The Interaction of Protein Disulfide Isomerase with a Substrate Protein at Different Stages along its Folding Pathway

Alistair Irvine

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy

Molecular Organisation and Assembly in Cells (MOAC)
Doctoral Training Centre

The University of Warwick

June 2010

DECLARATION

I hereby declare that the research submitted in this thesis was conducted by myself under the supervision of Prof. R. B. Freedman at the Department of Biological Sciences, University of Warwick and Dr. C. A. Blindauer, Department of Chemistry, University of Warwick

No part of this work has previously been submitted to be considered for a degree or other qualification. All sources of information have been specifically acknowledged in the form of references.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors Prof. Robert Freedman and Dr. Claudia Blindauer for the opportunity to undertake this PhD and for all their advice throughout the duration of the project.

Many thanks to postdoc Dr. Katrine Wallis for her patient assistance and advice in many aspects of my lab work. Thanks to Dr. Oksana Leszczyszyn for demonstration of NMR and mass spectrometry apparatus. Thanks also to many others in the Structural Biology and Chemical Biology laboratories for their suggestions and advice throughout my PhD.

Thank you to Dr. Lloyd Ruddock (University of Oulu) who kindly provided the original wild type BPTI construct used in this study and Dr. Edward Jack, who performed the mutagenesis on this construct.

I am also grateful to our collaborators from the University of Kent, where much of the NMR data was acquired. Dr. Michelle Rowe, Dr. Lee Byrnes and Dr. Richard Williamson assisted in the preparation, acquisition and processing of data. Dr. Mark Howard and Dr. Richard Williamson provided invaluable advice and feedback for the interpretation of data.

This multidisciplinary PhD was funded by the Engineering and Physical Sciences Research Council (EPSRC) through the Molecular Organisation and Assembly in Cells (MOAC) Doctoral Training Centre, under the direction of Prof. Alison Rodger. I would like to thank all my colleagues at MOAC for their support and friendship.

Finally I would like to thank my parents and wider family, which has had 3 new members since beginning my studies at the University of Warwick (William, Isaac and Lucy), with another one on the way.

ABSTRACT

Understanding the mechanisms through which proteins acquire their three dimensional structure is currently one of the most challenging tasks in structural biology. The formation of native disulfide bonds is an important step in the post-translational modification and folding of many proteins, helping to stabilise their structure. Protein disulfide isomerase (PDI) is a folding enzyme that catalyses thiol-disulfide exchange. As well as forming disulfide bonds in newly synthesised proteins, PDI also catalyses the rearrangement of intramolecular disulfides. The mechanisms through which PDI binds to substrate proteins are still not well understood.

In this study, interactions are examined between PDI and a model substrate protein, bovine pancreatic trypsin inhibitor (BPTI). Since PDI functions primarily as a folding enzyme its natural substrates will be unfolded or partly folded proteins. Here, recombinant BPTI constructs were prepared that represented different stages along the folding pathway of this small protein: unfolded, partly folded and natively folded BPTI.

A variety of biophysical techniques were then used to characterise each BPTI construct, both in isolation and in the presence of PDI. NMR spectra obtained at 5°C, including hydrogen deuterium exchange experiments, demonstrated the unfolded, partly folded and natively folded nature of each construct at low temperatures. The addition of PDI to each BPTI construct showed that, even at sub-stoichiometric concentrations, both the unfolded and partly-folded substrate proteins showed line broadening. In contrast, line broadening of natively folded BPTI required much higher concentrations of PDI.

NMR was also used to observe the effects of differently folded BPTI substrates binding to PDI. Focus was on the key bb'x binding region of PDI. Perturbations were observed even at low concentrations of unfolded and partly-folded substrate, whereas much larger concentrations were required for the natively folded protein. However, detailed investigations into the specific regions of binding suggest that the same key sites were involved at all stages of folding. Contrary to expectations, this small full length protein showed little binding to regions beyond the key b' domain.

The binding affinities between PDI and each BPTI substrate were estimated using surface plasmon resonance (SPR). As expected, PDI has a greater binding affinity to unfolded BPTI compared to the partly folded construct, with least affinity to the natively folded protein. However, the difference in affinity between unfolded and partly folded constructs was relatively small.

This is the first study to investigate the structural interaction of PDI with a partly folded, full length protein substrate. It is hoped that the findings of this study will contribute to a general understanding of oxidative protein folding in the endoplasmic reticulum (ER).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

BMRB	biological magnetic resonance data bank
BPTI	bovine pancreatic trypsin inhibitor
CD	circular dichroism
CV	column volume
E. coli	*Escherichia coli*
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
Ero1	endoplasmic reticulum oxidoreductin 1
ESI	electrospray ionisation
FAD	Flavin Adenine Dinucleotide
FID	free induction decay
FPLC	fast protein liquid chromatography
GAPDH	D-glyceraldehyde-3-phosphate dehydrogenase
GSH	glutathione
GSSG	glutathione disulfide
UGGT	UDP-glucose:glycoprotein glucosyltransferase
HBS-EP	HEPES buffered saline with EDTA and P20
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	hexa-histidine tag
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani
LMW	low molecular weight
MES	2-(N-morpholino)ethanesulfonic acid
MTP	microsomal triglyceride transfer protein
MWCO	molecular weight cut-off
NHS	N-hydroxysuccinimide

NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
OD	optical density
ORF	open reading frame
P4-H	collagen prolyl 4-hydroxylase
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PDI	protein disulfide isomerase
ppm	parts per million
RU	response units
SDS	sodium dodecyl sulphate
SPE	solid phase extraction
STD	saturation transfer difference
TFA	trifluoroacetic acid
TOCSY	total correlation spectroscopy
TOF	time of flight

Standard amino acid abbreviations used throughout.

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

Chapter 1. Introduction

This chapter introduces some general concepts involved in the study of protein folding, with an emphasis on the machinery used to facilitate folding in the endoplasmic reticulum (ER). The processes of oxidation and isomerisation of disulfide bonds are introduced along with protein disulfide isomerase (PDI), an important enzyme in catalysing these processes. As well as its function in catalysing thiol-disulfide exchange, an outline of the other functions of PDI is given. Different domains of PDI play interdependent roles in the structure and function of PDI. Current understanding of each domain is discussed.

PDI is the archetypal member of the PDI family of proteins. Current understanding of this family is introduced together with many of its member proteins. This family is part of the wider thioredoxin superfamily, which is also briefly introduced.

As well as PDI, the process of oxidative folding in the ER involves other proteins and chemicals used to deal with oxidising equivalents. The pathways involved are not clearly understood, but recent studies in this area are introduced.

Bovine pancreatic trypsin inhibitor (BPTI) is a model substrate protein for studies of PDI activity. This small protein is introduced along with its well characterised folding pathway. Previous studies involving PDI/BPTI interaction are reviewed.

The main biophysical technologies used in this study were NMR spectroscopy and surface plasmon resonance. The use of these techniques for studying protein structures and binding interactions are outlined.

Finally, the aims and scope of this study will be outlined, indicating the manner in which PDI/BPTI interactions will be studied using these technologies.

1.1. The Protein Folding Energy Landscape

In order to function correctly, most proteins require a specific three dimensional structure. Some of the earliest studies into folding showed that the polypeptide chain contains all the necessary information to specify its three dimensional conformation (Anfinsen 1973). However, despite an incredibly large number of possible conformations, proteins manage to achieve their natively folded state in an incredibly short timescale (Kubelka, Hofrichter et al. 2004). It has been demonstrated that many newly synthesised proteins require assistance to reach their folded states within a biologically relevant timescale (Hartl 1996). The mechanism by which they achieve this has been an immensely difficult problem, spanning many decades of study (Fersht 2008).

The folding of proteins is often described by a funnelling energy landscape, whereby they seek a conformational state that minimises their free energy, thereby increasing their stability (Bryngelson, Onuchic et al. 1995; Onuchic and Wolynes 2004), Figure 1.1. The ruggedness of the energy landscape arises due to the fact that protein structures are stabilised by thousands of weak interactions that cannot all be satisfied simultaneously during folding (Bartlett and Radford 2009).

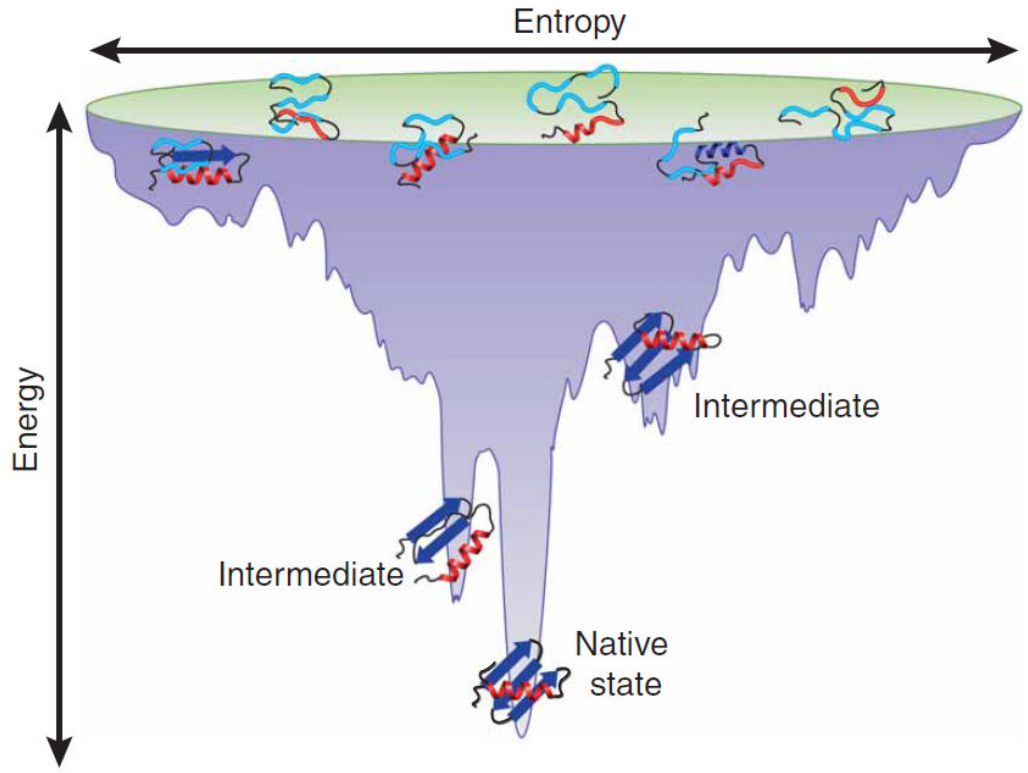


Figure 1.1: Schematic representation of a protein folding funnel energy landscape, modified from (Bartlett and Radford 2009). Nascent chain polypeptides can initially explore a wide range of conformations. Certain conformations occupy a lower energy state and are more stable, representing intermediates in the folding pathway. Often intermediates create an energy barrier that must be overcome before folding can proceed (*e.g.* some localised unfolding). The native state is usually the lowest energy and most stable state.

During the folding process, many metastable intermediates often include non-native interactions, which need to be overcome before the native state can be achieved (Hartl and Hayer-Hartl 2009). Although small single chain proteins can show distinct intermediates during protein folding (Creighton 1992), intermediate states are even more relevant to large proteins, which have a greater tendency to collapse into non-native conformations due to exposure of hydrophobic regions in an aqueous environment (Brockwell and Radford 2007).

In a cell folding occurs in a crowded environment, so unfolded or partly folded proteins are much more likely to aggregate compared to the relatively dilute conditions of most *in vitro* studies (Ellis and Minton 2006). However, the cellular environment also includes many molecular chaperones, which serve to prevent protein misfolding or aggregation (Ellis 2001; Saibil 2008).

Recent studies suggest that during the protein folding process molecular crowding appears to favour compact rather than expanded conformations (Charlton, Barnes et al. 2008; Engel, Westphal et al. 2008; Zhou, Rivas et al. 2008). To facilitate our understanding, molecular dynamic simulations of protein folding have recently begun to use bulk solvents that account for the macromolecular environment of the cell, including parameters of spatial confinement and molecular interactions (Cheung and Thirumalai 2007; Mittal and Best 2008; Echeverria and Kapral 2010).

1.2. Protein Folding in the Endoplasmic Reticulum

In eukaryotic cells, proteins destined for surface membranes or secretion must first enter the endoplasmic reticulum (ER) in an unfolded state and only exit once they have correctly folded and assembled. This process is performed for an estimated 33% of all human proteins (Chen, Zhang et al. 2005).

The main function of the ER is to perform post-translational modifications and folding of proteins. These processes are highly regulated, with the ER containing a wide variety of catalysts for both post-translational modifications, such as N-glycosylation (Helenius and Aebi 2004), and for protein folding, such as disulfide bond formation (Sevier and Kaiser 2002).

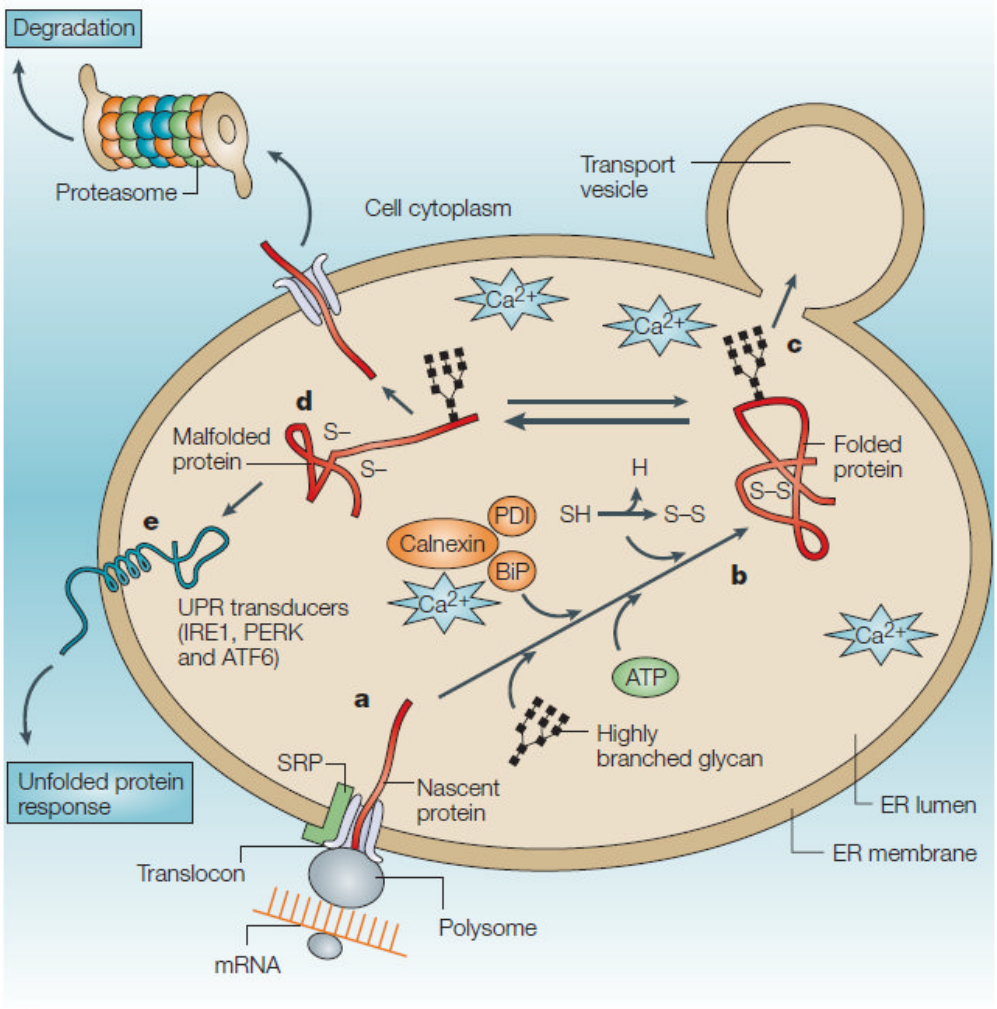


Figure 1.2: Schematic of protein folding and degradation pathways in the endoplasmic reticulum (Ma and Hendershot 2004). a) Secretory proteins are translocated into the ER lumen co-translationally through translocons in the ER membrane. b) Chaperones such as BiP and calnexin bind to nascent chain proteins, preventing misfolding or aggregation while monitoring glycosylation. Protein disulfide isomerase (PDI) facilitates the formation of disulfide bonds to stabilise the folded protein. c) Natively folded proteins exit the ER through transport vesicles. d) If a protein becomes misfolded, it will be targeted for degradation by the proteasome. e) If unfolded proteins accumulate, the unfolded protein response (UPR) is initiated by ER-stress sensors (IRE1, PERK, ATF6). SRP, signal-recognition particle; PERK, PKR-like ER kinase.

First, proteins are translated on membrane-bound polysomes and translocated in an extended, unfolded state through the translocon, a proteinaceous channel

in the ER membrane (Ma and Hendershot 2004), Figure 1.2a. These newly synthesised protein molecules are not immediately physiologically functional; they require folding into their native state.

Immediately as proteins are synthesised they begin to fold co-translationally in the ER, with folding being completed post-translationally. This combination has been shown to enhance the folding efficiency of proteins, particularly with multi-domain proteins (Netzer and Hartl 1997).

Unfolded or partially folded proteins in the ER have exposed hydrophobic regions, so are prone to aggregation. Small aggregates are potentially cytotoxic, so a quality control method preventing their accumulation is essential (Ellgaard and Helenius 2003). Chaperones are able to recognise proteins in this state, interact with them and prevent their aggregation (Fewell, Travers et al. 2001). A chaperone system exists in the ER that associates with the newly synthesised proteins to prevent their aggregation and help them fold and assemble correctly (Ellgaard, Molinari et al. 1999; Ma and Hendershot 2004), Figure 1.2b. The ER is also an oxidising and Ca^{2+}-rich environment, which facilitates the addition of N-linked glycans and disulfide bonds to proteins during their maturation (Ellgaard and Helenius 2001).

One set of chaperones are the lectin proteins calnexin (CNX) and calreticulin (CRT). Calnexin is a transmembrane protein whereas calreticulin is a soluble luminal protein (Jessop, Tavender et al. 2008). Lectins interact with glycoproteins, but CNX and CRT only specifically interact with glycoproteins that have monoglucosylated N-glycans (Hammond and Helenius 1994). They function to retain non-natively folded glycoproteins in the ER via a quality control system known as the calnexin/calreticulin cycle. This cycle is a multifunctional system, responsible for promoting the folding of glycoproteins, retaining non-native glycoproteins in the ER until they are correctly folded and targeting misfolded glycoproteins for degradation (Helenius and Aebi 2004).

When some nascent polypeptide chains enter the ER lumen, a process of N-linked glycosylation occurs via transfer of a triglucosylated, branched core oligosaccharide (Ellgaard and Helenius 2001). The calnexin/calreticulin cycle is initiated by the removal of two glucose residues by the sequential action of glucosidases I and II, generating a monoglucosylated glycoprotein. The remaining glucose effectively acts as a selective tag for incompletely folded proteins. The monoglucosylated glycoprotein then interacts with calnexin and calreticulin (Ellgaard and Helenius 2003). If disulfide bond formation is required, both chaperones can associate with the PDI family protein ERp57, via an extended domain known as the P domain (Solda, Garbi et al. 2006). The remaining glucose is cleaved by glucosidase II, which results in dissociation of calnexin and calreticulin. If natively folded, glycoproteins can exit the ER. However, non-native glycoproteins become substrates of UDP-glucose:glycoprotein glucosyltransferase (UGGT), which places a single glucose back on the glycan. This again targets the glycoprotein as a substrate for calnexin and calreticulin. UGGT effectively acts as a sensor of glycoprotein conformations. The deglucosylation-reglucosylation cycle continues until proper folding is achieved (Parodi 2000). A protein can only exit the cycle when UGGT fails to re-glucosylate it, either when its native conformation has been achieved or if the glycoprotein becomes targeted for degradation (Ellgaard and Helenius 2003). Overall, CRT and CNX together with glucosidase II and UGGT cooperate to increase the folding efficiency, to prevent premature oligomeric assembly and to prevent the export of misfolded glycoproteins from the ER (Ellgaard and Helenius 2001).

Another ER chaperone known as BiP belongs to the Hsp70 family of proteins (Flynn, Pohl et al. 1991). These proteins use ATP and ADP to regulate the binding and release of nascent proteins in an ATPase cycle (Bukau and Horwich 1998). During these cycles of binding and release, aggregation or

inappropriate interactions are prevented by BiP binding, whereas protein folding occurs during release (Hendershot, Wei et al. 1996).

The formation of native disulfide bonds is carried out by the ubiquitously expressed protein disulfide isomerase (PDI). Disulfide bonds in proteins are formed as covalent bonds between the thiol groups of two cysteine residues. Their main function is to stabilise protein structure. The formation of disulfide bonds requires an oxidising environment. Although precise figures vary, it is known that the ER is much more oxidising than the cytosol and provides an environment suitable for creating disulfide bonds (Raines 1997). By catalysing the oxidation and isomerisation of disulfide bonds, PDI can increase the rate of native disulfide bond formation by a factor of between 1000 to 6000 (Freedman, Hirst et al. 1994).

If the secretory protein is correctly folded, then it is transported out of the ER via transport vesicles and continues on the secretory pathway via the Golgi, Figure 1.2c. Alternatively, proteins that do not mature properly are removed from the folding pathway and targeted to the ER-associated degradation (ERAD) pathway, Figure 1.2d. This involves translocation to the cytosol for degradation by the proteasome, which requires both ER and cytosolic chaperones (Brodsky, Werner et al. 1999).

Alterations in any aspect of the ER environment can lead to the accumulation of unfolded proteins, which are prone to aggregation, and could interfere with the normal functioning of the ER (Lee 1992). When adverse physiological conditions that impact on protein folding are encountered in the ER, a signal-transduction cascade is activated, known as the unfolded protein response (UPR), Figure 1.2e. This serves to limit the accumulation of unfolded proteins and preserve the solubility of those that are present so that they can be targeted for degradation (Ma and Hendershot 2004). This coordinated response is an efficient method of maintaining a careful balance between retaining and degrading harmful proteins

while continuing the export of biologically active and essential proteins (Sitia and Braakman 2003).

Prolonged activation of the UPR, if failing to clear the ER of a high concentration of harmful mis-folded or mis-assembled proteins, results in the activation of apoptosis, which can serve to protect the host organism (Otomo, Ito et al. 1999).

1.3. Thiol Oxidation and Isomerisation of Disulfide Bonds

A vital step in the protein folding process is the formation of disulfide bonds between cysteine residue side chains. It occurs in the periplasm of Gram-negative bacteria and in the ER of eukaryotic cells. Figure 1.3 illustrates the initial oxidation of cysteine thiol groups to form disulfide bonds in a polypeptide chain, followed by a subsequent thiol-disulfide interchange (isomerisation) resulting in a different disulfide bonding configuration.

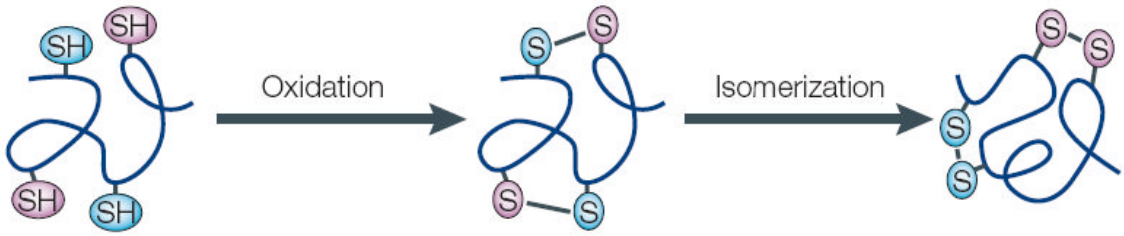


Figure 1.3: Thiol oxidation and isomerisation of disulfide bonds (Sevier and Kaiser 2002).

Early research into protein folding revealed that disulfide bond formation could occur spontaneously (Anfinsen 1973). However, disulfide bond formation occurred much more slowly *in vitro* compared to the environment of the cell. This suggested that enzymes may be involved in catalysing oxidative protein folding.

1.4. Protein Disulfide Isomerase

Protein disulfide isomerase (PDI) (E.C. 5.3.4.1, Swiss-Prot accession number P07237) is a multifunctional protein, able to catalyze disulfide bond formation, breakage, and rearrangement in all studied non-native protein and peptide substrates. Human PDI is encoded by the P4HB (prolyl 4-hydroxylase, beta polypeptide) gene (GeneID 5034). It is an ER localised protein with a KDEL ER retrieval motif (Pelham 1990). It is present in high concentrations (Lambert and Freedman 1985) and highly expressed in secretory tissues such as the liver and pancreas. In addition to its role in disulfide bond formation, PDI acts as a molecular chaperone (Cai, Wang et al. 1994) and forms the β-subunit of prolyl-4-hydroxylase (Koivu, Myllyla et al. 1987) and microsomal triglyceride transfer protein (Wetterau, Combs et al. 1991).

Christian Anfinsen and his group first identified the enzymatic acceleration of protein reactivation from reduced protein in 1963 (Goldberger, Epstein et al. 1963). They were later able to demonstrate that these enzymes could catalyse thiol-disulfide exchange reactions (Givol, Goldberger et al. 1964; Goldberger, Epstein et al. 1964; Givol, Delorenzo et al. 1965). Independently, another group also discovered the role of enzymes in disulfide bond formation (Venetianer and Straub 1963). PDI was purified from beef liver and confirmed as an enzyme which catalysed sulfhydryl-disulfide interchange in proteins (De Lorenzo, Goldberger et al. 1966), making it the first identified catalyst of protein folding. However, its current name was not used until 1975 (Hawkins and Freedman 1975). Later, PDI localisation to the ER was confirmed (Lambert and Freedman 1985) and was shown to exist in high concentrations (Hillson, Lambert et al. 1984). The sequence of rat PDI suggested a high degree of homology to the thioredoxin protein (Edman, Ellis et al. 1985; Creighton 1995a).

PDI degrades insulin by cleavage of the disulfide bonds using glutathione as a reducing agent and therefore was also known as glutathione-insulin

transhydrogenase (Noiva, Kimura et al. 1991). This function was originally thought to be performed by a separate enzyme (Hillson and Freedman 1980), but later glutathione-insulin transhydrogenase and PDI were identified as being the same protein (Bjelland, Wallevik et al. 1983).

1.4.1. PDI Catalysis of Thiol-disulfide Exchange

The principal function of PDI is as a thiol-disulfide oxidoreductase. PDI can catalyse three types of thiol-disulfide exchange reactions: oxidation, reduction and isomerisation, illustrated in Figure 1.4. These reactions involve the active sites of either the a or a' domains of PDI, which are defined by the amino acid sequence CGHC.

The oxidation of a substrate by PDI converts a substrate dithiol to a disulfide state (Figure 1.4A). The N-terminal active site cysteine thiolates are more reactive, so tend to form the mixed disulfides with the substrate. To complete the catalytic cycle, PDI needs to be reoxidised, *e.g.* using oxidised glutathione.

The reduction of a substrate by PDI converts a substrate disulfide bond into the dithiol state (Figure 1.4B). This reaction also involves an intermediate mixed disulfide between PDI and the substrate. The process oxidises PDI and to complete the catalytic cycle PDI needs to be returned to its reduced state, *e.g.* using reduced glutathione.

The isomerisation of a substrate by PDI involves the rearrangement of disulfides and thiols resulting in a substrate with a different configuration of disulfide bonds (Figure 1.4C). This can be achieved either via a single mixed disulfide (direct isomerisation), or via a combination of reduction and oxidation reactions by PDI. The isomerisation of disulfide bonds is an especially important role for PDI, since this process is often the rate limiting factor in protein folding (Huth, Perini et al. 1993; Land, Zonneveld et al. 2003).

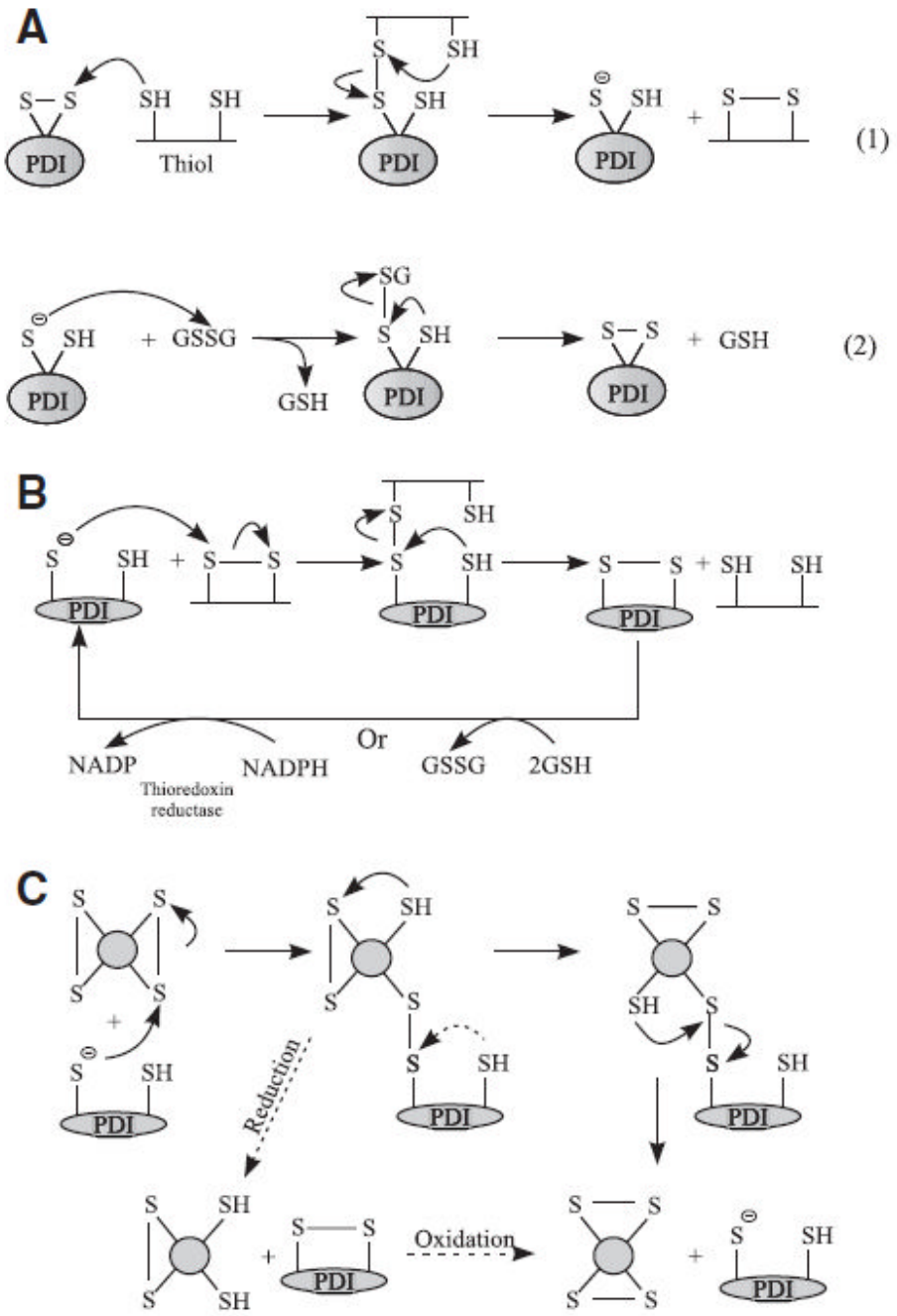


Figure 1.4: Schematics of the thiol-disulfide exchange reactions catalysed by PDI, modified from (Hatahet and Ruddock 2009). A) PDI oxidation of a peptide dithiol (reaction 1). Oxidised glutathione can then act as an electron acceptor to re-oxidise PDI (reaction 2). B) PDI reduction of a peptide disulfide. Reduced glutathione or NADPH (catalysed by thioredoxin reductase) can act as electron acceptors to reduce PDI. C) PDI isomerisation of peptide disulfide bonds. Isomerisation can take place either directly or via reduction-oxidation cycles. Thiol species are shown in their predominant ionisation state.

1.5. The Thioredoxin Superfamily

PDI contains four domains with the thioredoxin structural motif, and as such is classified as being a member of the thioredoxin superfamily (Appenzeller-Herzog and Ellgaard 2008). The thioredoxin fold consists of a four stranded β-sheet surrounded by three α-helices in a βαβαββα configuration (Martin 1995), Figure 1.5A. The protein of the same name consists of the thioredoxin fold, with an additional N-terminal βα, resulting in a βαβαβαββα configuration, Figure 1.5B (Weichsel, Gasdaska et al. 1996). The thioredoxin protein is a small 12 kDa protein found in all species (Powis and Montfort 2001). It is common to many protein families associated with sulfur-based redox reactions, including PDI, glutaredoxin and DsbA. Most of these enzymes contain a conserved active site motif of the form Cys-X-X-Cys, with the two intervening residues of the motif varying between families (Eklund, Gleason et al. 1991). There are also several charged residues surrounding the active site motif which are thought to influence its reactivity (Gruber, Cemazar et al. 2006).

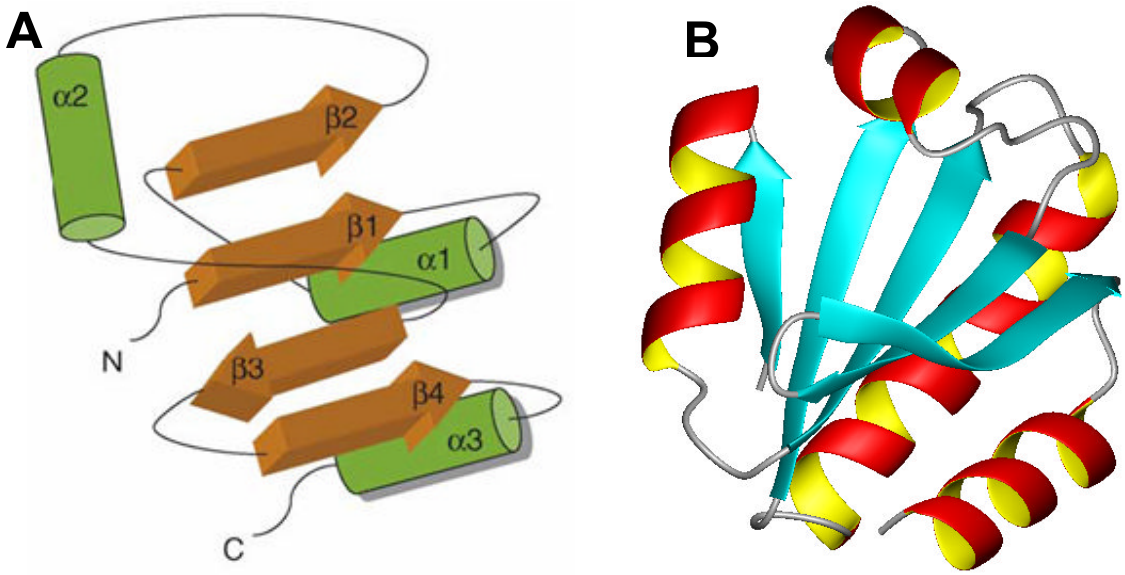


Figure 1.5: The thioredoxin fold in the protein thioredoxin. A) structure of the thioredoxin fold, illustrating its βαβαββα configuration, modified from (Gruber, Cemazar et al. 2006); B) structure of the reduced human thioredoxin protein, which consists of the thioredoxin fold plus an additional βα at the N-terminus, PDB ID 1ERT (Weichsel, Gasdaska et al. 1996).

1.6. Domain Architecture of PDI

The cDNA of PDI was first sequenced in 1985 from rat liver (Edman, Ellis et al. 1985). It encoded a total of 508 amino acids, which includes an N-terminal ER signal sequence and a C-terminal ER retention sequence. The mature protein is 491 amino acids long.

The current multi-domain architecture of PDI has been proposed using a combination of sequence similarities, comparing with superfamily members (*e.g.* thioredoxin) and experimental data, including proteolysis and expression of putative domains and domain combinations (Freedman, Gane et al. 1998; Alanen, Salo et al. 2003). The model has four distinct structural domains, a, b, b', and a', a highly acidic C-terminal extension and a 19 amino acid flexible interdomain region between the b' and a' domains, known as the x linker (Kemmink, Darby et al. 1997; Freedman, Gane et al. 1998), Figure 1.6. The catalytic activity of PDI depends on the CGHC motif in the a and a' domains. The catalytic a and a' domains are separated by two non-catalytic b and b' domains as well as the x-region.

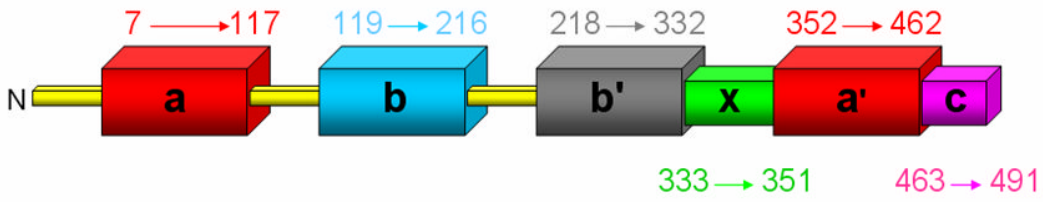


Figure 1.6: Human PDI domain architecture. Numbers indicate domain boundaries, with residue numbers representing mature human PDI (Sidhu 2008).

Despite intense efforts, no high resolution structure of full length human PDI has been achieved, although low resolution data has been achieved through small angle X-ray scattering of PDI in solution (Li, Hong et al. 2006). However, the

structures of isolated domains and domain combinations have been solved. These include the a domain, PDB ID 1MEK (Kemmink, Darby et al. 1996); the a' domain, PDB ID 1X5C (Tochio, Koshiba et al. Unpublished); the b domain, PDB ID 2BJX (Kemmink, Darby et al. 1997); the b'x construct, PDB ID 3BJ5 (Nguyen, Wallis et al. 2008); the bb' construct, PDB ID 2K18 (Denisov, Maattanen et al. 2009).

The solved structures of all four structural domains show that they each contain a thioredoxin fold. Hence, it has been postulated that these domains evolved from internal gene duplication, which have since lost sequence similarity but retained the overall domain fold. Studies on the molecular evolution of PDI suggested that it evolved from an ancestral thioredoxin-like double domain protein (Kanai, Toh et al. 1998; McArthur, Knodler et al. 2001).

A crystal structure for yeast PDI (Pdi1p) was recently published at 2.4 Å resolution, PDB ID 2B5E (Tian, Xiang et al. 2006), Figure 1.7. This has helped to establish a more detailed picture of the structure of human PDI. The structure contained one molecule of Pdi1p occupying the putative substrate-binding site of another molecule. It shows the four domains in a twisted U shape.

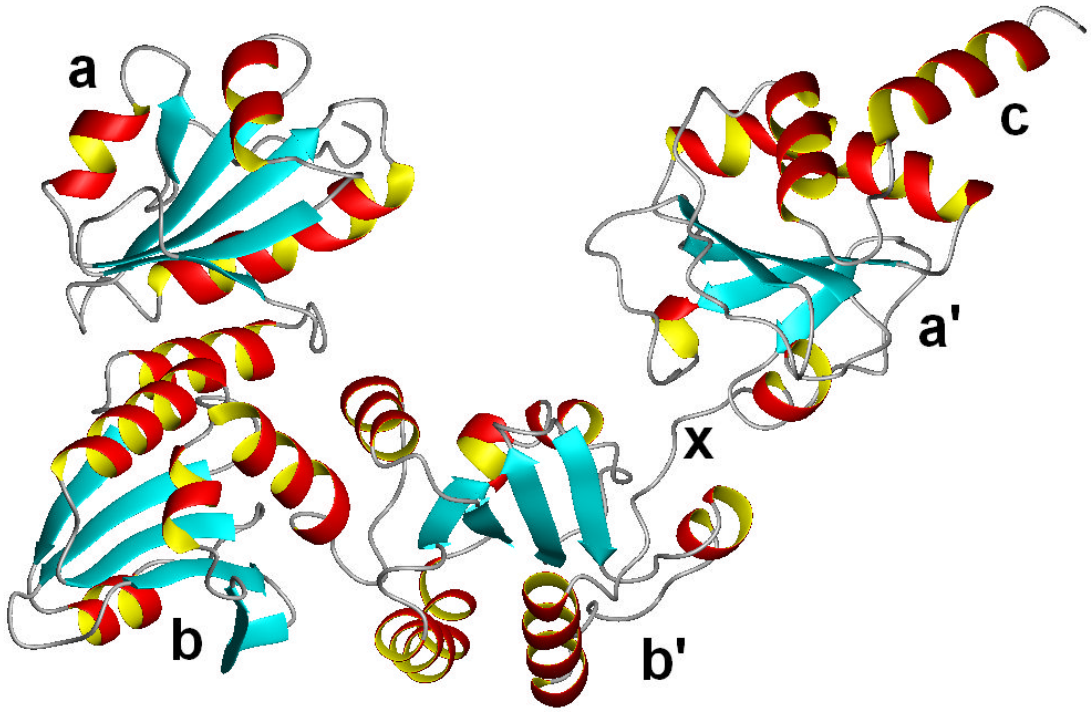


Figure 1.7: Ribbon diagram of Pdi1p, from PDB ID 2B5E (Tian, Xiang et al. 2006), illustrating its 'twisted U' structure. Each domain is labelled, along with the interdomain x linker.

In Pdi1p, the a and a' domains form the outstretched arms, with their catalytic sites facing one another across the cleft. The b and b' domains lie between the two catalytic domains, with the b domain displaced perpendicular to the plane of the molecule. The contact areas between the catalytic and non-catalytic domains (surface area ~200 $Å^2$) are small compared to the large contact area formed between the two non-catalytic domains (surface area ~700 $Å^2$). This provides a model whereby the two catalytic domains are flexible around a more rigid base formed by the b and b' domains (Tian, Xiang et al. 2006).

The catalytic a and a' domains contain hydrophobic patches surrounding the active sites, which appear to form internal hydrophobic surfaces lining the interior, suggesting sites for substrate binding. The yeast PDI structure also shows a hydrophobic pocket in the b' domain facing the catalytic sites (Tian, Xiang et al. 2006). These hydrophobic regions are illustrated in Figure 1.8.

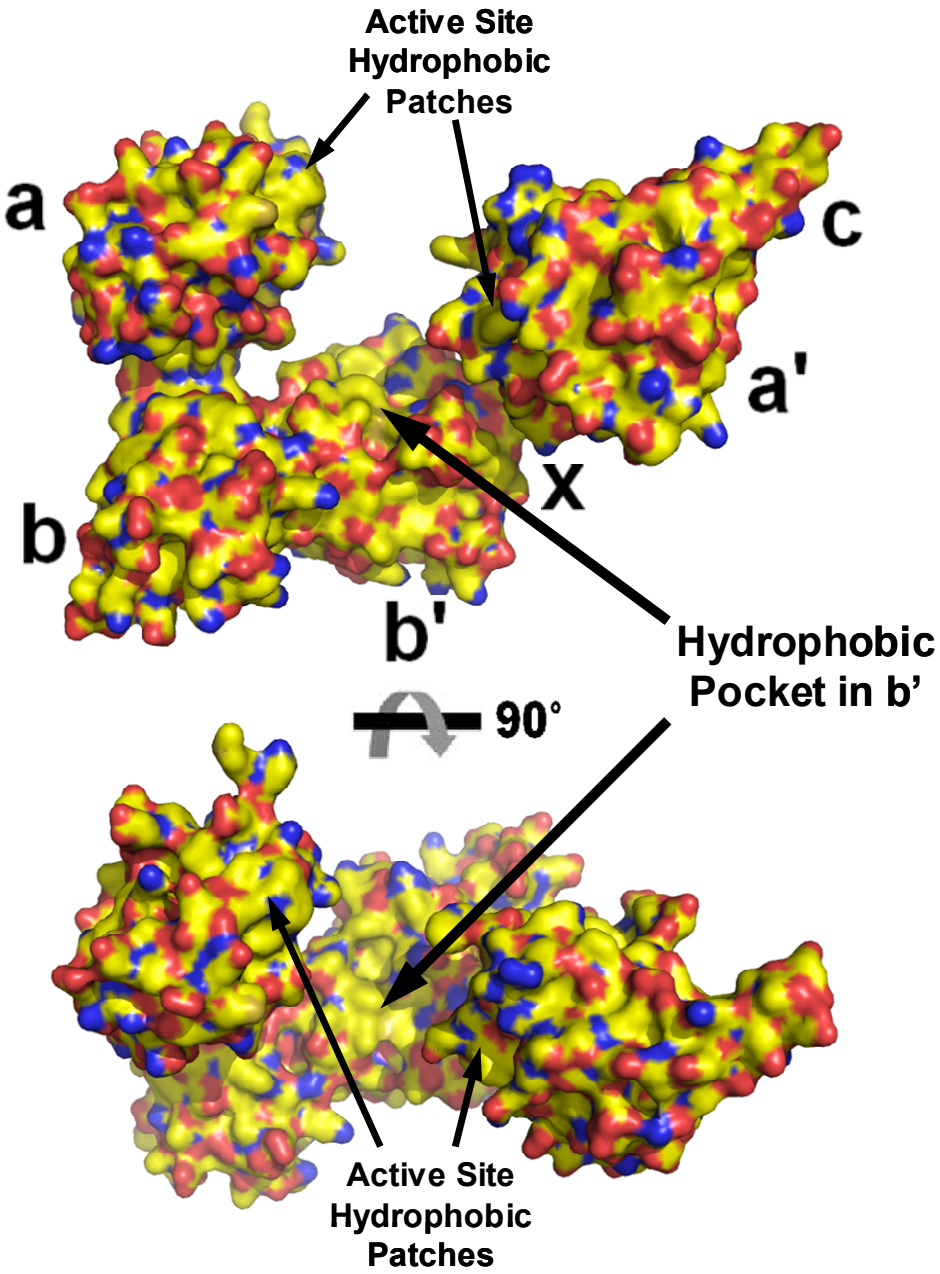


Figure 1.8: Surface diagram of Pdi1p, highlighting electrostatic potentials, from PDB 2B5E (Tian, Xiang et al. 2006). Red, acidic regions; blue, basic regions, yellow, non-polar regions. The top image has the same orientation as the ribbon diagram in Figure 1.7, with domains labelled; the bottom image has been rotated through 90°. A large non-polar region in the b' domain indicates the hydrophobic pocket thought to act as the main substrate binding region. Smaller non-polar regions in the a and a' domains indicate the hydrophobic patches surrounding the active sites.

1.6.1. The Catalytic Thioredoxin Domains of PDI

The catalytic a and a' domains share 36.8% identity with each other. They also have significant similarity with thioredoxin, 41% sequence similarity with the a domain and 32% with the a' domain. However, they have little similarity to the b and b' domains.

Using a redox assay with short peptide substrates, both the isolated a and a' domains were shown to have redox properties lying between a strong reductant (*e.g.* thioredoxin) and strong oxidant (*e.g.* DsbA) (Darby and Creighton 1995). These individual domains could act as either a reductant or oxidant, depending on the substrate and redox environment.

Both catalytic domains of human PDI have a CGHC active site motif. The active sites lie in such a way that the N-terminal active-site cysteine is more exposed, whereas the C-terminal active-site cysteine has limited solvent exposure.

During the process of disulfide bond formation, cysteines in the deprotonated state act as much better nucleophiles than when protonated. Hence, the pK_a values of cysteine residues are important in understanding catalysed disulfide bond formation (Karala, Lappi et al. 2010).

Studies have revealed that the pK_a of the N-terminal cysteine is in the range 4.4 to 6.7 (Hawkins and Freedman 1991; Kortemme, Darby et al. 1996; Ruddock, Hirst et al. 1996), considerably lower than the normal pK_a of a protein cysteine thiol, which averages around 8.3 (Bulaj, Kortemme et al. 1998). This results in the active thiolate state predominating at physiological pH. In contrast, the C-terminal active-site cysteine reportedly has a high pK_a (Hawkins and Freedman 1991; Nelson and Creighton 1994), resulting in the less active thiol state

predominating at physiological pH. Overall, this leads to less stable disulfide bonds being formed in PDI than in thioredoxin, which may explain the more oxidative nature of PDI. A recent study revealed that the side chains of Arg120, close to the a domain active site, and the corresponding Arg461, close to the a' domain active site, provide mechanisms to modulate the pK_a of the C-terminal active site cysteines (Karala, Lappi et al. 2010). It is believed that this has evolved to enable PDI to efficiently catalyse both oxidation and isomerisation reactions.

Activity observed using full length PDI on larger substrates, which require complex isomerisation, was not observed with individual a or a' domains, leading to the conclusion that other domains of PDI are required for these complex reactions (Darby and Creighton 1995).

1.6.2. The Non-Catalytic Thioredoxin Domains of PDI

Unlike the a and a' domains, the b and b' domains do not have the conserved thioredoxin–like active site residues. Therefore, these domains do not have oxidoreductase capability.

Structural alignment shows that the b and b' domains have only 19.3% identity to each other. It also reveals only 13% sequence identity between the a and b domains (Hatahet and Ruddock 2009). Despite showing only 27% sequence similarity to thioredoxin, the structure of the b domain indicates a thioredoxin-like fold (Kemmink, Dijkstra et al. 1999) and although the b' domain shows little sequence similarity with either the a or b domains, it too has a thioredoxin-like fold (Nguyen, Wallis et al. 2008).

The b' domain has been proposed as the main site of peptide binding for some time (Klappa, Ruddock et al. 1998). More recently, residues within the b' domain

of PDI were found to form a hydrophobic pocket involved in substrate binding (Byrne, Sidhu et al. 2009). This conserved binding pocket is also the primary region for substrate binding in other PDI family members (Ellgaard and Ruddock 2005) and is highlighted in the structure of yeast PDI (Pdi1p) in Figure 1.8. In contrast, the b domain has a much less direct role in substrate binding and is thought to have a mainly structural function, helping to stabilise and orientate the other domains.

Several studies have revealed that although the b' domain is essential for binding small peptides, other domains are required to bind larger substrates (Klappa, Ruddock et al. 1998; Wang, Wang et al. 2008).

1.6.3. Interdependence of PDI Thioredoxin Domains

In isolation, the catalytic a and a' domains are capable of catalysing simple thiol-disulfide exchange reactions, but show greatly reduced isomerase activity compared to the full length protein. For isomerisation reactions, it has been shown that the non-catalytic domains are also required. The b' domain in particular has been shown to be very important for isomerisation activity (Darby, Penka et al. 1998). The PDI fragment b'a'c was shown to be the minimum domain combination required that is able to catalyse disulfide bond rearrangement (Wang, Wang et al. 2008), illustrating the importance of the b' domain to the isomerase function of PDI. Furthermore, it was found that complex isomerisation reactions required all four PDI domains but did not require the C-terminal extension. This indicates a synergic relationship between all the thioredoxin domains of PDI in order to perform complex dithiol-disulfide oxidoreductase activity (Freedman, Klappa et al. 2002).

1.6.4. The x Region

The 19 residue x region of PDI provides a flexible link between the b' domain and the a' domain, making it the longest region between the four thioredoxin domains (Freedman, Gane et al. 1998). A recent study showed that the x region allowed PDI to adopt two distinct conformations (Byrne, Sidhu et al. 2009). In one conformation, known as the "capped" conformer, the x region covers the substrate binding site; in the other conformation, known as the "uncapped" conformer, the x region moves to expose the binding site (for illustration see Figure 6.3 in Section 6.2.4). A similar dynamic conformational exchange has been suggested for other PDI family members (Wang, Wang et al. 2008). Peptide ligands were shown to compete with x for binding, suggesting that x functions to gate access to the principal ligand-binding site of the b' domain (Byrne, Sidhu et al. 2009). The change in conformation also leads to a significant movement between the a' and b' domains, which could assist in inducing conformational changes in substrate proteins (Hatahet and Ruddock 2009).

1.6.5. The C-terminal Extension

The C-terminal extension of PDI forms an α-helix consisting of 29 residues, many of which are either glutamate or aspartate (Pihlajaniemi, Helaakoski et al. 1987). The functional importance of this acidic extension is not known, but it has been reported to have a low-affinity, high capacity calcium ion binding site (Macer and Koch 1988). The extension also contains the amino acid sequence KDEL, which is essential to retain PDI within the ER lumen. The isomerisation function of PDI does not require the C-terminal extension (Klappa, Ruddock et al. 1998). However, it is thought to stabilise the conformation of the a' domain through hydrophobic contacts (Tian, Xiang et al. 2006).

1.7. The PDI Family

For some time the PDI protein was considered to be the only enzymatic catalyst of thiol-disulfide exchange reactions in the ER. However, it is now known that PDI is just one protein in an ever-growing family. The PDI family, as it became known, is named after its archetypal and most abundant member, the PDI protein (Hatahet and Ruddock 2009). The PDI family of proteins are a subfamily of the thioredoxin superfamily.

PDI family members are believed to be the only enzymes which catalyse complex rate limiting isomerisation reactions (Weissman and Kim 1993). Proteins are classified as members of the family based on their structure, sequence similarity and ER localisation, rather than by physiological function. All contain at least one domain with a thioredoxin-like fold, which can be catalytic or non-catalytic in function. They also contain a cleavable N-terminal ER targeting signal sequence. Most PDI family members are ER localised with the aid of a C-terminal KDEL ER retention signal, or a similar variant of this motif. Since all PDI family members have distinct features and differing tissue distribution, they are all likely to have distinct physiological functions, although these may partially overlap (Hatahet and Ruddock 2009). This is supported by evidence that yeast PDI family members are not functionally interchangeable (Norgaard, Westphal et al. 2001).

Different organisms have different subfamilies of PDI. Whilst some clearly have functions similar to orthologs in other species, such as human PDI and yeast Pdi1p, many appear to be unique to a subset of organisms. Only 5 PDI family members have been identified in *S. cerevisiae*, whereas 20 have currently been defined in humans (Hatahet and Ruddock 2009), see Figure 1.9.

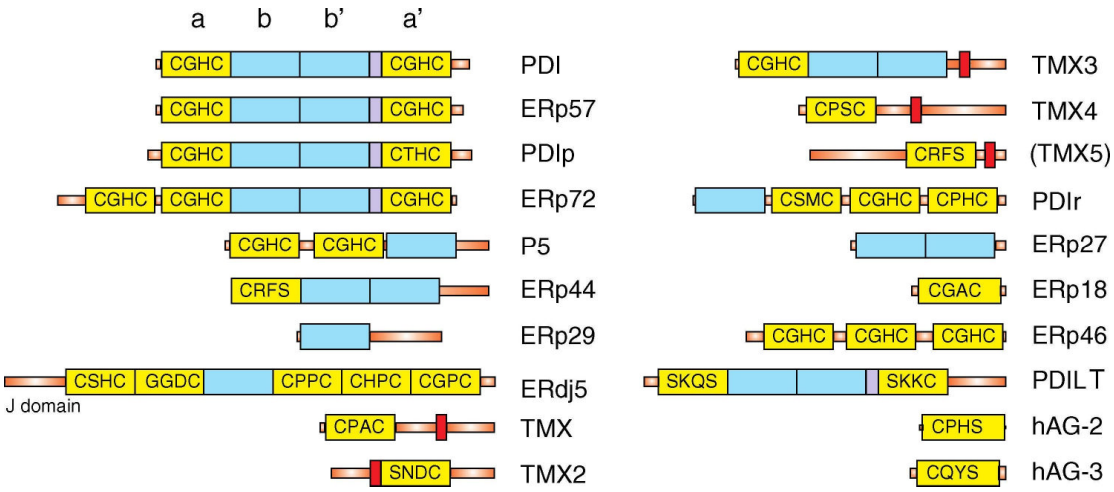


Figure 1.9: Illustration of human PDI family members. Yellow, catalytic domains likely to have a thioredoxin-like fold, with their active-site sequence inset; blue, non-catalytic domains likely to have a thioredoxin-like fold; red, transmembrane regions (Hatahet and Ruddock 2009).

Although detailed characterisation has yet to be undertaken on many PDI family members, it appears that only a small subset combine a generalised substrate binding domain with a catalytic domain, a combination that has been shown to be essential for efficient isomerisation by PDI (Darby, Penka et al. 1998). Indeed, many members of the PDI family (*e.g.* ERp27) do not have any active sites (Alanen, Williamson et al. 2006). Hence, perhaps the classification or nomenclature of this family of ER resident folding catalysts needs to be re-evaluated.

The only two human PDI family members that show the same domain architecture as PDI, *i.e.* abb'a'c, are ERp57 and PDIp. They all have two catalytically active a and a' domains with similar active site residues, see Figure 1.9. Other than PDI itself, ERp57 is the most characterised of all the PDI family members (Ferrari and Soling 1999).

Most PDI family proteins show a broad distribution of expression. However, two notable exceptions in humans are PDIp and PDILT. PDIp was reported to be expressed solely in acinar cells of the pancreas (DeSilva, Lu et al. 1996;

DeSilva, Notkins et al. 1997), while PDILT expression is highly specific to the testes after puberty and is the only PDI family member to show developmental control (van Lith, Hartigan et al. 2005; van Lith, Karala et al. 2007).

An interesting subset of the PDI family is the thioredoxin-related transmembrane proteins, which span the membrane of the ER. Five human transmembrane PDI family members have been proposed and their transmembrane regions identified, see Figure 1.9 (named with TMX prefix). All have type I transmembrane protein ER localisation signals (Appenzeller-Herzog and Ellgaard 2008), although only three of these have been published: TMX (Matsuo, Akiyama et al. 2001), TMX2 (Meng, Zhang et al. 2003) and TMX3 (Haugstetter, Blicher et al. 2005). Hence, there is a lot of scope to increase understanding of these ER resident, membrane bound proteins.

Another interesting human PDI family member is ERdj5. It is the only family member to contain a DnaJ domain (labelled J domain in Figure 1.9) (Cunnea, Miranda-Vizuete et al. 2003). This domain associates with BiP (an Hsp70 chaperone), thus implying a role for ERdj5 in ER-associated degradation (ERAD) (Hosoda, Kimata et al. 2003). Recent studies have revealed details of this role, which requires both the DnaJ domain and the redox activity of the catalytic domains for reductase activity of misfolded ER proteins (Dong, Bridges et al. 2008; Ushioda, Hoseki et al. 2008).

Until 2006, calsequestrin was the closest protein to PDI to have a full three dimensional structure (Wang, Trumble et al. 1998). It is a 40 kDa protein consisting of three thioredoxin-like domains and is involved in the regulation of Ca^{2+} ion channels in the sarcoplasmic reticulum of muscle cells (Kawasaki and Kasai 1994).

Although no structure for full length human PDI has been achieved, a full length structure of PDI from *Saccharomyces cerevisiae* (yeast) has been solved by X-

ray crystallography (Tian, Xiang et al. 2006). Like human PDI, yeast PDI (Pdi1p) incorporates all four thioredoxin domains, see Figure 1.10.

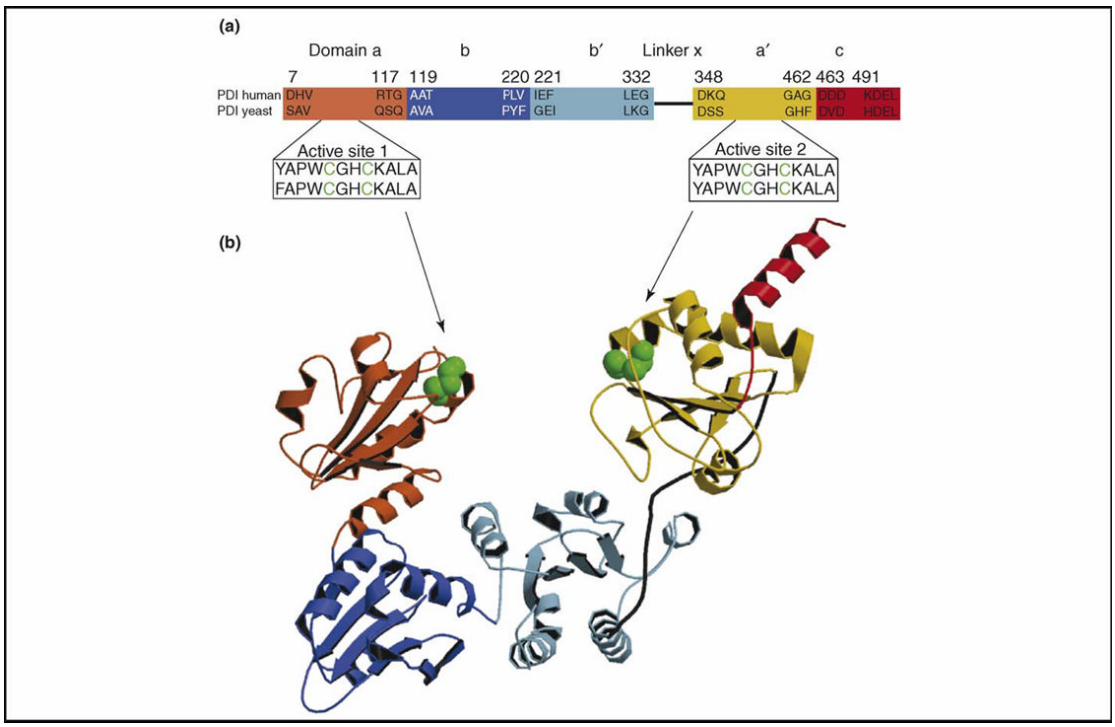


Figure 1.10: Alignment of human and yeast PDI domains. a) Domains showing boundary residues and residue numbers for mature human PDI, aligned with the homologous regions of yeast PDI. Residues flanking the active sites in the a and a' domains are highlighted. b) Crystal structure of yeast PDI (Pdi1p), PDB ID 2B5E. Active sites are highlighted green. Figure from (Gruber, Cemazar et al. 2006).

Sequence similarity between PDI and Pdi1p is high, so the structure has provided valuable insight into its human ortholog. The hydrophobic cleft is speculated to be large enough to accommodate a folded protein of approximately 100 residues, a figure which is important to consider when studying the interaction of PDI and substrate proteins (Tian, Xiang et al. 2006).

More recently, three crystal structures of human PDI family members have been solved. The small ERp29 protein (also known as ERp28) was resolved to 2.9 Å resolution (Barak, Neumann et al. 2009). ERp29 is an exceptional member of

the human PDI family in that as well as having a non-catalytic thioredoxin-like domain it also has a non-thioredoxin, entirely helical domain at its C-terminus, termed the D-domain (Liepinsh, Baryshev et al. 2001). It is very similar in structure and function to its ortholog Wind protein, a PDI-related protein found in *Drosophila melanogaster* (Lippert, Diao et al. 2007).

A crystal structure of ERp57 was recently solved in a complex with the substrate protein tapasin, to a 2.6 Å resolution (Dong, Wearsch et al. 2009). This is particularly interesting, given that ERp57 is a human PDI family member that shares the same domain architecture as PDI. Since its structure is solved with a bound natural substrate, this provides useful information about the nature of enzyme-substrate binding for all PDI family members. ERp57 is known to act in a complex to catalyse disulfide bond isomerisation specifically in N-glycosylated protein substrates (Russell, Ruddock et al. 2004). A knockout of ERp57 was found to be lethal in embryonic mice (Garbi, Tanaka et al. 2006). However, knockout of ERp57 in individual cells showed that very few proteins were affected (Solda, Garbi et al. 2006), indicating that ERp57 is involved in the folding of very few proteins or that other PDI family members are able to compensate for it.

A crystal structure of ERp44 was recently solved to 2.6 Å resolution (Wang, Wang et al. 2008). The protein consists of three thioredoxin-like domains (a, b and b'), with a C-terminal extension (Figure 1.9). The structure shows that the three thioredoxin domains form a clover shape, with the C-terminal extension occluding the substrate binding site and partially shielding the active site. It is thought that the C-terminus may act to mediate substrate access to the enzyme (Wang, Wang et al. 2008). ERp44 was initially identified as a ER-located binding partner for Ero1 (Anelli, Alessio et al. 2002). However, despite having a RDEL ER retention motif (Raykhel, Alanen et al. 2007), ERp44 is an unusual PDI family member in that it has since been reported to be co-localised to the ER, ER-Golgi intermediate compartment and *cis*-Golgi (Gilchrist, Au et al. 2006).

Interestingly, its only active site has an unusual motif, CRFS, since it lacks the C-terminal cysteine. It has been proposed that the absence of a second cysteine allows mixed disulfides between ERp44 and its substrate to persist for much longer, thus facilitating its proposed function to assist non-native protein transport from the ER (Anelli, Ceppi et al. 2007).

The only structural data available from full length human PDI comes from low resolution small angle X-ray scattering (SAXS) (Li, Hong et al. 2006). More recently, the same technique was used to characterise ERp72 (Kozlov, Maattanen et al. 2009).

Many PDI family proteins interact with their substrate primarily through a conserved binding pocket located in the non-catalytic b' domain (Ellgaard and Ruddock 2005). In some of these proteins, the binding site may have become specialised for the binding of particular substrates. For example, ERp57 is a PDI family member that forms a complex with either of the chaperone proteins calnexin or calreticulin (Russell, Ruddock et al. 2004). This complex then acts to catalyse disulfide bond isomerisation specifically in N-glycosylated protein substrates (Coe and Michalak 2010). The b domain seems to have a less crucial role in substrate binding and is thought to play a mainly structural role in multi-domain PDI proteins.

1.8. PDI Chaperone Function

Molecular chaperones are a functional class of proteins that assist the correct non-covalent assembly of other polypeptide-containing structures *in vivo*, but are not components of these final structures (Ellis 1993).

As well as its role in catalysis of thiol-disulfide exchange reactions, PDI can also exhibit molecular chaperone activity, *i.e.* it shows ability to prevent the

aggregation of substrate molecules, thus preventing them from following a non-productive folding pathway (Wang and Tsou 1993). In essence, PDI has two separate but related functions in the ER: an isomerase and a chaperone.

The chaperone function of PDI was first demonstrated using a protein that has no disulfide bonds, D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH), by showing that PDI prevents aggregation of this substrate (Cai, Wang et al. 1994). This anti-aggregation ability was also demonstrated using rhodanese as the substrate protein (Song and Wang 1995). Since these substrates contain no disulfide bonds, the chaperone activity was clearly independent of PDI's redox/isomerase function.

However, PDI chaperone activity has also been demonstrated during folding of proteins that contain disulfide bonds (Puig and Gilbert 1994; Yao, Zhou et al. 1997; Winter, Klappa et al. 2002). In studying of the refolding of proinsulin, the redox activity was found to be independent from the chaperone activity of PDI, since mutation of active site residues did not affect its chaperone activity (Winter, Klappa et al. 2002). A study of reduced and denatured acidic phospholipase A_2 ($APLA_2$) used alkylated PDI, which effectively prevents thiol-disulfide exchange, to show that the substrate could still become reactivated in the presence of both the alkylated PDI and a redox buffer (Yao, Zhou et al. 1997).

Some attempts have been made to identify the substrate binding site for PDI chaperone activity. It was found that PDI chaperone activity does not depend on the catalytic domain active sites, but it is inhibited by the presence of a peptide substrate for the b' domain (Quan, Fan et al. 1995). A similar study found that although the b' domain was necessary, other domains contributed to chaperone activity of PDI (Winter, Klappa et al. 2002). Likewise, domain combinations of ab and b'a'c of PDI were shown to lack chaperone function and this was speculated to result from the lack of protein substrate binding, thus more than just the b' domain was required (Sun, Dai et al. 2000).

When studying PDI-substrate interactions, it is important to consider the chaperone function of PDI, since this may influence the binding of substrates, especially those not considered candidates for thiol-disulfide exchange by PDI. However, since the action of PDI as a protein folding catalyst requires it to bind to non-native proteins, molecular chaperone like activity may be intrinsic to its function as a protein folding catalyst (Hatahet and Ruddock 2009). Certainly the literature seems to suggest that the regions required for substrate binding of chaperone activity correlate strongly with the regions required for substrate binding of protein thiol-disulfide exchange reactions.

1.9. PDI as a Subunit of Larger Protein Complexes

PDI is most commonly referred to in the context of its function in the oxidation, reduction and isomerisation of disulfide bonds. All of these disulfide related activities take place with PDI as a single peptide in a monomeric state. However, PDI is a multifunctional polypeptide that can also play a role within larger protein complexes. PDI serves as the β-subunit of both prolyl-4-hydroxylase, involved in the synthesis of collagens (Koivu, Myllyla et al. 1987), and microsomal triglyceride transfer protein, essential for the assembly of certain lipoproteins (Wetterau, Combs et al. 1991).

The main function of PDI in both of these protein complexes seems to be to maintain the highly insoluble α-subunits in a catalytically active, non-aggregated conformation (Lamberg, Jauhiainen et al. 1996; Kivirikko and Pihlajaniemi 1998). This function relates to the peptide binding and chaperone functions of PDI. It does not require the catalytic site cysteine residues (Vuori, Pihlajaniemi et al. 1992b; Lamberg, Jauhiainen et al. 1996).

1.9.1. Collagen Prolyl 4-Hydroxylase

Collagen prolyl 4-hydroxylase (P4-H) catalyses the formation of 4-hydroxyproline by hydroxylation of prolines (Kivirikko, Myllyla et al. 1989). This catalysis is essential in the formation of procollagen polypeptide chains into triple helical molecules.

P4-H has an approximate molecular weight of 240 kDa (Tuderman, Kuutti et al. 1975). It consists of two α-subunits of 63 kDa each (Pankalainen, Aro et al. 1970) and two β-subunits of 55 kDa each. The β-subunit was discovered to be PDI (Koivu, Myllyla et al. 1987). The α-subunit has been known to exist in two isomers (Annunen, Helaakoski et al. 1997), but more recently a third form was identified (van den Diepstraten, Papay et al. 2003). All α-subunits contained cysteine residues that form intrachain disulfide bonds required to maintain structure and enable tetramer assembly (Kivirikko, Myllyla et al. 1989).

Although PDI was identified as the β-subunit, it was found that PDI activity was not essential for tetramer assembly or P4-H enzyme activity (Vuori, Pihlajaniemi et al. 1992b). However, the solubility of the α-subunit is dependent upon its association with PDI (John, Grant et al. 1993). Hence, it is PDI's function as a molecular chaperone that ensures P4-H maintains its catalytic activity (Vuori, Pihlajaniemi et al. 1992a). This was supported by evidence that a PDI construct spanning from b' to a' domains (now known as b'xa') was sufficient for assembly of active P4-H (Pirneskoski, Ruddock et al. 2001). In addition, the C-terminal domain of PDI contains an ER retention signal, which is thought to be important in retaining P4-H within the ER (Vuori, Pihlajaniemi et al. 1992b).

1.9.2. Microsomal Triglyceride-transfer Protein

Microsomal triglyceride transfer protein (MTP) enhances the transport of triglycerides, cholesteryl esters and phospholipids between membrane vesicles via a shuttle mechanism (Atzel and Wetterau 1993; Wetterau, Lin et al. 1997). MTP is also involved in the assembly of some lipoproteins (Gordon and Jamil 2000; Hussain, Shi et al. 2003); the genetic disease abetalipoproteinemia, which results in a lack of apolipoprotein b (apoB) and triglycerides in blood plasma, arises from a mutation in the MTP gene (Wetterau, Aggerbeck et al. 1992). MTP is localised to the ER and Golgi apparatus (Swift, Zhu et al. 2003).

MTP is a 150 kDa protein (Wetterau, Aggerbeck et al. 1991), composed of two subunits, a 97 kDa α-subunit and a 55 kDa β-subunit (Wetterau and Zilversmit 1986). The β-subunit was identified as PDI (Wetterau, Combs et al. 1990). PDI is inactive with respect to its isomerase activity in the MTP complex (Wetterau, Combs et al. 1990; Wetterau, Combs et al. 1991). Moreover, the isomerase activity is not essential for its association with the larger α-subunit and for MTP activity, since PDI mutants lacking enzyme activity are fully functional in lipid transfer activity within the MTP complex (Lamberg, Jauhiainen et al. 1996). Lipoproteins were found to bind to MTP but not to PDI alone (Hussain, Shi et al. 2003). However, the PDI β-subunit of MTP is known to be vital to ensuring a soluble complex, since dissociated α-subunits formed inactive insoluble protein aggregates (Lamberg, Jauhiainen et al. 1996). Recently, a novel splice variant of MTP was identified that predominantly exists in adipocytes (fat cells) and are specifically localised to the Golgi apparatus of these cells (Mohler, Zhu et al. 2007).

1.10. Other Functions of PDI

It has been proposed that interaction with PDI is necessary for the export of misfolded secretory proteins via the ER-associated degradation (ERAD) pathway (Gillece, Luz et al. 1999), even for cysteine free proteins. This suggests that PDI, in its function as a chaperone, may act as a part of the ER quality control machinery, recognising terminally misfolded secretory proteins. This is supported by a recent study that suggests PDI is a component of the SPP-mediated ERAD machinery (Lee, Cho et al. 2010).

PDI has also been reported to be involved in many biological functions outside the ER, including cell adhesion (Turano, Coppari et al. 2002), although the mechanisms for PDI escaping the ER are unclear.

1.11. Other Processes Involved in Oxidative Folding

The PDI family are not the only ER resident proteins involved in the oxidation and isomerisation of disulfide bonds. Other substances are involved in either direct thiol-disulfide exchange of substrate proteins or in the transfer of oxidising equivalents to maintain the homeostasis of the ER. The manner in which these components interoperate is currently poorly understood.

1.11.1. Ero1

One enzyme known to be involved in the flow of oxidising equivalents for disulfide bond formation is Ero1 (endoplasmic reticulum oxidoreductin 1). Ero1 is a 65 kDa flavoenzyme localised to the ER membrane. Some studies suggest that PDI is an important intermediate in the transfer of disulfide bonds between Ero1 and substrate proteins (Frand and Kaiser 1999). Ero1 is thought to oxidise

reduced PDI via one internal pair of cysteines, with this pair in turn being re-oxidised by internal dithiol-disulfide exchange with another pair of cysteines within Ero1 (Frand, Cuozzo et al. 2000). A flavin adenine dinucleotide (FAD) cofactor is believed to be involved in the subsequent transfer of electrons, resulting in molecular oxygen being converted to hydrogen peroxide (Sevier and Kaiser 2008).

One model for disulfide formation is the flow of oxidising equivalents from molecular oxygen to Ero1, then to PDI before finally to the substrate protein. This model produces one molecule of hydrogen peroxide per disulfide bond made (Karala, Lappi et al. 2009). It was previously believed that the hydrogen peroxide produced would be highly toxic to the cell, but a recent study provided evidence that the hydrogen peroxide may be recycled and used as a substrate for disulfide bond formation. At physiological pH, the natively folded state is obtained more efficiently using peroxide than by the use of a glutathione redox buffer (Karala, Lappi et al. 2009).

1.11.2. Glutathione

Glutathione is a glutamate–cysteine–glycine tripeptide. It is the most abundant thiol-disulfide component of the eukaryotic cell and plays an active role in the process of protein disulfide bond formation (Bass, Ruddock et al. 2004). Glutathione is the principal redox buffer in the ER lumen, with the ratio of reduced and oxidised glutathione defining the redox potential. The ratio of reduced (GSH) to oxidised (GSSG) glutathione in the ER has been estimated to be between 1:1 and 3:1 (Bass, Ruddock et al. 2004). This is considerably more oxidising than the cytosolic ratio, estimated to be 30:1–100:1 (Hwang, Sinskey et al. 1992). GSH can be regenerated from GSSG by the enzyme glutathione reductase. A study of glutathione transport indicates there is a preferential transport of reduced glutathione into the ER lumen (Banhegyi, Lusini et al.

1999). It has also been demonstrated that Ero1 contributes to the oxidation of glutathione, suggesting a role for glutathione in buffering the ER against hyperoxidising conditions (Cuozzo and Kaiser 1999).

Glutathione can oxidise protein substrates directly and a recent study showed more than half of the microsomal glutathione was found to be present in mixed disulfides with protein (Bass, Ruddock et al. 2004). However, it has been shown that disulfide bond formation can still take place even in the absence of glutathione (Frand and Kaiser 1998), indicating the involvement of other chemicals and adding to the complexity of the system. Although cells lacking glutathione would be able to produce disulfide-bonded proteins via Ero1 and PDI alone, they would lack the ability to adapt rapidly to changes in disulfide-bond protein production and to other changes in redox conditions (Hatahet and Ruddock 2009).

There appears to be some functional cross-talk between PDI, Ero1 and glutathione pathways which is not yet fully understood (Frand, Cuozzo et al. 2000) and it seems likely that disulfide bonds are formed by complex interdependent pathways within the endoplasmic reticulum (Karala, Lappi et al. 2009). The strongly interconnected roles of PDI, Ero1, glutathione, hydrogen peroxide and molecular oxygen are currently areas of intense research.

1.12. Bovine Pancreatic Trypsin Inhibitor

Bovine pancreatic trypsin inhibitor (BPTI), also known as aprotinin, is a small monomeric single chain globular protein. The mature protein is 58 amino acids long with a molecular weight of 6.5 kDa. It was first isolated from bovine pancreas in 1936 (Kunitz and Northrop 1936). It is also known more generally as basic pancreatic trypsin inhibitor when from a non-bovine source, since it has 10 positively charged but only 4 negatively charged side chains, making it a strongly basic protein.

Physiologically, BPTI is known to inhibit digestive enzymes in the gut, known as serine proteases, including trypsin, chymotrypsin, plasmin and kallikrein (Mahdy and Webster 2004). It became the archetypal member of the Kunitz-type family of serine protease inhibitors. Kunitz-type inhibitors are protease inhibitors that work using a classic lock and key mechanism.

BPTI secondary structure is 18% helical and 25% β-sheet, with a large proportion retaining a coil structure (Kabsch and Sander 1983). The core of the protein contains an anti-parallel β-sheet, with two β-strands connected by a β-turn and a third, shorter β-strand in close proximity, together forming in a twisted hairpin. A surface exposed α-helix lies at the C-terminus with a short 3_{10} helix near the N-terminus, Figure 1.11A. The basic side chains of Lys15 and Arg16 on the exposed loop are thought to bind tightly to the active site of trypsin, inhibiting its enzymatic activity (Scheidig, Hynes et al. 1997).

An attractive feature of BPTI in the study of protein folding is that the mature protein has six cysteine residues, which are used to form three disulfide bonds in the native state, Figure 1.11B. The (30-51) disulfide binds a β-strand to the C-terminal α-helix, stabilising the protein core. The (5-55) disulfide binds the

helices at either end of the protein, while the (14-38) disulfide stabilises less structured regions of the protein.

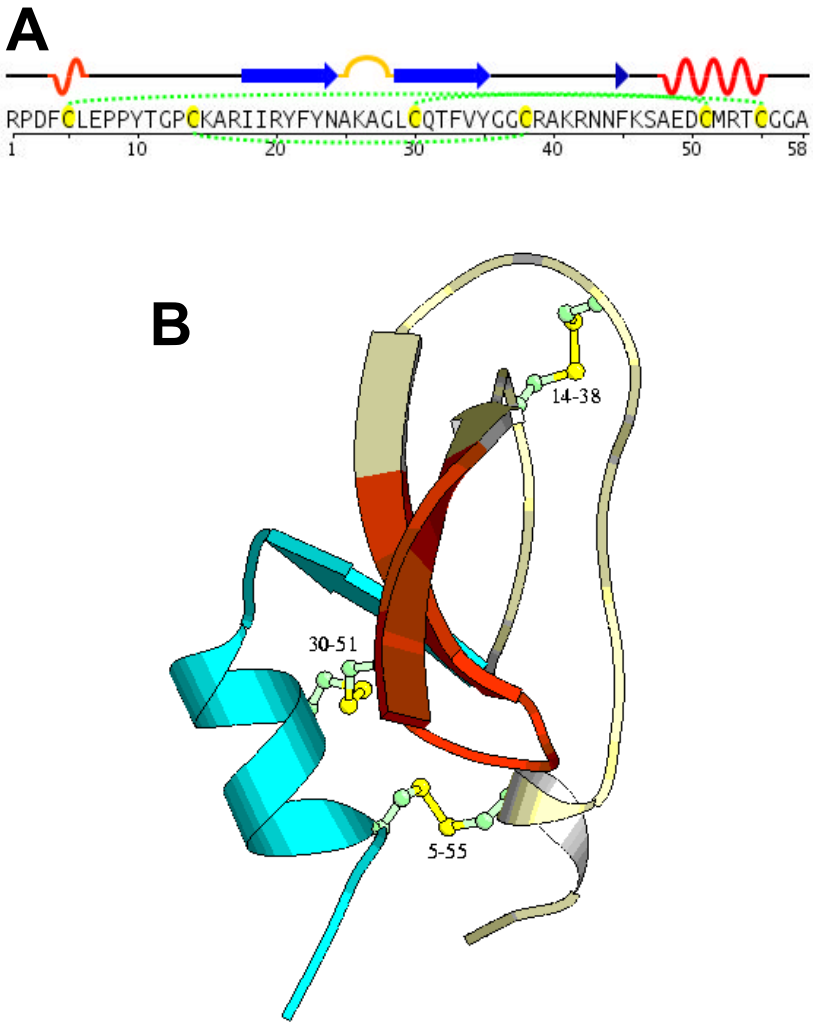


Figure 1.11: Mature BPTI. A) Primary and secondary structure elements. Red, helix; blue, β-strand; gold, β-turn; black, random coil. Cysteines are highlighted yellow in the primary sequence, with disulfide bonds indicated with green dotted lines. B) Ribbon diagram of native BPTI. Disulfide bonds are highlighted as ball and stick, labelled with the sequence numbers of the cysteines involved. Sequence numbers refer to mature BPTI. Based on PDB ID 1PIT.

Medicinally, BPTI has been used during surgical procedures to reduce bleeding, by slowing the breakdown of blood clots (fibrinolysis). It was shown to result in a reduction of 39% of requirements for blood transfusion during coronary artery

bypass graft surgery (Sedrakyan, Treasure et al. 2004). However, there have been concerns about its safety (Mahdy and Webster 2004), resulting in it being withdrawn from use. The debate about BPTI safety in surgical procedures is still ongoing (Dietrich 2009; Karkouti, Wijeysundera et al. 2010).

BPTI was one of the earliest protein crystal structures solved (Huber, Kukla et al. 1970), with a high resolution structure emerging a few years later (Deisenhofer and Steigemann 1975). Since this structure was available, BPTI became the first biological macromolecule to be modelled using molecular dynamics simulations (McCammon, Gelin et al. 1977).

BPTI was used as a model protein in the early days of the development of NMR to study biological molecules (Wuthrich and Wagner 1975; Brown, Demarco et al. 1978; Kumar, Ernst et al. 1980; Wuthrich, Wagner et al. 1980), becoming the first protein to have its NMR structure determined (Wagner and Wuthrich 1982; Havel and Wuthrich 1985). It was later used to develop NMR as a probe for investigating protein hydration (Otting, Liepinsh et al. 1991a) and to develop techniques for hydrogen deuterium exchange (Otting, Liepinsh et al. 1991b), resulting in a high quality NMR structure (Berndt, Guntert et al. 1992). It was further used as a model protein for identifying regions of internal mobility (Smith, Vanschaik et al. 1995).

In the context of this study, it is the folding pathway of BPTI that is of most interest. In particular, the mechanisms through which the protein forms its native disulfide bonds. During protein folding, the trapping of BPTI with a specific configuration of disulfide bonds is akin to trapping a clearly defined folding intermediate for this protein.

1.12.1. BPTI Folding Pathway

The folding pathway of BPTI, in terms of the combinations of disulfide bonds formed during folding, has been studied extensively. Early studies of folding intermediates were performed by Thomas Creighton in the 1970s (Creighton 1975), around the same time as Anfinsen and colleagues were developing some of the fundamental principles governing polypeptide self-assembly and protein folding (Anfinsen 1973; Anfinsen and Scheraga 1975). Meanwhile, a high resolution crystal structure of BPTI was revealed (Deisenhofer and Steigemann 1975). Different *in vitro* conditions effecting BPTI folding and unfolding were explored, including the use of denaturing agents (Creighton 1977; Creighton 1978; Creighton 1979; Creighton 1980) and the role that PDI may have in its folding was first explored (Creighton, Hillson et al. 1980). Throughout the 1980s, various biophysical techniques were used to characterise native BPTI and various conformations thought to be key to its folding pathway (Kosen, Creighton et al. 1980; Kosen, Creighton et al. 1981), showing that disulfide bonds can be used as effective probes for studying protein folding (Creighton 1986). This led to exploration of the folding pathway of other small proteins (Kaderbhai and Austen 1985; Pace and Creighton 1986). With well characterised intermediates now in place, attention turned to the kinetics of BPTI folding, in an attempt to establish the rate at which intermediates were formed (Goldenberg 1988). Other, less populated intermediate species of BPTI folding were also being investigated (Vanmierlo, Darby et al. 1991a). By this stage, a folding pathway for BPTI had been well established, with relative rates for each step in the process being estimated, Figure 1.12. It indicated that the optimum folding pathway necessitated the creation of non-native disulfide bonds and that isomerisation of these species was rate-limiting to establish the native structure.

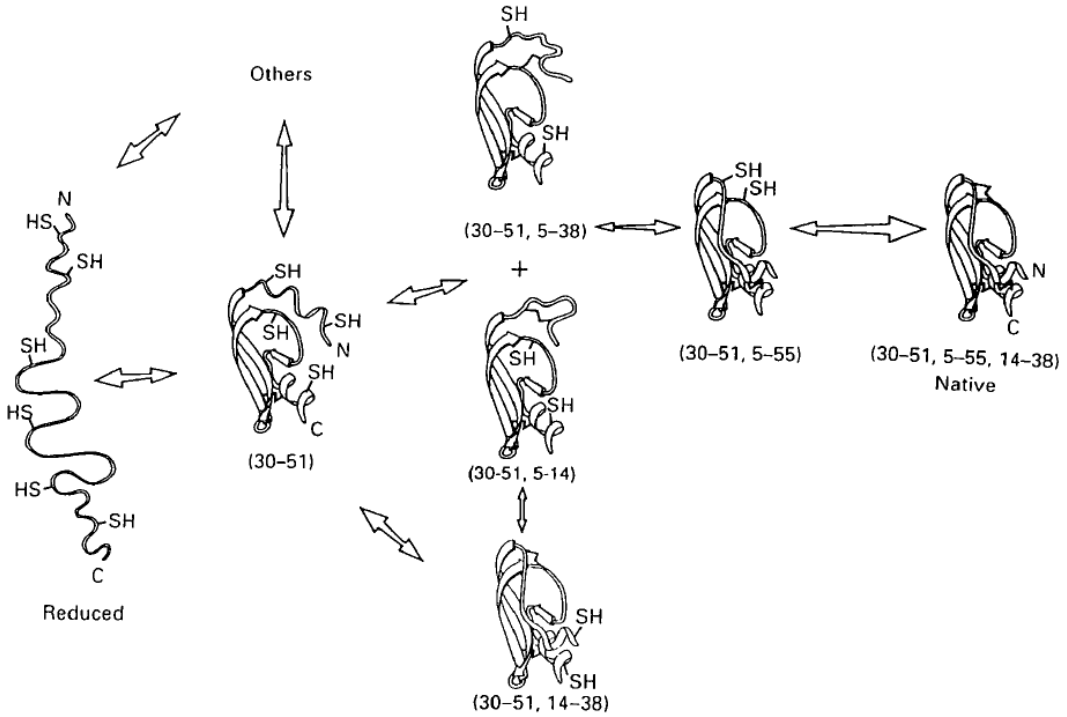


Figure 1.12: Model of the productive BPTI folding pathway, showing major disulfide intermediates in their approximate conformations (Darby, Morin et al. 1995). The non-productive quasi-native species (5-55, 14-38) has been omitted. The relative rate of each step is indicated by the thickness of the appropriate arrowhead; the wider the arrowhead, the greater the rate in that direction.

Meanwhile another group was independently studying intermediates of BPTI folding (Oas and Kim 1988). Controversially, they proposed an alternative folding pathway model, suggesting that only native disulfide bonds were predominant (Weissman and Kim 1991), Figure 1.13.

The discrepancies between the two models arise due to the manner in which kinetic folding intermediates are trapped by quenching. Whilst Creighton and colleagues used chemical quenchers, which may be inaccessible to buried thiols, Weissman and colleagues used an acid quenching technique. The controversy over the predominance of native disulfide species came to a head in a review article by Creighton and associated response by Weissman and Kim (Creighton 1992; Weissman and Kim 1992b). It should be noted that the

formation of non-native disulfides was never under dispute; rather, it was a debate about the dominant or transient nature of the non-native disulfide bonds in the BPTI folding pathway.

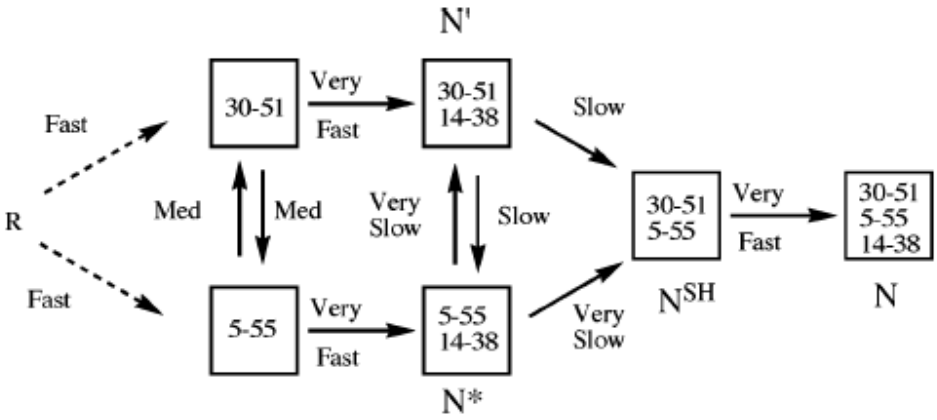


Figure 1.13: Alternative BPTI folding pathway model, emphasising the predominance of native disulfide bonds (Weissman and Kim 1991). The relative conversion rates between disulfide species are labelled. R is fully reduced BPTI; N is natively folded; N* is a kinetically trapped intermediate. Less populated species, including non-native intermediates, are not shown but are still required in transitions to the (30-51, 5-55) species.

Meanwhile, further characterisation of the conformational properties of intermediates involved the use of both one and two dimensional ^{1}H NMR (Vanmierlo, Darby et al. 1991a; Vanmierlo, Darby et al. 1991b; Darby, Vanmierlo et al. 1992; Staley and Kim 1992; Vanmierlo, Darby et al. 1992) and an X-ray crystal structure (Eigenbrot, Randal et al. 1992). These revealed that many intermediates have a native-like structure, which perhaps contributes to the difficulty of sequential native disulfide formation, due to steric hindrance of buried cysteines (Goldenberg 1992). The effects of other ER resident proteins on the folding of BPTI was investigated (Zapun, Creighton et al. 1992; Creighton, Bagley et al. 1993), showing the influence that PDI may have on the rate limiting steps of the process (see below). Further NMR studies focused on the influence of side chains on both the folding and structure of BPTI, revealing the importance of some aromatic rings (Kemmink and Creighton 1993; Kemmink,

Vanmierlo et al. 1993; Mendoza, Jarstfer et al. 1994; Yu, Weissman et al. 1995; Zhang and Goldenberg 1997).

Evidence of (30-51) as the most stable first disulfide bond in BPTI folding was revealed (Darby and Creighton 1993). The structure and dynamics of this critical one disulfide intermediate was further refined with the use of ^{1}H-^{15}N HSQC NMR of isotopically labelled intermediates (Vanmierlo, Darby et al. 1993). This, together with an independent NMR study of the same intermediate (Staley and Kim 1994), showed clear evidence that, despite only one disulfide being present, a large proportion of the protein showed native-like structure, with the first 14 N-terminal residues remaining unfolded. However, structure was only observed at low temperatures. Likewise, very similar structures were identified for both the (30-51, 5-14) and (30-51, 5-38) non-native intermediates using ^{1}H NMR (Vanmierlo, Kemmink et al. 1994). Interestingly, NMR of fully reduced and unfolded BPTI suggested that it retained some structure (Pan, Barbar et al. 1995). Structural studies of BPTI mutants using ^{13}C NMR were also performed (Hansen, Lauritzen et al. 1995).

Reviews comparing the folding pathways of several different proteins reveal common features in the formation of native disulfide bonds (Creighton, Zapun et al. 1995; Creighton 1995a; Creighton 1995b). A more recent comparison is available that relates these folding pathways to current understanding of the *in vivo* apparatus of prokaryotic and eukaryotic thiol-disulfide exchange (Mamathambika and Bardwell 2008).

1.13. PDI / BPTI Interaction Studies

Although a wide variety of peptides and proteins have been used as substrates for the study of PDI folding, BPTI has been a particularly popular candidate. As well as being a natural substrate for bovine PDI, it has the advantage of being a

small single domain protein containing three disulfide bonds. Furthermore, since the folding pathway of BPTI has been extensively characterised, the identification of soluble intermediates can be used to quantify the rate by which protein folding has been catalysed by PDI. It should be noted that bovine and human PDI are extremely similar (94.9% identity), as are BPTI and its human ortholog, so it is reasonable to consider BPTI a good model substrate for studies with human PDI.

PDI was first confirmed to catalyse the refolding of mature BPTI *in vitro* in 1980 (Creighton, Hillson et al. 1980). The study showed that the rate of those steps in the folding pathway that involved disulfide bond formation, reduction or isomerisation were significantly increased if they involved substantial conformational change of the substrate. Interestingly, even in this early study it was postulated that the intramolecular rearrangement of disulfides observed in BPTI folding alone would not be so prominent when PDI is present, since PDI bypasses the intramolecular disulfide rearrangement process by forming transient mixed disulfides with BPTI. Thus, intermediates involving non-native disulfides in the folding of BPTI may be replaced by complexes involving intermolecular disulfides with PDI (Creighton, Hillson et al. 1980). Another study showed that other components of the ER, including ATP and Ca^{2+}, did not add to or replace the increased rate of BPTI folding by PDI, indicating that such chemical additives were not involved in this catalytic process (Zapun, Creighton et al. 1992). Under certain conditions *in vitro*, one study found that the refolding of BPTI was reduced from several hours to about two minutes with the addition of PDI (Creighton, Bagley et al. 1993); a similar study reported a 3,500- to 6,000-fold increase in rates for the rate-limiting steps (Weissman and Kim 1993), which are both within the range for PDI catalysis now generally accepted (Freedman, Hirst et al. 1994). A much more recent study used BPTI to compare the catalytic activity of PDI with two other human PDI family proteins, ERp57 (referred to as ERp61) and ERp72 (Satoh, Shimada et al. 2005). They found that ERp57 and ERp72 accelerated the initial steps of BPTI folding more

efficiently than PDI. However, the rate limiting steps at the later stages of BPTI folding were catalysed less efficiently. Interestingly, they suggest that PDI and ERp72 can work cooperatively to catalyse BPTI folding more efficiently than with either enzyme alone (Satoh, Shimada et al. 2005). Another recent study used BPTI as a model substrate to show that the ability of PDI to act as a catalyst for disulfide bond formation or isomerisation was dependent on the pK_a of its active site cysteines, and that this pK_a could be altered by nearby arginine side chains (Karala, Lappi et al. 2010).

As well as using BPTI as a substrate for full length mature PDI, it has also been used in studies with various domain combinations of PDI, to identify how each domain contributes to the activities of PDI. One key study used BPTI to show that all the thioredoxin domains of PDI contributed to efficient thiol-disulfide exchange reactions of the enzyme, not just the catalytic domains (Darby, Penka et al. 1998).

As an alternative to human PDI, BPTI has also been used as a substrate in studies of yeast PDI. For example, the activity of various yeast PDI mutants was investigated by following the refolding of BPTI *in vitro* to identify the relative importance of the a and a' domains in the catalytic activity of the enzyme (Westphal, Darby et al. 1999).

There has been a lot of interest in the pathways involved in the transfer of oxidising equivalents to maintain homeostasis in the ER during protein folding. A recent study used BPTI as a substrate in investigating the role of hydrogen peroxide in this process (Karala, Lappi et al. 2009). They found that, at physiological pH, peroxide-mediated oxidation was much faster than that mediated by a glutathione redox buffer, showing that it can be used as a substrate for disulfide bond formation.

All of the above studies use the interaction of PDI with BPTI to elucidate either the effect of PDI on the kinetics of the BPTI folding pathway or to identify the functional significance of components of PDI to its catalytic activity. However, no previous studies have used BPTI or its folding intermediates to investigate the structural aspects of its binding to PDI.

1.14. Protein NMR Spectroscopy

1.14.1. Observing Nuclear Spin Frequencies

Spin is a fundamental property of atomic nuclei and elementary particles. Nuclear magnetic resonance (NMR) can be applied to atomic nuclei that have spin > 0, but it is most commonly applied to nuclei where spin = ½. The most common nuclei in biological molecules are ^{1}H, ^{12}C, ^{14}N and ^{16}O. Of these, however, only ^{1}H has a spin = ½. In order to observe other atoms, labelling with an isotope with spin = ½ is usually required. For studying recombinant proteins, this is usually achieved by growing cells in a minimal medium using isotopically labelled substances, *e.g.* ^{15}N labelled ammonium sulphate.

NMR active nuclei can be aligned to an external magnetic field. They precess with angular momentum around the external field at close to an angular frequency known as the Larmor frequency, which is unique for each isotope. Larmor frequencies are within the radio frequencies of the electromagnetic spectrum. In NMR, the different Larmor frequencies can be exploited to obtain different types of information about the biological molecule. A short radio frequency pulse is applied close to the Larmor frequency of the nuclear spins of interest, usually at 90^{o} to the external field, producing a transverse magnetisation of spins. A transient NMR signal, known as free induction decay (FID), can then be detected as the transverse spins relax towards equilibrium.

Although the resonance frequency will be close to the Larmor frequency of an isotope of interest, the precise frequency observed shifts depending on the chemical environment of the nucleus. This difference is known as a chemical shift and is influenced mainly by the electronic environment of the nucleus.

1.14.2. Converting FID to Frequency Signal

The free induction decay (FID) signal is very difficult to interpret in its raw state. Instead, the signal is Fourier transformed from the time domain signal of the FID to a frequency domain signal. Since the chemical shifts are very small compared to the resonance frequencies of the observed nuclei, the signal is normalised so that the differences in resonance frequencies between chemical environments, the chemical shifts, are highlighted. This is done by subtracting the frequency domain from a reference molecule, typically TMS (tetramethylsilane) or DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). Chemical shifts are typically in the order of a millionth of the resonance frequency of the observed nuclei, so are reported in parts per million (ppm) (Pascal 2008). This normalisation also means that chemical shift is independent of magnetic field (a chemical shift of 10 ppm is equivalent to 4 kHz on a 400 MHz spectrometer but 7 kHz on a 700 MHz spectrometer). In 2D NMR spectra, chemical shifts are typically displayed in ppm frequency with axes labelled ω_1 and ω_2.

1.14.3. Protein Structure from Frequency Signals

When using NMR to determine protein structure, one aspect of the chemical environment of nuclei that is utilised is their interactions with neighbouring nuclei. Each nucleus acts as a magnetic dipole and interactions between nuclei are

referred to as dipole-dipole interactions. These interactions can be transmitted either directly through space or indirectly through bonding electrons.

The atomic structure of small molecules can be identified using simple one dimensional NMR experiments. However, for large macromolecular structures such as proteins this produces many overlapping peaks that are impossible to resolve. Increasing dimensionality reduces the number of overlapping peaks. Multidimensional NMR relies on correlating multiple resonances. These are identified by cross-peaks in the NMR spectrum and provide information about their interaction. The cross-peaks between different ^{1}H atoms of amino acids tend to lie in different regions of certain NMR spectra, Figure 1.14. Using a variety of multidimensional NMR combining both through bond and through space experiments, the resonances associated with each amino acid in a protein sequence can be assigned.

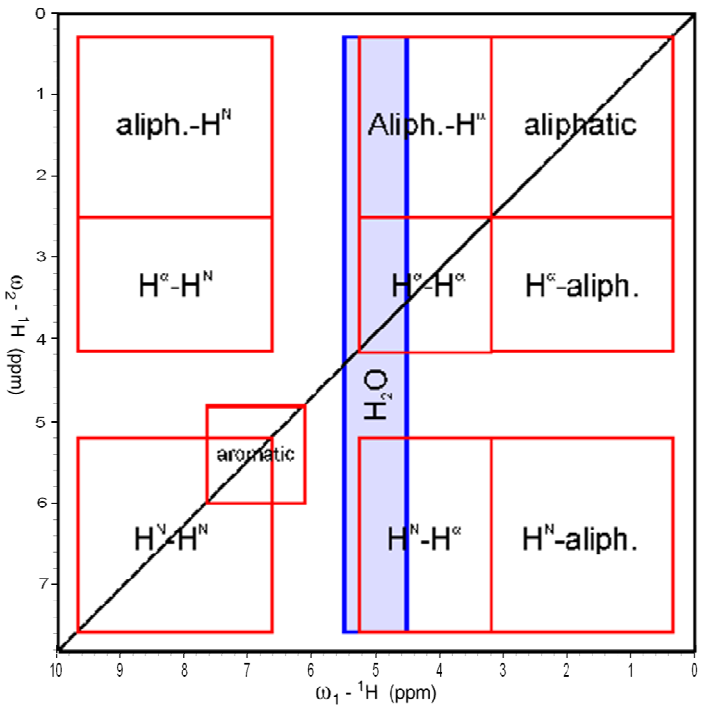


Figure 1.14: Spectral regions for two dimensional TOCSY and NOESY experiments, modified from (Birkbeck 2009).

1.14.4. Through Bond NMR Observations

Scalar coupling (J-coupling) is the coupling between two nuclear spins due to the influence of bonding electrons on the magnetic field running between the two nuclei. Hence it is a through-bond magnetic interaction between nuclei via paired electrons. NMR techniques such as COSY (correlation spectroscopy) use the resulting correlations to detect atoms that are linked by up to three chemical bonds. One benefit of this technique is that even when relatively insensitive heteronuclei are present bonding between atoms can still be inferred by transferring magnetisation via nearby sensitive ^{1}H nuclei.

A related technique, known as TOCSY (total correlation spectroscopy), magnetisation is dispersed over successive scalar couplings. This dispersion allows detection of nuclei over more than three bonds. Hence, it can show correlations between all protons within any amino acid residue, referred to as a spin system. A characteristic pattern of signals often enables the type of amino acid to be identified. However, some amino acids have almost identical spin systems.

1.14.5. Through Space NMR Observations

Through space interactions occur due to dipolar interactions between nuclei. Dipolar coupling is the direct interaction between two magnetic dipoles. It depends on the inverse cube of the internuclear distance. Since dipolar couplings are anisotropic interactions, in solution they average to zero as a result of rotational diffusion. However, the effect also causes the fluctuating local magnetic fields experienced by both nuclei to become correlated as the spins relax, a phenomenon known as cross relaxation. The resulting transfer of spin polarisations results in a measurable nuclear Overhauser effect (NOE). The NMR NOESY (Nuclear Overhauser Effect Spectroscopy) technique can be used

to measure NOEs in proteins. The cross-peaks in a resulting spectrum can be used to connect resonances from spins that are spatially close. NOESY spectra from large biomolecules can often be assigned using sequential walking (sequential assignment). Correlations may also be observed between nuclei that are close in three dimensional space but far apart in the primary sequence. NOEs are not normally observed for protons separated by more than 6 Å.

1.14.6. Heteronuclear Single Quantum Correlation

Heteronuclear Single Quantum Correlation (HSQC) is a two dimensional NMR experiment that detects correlations between ^{1}H nuclei and those of other NMR active nuclei, typically ^{15}N (Maudsley and Ernst 1977). The technique detects the transfer of magnetisation from each amide proton to its attached ^{15}N nucleus and back (Pascal 2008). The decoupled HSQC spectrum produces one peak for each ^{1}H-^{15}N pair. The resulting spectrum has a proton axis and a ^{15}N axis, with a single peak for each amino acid (excluding prolines). Additional peaks will be present from the side chains of amino acids that contain amine groups.

1.14.7. Hydrogen Deuterium Exchange

Hydrogen deuterium (H/D) exchange is a chemical reaction in which a covalently bonded hydrogen atom is replaced by a deuterium atom. Usually the protons being examined are the amides in the backbone of a protein. The protein sample is solubilised in a deuterium oxide solution. Using NMR observations, the method gives information about the solvent accessibility of various parts of the molecule, and thus the three dimensional surface of the protein. In particular, H/D exchange can provide insight into protein dynamics caused by structural fluctuations in the protein (Englander and Kallenbach 1983). HSQC spectra are recorded at a series of time points while the hydrogen is exchanging with the deuterium in the solvent. Since only the proton signal is

detected, the signal will decay as protons are exchanged for deuterium atoms. An amide buried in the hydrophobic core of a protein will exchange slowly, while an amide on the surface will exchange rapidly. Freely exposed amide hydrogens have been reported to exchange in about 1 s at pH 7 and 0°C. However, the protecting structure surrounding most backbone amides slow these rates down and spread them over many orders of magnitude, placing them on a convenient laboratory time scale (Krishna, Hoang et al. 2004). The precise rates of these exchanges can be informative about the dynamics of different parts of the protein. This can be particularly useful in studies of protein folding, since the extent to which different intermediates have folded can be elucidated from how solvent exposed they are (Wuthrich, Wagner et al. 1980). Furthermore, H/D exchange can aid in identifying which components of each intermediate protein have folded, thus identify detailed steps involved in establishing the native structure (Englander and Kallenbach 1983). The kinetics of the exchange process are dependent on environmental conditions such as pH and temperature (Woodward, Simon et al. 1982). A ten-fold change in exchange rate can be obtained by a change of one pH unit, or about 22°C in temperature (Krishna, Hoang et al. 2004).

1.14.8. Chemical Shift Mapping

Chemical shift mapping is a technique used to compare subtle differences between two NMR spectra. It is often used to compare a reference spectrum of a sample in controlled conditions with a spectrum where the sample conditions have been altered. For example, a chemical shift map can be produced to highlight the changes that occur in the HSQC spectrum of an enzyme after the addition of a substrate (Rajagopal, Waygood et al. 1997). The map identifies chemical shift perturbations for each assigned residue, thus indicating the extent to which they have been affected by the environmental change. In the case of substrate binding, the largest perturbations would be expected to occur at the

binding interface. When combined with other NMR techniques, detailed structural information about protein-protein interactions can be obtained (Schumann, Riepl et al. 2007).

Chemical shift mapping can be performed even when only one of the spectra being compared have been assigned. This is referred to as minimal shift mapping and involves mapping from each assigned peak in the source spectrum to the nearest peak in the target spectrum. It is convenient when comparing one reference spectrum with multiple spectra where the environment of the sample has been altered, *e.g.* comparing the HSQC spectrum of an enzyme alone with the HSQC spectra of the enzyme in the presence of different substrates.

1.14.9. Use of NMR Spectroscopy to Investigate Binding to PDI

The two most powerful biophysical techniques for investigating protein structures are X-ray crystallography and NMR spectroscopy. They are both regularly used to provide high resolution structures of biological molecules at atomic resolution.

Since NMR uses molecules in solution, it is not limited to those proteins that crystallise. Many attempts have been made to crystallise human PDI, but these have been unsuccessful, mainly due to its inherent flexibility, which makes crystal formation difficult. Even when successful, the crystallisation process often causes packing effects, whereas solution state NMR can observe a protein structure closer its native structure in the cell. NMR has a further advantage of being able to study dynamic properties such as protein flexibility and ligand binding (Baldwin and Kay 2009). This includes observations on a wide variety of timescales, including conformational change, chemical exchange and internal mobility (Mittermaier and Kay 2006). One major drawback of NMR

is the size limit of molecules that can be studied. In the case of PDI, this means that experiments are restricted to truncated versions of the enzyme; X-ray crystallography has no such limitation. Another major drawback of NMR is that data analysis tends to take much longer compared to crystallography data.

1.15. Studying Protein Interactions using Biosensors

A biosensor is an analytical device for the detection of a biological analyte that combines a biological component with a physicochemical transducer. Interaction between the target analyte (*e.g.* a substrate protein) and the biological material (*e.g.* an enzyme) produces a physicochemical change detected by the transducer, usually resulting in a signal proportional to the concentration of interacting analyte (Schasfoort and Tudos 2008). The most prominent example is the glucose biosensor, which uses enzyme-electrodes for the rapid self-diagnosis of blood glucose levels by diabetes sufferers (Barredo 2005).

1.15.1. Surface Plasmon Resonance as a Biosensor

One form of real-time biosensor is based on the phenomenon of surface plasmon resonance (SPR). It enables accurate analysis of protein binding specificity, kinetics and affinity. SPR biosensors work by immobilising one interacting partner to a surface before passing a liquid sample containing another interaction partner over the surface. Where interaction occurs, binding of molecules to the surface generates a response which is proportional to the bound mass. Sensors use optical techniques to detect a change in refractive index near the surface due to the interaction of the biological partners (Clow, Fraser et al. 2008). The application of SPR based sensors to monitoring biological interactions was first demonstrated in 1983 (Liedberg, Nylander et al. 1983). Recent publications offer good introductions into the theoretical and

practical use of SPR as a biosensor (Schasfoort and Tudos 2008; de Mol and Fischer 2010).

SPR biosensors can be used to determine a range of interaction characteristics, including the kinetics and affinity of an interaction and the concentration of interacting molecules present in a sample. Affinity analysis can provide quantitative data about the strength of binding between two interacting molecules in terms of a dissociation constant. Modern SPR biosensors can detect affinity binding events in the μM to sub-nM range (de Mol and Fischer 2010). The affinity of an interaction is calculated based on either rates of association and dissociation or the equilibrium level of binding. Kinetic analysis of binding can establish the rate of association or dissociation of an interaction. k_{on} and k_{off} rates can be ascertained using various concentrations of an analyte, after careful analysis of the time curves produced using various interaction models.

Although many applications use purified proteins, binding interactions with crude mixtures including nucleic acids, lipids, carbohydrates or even whole cells can be measured as they pass over the sensor. In such circumstances, the concentration of the actively binding molecule can be estimated from the response levels obtained.

One of the key advantages of SPR biosensors is that there is no need to label molecules with fluorescent or radioactive tags, thus avoiding the possibility that labels may change the native conformation of the molecule and affect its activity. Compared to other biophysical techniques, SPR biosensors require relatively small volumes of sample. Furthermore, the analyte is not destroyed during analysis and is often recovered for identification by mass spectrometry (Natsume, Nakayama et al. 2000; Williams and Addona 2000).

Despite all of the potential advantages of using SPR biosensors, some inherent disadvantages of this technology should be highlighted. The most obvious is the requirement of one of the interacting molecules to be immobilised to a surface, which could potentially cause steric hindrances that effect the interaction of binding partners. A variety of immobilisation strategies can be explored to minimise these effects. Another issue is that molecules can bind non-specifically to the surface, although appropriate controls should account for this. Mass transport effects can occur due to multiple binding of a single analyte molecule to many immobilised molecules; using appropriate flow rates for the analyte should avoid this effect. In order to get reproducible results, it is essential that the analyte can be removed from the surface without affecting the immobilised protein, a process known as regeneration. For some interactions, this regeneration can be very problematic.

1.15.2. Components of SPR Biosensors

All SPR instruments comprise three essential units integrated into one system: an optical unit, a liquid handling unit and a sensor surface. All configurations directly measure refractive index changes at the sensor surface in real-time (Schasfoort and Tudos 2008). The changes are usually due to binding events and are proportional to the change in mass at the chip surface.

Plasmons result from the oscillation of ionised gas and play a large role in the optical properties of metals. They are often confined to surfaces (surface plasmons) and interact strongly with light. Surface plasmons can be thought of as charged density waves propagating along the interface of a metal and an insulating media (de Mol and Fischer 2010). Generally, gold is used as the metal surface in sensor chips, since it combines favourable SPR characteristics with inertness in biomolecular interactions (Biacore 2003).

In the presence of an appropriate medium (*e.g.* a prism), polarised light striking a surface plasmon at a particular angle of incidence will become totally internally reflected. Despite being internally reflected, some of the photons of polarised light interact (resonate) with the free electrons of the metal surface, penetrating a short distance into a medium of a lower refractive index, thereby creating a type of wave known as an evanescent wave. This is a standing wave exhibiting exponential decay with distance. Resonance energy is transferred between the evanescent wave and the surface plasmons, causing a reduction in light intensity (Schasfoort and Tudos 2008). This allows the phenomenon of surface plasmon resonance to be observed.

In a typical SPR biosensor setup, a sensor chip is coated with a film of gold. During operation, one surface contains a prism and provides a surface for plasmon resonance to take place. On the other side of the gold film, where an evanescent wave will penetrate, molecules (*e.g.* proteins) are immobilised and an analyte is allowed to flow past in a controlled manner, Figure 1.15.

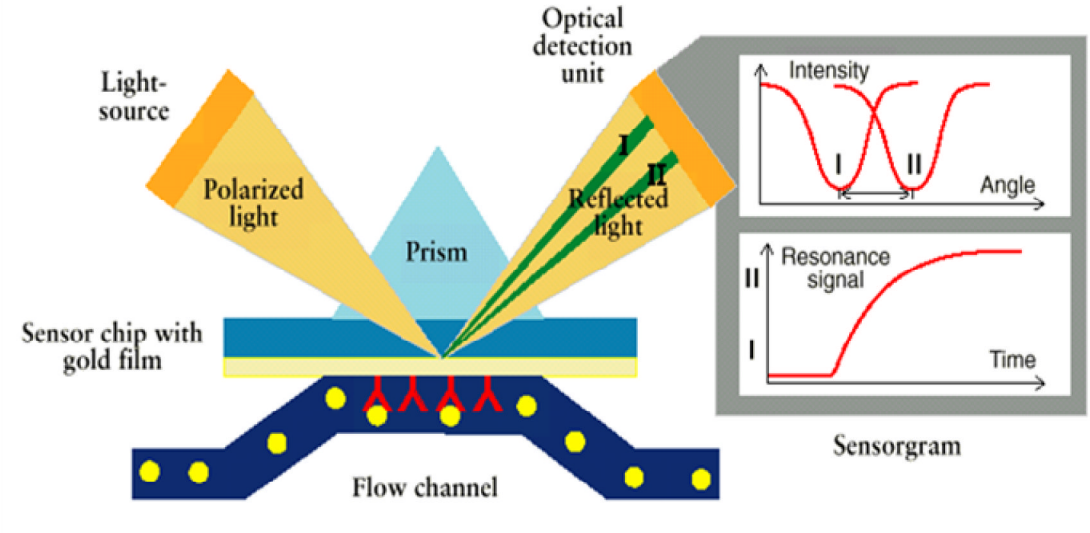


Figure 1.15: Schematic illustrating the principles of detection of a surface plasmon resonance signal (Paulo 2006).

The angle at which the maximum loss of the reflected light intensity occurs (the resonance angle) is dependent on the refractive indices of the media at both sides of the gold film. The refractive index of the prism never changes. However, on the other side the refractive index in the vicinity of the metal surface will change when accumulated mass (*e.g.* the analyte) bind to it (Schasfoort and Tudos 2008). Only refractive index changes very close to the surface affect the SPR signal, since the effective penetration depth of the evanescent wave is typically around 150 nm (Biacore 2003).

A linear relationship is found between resonance energy and mass concentration of biological molecules, such as proteins. Therefore, the SPR signal, expressed in resonance units, is effectively a measure of mass concentration at the sensor chip surface. This means that association and dissociation between the immobilised molecules and the analyte can be observed. Hence, rate constants and equilibrium constants can be calculated (de Mol and Fischer 2010).

1.15.3. The SPR Sensor Surface

In most commercial products, the sensor surface of a SPR biosensor is provided by a chip consisting of a glass surface coated with a thin layer of gold, onto which a range of functional groups can be attached. Thus, a sensor chip provides the physical conditions necessary to generate an SPR signal, with interactions being studied on the chip surface.

For interactions to take place, the sensor chip surface needs to be modified to immobilise molecules such that they can selectively capture the target analytes, but should not be prone to adsorbing other components present in the sample or buffer media (Schasfoort and Tudos 2008).

The most common sensor chips consist of a carboxymethylated dextran matrix covalently attached to the gold surface (Clow, Fraser et al. 2008). The molecule targeted for immobilisation needs to be chemically bonded to the dextran matrix so to provide a reproducible chip surface after extensive washing. The immobilisation must be done so that the analyte can access and bind to the immobilised molecule efficiently (Clow, Fraser et al. 2008).

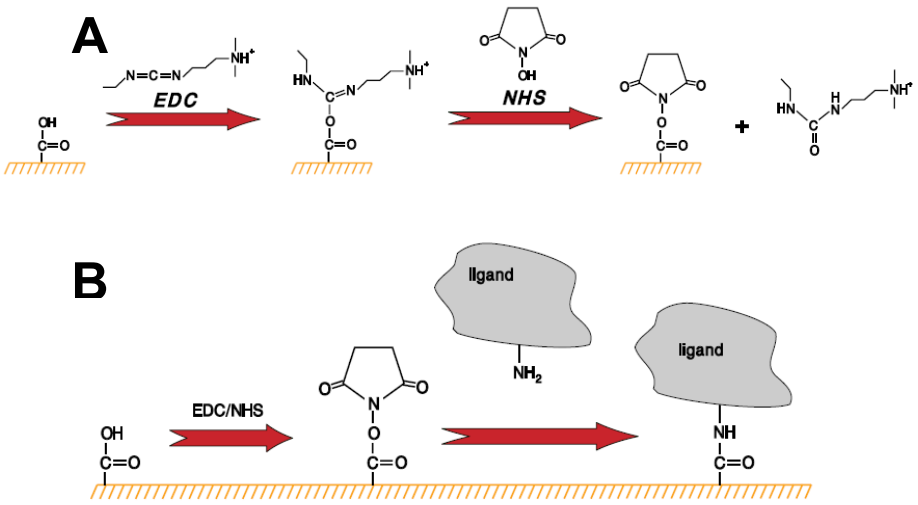


Figure 1.16: Immobilisation of protein (ligand) onto carboxymethylated sensor chip surface. A) Activation of chip surface using EDC and NHS; B) Ligand immobilisation onto EDC/NHS activated sensor chip via amine coupling (Biacore 2003).

The established procedure for immobilisation of a protein links amine groups in the protein to derivatised carboxyl groups, provided by the dextran, on the sensor chip surface (Johnsson, Lofas et al. 1991). This can be achieved in a two step process, Figure 1.16. First, the carboxyl activating agent EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) is bond to the carboxyl groups of dextran. A second activating agent, NHS (N-hydroxysuccinimide), is then bound to the EDC, Figure 1.16A. This provides a succinimide ester that can react with the amine groups of protein ligands in a process known as amine coupling, Figure 1.16B. The optimum pH for amine coupling can vary between proteins, so a process of pH scouting is often performed prior to immobilisation to identify

the most desirable pH. The amount of protein that has been immobilised is identified by the change in response units before and after immobilisation.

Amine coupling is only one mechanism of immobilising proteins. One potential issue is that amine coupling could bind proteins in a way that would affect the binding of substrates. A more uniform alignment to a chip surface could be achieved by site specific immobilisation via an affinity tag, *e.g.* using the His-tag of appropriate fusion proteins (Schmid, Keller et al. 1997). However, a disadvantage of the His-tag system is that metal dependent non-specific protein adsorption can occur (Clow, Fraser et al. 2008).

1.15.4. Detecting Binding of Analyte

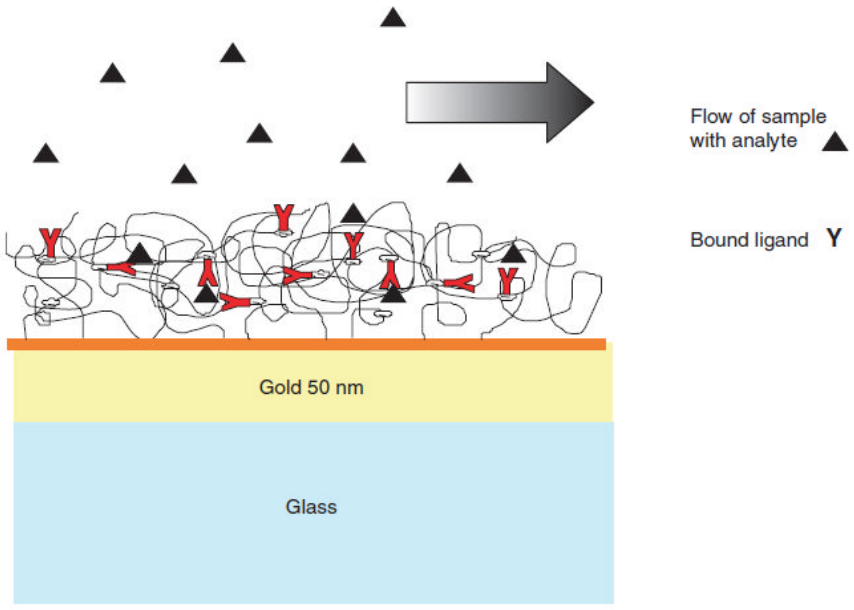


Figure 1.17: Schematic of the direct detection of analyte flowing over the sensor chip surface with bound protein (ligand). The analyte is captured by the ligand immobilised on the sensor surface. Accumulation of the analyte results in a refractive index change in the evanescent field shifting the SPR angle (Schasfoort and Tudos 2008).

When molecules in the analyte solution bind to the immobilised molecule on the surface the mass increases; when they dissociate the mass falls, Figure 1.17. When binding occurs biomolecules accumulate within the aqueous layer, changing the sensor surface, thus changing the refractive index (Schasfoort and Tudos 2008). This allows continuous, real-time monitoring of the association and dissociation of the interacting molecules (Paulo 2006), which is shown as a plot of resonance signal against time, known as a sensorgram, Figure 1.18. Changes in the amount of analyte bound can be detected down to concentrations corresponding to the picomolar to nanomolar range of analyte in the sample solution (Biacore 2003). After a set time of analyte associating with the immobilised protein, a buffer without analyte is used to dissociate bound analyte. If tightly bound, some analyte may remain bound to the immobilised protein. A regeneration buffer (*e.g.* 50 mM NaOH) is used to dissociate any remaining analyte, Figure 1.18. It is essential that the analyte can be removed from the sensor chip surface without affecting the immobilised protein. This makes binding analysis reproducible.

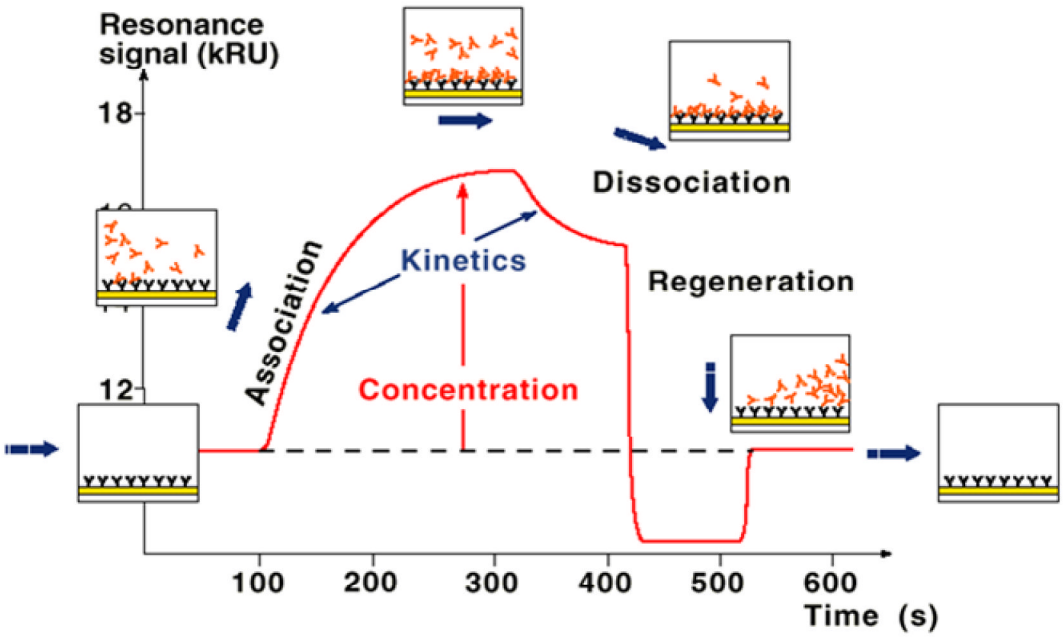


Figure 1.18: Typical sensorgram showing the association, dissociation and regeneration phases of analyte binding to an immobilised ligand (Paulo 2006).

1.16. Aims of this Study

This chapter has introduced general concepts of protein folding and mechanisms involved in protein folding in the endoplasmic reticulum. Disulfide bonds play a key role in the folding of many proteins, with oxidoreductase enzymes involved in catalysing their oxidation, reduction and isomerisation. Protein disulfide isomerase (PDI) is the most abundant and well studied of these enzymes, but represents only one member of the PDI family of proteins.

This study aims to learn more about the mechanisms through which PDI binds to substrate proteins by studying its interaction with a classic model protein, bovine pancreatic trypsin inhibitor (BPTI). The BPTI folding pathway has been studied extensively, with disulfide bonded intermediate states well characterised. The pathway includes intermediates with non-native as well as native disulfide bonds, making isomerisation necessary. This makes BPTI a good candidate for studying PDI activity, since the enzyme strongly catalyses these rate limiting steps.

This study aims to investigate the interaction between PDI and BPTI substrates using the technologies introduced in this chapter as well as other biophysical techniques. The aims can be broken down as follows:

- Create recombinant proteins representative of different stages of the folding pathway of BPTI.

- Characterise each BPTI construct using a variety of biophysical techniques. NMR will be used to identify which regions within each folding intermediate are folded and which remain unfolded. Hydrogen deuterium exchange will enable details to be elucidated about solvent

exposed regions of each BPTI construct, with a timecourse revealing the flexibility of each construct.

- The effects of titrating the PDI enzyme into solutions of each BPTI construct will be observed by NMR. What concentrations of PDI are necessary to observe perturbations in the BPTI spectra? Can regions of interaction be identified?

- The entire PDI enzyme is too large to be observed by NMR. However, the bb'x region of PDI is known to include the main binding regions of the enzyme. This construct has recently been studied by NMR and its structure is now well characterised. Thus, this study will add the various BPTI constructs to bb'x and observe any perturbations to the known spectrum. Does bb'x bind more strongly to one construct compared to another? Does binding occur in different parts of the binding region with different BPTI constructs? How do the observations of BPTI binding compare with binding of other substrates to PDI constructs from previous studies?

- Estimate the binding affinity between PDI and each BPTI substrate. How does the binding differ between substrates at different stages along the folding pathway of BPTI? How do these compare to binding by other PDI substrates studied previously?

It is hoped that the combination of approaches described above will facilitate understanding of the mechanisms of substrate binding to PDI and contribute to an overall understanding of the processes involved in oxidative protein folding in the endoplasmic reticulum.

Chapter 2. Materials and Methods

2.1. Wild Type BPTI and PDI Constructs

All proteins were expressed from a vector derived from pET-23b from Novagen, kindly provided by Dr. L Ruddock. This vector includes a T7 RNA polymerase promoter. Constructs were made by cloning the open reading frame (ORF) and inserting into the NdeI and BamHI sites of the multiple cloning site of the pET-23b vector.

Recombinant BPTI was expressed with a leading methionine, used as a start codon, but the wild type was otherwise identical to mature BPTI. PDI expression vectors included a hexa-histidine tag (His-tag), sequence MHHHHHHM, to facilitate purification. The His-tag was not removed after purification, but showed no evidence of affecting our observations of protein behaviour. All recombinant proteins were expressed in *E. coli* BL21 (DE3) pLysS strains.

2.2. Site Directed Mutagenesis of BPTI

Single point mutations were performed by site directed mutagenesis, mutating guanine to cytosine such that codon changes would be TGC→TCC or TGT→TCT. This would lead to the expression of serine instead of cysteine. Each of the desired mutant forms of BPTI required different combinations of single point mutations. However, many of the single point mutations were common to several of the desired constructs (*e.g.* C14S). A strategy was employed whereby the most common single point mutations would be performed first. These products were then used as templates to introduce

further single point mutations. Both forward and reverse primers were designed to incorporate mutations at each desired position, as listed below:

```
C5SForward : 5'-CGT CCG GAT TTT TCC CTG GAA CCA CCG TAT ACC-3'
C5SReverse : 5'-GGT ATA CGG TGG TTC CAG GGA AAA ATC CGG ACG-3'

C14SForward: 5'-CCG TAT ACC GGC CCT TCT AAA GCG CGG ATC ATT-3'
C14SReverse: 5'-AAT GAT CCG CGC TTT AGA AGG GCC GGT ATA CGG-3'

C38SForward: 5'-GTG TAC GGC GGT TCT CGA GCG AAA CGT AAT AAC-3'
C38SReverse: 5'-GTT ATT ACG TTT CGC TCG AGA ACC GCC GTA CAC-3'

C55SForward: 5'-GAC TGC ATG CGT ACC TCT GGT GGC GCT TAA TAA-3'
C55SReverse: 5'-TTA TTA AGC GCC ACC AGA GGT ACG CAT GCA GTC-3'
```

For each site directed mutagenesis, polymerase chain reaction (PCR) was performed using PfuUltra High-Fidelity DNA Polymerase (Agilent) to amplify the entire vector. 50 µl reactions were set up following manufacturers guidelines, which included 1.0 µl each primer at 100 ng/µl and 1.0 µl of DNA template at 100 ng/µl. The PCR cycling parameters were optimised such that parent DNA was denatured at 95°C for 30 s, followed by 30 cycles of 95°C for 30 s, 55°C for 1 minute to anneal primers, then 68°C for 4 minutes to extend primers. Reactions were incubated at 68°C for a further 10 minutes to ensure complete primer extension. After every mutagenesis, the desired mutation was confirmed by DNA sequencing.

2.3. Protein Expression

2.3.1. Expression of Unlabelled Protein

Ampicillin and chloramphenicol antibiotics were added to an LB agar plate. Frozen glycerol stocks of the appropriate construct were used to streak out cells onto an agar plate and grown overnight at 37°C. A single colony was used to inoculate 50 ml of LB in a 250 ml conical flask containing the appropriate antibiotics. This was incubated overnight at 37°C shaking at 180 rpm. The optical density (OD) of the overnight culture was measured at 600 nm and the appropriate volume used to inoculate a larger culture, typically six two litre flasks containing 400 ml LB each, such that the starting OD_{600} = 0.05. The culture was incubated at 37°C shaking at 180 rpm until OD_{600} = 0.4 (for BPTI) or OD_{600} = 0.5 (for PDI). Overexpression of the recombinant protein was induced with the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The culture was incubated for a further 4 hours at 37°C shaking at 180 rpm. Cells were then harvested by centrifugation at 6000 x g using a Beckman JLA-8.1000 rotor for 12 min at 4°C.

2.3.2. Expression of ^{15}N Labelled Protein

Whenever ^{15}N isotope labelling of samples was required, the expression protocol was identical except that LB was replaced by minimal medium as defined in Table 2.1.

Table 2.1: Minimal medium nutrients added to 1 litre of sterile water. For ^{15}N labelling of recombinant proteins, $(^{15}NH_4)_2SO_4$ was used, which consisted of 99% ^{15}N isotope (Cambridge Isotope Labs).

Nutrient	Amount (mg)
$(NH_4)_2SO_4$	1000
Glucose	4000
Na_2HPO_4	6800
KH_2HPO_4	3000
NaCl	500
Na_2SO_4	42
EDTA	50
$MnCl_2$	16
$FeCl_3$	5
$ZnCl_2$	0.5
$CuCl_2$	0.1
$CoCl_2$	0.1
H_3BO_3	0.1
$MgSO_4$	0.25
$CaCl_2$	44
d-Biotin	1
Thiamine	1

2.3.3. Lysis of Cell Culture

The cell pellet was resuspended in $^1/_{10}{}^{th}$ culture volume 20 mM sodium phosphate, 10 µg/ml DNase, pH 7.3. Cells were initially lysed by freezing overnight, then thawing in a water bath at room temperature. Further lysis occurred by sonication. A Bandelin Sonoplus HD 2070 sonicator was used to apply three bursts of 30 seconds sonication at 70% power with 90 second intervals, while keeping the culture on ice to prevent overheating.

2.4. Purification and Refolding of BPTI Constructs

2.4.1. Isolation of Inclusion Bodies

Lysed cells were centrifuged at 30000 x g for 12 min to pellet all insoluble cell debris. The pellet was resuspended in ¼ volume of inclusion body wash buffer (50 mM Tris-HCl, 10 mM EDTA, 0.5% Triton-X 100, pH 8.0) to break up and solubilise lipid membranes and membrane proteins. Centrifugation and resuspension of the pellet in the wash buffer were repeated, followed by a further centrifugation to pellet the inclusion bodies. The pellet was resuspended in the same volume of water and centrifuged once again. This was repeated to ensure all detergent and EDTA was removed. Inclusion body pellets could then be stored at -20°C if necessary.

2.4.2. Solubilisation of Inclusion Bodies

Inclusion bodies were solubilised by resuspending the pellet in solubilisation buffer (5 M guanidine hydrochloride, 50 mM tris-HCl, 20 mM DTT, pH 8.7) using $^{1}/_{200}{}^{th}$ volume of the total cell culture (*i.e.* 2400 ml culture resuspend in 12 ml solubilisation buffer). DTT was added to the buffer just before use to prevent air oxidation. The mixture was incubated at room temperature for 45 min. This reduced and denatured the protein molecules, causing the inclusion bodies to break up into monomeric BPTI.

2.4.3. Desalting of Solubilised Inclusion Bodies

The denaturant and reducing agent were removed from the BPTI solution using desalting columns. 5 ml HiTrap desalting columns (GE Healthcare) contain a

Superdex g-25 Superfine matrix and were used to effectively remove all small molecular weight substances. Columns were attached to an ÄKTApurifier 100 FPLC. Typically, four columns were attached in series to allow a larger volume of protein to be injected. The solvent used for desalting was 10 mM HCl, so that the protein remained in a reduced state at a low pH after elution. The columns were equilibrated with 3 CV of solvent. The solubilised inclusion body sample was centrifuged to pellet any remaining precipitate. 3 ml of supernatant was then loaded onto a 5 ml loop before being injected onto the column. Typically, a flow rate of 1.5 ml/min was used, ensuring that the maximum pressure (0.3 MPa) was not exceeded. The absorbances at 280 nm and 254 nm were recorded along with the conductance. The eluted material was collected in 1 ml fractions. Fractions correlating to observed peaks at 280 nm were analysed by SDS-PAGE to identify the BPTI protein. If required, the process was repeated, loading 3 ml of sample on each occasion. Separate HiTrap columns were maintained for each BPTI construct, to avoid cross-contamination.

2.4.4. Oxidation and Refolding of BPTI

Reduced BPTI was refolded to its fullest extent via an oxidation reaction. In the case of wild type BPTI, this generated the native protein. In the case of other BPTI constructs, the extent to which the protein proceeded through the folding pathway was determined by the combination of cysteines present, see Figure 1.12. Each sample was diluted such that the final concentration of protein would be 50 µM (0.3 mg/ml) after later addition of the refolding buffer containing oxidising and reducing agents. To prevent aggregation of the protein, it was important to dilute the BPTI before addition of the refolding agents. Separately, the following components were added together, such that the final concentrations (after later addition of BPTI) were as follows:

- Refolding buffer, 0.1 M Tris-HCl + 1mM EDTA, pH 8.7
- Oxidising agent, 0.5 mM glutathione disulfide (GSSG)

- Reducing agent, 2 mM glutathione GSH

Note that as well as using GSSG as the oxidising agent, reducing GSH is also added to facilitate the isomerisation of incorrectly paired disulfide bonds. Once prepared, the reagents were used immediately since they deteriorate rapidly. The protein was then added gradually to the refolding buffer while stirring. The reaction was incubated overnight at 4°C while stirring gently. The reaction mixture was centrifuged at 75,000 x g to pellet any aggregated protein.

2.4.5. Post-refolding Purification using Solid Phase Extraction

Once refolding was complete, the diluted protein was concentrated and the refolding buffer and agents were removed. This was achieved using solid phase extraction (SPE). Pre-packed 3 ml Grace Vydac SPE columns were used, containing reverse phase C-18 bonded to 13 µm silica beads with a 300 Å pore diameter. The relatively large pore size was essential for use with macromolecules such as proteins. Optimisation of experimental conditions indicated that these columns have a maximum binding capacity of 3.0 mg for BPTI. Protein samples were prepared by adding ¼ volume of sample buffer (20% acetonitrile, 2% TFA). The column was activated with 3 ml 50% acteonitrile, then equilibrated with 3 ml equilibration buffer (5% acetonitrile, 0.5% TFA). The sample was applied to the column either manually by syringe or by gravity flow overnight at 4°C, depending on the sample volume. The column was rinsed with 3 x 3 ml 5% acetonitrile to wash off weakly bound protein and to ensure all TFA was removed. Protein was eluted in 2 x 3 ml 50% acetonitrile.

2.4.6. Vacuum Evaporation and Lyophilisation

The sample eluted from SPE in 50% acetonitrile was put into a Duran vacuum desiccator (Schott), which was attached to a diaphragm pump and run under

vacuum overnight to evaporate off the volatile acetonitrile, leaving the protein in water. This was then frozen in liquid nitrogen and the water was sublimated by lyophilisation for 4-6 hours until only a white powder remained. The lyophilised protein was stored at -20°C.

2.5. Purification of PDI Constructs

2.5.1. Immobilised Metal Affinity Chromatography (IMAC)

Lysed cells were centrifuged at 30,000 x g for 12 min to pellet all insoluble cell debris. Initial purification of PDI constructs from the soluble portion of the cell lysate was performed using an IMAC column, whereby the His-tag of the recombinant protein binds to the nickel bound to a sepharose resin. IMAC columns were produced by adding Chelating Sepharose Fast Flow resin (GE Healthcare) to a syringe to produce a column volume (CV) of 5 ml. The column was rinsed with 10 CVs of water before adding 2 ml of 0.2 M nickel chloride. After rinsing with 5 CVs of water, 5 CVs of 20 mM sodium acetate, 0.5 M NaCl pH 3.0 was added to rinse out any weakly bound nickel. The column was equilibrated with 5 CVs of 20 mM sodium phosphate pH 7.3. The crude protein sample (total soluble portion of cell lysate) was then loaded onto the column. Loosely bound impurities were removed using 2 CVs of a low imidazole wash (25 mM imidazole, 0.5 M sodium chloride, 20 mM sodium phosphate, pH 7.3). This was followed by 5 CVs of a low salt solution (20 mM sodium phosphate pH 7.3). PDI was then eluted using 5 CVs of 50 mM EDTA, 20 mM sodium phosphate, pH 7.3.

2.5.2. Buffer Exchange by Dialysis

Dialysis was used to buffer exchange the eluted samples from the IMAC purification to remove the nickel and EDTA. Dialysis membrane with a molecular weight cut-off (MWCO) of 12-14 kDa (Medicell International) was used to buffer exchange into 20 mM sodium phosphate, pH 7.3.

2.5.3. Ion Exchange Chromatography

Further purification of the protein was performed using anion exchange chromatography. A 10 ml Source 30Q column (GE Healthcare) provided a resin with a strong positive charge. This was connected to an ÄKTApurifier 100 fast protein liquid chromatography (FPLC) system (GE Healthcare). The column was rinsed and equilibrated in 20 mM sodium phosphate, pH 7.3 (Buffer A). The protein sample was loaded onto the column via an injection loop. A gradient from 0-100% of 20 mM sodium phosphate, 0.5 M NaCl, pH 7.3 (Buffer B) was used to elute the protein over 10 CVs. 2 ml fractions were collected. Those corresponding to a peak in the chromatogram were analysed by SDS-PAGE to identify the desired protein.

2.5.4. Size Exclusion Chromatography

For preparations of the bb'x construct of PDI, size exclusion chromatography was used to separate the monomer from the dimer. A Superdex 75 10/300 GL gel filtration column (GE Healthcare) was connected to an ÄKTApurifier 100 FPLC (GE Healthcare). 20 mM sodium phosphate, 150 mM NaCl, pH 7.3 was used as the buffer. The column was equilibrated using 2 CVs of buffer (CV = 300 ml). Samples were concentrated using Centricon filter columns (10 kDa MWCO) to a volume of 1 ml. The sample was injected via a 2 ml injection loop

and buffer was consumed at a flow rate of 2.0 ml per minute. Protein elution was monitored at A_{280} and 5 ml fractions collected. Fractions showing absorbance peaks were analysed by native PAGE to verify monomer and dimer states.

2.6. Estimating Protein Concentration

Protein concentration was determined by measurement of the UV absorbance at 280 nm (A_{280}) using a Jenway 6305 spectrophotometer. The extinction coefficient (ε_{280}), expressed in $M^{-1}cm^{-1}$, for each protein was calculated by submitting the sequence of the mature protein (including His-tag, where appropriate) to the ProtParam tool of the Expasy Proteomics Server (available at http://www.expasy.ch/tools/protparam.html). The protein concentration was then calculated based on the Beer-Lambert law:

$$\text{Protein Concentration (M)} = A_{280} / (\varepsilon_{280} \times \text{path length of cuvette (cm)})$$

2.7. Alkylation using Iodoacetamide

Successful oxidation of wild type BPTI and BPTI mutants can be judged by the number of free thiols groups remaining. Free thiol groups can be alkylated by a carboxyamidomethylation reaction with iodoacetamide. When analysed by mass spectrometry, an additional 57.1 Da will be added to the mass of the protein for each free thiol group. 2.2 M iodoacetamide was prepared from powder into 100 mM sodium phosphate pH 7.0. Iodoacetamide is light sensitive, so was kept in the dark at all times by covering with aluminium foil. To fully dissolve, the iodoacetamide was kept at 65°C. BPTI samples were prepared by dilution to make 50 µl of 50 µM in 100 mM sodium phosphate pH 7.0 and kept on ice. Samples were used immediately to avoid air oxidation. The iodoacetamide was removed from the heat for 90 seconds to cool. The protein sample was added

and incubated at room temperature for exactly 30 seconds. Iodoacetamide, while binding preferentially to free thiol groups, will also show some binding to other amino acid side chains, so it was important not to let the reaction proceed for too long. The reaction was quenched by adding 33 µl of 20% acetonitrile, 2% TFA. This will quench the reaction by reducing the pH. The sample was frozen immediately in liquid nitrogen and stored at -80°C.

2.7.1. Alkylation to Trap Reduced Wild Type BPTI

The iodoacetamide reaction can also be used to stabilise the fully reduced wild type BPTI, if performed before the oxidation reaction. This fully reduced BPTI effectively represents a protein at the beginning of the folding pathway. As such, its structural interaction with PDI could be very informative. The protocol of alkylation to trap fully reduced BPTI was essentially as described above, but on a preparative scale. To reduce the amount of iodoacetamide required, the reaction used only 50 mM iodoacetamide, but the reaction was allowed to proceed for 20 min. Aliquots of the fully reduced protein were analysed by mass spectrometry to verify that alkylation was successful. As with fully folded BPTI, the reduced protein was further purified using solid phase extraction, see Section 2.4.5.

2.8. ESI Mass Spectrometry

2.8.1. Sample Preparation

All mass spectrometry samples of BPTI were prepared to a concentration of 50 µM in 50 µl of 50% acetonitrile. Normally, concentrating and buffer exchange of the sample was required. This was achieved by applying the sample to a PepClean C-18 Spin Column (Pierce). This contains a porous C-18 reversed-

phase resin that the protein can bind to during centrifugation in an aqueous buffer. The protein was then eluted by centrifugation with 50% acteonitrile.

2.8.2. Data Acquisition

Mass spectrometry was performed on a Bruker mircOTOF electrospray ionisation (ESI) time of flight (TOF) instrument. The control software was used to observe a window between 500 and 2500 mass/charge ratio (m/z). Each sample was injected via a syringe at a flow rate of 240 µl/h. Signals were recorded over 1.5 mins and an averaged spectrum was acquired. A mass/charge ruler facilitated the identification of the dominant mass species in various charge states. This was used to deconvolute the spectrum to mass versus intensity, independent of the charged state.

2.9. Electrophoresis Analysis

2.9.1. SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse protein samples under denaturing and reducing conditions.

2.9.1.1. SDS-PAGE for Analysis of BPTI Constructs

In order to achieve good resolution of bands near to the small molecular weight of BPTI (6.6 kDa), pre-cast gels were used for SDS-PAGE analysis. 12% NuPAGE Novex Bis-Tris Mini Gels (Invitrogen) were run using a 2-(N-morpholino)ethanesulfonic acid (MES) running buffer. Gels were typically run at 200 V for 35 min. SeeBlue Plus 2 Pre-Stained Standard (Invitrogen) provided

protein markers with a suitable range of molecular weights, including commercial pancreatic trypsin inhibitor (also known as aprotinin).

2.9.1.2. SDS-PAGE for Analysis of PDI Constructs

SDS-PAGE of PDI constructs were carried out according to the Tris-Glycine buffer system (Laemmli 1970). Discontinuous gels were cast consisting of a resolving gel at pH 8.8 and a stacking gel at pH 6.8. Typically, 12% acrylamide was used in the resolving gel. Gels were cast using the Mini Protean II gel kit (Bio-Rad). Samples were reduced with the addition of 10% β-mercaptoethanol to the sample loading buffer (1 M Tris-HCl pH 6.8, 40% glycerol, 0.8% SDS, 0.1% w/v bromophenol blue). A low molecular weight marker from Amersham GE Healthcare was used to estimate the molecular weight of each band. Gels were run at 60 V through the stacking gel, then at 200 V until the dye front ran off the gel (typically 60 min). Gels were stained using Coomassie Blue R250 dye.

2.9.2. Native PAGE

To distinguish between monomeric and dimeric states of various PDI constructs, PAGE was run under native conditions. The method was identical to SDS-PAGE of PDI constructs, except that SDS and β-mercaptoethanol were omitted from the gels, the loading buffer and the running buffer.

2.10. Circular Dichroism

Far-UV circular dichroism (CD) was performed on BPTI samples using a Jasco J-815 CD spectropolarimeter at 25°C, unless stated otherwise. Lyophilised samples were resuspended in 10 mM sodium phosphate, pH 7.3 to a concentration of 0.1 mg/ml (15 μM). Spectra were recorded between 260 nm

and 190 nm, averaged over 16 scans per sample and baseline corrected using buffer only. CD spectra were analysed using Dichroweb (Whitmore and Wallace 2008).

2.11. Surface Plasmon Resonance

Surface plasmon resonance was performed on a Biacore 2000 instrument from GE Healthcare. All experiments used a CM5 sensor chip, which contains a carboxymethyl dextran surface, onto which proteins were immobilised. All chemicals were supplied by GE Healthcare specifically for use with Biacore instruments. HBS-EP, a HEPES buffer containing EDTA and Surfactant P20, was used as a running buffer. The temperature was varied as described in the results.

2.11.1. Protein Immobilisation by Amine Coupling

All proteins were immobilised to the chip surface by amine coupling, whereby succinimide esters react spontaneously with primary amine groups to link covalently to the dextran matrix. Binding may occur at NH_2 of amino acid side chains as well as the N-terminus of the protein. Prior to immobilisation of the desired protein, pH scouting was performed to establish the most effective pH, which turned out to be pH 5.0 for all immobilised proteins.

Preparing the dextran surface of the CM5 sensor chip for amine coupling first requires reactions with 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and 0.05 M N-hydroxysuccinimide (NHS). This activates the carboxyl groups on the surface to give reactive succinimide esters. The succinimide esters react spontaneously with amine groups, allowing direct immobilisation of molecules. The protein was prepared to a concentration of 50 µg/ml in HBS-EP

buffer. The standard protocol for amine coupling was followed using the manufacturer's instructions. Once the protein immobilisation was complete, 1 M ethanolamine was used to deactivate excess reactive groups.

2.11.2. Binding Analysis

All protein samples used as analyte for binding analysis were prepared in HBS-EP running buffer to a volume of 370 µl for 2 replications. Prior to analysis, all flow cells were normalised using 40% glycerol to compensate for variations in the reflectance characteristics between individual sensor chips. The procedure followed the manufacture's instructions. Analyte was injected at a flow rate of 50 µl/min, which was fast enough to prevent mass transport effects. Regeneration of the chip after each analysis was performed using 400 µl of 50 mM NaOH.

2.11.3. Analysis of Biacore Data

BIAevaluation version 3.1 was used to process and analyse the raw sensorgram data. Sensorgrams were baseline corrected using the Y-transform function, by zeroing the signal in a region before the start of injection. Only the association and dissociation phases of each sensorgram were used for binding analysis. All curve fitting used simultaneous k_a/k_d calculations.

Fitting simultaneous k_a/k_d models to experimental data involves both generation of the rate equations for the model used and finding the values for parameters in the rate equations that best fit the experimental data.

During curve fitting calculations, differences are calculated between the experimental data points and the calculated points of the model, known as residuals. Using BIAevaluation, curve fitting optimises parameter values by minimising the sum of the squared residuals, which is performed by the

Marquardt-Levenberg algorithm. The algorithm starts with the initial values given for each parameter in the model, then iterates to minimise the sum of squared residuals (Biacore 1997).

The most appropriate fit of experimental data to a model is influenced by both the ability of the fitting algorithm to converge towards the true minimum in the sum of squared residuals function and by the number of parameters that can be varied in the model.

It is important to bear in mind that as the complexity of a model increases the ability to fit the equations to the experimental data improves, since the number of parameters involved increase and hence the degrees of freedom in the model increase. The simplest model that fits the data should always be regarded as the most appropriate.

For the data analysed in this study, many different models were tried. However, the two predefined BIAevaluation models that were examined closely were the 1:1 (Langmuir) Binding model and the Heterogeneous Ligand (Parallel Reactions) model. Use of these models for curve fitting to PDI/BPTI interaction sensorgrams and optimisation of the parameters for each model are discussed in section 7.4.

2.11.3.1. Definition of 1:1 (Langmuir) Binding Model

The 1:1 (Lamgmuir) Binding model is the simplest model, which assumes a 1:1 interaction between the analyte (A) and the immobilised ligand (B), forming the complex (AB).

Reaction: $A + B \leftrightarrow AB$

Parameters:

tOn – Sample injection start time (s), constant.

tOff – Sample injection end time (s), constant.

Conc – analyte concentration (M), constant.

Rmax – Maximum analyte binding capacity (Response Units), fit globally.

k_a – Association rate constant ($M^{-1}s^{-1}$), fit globally.

k_d – Dissociation rate constant (s^{-1}), fit globally.

RI – Bulk refractive index contribution (Response Units), fit locally.

Rate Equations:

A = Conc

B[0] = Rmax

dB/dt = - (ka*A*B - kd*AB)

AB[0] = 0

dAB/dt = (ka*A*B - kd*AB)

Total response:

AB + RI

Calculated Report Parameters:

K_A (1/M) = k_a/k_d. Equilibrium binding (affinity) constant, global.

K_D (M) = k_d/k_a. Equilibrium dissociation constant, global.

R_{eq} (Response Units) = k_a*Conc*R_{max}*/(k_a*Conc+k_d). Steady state binding level, local.

k_{obs} (1/s) = k_a*Conc+k_d. Observed rate constant (slope of ln(dR/dt) against t), local.

2.11.3.2. Definition of Heterogeneous Ligand (Parallel Reactions) Model

The Heterogeneous Ligand (Parallel Reactions) model describes an interaction between one analyte (A) and two independent ligands (B1 and B2), forming two complexes (AB1 and AB2). The calculated fitted curve is simply the sum of the two independent reactions. Heterogeneous ligand situations frequently occur in practice due to heterogeneous immobilisation of ligand as well as through heterogeneity in the ligand itself.

Reaction 1: A + B1 $\leftrightarrow$ AB1

Reaction 2: A + B2 $\leftrightarrow$ AB2

Parameters:
tOn – Sample injection start time (s), constant.
tOff – Sample injection end time (s), constant.
Conc – analyte concentration (M), constant.
Rmax – Maximum analyte binding capacity (Response Units), fit globally.
k_a1 – Association rate constant for A+B1=AB1 ($M^{-1}s^{-1}$), fit globally.
k_d1 – Dissociation rate constant for A+B1=AB1 (s^{-1}), fit globally.
k_a2 – Association rate constant for A+B2=AB2 ($M^{-1}s^{-1}$), fit globally.
k_d2 – Dissociation rate constant for A+B1=AB2 (s^{-1}), fit globally.
RI – Bulk refractive index contribution (Response Units), fit locally.

Rate Equations:
A = Conc
B1[0] = RMax1
dB1/dt = - (ka1*A*B1 - kd1*AB1)
B2[0] = RMax2
dB2/dt = - (ka2*A*B2 - kd2*AB2)

AB1[0] = 0
dAB1/dt = (ka1*A*B1 - kd1*AB1)
AB2[0] = 0
dAB2/dt = (ka2*A*B2 - kd2*AB2)
Total response:
AB1 + AB2 + RI

Calculated Report Parameters:
K_A1 (1/M) = k_a1/k_d1. Equilibrium binding (affinity) constant for A+B1, global.
K_A2 (1/M) = k_a2/k_d2. Equilibrium binding (affinity) constant for A+B2, global.

2.12. Nuclear Magnetic Resonance Spectroscopy

2.12.1. Sample Preparation

Unless otherwise stated, the NMR buffer used was 25 mM sodium phosphate, 100 mM NaCl, 10% D_2O, pH 6.5. This buffer was compatible with all BPTI and PDI constructs and allowed for direct comparison with previous PDI spectra.

BPTI samples were resuspended with the NMR buffer directly from lyophilised powder. This was centrifuged at 16,000 x g for 2 min to remove any precipitate and the pH was checked. Protein concentrations were measured but varied according to the solubility of each construct, which depended on each protein's stage through the folding pathway, and are specified for individual results. Experiments observing BPTI used 180 µl sample inserted into 3 mm Bruker Match NMR tubes.

Since bb'x alters due to lyophilisation, samples were prepared fresh and concentrated to 1.0-1.5 mM in 25 mM sodium phosphate, 150 mM NaCl, 10%

D_2O, pH 6.5 using Centricon filter column (10 kDa MWCO). 330 µl of sample were required for each experiment, which used 5 mm Shigemi tubes.

2.12.2. Data Acquisition of BPTI Samples

All BPTI experiments were performed on a Bruker AV II 700 spectrometer operating with a 16.45 Tesla ultra shielded superconducting magnet (700 MHz ^{1}H resonance frequency). It was operated using a TCI cryoprobe $^{1}H/^{13}C/^{15}N$ with Z gradients (Bruker). Operating temperatures varied and are specified for each result. Results were acquired using TopSpin version 2.1 (Bruker). Locking was performed against the deuterium signal. Shimming, matching and tuning were performed automatically using the appropriate commands. Water suppression was performed by using an excitation sculpting pulse sequence. The spectral width was set to 16 ppm and the power level was 5.90 db for a proton signal. The transmitter frequency offset, representing the precise chemical shift of the bulk water, and pulse time parameters were optimised for each sample. The centre of each spectrum was set to the transmitter frequency offset. A receiver gain adjustment was performed prior to data acquisition so that the optimum amount of signal would be detected by the receiver. For longer experiments, auto-shimming was activated to ensure a homogenous magnetic field was maintained. A 1D spectrum was acquired before and after longer experiments and compared to ensure the sample remained unaltered.

2.12.3. Data Processing of BPTI Data

Raw NMR data were processed using TopSpin version 2.1 (Bruker). The free induction decay (FID) signal was truncated using the qsin apodisation function, to optimise the signal-to-noise ratio. Zero filling of the truncated signal was used to avoid any degradation of the resolution. The FID was then Fourier transformed to change the signal from the time domain into the frequency

domain. Data were then phased and baseline corrected to optimise the symmetry of the peaks. The water peak from solvent was used to calibrate each spectra, with the precise ppm adjusted to account for variations in temperature (Wishart, Bigam et al. 1995).

2.12.4. Data Acquisition and Processing of PDI Samples

All HSQC spectrometry of ^{15}N labelled bb'x samples were performed alongside colleagues from the University of Kent. Samples were prepared with assistance from Dr. L. Byrne and Dr. R. Williamson by resuspending in NMR buffer and, after measuring the stock concentration, creating samples of the appropriate concentration. Data were acquired at 298 K (25°C) using a four channel Varian Unitylnova NMR spectrometer operating at 14.1 Tesla (600 MHz ^{1}H resonance frequency), equipped with a 5 mm HCN z-pulse field gradient probe. All experiments were solvent suppressed using WATERGATE (Piotto, Saudek et al. 1992) using a gradient field strength of 40-50 G cm^{-1}. Data processing was carried out by Dr. M. Rowe (University of Kent) using Varian and Bruker software.

2.12.5. 2D ^{1}H-^{1}H TOCSY / 2D ^{1}H-^{1}H NOESY

Both 2D TOCSY and 2D NOESY experiments were performed using 4096 data points in the F2 dimension (^{1}H) and 512 data points in the F1 dimension (^{1}H), with a mixing time of 60-80 ms. Total acquisition time of either experiment was approximately 8.5 hours. 2D TOCSY and 2D NOESY were normally run sequentially with the same sample.

2.12.6. ^{15}N-^{1}H HSQC

Unless otherwise stated, HSQC experiments were performed using 2048 data points in the F2 dimension (^{1}H) and 128 data points in the F1 dimension (^{15}N). Total acquisition time was approximately 15 min.

2.12.6.1. ^{15}N-^{1}H HSQC at Various Temperatures

Wherever HSQC spectra were collected at various temperatures, the lock signal was allowed to stabilise before proceeding. Shimming, matching and tuning were repeated at each new temperature.

2.12.6.2. ^{15}N-^{1}H HSQC of BPTI with PDI Titration

Titration experiments were performed with ^{15}N labelled wild type BPTI, ^{15}N labelled (30-51, 5-14) BPTI and ^{15}N labelled fully reduced BPTI at ratios ranging from 200:1 BPTI:PDI to 1:1 BPTI:PDI. For wild type and (30-51, 5-14) BPTI, HSQC spectra were obtained using 400 µM samples. Due to its limited solubility, HSQC spectra of reduced BPTI were acquired using 100 µM samples. The amino acid assignments from BPTI without PDI present were then mapped onto spectra with increasing PDI, thus identifying those BPTI residues least affected in the BPTI:PDI complex.

2.12.6.3. ^{15}N-^{1}H HSQC of bb'x with BPTI Titration

Separate titration experiments were performed for different BPTI constructs: wild type BPTI, (30-51, 5-14) BPTI and fully reduced BPTI. For wild type BPTI and (30-51, 5-14) BPTI, HSQC spectra were obtained using 250 µM bb'x, with bb'x:BPTI ratios of 25:1, 5:1 and 1:1. Due to the limited solubility of reduced

BPTI, HSQC spectra were obtained using 250 μM bb'x at bb'x:BPTI ratios of 25:1 and 8.3:1. An additional HSQC spectrum was obtained at a 2.3:1 ratio, but this required using bb'x at just 100 μM. The amino acid assignments from bb'x without BPTI were used to assign spectra with increasing BPTI, identifying residues that were significantly perturbed in the presence of the substrate.

2.12.6.4. ^{15}N-^{1}H HSQC with Hydrogen Deuterium Exchange

^{15}N-^{1}H HSQC experiments rely on a proton resonance. If the protein is resuspended in a deuterium oxide NMR buffer, the exchange of protons for deuterium will result in a loss of signal. Hence, the most solvent exposed regions of the protein will lose their HSQC peak first, whereas the buried regions of the protein will persist for longer. By comparing a timecourse of different BPTI constructs, the stability and extent of folding of each can be elucidated.

For hydrogen deuterium exchange, the NMR buffer was as before, except it now used 100% D_2O as the solvent (25 mM sodium phosphate, 100 mM NaCl, 100% D_2O, pH 6.5). ^{15}N labelled samples were prepared for wild type, (30-51, 5-14) and fully reduced BPTI. Since it is important to acquire HSQC spectra as quickly as possible after resuspending in the buffer, a dummy sample of ^{15}N labelled wild type BPTI was used for locking, shimming, matching, tuning and optimisation of parameters prior to the preparing the first experimental sample. All experiments were performed at 278K (5°C). To allow a greater frequency of HSQC spectra to be collected within the first hour of the timecourse, the acquisition time was decreased by halving the number of data points per scan. This decreased the HSQC acquisition time to approximately 5 min. Acquisition of the initial HSQC spectrum for each sample was started 5 min after the lyophilised protein was resuspended in the deuterium oxide NMR buffer. Spectra were obtained every 5 min until 20 min had elapsed, then every 10 min

until 1 hour had elapsed, then every hour until 7 hours elapsed time. Final spectra were recorded after 24 hours.

2.12.7. Data Analysis

Analysis of NMR data for all BPTI experiments was performed using Sparky version 3.1. Data analysis of bb'x experiments were carried out using CcpNMR Analysis version 2 (Vranken, Boucher et al. 2005). Initial experiments with bb'x alone were used as a control to analyse the effect of adding various BPTI constructs.

2.12.7.1. Peak Assignment of NMR Spectra from Previous Data

Wild type BPTI has been studied extensively by NMR. Previous NMR spectral data from Biological Magnetic Resonance Data Bank (BMRB) entry 5359 (Biamonti 1996) were used as a template for the assignment of initial spectra. Likewise, a previous study of bb'x (Byrne, Sidhu et al. 2009) provided spectral data for initial assignment of bb'x control spectra acquired in this study, using BMRB entry 15974.

2.12.7.2. *De Novo* Assignment of (30-51, 5-14) BPTI Backbone

^{1}H NMR analysis has been performed previously on the (30-51,5-14) non-native two disulfide intermediate of BPTI (Vanmierlo, Darby et al. 1991a). However, this was performed using different conditions, *e.g.* a pH of 4.6, so was not comparable for this study. To assign the backbone resonances for the (30-51, 5-14) BPTI, 3D HSQC-TOCSY and 3D HSQC-NOESY experiments were performed. These were used to identify each residue then establish a sequential

assignment. This data was then mapped to the corresponding peaks in a ^{15}N-^{1}H HSQC spectrum.

2.12.7.3. Chemical Shift Mapping

Analysis of the chemical shift between peaks of two HSQC spectra was performed by establishing a peak list for each spectrum, which identified the ^{1}H and ^{15}N resonances for each peak. The chemical shift difference was calculated as:

$$\text{combined shift difference} = |\ ^{1}H \text{ shift}\ | + 1/6\ |\ ^{15}N \text{ shift}\ |$$

The ^{15}N shift is divided by six because the chemical shift range of ^{15}N is approximately six times larger than for ^{1}H. The justification of this approach has been described in detail by previous studies (Garrett, Seok et al. 1997; Williamson, Carr et al. 1997). Visual depiction of chemical shifts was illustrated using MolMol version 2K.2.

Chapter 3. Preparation and Characterisation of BPTI Constructs

3.1. Introduction

The aim of this study was to investigate the structure of folding intermediates of BPTI and their interaction with PDI. The first stage in this process was to prepare all the appropriate constructs. Previous work in the Freedman group laboratory had produced well established protocols for the expression and purification of various PDI constructs. Details of expression of these constructs to study interaction with BPTI are outlined in the relevant NMR results chapters. Here, focus is on the expression and purification of the various BPTI constructs.

First, the expression system is explained. An outline of the mutations required for each of the folding intermediates is shown. Small culture expressions reveal which mutant constructs show sufficient expression. After some optimisation, large scale expression and purification is pursued.

For each construct, initial purification was performed through a series of washes designed to isolate inclusion bodies. Inclusion bodies were then solubilised using denaturant and reducing agent. The protein was then desalted to buffer exchange into 10 mM HCl. Where appropriate, the reduced protein was then oxidised by dilution using a buffer that included both oxidising and reducing agents, thus facilitating the processes of both oxidation and isomerisation. This resulted in each construct refolding to its fullest extent. Solid phase extraction was used to both concentrate the refolded protein and to remove the refolding agents. Finally, each construct was lyophilised prior to storage at -20°C.

Throughout this chapter, SDS-PAGE, chromatograms, mass spectra and circular dichroism are used to identify and characterise each of the BPTI constructs. Alkylation reactions were used prior to mass spectrometry to verify the oxidation state of each protein.

3.1.1. Expression Vector of Wild Type BPTI

The expression vector containing the ORF for recombinant wild type BPTI was kindly provided by Prof. Lloyd Ruddock from the University of Oulu, Finland. The expression vector was pET23b, which is a frame-shifted version of pET23a shown in Figure 3.1. The ORF for wild type mature BPTI was inserted into the multiple cloning site of the vector as described in section 2.1.

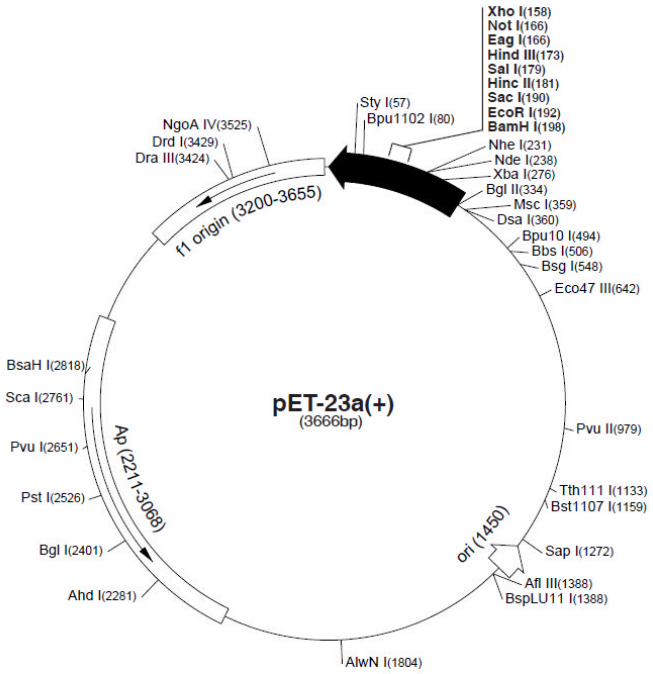


Figure 3.1: Vector map of pET23a (Novagen). Vector pET23b is identical, except it is a 3665bp plasmid, with 1 base pair subtracted after BamHI at 198.

3.1.2. BPTI Mutations to Trap Intermediates along the Folding Pathway

In order to generate partly-folded intermediates that could be analysed to elucidate their structure and interaction with PDI, mutations were performed to the wild type DNA. Figure 1.12 illustrates the folding pathway of BPTI, showing the combination of disulfide bonds used to define each intermediate. In order to trap each of the partly folded intermediates and prevent further oxidation, any thiol groups not involved in the disulfide bonds already formed were removed. This was performed by mutating out the appropriate cysteine residues. Serine is structurally very similar to cysteine, except that the thiol is replaced by a hydroxyl group. This study was essentially reconstructing many of the partly folded BPTI intermediates created by the Creighton group in the 1990s (see section 1.12). Figure 3.2 illustrates the desired effect of mutations to produce the (30-51, 5-14) intermediate, resulting in a folding intermediate similar to that created in a previous study (Vanmierlo, Kemmink et al. 1994). The lack of thiol groups at positions 38 and 55 prevent isomerisation to the (30-51, 5-55) or (30-51, 5-38) species.

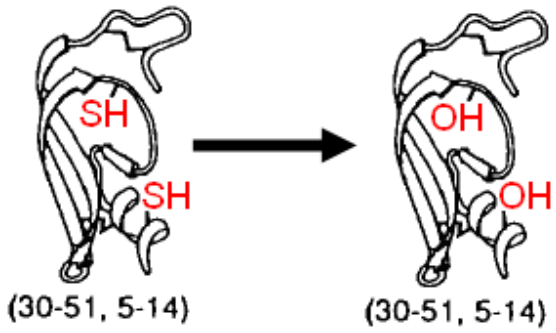


Figure 3.2: Double mutation of wild type BPTI. The mutations replaced cysteines with serines at specified locations. This prevented the formation of certain disulfide bonds, thus blocking the protein's progression through the folding pathway. The example shown has C38S and C55S mutations; when fully oxidised the (30-51, 5-14) non-native two disulfide bonded intermediate is formed.

The combinations of mutations required to trap each of the predominant species of the BPTI refolding pathway are outlined in Table 3.1. Site directed mutagenesis was performed for each single point mutation, changing guanine to cytosine such that codon changes would be TGC→TCC or TGT→TCT. Mutagenesis was performed in the Freedman group by Dr. Edward Jack, as described in Section 2.1.

Table 3.1: Mutations required to establish BPTI at various stages along its folding pathway. Non-native disulfide bonds are in italics.

Mutants (Cys to Ser)	**Disulfides Produced**	**Stage of Folding Pathway**
Wild Type	(30-51, 5-55, 14-38)	Native Protein
C14S C38S	(30-51, 5-55)	2 Disulfide Native Intermediate
C14S C55S	(30-51, *5-38*)	2 Disulfide Non-native Intermediate
C38S C55S	(30-51, *5-14*)	2 Disulfide Non-native Intermediate
C5S C14S C38S C55S	(30-51)	1 Disulfide Native Intermediate

3.2. Results

3.2.1. Verification by DNA Sequencing

Purified plasmids were sent for DNA sequencing for all BPTI constructs. The T7 promoter primer was used to sequence the BPTI insert in each case. This confirmed that all inserts contained the correct sequence needed to express the desired BPTI mutant.

3.2.2. Test Expression of BPTI Constructs

Small cultures were used to test the expression of each of the BPTI constructs specified in Table 3.1. The procedure was as specified in section 2.3.1, except that only 5 ml cultures were used for expression. After 4 hours expression, samples were centrifuged and the cell pellets were resuspended in buffer. These were frozen, thawed and sonicated to lyse cells. An aliquot of each sample was analysed by SDS-PAGE (Figure 3.3). Clear bands are observed for BPTI at 6 kDa for wild type BPTI and for the C38S C55S mutant.

All of the mutants of BPTI that did not express contain the C14S mutation. Although the DNA inserted into the expression vector of these constructs has been confirmed by sequencing, it is suspected that during mutagenesis of cysteine 14 to serine, using the PfuUltra high fidelity polymerase, another mutation occurred elsewhere in the vector that has prevented expression of these constructs. Some attempts were made to sub-clone these inserts into fresh pET23b vectors, which may have permitted expression of these constructs. However, these attempts were unsuccessful.

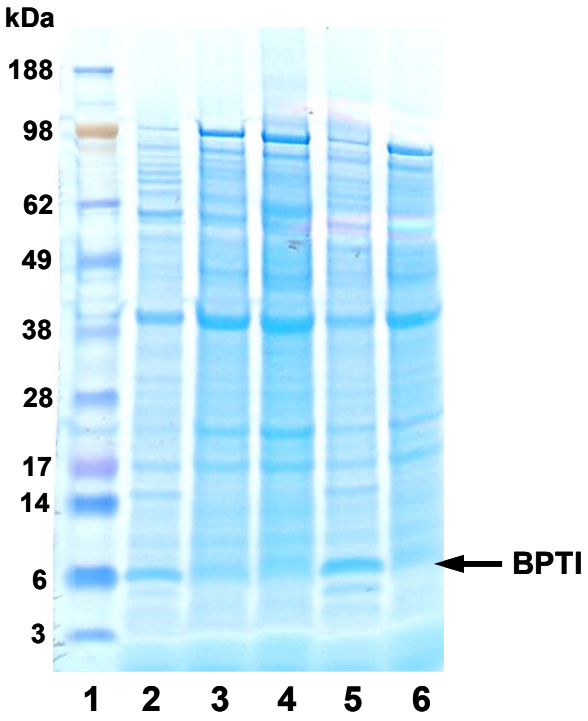


Figure 3.3: SDS-PAGE for test expression of various BPTI constructs. Lane 1, SeeBlue marker; lane 2, wild type BPTI; lane 3, C14S, C38S mutant BPTI; lane 4, C14S, C55S mutant BPTI; lane 5, C38S, C55S mutant BPTI; lane 6, C5S, C14S, C38S, C55S mutant BPTI (quad mutant). A distinct band can be seen at 6.6 kDa for both the wild type BPTI and for the C38S C55S mutant BPTI, but not for any of the constructs containing the C14S mutation.

3.2.3. Optimisation of Expression

To test if expression of BPTI constructs was induced with IPTG, two small cultures of wild type BPTI were grown to an OD_{600} = 0.4. At this stage, one culture was induced with 1 mM IPTG while the other culture remained without IPTG. Each culture was incubated at 37°C, with aliquots removed every hour between 3 and 6 hours, plus a final aliquot removed after overnight incubation. SDS-PAGE analysis of each aliquot is shown in Figure 3.4. It clearly shows increased expression in the culture induced with IPTG, with similar levels of protein expression occurring between 3-6 hours post-induction.

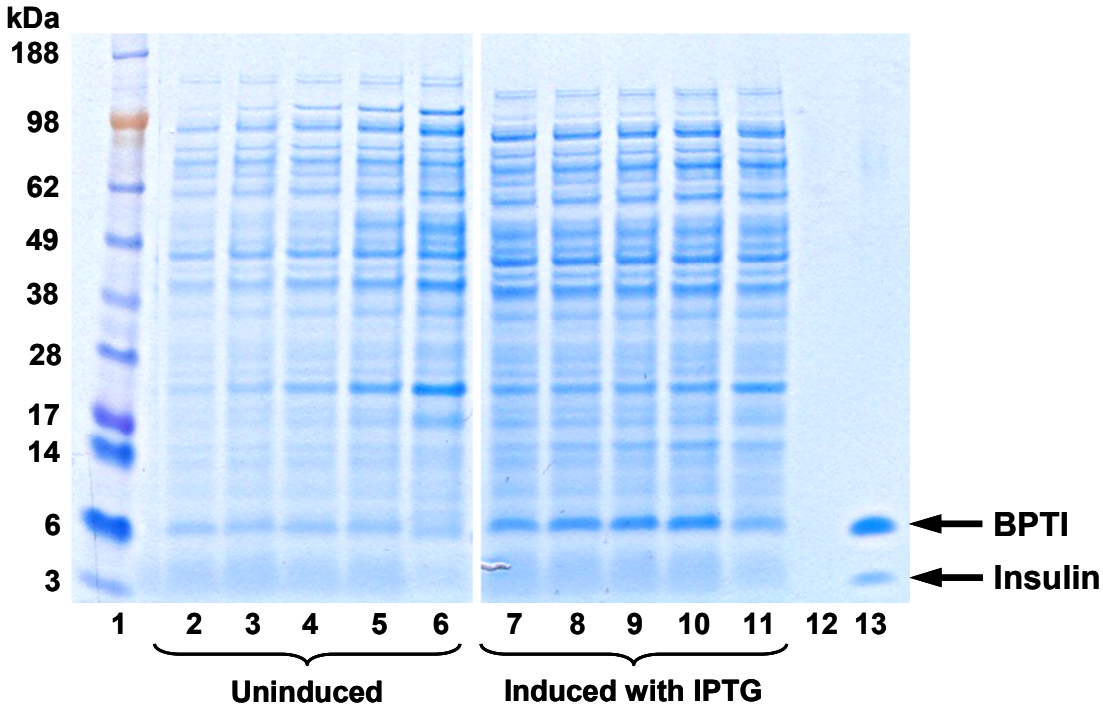


Figure 3.4: SDS-PAGE of a timecourse of wild type BPTI cultures expressed both in the absence and presence of IPTG. Lane 1, SeeBlue marker; Lanes 2-6, uninduced culture; lanes 7-11, culture induced with 1 mM IPTG. Aliquots of each culture were taken at various times after induction of the appropriate sample: 3 hours (lanes 2 and 7); 4 hours (lanes 3 and 8); 5 hours (lanes 4 and 9); 6 hours (lanes 5 and 10); overnight (lanes 6 and 11). As a control, a sample was prepared containing 5 µg of BPTI plus 5 µg of insulin. Since expression cultures did not express proteins below the molecular weight of BPTI, insulin, which gets reduced to separate α and β chains of approximately 3 kDa each, was used to verify that the bands observed at 6 kDa were BPTI, rather than a mixture of smaller proteins that were unable to be resolved in the gel. Clearly, the culture induced with IPTG shows larger levels of BPTI expression than the uninduced culture. However, small levels of BPTI expression are observed even in the uninduced sample. In the induced sample, there is little difference in amount of BPTI produced between 3-6 hours post-induction (lanes 7-10), with protein degradation occurring if the culture is left overnight (lane 11).

To further optimise their expression, small cultures of wild type BPTI and the C38S C55S mutant were expressed and analysed as described earlier, except

that the optical density at which cultures were induced with IPTG was varied. The SDS-PAGE (Figure 3.5) shows that both constructs had very similar levels of expression, irrespective of the OD_{600} at induction, between values of 0.3 to 0.6.

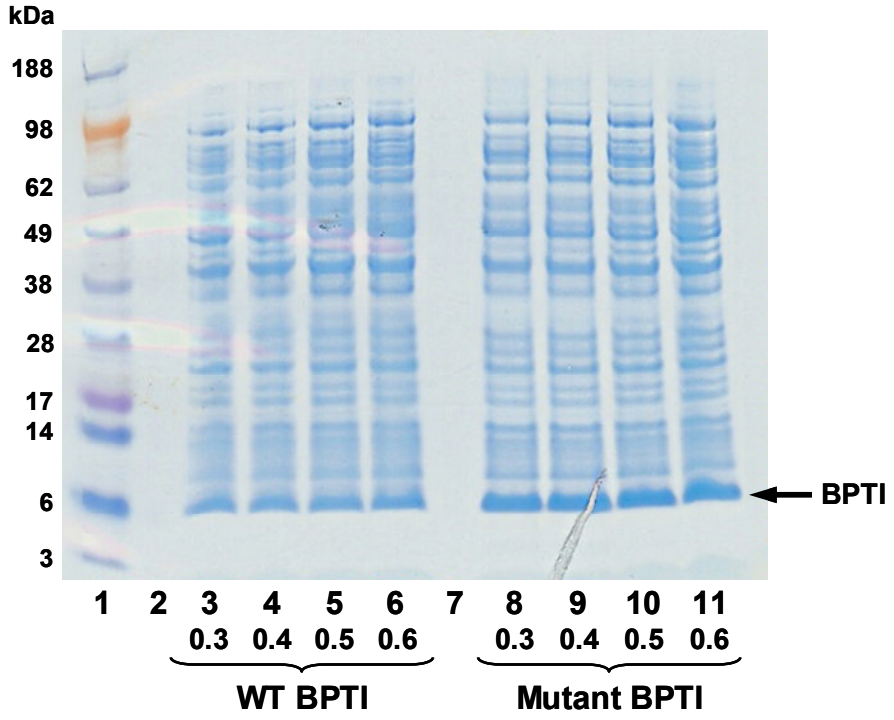


Figure 3.5: SDS-PAGE of cultures expressed after induction with IPTG at various optical densities. Lane 1, SeeBlue marker; lanes 3-6, wild type BPTI induced with IPTG at OD_{600} of 0.3 (lane 3), 0.4 (lane 4), 0.5 (lane 5) and 0.6 (lane 6). Lanes 8-11, C38S C55S mutant BPTI induced with IPTG at OD_{600} of 0.3 (lane 8), 0.4 (lane 9), 0.5 (lane 10) and 0.6 (lane 11). This indicates very similar expression levels at all of the different optical densities.

3.2.4. Isolation of Inclusion Bodies

All BPTI constructs were expressed as inclusion bodies. The protocol for expression and isolation of inclusion bodies was developed by the lab of Lloyd Ruddock (University of Oulu, Finland). Protocols for each purification step were developed as part of this study.

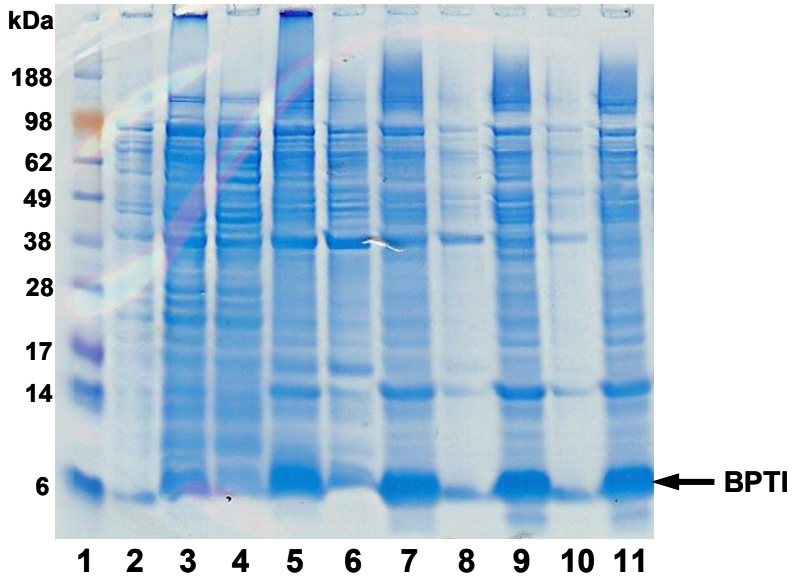


Figure 3.6: SDS-PAGE for isolation of wild type BPTI inclusion bodies. Lane 1, SeeBlue marker; lane 3, total expressed protein; lane 4, total soluble protein; lane 5, total insoluble protein; lane 6, supernatant after 1st wash; lane 7, pellet after 1st wash; lane 8, supernatant after 2nd wash; lane 9, pellet after 2nd wash; lane 10, supernatant after water rinse; lane 11, pellet after water rinse. Large bands are seen around 6.6 kDa in each of the pellet samples, indicating that the BPTI inclusion bodies have been successfully isolated.

Once expressed, the insoluble inclusion bodies needed to be purified from other insoluble cell debris. This followed the protocol defined in section 2.4.1. Figure 3.6 shows SDS-PAGE analysis of the isolation of inclusion bodies for wild type BPTI. All samples containing total expressed protein or pellet (after resuspension) show clear bands representing the desired protein, whereas samples containing the discarded supernatant (soluble proteins) show only very small amounts of BPTI. After sonication and centrifugation, the insoluble protein pellet was resuspended in a buffer containing detergent. This assisted in breaking up and solubilising lipid membranes and membrane proteins, thus separating them from the insoluble inclusion bodies during centrifugation. Although most detergent was discarded in the supernatant after centrifugation, it was important to ensure any trace detergent that may have remained in the inclusion body pellet was removed before proceeding, since the detergent could

bind to exposed hydrophobic regions of BPTI and interfere with downstream folding of the protein. To ensure all traces of detergent were removed, the inclusion bodies were resuspended in water twice, pelleting the inclusion bodies by centrifugation after each resuspension and discarding the supernatant. The isolated inclusion bodies mostly consist of BPTI, but some impurities persist.

3.2.5. Solubilisation of Inclusion Bodies and Desalting of BPTI

Once isolated, the inclusion bodies were solubilised using denaturant and reducing agent, as described in section 2.4.2. This caused the protein molecules to separate into monomers of reduced BPTI. The protein would have degraded rapidly if left in strong denaturing conditions, so it was buffer exchanged to 10 mM HCl using a desalting column. The low pH of this solution retained the protein in a reduced state. Figure 3.7 shows a typical chromatogram from the desalting of the solubilised inclusion bodies. Protein elution was detected by a peak in absorbance at 280 nm, with corresponding fractions being pooled. The salt is eluted later and is detected in the chromatogram by a peak in conductivity. SDS-PAGE was used to confirm the desired molecular weight of the protein (Figure 3.7, inset).

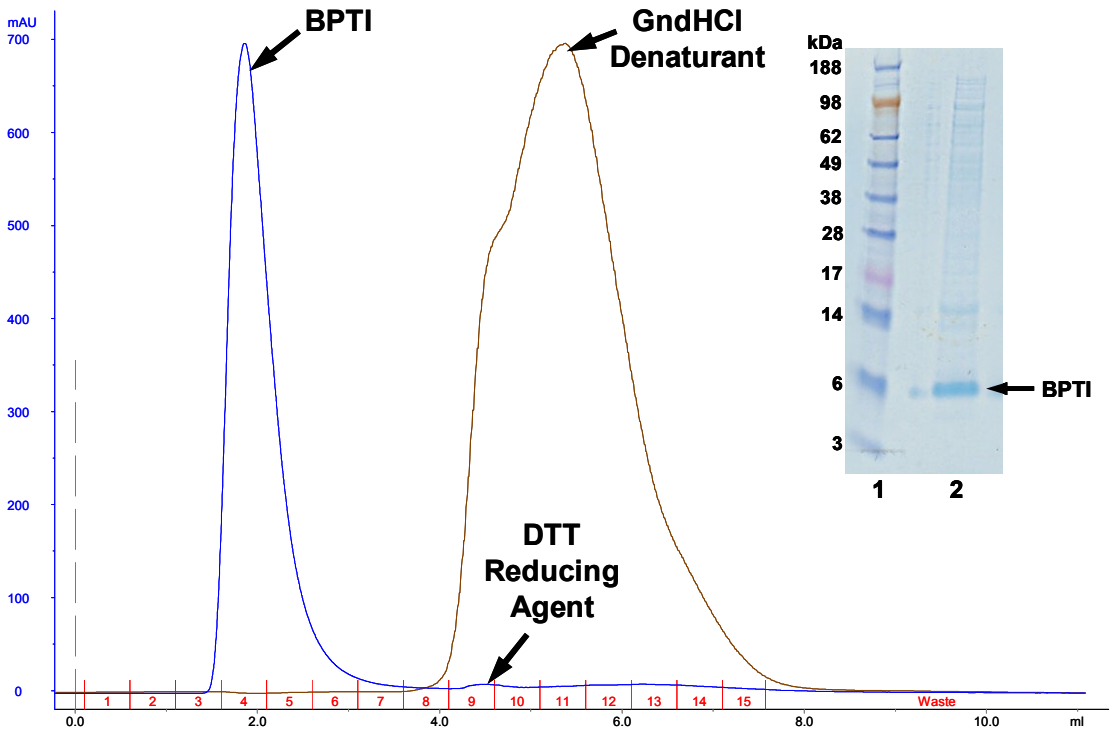


Figure 3.7: Chromatogram showing desalting of wild type BPTI using HiTrap desalting column. Absorbance at 280 nm (blue) indicates protein elution; high conductivity (brown) indicates the elution of salt. Inset: SDS-PAGE of eluted protein. Lane 1, SeeBlue marker; lane 2, pooled fractions 4 and 5 corresponding to A_{280} peak of the chromatogram.

3.2.6. Verification of Reduced BPTI by Mass Spectrometry

In order to confirm the protein was reduced BPTI, aliquots were analysed using ESI mass spectrometry. Based on the primary sequence of the protein, the molecular mass of the reduced protein was calculated as 6649 Da (using ProtParam, http://www.expasy.ch/tools/protparam.html). Standard mass spectrometry techniques can be used to identify if the main species is of the correct mass.

Once fully oxidised, the thiol groups of six cysteine residues would be involved in the formation of three disulfide bonds, resulting in the removal of six hydrogen atoms. Thus, the molecular mass of fully oxidised BPTI would be expected to be

six Daltons less than fully reduced wild type BPTI. Intermediates, which are partially oxidised, will have a difference of just two Daltons for each disulfide bond. Since this difference in mass is relatively small, it may be difficult to resolve peaks in a spectrum using standard mass spectrometry. To get a clearer idea of the success of reduction and oxidation reactions, an alkylation reaction adds to the overall mass for each free thiol present. Here, iodoacetamide was used for the alkylation process. The reaction is illustrated in Figure 3.8. Details are outline in Section 2.7.

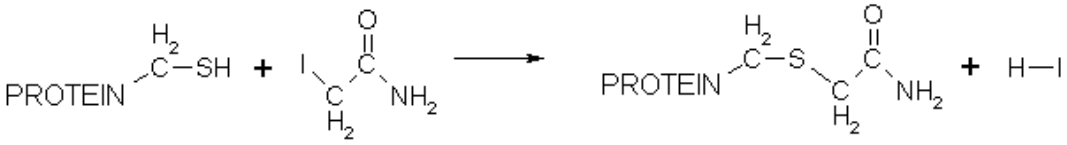


Figure 3.8: Iodoacetamide alkylation reaction.

Iodoacetamide has a molecular mass of 185.0 Da. When reacted with a free thiol group, iodine plus one hydrogen are removed. The addition in mass for each thiol group will be 185.0 Da – 126.9 Da (iodine) – 1 Da (hydrogen) = 57.1 Da. So, for fully reduced wild type BPTI, six free thiols will result in 57.1 Da x 6 = 342.6 Da added to the overall mass of the protein. Figure 3.9 shows mass spectra for aliquots of fully reduced wild type BPTI both unreacted (black) and after the iodoacetamide reaction (green). The correct difference in mass indicates that the majority of the protein has been fully reduced, with a small amount appearing to have only four free thiols.

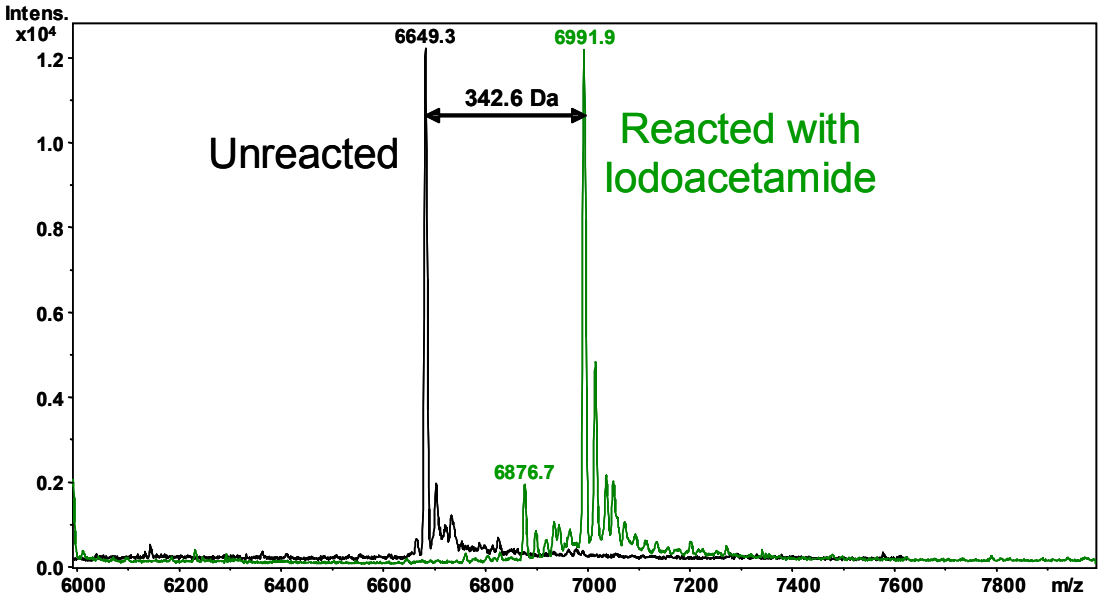


Figure 3.9: Deconvoluted ESI mass spectra of fully reduced wild type BPTI. The unreacted WT (black) is overlaid with a spectrum after an iodoacetamide reaction (green), confirming the reduced state of the protein.

3.2.7. Oxidation and Refolding of BPTI

Once reduced protein had been eluted from the desalting column, it was ready to be oxidised. The details of the oxidation reaction are specified in section 2.4.4. Since this reaction was performed in the presence of reducing as well as oxidising reagent, the process of isomerisation could also take place. This allowed refolding of each BPTI construct to continue until it could go no further along the folding pathway (see Figure 1.12). In the case of wild type BPTI, it refolded to the native structure. Figure 3.10 shows SDS-PAGE analysis of wild type BPTI before and after refolding. In this case, the reduced protein had been stored at -20°C, so needed to be thawed before use. Some aggregation of BPTI occurred during the refolding reaction, but this process had the advantage of causing remaining impurities also to aggregate, such that only very pure, fully oxidised BPTI remained. Centrifugation was used to pellet any aggregated protein and the supernatant (lane 7) was then taken forward.

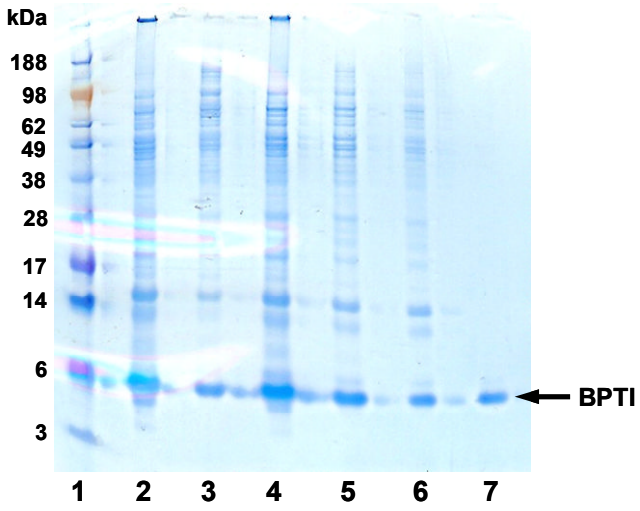


Figure 3.10: SDS-PAGE of wild type BPTI before and after refolding. Lane 1, SeeBlue marker; lane 2, total reduced; lane 3, reduced pellet; lane 4, reduced supernatant; lane 5, total refolded; lane 6, refolded pellet; lane 7, refolded supernatant.

By removing aliquots of the refolding reaction at different times, the progress of refolding was monitored. An alkylation reaction was used to prevent further oxidation and identify the oxidation state of each sample (section 0). Figure 3.11 shows mass spectra of wild type BPTI at various times through the refolding reaction. After 1 hour of oxidation, about 60% of the protein was completely oxidised (three disulfides), with most of the rest in the two disulfide state. This two disulfide species diminished and the fully oxidised species became increasingly dominant at later time intervals. After 4 hours, the two disulfide species disappeared completely (not shown). Allowing the oxidation reaction to proceed overnight at 4°C resulted in full oxidation of the protein.

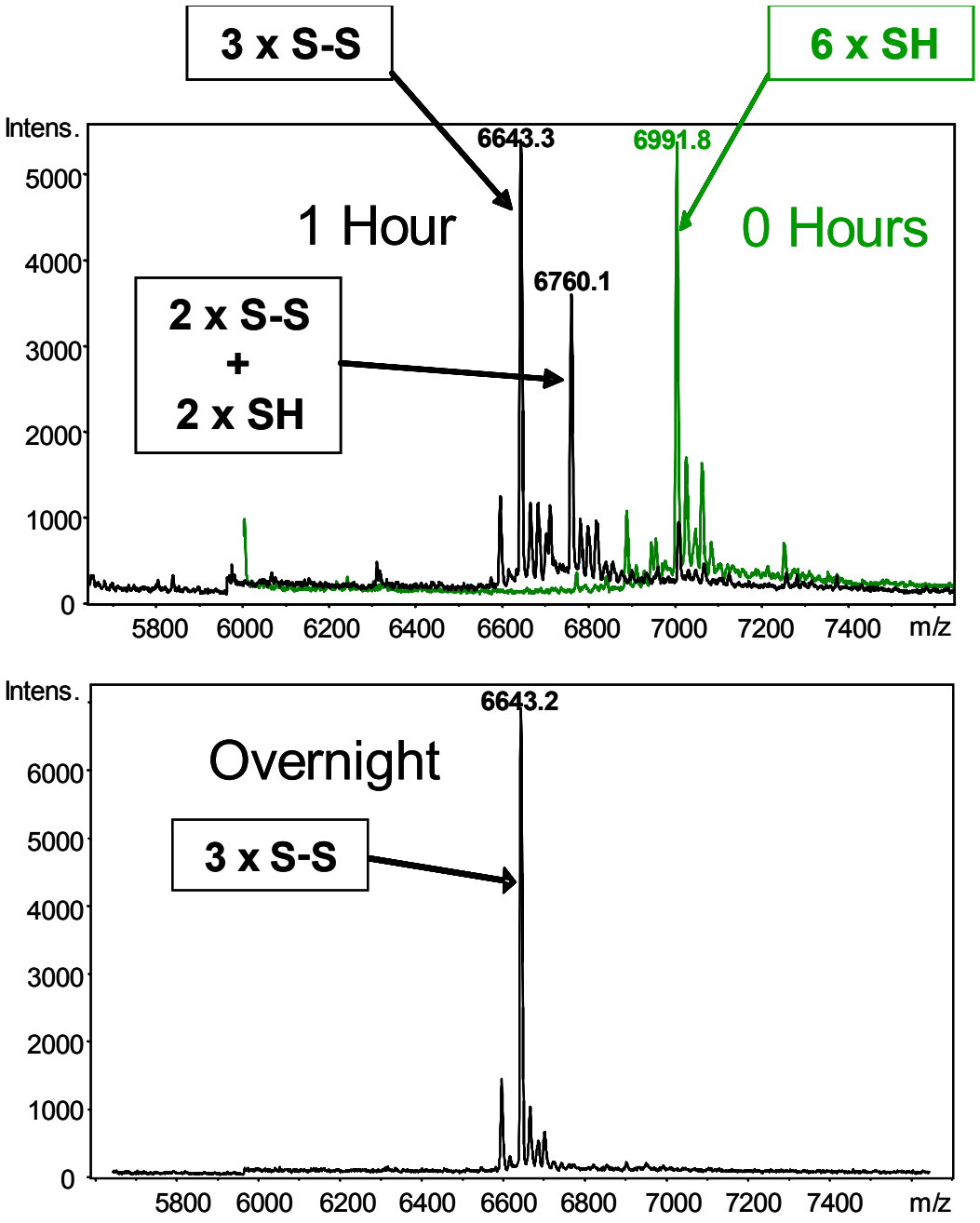


Figure 3.11: Deconvoluted ESI mass spectra of wild type BPTI at different stages of oxidation, after iodoacetamide reaction. The protein is fully reduced (top, green) at the start. After 1 hour, some protein has fully oxidised (top, black), but a significant proportion of the two disulfide species remains. When incubated overnight, all of the protein has fully oxidised (bottom).

3.2.8. Post-refolding Purification Using Solid Phase Extraction Chromatography

The refolded BPTI was further purified using solid phase extraction (SPE) chromatography. This served the following purpose:

- Removal of oxidising and reducing agents
- Concentrating of diluted protein
- Preparation for lyophilisation
- Further purification

The procedure was carried out as specified in section 2.4.5. Protein was eluted in 50% acteonitrile, which is suitable for downstream lyophilisation. SDS-PAGE analysis of steps through the purification process of wild type BPTI is shown in Figure 3.12. BPTI is observed before loading onto SPE column (lane 2) and again in the first elution fraction during SPE purification (lane 5).

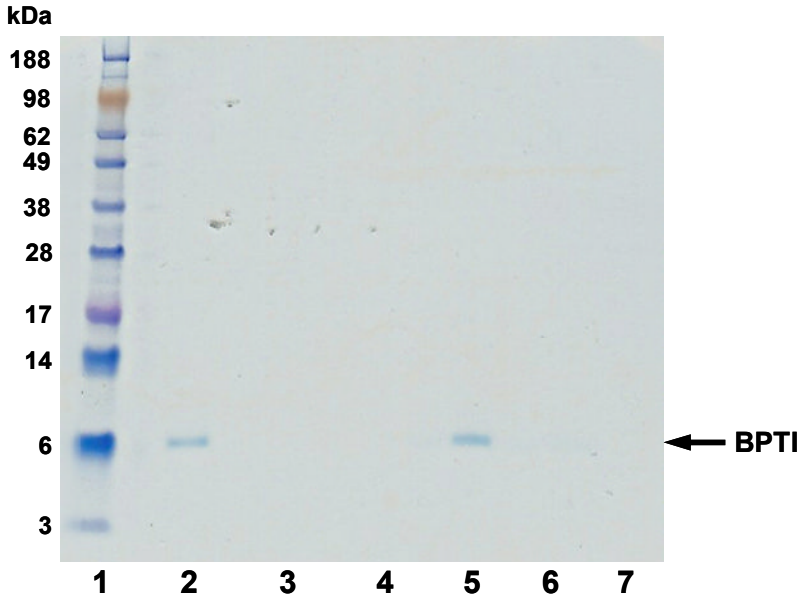


Figure 3.12: SDS-PAGE of SPE purification of refolded wild type BPTI. Lane 1, SeeBlue marker; lane 2, supernatant of refolded BPTI; lane 3, flow through from SPE; lane 4, rinse from SPE; lane 5, 1st elute from SPE; lane 6, 2nd elute from SPE; lane 7, 3rd elute from SPE.

Since different experiments require different solvent conditions and to avoid loss through the freeze/thaw cycle, BPTI was lyophilised prior to storage at -20°C. It has previously been shown that native BPTI can be lyophilised and resuspended without adverse affect on its structure (Desai, Osterhout et al. 1994). Samples from SPE elution were vacuum evaporated to remove the acteonitrile then lyophilised as described in section 2.4.6. Mass spectrometry of lyophilised samples verified that neither SPE nor lyophilisation affect the integrity of the samples (Figure 3.17).

3.2.9. Comparison of Recombinant Wild Type BPTI with Commercial BPTI

Before studying modified versions of BPTI it was important to establish if the recombinant wild type protein was similar to the commercially available product. Circular dichroism was performed with both commercially available trypsin inhibitor (purified from bovine pancreas) (Sigma) and with the recombinant wild type BPTI, which had been lyophilised as explained above.

Figure 3.13 shows that the commercial and recombinant proteins have very similar spectra. Dichroweb analysis showed the commercial product containing 0.28 α-helix; 0.33 β-sheet; 0.39 random coil (K2D method, NRMSD = 0.070). Likewise the recombinant protein contained 0.27 α-helix; 0.33 β-sheet; 0.41 random coil (K2D method, NRMSD = 0.114) (Whitmore and Wallace 2008). These results are also very similar to previous CD studies of native and refolded BPTI, as shown in Figure 3.14.

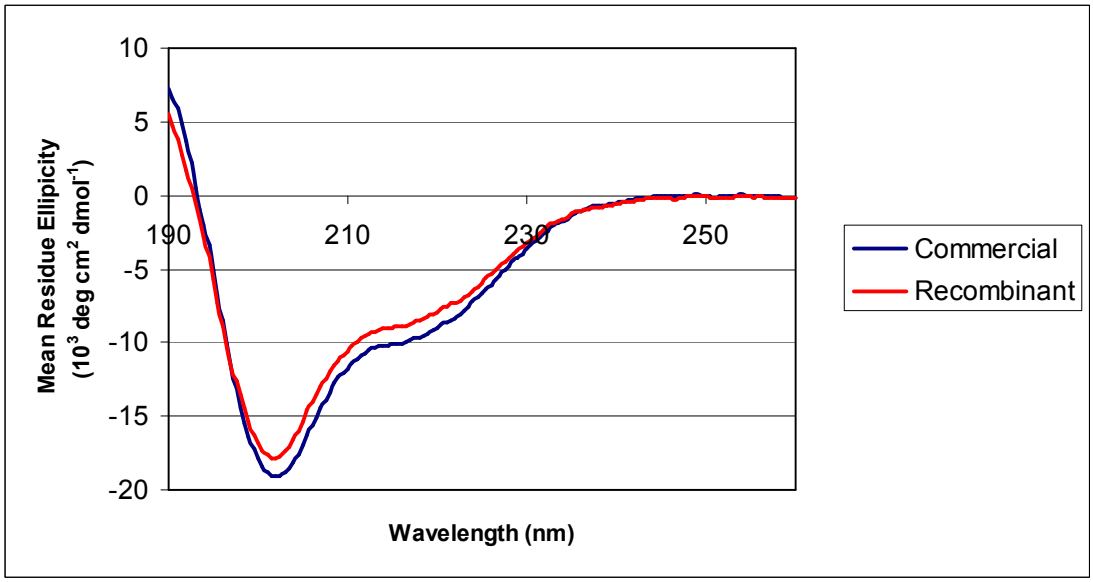


Figure 3.13: Far-UV circular dichroism of commercial (blue) and recombinant (red) wild type BPTI.

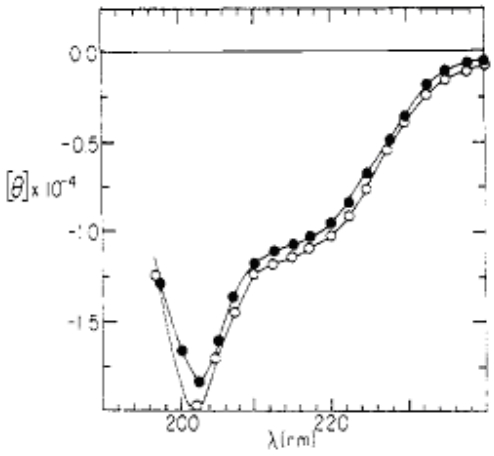


Figure 3.14: Far-UV circular dichroism of BPTI, showing the native protein (open circles) and the refolded protein (closed circles) (Kosen, Creighton *et al*. 1981).

The structure of recombinant wild type BPTI was also compared against commercial BPTI using NMR spectroscopy, see section 4.2.2 for details.

3.2.10. Expression and Purification of ^{15}N Labelled BPTI

Expression and purification of ^{15}N labelled wild type BPTI was identical to the processes for unlabelled protein, except that the cultures were grown in minimal medium using ^{15}N labelled ammonium sulphate (section 2.3.2).

Wild type BPTI contains 85 nitrogen atoms, so the overall mass would be expected to be approximately 85 Da larger than the unlabelled protein. Figure 3.15 shows ESI mass spectra of ^{15}N labelled wild type BPTI before oxidation, with and without alkylation. The unalkylated and alkylated samples show dominant species of molecular masses that are 82.5 Da and 83.0 Da respectively larger than the unlabelled protein (compare with Figure 3.9). This suggested that the protein was approximately 98% ^{15}N isotopically labelled, as expected.

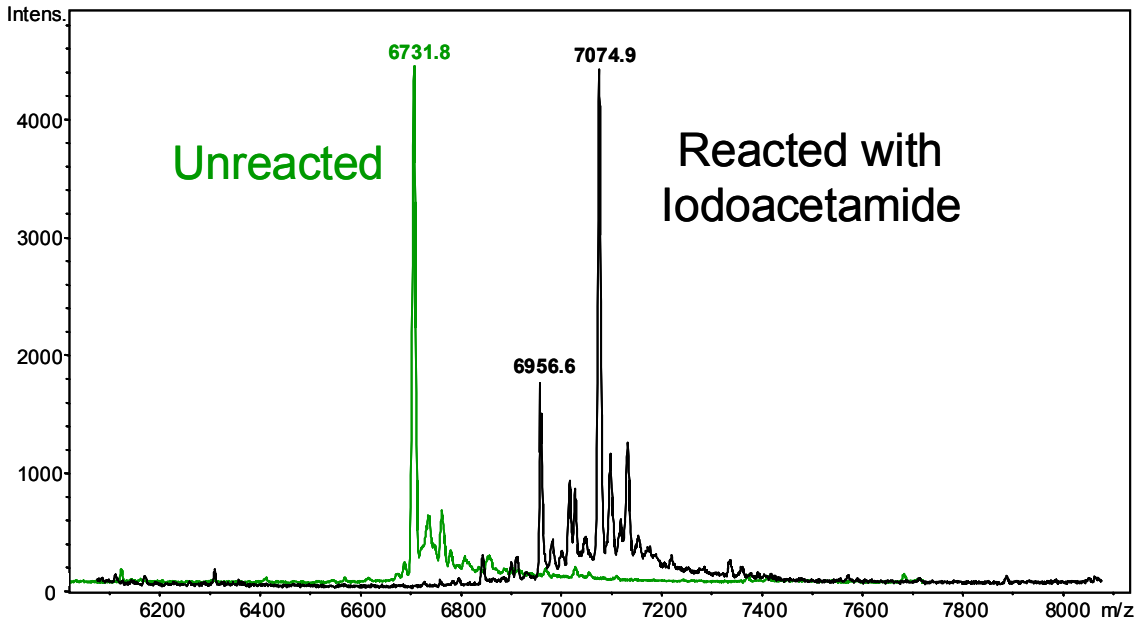


Figure 3.15: ESI mass spectra of reduced ^{15}N labelled wild type BPTI. Reacted (black) and unreacted (green) with iodoacetamide.

3.2.11. Trapping of Wild Type BPTI in the Reduced State

Fully reduced proteins are useful in the study of folding pathways, since they represent the starting point from which folding intermediates are derived, and are expected to have a high affinity to folding enzymes. Following desalting of solubilised inclusion bodies, wild type BPTI was trapped in the fully reduced state by alkylation of free thiol groups. Iodoacetamide was used as the alkylating agent in a process similar to that used for mass spectrometry analysis. However, conditions were optimised for preparative scale alkylation of the reduced BPTI (section 2.7.1). Since oxidation was not required, the alkylated reduced BPTI was then purified directly using solid phase extraction (section 2.4.5), then lyophilised (section 2.4.6) in the same manner as oxidised BPTI. ^{15}N labelled reduced BPTI followed the same process as the unlabelled protein, except that it was prepared using ^{15}N labelled wild type BPTI. Success of the preparation of samples was confirmed using mass spectrometry (data not shown).

One problem identified when lyophilised reduced BPTI was resuspended was the low solubility of the protein. It had a maximum solubility of just 100 μM in the NMR buffer, compared to the fully oxidised wild type which is still soluble at >1 mM. Significantly lower solubility is expected for a fully reduced protein, since it will have many more exposed hydrophobic regions and thus a higher tendency to aggregate.

3.2.12. Expression and Purification of (30-51, 5-14) BPTI Intermediate

Expression and purification of (30-51, 5-14) BPTI followed the same protocol as for wild type BPTI (sections 2.1 and 2.4). Likewise ^{15}N labelled protein was prepared using minimal medium.

Figure 3.16 shows a chromatogram for the desalting of ^{15}N labelled (30-51, 5-14) BPTI from solubilised inclusion bodies. The A_{280} signal indicates the early elution of the protein, possibly in several different conformations due to multiple peaks. The denaturant, indicated by high conductance, is eluted later along with the DTT reducing agent and small molecule contaminants.

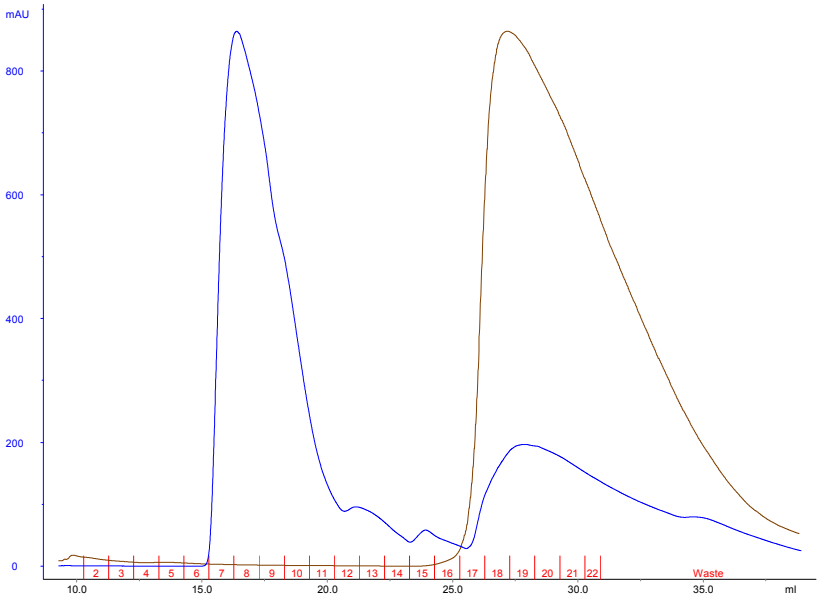


Figure 3.16: Chromatogram from desalting of ^{15}N labelled (30-51, 5-14) BPTI. Blue, absorbance at 280 nm; brown, conductance; red, eluted fractions.

The refolding of (30-51, 5-14) BPTI followed the same protocol as for wild type BPTI, but due to the reduced number of free thiol groups available (due to Cys to Ser mutations), the oxidation (and hence the refolding) can only proceed to a certain level, as defined by the folding pathway (Figure 1.12). Mass spectrometry confirmed successful oxidation and purification of the mutant sample, both unlabelled (Figure 3.17) and ^{15}N labelled (not shown).

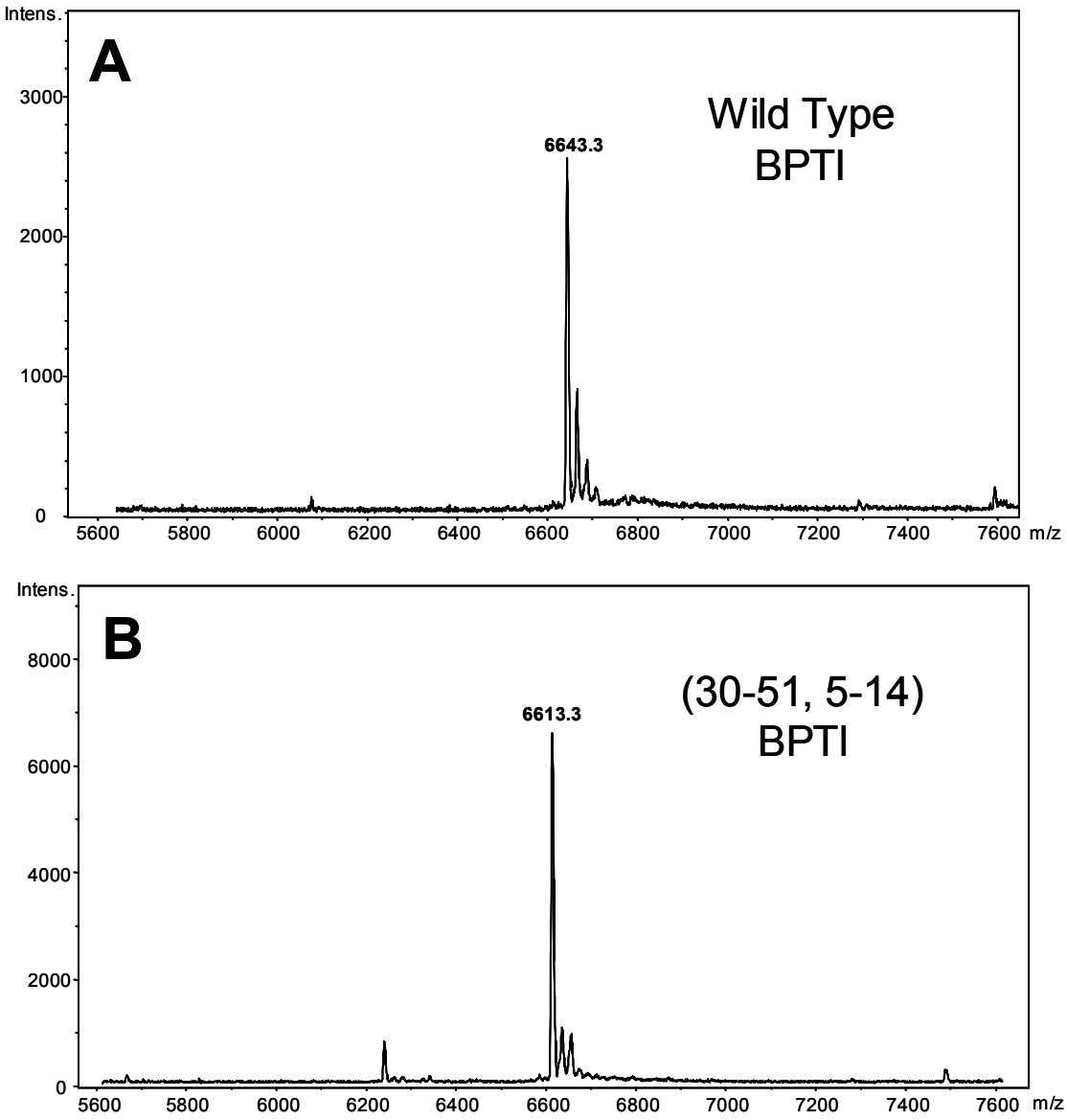


Figure 3.17: ESI mass spectra of fully oxidised BPTI. A) Wild type BPTI; B) (30-51, 5-14) BPTI, which contains C38S and C55S mutations.

3.2.13. Comparison of Different Constructs

As well as analysing BPTI constructs in isolation, it was also useful to compare the different constructs, so that any expected differences are observed. Figure 3.17 shows mass spectra for both wild type and (30-51, 5-14) BPTI after being fully oxidised. A difference of 30 Da is observed between the dominant peaks of

each spectrum. In the (30-51, 5-14) protein, two Cys to Ser mutations were made. At an atomic level, each Cys to Ser mutation is effectively equivalent to replacing a sulfur atom (32 Da) with an oxygen atom (16 Da). For two such mutations, this would result in a difference of 2 x 16 = 32 Da. However, in the fully oxidised protein, disulfide bond formation results in a single hydrogen atom being removed from each of the Cys involved, resulting in a reduction of 2 Da in mass for each disulfide bond. So the expected mass difference between wild type BPTI (3 disulfides) and (30-51, 5-14) BPTI (2 disulfides) was 30 Da, just as was observed from the mass spectra.

A comparison of secondary structure content between the various BPTI constructs can be observed using circular dichroism. Figure 3.18 shows CD spectra for the three constructs of BPTI. Some differences between the spectra are visible. However, quantitative analysis of secondary structure was unreliable, since reference data for unfolded proteins is not available (data not shown). This indicates that far-UV CD is not very useful for detecting differences between the different BPTI constructs, since large amounts of random coil secondary structure in this small protein make differences difficult to discriminate.

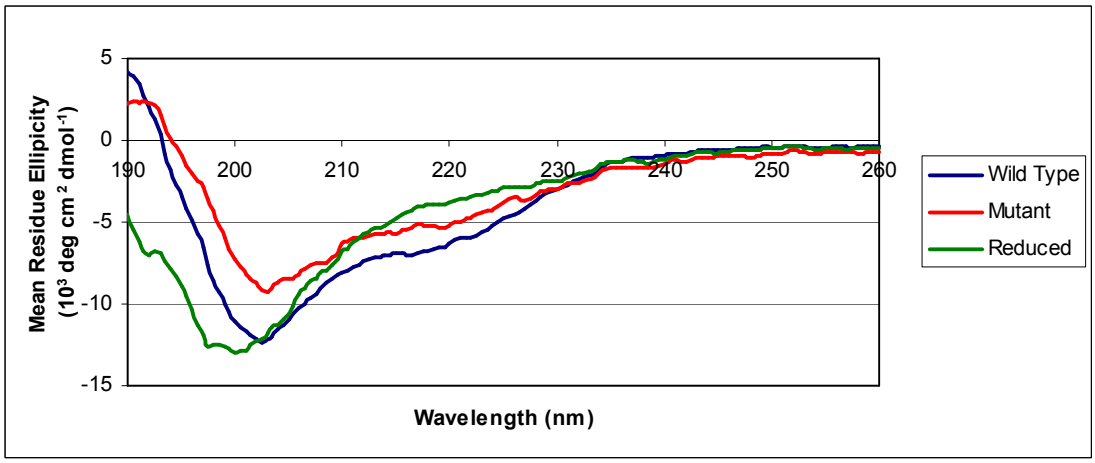


Figure 3.18: Far-UV circular dichroism of BPTI constructs at various stages of the folding pathway, all acquired at 25°C. Blue, fully oxidised wild type BPTI; red, (30-51, 5-14) mutant BPTI; green, reduced and alkylated wild type BPTI.

Circular dichroism of ^{15}N labelled proteins showed no significant difference in secondary structure compared to the unlabelled equivalent proteins, indicating that labelling did not significantly affect secondary structure (data not shown).

3.3. Discussion

This chapter has outlined the processes involved in the preparation of BPTI samples. The mutations required to produce various partly-folded intermediates were explained. Test expressions revealed that the (30-51, 5-14) two disulfide was the only one which showed acceptable levels of expression.

A protocol was established for both pre- and post-refolding purification of BPTI from inclusion bodies. The concentration and volumes of denaturant and reducing agent were optimised to ensure solubilisation of inclusion bodies in as small a volume as required. After exploring different purification techniques, including reverse phase purification using high performance liquid chromatography (HPLC), it was found that a desalting column provided the best option for further purification of the reduced constructs. Refolding was performed by dilution into a refolding buffer with oxidising and reducing agents. A number of options were also explored for treatment of the protein after refolding. Solid phase extraction is relatively uncommon for the purification of macromolecules such as proteins, but a suitable column and procedure was developed for the purification of BPTI. Volatile solvent was removed from the eluted protein by vacuum evaporation, before finally the protein could be lyophilised and stored at -20°C. Although the protocol was initially developed using the wild type protein, it was also sufficient for the purification of both (30-51, 5-14) and fully reduced BPTI.

The correct molecular masses of purified constructs were verified using ESI mass spectrometry. The oxidation state was confirmed by alkylation of free thiols with iodoacetamide prior to mass spectrometry.

3.3.1. Summary of Purified BPTI Constructs

This chapter has outlined the expression, refolding (where appropriate) and purification of three BPTI constructs representing different stages along this protein's folding pathway:

- Wild type BPTI, representing the natively folded mature protein, with three disulfide bonds. This will be referred to as wild type BPTI.
- A partly-folded BPTI, specifically the (30-51, 5-14) two disulfide bonded intermediate, which contains one native and one non-native disulfide bond. This will be referred to as (30-51, 5-14) BPTI.
- Fully unfolded BPTI, represented by the fully reduced and alkylated wild type mature protein. This will be referred to as reduced BPTI.

Chapter 4. NMR Spectroscopy of BPTI Constructs

4.1. Introduction

Before analyses of the interaction of each BPTI construct with PDI, a study of the structural characteristics of each recombinant construct in isolation was performed. Several NMR studies have already been performed with BPTI, with entries in the PDB data bank (Berndt, Guntert et al. 1992) and the Biological Magnetic Resonance Data Bank (BMRB) (Biamonti 1996). First, the recombinant wild type BPTI was compared to commercially available wild type BPTI. This allowed us to verify if the recombinant protein is indeed representative of the naturally occurring mature protein. Full assignment of the wild type protein was then performed. The compatibility of BPTI to conditions under which PDI has been studied was also investigated, so that interaction studies would be viable. NMR characterisation of (30-51, 5-14) BPTI and reduced BPTI followed in a similar manner, optimising conditions so that they are most informative for each construct. All NMR spectra shown focus on the fingerprint region, at ^{1}H between 6-11 ppm, where most information about the protein backbone can be elucidated.

4.2. NMR of Wild Type BPTI

4.2.1. TOCSY/NOESY NMR of Commercial BPTI at 36°C

The first stage in the process was to run ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY NMR experiments of commercially available native mature BPTI. These experiments served as a suitable starting point since they observe only proton resonances

and do not require any isotopic labelling of samples. NMR data was available from a previous study of commercially available mature BPTI, submitted to the BMRB data bank (Entry 5359) (Biamonti 1996). This study used 750 µM of BPTI solvated in 25 mM calcium chloride, 3 mM sodium azide, 90% H_2O, 10% D_2O, pH 5.8. Commercial BPTI used for this study was resuspended in the same solvent at the same pH so that results could be directly compared. Figure 4.1 shows an overlay of TOCSY and NOESY spectra recorded for commercial BPTI. In order to facilitate assignment of peaks, the appropriate ^1H data for assignments from the previous study were imported into the corresponding spectra. Since prolines do not have amide protons, these residues were not detected. The first residue, Arg1, was not available in the imported data. Interestingly, there is one residue that is not within the fingerprint region at all. The amide ^1H resonance for Gly37 is actually found at approximately 4 ppm. Inspection of the published structure of BPTI revealed that Gly37 is strongly shielded by the aromatic side chain of Tyr35, resulting in a chemical shift far upfield from where it may usually be expected. Visual inspection of individual peaks in the acquired spectra showed the close proximity of the imported assignments in the majority of cases. This allowed efficient assignment of peaks in both the acquired TOCSY and NOESY spectra. N-terminal residues up to Phe4 were not identified from the spectra. Likewise, residues Thr11, Arg39 and Lys46 were not identified. All other residues were identified in both the TOCSY and NOESY spectra. An assignment table from TOCSY and NOESY spectra is available in Appendix A.

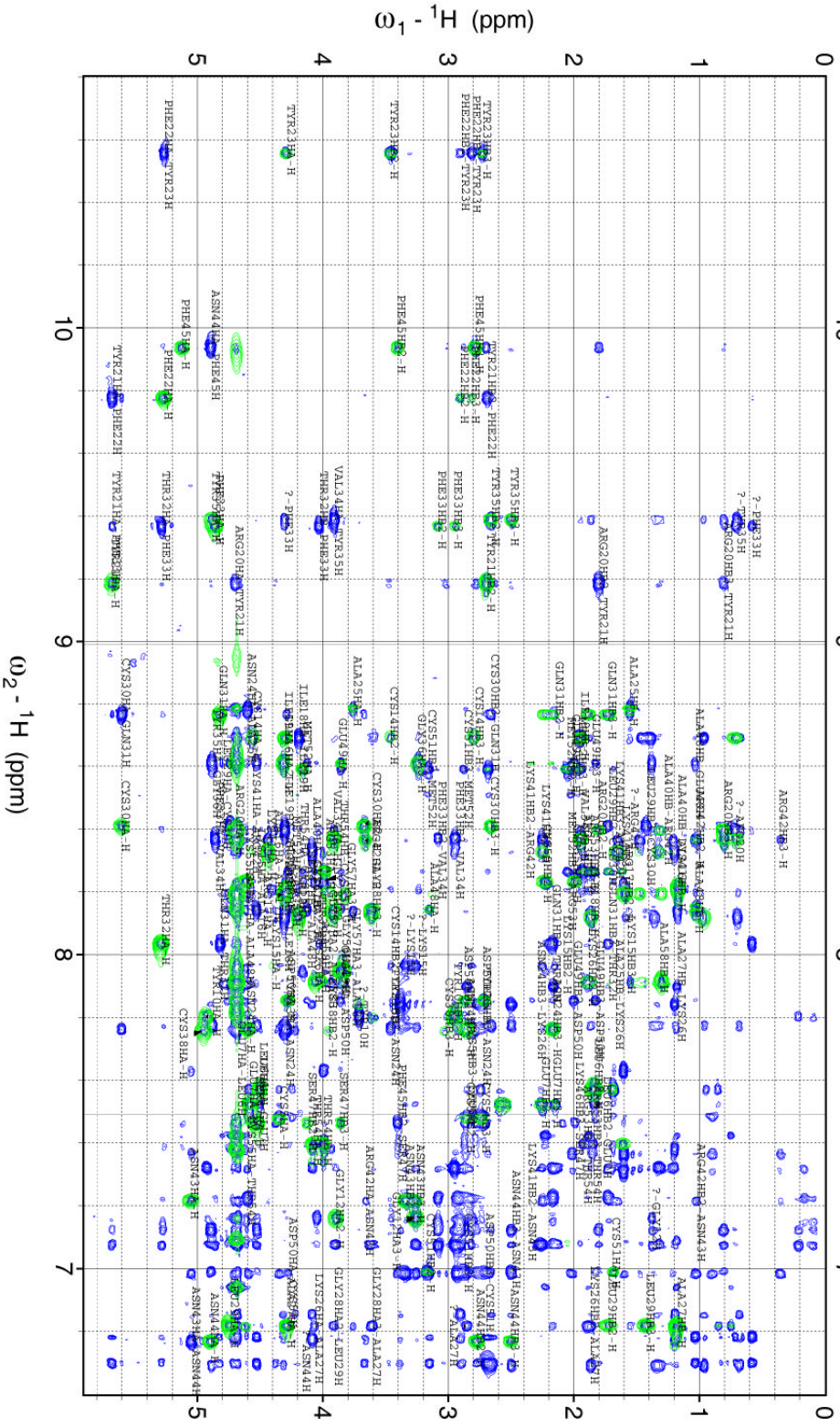


Figure 4.1: Overlay of TOCSY and NOESY spectra for commercial wild type BPTI. Green, TOCSY spectrum; blue, NOESY spectrum. NOESY assignments are labelled.

4.2.2. TOCSY/NOESY NMR of Recombinant Wild Type BPTI at 36°C

To ensure the recombinantly expressed wild type BPTI truly reflects the structure of the natural mature protein, its NMR spectra were compared with commercially available BPTI. This also confirmed that the purification and lyophilisation processes did not alter the protein conformation.

^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY NMR experiments were performed using the same solvent and with the same acquisition parameters as for the commercial protein.

Figure 4.2 shows an overlay of the recombinant and commercial TOCSY spectra. It shows a high degree of agreement between the two spectra, with all assigned peaks in the commercial protein spectrum also observed in the recombinant protein. This clearly indicates that the recombinant protein adopted the native state after purification and refolding and that its structure was not altered by any post-refolding treatment. Interestingly, some peaks that were not present in the commercial protein spectrum now appear, with relatively low intensity, in the spectrum for the recombinant protein. Comparing these peaks to assignments from previous NMR data of the mature protein allowed assignments of Asp3, Phe4, Thr11, Arg39 and Lys46. Overlay of commercial and recombinant BPTI NOESY spectra showed a similar pattern to the TOCSY spectra, with the same additional residues having their peaks assigned (not shown). This meant that only the prolines and Arg1 (plus the N-terminal Met used as a start codon) remained unassigned, with Gly37 outside the fingerprint region, as in the assignments of BMRB entry 5359 (Biamonti 1996). An assignment table from TOCSY and NOESY spectra is available in Appendix A.

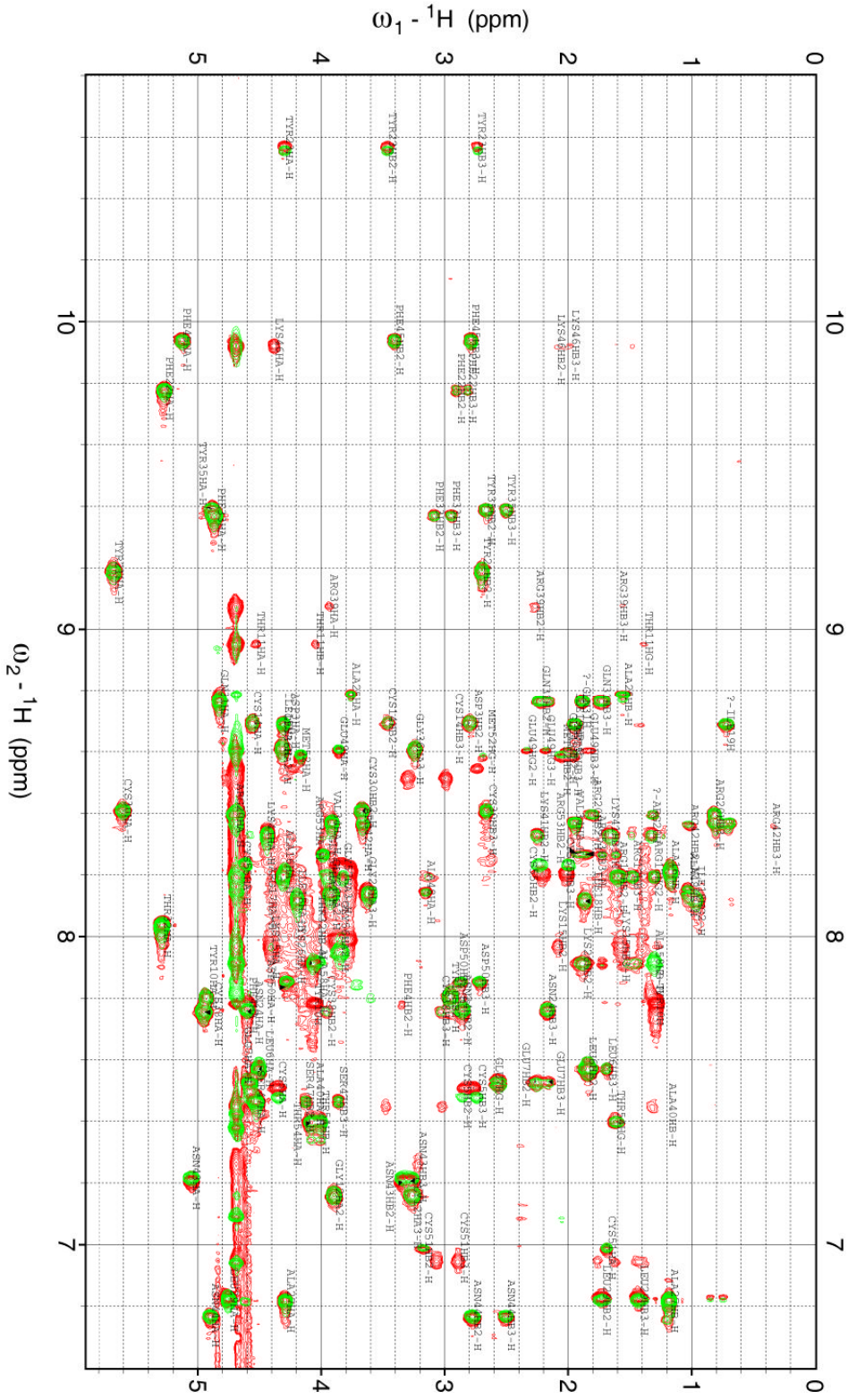


Figure 4.2: Overlay of TOCSY spectra from commercial (green) and recombinant (red) wild type BPTI samples. Assignments for the recombinant protein are labelled.

4.2.3. NMR Conditions Compatible with PDI

Since the main aim of this study was to investigate the structural interaction of PDI with BPTI it was important to establish a single NMR buffer that would be compatible with both proteins. Recent work with various PDI constructs used a buffer that gave optimal NMR spectra for these constructs (Nguyen, Wallis et al. 2008). The NMR buffer was 100 mM sodium chloride, 25 mM sodium phosphate, 90% H_2O, 10% D_2O, pH 6.5. This was used as the new buffer for NMR experiments with wild type BPTI, to observe if this protein was stable and yielded high quality NMR data under these conditions.

4.2.4. HSQC of Recombinant Wild Type BPTI at 36°C

^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) NMR experiments show interactions between the ^{15}N amide nitrogen atoms of the protein backbone and the hydrogen atoms that are covalently bonded to them. For a standard protein, this essentially results in a single resonance peak for each amino acid residue (except prolines), provided they remain in a single chemical environment. Some protein side chains that contain amide groups will also be visible.

^{15}N isotopically labelled wild type BPTI was prepared as outlined in section 2.3.2. An HSQC spectrum was then acquired using the parameters specified in section 2.12.6. Figure 4.3 shows the HSQC spectrum for wild type BPTI at 36°C.

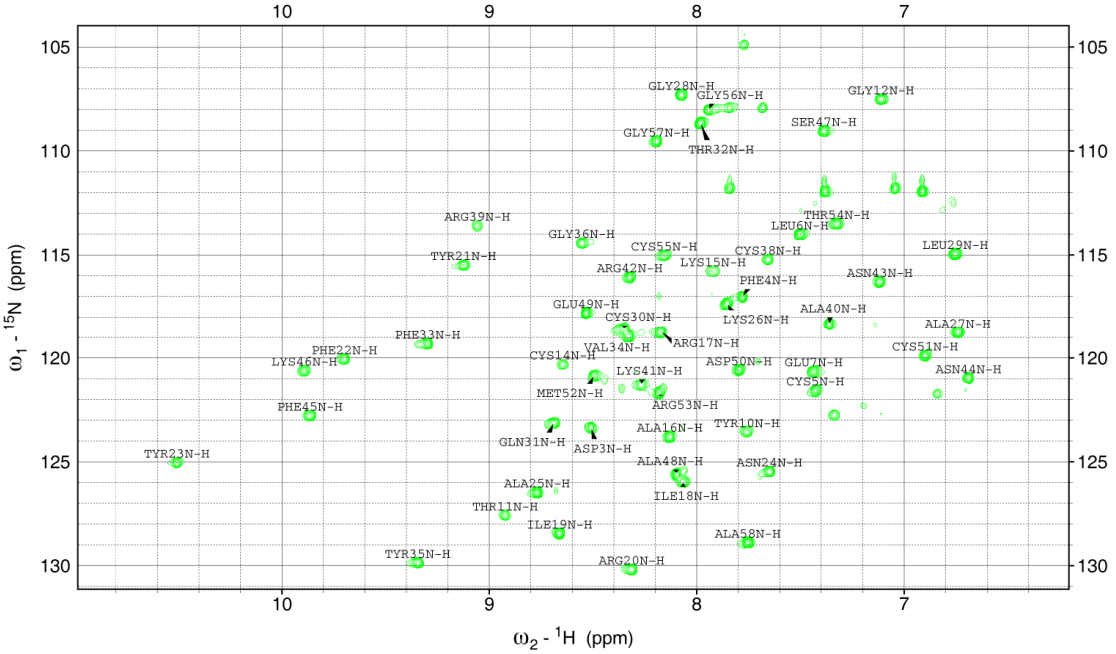


Figure 4.3: HSQC of recombinant wild type BPTI at 36°C.

The peaks of the spectrum are very well dispersed throughout the fingerprint region, with well resolved chemical shifts. This is indicative of a protein that is fully folded, whereby each residue is stable in a well defined and distinct chemical environment (Pascal 2008). Using both ^{15}N and ^{1}H data from previous NMR assignment of BPTI, BMRB entry 5359 (Biamonti 1996), assignment of peaks was performed. All residues were assigned except for the Arg1 and the prolines. Amides from side chains appear at a ^{15}N chemical shift of approximately 112 ppm and display a characteristic 'tear-drop' shape. A full list of these HSQC assignments is available in Appendix C.

4.3. NMR of (30-51, 5-14) BPTI Intermediate at Various Temperatures

NMR of wild type BPTI showed that the recombinantly expressed protein construct displays an almost identical structure to that of native mature BPTI. This further demonstrated that the expression, purification and refolding

processes yield the desired product, providing confidence that proteins representing alternative positions in the BPTI folding pathway will reflect the true structure of the folding intermediate.

Since a key aspect of this study is the interaction of BPTI with a folding enzyme, it is more desirable to investigate the structure of partly-folded or fully unfolded proteins, since they are the natural substrates of PDI. Using the same conditions as for the wild type, a ^{1}H-^{15}N HSQC was performed with (30-51, 5-14) BPTI, which is believed to be partly-folded. Figure 4.4 shows the HSQC spectrum for (30-51, 5-14) BPTI at 36°C. It shows much poorer dispersion of peaks, with most in the narrow range of 7.4-8.6 ppm. The accumulation of peaks in a narrow chemical shift range compared to the wild type spectrum of Figure 4.3 indicates that many of the residues are now in a very similar chemical environment. This is indicative of a less folded protein, whereby instead of strict positions in a well-defined structure, each residue is highly flexible and therefore able to experience many different chemical environments. On the NMR timescale, often only the average of these environments is observed, representing a single peak in the HSQC spectrum. In a completely unfolded and flexible protein, the averaged chemical shifts observed may be slightly different for each residue, due to the type of amino acid and its position in the polypeptide chain, which may impose certain steric hindrances. This results in a region of overlapping, poorly dispersed peaks that are impossible to distinguish. For (30-51, 5-14) BPTI, much of the spectrum in Figure 4.4 shows these characteristics of an unfolded protein. However, some peaks within the spectrum are clearly distinguishable, indicating that some residues have retained a unique chemical environment. This implies that some part or parts of (30-51, 5-14) BPTI remain folded, suggesting that the overall protein is indeed partly folded.

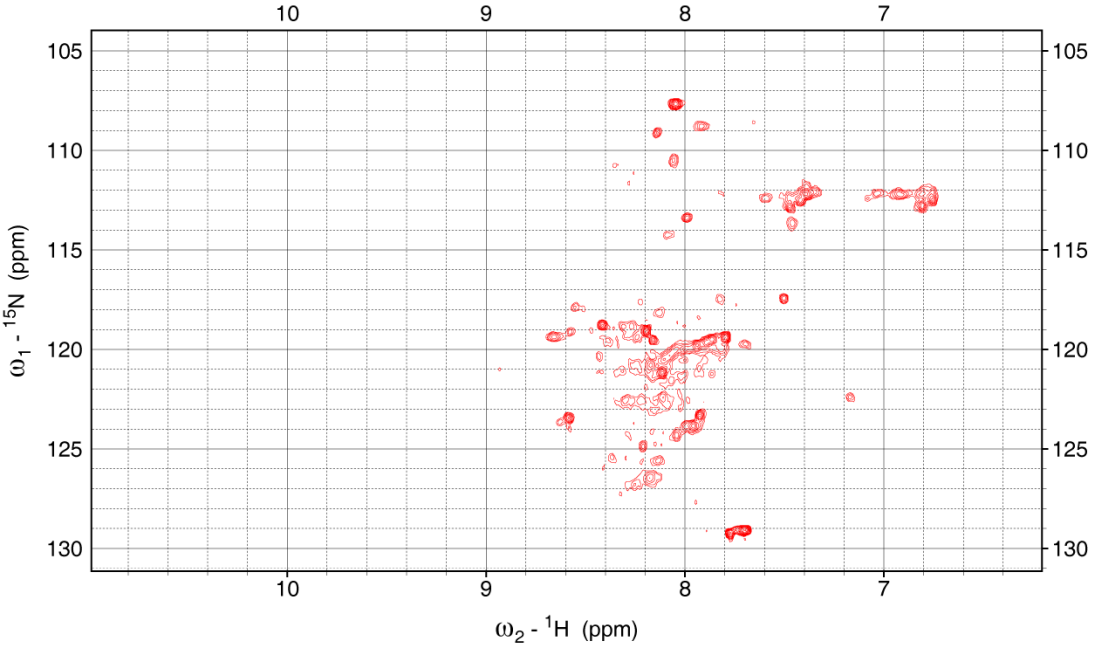


Figure 4.4: HSQC spectrum of (30-51, 5-14) BPTI at 36°C.

While Figure 4.4 shows strong evidence that (30-51, 5-14) BPTI represents a partly folded protein, it is not feasible to attempt to assign each of the peaks. However, such assignments are necessary in order to break down the behaviour of the protein to specific regions or key residues.

At first, this may seem to be an intractable problem. However, a previous study of similar partly folded BPTI constructs showed that increased resolution could be obtained for one dimensional ^{1}H NMR spectra by lowering the temperature (Vanmierlo, Kemmink et al. 1994). They proposed that lines broaden on increasing the temperature because of exchange between folded and unfolded molecules with an intermediate rate on the chemical shift timescale (Vanmierlo, Kemmink et al. 1994).

To explore the effect of temperature, a series of ^{1}H-^{15}N HSQC spectra were acquired for (30-51, 5-14) BPTI using the same parameters as before, except varying the temperature. Initially an HSQC spectrum was acquired at 36°C before the temperature was gradually lowered to acquire spectra at 30°C, 25°C,

20°C, 15°C, 10°C and 5°C. An increase in dispersion is clearly observed as temperature is reduced. Figure 4.5A shows that even at 20°C there is a significant gain in dispersion, with more distinct peaks visible. At 5°C, the HSQC spectrum displays peaks with good resolution, well dispersed over 7-10 ppm (Figure 4.5B). Indeed, this spectrum is much more characteristic of a folded protein, suggesting that the (30-51, 5-14) non-native two disulfide intermediate of BPTI has significant structure at 5°C.

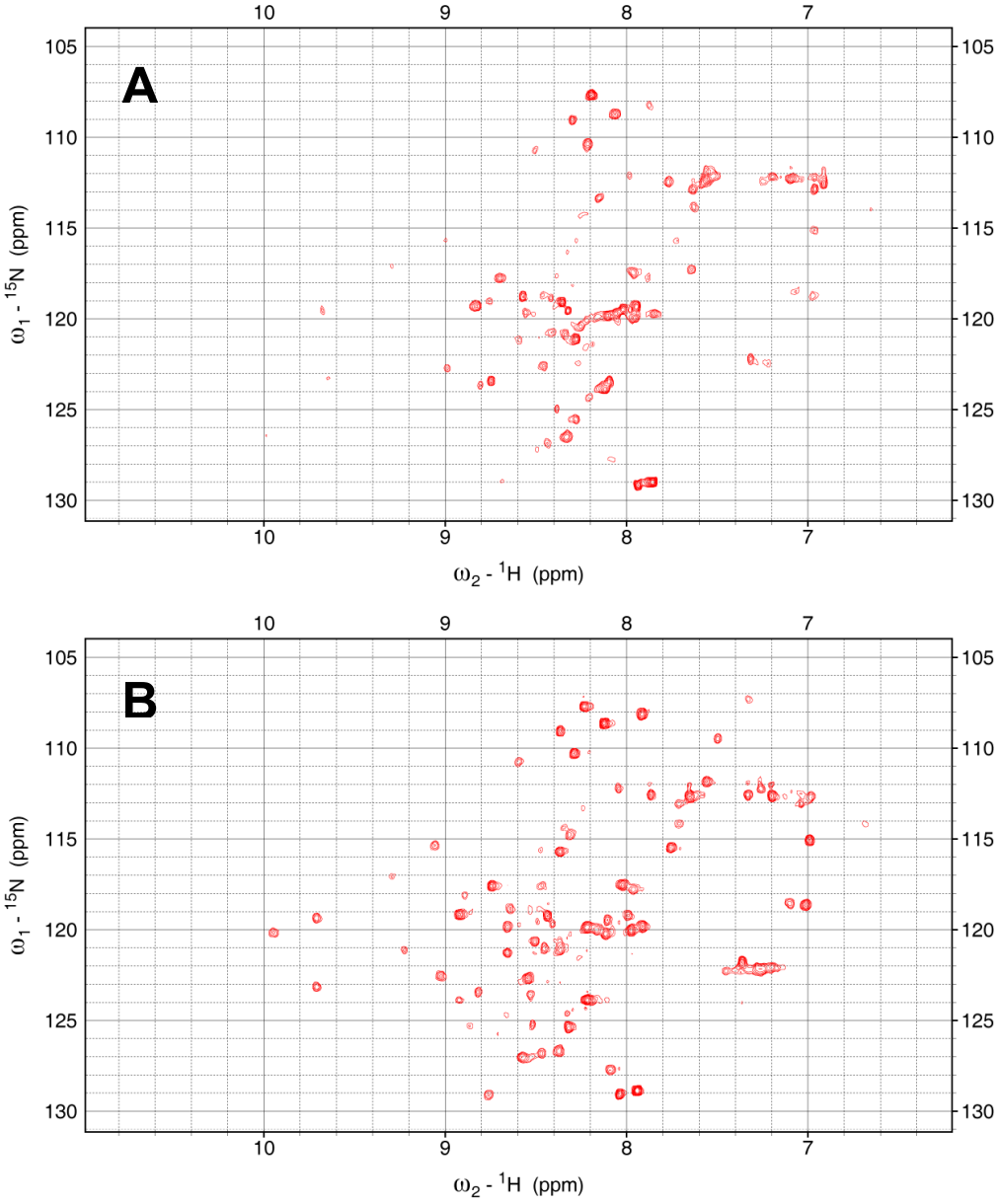


Figure 4.5: HSQC spectra of (30-51, 5-14) BPTI. A) 20°C; B) 5°C.

The effect of a change in temperature on secondary structure of (30-51, 5-14) BPTI was also observed using far-UV CD. Figure 4.6 shows that the protein has a very similar secondary structure profile at all temperatures. This suggests that the changes observed by NMR at various temperatures are due to flexibility and stability of the tertiary structure, rather than a change in secondary structure.

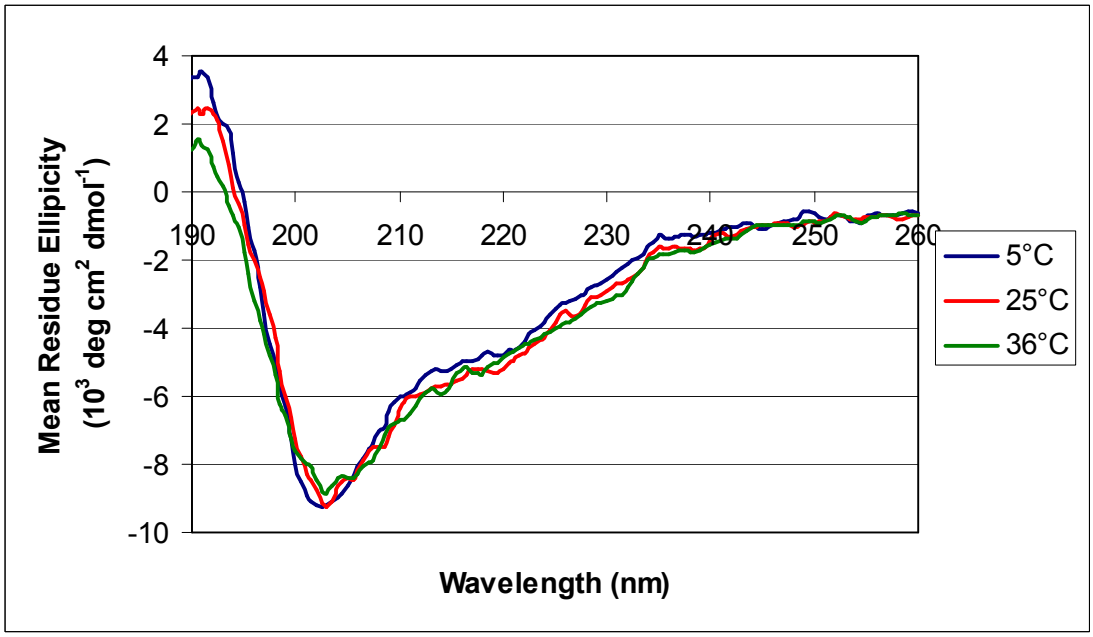


Figure 4.6: Far-UV circular dichroism of (30-51, 5-14) BPTI at various temperatures. Blue, 5°C; red, 25°C; green, 36°C.

To compare the (30-51, 5-14) BPTI HSQC spectrum acquired at 5°C with the wild type protein, a series of HSQC spectra was also acquired for wild type BPTI at various temperatures. These showed very little change in the wild type at lower temperatures (data not shown). This is not surprising since the wild type already displays stable native structure at 36°C, so lowering the temperature has little effect.

4.3.1. *De Novo* Assignment of (30-51, 5-14) BPTI NMR Spectra

Despite showing significant regions that appear to be folded, when overlaid with the wild type protein the HSQC spectrum of (30-51, 5-14) BPTI intermediate at 5°C displays a different overall structure to wild type BPTI (see Figure 4.10). Given this significant difference in spectra, the backbone of (30-51, 5-14) BPTI could not reliably be assigned through a simple comparison. Instead, separate NMR experiments were required for *de novo* assignment of the (30-51, 5-14)

BPTI spectrum. A ^{15}N labelled (30-51, 5-14) BPTI sample was used to perform 3D ^{15}N-^{1}H-^{1}H TOCSY and 3D ^{15}N-^{1}H-^{1}H NOESY experiments at 5°C. These experiments provide information similar to unlabelled samples used for 2D TOCSY and 2D NOESY, but have the additional ^{15}N dimension, thus making them much more informative and allowing distinction between peaks that overlap in the ^{1}H-^{1}H dimensions. Since there is a general characteristic pattern of chemical shifts associated with each amino acid (Roberts 1993), the 3D TOCSY spectrum was used to identify the amino acid residues in (30-51, 5-14) BPTI. A sequential walk across the "through space" interactions of the 3D NOESY spectrum identified the relative position of each amino acid in the primary sequence. These processes were performed iteratively so that as many peaks as possible could be assigned correctly. Figure 4.7 shows a strip plot for Arg53 to Gly56, a series of strips showing the sections of the 3D NOESY and 3D TOCSY spectra relevant to each residue, displayed at the appropriate ^{15}N chemical shift. Once the amino acid has been identified from the 3D TOCSY spectrum, the 3D NOESY is used to assign cross correlating peaks to establish the sequential assignment.

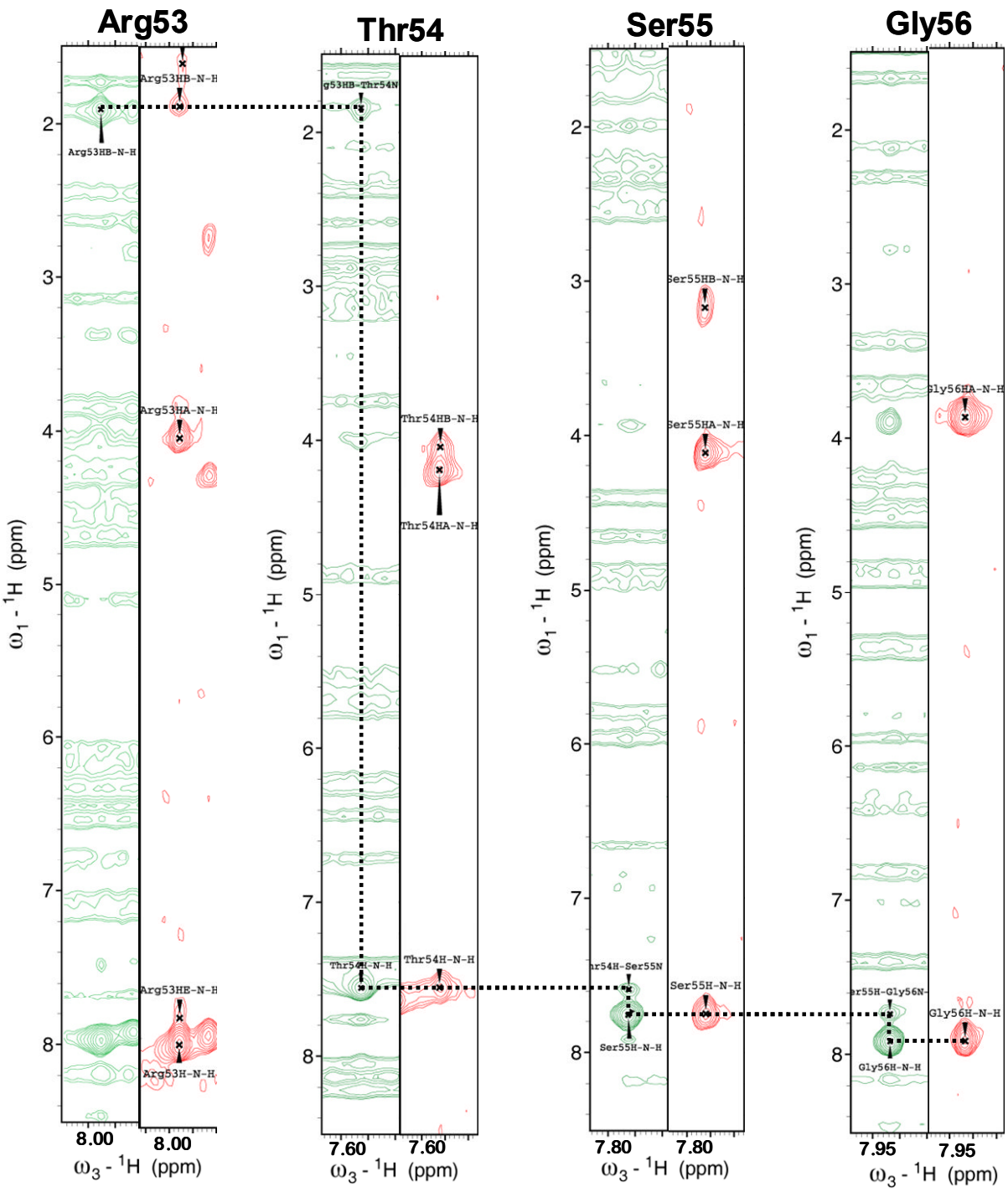


Figure 4.7: Strip plots of (30-51, 5-14) BPTI 3D spectra illustrating the process of sequential assignment of the protein backbone. Residues Arg53 to Gly56 are represented showing three NOESY cross-peaks. For each residue, the 3D NOESY is represented in green and 3D TOCSY is represented in red. The relevant assigned peaks are labelled in each spectrum.

In order to facilitate and visualise the assignment process, 2D ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY spectra were also acquired. Figure 4.8 shows an overlay of the 2D TOCSY and 2D NOESY spectra after assigned peaks from the 3D data had been mapped onto the corresponding peaks in the ^{1}H-^{1}H dimensions. In cases where assignments were unable to be identified in the 3D spectra, the 2D spectra were often able to highlight or clarify the relevant peaks. A previous study that assigned ^{1}H chemical shifts to a similar (30-51, 5-14) BPTI construct, under NMR conditions different to this study, was also used as a rough guide for assignment of resonances (Vanmierlo, Kemmink et al. 1994).

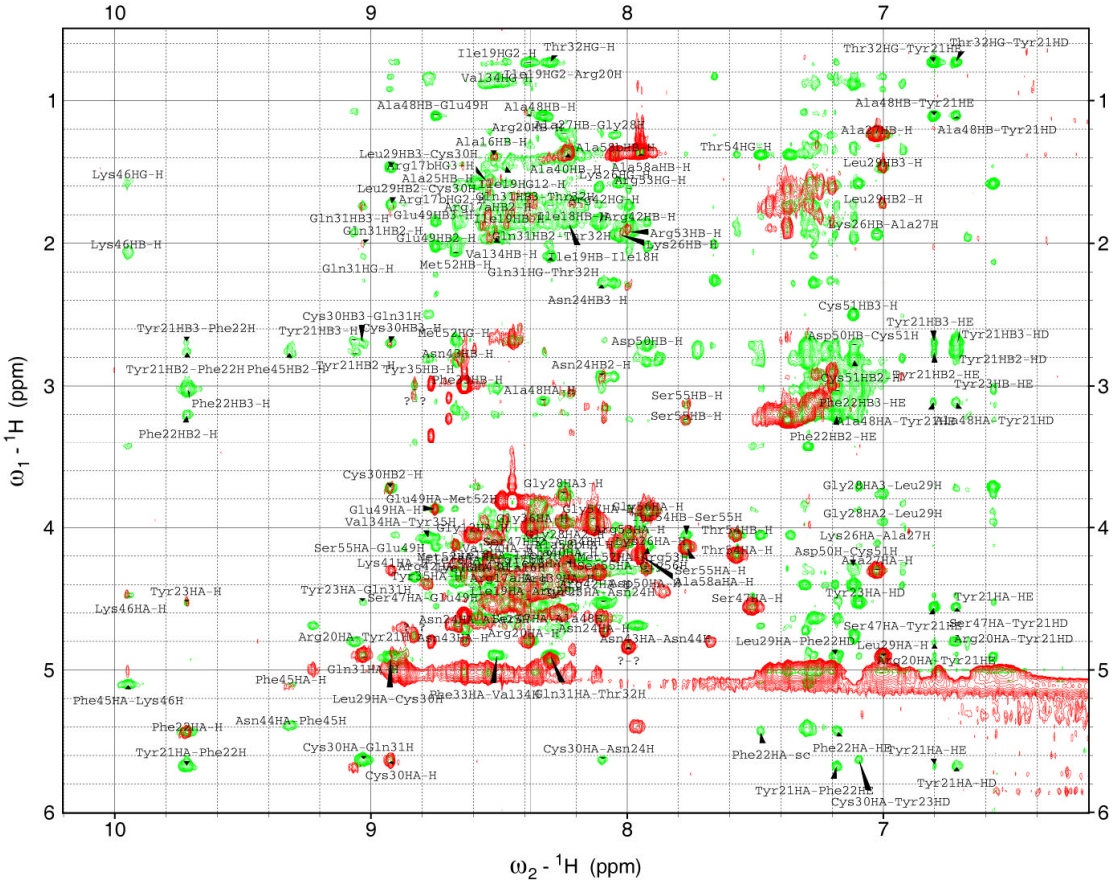


Figure 4.8: Overlay of 2D TOCSY (red) and 2D NOESY (green) spectra of (30-51, 5-14) BPTI at 5°C. Assignments of NOESY peaks are labelled.

4.3.2. Mapping 3D HQSC-TOCSY/ 3D HSQC-NOESY Assignments to HSQC Spectrum

Once peaks were defined in the 3D HSQC-TOCSY spectrum, labelling of the corresponding HSQC spectrum of (30-51, 5-14) BPTI was a straightforward process. The chemical shifts of the 3D HSQC-TOCSY correlated directly with those of the HSQC spectrum. Figure 4.9 shows the HSQC spectrum of (30-51, 5-14) BPTI, with labelled assignments.

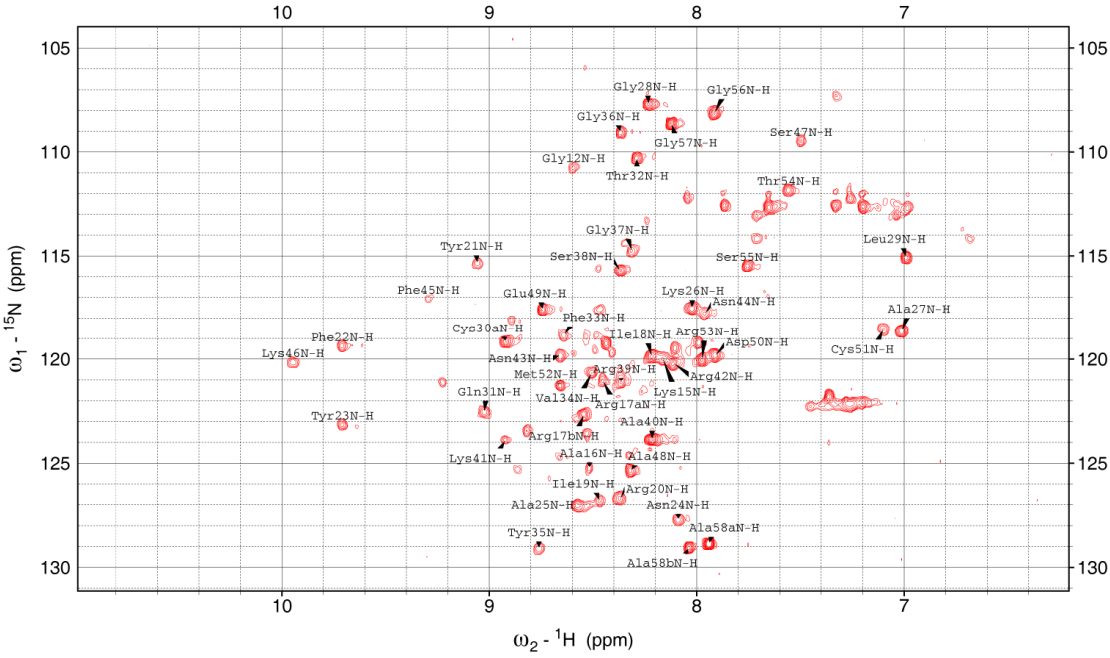


Figure 4.9: HSQC spectrum of (30-51, 5-14) BPTI at 5°C showing assigned peaks.

One of the most interesting features identified from the assignment of peaks is that all of the first 14 residues at the N-terminus of the protein are missing, except for Gly12. This indicates that this region of the protein is still very flexible, resulting in peak broadening and implying that the N-terminus of this mutant remains unfolded.

Native BPTI contains disulfides bonds between cysteines 30-51, 14-38 and 5-55. The 30-51 disulfide bond is typically the first to form and is buried within the core of the molecule. The other disulfide bonds assist in tethering more solvent exposed regions, restricting their flexibility (Figure 1.11). However, the (30-51, 5-14) BPTI intermediate does not contain the 14-38 or 5-55 native disulfide bonds, leaving much of the N-terminus in particular very flexible. Instead (30-51, 5-14) BPTI contains the non-native 5-14 disulfide bond, which is internal to the N-terminal region, creating a loop that may sterically hinder potentially stabilising interactions between the N-terminus and the rest of the molecule. In addition, it should be noted that even in native wild type BPTI, all but three of the first 17 N-terminal residues do not adopt regular secondary structure, making this region of the protein inherently more flexible.

In contrast, the C-terminal end produces the most intense and clearly defined peaks, suggesting a more clearly defined chemical environment for this region of the protein. Interestingly, the last residue on the C-terminus, Ala58, shows two clear peaks representing two distinct chemical environments for this residue. A full list of assignments for all spectra of (30-51, 5-14) BPTI at 5°C is available in Appendix B.

The difference between wild type and (30-51, 5-14) BPTI at 5°C can be visualised using an overlay of both HSQC spectra (Figure 4.10). Despite showing characteristics of a folded protein, the (30-51, 5-14) BPTI spectrum is still significantly different than the wild type spectrum at 5°C. The unfolded regions are large enough to affect the global structure of the protein, resulting in a different chemical environment for each residue.

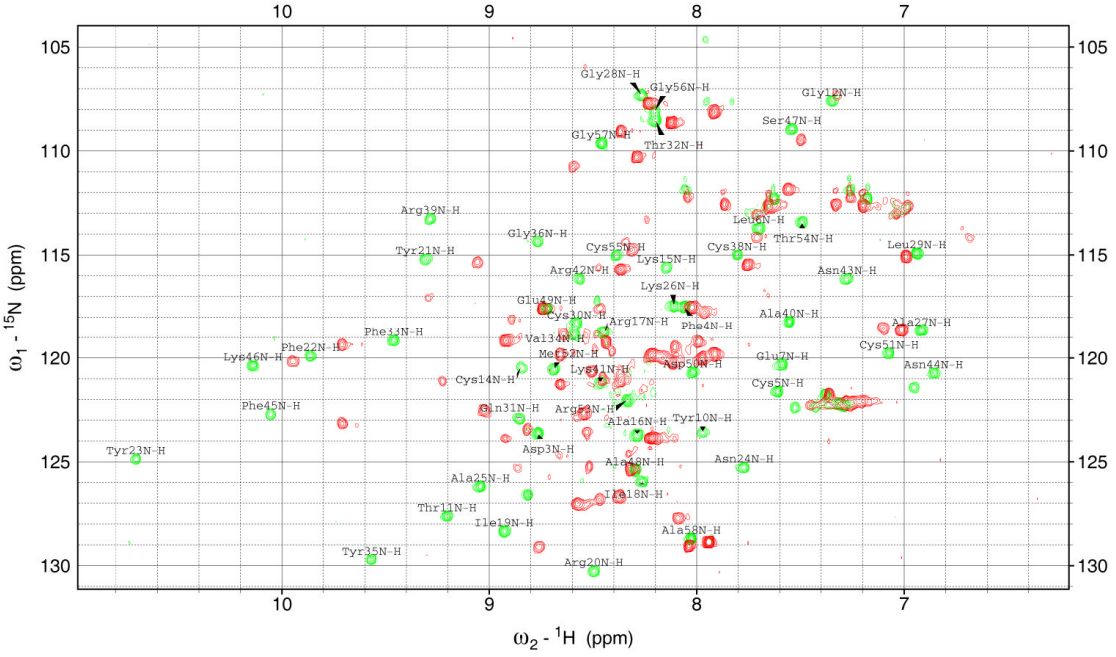


Figure 4.10: Overlay of HSQC spectra of wild type (green) and (30-51, 5-14) BPTI (red) at 5°C. Assignments of wild type BPTI are labelled.

In an effort to quantify the differences between the wild type and (30-51, 5-14) BPTI HSQCs, a chemical shift map was created comparing the chemical shifts of the two spectra (Figure 4.11). The process of calculating chemical shift changes was as outlined in section 2.12.7.3. Only residues with peaks assigned in both spectra were used in this analysis.

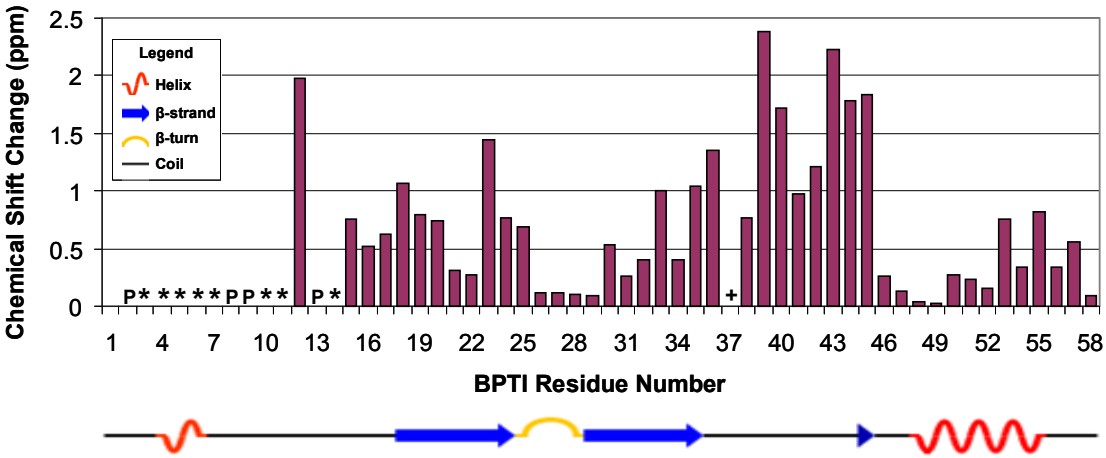


Figure 4.11: Change in chemical shift between wild type and (30-51, 5-14) BPTI HSQC spectra. The secondary structure associated with each residue is illustrated graphically. Residues not applicable to this analysis are represented as follows: P, proline; *, assigned in wild type but not in (30-51, 5-14) BPTI; +, assigned in (30-51, 5-14) but not in wild type BPTI. Arg1 was not detected in either HSQC spectra.

No obvious global pattern emerges, although the largest shifts occur in a coil region and a row of four small shifts correspond closely to the β-turn region of the protein, suggesting the environment of this structure may be least effected by the mutations from the wild type. As previously observed, only Gly12 of the first 14 residues at the N-terminus is assigned in (30-51, 5-14) BPTI. In wild type BPTI, Gly37 was observed upfield of the fingerprint region in the TOCSY spectrum due to chemical shielding by Tyr35 (see section 4.2.1), so it was not observed within the spectral range of the HSQC spectrum. Interestingly, however, Gly37 was observed in the HSQC spectrum of (30-51, 5-14) BPTI, suggesting that its NH proton was not shielded by Tyr35 in the partly-folded intermediate.

4.3.3. Mapping HSQC Assignments to Higher Temperatures

Once assignments had been made for (30-51, 5-14) BPTI at 5°C, this spectrum could be used as a template for mapping to less well defined spectra at higher

temperatures. The 5°C spectrum was used to assign available peaks in the HSQC spectrum acquired at 10°C. This in turn was used to assign the spectrum at 15°C, which was used to assign the spectrum at 20°C and so on until assignments were made, where possible, to the original spectrum at 36°C. No clear pattern emerges from the assigned residues, although most are towards the C-terminal region of the protein, with Ala58 still showing two clear peaks representing two distinct chemical environments for this residue. A full list of assignments for HSQC spectra of (30-51, 5-14) BPTI at 20°C and 36°C is available in Appendix C.

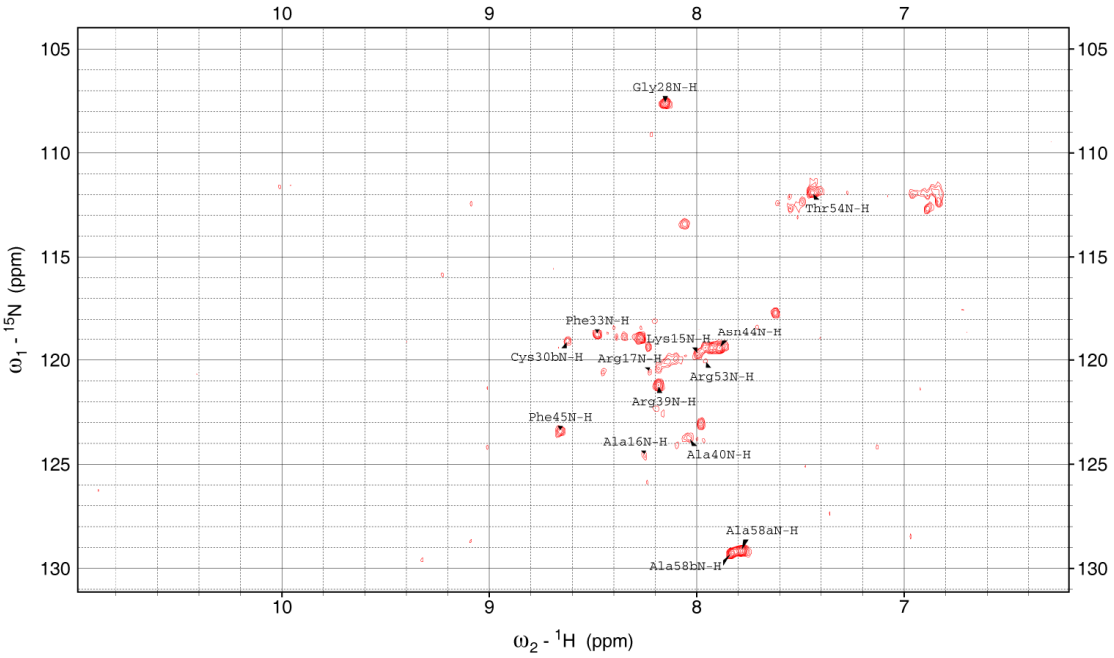


Figure 4.12: HSQC spectrum for (30-51, 5-14) BPTI at 36°C showing peak assignments.

4.4. NMR of Reduced Alkylated BPTI

So far both wild type, representing the fully folded protein, and (30-51, 5-14) BPTI, representing a folding intermediate, have been observed by NMR. However, by trapping wild type BPTI in a fully reduced state, a construct that would be a close representation to a fully unfolded BPTI should result. Together,

these constructs provide three distinct stages in the folding pathway of BPTI. Reduced BPTI was prepared by alkylation of the free thiol groups after reduction of wild type BPTI as specified in section 2.7.1.

4.4.1. NMR of Reduced Alkylated BPTI at 5°C

Since the partly-folded (30-51, 5-14) BPTI only showed good resolution of peaks at 5°C, this seemed a logical starting point for acquisition of an HSQC spectrum for reduced BPTI. Unfortunately, the solubility of reduced BPTI was limited to just 100 μM. Presumably this is due to large exposed hydrophobic regions of the unfolded protein causing aggregation even at relatively low molar concentrations. Initially, the same parameters were used to acquire an HSQC spectrum as for both wild type and (30-51, 5-14) BPTI at 5°C. However, t**Error! Reference source not found.**he resulting HSQC spectrum showed a relatively weak signal-to-noise ratio (not shown).

Since the molar concentration of the reduced BPTI sample was significantly lower than other BPTI constructs, the HSQC experiment was repeated, but with an increae in acquisition time. This was done by increasing the number of scans from 4 scans per experiment to 20 scans per experiment, resulting in a five fold increase in acquisition time. Figure 4.13 shows the improved signal-to-noise ratio that results from this enhanced spectrum. Even at 5°C, all visible peaks are contained within a narrow range of 8-9 ppm, plus a couple of small clusters around 112 ppm where aromatic side chains are usually visible.

As with (30-51, 5-14) BPTI at 36°C, this results from highly flexible residues of a largely unfolded protein experiencing very similar average chemical environments. Unlike (30-51, 5-14) BPTI, however, the unfolded nature of the protein remains even at low temperatures. For completeness, HSQC spectra were also acquired at 20°C and 36°C, but as expected this produced very

similar spectra, showing that it is also unfolded at higher temperatures (not shown).

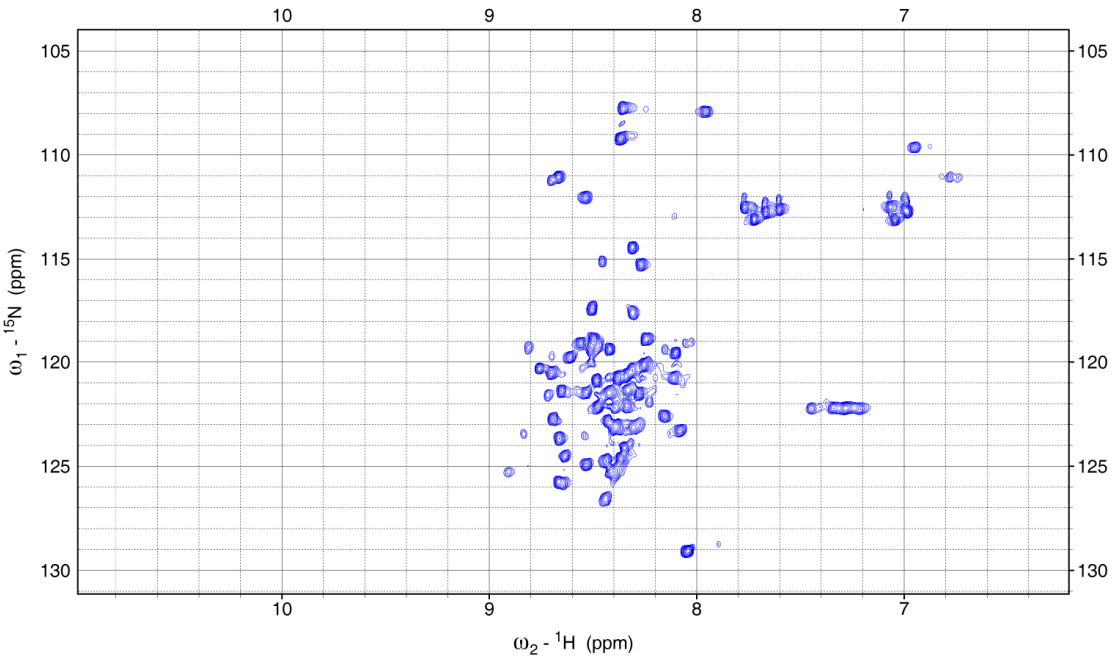


Figure 4.13: HSQC spectrum of reduced alkylated BPTI at 5°C, enhanced using longer data acquisition time.

4.5. Comparison of HSQC Spectra from BPTI Constructs

The differences between each of the BPTI constructs become clearly visible when the HSQC spectra of all three constructs at 5°C are overlaid (Figure 4.14). The well dispersed and well resolved peaks of the fully folded wild type BPTI (green) show a very different spectrum to the merged peaks of the unfolded reduced BPTI (blue). (30-51, 5-14) BPTI (red) shows a spectrum that lies somewhere in between the others, indicating an intermediate, partly-folded state.

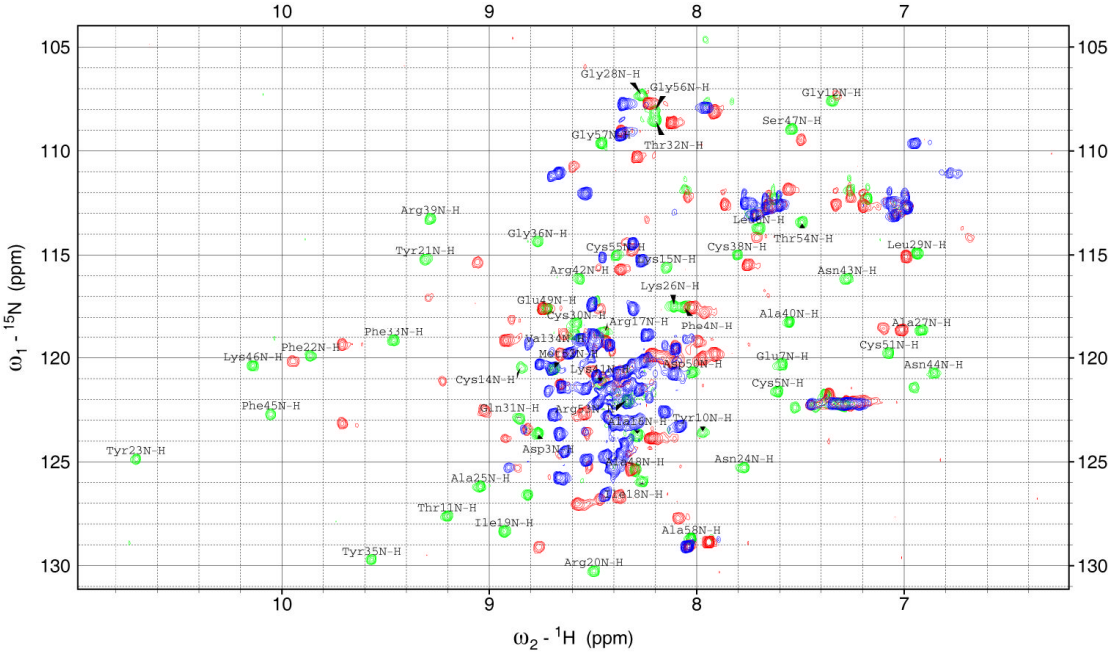


Figure 4.14: Overlay of HSQC spectra for BPTI constructs representing different stages along the protein folding pathway. Green, wild type BPTI; red, (30-51, 5-14) BPTI; blue, reduced BPTI. All spectra were acquired at 5°C. Assignments for wild type BPTI are labelled.

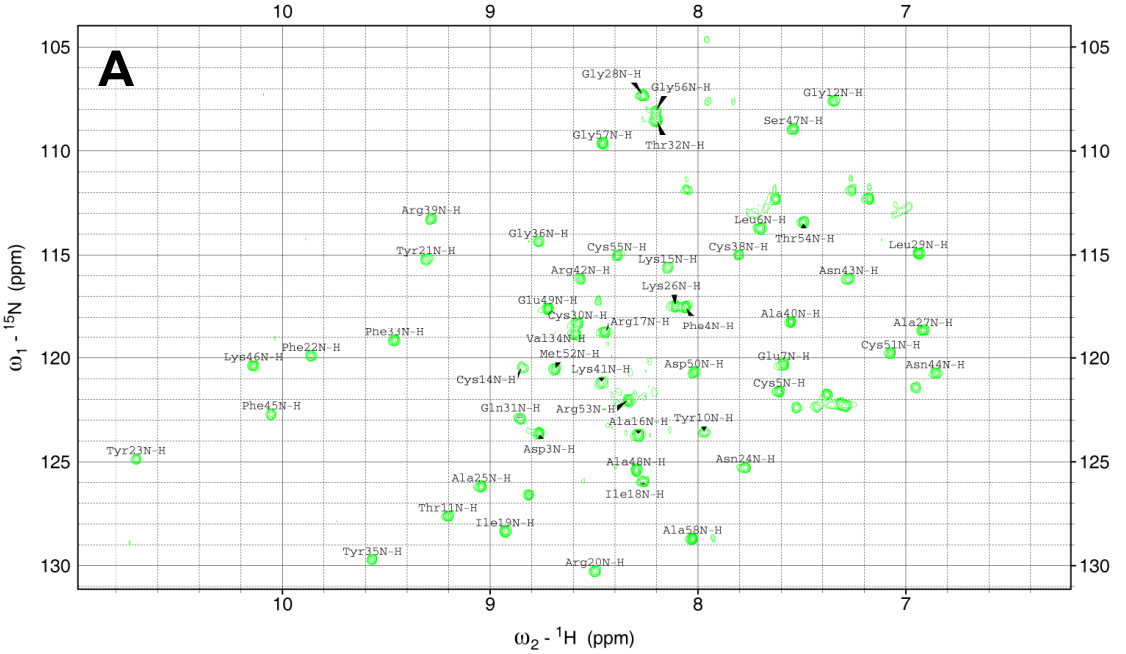
4.6. Hydrogen Deuterium Exchange of BPTI

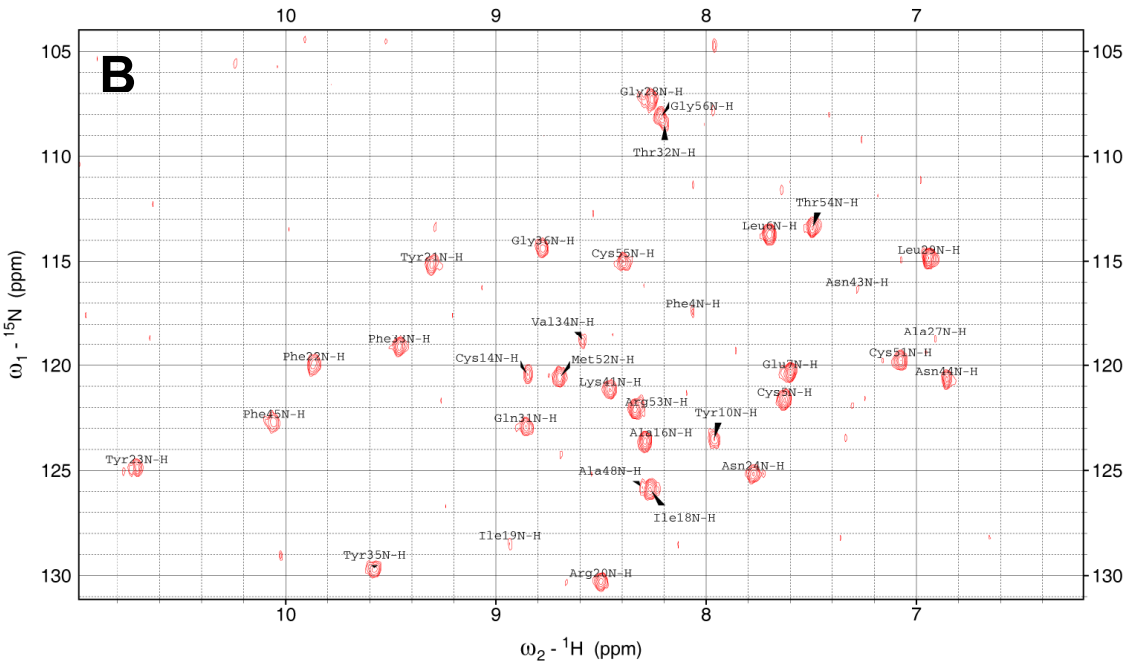
Since HSQC experiments rely on signals from amide protons, once the protein is dissolved in a deuterium oxide NMR buffer, the exchange of these protons for deuterium will result in a loss of signal. Hence, the most solvent exposed regions of the protein will lose their HSQC peaks first, whereas peaks associated with the buried regions of the protein will persist. A timecourse of HSQCs were acquired for each BPTI construct in order to observe the effect of exposure to a D_2O solvent. By comparing the timecourses of different BPTI constructs, we can see how relatively stably folded or unfolded each construct is.

Since the greatest distinction between structures occurs at 5°C, all hydrogen deuterium exchange experiments were also performed at 5°C. The NMR buffer

was identical to that used in previous HSQC experiments, except that the solvent is 100% D_2O, rather than 90% H_2O:10% D_2O. Each experiment was performed at a pH meter reading of 6.5, which equates to a pD of 6.9 (Glasoe and Long 1960). In order to minimise the time before data acquisition of a deuterated sample, a "dummy sample" was used to optimise instrument parameters, including locking, shimming, matching and tuning. Details of the experiment are outlined in section 2.12.6.4.

An overlay of HSQC spectra for hydrogen/deuterium exchange of wild type BPTI after 5 min and 24 hours is shown in Figure 4.15, along with a spectrum of wild type BPTI in 90% H_2O buffer. After 5 min, 35% of the residues have disappeared compared to those seen in H_2O solvent, but the majority remain visible in the HSQC spectra. After 24 hours, only 24% of the original peaks remain, nearly all of which are in helix or β-strand regions. Since random coil regions are inherently more flexible it is not surprising that these residues become surface exposed more easily.


A
Gly28N-H
Gly56N-H
Gly12N-H
Ser47N-H
Gly57N-H
Thr32N-H
Arg39N-H
Leu6N-H
Gly36N-H
Thr54N-H
Tyr21N-H
Cys55N-H
Cys38N-H
Leu29N-H
Lys15N-H
Arg42N-H
Asn43N-H
Lys26N-H
Glu49N-H
Cys30N-H
Arg17N-H
Phe4N-H
Ala40N-H
Ala27N-H
Phe33N-H
Val34N-H
Phe22N-H
Met52N-H
Cys51N-H
Lys46N-H
Glu7N-H
Lys41N-H
Asp50N-H
Asn44N-H
Cys14N-H
Cys5N-H
Phe45N-H
Gln31N-H
Arg53N-H
Tyr10N-H
Ala16N-H
Tyr23N-H
Asp3N-H
Ala48N-H
Asn24N-H
Ala25N-H
Ile18N-H
Thr11N-H
Ile19N-H
Ala58N-H
Tyr35N-H
Arg20N-H
105
110
115
120
125
130
10
9
8
7
ω1 - 15N (ppm)
ω2 - 1H (ppm)


B
Gly28N-H
Gly56N-H
Thr32N-H
Thr54N-H
Leu6N-H
Gly36N-H
Tyr21N-H
Cys55N-H
Leu29N-H
Asn43N-H
Phe4N-H
Val34N-H
Ala27N-H
Phe33N-H
Phe22N-H
Cys14N-H
Met52N-H
Cys51N-H
Glu7N-H
Lys41N-H
Asn44N-H
Cys5N-H
Arg53N-H
Phe45N-H
Tyr10N-H
Gln31N-H
Ala16N-H
Tyr23N-H
Asn24N-H
Ala48N-H
Ile18N-H
Ile19N-H
Tyr35N-H
Arg20N-H
105
110
115
120
125
130
10
9
8
7
ω1 - 15N (ppm)
ω2 - 1H (ppm)

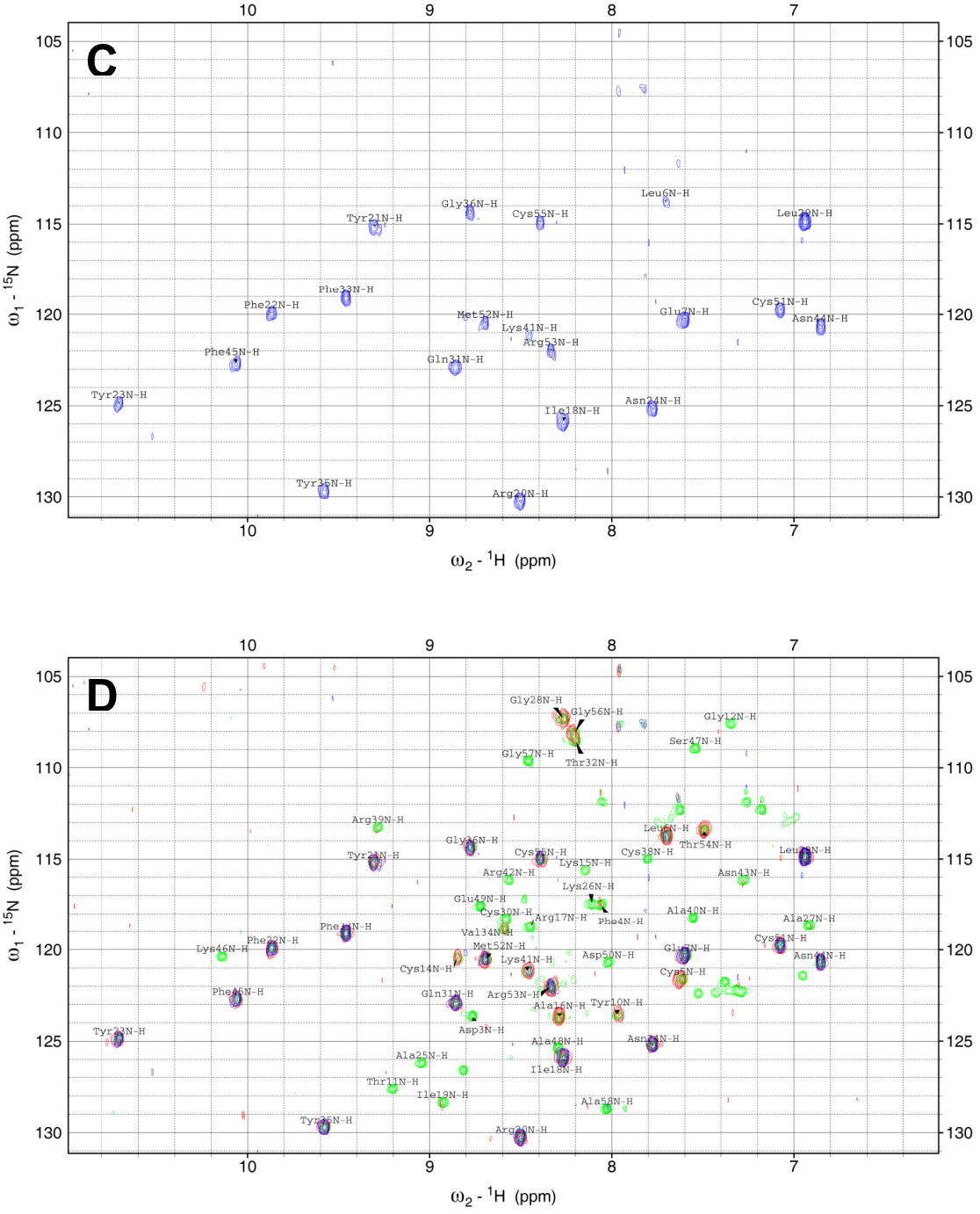
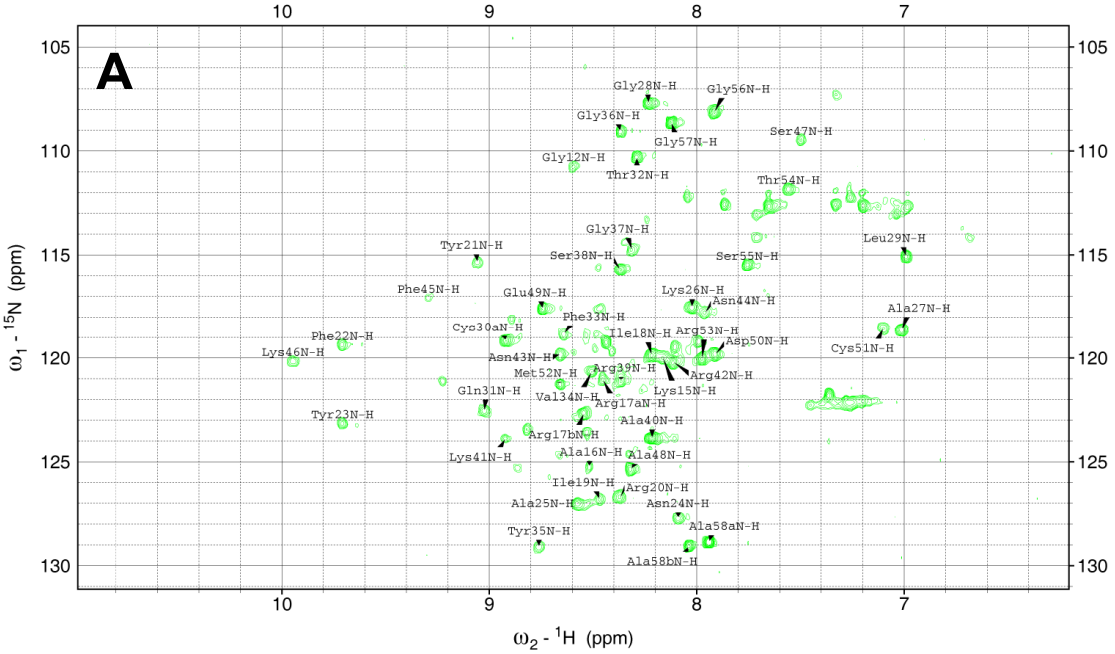
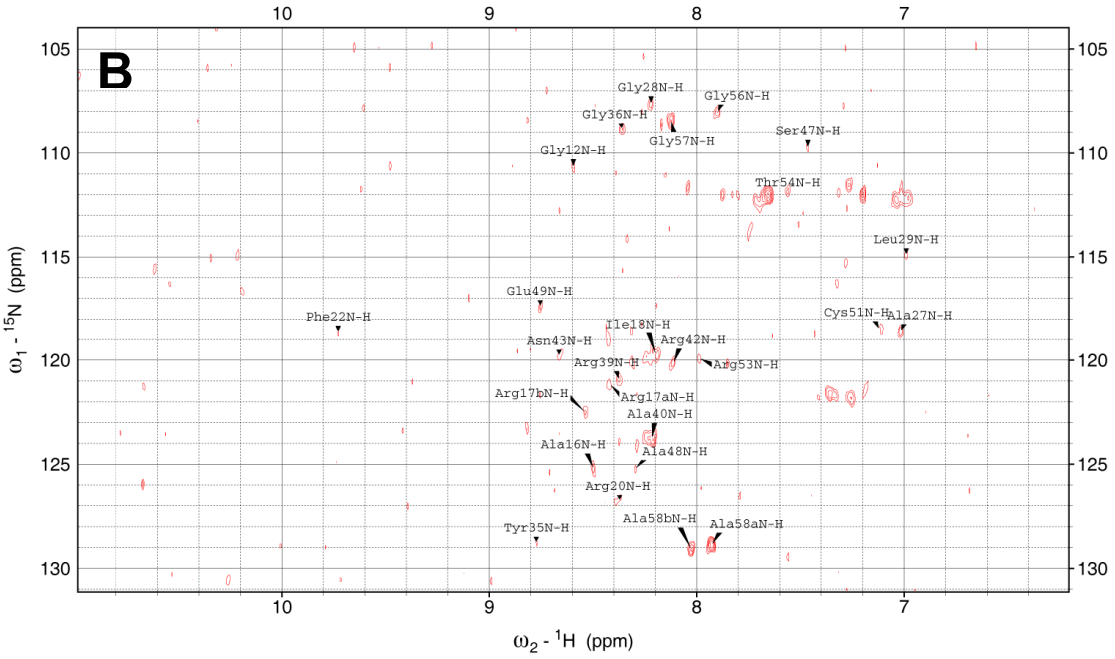


Figure 4.15: HSQC spectra for wild type BPTI with and without hydrogen deuterium exchange. A) in H_2O NMR buffer, with no deuterium exchange; B) after 5 min in D_2O NMR buffer; C) after 24 hours in D_2O NMR buffer; D) overlay of A, B and C spectra, illustrating residues that exchange rapidly, with assigned peaks for the non-exchanged spectrum labelled (green).


A
Gly28N-H
Gly56N-H
Gly36N-H
Gly57N-H
Ser47N-H
Gly12N-H
Thr32N-H
Thr54N-H
Gly37N-H
Leu29N-H
Tyr21N-H
Ser38N-H
Ser55N-H
Phe45N-H
Glu49N-H
Lys26N-H
Asn44N-H
Ala27N-H
Phe33N-H
Cys30aN-H
Ile18N-H
Arg53N-H
Phe22N-H
Asp50N-H
Lys46N-H
Cys51N-H
Asn43N-H
Arg39N-H
Met52N-H
Arg42N-H
Gln31N-H
Lys15N-H
Val34N-H
Arg17aN-H
Tyr23N-H
Ala40N-H
Arg17bN-H
Lys41N-H
Ala16N-H
Ala48N-H
Ile19N-H
Arg20N-H
Ala25N-H
Asn24N-H
Ala58aN-H
Tyr35N-H
Ala58bN-H
10
9
8
7
105
110
115
120
125
130
ω1 - 15N (ppm)
ω2 - 1H (ppm)


B
Gly28N-H
Gly56N-H
Gly36N-H
Gly57N-H
Ser47N-H
Gly12N-H
Thr54N-H
Leu29N-H
Glu49N-H
Phe22N-H
Cys51N-H
Ala27N-H
Ile18N-H
Asn43N-H
Arg42N-H
Arg39N-H
Arg53N-H
Arg17bN-H
Arg17aN-H
Ala40N-H
Ala16N-H
Ala48N-H
Arg20N-H
Ala58bN-H
Ala58aN-H
Tyr35N-H
10
9
8
7
105
110
115
120
125
130
ω1 - 15N (ppm)
ω2 - 1H (ppm)

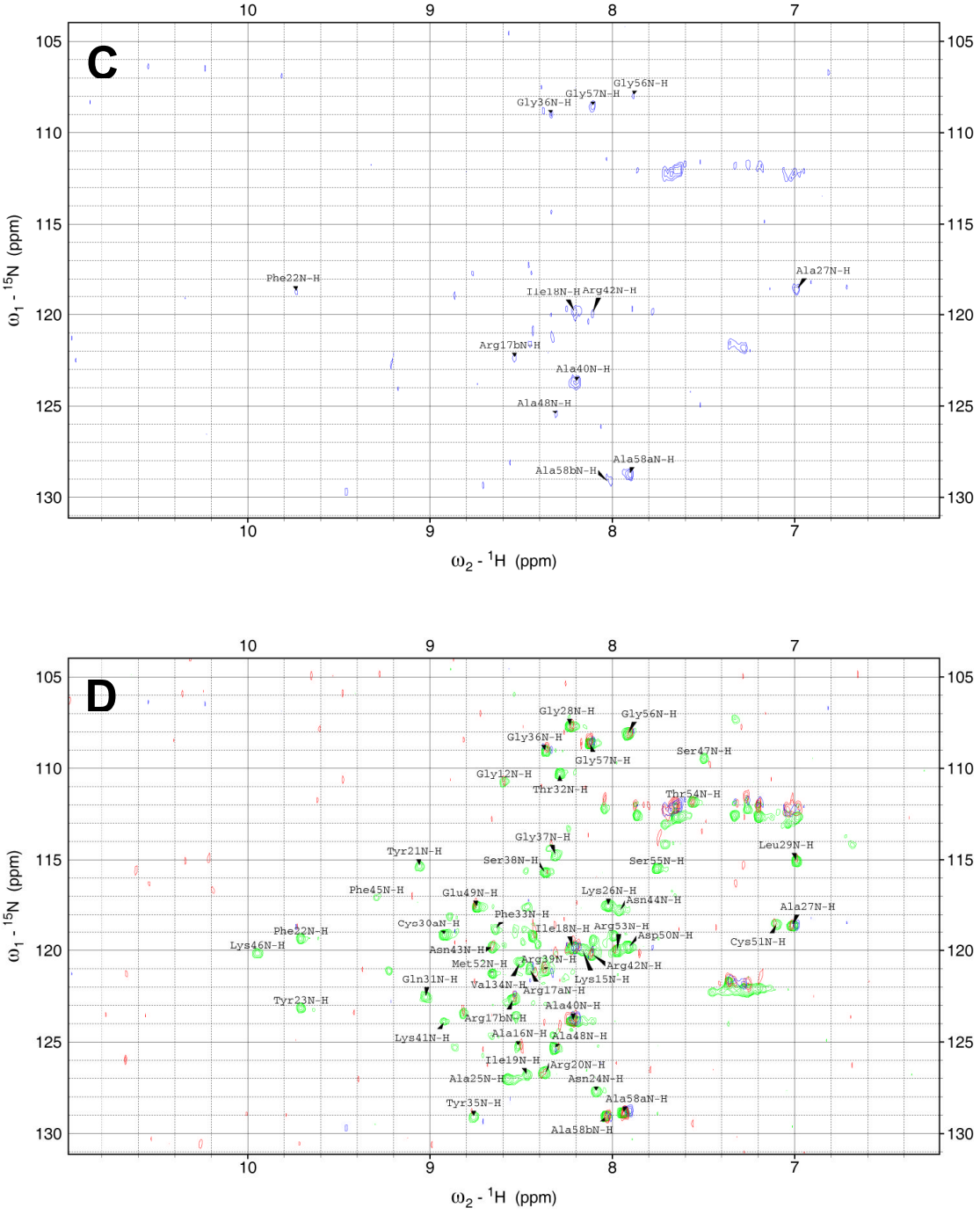


Figure 4.16: HSQC spectra for (30-51, 5-14) BPTI with and without hydrogen deuterium exchange. A) in H_2O NMR buffer, with no deuterium exchange; B) after 5 min in D_2O NMR buffer; C) after 24 hours in D_2O NMR buffer; D) overlay of A, B and C spectra, illustrating the rapid exchange of the majority of residues, with assigned peaks for the non-exchanged spectrum labelled (green).

An overlay of HSQC spectra for hydrogen/deuterium exchange of (30-51, 5-14) BPTI after 5 min and 24 hours is shown in Figure 4.16, along with a spectrum of (30-51, 5-14) BPTI in H_2O buffer. After 5 min, only 53% of the residues remain visible compared to those in the H_2O buffer and after 24 hours, this is reduced to just 29% of residues.

The assigned peaks of both wild type and (30-51, 5-14) BPTI were analysed to quantify their intensity for each of the HSQC spectra shown in Figure 4.15 and Figure 4.16. In each case, the peak height was identified and normalised to account for different levels of noise baseline (see Appendix D). The normalised peak heights were plotted against residue number for each spectrum. The resulting charts are shown in Figure 4.17, with a pictorial representation of the secondary structure shown below each.

Several observations can be made from Figure 4.17. First, the peak heights are generally a little higher for wild type than (30-51, 5-14) BPTI in the spectra before H/D exchange, and peaks of N-terminus of (30-51, 5-14) BPTI are missing. After 5 min of H/D exchange, many residues are no longer detected in the wild type. It may seem surprising that much of the natively folded protein exchanges so rapidly, but approximately one third of the protein lacks secondary structure (Kabsch and Sander 1983), and inspection of the remaining assigned peaks shows that most lie in helix or β-strand regions of the protein. Indeed, even after 24 hours, many of the helix and β-strand regions remain resistant to H/D exchange, suggesting they are in a stable position buried within the core of the molecule. Leu29 in particular seems to persist with relatively high peak intensity. This is located next to the Cys30-Cys51 disulfide bond, which is the most stable disulfide bond since it is buried in the core of the molecule.

After 5 min H/D exchange, fewer assigned peaks are identified in (30-51, 5-14) BPTI compared to the wild type, and nearly all that are present have a much lower peak height. This indicates that this partly-folded intermediate allows

much greater exposure of the core of the molecule to the deuterium solvent. This could be either due to its structure being less compact or the molecule being more flexible. Interestingly, although signal from C-terminal residues of the natively folded wild type BPTI appears to disappear after just 5 minutes in D_2O buffer, some signal from the C-terminal region of the mutant persists even after 24 hours in D_2O buffer, despite the signal from other residues disappearing. This suggests that in this particular partly-folded intermediate of BPTI, the C-terminal region has become buried into the central region of the protein.

In native BPTI, the 5-55 disulfide bond is the most surface exposed, hence is usually the last disulfide to be formed in the folding pathway (Figure 1.12). Typically, this would form a stable conformation whereby both the N- and C-termini are surface exposed in the native protein, see Figure 1.11. However, in the (30-51, 5-14) BPTI intermediate the native 5-55 disulfide bond is not present, leaving the C-terminus untethered. Additionally, this intermediate forms the non-native 5-14 disulfide bond, which is likely to sterically inhibit movement of the N-terminus into the vicinity of the C-terminus (hence the need for isomerisation to form the 14-38 native disulfide before the final 5-55 disulfide can be formed). Thus, the C-terminus of (30-51, 5-14) BPTI is free to explore its local environment and may become more stable by forming weak interactions with residues near the core of the protein.

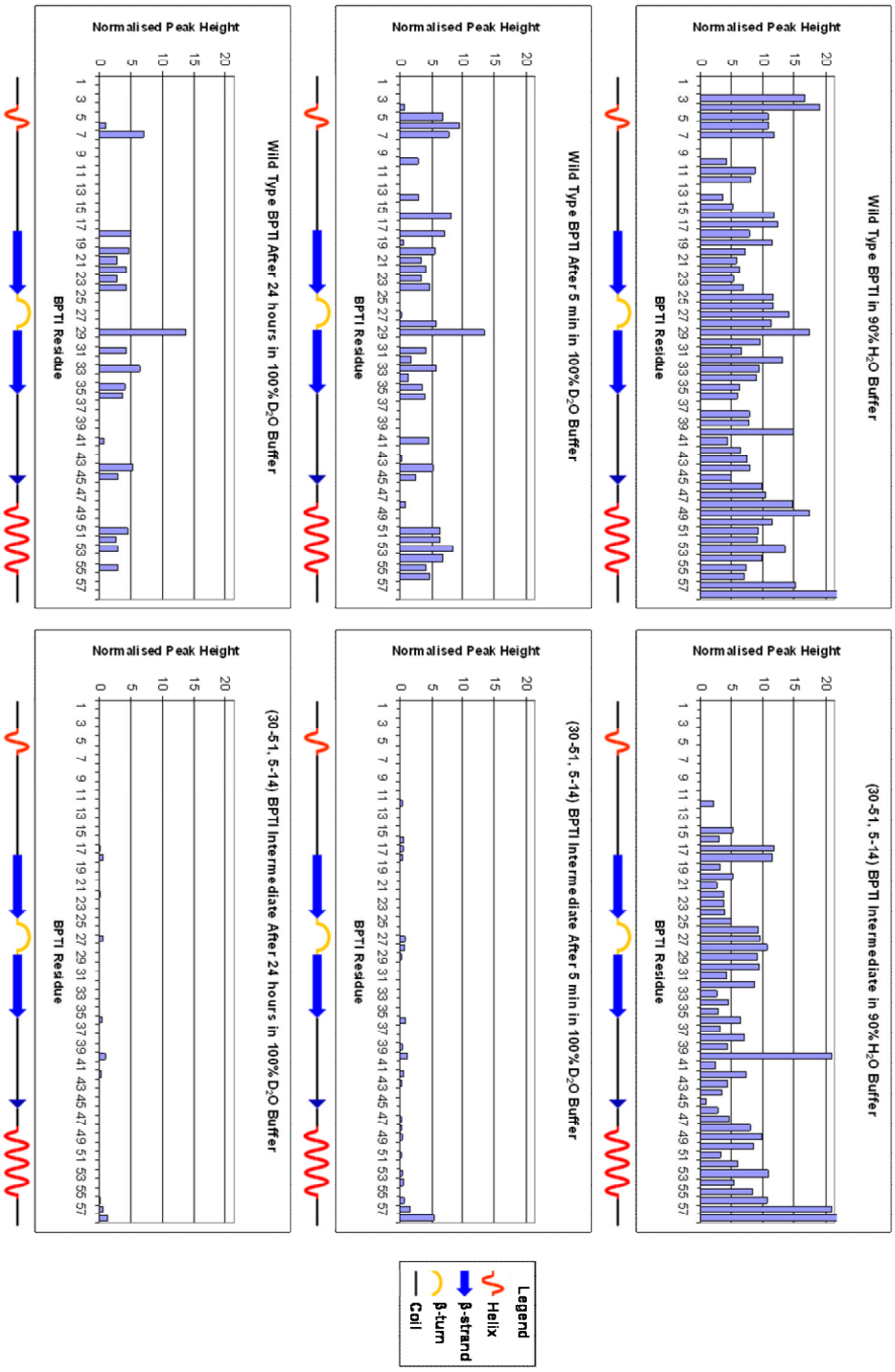


Figure 4.17: Comparison of peak heights for assigned residues of BPTI constructs. Left, wild type BPTI; right, (30-51, 5-14) BPTI intermediate. Top, in 90% H_2O buffer (effectively 0 min in D_2O buffer); middle, after 5 min in 100% D_2O buffer; bottom, after 24 hours in 100% D_2O buffer.

Figure 4.18 shows the HSQC spectrum of reduced BPTI after 5 min in D_2O NMR buffer. No clear peaks are observed, with nothing visible until the noise level of the spectrum is reached. This clearly indicates that the whole of the reduced alkylated BPTI molecule is fully exposed to solvent and is, therefore, fully unfolded.

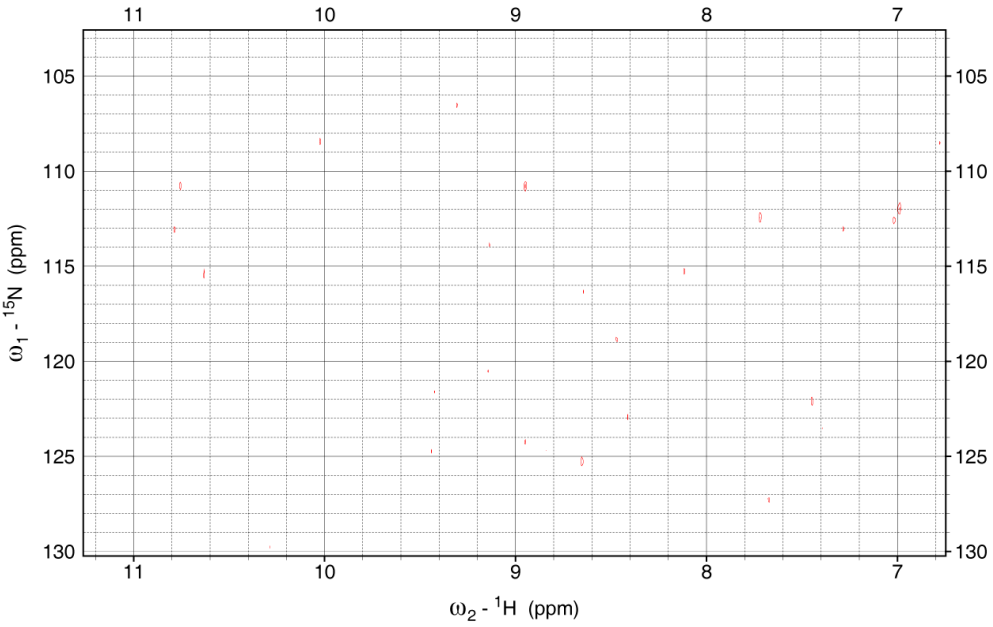


Figure 4.18: HSQC spectrum for reduced alkylated BPTI after 5 min in D_2O NMR buffer.

4.7. Discussion

The diagrammatic representation of the BPTI folding pathway (Figure 1.12) shows a general progression from the fully reduced protein, through intermediates with one or two disulfide bonds until the native, three disulfide bonded protein is reached. However, two keys points should be noted. First, the pathway is not a simple one way process. Rather, at any stage, including the native protein, there is always a certain amount of exchange between different disulfide bonding configurations. This is represented by the double-headed nature of each arrow showing a transition. It is the general increase in stability of

more folded intermediates that allows the protein to progress through each stage of the folding pathway. Secondly, within each stage of the pathway, represented by a particular disulfide bond configuration, there will exist a large range of valid structures. Instead of a single static structure, each stage is more accurately thought of as a dynamic ensemble of structures where one particular species may dominate. The results of this chapter have shown the dynamic nature of each of the three BPTI constructs.

The manner in which the dynamic nature of each construct is observed by NMR depends on the rate of exchange between chemical environments compared to the rate of acquisition of NMR data. Figure 4.19 illustrates this relationship in the simple case where two equally populated environments are in chemical exchange. If the rate of exchange is slow on the NMR timescale, both environments are observed as distinct peaks in the spectrum. If they are in fast exchange, only a single peak, representing an average of the two environments is observed. If exchange is on an intermediate timescale, coalescence occurs resulting in a broad peak with a large linewidth and reduced intensity (Roberts 1993).

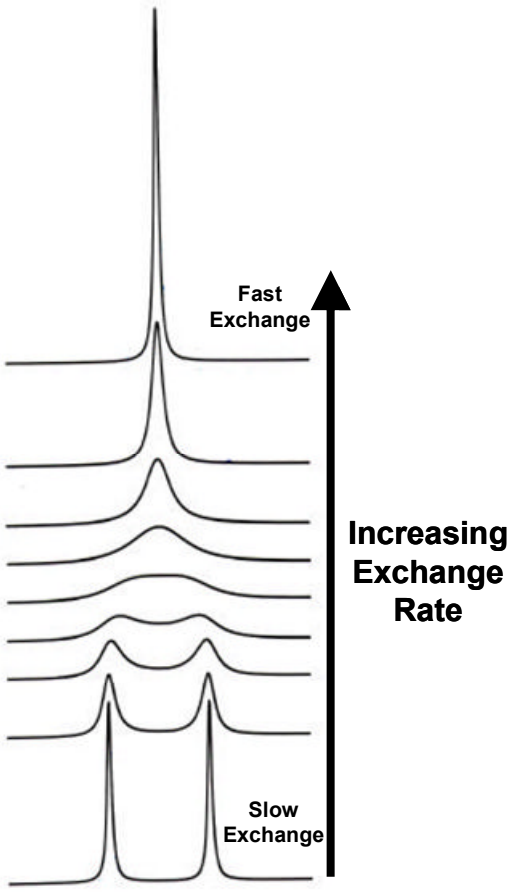


Figure 4.19: The effects of chemical exchange on NMR spectra at various exchange rates. modified from (Prestegard 2010).

Since wild type BPTI appears to be natively folded at all temperatures, the molecule is stable in a single conformation. Any changes from the native conformation would be transient and represent an insignificant proportion of the total population at any time, so are not effected by the chemical exchange rate. However, when comparing the HSQC spectra between 5 min and 24 hours of hydrogen deuterium exchange (Figure 4.15) some peaks disappear, implying that even in the fully folded protein non-native states are explored long enough to allow hydrogen deuterium exchange to take place.

The example shown in Figure 4.19 shows a simple case of chemical exchange between two equally weighted chemical environments. On the intermediate timescale, coalescence occurs resulting in line broadening and a small reduction in the amplitude of the signal. However, in the case of a very flexible molecule, such as the 14 N-terminal residues of (30-51, 5-14) BPTI, each amino group will experience many different chemical environments. The coalescence of all of these environments, if exchange is on the intermediate time scale, will result in line broadening and a reduction in amplitude so large that it is below the signal-to-noise ratio, thus no signal is detected from these residues.

The rest of the (30-51, 5-14) BPTI protein shows a distinctly different pattern. At 36°C, most peaks are displayed within a very narrow range of chemical shifts, implying that fast exchange is occurring, which is averaging many different environments of a largely unfolded protein. However, at 5°C peaks are much more disperse and well resolved, indicating they have a unique chemical environment. The lower temperature increases the stability of the dominant conformation, effectively slowing the exchange rate so that mostly a single chemical environment is observed for each residue (except at the N-terminus).

Interestingly, there is one exception to this observation, which is the last residue at the C-terminal end of the protein, Ala58. Here, two distinct peaks are observed for the residue. This implies that Ala58 exists in two conformations in slow exchange. This behaviour is not observed in wild type BPTI.

The fully reduced BPTI, which was completely or almost completely unfolded, shows many poorly resolved peaks. Like (30-51, 5-14) BPTI, the protein appears unfolded at 36°C. At 5°C, however, (30-51, 5-14) BPTI shows significant portions of folded protein, whereas reduced BPTI remains unfolded. Residues of the unfolded protein will experience many different chemical environments, occurring on a much faster timescale than for (30-51, 5-14) BPTI, resulting in peaks that represent an average of those environments for each

residue. Hydrogen deuterium exchange showed that all of the protein is surface exposed, confirming the unfolded nature of the protein.

This chapter has shown that recombinant wild type BPTI is extremely similar to commercially available natural mature BPTI through comparison of TOCSY and NOESY experiments. Analysis of HSQC spectra from ^{15}N isotopically labelled samples of wild type, (30-51, 5-14) and reduced BPTI has provided strong evidence that they represent natively folded, partly-folded and unfolded proteins respectively. Thus the available constructs act as good representations of substrates at different stages of folding for the study of binding with the PDI protein folding enzyme.

Chapter 5. NMR Spectroscopy of BPTI in the Presence of PDI

5.1. Introduction

So far, this study has focused on the expression, purification and characterisation of various BPTI constructs in isolation. Various NMR experiments demonstrated that wild type BPTI, (30-51, 5-14) BPTI and reduced BPTI represent native, partly-folded and unfolded proteins respectively. However the main interest of the study is to investigate how these constructs interact as substrates with the PDI folding enzyme.

In this chapter, further HSQC experiments were used to study ^{15}N labelled BPTI constructs. Here, however, changes were observed when increasing quantities of PDI were added to the substrate protein. By using ^{15}N labelled BPTI but unlabelled PDI, only resonances from the BPTI nuclei were observed, since each HSQC detects correlations with ^{15}N nuclei. Because PDI is a folding enzyme, it seemed reasonable to hypothesise that an unfolded protein, such as reduced BPTI, would bind much more readily than a fully folded protein, such as wild type BPTI. (30-51, 5-14) BPTI, being a partly-folded intermediate, would lie somewhere in between. The effects of PDI binding on the BPTI substrates were observed by NMR spectroscopy via HSQC experiments. Since (30-51, 5-14) BPTI was best observed at lower temperature, all HSQC experiments were performed at 5°C.

5.2. Expression and Purification of PDI

The expression and purification of each BPTI construct was shown in Chapter 3. Full length His-tagged PDI was expressed as outlined in section 2.1. Details of the purification protocols are given in section 2.5.

5.2.1. IMAC Purification

Since PDI constructs contain an N-terminal His-tag, purification can be performed using immobilised metal affinity chromatography (IMAC). Figure 5.1 shows SDS-PAGE analysis of IMAC purification for full length PDI. The total soluble protein applied to the IMAC column is shown in lane 3, showing good expression of PDI. It is very similar to the overall protein (including insoluble proteins), indicating that most of the protein is soluble (lane 2). Although a small amount does not bind (lane 4), the vast majority is eluted with the nickel when the column is washed with EDTA (lane 7). Only small amounts of PDI are observed in the wash steps (lanes 5 and 6).

5.2.2. Ion Exchange Purification

Since the protein was eluted from the IMAC column with nickel and EDTA, the eluted sample was dialysed overnight into 20 mM sodium phosphate (section 2.5.2). This was then loaded onto an anion exchange column. In a low salt environment, the protein will bind to the column. A gradient is then applied to gradually increase the salt content of the mobile phase, eventually resulting in the dissociation of PDI from the stationary phase. Figure 5.2 shows a typical chromatogram for ion exchange chromatography of PDI. The A_{280} signal shows PDI eluting at about 350 mM NaCl (70% of the salt gradient). The corresponding fractions were pooled and analysed using SDS-PAGE to confirm the correct molecular mass for PDI (Figure 5.1, lane 10).

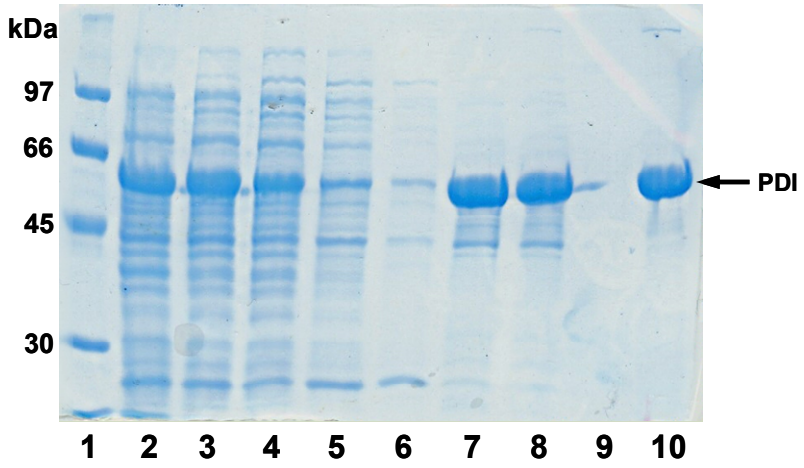


Figure 5.1: SDS-PAGE of PDI purification. Lane 1, LMW marker; lane 2, total expressed protein; lane 3, soluble protein; lane 4, IMAC flow through; lane 5, IMAC wash; lane 6, IMAC low salt; lane 7, IMAC elute; lane 8, dialysed into 20 mM sodium phosphate; lane 9, ion exchange flow through; lane 10, pooled fractions from ion exchange elute.

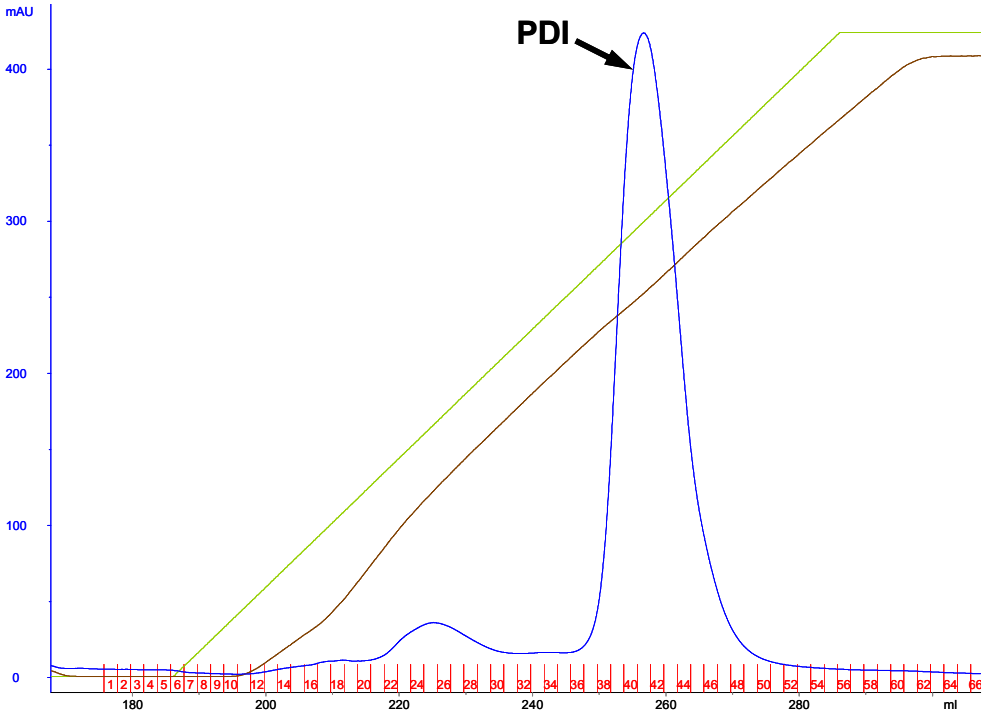


Figure 5.2: Chromatogram from ion exchange purification of full length PDI. Blue, absorbance at 280 nm; green, concentration of high salt buffer; brown, conductance; red, eluted fractions. At a low salt concentration, the protein is bound to the column; the protein is eluted at 70% of the salt gradient, corresponding to a 350 mM NaCl buffer.

5.2.3. Lyophilisation of PDI

Since all BPTI constructs were stored by lyophilisation, it was also desirable to store PDI in this manner. The protein eluted from the ion exchange column was in a high salt solution and was buffer exchanged into 10 mM ammonium bicarbonate. Since ammonium bicarbonate is a volatile salt, it was able to sublimate during the lyophilisation process. A native PAGE of samples before and after lyophilisation was run to observe if PDI was affected by this process (Figure 5.3). It shows that PDI is mainly monomeric, but with a small amount of dimer present at all stages through the process. SDS-PAGE analysis was also preformed and showed no signs of protein degradation from lyophilisation (data not shown).

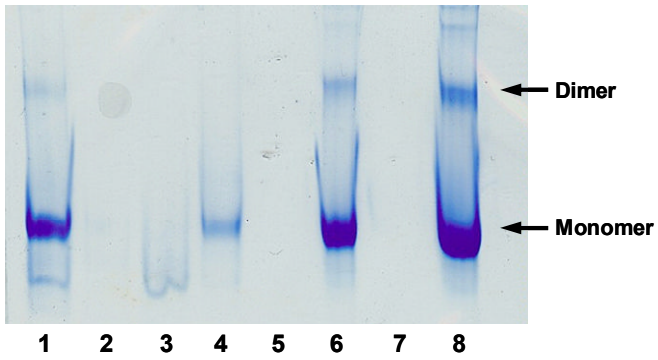


Figure 5.3: Native PAGE of PDI. Lane 1, dialysed into 20 mM sodium phosphate; lanes 2-5, fractions from ion exchange; lane 6, PDI buffer exchanged to 10 mM ammonium bicarbonate (before lyophilisation); lane 8, lyophilised PDI resuspended in NMR buffer. This indicates that PDI was not affected by the lyophilisation process. Whenever large samples are loaded, some dimer is observed, but the vast majority remains in the monomer state.

5.3. Titration of Full Length PDI into BPTI

It was important to make sure that conditions were compatible for both BPTI and PDI and that the two proteins did not aggregate after mixing. This possibility was already minimised by testing BPTI constructs in an NMR buffer previously used for NMR with PDI constructs.

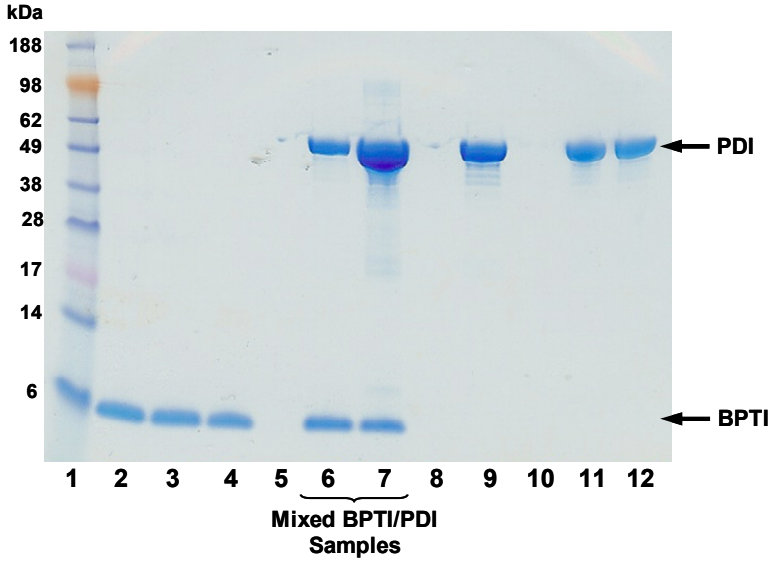


Figure 5.4: SDS-PAGE of (30-51, 5-14) BPTI, PDI and (30-51, 5-14) BPTI/PDI combined samples. Lane 1, SeeBlue marker; lanes 2-4, (30-51, 5-14) BPTI samples; lane 6, (30-51, 5-14) BPTI:PDI 5:1 (30 µM BPTI, 6 µM PDI); lane 7, (30-51, 5-14) BPTI:PDI 1:1 (30 µM each), lane 9, PDI ion exchange elute; lane 10, flow through from buffer exchange of PDI sample; lane 11, PDI after buffer exchange to 10 mM ammonium bicarbonate; lane 12, PDI after lyophilisation and resuspension in NMR buffer.

Figure 5.4 shows SDS-PAGE analysis of (30-51, 5-14) BPTI intermediate, PDI and a mixture of samples at different molar concentrations. It clearly shows monomeric PDI which can be purified, lyophilised and resuspended in NMR buffer without dimerisation or degradation occurring (lanes 9-12). As previously shown, BPTI samples were also prepared successfully in NMR buffer (lanes 2-4). The two proteins were also mixed successfully at different molar concentrations (lanes 6-7). Note that PDI (56 kDa) is more than 8 times larger in

molecular mass than BPTI (6.6 kDa), therefore will show up as a much larger band when at the same molar concentration (lane 7). Only at a 5 times higher molar concentration of BPTI does the PDI band appear a similar size (lane 6).

5.4. (30-51, 5-14) BPTI interaction with PDI

Since a partly-folded protein would be more likely to interact with PDI than one that is natively folded, it seemed logical to begin interaction studies using (30-51, 5-14) BPTI.

Before the addition of PDI, an HSQC spectrum was acquired with (30-51, 5-14) BPTI only. Since no prior knowledge was available with regard to the binding of PDI to a partly folded protein, initially the enzyme was added at a 1:1 molar ratio of BPTI:PDI. However, at this ratio it was found that very little was detected in the HSQC spectrum (data not shown). With a second sample, a sub-stoichiometric amount of PDI was added, at a BPTI:PDI ratio of 5:1. Surprisingly, even at this low sub-stoichiometric level of enzyme only a few peaks were visible in the HSQC spectrum, Figure 5.5. Overlay of the (30-51, 5-14) BPTI HSQC spectrum without PDI enabled the remaining peaks to be identified. Two clear peaks are visible for Ala58, with Gly57 and Gly36 also visible.

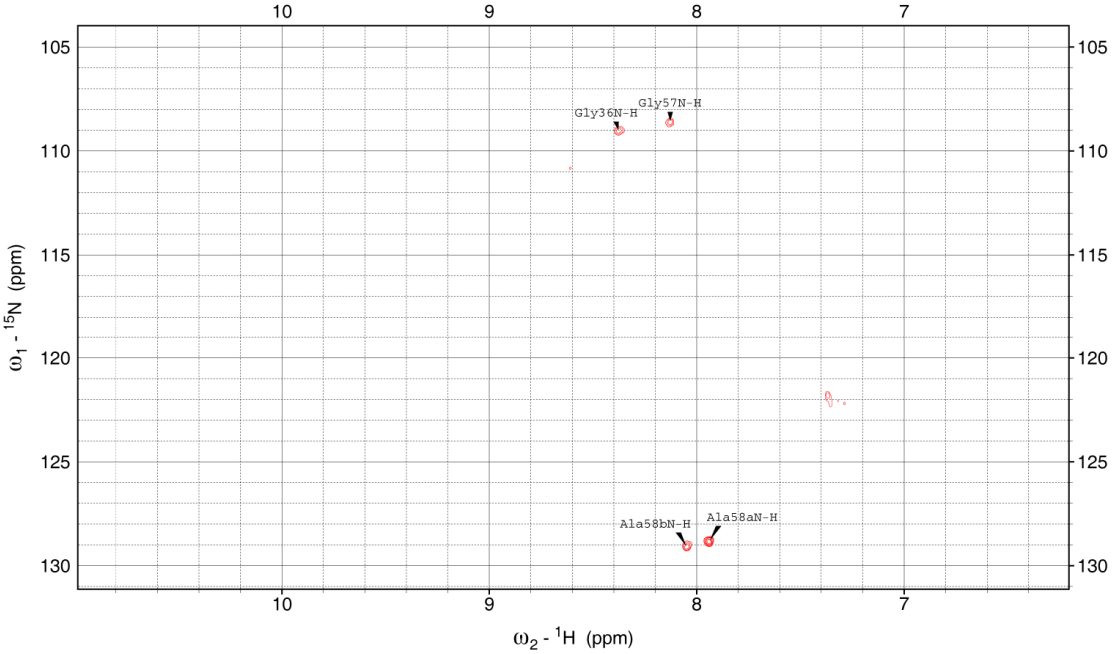


Figure 5.5: HSQC spectrum of (30-51, 5-14) BPTI in the presence of PDI at a BPTI:PDI ratio of 5:1.

Since BPTI is a relatively small protein (6.6 kDa), it is well suited to investigation using various NMR techniques, since such a small protein will have relatively fast molecular tumbling, enabling sharp peaks to be detected even at low temperatures. By comparison, PDI is a relatively large protein (56 kDa), resulting in a much slower molecular tumbling rate, making it too large to be detected using standard NMR techniques. When combined with ^{15}N labelled BPTI, the BPTI/PDI complex reaches 62.6 kDa, so is too large to be detected in the HSQC spectrum. So, generally speaking, peaks in the HSQC spectrum using ^{15}N labelled BPTI will only be visible for the unbound protein, but not when bound to the PDI enzyme. However, even in the bound state, regions of BPTI that remain flexible may still be observed, since they will still be able to show fast motions locally.

Figure 5.5 shows that almost all BPTI peaks have disappeared from the HSQC spectrum even at sub-stoichiometric concentrations of PDI. This suggests that on the timescale of observation of the HSQC signal, nearly all of the BPTI

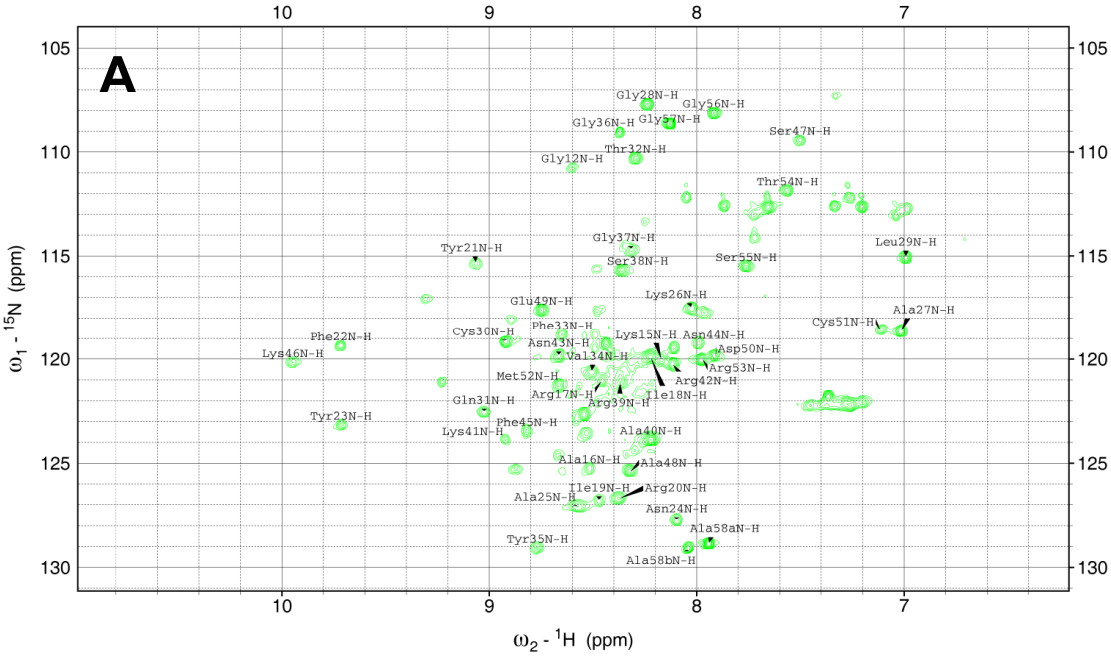
molecules have been bound by PDI, despite the enzyme being present at only one fifth of the molar concentration.

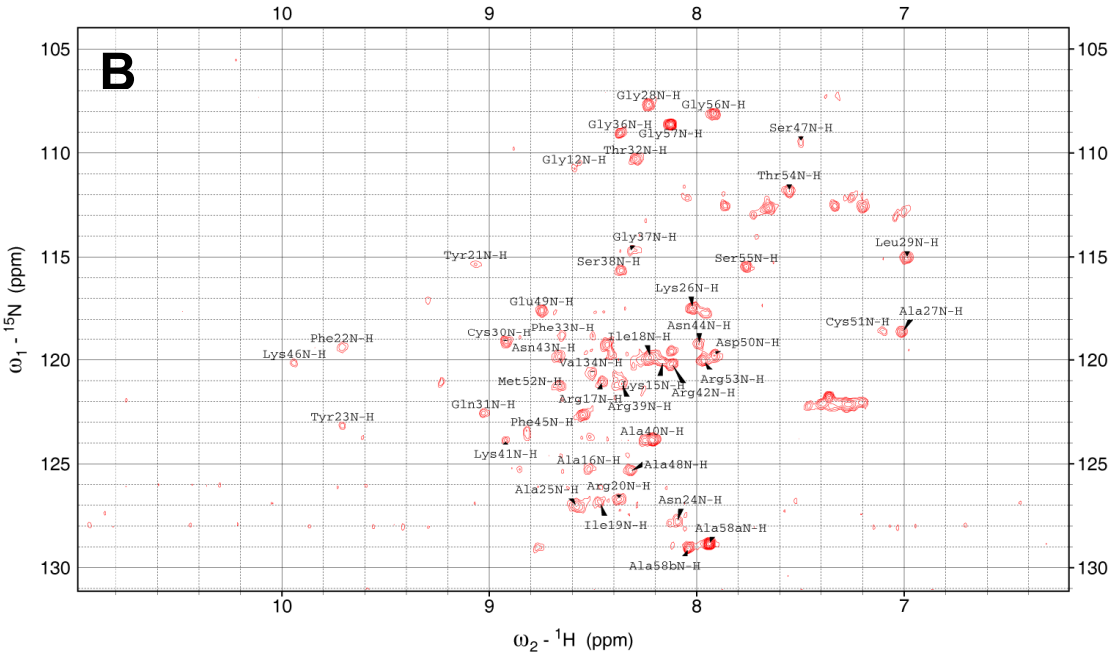
Theoretically, this could be explained if PDI had multiple binding sites. However, previous studies suggest that PDI only has a single binding site (Pirneskoski, Klappa et al. 2004). Rather, it seems that PDI is able to bind and unbind rapidly from its partly-folded substrate, such that all (30-51, 5-14) BPTI will have been bound within the NMR timescale. The drastic loss of signal even with PDI at one fifth of the molar concentration of BPTI may suggest that the enzyme is binding to at least 5 substrate molecules within the timescale of HSQC observation. Alternatively, binding may be less rapid, but exchange between bound and unbound BPTI may be on an intermediate timescale, resulting in some line broadening even in the unbound state.

Interestingly, however, a few peaks remain visible in the HSQC spectrum of Figure 5.5. This could simply be because these residues had a higher intensity in the unbound BPTI intermediate, and so when only a small fraction of the population remains in the unbound state, only these residues remain detected. Gly57 and Ala58 in particular showed relatively high peak heights in the HSQC spectrum of unbound (30-51, 5-14) BPTI at 5°C (see Appendix C).

Alternatively, those residues that remain visible may be able to experience some “local tumbling”, independent of the rest of the molecule. Perhaps the flexibility of the C-terminal region allows it to retain isotropy by moving unhindered by the rest of the molecule, thus permitting detection of HSQC signal from this region.

To get a better idea of the extent of PDI binding, HSQCs of ^{15}N labelled (30-51, 5-14) BPTI were obtained again, but this time titrating in smaller quantities of PDI enzyme. The enzyme was added at BPTI:PDI ratios of 200:1, 100:1, 50:1, 25:1, 10:1 and 5:1.


A
Gly28N-H
Gly56N-H
Gly36N-H
Gly57N-H
Ser47N-H
Thr32N-H
Gly12N-H
Thr54N-H
Gly37N-H
Tyr21N-H
Ser38N-H
Ser55N-H
Leu29N-H
Lys26N-H
Glu49N-H
Ala27N-H
Cys51N-H
Phe33N-H
Cys30N-H
Lys15N-H
Asn44N-H
Phe22N-H
Asn43N-H
Asp50N-H
Lys46N-H
Val34N-H
Arg53N-H
Met52N-H
Arg42N-H
Arg17N-H
Ile18N-H
Gln31N-H
Arg39N-H
Tyr23N-H
Phe45N-H
Lys41N-H
Ala40N-H
Ala16N-H
Ala48N-H
Ile19N-H
Arg20N-H
Ala25N-H
Asn24N-H
Ala58aN-H
Tyr35N-H
Ala58bN-H
10
9
8
7
105
110
115
120
125
130
ω1 - 15N (ppm)
ω2 - 1H (ppm)


B
Gly28N-H
Gly56N-H
Gly36N-H
Gly57N-H
Ser47N-H
Thr32N-H
Gly12N-H
Thr54N-H
Gly37N-H
Tyr21N-H
Ser38N-H
Ser55N-H
Leu29N-H
Lys26N-H
Glu49N-H
Ala27N-H
Cys51N-H
Phe33N-H
Cys30N-H
Ile18N-H
Asn44N-H
Phe22N-H
Asn43N-H
Asp50N-H
Lys46N-H
Val34N-H
Arg53N-H
Met52N-H
Lys15N-H
Arg42N-H
Arg17N-H
Arg39N-H
Gln31N-H
Tyr23N-H
Phe45N-H
Ala40N-H
Lys41N-H
Ala16N-H
Ala48N-H
Ala25N-H
Arg20N-H
Asn24N-H
Ile19N-H
Ala58aN-H
Ala58bN-H
10
9
8
7
105
110
115
120
125
130
ω1 - 15N (ppm)
ω2 - 1H (ppm)

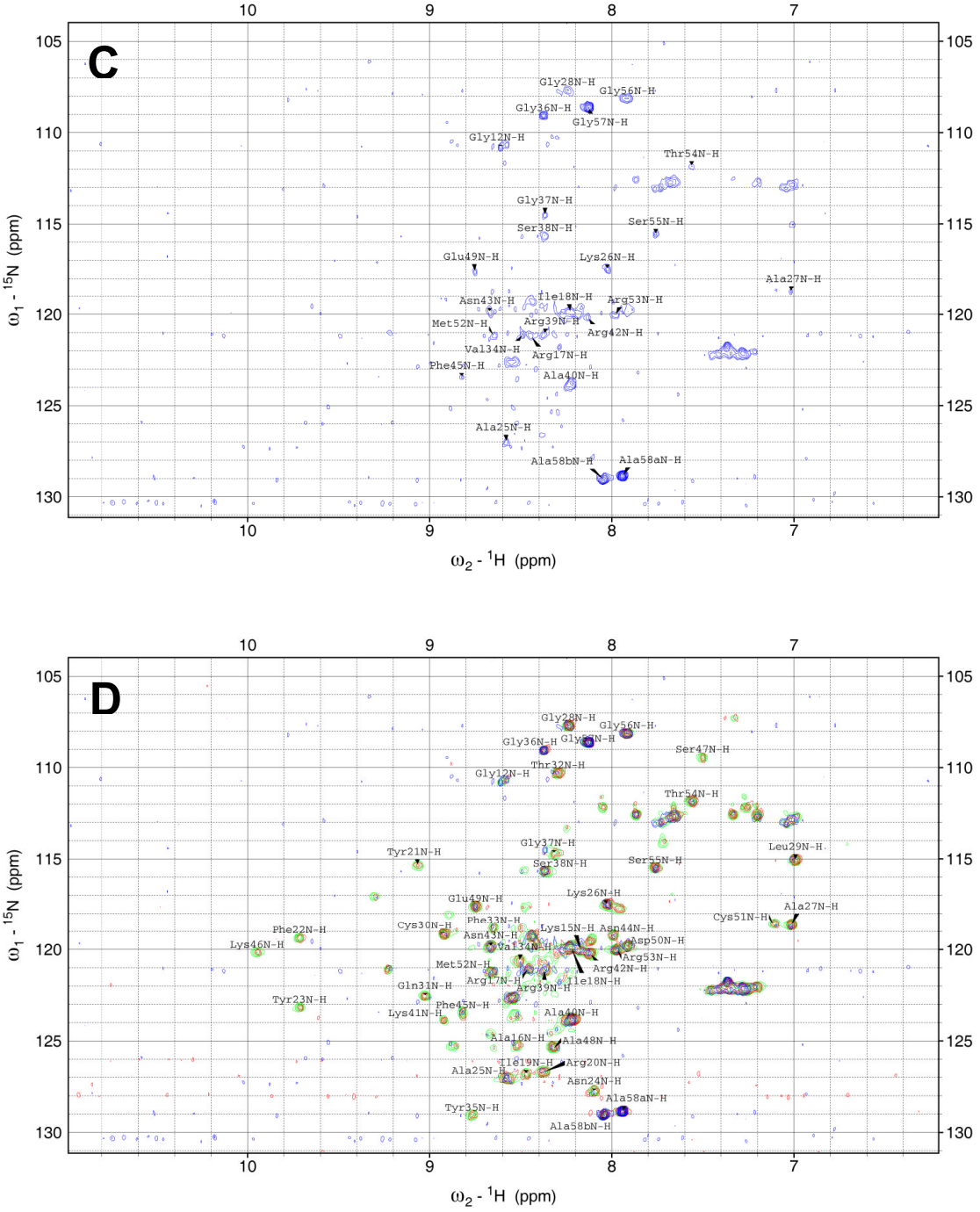


Figure 5.6: HSQC spectra of (30-51, 5-14) BPTI interaction with PDI. A) (30-51, 5-14) BPTI only; B) 50:1 ratio BPTI:PDI; C) 10:1 ratio BPTI:PDI; D) Overlay of spectra from A, B and C. Assigned peaks from (30-51, 5-14) BPTI only spectrum are labelled.

Comparison of each spectrum with a (30-51, 5-14) BPTI only control (no enzyme) showed very little difference at the 200:1 ratio (not shown). At 100:1

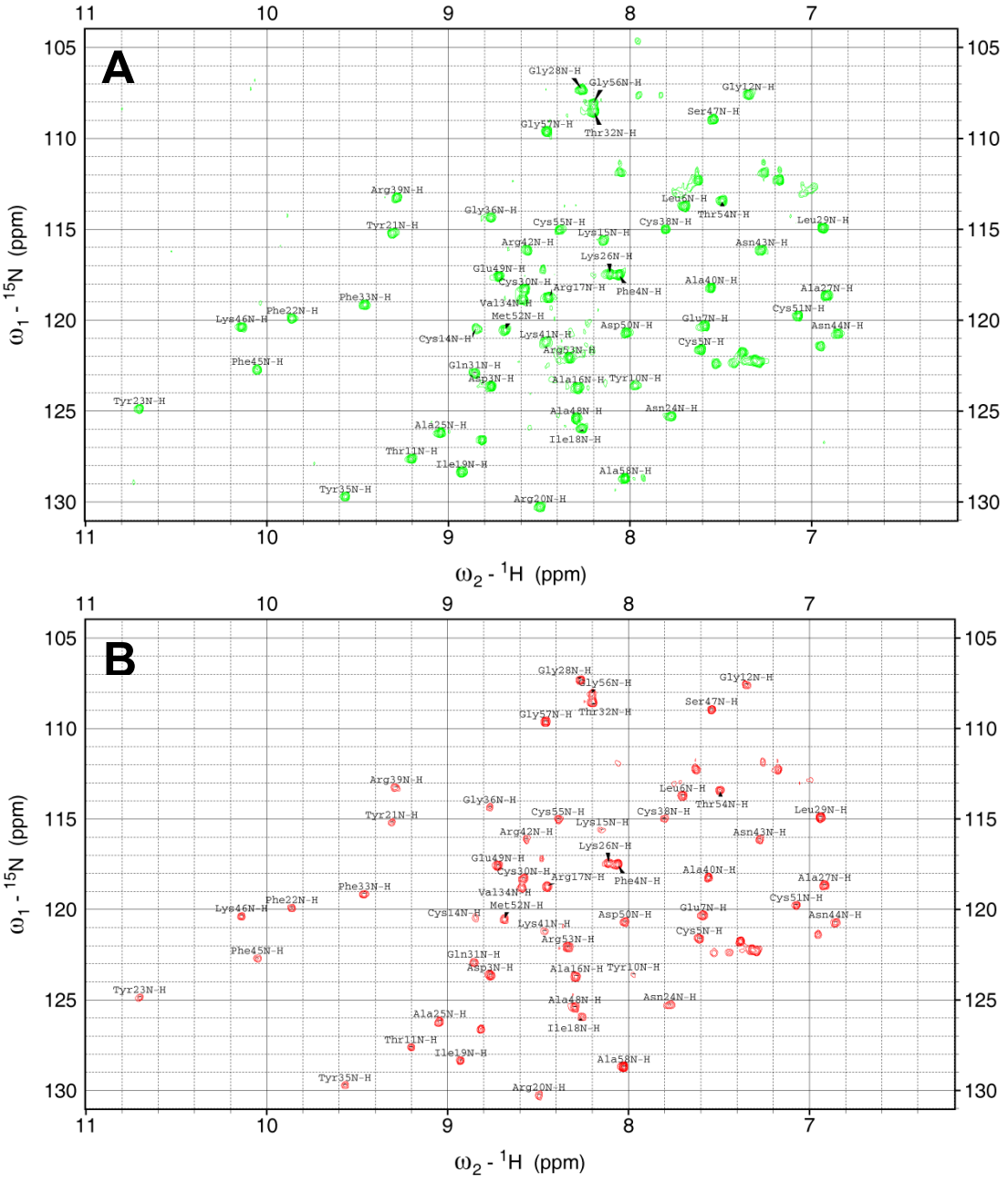
ratio, all assigned peaks are still present, but a reduction in intensity is clear. At 50:1, a further reduction of peak intensity is clear, with some assigned peaks no longer visible (Figure 5.6B). This trend continues as the molar concentration of PDI is increased. At 10:1 ratio, most assigned peaks are no longer visible, and most that remain are just above the noise level (Figure 5.6C). An overlay of HSQC spectra with various quantities of PDI illustrates the peaks that disappear most rapidly and those that persist even at higher concentrations (Figure 5.6D).

5.5. Wild Type BPTI interaction with PDI

Since PDI is a folding enzyme, it would be expected to have a much higher affinity to an unfolded or partly-folded substrate protein than to a protein in its native state. To investigate this, wild type BPTI was examined in a series of HSQCs whereby the PDI enzyme was titrated in at increasing molar concentrations. Since the partly-folded (30-51, 5-14) BPTI intermediate showed significant disappearance of peaks at BPTI:PDI ratio of 10:1 and wild type BPTI was expected to bind with less affinity, this seemed like a logical starting point.

As expected, the effect of PDI at BPTI:PDI 10:1 was much less than with (30-51, 5-14) BPTI (Figure 5.7B). All assigned peaks are still clearly present and well resolved, although when compared to wild type BPTI alone (Figure 5.7A), a reduction in peak intensity is evident. The assigned peaks are still present at a 5:1 ratio, with the intensities further reduced (not shown). It is not until a 1:1 ratio is reached that a dramatic reduction of visible peaks is observed (Figure 5.7C). The only remaining assigned peaks were for Asp3, Gly57 and Ala58. An overlay of HSQC spectra of wild type BPTI with various concentrations of PDI highlights the peaks that remain visible even at high concentrations (Figure 5.7D). As with (30-51, 5-14) BPTI, the C-terminal Gly57 and Ala58 residues remain visible. In addition, Asp3, the first residue detected at the N-terminus of the protein, also

remains visible, but with significantly reduced intensity. Chemical shift assignments for all BPTI with PDI titration spectra are shown in Appendix E.



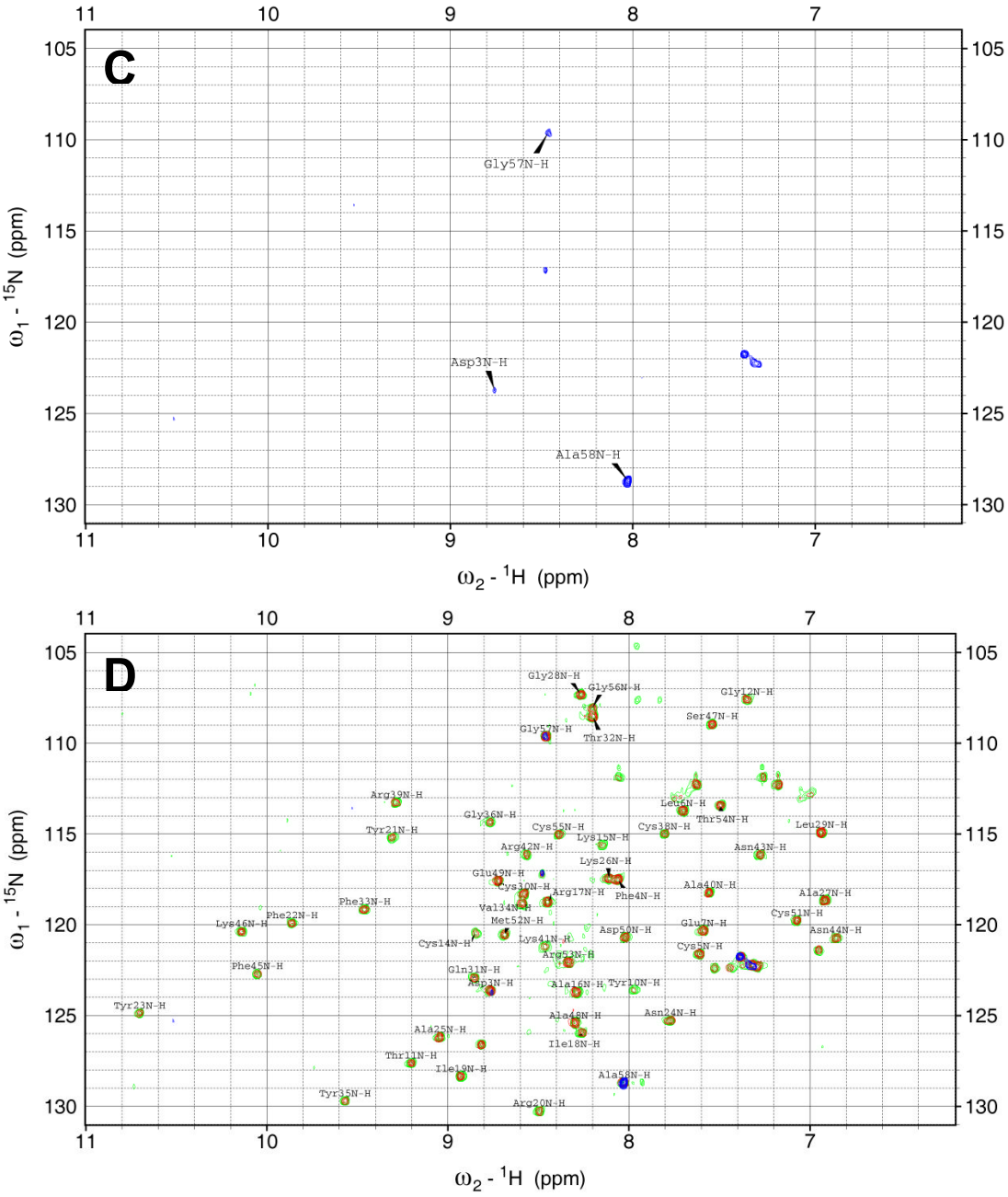


Figure 5.7: HSQC spectra of wild type BPTI interaction with PDI. A) BPTI only; B) 10:1 ratio BPTI:PDI; C) 1:1 ratio BPTI:PDI; D) Overlay of spectra from A, B and C. Assigned peaks from BPTI only spectrum are labelled.

To examine closer the effects of adding PDI, the heights of each assigned peak were plotted against sequence number of the mature protein. This was done for both wild type and (30-51, 5-14) BPTI. Since the greatest distinction between

constructs seems to occur at BPTI:PDI ratio of 10:1, these plots are shown and compared to BPTI before the addition of PDI (Figure 5.8). A full list of all normalised peak heights is available in Appendix F.

As expected, the spectra obtained before addition of PDI are very similar to those obtained for control experiments before hydrogen deuterium exchange (compare with Figure 4.17), showing good repeatability of this standard HSQC experiment between separate ^{15}N labelled sample preparations for both wild type and (30-51, 5-14) BPTI.

With wild type BPTI, all assigned peaks remain at the 10:1 ratio, but all are reduced in intensity. Interestingly, despite this reduction of intensity, due to some binding with PDI, the peak heights remain of the same relative magnitude throughout the protein. This is consistent with what may be expected, whereby unbound BPTI molecules continue to contribute to the chemical shift resonance signal for each peak, but the BPTI bound in the PDI/BPTI complex no longer contribute, resulting in an overall weaker intensity.

Since wild type BPTI is supposed to be natively folded, why should it show any signs of binding to PDI at all? As well as its primary role as a folding enzyme, PDI can also function as a molecular chaperone (Cai, Wang et al. 1994). Even though wild type BPTI exists predominantly in the native state, it is a dynamic molecule and will explore different conformations, occasionally exposing hydrophobic residues. This is especially true for BPTI, since it contains a high level of random coil structure. These events may enable PDI to bind, and are compatible with the results obtained for hydrogen deuterium exchange.

With the partly-folded (30-51, 5-14) BPTI intermediate, at a BPTI:PDI ratio of 10:1 most peaks have disappeared and the normalised peak height of most peaks that remain is very low, not much above the noise level of the HSQC

spectrum. This clearly shows that PDI has a much greater binding affinity for the (30-51, 5-14) BPTI intermediate than for wild type BPTI.

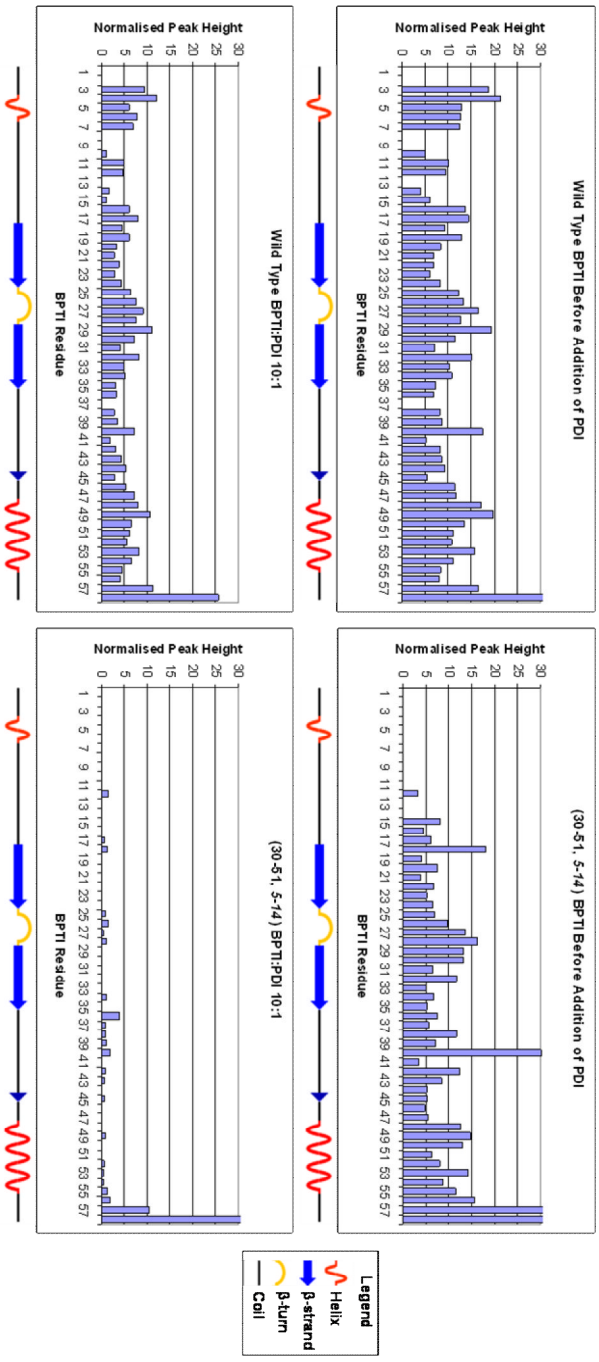


Figure 5.8: Comparison of peak heights for assigned residues of BPTI constructs before and after addition of PDI. Left, wild type BPTI; right, (30-51, 5-14) BPTI intermediate. Top, before addition of PDI; bottom, after addition of PDI to BPTI:PDI ratio 10:1.

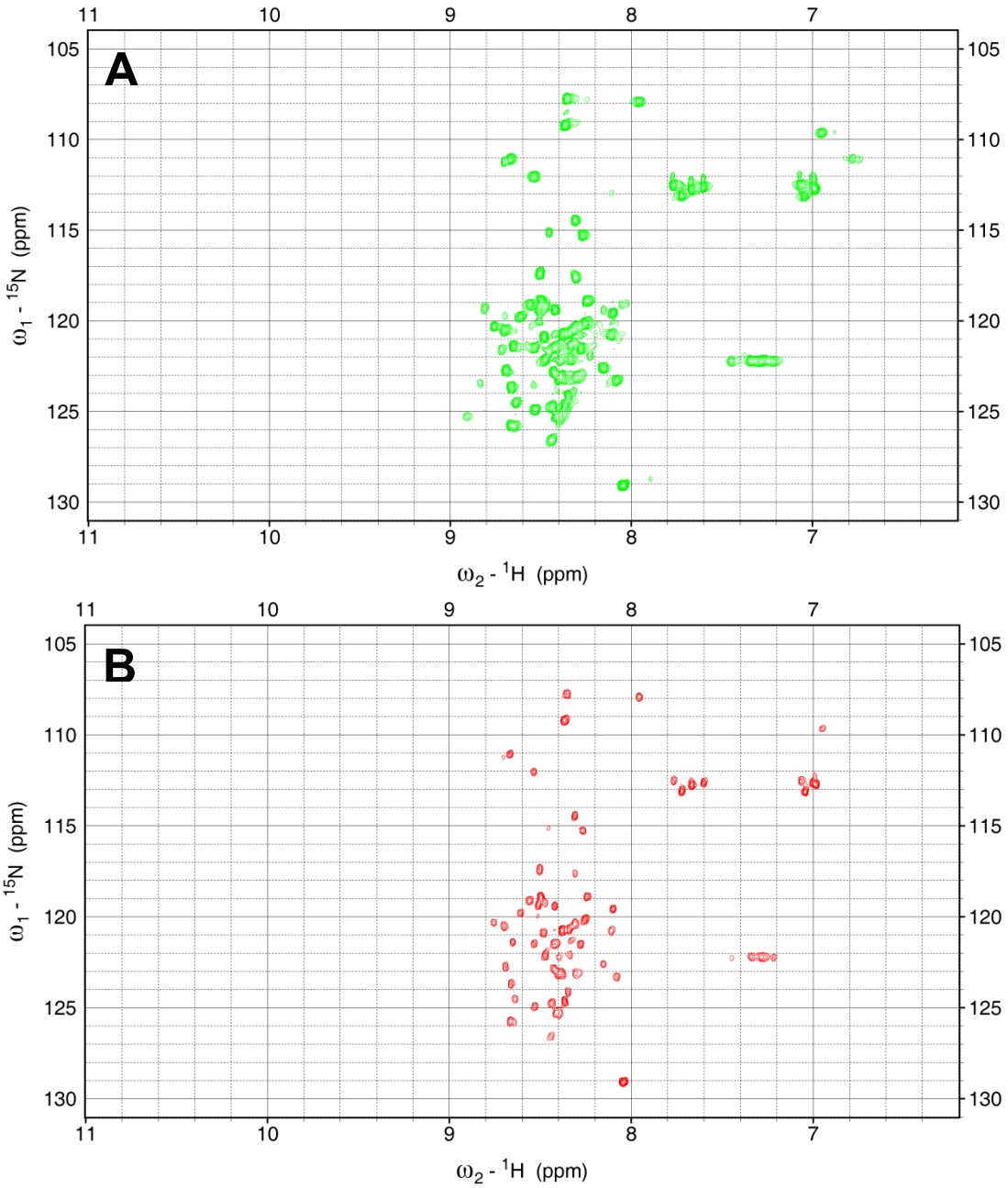
5.6. Reduced BPTI interaction with PDI

Since reduced BPTI is almost completely unfolded even at 5°C, it may be expected that this construct would have the largest affinity for PDI. HSQC spectra were acquired at reduced BPTI:PDI ratios of 100:1, 50:1, 25:1, 10:1 and 5:1. Since reduced BPTI only had a solubility of 100 μM in NMR buffer, acquisition times were increased five fold to maximise the number of visible peaks, as outlined in section 4.4.1.

At 100:1 ratio, a few peaks disappear and the remaining peaks are lower in intensity compared to the HSQC spectrum of reduced BPTI only. At 50:1 ratio, the same peaks remain but the intensity is further reduced (Figure 5.9B). This is comparable to the results for (30-51, 5-14) BPTI intermediate (Figure 5.6B). Ratios of reduced BPTI:PDI of 25:1 and 10:1 see a further reduction in intensity (not shown), but it is not until a 5:1 ratio before most of the original peaks disappear from the spectrum completely (Figure 5.6C). An overlay of HSQC spectra for reduced BPTI with various quantities of PDI is shown in Figure 5.6D.

It is interesting that most peaks are still visible at a reduced BPTI:PDI ratio of 10:1, whereas most have disappeared in the corresponding partly-folded (30-51, 5-14) BPTI spectrum. There could be a number of reasons for this. HSQC spectra for reduced BPTI were acquired at the solubility limit of 100 μM for this construct. At a BPTI:PDI ratio of 10:1, this means that only 10 μM of PDI was present. In contrast, the HSQC spectra of both wild type and (30-51, 5-14) BPTI were acquired at concentrations of 400 μM, resulting in 40 μM of PDI at the 10:1 ratio. However, both the rate of binding of BPTI to PDI and the fraction of BPTI bound in BPTI/PDI complex would be likely to increase with concentration. The more dilute reduced BPTI sample may require higher molar equivalents of PDI in order for binding to occur at the same rate and to the same extent, thus

compensating for the lower collision rate due to dilution. It could also be that the longer acquisition times required to observe peaks of the reduced BPTI spectra mean that lower intensity signals are able to be detected which would otherwise be invisible under the same conditions of (30-51, 5-14) BPTI.



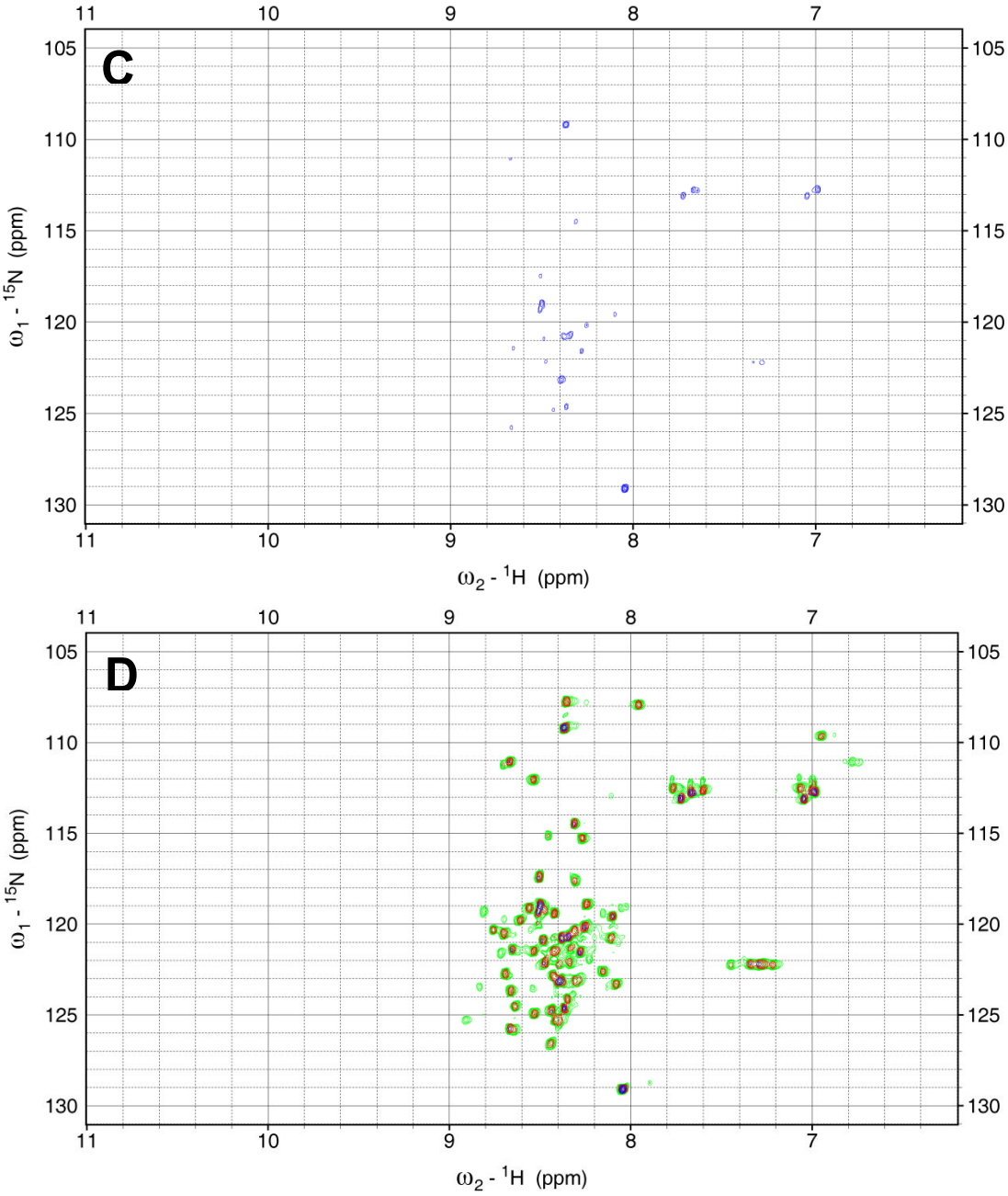


Figure 5.9: HSQC spectra of reduced BPTI interaction with PDI. A) Reduced BPTI only; B) 50:1 ratio BPTI:PDI; C) 5:1 ratio BPTI:PDI; D) Overlay of spectra from A, B and C.

For a direct comparison of wild type and (30-51, 5-14) BPTI to reduced BPTI, all experiments would need to be performed at 100 μM BPTI, so that both the molar equivalent and absolute concentration of PDI would be the same for each

construct. However, this would come at the cost of producing spectra with a lower signal-to-noise ratio for the potentially more soluble samples.

5.7. Discussion

One of the limitations of NMR as a technique for elucidating the structure of macromolecules is that standard methods are limited by the size of the molecule. Higher molecular weight proteins have slower molecular tumbling, resulting in faster relaxation of NMR signal (Yu 1999). A related phenomenon is chemical shift anisotropy, which occurs due to the instantaneous chemical shift of the nucleus depending on the orientation of the molecule. With small molecules, the effect is normally averaged away in solution. However, nuclei still experience a rapid variation of local field from this source, and this stimulates relaxation. This effect is more significant for slowly tumbling molecules (Neuhaus and Williamson 2000).

The introduction of isotopic labelling has enabled the structure of larger proteins to be elucidated by NMR (Cooke 1997), with multidimensional approaches such as HSQC allowing structures up to 35 kDa to be determined (Gronenborn and Clore 1995). Some larger protein structures have been solved using a segmentation approach, whereby isotopically labelled proteins are divided into overlapping regions for NMR analysis (Otomo, Ito et al. 1999).

After PDI was titrated into the ^{15}N labelled BPTI solution for the HSQC experiments performed here, the BPTI in a bound BPTI/PDI complex was much too large (~64.6 kDa), so slow tumbling of the complex results in the NMR signal disappearing before ^{15}N-^{1}H coherence was allowed to occur. The tumbling rate of the enzyme was slowed even further by the lower temperature at which the HSQC experiments were performed. Thus, the formation of BPTI/PDI complex could be determined by the reduction of peak intensity in the HSQC spectra.

This chapter has shown that PDI binds to all three BPTI constructs. The reduction of peak intensity correlates with the addition of PDI in a dose dependent manner. For both reduced BPTI and (30-51, 5-14) BPTI, peak intensity is significantly reduced at low sub-stoichiometric concentrations of PDI, with the majority of peaks disappearing with only one fifth molar equivalent PDI. This implies very rapid binding and unbinding of PDI on the NMR timescale. Binding to wild type BPTI only produces a significant loss of assigned peaks at a 1:1 ratio with PDI, indicating a much lower binding affinity for the natively folded protein.

Chapter 6. NMR Spectroscopy of bb'x Region of PDI in the Presence of BPTI

6.1. Introduction

In studying the interaction of PDI with substrate proteins at various stages along the folding pathway, the previous chapter focused on HSQC observations of each BPTI construct and showed the effects of adding PDI to each.

This chapter describes a similar investigation, but focuses on HSQC observations of PDI, titrating in the different BPTI constructs to see their effect on the enzyme. Since PDI is too large for direct NMR spectroscopy, it is desirable to analyse bb'x, which is known to be the main binding region of this enzyme (Klappa, Ruddock et al. 1998). By using ^{15}N labelled bb'x but unlabelled BPTI, only resonances from the bb'x nuclei were observed, since each HSQC detects correlations with ^{15}N nuclei. In contrast to BPTI, all HSQC spectra for bb'x were acquired at 25°C, since the higher temperature produces higher quality data for this larger protein construct.

In preparation, the expression and purification of ^{15}N labelled bb'x samples are described. An HSQC spectrum of bb'x alone was performed and analysed by comparison with a spectrum from a previous study. Separate bb'x samples were then prepared with increasing amounts of BPTI and their HSQC spectra were analysed by comparison with the bb'x only sample. Since bb'x (27.5 kDa) is significantly larger than BPTI (6.6 kDa), the effects of adding substrate were much more subtle than adding PDI to BPTI, with less visible differences between spectra.

6.2. Expression and Purification of bb'x

Expression of His-tagged bb'x was identical to the protocol for PDI (section 2.1). Since HSQC spectra were acquired, ^{15}N isotope labelling of bb'x was performed by growing cultures in minimal medium with ^{15}N labelled ammonium sulphate (section 2.3.2).

6.2.1. IMAC Purification of bb'x

The first step in the purification of His-tagged bb'x used IMAC purification. Using the N-terminal His-tag, the protein bound well to the column and eluted as expected once EDTA was added to the buffer. The collected fractions were analysed using SDS-PAGE to confirm the presence of purified bb'x in the elution (data not shown). The protein was then dialysed overnight into 20 mM sodium phosphate in preparation for ion exchange chromatography.

6.2.2. Ion Exchange Purification of bb'x

The dialysed protein was applied to an anion exchange column for further purification, just like full length PDI. However, unlike full length PDI, which was almost entirely monomeric, Figure 6.1 shows two large overlapping peaks, which were identified by native PAGE as monomer and dimer.

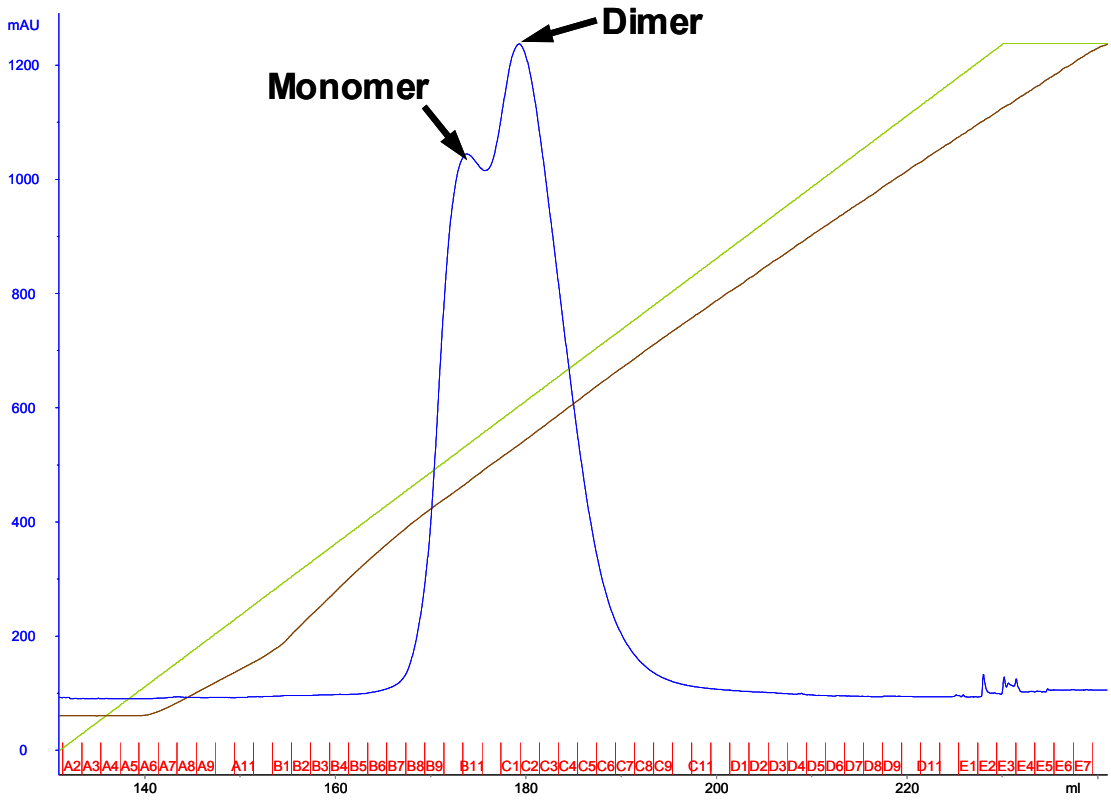


Figure 6.1: Chromatogram from ion exchange purification of bb'x. Blue, absorbance at 280 nm; green, concentration of high salt buffer; brown, conductance; red, eluted fractions. The monomer and dimer elute at slightly different salt concentrations, but there is a very large overlap between peaks.

6.2.3. Size Exclusion to Isolate bb'x Monomer

For the purposes of substrate interaction, only the monomer of bb'x is desired. To separate monomer and dimer, the appropriate fractions from ion exchange elution were pooled and concentrated to 1 ml so that it could be injected onto a gel filtration column. The protocol is detailed in section 2.5.4. The chromatogram of Figure 6.2 clearly shows good separation of monomer and dimer species. Native PAGE analysis of the fractions representing these peaks confirms them to be predominantly monomer and dimer, respectively. Fractions 34-38, containing bb'x monomer, were pooled and stored at 4°C.

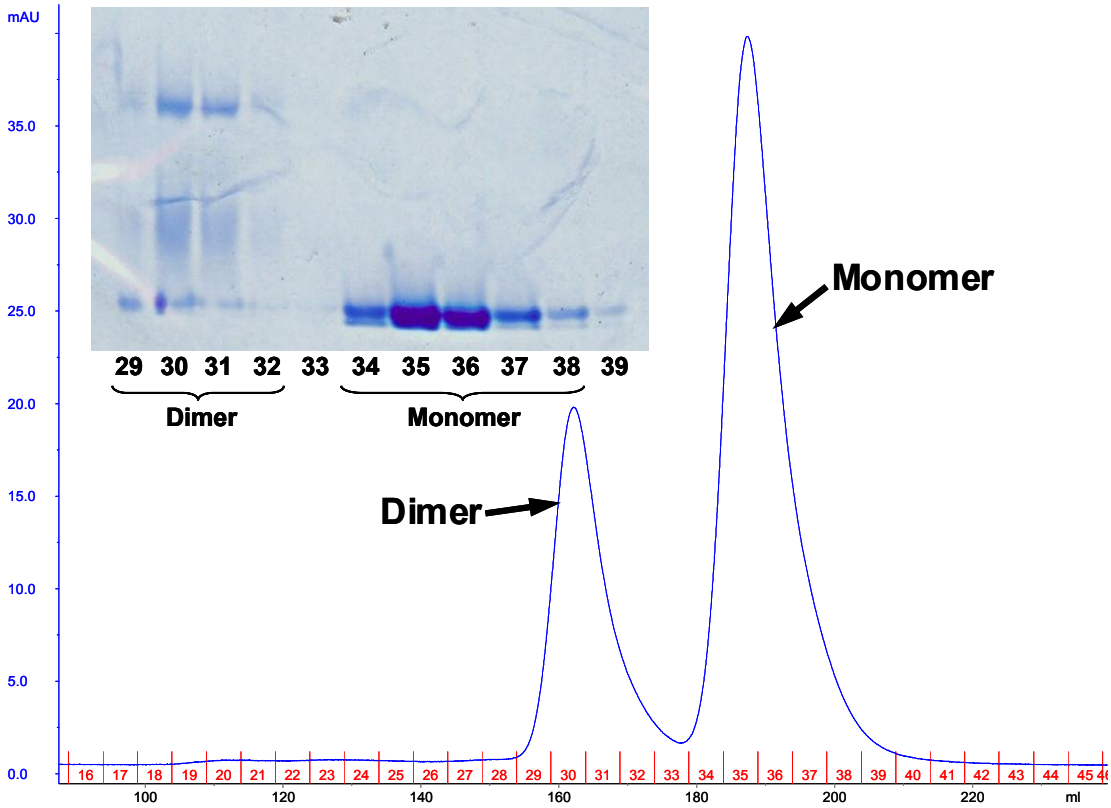


Figure 6.2: Chromatogram from gel filtration of bb'x. Blue, absorbance at 280 nm; red, eluted fractions. The absorbance signal indicates good separation of the monomer and dimer peaks. Inset: native PAGE of fractions eluted from gel filtration of bb'x. Fractions 29-32 show mainly dimer, whereas fractions 34-38 show mainly monomer.

6.2.4. Testing Lyophilisation using Fluorescence Spectroscopy

Since all BPTI constructs and full length PDI were stored by lyophilisation, the same procedure was followed for an aliquot of the bb'x monomer. However, other studies have shown that the freeze/thaw process resulted in dimerisation of the shorter b'x construct, see Figure 3 in (Wallis, Sidhu et al. 2009).

bb'x contains only one tryptophan residue, which is contained in the x region. So, intrinsic fluorescence was used to test if lyophilisation affects the conformation of bb'x. A change in fluorescence would indicate a change in the environment of the x region and hence a change in the conformation of bb'x. The change

occurs since, in the monomer state, x binds to the exposed hydrophobic region of the b' domain, whereas in a dimer the exposed regions of b' domains bind to each other, leaving the x regions free in solution, hence a more hydrophilic environment (Figure 6.3). The fluorescent emission spectra for purified capped monomer and uncapped dimer were previously illustrated in Figure 4B from (Wallis, Sidhu et al. 2009), which show maximum intensities at approximately 330 nm for the monomer and approximately 345 nm for the dimer.

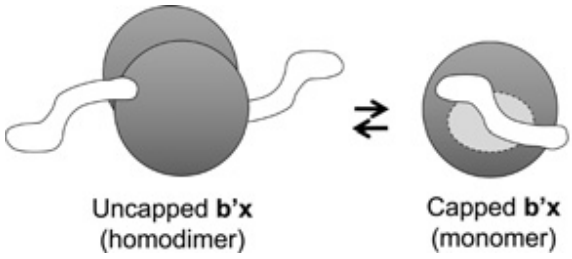


Figure 6.3: Model of interchange between uncapped homodimer and capped monomer of the b'x region of PDI, modified from (Byrne, Sidhu et al. 2009), caused by the exposed hydrophobic regions of the b' and x domains. The same model is valid with the addition of the b domain, *i.e.* the bb'x region.

After testing the effects of lyophilisation on bb'x, Figure 6.4 shows that a significant difference in fluorescence does occur. Before lyophilisation (blue), maximum intensity was reached at 332 nm, showing the sample is mainly monomeric. However, after lyophilisation there is a significant decrease in intensity and the maximum intensity is red-shifted to 336 nm, indicating that more dimer is present in the sample. This suggests that lyophilisation is not appropriate for the preparation of bb'x. It must be prepared fresh, stored at 4°C, and used before degrading.

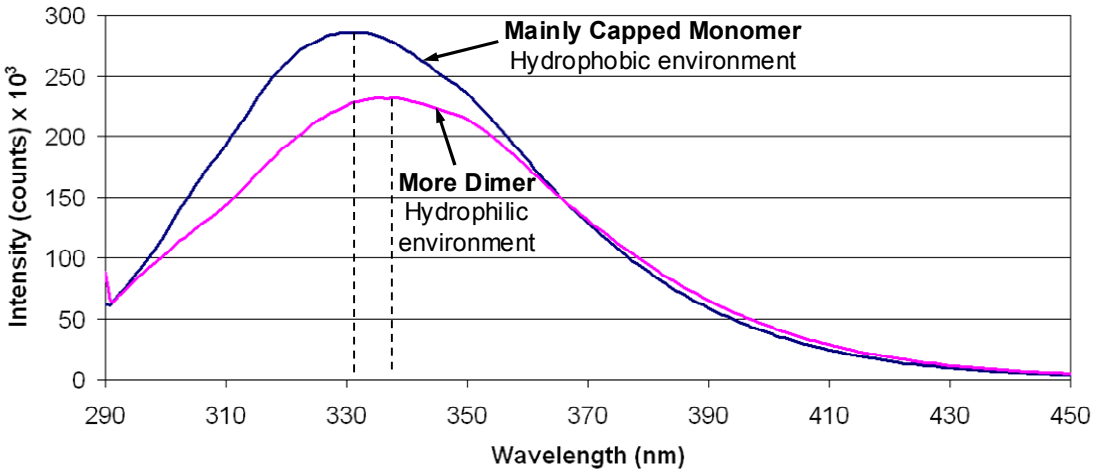


Figure 6.4: Intrinsic fluorescence of bb'x before and after lyophilisation. Freshly purified bb'x stored at 4°C (blue) shows fluorescence indicating the x-region of the protein is in a hydrophobic environment, thus will mainly exist as a capped monomer. Lyophilised bb'x, which was frozen during the lyophilisation process, when resuspended in buffer shows fluorescence indicating the x-region of the protein is in a more hydrophilic environment, suggesting some dimer formation of the protein.

6.3. HSQC Spectra of ^{15}N labelled bb'x with Substrate Titrations

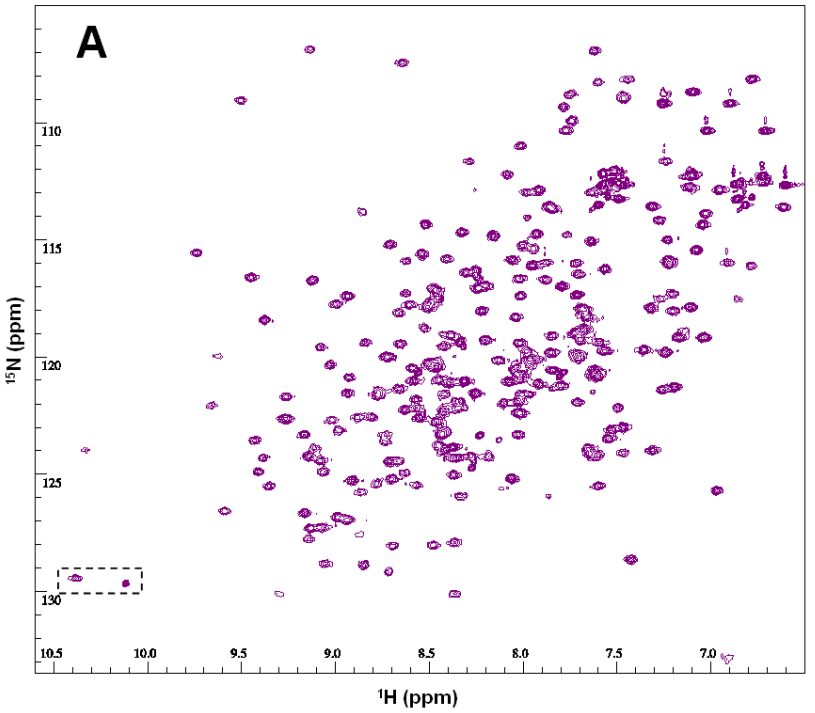
Acquisition of HSQC spectra of bb'x were performed as described in section 2.12.6.3. In contrast to HSQC spectra of BPTI, all bb'x NMR experiments were performed at 25°C, since previous data showed this temperature to provide a good balance between spectral resolution (better at higher temperatures) and protein stability (better at lower temperatures) (Byrne, Sidhu et al. 2009).

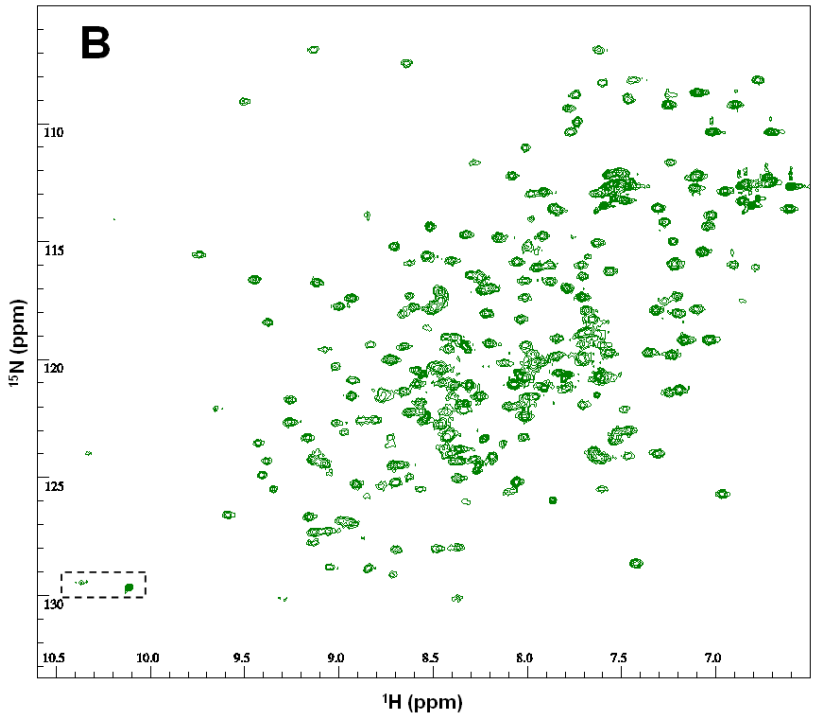
As a control, an HSQC spectrum was acquired for bb'x alone (Figure 6.5A). This was compared to a previously assigned HSQC experiment run under the same conditions (Byrne, Sidhu et al. 2009). An overlay showed that the two spectra were very similar, allowing rapid assignment of the control spectrum.

HSQC spectra could then be obtained for bb'x in the presence of various concentrations of BPTI. No prior knowledge was available for the interaction of

these proteins, apart from results for spectra of ^{15}N labelled BPTI detailed in the previous chapter. So, for both wild type and (30-51, 5-14) BPTI, separate samples were prepared at bb'x:BPTI ratios of 25:1, 5:1 and 1:1. The concentration of ^{15}N labelled bb'x was the same for each sample (250 μM), but the concentration of BPTI was varied at 10 μM, 50 μM and 250 μM respectively. Figure 6.5 shows example spectra of these titrations at a bb'x:BPTI ratio of 5:1 for wild type (Figure 6.5B) and (30-51, 5-14) BPTI (Figure 6.5C).

Since bb'x is relatively large (27.5 kDa) compared to the BPTI substrate (6.6 kDa), the overall effect of adding substrate was not visibly apparent, with only small differences observed when HSQC spectra of titrated samples were overlaid onto the control spectrum (Figure 6.5D).


A
110
115
120
125
130
^{15}N (ppm)
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
^{1}H (ppm)


B
110
115
120
125
130
^{15}N (ppm)
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
^{1}H (ppm)

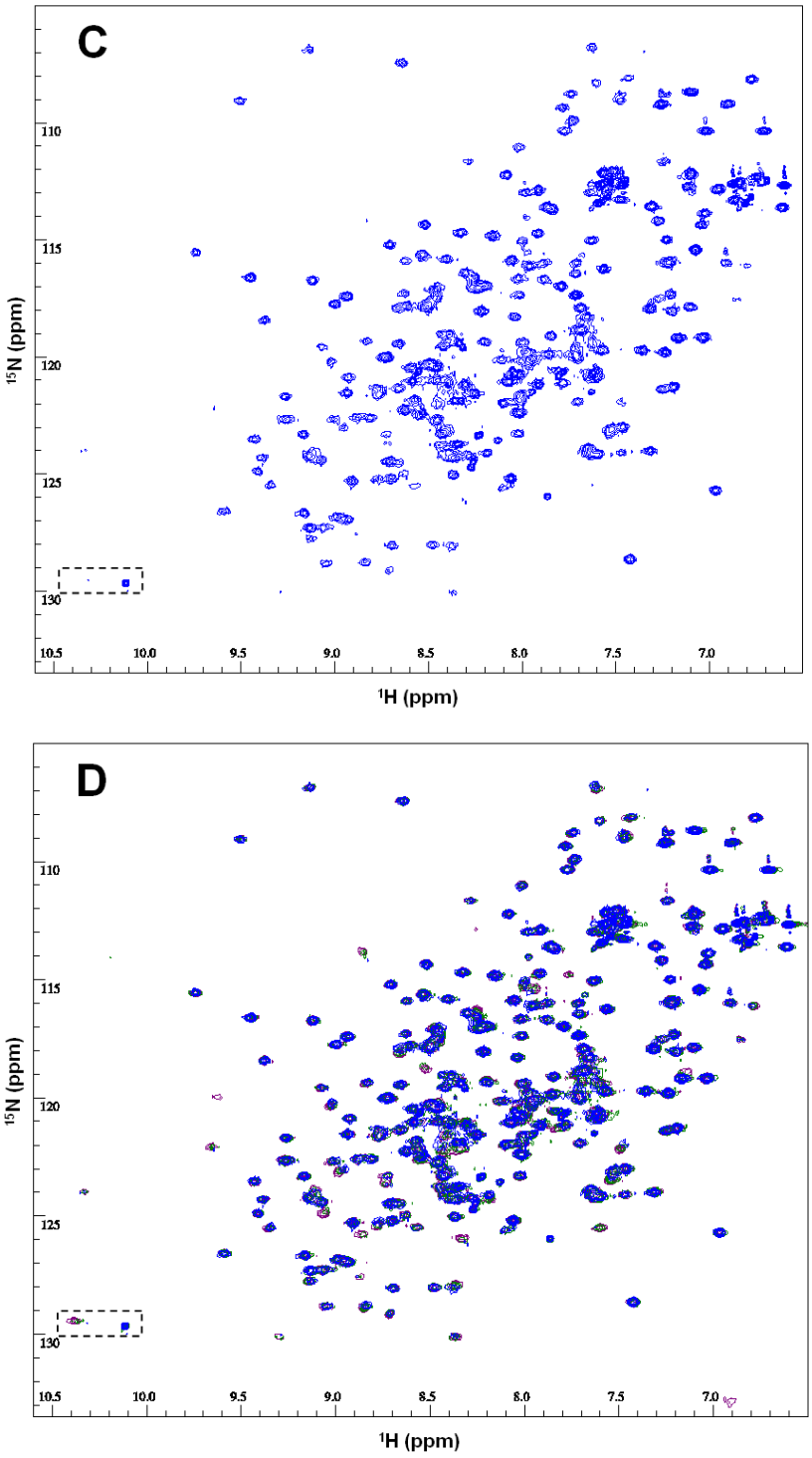


Figure 6.5: HSQC spectra of ^{15}N labelled bb'x at 25°C both in the absence of substrate and at a bb'x:BPTI ratio of 5:1. A) bb'x only; B) bb'x with wild type BPTI; C) bb'x with (30-51, 5-14) BPTI; D) Overlay of spectra from A, B and C. Peaks for indole resonances of Trp347 are highlighted with a dashed box.

6.4. Minimal Shift Mapping between bb'x with and without BPTI Substrate

One way to closely analyse the subtle differences between bb'x spectra before and after the addition of substrate is to estimate the perturbation of chemical shifts from assigned peaks of bb'x alone to the unassigned peaks of bb'x with the substrate added. Regions where large perturbations occur indicate a change in chemical environment and imply a significant role in substrate binding for this part of the protein. This technique has the advantage of being informative about significant regions of the protein that show a change in environment without having to go through laborious processes to accurately assign all chemical shifts in the each sample with substrate added. Thus, it is a useful technique for comparison of one control sample with many experimental spectra, as in the comparison of bb'x alone with bb'x plus substrate.

First, peaks were picked in each HSQC spectrum automatically using CCPNmr Analysis software and visual inspection ensured erroneous or duplicate peaks were removed. Then lists of ^{1}H and ^{15}N resonances were exported for each spectrum. These were compared with each assigned peak of the bb'x only control spectrum and in each case the nearest peak was identified. The change in chemical shift was then calculated for each assigned residue of bb'x, as described in section 2.12.7.3.

Once minimal shift maps had been calculated for bb'x at all bb'x:BPTI ratios, the maps of wild type and (30-51, 5-14) BPTI could be compared. At a 25:1 ratio, the addition of either BPTI construct shows only low levels of chemical shift perturbation. More interestingly, however, are the differences observed at the intermediate bb'x:BPTI ratio of 5:1 (Figure 6.6).

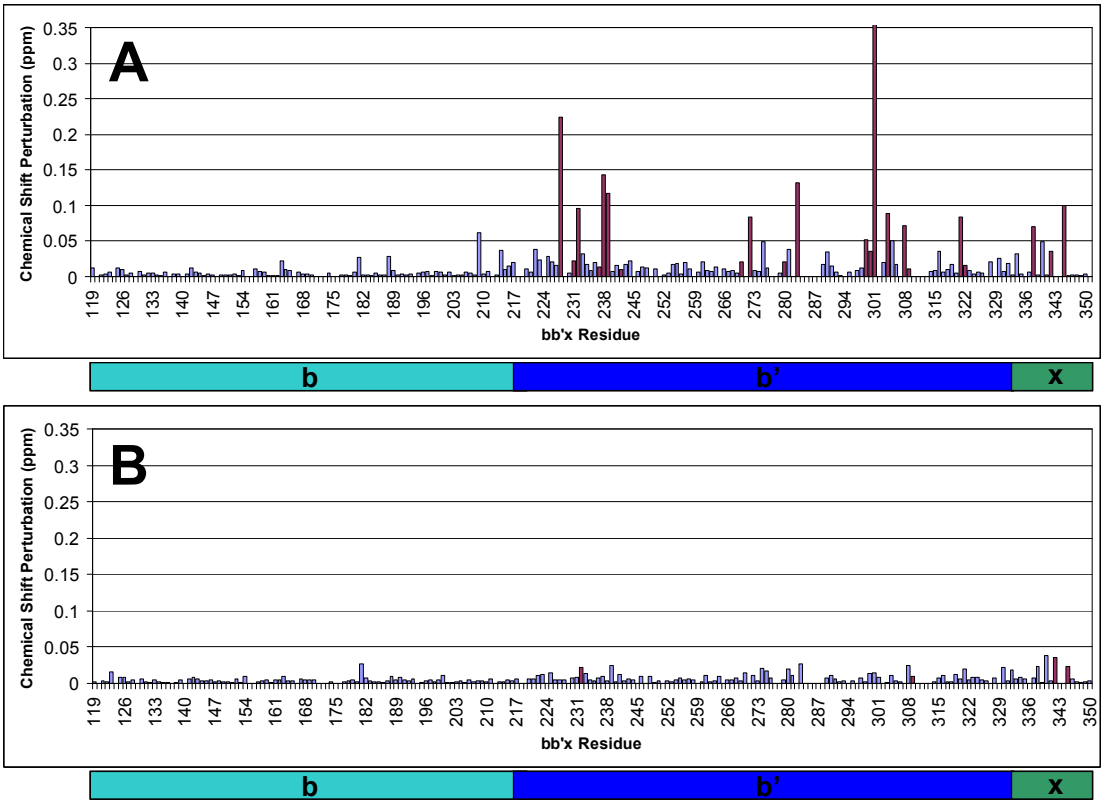


Figure 6.6: Chemical shift perturbations (to nearest peak) for bb'x alone compared to bb'x:BPTI at a 5:1 ratio. A) (30-51, 5-14) BPTI; B) wild type BPTI. Shift perturbations represent a change in chemical shift of an assigned peak in the bb'x only spectrum to the nearest identified peak in the spectrum of bb'x in the presence of BPTI. Since the scale of chemical shifts is larger in the ^{15}N axis than in the ^{1}H axis, chemical shift perturbations were calculated as a weighted sum, $^{1}/_{6}|\Delta^{15}N| + |\Delta^{1}H|$. Blue bars represent shifts of peaks identified in the target spectrum; maroon bars represent shifts from peaks that were unable to be identified in the target spectrum, presumed to be due to line broadening. The graphic below each graph indicates visually the position of each domain in the bb'x construct.

At this bb'x:BPTI ratio, the addition of wild type BPTI shows very little perturbation of chemical shifts in any part of the protein, suggesting that very little protein-substrate interaction is occurring. However, the addition of (30-51, 5-14) BPTI shows significant perturbation in some regions, suggesting bb'x interaction with this folding intermediate alters the chemical environment of these regions. Interestingly, the largest perturbations occur in the b' domain, but very little perturbation of chemical shifts occur in the b domain. This suggests

that even with this full length protein substrate (albeit a small protein) the b' domain dominates the binding of substrates with the b domain remaining largely unaffected. Visual comparison between the bb'x only spectrum and the spectrum of bb'x in the presence of (30-51, 5-14) BPTI indicated that many of the residues showing a large chemical shift perturbation no longer showed a visible peak after (30-51, 5-14) had been added, presumably due to line broadening (Figure 6.6, maroon bars).

By contrast, at the 1:1 ratio, the addition of either BPTI construct shows high levels of chemical shift perturbation (Figure 6.7), with the spectrum produced in the presence of the partly folded (30-51, 5-14) BPTI still showing larger changes in shift than the spectrum in the presence of wild type BPTI.

As with the addition of (30-51, 5-14) BPTI at lower concentration, the addition of wild type BPTI at a 1:1 ratio shows the vast majority of perturbations occurring in the b' domain, with comparatively few residues perturbed in the b domain. As before, when peaks between spectra were compared, many of the residues showing a large chemical shift perturbation seemed to be no longer represented by a visible peak in spectra containing BPTI (Figure 6.7, maroon bars). This is most likely due to binding of the substrate causing line broadening to such an extent that no peak is visible above the noise level, see Figure 6.8C. In analysing shift perturbations of residues with peaks identified even in the presence of BPTI (Figure 6.7, blue bars), it is interesting that of the two that are above 0.1 ppm, one is in the b domain, Leu188, suggesting that this may have an important role in substrate binding.

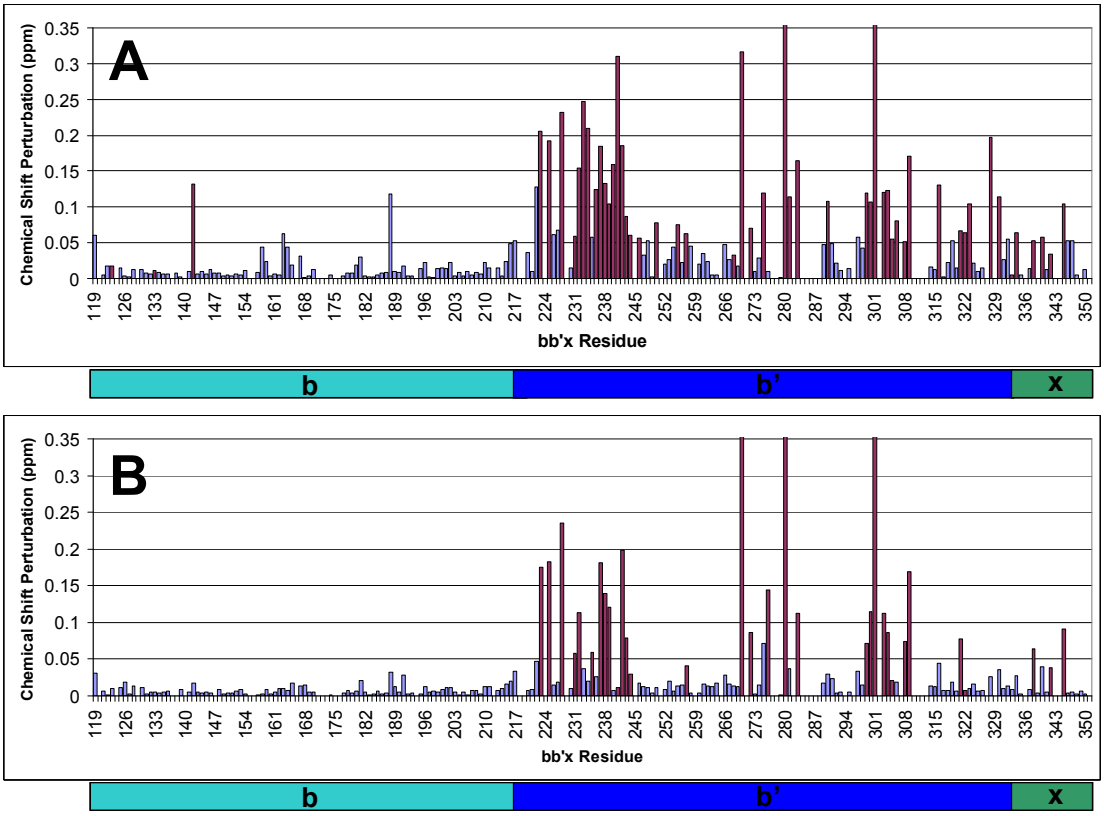


Figure 6.7: Chemical shift perturbations (to nearest peak) for bb'x alone compared to bb'x:BPTI at a 1:1 ratio. A) (30-51, 5-14) BPTI; B) wild type BPTI. Shift perturbations represent a change in chemical shift of an assigned peak in the bb'x only spectrum to the nearest identified peak in the spectrum of bb'x in the presence of BPTI. Since the scale of chemical shifts is larger in the ^{15}N axis than in the ^{1}H axis, chemical shift perturbations were calculated as a weighted sum, $^{1}/_{6}|\Delta^{15}N| + |\Delta^{1}H|$. Blue bars represent shifts of peaks identified in the target spectrum; maroon bars represent shifts from peaks that were unable to be identified in the target spectrum, presumed to be due to line broadening. The graphic below each graph indicates visually the position of each domain in the bb'x construct.

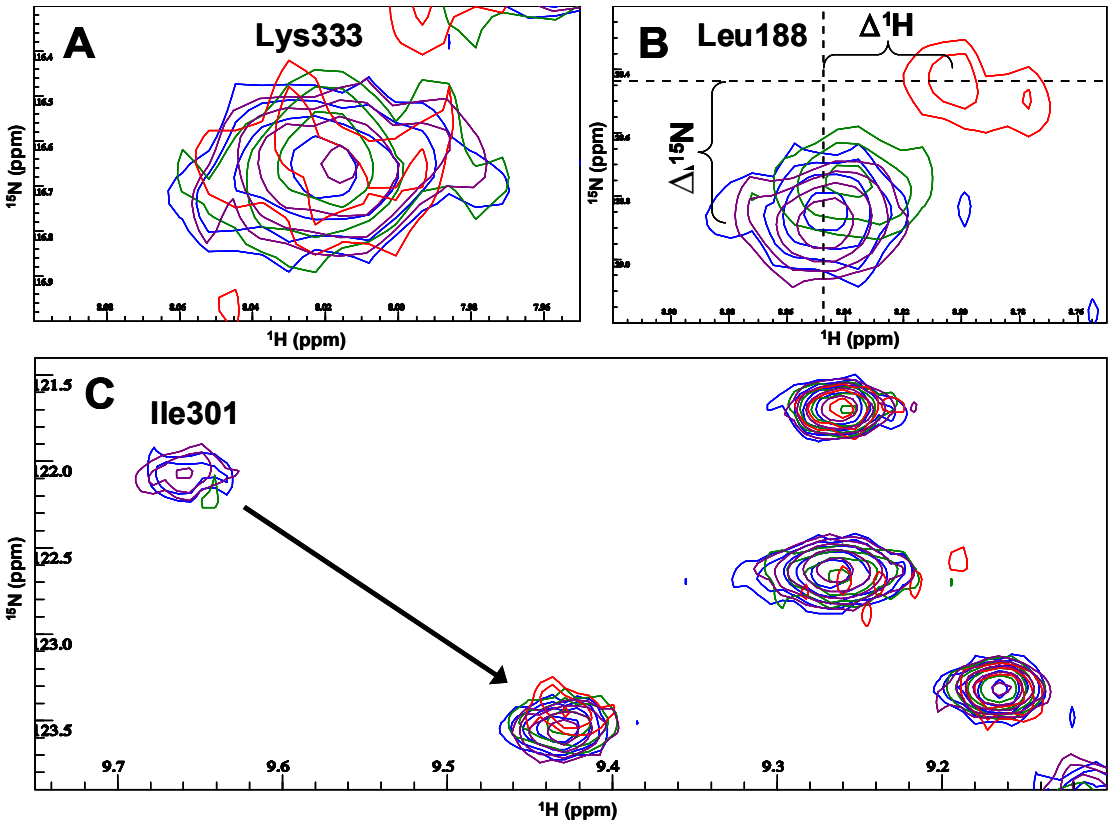


Figure 6.8: Chemical shifts of bb'x residues in the absence and presence of various concentrations of (30-51, 5-14) BPTI. Purple, bb'x only; blue, bb'x:BPTI ratio of 25:1; green, bb'x:BPTI ratio of 5:1; red, bb'x:BPTI ratio of 1:1. A) Lys333 shows no perturbation of chemical shift, suggesting that the backbone of this residue is not affected by substrate binding B) Leu188 shows increasing perturbation of chemical shift with increasing concentration of BPTI substrate, suggesting this residue may have a role in substrate binding. The extent of chemical shift perturbation is measured based on the change in shift from the peak of the bb'x only spectrum (purple) to the corresponding peak in the bb'x spectrum in the presence of substrate. An example is shown for bb'x in the presence of (30-51, 5-14) BPTI at bb'x:BPTI ratio of 1:1 (red). The shift perturbations are highlighted using dashed lines, showing the change in the ^{15}N axis, Δ^{15}N, and in the ^{1}H axis, Δ^{1}H. Since ^{15}N shifts are much larger than ^{1}H shifts and because the magnitude is important but not the direction of change, the chemical shift perturbation is calculated as a weighted sum: $^{1}/_{6}|\Delta^{15}N| + |\Delta^{1}H|$. A reduction in signal intensity is also seen with increasing concentration of substrate, probably due to line broadening. C) Line broadening of Ile301 results in loss of signal below the noise level of the spectrum at higher concentrations of BPTI substrate. Since minimal shift mapping detects the nearest peak in the unassigned spectrum, this will cause a large chemical shift perturbation to be calculated for this residue at higher concentrations (arrow).

Average minimal shifts across the whole bb'x region were calculated for both constructs at each bb'x:BPTI ratio, with residues showing the most significant perturbations identified and summarised (Table 6.1). Most large shifts were observed in the spectrum with a 1:1 ratio of (30-51, 5-14) BPTI and these were mapped to the crystal structure of bb', showing visually where the most significant perturbations are occurring (Figure 6.9). This highlights the significance of the b' domain (right) compared to the b domain (left) in binding of the (30-51, 5-14) BPTI substrate.

Table 6.1: Summary of minimal shift changes for bb'x in the presence of wild type and (30-51, 5-14) BPTI. All HSQC spectra were acquired using 250 µM bb'x.

	Wild Type BPTI			**(30-51, 5-14) BPTI**		
	Ratio of bb'x:BPTI			**Ratio of bb'x:BPTI**		
	25:1	5:1	1:1	25:1	5:1	1:1
Average Minimal Shift Change(ppm)	4.42E-03	6.80E-03	3.82E-02	5.42E-03	1.92E-02	5.67E-02
Total Shifts > 0.06	0	0	25	0	13	45
Total Shifts > 0.10	0	0	16	0	5	33
0.06 < Total Shifts < 0.10	0	0	9	0	8	12

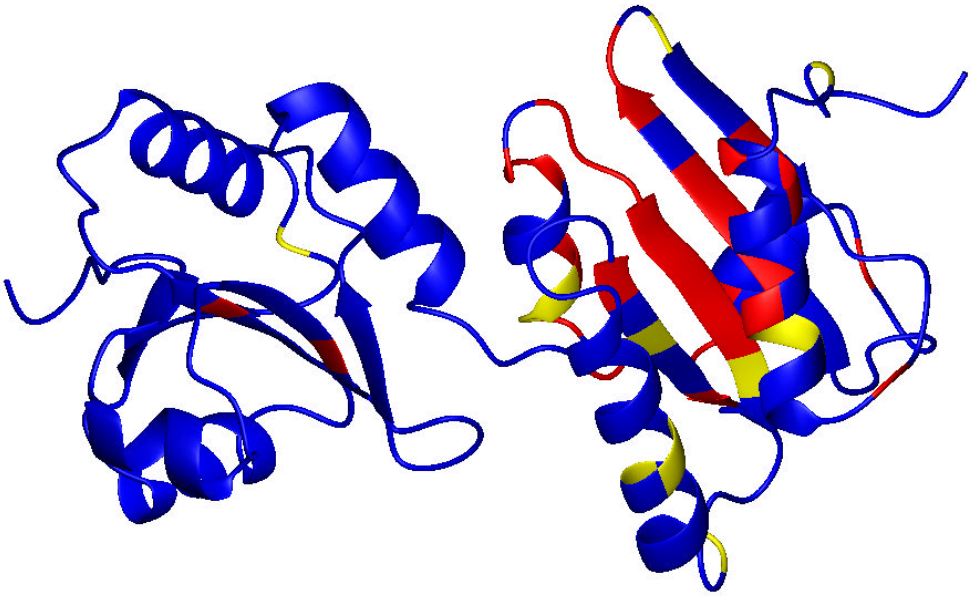
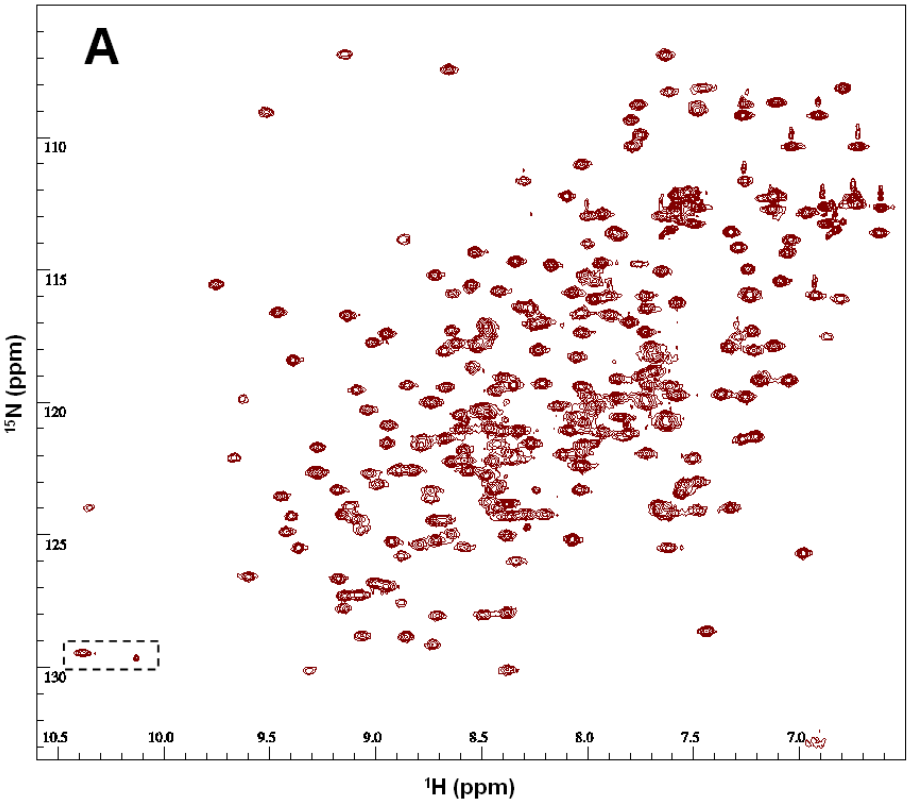
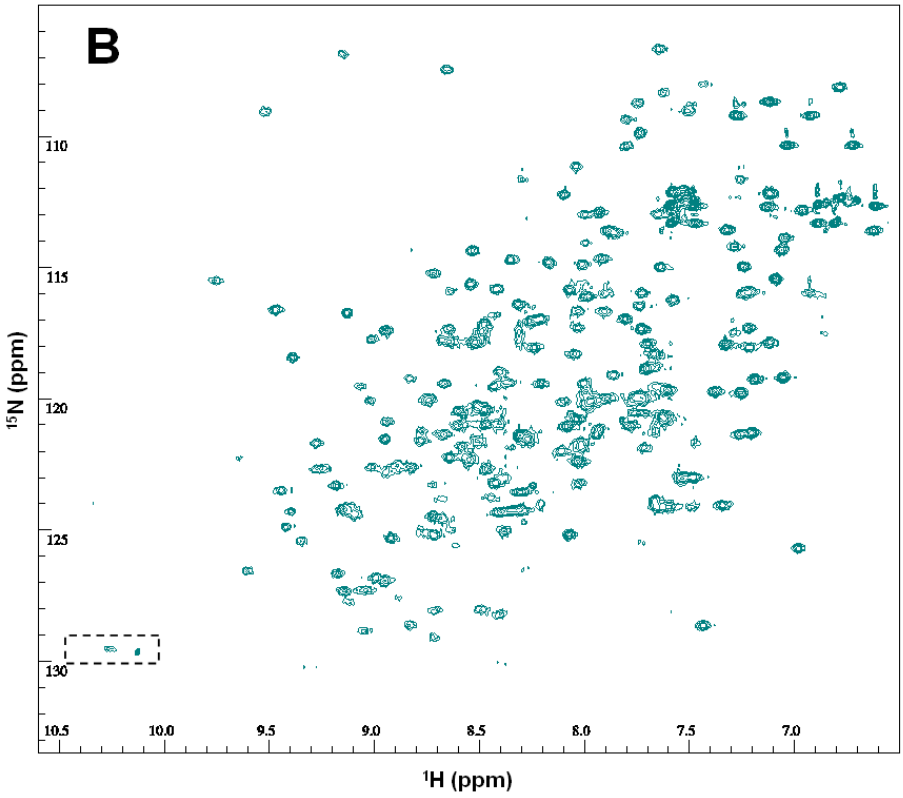


Figure 6.9: Illustration of bb' mapping regions of chemical shift perturbation occurring with the addition of (30-51, 5-14) BPTI, at bb'x:BPTI ratio of 1:1. Yellow, Perturbations > 0.06 ppm; red, perturbations > 0.1. Illustrated using structure of bb' from PDB ID 2K18 (Denisov, Maattanen et al. 2009).

6.5. HSQC Spectra of ^{15}N labelled bb'x with Titrations of Reduced BPTI

Due to the limited solubility of reduced BPTI, the range of bb'x:BPTI ratios used was different than for wild type or (30-51, 5-14) BPTI. As before, samples still contained 250 μM bb'x, with HSQC spectra acquired at ratios of 25:1 and 8.3:1, using reduced BPTI at concentrations of 10 μM and 30 μM respectively. To investigate a bb'x sample with a larger ratio of reduced BPTI, a sample was prepared at just 100 μM bb'x and 43 μM reduced BPTI (ratio 2.3:1). To compensate for the reduced molar concentration, a longer acquisition time was applied. Spectra for the 25:1 and 2.3:1 ratios are shown in Figure 6.10.


A
110
115
120
125
130
^{15}N (ppm)
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
^{1}H (ppm)


B
110
115
120
125
130
^{15}N (ppm)
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
^{1}H (ppm)

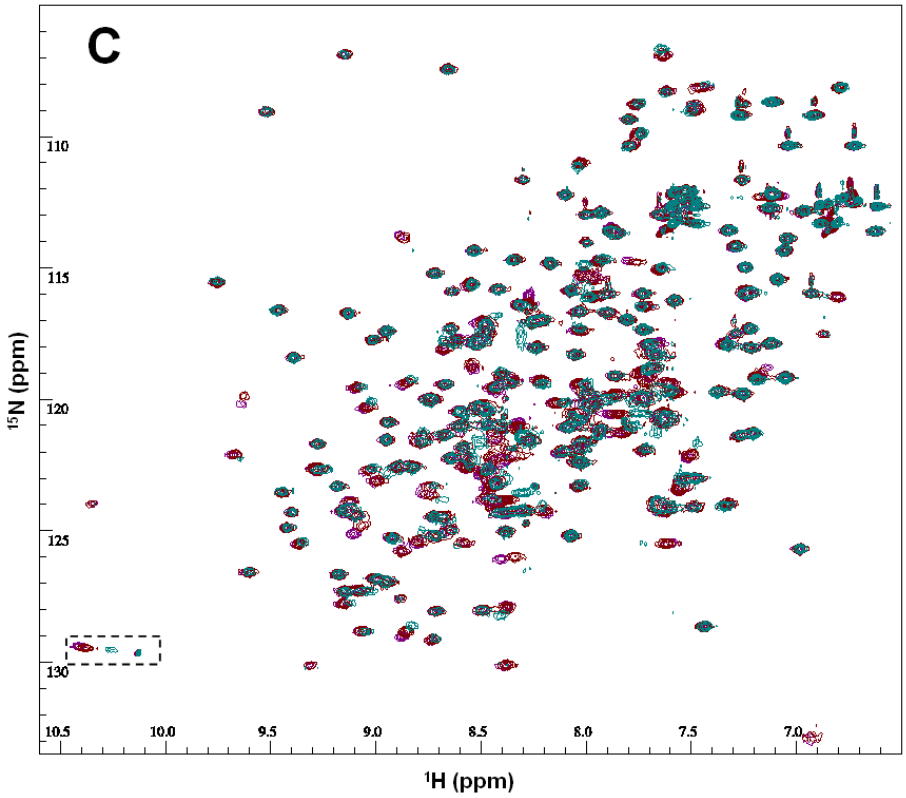


Figure 6.10: HSQC spectra of ^{15}N labelled bb'x at 25°C in the presence of reduced BPTI. A) at bb'x:BPTI ratio of 25:1 using 250 µM bb'x; B) at bb'x:BPTI ratio of 2.3:1 using 100 µM bb'x; C) Spectra from A and B overlaid on top of bb'x only spectrum (purple). Peaks for indole resonances of Trp347 are highlighted with a dashed box.

Analysis of spectra containing reduced BPTI followed the same procedures as for other BPTI constructs, but since the bb'x:BPTI ratio of 25:1 was the only ratio matching wild type and (30-51, 5-14) BPTI HSQC spectra, this was the main spectrum used for comparison with the other BPTI constructs.

Minimal shift maps of bb'x comparing before and after the addition of reduced BPTI were generated as described earlier. For the 25:1 ratio, the shift maps in the presence of reduced BPTI and (30-51, 5-14) BPTI were compared (Figure 6.11). Although the shift perturbations are small for both constructs, they are still noticeably larger in the presence of reduced BPTI compared with the partly folded protein, particularly towards the C-terminal end of the bb'x region. This

would suggest that bb'x has a slightly greater affinity to reduced BPTI than to the (30-51, 5-14) BPTI at this low concentration.

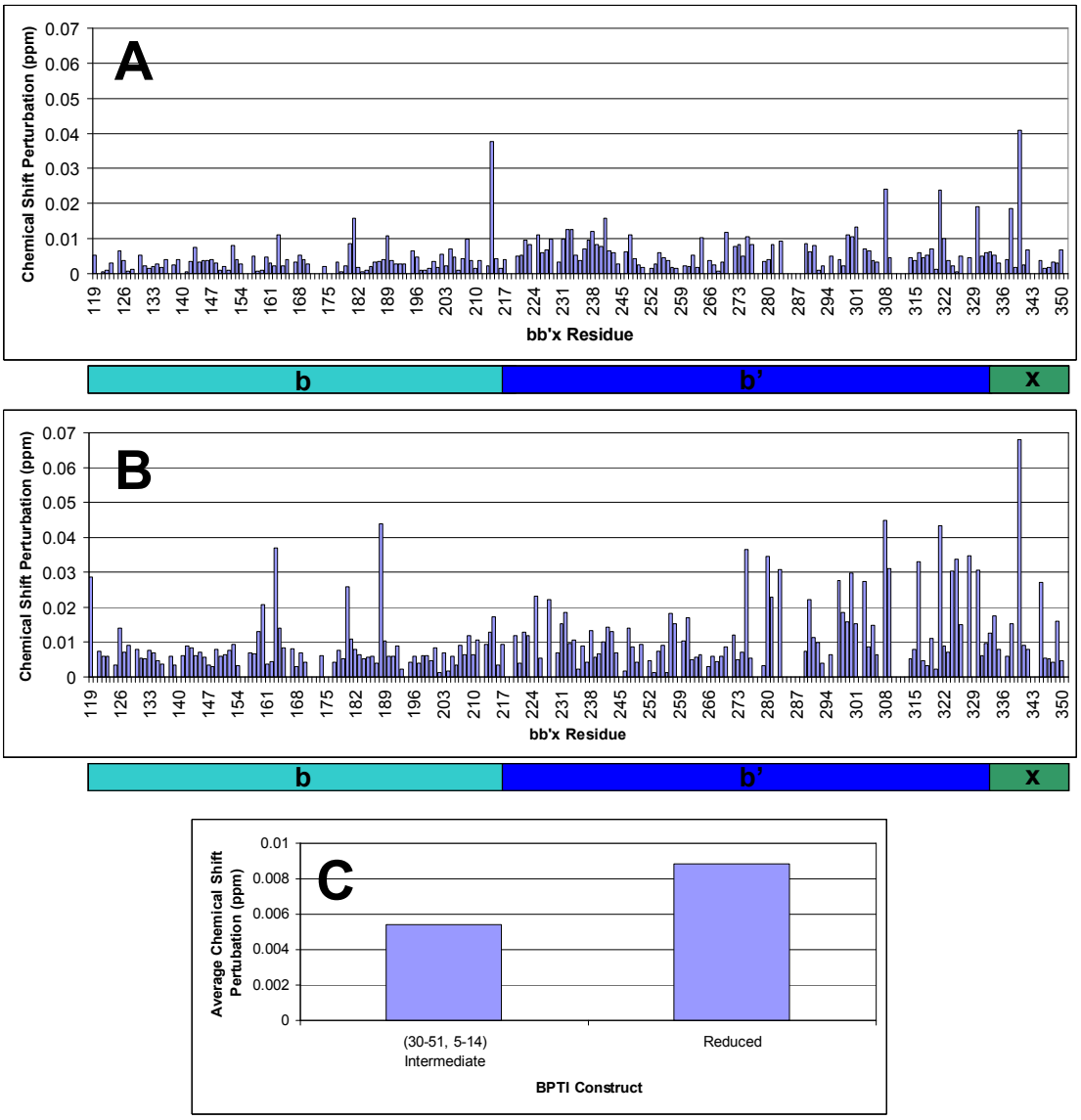


Figure 6.11: Chemical shift perturbations (to nearest peak) for bb'x alone compared to bb'x in the presence of BPTI at a bb'x:BPTI ratio of 25:1. A) using (30-51, 5-14) BPTI; B) using reduced BPTI; C) comparison of average perturbations from A and B. Shift perturbations represent a change in chemical shift of an assigned peak in the bb'x only spectrum to the nearest identified peak in the spectrum of bb'x in the presence of reduced BPTI. Since the scale of chemical shifts is larger in the ^{15}N axis than in the ^{1}H axis, chemical shift perturbations were calculated as a weighted sum, $^{1}/_{6}|\Delta^{15}N| + |\Delta^{1}H|$. The graphic below each graph indicates visually the position of each domain in the bb'x construct.

The minimal shift changes for bb'x with reduced BPTI at bb'x:BPTI ratios of 8.3:1 and 2.3:1 were also analysed. The results of all minimal shift data using reduced BPTI are summarised in Table 6.2.

Table 6.2: Summary of minimal shift changes for bb'x in the presence of reduced BPTI. HSQC spectra of bb'x with a bb'x:BPTI ratio of 25:1 and 8.3:1 were acquired using 250 µM bb'x, whereas the spectrum of 2.3:1 was acquired using 100 µM bb'x.

	Reduced BPTI		
	Ratio of bb'x:BPTI		
	25:1	8.3:1	2.3:1
Average Minimal Shift Change(ppm)	8.81E-03	1.27E-02	3.22E-02
Total Shifts > 0.06	1	7	28
Total Shifts > 0.10	0	2	13
0.06 < Total Shifts < 0.10	1	5	15

Figure 6.12 illustrates the differences in minimal shift perturbations in bb'x after addition of reduced BPTI at bb'x:BPTI ratios of 8.3:1 and 2.3:1. Generally, regions of relatively large shift change are visible in the b' domain, with only a few residues in the b domain showing a shift change. This is similar to observations seen in the presence of higher concentrations of wild type and (30-51, 5-14) BPTI. Another similarity is that many of the larger shift perturbations are for residues where peaks have not been identified in the spectrum containing reduced BPTI (maroon bars), probably due to line broadening. Leu188 shows the largest shift change in the b domain, just above 0.1 ppm for the 2.3:1 ratio, which is the largest change of all residues that have been identified in the spectrum containing reduced BPTI (blue bars). Although a broad trend may occur within each of the minimal shift maps in Figure 6.12, caution should be used when comparing then with each other, since spectra in

the presence of reduced BPTI were obtained at different concentrations of bb'x, due to the limited solubility of reduced BPTI, as explained above.

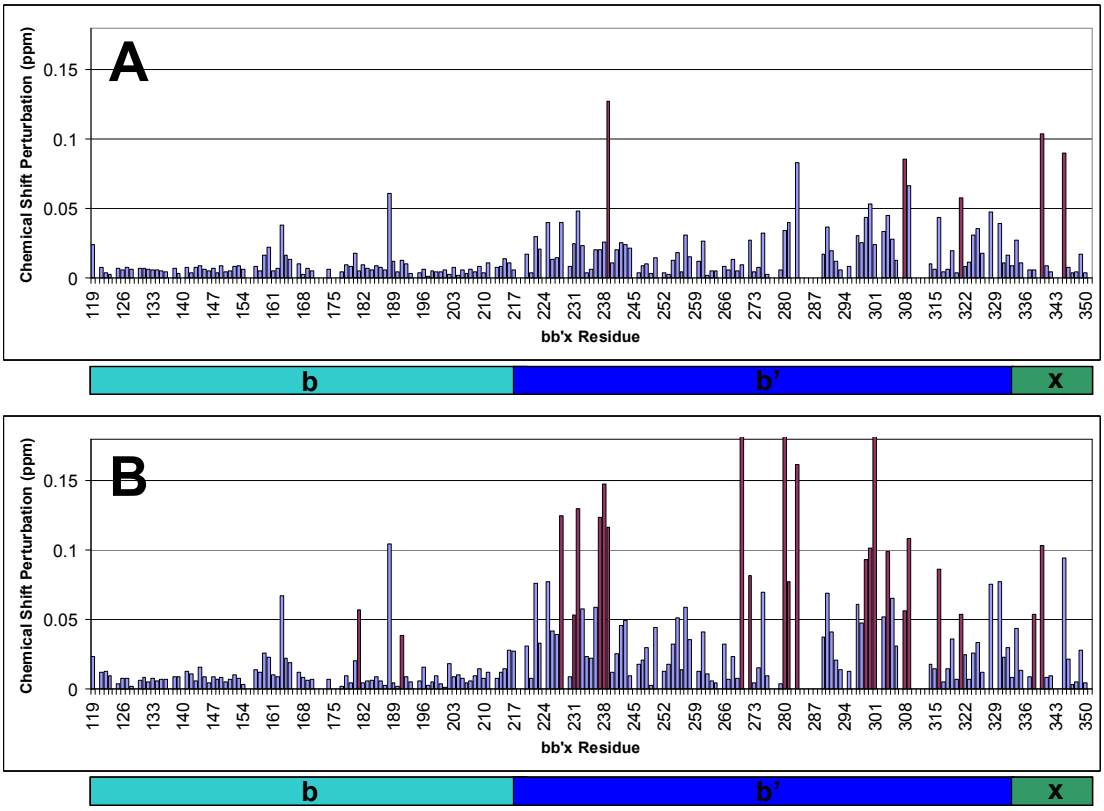


Figure 6.12: Chemical shift perturbations (to nearest peak) for bb'x alone compared to bb'x in the presence of reduced BPTI. A) bb'x:BPTI at 8.3:1 ratio using 250 µM bb'x; B) bb'x:BPTI at 2.3:1 ratio using 100 µM bb'x. Shift perturbations represent a change in chemical shift of an assigned peak in the bb'x only spectrum to the nearest identified peak in the spectrum of bb'x in the presence of reduced BPTI. Since the scale of chemical shifts is larger in the ^{15}N axis than in the ^{1}H axis, chemical shift perturbations were calculated as a weighted sum, $^{1}/_{6}|\Delta^{15}N| + |\Delta^{1}H|$. The graphic below each graph indicates visually the position of each domain in the bb'x construct.

6.6. Indole Resonances of Trp347 as Indicators of Substrate Binding

A recent study investigating the binding of peptides to various PDI constructs revealed that indole resonances of Trp347, identified as distinct peaks in HSQC spectra, could be used to monitor the binding of substrate to the b'x region (Byrne, Sidhu et al. 2009). In the absence of substrate, b'x and bb'x constructs of PDI were found to exist in two distinct forms, either as 'capped' monomer, whereby the x region binds to the b' domain, or as 'uncapped' dimer whereby dimerisation takes place via the interaction of b' domains (Nguyen, Wallis et al. 2008), see Figure 6.3.

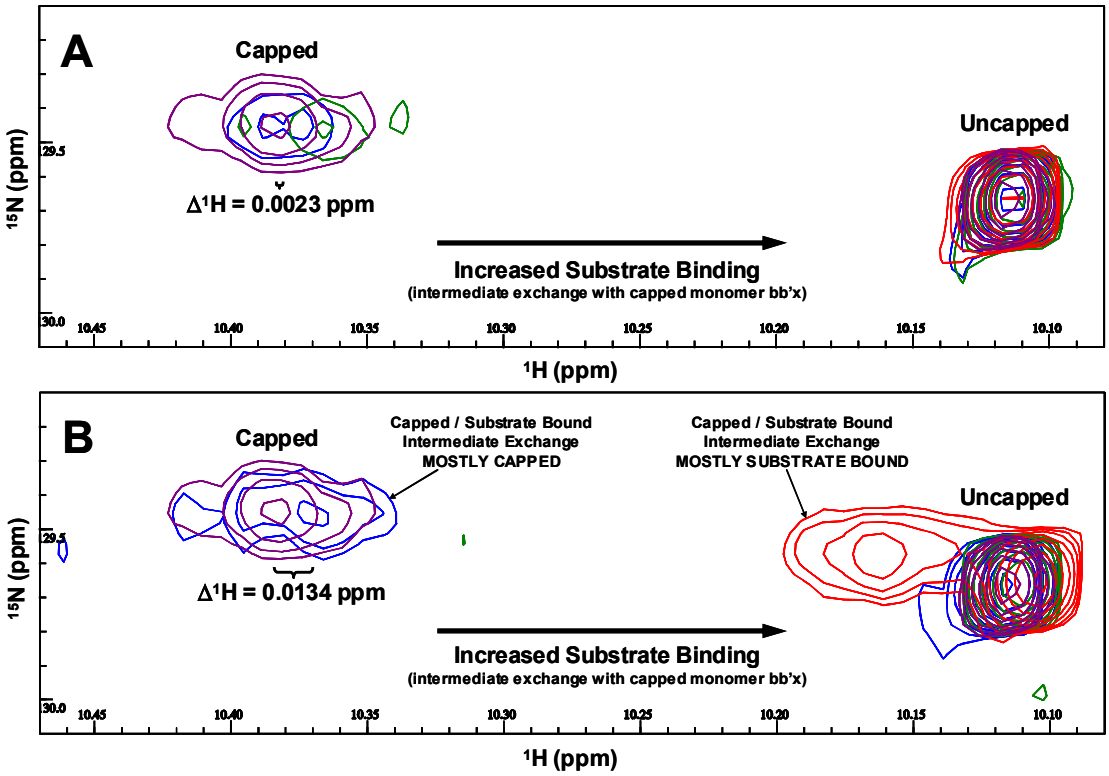


Figure 6.13: Perturbation of indole resonances of Trp347 from bb'x in the presence of various concentrations of BPTI construct. A) Wild type BPTI; B) (30-51, 5-14) BPTI intermediate. Purple, bb'x only (no BPTI); blue, bb'x:BPTI at 25:1; green, bb'x:BPTI at 5:1; red, bb'x:BPTI at 1:1. The change in proton chemical shift between bb'x only and bb'x in the presence of BPTI at a 25:1 ratio is labelled for the 'capped' peak of Trp347.

With Trp347, which is located in the x region, the two forms of bb'x can be observed as separate peaks in the HSQC spectrum, see purple contours in Figure 6.13. The downfield peak at ^{1}H = 10.38 ppm arises from the indole NH resonance of Trp347 in the monomeric capped conformer of bb'x; the upfield peak at ^{1}H = 10.11 ppm arises from the indole NH resonance of Trp347 in dimer bb'x.

Despite isolating the monomer form of bb'x by gel filtration during sample purification (see Section 6.2.3), the HSQC spectra show clear signs of both monomer and dimer bb'x. This is due to a conversion of bb'x from monomer to the dimer form before beginning the HSQC experiment (with sample stored at 4°C), during sample preparation and during acquisition of the HSQC data at 25°C.

The presence of two distinct peaks in the absence of substrate indicates that interconversion between the two conformations is slow on the NMR timescale (Byrne, Sidhu et al. 2009). However, substrates are able to displace the x region of the capped monomer, resulting in bb'x conversion into the uncapped form during substrate binding. Figure 6.13 suggests that for BPTI this interaction occurs on an intermediate timescale, resulting in broadened peaks being observed between the capped and uncapped states. The indole resonances of Trp347 are highlighted by dashed boxes in Figure 6.5 and Figure 6.10.

At a bb'x:BPTI ratio of 25:1 (Figure 6.13, blue contours), the wild type BPTI spectrum (Figure 6.13A) shows no sign of binding, since the uncapped monomer peak remains unchanged. By contrast, the (30-51, 5-14) BPTI spectrum (Figure 6.13B) indicates a monomer peak that has shifted slightly, with a reduced intensity, indicating a small amount of binding. It is not until a 5:1 ratio (Figure 6.13, green contours) that the wild type BPTI peak shifts slightly towards uncapped form, indicating the higher concentration required for wild type binding.

The (30-51, 5-14) BPTI spectrum at a 5:1 ratio shows almost no signal above the noise level for monomer bb'x, except for a tiny peak barely visible at ^{1}H = 10.31 ppm. Similarly, when the bb'x:BPTI ratio is increased to 1:1 (Figure 6.13, red contours), only the peak for bb'x dimer remains visible in the spectrum containing wild type BPTI. The disappearance of the bb'x monomer peaks are probably due to line broadening, indicating that interaction is occurring on an intermediate timescale.

Once the substrate unbinds from bb'x, it will presumably revert back to the capped conformer. If the conversion between substrate bound (uncapped) and unbound (capped) bb'x occurs on an intermediate timescale, this will result in line broadening and a reduction in peak intensity. At certain concentrations of wild type and (30-51, 5-14) BPTI, it appears that the NMR signal is reduced to the noise level, so no peak is identified.

At a bb'x:BPTI ratio of 1:1 using the (30-51, 5-14) BPTI construct, a clear peak is observed in the HSQC spectrum close to the uncapped species (Figure 6.13B, red contours). This indicates that the majority of bb'x monomer is now predominantly observed in the uncapped conformation, presumably due to binding of the (30-51, 5-14) BPTI substrate. This provides further evidence of the greater binding affinity to bb'x of this partly-folded substrate over natively folded BPTI.

The perturbations of indole resonances of Trp347 in the presence of reduced BPTI at various concentrations of bb'x are shown in Figure 6.14. Due to the limited solubility of reduced BPTI, it should be noted that a different set of bb'x:BPTI ratios were used. However, the lower ratio of 25:1 is comparable with those of wild type and (30-51, 5-14) BPTI, and the bb'x spectrum with reduced BPTI at this ratio shows a significant perturbation of the capped monomer peak (Figure 6.14, blue contours). This perturbation (0.0240 ppm) is much larger than

that observed with the same amount of (30-51, 5-14) BPTI, with very little perturbation visible in corresponding wild type peak. This comparison suggests that bb'x has a greater affinity for reduced BPTI than the more folded constructs.

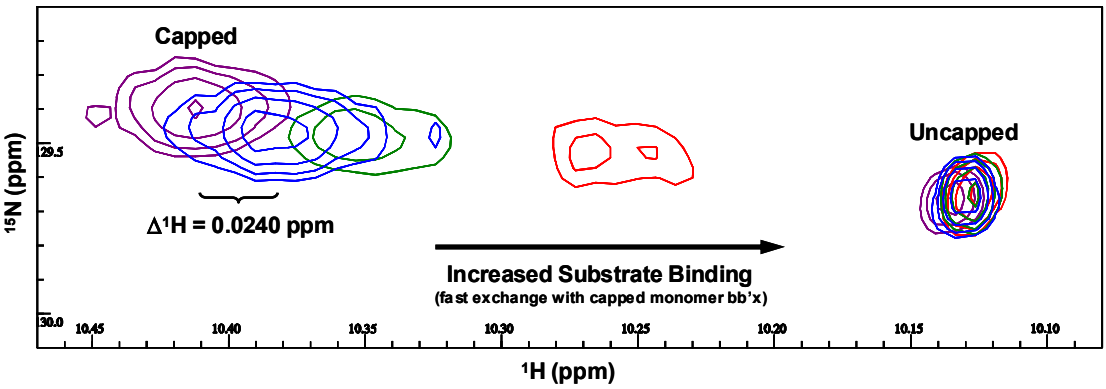


Figure 6.14: Perturbation of indole resonances of Trp347 from bb'x in the presence of various concentrations of reduced BPTI. Purple, bb'x only (no reduced BPTI); blue, bb'x: reduced BPTI at 25:1; green, bb'x: reduced BPTI at 8.3:1; red, bb'x:reduced BPTI at 2.3:1. The change in proton chemical shift between bb'x only and bb'x in the presence of reduced BPTI at a 25:1 ratio is labelled for the 'capped' peak of Trp347.

Although the effect of bb'x with reduced BPTI could only be measured up to a bb'x:BPTI ratio of 2.3:1, it is still enough for some interesting observations. Although some line broadening and a reduction of peak intensity is observed as the concentration of reduced BPTI is increased, all of the indole Trp347 peaks from the monomer still remain clearly visible in the spectrum. This contrasts with the spectra containing (30-51, 5-14) BPTI, where at bb'x:BPTI ratio of 5:1 the peak is barely visible. This suggests that reduced BPTI may have a faster exchange rate with bb'x than the (30-51, 5-14) BPTI intermediate. However, since the larger bb'x:BPTI ratio of 2.3:1 was performed using only 100 μM bb'x, it is difficult make any concrete conclusions with the available data.

6.7. Discussion

All BPTI substrates, at any stage along the folding pathway, indicate that binding to bb'x is focused mainly in the b'x region, with few chemical shift perturbations displayed in the b domain. This is in good agreement with previous PDI binding studies, such as the binding of bb' with peptides mastoparan and somatostatin as well as unfolded protein RNase A by (Denisov, Maattanen et al. 2009), as well as binding of Δ-somatostatin by (Byrne, Sidhu et al. 2009). Although b' has been known to be the primary binding domain of PDI for some time, it has been suggested that other domains will be involved with binding larger substrates (Klappa, Ruddock et al. 1998). Thus, it was anticipated that BPTI, being a full length protein, may require more regions of the PDI enzyme, including the b domain, for binding to take place. However, It was surprising that this full length protein, albeit rather small (6.6 kDa), has very little effect on the b domain of PDI upon binding, implying that this domain has little influence on the binding of the BPTI substrate. In contrast, a recent study of Ero1-Lα, a much larger substrate protein (approx 44 kDa), binding to PDI indicated that the minimal elements required included the b'xa' region of PDI (Wang, Li et al. 2009), suggesting larger protein substrates require more domains of the enzyme.

In order to compare and contrast more closely the effects of substrate binding on bb'x, chemical shift perturbations seen in the b'x region due to substrate binding from a variety of studies is shown in Figure 6.15. Since experiments were performed under different conditions and using substrates with different binding affinities, the criteria used to identify the most significant perturbations is different in each case. Lower concentrations of bb'x were used in experiments with BPTI substrates, hence smaller overall shift perturbations were observed. Here, the emphasis is on identifying similarities or differences in the positions of binding.

```
                  β1        α1             β2                   α2
                  ss      hhh  h         ss  sss            hhhhhhhhhh
(a)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(b)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(c)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(d)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(e)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(f)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(g)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(h)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(i)     219  LV  IEFTEQTAPK IFGGEIKTHI LLFLPKSVSD YDGKLSNFKT
(j)     241   F  GEIDGSVFAQ YVE-SGLPLG YLFY-NDEEE LEEYKPLFTE
                    hhhhhhh hhhh     ss ssss   hhhh hhhhhhhhhh

                           β3              α3              β4          β5
     hh          s  ssss     hh hhhhhhh     hhh   ssss  s        ssss
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
261  AAESFKGKIL FIFIDSDHTD NQRILEFFGL KKEECPAVRL ITLEEEMTKY
280  LAKKNRGLMN FVSIDA--RK FGRHAGNLNM KEQ-FPLFAI HDMT-EDLKY
     hhhhh    s sssss        hhhhhh           ssss ss      sss

                                       α4              βx          αx
                                 hh hhhhhh hhh  h     sss          hhh
311  KPES------EEL------TAE RITEFCHRFL EG
311  KPES------EEL------TAE RITEFCHRFL EG
311  KPES------EEL------TAE RITEFCHRFL EGKIKPHLMS QELPEDWDKQ P
311  KPES------EEL------TAE RITEFCHRFL EG
311  KPES------EEL------TAE RITEFCHRFL EGKIKPHLMS
311  KPES------EEL------TAE RITEFCHRFL EGKIKPHLMS QELPEDWDKQ P
311  KPES------EEL------TAE RITEFCHRFL EGKIKPHLMS QELPEDWDKQ P
311  KPES------EEL------TAE RITEFCHRFL EGKIKPHLMS QELPEDWDKQ P
311  KPES------EEL------TAE RITEFCHRFL EGKIKPHLMS QELPEDWDKQ P
326  GLPQLSEEAFDELSDKIVLESK AIESLVKDFL KG
     s     hhhhhhh         hhhh hhhhhhhhhh hh
```

Figure 6.15: Structural alignment of the b' domain of human PDI (black) and yeast PDI (blue) showing the ligand binding site as mapped by chemical shift perturbation. (a) Comparison of bb'x with bb' + Δ-somatostatin (yellow highlight > 0.13ppm; red highlight > 0.2ppm). (b) Comparison of b'x with b' + KFWWFS (yellow highlight > 0.15ppm; red highlight > 0.2ppm). (c) Comparison of b'x with b'x + Δ-somatostatin (yellow highlight > 0.09ppm; red highlight > 0.15ppm). (d) Comparison of bb'x with bb' as assigned by

(Denisov, Maattanen et al. 2007) (yellow highlight >0.15ppm; red highlight >0.3ppm). (e) Comparison of bb' with bb' + unfolded RNase (Denisov, Maattanen et al. 2009). (f) Comparison of bb'x with bb'x:(30-51, 5-14) BPTI at 5:1 ratio (yellow highlight >0.02ppm; red highlight >0.04ppm). (g) Comparison of bb'x with bb'x:wild type BPTI at 1:1 ratio (yellow highlight >0.02ppm; red highlight >0.04ppm). (h) Comparison of bb'x with bb'x:reduced BPTI at 8.3:1 ratio (yellow highlight >0.02ppm; red highlight >0.04ppm). (i) Peaks that seem to disappear due to line broadening, when comparing bb'x with bb'x:(30-51, 5-14) BPTI at 1:1 ratio (green highlight). Comparisons using BPTI substrates (f-i) indicate residues not assigned in the bb'x control spectrum (grey lettering), thus no chemical shift perturbation were calculated for these residues. (j) Hydrophobic ligand binding site identified in the crystal structure of yeast PDI (grey highlight). The secondary structure assignments for both the human and yeast structures are shown above and below the sequences in the corresponding colour (h, α-helix; s, β-sheet). Modified from Figure 6 in (Byrne, Sidhu et al. 2009) with assistance from Dr. Richard Williamson, University of Kent.

Generally, the region spanning β1 to β2 shows the greatest perturbations in all BPTI constructs, either via large minimal shifts or through line broadening (highlighted green) in this region. This follows a similar trend seen in all of the other bb'x ligand perturbations in Figure 6.15, emphasising the importance of this region. The reduced BPTI perturbations together with the line broadening of mutant BPTI at high concentrations suggest that the β4 to β5 region also acts as a particular focus of binding, which is in general agreement with other shift perturbation maps. Ligand binding does not appear to be localised to the small hydrophobic pocket proposed previously using homology modelling (Pirneskoski, Klappa et al. 2004), and this region has little overall influence on binding. A more recent study suggested that a large hydrophobic binding site is an effective way for PDI to bind a variety of substrates when unfolded and to release them once they acquire their native conformation with fewer hydrophobic residues exposed (Denisov, Maattanen et al. 2009).

Some subtle differences between the binding of BPTI constructs and others substrates can be observed in Figure 6.15. The perturbation of Ile236 by all BPTI constructs is only shared with the comparison of bb' with bb' + unfolded

RNase. Other interesting residue are Asp275 and Leu290, which show line broadening and large perturbation with all BPTI constructs, but were not perturbed significantly in other studies. Glu316 is perturbed with all BPTI, and also comparing bb'x to bb' of Denisov, but not in any others.

At the C-terminal end of b', Arg328 and Lys330 show perturbation, in good agreement with other constructs. In the x region, Ile334 and Ser340 show perturbation either side of βx.

An underlying assumption in the interpretation of chemical shift perturbations is that the extent of perturbation directly correlates with the importance of that residue or region in the binding process. This is not necessarily the case. Although perturbations in the b domain may be small compared to the b' domain, they could still have a significant substrate binding role. For example, Leu188 in the b domain of BPTI shows a noticeable chemical shift perturbation (Figure 6.8B), but this is still small compared to the large perturbations observed in the b' domain. Since the magnitude of shift perturbations caused by line broadening are not meaningful, then if only the remaining shifts are considered Leu188 appears relatively large compared to many in the b' domain. Leu188 also showed similar perturbations in ∆-somatostatin (Byrne, Sidhu et al. 2009) and may serve a key role in the substrate binding process.

Since HSQC spectra only observe backbone resonances of the protein or peptide sample, only perturbations in these chemical shifts are observed. Such a limitation may not fully reflect the role that amino acid side chains may have upon substrate binding.

The limitations of the minimal shift mapping technique itself should also be considered. Although it is a useful technique where the accurate assignments of spectra are difficult, it has the underlying assumption that the nearest peak represents the same residue. Visual inspection of minimal shift data indicates

that many residues show very large perturbations in bb'x because the peak has disappeared from the spectrum (due to line broadening). Thus the nearest peak actually represents the peak of a different residue (see Figure 6.8C). This can show up as extremely large perturbations in the minimal shift map, particularly if the peak is far away from the next nearest peak in the HSQC spectrum. Alternatively, in a very crowded region of the spectrum, a peak could have shifted a long way, beyond the next nearest peak, in which case the shift perturbation would be underestimated.

In this study, chemical shift perturbations that were likely to have been caused due to peak disappearance have been highlighted, but this is not normally the case in publications using the minimal shift approach. Thus, it is difficult to know the extent to which line broadening has influenced the shift map. A certain degree of subjectivity arises whenever very weak signals are detected. Deciding if a peak assigned in one spectrum has disappeared in another spectrum may depend on careful judgement about setting the noise level of the spectrum and deciding how far above the noise level a signal has to be before it can be designated as a genuine peak.

In this chapter, HSQC spectra of bb'x in the absence and presence of various BPTI constructs have been acquired and analysed. They show that BPTI binding generates largest perturbations in the b' domain, suggesting that this domain dominates binding of BPTI. Adding different concentrations of substrate protein indicated that reduced BPTI showed the greatest binding affinity, with (30-51, 5-14) BPTI showing clear signs of binding at lower concentrations than the wild type substrate.

Chapter 7. Binding Affinity of BPTI/PDI Interactions

7.1. Introduction

Previous chapters have focused on structural studies of the interaction between the PDI folding enzyme and the BPTI substrate at various stages along the folding pathway. In this chapter, insight was sought into the binding affinities between PDI and each BPTI construct. How does BPTI's affinity to PDI change as it goes through its folding pathway?

While several studies have investigated the nature of substrate binding to PDI (Pirneskoski, Klappa et al. 2004), specifically the role of domains in the binding process (Klappa, Ruddock et al. 1998), very little is known about the binding affinity or kinetics of binding with each substrate. There are few studies into the binding affinity of PDI with unfolded or partly-folded substrates, probably due to the difficulties associated with the stability of these proteins.

One technique for measuring protein-protein interactions is surface plasmon resonance (SPR). This involves immobilising one protein onto a chip surface and flowing the other in solution over the surface, allowing binding to occur and be measured via a change in mass on the chip surface.

7.1.1. Protein Constructs Used In Binding Studies

For binding studies, interest is mainly focused on how BPTI at various stages of folding interacts with the full length PDI enzyme. However, since NMR observations of PDI focus on the main substrate binding region, bb'x, it is also of interest to compare the binding to full length PDI with the binding to bb'x. This will indicate if the observations of bb'x binding reflect the true nature of binding

to the full length protein. PDI and bb'x were recombinantly expressed as described previously.

As a control, a protein of similar molecular mass to full length PDI (56kDa) was desired. Ovalbumin, after post-translational modifications, has a molecular weight of 44.3 kDa (Tai, Yamashita et al. 1977). It is considered a member of the serpin superfamily, but does not function as a serine protease inhibitor nor show any enzymatic activity (Hu and Du 2000). Although its function is unknown, it is presumed to be a storage protein (Gettins 2002). These characteristics were thought to make ovalbumin a suitable control for BPTI binding experiments. Ovalbumin purified from chicken egg white was used for immobilisation (Grade VI, Sigma).

7.1.2. Temperature Studies

Throughout this study, NMR experiments were performed at three different temperatures. Initially, NMR was performed at 36°C to compare the recombinant wild type BPTI with previous studies of the commercial protein. However, partly-folded BPTI required NMR at 5°C to get good resolution. In contrast, NMR of the bb'x region of PDI is best performed at 25°C. Since different temperatures were used for different types of experiment, it is desirable to identify the effects of temperature on BPTI/PDI interaction at each stage of the folding process. For this reason, separate sets of binding experiments were performed at 5°C, 25°C and 36°C.

7.2. Immobilisation of Proteins by Amine Coupling

PDI/BPTI binding studies could be performed in two ways, immobilising either the PDI or the BPTI constructs to the sensor chip. For this study, the PDI

constructs were immobilised, with various BPTI constructs used as the analyte. Ideally, the same experiments would also be performed with BPTI constructs immobilised and PDI as the analyte, to verify the results.

In order to be able to fully correlate binding experiments with the corresponding NMR experiments, both the full length PDI and the bb'x region of PDI were immobilised on the sensor chip surface.

Several different approaches are available for performing the immobilisation. By far the most common is amine coupling, whereby the sensor chip is pre-treated in a way that allows amine groups of the available protein to covalently bind to the chip surface (section 2.11.1). For this study, PDI and bb'x were bound on separate channels of the same chip, with another channel reserved as a blank. The details of the procedure are outlined in section 2.11.1. Figure 7.1 shows the sensorgram from bb'x immobilisation. After pre-treatment with EDC/NHS, bb'x was successfully bound to the sensor chip. Ethanolamine was then reacted with any remaining functional groups on the chip surface, displacing some bb'x that may not have become covalently bound. bb'x immobilisation resulted in a final value of 2582.2 response units (RU). Amine coupling of PDI was also successful, resulting in a final immobilisation value of 5591.4 RU.

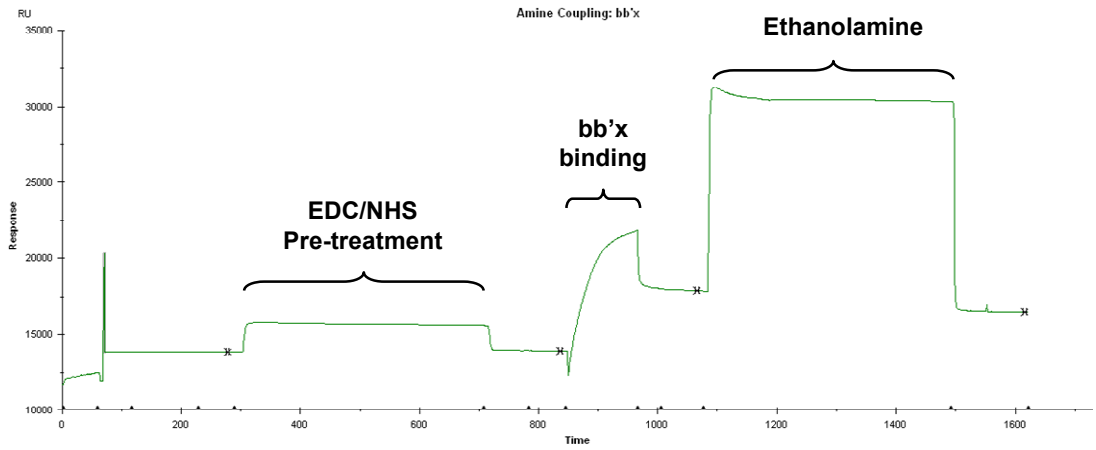


Figure 7.1: Sensorgram of bb'x immobilisation onto the sensor chip surface by amine coupling.

7.2.1. Preliminary Test to Check Binding to Immobilised Proteins

Before performing large sets of binding experiments with several different BPTI constructs, an initial test of binding was performed using (30-51, 5-14) BPTI as analyte flowing over the sensor chip with immobilised PDI constructs. Figure 7.2 shows the sensorgram from various concentrations of (30-51, 5-14) BPTI binding to the bb'x channel of the sensor chip. It shows very clear and distinct association and dissociation phases, with the response increasing in a dose-dependent manner. Also, it was crucially important to verify that the response level returned to the baseline value after regeneration (data not shown).

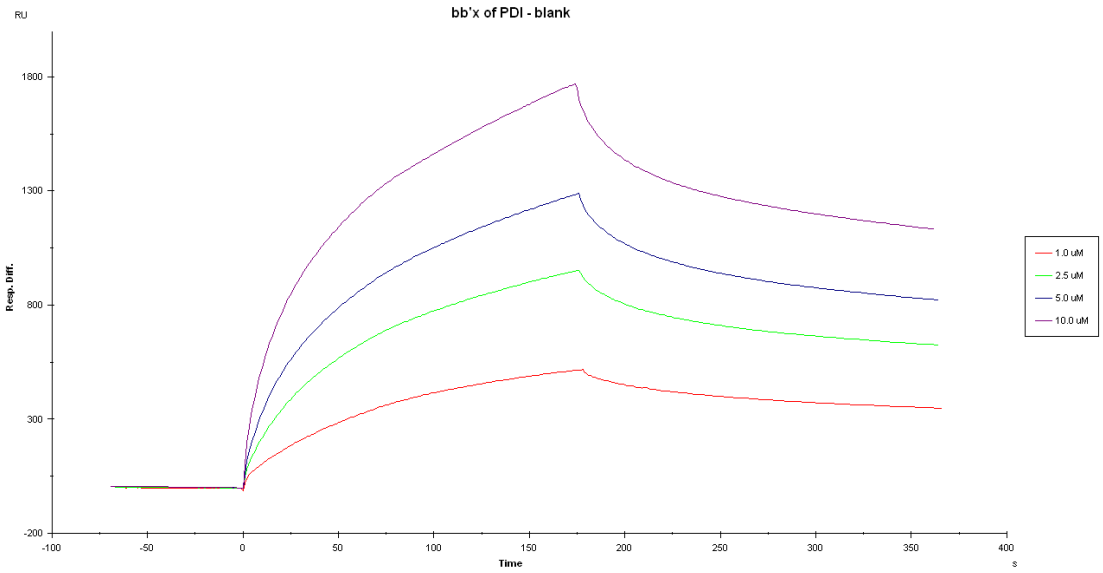


Figure 7.2: Sensorgram of (30-51, 5-14) BPTI as analyte binding to bb'x at various concentrations.

7.3. Experimental Design

Three different proteins were immobilised onto the sensor chip (PDI, bb'x and ovalbumin). Experiments were performed using three different analytes, representing different stages of folding of BPTI (wild type BPTI, (30-51, 5-14) BPTI, reduced BPTI). Binding analysis was performed at three different temperatures (5°C, 25°C and 36°C). This provided a total of 27 different experimental conditions.

There was little prior knowledge providing expected binding affinities for each condition, except for the preliminary experiment shown above. This meant that a wide range of different molar concentrations of analyte were used for each interaction. This also allowed for a potentially large difference in affinities between different interactions. In each case, lyophilised protein was resuspended in the running buffer. A buffer only negative control was used that did not contain analyte (0 μM). Six different molar concentrations of analyte were used for each experiment, at 1 μM, 2.5 μM, 5 μM, 10 μM, 15 μM and 20 μM. Each set of injections was repeated once to gauge the repeatability of each experiment.

7.4. Curve Fitting

The nature of the interactions between BPTI and PDI constructs were elucidated through the fitting of curves to kinetic data from the sensorgram. Using Biacore's BIAevaluation software, various curve fitting options were explored.

In all cases, simultaneous fitting of the association and dissociation phases of the kinetic data was attempted, since any valid kinetic model should be able to account for both aspects of binding. Curve fitting was also performed

simultaneously using all concentrations of the analyte, since the underlying processes by which binding occurs should be dose-independent. All parts of each curve were used in the fitting, except for small regions near the inject start and inject end, where perturbations often occur. For any model, the curve fitting process attempts to find the values for parameters in the rate equation that best fit the selected experimental data.

Since each lyophilised BPTI construct was resuspended directly into the running buffer, no change in bulk refractive index between running buffer and analyte was observed, so the refractive index difference parameter (RI) was set to a constant of zero.

7.4.1. Assessing Goodness of Fit

How well the fitted curves represent the data can be assessed in a number of ways. Firstly, visual inspection of the fitted curves can give an indication as to how reasonable the model is and identify clear systematic problems where the fitting may not match the data. Secondly, residual plots are provided, where residuals represent the difference between the experimental data and the fitted values for every point in the sensorgram. The scatter of residual plots indicates the noise of the signal. It is often easier to detect small systematic deviations from the residual plots than from the sensorgrams. Statistically, a chi^2 value represents the closeness of fit, where chi^2 approximates to the average squared residual per data point (Biacore 1997). However, the form of the fitted curve in relation to the experimental data (*i.e.* the shape of the residual plot) is often more valuable in assessing the goodness of fit (Biacore 1997).

7.4.2. Curve Fitting Models

The most basic fitting model assumes a one to one binding between the immobilised protein (ligand) and the analyte. With this model, single on (k_a) and off (k_d) rate constants are given, allowing simple calculation of a single dissociation constant (K_D) for the interaction.

Another fitting model takes account of the possibility of heterogeneity of the immobilised protein; it is referred to as "Heterogeneous ligand – Parallel reactions" in the BIAevaluation software. This fitting process implies that there is more than one mechanism by which the analyte (*i.e.* BPTI) can bind to the immobilised proteins. With this model, two sets of on and off rate constants are given, resulting in two separate dissociation constants. Since more parameters are involved, any solution when applying this model is more sensitive to initial parameters. The optimum solution for each BPTI binding was obtained by iteration through the model, using the previous on and off rates as initial parameters to the next iteration, until a minimum chi^2 value was obtained. Various reasons why such a model may be applicable to BPTI/PDI interaction are explored in the discussion section.

7.4.3. Curve Fitting for BPTI Binding Analysis

Curve fitting was attempted for all experiments using these two different models. Figure 7.3 shows a typical example of the results of fitting for each model. Visually, the curve fitting appeared better using the heterogeneous ligand model. However, this does not necessarily mean it is a more appropriate model, since better fits would be expected from a model that has more parameters and therefore more degrees of freedom. Indeed, recent reviews of biosensor literature show that it is a common mistake for users to fit their data using inappropriately complex models (Rich and Myszka 2008; Rich and Myszka

2010). Figure 7.4 summarises the dissociation constants obtained from two different binding models at 36°C. Despite showing a much better fit to the data, the model that assumes a heterogeneous ligand generally shows dissociation constants that are very similar to each other in every case except reduced BPTI binding to ovalbumin, where K_D1 is so small it is not visible on this scale (compare K_D1 and K_D2 in Figure 7.4B). A pattern of very similar K_D pairs in this model would suggest that the simpler one to one binding model may be more appropriate. Indeed, the summary of dissociation constants for BPTI binding at 36°C using the one to one model shows a remarkable similarity to the pattern of K_D pairs using the heterogeneous ligand model, despite the seemingly poor fit of individual curves (compare Figure 7.4A with Figure 7.4B).

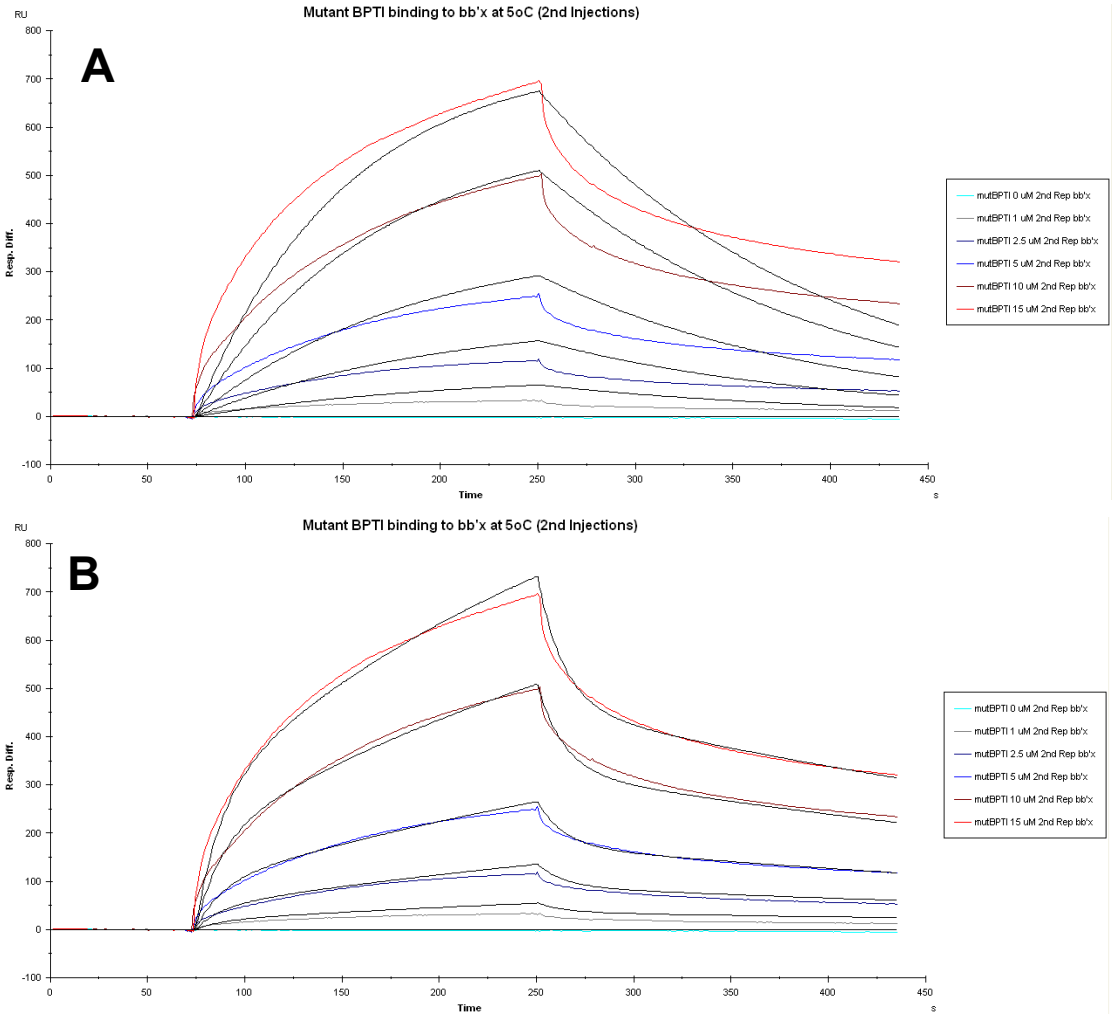


Figure 7.3: Example curve fitting to BPTI binding sensorgram data using different curve fitting models. A) 1:1 (Langmuir) binding model; B) model assuming heterogeneous ligand with parallel reactions.

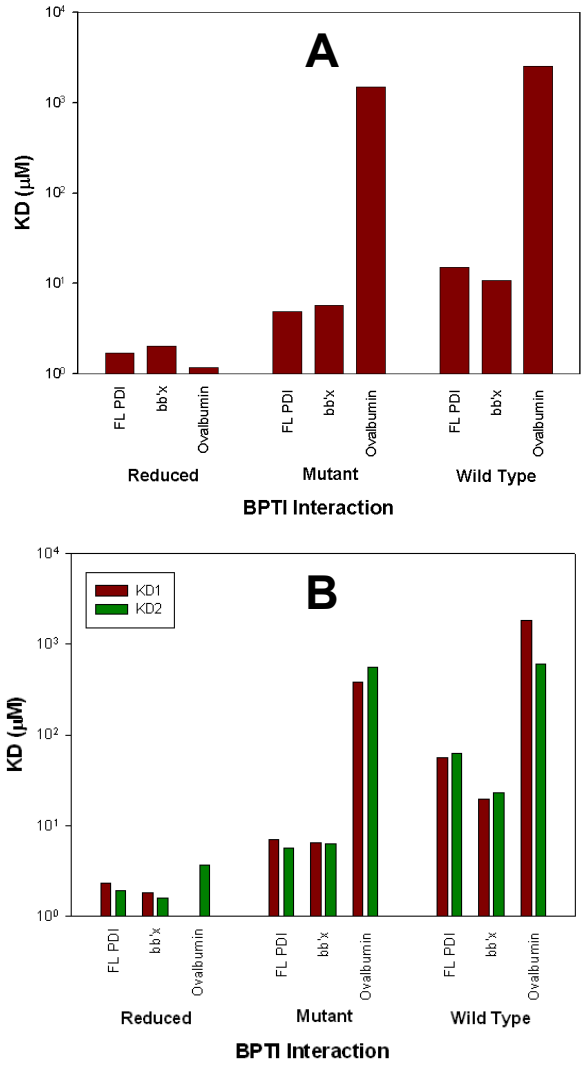


Figure 7.4: Different representations of BPTI binding at 36°C using different curve fitting models. A) 1:1 (Langmuir) binding model; B) model assuming heterogeneous ligand with parallel reactions.

7.5. BPTI Binding at Various Temperatures

The BPTI constructs representing different extents of folding were run over the sensor chip at different temperatures, as described above (section 7.3). The results after curve fitting assuming a 1:1 binding are summarised in Figure 7.5. The detailed results of the curve fitting analysis are shown in Appendix G.

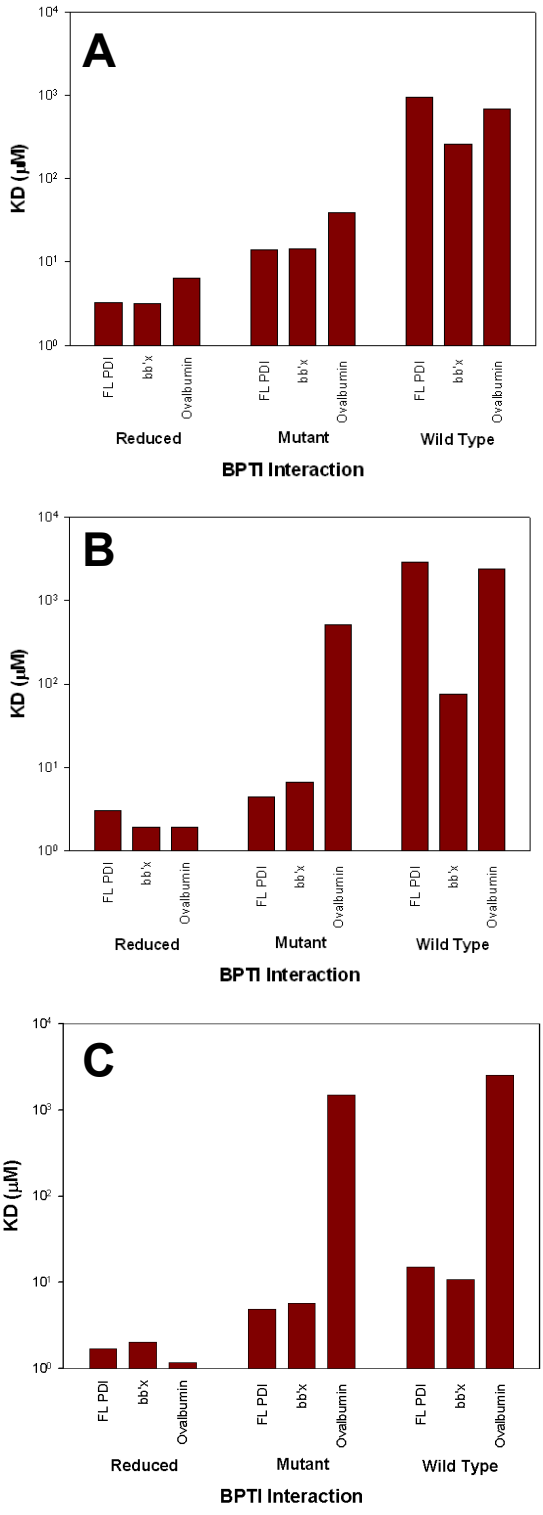


Figure 7.5: Average dissociation constants for reduced, (30-51, 5-14) mutant and wild type BPTI binding at various temperatures. BPTI was used as analyte and bound to immobilised full length PDI, bb'x region of PDI and ovalbumin. A) 5°C; B) 25°C; C) 36°C.

The curve fitting process provides separate global association and dissociation rate constants that can be separately analysed to observe trends in these kinetic rates. Figure 7.6 shows the association rate constants for the various interactions of BPTI. It shows a general trend similar to the overall dissociation constants shown in Figure 7.5, with the logarithmic axis reversed.

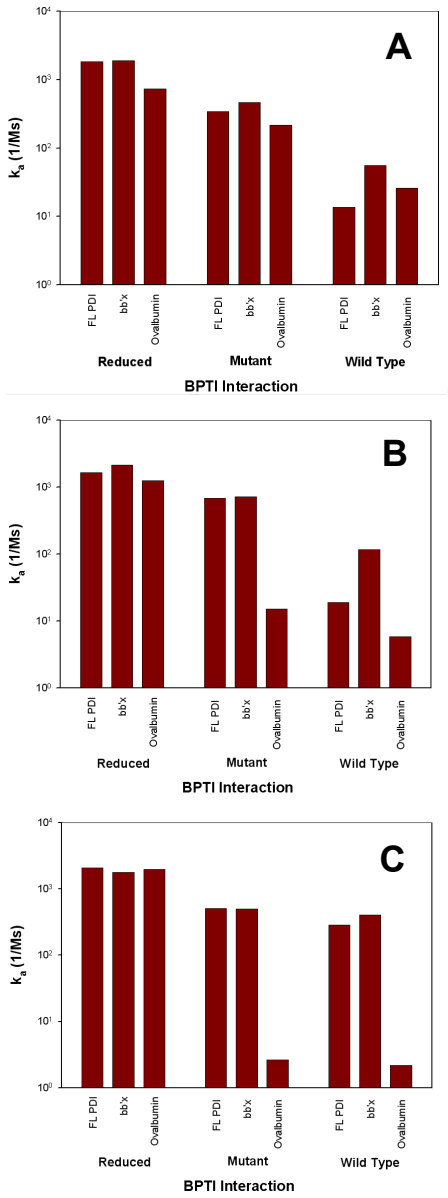


Figure 7.6: Average association rate constants for reduced, (30-51, 5-14) and wild type BPTI binding at various temperatures. A) 5°C; B) 25°C; C) 36°C.

Figure 7.7 shows the dissociation rate constants for the interactions. The variation is much smaller, with an axis displayed on a linear scale. This suggests that the majority of the differences observed between interactions are due to different 'on' rates rather than different 'off' rates.

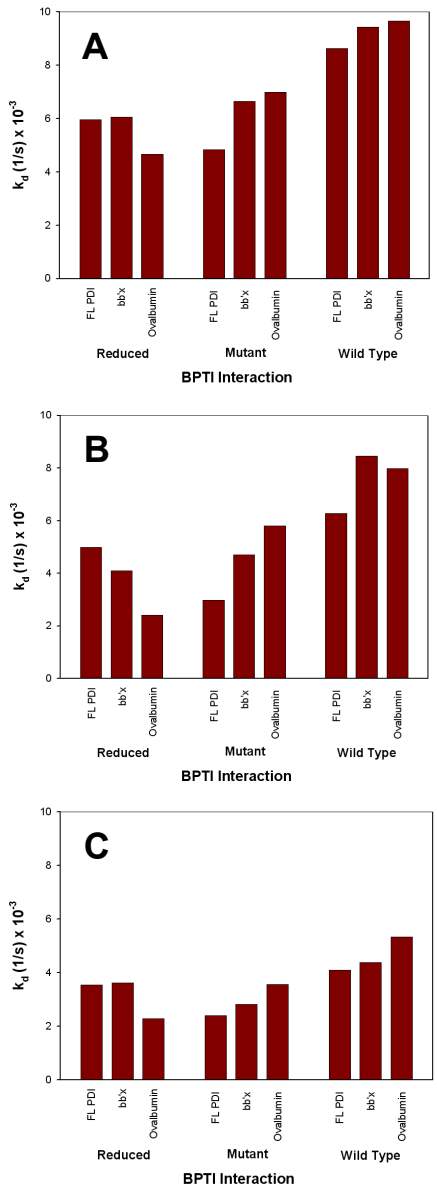


Figure 7.7: Average dissociation rate constants for reduced, (30-51, 5-14) and wild type BPTI binding at various temperatures. A) 5°C; B) 25°C; C) 36°C.

A general trend emerges at all temperatures that shows the reduced, unfolded BPTI with the strongest binding (lowest dissociation constant), with the partly-folded (30-51, 5-14) BTPI binding less well and the fully folded wild type BPTI showing the weakest binding. This trend may be expected, since it is reasonable to expect the less folded substrate to have a higher affinity to the folding enzyme. Although the wild type protein exists in an equilibrium that is heavily biased towards the native state, there will still be rare occasions when it is partly-unfolded, thus some binding to PDI may be expected. These rare unfolding events may take place more often at higher temperatures, as indicated by the higher affinity of wild type BPTI to PDI at 36°C.

Interestingly, at all temperatures, binding of reduced BPTI has just as high an affinity to ovalbumin as it does for PDI constructs. However, the partly folded intermediate shows much weaker binding to ovalbumin than to the PDI constructs. Since the partly-folded protein exposes only certain hydrophobic regions, the PDI enzyme binding site may be preferentially attracted to these regions, whereas ovalbumin is not. However, the fully unfolded, reduced BPTI will have significant exposed hydrophobic regions throughout the protein. It may be that such a protein has a tendency to bind to almost any protein.

There was no significant difference between binding of full length PDI and the shorter bb'x region of PDI for fully unfolded or partly folded BPTI substrates. This indicates that for this small protein, the bb'x region is sufficient for complete binding of the substrate. This provides greater confidence in interpretation of NMR results using the bb'x region of PDI, and that the observations likely reflect the behaviour of the full length folding enzyme.

There is a general trend that shows binding gets stronger at higher temperatures for all constructs binding to full length PDI or bb'x. Since proteins become more dynamic at higher temperatures, configurations that expose extra hydrophobic regions become more common, so hydrophobic interactions

become more frequent. So an increase in binding affinity at higher temperatures supports the theory that binding of PDI and substrate is primarily hydrophobic in nature.

7.6. Discussion

Although the method of curve fitting that best describes BPTI/PDI interactions is not clear, the general pattern that results from each model is consistent. The most unfolded protein, BPTI in the fully reduced form, shows the strongest affinity to PDI, with decreasing affinity the more folded the substrate becomes. Below, some reasons for the observed binding behaviour are explored. Suggestions are put forward for how the experimental process may be improved and alternative methods of elucidating binding affinities are given.

7.6.1. Alternative Conformations of PDI May Influence Binding

One theory supporting parallel binding to a heterogeneous ligand is the role of the x region of PDI in substrate binding. A recent study revealed that the x region modulates the exposure of the binding site of PDI (Nguyen, Wallis et al. 2008), and distinct species of ‘capped’ and ‘uncapped’ bb’x have been isolated (Wallis, Sidhu et al. 2009). These two species of PDI are likely to have different binding affinities, supporting a parallel binding model, with the ‘uncapped’ species exposing the binding site, thus having a much greater binding affinity. Such a model would be complicated by the ability of PDI to switch between the ‘capped’ and ‘uncapped’ states. Indeed, the rate of transition from ‘capped’ to ‘uncapped’ conformer may be determined by the difference in affinity for the b’ domain between the substrate and the x region.

7.6.2. Non-specific Binding of Substrate Hydrophobic Regions

PDI is thought to have only a single binding region, focused around the b' domain, with other domains influencing the binding process (Klappa, Ruddock et al. 1998). However, since binding is mainly based on hydrophobic interaction with its substrate, the region of binding on the substrate molecule will be based on its exposed hydrophobic regions, rather than a specific binding site. If all exposed hydrophobic regions had the same binding affinity, this would not be a problem, since a single binding constant could still be established. However, if different hydrophobic regions showed different affinities for PDI, a more complex binding picture would emerge. The binding affinity of each region would perhaps be proportional to its hydrophobicity. Such complex binding would be very difficult to deconvolute.

7.6.3. Calculating Affinities from Steady State Binding

Establishing the binding affinities could be easier if the injections had been allowed to proceed for longer, or at higher concentrations, such that steady-state equilibrium was established (for higher analyte concentrations). In this way, binding affinities could be calculated from the response units at steady state for each concentration, independent of the kinetics required to reach the steady state.

7.6.4. Alternative Form of Immobilisation

One of the major issues with the use of surface plasmon resonance as a technique for studying binding kinetics is that it is difficult to be certain that the process of immobilisation has not altered the protein's ability to bind its substrate. With amine coupling as an immobilisation strategy, the protein is

immobilised based on exposed amine groups. As well as the N-terminus of the protein, amine coupling could take place at various amino acid side chains. Thus, heterogeneity may exist in the nature of immobilisation, which may, in turn result in heterogeneity in binding due to different parts of the protein surface being more exposed than others.

Since the PDI constructs used in this study are expressed with an N-terminal His-tag, the SPR techniques used in this study could be improved by performing immobilisation using the His-tag of the recombinant PDI constructs, rather than by amine coupling. Using a sensor chip that has been functionalised with nitrilotriacetic acid (NTA), the His-tag of the protein can be non-covalently captured via nickel ions, see Figure 7.8 (Biacore 2003).

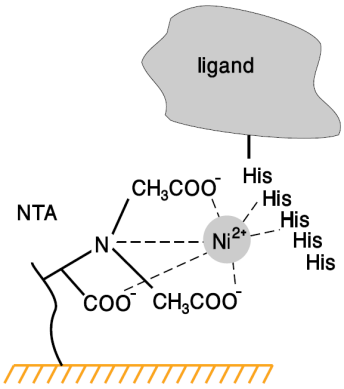


Figure 7.8: Immobilisation of ligand containing a His-tag via nickel chelated to NTA (Biacore 2003).

With this technique, more certainty can be put on the location in which the immobilisation takes place on the protein. Since the His-tag has only been added to the recombinant protein to facilitate purification, its immobilisation is much less likely to interfere with the natural binding properties of any PDI construct.

7.6.5. Alternative Methods of Probing Binding Interactions

7.6.5.1. Isothermal Titration Calorimetry

While SPR essentially detects binding based on a change of mass at the sensor chip surface, calorimetric techniques observe binding based on changes in heat.

Isothermal Titration Calorimetry (ITC) is a technique which allows the progress of a binding reaction to be followed thermodynamically. When two substances bind, energy is either released or absorbed. ITC measures changes in binding enthalpy and entropy as the binding reaction occurs, allowing the calculation of binding constants and reaction stoichiometry.

No ITC studies have been performed between PDI and BPTI. However, ITC has been used to show that the b' domain of ERp57 binds calnexin with micromolar affinity (Kozlov, Maattanen et al. 2006). A previous study used the related isothermal microcalorimetry technique to study the thermodynamics of the refolding of GAPDH assisted by PDI (Liang, Li et al. 2001). However, since GAPDH does not form disulfide bonds, this study focused on PDI's function as a chaperone, preventing aggregation, rather than as a folding enzyme.

Several studies have used calorimetry to study BPTI binding, mostly studying its interaction as an inhibitor of the trypsin enzyme. ITC was used to study the effects of Y35G replacement in BPTI on enzyme-inhibitor interaction with trypsin (Hanson, Domek et al. 2007) and it has also been used to study BPTI binding to the chaperone SecB (Panse, Swaminathan et al. 2000). Another calorimetric technique, differential scanning calorimetry (DSC), has also been used to study the effects of various mutations in BPTI (Liu, Breslauer et al. 1997; Grzesiak, Helland et al. 2000; Buczek, Krowarsch et al. 2002).

The main advantage of ITC over SPR is that it does not require the immobilisation of enzyme or substrate, thus eliminating any effects that immobilisation might have on the binding process. However, ITC tends to require larger volumes and higher concentrations of protein and can require considerable trial and error to optimise. Non-specific binding of unfolded or partly-folded BPTI to PDI could also become a problem, if different exposed hydrophobic regions of the unfolded protein displayed different affinities to the folding enzyme.

7.6.5.2. Estimating Binding Affinity from NMR Spectra

Another method of estimating the binding affinity between BPTI and PDI constructs may be through using certain NMR techniques. This can often be by observing ligand and titrating in the enzyme. Such experiments have recently been performed to observe binding to bovine serum albumin, where either a chemical shift or a linewidth is monitored as a function of changing solution composition (Fielding 2007). It has also been used to provide a crude estimate of peptide ∆-somatostatin binding affinity to the b'x region of PDI, focusing on data from the indole resonances of Trp347 (Byrne, Sidhu et al. 2009). Similarly, (Denisov, Maattanen et al. 2009) calculated dissociation constants for binding of various substrates to the bb' region of PDI by monitoring the chemical shift changes as a function on ligand concentration using a simple binding model. They performed a least-squares search by varying the values of K_D and the chemical shift of fully saturated protein. Standard deviations were derived by comparing cross-peaks in the HSQC spectra. Saturation transfer methods can also be employed, providing that the ligand is in fast exchange between free and bound states on the NMR time-scale (Fielding 2007).

7.6.6. Using Biosensors to Distinguish Substrate Specificity between PDI Family Members

SPR and other biosensor techniques could be applied to other PDI family members. Many PDI family members show distinct substrate specificity (Jessop, Watkins et al. 2009). The differences in binding affinities with different unfolded and partly-folded substrates could give insight into the different functions of each family member. For example, PDI could be compared with PDIp using both substrates that are ubiquitously expressed and those that are highly expressed in the pancreas.

The binding of substrates to ERp57 and PDI could be compared. Since ERp57 is thought to preferentially bind to N-glycosylated substrates, analytes could include both non-glycosylated and glycosylated protein substrates. However, since ERp57 only binds to N-glycosylated substrates when a complex with lectin partners calnexin or calreticulin, an affinity capturing mechanism could be employed to form these complexes (Biacore 2003). A similar capturing mechanism could be used to form a complex between ERdj5 and BiP, which can then be used to observe binding with misfolded proteins normally earmarked for the ERAD pathway.

One interesting comparison may be between PDI and ERp44. Since ERp44 has only one cysteine it has been speculated that a mixed disulfide with the substrate may persist. Using conditions that allow a mixed disulfide to form may provide a mechanism to test this hypothesis. If disulfides form and persist, the sensorgram may show a typical association phase, but a much shallower dissociation phase, since the substrate would remain covalently bound to the immobilised enzyme. With this kind of comparison, the precise redox conditions would become important and the effect of using running buffer in different pH conditions could be explored. Indeed, a variety of chemicals could be added to

the running buffer to observe the influence they have on the binding process, such as oxidising and reducing agents, Ca^{2+} ions or ATP.

Techniques are also available to attach lipid membranes and vesicles to the sensor chip surface (Biacore 2003), opening up the possibility that even substrate binding to PDI members with transmembrane domains could be investigated.

Chapter 8. Discussion

8.1. Introduction

Protein disulfide isomerase (PDI) has been studied for many years, with most emphasis put on its catalytic role in the oxidation and isomerisation of disulfide bonds. However, in order to catalyse these processes PDI must first bind to the substrate. Recently, more emphasis has been put on the nature of this binding. The earliest binding studies identified the b' domain of PDI as the primary binding site of the enzyme, but it was suggested that all domains influenced binding (Klappa, Ruddock et al. 1998). More recently, short peptides and a small unfolded protein have been used to try and localise the binding site to specific regions within the b' domain (Byrne, Sidhu et al. 2009; Denisov, Maattanen et al. 2009).

Although previous studies used a variety of different substrates to characterise binding to PDI, this is the first study to compare binding of the same substrate protein at different stages along its folding pathway. While previous studies focused on fully unfolded proteins or peptides, this is the first study to investigate the structural interaction of a partly-folded protein intermediate with the PDI folding enzyme. While all nascent proteins produced by the ribosome will initially be fully reduced, formation of initial disulfide bonds, where required, is often rapid. However, in a protein requiring multiple disulfide bonds, the initial configuration of disulfides is often incorrect, with disulfides often formed between cysteine residues in close proximity, rather than in the stable configuration required to establish the natively folded protein. In such circumstances, isomerisation of disulfide bonds is required and this process is often the rate limiting factor in protein folding (Huth, Perini et al. 1993; Land, Zonneveld et al. 2003). Given these circumstances, partly folded, rather than

completely unfolded proteins are likely to make up a large proportion, possibly even the majority, of substrates for the PDI enzyme and the wider PDI family of proteins.

For this study the protein bovine pancreatic trypsin inhibitor (BPTI) was chosen as a suitable substrate for studies of binding with PDI. BPTI is a small, well characterised protein containing three disulfide bonds (Deisenhofer and Steigemann 1975). The native protein has a particular configuration of disulfide bonds, but a folding pathway has been established that maps the order in which these bonds typically form, which includes the formation of non-native disulfides in folding intermediates (Darby, Morin et al. 1995), see Figure 1.12. This makes BPTI an ideal candidate as a substrate for studying PDI binding, since partly-folded intermediates, represented by their disulfide bond configuration, can be recombinantly expressed (Creighton, Darby et al. 1996).

One interesting feature of this study is that it has sought to characterise binding not only in terms of the changes to the PDI enzyme, but also in terms of the effects on the substrate, at different stages of its folding pathway, *i.e.* investigating how each BPTI substrate is affected by binding to PDI as well as how PDI is affected by binding of each BPTI substrate.

The main aims of this project were i) to prepare various BPTI constructs representing different stages along its folding pathway; ii) characterise each BPTI construct, using NMR and other techniques to establish the extent of folding; iii) observe changes in each BPTI construct upon addition of the PDI enzyme; iv) observe changes in the binding region of PDI upon addition of each BPTI construct; v) establish a binding affinity between PDI and each BPTI construct.

Previous chapters in this study used different techniques and different perspectives to investigate the nature of binding between PDI and various BPTI

constructs representing different stages along its folding pathway. In this chapter the results from all aspects of this study are brought together and used to make conclusions about the interaction of PDI with BPTI during the protein folding process. The findings from previous studies in this area are compared and contrasted with the results of this study, with the hope that they can add a valuable contribution to understanding of the processes involved in oxidative protein folding in the endoplasmic reticulum. Improvements that could be made to the experiments performed in this study are discussed. Different techniques that could be used to further characterise the enzyme-substrate interface are also suggested.

8.2. Preparation of Protein Samples

The first stage in this study was the mutagenesis of the mature wild type BPTI construct, such that specific cysteine residues were mutated to serines. This prevented the formation of disulfide bonds at mutated sites, thus arresting BPTI folding at particular stages along its folding pathway. Expression studies revealed that the (30-51, 5-14) two disulfide intermediate produced sufficient yield of protein to enable this intermediate to be investigated (Figure 3.3). This intermediate is particularly interesting because it contains one native disulfide bond (30-51) and one non-native disulfide bond (5-14), so under normal circumstances disulfide isomerisation would be required before folding to the native state could be completed (Figure 1.12).

A protocol for expression and purification of both wild type BPTI and (30-51, 5-14) BPTI were established and optimised. The constructs were expressed as inclusion bodies. After incubation in denaturing and reducing agents, each BPTI construct was desalted into 10 mM HCl to maintain the protein in reduced form. The mass of each construct was verified by ESI mass spectrometry. Both constructs were then refolded to their fullest extent by dilution into a refolding

buffer containing oxidising and reducing agents. Other preparations of wild type BPTI were alkylated via a reaction with iodoacetamide, resulting in a form of BPTI lacking disulfide bonds and equivalent to a fully reduced state. ^{15}N labelled versions of all three constructs were prepared in the same manner, using ^{15}N labelled ammonium sulphate. Mass spectrometry was once again used to verify the correct oxidation state and isotopic labelling of each construct.

Circular dichroism indicated that recombinant wild type BPTI had an almost identical secondary structure profile to commercially available, purified mature BPTI (Figure 3.13). However it was also important to compare their tertiary structures via NMR, to ensure that the recombinant product reflected the actual structure of BPTI. 2D TOCSY and 2D NOESY NMR confirmed that they have very similar structures (Figure 4.2). An HSQC spectrum of recombinant wild type BPTI was then acquired and assigned at 36°C (Figure 4.3).

Mature His-tagged PDI was prepared as specified in a previous study (Alanen, Salo et al. 2003). A previously developed protocol was also used for the preparation of bb'x (Byrne, Sidhu et al. 2009).

8.2.1. Soluble Expression of BPTI Constructs

In this study, all BPTI constructs were expressed as inclusion bodies, which were subsequently isolated and purified. The resulting BPTI monomers required refolding to reach the correct oxidative state and associated structure. This is in line with most previous studies using recombinant BPTI, which also expressed the protein as inclusion bodies.

However, recent developments in recombinant expression technology have opened up the possibility of expressing partly-folded protein intermediates as soluble proteins. This involves the use of specially adapted expression systems

that allow oxidative folding and isomerisation in the cytosol, such as the SHuffle™ competent cells recently developed by New England Biolabs.

SHuffle™ competent cells are *E. coli* cells capable of oxidising proteins in the cytoplasm to promote correct disulfide bond formation. Normally, reductases in the cytoplasm would reduce any disulfide bonds that have formed. However SHuffle™ competent cells have been adapted with deletions of the genes for these reductases (specifically glutaredoxin reductase and thioredoxin reductase) (de Marco 2009) . Thus, disulfide bonds formed in the cytoplasm are able to persist. Furthermore, the cells have also been modified to express an adapted version of DsbC that lacks the signal sequence that would normally target the enzyme towards the periplasm. Thus, DsbC is retained in the cytoplasm, where it functions to catalyse isomerisation of multi-disulfide bonded proteins. Cytoplasmic expression results in significantly higher yields of disulfide bonded proteins compared to periplasmic expression (de Marco 2009).

8.3. Structural Dynamics of the (30-51, 5-14) BPTI Partly-folded Intermediate

When an HSQC spectrum was acquired for (30-51, 5-14) BPTI at 36°C, it revealed poorly resolved peaks, characteristic of a protein that was mainly unfolded (Figure 4.4). However, when an HSQC spectrum of the same sample was acquired at 5°C, it showed well dispersed peaks, indicative of a well folded protein (Figure 4.5). This supported evidence from a previous study, which showed an increased resolution in ^{1}H NMR with BPTI intermediates as the temperature was decreased (Vanmierlo, Kemmink et al. 1994). It also strongly suggests that this construct was flexible at higher temperatures, but was stabilised by a reduction in temperature. However, the HSQC spectrum of (30-51, 5-14) BPTI at 5°C was still quite different from the wild type spectrum (Figure 4.10). Since a structure of this intermediate had not been determined in

the given conditions, a series of NMR experiments were performed to establish the backbone assignments. Once the HSQC spectrum could be assigned, it was clear that the N-terminal region of the protein was not visible. This was most likely due to line broadening caused by the N-terminal region being highly flexible, even at 5°C. This could be explained by the lack of a 5-55 disulfide bond, which would normally keep the N-terminus tethered to the C-terminus of the protein (see Figure 1.11). The non-native 5-14 disulfide creates a large flexible N-terminal loop with very little secondary structure that is highly dynamic and explores many different local chemical environments within the NMR timescale, causing line broadening in the HSQC spectrum and hence no signal is detected.

8.3.1. Is the C-terminus of (30-51, 5-14) BPTI Buried?

Hydrogen deuterium exchange experiments were performed on all BPTI constructs to establish the extent to which different regions of the protein were solvent exposed. As expected, the natively folded wild type protein displayed much less solvent exposure compared with the partly-folded (30-51, 5-14) BPTI intermediate (Figure 4.17). However, although signals from C-terminal residues of wild type BPTI disappeared after just 5 minutes in D_2O buffer, signals from the two C-terminal residues of the (30-51, 5-14) persisted even after 24 hours in D_2O buffer, despite the signal from other residues disappearing (Figure 4.17). This suggests that in this particular partly-folded intermediate of BPTI, the C-terminus has become buried into the central region of the protein. These C-terminal residues are also clearly distinguishable in the HSQC spectrum at 36°C (Figure 4.12), implying a unique chemical environment which would result from being buried in the centre of the molecule, even at a temperature where the rest of the molecule shows high flexibility. Also, the C-terminal of (30-51, 5-14) BPTI seems to be the one region of the HSQC spectrum that persists even after relatively high concentrations of PDI have been added (Figure 5.6C), suggesting

that even small amounts of unbound (30-51, 5-14) BPTI are enough to provide a noticeable signal for these residues. Together, these experiments provide evidence that the C-terminus of the (30-51, 5-14) BPTI partly-folded intermediate may be buried within the core of the molecule. Such a model is quite feasible, since the native 5-55 disulfide bond is missing, hence no longer constraining the movement of this region of the protein. Similarly, the N-terminus, which is in close proximity to the C-terminus in wild type BPTI, is kept away from the C-terminal region by the 5-14 non-native disulfide, removing this steric hindrance from the C-terminal. With such a reduction in restraints the C-terminus may by expected to be very flexible, but perhaps it forms weak bonds with proximal residues when able to bury within the core of the molecule.

8.4. Reduced Alkylated BPTI is Predominantly Unstructured

HSQC spectra were also acquired for ^{15}N isotopically labelled fully reduced and alkylated wild type BPTI. Despite complications due to the limited solubility of this construct, enhanced HSQC spectra clearly showed that the protein was predominantly unfolded and very dynamic even at low temperatures (Figure 4.13). This was supported by the hydrogen deuterium exchange experiment, which showed complete disappearance of all HSQC signals before they could be detected (after just 5 min) (Figure 4.18).

8.5. Recombinant BPTI Constructs Provide a Good Representation of Different Stages along the Protein's Folding Pathway

Analysis of HSQC spectra from ^{15}N isotopically labelled samples of wild type, (30-51, 5-14) and reduced BPTI provided strong evidence that they represent natively folded, partly-folded and unfolded proteins respectively (see overlay of

spectra in Figure 4.14). Thus the available constructs acted as good representations of substrates at different stages of folding for the study of binding with the PDI protein folding enzyme.

8.6. Interaction of PDI with Wild Type BPTI

As expected, natively folded wild type BPTI showed much weaker binding to PDI than the less folded BPTI constructs. However, an interaction was nevertheless evident from observations of HSQC spectra, both from the perspective of ^{15}N labelled wild type BPTI (Figure 5.7) and the perspective of the ^{15}N labelled bb'x region of PDI (Figure 6.7). SPR experiments suggested a dissociation constant of approximately 2.9 mM for binding between wild type BPTI and full length PDI at 25°C (Figure 7.5), which is comparable to an estimate of >2 mM for binding of folded RNase and bb' from a recent study (Denisov, Maattanen et al. 2009). Since PDI is primarily a folding enzyme responsible for oxidation and isomerisation of disulfide bonds in proteins during the folding process, it may seem unexpected that a natively folded protein, such as wild type BPTI, with all disulfide bonds intact, would interact at all with PDI. However, there are two main reasons why PDI interaction with a natively folded protein would be feasible. Firstly, as well as its primary role as a folding enzyme, PDI has been demonstrated to act as a chaperone (Cai, Wang et al. 1994; Song and Wang 1995). In the natural ER environment of PDI, protein aggregation can occur due to its very crowded environment, so the chaperone function of PDI, even for natively folded proteins, may be underrated (Ellis 2001). Secondly, even natively folded, highly stable proteins such as BPTI are dynamic and are associated with an ensemble of different conformations. On occasions this will include conformations that are partially unfolded and expose hydrophobic regions of the molecule, allowing PDI to bind to it. This is demonstrated by the hydrogen deuterium exchange experiment using wild type BPTI, which

suggested that after 24 hours most of the molecule had become solvent exposed (Figure 4.17).

8.7. Comparisons between Reduced BPTI and (30-51, 5-14) BPTI Interactions with PDI

The HSQC spectrum of reduced BPTI shows substantial line broadening, even in the presence of small concentrations of PDI (Figure 5.9). It seems to produce marginally larger perturbations when present in HSQC spectra of bb'x compared to the (30-51, 5-14) BPTI intermediate (Figure 6.11). However, due to the limited solubility of reduced BPTI, it is difficult to judge if this difference is significant. SPR binding analysis estimated dissociation constants of 3.0 μM for reduced BPTI and 4.4 μM for (30-51, 5-14) BPTI when binding to full length PDI at 25°C. Overall, the evidence suggests that fully reduced BPTI has a greater affinity to PDI than the two disulfide species, but this difference is relatively small.

8.8. Binding Affinities to PDI Constructs

It is interesting to compare the binding of BPTI constructs to similar studies of PDI binding. The dissociation constant for another full length protein, RNase A, binding to the bb' region of PDI was estimated to be 35 μM for the unfolded protein, compared with at least 2 mM for folded RNase A (Denisov, Maattanen et al. 2009). The same study estimated a K_D of 35 μM for the peptide somatostatin (similar to unfolded RNase) and an weaker binding of 130 μM for peptide mastoparan (Denisov, Maattanen et al. 2009). A separate study of a truncated version of somatostatin binding to the b'x region crudely estimated a dissociation constant between 0.1-1.0 mM (Byrne, Sidhu et al. 2009). These figures suggest that BPTI constructs bind much more tightly to the binding region of PDI than any of these other constructs (except when fully folded), or

they use multiple binding interactions with different regions of PDI. Perhaps the low level of secondary structure in BPTI allows hydrophobic regions to be readily exposed when it is in a partially folded state.

The affinity of PDI for substrates seems to be comparable to that of molecular chaperones. For example, the molecular chaperone BiP binds synthetic peptides with consensus motifs with 10–60 μM affinity (Blondelguindi, Cwirla et al. 1993). The relatively weak binding of PDI even to unfolded or partly-folded BPTI, along with the large size of the binding pocket in the b' domain, is consistent with a low degree of specificity for hydrophobic ligands (Denisov, Maattanen et al. 2009). The identification of several low-affinity substrate binding sites in PDI, rather than a single high-affinity site, could account for the broad specificity of PDI for non-native proteins (Hatahet and Ruddock 2007). A recent paper proposed that the b' domain acts as a large, multivalent hydrophobic binding site and is an effective way to bind a variety of substrates when unfolded, then to release them once they have acquired their native conformation, since fewer hydrophobic residues will be exposed in the native state (Denisov, Maattanen et al. 2009). Likewise, a recent review suggested that the combination of multiple low-affinity sites may be essential for the physiological function of PDI (Hatahet and Ruddock 2009).

8.8.1. Dependence of Binding on Mixed Disulfide Formation

Some caution should be taken when interpreting the results of this study and how accurately they reflect behaviour in the ER. The folding intermediate has been prevented in proceeding along the BPTI folding pathway by mutation of the cysteine residues not involved in its current disulfide bond configuration. However, this is likely to impact on the protein's binding to PDI, since the normal processes of oxidation and isomerisation would involve the formation a mixed disulfide with a least one of the cysteines that no longer exist (Hatahet and

Ruddock 2007). This means that off rates in kinetic analysis of binding are likely to be much higher for simple binding of an intermediate stuck at one point in the folding pathway than for an intermediate that is actively undergoing oxidation or isomerisation in the ER.

Likewise, in this study, wild type BPTI was trapped in the fully reduced state by alkylation of the free thiols, preventing oxidation. In practise, binding to PDI would result in a transient but important mixed disulfide in the process of oxidising the reduced substrate (Morjana and Gilbert 1991).

An alternative approach might be to generate mutations in both PDI and BPTI that would lead to the formation of a trapped mixed disulfide between the two proteins. For example, a PDI construct whereby all but one of the active site cysteines were mutated could be oxidised in the presence of a BPTI construct that would only have a single C14S mutation. According to the BPTI folding pathway, this would result in a (30-51, 5-55) intermediate, with the C38 remaining as a free thiol. When oxidised in the presence of the PDI mutant, a PDI/BPTI complex would form with a mixed disulfide bond between the remaining active site cysteine of PDI and the C38 of BPTI. To enable NMR of PDI, a truncated construct, *e.g.* abb'x or bb'xa', could be used. As performed in this study, the PDI and BPTI constructs could be ^{15}N labelled so that their effects on each other could be observed from the perspective of both the folding enzyme and the partly-folded substrate.

Another consideration with regard to the extent of substrate binding to PDI *in vivo* is the role of glutathione, which can compete with substrate binding by PDI (Lumb and Bulleid 2002).

8.9. Approximation to Physiological Conditions

The pH of the endoplasmic reticulum at rest is approximately 7.1 (Kim, Johannes et al. 1998), although it is highly permeable to H^+ equivalents, so is susceptible to alterations in the cytosolic pH. All NMR experiments in this study were performed at pH 6.5, since this was the pH used for the acquisition of previous PDI spectra. However, it is believed the structure would not alter with the relatively small increase in pH to that of the ER.

It should be noted that during the folding process *in vivo*, BPTI would actually be in the form of proBPTI, still containing N- and C-terminal extensions that are cleaved after export from the cell. Despite this sequence containing an extra cysteine residue which could theoretically result in an alternative protein folding pathway, one previous study suggests that this residue does not significantly alter the pathway of BPTI (Creighton, Bagley et al. 1993). This was contrary to an earlier study, which found that both the rate and yield of natively folded BPTI increased using proBPTI (Weissman and Kim 1992a). In any case, even if the folding pathway of BPTI is unaltered by these extensions, this should not be taken as a general rule and it remains plausible that the precursor of a protein may result in a different folding pathway than with the study of mature protein.

All protein refolding in this study was performed *in vitro* by dilution into a buffered solution with oxidising and reducing agents. This proved sufficient for the successful refolding of BPTI. However, the endoplasmic reticulum is a very crowded environment, and the effects of such protein crowding on the folding process need to be considered before the pathway can be said to reflect the folding pathway of BPTI *in vivo* (Ellis 2001).

8.10. Alternative NMR Techniques to Characterise PDI/BPTI Binding

8.10.1. Observing Substrate Folding in Real Time

Under certain conditions it may be possible to observe the refolding of BPTI in real time by NMR. Starting with fully reduced, but not alkylated, BPTI, short HSQC spectra could be acquired in quick succession to observe the refolding of BPTI to its native state. This could theoretically be done in both the absence and presence of PDI. However, the main obstacle would be the timescale of refolding, which is normally very rapid. Conditions would have to be optimised so that the folding process would be slowed considerably, perhaps by refolding at a lower pH or with a lower concentration of oxidising agent. A previous study used HSQC spectra to follow the kinetics of protein folding (Balbach, Forge et al. 1996). Here, the intensities and line shapes of the cross peaks in the spectrum reflected the kinetic time course of the folding events that occurred during the spectral accumulation.

8.10.2. Hydrogen Deuterium Exchange to Observe Enzyme-Substrate Interactions

This study used HSQC spectra to investigate H/D exchange of each of the BPTI construct in isolation (section 4.6). This was informative about both the structure and dynamics of BPTI at different stages along its folding pathway. The same technique could be used to study PDI/BPTI interactions by acquiring HSQC spectra of BPTI in the presence of PDI (Mandell, Baerga-Ortiz et al. 2005). By comparing the regions exchanged in the absence and presence of PDI, the areas of BPTI buried by binding could be identified, since they will be less solvent exposed when in the PDI/BPTI complex. Likewise, similar experiments

could be performed with the bb'x region of PDI in the absence and presence of BPTI constructs. Together, the data could reveal further detail about the interaction interface. NMR data could be complemented with measurements using mass spectrometry (Komives 2005).

8.10.3. Exchange-transferred NOE Spectroscopy

For a long time intramolecular NOEs have been used to aid the determination of protein structures. However, NOEs can also be transferred from one molecule to another in close proximity. Exchange-transferred NOE spectroscopy (et-NOESY) can be used to probe the conformation of small molecule ligands in association with high molecular weight proteins (Post 2003). The experiment is capable of giving information on the conformation of bound ligands (Neuhaus and Williamson 2000). It is applied to systems for which exchange is fast on the chemical shift timescale, so that ligand protons show a single resonance peak averaged over the free and bound states. An important consideration in all applications of et-NOESY is the demonstration of specific binding and the exclusion of nonspecific binding (Post 2003). The experiment relies on weakly binding ligands that remain on the enzyme for a relatively short time, so that NOEs between nuclei in the bound ligand are transferred to the more easily detected nuclei in the ligand free in solution by virtue of chemical exchange between the free and bound states (Evans 1995). However, ligands used in et-NOESY need to be small enough so that cross-peaks observed in the bound state are sufficiently larger than in the free state. This generally equates to an upper limit of 5 kDa for the ligand (Post 2003), so at 6.6 kDa BPTI would be a very challenging substrate.

8.10.4. Saturation Transfer Difference Spectroscopy

Another technique that exploits the intermolecular transfer of NOE is saturation transfer difference spectroscopy (STD). This technique also allows the detection of transient binding of ligands to macromolecular receptors, but does not have the same restriction with regard to the size of the ligand, so protein-protein interactions are more feasible (Post 2003). The STD method can be used to determine which part of the ligand molecule is responsible for binding, since the most strongly interacting groups of a ligand will show stronger STD effects. The method relies on the ability to selectively saturate protons of the macromolecular receptor by irradiating the spectral region containing “wings” of broad resonances of the macromolecule, which is also free of any smaller molecule signals (Mayer and Meyer 1999). Due to effective spin diffusion, saturation quickly propagates across the entire receptor. If the ligand is bound to the receptor, saturation will also spread onto the ligand. The result will be that the intensity of the ligand signal will be attenuated. Subtraction of the resulting spectrum from the reference spectrum (without saturation) yields the STD spectrum containing only signals of the binding ligands (Klein, Meinecke et al. 1999). STD can be particularly effective at identifying the contact surface of a protein-ligand interaction (Post 2003). As such, it could be a very useful technique in identifying more precisely the PDI-BPTI binding interface for BPTI at each stage along its folding pathway.

8.10.5. Water-Ligand Observed via Gradient Spectroscopy (WaterLOGSY)

WaterLOGSY is a variant of STD NMR spectroscopy that utilises the bound water at protein-ligand interfaces. It differs from STD spectroscopy in that the water is magnetised then transferred to the protein, whereas with STD the protein is magnetised directly (Meyer and Peters 2003). The WaterLOGSY

experiment utilises water molecules present at the protein-ligand interface and uses intermolecular NOE and chemical exchange to transfer magnetisation from bulk water to the protein. The magnetisation is then transferred to the bound ligand, which can leave the binding site carrying with it the magnetisation acquired from binding (Dalvit, Fogliatto et al. 2001). The result is that the resonances of non-binding ligands appear weaker than those of interacting ligands. A comparison can then be performed with a WaterLOGSY experiment of the ligand solution without the receptor protein present. Quantitative estimates of K_D can also be obtained (Fielding 2007). WaterLOGSY could potentially be used to complement STD experiments in further characterisation of the PDI-BPTI binding interface, comparing how each different BPTI construct binds to the enzyme and possibly as a method for verifying binding constants estimated using SPR.

8.11. Alternative Biophysical Approaches

8.11.1. Small Angle X-Ray Scattering

Small-angle X-ray scattering (SAXS) has been used to study properties of full length human PDI free in solution (Solda, Garbi et al. 2006), as well as another PDI family member (Kozlov, Maattanen et al. 2009). SAXS may be able to confirm or extend knowledge of substrate binding through crude measurements of the effect of binding a protein substrate. This technique has the advantage of being unlimited by either the size of the enzyme or the size of the substrate, so full length PDI could be observed in solution, both in the presence and absence of BPTI. However, the technique lacks the high resolution available from NMR or X-ray crystallography and it may be difficult to distinguish differences using such a small substrate protein.

8.12. Partly Folded Proteins in the Endoplasmic Reticulum

Nascent polypeptide chains often begin folding co-translationally and folding is completed post-translationally. Hence, when studying protein folding in the ER, most nascent proteins will be in a partly folded state, rather than being completely unfolded. Thus, many ER resident folding enzymes and quality control machinery will have partly folded proteins as their most predominant substrate. Despite this, most previous studies investigating the interaction of a substrate with folding enzymes have focused on using fully reduced, unfolded proteins or peptides as substrates.

PDI is an ER resident enzyme that functions both as a folding catalyst, assisting substrate proteins in reaching their native state, and as a chaperone, preventing protein misfolding and misassembly.

This study investigated the interactions between PDI and BPTI at different stages along its folding pathway. BPTI was prepared in both its natively folded and fully reduced forms. However, it was the preparation of a partly folded BPTI, representing an intermediate in the folding pathway of this protein, which was of greatest interest.

Structural characterisation of the folding intermediate by NMR indicated that the protein is highly flexible at physiological temperatures, indicating behaviour of a mainly unfolded protein. Interestingly, however, the stability of the protein increased markedly with a decrease in temperature. At 5°C, most of the protein adopts a well folded conformation, but the N-terminal region remains flexible. This verifies that at a low temperature the folding intermediate of BPTI is indeed a partly folded protein, a finding supported by comparison with NMR experiments of natively folded and fully reduced BPTI.

Interaction of the partly folded substrate with PDI revealed their binding affinity to be less than with the enzyme binding to a fully reduced protein, presumably due to having fewer exposed hydrophobic regions. However, the binding was still much stronger than PDI binding to the natively folded protein, which would have very few exposed hydrophobic regions.

NMR analysis of the structural changes in PDI during binding with the partly folded protein suggested that the regions of the enzyme affected by binding were similar to those affected by binding of both the reduced and natively folded substrate. These regions also broadly match the affects observed due to the binding of other substrates from previous studies.

Together, these findings imply that although binding affinity with PDI generally decreases as a substrate progresses along its folding pathway, the regions of binding on the enzyme remain unchanged.

This study focused on the interaction of a partly folded protein with the folding enzyme PDI. However, partly folded proteins will interact with many different enzymes in the crowded environment of the ER. Without a greater understanding of how partly folded proteins interact with the wide variety of folding enzymes and chaperones available in the ER, a full picture of how they are treated cannot be obtained. The outcomes of partly folded proteins are determined by interactions with proteins of both the folding and quality control machinery. The actions taken on partly-folded substrates by the ER machinery will determine its fate, either as a natively folded protein that is exported from the ER with appropriate signal sequence to reach its final destination or to become misfolded, where it may become degraded or lead to aggregate formation, Figure 8.1.

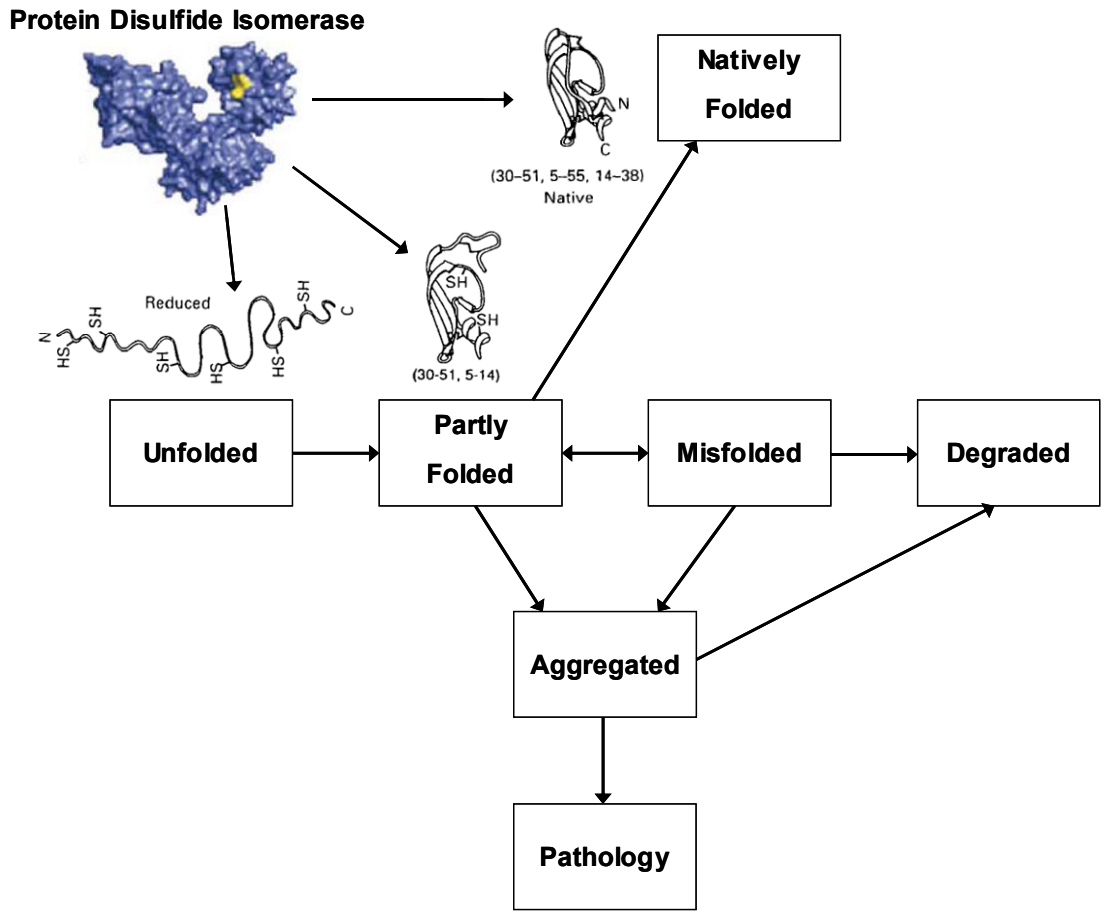


Figure 8.1: General protein folding pathways and consequences in the ER. The outcomes of partly-folded proteins are determined by interaction with proteins of the folding and quality control machinery. If not tightly controlled, protein misfolding can occur and the crowded environment of the ER can lead to aggregation. Misfolded proteins are often degraded, but aggregate formation can lead to pathology. One multifunctional enzyme that both assists protein folding to the native state (as an oxidoreductase) and prevents protein misfolding or aggregation (as a chaperone) is protein disulfide isomerase (PDI). In this study, the mechanisms through which PDI assists protein folding have been investigated by studying its interaction with a protein substrate at different stages along its folding pathway: in the fully reduced state, in a partly folded conformation and in the natively folded state (depicted graphically in the diagram). Understanding the nature of these interactions improves understanding of the general protein folding process and thus provides knowledge for how protein misfolding or aggregation could be prevented.

Understanding how partly-folded proteins interact with the crowded environment of the ER is crucial to understanding how proteins reach their native three

dimensional conformations, or how they may end up in a misfolded or aggregated state. This is the first study to investigate the structural interaction of PDI with a partly-folded protein substrate. It is hoped that the findings of this study will contribute to a general understanding of oxidative protein folding in the endoplasmic reticulum.

REFERENCES

Alanen, H. I., K. E. H. Salo, et al. (2003). "Defining the domain boundaries of the human protein disulfide isomerases." Antioxidants & Redox Signaling **5**(4): 367-374.

Alanen, H. I., R. A. Williamson, et al. (2006). "ERp27, a new non-catalytic endoplasmic reticulum-located human protein disulfide isomerase family member, interacts with ERp57." Journal of Biological Chemistry **281**(44): 33727-33738.

Anelli, T., M. Alessio, et al. (2002). "ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family." EMBO Journal **21**(4): 835-844.

Anelli, T., S. Ceppi, et al. (2007). "Sequential steps and checkpoints in the early exocytic compartment during secretory IgM biogenesis." EMBO Journal **26**(19): 4177-4188.

Anfinsen, C. B. (1973). "Principles that govern folding of protein chains." Science **181**(4096): 223-230.

Anfinsen, C. B. and H. A. Scheraga (1975). "Experimental and theoretical aspects of protein folding." Adv Protein Chem **29**: 205-300.

Annunen, P., T. Helaakoski, et al. (1997). "Cloning of the human prolyl 4-hydroxylase alpha subunit isoform alpha(II) and characterization of the type II enzyme tetramer - The alpha(I) and alpha(II) subunits do not form a mixed alpha(I)alpha(II)beta(2) tetramer." Journal of Biological Chemistry **272**(28): 17342-17348.

Appenzeller-Herzog, C. and L. Ellgaard (2008). "The human PDI family: Versatility packed into a single fold." Biochimica et Biophysica Acta **1783**(4): 535-548.

Atzel, A. and J. R. Wetterau (1993). "Mechanism of microsomal triglyceride transfer protein catalyzed lipid transport." Biochemistry **32**(39): 10444-10450.

Balbach, J., V. Forge, et al. (1996). "Protein folding monitored at individual residues during a two-dimensional NMR experiment." Science **274**(5290): 1161-1163.

Baldwin, A. J. and L. E. Kay (2009). "NMR spectroscopy brings invisible protein states into focus." Nature Chemical Biology **5**(11): 808-814.

Banhegyi, G., L. Lusini, et al. (1999). "Preferential transport of glutathione versus glutathione disulfide in rat liver microsomal vesicles." Journal of Biological Chemistry **274**(18): 12213-12216.

Barak, N. N., P. Neumann, et al. (2009). "Crystal structure and functional analysis of the protein disulfide isomerase-related protein ERp29." Journal of Molecular Biology **385**(5): 1630-1642.

Barredo, J.-L. (2005). "Enzyme biosensors". In Microbial enzymes and biotransformations, pp. 29-60 Totowa, N.J., Humana Press.

Bartlett, A. I. and S. E. Radford (2009). "An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms." Nature Structural & Molecular Biology **16**(6): 582-588.

Bass, R., L. W. Ruddock, et al. (2004). "A major fraction of endoplasmic reticulum-located glutathione is present as mixed disulfides with protein." Journal of Biological Chemistry **279**(7): 5257-5262.

Berndt, K. D., P. Guntert, et al. (1992). "Determination of a high quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with 3 crystal structures." Journal of Molecular Biology **227**(3): 757-775.

Biacore (1997). "Evaluating kinetic data". In BIAevaluation software handbook, pp. 41-43.

Biacore (2003). "Covalent immobilization and capturing methods". In Biacore sensor surface handbook, pp. 31-56.

Biamonti, C. (1996). "Structural and dynamic investigations of macromolecular recognition processes by nuclear magnetic resonance spectroscopy." Thesis, Rutgers University.

Birkbeck. (2009). "Spectral regions for two dimensional proton NMR." from http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/images/2dovrv_1.gif.

Bjelland, S., K. Wallevik, et al. (1983). "Immunological identity between bovine preparations of thiol-protein-disulfide oxidoreductase, glutathione-insulin transhydrogenase and protein disulfide isomerase." Biochimica et Biophysica Acta **747**(3): 197-199.

Blondelguindi, S., S. E. Cwirla, et al. (1993). "Affinity panning of a library of peptides displayed on bacteriophages reveals the binding-specificity of BiP." Cell **75**(4): 717-728.

Brockwell, D. J. and S. E. Radford (2007). "Intermediates: ubiquitous species on folding energy landscapes?" Current Opinion in Structural Biology **17**(1): 30-37.

Brodsky, J. L., E. D. Werner, et al. (1999). "The requirement for molecular chaperones during endoplasmic reticulum-associated protein degradation demonstrates that protein export and import are mechanistically distinct." Journal of Biological Chemistry **274**(6): 3453-3460.

Brown, L. R., A. Demarco, et al. (1978). "Influence of a single salt bridge on static and dynamic features of globular solution conformation of basic pancreatic trypsin inhibitor - 1H and 13C NMR studies of native and transaminated inhibitor." European Journal of Biochemistry **88**(1): 87-95.

Buczek, O., D. Krowarsch, et al. (2002). "Thermodynamics of single peptide bond cleavage in bovine pancreatic trypsin inhibitor (BPTI)." Protein Science **11**(4): 924-932.

Bukau, B. and A. L. Horwich (1998). "The Hsp70 and Hsp60 chaperone machines." Cell **92**(3): 351-366.

Bulaj, G., T. Kortemme, et al. (1998). "Ionization-reactivity relationships for cysteine thiols in polypeptides." Biochemistry **37**(25): 8965-8972.

Byrne, L. J., A. Sidhu, et al. (2009). "Mapping of the ligand-binding site on the b' domain of human PDI: interaction with peptide ligands and the x-linker region." Biochemical Journal **423**: 209-217.

Cai, H., C. C. Wang, et al. (1994). "Chaperone-like activity of protein disulfide isomerase in the refolding of a protein with no disulfide bonds." Journal of Biological Chemistry **269**(40): 24550-24552.

Charlton, L. M., C. O. Barnes, et al. (2008). "Residue-level interrogation of macromolecular crowding effects on protein stability." Journal of the American Chemical Society **130**(21): 6826-6830.

Chen, Y. J., Y. Zhang, et al. (2005). "SPD - A web-based secreted protein database." Nucleic Acids Research **33**: D169-D173.

Cheung, M. S. and D. Thirumalai (2007). "Effects of crowding and confinement on the structures of the transition state ensemble in proteins." Journal of Physical Chemistry B **111**(28): 8250-8257.

Clow, F., J. D. Fraser, et al. (2008). "Immobilization of proteins to biacore sensor chips using Staphylococcus aureus sortase A." Biotechnology Letters **30**(9): 1603-1607.

Coe, H. and M. Michalak (2010). "ERp57, a multifunctional endoplasmic reticulum resident oxidoreductase." International Journal of Biochemistry & Cell Biology **42**(6): 796-799.

Cooke, R. M. (1997). "Protein NMR extends into new fields of structural biology." Current Opinion in Chemical Biology **1**(3): 359-364.

Creighton, T. E. (1975). "Interactions between cysteine residues as probes of protein conformation - Disulfide bond between Cys-14 and Cys-38 of pancreatic trypsin inhibitor." Journal of Molecular Biology **96**(4): 767-776.

Creighton, T. E. (1977). "Effects of urea and guanidine-HCl on folding and unfolding of pancreatic trypsin inhibitor." Journal of Molecular Biology **113**(2): 313-328.

Creighton, T. E. (1978). "Experimental studies of protein folding and unfolding." Progress in Biophysics & Molecular Biology **33**(3): 231-297.

Creighton, T. E. (1979). "Electrophoretic analysis of the unfolding of proteins by urea." Journal of Molecular Biology **129**(2): 235-264.

Creighton, T. E. (1980). "Kinetic study of protein unfolding and refolding using urea gradient electrophoresis." Journal of Molecular Biology **137**(1): 61-80.

Creighton, T. E. (1986). "Disulfide bonds as probes of protein folding pathways." Methods Enzymol **131**: 83-106.

Creighton, T. E. (1992). "The disulfide folding pathway of BPTI." Science **256**(5053): 111-112.

Creighton, T. E. (1995a). "Disulfide-coupled protein folding pathways." Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **348**(1323): 5-10.

Creighton, T. E. (1995b). "Protein folding - An unfolding story." Current Biology **5**(4): 353-356.

Creighton, T. E., C. J. Bagley, et al. (1993). "On the biosynthesis of bovine pancreatic trypsin inhibitor (BPTI) - Structure, processing, folding and disulfide bond formation of the precursor in vitro and in microsomes." Journal of Molecular Biology **232**(4): 1176-1196.

Creighton, T. E., N. J. Darby, et al. (1996). "The roles of partly folded intermediates in protein folding." FASEB Journal **10**(1): 110-118.

Creighton, T. E., D. A. Hillson, et al. (1980). "Catalysis by protein disulfide isomerase of the unfolding and refolding of proteins with disulfide bonds." Journal of Molecular Biology **142**(1): 43-62.

Creighton, T. E., A. Zapun, et al. (1995). "Mechanisms and catalysts of disulfide bond formation in proteins." Trends in Biotechnology **13**(1): 18-23.

Cunnea, P. M., A. Miranda-Vizuete, et al. (2003). "ERdj5, an endoplasmic reticulum (ER)-resident protein containing DnaJ and thioredoxin domains, is expressed in secretory cells or following ER stress." Journal of Biological Chemistry **278**(2): 1059-1066.

Cuozzo, J. W. and C. A. Kaiser (1999). "Competition between glutathione and protein thiols for disulphide-bond formation." Nature Cell Biology **1**(3): 130-135.

Dalvit, C., G. Fogliatto, et al. (2001). "WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability." Journal of Biomolecular NMR **21**(4): 349-359.

Darby, N. J. and T. E. Creighton (1993). "Dissecting the disulfide-coupled folding pathway of bovine pancreatic trypsin inhibitor - Forming the 1st disulfide bonds in analogs of the reduced protein." Journal of Molecular Biology **232**(3): 873-896.

Darby, N. J. and T. E. Creighton (1995). "Functional properties of the individual thioredoxin-like domains of protein disulfide isomerase." Biochemistry **34**(37): 11725-11735.

Darby, N. J., P. E. Morin, et al. (1995). "Refolding of bovine pancreatic trypsin inhibitor via nonnative disulfide intermediates." Journal of Molecular Biology **249**(2): 463-477.

Darby, N. J., E. Penka, et al. (1998). "The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency." Journal of Molecular Biology **276**(1): 239-247.

Darby, N. J., C. P. M. Vanmierlo, et al. (1992). "Kinetic roles and conformational properties of the nonnative 2-disulfide intermediates in the refolding of bovine pancreatic trypsin inhibitor." Journal of Molecular Biology **224**(4): 905-911.

De Lorenzo, F., R. F. Goldberger, et al. (1966). "Purification and properties of an enzyme from beef liver which catalyzes sulfhydryl-disulfide interchange in proteins." J Biol Chem **241**(7): 1562-7.

de Marco, A. (2009). "Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli." Microbial Cell Factories **8**: 18.

de Mol, N. J. and M. J. Fischer (2010). "Surface plasmon resonance: a general introduction." Methods Mol Biol **627**: 1-14.

Deisenhofer, J. and W. Steigemann (1975). "Crystallographic refinement of structure of bovine pancreatic trypsin inhibitor at 1.5 A resolution." Acta Crystallographica Section B-Structural Science **31**: 238-250.

Denisov, A. Y., P. Maattanen, et al. (2009). "Solution structure of the bb' domains of human protein disulfide isomerase." FEBS Journal **276**(5): 1440-1449.

Denisov, A. Y., P. Maattanen, et al. (2007). "1H, 13C and 15N resonance assignments of the bb' domains of human protein disulfide isomerase." Biomolecular NMR Assignments **1**(1): 129-130.

Desai, U. R., J. J. Osterhout, et al. (1994). "Protein structure in the lyophilized state - A hydrogen isotope exchange NMR study with bovine pancreatic trypsin inhibitor." Journal of the American Chemical Society **116**(21): 9420-9422.

DeSilva, M. G., J. Lu, et al. (1996). "Characterization and chromosomal localization of a new protein disulfide isomerase, PDIp, highly expressed in human pancreas." DNA and Cell Biology **15**(1): 9-16.

DeSilva, M. G., A. L. Notkins, et al. (1997). "Molecular characterization of a pancreas-specific protein disulfide isomerase, PDIp." DNA and Cell Biology **16**(3): 269-274.

Dietrich, W. (2009). "Aprotinin: 1 year on." Current Opinion in Anesthesiology **22**(1): 121-127.

Dong, G., P. A. Wearsch, et al. (2009). "Insights into MHC Class I Peptide Loading from the Structure of the Tapasin-ERp57 Thiol Oxidoreductase Heterodimer." Immunity **30**(1): 21-32.

Dong, M., J. P. Bridges, et al. (2008). "ERdj4 and ERdj5 are required for endoplasmic reticulum-associated protein degradation of misfolded surfactant protein C." Molecular Biology of the Cell **19**(6): 2620-2630.

Echeverria, C. and R. Kapral (2010). "Macromolecular dynamics in crowded environments." Journal of Chemical Physics **132**(10): 9.

Edman, J. C., L. Ellis, et al. (1985). "Sequence of protein disulfide isomerase and implications of its relationship to thioredoxin." Nature **317**(6034): 267-270.

Eigenbrot, C., M. Randal, et al. (1992). "Structural effects induced by mutagenesis affected by crystal packing factors - the structure of a 30-51

disulfide mutant of basic pancreatic trypsin inhibitor." Proteins-Structure Function and Genetics **14**(1): 75-87.

Eklund, H., F. K. Gleason, et al. (1991). "Structural and functional relations among thioredoxins of different species." Proteins-Structure Function and Genetics **11**(1): 13-28.

Ellgaard, L. and A. Helenius (2001). "ER quality control: towards an understanding at the molecular level." Current Opinion in Cell Biology **13**(4): 431-437.

Ellgaard, L. and A. Helenius (2003). "Quality control in the endoplasmic reticulum." Nature Reviews Molecular Cell Biology **4**(3): 181-191.

Ellgaard, L., M. Molinari, et al. (1999). "Setting the standards: Quality control in the secretory pathway." Science **286**(5446): 1882-1888.

Ellgaard, L. and L. W. Ruddock (2005). "The human protein disulphide isomerase family: substrate interactions and functional properties." EMBO Reports **6**(1): 28-32.

Ellis, R. J. (1993). "The general concept of molecular chaperones." Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **339**(1289): 257-261.

Ellis, R. J. (2001). "Macromolecular crowding: obvious but underappreciated." Trends in Biochemical Sciences **26**(10): 597-604.

Ellis, R. J. and A. P. Minton (2006). "Protein aggregation in crowded environments." Biological Chemistry **387**(5): 485-497.

Engel, R., A. H. Westphal, et al. (2008). "Macromolecular crowding compacts unfolded apoflavodoxin and causes severe aggregation of the off-pathway intermediate during apoflavodoxin folding." Journal of Biological Chemistry **283**(41): 27383-27394.

Englander, S. W. and N. R. Kallenbach (1983). "Hydrogen exchange and structural dynamics of proteins and nucleic acids." Q Rev Biophys **16**(4): 521-655.

Evans, J. N. S. (1995). "The transferred nuclear Overhauser effect". In Biomolecular NMR spectroscopy, pp. 246-247. Oxford ; New York, Oxford University Press.

Ferrari, D. M. and H. D. Soling (1999). "The protein disulphide isomerase family: unravelling a string of folds." Biochemical Journal **339**: 1-10.

Fersht, A. R. (2008). "From the first protein structures to our current knowledge of protein folding: delights and scepticisms." Nature Reviews Molecular Cell Biology **9**(8): 650-654.

Fewell, S. W., K. J. Travers, et al. (2001). "The action of molecular chaperones in the early secretory pathway." Annu Rev Genet **35**: 149-91.

Fielding, L. (2007). "NMR methods for the determination of protein-ligand dissociation constants." Progress in Nuclear Magnetic Resonance Spectroscopy **51**(4): 219-242.

Flynn, G. C., J. Pohl, et al. (1991). "Peptide-binding specificity of the molecular chaperone BiP." Nature **353**(6346): 726-30.

Frand, A. R., J. W. Cuozzo, et al. (2000). "Pathways for protein disulphide bond formation." Trends in Cell Biology **10**(5): 203-210.

Frand, A. R. and C. A. Kaiser (1998). "The ERO1 gene of yeast is required for oxidation of protein dithiols in the endoplasmic reticulum." Molecular Cell **1**(2): 161-170.

Frand, A. R. and C. A. Kaiser (1999). "Ero1p oxidizes protein disulfide isomerase in a pathway for disulfide bond formation in the endoplasmic reticulum." Molecular Cell **4**(4): 469-477.

Freedman, R. B., P. J. Gane, et al. (1998). "Experimental and theoretical analyses of the domain architecture of mammalian protein disulphide isomerase." Biological Chemistry **379**(3): 321-328.

Freedman, R. B., T. R. Hirst, et al. (1994). "Protein disulfide isomerase: Building bridges in protein folding." Trends in Biochemical Sciences **19**(8): 331-336.

Freedman, R. B., P. Klappa, et al. (2002). "Protein disulfide isomerases exploit synergy between catalytic and specific binding domains." EMBO Reports **3**(2): 136-140.

Garbi, N., S. Tanaka, et al. (2006). "Impaired assembly of the major histocompatibility complex class I peptide-loading complex in mice deficient in the oxidoreductase ERp57." Nature Immunology **7**(1): 93-102.

Garrett, D. S., Y. J. Seok, et al. (1997). "Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system." Biochemistry **36**(15): 4393-4398.

Gettins, P. G. W. (2002). "Serpin structure, mechanism, and function." Chemical Reviews **102**(12): 4751-4803.

Gilchrist, A., C. E. Au, et al. (2006). "Quantitative proteomics analysis of the secretory pathway." Cell **127**(6): 1265-1281.

Gillece, P., J. M. Luz, et al. (1999). "Export of a cysteine-free misfolded secretory protein from the endoplasmic reticulum for degradation requires interaction with protein disulfide isomerase." Journal of Cell Biology **147**(7): 1443-1456.

Givol, D., F. Delorenzo, et al. (1965). "Disulfide interchange and the three-dimensional structure of proteins." PNAS **53**: 676-84.

Givol, D., R. F. Goldberger, et al. (1964). "Oxidation and disulfide interchange in the reactivation of reduced ribonuclease." J Biol Chem **239**: PC3114-16.

Glasoe, P. K. and F. A. Long (1960). "Use of Glass Electrodes to Measure Acidities in Deuterium Oxide." The Journal of Physical Chemistry **64**(1): 188-190.

Goldberger, R. F., C. J. Epstein, et al. (1963). "Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver." J Biol Chem **238**: 628-35.

Goldberger, R. F., C. J. Epstein, et al. (1964). "Purification and properties of a microsomal enzyme system catalyzing the reactivation of reduced ribonuclease and lysozyme." J Biol Chem **239**: 1406-10.

Goldenberg, D. P. (1988). "Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and cysteine-38 thiols." Biochemistry **27**(7): 2481-2489.

Goldenberg, D. P. (1992). "Native and nonnative intermediates in the BPTI folding pathway." Trends in Biochemical Sciences **17**(7): 257-261.

Gordon, D. A. and H. Jamil (2000). "Progress towards understanding the role of microsomal triglyceride transfer protein in apolipoprotein-B lipoprotein assembly." Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids **1486**(1): 72-83.

Gronenborn, A. M. and G. M. Clore (1995). "Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy." Critical Reviews in Biochemistry and Molecular Biology **30**(5): 351-385.

Gruber, C. W., M. Cemazar, et al. (2006). "Protein disulfide isomerase: the structure of oxidative folding." Trends in Biochemical Sciences **31**(8): 455-464.

Grzesiak, A., R. Helland, et al. (2000). "Substitutions at the P-1' position in BPTI strongly affect the association energy with serine proteinases." Journal of Molecular Biology **301**(1): 205-217.

Hammond, C. and A. Helenius (1994). "Quality-control in the secretory pathway - Retention of a misfolded viral membrane glycoprotein involves cycling between the ER, intermediate compartment, and Golgi-apparatus." Journal of Cell Biology **126**(1): 41-52.

Hansen, P. E., C. Lauritzen, et al. (1995). "Determination of structural changes in BPTI mutants using 13C NMR chemical shifts." Journal of Cellular Biochemistry: 72-72.

Hanson, W. M., G. J. Domek, et al. (2007). "Rigidification of a flexible protease inhibitor variant upon binding to trypsin." Journal of Molecular Biology **366**(1): 230-243.

Hartl, F. U. (1996). "Molecular chaperones in cellular protein folding." Nature **381**(6583): 571-580.

Hartl, F. U. and M. Hayer-Hartl (2009). "Converging concepts of protein folding in vitro and in vivo." Nature Structural & Molecular Biology **16**(6): 574-581.

Hatahet, F. and L. W. Ruddock (2007). "Substrate recognition by the protein disulfide isomerases." FEBS Journal **274**(20): 5223-5234.

Hatahet, F. and L. W. Ruddock (2009). "Protein disulfide isomerase: A critical evaluation of its function in disulfide bond formation." Antioxidants & Redox Signaling **11**(11): 2807-2850.

Haugstetter, J., T. Blicher, et al. (2005). "Identification and characterization of a novel thioredoxin-related transmembrane protein of the endoplasmic reticulum." Journal of Biological Chemistry **280**(9): 8371-8380.

Havel, T. F. and K. Wuthrich (1985). "An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution." Journal of Molecular Biology **182**(2): 281-294.

Hawkins, H. C. and R. B. Freedman (1975). "Randomly reoxidized soybean trypsin inhibitor and possibility of conformational barriers to disulfide isomerization in proteins." FEBS Letters **58**(1): 7-11.

Hawkins, H. C. and R. B. Freedman (1991). "The reactivities and ionization properties of the active site dithiol groups of mammalian protein disulfide isomerase." Biochemical Journal **275**: 335-339.

Helenius, A. and M. Aebi (2004). "Roles of N-linked glycans in the endoplasmic reticulum." Annual Review of Biochemistry **73**: 1019-1049.

Hendershot, L., J. Y. Wei, et al. (1996). "Inhibition of immunoglobulin folding and secretion by dominant negative BiP ATPase mutants." PNAS **93**(11): 5269-5274.

Hillson, D. A. and R. B. Freedman (1980). "Resolution of protein disulfide isomerase and glutathione-insulin transhydrogenase activities by covalent chromatography - Properties of the purified protein disulfide isomerase." Biochemical Journal **191**(2): 373-388.

Hillson, D. A., N. Lambert, et al. (1984). "Formation and isomerization of disulfide bonds in proteins - protein disulfide isomerase." Methods in Enzymology **107**: 281-294.

Hosoda, A., Y. Kimata, et al. (2003). "JPDI, a novel endoplasmic reticulum-resident protein containing both a BiP-interacting J-domain and thioredoxin-like motifs." Journal of Biological Chemistry **278**(4): 2669-2676.

Hu, H. Y. and H. N. Du (2000). "Alpha-to-beta structural transformation of ovalbumin: Heat and pH effects." Journal of Protein Chemistry **19**(3): 177-183.

Huber, R., D. Kukla, et al. (1970). "Basic trypsin inhibitor of bovine pancreas .1. Structure analysis and conformation of polypeptide chain." Naturwissenschaften **57**(8): 389-392.

Hussain, M. M., J. Shi, et al. (2003). "Microsomal triglyceride transfer protein and its role in apoB-lipoprotein assembly." Journal of Lipid Research **44**(1): 22-32.

Huth, J. R., F. Perini, et al. (1993). "Protein folding and assembly in vitro parallel intracellular folding and assembly - Catalysis of folding and assembly of the human chorionic gonadotropin alpha-beta dimer by protein disulfide isomerase." Journal of Biological Chemistry **268**(22): 16472-16482.

Hwang, C., A. J. Sinskey, et al. (1992). "Oxidized redox state of glutathione in the endoplasmic reticulum." Science **257**(5076): 1496-1502.

Jessop, C. E., T. J. Tavender, et al. (2008). "Substrate specificity of the oxidoreductase ERp57 is determined primarily by its interaction with calnexin and calreticulin." J Biol Chem.

Jessop, C. E., R. H. Watkins, et al. (2009). "Protein disulphide isomerase family members show distinct substrate specificity: P5 is targeted to BiP client proteins." Journal of Cell Science **122**(23): 4287-4295.

John, D. C. A., M. E. Grant, et al. (1993). "Cell-free synthesis and assembly of prolyl 4-hydroxylase - the role of the beta-subunit (PDI) in preventing

misfolding and aggregation of the alpha-subunit." EMBO Journal **12**(4): 1587-1595.

Johnsson, B., S. Lofas, et al. (1991). "Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors." Analytical Biochemistry **198**(2): 268-277.

Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure - Pattern-recognition of hydrogen-bonded and geometrical features." Biopolymers **22**(12): 2577-2637.

Kaderbhai, M. A. and B. M. Austen (1985). "Studies on the formation of intrachain disulfide bonds in newly biosynthesized bovine prolactin - Role of protein disulfide isomerase." European Journal of Biochemistry **153**(1): 167-178.

Kanai, S., H. Toh, et al. (1998). "Molecular evolution of the domain structures of protein disulfide isomerases." Journal of Molecular Evolution **47**(2): 200-210.

Karala, A. R., A. K. Lappi, et al. (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**(4): 883-892.

Karala, A. R., A. K. Lappi, et al. (2009). "Efficient Peroxide-Mediated Oxidative Refolding of a Protein at Physiological pH and Implications for Oxidative Folding in the Endoplasmic Reticulum." Antioxidants & Redox Signaling **11**(5): 963-970.

Karkouti, K., D. N. Wijeysundera, et al. (2010). "The Risk-Benefit Profile of Aprotinin Versus Tranexamic Acid in Cardiac Surgery." Anesthesia and Analgesia **110**(1): 21-29.

Kawasaki, T. and M. Kasai (1994). "Regulation of calcium channel in sarcoplasmic reticulum by calsequestrin." Biochemical and Biophysical Research Communications **199**(3): 1120-1127.

Kemmink, J. and T. E. Creighton (1993). "Local conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor and their roles in folding." Journal of Molecular Biology **234**(3): 861-878.

Kemmink, J., N. J. Darby, et al. (1996). "Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear 13C/15N NMR spectroscopy." Biochemistry **35**(24): 7684-7691.

Kemmink, J., N. J. Darby, et al. (1997). "The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules." Current Biology **7**(4): 239-245.

Kemmink, J., K. Dijkstra, et al. (1999). "The structure in solution of the b domain of protein disulfide isomerase." Journal of Biomolecular NMR **13**(4): 357-368.

Kemmink, J., C. P. M. Vanmierlo, et al. (1993). "Local structure due to an aromatic amide interaction observed by 1H nuclear magnetic resonance spectroscopy in peptides related to the N terminus of bovine pancreatic trypsin inhibitor." Journal of Molecular Biology **230**(1): 312-322.

Kim, J. H., L. Johannes, et al. (1998). "Noninvasive measurement of the pH of the endoplasmic reticulum at rest and during calcium release." PNAS **95**(6): 2997-3002.

Kivirikko, K. I., R. Myllyla, et al. (1989). "Protein hydroxylation - Prolyl 4-hydroxylase, an enzyme with 4 cosubstrates and a multifunctional subunit." FASEB Journal **3**(5): 1609-1617.

Kivirikko, K. I. and T. Pihlajaniemi (1998). Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. Advances in Enzymology. **72:** 325-398.

Klappa, P., L. W. Ruddock, et al. (1998). "The b' domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins." EMBO Journal **17**(4): 927-935.

Klein, J., R. Meinecke, et al. (1999). "Detecting binding affinity to immobilized receptor proteins in compound libraries by HR-MAS STD NMR." Journal of the American Chemical Society **121**(22): 5336-5337.

Koivu, J., R. Myllyla, et al. (1987). "A single polypeptide acts both as the beta-subunit of prolyl 4-hydroxylase and as a protein disulfide isomerase." Journal of Biological Chemistry **262**(14): 6447-6449.

Komives, E. A. (2005). "Protein-protein interaction dynamics by amide H/H-2 exchange mass spectrometry." International Journal of Mass Spectrometry **240**(3): 285-290.

Kortemme, T., N. J. Darby, et al. (1996). "Electrostatic interactions in the active site of the N-terminal thioredoxin-like domain of protein disulfide isomerase." Biochemistry **35**(46): 14503-14511.

Kosen, P. A., T. E. Creighton, et al. (1980). "Ultraviolet difference spectroscopy of intermediates trapped in unfolding and refolding of bovine pancreatic trypsin inhibitor." Biochemistry **19**(21): 4936-4944.

Kosen, P. A., T. E. Creighton, et al. (1981). "Circular dichroism spectroscopy of bovine pancreatic trypsin inhibitor and 5 altered conformational states - Relationship of conformation and the refolding pathway of the trypsin inhibitor." Biochemistry **20**(20): 5744-5754.

Kozlov, G., P. Maattanen, et al. (2009). "Structure of the Noncatalytic Domains and Global Fold of the Protein Disulfide Isomerase ERp72." Structure **17**(5): 651-659.

Kozlov, G., P. Maattanen, et al. (2006). "Crystal structure of the bb' domains of the protein disulfide isomerase ERp57." Structure **14**(8): 1331-1339.

Krishna, M. M. G., L. Hoang, et al. (2004). "Hydrogen exchange methods to study protein folding." Methods **34**(1): 51-64.

Kubelka, J., J. Hofrichter, et al. (2004). "The protein folding 'speed limit'." Current Opinion in Structural Biology **14**(1): 76-88.

Kumar, A., R. R. Ernst, et al. (1980). "A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete

proton-proton cross-relaxation networks in biological macromolecules." Biochemical and Biophysical Research Communications **95**(1): 1-6.

Kunitz, M. and J. H. Northrop (1936). "Isolation from Beef Pancreas of Crystalline Trypsinogen, Trypsin, a Trypsin Inhibitor, and an Inhibitor-Trypsin Compound." J Gen Physiol **19**(6): 991-1007.

Laemmli, U. K. (1970). "Cleavage of structural proteins during assembly of head of bacteriophage T4." Nature **227**(5259): 680-685.

Lamberg, A., M. Jauhiainen, et al. (1996). "The role of protein disulphide isomerase in the microsomal triacylglycerol transfer protein does not reside in its isomerase activity." Biochemical Journal **315**: 533-536.

Lambert, N. and R. B. Freedman (1985). "The latency of rat liver microsomal protein disulfide isomerase." Biochemical Journal **228**(3): 635-645.

Land, A., D. Zonneveld, et al. (2003). "Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage." FASEB Journal **17**(9): 1058-1067.

Lee, A. S. (1992). "Mammalian stress response: induction of the glucose-regulated protein family." Curr Opin Cell Biol **4**(2): 267-73.

Lee, S. O., K. Cho, et al. (2010). "Protein disulphide isomerase is required for signal peptide peptidase-mediated protein degradation." EMBO Journal **29**(2): 363-375.

Li, S. J., X. G. Hong, et al. (2006). "Annular arrangement and collaborative actions of four domains of protein disulfide isomerase - A small angle X-ray scattering study in solution." Journal of Biological Chemistry **281**(10): 6581-6588.

Liang, Y., J. Li, et al. (2001). "Thermodynamics of the folding of D-glyceraidehyde-3-phosphate dehydrogenase assisted by protein disulfide isomerase studied by microcalorimetry." European Journal of Biochemistry **268**(15): 4183-4189.

Liedberg, B., C. Nylander, et al. (1983). "Surface plasmon resonance for gas detection and biosensing." Sensors and Actuators **4**(2): 299-304.

Liepinsh, E., M. Baryshev, et al. (2001). "Thioredoxin fold as homodimerization module in the putative chaperone ERp29: NMR structures of the domains and experimental model of the 51 kDa dimer." Structure **9**(6): 457-471.

Lippert, U., D. Diao, et al. (2007). "Conserved structural and functional properties of D-domain containing redox-active and -inactive protein disulfide isomerase-related protein chaperones." Journal of Biological Chemistry **282**(15): 11213-11220.

Liu, Y., K. Breslauer, et al. (1997). ""Designing out" disulfide bonds: Thermodynamic properties of 30-51 cystine substitution mutants of bovine pancreatic trypsin inhibitor." Biochemistry **36**(18): 5323-5335.

Lumb, R. A. and N. J. Bulleid (2002). "Is protein disulfide isomerase a redox-dependent molecular chaperone?" EMBO Journal **21**(24): 6763-6770.

Ma, Y. J. and L. M. Hendershot (2004). "The role of the unfolded protein response in tumour development: Friend or foe?" Nature Reviews Cancer **4**(12): 966-977.

Macer, D. R. and G. L. Koch (1988). "Identification of a set of calcium-binding proteins in reticuloplasm, the luminal content of the endoplasmic reticulum." Journal of cell science **91**: 61-70.

Mahdy, A. M. and N. R. Webster (2004). "Perioperative systemic haemostatic agents." British Journal of Anaesthesia **93**(6): 842-858.

Mamathambika, B. S. and J. C. Bardwell (2008). "Disulfide-linked protein folding pathways." Annual Review of Cell and Developmental Biology **24**: 211-235.

Mandell, J. G., A. Baerga-Ortiz, et al. (2005). "Measurement of solvent accessibility at protein-protein interfaces." Methods Mol Biol **305**: 65-80.

Martin, J. L. (1995). "Thioredoxin - A fold for all reasons." Structure **3**(3): 245-250.

Matsuo, Y., N. Akiyama, et al. (2001). "Identification of a novel thioredoxin-related transmembrane protein." Journal of Biological Chemistry **276**(13): 10032-10038.

Maudsley, A. A. and R. R. Ernst (1977). "Indirect detection of magnetic resonance by heteronuclear 2-dimensional spectroscopy." Chemical Physics Letters **50**(3): 368-372.

Mayer, M. and B. Meyer (1999). "Characterization of ligand binding by saturation transfer difference NMR spectroscopy." Angewandte Chemie-International Edition **38**(12): 1784-1788.

McArthur, A. G., L. A. Knodler, et al. (2001). "The evolutionary origins of eukaryotic protein disulfide isomerase domains: New evidence from the amitochondriate protist Giardia lamblia." Molecular Biology and Evolution **18**(8): 1455-1463.

McCammon, J. A., B. R. Gelin, et al. (1977). "Dynamics of folded proteins." Nature **267**(5612): 585-590.

Mendoza, J. A., M. B. Jarstfer, et al. (1994). "Effects of Amino Acid Replacements on the Reductive Unfolding Kinetics of Pancreatic Trypsin Inhibitor." Biochemistry **33**(5): 1143-1148.

Meng, X. F., C. Zhang, et al. (2003). "Cloning and identification of a novel cDNA coding thioredoxin-related transmembrane protein 2." Biochemical Genetics **41**(3-4): 99-106.

Meyer, B. and T. Peters (2003). "NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors." Angewandte Chemie-International Edition **42**(8): 864-890.

Mittal, J. and R. B. Best (2008). "Thermodynamics and kinetics of protein folding under confinement." PNAS **105**(51): 20233-20238.

Mittermaier, A. and L. E. Kay (2006). "Review - New tools provide new insights in NMR studies of protein dynamics." Science **312**(5771): 224-228.

Mohler, P. J., M. Y. Zhu, et al. (2007). "Identification of a novel isoform of microsomal triglyceride transfer protein." Journal of Biological Chemistry **282**(37): 26981-26988.

Morjana, N. A. and H. F. Gilbert (1991). "Effect of Protein and Peptide Inhibitors on the Activity of Protein Disulfide Isomerase." Biochemistry **30**(20): 4985-4990.

Natsume, T., H. Nakayama, et al. (2000). "Combination of biomolecular interaction analysis and mass spectrometric amino acid sequencing." Analytical Chemistry **72**(17): 4193-4198.

Nelson, J. W. and T. E. Creighton (1994). "Reactivity and Ionization of the Active Site Cysteine Residues of DsbA, a Protein Required for Disulfide Bond Formation In Vivo." Biochemistry **33**(19): 5974-5983.

Netzer, W. J. and F. U. Hartl (1997). "Recombination of protein domains facilitated by co-translational folding in eukaryotes." Nature **388**(6640): 343-349.

Neuhaus, D. and M. P. Williamson (2000). "The Effects of Exchange and Internal Motion", pp. 141-182 In The nuclear Overhauser effect in structural and conformational analysis. New York; Chichester, John Wiley.

Nguyen, V. D., K. Wallis, et al. (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**(5): 1144-1155.

Noiva, R., H. Kimura, et al. (1991). "Peptide Binding by Protein Disulfide Isomerase, a Resident Protein of the Endoplasmic Reticulum Lumen." Journal of Biological Chemistry **266**(29): 19645-19649.

Norgaard, P., V. Westphal, et al. (2001). "Functional differences in yeast protein disulfide isomerases." Journal of Cell Biology **152**(3): 553-562.

Oas, T. G. and P. S. Kim (1988). "A Peptide Model of a Protein Folding Intermediate." Nature **336**(6194): 42-48.

Otomo, T., N. Ito, et al. (1999). "NMR observation of selected segments in a larger protein: Central-segment isotope labeling through intein-mediated ligation." Biochemistry **38**(49): 16040-16044.

Otting, G., E. Liepinsh, et al. (1991a). "Protein Hydration in Aqueous Solution." Science **254**(5034): 974-980.

Otting, G., E. Liepinsh, et al. (1991b). "Proton Exchange with Internal Water Molecules in the Protein BPTI in Aqueous Solution." Journal of the American Chemical Society **113**(11): 4363-4364.

Pace, C. N. and T. E. Creighton (1986). "The Disulfide Folding Pathway of Ribonuclease T1." Journal of Molecular Biology **188**(3): 477-486.

Pan, H., E. Barbar, et al. (1995). "Extensive Nonrandom Structure in Reduced and Unfolded Bovine Pancreatic Trypsin Inhibitor." Biochemistry **34**(43): 13974-13981.

Pankalainen, M., H. Aro, et al. (1970). "Protocollagen Proline Hydroxylase - Molecular Weight, Subunits and Isoelectric Point." Biochimica et Biophysica Acta **221**(3): 559-565.

Panse, V. G., C. P. Swaminathan, et al. (2000). "Thermodynamics of substrate binding to the chaperone SecB." Biochemistry **39**(9): 2420-2427.

Parodi, A. J. (2000). "Protein glucosylation and its role in protein folding." Annual Review of Biochemistry **69**: 69-93.

Pascal, S. M. (2008). NMR primer: An HSQC-based approach with vector animations, pp. 1-9. Chichester, IM Publications.

Paulo, J. (2006). Biacore T100: An overview of underlying concepts, basic operation, experimental techniques and potential applications, Biacore.

Pelham, H. R. B. (1990). "The Retention Signal for Soluble Proteins of the Endoplasmic Reticulum." Trends in Biochemical Sciences **15**(12): 483-486.

Pihlajaniemi, T., T. Helaakoski, et al. (1987). "Molecular cloning of the beta-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene." EMBO journal **6**(3): 643-649.

Piotto, M., V. Saudek, et al. (1992). "Gradient Tailored Excitation for Single Quantum N-Spectroscopy of Aqueous Solutions." Journal of Biomolecular NMR **2**(6): 661-665.

Pirneskoski, A., P. Klappa, et al. (2004). "Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase." Journal of Biological Chemistry **279**(11): 10374-10381.

Pirneskoski, A., L. W. Ruddock, et al. (2001). "Domains b ' and a ' of protein disulfide isomerase fulfill the minimum requirement for function as a subunit of prolyl 4-hydroxylase - The N-terminal domains a and b enhance this function and can be substituted in part by those of ERp57." Journal of Biological Chemistry **276**(14): 11287-11293.

Post, C. B. (2003). "Exchange-transferred NOE spectroscopy and bound ligand structure determination." Current Opinion in Structural Biology **13**(5): 581-588.

Powis, G. and W. R. Montfort (2001). "Properties and biological activities of thioredoxins." Annual Review of Biophysics and Biomolecular Structure **30**: 421-455.

Prestegard, J. H. (2010). "Chemical Exchange Diffusion." 2010, from http://tesla.ccrc.uga.edu/courses/bionmr/lectures/pdfs/ChemicalExchangediffusion2010.pdf.

Puig, A. and H. F. Gilbert (1994). "Protein Disulfide Isomerase Exhibits Chaperone and Anti-Chaperone Activity in the Oxidative Refolding of Lysozyme." Journal of Biological Chemistry **269**(10): 7764-7771.

Quan, H., G. B. Fan, et al. (1995). "Independence of the Chaperone Activity of Protein Disulfide Isomerase from Its Thioredoxin-Like Active Site." Journal of Biological Chemistry **270**(29): 17078-17080.

Raines, R. T. (1997). "Nature's transitory covalent bond." Nature Structural Biology **4**(6): 424-427.

Rajagopal, P., E. B. Waygood, et al. (1997). "Demonstration of protein-protein interaction specificity by NMR chemical shift mapping." Protein Science **6**(12): 2624-2627.

Raykhel, I., H. Alanen, et al. (2007). "A molecular specificity code for the three mammalian KDEL receptors." Journal of Cell Biology **179**(6): 1193-1204.

Rich, R. L. and D. G. Myszka (2008). "Survey of the year 2007 commercial optical biosensor literature." Journal of Molecular Recognition **21**(6): 355-400.

Rich, R. L. and D. G. Myszka (2010). "Grading the commercial optical biosensor literature-Class of 2008: 'The Mighty Binders'." Journal of Molecular Recognition **23**(1): 1-64.

Roberts, G. C. K. (1993). "Resonance assignment strategies for small proteins". In NMR of macromolecules: a practical approach, pp. 71-99 Oxford, IRL Press.

Ruddock, L. W., T. R. Hirst, et al. (1996). "pH-dependence of the dithiol-oxidizing activity of DsbA (a periplasmic protein thiol:disulphide oxidoreductase) and protein disulphide isomerase: Studies with a novel simple peptide substrate." Biochemical Journal **315**: 1001-1005.

Russell, S. J., L. W. Ruddock, et al. (2004). "The primary substrate binding site in the b' domain of ERp57 is adapted for endoplasmic reticulum lectin association." Journal of Biological Chemistry **279**(18): 18861-18869.

Saibil, H. R. (2008). "Chaperone machines in action." Current Opinion in Structural Biology **18**(1): 35-42.

Satoh, M., A. Shimada, et al. (2005). "Differential cooperative enzymatic activities of protein disulfide isomerase family in protein folding." Cell Stress & Chaperones **10**(3): 211-220.

Schasfoort, R. B. M. and A. J. Tudos (2008). Handbook of surface plasmon resonance. Cambridge, Royal Society of Chemistry.

Scheidig, A. J., T. R. Hynes, et al. (1997). "Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): Engineering of inhibitors with altered specificities." Protein Science **6**(9): 1806-1824.

Schmid, E. L., T. A. Keller, et al. (1997). "Reversible oriented surface immobilization of functional proteins on oxide surfaces." Analytical Chemistry **69**(11): 1979-1985.

Schumann, F. H., H. Riepl, et al. (2007). "Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions." Journal of Biomolecular NMR **39**(4): 275-289.

Sedrakyan, A., T. Treasure, et al. (2004). "Effect of aprotinin on clinical outcomes in coronary artery bypass graft surgery: A systematic review and meta-analysis of randomized clinical trials." Journal of Thoracic and Cardiovascular Surgery **128**(3): 442-448.

Sevier, C. S. and C. A. Kaiser (2002). "Formation and transfer of disulphide bonds in living cells." Nature Reviews Molecular Cell Biology **3**(11): 836-847.

Sevier, C. S. and C. A. Kaiser (2008). "Ero1 and redox homeostasis in the endoplasmic reticulum." Biochimica Et Biophysica Acta-Molecular Cell Research **1783**(4): 549-556.

Sidhu, A. (2008). "Structural Studies on Protein Disulfide Isomerase." Thesis, University of Warwick.

Sitia, R. and I. Braakman (2003). "Quality control in the endoplasmic reticulum protein factory." Nature **426**(6968): 891-4.

Smith, P. E., R. C. Vanschaik, et al. (1995). "Internal Mobility of the Basic Pancreatic Trypsin Inhibitor in Solution - A Comparison of NMR Spin Relaxation Measurements and Molecular Dynamics Simulations." Journal of Molecular Biology **246**(2): 356-365.

Solda, T., N. Garbi, et al. (2006). "Consequences of ERp57 deletion on oxidative folding of obligate and facultative clients of the calnexin cycle." Journal of Biological Chemistry **281**(10): 6219-6226.

Song, J. L. and C. C. Wang (1995). "Chaperone-like activity of protein disulfide isomerase in the refolding of rhodanese." European Journal of Biochemistry **231**(2): 312-316.

Staley, J. P. and P. S. Kim (1992). "Complete Folding of Bovine Pancreatic Trypsin Inhibitor with Only a Single Disulfide Bond." PNAS **89**(5): 1519-1523.

Staley, J. P. and P. S. Kim (1994). "Formation of a Native-Like Subdomain in a Partially Folded Intermediate of Bovine Pancreatic Trypsin Inhibitor." Protein Science **3**(10): 1822-1832.

Sun, X. X., Y. Dai, et al. (2000). "Contributions of protein disulfide isomerase domains to its chaperone activity." Biochimica et Biophysica Acta **1481**(1): 45-54.

Swift, L. L., M. Y. Zhu, et al. (2003). "Subcellular localization of microsomal triglyceride transfer protein." Journal of Lipid Research **44**(10): 1841-1849.

Tai, T., K. Yamashita, et al. (1977). "Structure of the Carbohydrate Moiety of Ovalbumin Glycopeptide III and Difference in Specificity of Endo-Beta-N-Acetylglucosaminidases CII and H*." Journal of Biological Chemistry **252**(19): 6687-6694.

Tian, G., S. Xiang, et al. (2006). "The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites." Cell **124**(1): 61-73.

Tochio, N., S. Koshiba, et al. (Unpublished). "The solution structure of the second thioredoxin-like domain of human protein disulfide isomerase."

Tuderman, L., E. R. Kuutti, et al. (1975). "Affinity Column Procedure Using Poly(L-Proline) for Purification of Prolyl Hydroxylase - Purification of Enzyme from Chick-Embryos." European Journal of Biochemistry **52**(1): 9-16.

Turano, C., S. Coppari, et al. (2002). "Proteins of the PDI family: Unpredicted non-ER locations and functions." Journal of Cellular Physiology **193**(2): 154-163.

Ushioda, R., J. Hoseki, et al. (2008). "ERdj5 is required as a disulfide reductase for degradation of misfolded proteins in the ER." Science **321**(5888): 569-572.

van den Diepstraten, C., K. Papay, et al. (2003). "Cloning of a novel prolyl 4-hydroxylase subunit expressed in the fibrous cap of human atherosclerotic plaque." Circulation **108**(5): 508-511.

van Lith, M., N. Hartigan, et al. (2005). "PDILT, a divergent testis-specific protein disulfide isomerase with a non-classical SXXC motif that engages in

disulfide-dependent interactions in the endoplasmic reticulum." Journal of Biological Chemistry **280**(2): 1376-1383.

van Lith, M., A. R. Karala, et al. (2007). "A developmentally regulated chaperone complex for the endoplasmic reticulum of male haploid germ cells." Molecular Biology of the Cell **18**(8): 2795-2804.

Vanmierlo, C. P. M., N. J. Darby, et al. (1992). "The Partially Folded Conformation of the Cys-30 Cys-51 Intermediate in the Disulfide Folding Pathway of Bovine Pancreatic Trypsin Inhibitor." PNAS **89**(15): 6775-6779.

Vanmierlo, C. P. M., N. J. Darby, et al. (1993). "Partially Folded Conformation of the (30-51) Intermediate in the Disulfide Folding Pathway of Bovine Pancreatic Trypsin Inhibitor - 1H and 15N Resonance Assignments and Determination of Backbone Dynamics from 15N Relaxation Measurements." Journal of Molecular Biology **229**(4): 1125-1146.

Vanmierlo, C. P. M., N. J. Darby, et al. (1991a). "(14-38, 30-51) Double Disulfide Intermediate in Folding of Bovine Pancreatic Trypsin Inhibitor - A 2-Dimensional 1H Nuclear Magnetic Resonance Study." Journal of Molecular Biology **222**(2): 353-371.

Vanmierlo, C. P. M., N. J. Darby, et al. (1991b). "2-Dimensional 1H Nuclear Magnetic Resonance Study of the (5-55) Single Disulfide Folding Intermediate of Bovine Pancreatic Trypsin Inhibitor." Journal of Molecular Biology **222**(2): 373-390.

Vanmierlo, C. P. M., J. Kemmink, et al. (1994). "1H NMR Analysis of the Partly Folded Nonnative 2-Disulfide Intermediates (30-51,5-14) and (30-51,5-38) in the Folding Pathway of Bovine Pancreatic Trypsin Inhibitor." Journal of Molecular Biology **235**(3): 1044-1061.

Venetianer, P. and F. B. Straub (1963). "Enzymic Formation of the Disulfide Bridges of Ribonuclease." Acta Physiol Acad Sci Hung **24**: 41-53.

Vranken, W. F., W. Boucher, et al. (2005). "The CCPN data model for NMR spectroscopy: Development of a software pipeline." Proteins-Structure Function and Bioinformatics **59**(4): 687-696.

Vuori, K., T. Pihlajaniemi, et al. (1992a). "Characterization of the Human Prolyl 4-Hydroxylase Tetramer and Its Multifunctional Protein Disulfide Isomerase Subunit Synthesized in a Baculovirus Expression System." PNAS **89**(16): 7467-7470.

Vuori, K., T. Pihlajaniemi, et al. (1992b). "Site Directed Mutagenesis of Human Protein Disulfide Isomerase - Effect on the Assembly, Activity and Endoplasmic Reticulum Retention of Human Prolyl 4-Hydroxylase in Spodoptera Frugiperda Insect Cells." EMBO Journal **11**(11): 4213-4217.

Wagner, G. and K. Wuthrich (1982). "Sequential Resonance Assignments in Protein 1H Nuclear Magnetic Resonance Spectra - Basic Pancreatic Trypsin Inhibitor." Journal of Molecular Biology **155**(3): 347-366.

Wallis, A. K., A. Sidhu, et al. (2009). "The ligand-binding b' domain of human protein disulphide isomerase mediates homodimerization." Protein Science **18**(12): 2569-2577.

Wang, C. C. and C. L. Tsou (1993). "Protein Disulfide Isomerase Is Both an Enzyme and a Chaperone." FASEB Journal **7**(15): 1515-1517.

Wang, L., S. J. Li, et al. (2009). "Reconstitution of Human Ero1-L alpha/Protein Disulfide Isomerase Oxidative Folding Pathway in Vitro: Position-dependent differences in role between the a and a' domains of protein disulfide isomerase." Journal of Biological Chemistry **284**(1): 199-206.

Wang, L. K., L. Wang, et al. (2008). "Crystal structure of human ERp44 shows a dynamic functional modulation by its carboxy-terminal tail." Embo Reports **9**(7): 642-647.

Wang, S. R., W. R. Trumble, et al. (1998). "Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum." Nature Structural Biology **5**(6): 476-483.

Weichsel, A., J. R. Gasdaska, et al. (1996). "Crystal structures of reduced, oxidized, and mutated human thioredoxins: Evidence for a regulatory homodimer." Structure **4**(6): 735-751.

Weissman, J. S. and P. S. Kim (1991). "Reexamination of the Folding of BPTI - Predominance of Native Intermediates." Science **253**(5026): 1386-1393.

Weissman, J. S. and P. S. Kim (1992a). "The Pro Region of BPTI Facilitates Folding." Cell **71**(5): 841-851.

Weissman, J. S. and P. S. Kim (1992b). "The Disulfide Folding Pathway of BPTI - Response." Science **256**(5053): 112-114.

Weissman, J. S. and P. S. Kim (1993). "Efficient Catalysis of Disulfide Bond Rearrangements by Protein Disulfide Isomerase." Nature **365**(6442): 185-188.

Westphal, V., N. J. Darby, et al. (1999). "Functional properties of the two redox-active sites in yeast protein disulphide isomerase in vitro and in vivo." Journal of Molecular Biology **286**(4): 1229-1239.

Wetterau, J. R., L. P. Aggerbeck, et al. (1992). "Absence of Microsomal Triglyceride Transfer Protein in Individuals with Abetalipoproteinemia." Science **258**(5084): 999-1001.

Wetterau, J. R., L. P. Aggerbeck, et al. (1991). "Structural Properties of the Microsomal Triglyceride Transfer Protein Complex." Biochemistry **30**(18): 4406-4412.

Wetterau, J. R., K. A. Combs, et al. (1991). "Protein Disulfide Isomerase Appears Necessary to Maintain the Catalytically Active Structure of the Microsomal Triglyceride Transfer Protein." Biochemistry **30**(40): 9728-9735.

Wetterau, J. R., K. A. Combs, et al. (1990). "Protein Disulfide Isomerase Is a Component of the Microsomal Triglyceride Transfer Protein Complex." Journal of Biological Chemistry **265**(17): 9800-9807.

Wetterau, J. R., M. C. M. Lin, et al. (1997). "Microsomal triglyceride transfer protein." Biochimica et Biophysica Acta **1345**(2): 136-150.

Wetterau, J. R. and D. B. Zilversmit (1986). "Localization of Intracellular Triacylglycerol and Cholesteryl Ester Transfer Activity in Rat Tissues." Biochimica et Biophysica Acta **875**(3): 610-617.

Whitmore, L. and B. A. Wallace (2008). "Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases." Biopolymers **89**(5): 392-400.

Williams, C. and T. A. Addona (2000). "The integration of SPR biosensors with mass spectrometry: possible applications for proteome analysis." Trends in Biotechnology **18**(2): 45-48.

Williamson, R. A., M. D. Carr, et al. (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**(45): 13882-13889.

Winter, J., P. Klappa, et al. (2002). "Catalytic activity and chaperone function of human protein disulfide isomerase are required for the efficient refolding of proinsulin." Journal of Biological Chemistry **277**(1): 310-317.

Wishart, D. S., C. G. Bigam, et al. (1995). "1H, 13C and 15N Chemical Shift Referencing in Biomolecular NMR." Journal of Biomolecular NMR **6**(2): 135-140.

Woodward, C., I. Simon, et al. (1982). "Hydrogen Exchange and the Dynamic Structure of Proteins." Molecular and Cellular Biochemistry **48**(3): 135-160.

Wuthrich, K. and G. Wagner (1975). "NMR Investigations of Dynamics of Aromatic Amino Acid Residues in Basic Pancreatic Trypsin Inhibitor." FEBS Letters **50**(2): 265-268.

Wuthrich, K., G. Wagner, et al. (1980). "Correlations between Internal Mobility and Stability of Globular Proteins." Biophysical Journal **32**(1): 549-560.

Yao, Y., Y. C. Zhou, et al. (1997). "Both the isomerase and chaperone activities of protein disulfide isomerase are required for the reactivation of reduced and denatured acidic phospholipase A2." EMBO Journal **16**(3): 651-658.

Yu, H. T. (1999). "Extending the size limit of protein nuclear magnetic resonance." PNAS **96**(2): 332-334.

Yu, M. H., J. S. Weissman, et al. (1995). "Contribution of Individual Side Chains to the Stability of BPTI Examined by Alanine-Scanning Mutagenesis." Journal of Molecular Biology **249**(2): 388-397.

Zapun, A., T. E. Creighton, et al. (1992). "Folding In Vitro of Bovine Pancreatic Trypsin Inhibitor in the Presence of Proteins of the Endoplasmic Reticulum." Proteins-Structure Function and Genetics **14**(1): 10-15.

Zhang, J. X. and D. P. Goldenberg (1997). "Mutational analysis of the BPTI folding pathway. 1. Effects of aromatic->leucine substitutions on the distribution of folding intermediates." Protein Science **6**(7): 1549-1562.

Zhou, H. X., G. N. Rivas, et al. (2008). "Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences." Annual Review of Biophysics **37**: 375-397.

APPENDIX A – CHEMICAL SHIFT RESONANCES FOR WILD TYPE BPTI ASSIGNMENTS AT 36°C

Chemical shift data (in ppm) are given for both commercial and recombinant wild type BPTI at 36°C. Assignment tables are based on ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY assignments.

	Commercial BPTI				Recombinant BPTI			
Amino Acid	**NH**	**CαH**	**CβH**	**CγH**	**NH**	**CαH**	**CβH**	**CγH**
Arg1	-	-	-	-	-	-	-	-
Pro2	-	-	-	-	-	-	-	-
Asp3	-	-	-	-	8.697	4.250	2.759	-
Phe4	-	-	-	-	7.779	4.591	3.347	-
Cys5	7.477	4.348	2.859 2.742	-	7.509	4.354	2.843 2.770	-
Leu6	7.574	4.507	1.840	-	7.569	4.512	1.839 1.683	-
Glu7	7.526	4.584	2.252 2.168	-	7.528	4.586	2.257 2.165	2.569
Pro8	-	-	-	-	-	-	-	-
Pro9	-	-	-	-	-	-	-	-
Tyr10	7.808	4.933	2.954	-	7.802	4.930	2.955	-
Thr11	-	-	-	-	8.953	4.527	4.048	1.387
Gly12	7.164	3.896 3.261	-	-	7.160	3.895 3.260	-	-
Cys14	8.702	4.559	3.460 2.794	-	8.697	4.558	3.461 2.798	-
Lys15	7.967	4.407	2.090 1.577	-	7.968	4.407	2.085 1.575	-
Ala16	8.221	4.294	1.172	-	8.210	4.296	1.171	-
Arg17	8.205	4.315	1.594	-	8.190	4.312	1.594 1.476	1.303
Ile18	8.130	4.196	1.865	-	8.116	4.195	1.867	0.963
Ile19	8.694	4.307	1.946	-	8.689	4.307	1.947	-

Arg20	8.399	4.699	1.804 0.809	-	8.394	4.697	1.807 0.812	-
Tyr21	9.184	5.678	2.698	-	9.186	5.679	2.700	-
Phe22	9.778	5.269	2.906 2.818	-	9.777	5.271	2.906 2.816	-
Tyr23	10.560	4.294	3.459 2.731	-	10.570	4.297	3.463 2.738	-
Asn24	7.774	4.601	2.859 2.167	-	7.779	4.603	2.862 2.169	-
Ala25	8.794	3.761	1.554	-	8.789	3.760	1.554	-
Lys26	7.910	4.069	1.884	-	7.913	4.067	1.884	-
Ala27	6.822	4.290	1.181	-	6.817	4.292	1.184	-
Gly28	8.141	3.917 3.611	-	-	8.143	3.919 3.628	-	-
Leu29	6.828	4.743	1.728 1.427	-	6.828	4.755	1.730 1.429	-
Cys30	8.417	5.606	3.670 2.665	-	8.406	5.607	3.672 2.665	-
Gln31	8.771	4.824	2.180 1.729	-	8.763	4.824	2.223 1.729	-
Thr32	8.040	5.289	4.035	-	8.035	5.291	4.035	-
Phe33	9.373	4.862	3.084 2.951	-	9.375	4.867	3.088 2.949	-
Val34	8.386	3.915	1.949	-	8.368	3.915	1.947	-
Tyr35	9.389	4.879	2.662 2.500	-	9.388	4.874	2.664 2.501	-
Gly36	8.616	4.316 3.246	-	-	8.616	4.317 3.241	-	-
Gly37	4.334	4.225 2.907	-	-	4.307	4.219 2.903	-	-
Cys38	7.764	4.951	3.958 3.031	-	7.755	4.951	3.955 3.029	-
Arg39	-	-	-	-	9.075	3.937	2.271 1.594	-
Ala40	7.422	4.085	1.203	-	7.422	4.080	1.200	-
Lys41	8.327	4.439	2.252 1.654	-	8.331	4.439	2.251 1.657	-
Arg42	8.366	3.664	1.025 0.364	-	8.358	3.658	1.014 0.341	-
Asn43	7.214	5.055	3.335 3.265	-	7.212	5.052	3.335 3.285	-
Asn44	6.776	4.897	2.774 2.504	-	6.762	4.898	2.775 2.503	-
Phe45	9.939	5.122	3.405	-	9.938	5.125	3.406	-

			2.784				2.787	
Lys46	-	-	-	-	9.913	4.383	2.092 1.983	-
Ser47	7.468	4.533	4.124 3.861	-	7.466	4.531	4.124 3.862	-
Ala48	8.148	3.151	1.033	-	8.140	3.155	1.035	-
Glu49	8.613	3.865	2.007 1.830	-	8.603	3.858	2.005 1.834	2.334 2.184
Asp50	7.856	4.282	2.875 2.714	-	7.849	4.281	2.870 2.719	-
Cys51	6.992	1.687	3.172 2.881	-	6.985	1.687	3.175 2.892	-
Met52	8.593	4.163	2.052 1.963	-	8.581	4.170	2.052 1.963	2.692
Arg53	8.269	3.983	1.923 1.863	-	8.267	3.986	1.918 1.869	-
Thr54	7.392	4.072	4.000	-	7.398	4.078	4.005	1.615
Cys55	8.239	4.612	2.241 2.004	-	8.204	4.643	2.221 2.024	-
Gly56	7.998	3.849	-	-	7.997	3.848	-	-
Gly57	8.198	3.944 3.824	-	-	8.211	3.950 3.842	-	-
Ala58	7.923	4.023	1.301	-	7.919	4.025	1.303	-

APPENDIX B – CHEMICAL SHIFT RESONANCES FOR (30-51, 5-14) BPTI INTERMEDIATE ASSIGNMENTS AT 5°C

Chemical shift data (in ppm) are given for spectral assignments of (30-51, 5-14) BPTI partly-folded intermediate at 5°C. The assignment table is derived from a combination of 3D HSQC-TOCSY, 3D HSQC-NOESY, 2D ^{1}H-^{1}H TOCSY, 2D ^{1}H-^{1}H NOESY and ^{15}N-^{1}H HSQC spectra. Peak lists of the assignments for each spectrum are provided on the accompanying CD.

Amino Acid	N	NH	CαH	CβH	CγH	CδH	CεH
Gly12	110.8	8.602	4.051	-	-	-	-
Lys15	120.0	8.175	4.316	1.722	-	-	-
Ala16	125.2	8.520	4.271	1.387	-	-	-
Arg17a	121.1	8.459	4.386	1.694 1.475	-	-	-
Arg17b	122.7	8.545	4.288	1.794	1.648 1.570	-	-
Ile18	119.9	8.224	4.246	1.753	1.511 1.226	-	-
Ile19	126.8	8.473	4.481	1.856	1.466 0.730	-	-
Arg20	126.6	8.381	4.802	1.084	-	-	-
Tyr21	115.4	9.062	5.512	2.761 2.692	-	6.716	6.776
Phe22	119.4	9.717	5.430	3.202 3.025	-	7.180	7.204
Tyr23	123.1	9.713	4.520	3.013 2.735	-	7.097	6.571

Asn24	127.7	8.095	4.629	2.928 2.279	-	-	-
Ala25	127.0	8.580	3.707	1.578	-	-	-
Lys26	117.5	8.030	4.108	1.934	-	-	-
Ala27	118.6	7.021	4.299	1.240	-	-	-
Gly28	107.8	8.240	3.958 3.760	-	-	-	-
Leu29	115.0	6.998	4.900	1.721 1.465	-	-	-
Cys30	119.2	8.923	5.632	3.727 2.702	-	-	-
Gln31	122.7	9.029	4.888	1.998 1.738	-	-	-
Thr32	110.3	8.299	4.938	4.047	-	-	-
Phe33	118.8	8.640	4.850	3.195 3.026	-	-	-
Val34	120.7	8.507	4.069	1.941	0.830 0.875	-	-
Tyr35	129.1	8.772	4.388	2.821 2.504	-	6.940	-
Gly36	108.9	8.373	3.990 3.980	-	-	-	-
Gly37	114.7	8.315	3.878 3.842	-	-	-	-
Ser38	115.7	8.369	4.420	3.945 3.896	-	-	-
Arg39	121.0	8.362	4.434	1.785	-	-	-
Ala40	123.9	8.224	4.256	1.362	-	-	-
Lys41	124.0	8.920	4.297	1.771	-	-	-
Arg42	120.2	8.141	4.316	1.847	-	-	-
Asn43	119.8	8.660	4.623	2.795	-	-	-

Asn44	117.7	7.962	5.384	-	-	-	-
Phe45	117.1	9.316	5.108	2.766	-	7.294	-
Lys46	120.2	9.952	4.470	2.069	-	-	-
Ser47	109.4	7.506	4.553	4.134 3.874	-	-	-
Ala48	125.3	8.327	3.111	1.106	-	-	-
Glu49	117.6	8.748	3.865	2.019 1.852	-	-	-
Asp50	119.8	7.718	4.288	2.717 2.721	-	-	-
Cys51	118.5	7.110	-	2.992 2.657	-	-	-
Met52	121.3	8.666	4.120	2.045 2.051	-	-	-
Arg53	120.0	7.984	4.047	1.888	-	-	7.795
Thr54	111.9	7.568	4.188	4.049	-	-	-
Ser55	115.5	7.764	4.136	3.240 3.133	-	-	-
Gly56	108.1	7.919	3.886	-	-	-	-
Gly57	108.7	8.129	3.931 3.565	-	-	-	-
Ala58a	128.9	7.947	4.194	1.372	-	-	-
Ala58b	129.2	8.045	4.183	1.381	-	-	-

APPENDIX C – CHEMICAL SHIFT RESONANCES FOR HSQC SPECTRA AT VARIOUS TEMPERATURES

Chemical shift data (in ppm) are given for ^{15}N-^{1}H HSQC assignments of both wild type and (30-51, 5-14) BPTI partly-folded intermediate at various temperatures.

Wild Type BPTI

	5°C		36°C	
Amino Acid	**N**	**NH**	**N**	**NH**
Arg1	-	-	-	-
Pro2	-	-	-	-
Asp3	123.630	8.764	123.360	8.509
Phe4	117.510	8.060	117.015	7.780
Cys5	121.612	7.609	121.564	7.428
Leu6	113.719	7.700	114.008	7.503
Glu7	120.326	7.587	120.645	7.436
Pro8	-	-	-	-
Pro9	-	-	-	-
Tyr10	123.596	7.971	123.573	7.758
Thr11	127.620	9.203	127.581	8.924
Gly12	107.590	7.347	107.517	7.110
Pro13	-	-	-	-
Cys14	120.471	8.844	120.312	8.645
Lys15	115.610	8.145	115.778	7.922
Ala16	123.711	8.287	123.796	8.130
Arg17	118.742	8.449	118.744	8.173
Ile18	125.958	8.262	125.966	8.067
Ile19	128.352	8.927	128.452	8.662
Arg20	130.263	8.495	130.181	8.314
Tyr21	115.202	9.307	115.489	9.124
Phe22	119.902	9.863	120.056	9.699
Tyr23	124.867	10.702	125.050	10.506

Asn24	125.282	7.773	125.474	7.650
Ala25	126.191	9.043	126.478	8.769
Lys26	117.459	8.109	117.350	7.854
Ala27	118.658	6.917	118.744	6.740
Gly28	107.326	8.265	107.295	8.074
Leu29	114.922	6.938	114.957	6.754
Cys30	118.304	8.580	118.578	8.353
Gln31	122.928	8.856	123.126	8.689
Thr32	108.526	8.198	108.653	7.982
Phe33	119.153	9.461	119.310	9.297
Val34	118.848	8.590	118.956	8.330
Tyr35	129.711	9.568	129.866	9.345
Gly36	114.361	8.766	114.431	8.553
Gly37	-	-	-	-
Cys38	115.000	7.804	115.214	7.656
Arg39	113.264	9.285	113.597	9.058
Ala40	118.225	7.553	118.335	7.361
Lys41	121.204	8.464	121.293	8.268
Arg42	116.143	8.565	116.067	8.323
Asn43	116.159	7.277	116.296	7.118
Asn44	120.730	6.856	120.972	6.690
Phe45	122.714	10.054	122.787	9.865
Lys46	120.370	10.139	120.610	9.892
Ser47	108.942	7.541	109.027	7.387
Ala48	125.414	8.297	125.591	8.097
Glu49	117.583	8.721	117.800	8.533
Asp50	120.698	8.019	120.586	7.796
Cys51	119.771	7.073	119.886	6.899
Met52	120.550	8.687	120.855	8.495
Arg53	122.057	8.333	121.704	8.183
Thr54	113.427	7.493	113.506	7.325
Cys55	115.027	8.388	115.029	8.156
Gly56	108.082	8.203	108.018	7.942
Gly57	109.635	8.459	109.541	8.194
Ala58	128.702	8.026	128.895	7.754

(30-51, 5-14) BPTI Intermediate

	5°C		20°C		36°C	
Amino Acid	**N**	**NH**	**N**	**NH**	**N**	**NH**
Gly12	110.755	8.594	110.738	8.506	-	-
Lys15	119.976	8.164	119.713	8.058	119.710	7.995
Ala16	125.236	8.518	124.965	8.384	124.573	8.251
Arg17a	121.038	8.453	120.867	8.340	120.571	8.227
Arg17b	122.658	8.544	-	-	-	-
Ile18	119.872	8.217	119.819	8.104	-	-
Ile19	126.807	8.468	126.865	8.437	-	-
Arg20	126.682	8.373	126.528	8.332	-	-
Tyr21	115.357	9.057	115.659	8.997	-	-
Phe22	119.361	9.709	119.593	9.677	-	-
Tyr23	123.151	9.708	-	-	-	-
Asn24	127.715	8.089	127.754	8.081	-	-
Ala25	127.026	8.576	127.210	8.494	-	-
Lys26	117.525	8.023	117.469	7.964	-	-
Ala27	118.636	7.012	118.729	6.974	-	-
Gly28	107.703	8.232	107.669	8.194	107.626	8.152
Leu29	115.073	6.991	115.118	6.963	-	-
Cys30a	119.165	8.919	119.303	8.827	-	-
Cys30b	-	-	119.023	8.755	119.088	8.625
Gln31	122.550	9.024	122.732	8.988	-	-
Thr32	110.296	8.289	110.347	8.214	-	-
Phe33	118.844	8.641	118.816	8.569	118.781	8.480
Val34	120.659	8.506	120.734	8.408	-	-
Tyr35	129.100	8.761	-	-	-	-
Gly36	109.058	8.367	109.026	8.296	-	-
Gly37	114.758	8.312	114.355	8.256	-	-
Ser38	115.701	8.366	115.689	8.277	-	-
Arg39	121.096	8.366	121.134	8.278	121.211	8.184
Ala40	123.874	8.215	123.861	8.118	123.757	8.037
Lys41	123.878	8.922	123.669	8.806	-	-
Arg42	120.224	8.113	-	-	-	-

Asn43	119.841	8.657	119.658	8.555	-	-
Asn44	117.769	7.963	119.346	7.946	119.428	7.891
Phe45	117.061	9.293	123.415	8.745	123.424	8.658
Lys46	120.163	9.946	-	-	-	-
Ser47	109.444	7.499	-	-	-	-
Ala48	125.362	8.318	125.543	8.278	-	-
Glu49	117.595	8.743	117.778	8.698	-	-
Asp50	119.813	7.913	119.684	7.845	-	-
Cys51	118.549	7.100	118.506	7.068	-	-
Met52	121.269	8.658	121.159	8.589	-	-
Arg53	120.040	7.974	119.927	7.956	120.034	7.957
Thr54	111.834	7.559	112.119	7.521	111.908	7.442
Ser55	115.480	7.755	115.676	7.722	-	-
Gly56	108.128	7.915	108.222	7.875	-	-
Gly57	108.633	8.123	108.742	8.064	-	-
Ala58a	128.862	7.941	128.994	7.856	129.151	7.785
Ala58b	129.062	8.038	129.157	7.937	129.285	7.832

APPENDIX D – HSQC ASSIGNED PEAK HEIGHTS FROM HYDROGEN DEUTERIUM EXCHANGE

Summary of peaks assigned in ^{15}N-^{1}H HSQC spectra of wild type BPTI and (30-51, 5-14) BPTI intermediate in NMR buffer using solvents of 90% H_2O, 100% D_2O after 5 min and 100% D_2O buffer after 24 hours. Data show the height of each assigned peak, after normalisation to account for the signal-to-noise ratio of each spectrum.

			Height of Assigned Peaks (Normalised)					
			Wild Type			(30-51,*5-14*) Intermediate		
Amino Acid	**Secondary Structure**	**Disulfides**	**90% H_2O Buffer**	**D_2O After 5 min**	**D_2O After 24 hours**	**90% H_2O Buffer**	**D_2O After 5 min**	**D_2O After 24 hours**
Arg1	coil							
Pro2	coil							
Asp3	coil		16.683					
Phe4	3_{10} helix		19.093	0.586				
Cys5	3_{10} helix	5-55	10.837	6.804				
Leu6	3_{10} helix		10.779	9.456	0.966			
Glu7	coil		11.808	7.822	7.063			
Pro8	coil							
Pro9	coil							
Tyr10	coil		4.159	2.907				
Thr11	coil		8.739					
Gly12	coil		8.030			2.029	0.297	
Pro13	coil							
Cys14	coil	14-38	3.436	2.899				
Lys15	coil		5.117			5.039		

Ala16	coil		11.776	8.020		2.872	0.499	
Arg17	coil		12.418			11.811	0.553	0.089
Ile18	β-strand		7.847	7.053	5.041	11.482	0.263	0.594
Ile19	β-strand		11.448	0.419		3.052		
Arg20	β-strand		7.169	5.493	4.721	5.156	-0.042	
Tyr21	β-strand		5.721	3.343	2.756	2.611		
Phe22	β-strand		6.262	4.085	4.167	3.676	0.008	0.163
Tyr23	β-strand		5.333	3.291	2.832	3.712		
Asn24	β-strand		6.904	4.674	4.257	3.896		
Ala25	β-turn		11.551			4.936		
Lys26	β-turn		11.687			9.280		
Ala27	β-turn		14.251	0.161		9.535	0.776	0.641
Gly28	β-turn		11.264	5.701		10.654	0.607	-0.070
Leu29	β-strand		17.494	13.417	13.762	9.104	0.181	
Cys30	β-strand	30-51	9.574			9.388		
Gln31	β-strand		6.580	4.101	4.201	4.152		
Thr32	β-strand		13.012	1.694		8.638		
Phe33	β-strand		9.338	5.727	6.499	2.483		
Val34	β-strand		8.930	1.270		4.413		
Tyr35	β-strand		6.286	3.411	4.021	2.686	0.079	
Gly36	coil		5.865	3.889	3.660	6.424	0.855	0.456
Gly37	coil					2.998		
Cys38	coil	14-38	7.807			7.066		
Arg39	coil		7.693			4.367	0.382	
Ala40	coil		15.024			21.096	1.096	0.930
Lys41	coil		4.317	4.576	0.754	2.342		

Arg42	coil		6.373			7.275	0.544	0.262
Asn43	coil		7.441	0.133		4.272	0.175	
Asn44	coil		7.866	5.141	5.325	3.309		
Phe45	β-strand		4.877	2.471	3.025	0.776		
Lys46	coil		9.938			2.780		
Ser47	coil		10.382			4.622	0.104	
Ala48	α-helix		14.787	0.750		7.990	0.097	0.078
Glu49	α-helix		17.460			9.939	0.381	
Asp50	α-helix		11.536			8.479		
Cys51	α-helix	30-51	9.164	6.326	4.474	3.226	0.137	-0.013
Met52	α-helix		9.138	6.297	2.629	5.840		
Arg53	α-helix		13.612	8.373	3.031	10.894	0.344	
Thr54	α-helix		9.907	6.816		5.262	0.430	
Cys55	α-helix	5-55	7.270	4.175	2.937	8.333		
Gly56	coil		6.933	4.738		10.693	0.607	0.083
Gly57	coil		15.140			21.087	1.569	0.683
Ala58	coil		28.425			59.920	5.323	1.248
Total Peaks Identified			52	34	20	45	24	13
% of Total Amino Acids			89.7	58.6	34.5	77.6	41.4	22.4
% of Amino Acids in H_2O Buffer			100.0	65.4	38.5	100.0	53.3	28.9

APPENDIX E – CHEMICAL SHIFT RESONANCES FOR HSQC SPECTRA OF BPTI CONSTRUCTS WITH PDI TITRATIONS

Chemical shift data (in ppm) are given for ^{15}N-^{1}H HSQC assignments of both wild type and (30-51, 5-14) BPTI partly-folded intermediate with various molar equivalent concentrations of PDI.

Wild Type BPTI:PDI

1:1 Ratio

Amino Acid	N	NH
Asp3	123.725	8.758
Gly57	109.616	8.464
Ala58	128.692	8.027

(30-51, 5-14) BPTI:PDI

	50:1 Ratio		10:1 Ratio	
Amino Acid	N	NH	N	NH
Gly12	110.754	8.590	110.796	8.611
Lys15	120.089	8.163		
Ala16	125.259	8.520		
Arg17	121.049	8.457	121.187	8.444
Ile18	119.878	8.223	119.874	8.228
Ile19	126.875	8.465		
Arg20	126.710	8.375		
Tyr21	115.346	9.067		
Phe22	119.387	9.711		
Tyr23	123.170	9.707		
Asn24	127.800	8.094		
Ala25	127.044	8.579	126.994	8.578
Lys26	117.455	8.020	117.478	8.025
Ala27	118.661	7.010	118.782	7.017

Gly28	107.651	8.231	107.661	8.242
Leu29	115.041	6.986		
Cys30	119.110	8.921		
Gln31	122.562	9.028		
Thr32	110.282	8.295		
Phe33	118.840	8.648		
Val34	120.644	8.503	121.109	8.489
Gly36	109.042	8.368	109.072	8.374
Gly37	114.741	8.320	114.529	8.367
Ser38	115.638	8.362	115.645	8.368
Arg39	121.144	8.361	121.073	8.370
Ala40	123.864	8.215	123.767	8.223
Lys41	123.874	8.922		
Arg42	120.181	8.120	120.182	8.131
Asn43	119.828	8.665	119.877	8.670
Asn44	119.244	7.988		
Phe45	123.409	8.820	123.436	8.822
Lys46	120.151	9.940		
Ser47	109.454	7.498		
Ala48	125.333	8.320		
Glu49	117.576	8.747	117.626	8.751
Asp50	119.789	7.915		
Cys51	118.637	7.097		
Met52	121.233	8.657	121.172	8.646
Arg53	120.048	7.965	120.035	7.980
Thr54	111.829	7.554	111.847	7.562
Ser55	115.479	7.761	115.559	7.759
Gly56	108.098	7.917	108.125	7.914
Gly57	108.630	8.124	108.644	8.127
Ala58a	128.852	7.940	128.856	7.941
Ala58b	129.085	8.040	129.079	8.043

APPENDIX F – HSQC ASSIGNED PEAK HEIGHTS FROM BPTI WITH PDI TITRATIONS

Summary of peaks assigned in ^{15}N-^{1}H HSQC spectra of wild type BPTI and (30-51, 5-14) BPTI intermediate with and without the addition of PDI. Data show the height of each assigned peak, after normalisation to account for the signal-to-noise ratio of each spectrum.

			Height of Assigned Peaks (Normalised)			
			Wild Type		(30-51,*5-14*) Intermediate	
Amino Acid	**Secondary Structure**	**Disulfides**	**Without PDI**	**BPTI:PDI 10:1**	**Without PDI**	**BPTI:PDI 10:1**
Arg1	coil					
Pro2	coil					
Asp3	coil		18.770	9.432		
Phe4	3_{10} helix		21.362	12.243		
Cys5	3_{10} helix	5-55	12.874	6.177		
Leu6	3_{10} helix		12.824	7.910		
Glu7	coil		12.417	6.874		
Pro8	coil					
Pro9	coil					
Tyr10	coil		5.000	1.017		
Thr11	coil		10.217	4.922		
Gly12	coil		9.548	4.740	3.206	1.463
Pro13	coil					
Cys14	coil	14-38	4.056	1.747		
Lys15	coil		6.117	1.192	8.033	
Ala16	coil		13.774	6.095	4.600	
Arg17	coil		14.484	8.097	5.998	0.759
Ile18	β-strand		9.331	4.621	18.074	1.371
Ile19	β-strand		12.973	6.080	4.028	
Arg20	β-strand		8.434	3.330	7.608	
Tyr21	β-strand		6.919	2.883	3.850	
Phe22	β-strand		6.873	3.857	6.718	
Tyr23	β-strand		6.026	2.843	5.268	
Asn24	β-strand		8.293	4.278	6.555	

Ala25	β-turn		12.226	6.354	6.929	0.933
Lys26	β-turn		13.468	7.618	9.793	1.512
Ala27	β-turn		16.550	9.214	13.737	0.499
Gly28	β-turn		12.654	7.610	16.365	1.082
Leu29	β-strand		19.471	11.117	13.253	
Cys30	β-strand	30-51	11.336	7.118	13.200	
Gln31	β-strand		7.202	4.026	6.454	
Thr32	β-strand		15.114	8.236	11.661	
Phe33	β-strand		10.362	5.060	4.959	
Val34	β-strand		10.878	5.134	6.842	1.102
Tyr35	β-strand		7.264	2.939	5.217	
Gly36	coil		7.003	3.154	7.686	3.877
Gly37	coil				5.623	0.813
Cys38	coil	14-38	8.246	2.856	11.791	0.968
Arg39	coil		8.541	3.580	7.249	0.979
Ala40	coil		17.537	7.206	30.115	1.952
Lys41	coil		5.082	2.039	3.461	
Arg42	coil		8.135	2.945	12.428	0.769
Asn43	coil		8.693	4.398	8.396	0.672
Asn44	coil		9.292	5.370	5.228	
Phe45	β-strand		5.449	2.795	5.237	0.632
Lys46	coil		11.491	5.329	4.708	
Ser47	coil		11.664	7.192	5.439	
Ala48	α-helix		16.993	8.084	12.505	
Glu49	α-helix		19.847	10.579	14.826	0.969
Asp50	α-helix		13.676	6.572	13.072	
Cys51	α-helix	30-51	10.920	6.073	6.249	
Met52	α-helix		10.719	5.558	8.142	0.744
Arg53	α-helix		15.673	8.255	14.210	0.523
Thr54	α-helix		11.057	6.566	8.669	0.432
Cys55	α-helix	5-55	8.422	4.485	11.531	1.241
Gly56	coil		7.885	4.212	15.558	1.860
Gly57	coil		16.643	11.380	33.376	10.420
Ala58	coil		31.721	25.726	70.255	31.454
Total Peaks Identified:			52	52	45	24
% of Total Amino Acids:			89.7	89.7	77.6	41.4
% of Amino Acids in Sample without PDI:			100.0	100.0	100.0	53.3

APPENDIX G – SURFACE PLASMON RESONANCE CURVE FITTING FOR PDI/BPTI INTERACTIONS

Analysis data from surface plasmon resonance of various BPTI constructs binding to immobilised proteins after curve fitting using 1:1 (Langmuir) binding model. Kinetic data was fitted simultaneously to association and dissociation phases of the sensorgram data using BIAevaluation analysis software (Biacore).

Reduced BPTI at 5°C

		Immobilised Protein		
		FL PDI	**bb'x**	**Ovalbumin**
	Replicate			
k_a (1/Ms)	1st	1830	1870	754
	2nd	1830	1910	715
	Average	1830	1890	734.5
k_d (1/s)	1st	5.94E-03	5.95E-03	4.58E-03
	2nd	6.01E-03	6.16E-03	4.77E-03
	Average	5.98E-03	6.06E-03	4.68E-03
K_D (M)	1st	3.24E-06	3.17E-06	6.07E-06
	2nd	3.29E-06	3.22E-06	6.67E-06
	Average	3.27E-06	3.20E-06	6.37E-06
	Standard Deviation	3.54E-08	3.54E-08	4.24E-07
	Standard Error	2.50E-08	2.50E-08	3.00E-07
Chi^2	1st	15300	3670	10300
	2nd	11800	2350	9280
	Average	13550	3010	9790
R_{max}	1st	2520	1240	2760
	2nd	2370	1160	2830

(30-51, 5-14) BPTI at 5°C

		Immobilised Protein		
		FL PDI	**bb'x**	**Ovalbumin**
	Replicate			
k_a (1/Ms)	1st	336	492	126
	2nd	354	431	305
	Average	345	461.5	215.5
k_d (1/s)	1st	4.57E-03	6.40E-03	7.03E-03
	2nd	5.11E-03	6.89E-03	6.95E-03
	Average	4.84E-03	6.65E-03	6.99E-03
K_D (M)	1st	1.36E-05	1.30E-05	5.57E-05
	2nd	1.44E-05	1.60E-05	2.28E-05
	Average	1.40E-05	1.45E-05	3.93E-05
	Standard Deviation	5.66E-07	2.12E-06	2.33E-05
	Standard Error	4.00E-07	1.50E-06	1.65E-05
Chi^2	1st	10700	2250	3130
	2nd	6030	1320	1300
	Average	8365	1785	2215
R_{max}	1st	3910	1420	4920
	2nd	3620	1540	2100

Wild Type BPTI at 5°C

		Immobilised Protein		
		FL PDI	bb'x	Ovalbumin
	Replicate			
k_a (1/Ms)	1st	21	88.4	42.8
	2nd	6.09	23.6	8.87
	Average	13.545	56	25.835
k_d (1/s)	1st	8.29E-03	8.95E-03	9.15E-03
	2nd	8.97E-03	9.91E-03	1.02E-02
	Average	8.63E-03	9.43E-03	9.68E-03
K_D (M)	1st	3.94E-04	1.01E-04	2.14E-04
	2nd	1.47E-03	4.20E-04	1.15E-03
	Average	9.32E-04	2.61E-04	6.82E-04
	Standard Deviation	7.61E-04	2.26E-04	6.62E-04
	Standard Error	5.38E-04	1.60E-04	4.68E-04
Chi^2	1st	5660	1760	1650
	2nd	3210	939	813
	Average	4435	1349.5	1231.5
R_{max}	1st	27700	3850	8140
	2nd	90800	13800	37700

Reduced BPTI at 25°C

		Immobilised Protein		
		FL PDI	bb'x	Ovalbumin
	Replicate			
k_a (1/Ms)	1st	1630	2090	1260
	2nd	1690	2210	1260
	Average	1660	2150	1260
k_d (1/s)	1st	4.89E-03	3.98E-03	2.39E-03
	2nd	5.07E-03	4.20E-03	2.43E-03
	Average	4.98E-03	4.09E-03	2.41E-03
K_D (M)	1st	3.00E-06	1.90E-06	1.90E-06
	2nd	3.00E-06	1.90E-06	1.93E-06
	Average	3.00E-06	1.90E-06	1.92E-06
	Standard Deviation	0.00E+00	0.00E+00	2.12E-08
	Standard Error	0.00E+00	0.00E+00	1.50E-08
Chi^2	1st	13700	2260	14900
	2nd	11800	1670	13600
	Average	12750	1965	14250
R_{max}	1st	2590	1170	2220
	2nd	2430	1070	2180

(30-51, 5-14) BPTI at 25°C

		Immobilised Protein		
		FL PDI	bb'x	Ovalbumin
	Replicate			
k_a (1/Ms)	1st	615	692	7.36
	2nd	757	745	22.9
	Average	686	718.5	15.13
k_d (1/s)	1st	2.87E-03	4.69E-03	5.74E-03
	2nd	3.08E-03	4.71E-03	5.86E-03
	Average	2.98E-03	4.70E-03	5.80E-03
K_D (M)	1st	4.67E-06	6.78E-06	7.79E-04
	2nd	4.08E-06	6.33E-06	2.57E-04
	Average	4.38E-06	6.56E-06	5.18E-04
	Standard Deviation	4.17E-07	3.18E-07	3.69E-04
	Standard Error	2.95E-07	2.25E-07	2.61E-04
Chi^2	1st	14400	2160	5840
	2nd	7270	1380	4620
	Average	10835	1770	5230
R_{max}	1st	3120	1070	71700
	2nd	2340	940	22100

Wild Type BPTI at 25°C

		Immobilised Protein		
		FL PDI	bb'x	Ovalbumin
	Replicate			
k_a (1/Ms)	1st	1.13	94.4	9.66
	2nd	36.7	142	2.05
	Average	18.915	118.2	5.855
k_d (1/s)	1st	6.40E-03	8.48E-03	7.66E-03
	2nd	6.17E-03	8.44E-03	8.29E-03
	Average	6.29E-03	8.46E-03	7.98E-03
K_D (M)	1st	5.66E-03	8.98E-05	7.93E-04
	2nd	1.68E-04	5.96E-05	4.05E-03
	Average	2.91E-03	7.47E-05	2.42E-03
	Standard Deviation	3.88E-03	2.14E-05	2.30E-03
	Standard Error	2.75E-03	1.51E-05	1.63E-03
Chi^2	1st	4050	990	2930
	2nd	3460	861	2630
	Average	3755	925.5	2780
R_{max}	1st	358000	2690	37600
	2nd	10400	1740	184000

Reduced BPTI at 36°C

		Immobilised Protein		
		FL PDI	bb'x	Ovalbumin
	Replicate			
k_a (1/Ms)	1st	2150	1790	2110
	2nd	2040	1750	1840
	Average	2095	1770	1975
k_d (1/s)	1st	3.27E-03	3.35E-03	1.99E-03
	2nd	3.81E-03	3.87E-03	2.59E-03
	Average	3.54E-03	3.61E-03	2.29E-03
K_D (M)	1st	1.52E-06	1.88E-06	9.47E-07
	2nd	1.87E-06	2.21E-06	1.41E-06
	Average	1.70E-06	2.05E-06	1.18E-06
	Standard Deviation	2.47E-07	2.33E-07	3.27E-07
	Standard Error	1.75E-07	1.65E-07	2.32E-07
Chi^2	1st	8950	2250	10900
	2nd	7580	1880	7200
	Average	8265	2065	9050
R_{max}	1st	1700	955	1180
	2nd	1830	1010	1320

(30-51, 5-14) BPTI at 36°C

		Immobilised Protein		
		FL PDI	bb'x	Ovalbumin
	Replicate			
k_a (1/Ms)	1st	420	438	1.82
	2nd	590	564	3.43
	Average	505	501	2.625
k_d (1/s)	1st	2.40E-03	2.79E-03	3.44E-03
	2nd	2.40E-03	2.83E-03	3.70E-03
	Average	2.40E-03	2.81E-03	3.57E-03
K_D (M)	1st	5.72E-06	6.37E-06	1.89E-03
	2nd	4.07E-06	5.02E-06	1.08E-03
	Average	4.90E-06	5.70E-06	1.49E-03
	Standard Deviation	1.17E-06	9.55E-07	5.73E-04
	Standard Error	8.25E-07	6.75E-07	4.05E-04
Chi^2	1st	16500	4690	8740
	2nd	10000	2970	2260
	Average	13250	3830	5500
R_{max}	1st	4400	2140	262000
	2nd	3130	1630	116000

Wild Type BPTI at 36°C

		Immobilised Protein		
		FL PDI	bb'x	Ovalbumin
	Replicate			
k_a (1/Ms)	1st	216	371	2.6
	2nd	357	441	1.77
	Average	286.5	406	2.185
k_d (1/s)	1st	4.06E-03	4.12E-03	5.27E-03
	2nd	4.13E-03	4.62E-03	5.39E-03
	Average	4.10E-03	4.37E-03	5.33E-03
K_D (M)	1st	1.88E-05	1.11E-05	2.02E-03
	2nd	1.16E-05	1.05E-05	3.05E-03
	Average	1.52E-05	1.08E-05	2.54E-03
	Standard Deviation	5.09E-06	4.24E-07	7.28E-04
	Standard Error	3.60E-06	3.00E-07	5.15E-04
Chi^2	1st	16500	3690	4900
	2nd	7800	2180	3980
	Average	12150	2935	4440
R_{max}	1st	4890	1770	146000
	2nd	3050	1470	222000