.17	59.67	70.57
319 Ala	9.22	125.27	56.14	18.03
320 Glu	9.30	118.23	60.31	29.21
321 Arg	7.61	119.58	58.04	29.94
322 Ile	8.44	121.74	65.50	38.22
323 Thr	8.50	116.83	-	67.87
324 Glu	7.99	121.52	59.94	30.35
325 Phe	7.84	119.66	59.33	38.89
326 Cys	7.62	115.90	64.41	27.46
327 His	8.54	118.17	60.43	30.45
328 Arg	8.66	118.84	59.93	30.21
329 Phe	8.24	120.62	60.47	37.83
330 Leu	8.13	123.59	57.50	41.35
331 Glu	7.68	116.28	56.33	30.64
332 Gly	7.66	108.45	46.36	
333 Lys	7.92	116.52	56.22	33.62
334 Ile	7.82	118.61	59.22	38.73
335 Lys	8.64	127.90	54.43	32.42
336 Pro			-	-
337 His	-	-	-	-
338 Leu	-	-	-	-
339 Met	-	-	-	-
340 Ser	-	-	-	-
341 Gln	-	-	-	-
342 Glu	-	-	-	-
343Leu	-	-	-	-
344 Pro			62.93	32.09

Residue	H <sub>N</sub>	N <sup>H</sup>	C <sub>α</sub>	C <sub>β</sub>
<b>345 Glu</b>	8.40	120.39	57.31	29.94
<b>346 Asp</b>	8.20	119.12	54.14	40.60
<b>347 Trp</b>	7.71	120.57	57.95	29.14
<b>348 Asp</b>	7.97	120.30	54.10	40.21
<b>349 Lys</b>	7.52	120.38	55.87	32.88
<b>350 Gln</b>	8.06	122.91	53.55	28.80
<b>351 Pro</b>			-	-

#### 4.3 0.3 mM *b'*x (37°C) Methyl Sidechain Resonance Assignment Tables

Residue	221 Ile	227 Thr	228 Ala	231 Ile	236Ile	238 Thr	240 Ile	241 Leu
<b>C<sub>α</sub></b>	59.95	62.03	56.83	64.24	61.85	61.94	58.97	53.66
<b>H<sub>α</sub></b>	4.30	4.93	3.69	3.43	4.07	4.15	4.87	4.96
<b>C<sub>β</sub></b>	41.49	69.84	16.20	38.71	37.90	69.54	40.15	43.55
<b>H<sub>βa</sub></b>	1.67	4.01	1.06	1.32	1.79	4.10	0.90	0.52
<b>H<sub>βb</sub></b>								1.10
<b>C<sub>γ1</sub></b>	28.01	23.56		28.26	27.99		30.02	28.90
<b>H<sub>γ1a</sub></b>	1.04	0.90		0.64	1.02		0.79	1.15
<b>H<sub>γ1b</sub></b>	1.51			1.32	1.54		1.34	
<b>C<sub>γ2</sub></b>	17.77			16.61	18.59	22.43	14.86	
<b>H<sub>γ2</sub></b>	0.78			-0.03	0.87	1.13	-0.25	
<b>C<sub>δ1</sub></b>	14.51			14.20	13.30		14.60	26.43
<b>H<sub>δ1</sub></b>	0.82			0.51	0.76		0.58	0.29
<b>C<sub>δ2</sub></b>								26.03
<b>H<sub>δ2</sub></b>								0.52

Residue	242 Leu	244 Leu	248 Val	255 Leu	260 Thr	261 Ala	262 Ala	269 Ile
<b>C<sub>α</sub></b>	52.87	50.89	61.95	57.55	66.64	54.83	53.92	60.39
<b>H<sub>α</sub></b>	4.21	5.01	4.19	3.72	3.84	4.02	1.84	4.27
<b>C<sub>β</sub></b>	41.81	44.72	32.66	40.60	68.60	18.60	19.54	43.33
<b>H<sub>βa</sub></b>	1.31	0.98	2.25	0.77	4.23	1.61	0.61	1.59
<b>H<sub>βb</sub></b>	1.44	1.95		1.47				
<b>C<sub>γ1</sub></b>	26.82	25.04	21.31	26.38	21.42			27.61
<b>H<sub>γ1a</sub></b>	1.50	0.70	1.06	1.10	1.12			1.52
<b>H<sub>γ1b</sub></b>								1.52
<b>C<sub>γ2</sub></b>			21.77					17.40
<b>H<sub>γ2</sub></b>			1.19					0.39
<b>C<sub>δ1</sub></b>	23.99	26.41		21.77				13.31
<b>H<sub>δ1</sub></b>	0.62	0.21		0.05				0.35
<b>C<sub>δ2</sub></b>	25.21	26.46		24.43				
<b>H<sub>δ2</sub></b>	0.67	1.48		0.34				

Residue	270 Leu	272 Ile	274 Ile	279 Thr	285 Leu	290 Leu	297 Ala	298 Val
<b>C<sub>α</sub></b>	54.57	58.97	59.94	65.98	60.43	53.47	51.63	59.38
<b>H<sub>α</sub></b>	4.84	5.43	4.29	4.02	3.75	4.46	4.79	5.69
<b>C<sub>β</sub></b>	43.16	42.51	42.39	68.75	40.73	45.80	23.50	35.46
<b>H<sub>βa</sub></b>	1.58	1.70	1.16	4.15	1.77	0.82	1.37	1.73
<b>H<sub>βb</sub></b>	1.94				1.77	0.95		
<b>C<sub>γ1</sub></b>	27.07	27.60	28.82		29.08	26.84		21.95
<b>H<sub>γ1a</sub></b>	1.50	0.89	0.47		1.48	0.95		0.72
<b>H<sub>γ1b</sub></b>		1.41	1.07					
<b>C<sub>γ2</sub></b>		19.02	18.20	21.54				21.95
<b>H<sub>γ2</sub></b>		0.74	0.47	1.28				0.82
<b>C<sub>δ1</sub></b>	24.39	15.70	14.99		25.53	24.21		
<b>H<sub>δ1</sub></b>	0.85	0.63	-0.12		0.62	-0.14		
<b>C<sub>δ2</sub></b>	25.28				25.92	25.28		
<b>H<sub>δ2</sub></b>	0.98				0.85	0.32		

Residue	300 Leu	302 Thr	303 Leu	308 Thr	317 Leu	318 Thr	319 Ala	322 Ile
<b>C<sub>α</sub></b>	54.40	59.39	54.48	62.43	53.10	59.15	56.03	65.51
<b>H<sub>α</sub></b>	5.43	4.36	4.74	4.58	4.55	5.39	3.98	3.91
<b>C<sub>β</sub></b>	47.25	70.50	42.46	70.37	41.25	70.92	18.05	38.22
<b>H<sub>βa</sub></b>	1.18	4.67	1.71	4.09	1.50	4.31	1.53	1.93
<b>H<sub>βb</sub></b>	1.36		1.88		1.67			
<b>C<sub>γ1</sub></b>	27.19	22.56	27.60	22.20	26.06	21.98		28.81
<b>H<sub>γ1a</sub></b>	1.18	1.25	1.63	1.04	1.67	1.41		1.16
<b>H<sub>γ1b</sub></b>								-
<b>C<sub>γ2</sub></b>								18.60
<b>H<sub>γ2</sub></b>								1.10
<b>C<sub>δ1</sub></b>	24.29		25.95		24.00			12.90
<b>H<sub>δ1</sub></b>	-0.27		0.92		0.65			0.90
<b>C<sub>δ2</sub></b>	26.15		25.15		26.13			
<b>H<sub>δ2</sub></b>	-0.47		0.95		0.73			

Residue	330 Leu	334 Ile
<b>C<sub>α</sub></b>	57.69	59.22
<b>H<sub>α</sub></b>	3.42	4.54
<b>C<sub>β</sub></b>	41.35	38.73
<b>H<sub>βa</sub></b>	1.43	2.07
<b>H<sub>βb</sub></b>	2.09	
<b>C<sub>γ1</sub></b>	26.64	26.40
<b>H<sub>γ1a</sub></b>	1.91	1.18
<b>H<sub>γ1b</sub></b>		1.41
<b>C<sub>γ2</sub></b>		18.60
<b>H<sub>γ2</sub></b>		1.37
<b>C<sub>δ1</sub></b>	26.36	12.12
<b>H<sub>δ1</sub></b>	0.90	1.02
<b>C<sub>δ2</sub></b>	22.35	
<b>H<sub>δ2</sub></b>	1.03	

### 5.1 Derivation of Dissociation Constant ( $K_d$ )

During a titration we know the total concentrations of ligand [L] and protein [P]:-

$$[L]_{\text{total}} = [L] + [PL] \text{ Complex}$$

$$[P]_{\text{total}} = [P] + [PL] \text{ Complex}$$

We also know the chemical shifts observed  $(\Delta)_{\text{obs}}$  in fast exchange state are a weighted average of the shifts of free  $(\delta)_{\text{free}}$  and bound  $(\delta)_{\text{bound}}$  states:-

$$(\Delta)_{\text{obs}} = (\Delta)_{\text{free}} f_{\text{free}} + (\Delta)_{\text{bound}} f_{\text{bound}}$$

where  $f_{\text{free}}$  and  $f_{\text{bound}}$  are the fractions of free and bound respectively. Because:-

$$f_{\text{free}} + f_{\text{bound}} = 1$$

The fraction of ligand bound can be expressed as:-

$$f_{\text{bound}} = ((\Delta)_{\text{obs}} - (\Delta)_{\text{free}}) / ((\Delta)_{\text{bound}} - (\Delta)_{\text{free}})$$

And  $K_d$  can be derived as follows:-

$$\Delta = \Delta_o \frac{(K_d + [L] + [P]) - \sqrt{((K_d + [L] + [P])^2 - 4 [P][L])}}{2[P]}$$

**5.2 0.3 mM *b'*x + 0.9 mM  $\Delta$ -som (37°C) Methyl Sidechain Resonance Assignment Tables**

Residue	221 Ile	227 Thr	228 Ala	231 Ile	236Ile	238 Thr	240 Ile	241 Leu
<b>C<sub><math>\alpha</math></sub></b>	59.97	62.09	56.79	64.31	61.97	61.91	58.96	53.66
<b>H<sub><math>\alpha</math></sub></b>	4.30	4.92	3.77	3.42	4.07	4.15	4.87	4.98
<b>C<sub><math>\beta</math></sub></b>	41.49	69.84	16.20	38.71	37.50	69.55	40.32	43.55
<b>H<sub><math>\beta</math>a</sub></b>	1.65	4.00	1.12	1.30	1.79	4.10	0.90	0.49
<b>H<sub><math>\beta</math>b</sub></b>								1.04
<b>C<sub><math>\gamma</math>1</sub></b>	28.00	23.56		-	28.00		-	28.60
<b>H<sub><math>\gamma</math>1a</sub></b>	1.03	0.89		-	1.01		-	1.17
<b>H<sub><math>\gamma</math>1b</sub></b>	1.54			-	1.54		-	
<b>C<sub><math>\gamma</math>2</sub></b>	17.80			17.01	18.60	22.42	14.86	
<b>H<sub><math>\gamma</math>2</sub></b>	0.81			0.00	0.87	1.15	-0.25	
<b>C<sub><math>\delta</math>1</sub></b>	14.91			14.20	13.30		14.65	25.77
<b>H<sub><math>\delta</math>1</sub></b>	0.84			0.57	0.76		0.57	0.33
<b>C<sub><math>\delta</math>2</sub></b>								26.18
<b>H<sub><math>\delta</math>2</sub></b>								0.54

Residue	242 Leu	244 Leu	248 Val	255 Leu	260 Thr	261 Ala	262 Ala	269 Ile
<b>C<sub><math>\alpha</math></sub></b>	-	50.86	61.91	57.42	66.40	54.77	53.95	60.78
<b>H<sub><math>\alpha</math></sub></b>	-	5.00	4.20	3.73	3.83	4.02	1.86	4.29
<b>C<sub><math>\beta</math></sub></b>	-	44.72	32.64	40.65	68.55	18.60	19.54	43.72
<b>H<sub><math>\beta</math>a</sub></b>	-	0.98	2.25	0.81	4.23	1.62	0.61	1.59
<b>H<sub><math>\beta</math>b</sub></b>	-	1.95		1.48				
<b>C<sub><math>\gamma</math>1</sub></b>	-	25.29	21.30	27.17	21.43			27.26
<b>H<sub><math>\gamma</math>1a</sub></b>	-	0.67	1.06	1.12	1.12			0.80
<b>H<sub><math>\gamma</math>1b</sub></b>								1.52
<b>C<sub><math>\gamma</math>2</sub></b>			21.71					16.99
<b>H<sub><math>\gamma</math>2</sub></b>			1.19					0.40
<b>C<sub><math>\delta</math>1</sub></b>	-	26.40		21.85				13.36
<b>H<sub><math>\delta</math>1</sub></b>	-	0.17		0.06				0.35
<b>C<sub><math>\delta</math>2</sub></b>	-	26.40		24.50				
<b>H<sub><math>\delta</math>2</sub></b>	-	1.48		0.33				

Residue	270 Leu	272 Ile	274 Ile	279 Thr	285 Leu	290 Leu	297 Ala	298 Val
<b>C<sub>α</sub></b>	54.19	58.95	59.54	65.91	60.46	53.88	51.58	59.66
<b>H<sub>α</sub></b>	4.86	5.47	4.30	4.03	3.68	4.43	4.79	5.70
<b>C<sub>β</sub></b>	43.16	42.46	42.40	68.77	40.76	45.84	23.23	35.49
<b>H<sub>βa</sub></b>	1.62	1.73	1.18	4.15	1.76	0.81	1.38	1.72
<b>H<sub>βb</sub></b>	1.82				-	1.00		
<b>C<sub>γ1</sub></b>	27.47	27.59	28.57		28.82	26.70		-
<b>H<sub>γ1a</sub></b>	1.50	-	0.54		1.45	1.00		-
<b>H<sub>γ1b</sub></b>		0.92	1.10					
<b>C<sub>γ2</sub></b>		19.10	17.79	21.55				21.95
<b>H<sub>γ2</sub></b>		0.79	0.48	1.28				0.80
<b>C<sub>δ1</sub></b>	24.80	15.66	15.00		25.50	24.20		
<b>H<sub>δ1</sub></b>	0.84	0.66	-0.01		0.58	-0.15		
<b>C<sub>δ2</sub></b>	24.80				25.10	25.28		
<b>H<sub>δ2</sub></b>	0.97				0.84	0.36		

Residue	300 Leu	302 Thr	303 Leu	308 Thr	317 Leu	318 Thr	319 Ala	322 Ile
<b>C<sub>α</sub></b>	54.40	59.41	-	62.42	52.68	59.09	55.98	65.11
<b>H<sub>α</sub></b>	5.40	4.35	-	4.47	4.46	5.38	3.99	3.91
<b>C<sub>β</sub></b>	47.65	70.53	-	70.37	41.63	70.81	18.06	38.30
<b>H<sub>βa</sub></b>	1.17	4.67	-	4.08	1.34	4.27	1.53	1.92
<b>H<sub>βb</sub></b>	1.36		-		1.49			
<b>C<sub>γ1</sub></b>	27.15	22.56	-	21.79	26.94	21.98		28.81
<b>H<sub>γ1a</sub></b>	1.17	1.25	-	1.06	1.50	1.41		1.14
<b>H<sub>γ1b</sub></b>								1.95
<b>C<sub>γ2</sub></b>								18.60
<b>H<sub>γ2</sub></b>								1.10
<b>C<sub>δ1</sub></b>	24.34		-		23.72			12.53
<b>H<sub>δ1</sub></b>	-0.27		-		0.65			0.89
<b>C<sub>δ2</sub></b>	26.04		-		26.14			
<b>H<sub>δ2</sub></b>	-0.47		-		0.72			

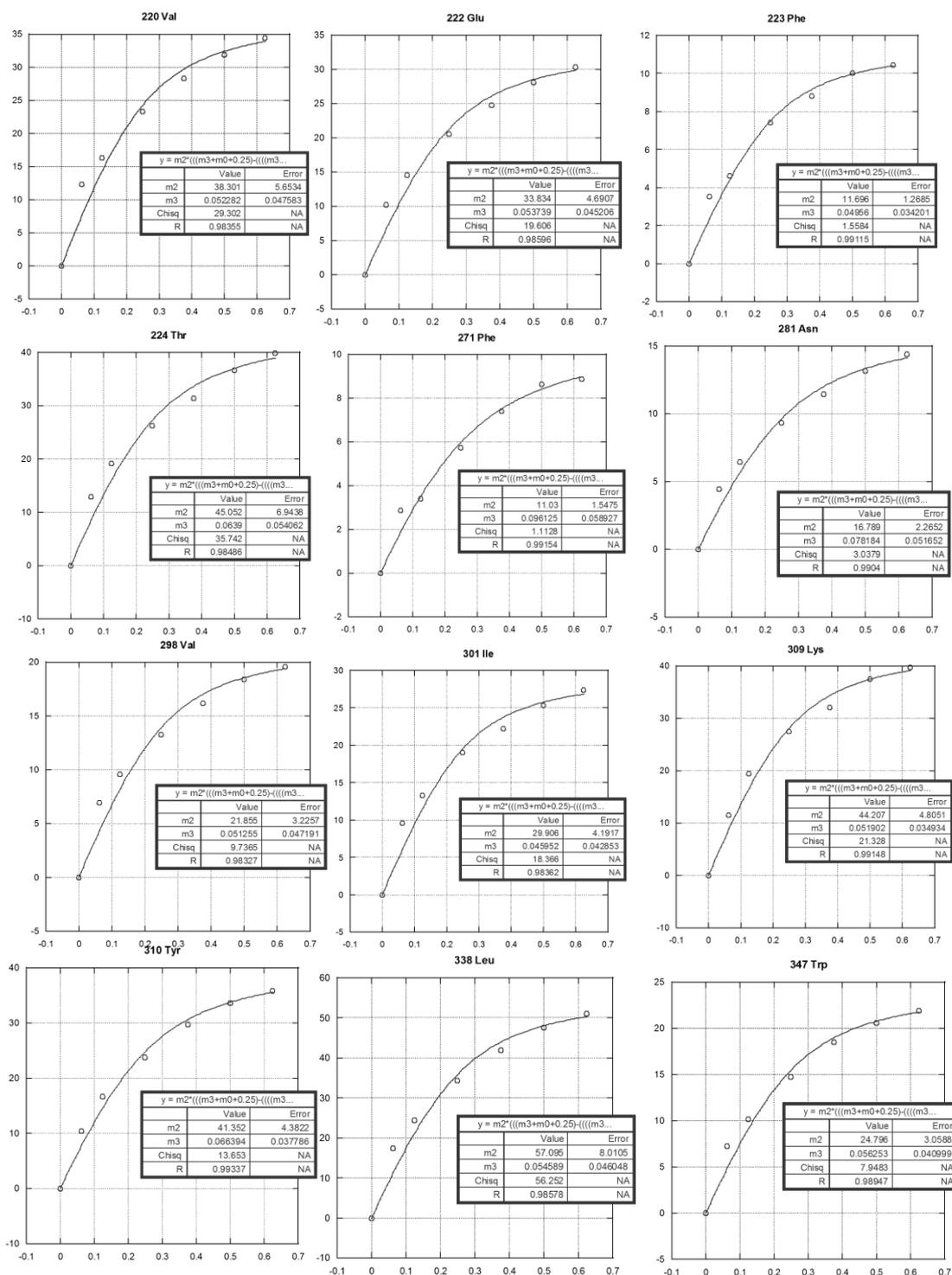
<b>Residue</b>	<b>330 Leu</b>	<b>334 Ile</b>
<b>C<sub>α</sub></b>	57.39	59.22
<b>H<sub>α</sub></b>	3.46	4.54
<b>C<sub>β</sub></b>	41.31	38.93
<b>H<sub>βa</sub></b>	1.45	2.03
<b>H<sub>βb</sub></b>	2.10	
<b>C<sub>γ1</sub></b>	26.64	26.34
<b>H<sub>γ1a</sub></b>	1.94	1.16
<b>H<sub>γ1b</sub></b>		1.41
<b>C<sub>γ2</sub></b>		18.36
<b>H<sub>γ2</sub></b>		1.30
<b>C<sub>δ1</sub></b>	22.39	12.48
<b>H<sub>δ1</sub></b>	0.91	1.00
<b>C<sub>δ2</sub></b>	26.24	
<b>H<sub>δ2</sub></b>	1.04	

### 5.3 Representative fits of 0.25 mM <sup>15</sup>N W111F *abb*'x (reduced) + <sup>19</sup>F Δ-som:

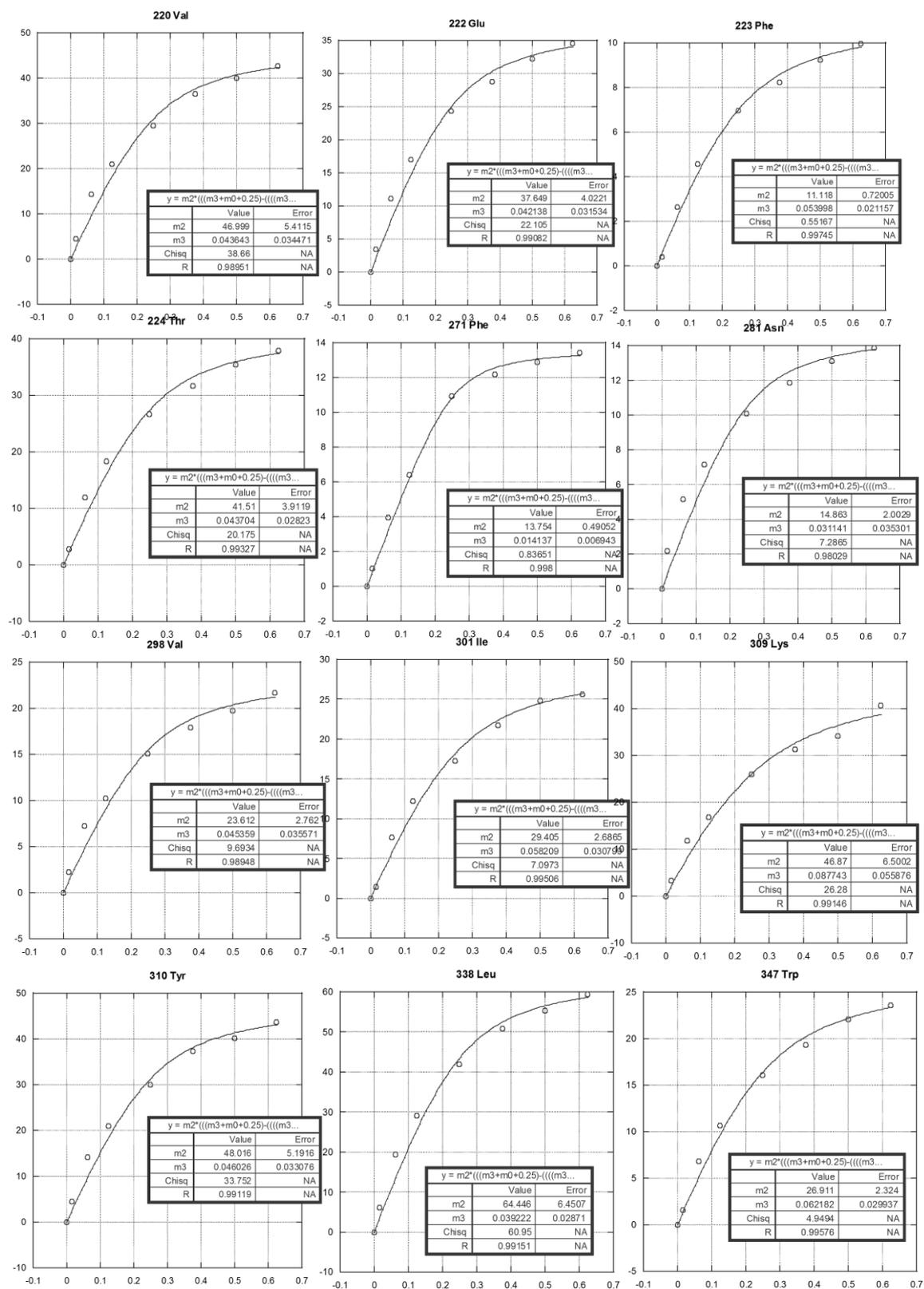
#### (a) + F1 Δ-som (protein observed)

Peptide concentration is shown along the x-axis (mM), delta chemical shift is shown along the y-axis (Hz).  
Kd is given as m3 value (mM).

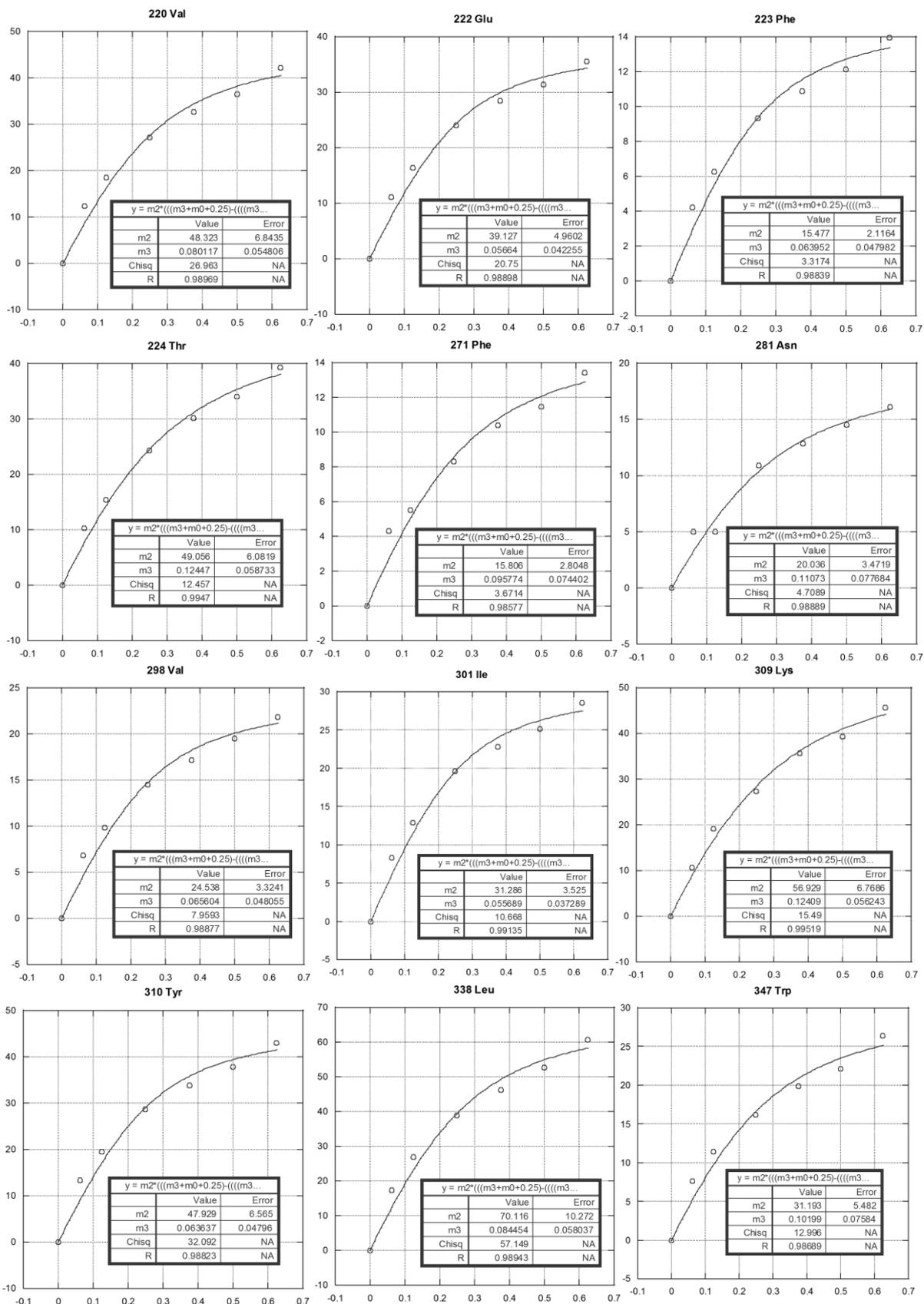
Kaleidagraph fit equation:  $m_2 * (((m_3 + m_0 + 0.25) - (((m_3 + m_0 + 0.25)^2 - (4 * 0.25 * m_0))^0.5)) / (2 * 0.25)); m_2 = 100; m_3 = 0.01;$



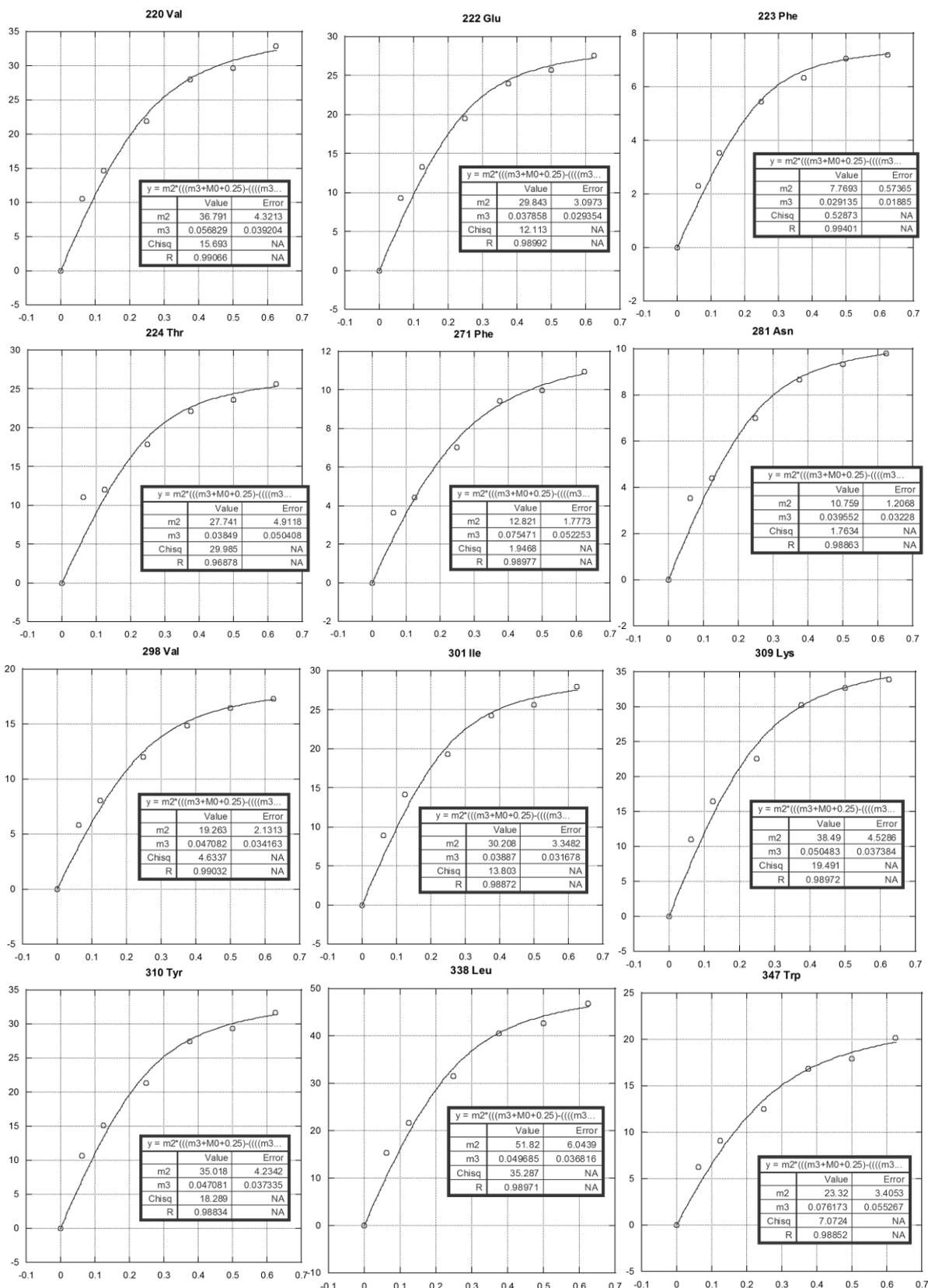
(b) + F2 Δ-som (protein observed)



(c) + F3 Δ-som (protein observed)

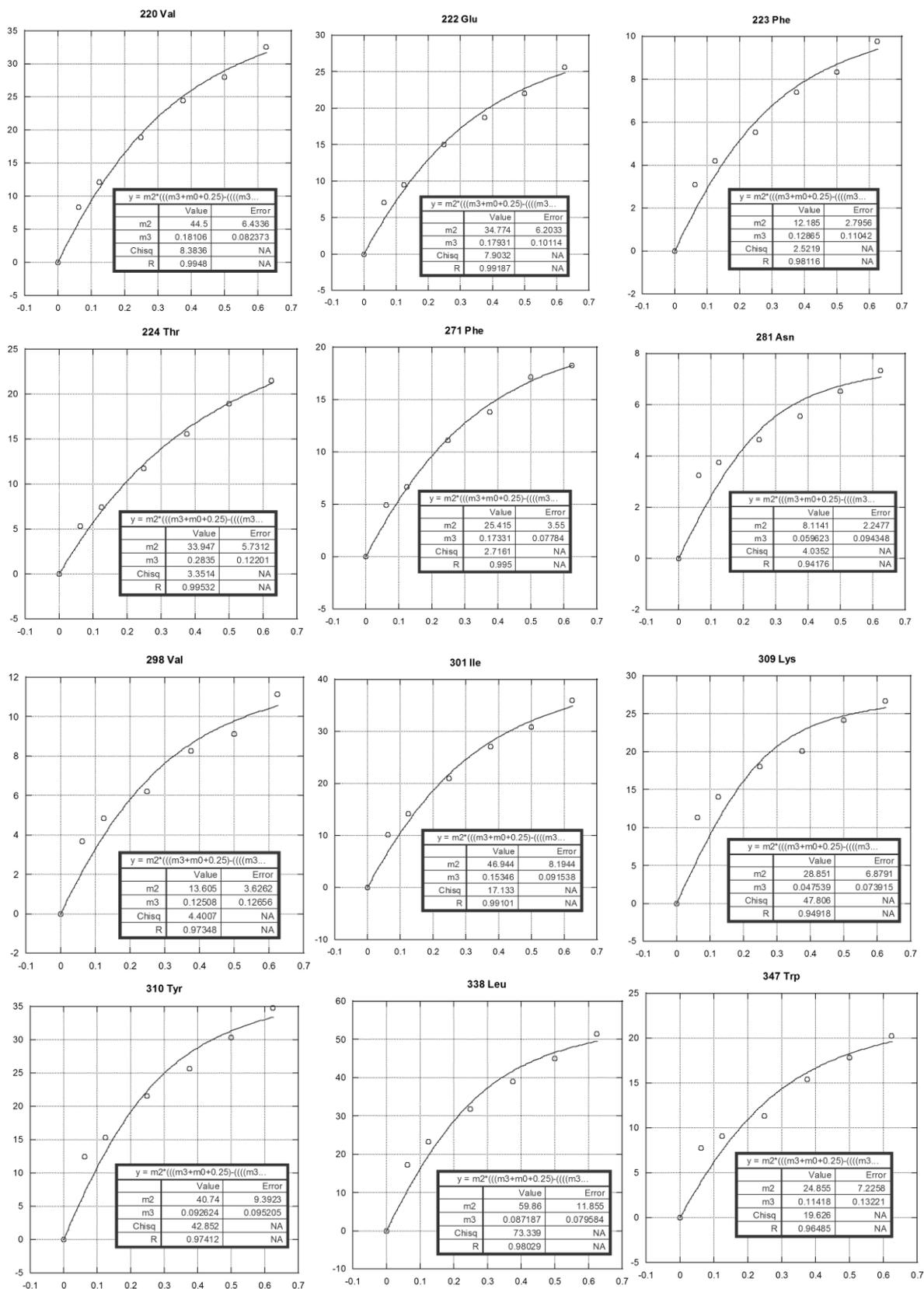


(d) + F1,2,3 Δ-som (protein observed)

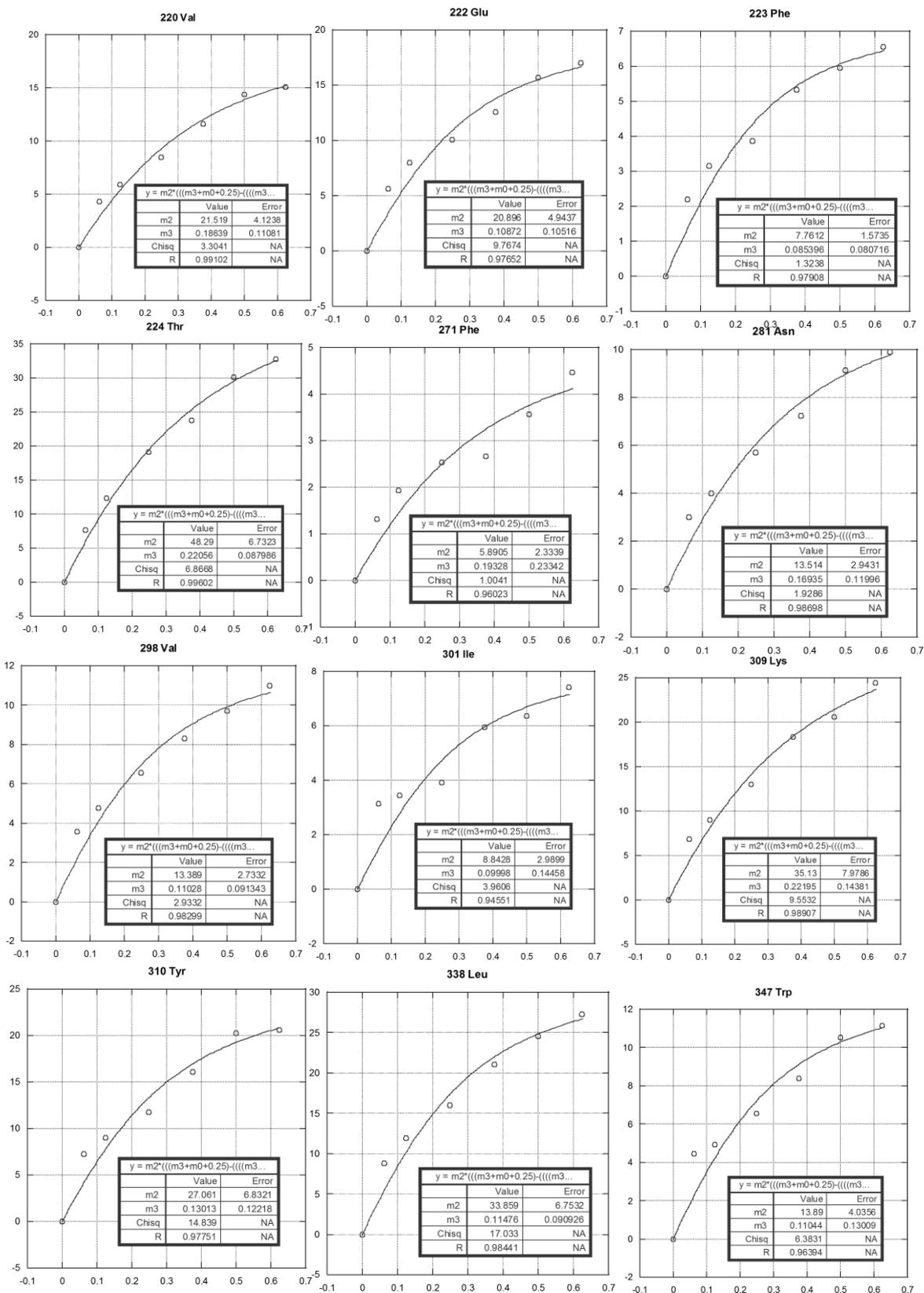


## 5.4 Representative fits of 0.25 mM <sup>15</sup>N W111F *abb*'x (reduced) + Δ-som alanine substitutions:

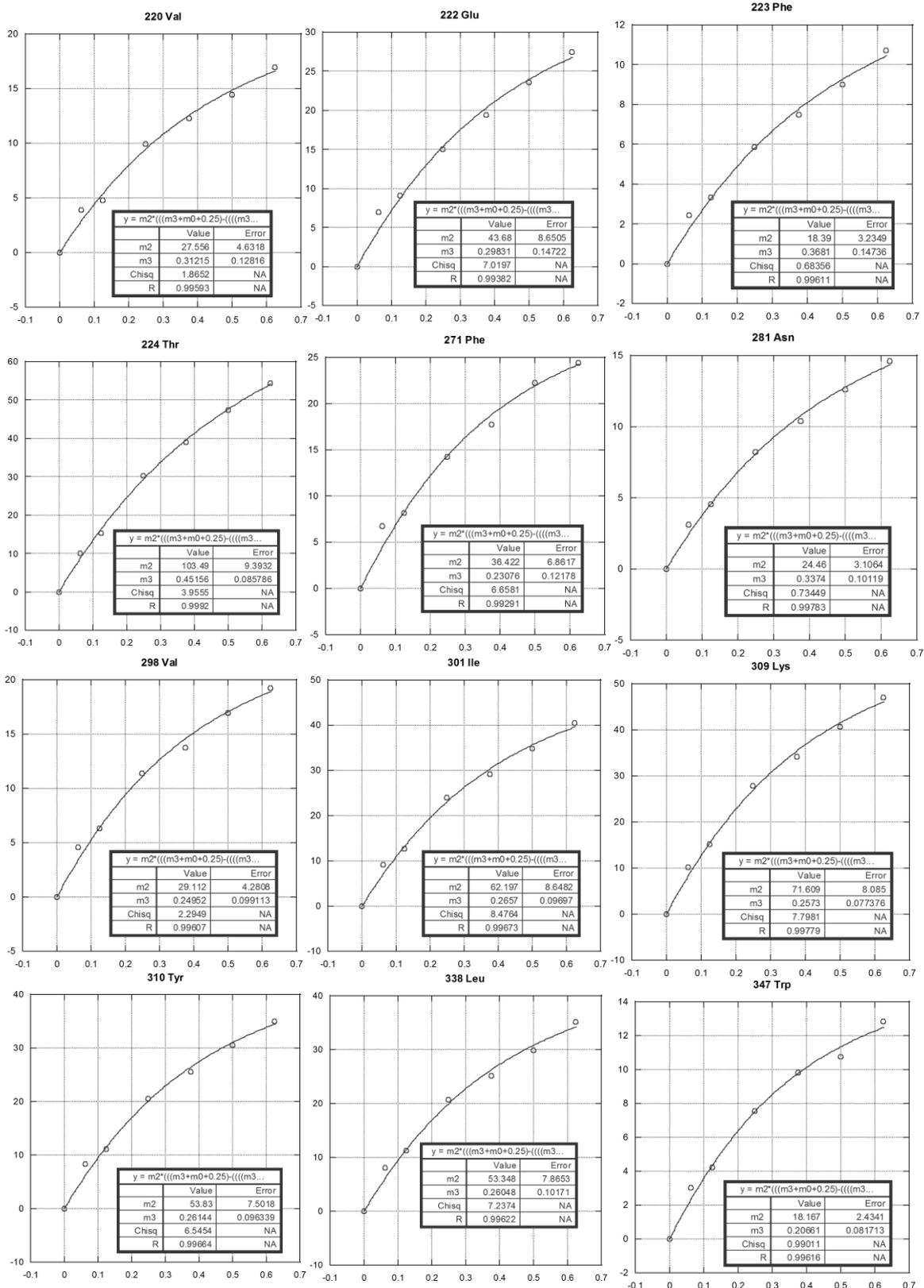
### (a) + Phe 6 → Ala Δ-som



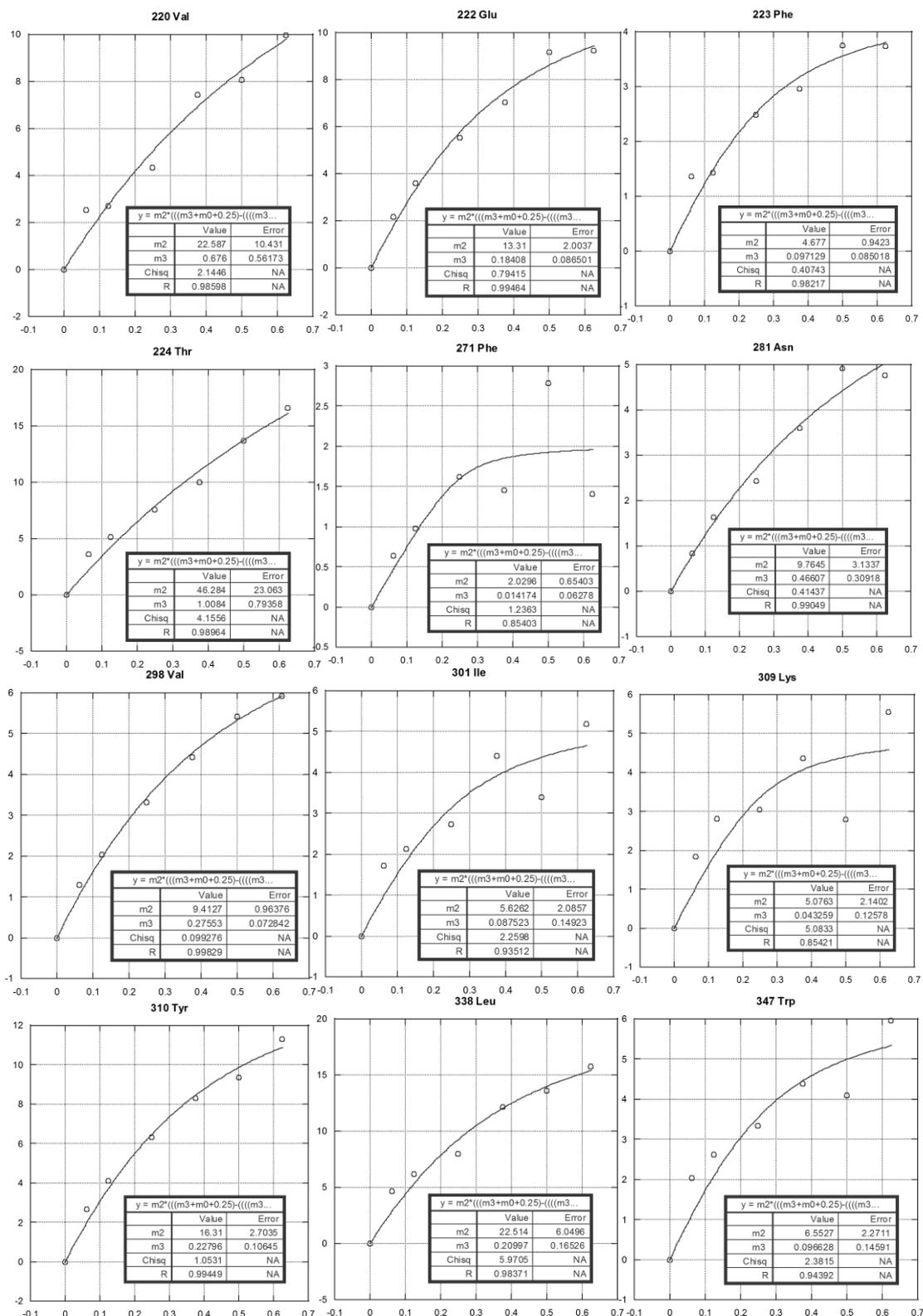
## (b) + Phe 7 → Ala Δ-som



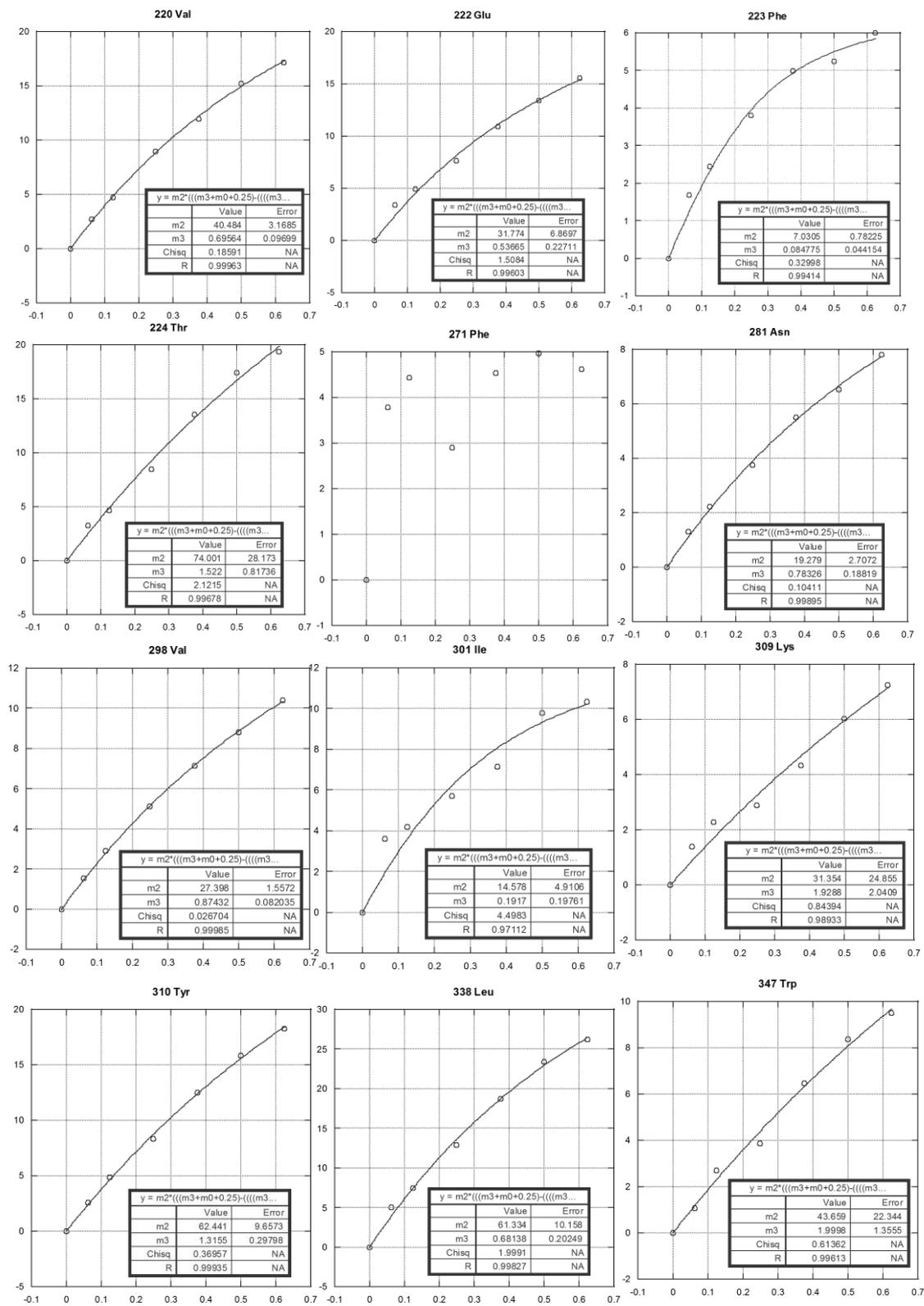
(c) + Phe 11 → Ala Δ-som



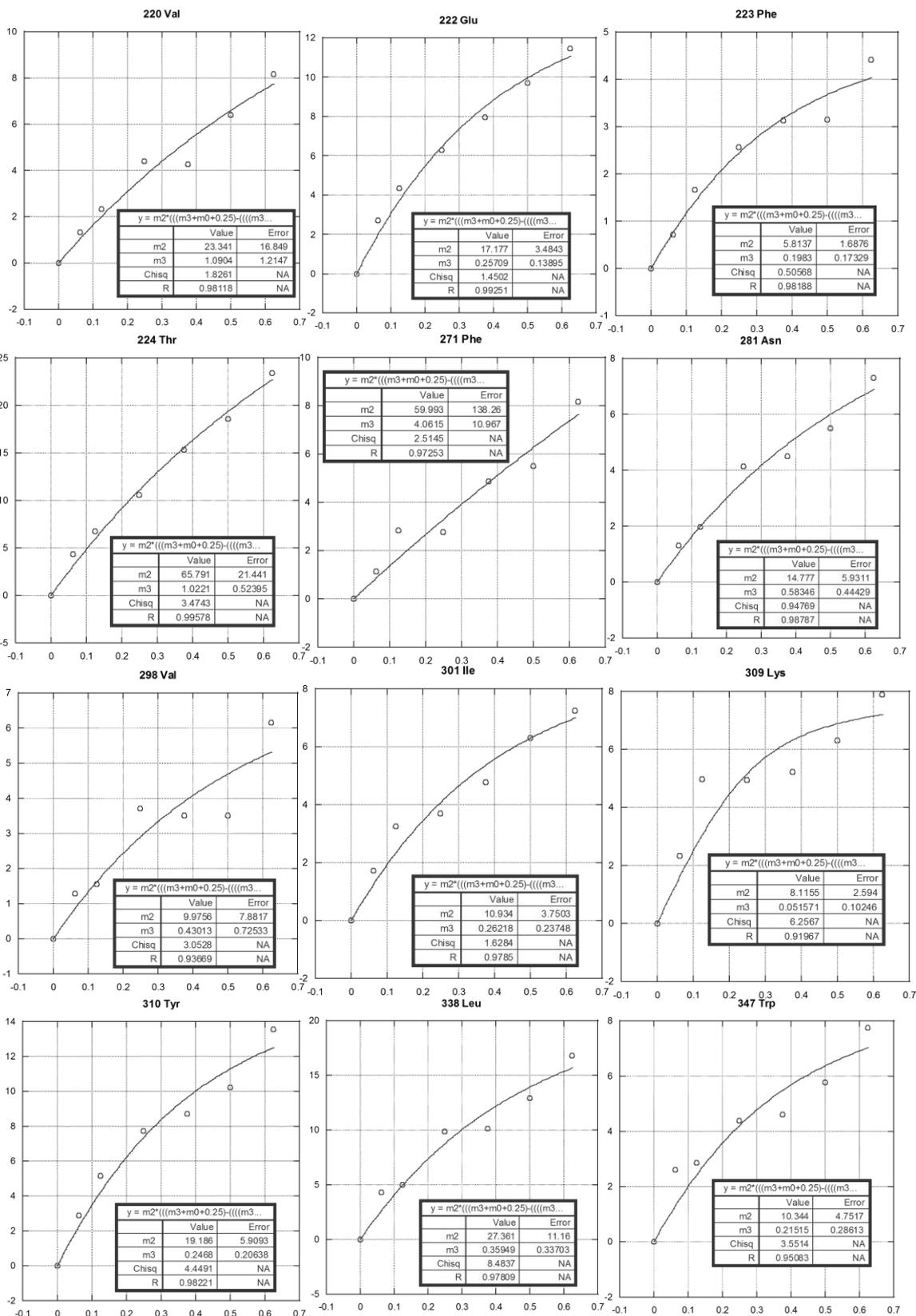
(d) + Phe 6,7 → Ala Δ-som



## (e) + Phe 6,11 → Ala Δ-som



(f) + Phe 7,11 → Ala Δ-som



### 6.1 Table of Redox Potential Samples

All peak intensities were made relative to the fully reduced control spectra (*highlighted in green*).

Redox Potential (mV)	[GSH] /mM	[GSSG] /mM	[DTTred] /mM	[DTTox] /mM
-299			5	0
-256			4	1
-238			2.5	2.5
-232			2	3
-229			1.75	3.25
-224			1.35	3.65
-219			1	4
-208			0.5	4.5
-198			0.25	4.75
-190	4.8	0.2		
-189			0.125	4.875
-182			0.075	4.925
-179	4.6	0.4		
-176			0.001	4.999
-161	3.9	1.1		
-156	3.75	1.25		
-146	3	2		
-138	2.5	2.5		
-126	1.8	3.2		
-108	1	4		
-103	0.85	4.15		
-97	0.7	4.3		
-68	0.25	4.75		
-44	0.1	4.9		
-38	0	5		

## 6.2 <sup>15</sup>N W111F *abb*'x Redox Potential Raw Data (F<sub>red</sub>)

Redox Potentials in red collected using DTT<sup>red</sup>:DTT<sup>ox</sup>, in black using GSH:GSSG

Kaleidagraph Equation Fit:  $m2 + ((m3 * m0^{m4}) / (m5^{m4} + m0^{m4}))$ ;  $m2=1$ ;  $m3=1$ ;  $m4=2$ ;  
 $m5=100$

Redox Potential (mV)	14 Lys	21 Leu	31 Phe	32 Tyr	33 Ala	35 Trp	40 Lys
<b>-299</b>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
<b>-256</b>	1.005	0.999	0.918	1.098	1.082	0.991	1.060
<b>-238</b>	1.019	0.935	0.929	0.958	0.936	1.094	1.028
<b>-232</b>	1.036	1.004	0.928	1.000	0.982	0.974	1.040
<b>-229</b>	1.006	0.963	0.913	1.127	0.944	1.062	1.077
<b>-224</b>	1.034	1.012	0.880	0.901	0.918	0.998	1.009
<b>-219</b>	0.997	1.041	0.858	1.126	1.007	1.024	1.071
<b>-208</b>	1.012	1.028	0.873	0.965	1.065	1.031	1.030
<b>-198</b>	1.024	0.995	0.904	1.100	1.094	1.081	1.090
<b>-189</b>	0.950	0.962	0.989	1.050	1.002	1.072	1.047
<b>-182</b>	0.831	0.819	0.766	0.719	0.887	0.867	0.903
<b>-176</b>	0.640	0.629	0.626	0.550	0.656	0.646	0.666
<b>-190</b>	0.623	0.759	0.663	0.506	0.299	0.557	0.553
<b>-179</b>	0.583	0.765	0.659	0.587	0.643	0.520	0.549
<b>-161</b>	0.342	0.548	0.366	0.284	0.336	0.293	0.298
<b>-156</b>	0.356	0.468	0.345	0.356	0.192	0.235	0.351
<b>-146</b>	0.212	0.374			0.712	0.254	0.173
<b>-138</b>	0.178	0.263	0.418	0.275	0.197	0.000	0.000
<b>-126</b>	0.153	0.000	0.000	0.000	0.664	0.000	0.000
<b>-108</b>	0.000	0.000	0.000	0.000	0.452	0.000	0.000
<b>-103</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>-97</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>-68</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>-44</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>-38</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Redox Potential (mV)	41 Ala	42 Leu	43 Ala	65 Val	66 Asp	67 Ala	70 Glu
<b>-299</b>	1.000	1.000	1.000	1.000	1.000	1.000	1.000
<b>-256</b>	1.110	1.016	1.056	0.848	1.022	1.052	0.973
<b>-238</b>	1.056	1.027	0.995	0.913	1.036	1.047	0.995
<b>-232</b>	1.079	1.039	1.052	0.872	0.957	0.981	0.992
<b>-229</b>	0.983	1.044	0.962	0.970	1.120	1.057	0.988
<b>-224</b>	1.065	1.038	1.041	0.835	1.024	0.954	0.988
<b>-219</b>	1.070	1.083	1.082	0.939	1.013	0.973	1.043
<b>-208</b>	1.063	1.072	1.072	0.950	0.981	0.943	1.035
<b>-198</b>	1.027	0.992	1.060	0.824	0.997	0.959	1.003
<b>-189</b>	0.990	0.974	1.047	0.905	0.917	0.921	0.957
<b>-182</b>	0.887	0.904	0.871	0.706	0.876	0.744	0.893
<b>-176</b>	0.665	0.634	0.675	0.524	0.665	0.564	0.758
<b>-190</b>	0.566	0.539	0.554	0.613	0.669	0.537	0.900
<b>-179</b>	0.522	0.445	0.550	0.678	0.570	0.504	0.863
<b>-161</b>	0.340	0.288	0.339	0.303	0.338	0.398	0.746
<b>-156</b>	0.257	0.257	0.339	0.469	0.166	0.277	0.731
<b>-146</b>	0.208	0.166	0.240	0.371	0.000	0.376	0.650
<b>-138</b>	0.168	0.175	0.269	0.265	0.000	0.000	0.663
<b>-126</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.604
<b>-108</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.134
<b>-103</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.043
<b>-97</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>-68</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>-44</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<b>-38</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Redox Potential (mV)	78 Gly	81 Gly	84 Thr	90 Asn
<b>-299</b>	1.000	1.000	1.000	1.000
<b>-256</b>	1.097	1.017	0.925	0.886
<b>-238</b>	1.118	1.056	0.940	1.013
<b>-232</b>	1.131	1.010	0.988	0.913
<b>-229</b>	1.096	1.039	0.979	0.989
<b>-224</b>	1.157	1.015	0.966	0.989
<b>-219</b>	1.200	1.037	0.871	1.056
<b>-208</b>	1.129	1.040	1.031	0.978
<b>-198</b>	1.025	0.983	1.041	1.025
<b>-189</b>	1.059	1.036	0.920	1.023
<b>-182</b>	0.917	0.859	0.739	0.847
<b>-176</b>	0.808	0.690	0.565	0.622
<b>-190</b>	0.649	0.680	0.648	0.498
<b>-179</b>	0.611	0.615	0.492	0.452
<b>-161</b>	0.328	0.332	0.243	0.276
<b>-156</b>	0.370	0.325	0.325	0.257
<b>-146</b>	0.235	0.112	0.369	0.313
<b>-138</b>	0.203	0.164	0.217	0.000
<b>-126</b>	0.000	0.136	0.487	0.000
<b>-108</b>	0.000	0.000	0.388	0.000
<b>-103</b>	0.000	0.000	0.000	0.000
<b>-97</b>	0.000	0.000	0.000	0.000
<b>-68</b>	0.000	0.000	0.000	0.000
<b>-44</b>	0.000	0.000	0.000	0.000
<b>-38</b>	0.000	0.000	0.000	0.000

7.1 Fits of 0.25 mM  $^{15}\text{N}$  6-F-Trp W111F *abb'*x (reduced) +  $\Delta$ -som: