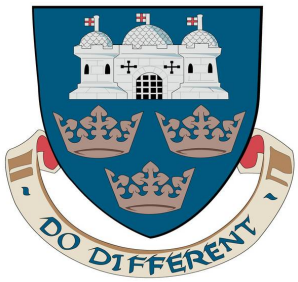
Determining the molecular basis for interactions between WW domains of WWP2 and the target regions OCT4 and Smad7



Danielle de Bourcier

A thesis submitted for the degree of

Doctor of Philosophy

School of Chemistry

University of East Anglia

September 2018

Declaration

I declare that the work contained in this thesis, submitted by me for the degree of Doctor of Philosophy, is my own original work, except where due reference has been made to other authors or co-workers and has not been previously submitted by me for a degree at this or any other university.

Danielle de Bourcier
September 2018

Abstract

WWP2, a NEDD4 E3 ubiquitin ligase, contains four WW domains which display a preference for binding protein substrates with proline-rich motifs. Through these WW domains, WWP2 interacts with different substrates such as PTEN (a tumour suppressor gene often lost in cancers), Smads (involved in the TGFβ signalling pathway, often deregulated in various cancers), and the stem-cell specific transcription factor OCT4 (involved in cancer cell proliferations, dedifferentiation, and chemotherapy resistance). Ubiquitination of the TGFβ receptors is mediated by WWP2-Smad7 complexes, marking them for degradation in the proteasome. WWP2-OCT4 complexes result in degradation of OCT4 in the proteasome. Downregulation of OCT4 occurs during differentiation; however, some malignant cells regain the ability to express OCT4. Overexpression of WWP2 in animal models causes increased tumour spread, and human cancer tissue samples present altered isoform expression patterns for WWP2.

This thesis investigates WW domain structural characteristics and the interactions between these domains and two of their targets, Smad7 and OCT4. Using Nuclear Magnetic Resonance (NMR) spectroscopy, we show that the second WW domain (WW2 – expressed as a fusion protein with solubility tag GB1) is in an intermediate exchange regime when isolated, indicating that it is not fully folded. By observing ligand interaction using NMR, peptides derived from Smad7 and OCT4 are found to bind to the WW2 domain and stabilise its structure. The two peptides have a similar, but not identical, effect on the structure and reveal the respective binding sites of WW2. In addition, NMR spectroscopy has provided structural information of the isolated GB1-WW3 domain. Isothermal Titration Calorimetry (ITC) confirmed interactions between WW2 and its targets and indicates that Smad7 binds with a higher affinity than OCT4.

Contents

List of Figures

List of Tables

Abbreviations

aa	amino acid
Abs, A	Absorbance
AMP	Adenosine Monophosphate
APS	Ammonium persulphate
ARIA	Ambiguous Restraints for Iterative Assignment
ATP	Adenosine triphosphate
BMP	Bone Morphogenetic Protein
BMRB	Biological Magnetic Resonance Bank
CASP	Critical Assessment of Structure Prediction
CCPN	Collaborative Computing Project for NMR
Co-Smad	Common mediator Smads
CSP	Chemical Shift Perturbation
CTD	C-terminal Transactivation Domain
CV	Column Volume
D_2O	Deuteriumoxide
DANGLE	Dihedral ANgles from Global Likelihood Estimates
DNA	Deoxyribonucleic Acid
DTT	Dithioltheritol
DUB	Deubiquitinating Enzymes
ECC	Embryonal Carcinoma Cells
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial-Mesenchymal Transition
ESC	Embryonic Stem Cells
FID	Free Induction Decay
FT	Fourier Transform

GB1	B1 domain of streptococcal protein G
GDNF	Glial cell line-Derived Neurotrophic Factor
GST	Glutathione-S-transferase
HECT	Homologous to the E6-associated protein Carboxyl Terminus
HF	High Fidelity
HSQC	Heteronuclear Single Quantum Coherence
ICM	Inner Cell Mass
IMAC	Immobilised Metal Ion Affinity Chromatography
iPSC	Induced Pluripotent Stem Cells
IPTG	Isopropyl β-D-1-thiogalactopyranoside
I-Smad	Inhibitory Smad
ISPA	Isolated Spin Pair Approximation
ITC	Isothermal Titration Calorimetry
$\mathbf{K_D}$	Dissociation constant
LB	Lysogeny Broth
Mad	Mothers against dpp
MEM	Minimal Essential Media
MH	Mad-homology
MWCO	Molecular Weight Cut-Off
mWWP2	Murine WWP2
NEDD4	Neural precursor cell Expressed Developmentally Downregulated protein 4
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
$\mathbf{NTD^A}$	N-terminal N-Transactivation Domain
OCT4	Octamer-binding Transcription factor 4
O.D.	Optical Density
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
POU	Pit-Oct-Unc
$\mathbf{POU^H}$	Pit-Oct-Unc homeodomain

POU^{S}	Pit-Oct-Unc specific
ppm	Parts per million
PPxY	Proline-Proline-x-Tyrosine, where x = any amino acid
PTEN	Phosphatase and Tensin homolog
PTM	Post Translational Modification
RF	Radio Frequency
RING	Really Interesting New Gene
RNA	Ribonucleic acid
rpm	Revolutions per minute
RCSB	Research Collaboratory for Structural Bioinformatics
R-Smad	Receptor Smad
S/N	Signal-to-noise ratio
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SUMO	Small Ubiquitin-like Modifier
TβRI	TGFβ Receptor type I
TβRII	TGFβ Receptor type II
TE	Trophectoderm
TEMED	Tetramethylethylenediamine
TGFβ	Transforming Growth Factor β
Tris	Tris-(hydroxymethyl)-aminomethane
TROSY	Transverse Relaxation Optimised Spectroscopy
Ub	Ubiquitin
UBC	Ubiquitin-Conjugating
UBL	Ubiquitin-Like
UEA	University of East Anglia
UEV	Ubiquitin E2 Variant
ULP1	Ubiquitin-Like Protease-1
UV	Ultraviolet
WW	Tryptophan-Tryptophan
WWP2	WW domain containing Protein 2

Mum, thank you for supporting me in every way, always. You are a brave, strong and inspiring woman and beat cancer whilst I was completing this PhD. This thesis is dedicated to you.

Acknowledgements

Firstly, I would like to thank my supervisory team, Dr. Tharin Blumenschein, Dr. Andrew Chantry and Prof. Nick Le Brun, for their ongoing guidance, advice and support, which has been invaluable throughout my PhD. This was an amazing opportunity which has helped me grow not just as a researcher but also as a person. I would also like to express my gratitude to the Big C, Norfolk's cancer charity, for providing the necessary funding for this PhD.

Thank you to the fellow scientists who have offered support over the years in the form of coffee breaks, nights out or late night whatsapp chats. You are too many to name but special thanks go to Sarah, Jess, Anna and Jenny. Special thanks also go to Dr. Derek Warren for helping me during the darkest hour of thesis printing.

Massive thanks go to Sam and Dave. You've been the best Bigtime Super Funclub members I could have ever hoped for and are the reasons I kept my sanity. An additional thanks must also go to Dave for his unwavering support when I broke LaTex (repeatedly). There are so many others I would like to thank for their friendships which have made my time at UEA so wonderful: Ian, Rupert, Craig and Brodie to name just a few, you guys are awesome. And to the four-legged friends I met on BorrowMyDoggy; Charlie, Lucy, Ozzy and Basil, thanks for the walks, cuddles and for helping to clear my mind.

For Mum and Dad, I am forever grateful. Your love and support has been more helpful than you can imagine. Dad, you have always inspired me to ask questions which led me to want to know all the answers; Mum, you have always believed in me and I appreciate that you were on the other end of the phone whenever I needed a chat (sometimes several times a day). I am also grateful for my brother

Sean, sister Rochelle and grandparents Joan, June and Don, for always having my back!

Finally, I would like to thank my wonderful fiancé William Jakes. He has supported me through both my undergraduate and postgraduate degrees. He has stood by me and consoled me through all of the hard times and the stress, he has shared in all my triumphs and he has listened to me talk/rant for hours about NMR and stubborn proteins. I would be lost without him.

Chapter 1

Introduction

Cancer is not just one disease but many diseases. There are over 200 types of cancer and it was the cause of more than one in four of all deaths in the UK in 2016 [Cancer Research UK, 2016]. It affects people of all ages and one in two people in the UK will face a cancer diagnosis at some point in their lifetime [Cancer Research UK, 2016]. An individual's risk of developing cancer depends on many factors, including age, lifestyle and genetic make-up. Cancer occurs when changes to oncogenes and tumour suppressor genes lead to abnormal and uncontrollable cell division and proliferation. Normal cells have limited life spans and their growth and differentiation must be strictly regulated to avoid the occurrence of unsuitable types and quantities of cells. The cell-division cycle is controlled by genes which either promote or suppress cell proliferation, known as proto-oncogenes and tumour suppressor genes respectively [Lee and Muller, 2010].

The proliferation of cancer cells proceeds in an unregulated manner to form clones of cells which can expand irregularly. This unrestrained growth is known as a tumour or neoplasm. Cancer cells are unlike other cells in the sense that they do not remain confined to one area and the spread of cancer to another part of the body is known as metastasis [Reynolds and Schecker, 1995]. Tumour metastasis is the primary cause of mortality in cancer patients and remains the key challenge for cancer therapy homeostasis.

The ubiquitin system has implications in cancer-relevant processes and one of three enzymes involved in this pathway, namely the E3 ubiquitin ligases, have been directly linked to the development of cancer. This is due to the ability of E3 ligases to trigger the degradation of oncogenes or tumour suppressors.

This introduction outlines the ubiquitin system and the processes of protein degradation at the proteasome. Descriptions of the E3 ubiquitin ligases in terms of the two main classes, RING family E3s and HECT family E3s, are followed by a brief introduction to the HECT E3 ligase sub-group known as the NEDD4 (Neural precursor cell Expressed Developmentally Downregulated protein 4) sub-family and its group member WWP2 (WW domain containing protein 2). Two potential target proteins, Smad7 and OCT4, are discussed in terms of the TGFβ pathway and stem cell pluripotency respectively, and the aims of the thesis are outlined.

1.1 Ubiquitination

1.1.1 Protein degradation

Protein levels within the cell are finely tuned and are determined not only by rates of synthesis but also by rates of degradation. Proteins experience different half-lives depending on their role and cell regulation depends on specific proteins having differential rates of degradation [Lecker *et al.*, 2006]. Protein degradation is an important mechanism for cell-cycle control with the ability to affect enzyme levels in eukaryotes and was first reported by Schimke in the 1960s and 1970s [Schimke, 1964]; [Schimke and Doyle, 1970].

Certain proteins within the cell have critical roles and so it is essential that the protein degradation mechanism is highly selective whilst also having the ability to balance the regulatory system requirements. The ubiquitin-proteasome pathway is a highly specific pathway which targets eukaryotic proteins or cellular molecules which are tagged with ubiquitin molecules for degradation at the

proteasome [Hershko and Ciechanover, 1998]; [Rock *et al.*, 1994]. Proteasomes are multi-protein complexes found in the cytosol and the nucleus [Coux *et al.*, 1996], and are only involved in the degradation of specific proteins which have been modified in such a way that they are susceptible to degradation. The covalent attachment of polyubiquitin chains is one such modification, which labels the protein post-translationally for degradation in a proteasomal manner, a concept first proposed in 1980 [Hershko *et al.*, 1980]. The ubiquitin pathway is also reversible and there are ∼100 deubiquitinating enzymes (DUBs) in the human genome, with the ability to remove ubiquitin from substrate proteins; hence, preventing degradation via this pathway [Grabbe *et al.*, 2011]. Therefore, for successful degradation of a protein, it requires labelling with ubiquitin and must also evade the DUBs which remove ubiquitin [Huang and Dixit, 2016].

1.1.2 The ubiquitin system

The ubiquitin-proteasome pathway has been implicated in a number of human diseases such as cancer, metabolic syndromes, neurodegenerative diseases, autoimmunity, inflammatory disorders, infection and muscle dystrophies [Adams, 2002] but also highlights a promising therapeutic target [Popovic *et al.*, 2014].

Ubiquitin is a small (76-amino acid) protein of 8.5 kDa [Schlesinger *et al.*, 1975], found to be highly conserved in almost all eukaryotic cells and is involved in the selective and irreversible deactivation of proteins [Ciechanover *et al.*, 1985]. This process, known as ubiquitination, is a post-translational modification (PTM) and functions by targeting proteins for degradation via covalent ligation to ubiquitin. The polyubiquitin-substrate complex is produced by one of two mechanisms; conjugation of a polyubiquitin chain to a substrate, or by the ligation of a single ubiquitin molecule to a substrate followed by chain extension [Nguyen *et al.*, 2014]. Proteolysis is crucial for a vast number of cellular processes and so a deeper understanding of the ubiquitin-mediated system is of great significance. The ligation of ubiquitin to substrate proteins involves the sequential action of three enzymes; ubiquitin-activating enzyme E1,

ubiquitin-conjugating enzyme E2 and ubiquitin-protein ligase E3 [Hershko *et al.*, 2000] and is illustrated in figure 1.1.

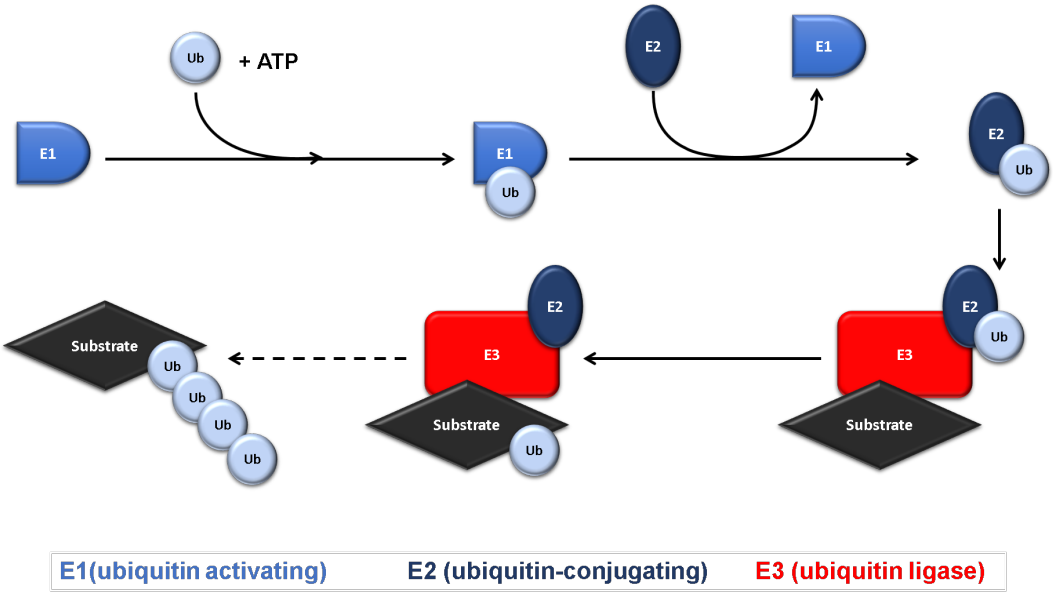


Figure 1.1: The ubiquitin pathway involved in the regulation of protein degradation and function, adapted from [Huang and Dixit, 2016]. The ubiquitin protein is first attached to ubiquitin-activating enzyme E1 in the presence of ATP. The activated ubiquitin is then passed to the E2 ubiquitin-conjugating enzyme. The E3 ubiquitin ligase recognizes a protein substrate and recruits the E2-ubiquitin complex whilst catalysing the ubiquitin transfer from the E2 to the substrate. Repeating this cycle many times leads to poly-ubiquitination of the substrate which in turn is recognised and degraded by the 26S proteasome.

Free ubiquitin molecules produced by ribosomal protein or deubiquitylating enzymes (DUBs) are activated by energy-dependent ubiquitin-activating enzyme E1 during the initial step. E1 uses adenosine triphosphate (ATP) as a source of energy to produce a thioester bond between its active site cysteine residues and the carboxyl-terminal glycine residue of ubiquitin, explained in further detail below. The activated ubiquitin is then passed from the E1 to an active site cysteine of the ubiquitin-conjugating enzyme E2. A ubiquitin protein ligase E3 then catalyses the final step by recruiting the E2-ubiquitin complex, and binding a substrate protein which has a recognisable subunit before labelling the substrate with ubiquitin [Hershko, 1983]; [Hershko *et al.*, 2000].

The fate of the substrates depends on where and how (figure 1.2) the

ubiquitination takes place.

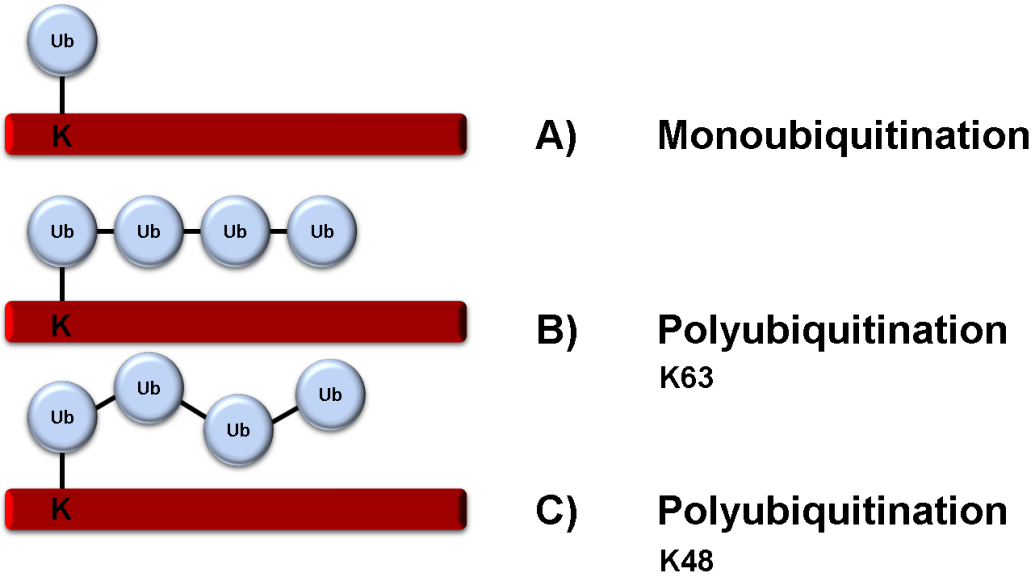


Figure 1.2: **(A)** Monoubiquitination, or ligation of a single ubiquitin molecule via a lysine, is essential for endocytosis, endosomal sorting, histone regulation, DNA repair, virus budding and nuclear export. Multi-monoubiquitination can also take place, where several single ubiquitin molecules form a linkage with the substrate lysines, implicated in receptor endocytosis. Polyubiquitination is where more than one ubiquitin molecules are attached to a lysine via a chain. In particular, polyubiquitination is involved with DNA repair, endocytosis and activation of protein kinases when linked via Lys63 **(B)**; and when linked via Lys48, targets proteins for degradation at the proteasome **(C)**. Adapted from [Haglund *et al.*, 2003].

A single ubiquitin molecule attached to a lysine in a target substrate is called monoubiquitination, and influences several processes such as; endocytosis, endosomal sorting, histone regulation, DNA repair, virus budding and nuclear export [Haglund *et al.*, 2003]; [Hurley *et al.*, 2006]. When multiple ubiquitin molecules assemble into a chain by forming an isopeptide bond between the carboxyl group of the C-terminal glycine residue of target proteins (or new ubiquitin molecules) and the amino group of internal lysine residues of ubiquitin, it is known as polyubiquitination (figure 1.3B). This modification can result in receptor endocytosis [Passmore and Barford, 2004].

Polymerisation can occur at any of ubiquitin's seven lysine positions (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 or Lys63), allowing diverse structures to form

which cause different substrate fates [Komander, 2009]. The most common sites for ubiquitin-linked chains are Lys48 and Lys63, which are labelled on the ubiquitin structure in figure 1.3A. When linked via Lys63, the polyubiquitin chain is involved in processes such as DNA repair, endocytosis and protein kinase activation, as opposed to a linkage via Lys48 which tags the protein for degradation at the 26S proteasome [Haglund *et al.*, 2003]; [Pickart, 1997]; [Tenno *et al.*, 2004].

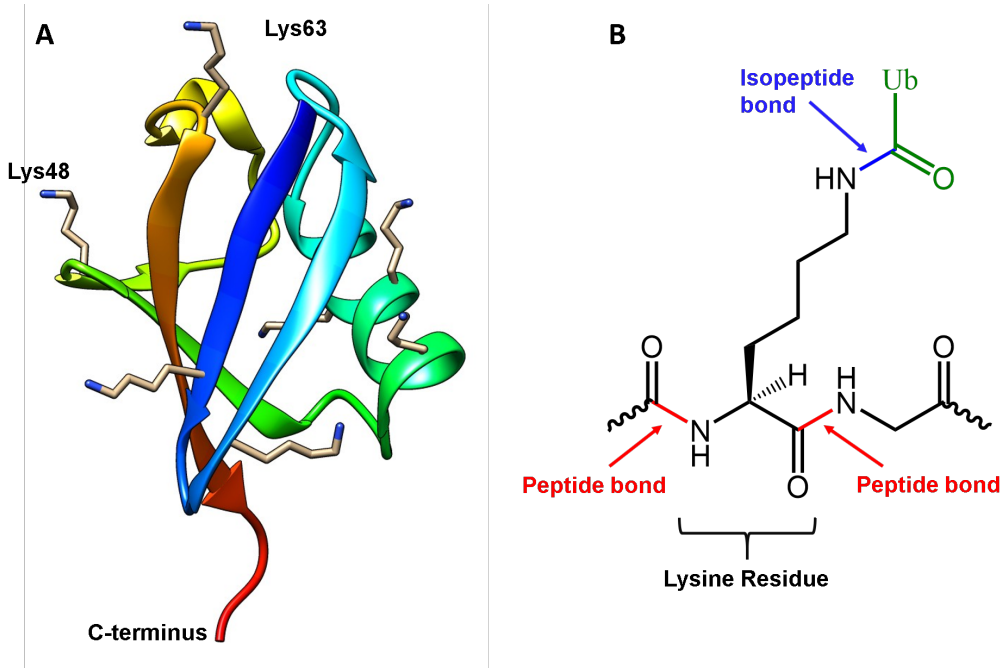


Figure 1.3: **(A)** The structure of ubiquitin (PDB: 1UBQ) showing lysine side chains, with Lys48 (involved in protein degradation) and Lys63 (involved in DNA repair, endocytosis and protein kinase activation) labelled. **(B)** A representation of the isopeptide bond between ubiquitin and a target protein or another ubiquitin molecule. The ubiquitin molecule is shown in green, and its C-terminal glycine carboxyl group forms an isopeptide bond (blue) with the epsilon amino group of a lysine residue from the target.

The 26S proteasome is a large and abundant proteolytic enzyme found in the nucleus and cytoplasm of all eukaryotic cells which has the ability to recognise proteins which are designated for degradation after being labelled with polyubiquitin. It is made up of two sub-complexes: a 20S catalytic central cylindrical core complex and a 19S regulatory complex which is bound to either or both ends of the 20S cylinder [Walz *et al.*, 1998]; [Rechsteiner, 1998], as shown in figure 1.4.

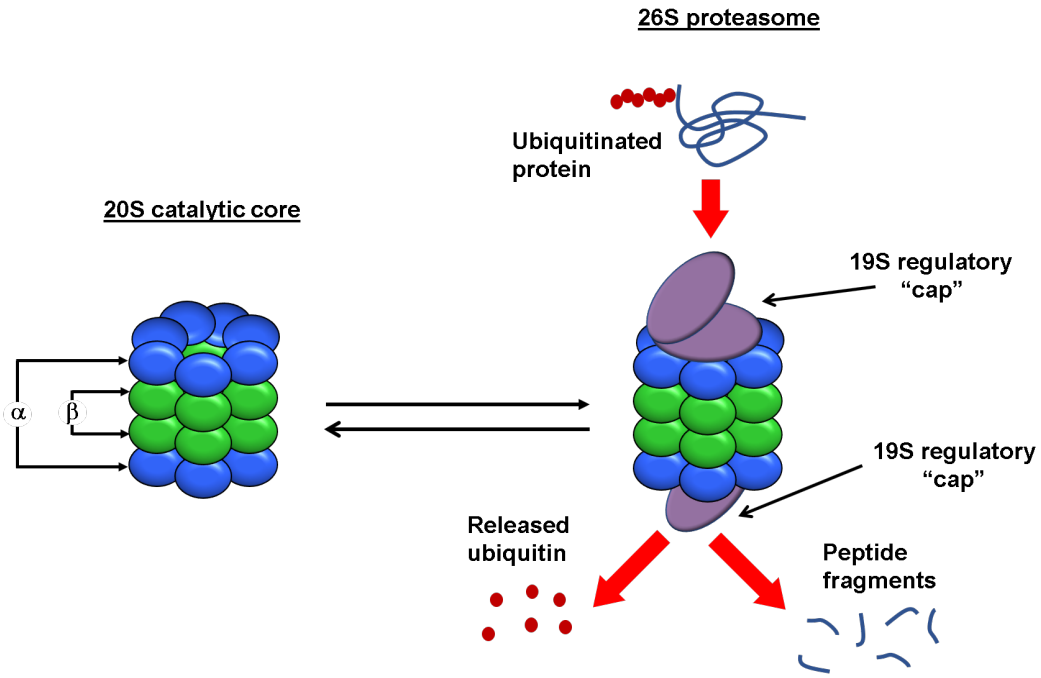


Figure 1.4: A simplified schematic of the 26S proteasome adapted from [Adams, 2002] and [Liggett *et al.*, 2010]. The 26S proteasome is a multiprotein complex made up of a proteolytically active 20S core particle (a barrel structure of four stacked rings, with labelled α and β rings) that is capped by either one or two 19S regulatory particles. The 19S regulatory units control access to the proteolytic core on recognising ubiquitinated proteins. α and β subunits of the 20S core are coloured in blue and green respectively, and the 19S regulatory cap is in purple. Ubiquitinated proteins enter the 20S catalytic core through the 19S regulatory caps and are subsequently degraded.

The 20S complex has a hollow barrel structure with four seven-membered rings (made of α and β subunits) stacked on top of each other [Lowe *et al.*, 1995]; [Groll and Huber, 2003]. The active sites point inwards towards the hollow and a small pore at either end of the structure allows small peptides or denatured proteins to enter into the hollow [Thrower *et al.*, 2000]. The 19S complex recognises the ubiquitinated target protein (as well as other potential substrates of the proteasome) and unfolds it, before inserting it into the 20S cylindrical core complex where it is disassembled into constituent amino acids for reuse [Glickman and Ciechanover, 2002]; [Hershko *et al.*, 2000]; [Lee *et al.*, 2011]. The majority of proteins targeted by the ubiquitin system are regulatory proteins and can be involved in a number of processes, including but not limited to: DNA

repair, signal transduction, transcriptional regulation, antigen processing, stress response and apoptosis. The 26S proteasome also targets proteins which are misfolded due to mutations or damage, and whose abnormal conformations could be harmful to the cell [Voorhees *et al.*, 2003]. For the proteasome to recognise a protein, it must be tagged with a chain of at least 4 ubiquitin molecules. A chain length of only 2 ubiquitin molecules causes the affinity to decrease by ∼100-fold. Whilst the signal strength is dependent on the number of ubiquitin molecules, it is a non-linear relationship, meaning that increasing the chain length beyond 4 ubiquitin molecules does impact positively on the affinity, but to a lesser extent with each addition. This indicates that the distinct conformation of the ubiquitin complex could play a part in the increase in affinity as opposed to the increasing concentration of ubiquitin molecules [Pickart, 1997]; [Thrower *et al.*, 2000].

1.1.3 E1 ubiquitin-activating enzymes

Ubiquitin-activating enzymes, labelled as E1 in figure 1.1, are involved in the first step of the ubiquitin pathway, the initial activation of ubiquitin to a high energy intermediate. They are one of three enzymes involved in the ubiquitin system; however, there are just two genes that code for the ubiquitin activator present in humans [Huang and Dixit, 2016]. The proteasome and E1 enzyme are considered to be broad targets due to their lack of specificity which could result in detrimental side-effects with respect to drug development [Huang and Dixit, 2016].

The ubiquitin-activating enzyme catalyses the activation of ubiquitin as shown in equation 1.1; [Handley *et al.*, 1991]:

$$i) E1 + Ub + ATP \overset{Mg^2+}{\rightleftharpoons} E1_{AMP\sim Ub} + PP_i \tag{1.1}$$

$$ii) E1_{AMP\sim Ub} \rightleftharpoons E1^{s\sim Ub} + AMP$$

$$iii) E1^{s\sim Ub} + Ub + ATP \rightleftharpoons E1^{s\sim Ub}_{AMP\sim Ub} + PP_i$$

where Ub = ubiquitin, s = thioester and PP_i = pyrophosphate.

Each E1 has two active sites allowing two ubiquitin molecules to bind at one time. A ubiquitin-adenylate intermediate is formed as ATP hydrolysis facilitates the attachment of AMP to the carboxyl group of the ubiquitin C-terminal glycine, with pyrophosphate released [Haas and Rose, 1982]. The adenylated ubiquitin molecule is then open to attack by the catalytic cysteine SH group after a reconfiguration of the E1 structure around the active site [Olsen *et al.*, 2010]. The cysteine group forms a thioester bond with the adenylated ubiquitin carboxyl terminus and releases AMP. A second ubiquitin molecule is then adenylated by the E1 and the E1 enzyme recruits the E2 conjugator [Haas and Rose, 1982].

The chemistry of the E1-catalysed reaction was resolved over 30 years ago and relies on the generation of a ubiquitin-adenylate intermediate, still tightly bound to an ATP-binding domain on the E1. This leaves the terminal carboxyl group on the ubiquitin molecule open to attack by the catalytic cysteine, creating a high-energy thioester bond which is then targeted by the next enzyme in the cascade [Haas and Rose, 1982], the E2 ubiquitin-conjugating enzyme. The structural events that accompany this activation are unknown but solved crystal structures of the ubiquitin-like protein SUMO in complex with its E1 enzyme [Olsen *et al.*, 2010] indicates that the conformation of the active site cysteine residue region adjusts between adenylation of the protein and formation of the thioester conjugate to allow the formation of the high-energy bond [Olsen and Lima, 2013].

1.1.4 E2 ubiquitin-conjugating enzymes

E2s, or ubiquitin-conjugating enzymes, are more diverse than the first step in the ubiquitin system, containing around 35 active E2s in humans [van Wijk and Timmers, 2010]; [Stewart *et al.*, 2016]; [Huang and Dixit, 2016]. Whilst the specificity of E2 enzymes is slightly higher, they possess a highly conserved enzymatic core, adjacent to the cysteine residue that forms the thioester bond with ubiquitin, making the development of selective E2 inhibitors challenging [Hershko *et al.*, 2000]; [Huang and Dixit, 2016].

Active E2s contain a ∼150 residue ubiquitin-conjugating domain (UBC) which contains the active site cysteine, allowing interaction with the E1s. This conserved domain is sandwiched between variable N-terminal and C-terminal tails, and these tails, along with sequence insertions, allow selectivity for interactions with E3s, therefore impacting the function of the E2s [Martinez-Noel *et al.*, 2001]. Ubiquitin E2 variant (UEV) proteins also contain a UBC domain but not the active site cysteine.

In the ubiquitin pathway, the binding of ubiquitin to an E1 triggers the passing of activated ubiquitin from E1 to an active site cysteine of the ubiquitin-conjugating enzyme E2 [Hershko, 1983]. This catalytic cysteine is found in a groove surrounded by conserved residues which are important for the formation of thioester and isopeptide bonds [Wenzel *et al.*, 2011]. The ubiquitin E2s must also selectively target E1s conjugated to ubiquitin, and not the ubiquitin-like (UBL) modifiers which are structurally similar, using highly conserved residues not present in the UBL E2s [Ye and Rape, 2009].

E2s also have a role in ubiquitin chain assembly [Pickart and Eddins, 2004], in particular for the RING E3s [Ye and Rape, 2009], and can determine the linkage specificity and ubiquitin chain length. Once a substrate has been labelled with a ubiquitin molecule, the E2-E3 pair then switches to a chain elongation mechanism, where further ubiquitins can be added to the substrate protein. It is likely that the switch occurs due to the dedicated role of a conjugated E2 [Ye and Rape, 2009] [Christensen *et al.*, 2007]. Some E2s are able to increase the rate of ubiquitin chain formation by forming short ubiquitin chains in advance of the ubiquitin elongation step [Ye and Rape, 2009].

1.1.5 E3 ubiquitin ligases

In the ubiquitin-conjugation cascade, the substrate specificity of the enzymes increases as the ubiquitination pathway progresses [Ciechanover *et al.*, 2000]. E1 enzymes have the ability to interact with one of the E2s which can, in turn, interact with one of ∼700 E3 ligases found in the human genome, allowing the

ubiquitination of thousands of substrate proteins [Hershko *et al.*, 2000]. The size and structure of E3 ligases are varied, causing the identification of new E3 enzymes to be demanding [Li *et al.*, 2008]. Despite this, several E3s are now characterised and indicate the presence of several significant E3 domain structures [Iconomou and Saunders, 2016]. This domain sequence can also be found in many mammalian, viral and plant genomes, which suggests that some of these could also be E3s [Scheffner and Staub, 2007]; [Ardley and Robinson, 2005]. In addition to the ∼700 ubiquitin E3 ligases, there are well over 100 deubiquitinating enzymes (DUBs) with the ability to remove ubiquitin from substrate proteins, leading to a reasonably high risk of a mutation occurring in one of these proteins with the capability to cause disease [Popovic *et al.*, 2014].

The E3s are a widely diverse [Hershko and Ciechanover, 1998] but highly conserved group of proteins which are responsible for substrate specificity and regulation of ubiquitination. At least one of the common domain structures is generally present in each E3 protein [Ardley and Robinson, 2005]. The E3s are involved in the regulation of cell proliferation, cell cycle arrest and apoptosis [Shi and Grossman, 2010], and notably have the ability to mediate the specificity of substrate ubiquitination, making this enzyme a key target for drug discovery [Huang and Dixit, 2016]. Dysfunction of the E3 ligases is often linked with neurodegenerative disease and cancer [Nalepa *et al.*, 2006].

The E3s can largely be categorised into three distinct families: RING (Really Interesting New Gene) finger, HECT (Homologous to E6-associated protein C-Terminus) and U-box (UFD2 homology) [Pickart and Eddins, 2004]. These families have adopted distinct mechanisms for achieving the transfer of ubiquitin from the E2 onto a substrate protein. The HECT domain E3 ligases have a catalytic cysteine residue and they attach themselves to the C-terminus of activated ubiquitin by forming a covalent thioester intermediate, before the ubiquitin is transferred onto the lysine side chain of a substrate protein [Christensen and Klevit, 2009]. RING E3 ligases do not possess an active-site cysteine residue and instead act as Zn^{2+} regulated scaffolds for E2 enzymes, enabling the ubiquitin to transfer directly onto the substrate protein [Christensen and Klevit, 2009]. They have also been

described as having the ability to allosterically activate E2s [Özkan *et al.*, 2005]. U-box is a modified RING motif without the full complement of Zn^{2+} binding ligands [Bernassola *et al.*, 2008].

1.1.5.1 RING finger E3 ligases

The RING finger E3 ligases are the largest family (over 600) of E3s [Li *et al.*, 2008] and are involved in a number of different cellular functions [Deshaies, 1999]; [Borden, 2000]. They contain a RING finger domain (∼70-residues) [Borden and Freemont, 1996]; [Jackson *et al.*, 2000] which gets its name as it was discovered in the protein sequence product of a gene called Really Interesting New Gene 1 (RING1). This domain consists of 7 specifically placed cysteines and a histidine residue, forming a folded domain [Lorick *et al.*, 1999], as illustrated in figure 1.5. This structure then binds to zinc ions at a ratio of 4:1, Cys/His:Zn^{2+}, (known as a cross-brace formation) [Borden and Freemont, 1996], and assists with the transfer of ubiquitin to substrates or to RING finger proteins [Hegde, 2004].

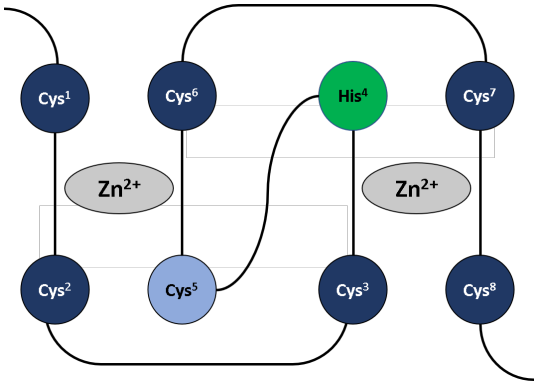


Figure 1.5: A schematic diagram of the RING finger domain cross-brace formation. Zinc ions are each coordinated tetrahedrally by the four binding ligands. Zinc atoms are shown in grey, cysteines in blue and histidine in green. Cys^5 is shown in light blue as in certain RING E3s a histidine residue is present in its place. Adapted from [Jackson *et al.*, 2000] and [Borden and Freemont, 1996].

Of the 7 cysteine residues, none are catalytic cysteines and so cannot bind to ubiquitin. Instead, they act as a scaffold, interacting with E2s to bring them into close proximity with the substrate so that the transfer of ubiquitin can occur

[Glickman and Ciechanover, 2002]. Cys^5, in figure 1.5, is shown in light blue as in certain RING E3s a histidine residue is present in its place, although the cross-brace structure is not affected [Borden and Freemont, 1996]; [Jackson *et al.*, 2000].

A RING finger E3 can form different linkages to ubiquitin, depending on which E2 it is paired with [Metzger *et al.*, 2014] and can also form either single-subunits or multi-subunits which allow them to carry out certain roles. Single-subunit RING finger E3s contain the RING finger domain as well as a substrate recognition site, whereas the formation of heterodimers, or multi-subunit RING finger E3s, are required for certain E3s to carry out their role. In particular, this is important for E3s which lack the ability to bind E2s [Metzger *et al.*, 2014]. RING finger E3s also include a number of enzymes which have been classed as tumour suppressors or oncoproteins [Bernassola *et al.*, 2008]. The U-box domain family holds the same conformation but cannot coordinate to zinc ions. Instead, they stabilise the structure by forming salt bridges.

1.1.5.2 HECT E3 ligases

HECT type E3 ligases are a subfamily of ∼30 members, which contain a conserved HECT domain. They are implicated in protein trafficking, immune response and cellular proliferation [Rotin and Kumar, 2009] and can be divided further into three families of HECT E3 ligases [Rotin and Kumar, 2009]; the HERC family (HERC1 to 6), which contain a regulator of chromosome condensation 1 (RCC1)-like domain, the NEDD4 family (NEDD4, NEDD4L, SMURF1/2, ITCH, WWP1, WWP2, NEDL1 and NEDL), which contain between two and four WW domains and finally the HECT proteins which fit neither or the previous groups (E6AP, HECTD2, KIAA0614, TRIP12, G2E3, EDD, HACE1, HECTD1, UBE3B, UBE3C, KIAA0317, HUWE1 and HECTD3) and instead contain various domains including RING, DOC and zinc-finger domains [Rotin and Kumar, 2009]. All of the HECT E3 ligases contain a HECT domain which occurs at the C-terminus and is much larger than the RING motif,

containing ∼350 amino acids [Rotin and Kumar, 2009]. The HECT domain regulates a variety of processes, for example the NEDD4 ITCH E3 ligase is implemented in the regulation of the immune system [Gao *et al.*, 2004], whereas the Smurf1 E3 ligase is a negative regulator of the bone morphogenetic protein (BMP) signaling pathway [Cao *et al.*, 2014]. Other roles displayed by the HECT domain include membrane trafficking [Garcia-Gonzalo and Rosa, 2005] and regulating degradation [Chen *et al.*, 2005].

The structure of the HECT domain consists of two lobes, with the catalytic cysteine residue at the C-terminal end, and the E2 binding domain at the N-terminal lobe [Huibregtse *et al.*, 1995]. The HECT type E3s function in a similar way to the E1s and E2s which contain an essential cysteine residue in the HECT domain [Christensen and Klevit, 2009]. An intermediate thioester bond forms between the thiol group of the catalytic cysteine residue and the carboxyl group of the ubiquitin C-terminal glycine residue before the ubiquitin is transferred to the target protein [Huang *et al.*, 1999]; [Bernassola *et al.*, 2008]; [Scheffner *et al.*, 1995]; [Kim and Huibregtse, 2009]; [Spratt *et al.*, 2014]. The two lobes are thought to be connected by a flexible hinge region, which allows them to move closer together when transferring ubiquitin [Ogunjimi *et al.*, 2005]; [Verdecia *et al.*, 2003]. When a multi-ubiquitin chain is formed on a substrate protein, it tags that protein for degradation at the proteasome.

E6-AP1 (E6-associated protein 1) was one of the first E3s to be identified and characterised before many other proteins which contained a HECT (homologous to E6-associated protein C-terminus) domain followed suit [Ardley and Robinson, 2005]. HECT E3s play an important role in numerous biological processes such as protein trafficking, the immune response, and in several signaling pathways that regulate cellular growth and proliferation [Bernassola *et al.*, 2008].

This thesis explores a member of the HECT E3 ligase family which is of interest due to the frequent mutation, absence or defects of this subfamily in many diseases such as cancer. In particular, NEDD4 E3s have been discovered as crucial regulators of cancer development and therapy and hence will be a focus of this study.

1.2 The NEDD4 family

Within the HECT E3 ligases is a small sub-group of nine known as the NEDD4 (Neural precursor cell Expressed Develop-mentally Down-regulated 4) family. Figure 1.6B shows schematic diagrams for each of the NEDD4 family members, highlighting the differences in configurations of the C2 and WW domains. The NEDD4 sub-group of E3 ligases is characterised by three functional domains (figure 1.6A). The first is an N-terminal Ca^{2+} dependent, phospholipid-binding C2 domain involved in membrane interaction [Ingham *et al.*, 2004]; [Dunn *et al.*, 2004]; [Rizo and Südhof, 1998]. Next, the central region contains between two and four WW domains which are able to recognise target substrates [Ingham *et al.*, 2005]; [Riling *et al.*, 2015]. Finally, the C-terminal HECT domain is a highly conserved domain of roughly 40 kDa which conjugates ubiquitin to the target [Kanelis *et al.*, 2001].

The C2 domain binds to phospholipids when in the presence of Ca^{2+} [Kanelis *et al.*, 2001]; however, there are many other C2 containing proteins which contain non-Ca^{2+} binding C2 domains [Cho, 2001]. WW domains of ∼35-40 amino acids are crucial for the specific interaction with target substrates and are named as a result of two Tryptophan (W) residues found within its sequence which are spaced 21 amino acids apart. These domains display a preference for protein substrates which contain PPxY motifs (Proline-Proline-x-Tyrosine where x = any amino acid) [Ingham *et al.*, 2005]. The HECT domain acts as a carrier for ubiquitin and directly conjugates ubiquitin to the target lysine residue using its active-site cysteine residue [Kanelis *et al.*, 2001]; [Zhang *et al.*, 2016].

The compositions of the NEDD4 family range from 748 aa (amino acids) (Smurf2) to 1606 aa (NEDL1) in size (figure 1.6B). The varying number of WW domains in each ligase is linked to its function, for example, Smurf1 and Smurf2 have two and three WW domains respectively. Smurf1 binds to and promotes the degradation of Smad1 and Smad5 [Zhu *et al.*, 1999], as opposed to Smurf2 which inhibits the TGFβ pathway by adding a single ubiquitin molecule to Smad3 [Tang *et al.*, 2011]. Along with the varying compositions shown in figure 1.6,

differing natural isoforms also exist, suggesting additional functions and binding partners are possible [Chen and Matesic, 2007]; [Dunn *et al.*, 2004]; [Flasza *et al.*, 2002].

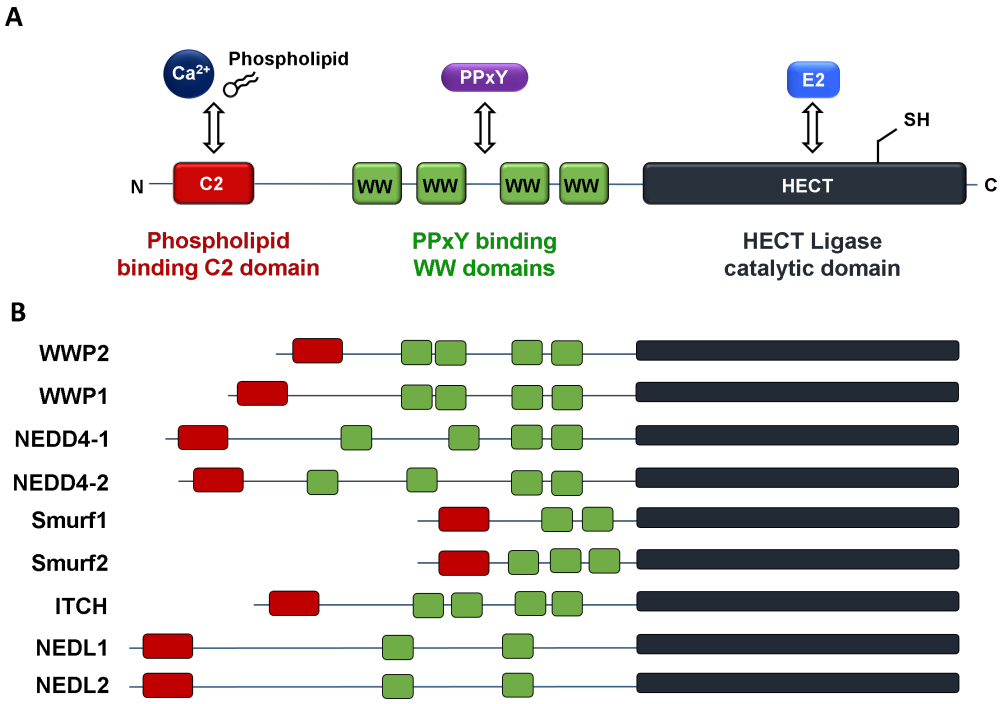


Figure 1.6: **(A)** [A schematic representation of the general structure of the NEDD4 E3 ligase family showing the N-terminal C2 domain, between two and four WW domains and a catalytic HECT domain. The C2 domain binds Ca^{2+} and phospholipids, which targets NEDD4 ligases to intracellular membranes. The WW domains (Trp-Trp) bind conserved PPxY sequences on substrates and hence play a role in controlling substrate selectivity. The HECT domain contains a catalytic cysteine residue which forms an active thioester complex with ubiquitin. **(B)** Simple schematic diagrams showing the structural domains of the nine Nedd4-like E3 ligases.

Of the nine NEDD4 ligase family, WWP1, WWP2, Smurf1, Smurf2 and NEDD4L are all confirmed as being over-expressed in cancers [Chen and Matesic, 2007]; [Wang *et al.*, 2007]; [Qin *et al.*, 2016]; [Tanksley *et al.*, 2013]; [Xu *et al.*, 2016]. Various targets of these ligases are known, including Octamer binding Transcription factor 4 (OCT4), the Smads (TGFβ signaling pathway) and the tumour suppressor Phosphatase and Tensin homolog (PTEN) [Wang *et al.*, 2007]; [Xu *et al.*, 2004]; [Tanksley *et al.*, 2013]; [Kwan *et al.*, 2011]; [Kwon *et al.*, 2013]; [Maddika *et al.*, 2011]; [Yang *et al.*, 2015].

1.2.1 WWP2 (WW domain containing Protein 2)

Human WWP2 (WW domain containing protein 2) was originally identified when screening for WW domain-containing proteins [Pirozzi *et al.*, 1997]. It is a HECT-type E3 ligase, one of nine NEDD4-like proteins, which contains four WW domains [Xu *et al.*, 2004], and is found to be expressed in undifferentiated ESCs (Embryonic Stem Cells) and ECCs (Embryonal Carcinoma Cells). When overexpressed in animal models, an increase in tumour spread was observed, and altered expression of its isoform WWP2-N was observed in prostate cancer and melanoma [Soond *et al.*, 2013].

Several WWP2 substrates linked to tumorigenesis have been encountered such as the PTEN tumour suppressor [Maddika *et al.*, 2011], key signaling components within the transforming growth factor-β (TGFβ) pathway known as Smads [Choi *et al.*, 2015], and embryonic transcription factor, OCT4 [Xu *et al.*, 2004]; [Choi *et al.*, 2015]. PTEN, a lipid phosphatase, is often found to be mutated in human cancer cells and experiences degradation at the proteasome which is induced by WWP2 [Maddika *et al.*, 2011]. Low PTEN expression in endometrial tumours led to overexpression of WWP2 suggesting an oncogenic role [Yang *et al.*, 2015]. Stem cell specific transcription factor OCT4, which is often up-regulated in cancer stem cells [Hu *et al.*, 2010], was also found to interact with WWP2 [Xu *et al.*, 2009]. WWP2 interacts with various Smad proteins responsible for signaling activity of the TGFβ pathway. TGFβ can cause cells to undergo apoptosis, or it can stimulate cells to undergo Epithelial-Mesenchymal Transition (EMT), a process which converts static epithelial cells into highly invasive mesenchymal cells permitting tumorigenic metastasis to take place [Soond and Chantry, 2011]. E3 ubiquitin ligase WWP2 is also reported to have a key involvement in craniofacial development [Zou *et al.*, 2011] and chondrogenesis [Maddika *et al.*, 2011]. Overexpression of WWP2 has been recognised in liver tumour cells and when blocked, liver cancer cell adhesion, invasion, and migration was inhibited [Qin *et al.*, 2016]; [Xu *et al.*, 2016].

Figure 1.7 shows the naturally occurring isoforms WWP2-C and WWP2-N as well as the domain constructs WW2, WW3 and WW2-3 which are a focus of this thesis.

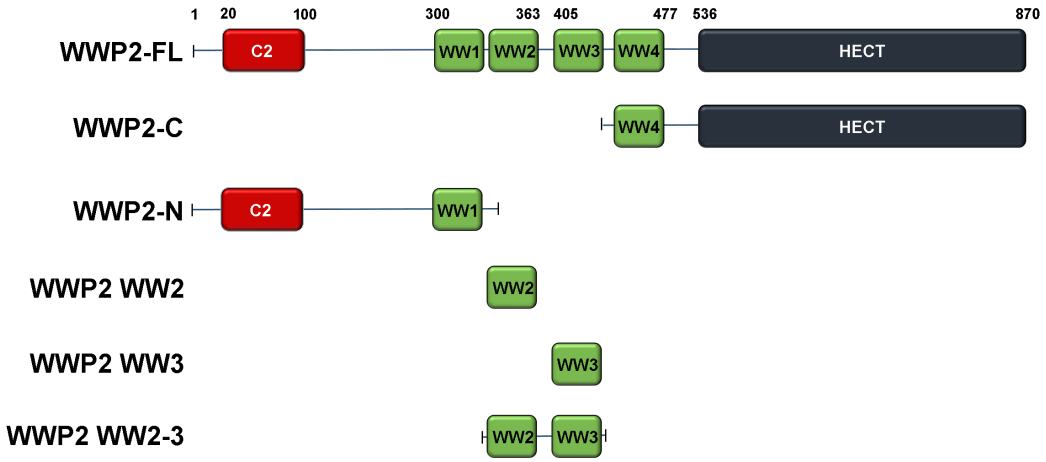


Figure 1.7: A schematic diagram of WWP2 showing the C2, WW and HECT domain boundaries. Naturally occurring isoforms WWP2-C and WWP2-N are shown with domain constructs WW2, WW3 and WW2-3 highlighted as they are a focus of this report.

The sequence alignment of WW domains present in the full length WWP2 construct are shown in figure 1.8, as determined using the ClustalW server [Larkin *et al.*, 2007]; [Goujon *et al.*, 2010]. Secondary structure was predicted using the PSIPRED webserver v3.3 [Jones *et al.*, 2018]; [Buchan *et al.*, 2013] and orange arrows represent predicted β strands.

E3 ligases are the second most prevalent cancer-related functional gene family next to protein kinases [Shi and Grossman, 2010] and it is only recently that various HECT E3 ubiquitin ligases, such as WWP2, have emerged as crucial regulators of cancer development. Tissue-specific expression patterns and differential preference for Smad proteins [Soond and Chantry, 2011] was observed for isoforms of WWP2 and these isoforms have potential as biomarkers or therapeutic targets for cancer [Soond *et al.*, 2013].

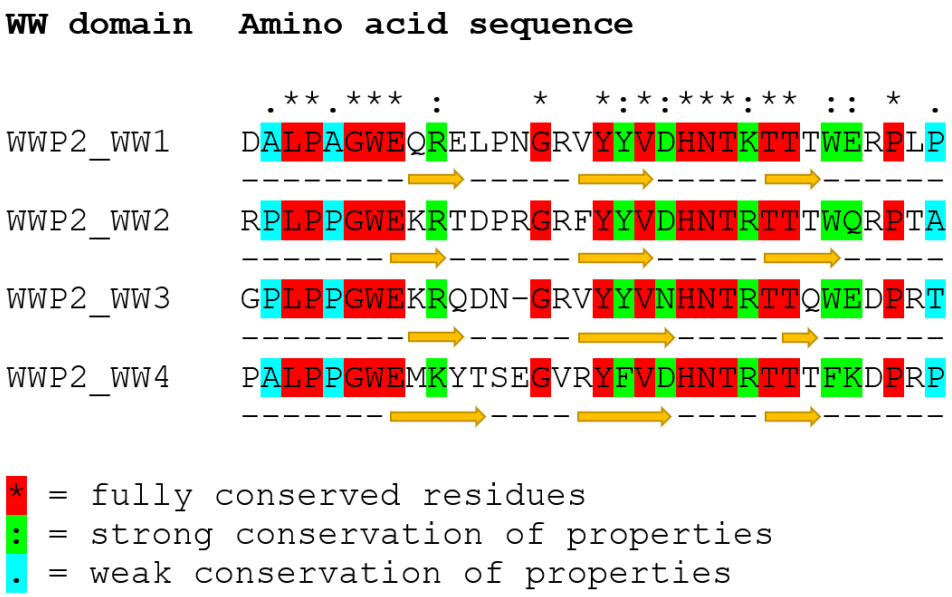


Figure 1.8: The Sequence alignment of WW domains present in the full length WWP2 construct. Secondary structure was predicted using the PSIPRED webserver v3.3 [Jones *et al.*, 2018]; [Buchan *et al.*, 2013] and orange arrows represent predicted β strands. Regions highlighted in red are fully conserved residues, in green are residues with strongly related properties and in blue are residues with weakly related properties, according to the ClustalW server [Larkin *et al.*, 2007]; [Goujon *et al.*, 2010].

1.2.2 WW domains

Many signaling proteins contain domains capable of mediating specific protein-protein interactions, by recognising short peptide motifs of their binding partners. In particular, the WW domains are protein-protein interaction domains of between 30 and 40 amino acids in length and have a preference for binding small proline-rich sequences, commonly the PPxY (where P is proline, Y is tyrosine and x is any amino acid) motif [Sudol *et al.*, 1995]; [Sudol, 1996b]; [Panwalkar *et al.*, 2016]. They have also been found to interact with phospho-serine/threonine residues in certain substrates [Sudol *et al.*, 1995]; [Lu *et al.*, 1999]. WW domain-containing proteins have roles in a variety of cellular processes, including transcription, RNA processing, protein trafficking and receptor signaling and have been linked to cancer, hereditary disorders,

Alzheimer's and Huntington's disease. They are named after their two highly conserved tryptophan residues (WW) spaced 20-22 amino acids apart [Bork and Sudol, 1994]; [Sudol, 1996a]. They assemble into a single three stranded β sheet containing a conserved proline residue [Bork and Sudol, 1994]; [Sudol *et al.*, 1995]; [Zarrinpar and Lim, 2000] three residues C-terminal to the second tryptophan [Staub and Rotin, 1996] and two ligand-binding grooves. The two tryptophan residues are located on opposite sides of the β sheet [Chong *et al.*, 2010], one of which is in the hydrophobic core, providing domain stability, and the other is within a target protein substrate binding pocket [Panwalkar *et al.*, 2016]. There are instances where both tryptophans are not present in a WW domain and instead, one may be replaced by an aromatic residue phenylalanine or tyrosine [Sudol *et al.*, 2005]; [Bork and Sudol, 1994].

The WW domains have been split into four groups, defined by their proline-based ligand binding specificity [Espanel and Sudol, 1999]. Group I domains interact with the core PPxY motif sequence as described above. Group II domains bind to a proline region which usually surrounds a leucine residue (PPLP motif) [Bedford *et al.*, 1997]; [Bedford *et al.*, 1998]. Group III bind to the proline rich region which encompass arginine or lysine residues (PR motif) [Sudol *et al.*, 2001] and finally, group IV binds to a phosphorylated serine or threonine residues followed by a proline (pSP/pTP) [Sudol *et al.*, 2001]; [Sudol and Hunter, 2000]. The group II and group III WW domains are merged in some literature as they interact with similar proteins [Ingham *et al.*, 2005].

Different WW domains from the same protein possess diverse substrate specificity, and so it is likely that the NEDD4 family proteins are able to specifically select different substrate proteins via their WW domains despite only minor differences in their structures. Most HECT-type E3 ligases tend to contain multiple WW domains, hence broadening the variety of substrates available for interaction [Ilsley *et al.*, 2002]; [Ingham *et al.*, 2005]. In the past, structural studies of WW domains were often problematic due to common undesirable properties such as insolubility, tendency to aggregate and conformational flexibility [Umadevi *et al.*, 2005]. However, with the use of solubility tags, many

WW domain structures have been elucidated.

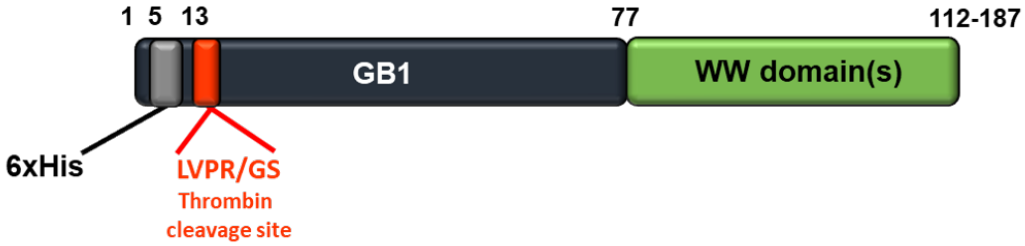


Figure 1.9: A schematic representation of the GB1-WW domain(s) construct. A hexa-His tag (grey) is present at the beginning of the construct to enable nickel-affinity purification. This region is later cleaved using a thrombin cleavage site (red). The GB1 domain (dark blue) is used to increase solubility and is described in more detail in section 3.1.3. Finally, the length of the WW domain region (green), is dependant on which WW domains are included in the construct. His-GB1-WW2 is 112 amino acids in length, as opposed to the tandem His-GB1-WW2-3 construct, which is 187 amino acids in length.

Figure 1.9 shows a schematic diagram of the GB1-WW domain(s) construct. WW domains were expressed as a fusion protein with a GB1 protein tag for solubility (see section 3.1.3) and an additional hexa-histidine tag (6 x His), which is used to facilitate the purification of the protein. The use of polyhistidine tags [Smith *et al.*, 1988] is now a common method for purification of recombinant protein [Carson *et al.*, 2007]. Their mechanism works by a sequence of histidine residues (commonly 6) binding to a nickel-affinity column and hence allowing impurities to be removed from the sample through washes. The hexa-histidine tag and fusion protein can then be eluted by introducing a high concentration of competitively binding imidazole, which has a very similar structure to histidine. The variation in length of WW domain depends on which domains are present.

Figure 1.10 shows four example WW domain RCSB PDB [Berman *et al.*, 2000] (`http://www.rcsb.org`) structures in complex with ligands (A-D), each shown in two orientations. WW domains are displayed in green and peptide structures are shown with amino acid labels in blue. The majority of WW domain complexes present on the RCSB PDB [Berman *et al.*, 2000] were NMR solution state structures, with few crystal structures.

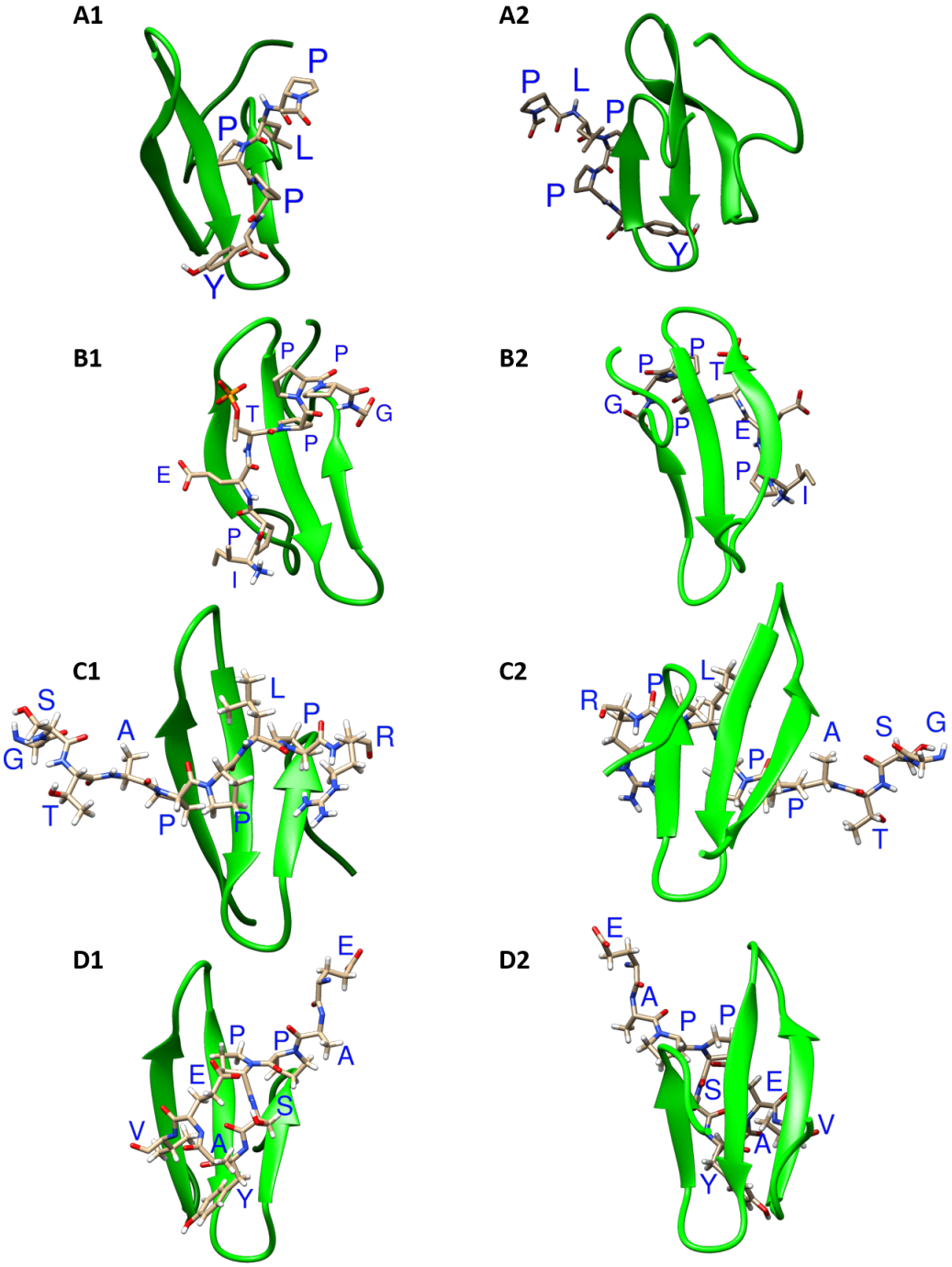


Figure 1.10: Example lowest energy WW domain PDB structures in complex with ligands, with 1 and 2 being different orientations. WW domains are shown in green and peptide structures are shown with amino acid labels in blue. **A** Structure of YAP65 WW1 domain complexed to Acetyl-PLPPY, derived using solution NMR. [PDB=1K9R], [Pires *et al.*, 2001]. **B** Structure of the WW domain of PIN1 in complex with a human phosphorylated Smad3 derived peptide, derived using solution NMR. [PDB=2LB3], [Aragón *et al.*, 2011]. **C** Solution structure of the first WW domain of FBP11 / HYPA (FBP11 WW1) complexed with a PL (PPLP) motif peptide ligand [PDB=2DYF], [Kato *et al.*, 2006]. **D** Crystal Structure of the complex of 3rd WW domain of Human Nedd4 and 1st PPXY Motif of ARRDC3 [PDB=4N7H], [Qi *et al.*, 2014].

Figure 1.10**A** shows the structure of YAP65 WW1 domain when complexed to Acetyl-PLPPY peptide [PDB ID=1K9R], [Pires *et al.*, 2001]. The structure was derived using solution NMR and the final ensemble of 20 lowest energy structures had a RMSD of 0.13 Å.

Figure 1.10**B** shows the structure of the WW1 domain of PIN1 in complex with a human phosphorylated Smad3 derived peptide [PDB ID=2LB3], [Aragón *et al.*, 2011]. The structure was derived using solution NMR and the final ensemble of 20 lowest energy structures had a RMSD of 0.20 Å.

The solution structure of the WW1 domain of FBP11 / HYPA (FBP11 WW1) complexed with a PL (PPLP) motif peptide ligand is shown in figure 1.10**C** [PDB ID=2DYF], [Kato *et al.*, 2006]. This structure was also derived using solution NMR and the 20 lowest energy structures had a RMSD of 0.02 Å.

Finally, figure 1.10**D** shows the crystal structure of the WW3 domain of Human Nedd4 when in complex with the 1st PPXY Motif of ARRDC3 [PDB ID=4N7H], [Qi *et al.*, 2014]. This structure was derived using x-ray diffraction and yielded a resolution of 1.70 Å.

1.3 Smad7

1.3.1 The TGFβ signaling superfamily

The transforming growth factor-β (TGFβ) superfamily of growth factors is a group of more than 50 structurally related polypeptide growth factors, split into subgroups including the Bone Morphogenic Proteins (BMPs), Activins, Inhibins, Glial cell line-Derived Neurotrophic Factor (GDNF) and the TGFβ subfamily (TGFβ_1, TGFβ_2 and TGFβ_3) [Massagué, 1998]; [Böttner *et al.*, 2000]. The group was named after the first of its members, TGFβ_1, originally described by Assoian in 1983 [Assoian *et al.*, 1983]. Each of these growth factors were identified mainly through their roles in development as they are able to regulate tissue differentiation

by controlling cell proliferation, differentiation, migration and apoptosis, as well as having important functions during embryonic development [Böttner *et al.*, 2000]; [Derynck *et al.*, 2001]; [Massagué, 1998]; [Whitman, 1998]; [Heldin *et al.*, 2009].

The TGFβ growth factors contain two identical subunits, usually linked by a disulfide bond. The TGFβ signaling pathway involves binding of the dimer to membrane receptor serine/threonine kinases. There are roughly 500 human serine/threonine kinases, with only twelve at the cell surface. All twelve function as receptors for the TGFβ family and are hence known as the TGFβ receptor family. The receptors can be split into the type I TGF-β membrane receptor (TβRI) subfamily (with 7 members in humans) and the type II TGF-β membrane receptor (TβRII) family (with 5 members in humans) at $\sim$55 kDa and $\sim$70 kDa respectively [Macias *et al.*, 2015].

The TGFβ subfamily has been implicated both as tumour suppressors and tumour enhancers [Derynck *et al.*, 2001]. Disruption or mutations to certain components in the TGFβ pathway have been associated with several human diseases, including cancer [de Caestecker *et al.*, 2000]; [Derynck *et al.*, 2001].

1.3.2 The Transforming Growth Factor β (TGFβ) signaling pathway

The TGFβ group consists of ligands and receptors which are involved in signaling through well characterised downstream mediators known as the Smads. Following ligand binding by a member of the TGFβ family, receptor serine/threonine kinases are able to directly phosphorylate and activate the Smads [Massagué and Chen, 2000]. The Smad nomenclature is derived from the combination of the Mad (Mothers against dpp) gene in Drosophila and the related Sma genes in *Caenorhabditis elegans* [Sekelsky *et al.*, 1995]; [Derynck *et al.*, 1996]; [Savage *et al.*, 1996]. The human genome encodes eight different types of Smad; Smad1 to Smad8. These can be subdivided into three categories: receptor-activated Smads (R-Smads: Smad1, Smad2, Smad3, Smad5, Smad8),

which become phosphorylated by the receptors; common mediator Smads (Co-Smads: Smad4), which bind to activated R-Smads; and inhibitory Smads (I-Smads: Smad6, Smad7), which are induced by members of the TGFβ family [Mehra *et al.*, 2000]; [Kitisin *et al.*, 2007].

The TGFβ signaling pathway is triggered by ligands of the TGFβ family binding to receptors and is characterised in general terms in figure 1.11.

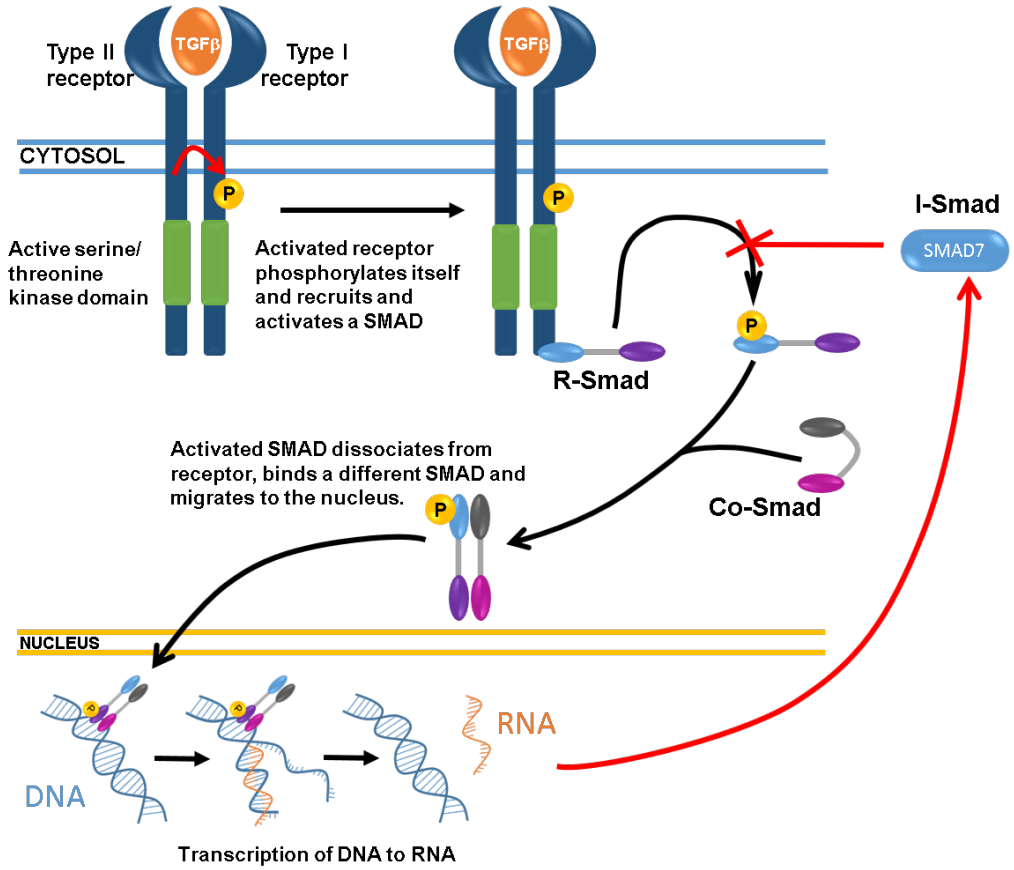


Figure 1.11: A schematic diagram showing the self-phosphorylation of the receptor serine/threonine kinases and the recruitment and activation of a cytoplasmic gene regulatory protein (R-Smad). The R-Smad is then shown to separate from the receptor and bind to a Co-Smad. These complexes migrate to the nucleus, where they instigate transcription of specific target genes. Negative regulation of the TGFβ signaling can be seen where the activation of the Smad gene regulatory protein is prevented by the I-Smad. Adapted from van Laar and Huizinga [2005].

The R-Smads are activated by relatively specific TGFβ ligands. The ligand

binds to the type II TGF-β membrane receptor (TβRII), which forms a complex with the type I receptor. The type II receptor then phosphorylates the type I receptor at various serine and threonine residues [Shi and Massagué, 2003]; [Wrana *et al.*, 1994]. The type I receptor kinases recruit the R-Smads and phosphorylate them, activating them and instigating the formation of a complex with a Co-Smad. This complex then migrates to the nucleus where it regulates gene expression. Negative regulation of the TGFβ signaling can be seen where the activation of the Smad gene regulatory protein is inhibited by the I-Smad.

When applied to the TGFβ signaling pathway, the TGFβ subfamily bind to the membrane receptor and trigger a signaling cascade involving R-Smads 2 and 3, Co-Smad 4 and the I-Smad 7. Smad7 competes with the R-Smads for receptor interaction and marks these receptors for degradation [Hayashi *et al.*, 1997].

All Smads are made up of two globular domains separated by a proline-rich linker region which bind WW domains [Massagué, 1998]; [Attisano and Wrana, 1998]. The C-terminal Mad homology 2 (MH2) domain is highly conserved in all eight of the Smads and acts by mediating Smad oligomerisation and Smad-receptor interaction. There is also a phosphorylation site SSxS motif (Ser-Ser-Val/Met-Ser sequence) at the C-terminus of the MH2 domain of R-Smads [Kretzschmar *et al.*, 1997]. A highly conserved N-terminal Mad homology 1 (MH1) is present in R-Smads and Co-Smads and functions by binding DNA and hence regulates transcription [Macias *et al.*, 2015]. The MH1 and MH2 domains are autoinhibitory, therefore the R-Smads require TGFβ stimulation for the signaling cascade to occur. The N-terminus sequence of the I-Smads is somewhat dissimilar to both the R-Smads and the Co-Smads [Ten Dijke and Hill, 2004].

1.3.3 Smad7 and the TGFβ pathway

Several of the NEDD4 HECT E3 ligases, including Smurf1, Smurf2, WWP1 and Nedd4-2 are able to negatively regulate the TGFβ signaling pathway [Chen and Matesic, 2007]; [Izzi and Attisano, 2004]. As mentioned above, the I-Smad, Smad7, has the ability to inhibit the TGFβ mechanism, which is crucial for healthy cells

due to its oncogenic effects. Figure 1.11 illustrates Smad7, which is expressed in response to TGFβ, forming a negative feedback loop by inhibiting TGFβ signaling [Yan and Chen, 2011]; [Yan *et al.*, 2009]; [Bai and Cao, 2002]; [Hata *et al.*, 1998]; [Quan *et al.*, 2002]; [Kavsak *et al.*, 2000]; [Nakao *et al.*, 1997].

Smad7 competes with the R-Smads for receptor binding to the activated type I receptor, [Hayashi *et al.*, 1997] preventing phosphorylation of the R-Smads and hence inhibiting the TGFβ signaling. The inhibitory role of Smad7 in the TGFβ signaling pathway makes it an interesting target and its expression pattern is found to be altered in a number of different cancers [Stolfi *et al.*, 2013]; [Yan *et al.*, 2009]. Smad7 also recruits several HECT type NEDD4 E3 ubiquitin ligases to the TGFβ receptor [Inoue and Imamura, 2008] through an interaction between the proline-rich linker region of Smad7 and the WW domains of the E3 ligases. Once recruited to the receptor, the E3 ligase HECT domain polyubiquitinates Smad7 and the TGFβ receptor and hence labels them both for degradation at the proteasome [Kavsak *et al.*, 2000].

1.3.4 Regulation of TGFβ signaling by WWP2

Involvement of WWP2 in the regulation of TFGβ signaling pathways occurs via its interaction with Smad2, Smad3 and Smad7 proteins. The three naturally occurring isoforms of WWP2, shown above in section 1.2.1, interact with the Smads with different selectivity. The WWP2-C isoform only interacts with Smad7, as opposed to the WWP2-N isoform which binds Smad2 and Smad3 [Soond and Chantry, 2011]. The WWP2-N isoform was found to interact with R-Smads, Smad2 and Smad3 but no interaction was observed for Smad7. The WWP2-N isoform does not contain the HECT ubiquitin ligase domain and instead formed a complex with WWP2-FL with TGFβ regulation, allowing degradation of Smad2 and Smad3 by upregulating WWP2-FL activity [Soond and Chantry, 2011]. Upon TGFβ stimulation, the two isoforms dissociate. The interaction of WWP2-N with R-Smads suggests a specificity for WW1 recognising the Smad2 and Smad3 proline rich linker region. WWP2-C appears

to only interact with Smad7, leading to its degradation at the proteasome; however, no interaction is seen with the R-Smads. Therefore, it is likely that the WW domain present in the WWP2-C isoform, WW4, preferentially binds Smad7 over R-Smads. WWP2-FL causes substantial degradation of Smad7 and degradation of Smad3 to a lesser extent. The presence of TGFβ also led to an increase in the rate of ubiquitin-mediated degradation of Smads by WWP2 [Soond and Chantry, 2011].

1.3.5 Smad7 structure and constructs

Figure 1.12 shows a schematic diagram of Smad7 and its domains, where the MH1 and MH2 domains are separated by a proline-rich linker region [Wrana and Attisano, 2000]. This linker is of particular importance as it contains the WW domain binding PPxY motif [Aragón *et al.*, 2012].

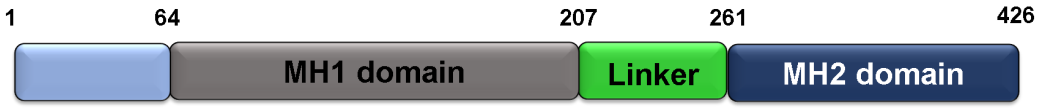


Figure 1.12: A schematic diagram of the structure of Smad7. The MH1 domain which has a role in DNA binding is shown in grey, the MH2 domain which mediates interaction with a wide variety of proteins is shown in Navy, and a linker region containing a PPxY motif is shown in green.

A short peptide construct derived from the linker region of Smad7 was designed by Lloyd Wahl [Wahl, 2016], to include the region of interest, the PPxY motif. The sequence expressed has been commonly used in WW domain ligand binding experiments, but additional residues were included at the N-terminus to increase the molecular weight [Wahl, 2016], allowing for the dialysis of Smad7 peptide using a commercially available dialysis membrane. Bacterial expression of small peptides can be problematic due to solubility difficulties as well as poor expression in cell culture. A reason for this is their unstructured state in solution which can be susceptible to protease activity [Kuliopulos *et al.*, 1994]. A solution to this problem is to express the Smad7 peptide as a fusion protein construct, which

assists the peptide stability during expression and the solubility of the expressed peptide. Smad7 was expressed as a fusion protein with a SUMO (Small Ubiquitin-like Modifier) protein tag and an additional hexa-histidine tag (6 x His), which was used to facilitate the purification of the protein [Butt *et al.*, 2005]. In this case, the polyhistidine tag was applied to the N-terminus of a SUMO tag using SUMO-fusion technology [Malakhov *et al.*, 2004]. SUMO proteins are reversibly attached to target proteins and promote folding and structural stability of the fusion protein, allowing recombinant expression of stable peptide sequences [Marblestone *et al.*, 2006]; [Butt *et al.*, 2005]; [Zuo *et al.*, 2005]. The His-SUMO tag can be removed using a natural SUMO protease, known as ULP1 [Li and Hochstrasser, 1999], which recognises the SUMO tag by its fold and cleaves at the C-terminal Gly-Gly motif, leaving just the required peptide [Li and Hochstrasser, 1999].

Figure 1.13: A schematic diagram of His-SUMO-Smad7 structure. The polyhistidine tag is shown in Navy, the SUMO tag in turquoise and the Smad7 peptide in green.

Figure 1.13 shows a schematic diagram of the His-SUMO-Smad7 construct. The SUMO and His tag can then be efficiently removed using naturally occurring SUMO protease, ULP1 (Ubiquitin-like-specific protease 1), which recognises the tertiary structure of SUMO and cleaves the tag, enabling the target peptide to be generated.

The interactions between a peptide of Smad7, which incorporates the PPxY motif, and the WW domains of WWP2 will be explored in this thesis using NMR and ITC.

1.4 OCT4 (Octamer binding transcription factor 4)

1.4.1 Pluripotent stem cells

Stem cells have the potential to develop into many different cell types in the body and have two defining characteristics; they are self-renewing, and they can differentiate. Stem cells undergo self-renewing cell divisions in which at least one of the daughter cells is also a stem cell. This ability to proliferate, creating identical copies of themselves, is known as self-renewal. In embryos, this process involves stem cells undergoing symmetrical cell divisions to give two daughter cells, both of which have stem cell characteristics. Stem cells can also give rise to multiple different cell types with specialised functions and typically generate an intermediate cell type before they differentiate into a specialised cell. Pluripotency is the ability of cells to undergo indefinite self-renewal and differentiate into all specialized cell lineages. This ability to differentiate makes stem cells an important tool for research into diseases as well as clinical treatments in the future [Reya *et al.*, 2001].

There are three types of stem cell; embryonic stem cells, which are grown in the laboratory from cells found only at the earliest stages of development; tissue stem cells, which are found among differentiated cells in a tissue or organ; and induced pluripotent stem cells (iPSCs), which are similar to embryonic stem cells but made from specialised adult cells [Reya *et al.*, 2001]. Embryonic stem cells are pluripotent which means that they have the potential to develop into any type of cell or tissue with the exception of those which make up a placenta or embryo. These cells are able to replicate indefinitely as well as maintain their pluripotent potential to differentiate into each of the more than 200 cell types of the adult body.

All mammals start out as totipotent stem cells which have the potential to divide. After multiple divisions, the cells begin to specialise and form a blastocyst

(hollow sphere of cells). This blastocyst is made up of an outer layer known as the trophectoderm (TE) and an inside layer called the inner cell mass (ICM). The ICM cells are pluripotent stem cells which self-renew their populations by asymmetric cell division. A stem cell is capable of dividing into two different daughter cells, one being pluripotent and the other differentiated [Hongbao and Shen, 2007]. These cells are able to give rise to all cell types of the three embryonic germ layers (ectoderm, mesoderm and endoderm) and the germ cell lineage. As these cells specify sequentially down a line of cell types, they become more limited in their differentiation ability. Therefore, by the time cells are present in tissues of the germ layers, they can only generate cells of that tissue. These cells are called multipotent stem cells.

There is still much to be discovered about what drives ESCs to differentiate into different cell types and what governs their pluripotency. Embryonic development is controlled by regulatory genes, some of which regulate the transcription of other genes [Schöler, 1991]. These regulators control the production of a protein (repressors or activators) which in turn mediate phenotypic changes during stem cell differentiation.

Four key genes are required for the induction of a pluripotent cell from a somatic cell: KLF4, SOX4, OCT4 and MYC [Kim *et al.*, 2009].

1.4.2 Octamer binding transcription factor 4 (OCT4)

OCT3/4 or OCT4, is an octamer binding transcription factor encoded by the gene POU5F1 (POU domain, class 5, transcription factor 1), and was discovered in 1990 [Okamoto *et al.*, 1990]; [Rosner *et al.*, 1990]. Transcription factors are used to control the transcription of genetic information from DNA to messenger RNA and this function can be performed by a single protein or by a combination of proteins in a complex. OCT4 is crucial for development and pluripotency [Nichols *et al.*, 1998] and has been found in embryonic stem cells but not in their differentiated daughters [Solter, 2000].

OCT4 belongs to class V of the POU transcription factor family. The acronym POU is derived from the names of the first four transcription factors in which it was found: the Pituitary-specific Pit-1, the Octamer transcription factor proteins Oct1 and Oct2 and the neural Unc-86 transcription factor (Pit-Oct-Unc) [Pan *et al.*, 2002]. POU is a family of proteins with highly conserved DNA binding homeodomains which have a common sequence of 60 amino acids. This motif is found in the majority of eukaryotic organisms which perform basic regulatory functions and is useful for studies attempting to understand specific protein-DNA interactions [Sturm *et al.*, 1988]. Transcription factors contain DNA-binding domains which are able to bind to particular sequences of DNA adjacent to the genes that they regulate. The general structure of a homeodomain can be described as 'Helix-Turn-Helix'. This DNA-binding motif is ∼20 amino acids long and is formed by two helices connected by a short turn. The second helix acts as a recognition helix which binds in a sequence-specific manner in the major groove.

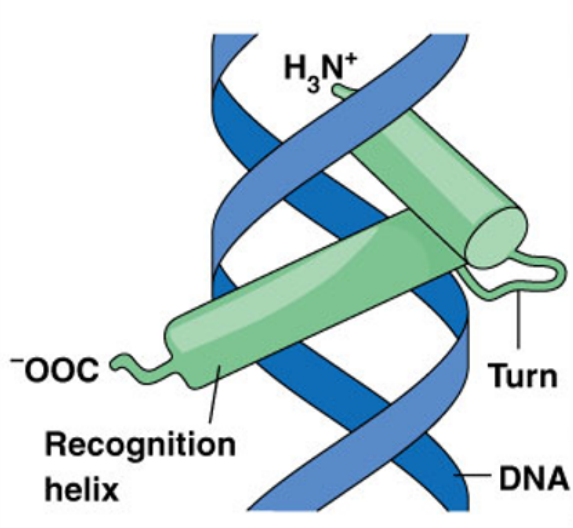


Figure 1.14: The Helix-Turn-Helix DNA binding motif contains two alpha helices (shown in green) which are joined by a short flexible turn. This motif binds DNA by recognising the sequence in the major groove. Figure taken from [Department of Biology, Memorial University of Newfoundland, 2018].

1.4.3 Pluripotency in embryo development

OCT4 is expressed in pluripotent cells, including early embryos, ESCs, ECCs, and embryonic germ (EG) cells and plays a key role in regulating the pluripotent and self-renewing state of stem cells [Nichols *et al.*, 1998]; [Wu and Schöler, 2014]. This embryonic transcription factor is expressed in ECCs and has also been detected in adult human stem cells such as bone marrow but not regular somatic cell types. This suggests that it has an important role in maintaining cellular pluripotency.

OCT4 is down-regulated during differentiation; however, it is now known that some malignant cells regain the ability to express OCT4 [Xu *et al.*, 2009]. Research into the function and regulation of OCT4 expression in mice has revealed that knocking out the OCT4 gene results in their death [Xu *et al.*, 2009]. This occurs because of the lack of inner cell mass formations which suggests that the gene is essential for the establishment of pluripotent cell lineages in early mammalian development. To sustain self-renewal of ESCs, the level of expression of OCT4 must be controlled and the precise level of OCT4 governs the fate of ESCs. Therefore, suppression of expression of OCT4 in ESCs causes differentiation into trophectoderm cells, and when OCT4 is overexpressed the differentiation is driven into mesodermal and endodermal lineages [Liao and Jin, 2010]. Total loss of OCT4 eliminates pluripotency, however interestingly, when OCT4 levels are only slightly deficient, pluripotency is improved in ESCs. It has been suggested that OCT4 is a "gatekeeper into and out of the reprogramming expressway that can be directed by altering the experimental conditions" [Sterneckert *et al.*, 2012].

1.4.4 OCT4, a target for WWP2

E3 ubiquitin ligase WWP2, which has been found to have tumorigenic and metastatic roles, is reported to specifically interact with the stem cell specific transcription factor OCT4 [Xu *et al.*, 2009], often up-regulated in cancer stem

cells [Hu *et al.*, 2010]. GST pull-down experiments have shown that the C2 and HECT domains of WWP2 are not capable of binding to OCT4 as opposed to the tryptophan WW domain region which showed a clear interaction [Xu *et al.*, 2004]. This interaction indicates that it is the WW domains of WWP2 which interact with the PPxY motif of OCT4.

In 2009, Xu *et. al* discovered that the second WW domain (WW2) was found to interact with the highest affinity of the four domains [Xu *et al.*, 2009], marking WW2 as an attractive therapeutic target for anticancer drugs.

WWP2 has the ability to enhance ubiquitin modification of OCT4 and is effective *in vitro* and *in vivo* [Xu *et al.*, 2004]; [Xu *et al.*, 2009]. The study by Xu *et. al* suggested that post-translational modification by ubiquitin may play a role in maintaining the OCT4 protein level in embryonic stem cells and was found to enhance degradation of OCT4 through the 26S proteasome in a dosage-dependent manner [Xu *et al.*, 2009]. WWP2 was also able to regulate its own ligase activity by auto-ubiquitination [Liao and Jin, 2010].

Previous research has marked WW2 as an attractive therapeutic target for anticancer drugs. An additional aim of this study was to identify whether the WW2 region binds with a higher affinity to the target substrate OCT4 than the tandem WW2-3 region.

1.4.5 Structure of OCT4

The octamer-binding subgroup of the POU transcription factor family binds the octamer sequence motif AGTCAAAT, located in the promoter or enhancer regions of its target genes, regulating their expression [Pan *et al.*, 2002]. The DNA binding domain is located in the centre of the protein sequence, between the N-terminal and C-terminal transactivation domains [Ambrosetti *et al.*, 2000].

The structure of the POU family consists of two subunits, comprised of a POU-specific (POU^S) domain with four α-helices and a POU homeodomain (POU^H)

with three α-helices. The carboxy-terminal (POU^H) subunit is a homeobox domain containing 60 amino acid residues and the larger amino-terminal subunit (POU^S) is a POU-specific domain with 75 amino acid residues.

The human OCT4 gene can generate at least three transcripts (OCT4A, OCTB and OCT4B1). OCT4A is of importance due to its role in embryonic stem cell self-renewal and pluripotency as opposed to the latter two which lack the ability to maintain ESC self-renewal and are thought to have a different role to OCT4A. The structure of OCT4A (shown in figure 1.15) illustrates the N-terminal N-transactivation domain (NTD^A), the POU-specific (POU^S) domain, the flexible linker region joining the POU^S to the POU homeobox domain (POU^H) and finally the C-terminal transactivation domain (CTD). The PPxY motif is located in the N-transactivation domain (NTD^A) which is predicted to be an intrinsically unfolded region and is highlighted in figure 1.15 .

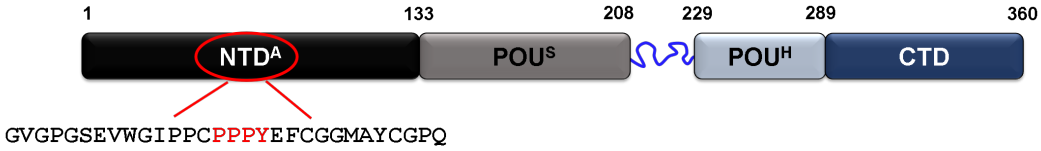


Figure 1.15: A schematic diagram of OCT4 structure. The N-transactivation domain (NTD^A) is shown in black, the POU-specific region in grey, the POU homeodomain in light blue, the C-transactivation domain in Navy and the flexible linker region in blue. The PPxY motif is located in the NTD^A domain and the sequence of the OCT4 peptide is shown below.

The POU domains bind to DNA using the 'Helix-Turn-Helix' conformation [Pan *et al.*, 2002]. Figure 1.16 shows the crystal structure of OCT4-1 POU homeodomain and POU-specific domain forming a complex around DNA. The POU^S domain interacts with the ATGC sequence in the 5' half of the octamer site with low affinity and is separated from the (POU^H) domain, which binds the AAAT sequence [Klemm and Pabo, 1996] by a non-conserved region of 15-55 amino acids, the linker region. The POU domains interact with the major groove of the DNA and bind to opposite sides of the double helix [Klemm and Pabo, 1996].

The linker region of different OCT proteins is variable in length and sequence

[Sturm *et al.*, 1988]. The crystal structure was also available for the POU subdomains of the OCT6 protein [Jauch *et al.*, 2011] but for both OCT1 and OCT6 the linker regions were not visible, suggesting they were flexible and unstructured. In contrast, the OCT4 POU domain was crystallised when bound to the PORE DNA-binding motif and the N-terminal of the linker was visible showing an α helix structure [Esch *et al.*, 2013]. This conserved region, only present in OCT4 [Esch *et al.*, 2013], indicates that this group has a distinct functional role.

Figure 1.16: 1GT0 PDB structure [Reményi *et al.*, 2003] showing the crystal structure of OCT4-1 POU homeodomain and OCT4-1 POU-specific domain forming a complex with DNA. Residues which were not visible in the electron density map are shown as a dotted line. Figure exported from Chimera [Pettersen *et al.*, 2004].

Unless otherwise specified, where OCT4 is mentioned in this thesis, it is referring to a PPxY motif containing peptide of OCT4A as a His-SUMO fusion protein (figure 1.17). The His-SUMO fusion system is described in section 1.3.5.

Figure 1.17: Schematic diagram of OCT4A structure. The polyhistidine tag is shown in Navy, the SUMO tag in turquoise and the OCT4 peptide in blue.

1.5 Methods to investigate protein structure and protein-protein interactions

The structure of proteins can be determined in several different ways, commonly by X-ray crystallography or NMR spectroscopy. These techniques provide experimental data regarding the structure of the molecule.

The majority of structures in the Protein Data Bank (PDB) are determined from the X-ray diffraction of protein crystals, obtained using crystallography. This technique uses a crystallised form of a purified protein, which is placed in an intense beam of X-rays (usually monochromatic) and the proteins cause the X-ray beam to be diffracted to produce a pattern of reflections. Data is collected with the crystal in numerous different orientations, allowing intensity data for each spot to be recorded for each different placement [Drenth, 2007]. This data is analysed to determine the distribution of electrons in the protein and so the location of each atom can be acquired from the resulting map of electron density. X-ray crystallography is an impressive method able to deliver comprehensive atomic information. However, to crystallize a protein is challenging and can limit the types of proteins available for analysis. The technique works well when resolving the structures of inflexible proteins due to the fact that the molecules are aligned in the same precise orientation and hence are able to produce ordered crystals [Drenth, 2007]. In the case of certain proteins, only a single domain may

crystallise and so the structure of the protein as a whole cannot be determined. Highly flexible proteins cannot be observed using this method. Once conditions for crystallising an ordered protein are optimised, protein samples of high purity and stability are required for crystallography, as contaminants can disrupt the structure or resolution acquired. Due to previous issues encountered by the laboratory group when attempting to crystallise domains of WWP2 [Wahl, 2016], as well difficulties documented by Jiang and Zheng [Jiang *et al.*, 2015], it was determined that this method would likely be problematic and time-consuming. Instead, other methods for structure determination were explored.

Another commonly used method for determining protein structure is Nuclear Magnetic Resonance (NMR) spectroscopy. This method is advantageous as it allows the three-dimensional structures of proteins to be obtained whilst in the solution phase [Cavanagh *et al.*, 2010], therefore information for flexible proteins can still be collected provided that peak dispersion allows individual peaks to be revealed. It is also a useful technique for investigating kinetic reactions and the atomic properties of proteins. A sample of purified protein is placed in a strong magnetic field, and a sequence of electromagnetic pulses are employed. The resonances detected can be analysed to show which nuclei are physically close to one another as well as characterise the conformation of molecules. This data can be used to create a protein model with the locations of the atoms. Proteins up to around 25 kDa can be investigated using NMR, whereas larger proteins can cause difficulties due to peak overcrowding in the spectra [Sarkar, 1996] and peak broadening due to faster decay of the NMR signal as a result of the molecular tumbling. NMR samples must be of high purity and concentration so that signal can be detected with shorter experiment times and spectra are not altered by artefacts or contamination. Stability is a desired attribute of a protein sample as triple resonance experiments can take several days each to acquire enough signal for assignment and these samples are costly due to the requirement of isotopic labelling.

While X-ray crystallography gives a static picture of the structure of proteins, NMR has the ability to provide dynamic information about the target protein in

solution. This thesis focuses on the use of NMR to study protein structure and interactions. The theory and methods used are described in more detail below.

1.6 Nuclear Magnetic Resonance (NMR) Spectroscopy

1.6.1 NMR theory

Nuclear magnetic resonance (NMR) was first reported in 1946 by Bloch [1946] and by Purcell [1946] and since then its applications to science have been continuously expanding. Spectroscopy is the study of the interaction of electromagnetic radiation with matter. NMR spectroscopy uses the magnetic properties of certain atomic nuclei [Rule and Hitchens, 2006] and records the absorption of energy between quantised energy levels [Keeler, 2011] to study physical, chemical, and biological properties of matter. Once the covalent structure of a compound is known, it is possible to use a variety of NMR techniques to determine the conformation when in solution.

The foundation of all NMR experiments relies on a mechanical property of the nucleus – the angular momentum of the spinning nuclear charge, known as the nuclear spin quantum number (I) [Cavanagh *et al.*, 2010]. Protons, electrons, and neutrons all possess spin and each individual unpaired electron, proton, and neutron has a spin of $\frac{1}{2}$ [Levitt, 2013]. The most useful nuclei for NMR are those with spin $I{=}\frac{1}{2}$, as these nuclei have a spherical charge distribution with magnetic properties that make them NMR active. Nuclei with $I{>}\frac{1}{2}$ have a non-zero magnetic moment and are therefore also NMR active; however, they also have an electric quadrupole moment (Q) which results in very broad spectral peaks and so data from these nuclei are more difficult to interpret. ^{14}N has a spin of $I{=}1$ and so is a quadrupolar nucleus which experiences peak broadening [Kwan *et al.*, 2011]. Nuclei with $I{=}0$ are not NMR active, for example naturally abundant ^{12}C, and will not be able to interact with the magnetic field.

The commonly used nuclei in protein NMR are ^{1}H, ^{15}N and ^{13}C, due to their spin $I = \frac{1}{2}$. ^{1}H is naturally abundant; however, ^{15}N and ^{13}C isotopes only account for such small percentages of the naturally abundant forms of these elements. Therefore, recombinantly expressed proteins for study by NMR must be isotopically labelled with magnetically active isotopes by expressing in an environment with only those isotopes available [Keeler, 2011]; [Levitt, 2013]. This labelling is an expensive process and so only the nuclei required for the sample are labelled, for example, an NMR titration using [^{1}H, ^{15}N]-HSQC spectra will only require ^{15}N isotopic labelling.

Nuclei with spin I are (2 I + 1) fold degenerate in the absence of a magnetic field, for example, a hydrogen nucleus ($I = \frac{1}{2}$) has two possible spin states. NMR spectrometers contain extremely strong magnets of between 1 and 20 Tesla (T). When a nucleus is placed inside the NMR spectrometer, a magnetic field is generated around and within the sample and nuclei with nuclear angular momentum, such as ^{1}H, ^{15}N and ^{13}C, will orient in one of multiple different arrangements depending on the nuclei's magnetic moment. The effect on the alignment of the nuclei due to the magnetic field is known as the Zeeman Effect. NMR is the spectroscopy of the nuclear Zeeman sublevels [Levitt, 2013].

The energy difference between the energy states increases as the strength of the magnetic field increases [Kwan *et al.*, 2011]. Therefore, large magnetic fields are usually employed to increase the splitting energy and the initial population difference between the two spin states, consequently increasing net movement between states and improving sensitivity. Even with these strong magnetic fields, the energy separation is small and so radiofrequency radiation can be used to excite the nucleus and 'flip' its alignment with the field [Kwan *et al.*, 2011]. The difference in energy between the low energy and high energy arrangements of the nuclear spin is measured and resonant absorption will occur when the electromagnetic radiation of the correct frequency exactly matches the energy separation [Jacobsen, 2007].

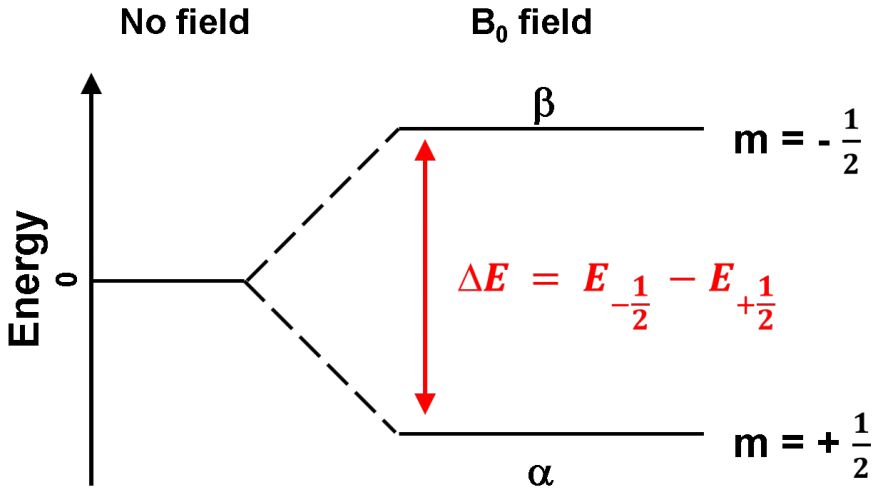


Figure 1.18: An energy diagram showing the Zeeman splitting of spin energy states for a nucleus with spin I = $\frac{1}{2}$ as an external magnetic field is applied.

Figure 1.18 shows the nuclear Zeeman levels of ^{1}H nuclei with I=$\frac{1}{2}$ against the applied magnetic field. As mentioned above, for hydrogen nuclei there are two possible spin states, these are the high energy state (β) and the low energy state (α) [Kwan *et al.*, 2011]. In the absence of a magnetic field there is no net magnetisation, instead, the orientations of the nuclei are random (figure 1.19A). When a magnetic field is applied, the degeneracy is lifted, and nuclear spins align parallel with the field (figure 1.19B) [Keeler, 2011]. The energy difference between these spin states is known as the 'splitting energy' (ΔE) and is a function of the external magnetic field B_0; [Akitt and Mann, 2000].

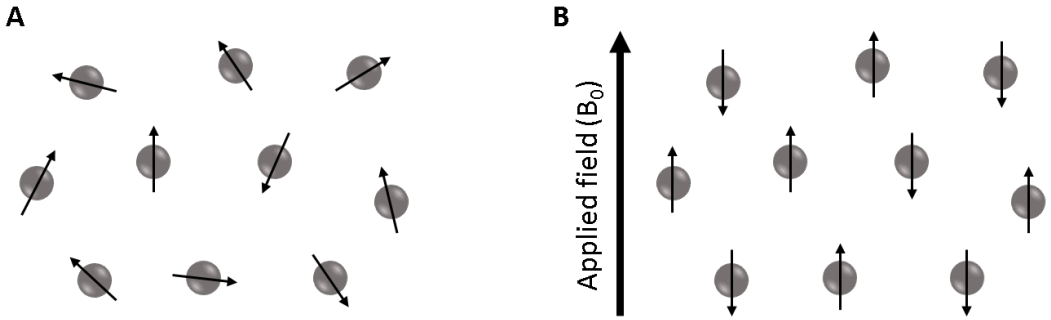


Figure 1.19: The distribution of nuclear spins A) without an applied magnetic field – hence randomly orientated, and B) after a magnetic field is applied – nuclear spins align themselves parallel to the field.

After the magnetic field is applied, the alignment of the nuclei is dependent on the temperature and the Boltzmann distribution. This means that whilst nuclei will be aligned in both the high and low states, they have a slight preference to align with the field at room temperature rather than against it as it takes energy to oppose the field. When all of the atoms in a sample are observed, there is a bulk magnetisation vector (a sum of the individual magnetic moments) which aligns with the magnetic field in the low energy state [Keeler, 2011]. The Boltzmann distribution equation is displayed in equation 1.2

$$\frac{N_{\text{Upper}}}{N_{\text{Lower}}} = e^{\frac{-\Delta E}{kT}} = e^{\frac{-h\upsilon}{kT}} \tag{1.2}$$

where N_{Upper} and N_{Lower} are the relative proportions of nuclei in the upper and lower spins states and k is the Boltzmann constant (1.3805×10^{-23} J K^{-1}). ^{1}H is the most commonly studied nuclei and contains a single proton, hence $I = \frac{1}{2}$ and so is, therefore, able to behave as if the positively charged nucleus was spinning on its axis [Rule and Hitchens, 2006].

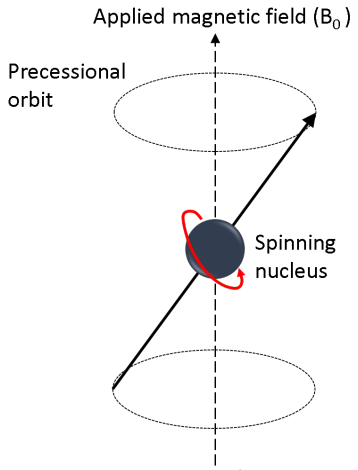


Figure 1.20: Larmor precession - Spin $\frac{1}{2}$ nucleus spinning on its axis.

The spinning charge is able to create a tiny magnetic field [Jacobsen, 2007] and when positioned within a strong external magnetic field, the magnetic nucleus

tries to align with it [Kwan *et al.*, 2011]. The spinning nucleus also has angular momentum and the torque exerted by the external magnetic field causes a circular motion called precession [Jacobsen, 2007] as shown in figure 1.20.

The rate of this precession is proportional to the strength of the external magnetic field and the nuclear magnetic strength (equation 1.3):

$$v_0 = \frac{\gamma B_0}{2\pi} \tag{1.3}$$

This equation is the Larmor Equation where v_0 is the Larmor frequency (Hz) and γ is the gyromagnetic ratio (Hz/T), which describes the ratio of mechanic and magnetic properties of the nucleus (the nuclear magnetic strength) and depends on the type of nucleus. B_0 is the strength of the external magnetic field in Tesla (T).

When a radio frequency field (B_1) is applied, the individual spins transition from the low energy state to the high energy state [Keeler, 2011] and this is known as a pulse. Before the pulse is applied, the position of the bulk magnetisation vector is along the z-axis.

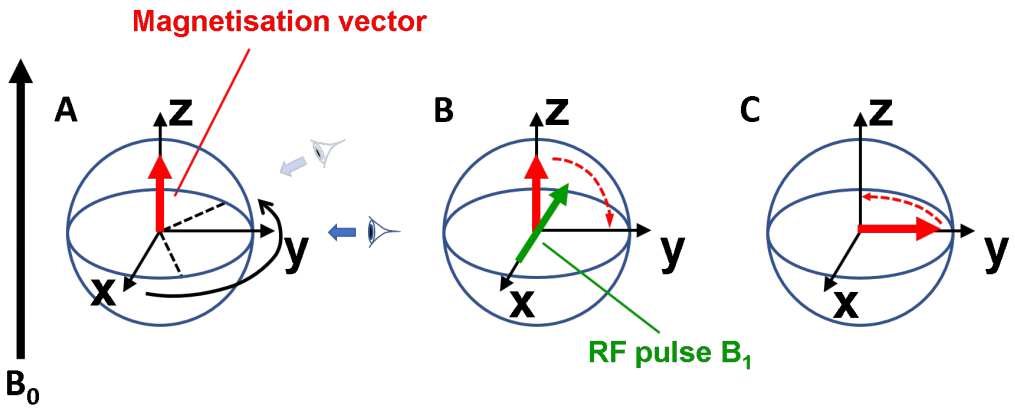


Figure 1.21: **(A)** Position of the bulk magnetisation vector (in red) along the z-axis, displayed in the rotating frame (rotating at or near the Larmor frequency) **(B)** The effect of a 90°x Radiofrequency (RF) pulse (B_1). The magnetisation vector is rotated to the y-axis. **(C)** The bulk magnetisation then rotates in the xy plane. Adapted from [Keeler, 2011].

Figure 1.21A displays the vector in the rotating frame, where the xy plane

rotates about the z-axis (rotating at or near the Larmor frequency). The rotating frame is used to simplify the complex motions before, during and after an RF pulse, by effectively removing the effect of B_0.

If a 90° x Radiofrequency (RF) pulse (B_1) is applied to the sample, the bulk magnetisation is rotated onto the y-axis, as illustrated in figure 1.21B, before it starts rotating about the z-axis in the xy plane (figure 1.21C) [Keeler, 2011]. This causes the magnetisation to oscillate in the x and y axes as a function of evolution time and produces sine and cosine waves (figure 1.22).

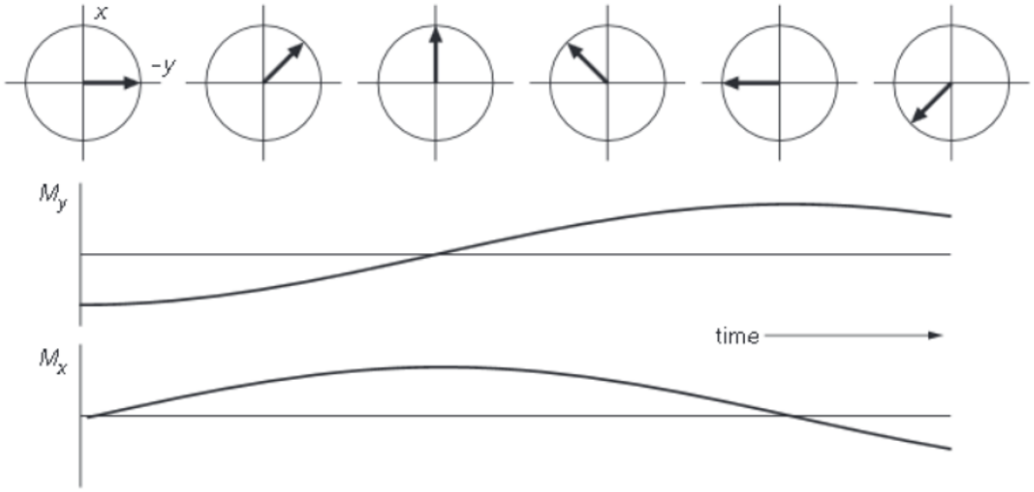


Figure 1.22: The bulk magnetisation as it rotates around the xy axis, from the z-axis orientation, displayed as a function of evolution time [Keeler, 2011].

Despite individual NMR active nuclei having the same magnetic moment as each other, for example, protons all have the same magnetic moment, each can give rise to slightly different frequencies. This is due to the effective magnetic field experienced by the nuclei being the sum of the large magnetic field (B_0) and the smaller 'shielding' or 'de-shielding' magnetic fields [Levitt, 2013]. The 'shielding' effect of the surrounding electrons is able to create its own small magnetic field which reduces the magnetic field experienced at the nucleus. This causes the energy gap to be reduced; therefore, the resonance frequency is also reduced [Rule and Hitchens, 2006]. This means that the local chemical environment of a nucleus influences the frequency it experiences, making it distinctive from nuclei in varying

environments. This shift of the NMR frequency due to the chemical environment is called the chemical shift, and it explains why NMR is a direct probe of chemical structure. Different atoms within a molecule can be identified by their chemical shift in parts per million (ppm), described in more detail below. The intensity of the NMR signal is also important as the areas of the peaks are proportional to the size of the population of nuclei producing the same resonance frequency [Jacobsen, 2007].

1.6.1.1 Chemical shift

The precise frequency of precession is dependent on the effective magnetic field at the nucleus, and therefore the resonance will change between spectrometers with different field strengths. Due to the linear relationship between field strength and resonance frequency [Keeler, 2011], it would then not be possible to directly compare spectra using the Hz scale. Instead, a normalised scale, known as the parts per million (ppm) δ scale is used, which expresses the frequency independent to the spectrometer frequency (see equation 1.4).

$$\delta(ppm) = \frac{v_{sample} - v_{ref}}{v_{ref}} \text{ x } 10^6 \tag{1.4}$$

where δ is the chemical shift in ppm, v_{sample} is the observed chemical shift (in Hz) and v_{ref} is the observed chemical shift of a reference compound.

The ppm scale allows spectra acquired for one sample at two different frequencies, for example, 500 MHz and 800 MHz, to be overlaid and resonances will appear at the same ppm. By using a higher frequency, the spectral resolution is improved as the resonance frequencies will be more spread out, for example, 0.1 ppm corresponds to 50 MHz or 80 MHz when acquired in a 500 MHz or 800 MHz spectrometer respectively. The sensitivity of the experiment is also improved at higher field strengths due to an increase in the energy gap between low and high energy states. This can be explained as a larger energy gap leads to more nuclei being present in the lower energy state,

producing a stronger bulk magnetisation.

1.6.1.2 One and two-dimensional NMR techniques

One-dimensional (1D) NMR of proteins produces spectra which are too complicated to interpret. This is due to heavily overlapping signals, and by introducing additional dimensions to the spectra, these spectral overlaps can be resolved. Two-dimensional (2D) NMR spectroscopy uses a simple series of 1D experiments collected with different timings. They have the same initial and final steps as for 1D experiments; preparation and detection, as shown in figure 1.23A. In addition to these steps, 2D experiments (figure 1.23B) have an indirect evolution time t_1 and a mixing sequence.

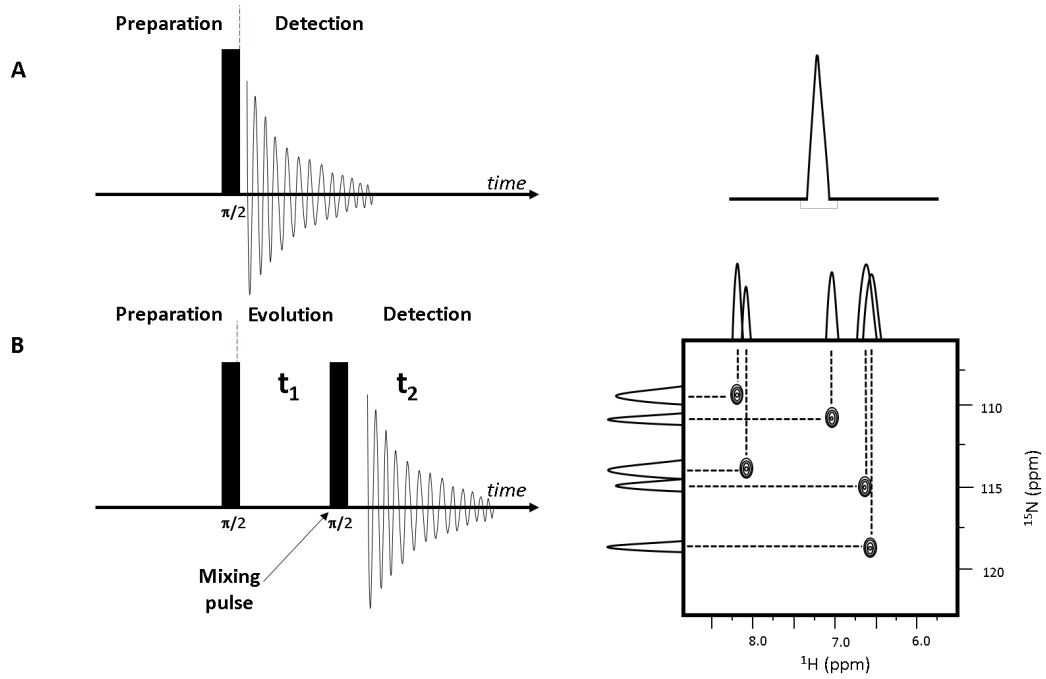


Figure 1.23: Basic NMR schemes **(A)** One-dimensional NMR pulse sequence showing the FID over a single time dimension and the resulting spectrum after Fourier Transform (FT). **(B)** Two-dimensional NMR pulse sequence with a variable time slot t_1, and an example [^{1}H, ^{15}N]-HSQC spectrum after FT and processing.

The initial step is where the sample equilibrates after being introduced to

the B_0 field, and the bulk magnetisation vector moves to the z axis. Next, in the preparation stage, one or more RF pulses, close to the resonant frequency of that nucleus, are applied to excite the nuclei into the higher energy state. This is typically a 90° pulse and generates transverse magnetisation, shown above in figure 1.21. The evolution time (t_1) then allows magnetisation to evolve before further pulses (mixing pulses) are applied. Finally, during the detection stage, the signal is recorded as an oscillating signal generated by the relaxation of the nuclei to their ground state as a function of t_2 and is known as the free induction decay (FID).

The pulse sequences are then repeated with increasing values of t_1 and after each, FIDs are recorded as functions of t_2 [Croasmun, 1994]. The FIDs at each t_1 value are then added together. Whilst this increases the time taken to acquire the experiment, the signal-to-noise (S/N) ratio [Bruch, 1996] and hence the sensitivity is increased. This is because noise in the spectrum appears randomly, as opposed to signals from the sample which occur in the same position each time; therefore, when added together the signal becomes much stronger than the noise. This is known as time averaging [Bruch, 1996].

The resultant free-induction decay (FID) is not in an analysable form. Modern spectrometers operate in the pulsed Fourier Transform (FT) mode which allows the entire spectrum to be excited at the same time and can be recorded quickly rather than having to sweep through different frequencies [Jacobsen, 2007]. The signal is generated by the radiofrequency pulse and is picked up by a receiver coil as a decaying oscillation at the resonance frequency of the spin, w. The 'phase' must then be set electronically by using the correct phase for the reference frequency w_0. A Fourier transform is a mathematical operation which converts the signal from being as a function of time to a function of frequency [Akitt and Mann, 2000].

In general, 2D NMR experiments can be separated into two groups, homonuclear and heteronuclear. Both of these groups can present coupling information either through bond (for instance COSY) or through space (NOESY). Fourier transform (FT) is applied in both dimensions, t_1 and t_2, allowing a 2D frequency correlation map to be produced. By changing the

preparation time and pulses applied it is possible to record many different types of 2D NMR spectra.

The example [^{1}H, ^{15}N]-HSQC shown in figure 1.23B, demonstrates how a crowded region in a 1D spectrum (shown above and to the left of the spectrum) can be more clearly distinguished through the process of adding a second heteronuclear ^{15}N dimension.

1.6.1.3 The HSQC spectrum

The HSQC (Heteronuclear Single Quantum Coherence) spectrum is the most widely used 2D NMR experiment for proteins and shows the frequency correlation of directly bound nuclei, commonly H-N or H-C. Depending on the desired correlation, samples are isotopically labelled with spin $I = \frac{1}{2}$ nuclei, ^{13}C or ^{15}N as described above.

The [^{1}H, ^{15}N]-HSQC spectrum is typically one of the first multidimensional NMR experiments performed on a protein sample and is a useful tool for determining whether subsequent experiments would be effective. ^{1}H-^{15}N correlations are particularly useful as the resulting spectrum displays a single peak for each individual amide proton in the protein backbone, with the exception of proline residues which lack this amide. A peak is therefore present for all other amino acid residues, as well as additional peaks for each residue containing an NH or NH_2 group in their sidechain (glutamine, asparagine and tryptophan), which occur due to the proton coupling at the shared nitrogen atom.

The [^{1}H, ^{15}N]-HSQC is able to observe the heteronuclear correlations in a relatively short amount of time and so can quickly provide useful information about the protein sample. Each protein will typically display a distinctive pattern of amide peaks, allowing prompt identification of future samples, differing conditions, as well as sample stability throughout experiment sets and the signal to noise ratio can determine if the protein sample is of sufficiently high

concentration. The peak dispersion can also indicate the degree of folding for a sample. Amino acid residues with no secondary structure exhibit typical chemical shifts known as "random-coil" chemical shifts [Wishart *et al.*, 1995]. Therefore, folded proteins tend to give well-dispersed peaks, shifted away from the random-coil values which overlap within the centre of the spectrum.

The HSQC takes advantage of the higher gyromagnetic ratio of ^{1}H, 10 times that of ^{15}N. This causes the nitrogen signal to be weaker since the bulk magnetisation vector is smaller. To overcome this, the magnetisation is transferred from the ^{1}H nucleus to the ^{15}N nucleus, where it evolves, before being transferred back to ^{1}H for detection.

The [^{1}H, ^{15}N]-HSQC pulse sequence, shown in figure 1.24, is a basic example of a 2D HSQC experiment. During the preparation stage (A), a spin echo sequence transfers magnetisation from the abundant proton (^{1}H) to the chemically bonded ^{15}N. The evolution time (B) sees a 180° pulse applied only to the ^{1}H, which refocuses the offset of this spin and the evolution of the coupling. The next stage (C) is the mixing time, where magnetisation is transferred from the ^{15}N spin back to the ^{1}H where it is detected [Keeler, 2011]. Fourier transform is then applied to the acquired FIDs to yield frequencies.

Multiple repetitions of this pulse sequence are completed and the t_1 evolution time is increased each time. This creates a series of data points which can be Fourier transformed to yield frequencies. The HSQC spectrum is then displayed as a 2D contour plot, where each peak can be assigned to an amino acid residue.

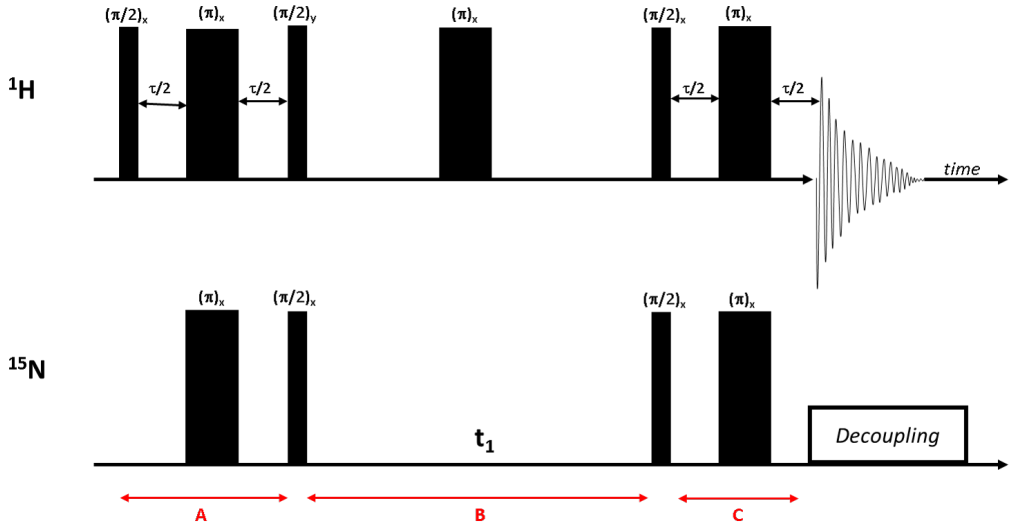


Figure 1.24: A basic 2D [^{1}H, ^{15}N]-HSQC pulse sequence, highlighting three regions. **A** preparation, **B** evolution and **C** mixing time. Adapted from *Understanding NMR Spectroscopy*, James Keeler [Keeler, 2011].

1.6.1.4 Triple resonance experiments and backbone assignment

Triple resonance spectra are generally three-dimensional (3D) experiments which pulse 3 different nuclei and are useful in protein NMR as they permit protein backbone assignment as well as aliphatic or aromatic sidechain atom assignment. Protein samples are typically labelled using naturally abundant ^{1}H and isotopically labelled ^{13}C and ^{15}N. The need for double labelled proteins is an apparent drawback, as they are expensive to prepare [Choudhary *et al.*, 2015]. They do however offer higher levels of peak dispersion which reduces the problem of overlapping peaks [Rule and Hitchens, 2006].

The pulse sequence of a three-dimensional NMR experiment (figure 1.25A) adopts an additional evolution time (t_2) and mixing time, to the 2D NMR pulse sequence shown in 1.23A.

The nomenclature for triple resonance experiments is very descriptive and the nuclei used for magnetisation are listed in order of their use. Nuclei in brackets are representative of nuclei which are not detected in the spectrum but are only

for transfer. For example, in the HN(CA)CO experiment (figure 1.25B), the magnetisation is transferred from ^{1}H to ^{15}N and then via the N-Cα J-coupling (bond) to the ^{13}Cα. It is then transferred to the ^{13}CO via the ^{13}Cα-^{13}CO J-coupling [Clubb *et al.*, 1992]. It is finally detected as the magnetisation is transferred back the same way, ending up back to ^{1}H. The chemical shift is only evolved on ^{1}H, ^{15}N and ^{13}CO and not on the ^{13}Cα, hence why CA is in brackets [Clubb *et al.*, 1992] in the experiment name. This experiment results in a three-dimensional spectrum.

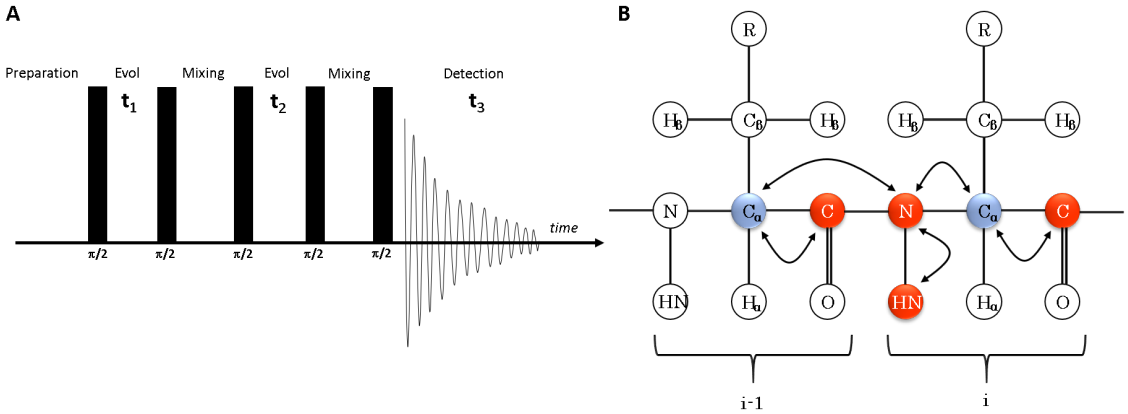


Figure 1.25: **(A)**. Simplified three-dimensional NMR pulse sequence showing a second evolution time, t_2. **B**. Schematic diagram of an HN(CA)CO NMR experiment showing the transfer of magnetisation from ^{1}H to ^{15}N, and then through the $^{13}C_{\alpha}$ to the ^{13}CO. It is then transferred back via the same route to the ^{1}H for detection. (Adapted from www.Protein-nmr.org).

For backbone assignment, HNCO, HNCACO, HNCACB and CBCACONH experiments can be used [Cavanagh *et al.*, 2010]. The magnetisation transfer path for each of these experiments is illustrated in figure 1.26.

As mentioned above, triple resonance experiments can provide information on the sequential links between residues. When a chain of sequential assignments has been obtained, it can be located within the known protein sequence using characteristic amino acid chemical shift values. Two of the key three-dimensional experiments used in backbone assignment are the HNCACB and CBCA(CO)NH. These experiments can be used to correlate the Cα and Cβ resonances with the corresponding amide resonances.

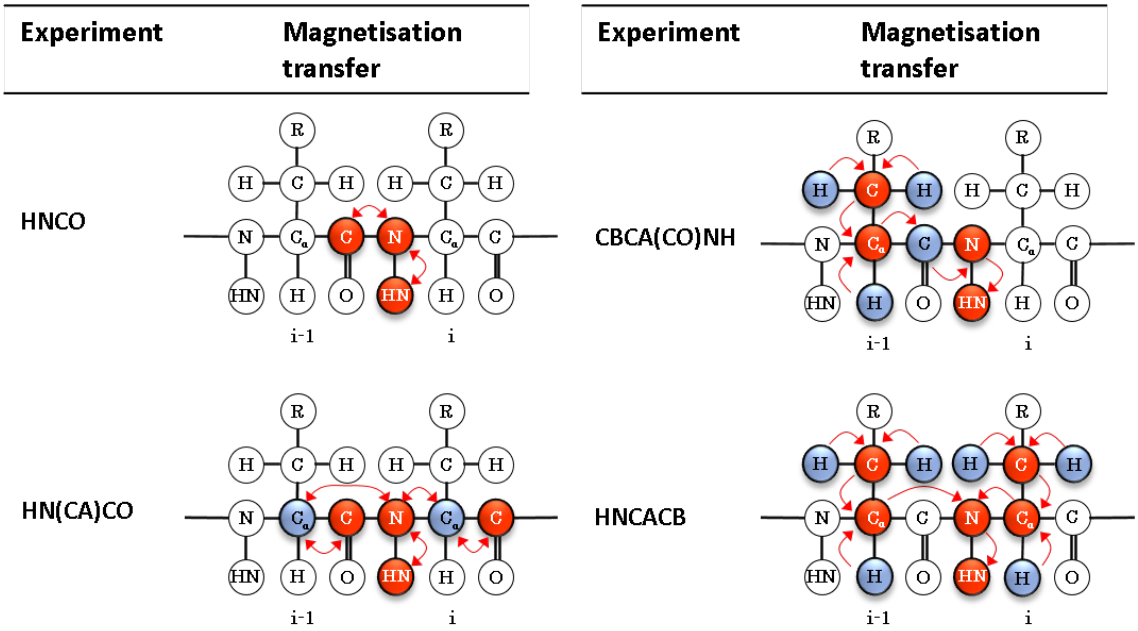


Figure 1.26: The coherence transfer for triple resonance experiments used in protein backbone assignment. Detected nuclei are shown in red and if magnetisation was transferred via a nucleus it is shown in blue. Red arrows show the path of magnetisation. (Adapted from www.Protein-nmr.org).

The HNCACB is a three dimensional spectrum that correlates the ^{1}H and ^{15}N resonances of a backbone amide residue i to the intra-residue α and β carbon resonances (i), as well as the α and β carbons of the previous residue (i-*1*). The i-*1* Cα and Cβ resonances generally appear as less intense peaks when compared to i residue peaks and so when aligned with each other, the residues can be matched and sequentially assigned.

The CBCA(CO)NH experiment can further aid backbone assignment as it shows only α and β carbons from the previous residue amide (the i-*1* residue). An example is shown in figure 1.27.

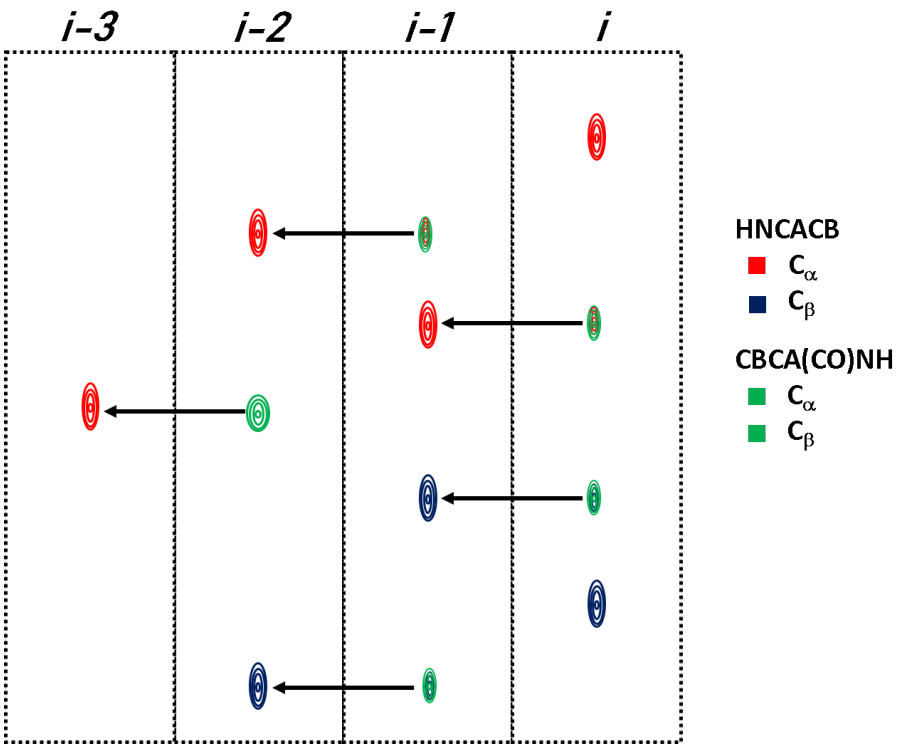


Figure 1.27: A schematic diagram showing the sequential backbone assignment using HNCACB and CBCA(CO)NH NMR spectra. HNCACB spectra show the Cα (red peaks) and Cβ (blue peaks) nuclei of the *i* residue and weaker peaks showing the *i-1* residue Cα and Cβ. The CBCA(CO)NH experiment observes Cα and Cβ peaks from the *i-1* residue (green peaks).

The HNCACB spectrum provides chemical shift information for the Cα and Cβ of the *i* and previous residue *i-1*. Each strip should have up to four peaks, two from the *i* residue and two from the *i-1* residue. In this experiment, the two peaks have opposite signs, for example, in this case Cα is shown in red and Cβ in blue. The CBCACONH spectrum peaks in figure 1.27 are shown in green and each strip contains up to two peaks, the Cα atom and the Cβ atom of the *i-1* amide. Two *i-1* peaks, corresponding to the Cα and Cβ in the *i* residue, can be matched to the *i* peaks of the *i-1* residue.

Weaker *i-1* HNCACB peaks overlap with the stronger CBCA(CO)NH peaks. This can be especially useful when the *i-1* peaks are so weak that they are hidden by the noise, for example, in figure 1.27 in the *i-2* strip, the previous residue (glycine) peak is not present in the HNCACB and so the CBCA(CO)NH peak can

be used to set the sequential assignment instead. They are also useful when both the *i* and *i-1* HNCACB peaks have the same intensity.

When a chain of sequential assignments has been set, the residues can be assigned by comparing the chemical shifts of the resonances with a database showing the most common range of shifts for a particular amino acid. Residues which have distinctive shifts, for example a glycine residue has only one Cα peak and no Cβ (shown in strip *i-3* of figure 1.27). Other experiments, such as HNCACO and HNCO can be used to assist with the sequential backbone assignment, using the CO peaks of the *i* and *i-1* residues.

Examples of GB1-WW2:Smad7 spectra, CBCA(CO)NH and HNCACB are shown in figure 1.28. Each strip is taken from a different nitrogen plane (figure exported from CCPN Analysis [Vranken *et al.*, 2005]). As explained above, the HNCACB shows correlations between a backbone amide residue *i* and the α and β carbon resonances of the *i* and (*i-1*) residues. In this experiment, magnetisation is transferred between Cα and Cβ in such a way that the two peaks have opposite signs, shown here as Cα in blue and Cβ in green. In figure 1.28, the CBCACONH spectrum, shown in brown, shows that in each strip of the spectrum are two peaks, the Cα atom and the Cβ atom of the previous amide *i-1* residue. Two peaks which correspond to the Cα and Cβ peaks are therefore visible for each residue and overlap the weaker HNCACB peaks, as shown on the right-hand side of figure 1.28.

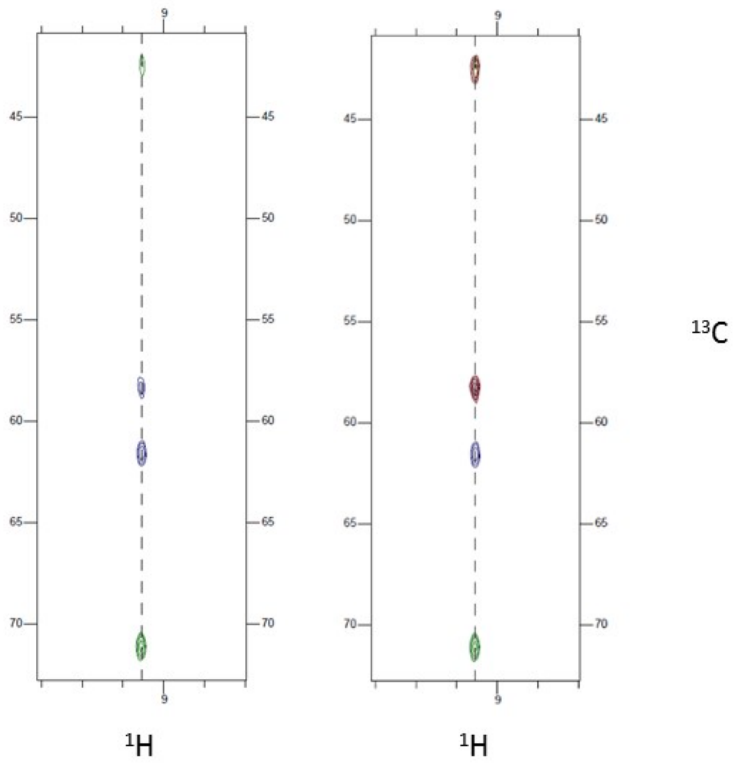


Figure 1.28: (**Left**) GB1-WW2 HNCACB. Green peaks are Cβ, blue peaks are Cα. (**Right**) GB1-WW2 HNCACB overlaid with CB(CA)CONH (brown). The CB(CA)CONH peaks in brown are shown overlapping the weaker *i-1* residues of the HNCACB (the top two peaks on the left image). Exported from CCPN Analysis [Vranken *et al.*, 2005].

Figure 1.29 shows the sequential peaks from residue 56 (CBCACONH) in brown and 55 (HNCACB) in green and blue. The chemical shifts of the i peaks in the HNCACB spectra (of residue Tyr55) match the *i-1* peaks (of residue Thr56) in the CBCACONH spectra and so a sequential link can be set.

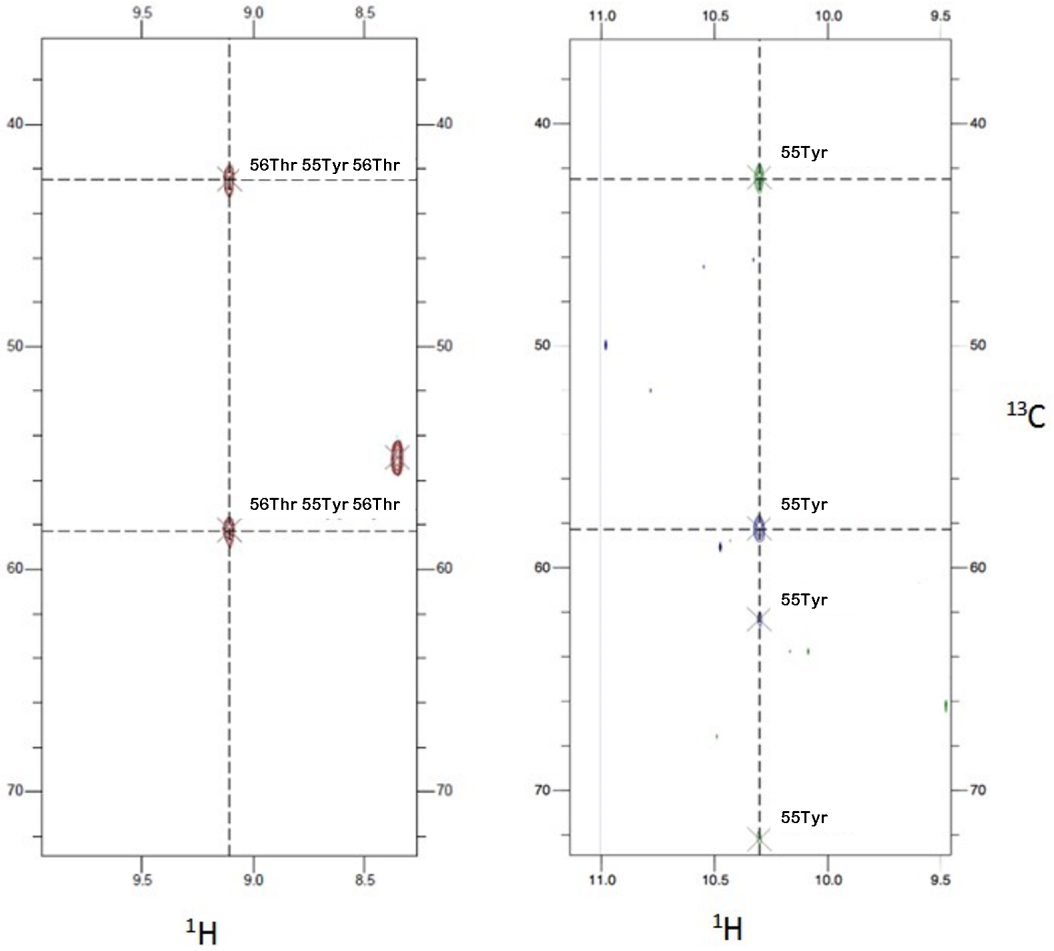


Figure 1.29: ^{1}H, ^{13}C GB1-WW2 *i* and *i-1* HNCACB (shown in green and blue) and *i-1* CBCACONH (shown in brown). The *i-1* peaks of residue Thr56 in the CBCACONH match to the *i* peak of residue Tyr55 in the HNCACB spectrum. Exported from CCPN Analysis.

CC(CO)NH, H(CCO)NH, H-CCH-TOCSY and HCCH-TOCSY spectra were used along with a [^{1}H, ^{13}C]-HSQC to assign aliphatic side chain ^{1}H and ^{13}C resonances. The aromatic ^{13}C-TOCSY and aromatic ^{13}C-TROSY were used to assign aromatic side chain atoms.

1.6.1.5 Conformational Exchange

Based on a number of factors including flexibility, pH, temperature or ligand availability, the chemical environment of a nucleus can change. As described previously, different chemical environments correspond to different chemical shifts and when proteins switch between conformations on a microsecond to millisecond time-scale, it is known as conformational exchange. This change in the local chemical environment can have a considerable effect on the appearance of its NMR spectra as is illustrated in figure 1.30.

When nuclei move slowly (or do not move) between two states, A and B, with respect to the change in frequency ($v_{(\mathrm{A})} - v_{(\mathrm{B})}$) (Hz) of the states, two individual environments are observed as two sharp peaks. This is known as slow exchange and is represented by the equation $k_{\mathrm{ex}} << |v_{(\mathrm{A})} - v_{(\mathrm{B})}|$. At the other extreme, A and B exchange quickly between the two states with respect to the change in frequency of the two states. This is called fast exchange and is defined by $k_{\mathrm{ex}} >> |v_{(\mathrm{A})} - v_{(\mathrm{B})}|$. In this exchange process, the peak observed in an NMR spectrum is the average of the population-weighted states. When the rate of exchange is roughly the same as the frequency difference between states, an intermediate exchange process occurs ($k_{\mathrm{ex}} \sim |v_{(\mathrm{A})} - v_{(\mathrm{B})}|$). This can be problematic in NMR spectroscopy as it can cause significant peak broadening, leading to peaks not being observable in the spectrum [Williamson, 2013].

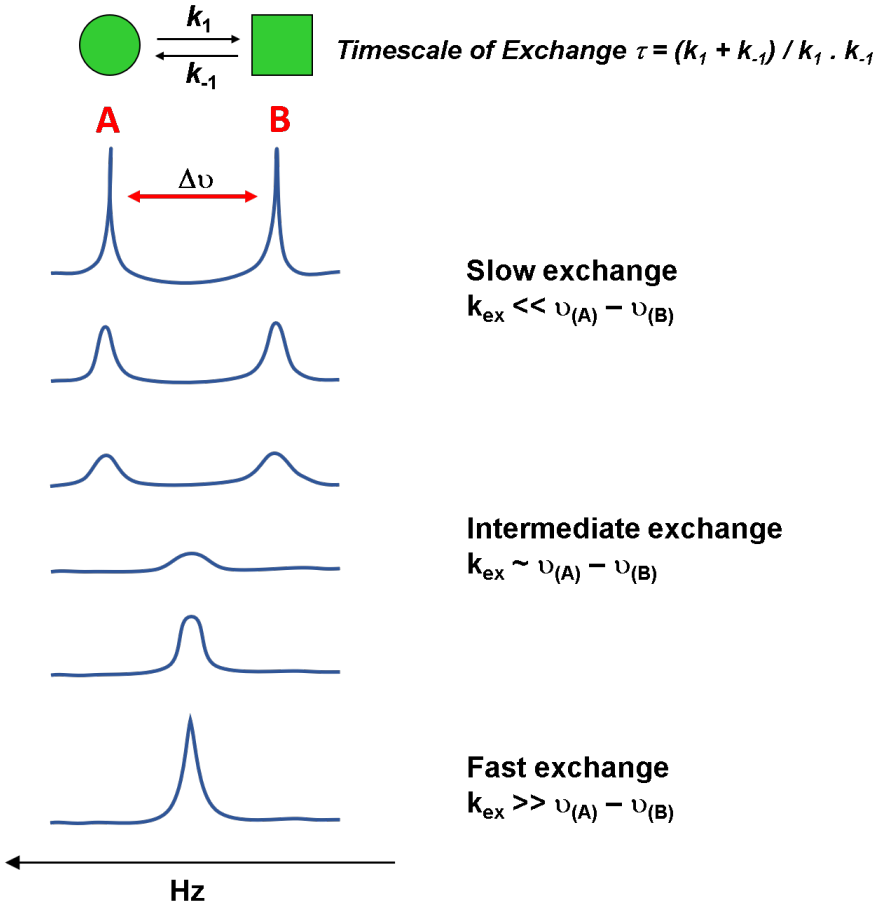


Figure 1.30: A simplified illustration of conformational exchange between two nuclei (A and B) and the effect on NMR spectra. The timescale of exchange (τ) is defined where k_1 and k_{-1} are the forward and reverse rate constants, k_{ex} is the rate of exchange, Δv is the frequency change in Hz. Slow exchange occurs when the rate of exchange is much lower than the difference in frequency between the two states (A and B). Fast exchange is when the opposite occurs; a higher rate of exchange compared to the frequency between the two states. When the rate of exchange is similar to the frequency between states, it is known as intermediate exchange, which significantly broadens peaks, often to the extent they cannot be observed in an NMR spectrum. Figure adapted from [Rule and Hitchens, 2006].

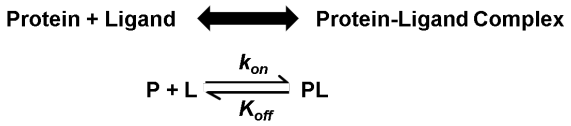
1.6.2 Protein-ligand binding

Protein-ligand interactions are commonly studied using the method described below in section 1.7, Isothermal Titration Calorimetry. Nuclear Magnetic Resonance spectroscopy can be a useful alternative as it can offer insight into the

interaction between a protein and ligand at the atomic level. This is achieved by tracking the change in resonance of the backbone amide in [^{1}H, ^{15}N]-HSQC spectra after addition of a ligand [Fielding, 2003]. The backbone amide resonances appear in different positions upon experiencing changes in their local magnetic field. A direct protein-ligand interaction can cause significant changes to the [^{1}H, ^{15}N]-HSQC as the ligand is titrated into the sample. For fast exchange processes, the amount that the resonances move relates to a ligand binding curve and at saturation, the resonance will stop moving. As explained below, the dissociation constant can then be calculated for each residue. For peaks which disappear and reappear in a spectrum (in a slow exchange process), it is much harder to calculate the dissociation constant.

Many factors such as binding affinities, concentrations, and dissociation rates, all play a part in determining the interaction between a protein and a ligand. NMR and ITC can be used to help determine the dissociation constant and hence the binding affinity of a protein-ligand interaction. Equation 1.5 shows a simple reaction where protein (P) and ligand (L) form a complex (binding rate: k_{on}) and dissociate (dissociation rate: k_{off}).

(1.5)



where k_{on} and k_{off} are the binding and dissociation rates respectively.

The equation for the equilibrium constant (K_{eq}) of a ligand binding to a protein is shown in equation 1.6

$$K_{eq} = \frac{[PL]}{[P][L]} = \frac{k_{on}}{k_{off}} \tag{1.6}$$

where [P] is the concentration of free protein, [L] is the concentration of free ligand and [PL] is the concentration of the protein-ligand complex. The dissociation constant (K_d) is the inverse of (K_{eq}), as displayed in equation 1.7 and has the units M (or mol/L).

$$K_d = \frac{[P][L]}{[PL]} \tag{1.7}$$

The rate at which a ligand enters a binding site is often controlled by a diffusion process and is dependent on the ligand. The dissociation rate is mainly determined by the concentration of bound ligand and the strength of the binding interaction. The K_d, or dissociation constant, of a protein-ligand complex, can allow comparisons between different interaction affinities and has a relationship where lower dissociation constant values correspond to stronger binding affinities.

Weak interactions, for example, ubiquitin:ubiquitin receptor reactions, occur at $K_d > 10^{-4}$ M. The majority of protein interactions occur at a moderate affinity, or $10^{-4} < K_d < 10^{-8}$ M. The highest affinity interactions, for example, antigen:antibody interactions, can occur at $K_d < 10^{-9}$ M.

1.7 Isothermal Titration Calorimetry (ITC)

Isothermal Titration Calorimetry (ITC) is a technique capable of measuring the heat absorbed or generated (enthalpy change ΔH) during the interaction between molecules at a constant temperature. For the investigation of protein-ligand interactions, the method usually involves a protein in a cell, a reference cell and a ligand in a syringe.

The ligand is precisely and gradually titrated into the cell, the sample cell becomes warmer during an exothermic reaction, and the calorimeter measures and records the enthalpy change. Binding causes a downward peak in the raw

heat plot which is directly proportional to the amount of binding (figure 1.31A) and as the temperature returns to that of the reference cell, the signal returns to its starting position. When peaks level out, saturation is achieved, and peaks observed after the saturation point are representative of the heat of dilution. ITC is particularly useful, and despite heat changes being of the order 10 x 10^{-10} °C, the ΔH as a function of concentrations can be used to determine the binding affinity, enthalpy, entropy and stoichiometry of an interaction, all in the same experiment, as illustrated in figure 1.31B. The binding isotherm or integrated heat plot of an ITC experiment contains various information. It is achieved by integrating the area of each peak and plotting the result against the molar ratio of ligand to protein. The K_d can be determined from the gradient of the binding curve, where gradient = $K_a = \frac{1}{K_d}$. The stoichiometry can be identified by finding the molar ratio at the centre of the binding curve.

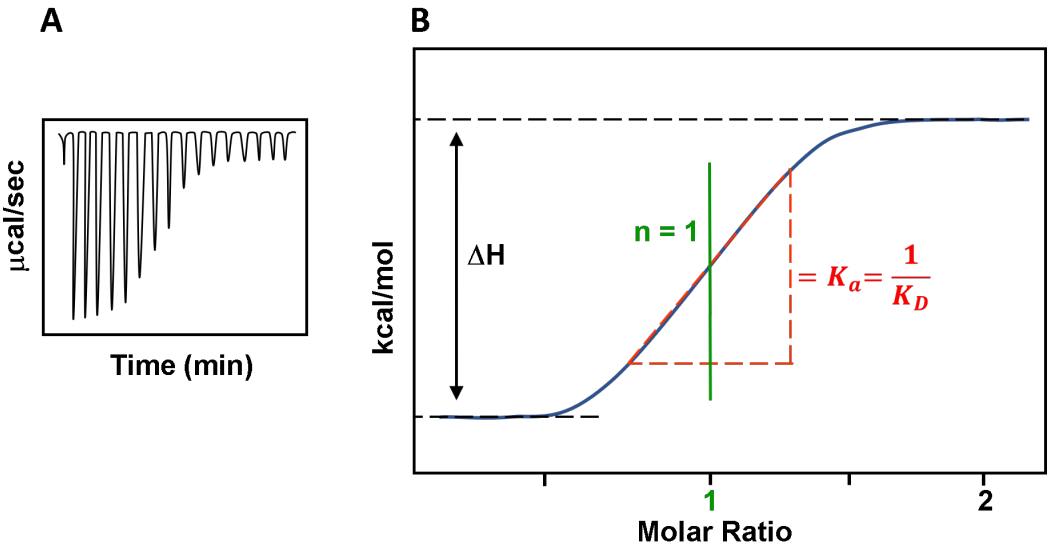


Figure 1.31: **(A)** An example of an exothermic raw heat plot showing downward peaks at each injection point representing the heat change. Saturation is confirmed when additional ligand does not affect the heat change, other than by heat of dilution, and so steady values are obtained at the end of the titration. **(B)** A simplified integrated heat plot of a 1:1 binding ITC experiment. The binding mechanism can be determined by the enthalpic response, ΔH shown in black. The stoichiometry (n) is determined by the midpoint of the curve and the K_d is calculated from the gradient of the binding curve. Adapted from [Malvern Panalytical, 2018].

The equation for the standard Gibbs energy of binding (ΔG) is shown in

equation 1.8.

$$\Delta G = -RTlnK_a = RTlnK_d \tag{1.8}$$

where T is the temperature and R is the gas constant (8.314 J K^{-1} mol^{-1}). Negative Gibbs energy and small K_d values correspond to stronger binding of complexes. The value for ΔG can also be used to determine the entropic change of binding by using the known ΔH, temperature (T) and equation 1.9, where ΔS is the entropy change.

$$\Delta G = \Delta H - T\Delta S \tag{1.9}$$

The data can be fitted in Malvern Analysis software using a fixed stoichiometry (if known) and protein and ligand concentrations to determine the K_d, and ΔH, allowing ΔG and ΔS to be calculated.

1.8 Aims of the thesis

As discussed earlier in this chapter, E3 ubiquitin ligase WWP2 is involved in several pathways related to cell proliferation and has been shown to increase tumour spread following overexpression. The different binding partners of WWP2 are involved in regulating cell proliferation and differentiation, resulting in the involvement of WWP2 in cancer. The specificity of WWP2 for certain substrates through its WW domains, as well as its roles in tumorigenesis and metastasis in various forms of cancer, clearly represents a potentially attractive therapeutic target for anticancer drugs. Designing new therapeutic inhibitors requires molecular details of the interactions between a protein and its targets. Therefore, this thesis aimed to investigate changes in the WW domains of

WWP2 upon binding to substrate proteins and then map the binding sites.

In order to understand the interactions between WWP2 and these binding partners, the structural characteristics of WW2, WW2-3 and WW3 domains as well as the molecular basis for interactions between WW domains and the target regions OCT4 and Smad7 were investigated. To achieve this, Nuclear Magnetic Resonance (NMR) spectroscopy was used to attain backbone assignments, and where possible structural information for WW2, WW2-3 and WW3 domains. NMR titrations characterised the molecular basis for the interactions between WW2 and binding partners OCT4 and Smad7, in addition to WW3 and binding partner OCT4. Determination of K_d values for WW2 with OCT4 and Smad7 will be achieved using a combination of isothermal titration calorimetry (ITC) experiments and NMR titration data. Developments in knowledge regarding the structure and interactions of WWP2 with its target protein substrates are extremely important for determining how it operates and how it can be targeted.

Chapter 2

Materials and Methods

2.1 Media and buffers

LB

- 10 g/L tryptone
- 5 g/L yeast extract
- 10 g/L NaCl

pH 7.4, autoclaved on cycle 1 - Media Run.

M9 Salts

- 42 mM Na_2HPO_4
- 22 mM KH_2PO_4
- 8.5 mM NaCl
- 18.7 mM NH_4Cl
- Appropriate antibiotics

pH 7.8, autoclaved on cycle 1 - Media Run.

Minimal Essential Media (MEM)

To M9 Salts add sterile:

- 0.4% glucose
- 2 mM $MgSO_4$
- 10 mM $CaCl_2$
- 10 mM fresh $FeSO_4$
- 1 x micro-nutrient solution
- 1 x MEM vitamin mix (Sigma Aldrich)
- Appropriate antibiotics

Isotopically enriched MEM was made using ^{15}N NH_4Cl where nitrogen labelling is required, or ^{15}N NH_4Cl and ^{13}C glucose for nitrogen and carbon labelled proteins.

1000 x stock solution of micronutrient solution

- 3 μM $(NH_4)_2MoO_4$
- 4 μM H_3BO_3
- 30 μM $CaCl_2$
- 10 μM $CuSO_4$
- 80 μM $MnCl_2$
- 10 μM $ZnSO_4$

TAE Buffer (Tris-acetate-EDTA)

- 40 mM Tris
- 20 mM acetic acid
- 1 mM EDTA

2 x SDS loading buffer

- 0.312 M Tris pH 6.8
- 10% SDS
- 25% β-mercaptoethanol
- 0.05% Bromophenol Blue

Ammonium bicarbonate buffer

- 2 mM ammonium bicarbonate

NMR buffer

- 20 mM Na_2HPO_4
- 50 mM NaCl
- pH 6.8

NMR titrations also required 15 mM DTT due to peptide cysteines.

10 x Tris-Glycine running buffer

- 250 mM Tris
- 1.92 M Glycine
- 1% SDS

Tris-HCl buffer

- 25 mM Tris-HCl pH 8
- 0.5 M NaCl

Gel Filtration buffer

- 50 mM Tris-HCl pH 7.5
- 150 mM NaCl

Nickel-NTA affinity chromatography buffers

Table 2.1: Nickel-NTA affinity chromatography buffers

	Buffer X: 0.5 M NaCl, 20 mM Tris-HCl, pH 8
Wash buffer A	Buffer X + 1 mM imidazole
Wash buffer B	Buffer X + 5 mM imidazole
Wash buffer C	Buffer X + 10 mM imidazole
Wash buffer D	Buffer X + 15 mM imidazole
Elution buffer E	Buffer X + 250 mM imidazole

Polyacrylamide gels

Table 2.2: SDS-PAGE gel recipe – volumes of each solution required to make two 0.75 mm thick gels. All resolving gels used the same 5% acrylamide stacking gel.

	Resolving Gel			**Stacking Gel**
	10%	**12%**	**15%**	**5%**
dH_2O	7.7 mL	6.9 mL	5.7 mL	5.8 mL
1.5 M Tris pH 8.8	4.0 mL	4.8 mL	6.0 mL	-
1 M Tris pH 6.8	-	-	-	1.5 mL
40 % Acrylamide /Bis	5.33 mL	4.8 mL	8.0 mL	2.5 mL
10% SDS	160 μL	160 μL	160 μL	100 μL
10% APS	160 μL	160 μL	160 μL	100 μL
TEMED	16 μL	16 μL	16 μL	10 μL

Tricine gels

Table 2.3: Tricine gel recipe – showing volumes of each component required for two tricine gels.

	Resolving gel	**Stacking gel**
dH_2O	3.1 mL	6.7 mL
40% (w/v) Acylamide 19:1	5 mL	1.05 mL
Gel buffer	5 mL	2.5 mL
70% (v/v) glycerol	2 mL	-
10% (w/v) APS	66 μL	80 μL
TEMED	6.6 μL	8 μL

Tricine gel buffers

Table 2.4: Tricine gel buffer recipes.

	Gel buffer	**Cathode buffer**	**Anode buffer**
Tris HCl	18.15 g	-	-
Tris base	-	24.2 g	48.4 g
Tricine	-	35.84 g	-
SDS	150 mg	2 g	-
dH_2O	50 mL	200 mL	200 mL
pH	8.45	8.25	8.9

2.2 DNA techniques

2.2.1 WW domain constructs

Table 2.5: WW domain constructs.

Name	Protein construct
WW2	RPLPPGWEKRTDPRGRFYYVDHNTRTTTWQRPTA
WW3	GPLPPGWEKRQDN-GRVYYYVNHNTRTTQWEDPRT
WW2-3	RPLPPGWEKRTDPRGRFYYVDHNTRTTTWQRPTA EYVRNYEQWQSQRNQLQGAMQHFSQRFLYQSSSAS TDHDPLGPLPPGWEKRQDNGRVYYYVNHNTRTTQW EDPRTQ

2.2.2 Bacterial expression vectors

2.2.2.1 pSKDuet01

The pSKDuet01 plasmid was used for the expression of WW domain protein constructs of WWP2. This plasmid has resistance to kanamycin and was a gift from Hideo Iwai (Addgene plasmid # 12172) [Iwai *et al.*, 2006]. pSKDuet01 contains a hexa-His-tag for purification and GB1 solubility tag (Protein G Immunoglobulin-binding domain 1). This plasmid expresses the fusion protein in bacteria by using the lactose inducible T7 RNA polymerase promoter and isopropyl β-D- thiogalactoside (IPTG) to induce. The original vector contained a gene encoding a DnaE Nintein (NpuDnaE) which was removed and replaced with WW domain constructs. Figure 2.1 was taken from Addgene.

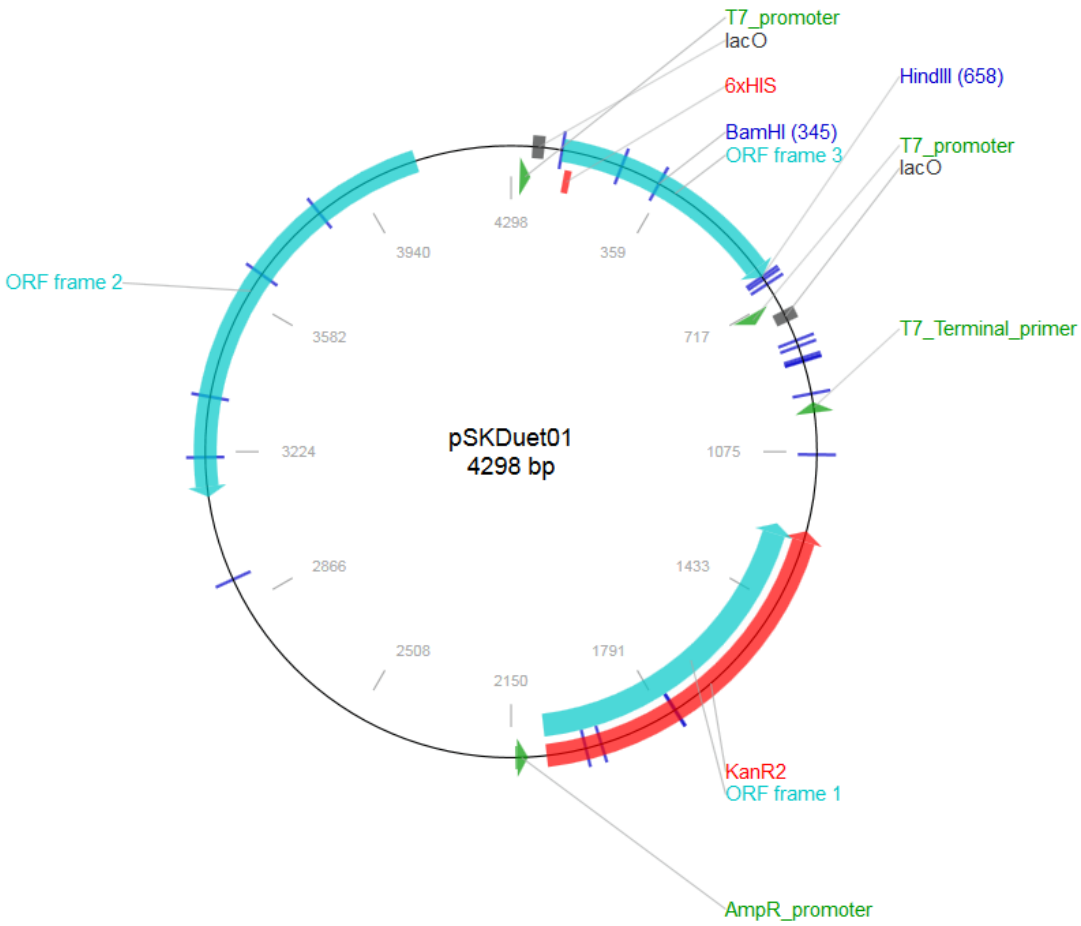


Figure 2.1: Plasmid map of pSKDuet01 (Addgene plasmid # 12172) which was a gift from Hideo Iwai.

The hexa-His-GB1 tagged tandem domain WW2-3 was obtained from the Chantry laboratory, UEA.

2.2.3 In-Fusion Cloning

Base vector pSKDuet01 (Addgene plasmid # 12172) [Iwai *et al.*, 2006] was linearised using restriction enzymes BamHI (Promega) and HindIII (Promega) and was obtained from the Chantry laboratory, along with PCR primers. Forward

(fwd) and reverse (rev) primers used to generate the WW2 and WW2-3 constructs, are shown in table 2.6 and were designed to contain additional base pairs (roughly 15 bp), which are complementary to the ends of the linearised pSKDuet01, also shown in table 2.6. Phusion High fidelity (HF) DNA polymerase (Finnzymes 10024537) was used to amplify the gene of interest from pRK5-HA-FL-WWP2 by PCR, as shown in tables 2.7 and section 2.2.3.2.

Table 2.6: Primer sequences for High Fidelity Phusion Reactions

	Primers
WW2	Fwd: 5'-CGTAACGGAAGGATCCGAGCGGCCCCTTCCTC-3' Rev: 5'-ATGCGGCCGCAAGCTTTTAGCGCACGTACTCCGCG-3'
WW2-3	Fwd: 5'-CGTAACGGAAGGATCCGAGCGGCCCCTTCCTC-3' Rev: 5'-ATGCGGCCGCAAGCTTTTACTGGGTCCGGGGATCC-3'
pSKDuet01 vector specific	Fwd: 5'-CGTAACGGAAGGATCC-3' Rev: 5'-ATGCGGCCGCAAGCTTTTA-3'

2.2.3.1 20 μL Polymerase Chain Reaction (PCR)

The polymerase chain reactions used the high fidelity (HF) Phusion DNA Polymerase (Finnzymes 10024537). Initially a 20 μL reaction was set up to ensure the reaction worked: 4 μL 5xHF Phusion buffer, 0.4 μL dNTPs at 10 mM, 0.25 μL pRK5-HA-FL-WWP2 Template DNA (2034.5 ng/μL), 1 μL of both forward and reverse primers at 10 μM, 0.2 μL Phusion DNA Polymerase were mixed and made up to 20 μL in a PCR tube with 13.15 μL ddH_2O. A thermocycler (MJ Research PTC200) was used to perform all PCR reactions and used the PCR cycle programs shown in table 2.7.

Table 2.7: PCR program reaction steps for inserting WW domains into pSKDuet01.

	Duration	**WW2**	**WW2-3**
Initial denaturation	30 s	98 °C	98 °C
Denaturation	8-10 s	98 °C	98 °C
Annealing	20 s	66.6 °C	62 °C
Extension	10 s	72 °C	72 °C
Final extension	8 minutes	72 °C	72 °C
Hold		4 °C	4 °C

2.2.3.2 100 μL Polymerase Chain Reaction (PCR)

Successful reactions were upscaled to 100 μL PCR Phusion reactions: 20 μL 5xHF Phusion buffer, 2 μL dNTPs at 10 mM, 0.25 μL pRK5-HA-FL-WWP2 Template DNA (2034.5 ng/μL), 5 μL of both forward and reverse primers at 10 μM, 1 μL Phusion DNA Polymerase were mixed and made up to 100 μL in a PCR tube with 66.75 μL ddH_2O. The same PCR cycle programs were used as in table 2.7.

PCR products (100 μL sample) were combined with 15 μL DNA Gel Loading Dye (6X) and the whole sample was loaded alongside a 5 μL control of pRK5-HA-FL-WWP2 template DNA sample with 1 μL DNA gel loading dye (6X) and 1 kb Plus DNA Ladder. Gels were run for 1 hour at 100 V and then imaged on a BioRad ChemiDoc XRS scanner. The gel was then illuminated by UV light and the bands of correct height were excised, weighed and purified using QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions. The concentration of the purified PCR products were then calculated using the absorbance at 260 nm, measured with a Nanodrop ND-1000 Spectrophotometer.

2.2.3.3 Agarose gels

1% w/v agarose gels were prepared by heating 0.6 g agarose with 60 mL 1 x TAE Buffer until all agarose had dissolved. When the solution had cooled to ∼40°C, 6 μL SYBR Safe (ThermoFisher) was added and poured into a 60 mL gel tray to set. PCR products (20 μL sample) were combined with 4 μL DNA Gel Loading Dye (6X) and the whole sample was loaded alongside 5 μL pRK5-HA-FL-WWP2 Template DNA sample with 1 μL DNA Gel Loading Dye (6X) and 1 kb Plus DNA Ladder. Gels were run for 1 hour at 100 V and then imaged on a BioRad ChemiDoc XRS scanner.

2.2.3.4 In-Fusion

The amplified insert was cloned into the pSKDuet01 vector using the In-Fusion HD Cloning kit (Clontech Laboratories, Inc. 639648). A 10 μL In-Fusion cloning reaction was set up as in table 2.8.

Table 2.8: 10 mL In-Fusion cloning reaction protocol.

	10 mL In-fusion reaction
Gel purified PCR fragment	1 μL (10 ng)
Gel purified linearized vector (pSKDuet01)	6 μL (50 ng)
5X In-Fusion HD Enzyme Premix	2 μL
Deionised Water	1 μL

The reaction was mixed and then incubated for 15 minutes at 50 °C, and then placed on ice, ready for transformation into Stellar™ Cells or storage at -20 °C. A 1 μL aliquot of this reaction was used for transformation into Stellar™ Competent Cells, a high efficiency transformation strain, provided with the kit. Stellar™ genotype: F-, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Θ80d lacZΔ M15, Δ(lacZYA-argF) U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ-.

2.2.3.5 Bacterial transformation

An aliquot of competent cells was left to thaw on ice. Plasmid DNA (1 μL) was added and incubated on ice for 30 minutes before heat shocking for 60 seconds in a 42 °C water bath. After incubation on ice for a further 10 minutes, preheated LB ($\sim$700 μL) or SOC medium (provided with In-Fusion HD Cloning kit (Clontech Laboratories, Inc. 639648)) was added. The samples were then incubated at 37 °C for an hour. 200 μL of the competent cells were spread on an LB agar plate with appropriate antibiotics. The remaining cells were centrifuged (Mikro 200 benchtop centrifuge, 1 minute, 8,000 x g), then concentrated by removing $\sim$700 μL of the media and re-suspending the cell pellet. The remainder of the cells ($\sim$200 μL) were then plated onto a fresh LB agar plate. Both plates were incubated overnight at 37 °C.

WW domain constructs in pSKDuet01 were transformed into BL21(DE3) *Escherichia coli* (*E. coli*) containing kanamycin at 50 μg/mL concentration.

OCT4 and Smad7 peptide were expressed as fusion proteins with the SUMO tag, using a modified pET21d(+) vector (a gift from Dr Robin Maytum, University of Bedfordshire). SUMO-OCT4 and SUMO-Smad7 plasmids were transformed into *E. coli* and were expressed with ampicillin at 100 μg/mL concentration.

During test expressions with Rosetta2(DE3)pLysS, BL21(DE3)pLysS and BL21-CodonPlus-RP cells, Chloramphenicol at 35 μg/mL was also added.

2.2.3.6 Plasmid DNA purification

4 mL LB overnight starter cultures with the appropriate antibiotics were prepared using single colonies from transformations in either Stellar Competent Cells or XL1 Blue competent cells. Cultures were grown overnight at 37 °C, 180 rpm. Samples were centrifuged, and plasmid DNA was purified from the cell pellets using a Nucleospin Plasmid DNA purification Kit following manufacturer's instructions. Where more DNA is required, NucleoBond® Xtra

Midi or NucleoBond® Xtra Maxi were used according to the manufacturer's instructions.

2.2.4 Sequencing

Each plasmid DNA sample was prepared for sequencing using a concentration of 50-100 ng/μL of DNA in samples made up to 15 μL with nuclease-free water. Where standard primers from the Eurofins Genomics website were not appropriate for the construct, correct primers at 10 pmol/μL in a total volume of 15 μL were sent to Eurofins Genomics along with the sequencing samples.

2.3 Protein techniques

2.3.1 Protein expression

2.3.1.1 Overexpression of recombinant protein in Lysogeny Broth (LB) media (1 L)

A single colony, from a fresh transformation plate, was picked and used to inoculate 12 mL of LB media with appropriate antibiotics. This was grown in a shaking incubator at 37 °C overnight, 180 rpm. 2 x 500 mL flasks of LB were prepared, and a small aliquot of each flask was saved to be used as a blank when determining Optical Density (OD) during the overexpression. 6 mL of the saturated bacterial culture was transferred to each 500 mL flask of LB along with the appropriate antibiotics. The cells were incubated at 37 °C, 180 rpm. Aliquots (1 mL) were taken at half-hour intervals and the OD was measured at 600 nm. When the OD_{600} reached between 0.8 and 1, cells were induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG, amount determined during test expression) and were grown for a further three hours at 37 °C, 180 rpm. The cells were harvested by centrifugation at 5,000 x g (Beckman J26 XP,

rotor JLA-9.1), 4 °C for 15 minutes. The media was removed and LB (20 mL) was used to resuspend each pellet. The cells were centrifuged again at 5,000 x g (Beckman J26 XP, rotor JA-12), 4 °C for 15 minutes and the media was aspirated. The bacterial pellet was stored at -20 °C. SDS-PAGE was used to confirm the expression.

2.3.1.2 Small scale test expression

Bacterial cells were transformed on LB agar plates as described in section 2.2.3.5. For GB1-WW2, a test expression was performed to optimise recombinant protein expression. Bacterial cells used included: *E. coli* BL21(DE3), BL21(DE3*), Rosetta2(DE3)pLysS, BL21(DE3)pLysS and Bl21-CodonPlus-RP. One colony from each cell-line transformation was used to inoculate 3-6 mL LB containing appropriate antibiotics. Cultures were incubated for ∼4 hours at 37 °C, 180 rpm until OD_{600} reached 0.8-1.2. 1 mL samples were taken from each starter culture and prepared for SDS-PAGE. The cultures were then induced using 0.4 mM IPTG (Isopropyl-β-D-1-thiogalactopyranoside) and incubated at 37 °C for a further 3 hours, 180 rpm. The optimum cell line was determined using SDS-PAGE analysis.

The optimised conditions were also successful for GB1-WW2-3 and GB1-WW3 expression and so for full-scale expression of GB1 tagged WW domains, 0.4 mM IPTG was used for induction and left for 3 hours at 37 °C.

Expression of fusion proteins SUMO-OCT4 peptide and SUMO-Smad7 peptide used the following conditions, which were optimised by previous group members for SUMO tagged peptides; 0.4 mM IPTG, ∼16 h induction at 30 °C.

2.3.1.3 Overexpression of isotopically labelled recombinant proteins

Plasmids were transformed into the competent cells and spread on agar plates containing appropriate antibiotics. A single colony from a fresh transformation

was used to inoculate an 8 mL LB starter culture which was then grown overnight at 37 °C, 180 rpm.

100 mL of unlabelled MEM was inoculated using 5 mL saturated starter culture. The MEM culture was grown for $\sim$ 2-3 hours (or until it appeared saturated) at 37 °C and cells were then transferred to a sterile centrifuge tube and harvested by centrifugation at 5,000 x g (Beckman J26 XP, rotor JA-16.250), 25 °C for 10 min.

The cell pellet was then resuspended in 10 mL of isotopically enriched MEM. Once resuspended, the culture was distributed into the full culture volume (n x 500 mL) and grown at 37 °C until OD_{600} = 0.6 - 0.8. Expression was then induced with IPTG (amount determined during test expression). The culture was then centrifuged at 5,000 x g (Beckman J26 XP, rotor JLA-9.1), 4 °C for 10 min, and the recombinant protein was purified as outlined in section 2.3.2.1.

2.3.1.4 Preparation of ULP1 (Ubiquitin-like-specific protease 1)

His-UPL1 protease was a gift from Dr. Robin Maytum, University of Bedfordshire and was cloned in pJC20 expression vector. ULP1 was transformed in Rosetta2(DE3) competent cells, as in section 2.2.3.5, using ampicillin (100 μg/mL) and chloramphenicol (35 μg/mL). A single colony from a fresh transformation was used to inoculate 8 mL LB starter culture which was left to grow at 37 °C, 180 rpm.

1 L LB was inoculated using 1 mL saturated starter culture. When OD_{600} reached between 0.8 and 1, cells were induced by adding 0.5 mM IPTG and were grown for a further three hours at 37 °C, 180 rpm. The cells were harvested by centrifugation at 5,000 x g (Beckman J26 XP, rotor JLA-9.1), 4 °C for 15 minutes and the media was discarded. The bacterial pellet was frozen at -20 °C. SDS-PAGE analysis was used to confirm the expression. ULP1 was then purified using Ni-NTA affinity chromatography as in section 2.3.2.1. Protein samples were pooled and purified using a disposable PD10 desalting column (GE Healthcare).

Equilibration buffer (75 mM Tris-HCl, pH 8), 1 mM EDTA and 0.5 mM DTT) was prepared and the used to equilibrate the column by filling the column completely five times and discarding flow-through. 2.5 mL of sample was added to the column and allowed to enter the column. Where there was not 2.5 mL of sample, the volume was adjusted to 2.5 mL using equilibration buffer. Once the sample had entered the bed, the flow-through was discarded and the sample was eluted using 3.5 mL of buffer.

500 μL aliquots with 50 % glycerol were frozen in liquid Nitrogen and stored at -80 °C. A 500 μL aliquot of ULP1 was tested to confirm activity.

2.3.2 Purification techniques

2.3.2.1 Nickel-NTA Immobilized metal affinity chromatography (IMAC)

His-tagged proteins were purified using Super-Ni-NTA affinity resin (Generon). Cell pellets were thawed on ice and then re-suspended in 20 mL ice-cold binding buffer X (20 mM Na_2HPO_4, 0.5 M NaCl and 1 mM imidazole at pH 7.4). The cell suspension was sonicated using a Sonics Vibracell VCX 130, three times for 2 minutes (Pulse 02 01 Amp 1 80%) whilst held on ice. The cell lysate was then centrifuged for 45 minutes at 4 °C, 30,000 x g (Beckman J26 XP, rotor JA-20). The supernatant was filtered through a 0.45 μm filter before loading it onto the Ni-NTA column. The flow-through was collected on ice and then the column was washed twice with 7 mL Buffer A, twice with 7 mL Buffer B, twice with 7 mL Buffer C and twice with 7 mL Buffer D (see table 2.1 for buffer compositions). All samples were collected on ice. The protein was then eluted by washing the column four times with elution Buffer E, storing each aliquot separately on ice. Samples from all fractions were run on an SDS-PAGE gel.

During the test purification of GB1-WW2-3 with urea, Nickel-NTA affinity chromatography buffers were prepared as previously, along with an additional 6 M

urea in each to denature proteases. When using urea, purification took place at room temperature and samples were not collected on ice as cold temperatures cause the urea to precipitate. Samples were thoroughly dialysed to remove urea straight after purification.

2.3.2.2 Size-exclusion chromatography

The HiLoad 16/600 Sepharose Superdex75 grade column (GE Healthcare) was connected to the ÄKTA-FPLC and washed with 2 column volumes (120 mL) of water before equilibration with 2 column volumes of gel filtration buffer (see section 2.1) at 1 mL/min. The sample was concentrated to 1 mL using a Vivaspin® 500 centrifugal concentrator (Sartorius), and injected onto the column. The sample was gel filtered at 1 mL/min and fractions were collected in 2 mL aliquots. SDS-PAGE was used to analyse alternate fractions before pooling the desired fractions.

2.3.3 Digestion of Fusion tags

2.3.3.1 Thrombin digestion

WW domains were expressed as GB1 fusion proteins and had an N-terminus polyhistidine tag (6xHis-tag) with a thrombin cleavage site, as shown in figure 2.2.

Figure 2.2: Schematic diagram showing an example WW domain construct with N-terminus polyhistidine tag and GB1 solubility tag, with the thrombin cleavage site highlighted.

The His-tag was removed after purification with a Ni^{2+} IMAC column (section 2.3.2.1) using the thrombin cleavage site. 2 units of human Thrombin (Sigma-Aldrich) was added to the sample per mg of purified protein. Using

10 kDa MWCO dialysis tubing, the protein/thrombin mix was dialysed into Tris-HCl buffer at room temperature until a sufficient amount of protein was digested, as determined by SDS-PAGE. The digested His-tag was then removed by Ni-NTA affinity chromatography.

2.3.3.2 ULP-1 digestion

Peptides of Smad7 and OCT4 were expressed as fusion proteins with a His-SUMO (Small Ubiquitin-like Modifier) tag. ULP1 (Ubiquitin-like-specific protease 1) was used to digest the His-SUMO tag. Purified protein was dialysed into Tris-HCl buffer at 4 °C using 10 kDa MWCO dialysis tubing. His-ULP1 was then added to the sample (500 μL per 1 L culture volume, expressed and purified as in section 2.3.1.4) and incubated at room temperature until SDS PAGE confirmed digest. The digested His-SUMO tag was removed by Ni-NTA affinity chromatography.

2.4 General biological techniques

2.4.1 Preparation of competent cells, $CaCl_2$ method

To prepare *Escherichia coli* strains for efficient transformation, 5 mL LB culture was inoculated with fresh colonies picked from LB-agar plates and incubated overnight at 37 °C, 180 rpm. 1 mL of the overnight culture was used to inoculate 100 mL of fresh LB. This cell culture was grown at 37 °C, 180 rpm until ($OD_{600} = \sim 0.4$) and the cells were collected by centrifugation at 5000 x g for 5 minutes at 4 °C. Cells were resuspended in 50 mL ice-cold, sterile 0.1 M $CaCl_2$ and held on ice for 1-2 hours. Cells were collected as previously by centrifugation and gently resuspended in 10 mL ice-cold, sterile 0.1 M $CaCl_2$. 10% (v/v) ice-cold, sterile glycerol was added to the cells and mixed well before incubation on ice for 30 minutes. 100-200 μL aliquots were prepared in sterile 1 mL microcentrifuge tubes, frozen in liquid nitrogen and stored at -70 °C.

2.4.2 SDS Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE gels were made using the recipe shown in table 2.2 . Gel running equipment was set up with 1 x Tris-Glycine running buffer (diluted from 10 x stock in section 2.1) in the upper compartment and used 1 x Tris-Glycine running buffer filled the lower compartment. Samples were prepared by re-suspending the pellet in water and adding an equal volume of 2 x SDS loading buffer. The samples were then boiled for 10 minutes and centrifuged at 14,000 x g for 10 minutes. 8 μL of each sample on 15 well gels, and 10 μL on 10 well gels (unless stated otherwise), with a 5 μL aliquot of molecular weight marker (ThermoScientific Unstained molecular Weight Marker, #26610) were loaded into the wells and run for $\sim$1 hour at 200V. The gels were then stained using InstantBlue (Expedeon) Coomassie blue stain for at least 1 hour, or until protein bands were clearly visible.

The ThermoScientific Unstained molecular Weight Marker used contains seven proteins spanning 14.4 to 116kDa for use as size standards for comparison with SDS-gels obtained.

2.4.3 Tricine gels

Tricine gels (used for resolving peptides and low MW proteins) were used to attempt to visualise peptides Smad7 and OCT4. Gels were prepared using the recipe in table 2.3 and were run using the Anode and Cathode buffers listed in table 2.4 [Schägger, 2006].

8 μL of each sample on 15 well plates, and 10 μL on 10 well plates, with a 5 μL aliquot of molecular weight marker (ThermoScientific PageRuler™ Unstained Low Range Protein Ladder, #26632) were loaded into the wells and run for $\sim$1 hour at 200V.

2.4.4 Silver staining

Silver staining protocol is used to detect peptides on tricine gels due to its enhanced sensitivity when compared to InstantBlue (Expedeon).

Tricine gels were removed from the running buffer and transferred immediately into 50% (v/v) methanol for 10 minutes. 50% methanol was replaced with 5% methanol for 10 minutes before gels were washed briefly with dH_2O and 2 μM DTT was added. After shaking for an additional 10 minutes, gels were washed gently with dH_2O and poured off. 0.1% $AgNO_3$ was added and left to shake for >20 minutes. This solution was poured off, gels were rinsed quickly with dH_2O three times, and fresh developing solution (7.5 g Na_2CO_3, 125 μL 35% (v/v) formaldehyde, dH_2O to 250 mL) was added. To prevent over-staining, once bands were clearly visible, solid citric acid was added until the reaction stopped.

2.4.5 Measuring protein concentration

Protein concentration was calculated using a UV visible spectrometer or Nanodrop at 280 nm. Both methods make use of the Beer-Lambert law and predicted extinction coefficients (calculated from ExPASy Protparam tool (https://web.expasy.org/protparam/) to calculate the concentration of a protein sample. The Beer-lambert Law can be seen in equation 2.1:

$$A = \epsilon \cdot c \cdot l \tag{2.1}$$

Where A is the absorbance, ϵ is the molar absorptivity (L mol^{-1} cm^{-1}), c is the concentration of the compound in solution (mol L^{-1}) and l is the path length of the sample (cm).

2.4.6 Protein lyophilization

Peptides were lyophilized in a freeze drier after dialysis into volatile ammonium bicarbonate buffer and flash freezing in liquid nitrogen. The tubes were then sealed with parafilm and stored at 4 °C.

2.5 Nuclear Magnetic Resonance (NMR)

2.5.1 NMR sample preparation

After purification, samples were either concentrated by freeze-drying (section 2.4.6) or spin concentrating. Freeze-dried peptides Smad7 and OCT4 were added directly to NMR sample buffers as opposed to WW domain proteins which were concentrated after dialysis into NMR buffer. The protein was concentrated using Vivaspin® 20, 5 kDa MWCO, centrifugal concentrators (Sartorius) until the sample concentration was roughly 1 mM (or when 1 mM concentration could not

be achieved, the sample was concentrated to volume ∼500 μL). 430 μL of protein sample was made up to 500 μL by the addition of D_2O to a final concentration of 10% v/v, 200 μM DSS and 0.03% NaN_3. The sample was filtered using a 0.22 μm spin filter (Corning) and transferred to a 5 mm NMR tube.

2.5.2 NMR pH determination

The pH of NMR samples was determined and adjusted to between 6.8 and 7.2 using a modified version of the method described by Baryshnikova *et al.* [2008] which uses 1D ^{1}H-NMR chemical shifts. Equation 2.2, was used to calculate pH using imidazole peaks in the 1D spectrum.

$$pH = pKa - log_{10}(\frac{\delta_{\mathrm{Obs}} - \delta_{\mathrm{A}}}{\delta_{\mathrm{HA}} - \delta_{\mathrm{Obs}}}) \tag{2.2}$$

where δ_{Obs} is the observed chemical shift (ppm) of the imidazole peak and δ_{HA} and δ_{A} are the predetermined chemical shifts with respect to the protonated and deprotonated states respectively.

2.5.3 NMR spectral processing

The NMRPipe software package [Delaglio *et al.*, 1995] was used to process and optimise acquired spectra by applying zero-filling, linear prediction, Fourier transformation and phase adjustment. The proton chemical shifts were referenced using the frequency of the DSS standard peak in each sample. Indirectly detected dimensions were referenced using a script written in house by Dr James Tolchard which used IUPAC standard gyromagnetic ratios.

2.5.4 Backbone assignments

All processed spectra were analysed in CCPN Analysis software [Vranken *et al.*, 2005]. Protein backbone assignment is described in section 1.6.1.4.

2.5.5 Side chain assignments

Aliphatic side-chain assignments were determined using H-CCH-TOCSY and HCCH-TOCSY spectra in combination with a [^{1}H, ^{13}C]-HSQC. Aromatic side chains (histidine Hδ1, tryptophan Hδ1, tyrosine Hδ, ϵ and phenylalanine Hδ, ϵ) were assigned using 2D HBCBCGCDHD and 2D HBCBCGCDCEHE experiments [Yamazaki *et al.*, 1993]. Aromatic CH correlations were assigned using a ^{13}C-TROSY-Aromatic spectrum. All spectra for each set of experiments were analysed in a single CCPN Analysis [Vranken *et al.*, 2005] project.

2.5.6 WW3 structure calculation

^{13}C-NOESY and ^{15}N-NOESY experiments were acquired to determine through-space distance restraints. NOE (Nuclear Overhauser Effect) peaks were picked manually in CCPN Analysis software to avoid noise. Shift lists, NOESY peak lists and the amino acid sequence were exported in NMRView format with the XPLOR naming system. Protein backbone Phi and Psi angles and secondary structure assignments were predicted from the amino acid sequence, experimental chemical shifts and database values using a program within CCPN analysis software [Vranken *et al.*, 2005] called DANGLE (Dihedral ANgles from Global Likelihood Estimates) [Cheung *et al.*, 2010]. These predicted dihedral angles were then exported in CNS format.

Files were then uploaded to the WeNMR [Wassenaar *et al.*, 2012] ARIA (Ambiguous Restraints for Iterative Assignment) Webportal [Nilges, 2016] `http://enmr.chemie.uni-frankfurt.de/portal/aria.html` for structure

prediction. The webportal is based on the ARIA version 1.2 package [Linge *et al.*, 2001]; [Rieping *et al.*, 2006]; [Linge *et al.*, 2003] with several modifications, and used 20 final refined structures for each run. Tolerances used were ^{1}H=0.05, ^{15}N=0.4 and ^{13}C=0.02 and mixing times for ^{15}N and ^{13}C-NOESY experiments were 120 and 100 ms respectively. Rotation correlation time was estimated as 6.8 (ns) using the molecular weight of the protein input into the Nick Anthis Protein Correlation Time Calculator [Anthis, 2013]. This value is approximate as the shape of the protein is not taken into account. The empirical formula value was chosen as it incorporates a best-fit dataset from the NESG website [NESG Wiki, 2011].

ARIA assignment of NOE cross-peaks is automated and begins by determining every potential assignment for each peak. The volume of each peak is then transformed into a distance restraint, calculated using the isolated spin pair approximation (ISPA) shown in equation 2.3 [Thomas *et al.*, 1991].

$$r_{ij} = r_{ref}(a_{ref}/a_{ij})^{\frac{1}{6}} \tag{2.3}$$

where,
r_{ij} is the inter-proton distance,
a_{ij} is the corresponding 2D NOE peak intensity,
r_{ij} is the inter-proton distance
r_{ref} is a known inter-proton distance
a_{ref} is a known peak intensity

The use of this equation assumes that the mixing time is short so that each peak intensity is the result of relaxation between just two spins and also that the internal motions are negligible [Keepers and James, 1984].

ARIA then performs several iterations to identify incorrect assignments as well

as noise. The list of restraints can then be corrected by removing noise peaks and peak assignments which are unlikely, to create a new structure ensemble using the new list, which is then analysed again in the next iteration. This is repeated for seven cycles and iteration 8 is the output [Rieping *et al.*, 2006].

2.5.7 NMR parameters

Table 2.9: NMR acquisition parameters – where a range is shown, parameters were optimised for the sample.

		Increments			**Spectral width (ppm)**		
Experiment	**Scans**	**^{1}H**	**^{15}N**	**^{13}C**	**^{1}H**	**^{15}N**	**^{13}C**
[^{1}H, ^{15}N]-HSQC	8-64	2048	128-256	-	15	30	-
[^{1}H-^{13}C]-HSQC	32	1024	-	512	15	-	80
HNCACB	32	2048	48-64	132-162	15	30	75
CBCA(CO)NH	24	2048	48-64	128	15	30	75
HNCO	8	2048	48-64	128	15	30	22
HNCACO	32	2048	48-64	124	15	32	22
CC(CO)NH	32	2048	42-48	164	15	32	75
HCCCONH	32	2048	42-48	164 (^{1}H)	15	32	13 (^{1}H)
TROSY-AR	32	2048	-	256	15	-	40
^{13}C NOESY	16	2048	196 (^{1}H)	64	15	13 (^{1}H)	75
^{15}N NOESY	16	2048	48	196 (^{1}H)	15	32	13 (^{1}H)
NOESY AR	16	2048	512 (^{1}H)	1	15	13 (^{1}H)	75
HBCBCGCDHD	512	2048	-	96	14	-	30
HBCBCGCDCEHE	512	2048	-	78	14	-	30
HCCH-TOCSY	8	2048	196 (^{1}H)	64	15	13 (^{1}H)	75
H-CCH-TOCSY	8	2048	196 (^{13}C)	64	15	75 (^{13}C)	75

NMR spectra were acquired on either a Bruker Avance III 800 MHz spectrometer or a Bruker Avance I 500 MHz spectrometer. Spectral parameters are displayed in table 2.9.

2.5.8 NMR titrations

2.5.8.1 Peptide constructs

OCT4 peptide, sequence displayed in table 2.10, was either synthesised and purified by PeptideSynthetics or expressed recombinantly before purification. Smad7 peptide, sequence also displayed in table 2.10, was expressed recombinantly before purification.

Table 2.10: Amino acid sequence of OCT4 peptide.

	Peptide amino acid sequence
OCT4 peptide	GVGPGSEVWGIPPCPPPYEFCGGMAYCGPQ
Smad7 peptide	SRLCELESPPPPYSRYPMD

2.5.8.2 GB1-WW2:OCT4 peptide titration

GB1-WW2 was titrated with a synthetic OCT4 peptide sample (PeptideSynthetics). Due to initial peptide solubility issues, the peptide was added to the GB1-WW2 sample by mass. A [^{1}H, ^{15}N]-HSQC was acquired after each addition and ratios were calculated. A 0.4 mM sample of ^{15}N GB1-WW2 was prepared as in section 2.5.1 and OCT4 peptide was added directly to the NMR tube at each step. The tube was then inverted 10 times to ensure thorough mixing. Table 2.11 shows the concentration of the OCT4 peptide at each ratio.

Table 2.11: GB1-WW2 titration with OCT4 peptide by mass.

Ratio GB1-WW2:OCT4	**OCT4 peptide concentration (mM)**
1:0	0
1:0.375	0.15
1:1.125	0.45
1:1.75	0.7
1:2.75	1.1
1:4.5	1.8
1:7	2.8
1:8.75	3.5

2.5.8.3 GB1-WW2:Smad7 peptide titration

The ^{15}N His-GB1-WW2 sample was expressed and purified as in section 2.3.1.3. A recombinant Smad7 peptide sample, expressed and purified by a previous group member, was provided in water. Two 600 μL NMR samples were prepared, the start titration point at 1:0 molar ratio of His-GB1-WW2:Smad7 peptide, and the end titration point at 1:10 molar ratio, both with 0.16 mM ^{15}N labelled His-GB1-WW2. The final titration point with 1:10 molar ratio, contained 1.6 mM Smad7 peptide. The ^{15}N labelled His-GB1-WW2 sample was previously dialysed into NMR buffer and so only required the addition of 10% D_2O, 2% NaN_3, 2% DSS and 15 mM DTT. The Smad7 peptide sample also required the addition of 20 mM NaHPO4 and 50 mM NaCl to ensure the buffer remained constant throughout the titration experiments.
A [^{1}H, ^{15}N]-HSQC was acquired for both samples, followed by [^{1}H, ^{15}N]-HSQC's for each subsequent titration point. Peptide concentration was increased in the initial

1:0 sample by removing an aliquot and replacing it with a calculated amount of the 1:10 sample. Table 2.12 shows the volumes added and removed from each sample. Spectra were processed as in 2.5.3 and analysed in CCPN Analysis software.

2.5.8.4 GB1-WW3 OCT4 peptide titration

^{15}N GB1-WW3 was expressed and purified as in section 2.3.1.3 and unlabelled lyophilised OCT4 peptide was prepared as described in section 2.4.6. GB1-WW3 was dialysed into NMR buffer and OCT4 peptide was resuspended in NMR buffer. Both samples were then adjusted to contain 10% D_2O, 2% NaN_3, 2% DSS and 15 mM DTT. GB1-WW3 (0.3 mM) was titrated with recombinant OCT4 peptide using two 500 μL NMR samples per titration prepared as indicated in table 2.13, one at the start concentration of OCT4 (0 mM) and one at the final concentration (3.26 mM).

A [^{1}H, ^{15}N]-HSQC was acquired for both samples, followed by [^{1}H, ^{15}N]-HSQC's for each subsequent titration point. Peptide concentration was increased in the initial 1:0 sample by removing an aliquot and replacing it with a calculated amount of the 1:10 sample. Spectra were processed as in 2.5.3 and analysed in CCPN Analysis software.

Table 2.12: GB1-WW2 titration with recombinant Smad7 peptide. Data points were achieved by the removal of the indicated volume from the initial sample (1:0) and replacing with the same volume of the 1:10 sample.

Protein:	GB1-WW2				[protein] =	0.16	mM (this protein/buffer solution is stockA)	
Ligand:	Smad7 Peptide				[ligand] =	1.6	mM in stockA; this solution is stockB	
Buffer:	20 mM sodium phosphate, 50 mM NaCl, pH 6.8					in 500 µL		
sample #	volume of protein solution /uL	volume of protein/ligand solution /uL	final [protein] /mM	final [ligand] /mM	volume out/in /uL	molar ratio ligand:protein		
1 (stockA)	500	0	0.16	0	0.00	0		
2			0.16	0.016	5.00	0.1		
3			0.16	0.032	5.05	0.2		
4			0.16	0.08	15.31	0.5		
5			0.16	0.16	26.32	1		
6			0.16	0.32	55.56	2		
7			0.16	0.64	125.00	4		
9			0.16	0.96	166.67	6		
10			0.16	1.28	250.00	8		
11 (Stock B)	0	1143.89	0.16	1.6	500.00	10		
	total volume of protein/ligand soln needed for titration =				1143.89			

Table 2.13: GB1-WW3 titration with OCT4 peptide. Data points were achieved by the removal of the indicated volume from the initial sample (1:0) and replacing with the same volume of the 1:10 sample.

Protein:	WW3		[protein] =	0.3	mM (this protein/buffer solution is stockA)
Ligand:	OCT4 Peptide		[ligand] =	3.26	mM in stockA; this solution is stockB
Buffer:	20mM sodium phosphate, 50mM NaCl, pH 6.8			in 500 uL	

sample #	volume of protein solution /uL	volume of protein/ligand solution /uL	final [protein] /mM	final [ligand] /mM	volume out/in /uL	molar ratio ligand:protein
1 (stockA)	500	0	0.3	0	0.00	0
2			0.3	0.075	11.50306748	0.25
3			0.3	0.15	11.77394035	0.5
4			0.3	0.225	12.05787781	0.75
5			0.3	0.3	12.35584843	1
6			0.3	0.45	25.33783784	1.5
7			0.3	0.75	53.38078292	2.5
9			0.3	1.5	149.4023904	5
10			0.3	2.25	213.0681818	7.5
12 (stockB)	0	500.00	0.3	3.26		10.86666667
	total volume of protein/ligand soln needed for titration =				477.38	

2.5.8.5 Dissociation constant calculation

CCPN Analysis was used to calculate the K_d of residues experiencing fast-exchange upon ligand binding. The change in chemical shift (ppm) was weighted according to the spectral dispersion and then fit using the following equation:

$$y = A(B + x - \sqrt{(B + x)^2 - 4x}) \tag{2.4}$$

where,
$A = \Delta\delta_\infty/2$
$B = 1 + K_d/a$
$x = b/a$
$y = \Delta\delta_{obs}$
a = total protein concentration
b = total ligand concentration
$\Delta\delta_{obs}$ = change in chemical shift
$\Delta\delta_\infty$ = difference between the chemical shift at the start point and saturation point.

Peaks which fit poorly with the equation above or were not in fast-exchange were ignored and the binding site was identified by significant peak trajectories upon binding to the ligand. Final K_d values were calculated as an average of the binding site residues.

2.5.8.6 WW2 temperature series

A temperature series of [^{1}H, ^{15}N]-HSQCs of GB1-WW2, starting at 298 K and decreasing to 283 K was acquired using a Bruker Avance I 500 MHz spectrometer. 30 minutes was left between each temperature reduction to allow the sample to equilibrate.

2.6 Isothermal Titration Calorimetry

2.6.1 ITC sample preparation

Isothermal titration calorimetry (ITC) was carried out using a MicroCal PEAQ-ITC calorimeter (Malvern) at the John Innes Centre, Norwich, UK. Protein (GB1-WW2) and peptides (Smad7 and OCT4) were dialysed three times into the same 5 L buffer (20 mM Na_2HPO_4, 50 mM NaCl, pH 6.8) The concentration of each sample was confirmed using the absorbance at 280 nm and adjusted using the final dialysis buffer.

The sample cell was filled with ∼300 μL of GB1-WW2 protein and the syringe was loaded with ∼70μL of ligand. Once the experiment begins, the heat signal (the baseline) is left to stabilise with the rotating syringe placed down into the cell, ensuring that the two samples will be completely mixed after each addition. The area of each injection peak is then calculated using the change from the baseline. If an interaction takes place, and chosen parameters allow, throughout the titration, the area of each injection point should reduce as the protein becomes saturated with the ligand. Controls of buffer into buffer, ligand into buffer and buffer into ligand allow the heat of dilution to be subtracted from the results. After saturation has been reached, the peaks detected are equal to the ligand dilution.

2.6.2 ITC parameters

ITC experiments were carried out at 25 °C and involved a series of 18 injections of 2 μL each, plus an initial injection of 0.4 μL (for a total of 19 injections). The initial injection is discarded before data analysis as it typically shows misleading data, for example, a smaller peak area than expected, caused by diffusion of the ligand into the cell through the needle tip during temperature equilibration or small air bubbles present in the needle. Injection duration was 4 s every 150 s and with a stir speed of 750 rpm.

Buffer-buffer, Buffer-protein, peptide-buffer control experiments were also carried out and were subtracted from results using the MicroCal PEAQ-ITC Control software. All analysis and experiment design were carried out within the MicroCal PEAQ-ITC Control software.

Chapter 3

Structure determination of GB1-WW3

3.1 Introduction

3.1.1 Structure determination

Structural information for determining a protein structure is obtained primarily through distance restraints and torsion angles. The Nuclear Overhauser Effect (NOE) is defined as the transfer of magnetisation from one nuclei to another, through space, via cross-relaxation [Freeman and Anderson, 1962] and is observed as an intensity change of one resonance when the intensity of a neighbouring resonance (not necessarily directly bonded) is altered. Typically, the maximum distance between two protons able to generate an observable NOE peak is 6Å in proteins. ^{15}N-NOESY-HSQC correlates nearby proton nuclei to amide proton resonances and ^{13}C-NOESY-HSQC correlates nearby protons to the CH_x group protons. The intensity of the NOE peaks is proportional to the distance of the corresponding proton from the CH or amide group protons and are labelled with the chemical shift (δ) of each nucleus. This allows a set of distance restraints

between specific nuclei to be determined, making the calculation of a structure ensemble possible [Cavanagh *et al.*, 2010]; [Teng *et al.*, 2005].

Various programs exist to aid the process of structure calculation and in this project, Ambiguous Restraints from Iterative Assignment (ARIA) was used [Rieping *et al.*, 2006]. The software calculates distance restraints from NOESY peaks and uses them in conjunction with dihedral restraints, which describe the rotational freedom of the amide and carboxyl groups around the amino acid α-carbon. The dihedral restraints were generated using Dihedral ANgles from Global Likelihood Estimates (DANGLE) software embedded in CCPN Analysis. Backbone torsion angles are predicted using the amino acid sequence and backbone chemical shifts. As the chemical shift is closely related to the secondary structure, the chemical shift can be matched to a database of high-resolution structures to predict the backbone torsion angles of that region [Shen *et al.*, 2009].

ARIA software is able to automate the process of NOE assignment before using these assignments for structure calculation [Rieping *et al.*, 2006]. This avoids the manual assignment of hundreds to thousands of NOE peaks saving significant time. Peaks are initially assigned ambiguously to multiple protons and several rounds of iterative structure calculations allow ARIA to discard unlikely assignments.

3.1.2 WW3 domain

Relatively little is known about the third WW domain of WWP2. In 2009, Xu *et al.* described the interactions between the various domains of WWP2 and His-tagged binding partner OCT4 [Xu *et al.*, 2009]. The findings showed that the C2 and HECT domains were unable to bind His-OCT4 as opposed to the WW domain region which associated with OCT4 in GST pull-down experiments. Subsequent pull-downs with the individual domains WW1, WW2, WW3 and WW4 indicated that an interaction occurred across all four domains; however, significant preference for the second WW2 domain was observed. The affinity of the interaction between the WW domains WW1, WW3 and WW4, for the OCT4

peptide, is therefore expected to be lower than that of WW2. When compared with the WW2 sequence, the remaining three WW domains shared the following sequence identity; WW1 = 73%, WW3 = 74% and WW4 = 60 %. With 74% sequence identity to the WW2 domain, it was of interest to quantify the differences in binding affinity between the WW2 domain and WW3 domain interactions with the OCT4 peptide.

3.1.3 GB1 solubility tag

For structural or dynamic studies using NMR, protein samples are required at high concentrations. It is also important that this concentration does not lead to aggregation or precipitation of the sample as this leads to poor quality NMR spectra. Previous members of the laboratory group successfully expressed GST-WW1-4; however, cleavage of the GST tag led to aggregation. Therefore, a tag that would aid solubility whilst not interfering with protein structure, and hence would not need to be removed for NMR experimentation, was required. The laboratory group determined that expression and purification of WW domains with characteristics suitable for NMR could be achieved by expression with fusion partner GB1 (B1 domain of streptococcal protein G). GB1 is a small protein which can increase expression level and improve sample solubility and stability [Cheng and Patel, 2004]; [Zhou and Wagner, 2010]. The 56-residue construct is also reported to be able to retain the secondary structure of the target protein [Wishart *et al.*, 1995], meaning the NMR experiments can be acquired for the sample without needing to cleave the GB1 tag. Therefore, the expression of WW domain constructs as fusion proteins with GB1 was established and purification was optimised.

3.1.4 Experimental aims

The aim of this chapter is to provide information on the interaction between GB1-WW3 and OCT4 peptide, as well as the structure of recombinant GB1-WW3,

using solution-state NMR.

For structure calculations, a number of NMR experiments had been acquired and processed previously by various group members and the majority of the backbone assignment had been completed. A ^{15}N, ^{13}C labelled GB1-WW3 NMR sample was prepared for additional experiments allowing side-chain assignment. These included H-CCH-TOCSY and HCCH-TOCSY spectra for aliphatic side-chain assignments, 2D HBCBCGCDHD and 2D HBCBCGCDCEHE experiments [Yamazaki *et al.*, 1993] for assignment of aromatic side-chains (histidine Hδ1, tryptophan Hδ1, tyrosine Hδ,ϵ and phenylalanine Hδ,ϵ) and aromatic CH correlations using a ^{13}C-TROSY-Aromatic spectrum. ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC experiments were also necessary to determine through-space distance restraints, which are required for structure determination. WeNMR [Wassenaar *et al.*, 2012] ARIA (Ambiguous Restraints for Iterative Assignment) is a Webportal [Nilges, 2016] for calculating structure ensembles and the output structures are then evaluated using the Protein Structure Validation Suite (PSVS).

3.2 Results

3.2.1 ^{15}N, ^{13}C labelled GB1-WW3 expression and purification

His-GB1-WW3 was overexpressed in BL21(DE3)* (*E. coli*) using conditions optimised by a member of the Chantry laboratory group (0.8 mM IPTG, overnight induction at 30 °C), in 3 L ^{15}N, ^{13}C labelled Minimal Essential Medium (MEM). The protein was then purified using Ni-NTA Immobilized Metal Affinity Chromatography (IMAC). Figure 3.1A shows His-GB1-WW3 expression, pre and post-induction and figure 3.1B shows subsequent Ni-NTA column purification.

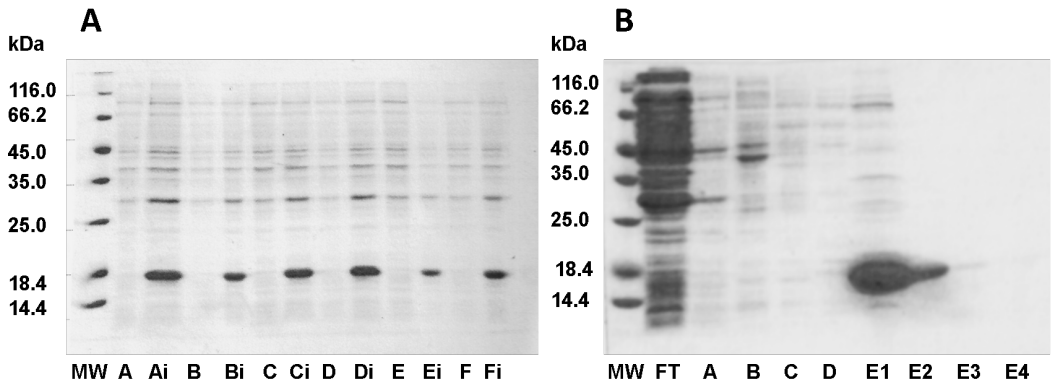


Figure 3.1: 15% SDS-PAGE analysis of **(A)** ^{15}N, ^{13}C His-GB1-WW3 expression with pre and post-induction samples. A clear His-GB1-WW3 band at ∼18kDa is present in all post-induction samples (marked with an i) indicating expression was successful. **(B)** Purification of ^{15}N, ^{13}C labelled His-GB1-WW3 using a Ni-NTA column. Purified protein can be seen in the elution lanes E1 and E2. Small levels of contaminants are visible in sample E1; however, these are much lower in concentration than the His-GB1-WW3 protein. Imidazole wash concentrations: wash A = 1 mM, wash B = 10 mM, wash C = 20 mM, wash D = 30 mM, elution E = 250 mM. 1 μL samples loaded.

Purified His-GB1-WW3 was present in the E1 and E2 samples and the E1 sample had low-level contaminants which would be negligible in NMR experiments due to the high concentration of protein. Two units of thrombin were added to the elution aliquots for every mg of protein present. The samples were dialysed into Tris-HCl buffer at room temperature until the digest was complete (figure 3.2A).

The successfully digested protein was purified using a Ni-NTA column to remove the polyhistidine tag from the sample as shown in figure 3.2B, where the protein is in the flow through and wash samples and the His tag is separated in the elution. A benzamidine purification column was used to remove the thrombin from the sample. Thrombin binds to the column and hence is separated from the protein and is removed in the elution step. Figure 3.2C shows the benzamidine purification samples. GB1-WW3 was present in the wash samples (A, B, C and D) and so were combined and concentrated and prepared as two NMR samples; one for structure calculation and a second for titration with OCT4 peptide.

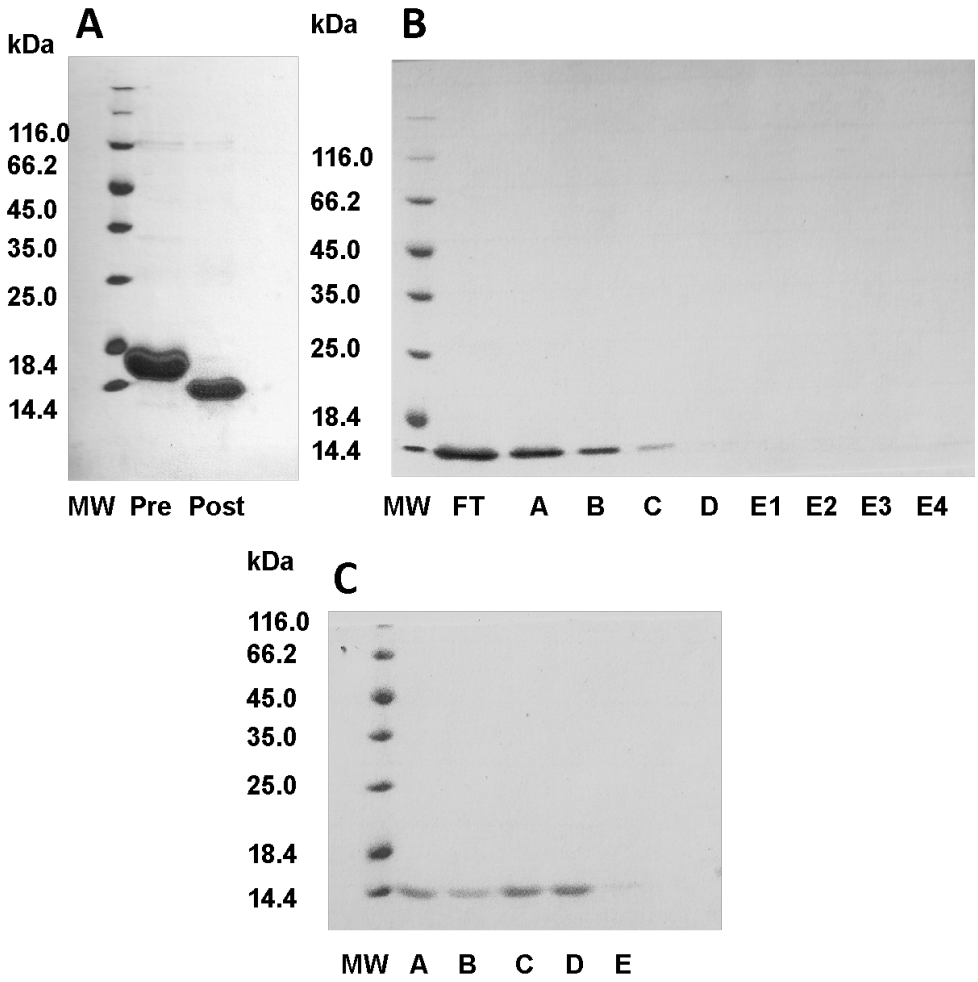


Figure 3.2: 15% SDS-PAGE analysis of **(A)** His-GB1-WW3 thrombin digest showing the sample pre and post-digestion. **(B)** Purification of GB1-WW3 using Ni-NTA column to remove the polyhistidine tag. **(C)** Removal of thrombin from the sample using a benzamidine column. Purified GB1-WW3 protein was present in washes A, B, C and D. 10 μL samples loaded.

3.2.2 Backbone assignment

Previous work on the GB1-WW3 structure had yielded a 92% backbone assignment (90 of 98 residues), equivalent of full assignment as the missing assignments included only the initial three N-terminal residues, which are commonly very weak and are subsequently not observed, and five proline residues

which are unobservable due to their lack of amide proton. A full list of the assigned residues and resonances is included in Appendix A. When studying the spectra, 67L was found to be misassigned, and so all of the NH, Cα and Cβ resonance assignments were then checked using the backbone assignment method described in section 1.6.1.4. The previously acquired spectra ([^{1}H, ^{15}N]-HSQC, HNCACB, CBCACONH, HNCO, HNCACO) were used in combination with newly acquired H-CCH-TOCSY and HCCH-TOCSY spectra to confirm current backbone assignments and make adjustments where necessary. CCPN Analysis [Vranken *et al.*, 2005] software was used to analyse all spectra and figure 3.3 shows fully assigned GB1-WW3 [^{1}H, ^{15}N]-HSQC. Peaks which are labelled but not visible in this [^{1}H, ^{15}N]-HSQC figure are below the contour level, which was chosen for best observation of the majority of peaks.

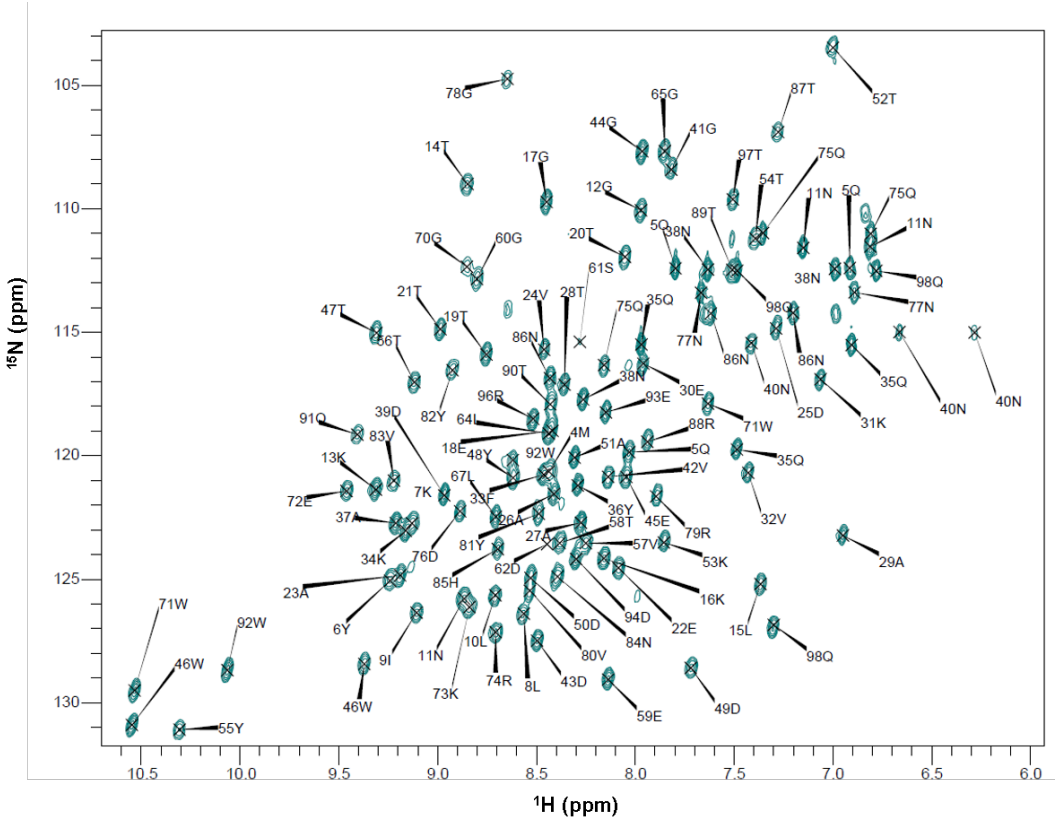


Figure 3.3: A [^{1}H, ^{15}N]-HSQC spectrum of ~2 mM ^{15}N, ^{13}C labelled GB1-WW3. 92% of residues were assigned and labels are present for backbone and side-chain resonances where possible. This experiment was performed using a Bruker Avance III 800 MHz spectrometer at 298 K. Sample buffer: 20 mM sodium phosphate buffer, 50 mM NaCl, pH 6.8.

3.2.3 Side-chain assignments

The following experiments had also been previously acquired: H(CC)(CO)NH, CC(CO)NH, ^{15}N TOCSY and [^{1}H-^{13}C]-HSQC. Aliphatic side-chain assignment had been previously attempted using the H(CC)(CO)NH, CC(CO)NH and [^{1}H-^{13}C]-HSQC experiments but was not complete. For more thorough aliphatic side-chain assignment, H-CCH-TOCSY and HCCH-TOCSY experiments were used to assign as many of the remaining side-chain ^{1}H and ^{13}C resonances as possible.

As the Cα and Cβ carbon values were already known after the backbone assignment, CH cross-peaks from residues which have unique Cα and Cβ's could be identified in the [^{1}H-^{13}C]-HSQC. The TOCSY spectra correlate each ^{1}H to the CH group of the sidechain (HCCH-TOCSY) and from each ^{13}C to the CH group of the sidechain (H-CCH-TOCSY). Therefore, by navigating from the CH peaks in the [^{1}H-^{13}C]-HSQC to the HCCH and H-CCH-TOCSY spectra, the chemical shift values of the ^{1}H and ^{13}C nuclei within that side-chain could be assigned.

HBCBCGCDHD, HBCBCGCDCEHE ^{13}C-TROSY and ^{13}C-NOESY-HSQC-Aromatic experiments were also acquired, centred in the ^{13}C aromatic region, and were used to assign the majority of the aromatic side-chains.

3.2.4 Structure determination

^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC experiments were also acquired to determine through-space distance restraints. NOE (Nuclear Overhauser Effect) peaks were picked manually in CCPN Analysis software. The assignment shift list, NOESY peak lists and the amino acid sequence were then uploaded to the WeNMR ARIA (Ambiguous Restraints for Iterative Assignment) Webportal [Nilges, 2016] along with dihedral restraints calculated from CCPN Analysis. Assignment of NOE peaks was accomplished within ARIA and peak volumes were used to calculate distance restraints. 20 final refined structures were acquired for each run.

ARIA assignment of NOE cross-peaks is automated and begins by determining every potential assignment for each peak. The volume of each peak is then transformed into a distance restraint. Several iterations are then performed, each with 50 structures per iteration, to identify incorrect assignments and create an updated assignment list. The final output is generated after the 8th iteration, which uses 100 structures, and produces an ensemble of the 20 lowest energy structures. Where violations occurred, peaks were manually investigated and where assignments for NOEs could be clearly identified, they were assigned.

Figure 3.4 shows the ARIA ensemble (20 models) overlaid and aligned at the N-terminus GB1 domain, exported from Chimera software [Pettersen *et al.*, 2004]. Predicted secondary structure information was highlighted as followed: α helix is shown in red, β sheet in blue and random coil regions in light grey.

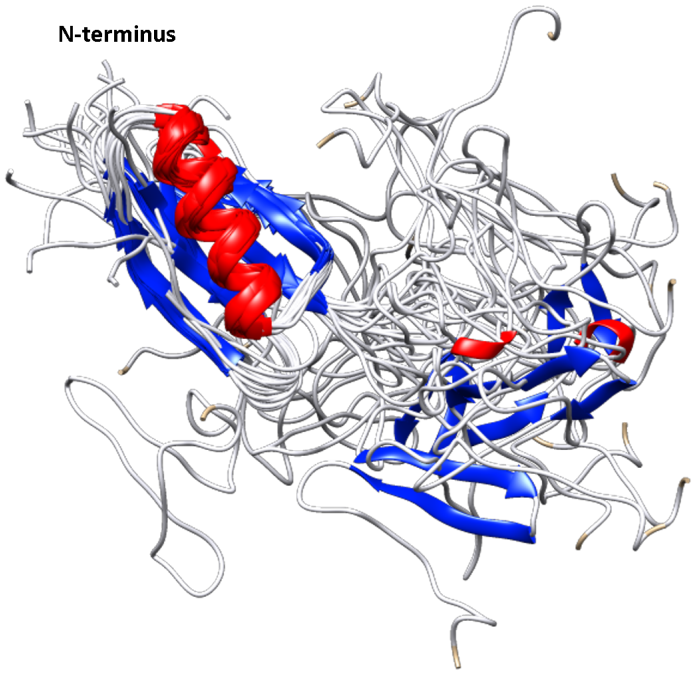


Figure 3.4: The GB1:WW3 ribbon diagram of ARIA ensemble (20 models) aligned at the N-terminus GB1 domain exported from Chimera. There is a disordered linker region between the two domains and so the WW3 domain is therefore not in a fixed position. α helix is shown in red, β sheet in blue and random coil regions in light grey. Figure exported from Chimera [Pettersen *et al.*, 2004].

Immediately it is clear that the GB1 domain at the N-terminus shows more consistent structures than the WW3 domain at the C-terminus. The flexible linker region between the two domains, with few NOE restraints, leads to the WW3 domain being oriented in different positions in each model. Despite this, it is clear that the β sheets are present in very few of the models displayed here, meaning that the Chimera software is not recognising many of the β sheets expected. This is due to inaccuracies in the structure calculation, leading to regions which do not match a β sheet arrangement.

The root-mean-square deviation (RMSD) of all backbone atoms of the ensemble for the GB1-WW3 is 7.1 Å which was calculated using the Protein Structure Validation Suite (PSVS) [Bhattacharya *et al.*, 2007]. As mentioned above, when the GB1 domain is aligned, the WW3 domains are spread out in different orientations. This is, therefore, a less accurate way of measuring the RMSD of the protein. Instead, the two individual domains were split, GB1 (residue 1-61) and WW3 (residue 69-98) ignoring the linker region (residue 62-68) and uploaded to PSVS separately. The RMSD average of the GB1 domain alone was found to be 1.9 Å, while the RMSD average of the WW3 domain was 2.9 Å.

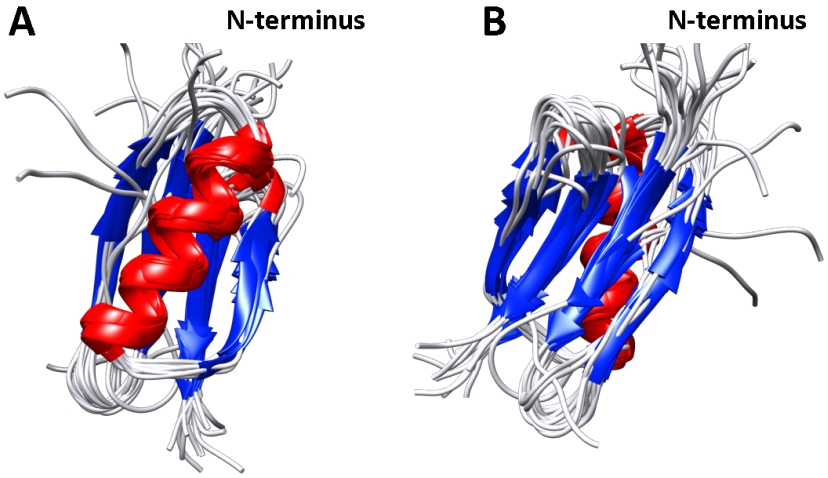


Figure 3.5: The ribbon diagram of refined ARIA ensemble (20 models) showing only the N-terminus GB1 domain exported from Chimera. α helix is shown in red, β sheet in blue and random coil regions in light grey. (A) and (B) are alternative orientations of the same structure. Figure exported from Chimera [Pettersen *et al.*, 2004].

When examined as an individual domain, the GB1 structure here (figure 3.5 holds a similar conformation to other GB1 domain structures found in the Protein Data Bank (PDB). This structure contains a four-stranded β sheet, with a parallel α helix in the middle of the sequence. Figure 3.6 shows the overlay of the GB1 domain (residue 1-61) (coloured as previously) and the solution structure of GB1 domain Protein G (PDB:2J52) (shown in green) from the PDB [Wilton *et al.*, 2008].

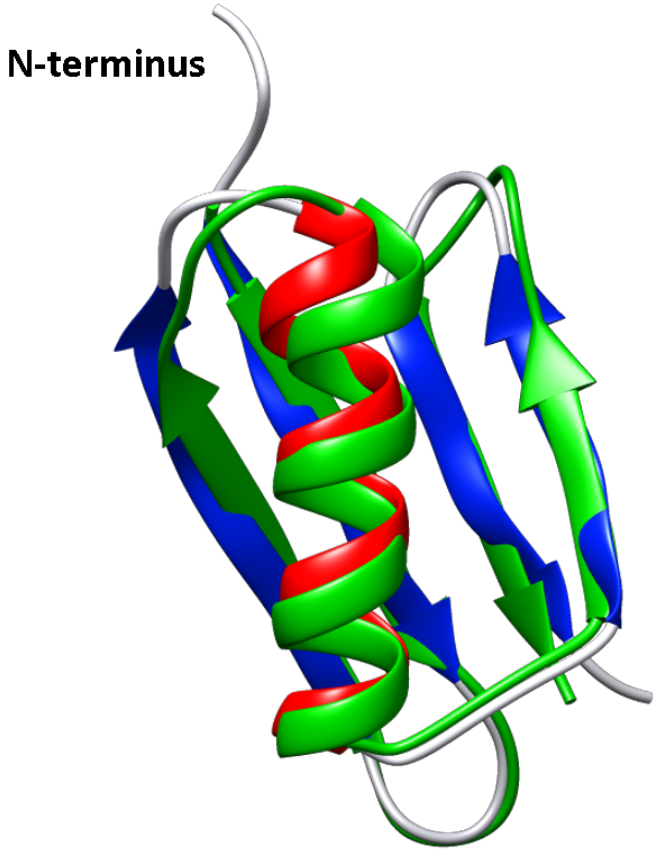


Figure 3.6: The ribbon diagram of the GB1 structure overlaid with the solution structure of GB1 domain Protein G (PDB:2J52). As previously, the GB1 secondary structure is shown as follows: α helix in red, β sheet in blue and random coil regions in light grey. The PDB structure is shown in green. Figure exported from Chimera [Pettersen *et al.*, 2004].

The structure calculated here is fairly similar to the deposited structure for the GB1 domain; however, it does require further refinement. The secondary structure

elements are all present and fold in a very similar way.

When the WW3 domain (residue 69-98) ARIA ensemble (20 models) was overlaid and aligned, far less similarity was observed. Whilst three strands were generally observed, very few were recognised as the expected β structure and a helical loop was also present in two models. Observation of the lowest energy model clearly shows that the C-terminus region is problematic with the third β sheet not forming as expected, as shown in figure 3.17B. Figures were exported from Chimera software [Pettersen *et al.*, 2004].

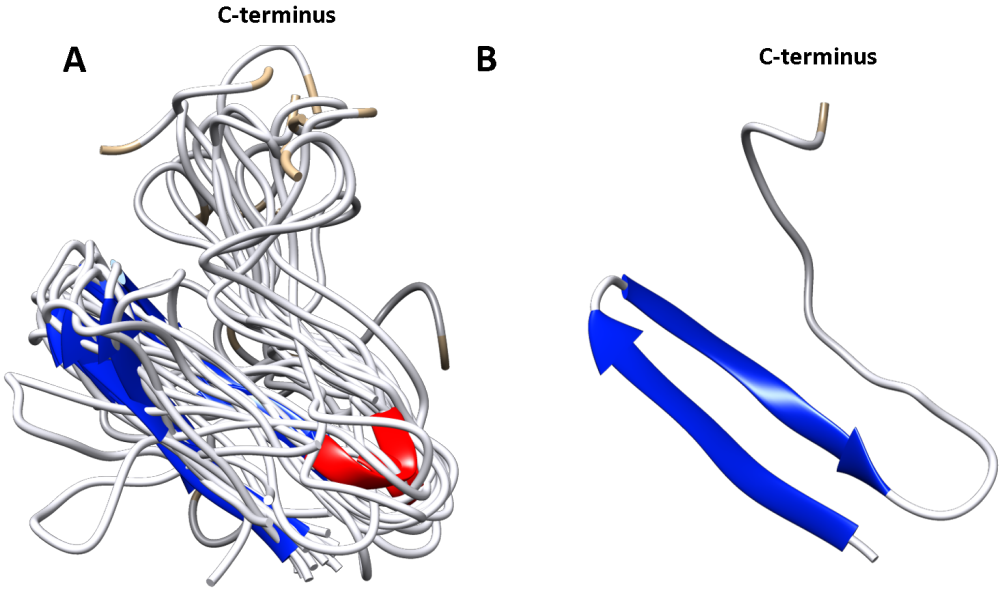


Figure 3.7: **(A)** The ribbon diagram of refined ARIA ensemble (20 models) showing only the C-terminus WW3 domain aligned. α helix is shown in red, β strand in blue and random coil regions in light grey. Whilst there are general similarities, the issues with the structure mean that the expected β sheets are not recognised by Chimera in the majority of structures. **(B)** The ribbon diagram of the lowest energy C-terminus WW3 domain showing two β strands. Inaccuracies in the third strand led to the secondary structure not being predicted. Figure exported from Chimera [Pettersen *et al.*, 2004].

3.2.5 Recombinant OCT4 peptide expression, purification

For titration of GB1-WW3 with OCT4 peptide, OCT4 peptide was expressed recombinantly as a SUMO fusion protein (plasmid provided by Jessica Watt from

the Chantry laboratory). SUMO-OCT4 was transformed in BL21(DE3) (*E. coli*) and induced with 0.8 mM IPTG overnight at 30 °C. Pre and post samples were analysed using SDS-PAGE to confirm induction was successful. The induced cell culture was then sonicated for cell lysis and purified using a gravity flow Ni NTA column. Purification was confirmed using SDS-PAGE (figure 3.8).

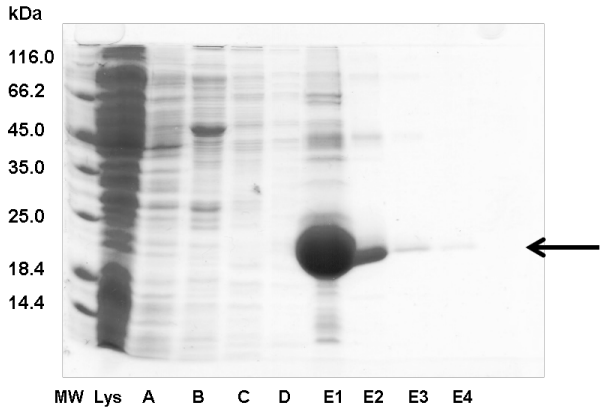


Figure 3.8: An example of a 15% SDS-PAGE gel showing Ni-NTA purification of unlabelled SUMO-OCT4 fusion protein (indicated by an arrow at ∼22 kDa - SUMO-OCT4 runs higher than its molecular weight of ∼15 kDa). The highest concentration of SUMO-OCT4 was found in the elution fractions E1 and E2. Imidazole wash concentrations: wash A = 1 mM, wash B = 10 mM, wash C = 20 mM, wash D = 30 mM, elution E = 250 mM. 10 μL samples were loaded.

Sample E1 shows there are some impurities in the sample; however, the concentration of OCT4 peptide is significantly higher and so the sample was suitable for the NMR titration.

Samples E1 and E2 were combined and dialysed into Tris-Cl buffer, before 500 μL ULP1 per 1 L culture volume, was added to the sample and was left at room temperature overnight with agitation for digestion of the SUMO tag. Analysis with SDS-PAGE showed incomplete digestion and so the sample was left for longer with aliquots taken at regular intervals until the digestion was complete (figure 3.9A).

After digestion, a SUMO tag band could be seen at a slightly lower molecular weight, consistent with the cleavage of the OCT4 peptide. The lower band,

which appears towards the bottom of the post-digest lane was originally suspected to be the OCT4 peptide running with the gel front. This theory was discounted following SDS-PAGE analysis of the Ni-NTA column purification of the post-digestion sample (see figure 3.9B), as this lower band (marked with a red arrow) binds to the column until the elution buffer is added, indicating that the His-tag is still present. Therefore, it is more likely that this band is a His-tagged SUMO degradation band.

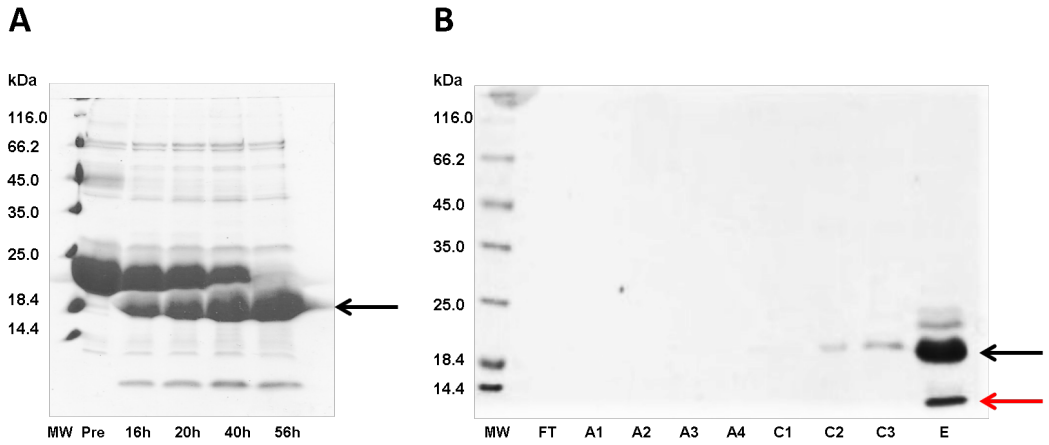


Figure 3.9: **(A)** 15% SDS-PAGE analysis showing pre and post-digest of SUMO-OCT4 peptide. Full digestion was accomplished after 56 hours and is marked with a black arrow. **(B)** 15% SDS-PAGE analysis showing Ni-NTA purification of unlabelled OCT4 peptide. The purified peptide is present in the FT and washes but cannot be observed using Coomassie brilliant blue. The His-SUMO tag is in elution fraction E, indicated by a black arrow. The red arrow shows a SUMO degradation band which still contains the His-tag. Imidazole wash concentrations: wash A = 1 mM, C = 20 mM, elution E = 250 mM. 10 μL samples loaded.

The peptide was not observed on the gel, likely due to its small size and poor ability for binding Coomassie brilliant blue. Silver staining and Tricine gels were attempted as an alternative for visualising the OCT4 peptide but were unsuccessful. The absorbance at 280 nm was therefore used to determine which fractions contained OCT4 peptide and confirmed that protein was present in the flow through and A1 wash. These two fractions were pooled to collect the OCT4 peptide and dialysed into ammonium bicarbonate buffer using low MWCO dialysis membrane and freeze dried to concentrate.

3.2.6 NMR titrations of ^{15}N GB1-WW3 with recombinant OCT4 peptide

Two NMR samples were prepared for the titration of GB1-WW3 with recombinant OCT4 peptide. The first sample (molar ratio 1:0) did not contain the OCT4 peptide and the second sample (the final titration point sample) contained a 10-fold concentration of peptide added by weight, yielding a molar ratio of 1:10. Both samples were adjusted so that the pH was ∼6.8. [^{1}H, ^{15}N]-HSQC spectra were acquired initially for the starting titration point sample and again for the end titration point sample. Aliquots of the 1:0 molar ratio starting titration point sample were then removed and replaced with aliquots of the end titration point sample at molar ratio 1:10, allowing the concentration of OCT4 to increase without affecting the concentration of ^{15}N labelled GB1-WW3. The NMR tube was inverted 5 times and then spun in a hand centrifuge to ensure the sample was well mixed and that all of the sample accumulated at the bottom after each addition. A superimposition of the ten [^{1}H, ^{15}N]-HSQC spectra, ranging from molar ratio 1:0 GB1-WW3:OCT4, up to 1:10, can be seen in figure 3.10. Assignments were propagated from GB1-WW3 assignments and the residue assignments at molar ratio 1:0 are shown. The concentration of OCT4 peptide at each titration point can be seen in table 2.5.8.4.

The titration showed a number of WW3 domain peaks shifting upon the addition of OCT4 peptide, indicating that an interaction took place. These changes occur due to the effect of other atoms on the chemical shift of each amide peak. When a residue experiences a different environment, its own chemical shift is altered too. Peaks which experience the largest changes are therefore more likely to be involved in the binding of WW3 to OCT4 peptide as their environment has changed most substantially. The residues displaying peak migration with the addition of peptide are in fast exchange and the peak trajectory can be tracked and measured. Fast exchange appears in this way due to the conformation of the protein swapping between the bound and unbound. As more peptide is added, a higher population will be in the bound state,

causing the peak to shift as the weighted average of the two conformations is observed. Slow exchange processes were also observed, where peaks disappear and then reappear elsewhere in the spectrum.

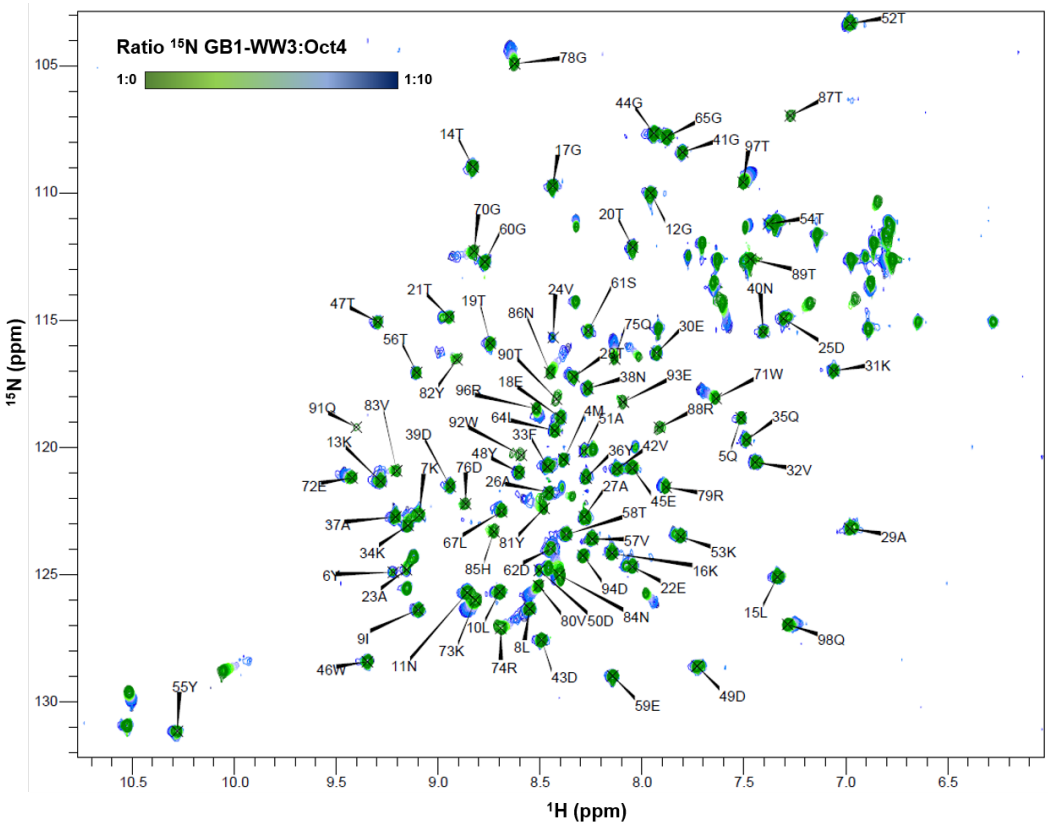


Figure 3.10: Overlaid [^{1}H, ^{15}N]-HSQC spectra of GB1-WW3 titrated with unlabelled recombinant OCT4 peptide. GB1 tag peaks are stationary. Overall there is a small decrease in intensity and number of peaks as the titration progresses. Assignments were propagated from GB1-WW3 assignments and peak labels, therefore, show starting titration points. The experiments were performed using a Bruker Avance III 800 MHz spectrometer at 298 K. Sample buffer: 20 mM Na_2HPO_4, 50 mM NaCl, 15 mM DTT, pH 6.8. The concentration of GB1-WW3 was 0.3 mM throughout.

A large excess of OCT4 peptide was used in an attempt to ensure the saturation point of the interaction was reached. GB1 tag peaks remained stationary throughout the titration, indicating that the tag did not interact with the peptide. Detailed regions of the titration can be seen in figure 3.11**A**, showing GB1 peaks 28T, 38N and 61S are stationary throughout the titration.

86N moved in fast exchange as OCT4 peptide concentration increases and 24V increased in intensity. An additional peak merging out to the left of 28T also gained intensity. Figure 3.11**B** displays 88R disappearing as more OCT4 peptide was added and 71W moving in fast exchange. Red arrows highlight the direction of movement for peaks experiencing fast exchange processes.

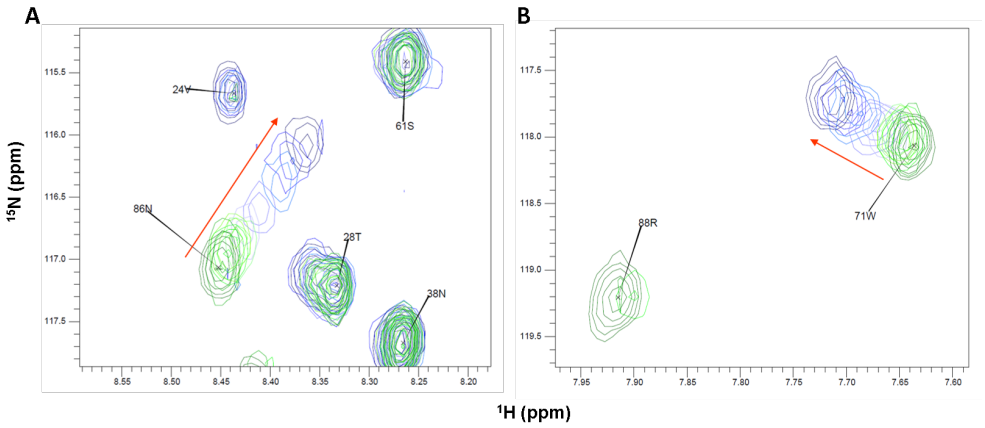


Figure 3.11: Detailed regions from GB1-WW3:OCT4 peptide [^{1}H, ^{15}N]-HSQC titration spectra. Assignments were propagated from GB1-WW3 assignments. (**A**) shows that GB1 peaks 28T, 38N and 61S did not move during the titration. 24V stayed in the same position; however, the intensity did increase as OCT4 peptide was added. 86N can be observed moving in intermediate exchange (peak broadening present), with a red arrow highlighting the direction of movement as OCT4 peptide concentration increases. (**B**) shows residue 88R is present at the start of the titration but disappears as more OCT4 peptide is added. 71W moves in fast exchange with its direction highlighted with a red arrow. The experiments were performed using a Bruker Avance III 800 MHz spectrometer at 298 K. Sample buffer: 20 mM Na_2HPO_4, 50 mM NaCl, 15 mM DTT, pH 6.8. The concentration of GB1-WW3 was 0.3 mM throughout.

The increasing intensity of 24V could indicate that the residue was originally in multiple conformations but becomes more stable as the OCT4 peptide is added. In contrast, there are multiple WW domain residues which disappear over the course of the titration. Whilst some additional peaks are observed appearing towards the end of the titration, there is an overall decrease in the number of peaks by 5. Residues involved in slow exchange processes, which disappear as OCT4 peptide is added, may be appearing elsewhere in the spectrum. The overall decrease of 5 peaks after the addition of 10-fold OCT4 peptide suggests that some residues were experiencing an intermediate exchange process when partially bound to OCT4 peptide.

Some peaks were observed to move after the first titration point, at ratio 1:0.25, as can be seen clearly for residue 88R in figure 3.11**(B)**. Peak positions were also observed moving up to the 1:10 ratio suggesting that the saturation point was not achieved. Therefore, it is unknown whether the missing residue peaks would have appeared after full saturation of peptide in a more stable conformation.

Shift trajectories upon addition of the OCT4 peptide were plotted against residue number and are displayed in the graph in figure 3.12. The PSIPRED webserver v3.3 [Jones *et al.*, 2018] was used to predict areas of secondary structure [Buchan *et al.*, 2013] and figure 6.11 shows three predicted β sheet regions in yellow on the schematic diagram.

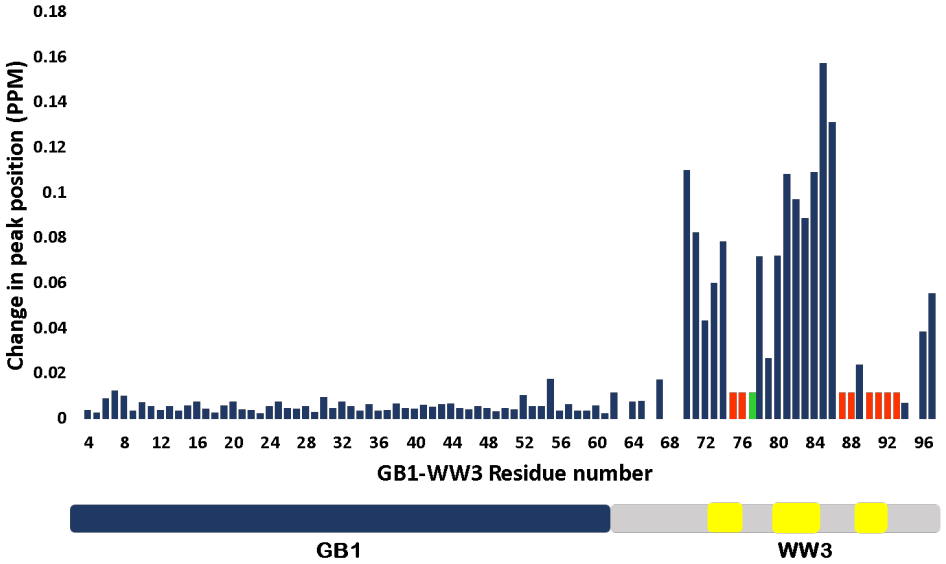


Figure 3.12: The trajectories (in ppm) of GB1-WW3 backbone amide peaks at the start and end of the titration with OCT4 peptide. GB1 peaks (1-61) remain fairly stable throughout the titration, as opposed to the WW domain residues (62 and above), which show larger trajectories. The residues which could not be assigned in the [^{1}H, ^{15}N]-HSQC were plotted at 0.01 ppm in green. Residues which were assigned in the initial titration point, but could not be followed to the end (e.g. disappearing or slow exchange peaks), are shown in red, also at 0.1 ppm. Proline residues cannot be observed and are represented as gaps. Secondary structure prediction of the WW3 domain using the PSIPRED webserver v3.3 predicted areas of secondary structure [Jones *et al.*, 2018]; [Buchan *et al.*, 2013] and the three yellow regions represent predicted β strands.

Shift changes in the spectra are not evenly weighted due to the spectral dispersion of nitrogen being larger than that of hydrogen. Therefore, shift

changes were weighted appropriately to allow for this by dividing nitrogen shift values by 10. The change in peak position was then calculated using Pythagoras' theorem (equation 3.1).

$$a^2 + b^2 = c^2 \tag{3.1}$$

where a = difference in ^{1}H shift values, b = difference in adjusted ^{15}N shift values and c = change in peak position.

Peak positions of GB1 peaks (1-61) were fairly stable throughout the titration, indicating OCT4 did not interact with the tag. The majority of WW domain residues (62 and above) with full residue assignments throughout the titration, show much larger trajectories than the GB1 tag. Residues which were not present in any of the [^{1}H, ^{15}N]-HSQC spectra are plotted at 0.01 ppm in green (the unassigned 77N). Residues which were assigned in the initial titration point but could not be assigned at the end point (e.g. disappearing or slow exchange peaks) are shown in red. These residues could not be followed as the titration progressed and were likely not in fast exchange, for example appearing elsewhere in the spectrum (slow exchange) or subjected to peak broadening (intermediate exchange). Proline residues cannot be observed in [^{1}H, ^{15}N]-HSQC spectra and so are represented as gaps (63P, 66P, 68P, 69P, 95P).

The peaks displaying the largest (in excess of 0.1 ppm) measurable trajectories (70G, 81Y, 84N, 85H and 86N) were generally observed to be in the intermediate exchange regime, causing peak broadening and loss of intensity during the titration before beginning to regain it towards the end. This can be explained as the exchange process is governed by the frequency difference between bound and unbound states in relation to the binding on and off rates. For this protein, when the resonances of the bound and unbound states were close together, they experience fast exchange as defined by $k_{\text{ex}} >> |v_{(\text{A})} - v_{(\text{B})}|$. This was observed for a number of residues in this titration and can be identified

by a peak at the average of the population-weighted states, which migrates as the population of the bound state increases (e.g. as more peptide is added). Peaks experiencing intermediate exchange are defined by ($k_{ex} \sim |v_{(A)} - v_{(B)}|$), where the rate of exchange is roughly the same as the frequency difference between states. Therefore, peak trajectories for these residues are larger and demonstrate a greater involvement in the binding interaction with OCT4 peptide. Figure 3.13 shows residue 70G experiencing intermediate exchange; however, as saturation was not achieved in this titration, the intensity was not able to return to the initial level.

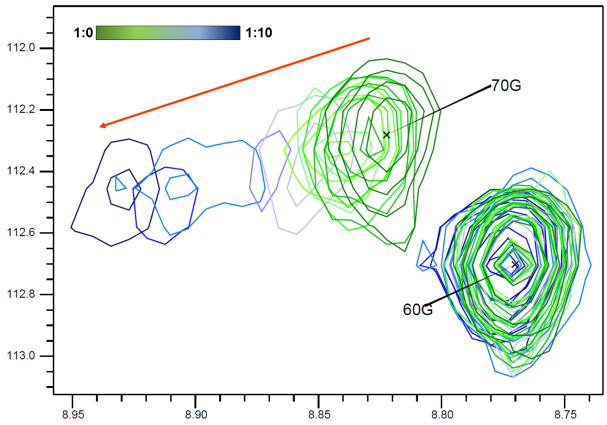


Figure 3.13: [^{1}H-^{15}N]-HSQC spectra of GB1-WW3 titrated with unlabelled recombinant OCT4 peptide showing residue 70G in intermediate exchange. The experiments were performed using a Bruker Avance III 800 MHz spectrometer at 298 K. Sample buffer: 20 mM Na_2HPO_4, 50 mM NaCl, 15 mM DTT, pH 6.8. The concentration of GB1-WW3 was 0.3 mM throughout.

The disappearance of several peaks throughout the titration is likely due to peak broadening as a result of being in intermediate exchange. If binding had reached the saturation point, it was expected that these peaks would have regained intensity as they moved into the bound conformation.

Individual residues which were present throughout the titration and also displayed fast exchange processes, were used to estimate the K_d using CCPN Analysis software. The change in resonance position and the saturation point for each residue were fitted within the software. Figure 3.14 shows the CCPN Analysis software output of resonance shift changes in ppm vs the molar ratio of protein to ligand for four residues in fast exchange.

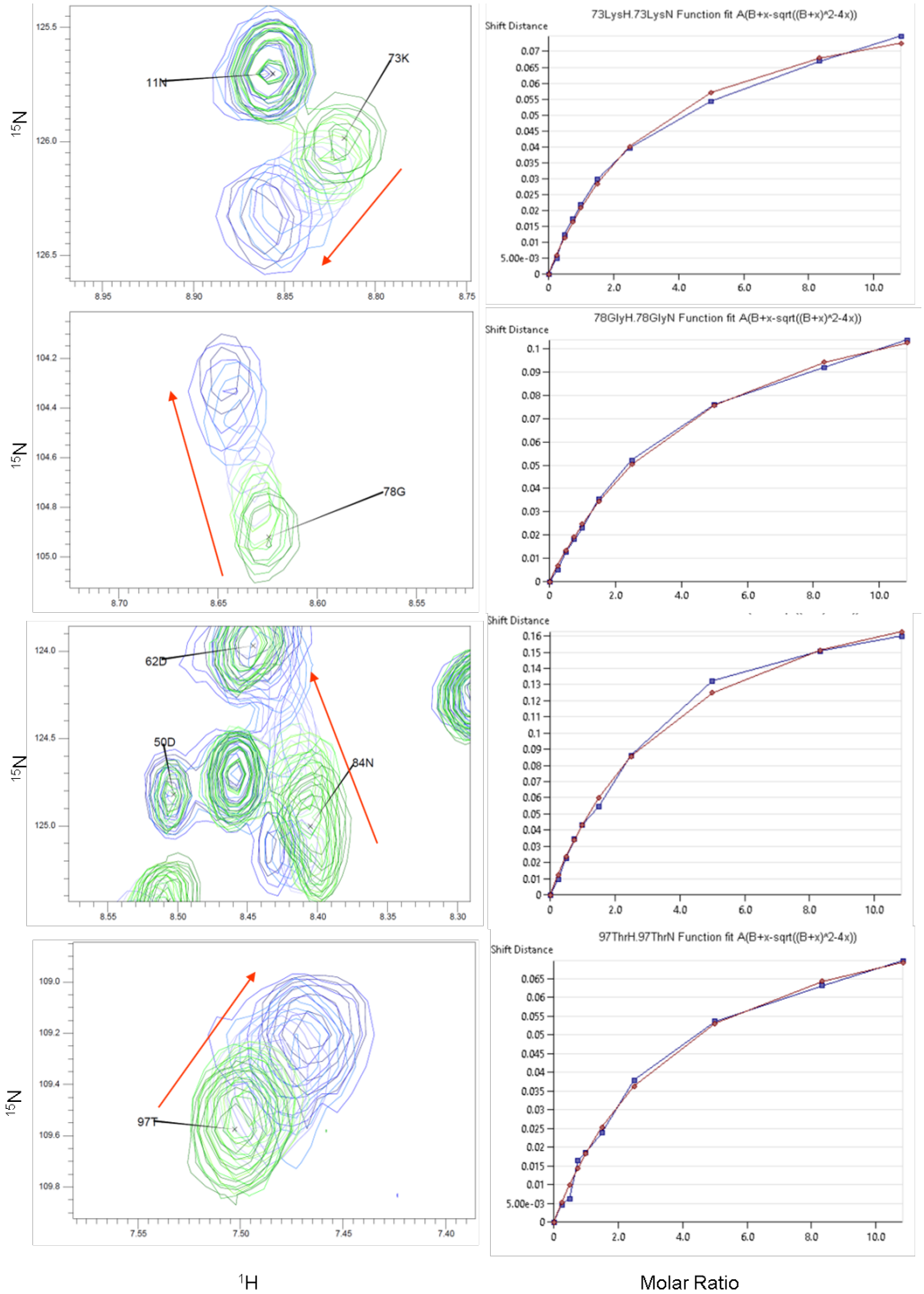


Figure 3.14: The CCPN Analysis software K_d fit, shown in red, for two fast exchange residues from the GB1:WW3:OCT4 titration. The estimated dissociation constants for these residues are: 73K: 0.82 ± 0.07 mM, 78G: 1.23 ± 0.07 mM, 84N: 0.96 ± 0.09 mM and 97T: 0.99 ± 0.11 mM.

Initially, the addition of small concentrations of OCT4 peptide appears to change the resonance position, creating a sharp increase in shift distance, gradually tapering off to a gentle curve as the OCT4 peptide concentration nears saturation. As there is still peak movement between the last two titration points in all four graphs, saturation was not achieved.

The graphs display acceptable fits when compared with the shift distance graphs, and peak migration shows fast exchange is occurring. The direction of peak migration is illustrated with a red arrow for each fast exchange resonance change. The K_d value is estimated using equation 2.4, in section 2.5.8.5, which is appropriate for peaks that are in fast exchange.

Table 3.1: K_d values and errors calculated for WW3 residues experiencing a fast exchange migration when titrated with OCT4 peptide, as determined by CCPN Analysis. Individual K_d errors used the built-in bootstrap method and indicate the fit error. The average K_d was calculated for all of the residues shown and standard deviation and error propagation values were stated.

Residue	**K_d (mM)**	**K_d error (mM)**
67Leu	0.95	0.24
71Trp	1.24	0.12
72Glu	1.51	0.11
73Lys	0.82	0.07
75Gln	1.12	0.12
78Gly	1.23	0.07
79Arg	0.99	0.11
80Val	1.44	0.24
84Asn	0.96	0.09
89Thr	1.01	0.14
96Arg	1.43	0.14
97Thr	0.99	0.11
98Gln	1.01	0.14
Average	1.13	0.039 (propagated)
Standard Deviation	0.21	

Table 3.1 shows the individual K_d values for residues with a trajectory of > 0.02 ppm throughout the GB1-WW3 titration with OCT4 peptide with individual K_d errors indicating the fit errors. The average of the dissociation constants was found to be 1.13 ± 0.21 mM. Standard deviation is given as the error for the average K_d value.

3.3 Discussion

This chapter attempted to gain structural insight into the third WW domain of WWP2 (WW3) and investigate the interaction between WW3 and a peptide of OCT4 using solution-state NMR. Structure calculation of the WW3 domain proved difficult despite the GB1 domain appearing much better defined. NMR titrations with GB1-WW3 and OCT4 peptide indicated that an interaction takes place, which was expected following previous research by Xu *et al.*, showing that WW3 and OCT4 interacted, just at a lower affinity to the WW2 domain [Xu *et al.*, 2009].

The [^{1}H, ^{15}N]-HSQC peaks for GB1-WW3 were well dispersed indicating a well-defined structure. The backbone had been previously fully assigned; however, on closer inspection, one backbone residue, 67L, sandwiched between two proline residues, was incorrectly assigned. The correct resonance position for 67L was discovered and, with additional spectra, side-chain assignments could be revised and assigned as comprehensively as possible. Ultimately, the backbone was fully assigned at 92% due to the presence of proline residues and the absence of the three N-terminal residues. Aliphatic and aromatic side-chain residues were assigned as thoroughly as possible, with 85 residues fully assigned, 13 partially assigned and 2 with no assignments (the two residues at the N-terminus). The majority of the partially assigned residues were located in the WW domain.

The CSI 3.0 web server [Hafsa *et al.*, 2018]; [Hafsa *et al.*, 2015] [Berjanskii and Wishart, 2005] uses backbone chemical shift data to predict protein secondary structures. For GB1-WW3, CSI 3.0 predicted the GB1 structure as

two β strands, an α helix, followed by an additional two β strands, which agrees with PDB structures of this domain. The structure of the GB1 domain calculated here contains these secondary structure regions and holds a similar structure to deposited solution NMR structures. It does however still require further refinement as the structural ensemble of the domain shows discrepancies and cannot be superimposed.

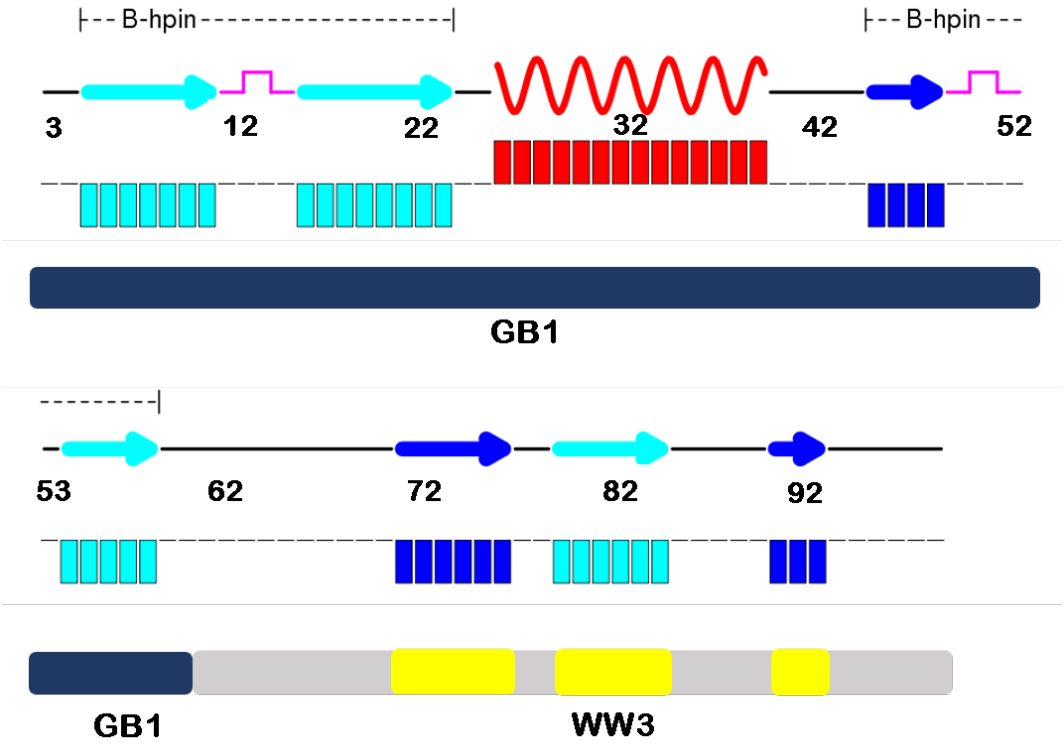


Figure 3.15: The output from the CSI web server, using GB1-WW3 chemical shifts [Hafsa *et al.*, 2015]. Edge and interior β strands are shown in dark and light blue respectively, α helices in red, turns in pink and coil regions in black. The GB1 domain (residue 1-61) is predicted to be made up of an α helix, sandwiched between two β hairpins. The linker region (residue 62-6) is predicted to be random coil and the WW3 domain (67-98) is predicted to be three β strands. Figure adapted from [Hafsa *et al.*, 2018].

The WW domain is predicted to have three β strands, agreeing with several solved WW domain structures from the NEDD4 family. Three β strands is a common conformation for WW domains [Bork and Sudol, 1994]; [Sudol *et al.*, 1995]; [Zarrinpar and Lim, 2000] and an example WW domain from NEDD4-2 protein is displayed in figure 3.16. This solution structure was obtained from the

PDB (1WR3) [Wilton *et al.*, 2008], and forms three parallel β strands.

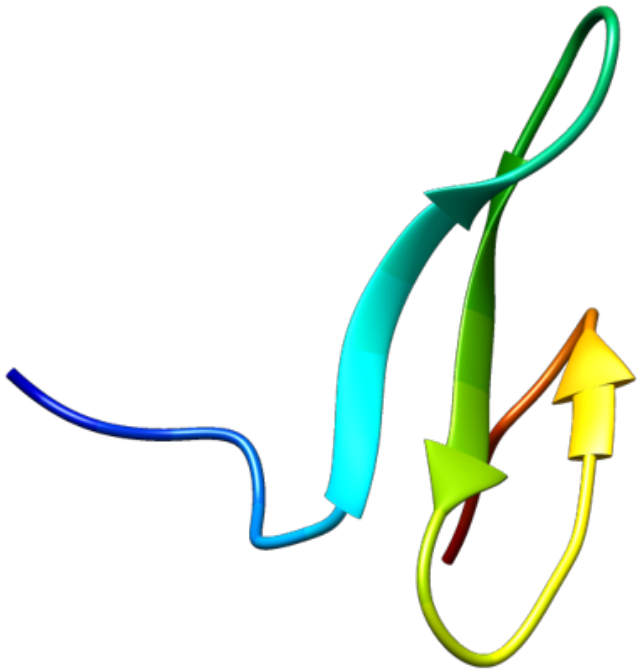


Figure 3.16: The PDB solution structure of the WW1 domain of NEDD4-2 (PDB:1WR3) showing three β strands [Wilton *et al.*, 2008]. Figure exported from Chimera[Pettersen *et al.*, 2004].

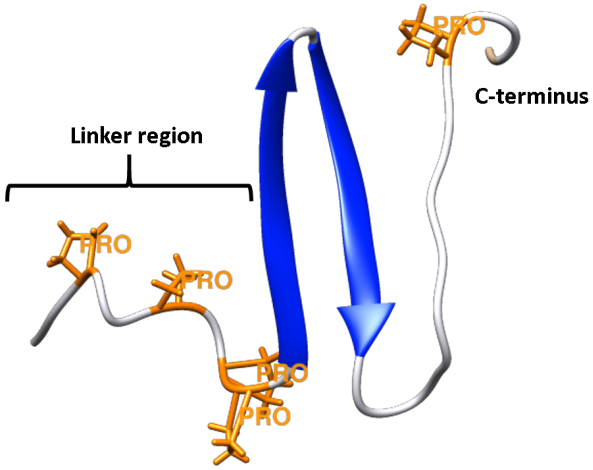


Figure 3.17: The C-terminus of the GB1-WW3 ribbon diagram showing the linker region and the WW3 domain with proline residues highlighted in orange for clarity. The majority of the proline residues are present in the flexible linker region. Figure exported from Chimera [Pettersen *et al.*, 2004].

The structure of WW3 calculated in this chapter was problematic, with models yielding various different structures. The structure has been gradually refined and whilst it has improved it requires more refinement before publication. Whilst the WW domain contained five proline residues, the majority of these were located in the linker region (see figure 3.17) and so would not have affected the number of restraints for the third β sheet.

A possible reason for the problems with the C-terminus could be that the third β sheet is not fully folded. In 2016, Panwalkar *et al.* published a paper describing the interaction between a WW domain of human NEDD4-1 and a target peptide α-hENaC [Panwalkar *et al.*, 2016]. Here it was revealed following chemical shift perturbation titrations, residues in the third β strand (β3) region underwent the most significant chemical shift change in the domain upon binding the peptide. Of the three β strands, β3 was also found to undergo the largest reduction of mobility in the presence of the target peptide [Panwalkar *et al.*, 2016]. When coupled with folding studies which demonstrate that β3 is also the first strand to dissociate [Davis and Dyer, 2013]; [Petrovich *et al.*, 2006], this data could indicate that β3 is a more flexible strand with the ability to sample non-native conformers and subsequently promote peptide recognition [Panwalkar *et al.*, 2016]. Sequence alignment of the WW domain of human NEDD4-1 with the WW3 domain of WWP2 yielded 58.8% identity and 73.5% similarity. This suggests that the third β strand of the WW3 domain may also be more flexible than the previous two β stands. To test this theory, NMR relaxation experiments could be acquired to investigate the dynamic behaviour of the WW3 domain.

Another possible reason for a problematic C-terminus could be due to incorrect side-chain assignments in this region which are forcing some false distance constraints to be used during the structure calculation, or assignments which are missing altogether and hence not providing enough constraints to generate a refined structure.

NMR titrations of GB1-WW3 with OCT4 peptide indicated that an interaction took place based on changes in [^{1}H, ^{15}N]-HSQC resonance values upon addition of the ligand. Several of the affected peaks displayed clear fast exchange processes,

where the resonances slowly migrate as OCT4 was added, whilst other peaks were observed appearing or disappearing as the titration progressed.

Some WW3 peaks started to move immediately after the initial titration point at molar ratio 1:0.25 GB1-WW3:OCT4 and continued to move as the OCT4 concentration was increased. Despite attempting to use a large excess of OCT4, the titration did not reach saturation and it is possible that the peptide concentration was lower than expected. The freeze-dried peptide was added to the titration by mass, meaning that salts or moisture could have been present in the peptide sample which would lead to a lower concentration than expected. As saturation was not reached, it remains unknown whether a more stable structure would have arisen if the protein was fully in the bound state.

The GB1 tag resonances remained stationary throughout the titration confirming that the tag did not interact with the OCT4 peptide. One residue, 24V, did increase in intensity as the concentration of OCT4 was increased; however, the position of the peak was not altered. This suggests that this residue was in slow exchange between two conformations and as the ligand was added, the second conformation became less occupied, causing an increase in intensity at the position of the first conformation.

Through the course of the titration, several peaks disappear and appear. The number of disappearing peaks in this titration was higher than the number of appearing peaks, leading to an overall decrease in the number of peaks by 5. This could be a result of intermediate exchange between bound and unbound states causing peak broadening due to saturation not being achieved.

From the plotted shift trajectories, it was clear that a number of peaks disappeared throughout the titration and hence could not be tracked (figure 3.12 shown in red). The WW domain region displayed much larger trajectories for a number of peaks, up to the value of 0.16 ppm. Peaks which disappeared throughout the titration also represent a structural modification upon ligand binding, indicating that the majority of the WW3 domain was affected by the interaction.

The K_d values estimated using the CPCPN Analysis software were indicative of relatively weak binding, in the mM range. Errors in the calculated K_d values were relatively high which is to be expected as saturation was not reached. The increase in trajectory as OCT4 peptide was added was initially a large change, gradually lessening as more of the WW2 protein was bound.

The average K_d for peaks in fast exchange was 1.13 ± 0.21 mM. When compared to K_d values for WW domains of WWP2 and a small peptide region of OCT4, determined by Jiang *et al.* using ITC, the value for WW3 is significantly different from the result achieved here [Jiang *et al.*, 2015]. This research used a much smaller peptide of OCT4 (T1: RRPCPPPYEFC) inclusive of the PPxY motif and the K_d values are reproduced in table 3.2. Determination of K_d using NMR titration data or isothermal titration calorimetry is expected to yield the same result.

Table 3.2: The dissociation constants and errors for WWP2 WW domains and a small PPxY containing peptide of OCT4 [Jiang *et al.*, 2015].

Protein	**WW2**	**WW3**	**WW4**
Apparent K_d μM	53 ± 8	42 ± 11	29 ± 17

The dissociation constants obtained by Jiang *et al.* [2015] were of the μM range, as opposed to here, in the mM range. These vast differences may be due to the smaller peptide construct being more accessible to binding, as well as the different conditions of the experiments. ITC experiments by Jiang *et al.* used protein and peptide in 25 mM Tris–HCl, and 150 mM NaCl at pH 8.0 [Jiang *et al.*, 2015] and unfortunately did not state which WW domain boundaries were used. The conditions used in NMR titrations for this thesis were performed in Phosphate buffer (20 mM Na_2HPO_4, 50 mM NaCl, 15 mM DTT, pH 6.8). As the sample conditions varied quite substantially, data across the two sets of conditions were less comparable. If the conditions were kept the same across the two separate studies, it is expected that the domain binding affinities in a single study would be relatively proportionate to the other.

Whilst an interaction between WW3 and OCT4 peptide was confirmed using the NMR titration results, the affinity of this interaction was not calculated with

great confidence due to saturation of the WW3 domain not being reached. In the future, ITC experiments between GB1-WW3 and the same OCT4 construct could be acquired and compared with the data for the shorter OCT4 construct. A double labelled NMR sample of GB1-WW3 in the presence of excess OCT4 could also allow assignment of the peaks which disappeared and reappeared in a slow exchange process.

Despite much time and effort being invested in to investigating problems and gradually refining the GB1-WW3 structure, further refinement is required for a structure of publication quality. The GB1 domain structure is further along in the refinement process; however, a number of issues remain for the WW3 domain. These issues are likely caused by incorrect or missing assignments, particularly in the region of the third β strand. Further work on WW3, including ensuring correct assignments and NOE refinement, could allow successful calculation of the structure.

Chapter 4

Analysis of WW2 and tandem WW2-3 domains

4.1 Introduction

Proteins within the NEDD4 family of HECT E3 ligases are known to encompass between two and four WW domains. These WW domains are involved in the recognition of different substrates [Ingham *et al.*, 2005] and their subsequent binding and positioning. Research has also shown that they have a role to play in autoinhibition of the HECT domain through blocking E2-E3 ligase trans-thiolation [Riling *et al.*, 2015].

An important question to answer is how the respective structures of the WWP2 WW domains, each with specific affinities for individual substrate targets, contribute to their activity. It is therefore advantageous to determine the structure of each domain individually, leading to an understanding of the structural and functional characteristics, as well as to determine if two-domain constructs offer an increased affinity for binding. The second WW domain (WW2) and the tandem second to third domain (WW2-3) were initially selected as a focus for this study, due to lack of structural information for these domains

and research suggesting their involvement in binding specific targets with implications in cancer.

4.1.1 WW2 domain

The second WW domain is of interest due to its specificity for certain binding partners and their involvement in pathways related to cancer. An example of the specificity of WW domains of human ubiquitin ligase WWP2 was described in a paper in 2009, where GST pull-down experiments described the second WW domain (WW2) of WWP2 as having highly enhanced interactions with substrate partner OCT4 when compared to the other WW domains (WW1, WW3 and WW4) [Xu *et al.*, 2009]. NEDD4-like E3 ligases, NEDD4L and SMURF1, have WW2 domains which were also discovered to bind the PPxY motif in Smad7 [Aragón *et al.*, 2012]. The NEDD4L WW2 domain and SMURF1 WW2 domain have 56.76% and 51.35% sequence identity respectively in relation to the WWP2 WW2 domain. The indication that WW2 is the central binding domain for the interaction between WWP2 and OCT4, as well as its potential involvement in the binding of Smad7, marks WW2 of interest for further investigation.

4.1.2 WW2-3 tandem domain

WW domains have also been proven to recognise certain target proteins as a pair of WW domains [Sudol *et al.*, 2005] and several tandem domains from recombinant fragments of NEDD4L, SMURF1 and SMURF2, were observed to interact with a peptide of Smad7 [Aragón *et al.*, 2012]. The WW2-3 linker region of WWP2 was also found to have a role in inhibiting E3 ligase activity [Chen *et al.*, 2017]. Therefore, analysis of the WWP2 tandem domain containing the WW2, WW3 domains and linker region is investigated here to elucidate if the binding affinity of the interaction is affected by having more than one WW domain available for interaction.

4.1.3 Effect of temperature on protein conformational exchange

Proteins switching between conformations on a microsecond to millisecond time-scale is known as conformational exchange and causes changes in the local chemical environment of the nuclei. This change can have a considerable effect on the appearance of its NMR spectra and is particularly problematic in NMR when proteins are in an intermediate exchange regime, which can cause significant peak broadening to the point where the peaks are not visible in the spectrum.

The rate of change between different conformations is impacted largely by the temperature. This temperature dependence can be described by the Arrhenius equation (equation 4.1), where k is the rate constant (s^{-1}), T is the temperature (Kelvin), A is a chemical reaction constant (s^{-1}), E_a is the activation energy (kJ mol^{-1}) and R is the universal gas constant (kJ $mol^{-1}K^{-1}$).

$$k = Ae^{\frac{-E_a}{RT}} \tag{4.1}$$

When the temperature of the sample is lowered, this causes the protein to experience a decrease in kinetic energy, pushing the exchange process into a relatively slow exchange. This could allow residue peaks which are experiencing peak broadening as a result of intermediate conformational exchange to become observable. An alternative option is to raise the temperature during NMR experiments in an attempt to move the exchange into a fast exchange process, allowing the weighted average chemical shift of all of the conformations present to be observed.

4.1.4 Experimental aims

The aim of this chapter is to describe the expression and purification of the second WW domain (WW2) and tandem second and third domain (WW2-3), as fusion proteins with His-GB1 tags. This was optimised to allow the fusion proteins to be overexpressed, sufficiently purified and concentrated for analysis by NMR spectroscopy. A [^{1}H, ^{15}N]-HSQC spectrum was acquired for His-GB1-WW2-3, to assess resolution and peak dispersion, with the aim of determining if the domain was folded. As discussed in section 1.5, solution-state NMR spectroscopy is the only technique which allows the three-dimensional structure of a protein to be determined whilst in the solution phase. The second WW domain was investigated by NMR spectroscopy, with the aim of assigning the backbone and providing structural information for the isolated domain.

4.2 Results

4.2.1 Preparation of His-GB1-WW2 construct

WW2 was subcloned into plasmid pSKDuet01 using linearized pSKDuet01, WW1-2 Reverse Primer 5'-3', WW2-3 forward Primer 5'-3' and pRK5-HA-WWP2-FL. A 100 μL high fidelity (HF) Phusion PCR reaction was performed and the correct sized band was gel extracted after running the resultant sample on an agarose gel (figure 4.1). Here, the band at $\sim$ 120 bp correlates to WW2 insert and was extracted using a QIAquick Gel Extraction Kit and following manufacturers protocol.

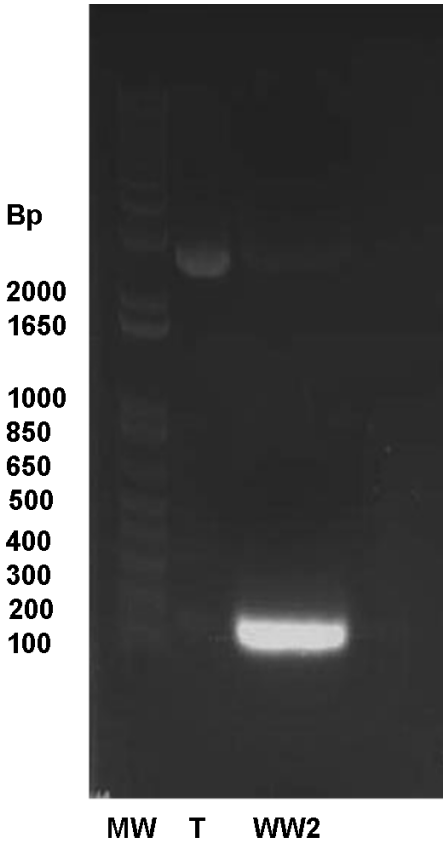


Figure 4.1: An agarose gel showing the result of PCR Phusion reaction of WW2 with control T, template pRK5-HA-WWP2-FL.

An In-Fusion® HD Cloning Kit was used to clone His-GB1-WW2 fragment into gel extracted linearised vector pSKDuet01, provided by Jessica Watt of the Chantry laboratory at UEA. Stellar competent cells were then used for transformation. 8 x 20 μL PCR reactions were completed for PCR colony screening. Figure 4.2 shows that all of the colonies contained a band at the correct size.

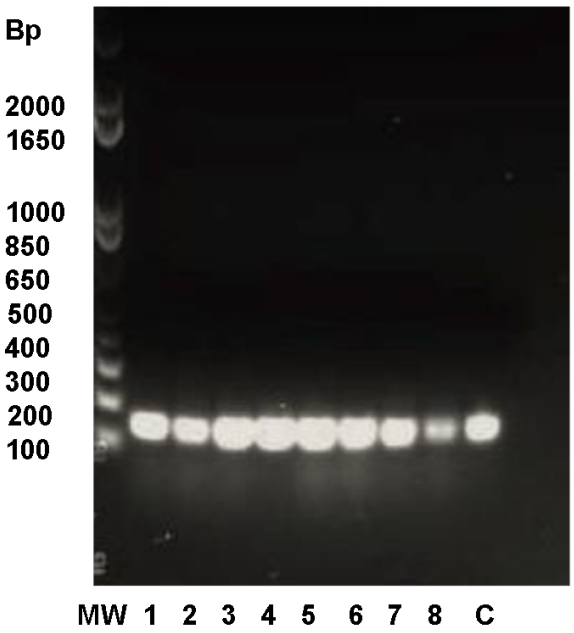


Figure 4.2: An agarose gel of PCR screening for eight GB1-WW2 colony samples (lanes 1-8) and control (C).

The negative control sample also displayed a band at the size of His-GB1-WW2 which indicates that a contaminant may have been present in another component of the PCR reaction. DNA extraction and purification of sample 2, 3, 4 and 5 were performed and sent for sequencing (Eurofins genomics). Samples 3, 4 and 5 were found to have the correct sequence and sample 2 was discarded as a mutation was identified. DNA extraction and purification on a midi scale were then performed on sample 3 (2.9 μg/μL).

His-GB1-WW2 was transformed into different cell lines for small scale test expression using competent *E. coli* cell lines, BL21 (DE3), BL21-CodonPlus(DE3)-RP, Rosetta 2(DE3)pLysS and BL21 (DE3)pLysS. Expression conditions were 37 °C, 200 rpm until OD_{600} reached 0.8-1.2 at which point 0.8 mM IPTG was added and cultures were left 3 hours at 37°C for induction. Small aliquots from each sample were taken before and after induction, centrifuged, and the insoluble pellet was resuspended in SDS loading

buffer and analysed using a SDS-PAGE analysis. The 12% SDS-PAGE gel is shown in figure 4.3.

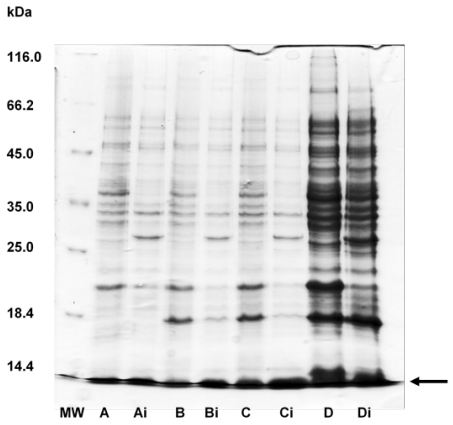


Figure 4.3: 12% SDS-PAGE gel showing GB1-WW2 domain competent cell test expression pre and post induction with IPTG. The relevant protein band is highlighted with an arrow. **A**: BL21(DE3), **Ai**: BL21(DE3)induced, **B**: Rosetta2(DE3)pLysS **Bi**: Rosetta2(DE3)pLysS induced, **C**: BL21(DE3)pLysS, **Ci**: BL21(DE3)pLysS induced, **D**: CodonPlus(DE3)-RP, **Di**: CodonPlus(DE3)-RP induced.

Despite the poor separation between the protein of interest and the front of the gel due to the low acrylamide percentage, it was possible to see that all strains expressed the GB1-WW2 construct. Post induction samples were resuspended in a larger volume than the pre-induction samples due to a larger cell pellet size after centrifugation. In this case, the post induction samples appear less concentrated than the pre-samples, which was taken into account when analysing the gel. BL21(DE3) (shown as A and Ai in figure 4.3) was selected for expression of His-GB1-WW2 as it produced similar expression levels with reduced levels of contaminants to Rosetta2(DE3) and BL21(DE3)pLysS. It is also simpler to use as it does not require choloramphenicol.

4.2.2 ^{15}N labelled His-GB1-WW2 expression and purification

His-GB1-WW2 was overexpressed using previous conditions, in 1 L ^{15}N labelled Minimal Essential Medium (MEM) and the protein was then purified using Ni-NTA Immobilized Metal Affinity Chromatography (IMAC). Figure 4.4 shows an 18% SDS gel with samples taken at each stage of the purification, with a clear His-GB1-WW2 protein band in the elution lanes.

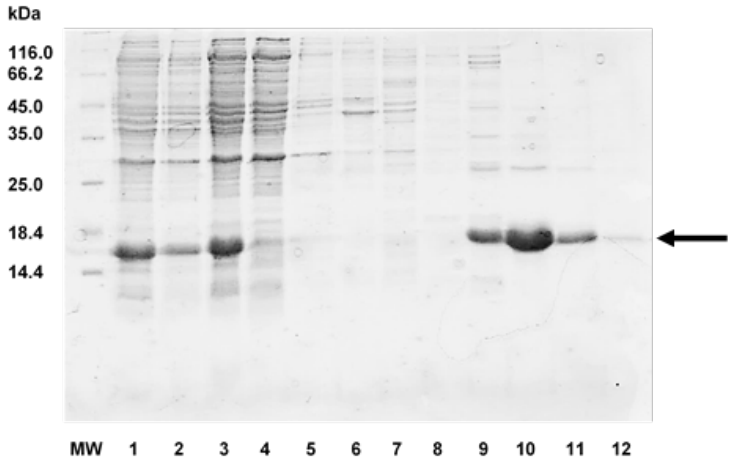


Figure 4.4: SDS-PAGE analysis showing purification of ^{15}N labelled His-GB1-WW2 domain using a Ni-NTA column. Purified protein can be seen in the elution bands in lanes 9, 10 and 11. Imidazole wash concentrations: wash A = 1 mM, wash B = 10 mM, wash C = 20 mM, wash D = 30 mM, elution E = 250 mM. 10 μL samples loaded. **lane 1**: Pre-sonication, **lane 2**: Lysate, **lane 3**: Supernatant, **lane 4**: Column flow-through, **lane 5**: wash A, **lane 6**: wash B, **lane 7**: wash C, **lane 8**: wash D, **lane 9**: elution 1, **lane 10**: elution 2, **lane 11**: elution 3, **lane 12**: elution 4. 10 μL samples loaded.

The E1, E2 and E3 elution samples had low-level contaminants which would be negligible in NMR experiments and so they were combined and dialysed into NMR buffer. The absorbance of the combined sample at 280 nm was measured and the concentration of the sample was calculated. The sample was then concentrated to 500 μL which yielded a concentration of 0.55 mM and an NMR sample was prepared. The [^{1}H, ^{15}N]-HSQC spectrum shown in figure 4.5 was acquired using a Bruker Avance I 500 MHz spectrometer. This spectrum provided information on

the conformational state of the protein and whether or not it was folded, therefore determining if it was a suitable sample for NMR spectroscopy [Cavanagh *et al.*, 2010]; [Rehm *et al.*, 2002]. The spectrum showed well-dispersed peaks of good intensity and resolution, which initially indicated that the protein is likely well folded. For backbone assignment, triple resonance experiments were required and so ^{15}N, ^{13}C labelled samples were prepared.

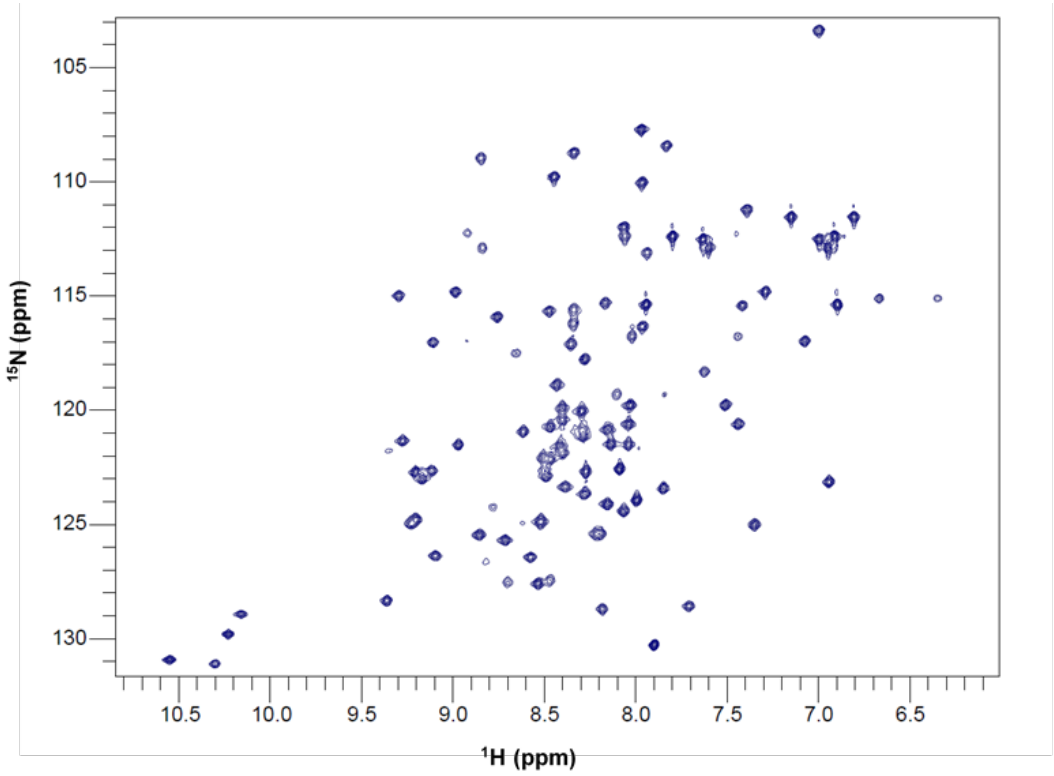


Figure 4.5: [^{1}H, ^{15}N]-HSQC of His-GB1-WW2 at a concentration of 0.55 mM, acquired at 298 K, 500 MHz. Peaks are well dispersed and resolved, indicating folded protein.

4.2.3 ^{15}N, ^{13}C labelled His-GB1-WW2 expression, purification and assignment

[^{1}H, ^{15}N]-HSQC and [^{1}H, ^{13}C]-HSQC, HNCACB, CBCACONH, HNCO and HNCACO experiments were acquired using a Bruker Avance III spectrometer at 800 MHz. These experiments were used to manually assign the backbone of His-

GB1-WW2, as shown in figure 4.6.

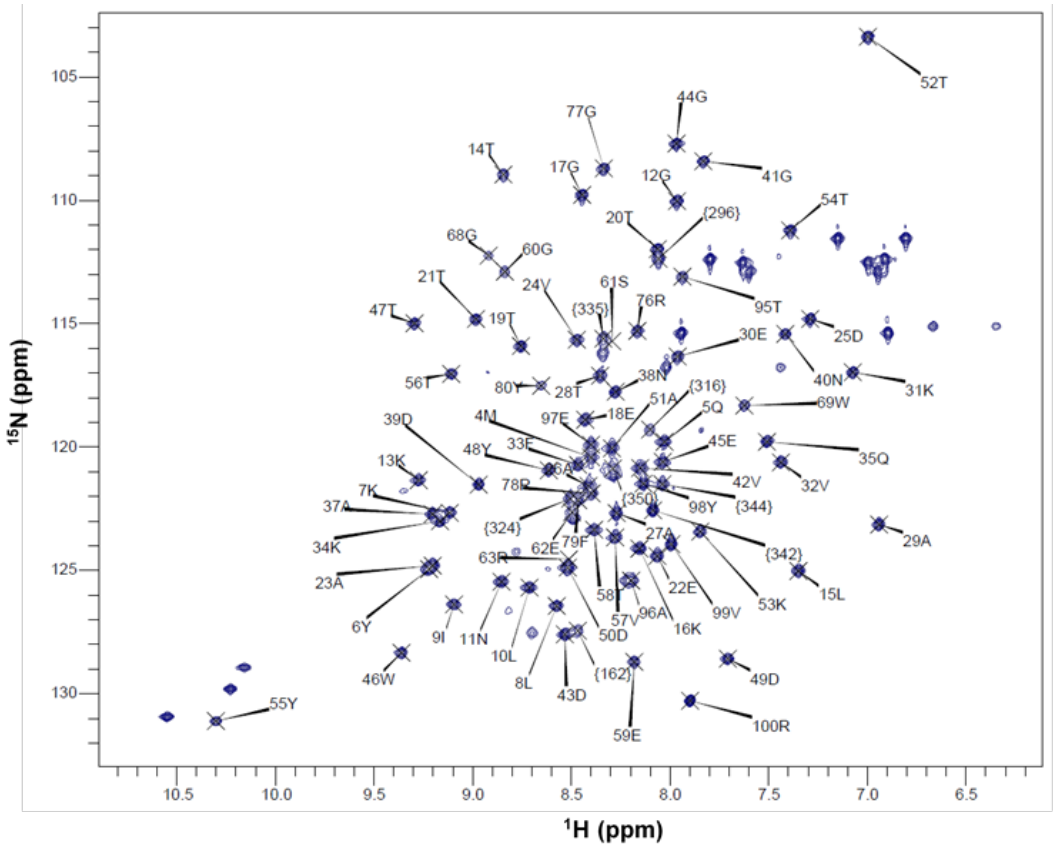


Figure 4.6: The assigned [^{1}H, ^{15}N]-HSQC, acquired at 800 MHz, for His-GB1-WW2 at a concentration of 0.55 mM, 298 K.

75% of His-GB1-WW2 (75 of 100 residues) was assigned, using the method described in section 1.6.1.4. Of the unassigned residues, five proline residues, all from the WW2 domain region (Pro64, Pro66, Pro67, Pro75 and Pro94), were not expected to be observed in the [^{1}H, ^{15}N]-HSQC due to their lack of amide proton.

Glutamine and asparagine residues contain an amide group in their side chains and are observed as two peaks with matching ^{15}N ppm values, found in the upper right-hand side of the [^{1}H, ^{15}N]-HSQC. Tryptophan residues have a single side chain peak located in the bottom left of a [^{1}H, ^{15}N]-HSQC spectrum. Whilst the three tryptophan residues are not assigned in this spectrum, the far left bottom

peak is 46W as it can be propagated from other GB1-tagged assignments. The remaining two tryptophan peaks will be 69W and 91W.

The number of peaks in the [^{1}H, ^{15}N]-HSQC should be similar to the number of residues present in the protein, taking into account additional side chain peaks and discounting proline residues. The number of peaks in this spectrum was not as high as expected, hence full assignment was not possible. A potential explanation for this is intermediate conformational exchange, which can lead to peak broadening to the extent that affected peaks are not visible (section 1.6.1.5).

4.2.4 His-GB1-WW2 thrombin digestion

His-GB1-WW2 was expressed using the same conditions as used previously and the protein was then purified using Ni-NTA IMAC. Two units of thrombin were added to the elution aliquots for every mg of protein present. The samples were dialysed into Tris-HCl buffer at room temperature until the digest was complete. Figure 4.7 shows Ni-NTA purification of the digested GB1-WW2 sample following removal of thrombin using a benzamidine column.

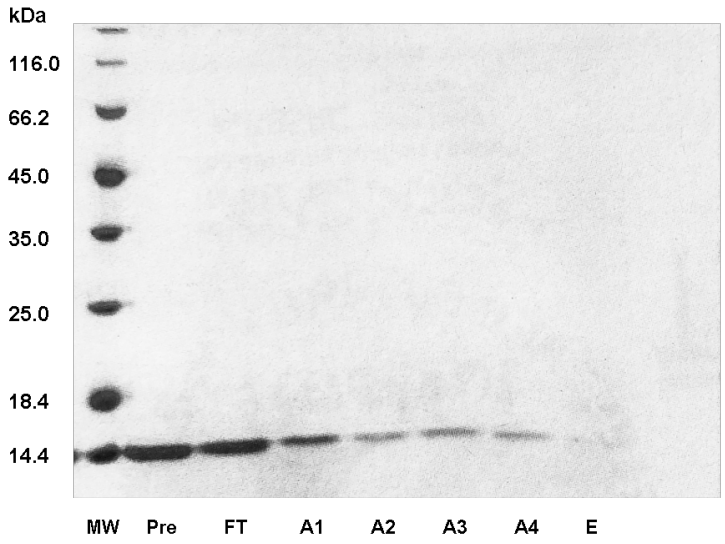


Figure 4.7: Post-digest GB1-WW2 purification using Ni-NTA column. Lanes show the pre-purification sample, the flow through, washes A1-A4 and an elution. 10 μL samples loaded.

4.2.5 ^{15}N, ^{13}C labelled GB1:WW2 backbone assignment

A ^{15}N, ^{13}C His-GB1-WW2 sample was expressed and purified as previously. The histidine tag was cleaved using thrombin, the protein was repurified and an NMR sample was prepared. A [^{1}H, ^{15}N]-HSQC of GB1:WW2 was acquired and compared with the undigested His-GB1-WW2 [^{1}H, ^{15}N]-HSQC. Figure 4.8 shows the overlaid [^{1}H, ^{15}N]-HSQC spectra, with the fusion tagged protein His-GB1-WW2 in **Blue**, and the digested GB1:WW2 shown in **Red**.

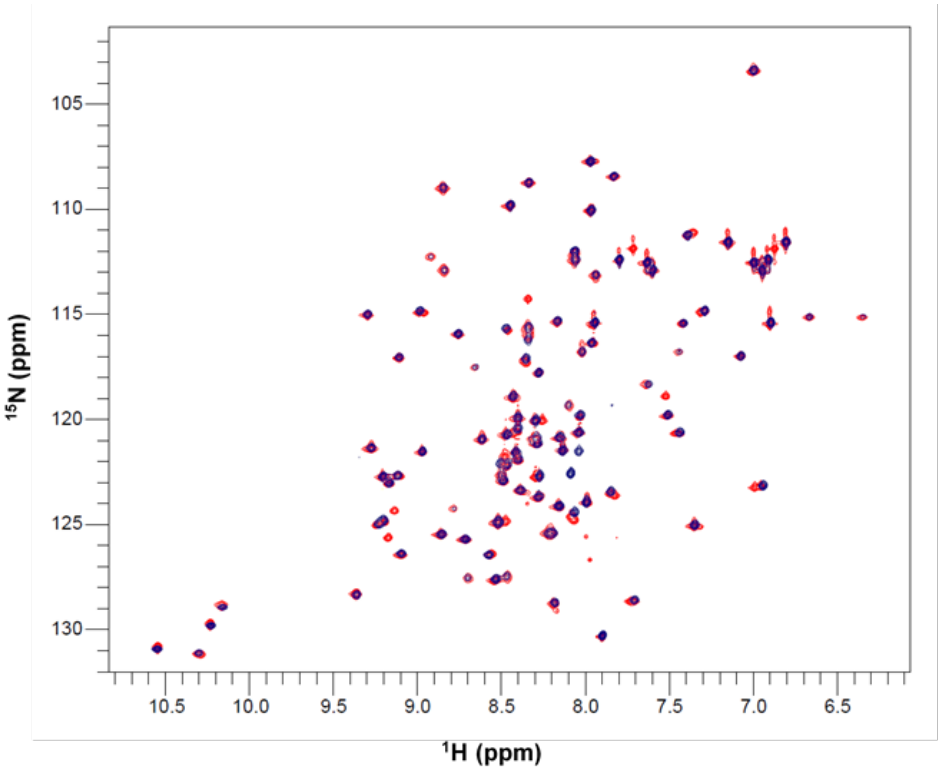


Figure 4.8: [^{1}H, ^{15}N]-HSQC showing His-GB1-WW2 (blue) and digested GB1-WW2 (red). Disappearing peaks are likely to be due to the His-tag cleavage. Peaks which appear in the post-digest GB1-WW2 HSQC are interesting and could indicate that conformational exchange was affected by the polyhistidine tag. This experiment was performed using a Bruker Avance III 800 MHz spectrometer at 298 K. Sample buffer: 20 mM sodium phosphate buffer, 50 mM NaCl, pH 6.8.

Additional peaks are observed post-digestion, despite the loss of peaks from the residues present in the tag. In the digested sample, two peaks, at around

8.1 ppm (^{1}H) and 122.0 ppm (^{15}N), were no longer present in the spectrum, whilst an additional ten peaks were observed appearing throughout the spectrum (two of which are in the side chain region). This increase in peak number, despite a reduction in the number of observable residues, indicates that the His-tag was indeed interfering with the conformation of the protein. Therefore, future NMR experiments were obtained using a digested sample as this reduced the effects of intermediate exchange.

[^{1}H, ^{15}N]-HSQC, HNCACB and CBCACONH experiments were acquired using a Bruker Avance III spectrometer at 800 MHz. These experiments were used to manually assign the backbone of GB1-WW2, as shown in figure 4.9. A full list of the assigned residues and resonances is included in Appendix B.

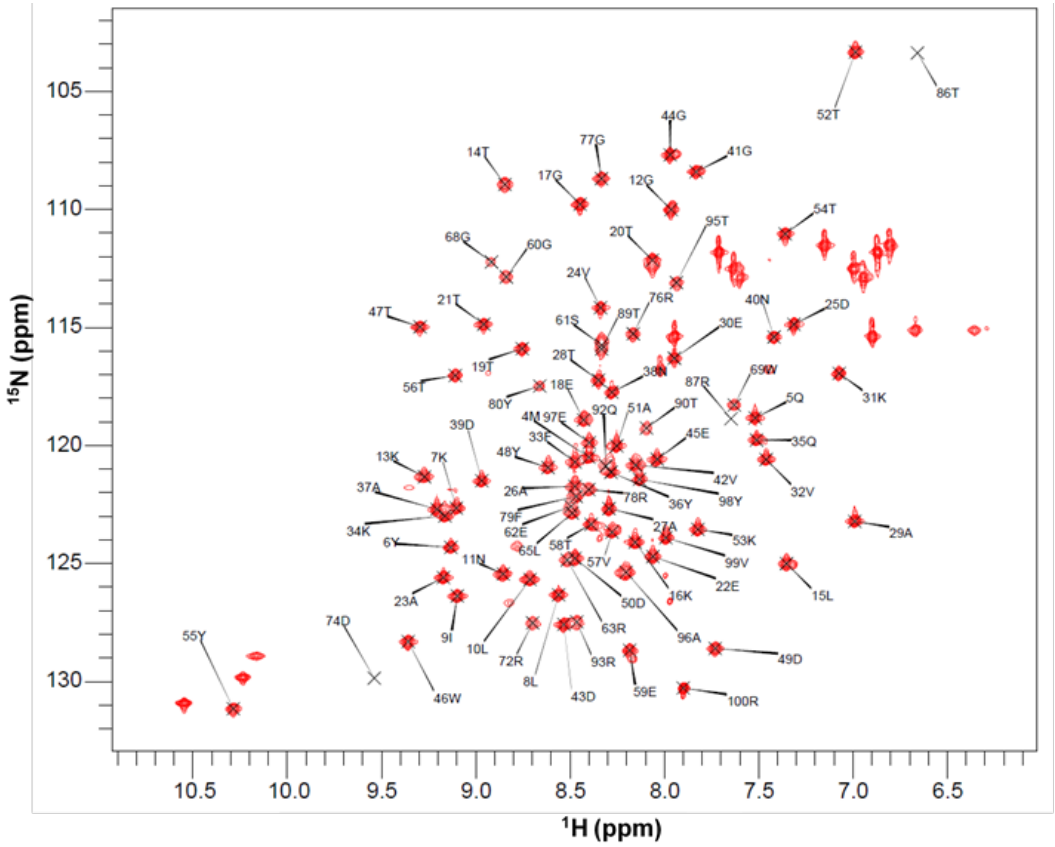


Figure 4.9: [^{1}H, ^{15}N]-HSQC showing GB1-WW2 assignments. 82% of the residues were successfully assigned. This experiment was performed using a Bruker Avance III 800 MHz spectrometer at 298 K. Sample buffer: 20 mM sodium phosphate buffer, 50 mM NaCl, pH 6.8.

82% of GB1:WW2 (82 of 100 residues) could be assigned, using the method described in section 1.6.1.4. Of the unassigned residues, one was from the GB1 tag region and fourteen from the WW2 domain, including the five WW domain proline residues, (Pro64, Pro66, Pro67, Pro75 and Pro94). 86T and 74D peaks are not visible in figure 4.9 as they were below the contour level which was optimised for the majority of the peaks. This indicates that these two peaks are experiencing intermediate exchange and hence the peaks are broadened which is why they cannot be viewed at the same contour levels as the rest of the peaks.

4.2.6 GB1:WW2 [^{1}H, ^{15}N]-HSQC temperature series

The ^{1}H chemical shift is a sensitive parameter which is related to the conformation of a protein. Proteins often exchange between various available conformations, and when in fast exchange the observed chemical shift is the weighted average chemical shift of all of the conformations present. When the temperature decreases, the different population levels of each conformation can be modified, hence the observed chemical shift values may also change. In addition, at lower temperatures, the effect of intermediate exchange could be lessened, as described in section 4.1.3, which could allow broadened peaks to then be observable by solution NMR.

Taking this into account, a series of [^{1}H, ^{15}N]-HSQC experiments were acquired at temperatures 278-298 K, decreasing in 5 K increments. Figure 4.10 shows the superimposition of [^{1}H, ^{15}N]-HSQCs at five different temperatures.

All peaks shifted position throughout the series, and some peak intensities were also affected. Contour levels were optimised for best viewing of the majority of peaks. Side chain residues were observed to have moved more dramatically. The number of peaks present did not increase as the temperature was lowered, indicating that the missing peaks remained in intermediate exchange.

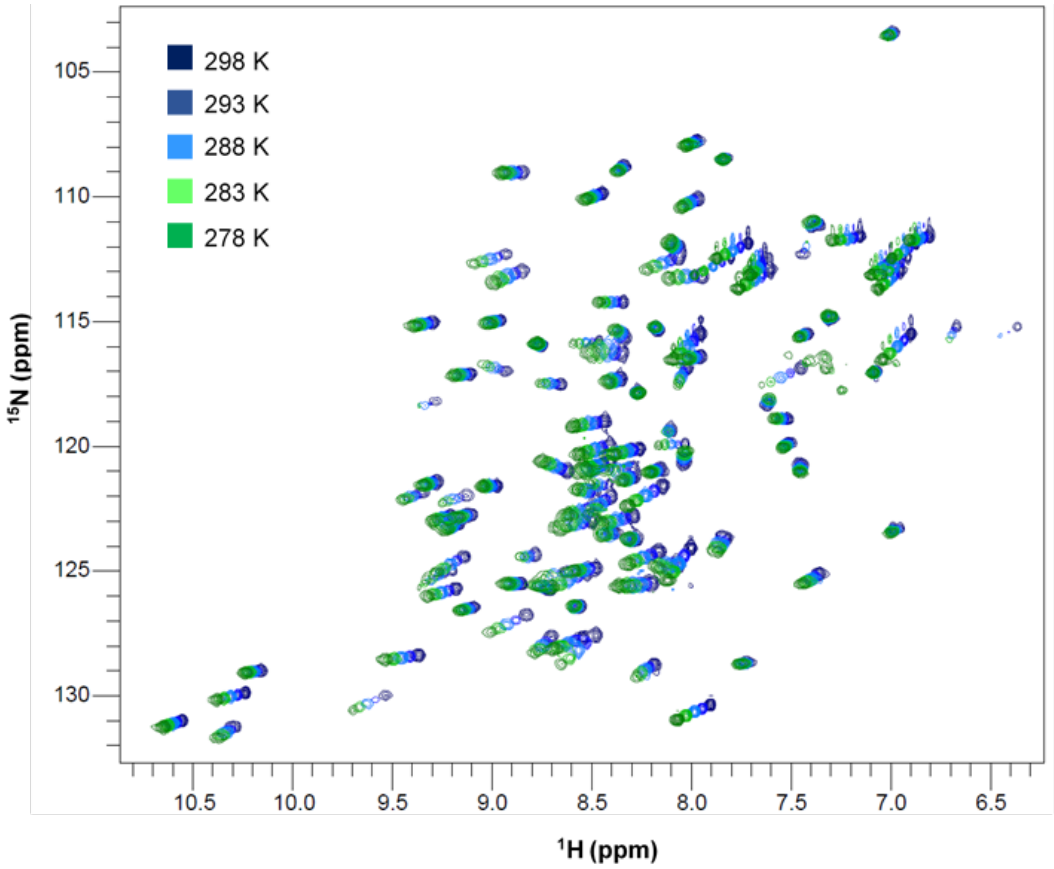


Figure 4.10: Superimposed [^{1}H, ^{15}N]-HSQCs of GB1-WW2 throughout a temperature series. Spectra were acquired at 298 K, 293 K, 288 K, 283 K and 278 K. Spectra were acquired using a Bruker Avance I 500 MHz spectrometer.

4.2.7 WW domain disorder prediction

When in the absence of a binding partner or in certain physiological conditions, it is possible that protein regions, or domains, can lack a stable structure. These proteins or domains are described as disordered, as they repeatedly interchange between multiple conformations, a common feature of proteins with a regulatory role. Due to the proposed intermediate exchange, the GB1-WW2 construct and also the WW1-4 domain region were investigated using a disorder predictor. There are many different disorder predictors available, each making use of a variety of methodologies.

Here, a meta-predictor approach was used, a technique which incorporates predictions from numerous prediction software [Li *et al.*, 2015]. MetaDisorder is a method which compiles 13 disorder predictors which perform well in CASP (Critical Assessment of Structure Prediction) experiments, weighted by method accuracy. This method was selected because it obtained the greatest accuracy value in a review of predictors for intrinsically disordered proteins [Li *et al.*, 2015].

Protein sequences of the GB1-WW2 and WW1-4 constructs were submitted to the GeneSilico MetaDisorder web service [Kozlowski and Bujnicki, 2012] (accessible: http://iimcb.genesilico.pl/metadisorder/metadisorder.html) which predicts disorder and the results are displayed in figure 4.11.

At the top of figure 4.11, the disorder prediction of the first to fourth WW domains (WW1-4) shows a clear dip into the ordered region for WW1, WW3 and WW4 domains, as opposed to WW2, which remains fairly stable on the border between predicted ordered and disordered regions. The linker region between WW2 and WW3 was predicted to have areas with a tendency towards high levels of disorder. The bottom of figure 4.11 shows the GB1-WW2 construct disorder prediction. The GB1 tag and the middle portion of the WW2 domain are predicted to be ordered, as opposed to a relatively higher disorder tendency for the region where the GB1 domain meets the WW2 domain.

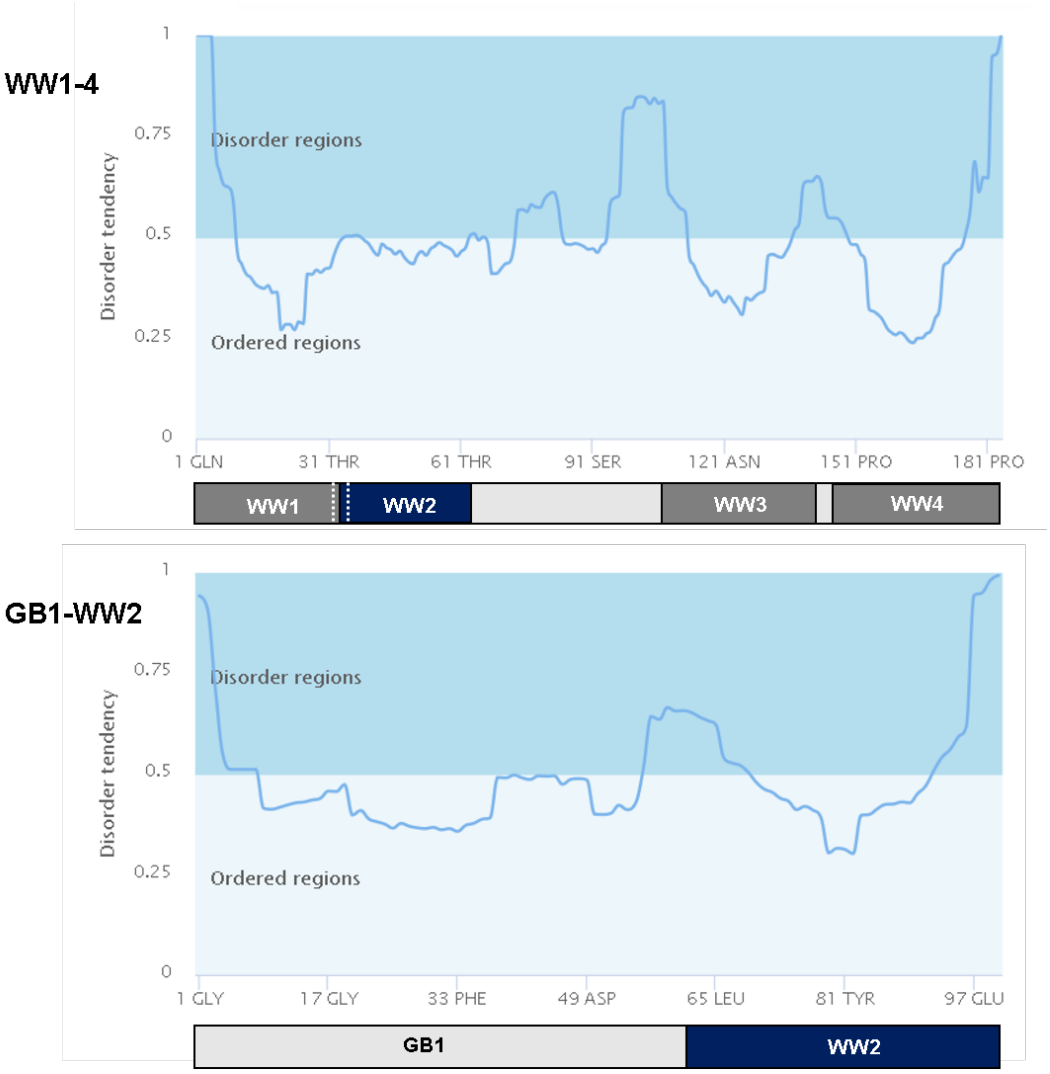


Figure 4.11: Disorder prediction plot showing WW domains 1-4 (top) and GB1-WW2 domain (bottom). Top: Domain boundaries are WW1 = residue 1-33; WW2 = residue 30-63; WW3 = 105-137; WW4 = 144-177. Bottom: Domain boundaries are GB1 = 1-61, WW2 = 62-100. WW1, WW3 and WW4 regions showed clear ordered predictions, as opposed to the WW2 region which indicated more of a borderline ordered/disordered region.

4.2.8 His-GB1-WW2-3 expression and purification

The His-GB1-WW2-3 plasmid, provided by Lloyd Wahl of the Chantry laboratory at UEA, was sent for sequencing and was found to contain two mutations. This led to the subcloning of WW2-3 into pSKDuet01 from the

pRK5-HA-FL-WWP2 Template DNA. A high fidelity (HF) Phusion PCR reaction was performed and the correct sized band was gel extracted. In this gel (figure 4.12A) two bands were present. The top band (∼350 bp) was of the correct size and so was gel extracted. The lower band, which was consistent with the weight of WW2, was ignored.

DNA was cloned using the In-Fusion cloning technique, followed by the transformation of His-GB1-WW2-3 into Stellar competent cells. Four colonies were selected, the plasmid was purified using Nucleospin Plasmid DNA purification kit and each was sent for sequencing. A colony with the correct sequence was selected, the plasmid re-purified (stock of 25.4 μg/μL) and the sequence confirmed.

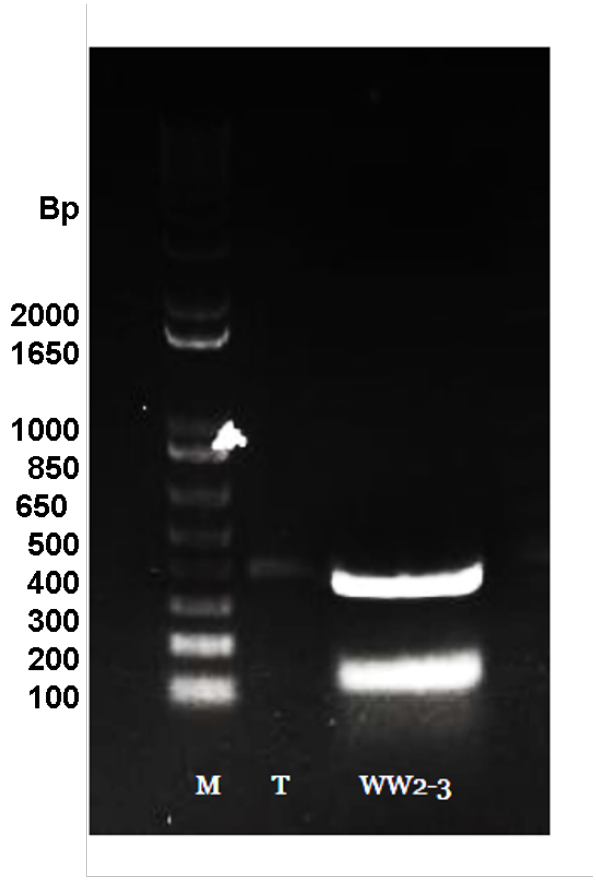


Figure 4.12: An agarose gel showing the result of HF Phusion reaction of WW2-3 where T is template pRK5-HA-WWP2-FL.

Test expressions in *E. coli* cell lines **(B)** BL21 (DE3), **(R)** Rosetta2(DE3)pLysS, **(P)** BL21(DE3)pLysS and **(C)** BL21-CodonPlus(DE3)-RP were carried out for the His-GB1-WW2-3 construct. Cultures were inoculated with a single colony from each cell line with appropriate antibiotics and induced at 37 °C for three hours with 0.8 mM IPTG. The cells were centrifuged and re-suspended in SDS loading buffer and run on a 12% SDS-PAGE gel (figure 4.13).

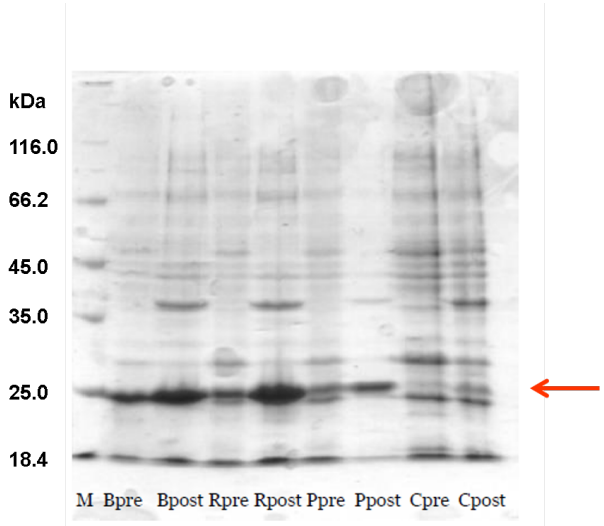


Figure 4.13: A 12% SDS-PAGE gel of WW2-3 test expression in **(B)** (BL21 (DE3)), **(R)** (Rosetta2(DE3)pLysS), **(P)** (BL21(DE3)pLysS) and **(C)** (BL21-CodonPlus(DE3)-RP) pre and post induction with IPTG. His-GB1-WW2-3 is marked with a red arrow.

BL21(DE3) was chosen for expression as it had similar expression levels to Rosetta2(DE3)pLysS with similar contaminant levels but did not require chloramphenicol. The post induction sample of BL21(DE3)pLysS did not load well due to viscosity and so the apparent lower level of contaminants in this sample was likely due to a lower concentration.

Another unlabelled His-GB1-WW2-3 sample was expressed using 0.8 mM IPTG (figure 4.14A) and overnight incubation at 30 °C to reduce the proteolytic activity suspected of causing degradation in the test expressions. The protein was purified at 4 °C using a Ni-NTA column with protease inhibitors to minimise degradation (figure 4.14B).

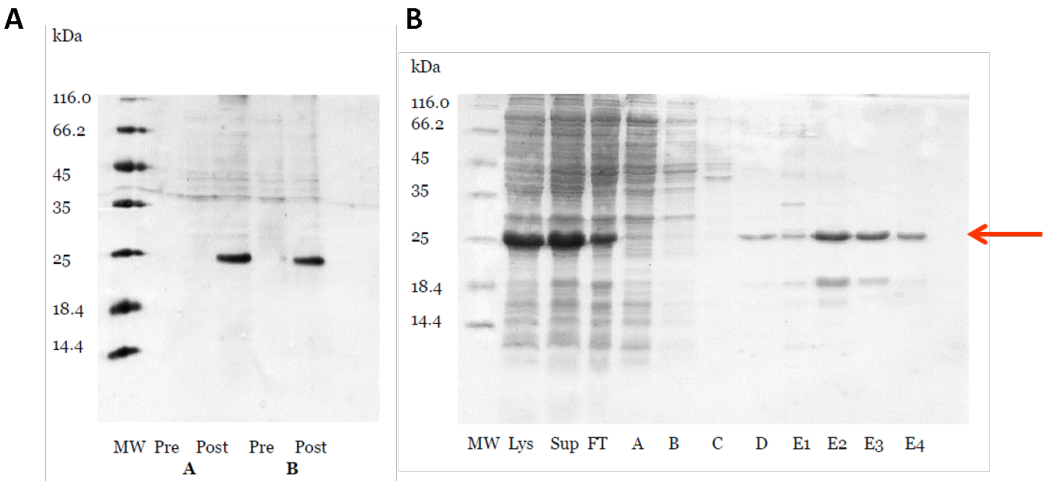


Figure 4.14: **(A)** 15% SDS-PAGE gel showing pre and post induction samples of unlabelled His-GB1-WW2-3. 10 μL samples loaded. **(B)** His-GB1-WW2-3 purification using a Ni-NTA column showing His-GB1-WW2-3 (~27 kDa) elution bands (marked by a red arrow) in E1, E2 and E3 with a lower degradation band at around ~18 kDa.

Degradation bands were not present in the expression gel but after a freeze-thaw cycle and purification, a lower band at ~18 kDa was visible in the elution. This indicated that the degraded protein contained the His tag and so could not be removed by a Ni-NTA column. Therefore, E2, E3 and E4 samples were combined and gel filtration was used to attempt removal of the lower band. Two clear peaks could be seen in the absorbance spectrum and alternate samples from each peak were run on a 15% SDS-PAGE gel (figure 4.15).

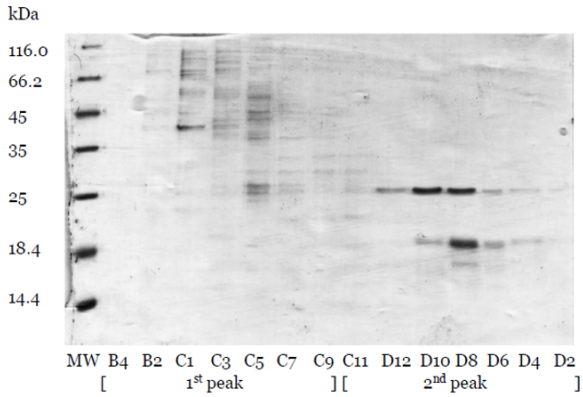


Figure 4.15: 15% SDS-PAGE gel of His-GB1-WW2-3 gel filtration. The lower contaminant band was still present in the same fractions as the His-GB1-WW2-3.

The lower contaminant band was present in the same fractions as GB1-WW2-3, but fraction D10 had a significantly higher concentration of His-GB1-WW2-3 when compared to the degraded protein band. Due to limitations of NMR, the concentration of this single fraction would not be sufficient. All fractions containing protein in the second peak were combined for testing thrombin digestion.

4.2.9 His-GB1-WW2-3 thrombin digestion

2 units of thrombin per mg of protein were added to a previously prepared sample of His-GB1-WW2-3 (figure 4.16). The samples were dialysed into Tris-HCl buffer at room temperature until the digest was complete.

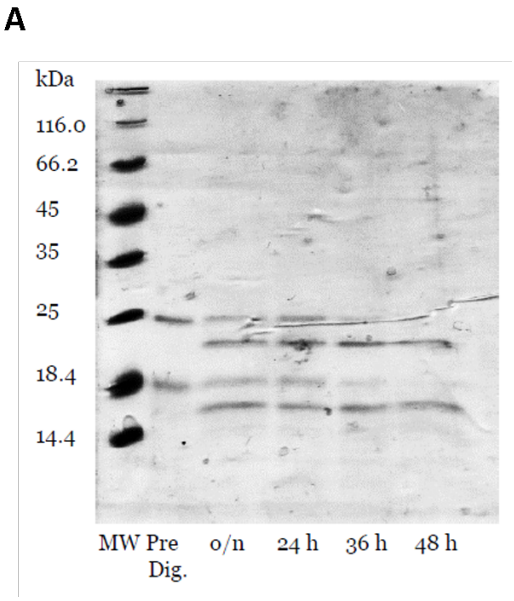


Figure 4.16: 15% SDS-PAGE gel showing unlabelled His-GB1-WW2-3 pre and post thrombin digest at different time points. 8 μL samples loaded.

Samples were taken after 16, 24, 36 and 48 hours until digestion was complete. The molecular weight of both bands had decreased and showed almost full digestion of the His tag after 48 hours. Whilst this sample was useful to test digestion with thrombin, the protein could not be used further due to the degradation observed.

4.2.10 100 mL His-GB1-WW2-3 test purification with urea

A 100 mL His-GB1-WW2-3 culture was expressed in ^{1}H, ^{15}N MEM. The protein was purified using a Ni-NTA column with 6M urea to denature proteases which could have caused the degradation observed previously (figure 4.17).

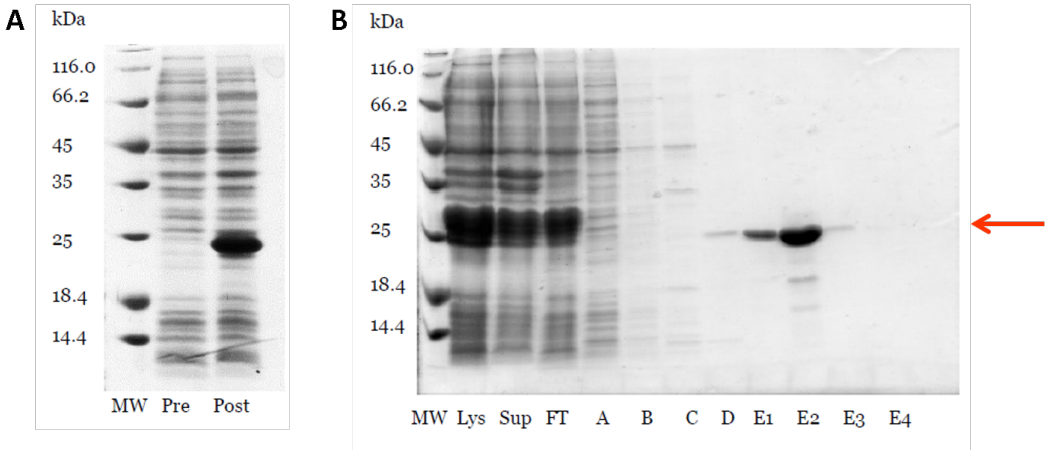


Figure 4.17: **(A)** 15% SDS-PAGE gel showing pre and post IPTG induction His-GB1-WW2-3. **(B)** 15% SDS-PAGE gel showing purification of His-GB1-WW2-3. His-GB1-WW2-3 is indicated by a red arrow. 10 μL samples loaded.

A small amount of degradation took place, as can be seen by a faint band appearing in the E2 lane. There is a clear distinction between the His-GB1-WW2-3 concentration and the degradation band in this gel and so fractions E1 and E2 were combined and dialysed into Tris-Cl buffer multiple times to remove the urea. The sample contained a small amount of precipitate and was compared to a soluble sample taken after centrifugation using SDS-PAGE (figure 4.18A). His-GB1-WW2-3 was visible in both samples and so the soluble fraction was digested with thrombin for 24 hours (figure 4.18B). Full digestion was not complete and so the sample was left a further 24 hours before purification 4.18C).

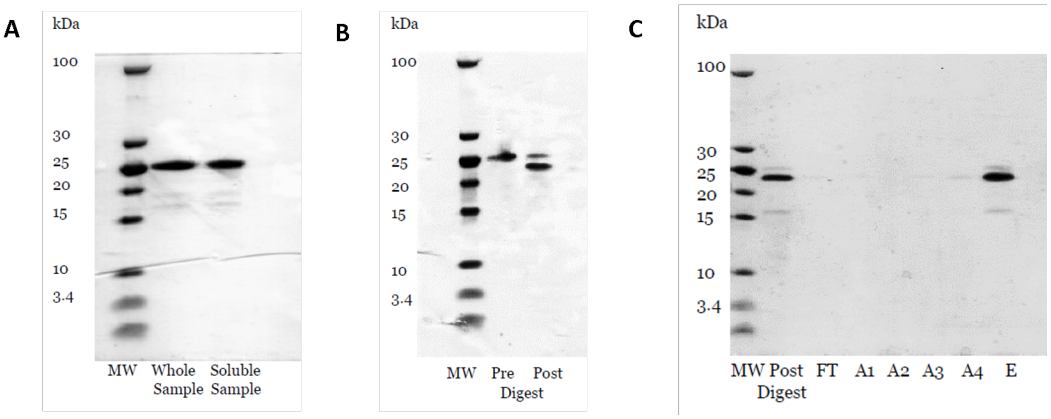


Figure 4.18: **(A)** A tricine gel showing the whole post purification sample of unlabelled GB1-WW2-3 and the soluble only sample post centrifugation. **(B)** A tricine gel showing His-GB1-WW2-3 samples pre and post-digestion with thrombin. Full digest had not yet completed. **(C)** A tricine gel showing post-digest sample after 48 hours and purification fractions of GB1-WW2-3. The elution fraction contained His-GB1-WW2-3 which suggests that the histidine tag was still present. 8 μL samples loaded.

His-GB1-WW2-3 bound to the Ni-NTA column and was collected during the elution step with a high concentration of imidazole. This indicated that the histidine tag had not been successfully digested; however, an NMR sample was prepared to determine if the construct was fully folded.

4.2.11 NMR experiments of His-GB1-WW2-3

Due to the small culture size of the GB1-WW2-3 urea test, the concentration of the NMR sample was low at just 0.05 mM. This required longer experiment times to get sufficient signal and a [^{1}H, ^{15}N]-HSQC was acquired for the sample (figure 4.19).

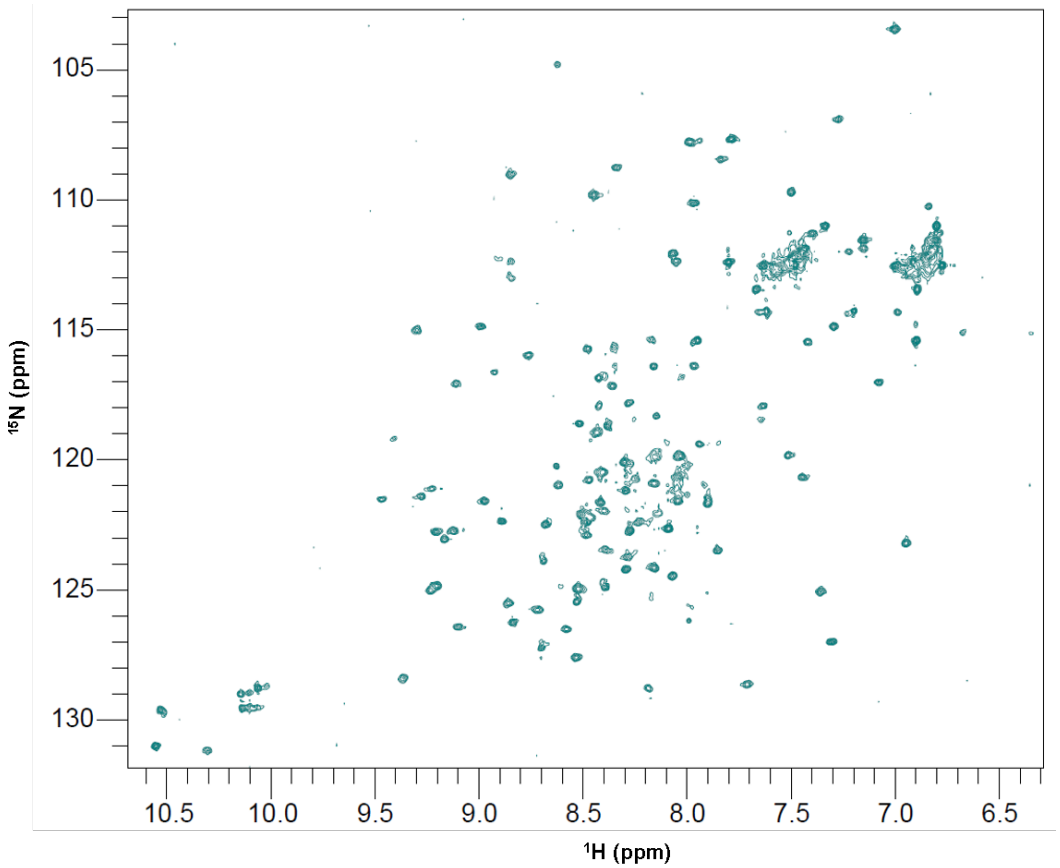


Figure 4.19: [^{1}H, ^{15}N]-HSQC of GB1-WW2-3 at a concentration of 0.05 mM, acquired at 298 K, 800 MHz. The clustered region of peaks in the middle of the spectrum indicated misfolded protein or aggregation in the sample.

The His-GB1-WW2-3 [^{1}H, ^{15}N]-HSQC showed some limited dispersion in the central, random coil region of the spectrum (8-9 ppm for ^{1}H), which is indicative of misfolded/aggregated protein. When overlaid with assigned GB1-WW2 (in red) and GB1-WW3 (in blue) [^{1}H, ^{15}N]-HSQC spectra (figure 4.20), it was evident that the number of peaks in the GB1-WW2-3 sample (in green) did not match the expected number and so backbone assignment was not attempted.

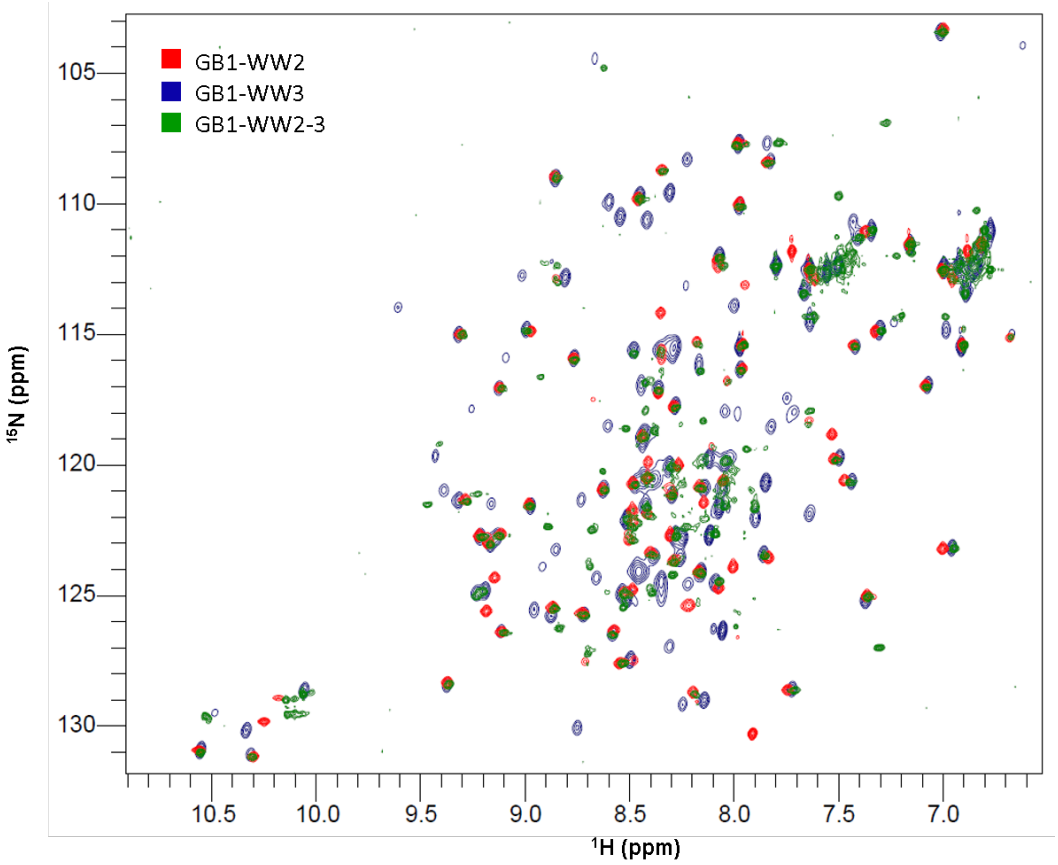


Figure 4.20: [^{1}H, ^{15}N]-HSQC overlay of GB1-WW2 (red), GB1-WW3 (blue) and GB1-WW2-3 (green). The number of peaks in the GB1-WW2-3 spectra was significantly less than the expected number of residues which would be observable.

4.3 Discussion

Previous research had highlighted the second WW domain of WWP2 as the main domain involved in the binding of OCT4 [Xu *et al.*, 2009], and that tandem WW domains had in certain cases, recognised target proteins with a higher binding affinity than individual WW domains [Aragón *et al.*, 2012]. This chapter aimed to provide structural information for the WW2 domain and explore the WW2-3 tandem domain.

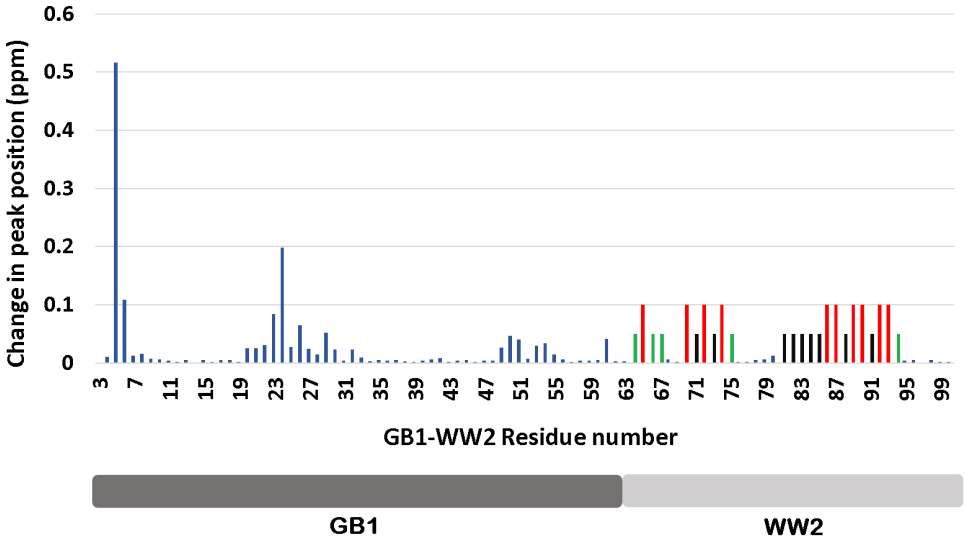


Figure 4.21: Changes in peak position (ppm) of GB1-WW2 backbone amide peaks before and after digestion of the His-tag. Residues which could be assigned in the [^{1}H, ^{15}N]-HSQC before and after His-tag digestion are shown in blue. Residues which appeared and could be assigned after His tag digestion are shown in red at a fixed height of 0.1 ppm for comparison. Residues which were unassigned in both [^{1}H, ^{15}N]-HSQC spectra are shown in black and prolines are shown in green at 0.05 ppm for clarity.

NMR assignments of GB1-WW2, in the presence and absence of the polyhistidine tag, revealed that the tag was having an effect on the number of assignments visible in the [^{1}H, ^{15}N]-HSQC spectra. Of particular interest, was the increase in peaks after digestion of the histidine tag. Thrombin digestion cleaved a total of 17 residues from the beginning of the His-GB1-WW2 sequence and yet an additional 10 peaks were present in the post thrombin digest [^{1}H, ^{15}N]-HSQC. This suggests that the tag was impacting the intermediate conformational exchange of the protein. A comparison of the peak positions before and after digestion of the His-tag (figure 4.21) saw that regions of the GB1 tag were observed to have moved. The WW2 domain was missing 19 observable residue assignments in the His-tagged [^{1}H, ^{15}N]-HSQC and WW2 residues which were present in both spectra were not observed to move. Red bars in figure 4.21 represent assigned residue peaks which appeared after digestion and were given the value 0.1 ppm. Here it is clear that the WW domain becomes more stable after cleavage of the His-tag as it moves out of intermediate exchange. Black bars

represent residues which are absent from both spectra and green bars are proline residues, both of which were given the value 0.05 ppm for clarity.

Looking at the GB1 region in more detail, three areas seem to have been affected most by the presence of the His-tag (residues 5, 6, 23, 24, 50 and 51). When these residues were highlighted on the GB1 domain NMR solution structure (figure 4.22), which was taken from the GB1-WW3 structure elucidated in chapter 3, it is clear that the three regions lie on the same side of GB1, furthest from the WW2 domain.

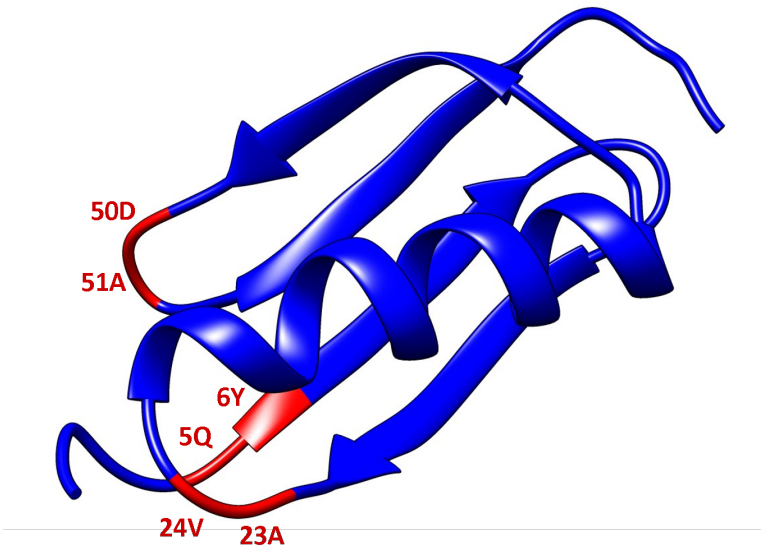


Figure 4.22: GB1 domain NMR solution structure taken from GB1-WW3 structure calculated in chapter 3. Residues which were observed to have moved position the most in the [^{1}H, ^{15}N]-HSQCs taken before and after digestion of the His-tag are highlighted in red.

The polyhistidine tag is small with little charge and many proteins have been observed to crystallise with an intact His-tag, with little or no impact on the protein structure [Carson *et al.*, 2007]. However, in this case, there is clear evidence that the tag has had an effect on GB1-WW2. This is potentially due to the tag being too close to the start of the GB1 domain and hence interfering with the environment of the GB1 domain residues in close proximity to the tag. Despite the change in amide peak position, the Cα and Cβ peaks of the affected residues in the presence and absence of the tag were compared and found to be the same and hence the

overall fold of the GB1 domain was not affected .

A large amount of data was still missing from the WW2 domain after digestion; however, a substantial number of peaks appeared suggesting that the removal of the tag brought some residues out of intermediate exchange. Before digestion of the His-tag, 75% of GB1-WW2 was assigned, as opposed to 82% afterwards. The assigned GB1-WW2 after digestion had 82 of 100 residues assigned. Of the missing eighteen residues, five were proline residues from the WW2 domain (Pro64, Pro66, Pro67, Pro75 and Pro94) and two were N-terminus residues. This left an additional eleven residues which were expected to be observed but were not assignable, which could be a result of certain residues still experiencing intermediate exchange. Four peaks were present in the [^{1}H, ^{15}N]-HSQC taken after digest of the tag, but these could not be assigned. Two were not observable in the triple resonance experiments and were therefore likely not real peaks. The remaining two did have very weak peaks in the triple resonance spectra but did not match any of the missing residues.

It was hoped that reducing the temperature of the digested GB1-WW2 sample would move the conformational exchange from an intermediate to a slow exchange regime and allow the missing residues to be assigned. Unfortunately, the temperature series failed to introduce any additional peaks. The side chain residues which were present appeared to move more dramatically as the temperature decreased, which could be due to the exchangeable protons (NH_2) and therefore changes in the degree of hydrogen bonding. The absence of additional peaks indicated that the missing residues were still experiencing intermediate exchange. An alternative option could have been raising the temperature, forcing the protein into a fast exchange process. As mentioned previously, this would lead to a single peak being present for each residue which would be observed at the weighted average chemical shift of all of the conformations present. However, this experiment was not attempted as the sample was required for further experiments and raising the temperature would increase the risk of denaturation.

Disorder predictions of the WW1-4 region suggested that the WW1, WW3 and WW4 domains would be ordered, whilst the WW2 domain lay on the border

between the ordered and disordered zone. When the GB1-WW2 construct underwent disorder prediction, the GB1 domain was mainly in the ordered zone, with a borderline ordered/disordered section towards the end of the GB1 sequence. There is then a section of predicted disorder in the region joining GB1 to WW2, which could be affecting the stability of the GB1-WW2 construct.

The tandem His-GB1-WW2-3 construct proved difficult to successfully express and purify. Despite a number of different approaches, including purification at low temperatures and in the presence of urea, at least a small amount of degradation was present in the final samples. The degradation band was observed by SDS-PAGE analysis at ∼18 kD. It had previously been noted that GB1-WW2 ran slightly high when analysed with SDS-PAGE, which led to the assumption that a similar characteristic could be observed for His-GB1-WW2-3. Bearing in mind that the degradation band still contained the His-tag, as it was present in the elution lanes of the purification step, it could be suggested that the construct was cleaving in the linker region between WW2 and WW3. The molecular weight of His-GB1-WW2-3 is ∼21 kD, compared to ∼17 kD which is the molecular weight of the His-GB1 tag along with WW2 and partial linker region (up to the start of the predicted highly disordered region – marked 91 Ser). The [^{1}H, ^{15}N]-HSQC acquired for the sample showed a high density of peaks around the 8.0-9.0 ppm (^{1}H) (random coil) region indicating that the protein had partially folded regions or aggregation in the sample. In addition, when overlaid with the GB1-WW2 and GB1-WW3 assignments, it was clear that a large number of peaks were missing from the spectra, particularly due to the fact that the linker region, at 41 residues long, was also present. Therefore, further experiments were not attempted for the tandem His-GB1-WW2-3 construct.

Assignment of the His-GB1-WW2 domain yielded only 72% of backbone assignments; however, an additional seven peaks were found and assigned after digestion of the His-tag. Unfortunately, 82% backbone assignment was less than required for calculating the protein structure and so the research focus moved towards studying the interactions of this domain with peptide regions of its targets, OCT4 and Smad7.

Chapter 5

Studying the interaction between GB1-WW2 and target region OCT4

5.1 Introduction

Previous studies have shown that E3 ubiquitin ligase mWWP2, a murine ubiquitin ligase with 96% sequence homology to human WWP2, can specifically interact with OCT4 through its WW domains, promoting the ubiquitination of OCT4 *in vitro* and *in vivo*; [Xu *et al.*, 2004]; [Xu *et al.*, 2009]. Liao and Jin proved in 2010 that mWWP2 (WWP2 from mouse) plays a role in the ubiquitination and degradation of OCT4 at the proteasome. There, it was determined that ubiquitination of OCT4 occurs through the Lys-63 linkage of ubiquitin, which mediates the degradation of OCT4 through the 26S proteasome [Liao and Jin, 2010]. In 2009, Xu et al. went on to determine that human WWP2 specifically interacts with OCT4 and promotes its ubiquitination *in vitro* and *in vivo*; [Xu *et al.*, 2009].

Murine mWWP2 was also found to be involved in controlling OCT4 levels

during differentiation of ECCs and is able to catalyse OCT4 ubiquitination as well as auto-ubiquitination, in a dosage-dependent manner, ensuring the levels of OCT4 are regulated to preserve pluripotency [Xu *et al.*, 2004]; [Liao and Jin, 2010]. When OCT4 is overexpressed, differentiation of ES cells increases, in contrast to a reduction in OCT4 expression, which leads to loss of pluripotency and dedifferentiation to the trophectoderm [Niwa *et al.*, 2000]. Downregulation of WWP2 expression, which was inhibited by specific RNA interference (RNAi), was found to increase OCT4 levels in undifferentiated cells, indicating its significance as a regulator of OCT4 in ES cells [Xu *et al.*, 2009].

The study by Xu et al. in 2004, described a HECT domain of mWWP2 which is essential for the ligation of ubiquitin to OCT4; however, neither the HECT or C2 domain are involved in OCT4-mWWP2 binding. GST pull-down studies proved that a construct containing just the four WW domains (WW1-4) interacted with OCT4 and hence demonstrated that it is the WW domain region which is involved in the binding to OCT4. This was consistent with data obtained for other HECT E3 ubiquitin ligases [Xu *et al.*, 2004]. In 2009, it was reported that the second of the four sequential WW domains (WW2) of human WWP2 was found to have highly enhanced interactions with OCT4 when compared with the other three WW domains [Xu *et al.*, 2009] and so the second WW domain (WW2) will be the focus of this chapter. Unless otherwise specified, where WWP2 is mentioned in this thesis, it is referring to human WWP2.

NMR titrations allow the exploration of interactions at an atomic level and so can be used to monitor changes in a target protein upon binding to a ligand. Specific residues which are involved in binding can be identified and their affinities can be quantified by observing changes in backbone amide [^{1}H, ^{15}N]-HSQC chemical shift positions. Changes in chemical shifts occur when the environment of an atom differs, which can arise due to conformational changes, where ligand binding affects another site on the same protein, as well as a direct interaction with the binding substrate, changing the local magnetic field experienced by the residue involved.

5.1.1 Experimental aims

The aim of this chapter is to further investigate the interaction between the WW2 domain of human WWP2, as a fusion protein with GB1, and a peptide from substrate protein OCT4. The OCT4 peptide was designed to encompass the PPxY motif, which is found in the NTD^{A} domain of OCT4A isoform (section 1.4). It was either expressed as a recombinant fusion protein with His-SUMO tag or purchased as a synthetic peptide from PeptideSynthetics. The acquisition of triple resonance NMR experiments, using a sample of ^{15}N, ^{13}C labelled GB1-WW2, fully saturated with unlabelled OCT4 peptide, allowed resonance assignment to be accomplished as the interaction with OCT4 stabilised the protein, which was previously in intermediate exchange. NMR titrations and ITC experiments were acquired and aimed to determine the affinity of the interaction between OCT4 peptide and WW2.

5.2 Results

5.2.1 GST-OCT4 pull-downs

Expression of GST-OCT4 was optimised for pull-downs with WWP2 constructs containing different domains. GST-OCT4 was successfully expressed, however, during the initial purification steps, the GST tag cleaved from the OCT4. Purification of a new culture was attempted with great caution at 4 °C and with short incubation times on the glutathione column. The GST-OCT4, however, was very unstable and degraded regardless. Due to GST-OCT4 instability and time restraints, this construct was abandoned. The project instead focused on an OCT4 peptide construct expressed as a fusion protein with a SUMO tag, which was used to study the interaction between OCT4 and the WW domains.

5.2.2 NMR titration of GB1-WW2 with synthetic OCT4 peptide

For NMR titrations of GB1-WW2 with OCT4 peptide, ^{15}N GB1-WW2 was expressed and the protein was purified as in chapter 4 and a 0.4 mM sample was prepared as in section 2.5.1. A short synthetic peptide of OCT4 containing the PPxY motif was used (sequence in section 2.5.8.2) so that titration data could be obtained quickly without having to optimise expression and purification of recombinant peptide. The OCT4 peptide was titrated into a 0.4 mM sample of ^{15}N GB1-WW2 by mass, mixed by inversion to avoid solubility problems and a [^{1}H, ^{15}N]-HSQC was acquired after each addition. Nine titration points were recorded, from 1:0 protein to peptide ratio, up to 1:8.75, as shown in table 2.11 in section 2.5.8.2.

Figure 5.1 shows the superimposition of the nine spectra, ranging from ratios of 1:0 GB1-WW2 to OCT4, up to 1:8.75, with peaks labelled by resonance assignment of a fully saturated GB1-WW2:OCT4 sample (described in section 5.2.4). At the titration start point, the sample contained GB1-WW2 at a concentration of 0.4 mM. At this starting point, a number of peaks were missing from the [^{1}H, ^{15}N]-HSQC due to intermediate exchange within the protein (explained in section 1.6.1.5).

This occurs when regions or residues are in undefined conformations and so they experience differing environments, causing peaks to become broadened until they are not visible in the spectrum. Several peaks are observed to shift position or disappear throughout the titration and the disappearing peaks were potentially present elsewhere in the spectrum at the end of the titration.

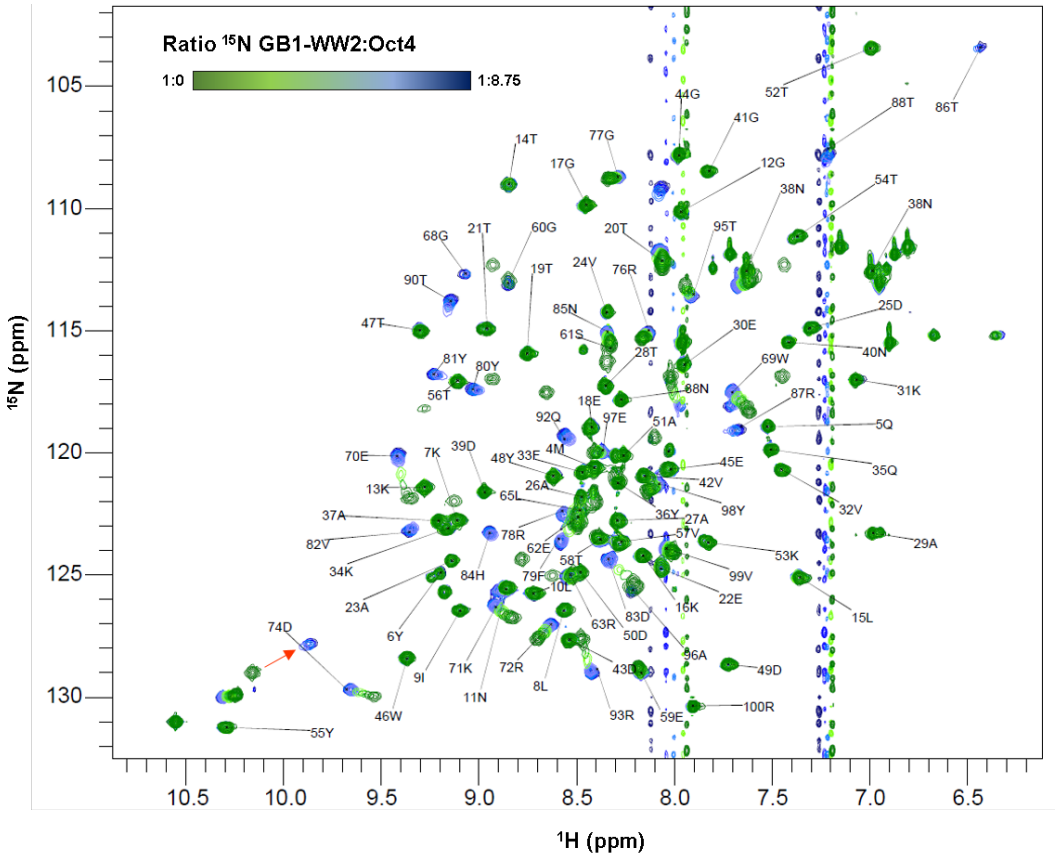


Figure 5.1: An overlay of the [^{1}H, ^{15}N]-HSQC spectra showing 0.4 mM GB1-WW2 and unlabelled OCT4 titration with GB1-WW2 peaks labelled at the saturated position. The 1:0 GB1-WW2 to OCT4 spectrum starts in green and changes through blue to navy as the concentration of OCT4 is increased to 1:8.75. The small red arrow follows the slow exchange movement of a tryptophan side chain peak. The experiments were performed using a Bruker Avance I 500 MHz spectrometer at 298 K. Sample buffer: 20 mM Sodium phosphate buffer, 50 mM NaCl, 15 mM DTT, pH 6.8.

Despite peaks moving and disappearing as the titration progresses, the overall number of peaks increased on saturation of GB1-WW2 with OCT4 peptide and seven additional peaks were present when compared with the initial spectrum of GB1-WW2 alone. This indicated that the protein was moving out of intermediate exchange, consistent with an interaction taking place. The residues with newly visible peaks were in well-defined conformations and the protein was more stable. GB1 peaks were not affected by the addition of the peptide indicating that the tag did not interact with the ligand.

It was noted that whilst some peak movement was observed at ratio 1:0.375, peak positions did not change significantly until ratio of 1:2.75 (sky blue). This could be explained by peaks experiencing slow exchange, where their intensity may change but not the peak position until a large % of the ligand is bound, at which point the conformation jumps to a new position. When the ratio was increased to 1:4.5, migrating peaks reached their end destination and intensity did not increase upon further addition of OCT4. A further two titration points were acquired 1:7 and 1:8.75 to confirm saturation. Three unassigned tryptophan side chain peaks can be seen in the characteristic region in the bottom left of figure 5.1. The bottom peak is stationary throughout the titration and hence is likely to be GB1 peak 46W. This was confirmed by overlaying a previously assigned [^{1}H, ^{15}N]-HSQC of GB1-WW3. Of the two remaining WW2 tryptophan side chain peaks, 69W and 91W, one is located above 55Y, observed moving in fast exchange, and the other is to the right of this and is moving in slow exchange (present in one position at the beginning of the titration and in a different position at the end - movement marked by a red arrow).

Figure 5.2 shows zoomed regions of the GB1-WW2:OCT4 titration. In all four images, A-D, GB1 peaks (1-61) are unchanging in position, proving that these OCT4 peptide residues do not interact with the GB1 domain. The detailed regions also highlight previously unassignable WW2 peaks appearing as OCT4 peptide is added. Figure 5.2(**A**) shows the GB1 peak 56T remaining stationary throughout the titration, and the previously unassignable WW2 residue peaks, 80Y and 81Y, appearing as OCT4 peptide is added. 5.2(**B**) shows the GB1 peaks 16K and 22E are not affected by OCT4 peptide as opposed to 99V which can be seen to shift slightly in fast exchange. Figure 5.2(**C**) shows 70E moving in intermediate exchange as OCT4 peptide is added, with broadened peaks in the intermediate region between the start and end point. The disappearing peak in figure 5.2(**D**) could not be assigned and so it was not possible to match this with an appearing peak.

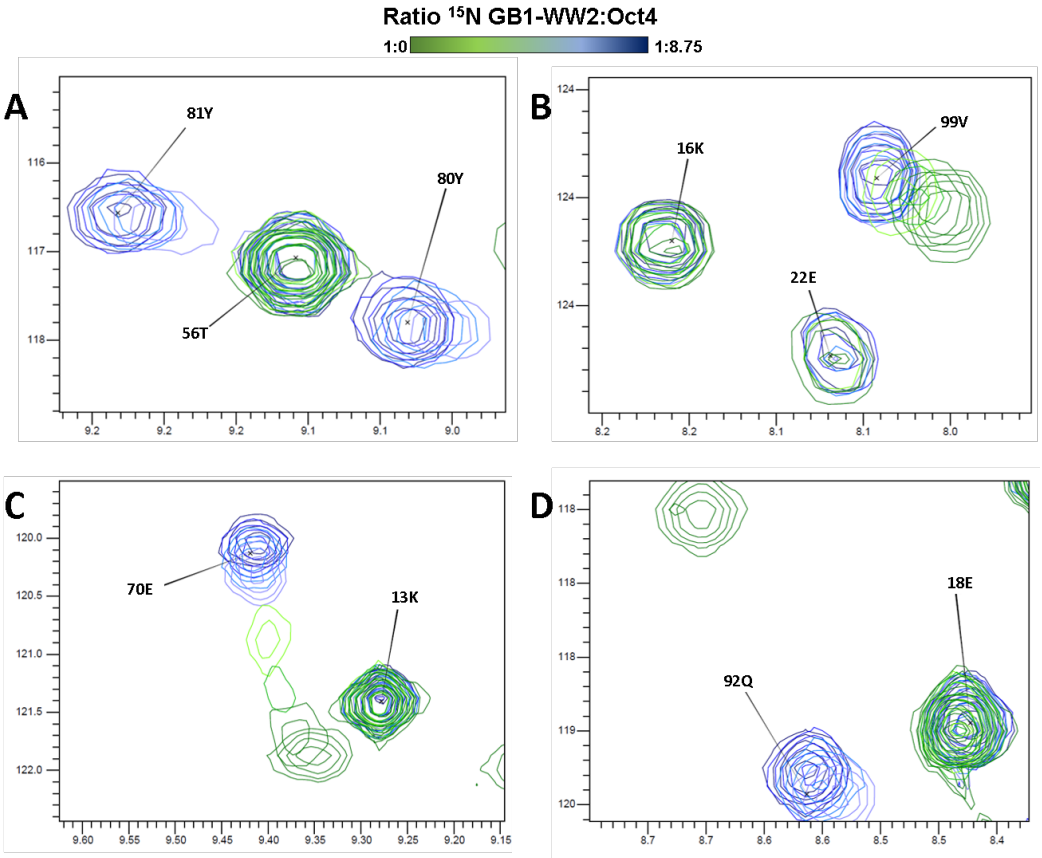


Figure 5.2: Detailed regions from GB1-WW2:OCT4 titration [^{1}H, ^{15}N]-HSQC's showing stationary GB1 peaks and migrating, disappearing and appearing WW2 peaks. (**A**) GB1 peak 56T remains stationary throughout the titration, but previously unassignable WW2 residue peaks, 80Y and 81Y, appear as OCT4 peptide is added. (**B**) GB1 peaks 16K and 22E are not affected by OCT4 peptide, 99V can be seen to shift slightly in fast exchange. (**C**) 13K does not move during the titration and 70E moves in intermediate exchange. (**D**) GB1 peak 18E is not affected and 92Q appears at the end of the titration. Another peak in the top left is observed to disappear with the addition of OCT4 peptide. The experiments were performed using a Bruker Avance I 500 MHz spectrometer at 298 K. Sample buffer: 20 mM Sodium phosphate buffer, 50 mM NaCl, 15 mM DTT, pH 6.8. The concentration of GB1-WW2 was 0.4 mM.

Figure 5.3 shows the shift trajectories for the peaks which were assigned in both spectra, plotted against residue number. The weighting of shift changes were adjusted due to the difference in the gyromagnetic ratio between hydrogen and nitrogen, as previously described in section 3.2.6.

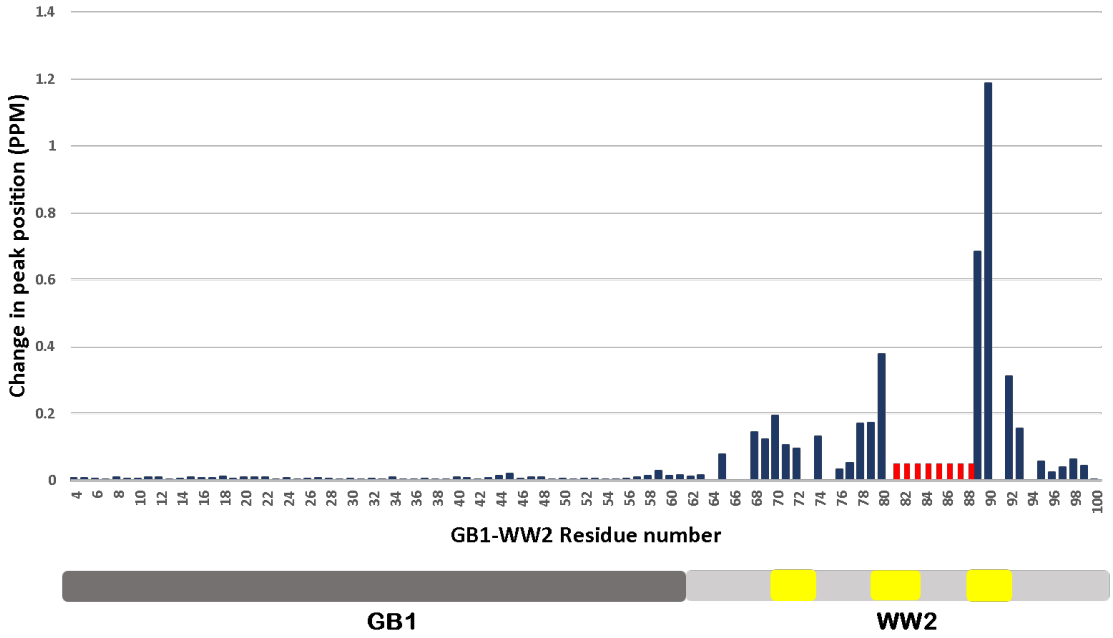


Figure 5.3: The trajectories (in ppm) of GB1-WW2 backbone amide peaks which could be assigned in the [^{1}H, ^{15}N]-HSQC at the start and end of the titration with OCT4 peptide. GB1 peaks (1-61) remain stable throughout the titration, as opposed to the WW domain residues (62 and above), which show larger trajectories. Residues which were only present after the addition of OCT4 peptide are marked in red at 0.05 ppm for comparison. Unassignable proline residues are marked as gaps in the graph. Underneath is a schematic showing the GB1 region in grey and the WW2 domain in light grey. The secondary structure prediction of WW2 domain is shown in yellow, with the three highlighted elements representing predicted β strands.

The peak positions of GB1 peaks (residues 1-61) were stable throughout the titration, indicating the tag did not interact with the OCT4 peptide. The WW domain residues (62 and above) show generally much larger trajectories than the GB1 tag. As a number of WW2 residues were missing from the GB1-WW2 assignments in the absence of OCT4 peptide, as well as issues with propagating some WW2 peak assignments from other samples, the trajectory data available was limited. Instead, residues which were only visible after the addition of OCT4 peptide are marked in red, as these residues were previously in intermediate exchange but when saturated with OCT4 they moved to a stable conformation. This indicates that this region folds upon binding OCT4; however, since the peak movement could not be followed, the change in peak position cannot be quantified. The PSIPRED webserver v3.3 [Jones *et al.*, 2018] was used to predict

areas of secondary structure [Buchan *et al.*, 2013]. The three yellow regions highlighted on the schematic diagram in figure 5.9, represent predicted β strands. The region 70-73 is difficult to interpret due to lack of residue data but 79-82 contains residues which only appear after the addition of OCT4. The third region 88-91, contains the two largest trajectories observed, suggesting that these residues are part of the region which is involved in the binding of OCT4.

5.2.3 Recombinant OCT4 peptide expression, purification and preparation of GB1-WW2:OCT4 NMR sample

For assignment of saturated GB1-WW2:OCT4 peptide, a large sample of OCT4 peptide was required, and for that reason, the peptide was expressed recombinantly as a SUMO fusion protein (plasmid provided by Jessica Watt from the Chantry laboratory at UEA). SUMO-OCT4 was transformed in BL21 (DE3) E. coli and induced with 0.8 mM IPTG overnight at 30 °C (typical expression gel shown in figure 5.4).

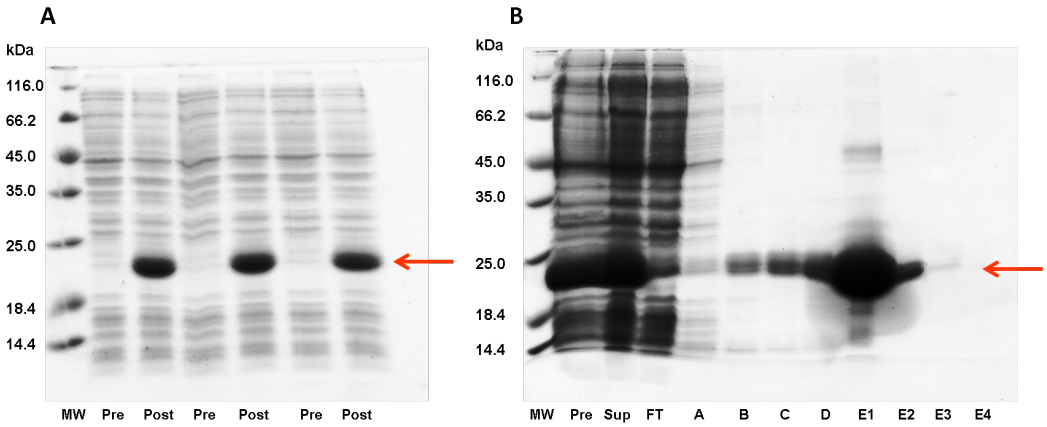


Figure 5.4: Example 15% SDS-PAGE analysis showing; **A** pre and post-induction samples of SUMO-OCT4 peptide (indicated by an arrow at ~25 kDa - SUMO-OCT4 runs higher than its molecular weight of ~15 kDa). **B** Ni-NTA purification of unlabelled SUMO-OCT4 peptide. Due to high concentrations of SUMO-OCT4, small amounts of purified protein were collected during the wash fractions B, C and D. The highest concentration of SUMO-OCT4 was found in the elution fraction E1, and so E1 and E2 were combined and taken forwards. Imidazole wash concentrations: wash A = 1 mM, wash B = 10 mM, wash C = 20 mM, wash D = 30 mM and elution E = 250 mM. 10 μL samples were loaded.

The protein was purified using a gravity flow Ni NTA column and purification was confirmed using SDS-PAGE (figure 5.4B). Washes A, B, C and D contain 1 mM, 10 mM, 20 mM and 30 mM imidazole respectively. The Elution wash (E) contains 250 mM imidazole. Sample E1 shows there are some impurities in the sample; however, the difference in concentration between impurities and protein meant that the sample could still be used. The large quantity of SUMO-OCT4 protein could have caused the leakage of purified protein into washes B, C and D due to the column being near binding capacity. Samples E1 and E2 were combined and dialysed into Tris-Cl buffer. A 500 μL aliquot of ULP1, as described in section 2.3.1.4, was added to the sample (500 μL per 1 L culture volume) and left at room temperature overnight with agitation to digest the SUMO tag (figure 5.5A).

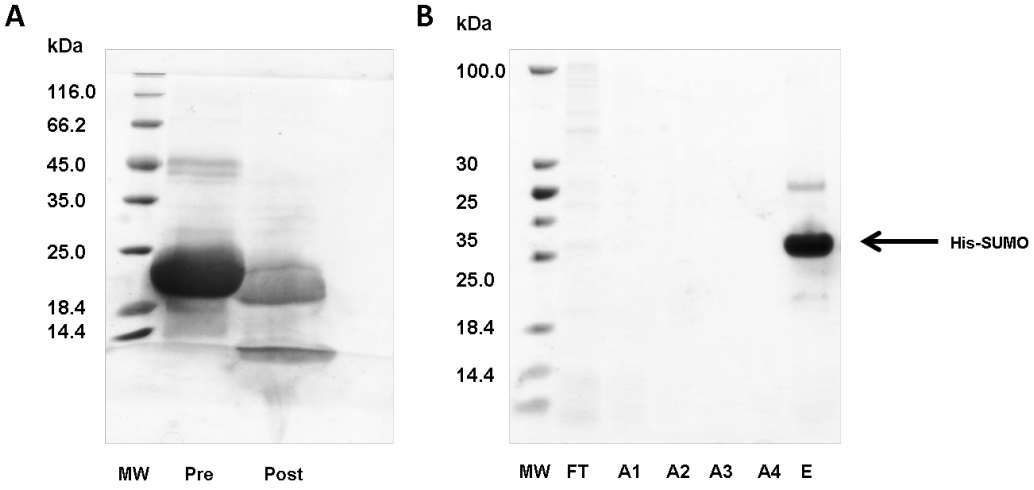


Figure 5.5: **A** 15% SDS-PAGE analysis showing pre and post-digest of SUMO-OCT4 peptide. **B** Tricine gel showing Ni-NTA purification of unlabelled OCT4 peptide. The purified peptide was collected in the FT and washes and the His-SUMO tag is in elution fraction E. Imidazole wash concentrations: wash A = 1 mM, wash B = 10 mM, wash C = 20 mM, wash D = 30 mM, elution E = 250 mM. 10 μL samples loaded.

After digestion, a SUMO tag band could be seen at a slightly lower molecular weight, consistent with the cleavage of the OCT4 peptide. The lower band, which appears towards the bottom of the post-digest lane is a contaminant running with the gel front and not the OCT4 peptide. The peptide could not be easily visualised due to its small size binding less Coomassie brilliant blue. Silver staining and

Tricine gels were attempted as an alternative for visualising the OCT4 peptide but were unsuccessful. The post-digestion sample was repurified to remove the His-SUMO tag (seen in lane E of figure 5.5B), whilst flow through and washes A1, A2, A3 and A4 were pooled to collect the OCT4 peptide.

After removal of the SUMO tag by Ni-NTA purification, the peptide was dialysed into ammonium bicarbonate buffer and freeze-dried to concentrate. Since some solubility issues were initially encountered when resuspending the OCT4 peptide, the physicochemical properties were estimated using INNOVAGEN Peptide Solubility calculator:

(https://pepcalc.com/peptide-solubility-calculator.php).

The estimation suggested that OCT4 peptide had an isoelectric point of 0.83 and hence, the pH of this acidic peptide needed to be raised. The peptide was stored lyophilised until it was required, at which point it was resuspended in NMR buffer with small additions of NaOH until no precipitate remained. SDS-PAGE was then used to confirm that the sample had not precipitated or degraded.

5.2.4 NMR resonance assignment

A [^{1}H, ^{15}N]-HSQC was acquired for a 0.4 mM of ^{15}N, ^{13}C labelled GB1-WW2 and 0.7 mM of unlabelled OCT4 (the yield of a 1 L culture), on a 500 MHz Bruker Avance I NMR spectrometer at 25 °C. When compared to the saturated spectra from GB1-WW2:OCT4 titration, several peaks were seen in an intermediate state between GB1-WW2 and the saturated spectrum, hence it was determined that it had not reached saturation.

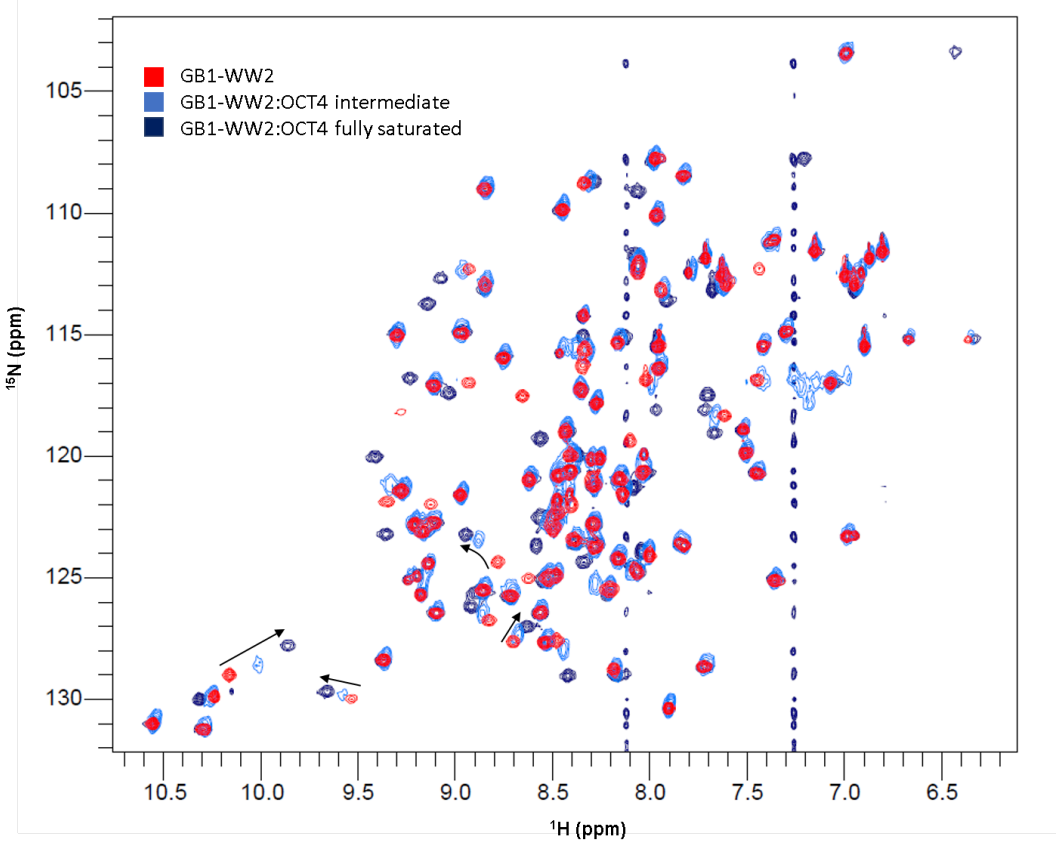


Figure 5.6: An overlay of the [^{1}H, ^{15}N]-HSQC spectra showing GB1-WW2 alone (red), GB1-WW2:OCT4 fully saturated from the titration end point (navy) and ^{15}N, ^{13}C labelled GB1-WW2:OCT4 sample (blue). Arrows show the peak movement as OCT4 peptide is added. It is clear that the ^{15}N, ^{13}C labelled GB1-WW2:OCT4 sample is in an intermediate position between unsaturated and saturated samples. Arrows highlight regions where peak migration is clear. The experiments were performed using a Bruker Avance I 500 MHz spectrometer at 298 K. Sample buffer: 20 mM Sodium phosphate buffer, 50 mM NaCl, 15 mM DTT, pH 6.8.

Figure 5.6 shows an overlay of titration points GB1-WW2 (red) and GB1-WW2:OCT4 fully saturated (navy), with the ^{15}N, ^{13}C GB1-WW2:OCT4 sample (blue). Arrows show several examples of the migration of GB1-WW2 peaks from before addition of OCT4 peptide, to fully saturated GB1-WW2. It is observed that several peaks which move with addition of OCT4 are in an intermediate state, somewhere between WW2 alone and the saturated sample. This indicates that the sample was only partially saturated with OCT4 peptide.

This sample could not be used for assignment of GB1-WW2 in an excess of

OCT4 peptide as it had not reached saturation and so more OCT4 peptide was required. SDS-PAGE analysis showed degradation bands in the previous OCT4 peptide purification and so it is likely the concentration was lower than expected and hence, was unable to saturate the GB1-WW2 sample.

Future preparations were repeated on a larger scale and purification was carried out at 4 °C to avoid degradation. Recombinant OCT4 peptide was added to the NMR sample along with 2 mg of synthetic peptide, which produced a [^{1}H, ^{15}N]-HSQC resembling the saturated titration end point spectrum. HNCACB and CBCACONH experiments were then acquired in a 500 MHz Bruker Avance I NMR spectrometer at 25 °C and were used to manually assign the backbone of GB1-WW2 in the presence of OCT4 peptide (figure 5.7), using the method described in 1.6.1.4.

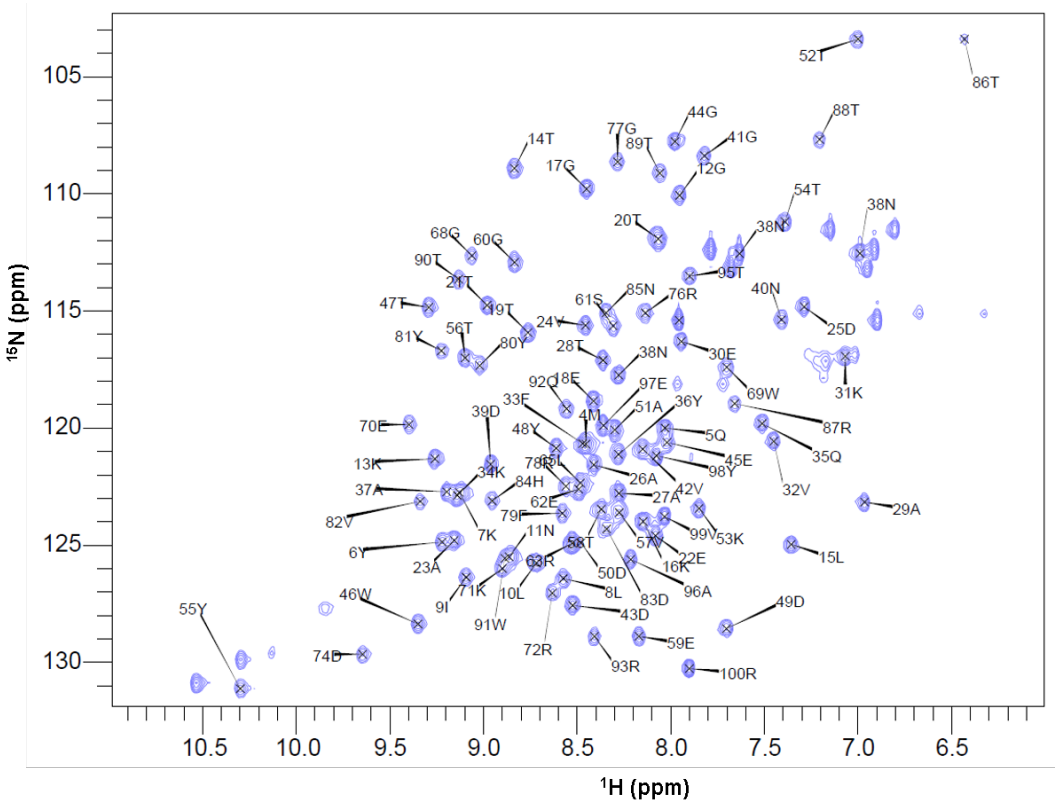


Figure 5.7: A [^{1}H, ^{15}N]-HSQC spectrum of ^{15}N, ^{13}C labelled GB1-WW2 in the presence of an excess of unlabelled OCT4 peptide. 93% of residues were assigned and labels are present for these assignments. This experiment was performed using a Bruker Avance I 500 MHz spectrometer at 298 K. Sample buffer: 20 mM Sodium phosphate buffer, 50 mM NaCl, 15 mM DTT, pH 6.8.

93% of GB1-WW2:OCT4 was assigned, leaving 7 residues unassigned, 5 of which were proline residues from the WW2 domain (Pro64, Pro66, Pro67, Pro75 and Pro94), 2 N-terminus residues and residue 73Thr. Residues at the N-terminus may not be observed due to the quick exchange of protons in the N-terminal NH_2 groups. A full list of the assigned residues and resonances is included in Appendix C.

Some slight degradation was observed when the [^{1}H, ^{15}N]-HSQC taken after the HNCACB experiment (red) was overlaid with the [^{1}H, ^{15}N]-HSQC spectra acquired before (blue) (see figure 5.8).

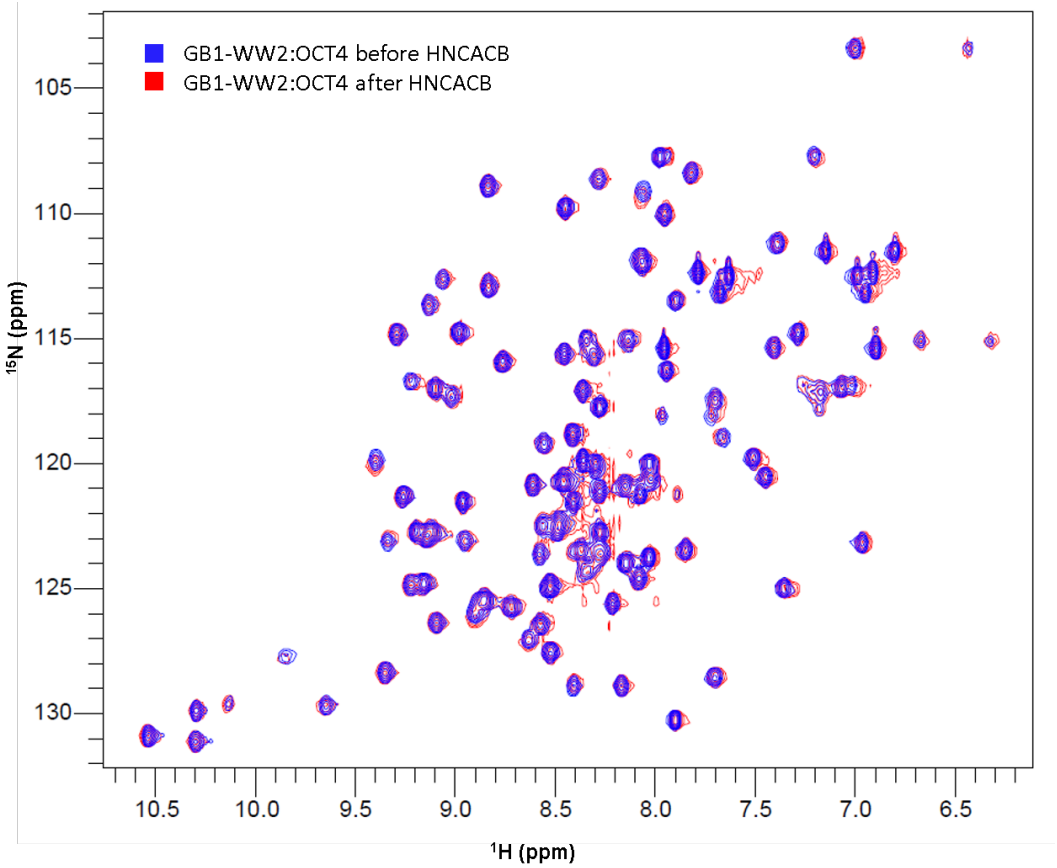


Figure 5.8: An overlay of ^{15}N, ^{13}C labelled GB1-WW2:OCT4 [^{1}H, ^{15}N]-HSQC spectra acquired before (blue) and after (red) acquisition of the HNCACB spectrum. Contour levels were set to the same value. Small changes in peaks were observed after the HNCACB. The experiments were performed using a Bruker Avance I 500 MHz spectrometer at 298 K. Sample buffer: 20 mM Sodium phosphate buffer, 50 mM NaCl, 15 mM DTT) pH 6.8.

These degradation peaks were minor at this stage and below the necessary contour level and so this NMR experiment could still be used for backbone assignment. Planned future experiments, HNCO and HNCACO, which can be used to assist with backbone assignment, were not acquired.

Figure 5.9 shows the comparison (in ppm) of all GB1-WW2 backbone amide peaks which are assigned in the GB1-WW2 [^{1}H, ^{15}N]-HSQCs taken in the presence and absence of OCT4 peptide.

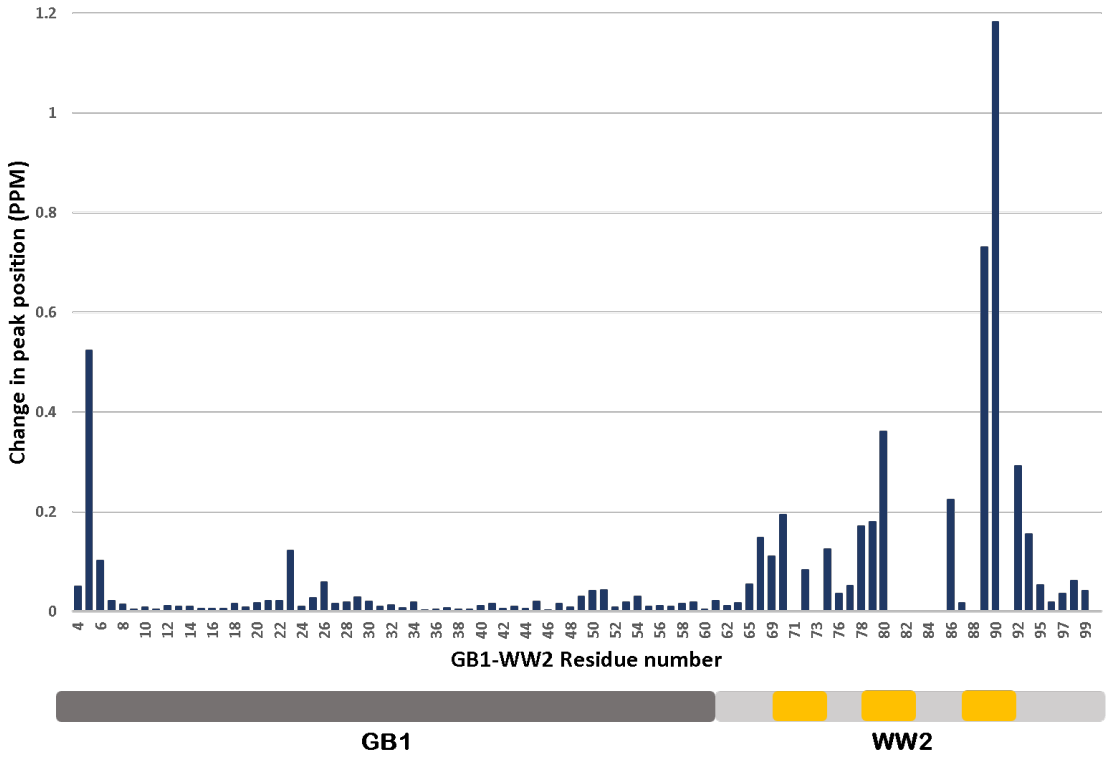


Figure 5.9: A comparison (in ppm) of all GB1-WW2 backbone amide peaks which are assigned in both the GB1-WW2 [^{1}H, ^{15}N]-HSQC and the saturated bound GB1-WW2:OCT4 [^{1}H, ^{15}N]-HSQC. GB1 peaks (1-61) are generally stable, with the exception of the N-terminus and residue 23. The WW domain residues (62 and above) show larger trajectories. Underneath is a schematic showing the GB1 region in grey and the WW2 domain in light grey. Secondary structure prediction of the WW2 domain is shown in yellow, with the three highlighted elements representing predicted β strands.

Shift changes were weighted as described previously (section 3.2.6). The WW domain residues (62 and above) generally have much larger trajectories than the

GB1 tag and some residues are missing data due to lack of GB1-OCT4 assignments in the absence of OCT4 peptide. The largest peak differences in the WW2 domain match the titration peak trajectory data, reiterating that this region is involved in the binding of OCT4.

The peak positions of GB1 peaks (residues 1-61) were expected to be stable, as with the titration, which would confirm that the tag did not interact with the OCT4 peptide. In contrast to the titration data, some differences were observed at the N-terminus and residue 23.

SDS-PAGE analysis of the digest and benzamidine purification gel for this sample (shown in figure 5.10) was inspected and a small amount of undigested His tagged protein was found to still be present. This could have been caused by a problem with the Ni-NTA column used.

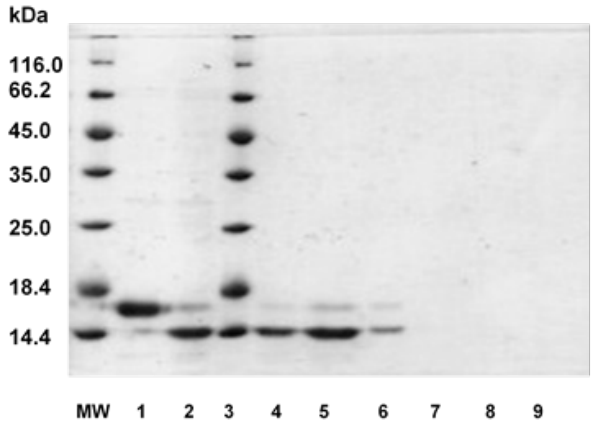


Figure 5.10: His-GB1-WW2 digest using thrombin and subsequent thrombin removal using a benzamidine column. 10 μL samples loaded. **1**: Pre-digestion, **2**: Post-digestion, **3**: MW marker, **4**: Wash A1, **5**: Wash A2, **6**: Wash A3, **7**: Elution 1, **8**: Elution 2, **9**: Elution 3.

CCPN Analysis was unable to provide a good K_d fit. Instead, an estimation of K_d was determined for residue 74D which experienced fast exchange during the GB1-WW2:OCT4 titration by modelling the fit of the peak trajectories to a binding curve in Excel, as shown in figure 5.11. The peak trajectories from the NMR titration are shown in blue and the K_d fit is shown in red. This graph shows

that saturation was achieved as the last three titration points are roughly in line with each other. With a K_d of 150 μM, a reasonably good fit is achieved. The value of K_d is equal in magnitude to the concentration of free ligand at which half the protein molecules are bound to ligand. The graph displays the total ligand concentration on the x axis, meaning the free ligand plus the bound ligand.

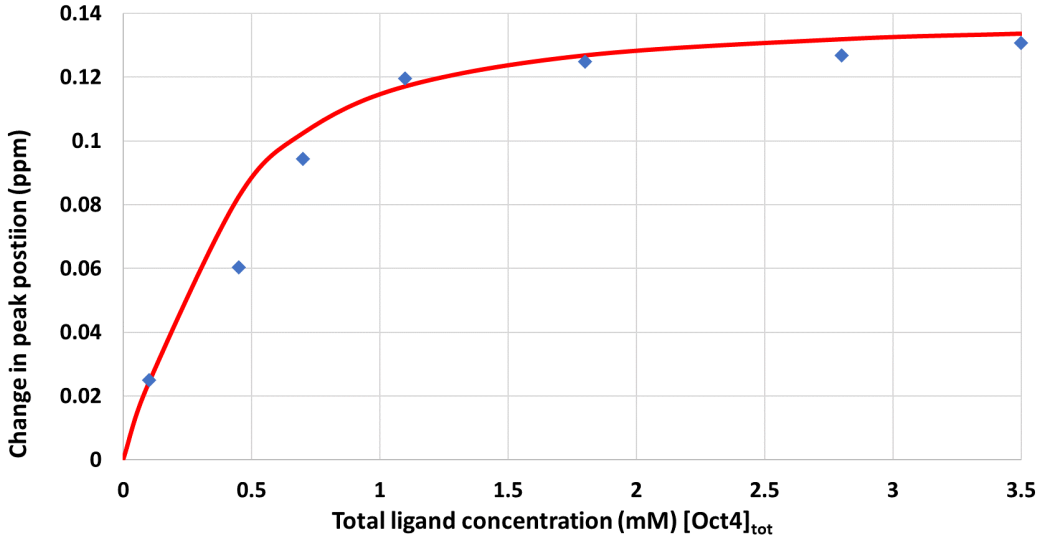


Figure 5.11: The K_d fit for GB1-WW2:OCT4 peptide, achieved by modelling the titration trajectory of fast exchange residue 74D using Excel. The manual K_d fit is shown in red and the peak trajectories for residue 74D are shown in blue. The resultant K_d was 150 μM; however, this is only a manual approximation.

5.2.5 Isothermal Titration Calorimetry experiments

Isothermal titration calorimetry, described in section 1.7, was used to further investigate interactions between the WW2 domain and a peptide of OCT4. The experiment was designed using Malvern MicroCal PEAQ-ITC Analysis Software and the binding parameters were estimated as 1 binding site and $K_d = 1$ μM. The concentration of GB1-WW2 was calculated, using the absorbance at 280 nm, to be 30 μM and so OCT4 was resuspended, aiming for a 300 μM sample, as suggested for an ideal binding curve. Due to difficulties with peptide solubility at high concentrations, the OCT4 peptide concentration was $\sim$ 210 μM after resuspending the freeze-dried sample. Some precipitate was visible in the sample and so it was

centrifuged, and the supernatant was transferred to a fresh microcentrifuge tube before the concentration was measured. Because of this lower concentration of ligand, the GB1-WW2 sample was diluted to $\sim$ 23 μM in an attempt to keep a similar ratio of GB1-WW2:OCT4. Both samples were dialysed into the same batch of Tris-Cl buffer before concentrations were finely adjusted.

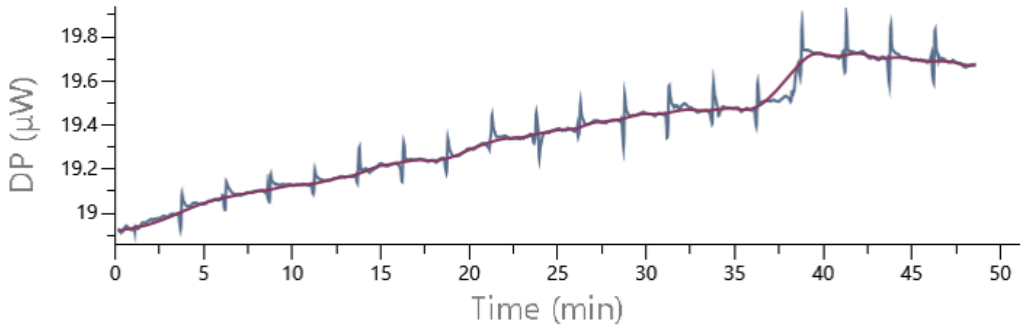


Figure 5.12: Control ITC experiment showing heat of dilution for OCT4 peptide titrated into buffer. The mean of the control dataset was subtracted from GB1-WW2:OCT4 titrations using Malvern MicroCal PEAQ-ITC Analysis Software.

Three GB1-WW2:OCT4 ITC titrations were carried out, along with three control experiments to determine the heat of dilution of OCT4 and GB1-WW2 and to ensure that the buffers were matched. This was achieved by injecting OCT4 into buffer and buffer into GB1-WW2 respectively, as well as a buffer into buffer control. The heat of dilution of GB1-WW2 was minimal and so could be ignored, whereas the heat of dilution of OCT4 (figure 5.12) was subtracted from the GB1-WW2:OCT4 ITC data. The observed jump in the baseline of the OCT4 into buffer control was not of concern as it is a common occurrence if a small bubble is present in the syringe.

Figure 5.13 shows a representative raw plot of 23.8 μM GB1-WW2 titrated with 210 μM OCT4 peptide at 25 °C, and below an image of the integrated heat change. In the raw heat plot, each injection point can be identified by a distinct drop in DP, the area of which represents the heat exchanged. The first injection is a smaller volume as it is less accurate due to ligand diffusing into the cell during the temperature equilibration. This interaction was relatively weak, and

due to concentrations of protein and peptide, saturation was not achieved. Because of this, a K_d could not be determined using the Malvern MicroCal PEAQ-ITC Analysis Software.

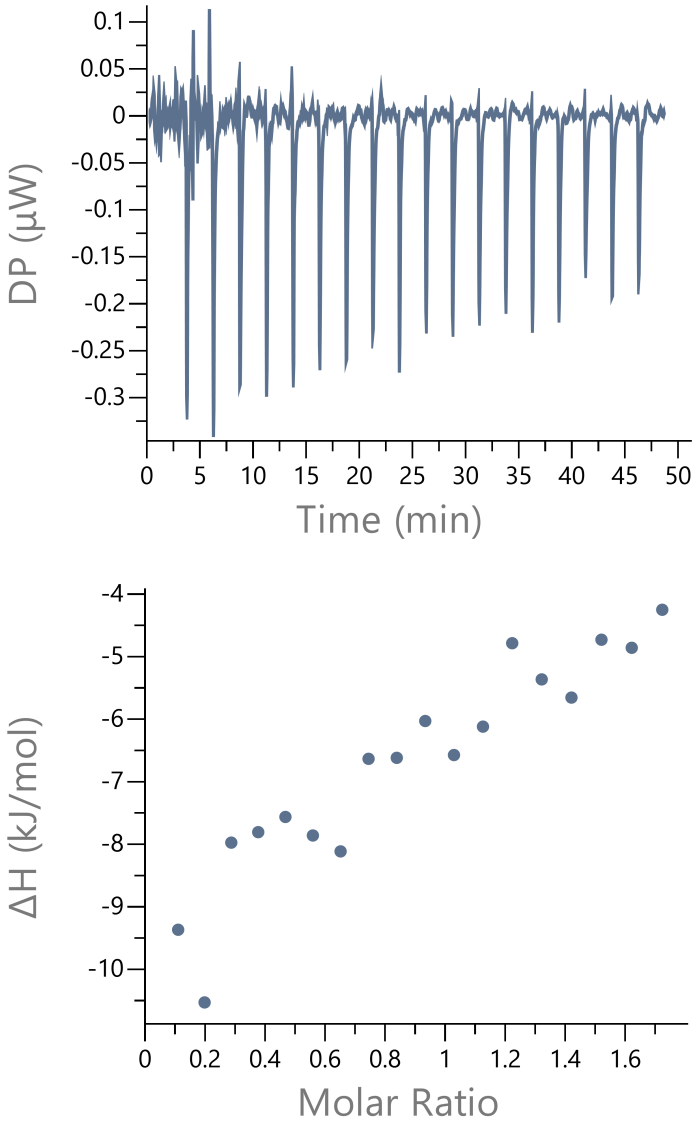


Figure 5.13: Representative ITC measurements of the binding of 23.8 μM GB1-WW2 to ~210μM OCT4 peptide at 25 °C. The top panel shows a representation of the heat differences upon injection of peptide, and the lower panel shows the integrated heats of injection, with control (OCT4 peptide titrated into buffer) experiment data subtracted using Malvern MicroCal PEAQ-ITC Analysis Software. Experiments were performed on a MicroCal PEAQ-ITC calorimeter (Malvern) at the John Innes Centre. Parameters are shown in section 2.6.2.

As the software could not estimate a K_d, the fit was simulated in Microsoft Excel using the general binding equation, with the assumption that ΔH was -10.5 at the start of the experiment and 0 at saturation point (figure 5.14). The % bound was calculated from experimental ΔH values, shown here as blue points, which was compatible with a simulated line of best fit, where K_d was 15 μM. This graph also highlights that the ligand concentration had not reached saturation and the curve was slightly more rounded than the ITC data.

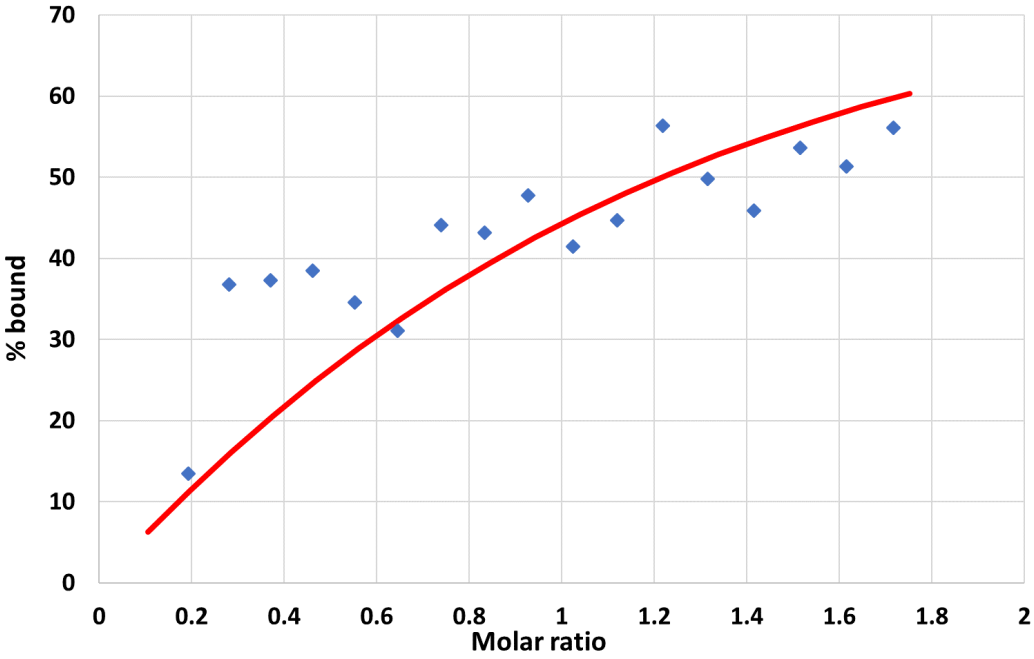


Figure 5.14: The manual K_d fit of the GB1-WW2:OCT4 ITC data, with % bound calculated from experimental ΔH values (blue points) and the simulated line of best fit (red line) with a K_d value of 15 μM.

5.3 Discussion

This chapter investigated the second WW domain of WWP2 and its interaction with a peptide from OCT4. The NMR titration and ITC results demonstrate that an interaction occurs between the second WW domain of WWP2 and a peptide of target region OCT4, an outcome which was expected following research by Xu *et al.* [Xu *et al.*, 2009]. ITC experiments by Jiang et al. had also confirmed an interaction between WW domains of WWP2 and a smaller peptide of OCT4 (T1: RRPCPPPYEFC) [Jiang *et al.*, 2015] and K_d was calculated as 53 ± 8 μM . The PPxY motif was present in this construct and interacted with WW2 at a similar affinity to the WW3 and WW4 domain [Jiang *et al.*, 2015], a characteristic not observed in the paper by Xu *et al.* [Xu *et al.*, 2009].

NMR titrations of GB1-WW2 with a synthetic peptide of OCT4 saw the total number of [^{1}H, ^{15}N]-HSQC peaks rise by 7 over the course of the titration. The GB1-WW2 construct was not fully folded in the absence of OCT4 and many of the residue peaks were not visible due to broadening caused by intermediate exchange. The fact that peaks appeared during the course of the titration implies that GB1-WW2 is in a more stable conformation in the presence of OCT4 peptide. In addition, a number of peaks were observed to move, appear and disappear as the titration progressed, suggesting an interaction between the protein and the peptide. When the trajectories of the peaks in the titration were mapped, the WW2 domain displayed a clear interaction. Several peaks were observed to shift quite significantly and perhaps just as importantly, a region of 8 consecutive peaks in the WW2 region was not present in the absence of OCT4 but appeared as the peptide was added. This indicates large changes in this region, despite the fact that they cannot be quantified. The intensity and position of peaks reached a plateau at ratio 1:4.5, confirming that saturation of GB1-WW2 with OCT4 peptide was obtained. GB1 tag residues were stationary throughout the titration confirming that the tag did not interact with the peptide. NMR titrations were used to estimate the K_d of the interaction between GB1-WW2 and OCT4 peptide. Although the fit in CCPN Analysis was not successful, manually the K_d could be approximated at ∼150 μM.

The expression and purification approach of recombinant OCT4 peptide was adapted from previous SUMO-tagged proteins and yielded in the region of 1 mg freeze dried peptide per litre of culture. The solubility of the freeze-dried protein for ITC was an

issue and resuspended samples for ITC required centrifugation to remove all precipitates as their presence is problematic in ITC experiments and can severely affect the quality of the baseline. Interestingly, the baselines of the second and third ITC runs were less stable, perhaps indicating that the OCT4 peptide was precipitating slowly over time.

Backbone assignment of 0.4 mM of ^{15}N, ^{13}C labelled GB1-WW2 and 0.7 mM of unlabelled recombinant OCT4 peptide yielded a 92% assignment, representative of full assignment of each observable residue except for 73Thr. 82% of GB1-WW2 had been assigned previously meaning a 11% increase in residue assignment. Due to the lack of assignment of GB1-WW2, a comparison between peak positions before and after saturation with OCT4 is less useful; however, when the ^{15}N, ^{13}C labelled GB1-WW2 sample saturated with OCT4 was compared to peaks acquired on a sample of GB1-WW2 in the absence of OCT4, it can be clearly observed that the majority of the GB1 tag region has a much smaller change in peak position with the addition of OCT4 peptide when compared with the WW2 region. Unfortunately, several of the residues of GB1-WW2 which were unable to be assigned were within WW2, the region of interest. There are, however, clear regions of larger changes here, in particular, 89T and 90T which change by up to 1.2 ppm.

Throughout the backbone assignment experiments, minor changes were observed with [^{1}H, ^{15}N]-HSQC backbone residue peaks (as shown in figure 5.8). The appearance of small, low intensity peaks around the centre of the spectrum, suggested minor degradation in the sample. Due to the low intensity of the degradation, and because residue peaks were clear when contours were adjusted, it was determined that the triple resonance experiments acquired could be used and were sufficient to allow assignment of the [^{1}H, ^{15}N]-HSQC backbone residues. Plans for future triple resonance experiments were cancelled as the experiments already acquired were adequate for backbone assignment and it was expected that degradation would continue in this sample.

The ^{15}N, ^{13}C saturated GB1-WW2:OCT4 [^{1}H, ^{15}N]-HSQC, which contained a mixture of recombinant and synthetic OCT4 (section 5.2.4), was overlaid with the GB1-WW2:OCT4 (recombinant) titration end point [^{1}H, ^{15}N]-HSQC (see figure 5.15). There were some differences in peak position between the two saturated samples, with an example showing 24V peaks in two different positions. The ^{15}N, ^{13}C GB1-WW2:OCT4 sample has a peak in both positions, the weaker of which overlaps

with the 24V peak from the titration spectra. The stronger of the two peaks in this spectrum was assigned as 24V. The pH of the titration sample and the saturated OCT4 sample, were 7.27 and 7.29 respectively, indicating that the changes between the samples were not caused by pH, buffer composition or experimental temperature.

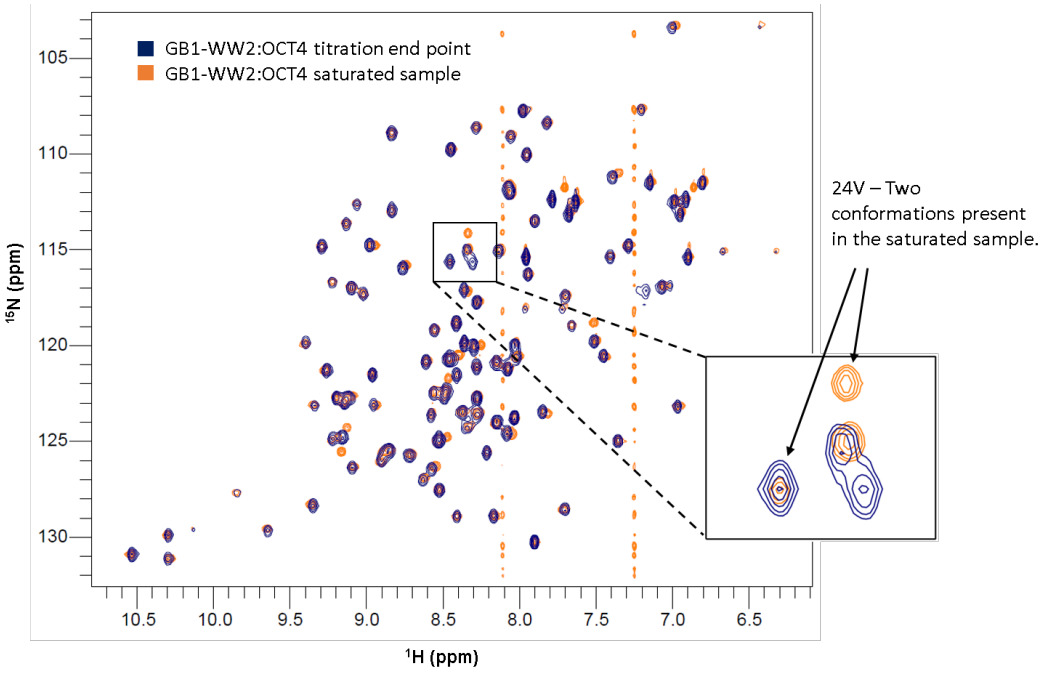


Figure 5.15: An overlay of saturated ^{15}N, ^{13}C labelled GB1-WW2:OCT4 [^{1}H, ^{15}N]-HSQC and the ^{15}N GB1-WW2:OCT4 titration endpoint [^{1}H, ^{15}N]-HSQC. A close up of the boxed region highlights that 24V is present in two conformations in the saturated sample, as opposed to the titration end point which has just one.

Comparison of GB1 peak positions in the saturated GB1-WW2:OCT4 spectrum and the GB1-WW2 in the absence of OCT4 spectrum, showed that certain residues such as 5Q, 6Y and 23A were in different positions. Initially, it appeared that the GB1 tag may have been experiencing a secondary interaction with OCT4 peptide; however, this was discounted as no changes were seen for a titration of GB1-WW2:OCT4. Instead, when the SDS-PAGE analysis of the final sample was analysed, it was found to contain a small amount of undigested His tagged protein (figure 5.10). When explored further, the GB1 residues with the largest changes in peak positions were located in the same area of the structure, as shown in figure 5.16. This figure shows the GB1 domain taken from the NMR solution structure of GB1-WW3 as determined in chapter 3.

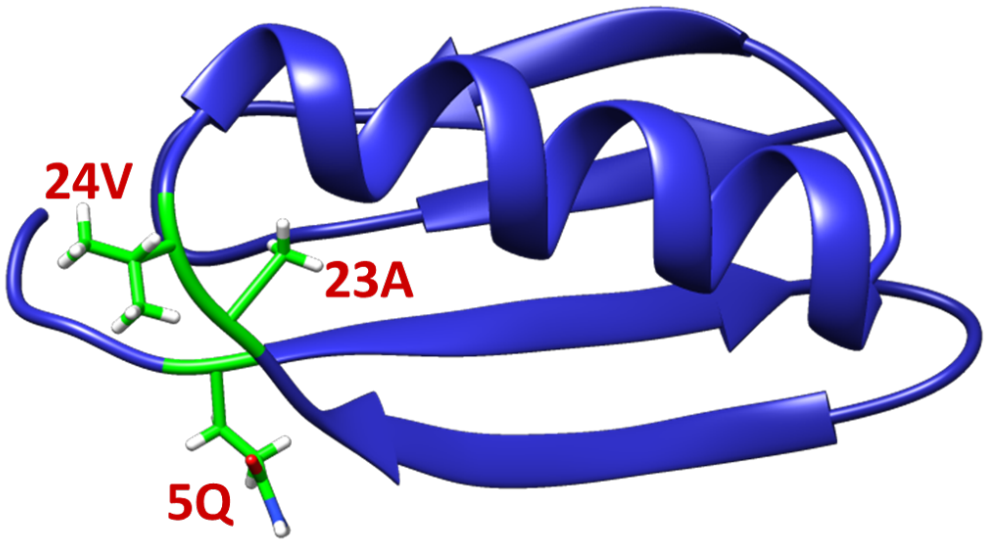


Figure 5.16: GB1 domain NMR solution structure taken from GB1-WW3 structure calculated in chapter 3. Residues which were observed to have moved position by the most (5Q, 23A and 24V) in [^{1}H, ^{15}N]-HSQC's taken before and after the addition of OCT4 peptide are highlighted in green.

The change in peak position of GB1 residues is likely due to the presence of the polyhistidine tag in the sample which could be seen by SDS-PAGE analysis of the final sample purification step (figure 5.10). The presence of the histidine region had an effect on the environment of nearby residues, hence causing them to experience a different chemical shift. As the region affected was at the opposite end of the GB1 structure to the WW2 domain, the changes in GB1 resonances were not of concern.

The results of the GB1-WW2:OCT4 ITC analysis, unfortunately, could not be used to determine K_d using the Malvern MicroCal PEAQ-ITC Analysis Software; however, an interaction was evident due to the decreasing heat exchange as more OCT4 peptide was added. This is not representative of the heat of dilution, which would be observed as regular heats of exchange for each injection. The interaction was weak and a higher concentration of OCT4 peptide would have been required to reach saturation and determine a K_d using the Malvern analysis software. This was not possible due to peptide solubility issues. Instead, a K_d value was estimated by modelling in Microsoft Excel. The resultant K_d value of 15 μM, corresponded to a reasonably low binding affinity which would agree with the ITC data, whilst extremely approximate.

Ultimately, an interaction between WW2 and OCT4 peptide was confirmed. Whilst a lot of assumptions have been made, the K_d for GB1-WW2 and OCT4 could be described as in the region of 15-150 μM.

The weak binding observed did not agree with GST pull-down data acquired by Xu et. al [Xu *et al.*, 2009], or ITC data from Jiang et al. [Jiang *et al.*, 2015]; however, saturation was not reached during ITC. If the concentration of the larger OCT4 peptide construct used here could be increased by buffer adjustment to allow saturation to occur, a more accurate K_d could be determined.

Chapter 6

Studying the interaction between GB1-WW2 and target region Smad7

6.1 Introduction

Many of the NEDD4 HECT E3 ligases, which were mentioned in Chapter 1, are involved in the regulation of the TGFβ signaling pathway [Chen and Matesic, 2007]. In particular, Smad7 has been reported as a key player in the negative regulation of activated TGF-β receptors [Ebisawa *et al.*, 2001]; [Kavsak *et al.*, 2000]. The proline-rich linker region of Smad7 interacts with WW domains of the NEDD4 E3 ligases and recruits the E3s and Smad7 to the TGFβ receptor. The HECT domain of the E3 ligase then polyubiquitinates the TGFβ receptors and Smad7, leading to their degradation and the interruption of the signaling pathway. Previous studies showed that various isoforms of E3 ligase WWP2 bind to the Smad7 in different ways, with the ability to modulate levels of activated and inhibitory Smads [Soond and Chantry, 2011]. With TGFβ stimulation, the full-length WWP2 construct interacted with various Smads; however, it was the interaction with inhibitory Smad7 which was found to be the favoured target [Soond and Chantry, 2011].

The third and fourth WW domains (WW3 and WW4) of WWP2 were found to bind Smad7 in GST pull-down assays [Soond and Chantry, 2011] and many interactions between WW domains of different NEDD4 E3 ligases (hNEDD4L, hSmurf1, hSmurf2 and hYAP) and Smad7 have been identified previously [Aragón *et al.*, 2012]. The interaction of the second WW domain of WWP2 (WW2) with a PPxY motif containing OCT4 peptide, described in chapter 5, was expected following research by Xu et al. in 2009. The interaction between WW2 and Smad7, however, had not been determined previously. In this chapter, the interaction between the second WW domain of WWP2 (WW2) and Smad7 peptide is explored.

6.1.1 Experimental aims

The aim of this chapter is to describe the interaction between the WW2 domain of WWP2 and a PPxY motif-containing peptide from the linker region of protein Smad7. NMR titrations of the recombinant GB1-WW2 and recombinant Smad7 peptide allowed the interaction between the two regions to be observed as the concentration of Smad7 increases. Triple resonance NMR experiments were acquired using a sample of ^{15}N, ^{13}C labelled GB1-WW2, fully saturated with unlabelled Smad7, in an attempt to fully assign the WW2 domain, including residues which were previously in intermediate exchange. Isothermal titration calorimetry was then used to quantify an interaction between the protein and peptide.

6.2 Results

6.2.1 ^{15}N labelled GB1-WW2 domain synthesis and purification

^{15}N labelled His-GB1-WW2 sample was expressed as described in the previous chapter, induced with 0.8 mM IPTG with overnight incubation at 30°C. Figure 6.1 shows a 15% SDS-PAGE gel of ^{15}N labelled GB1-WW2 samples, pre and post-induction. The band at $\sim$ 16 kDa is GB1-WW2 as it is known that this protein

runs at a higher molecular weight on SDS-PAGE gels then its actual molecular weight of 13.3 kDa.

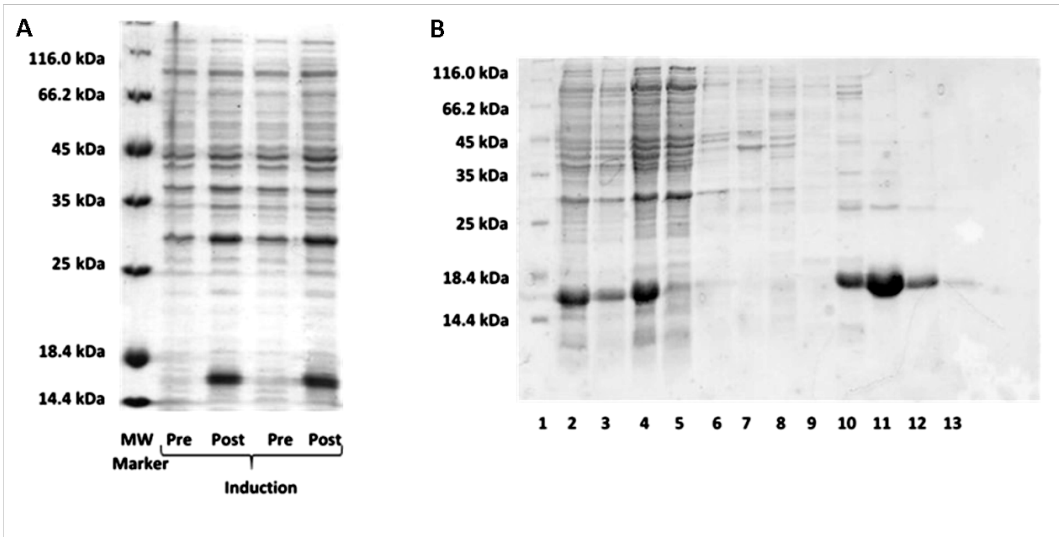


Figure 6.1: (**A**) 15% SDS-PAGE gel showing induction of ^{15}N labelled GB1-WW2. Samples were taken before and after induction with IPTG. 10 μL samples loaded. (**B**) 15% SDS-PAGE gel of ^{15}N GB1-WW2 purification by Ni-NTA, showing clear bands in E1, E2 and E3 (lanes 10-12). 10 μL samples loaded. Lanes: 1 = MW marker, 2 = pre sonication 3 = lysate, 4 = supernatant, 5 = flow through, 6 = wash A, 7 = wash B, 8 = wash C, 9 = wash D, 10 = first elution wash (E1), 11 = second elution wash (E2), 12 = third elution wash (E3) and 13 = fourth elution wash (E4). Imidazole wash concentrations: A = 1 mM, B = 10 mM, C = 20 mM, D = 30 mM, E = 250 mM.

The culture was then harvested, the protein was purified as described in section 2.3.2, using a Ni-NTA column and the samples were analysed using SDS-PAGE analysis. Figure 6.1 also shows the samples throughout the purification. Washes A, B, C and D contain 1 mM, 10 mM, 20 mM and 30 mM imidazole respectively. The Elution wash (E) contains 250 mM imidazole. Samples E1, E2 and E3 were combined and dialysed into NMR buffer. The absorbance of the 21 mL combined sample at 280 nm was A = 1.326 and the sample concentration was calculated to be 0.049 mM.

The sample was concentrated to 1 mL using a 5000 MWCO spin concentrator and the concentration was calculated using absorption at 280 nm to be 0.5 mM. A [^{1}H, ^{15}N]-HSQC of ^{1}H, ^{15}N His-GB1-WW2, at 500 MHz, indicated that the protein was well folded and could be used for a titration with a sample of lyophilised recombinant Smad7 peptide which was prepared by a previous member of the group.

6.2.2 NMR titrations of ^{15}N, His-GB1-WW2 with recombinant Smad7 peptide

NMR samples for the titration of recombinant Smad7 peptide into GB1-WW2 were prepared as described in section 2.5.8.3. [^{1}H, ^{15}N]-HSQC spectra were acquired for both the starting titration point sample and the end titration point sample. Aliquots of the 1:0 molar ratio starting titration point sample were then removed and replaced with aliquots of the end titration point sample at molar ratio 1:10. Therefore, the concentration of Smad7 was increasing but the ^{15}N His-GB1-WW2 concentration stayed the same. Between each addition, the NMR tube was inverted 5 times and then spun in a hand centrifuge to ensure all of the sample accumulated at the bottom. A superimposition of the ten [^{1}H, ^{15}N]-HSQC spectra, ranging from molar ratio 1:0 His-GB1-WW2:Smad7, up to 1:10, can be seen in figure 6.2. Table 2.12 in section 2.5.8.3 shows Smad7 concentrations at each titration point.

The chemical shift of each NH peak in the [^{1}H, ^{15}N]-HSQC is dependent on the environment it is in and so when other atoms are involved, the environment can change, hence altering the chemical shift for that peak. The residues which are most affected by the addition of a target region will experience the biggest changes to their local environment and hence display substantial changes in their chemical shift. The high excess of peptide used is to ensure that the interaction is fully saturated and so the bound state is observed.

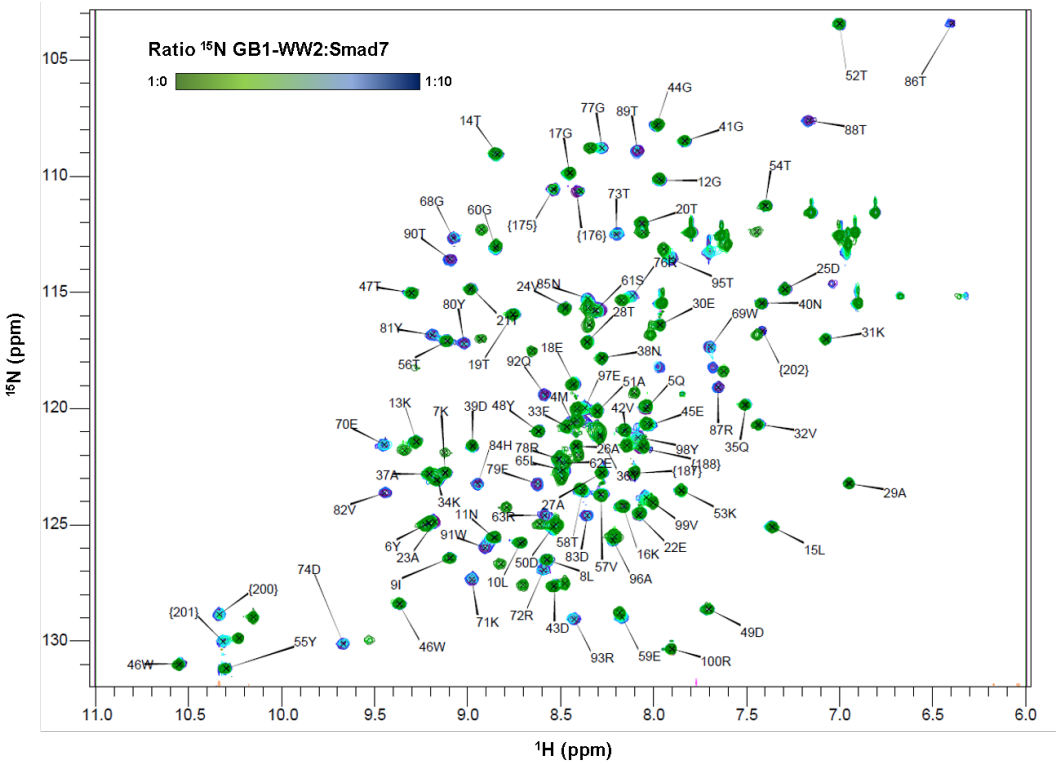


Figure 6.2: Overlaid [^{1}H, ^{15}N]-HSQC spectra of His-GB1-WW2 titrated with unlabelled Smad7 peptide. GB1 tag peaks are generally stationary. Some of the WW2 domain peaks disappear as the titration progresses; however, more peaks are observed when saturated with Smad7 peptide. Assignment labels show the final titration point and are described later in the chapter. The experiments were performed using a Bruker Avance I 500 MHz spectrometer at 298 K. Sample buffer: 20 mM Sodium phosphate buffer, 50 mM NaCl, 15 mM DTT, pH 6.8. The concentration of GB1-WW2 was 0.16 mM.

Peak positions started to change slightly as the molar ratio reached 1:0.5 and final peak positions were reached by ratio 1:2. Intensities of some peaks continued to increase slightly until ratio 1:6, at which point maximum intensity was achieved. Two further titration points, 1:8 and 1:10 were also acquired to confirm saturation. Peak assignments were propagated from a combination of His-GB1-WW2 assignments (determined in section 4.2.5) and GB1-WW2:Smad7 assignment, described later in this chapter (section 6.2.3.4).

As the concentration of Smad7 increases throughout this titration, several of the peaks appear and disappear. The majority of the GB1 tag residue peaks remained stationary; however, some of the assigned WW2 domain peaks disappear as the titration

progresses. In addition to this, a higher number of peaks were present at the end of the titration than were visible at the beginning. This indicates that the His-GB1-WW2 protein, which was previously in intermediate exchange, is gaining stability to form a more stable structure when in its bound complex.

Figure 6.3 focuses on four areas of the [^{1}H, ^{15}N]-HSQC's throughout the titration. In the first detailed region, **A**, WW2 Trp residue side chain peaks {200} and {201} are in slow exchange and so disappear and reappear elsewhere in the spectrum as the titration progresses. In this particular case, it is clear to see where the peaks are migrating to. In other regions where peaks disappear but do not reappear in the local vicinity, it can be much harder to follow the peak movement. The 46W side chain peak and residue 55T are not observed to move throughout the titration indicating that these GB1 residues are not affected by the presence of Smad7.

In figure 6.3**B**, residue 17G, in the GB1 tag region, is not affected, but as the titration progresses, two new peaks appear. Previously, in chapter 4, GB1-WW2 was observed to be in an intermediate exchange regime and hence a number of peaks were missing from the [^{1}H, ^{15}N]-HSQC spectrum. These could be peaks which disappeared in a different part of the spectrum or they could be additional peaks which have appeared as the His-GB1-WW2 is coming out of intermediate exchange. These new peaks will therefore not have been assigned previously and so assignments could not be propagated. Figure 6.3**C** shows that 5Q is not affected by the addition of Smad7, but that the peak above (90T) disappears as the titration progresses.

Finally, in figure 6.3**D**, 20T is not affected and an unassigned peak under the 20T residue disappears. 95T moves in intermediate exchange which is observed as slight migration as Smad7 is added. 73T appears as the titration progresses.

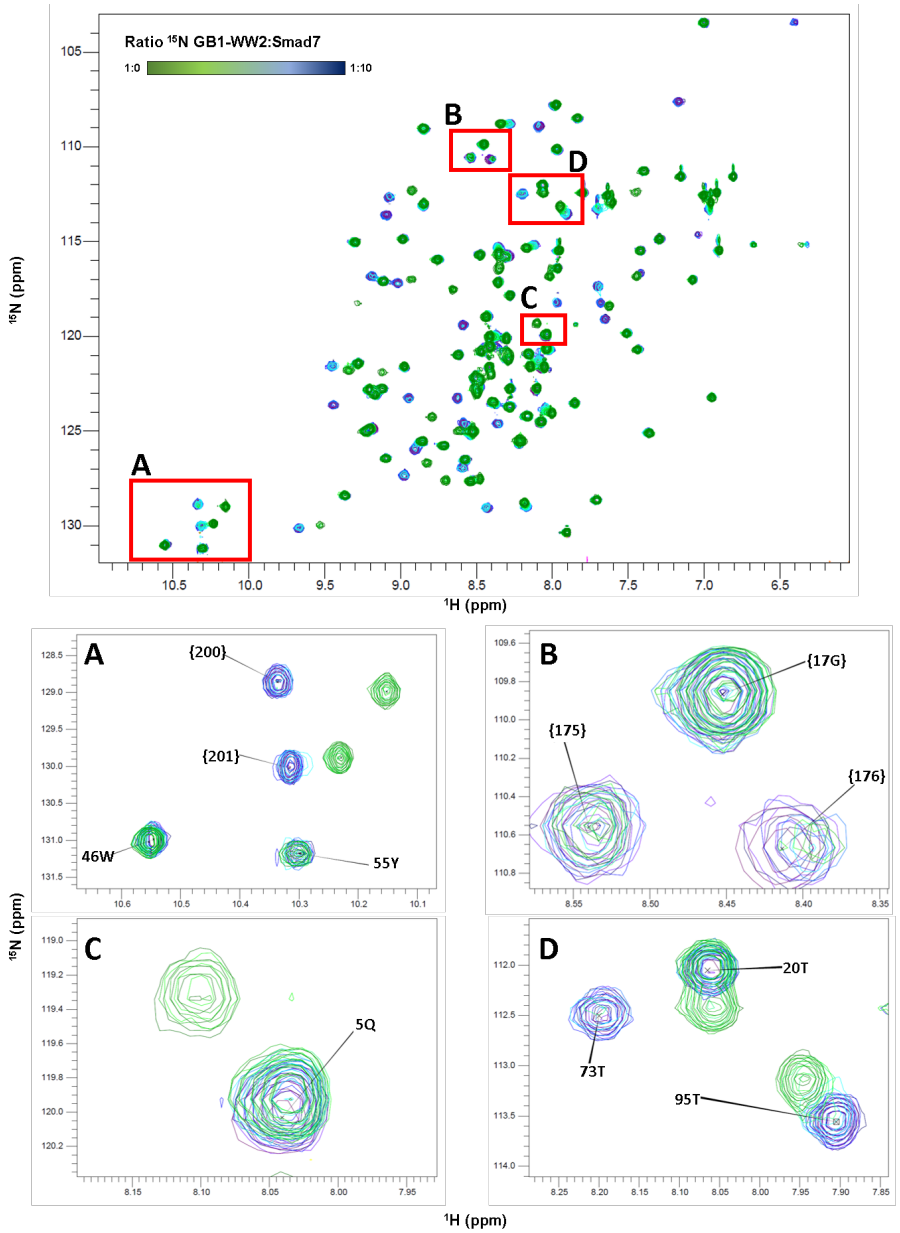


Figure 6.3: Detailed regions from GB1-WW2:Smad7 [^{1}H, ^{15}N]-HSQC titration spectra highlighting regions of appearing, disappearing and migrating peaks. Generally, GB1 peaks remained stationary throughout the titration. Assignments were propagated from other GB1-WW2 assignments. (**A**) shows WW2 Trp residue side chain peaks {200} and {201} disappearing and reappearing elsewhere in the spectrum as the titration progresses. GB1 peaks; 46W side chain peak and 55T, are stationary. (**B**) shows GB1 tag residue 17G is also not affected, and two peaks appeared after the addition of Smad7. (**C**), shows that GB1 peak 5Q is not affected by the addition of Smad7, but that the unassigned peak above (90T) disappears as the titration progresses. (**D**) shows 20T is not affected but an unassigned peak just below it disappears. 95T migrates slowly as it is in intermediate exchange and residue 73T is seen to appear as the titration progresses. The experiments were performed using a Bruker Avance I 500 MHz spectrometer at 298 K. Sample buffer: 20 mM Sodium phosphate, 50 mM NaCl, 15 mM DTT, pH 6.8. The concentration of GB1-WW2 was 0.16 mM.

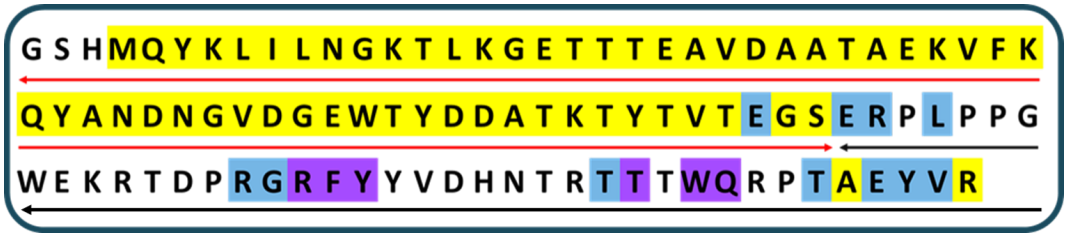


Figure 6.4: The GB1-WW2 sequence, with GB1 indicated with a red arrow and WW2 a black arrow. Residues highlighted in yellow were stationary, blue residues show peaks which migrate, and purple residues had peaks which disappeared. Unassigned residues have not been highlighted.

The GB1-WW2 sequence was illustrated in an assignment graph, shown in figure 6.4, with GB1 indicated with red arrows and WW2 in black. Residues highlighted in yellow were stationary throughout the titration, blue residues show peaks which migrate throughout the course of the titration and purple residues had peaks which disappeared as more Smad7 peptide was added. Residues which have not been highlighted could not be assigned. As expected, the vast majority of the GB1 domain remains stationary which implies that the GB1 tag is not interfering with the interaction between the WW2 domain and Smad7. The WW2 domain had peaks which moved and also peaks which disappeared as more Smad7 was added. Throughout the titration, 1 unassigned peak was also observed disappearing, 2 moved, and 22 peaks appeared. The fact that more peaks appeared could indicate that WW2 bound to Smad7 is not in intermediate exchange and can, therefore, be fully assigned.

Many of the WW2 domain amide peaks were observed to move position during the titration with Smad7 peptide in figure 6.5, shown earlier in this chapter. Disappearing residues are likely reappearing elsewhere in the spectrum due to a slow chemical exchange rate. Shift trajectories upon addition of Smad7 peptide were plotted against residue number and displayed in the graph in figure 6.11. Shift changes were evenly weighted due to the difference in the gyromagnetic ratio between hydrogen and nitrogen (described in section 3.2.6). Peak positions of GB1 peaks (1-61) were generally stable except for a few residues which show small shifts. 23Ala, 4Met and 6Tyr all show slight changes in peak position. As shown in chapter 4, the presence of the polyhistidine tag was able to affect the structure of GB1-WW2.

As this titration was acquired using the GB1-WW2 construct still containing the polyhistidine tag, it is likely that the tag is interfering with these residues. As mentioned in the discussion section of chapter 5, the residues affected are all at the N-terminus of

the GB1 structure and are located near to the His-tag region. The WW domain region (62 and above) contains slightly higher trajectories than the GB1 region. Perhaps of more interest, however, is the appearance of a considerable number of additional peaks. These peaks were previously experiencing intermediate exchange and due to stabilisation of the structure, were able to be assigned.

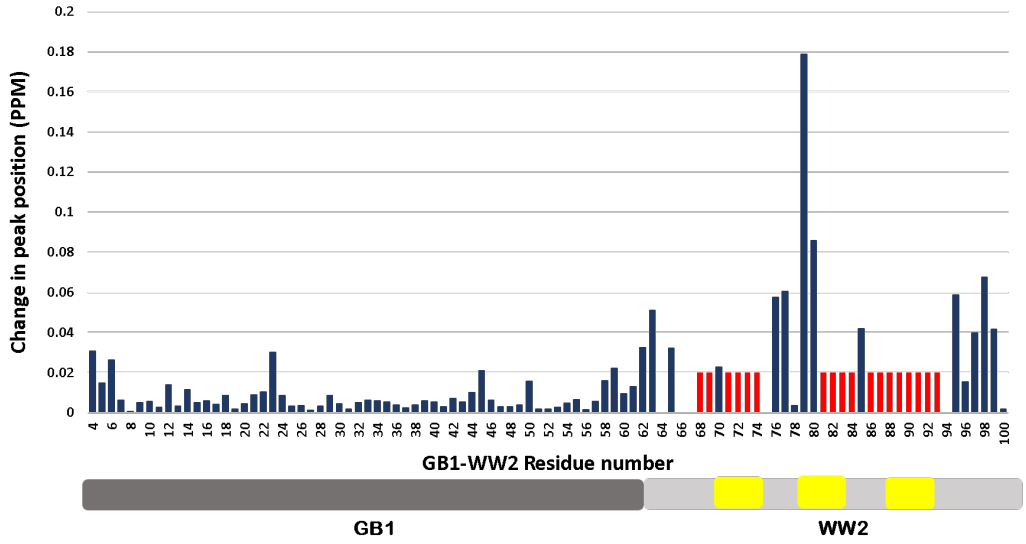


Figure 6.5: The trajectories (in ppm) of all His-GB1-WW2 backbone amide peaks throughout the [^{1}H, ^{15}N]-HSQC titration with Smad7 peptide. GB1 peaks (1-61) are relatively stable throughout the titration, except for a few residues which migrate slightly (23Ala , 4Met and 6Tyr). The WW domain residues (62 and above) generally show larger trajectories. Residues which were only present after the addition of Smad7 peptide are marked in red at 0.02 ppm for comparison. Underneath the graph is a schematic diagram showing the GB1 region in dark grey and the WW2 domain in light grey. Secondary structure prediction using the PSIPRED webserver v3.3 predicted areas of secondary structure [Jones *et al.*, 2018]; [Buchan *et al.*, 2013] and the three yellow regions represent predicted β strands.

A double labelled sample of ^{15}N, ^{13}C, GB1-WW2 with an excess of unlabelled Smad7 peptide was prepared and used for triple resonance experiments to assign the backbone.

6.2.3 NMR triple resonance assignment

6.2.3.1 ^{15}N, ^{13}C, GB1-WW2 expression and purification

A 2 L double-labelled sample was prepared for triple resonance NMR experiments for the assignment of the GB1-WW2 backbone when in the presence of Smad7. GB1-WW2 was expressed as above, using ^{15}N, ^{13}C Minimal Essential Media (MEM). The sample was purified using a Ni-NTA column, the hexahistidine tag was digested and the sample was re-purified.

Figure 6.6 shows the pre and post-induction samples and purification using a Ni-NTA column. Washes A, B, C and D contain 1 mM, 10 mM, 20 mM and 30 mM imidazole respectively. The Elution (E) contains 250 mM imidazole. The majority of the protein could be clearly seen in E1 and E2 (lanes 9 and 10) and so these samples were combined and dialysed into Tris-Cl buffer with 3 changes. The Elution sample E1 shows contaminants were present; however, this did not prove to be a problem during NMR experiments due to the high concentration of ^{15}N, ^{13}C GB1-WW2 protein.

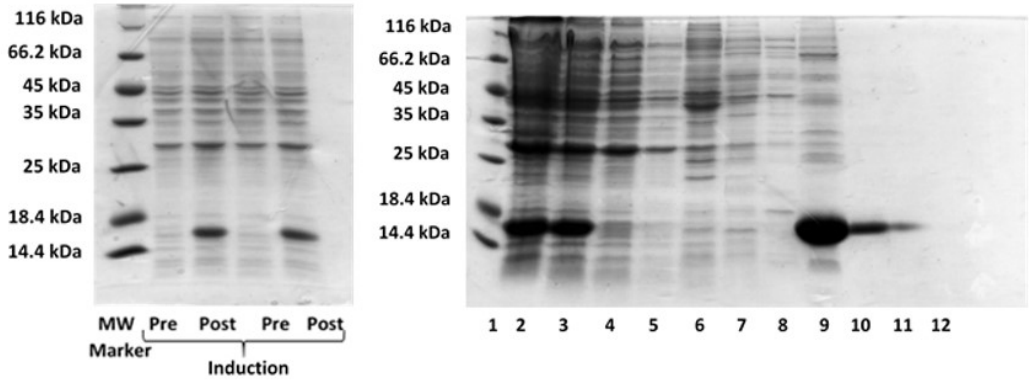


Figure 6.6: (**Left**) 15% SDS-PAGE gel showing induction of ^{15}N, ^{13}C GB1-WW2. Samples were taken before and after induction with IPTG. 10 μL samples loaded. (**Right**) 15% SDS-PAGE gel of ^{15}N, ^{13}C GB1-WW2 purification using a Ni-NTA column. The majority of the protein is in E1 (lane 9); however, there is also some present in E2 (lane 10). 10 μL samples loaded.

60 μL thrombin was added to the sample to digest the hexahistidine tag and samples were analysed using SDS-PAGE analysis to determine when digestion was complete (figure 6.7 **Left**). A benzamidine purification column was used to remove the thrombin from the sample and a Ni-NTA column was used to remove the digested His-tag (figure

6.7 **Right**). The thrombin binds to the benzamidine column until the elution, whereas the sample is collected during the washes. The digested His-tag binds to the Ni-NTA column and is eluted in the high imidazole elution wash. Some of the GB1-WW2 protein was still present in both the Benzamidine elution and the Ni-NTA elution. This could be due to non-specific binding to the column. Fractions FT, A1, A2, A3 and A4 were dialysed into NMR buffer. The two elution samples were dialysed into Tris-Cl buffer and then purified again using benzamidine and Ni-NTA columns (see figure 6.8). The purified ^{15}N, ^{13}C labelled GB1-WW2 was then dialysed into NMR buffer.

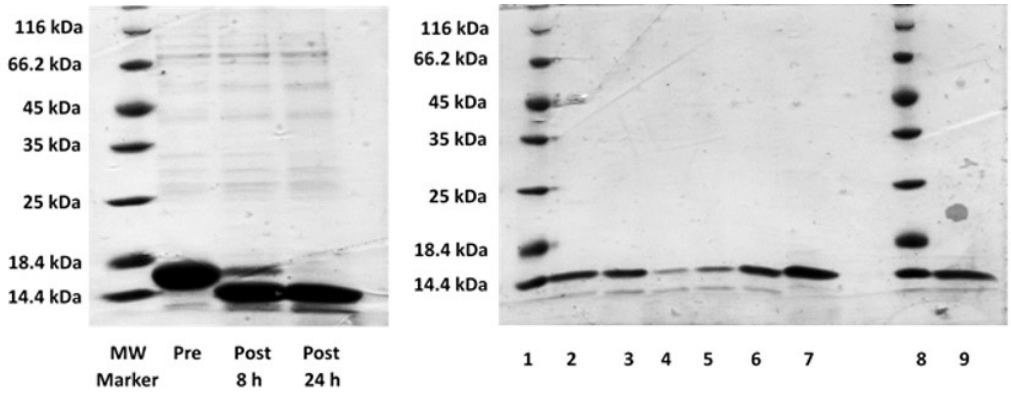


Figure 6.7: (**Left**) 15% SDS-PAGE gel of thrombin digestion of ^{15}N, ^{13}C GB1-WW2. Samples are pre digestion, after 8 hours and after 24 hours the sample appears to have fully digested. 10 μL sample loaded. (**Right**) 15% SDS-PAGE gel of ^{15}N, ^{13}C GB1-WW2 purification using a benzamidine column for thrombin removal and Ni-NTA to remove the His-tag. 10 μL sample loaded. Lanes: 1 = MW marker, 2 = flow through, 3 = wash A1, 4 = wash A2, 5 = wash A3, 6 = wash A4, 7 = elution, 8 = MW marker, 9 = Benzamidine elution.

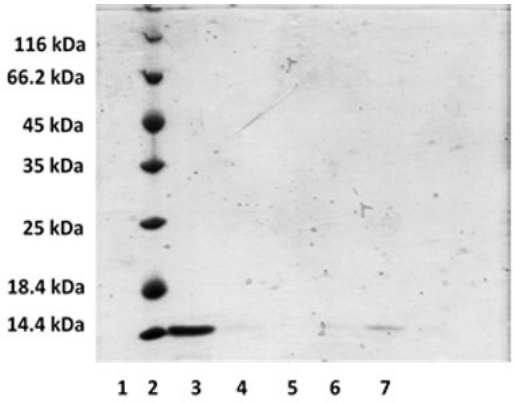


Figure 6.8: 15% SDS-PAGE gel of repurification of elution samples (E and E Benz.) from ^{15}N, ^{13}C GB1-WW2 purification. 10 μL samples loaded. Lanes: 1 = Benzamidine elution, 2 = MW Marker, 3 = FT, 4 = wash A1, 5 = wash A2, 6 = wash A3, 7 = elution.

6.2.3.2 Expression and purification of SUMO-Smad7

The SUMO-Smad7 peptide construct, designed by a previous member of the group, was transformed in BL21(DE3)* *E. coli* and expressed in 6 L of LB using 0.8 mM IPTG at 30°C overnight (as described in section 2.3) to induce expression as is shown on the left in figure 6.9. The protein expressed well and gave high yields after purification which can be seen on the right of figure 6.9. The digestion was also successful as can be seen by the reduction in molecular weight of the post-digestion samples, but the peptide could not be easily visualised due to its small size binding less Coomassie brilliant blue (figure 2.3 Bottom).

After the cleaved SUMO tag was removed using a Ni-NTA column, the peptide was concentrated by dialysing into ammonium bicarbonate, using 0.5-1 kDa MWCO dialysis tubing and then lyophilised.

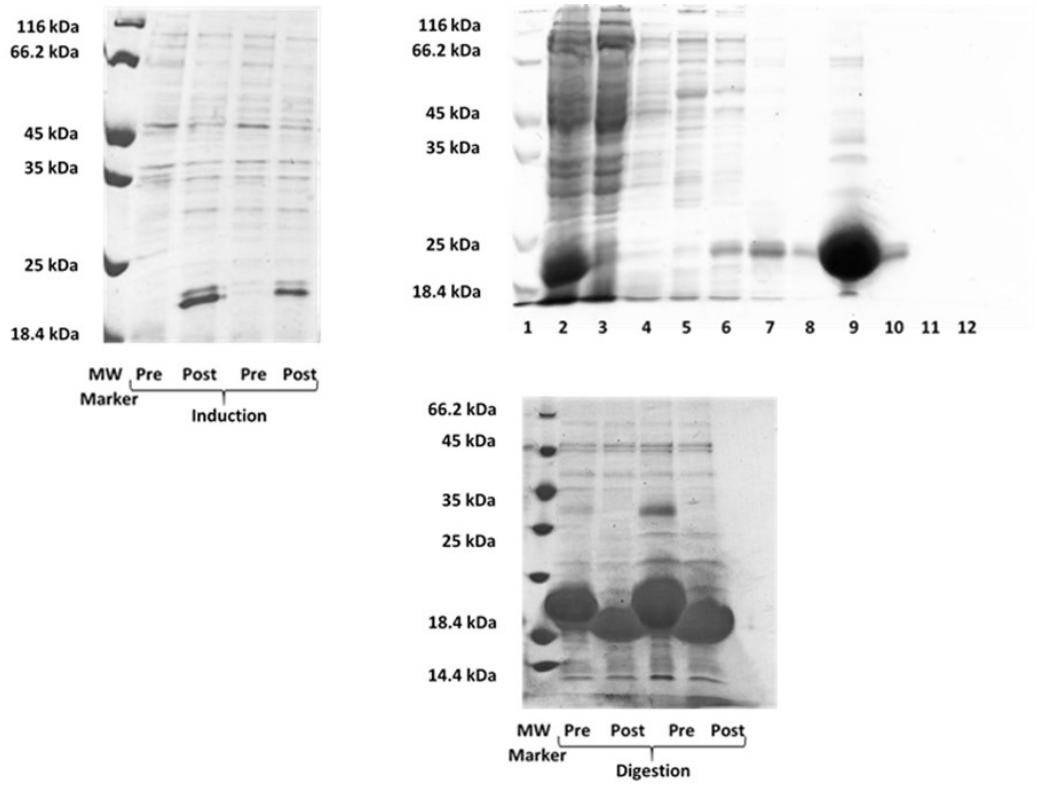


Figure 6.9: (**Left**) 15% SDS-PAGE gel of SUMO-Smad7 pre and post-induction. 10 μL sample loaded. (**Right**) 15% SDS-PAGE gel of SUMO-Smad7 purification by Ni-NTA. 10 μL sample loaded. Lanes: 1 = MW Marker, 2 = supernatant, 3 = flow through, 4 = wash A, 5 = wash B, 6 = wash C, 7 = wash D, 8 = wash D3, 9 = first elution (E1), 10 = second elution (E2), 11 = third elution (E3), 12 = fourth elution (E4). 10 μL sample loaded. (**Bottom**) 15% SDS-PAGE gel of SUMO-Smad7 pre and post-digestion with ULP1 SUMO protease. 10 μL sample loaded.

6.2.3.3 GB1-WW2:Smad7 NMR

The data acquired during the titration of ^{15}N labelled His-Gb1-WW2 and Smad7 showed that the interaction was saturated at a ratio of 1:3.49 or [His-GB1-WW2]:[Smad7] = 0.16 mM:0.558 mM. Due to limitations in peptide solubility, concentrating the peptide to very high concentrations was avoided. Instead, two 500 μL samples were prepared, one of ^{15}N labelled His-GB1-WW2 and one of unlabelled Smad7 peptide, each at double the required concentration so they could be combined to give the correct molar ratio. To determine the molar ratio required for saturation the following equation 6.1 was used.

$$K_{\mathrm{d}} = \frac{[WW2][Smad7]}{[WW2.Smad7]} = \frac{[P][L]}{[PL]} \tag{6.1}$$

Where P is unbound protein, L is unbound ligand and PL is bound protein and ligand.

The 45 mL sample of ^{15}N, ^{13}C GB1-WW2, was at a concentration of 2.488 x10^{-5} M and was concentrated to twice the required concentration. Optimum Smad7 peptide concentration was calculated to be 3.49 mM and so a sample containing twice the concentration required was prepared using 4.57 mg of freeze-dried peptide, dissolved in 500 μL NMR buffer. 250 μL of ^{15}N, ^{13}C labelled GB1-WW2 sample was added to 250 μL of the Smad7 sample to give a 500 μL sample at the correct ratio.

6.2.3.4 GB1-WW2:Smad7 resonance assignment

A [^{1}H, ^{15}N]-HSQC and HNCACB, CBCACONH, HNCO and HNCACO experiments were acquired. These experiments were used to set sequential links between residues so that each of the resonances can be linked to the previous residue and finally all the residues can be sequentially assigned using characteristic ppm values for each amino acid. Initially, the 2D [^{1}H, ^{15}N]-HSQC was compared to a partially assigned spectrum of GB1-WW2, as shown in the previous chapter, and assignments for the GB1 tag and a number of WW2 domain assignments could be propagated onto the 2D [^{1}H, ^{15}N]-HSQC of GB1-WW2:Smad7. The HNCO and HNCACO spectra were also used in a similar way to set sequential links using the CO groups. CCPN Analysis was used for the assignment of ^{15}N, ^{13}C labelled GB1-WW2 in the presence of Smad7 peptide.

92 of 100 residues were assigned in the 2D [^{1}H, ^{15}N]-HSQC of GB1-WW2:Smad7 (figure 6.10). A full list of the assigned residues and resonances is included in Appendix D. Five of the residues in the WW2 region are not present in the HSQC as they are prolines and do not have an NH bond. Three N-terminal residues are also not assigned. It is common that the first few residues of a protein do not show visible peaks due to proton exchange. Disregarding peaks which are not visible in the [^{1}H, ^{15}N]-HSQC, or expected to be visible, the remaining protein backbone of ^{15}N, ^{13}C GB1-WW2 in the presence of unlabelled Smad7 was fully assigned.

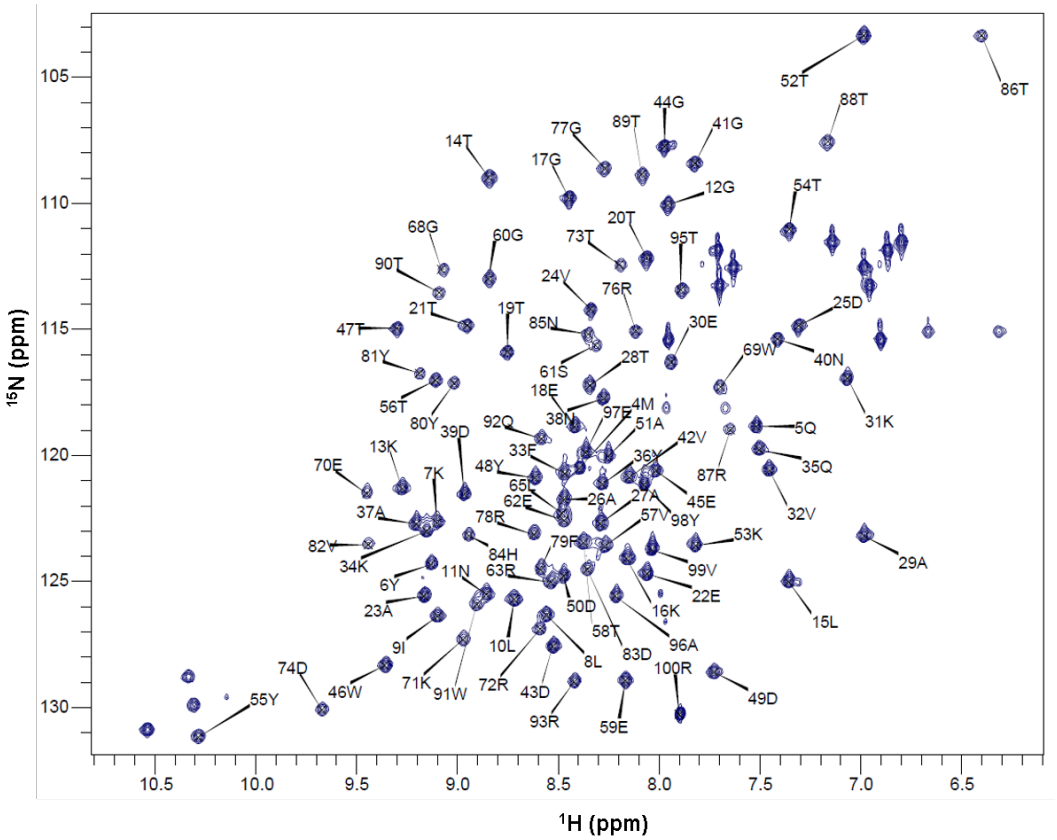


Figure 6.10: [^{1}H, ^{15}N]-HSQC spectrum of 0.5 mM ^{15}N, ^{13}C labelled GB1-WW2 in the presence of 1.75 mM unlabelled Smad7 peptide. 92% of residues were assigned and labels are present for these assignments. Side chain peak assignments are not shown. This experiment was performed using a Bruker Avance III 800 MHz spectrometer at 298 K. Sample buffer: 20 mM sodium phosphate, 50 mM NaCl, 15 mM DTT, pH 6.8.

6.2.3.5 Effect of Smad7 peptide on residue peak position

Many of the WW2 domain amide peaks were observed to move position after the saturation with Smad7 peptide in figure 6.2, shown earlier in this chapter. Disappearing residues are likely reappearing elsewhere in the spectrum due to a slow chemical exchange rate. Changes in shift position after the addition of Smad7 peptide were plotted against residue number and displayed in the graph in figure 6.11.

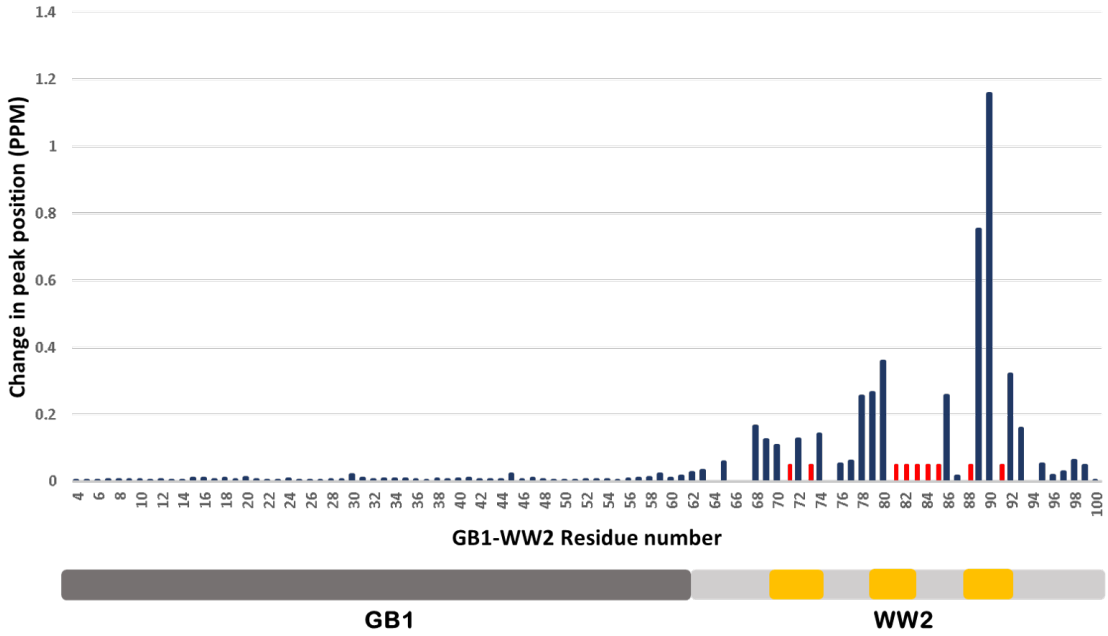


Figure 6.11: The change is peak position (in ppm) of all GB1-WW2 backbone amide peaks which are assigned in both the GB1-WW2 [^{1}H, ^{15}N]-HSQC and the GB1-WW2:Smad7 [^{1}H, ^{15}N]-HSQC after titration of the Smad7 ligand. GB1 peaks (1-61) are stable, as opposed to the WW domain residues (62 and above), which show larger trajectories. Residues which were only present after the addition of Smad7 peptide are marked in red at 0.05 ppm for comparison. Underneath the graph is a schematic diagram showing the GB1 region in grey and the WW2 domain in light grey. Secondary structure prediction using the PSIPRED webserver v3.3 [Jones *et al.*, 2018] predicted areas of secondary structure [Buchan *et al.*, 2013] and the three yellow regions represent predicted β strands.

Shift changes were evenly weighted as previously described in section 3.2.6. Peak positions of GB1 peaks (1-61) were stable, indicating Smad7 did not interact with the tag. This was expected as the histidine tag, which had previously caused problems with intermediate exchange (chapter 4), was removed prior to the acquisition of NMR experiments in both the GB1-WW2 and the GB1-WW2:Smad7 peptide samples. The WW domain residues (62 and above) generally show much larger trajectories than the GB1 tag.

A number of WW2 residues were missing from the GB1-WW2 assignment, and so cannot be compared to the fully assigned GB1-WW2:Smad7 assignments. The GB1-WW2 structure is more stable in the presence of Smad7 peptide and is no longer in intermediate exchange. The PSIPRED webserver v3.3 was used to predict areas of secondary structure [Jones *et al.*, 2018]; [Buchan *et al.*, 2013] and figure 6.11 shows

predicted β sheet regions in yellow on the schematic diagram. Region 70-73 and 79-82 show an increased change in peak position when compared to the GB1 domain, but residue data in these regions are missing. The third region 88-91, contains the two largest trajectories observed, suggesting that this region is involved in the binding of Smad7.

6.2.4 Isothermal Titration Calorimetry experiments

Isothermal titration calorimetry, described in section 1.7, was used to determine the K_d of the interaction between GB1-WW2 and Smad7 peptide. The experiment was designed using Malvern MicroCal PEAQ-ITC Analysis Software and the binding parameters were estimated as 1 binding site and $K_d = 1$ μM. The concentration of GB1-WW2 was calculated using the absorbance at 280 nm to be 30 μM and so Smad7 was resuspended, aiming for a 300 μM sample, as suggested by the software to give an ideal binding curve. The final concentrations of GB1-WW2 and Smad7 were 30.2 mM and 336 mM respectively and both were dialysed into the same batches of Tris-Cl buffer three times before concentrations were finely adjusted.

Three GB1-WW2:OCT4 ITC titrations were carried out, along with three control experiments to determine the heat of dilution of Smad7, GB1-WW2 and to ensure that the buffers matched. A Smad7 into buffer control and buffer into GB1-WW2 control were used to determine the heat of dilution of the two ITC components and a buffer into buffer control was also used to verify matching buffers. The heat of dilution of GB1-WW2 was minimal and so could be ignored, whereas the heat of dilution of Smad7 (figure 6.12) was significant and so was subtracted from the GB1-WW2:Smad7 titration data before analysis. The slight baseline drift is likely due to small fluctuations in temperature in the room and so was not a concern. It is far more important that the pre and post-injection baseline is steady, which in this case it is, so that the Malvern MicroCal PEAQ-ITC Analysis Software is able to integrate the peaks accurately. The DP is the recorded power differential between the reference and sample cells to maintain a zero temperature difference between the cells.

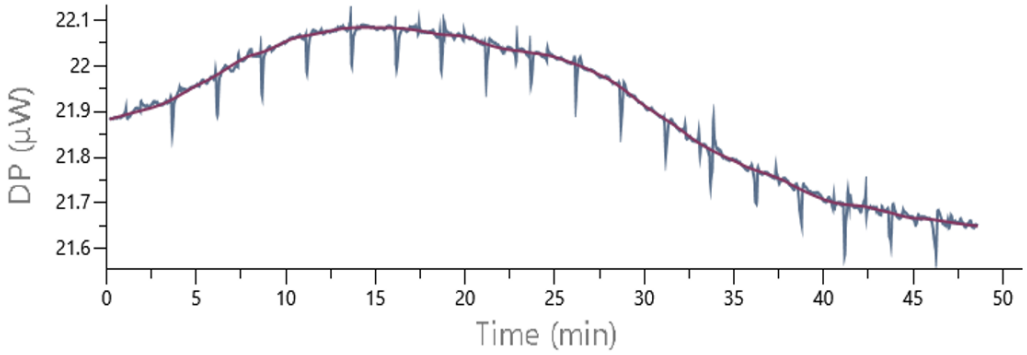


Figure 6.12: Control ITC experiment showing heat of dilution for Smad7 peptide titrated into buffer. The mean of the control dataset was subtracted from GB1-WW2:Smad7 titrations using Malvern MicroCal PEAQ-ITC Analysis Software.

A representative raw plot of 30.2 μM GB1-WW2 titrated with 336 μM Smad7 peptide at 25 °C is shown in figure 6.13, and below an image of the integrated heat change. Injection points are clearly recognised by a sharp drop in DP, the area of which represents the heat exchanged. Here, saturation was achieved after around 34 minutes and continued to produce similarly sized peaks until the end of the experiment at 50 minutes.

Malvern MicroCal PEAQ-ITC Analysis Software was used to calculate the K_d after subtracting the Smad7 peptide into buffer control. K_d was calculated to be 2.28 $\pm$ 0.18 μM. The first injection is only 0.4 μL which reduces the inaccuracies experienced during temperature equilibration as the ligand diffuses out of the needle into the cell. The remaining injections are each 2 μL. At the beginning of the titration, there is a large excess of potential binding sites which are unpopulated, which leads to larger heats as it is likely that all Smad7 peptide interacts with the GB1-WW2. As the titration proceeds, there are less potential binding sites free and so whilst the same amount of Smad7 peptide is added, less will be able to bind.

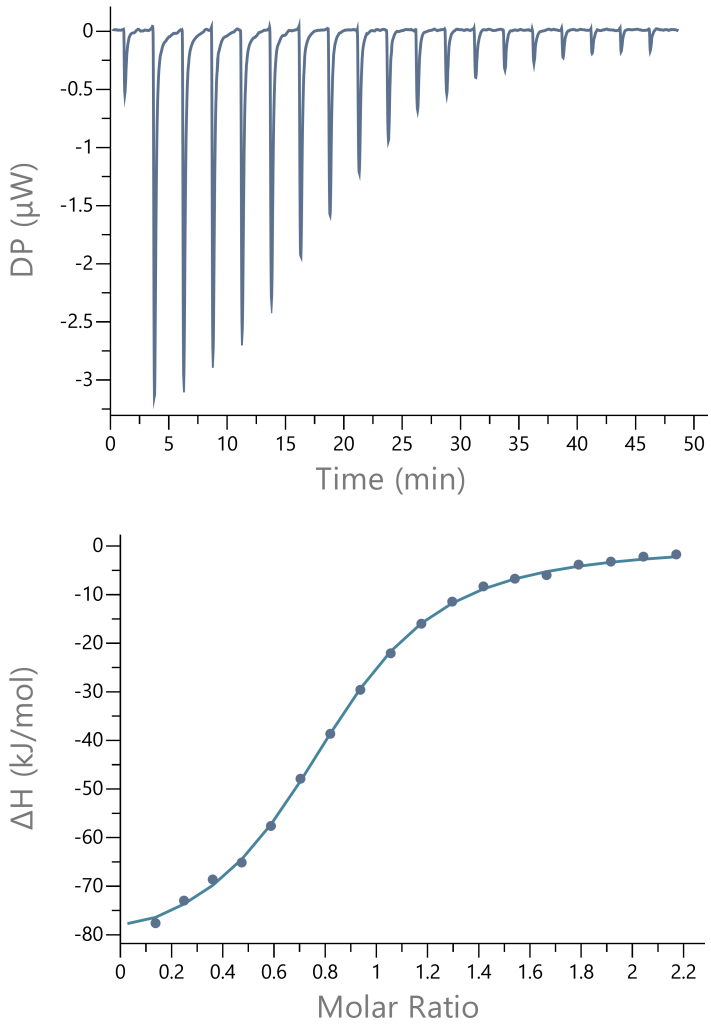


Figure 6.13: Representative ITC measurements of the binding of 30.2 μM GB1-WW2 to 336μM Smad7 peptide at 25 °C. The top panel shows a representation of the heat differences upon injection of peptide, and the lower panel shows the integrated heats of injection, with control (Smad7 peptide titrated into buffer) experiment data subtracted using Malvern MicroCal PEAQ-ITC Analysis Software. Experiments were performed on a MicroCal PEAQ-ITC calorimeter (Malvern) at the John Innes Centre. Parameters are shown in section 2.6.2.

Thermodynamic parameters obtained from three GB1-WW2:Smad7 ITC experiments, along with the standard deviation errors, are shown in table 6.1. The N-value equals the stoichiometry (binding ratio) of the interaction only if the concentrations used are accurate. The lower N-value observed here could be due to inaccurate concentrations, in particular, less GB1-WW2 than expected in the cell

and/or more Smad7 in the syringe. These inaccuracies in concentration will also lower the ΔH value (heat of binding) and the K_d (dissociation constant) will be higher.

Table 6.1: Mean thermodynamic parameters obtained from GB1-WW2:Smad7 peptide ITC experiments, as determined using Malvern MicroCal PEAQ-ITC Analysis Software, after the subtraction of the Smad7 peptide titrated into buffer control.

	Mean	**SD**	**Units**
N	0.82	$\pm$0.06	Number of sites
$\mathbf{K_d}$	2.28	$\pm$0.18	μM
$\mathbf{\Delta H}$	-19.87	$\pm$0.85	kcal/mol
$\mathbf{\Delta G}$	-7.70	$\pm$0.05	kcal/mol
$\mathbf{-T\Delta S}$	12.20	$\pm$0.96	kcal/mol

The value of K_d was determined to be 2.28 $\pm$ 0.18 μM which suggests moderate binding for a biological reaction. A negative value of ΔG (Gibbs free energy), indicates that the binding interaction is favourable and the positive -TΔS value (entropy) indicates that during the interaction there is a decrease in entropy.

6.3 Discussion

This chapter looked at the interaction between the second WW domain of WWP2 and a peptide from Smad7, due to the involvement of Smad7 in the negative regulation of activated TGF-β receptors [Ebisawa *et al.*, 2001][Kavsak *et al.*, 2000]. The NMR titration and ITC results both indicate a reaction taking place between the protein and ligand. Previous research had shown that isoforms of E3 ligase WWP2 could interact with Smad7 in different ways, and pull-down studies had confirmed an interaction between WW4 (fourth WW domain of WWP2) and Smad7 [Soond and Chantry, 2011].

Initially, NMR titrations were acquired to detect movement of peaks, and hence determine if the structure of His-GB1-WW2 changed at all upon addition of Smad7. The total number of [^{1}H, ^{15}N]-HSQC peaks rose by 4 over the course of the titration and intensity of many peaks increased, implying that GB1-WW2 was in a more stable conformation in the presence of Smad7 peptide. A number of peaks also moved, appeared and disappeared as the titration progressed, indicating an interaction

between the protein and the peptide. Stable peak positions and intensity was reached at a ratio of 1:6 (His-GB1-WW2:Smad7), suggesting that saturation of GB1-WW2 was obtained as the remaining titration points did not affect the residue peak positions.

GB1 tag residues were all assigned except for the immediate N-terminal residues. They were also stationary except for one residue which showed very slight migration starting at ratio 1:0.5 and had moved by just 0.02 ppm by ratio 1:2. This titration was performed before the digestion of His-GB1-WW2 was optimised, and the His tag had been found to interfere with the conformation of GB1-WW2 (section 4.2.5). When the saturated titration end point of His-GB1-WW2:Smad7 peptide was compared with the digested GB1-WW2 in the presence of Smad7, an additional peak was present, despite the cleavage of 17 residues. This agrees with the previous suggestion that the His-tag wasn't folded and hence interfered with the conformation of GB1-WW2, leading to peaks in the [^{1}H, ^{15}N]-HSQC experiencing intermediate exchange.

The SUMO-Smad7 peptide construct was expressed and the protein was purified using the same conditions as SUMO-OCT4 peptide; however, it gave higher yields and a more stable peptide after digestion and lyophilising. There were no solubility issues when resuspending and purification could be performed at room temperature with fractions collected on ice.

92 of 100 residues were assigned in the 2D [^{1}H, ^{15}N]-HSQC of 0.5 mM ^{15}N, ^{13}C-labelled GB1-WW2 in the presence of 1.75 mM unlabelled Smad7 peptide, with only proline and N-terminus residues unassigned. Due to the incomplete assignment of GB1-WW2, the effect of Smad7 on all GB1-WW2 residues is not quantifiable in terms of peak position. It does, however, suggest that these regions are important as the intermediate exchange experienced in the absence of Smad7 is stabilised which implies that folding upon binding occurs. Peaks in the GB1 tag region have little to no peak movement in the presence of Smad7 indicating that there is no interaction between the peptide and the tag. The region of highest change includes the residues 89T and 90T, a region which also corresponds with a predicted β sheet.

The ITC results for GB1-WW2 and Smad7 peptide determined the K_d to be 2.28 $\pm$ 0.18 μM, which is in the same region as other WW domains from the NEDD4 E3 ligases with Smad7, for example, NEDD4L: WW1 (23.6 μM), WW2 (4.2 μM), WW3 (8.0 μM) and WW4 (12.4 μM), determined by Aragón et al. [2012] using ITC at

15 °C. Buffers were successfully matched, which was evident by the end of the experiments, where only the heat of dilution was present (all peaks were the same size).

The Gibbs free energy (ΔG = -7.70 kcal.mol^{-1}), enthalpy (ΔH = -19.87 kcal.mol^{-1}) and entropy (-TΔS = 12.20 kcal.mol^{-1}) values help to illustrate the binding of GB1-WW2 to Smad7 peptide. The positive -TΔS (entropy) value indicates that during the interaction there is a decrease in entropy. This is consistent with the protein folding upon binding and agrees with the data acquired during the NMR titrations. The negative (-ΔG) value indicates that the reaction is favourable, and the high enthalpy change (ΔH) value elucidates an exothermic reaction driven by enthalpy as opposed to entropy and the protein becomes more stable as it binds and folds.

The interaction between WW2 and Smad7 peptide was confirmed using both ITC and NMR titrations and the backbone of WW2 was assigned in the presence of Smad7 peptide. The K_d of 2.28 μM is typical of a moderate biological protein-protein interaction. Previous research by Lloyd Wahl [Wahl, 2016] determined K_d values for WW3 and WW4 domains of WWP2 from NMR titrations. The K_d values calculated for these domains are much higher than the K_d from ITC experiments of GB1-WW2 and Smad7 peptide and are shown in table 6.2.

Table 6.2: K_d values for the interaction of Smad7 peptide with WW3 and WW4 domains of WWP2, previously determined from NMR titration data by Lloyd Wahl [Wahl, 2016]. Standard deviation is given as the K_d error.

	K_d of Smad7 peptide
WW3	227 $\pm$27 μM
WW4	160 $\pm$59 μM

The affinity of the interaction between Smad7 peptide and WW2 is much higher, as indicated by the lower dissociation constant; however, the WW3:Smad7 and WW4:Smad7 dissociation constants were determined to be unreliable as saturation was not reached during the titrations.

Chapter 7

Discussion

As stated in section 1.8, the main aims of this thesis were to investigate interactions between WW domains of WWP2, an E3 ubiquitin ligase involved in pathways related to cell proliferation, and target regions OCT4 and Smad7, which have been linked to regulating cell proliferation and differentiation. Understanding the specificity of WW domains for certain binding partners provides direction for therapeutic targets with potential as anticancer drugs.

This chapter includes a comparison between the interactions of GB1-WW2 and the two target proteins OCT4 and Smad7, an evaluation of the different binding modes of GB1-WW2 and GB1-WW3 with OCT4 peptide, and an overview of the conclusions established throughout this project.

7.1 Interactions of WW2 and target proteins OCT4 and Smad7

The interactions between the WW2 domain of WWP2 and the two target regions from OCT4 and Smad7 were discussed separately in chapters 5 and 6 respectively. Isothermal Titration Calorimetry (ITC) was used to assess the binding affinity of WW2 for the two target regions. Unfortunately, binding saturation was not achieved

for the ITC experiment of GB1-WW2 with OCT4 peptide and so Malvern analysis software could not recognise the binding. Because of this, the K_d was estimated as $\sim$15 μM by fitting the binding curve in Excel. The binding curve, despite not reaching saturation, was found to fit well with the K_d fit. ITC of GB1-WW2 and Smad7 did reach saturation point and so Malvern analysis software calculated the K_d to be 2.28 ± 0.18 μM.

The titration of GB1-WW2 with synthetic OCT4 and recombinant Smad7 peptides were not fully comparable as the Smad7 titration used a polyhistidine-tagged GB1-WW2 construct (His-GB1-WW2). As previously described in chapter 6, the histidine tag interfered with the conformation of GB1-WW2, resulting in just 75% of the backbone being available for assignment at the start of the titration. By the end titration point, an additional 17 residues had appeared in the spectrum, allowing 92% of the construct to be assigned. The residues that could not be assigned were 3 N-terminal peaks and 5 proline residues. The titration with OCT4 peptide gained 7 peaks throughout the course of the titration, also yielding 92% assignment. The proline residues and two N-terminal residues were unobserved, as was residue 73Thr, which was likely still experiencing conformational exchange.

Peak movement was observed for the GB1-WW2:OCT4 titration from the first titration point at molar ratio 1:0.375 (0.4 mM:0.15 mM) for GB1-WW2:OCT4. The vast majority of peak movement occurred at the molar ratio 1:2.75 (0.4 mM:1.1 mM) and had reached their final positions at molar ratio 1:4.5 (0.4 mM:1.8 mM). Peaks were observed to decrease very slightly in intensity in the His-GB1-WW2:Smad7 titration at molar ratio 1:0.5 (0.16 mM:0.08 mM), peak positions reached their final position at molar ratio 1:2 (0.16 mM:0.32 mM) and saturation was reached at ratio 1:6 (0.16 mM:0.96 mM). GB1-WW2 varied in concentration between the two titrations, with 0.4 mM in the titration with OCT4 peptide and 0.16 mM in the titration with Smad7 peptide. A lower concentration of protein (GB1-WW2) would affect the binding as higher molar ratios of peptide would be required before residues were observed to move due to the dilution effect. For direct comparison of individual data points, it would be ideal to prepare all titration samples with uniform protein and buffer concentrations.

The trajectories observed for the GB1-WW2:Smad7 titration were far smaller than for the GB1-WW2:OCT4 titration; however, this is likely due to the large number of

peaks which were unobserved at the beginning of the Smad7 titration. Whilst the peaks which appear as the titration progresses cannot be tracked and hence their trajectories are not visualised on the graph, they are still involved in the binding. When in the free state, these residues are in multiple conformations and are broadened due to intermediate exchange; however, when bound, the resonances stabilise and are observable as they form a single, fully folded structure. When the largest peak trajectory from the GB1-WW2:Smad7 titration (79Phe) was compared with the GB1-WW2:OCT4 titration, both had a trajectory of ∼0.18 ppm.

When the bound states of ^{15}N, ^{13}C labelled GB1-WW2 with each peptide were compared, there were some differences in peak positions indicating that they did not bind in an identical way. Figure 7.1 shows the changes in peak positions of the residues in the WW2 domain (residues 61-100).

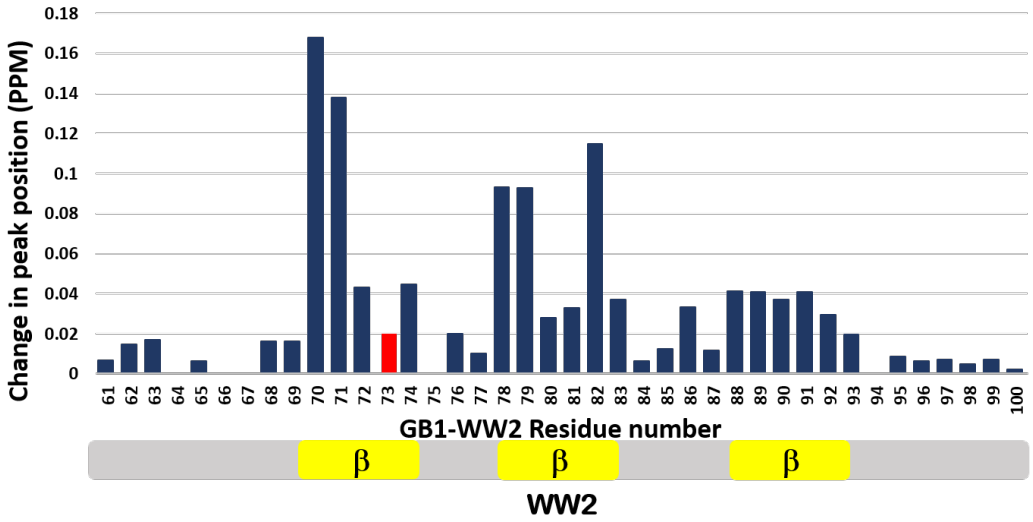


Figure 7.1: The difference in [^{1}H, ^{15}N]-HSQC peak positions (ppm) of the WW2 domain in the presence of OCT4 or Smad7 peptide. The WW2 schematic diagram showing predicted β strands – using PSIPRED webserver v3.3 [Buchan *et al.*, 2013].

The majority of the WW2 residues interact with the two peptides in a similar way, with small differences of under 0.03 ppm. However, residues which align with the three highlighted β sheet regions have more substantial differences and could highlight how the binding regions differ.

The gaps present in figure 7.1 are proline residues and hence were not assigned

in the [^{1}H, ^{15}N]-HSQCs. Residue 73T, highlighted in red at 0.02 ppm, could not be observed in the GB1-WW2 [^{1}H, ^{15}N]-HSQC or when in the presence of OCT4 peptide. It was however assigned in the presence of Smad7, indicating that this binding partner stabilises the WW2 domain. The location of this peak is highlighted in figure 7.2, where the GB1-WW2:OCT4 [^{1}H, ^{15}N]-HSQC (blue) is overlaid with the GB1-WW2:Smad7 [^{1}H, ^{15}N]-HSQC (red).

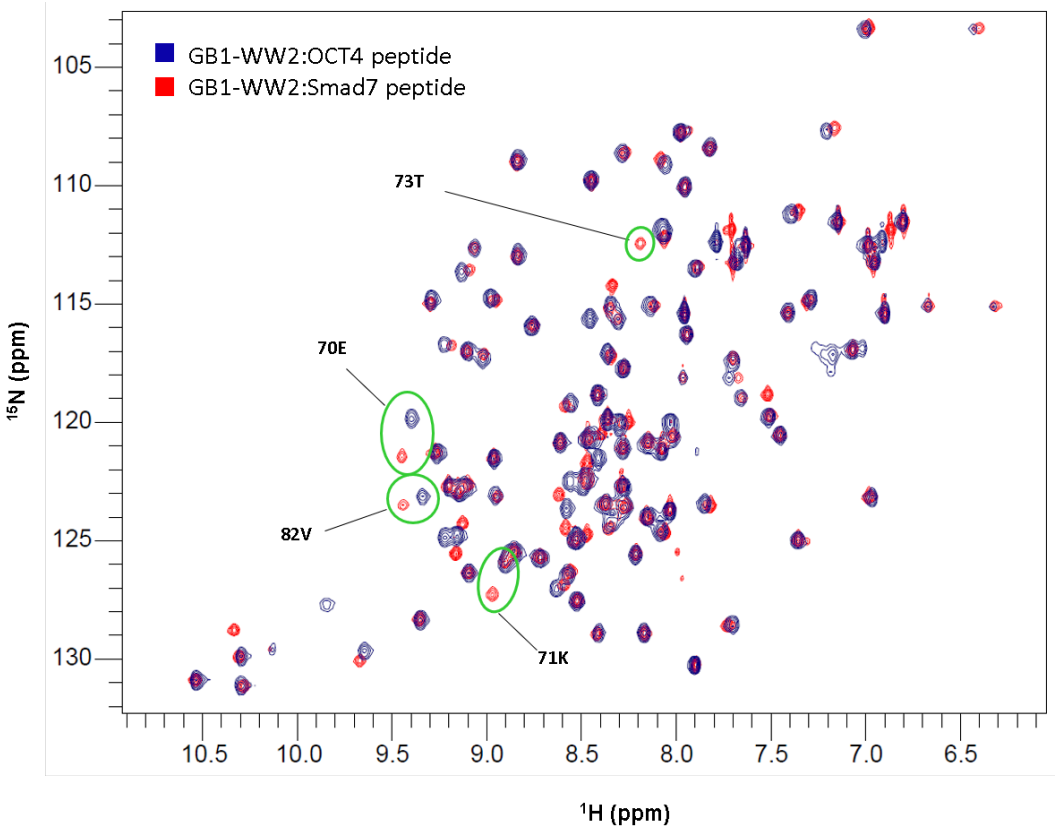


Figure 7.2: An overlay of GB1-WW2:Smad7 and GB1-WW2:OCT4 [^{1}H, ^{15}N]-HSQCs highlighting residues with the largest differences. Residue 73T is unassigned in the GB1-WW2:OCT4 [^{1}H, ^{15}N]-HSQC but is present for the GB1-WW2:OCT4 sample. Residue peaks for 70E, 71K and 82V had the largest observed changes in peak position.

It is clear that several GB1-WW2 resonances respond to different binding partners in unique ways. Other residues displaying the largest differences in position are highlighted with a green circle and labelled (figure 7.2).

The GB1 domain of the GB1-WW2:OCT4 bound sample displayed slight differences

in position to the GB1-WW2:Smad7 sample in the same regions discussed in previous chapters 4 and 5 (5Q and 23A region). This suggested that a portion of undigested polyhistidine-tagged protein remained in the GB1-WW2:OCT4 sample, which also led to some GB1 peaks being observed in both the His-tagged and digested peak positions (chapter 5).

The K_d estimation for GB1-WW2:OCT4, determined by fitting the binding curve in excel, had large uncertainties. Therefore, estimations were also made using a residue (74D) experiencing fast exchange processes in the GB1-WW2:OCT4 titration. The K_d estimated from the NMR titration was $\sim$150 μM. Whilst a lot of assumptions have been made, the K_d for GB1-WW2 and OCT4 could be described as in the region of 15-150 μM.

As mentioned above, the K_d of GB1-WW2:Smad7 was calculated using ITC to be 2.28 $\pm$ 0.18 μM. Therefore, it is likely that the interaction of GB1-WW2 with Smad7 peptide is between 10 and 100 times tighter than with OCT4 peptide.

The observations made during the two titration series also demonstrate that the interaction of GB1-WW2 with Smad7 peptide had a higher binding affinity than the interaction with OCT4 peptide. Slow exchange is an indicator of stronger binding as the lifetime of the complex is longer, as opposed to intermediate-fast exchange which is observed for complexes with weaker binding affinities as the exchange rate between the free and bound state (k_{off}) is faster. The GB1-WW2:Smad7 titration saw peaks which appeared and disappeared in slow exchange. In contrast, several peaks were observed in intermediate to fast exchange regimes for the GB1-WW2:OCT4 titration. The saturated bound GB1-WW2:Smad7 sample also allowed full assignment, whereas residue 73T was in an intermediate exchange regime in the saturated bound GB1-WW2:OCT4 peptide spectrum and so could not be observed.

7.2 Interactions of WW2 and WW3 with target protein OCT4

The interactions between the WW2 and WW3 domains of WWP2 and the target region from OCT4 were investigated using solution NMR titrations and discussed in

chapters 3 and 5 respectively. In both titrations, the GB1 tag region did not experience any changes in peak position, confirming that no interaction occurred between the peptide and the tag. Titrations of WW2 with synthetic OCT4 peptide saw migration of residue peaks from the initial titration point at molar ratio 1:0.375 (0.4 mM:0.15 mM). The majority of peak movement occurred at the molar ratio 1:2.75 (0.4 mM:1.1 mM) and binding saturation was reached at molar ratio 1:4.5 (0.4 mM:1.8 mM). NMR titrations of WW3 with recombinant OCT4 peptide also saw peaks move from the first titration point at molar ratio 1:0.25 (0.3 mM:0.075 mM); however, peaks continued to move up until the final titration point at molar ratio 1:10(0.3 mM:3 mM). Therefore, saturation was not achieved for the GB1-WW3:OCT4 titration.

In isolation, the GB1-WW2 construct experiences conformational exchange to the extent that only 82% assignment could be achieved. Upon titration with OCT4, an additional seven peaks emerged, indicating that the construct is more stable when in the presence of OCT4 peptide. When GB1-WW3 is isolated it can be fully assigned except for residue 77N; however, upon titration with OCT4 peptide, despite peaks appearing and disappearing throughout the titration, there was an overall decrease of 5 peaks. This indicated that these peaks shifted towards intermediate exchange processes and were so broad that they could not be observed.

The K_d for GB1-WW2 and OCT4 peptide could be described as in the region of 15-150 μM, as explained above in section 7.1. In contrast, the K_d of GB1-WW3 and OCT4 peptide was estimated to be 1.13 ± 0.24 mM. Whilst the K_d of GB1-WW3:OCT4 would involve large errors as saturation was not achieved, it is reasonable to estimate that the binding is of the mM order and hence binding is relatively weak. When the two domains are compared, the WW2 domain is likely to bind to OCT4 with an affinity 10-100 times higher than the WW3 domain.

This research agrees with data released by Xu *et al.* in 2009, where pull-down assays indicated that the WW2 domain interacted with a His-tagged construct of OCT at a higher affinity than the remaining WW domains (WW1, WW3 and WW4) [Xu *et al.*, 2009]. In contrast, when interactions between a smaller OCT4 peptide construct (T1: RRPCPPPYEFC) and the WW domains were investigated by Jiang *et al.* in 2015, similar K_d values were observed across all of the WW domains [Jiang *et al.*, 2015]. This could be because the OCT4 peptide used by Jiang *et al.* is too short and so the specificity

of the WW2 domain is therefore lost.

It is expected that under set conditions, a study of the different binding domains with peptide would give comparable results. The results by Jiang *et al.* were acquired using ITC with protein and peptide in buffer (25 mM Tris-HCl, 150 mM NaCl, pH 8.0) as opposed to here, where NMR titrations were performed in buffer (20 mM Na_2HPO_4, 50 mM NaCl, 15 mM DTT, pH 6.8. The variation in pH could play a large role in the differences observed in the K_d values; however, Xu *et al.* used a pH of 7.6 for GST pull-downs and was still able to observe the specificity for the WW2 domain [Xu *et al.*, 2009]. It is more likely that the small size of the peptide used by Jiang *et al.* led to the binding affinity being lost and instead all domains were able to bind the small domain in a roughly equal manner.

7.3 Conclusions

Analysis of the WW2 domain established that when isolated as a fusion protein with GB1, conformational exchange caused extensive peak broadening for a number of residues. Therefore, full resonance assignment of GB1-WW2 could not be achieved, and its structure could not be determined as anticipated. The removal of the polyhistidine-tag did permit 82% of the backbone resonances to be assigned; however, lowering the temperature of the sample to 278 K was not successful at pushing the exchange process into slow exchange.

Using solution NMR titrations and ITC, interactions between WW2 and targets OCT4 and Smad7, were confirmed. Interestingly, when in the presence of peptides from the target proteins OCT4 and Smad7, the WW2 domain bound the to the targets which in turn stabilised its structure. Smad7 and OCT4 peptides each had a similar, but not identical, effect on the structure of WW2 and provided an insight into the respective binding sites. The binding affinities of the two interactions were significantly different, with the interaction of GB1-WW2 with Smad7 peptide being between 10 and 100 times stronger than with OCT4 peptide.

With respect to the interaction of WW2 and WW3 domains with OCT4 peptide, the WW2 domain is estimated as having an affinity 10-100 times higher than that of the

WW3 domain, with both WW domains experiencing significant changes in conformation during the binding process. Structure calculations for the GB1-WW3 fusion protein proved difficult and despite gradual improvement, is not yet fully refined. The GB1 domain is more consistent with its secondary structure matching the literature and CSI predictions. The WW3 domain, which is predicted to be a triple stranded β sheet, is proving more difficult to elucidate, potentially due to insufficient or incorrect assignments in the region of the third β sheet. As mentioned in chapter 3, the third β sheet of a WW domain of E3 ligase human NEDD4-1 was suspected of being a more flexible strand which can sample non-native conformers and gains its rigidity when bound to a target ligand [Panwalkar *et al.*, 2016]. It is possible that the WW3 domain is experiencing the same flexibility in the third β strand and making structure calculations for this region more challenging.

Difficulties were also encountered with the expression and purification of the tandem WW2-3 construct due to degradation at the linker region. Once successfully purified under denaturing conditions, the construct was evaluated using solution NMR. The resultant [^{1}H, ^{15}N]-HSQC demonstrated that the protein had unfolded regions as well as a large number of peaks not visible in the spectrum, indicative of conformational exchange. Due to the instability of the construct, further experiments were not attempted.

The K_d values estimated in this project are different from the values in the literature. One factor which would affect dissociation constants would be the construct used for ligand binding. The PPxY containing OCT4 peptide construct was located in the NTD^{A} domain and so depending on the fold of that domain, the peptide created may not interact with the same affinity as in the natural full-length construct. The Smad7 peptide used in this project was derived from the linker region of the protein, meaning it is likely that this construct is more representative of the interaction with the full Smad7 protein.

Ultimately, this thesis has demonstrated how WW domain regions of WWP2 are able to interact specifically with binding partners OCT4 and Smad7 and provided insight into the binding sites involved in these interactions. A greater understanding of the structure and interactions of WWP2 and its protein substrates is crucial for determining how this protein can be targeted in the future for therapeutic applications.

References

Adams, J. (2002). Proteasome inhibitors as new anticancer drugs. *Current Opinion in Oncology*, 14(6):628–634.

Akitt, J. W. and Mann, B. E. (2000). *NMR and Chemistry: An introduction to modern NMR spectroscopy*. CRC Press.

Ambrosetti, D.-C., Scholer, H. R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the HMG and POU domains mediates the synergistic action of Sox2 and Oct-3 on the FGF-4 enhancer. *Journal of Biological Chemistry*, 275:23387–23397.

Anthis, N. (2013). Nick Anthis Protein Correlation Time Calculators. `http://nickanthis.com/tools/tau.html`. [Online; Accessed August 2018].

Aragón, E. *et al.* (2011). A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes & Development*, 25(12):1275–1288.

Aragón, E. *et al.* (2012). Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF-β pathways. *Structure*, 20(10):1726–1736.

Ardley, H. C. and Robinson, P. A. (2005). E3 ubiquitin ligases. *Essays in Biochemistry*, 41:15–30.

Assoian, R. K., Komoriya, A., Meyers, C. A., Miller, D. M., and Sporn, M. B. (1983). Transforming growth factor-beta in human platelets. identification of a major storage site, purification, and characterization. *Journal of Biological Chemistry*, 258(11):7155–7160.

Attisano, L. and Wrana, J. L. (1998). Mads and Smads in TGFβ signalling. *Current Opinion in Cell Biology*, 10(2):188–194.

Bai, S. and Cao, X. (2002). A nuclear antagonistic mechanism of inhibitory Smads in transforming growth factor-β signaling. *Journal of Biological Chemistry*, 277(6):4176–4182.

Baryshnikova, O. K., Williams, T. C., and Sykes, B. D. (2008). Internal pH indicators for biomolecular NMR. *Journal of Biomolecular NMR*, 41(1):5–7.

Bedford, M. T., Chan, D. C., and Leder, P. (1997). FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. *The EMBO Journal*, 16(9):2376–2383.

Bedford, M. T., Reed, R., and Leder, P. (1998). WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. *Proceedings of the National Academy of Sciences*, 95(18):10602–10607.

Berjanskii, M. V. and Wishart, D. S. (2005). A simple method to predict protein flexibility using secondary chemical shifts. *Journal of the American Chemical Society*, 127(43):14970–14971.

Berman, H. M. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242.

Bernassola, F., Karin, M., Ciechanover, A., and Melino, G. (2008). The HECT Family of E3 Ubiquitin Ligases: Multiple Players in Cancer Development. *Cancer Cell*, 14(1):10–21.

Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007). Evaluating protein structures determined by structural genomics consortia. *Proteins: Structure, Function, and Bioinformatics*, 66(4):778–795.

Borden, K. L. (2000). Ring domains: master builders of molecular scaffolds? *Journal of Molecular Biology*, 295(5):1103–1112.

Borden, K. L. and Freemont, P. S. (1996). The RING finger domain: a recent example of a sequence—structure family. *Current Opinion in Structural Biology*, 6(3):395–401.

Bork, P. and Sudol, M. (1994). The WW domain: a signaling site in dystrophin? *Trends in Biochemical Sciences*, 19(12):531–533.

Böttner, M., Krieglstein, K., and Unsicker, K. (2000). The transforming growth factor-βs: structure, signaling, and roles in nervous system development and functions. *Journal of Neurochemistry*, 75(6):2227–2240.

Bruch, M. (1996). *NMR spectroscopy techniques*. CRC Press.

Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, 41(W1):W349–W357.

Butt, T. R., Edavettal, S. C., Hall, J. P., and Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification*, 43(1):1–9.

Cancer Research UK (2016). Cancer Statistics for the UK. `http://www.cancerresearchuk.org/cancer-info/cancerstats/keyfacts/Allcancerscombined`. [Online; Accessed June 2018].

Cao, Y. *et al.* (2014). Selective small molecule compounds increase BMP-2 responsiveness by inhibiting Smurf1-mediated Smad1/5 degradation. *Scientific Reports*, 4:4965.

Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., and DeLucas, L. J. (2007). His-tag Impact On Structure. *Acta Crystallographica Section D: Biological Crystallography*, 63(3):295–301.

Cavanagh, J. *et al.* (2010). *Protein NMR Spectroscopy: Principles and Practice*. Elsevier.

Chen, C. and Matesic, L. E. (2007). The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer and Metastasis Reviews*, 26(3-4):587–604.

Chen, Z. *et al.* (2005). Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature*, 436(7051):725–730.

Chen, Z. *et al.* (2017). A tunable brake for HECT ubiquitin ligases. *Molecular Cell*, 66(3):345–357.

Cheng, Y. and Patel, D. J. (2004). An efficient system for small protein expression and refolding. *Biochemical and Biophysical Research Communications*, 317(2):401–405.

Cheung, M.-S., Maguire, M. L., Stevens, T. J., and Broadhurst, R. W. (2010). DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance*, 202(2):223–233.

Cho, W. (2001). Membrane targeting by C1 and C2 domains. *Journal of Biological Chemistry*, 276(35):32407–32410.

Choi, B. H., Che, X., Chen, C., Lu, L., and Dai, W. (2015). WWP2 is required for normal cell cycle progression. *Genes & Cancer*, 6(9-10):371–377.

Chong, P. A., Lin, H., Wrana, J. L., and Forman-Kay, J. D. (2010). Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity. *Proceedings of the National Academy of Sciences*, 107(43):18404–18409.

Choudhary, M. I. *et al.* (2015). *Solving problems with NMR spectroscopy.* Academic Press.

Christensen, D. E., Brzovic, P. S., and Klevit, R. E. (2007). E2–BRCA1 RING interactions dictate synthesis of mono-or specific polyubiquitin chain linkages. *Nature Structural and Molecular Biology*, 14(10):941–948.

Christensen, D. E. and Klevit, R. E. (2009). Dynamic interactions of proteins in complex networks: identifying the complete set of interacting E2s for functional investigation of E3-dependent protein ubiquitination. *The FEBS Journal*, 276(19):5381–5389.

Ciechanover, A., Orian, A., and Schwartz, A. L. (2000). The ubiquitin-mediated proteolytic pathway: Mode of action and clinical implications. *Journal of Cellular Biochemistry*, 77(34):40–51.

Ciechanover, A., Wolin, S. L., Steitz, J. A., and Lodish, H. F. (1985). Transfer RNA is an essential component of the ubiquitin-and ATP-dependent proteolytic system. *Proceedings of the National Academy of Sciences*, 82(5):1341–1345.

Clubb, R. T., Thanabal, V., and Wagner, G. (1992). A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^{1}H^{N}$, ^{15}N, and $^{13}C'$chemical shifts in ^{15}N-^{13}C-labelled proteins. *Journal of Magnetic Resonance*, 97:213–217.

Coux, O., Tanaka, K., and Goldberg, A. L. (1996). Structure and functions of the 20S and 26S proteasomes. *Annual Review of Biochemistry*, 65(1):801–847.

Croasmun, W. R. (1994). *Two-dimensional NMR spectroscopy: applications for chemists and biochemists*, volume 15. John Wiley & Sons.

Davis, C. M. and Dyer, R. B. (2013). Dynamics of an ultrafast folding subdomain in the context of a larger protein fold. *Journal of the American Chemical Society*, 135(51):19260–19267.

de Caestecker, M. P., Piek, E., and Roberts, A. B. (2000). Role of Transforming Growth Factor-β signaling in cancer. *Journal of the National Cancer Institute*, 92(17):1388–1402.

Delaglio, F. *et al.* (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR*, 6(3):277–293.

Department of Biology, Memorial University of Newfoundland (2018). Regulation of gene expression, principles of cell biology. `http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-23/CB23.html`. [Online; Accessed August 2018].

Derynck, R., Akhurst, R. J., and Balmain, A. (2001). TGF-β signaling in tumor suppression and cancer progression. *Nature Genetics*, 29(2):117–129.

Derynck, R. *et al.* (1996). Nomenclature: vertebrate mediators of TGFβ family signals. *Cell*, 87(2):173.

Deshaies, R. (1999). SCF and cullin/RING H2-based ubiquitin ligases. *Annual Review of Cell and Developmental Biology*, 15:435–467.

Drenth, J. (2007). *Principles of Protein X-Ray Crystallography*. Springer Science & Business Media.

Dunn, R., Klos, D. A., Adler, A. S., and Hicke, L. (2004). The C2 domain of the Rsp5 ubiquitin ligase binds membrane phosphoinositides and directs ubiquitination of endosomal cargo. *Journal of Cell Biology*, 165(1):135–144.

Ebisawa, T. *et al.* (2001). Smurf1 interacts with transforming growth factor-β type i receptor through Smad7 and induces receptor degradation. *Journal of Biological Chemistry*, 276:12477—-12480.

Esch, D. *et al.* (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nature Cell Biology*, 15(3):295–301.

Espanel, X. and Sudol, M. (1999). A single point mutation in a group I WW domain shifts its specificity to that of group II WW domains. *Journal of Biological Chemistry*, 274(24):17284–17289.

Fielding, L. (2003). NMR methods for the determination of protein-ligand dissociation constants. *Current Topics in Medicinal Chemistry*, 3(1):39–53.

Flasza, M., Gorman, P., Roylance, R., Canfield, A. E., and Baron, M. (2002). Alternative splicing determines the domain structure of WWP1, a Nedd4 family protein. *Biochemical and Biophysical Research Communications*, 290(1):431–437.

Freeman, R. and Anderson, W. (1962). Use of weak perturbing radio-frequency fields in Nuclear Magnetic Double Resonance. *The Journal of Chemical Physics*, 37(9):2053–2073.

Gao, M. *et al.* (2004). Jun turnover is controlled through JNK-dependent phosphorylation of the E3 ligase ITCH. *Science*, 306(5694):271–275.

Garcia-Gonzalo, F. and Rosa, J. (2005). The HERC proteins: functional and evolutionary insights. *Cellular and Molecular Life Sciences CMLS*, 62(16):1826–1838.

Glickman, M. H. and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiological Reviews*, 82(2):373–428.

Goujon, M. *et al.* (2010). A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Research*, 38(2):W695–W699.

Grabbe, C., Husnjak, K., and Dikic, I. (2011). The spatial and temporal organization of ubiquitin networks. *Nature Reviews Molecular Cell Biology*, 12(5):295–307.

Groll, M. and Huber, R. (2003). Substrate access and processing by the 20S proteasome core particle. *The international journal of biochemistry & cell biology*, 35(5):606–616.

Haas, A. and Rose, I. (1982). The mechanism of ubiquitin activating enzyme. a kinetic and equilibrium analysis. *Journal of Biological Chemistry*, 257(17):10329–10337.

Hafsa, N. E., Arndt, D., and Wishart, D. S. (2015). CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic Acids Research*, 43(W1):W370–W377.

Hafsa, N. E., Arndt, D., and Wishart, D. S. (2018). Csi 3.0 webserver. `http://csi3.wishartlab.com/cgi-bin/index.php`. [Online; Accessed September 2018].

Haglund, K., Di Fiore, P. P., and Dikic, I. (2003). Distinct monoubiquitin signals in receptor endocytosis. *Trends in Biochemical Sciences*, 28(11):598–604.

Handley, P. M., Mueckler, M., Siegel, N. R., Ciechanover, A., and Schwartz, A. L. (1991). Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1. *Proceedings of the National Academy of Sciences*, 88(1):258–262.

Hata, A., Massagué, J., and Shi, Y. (1998). TGF-β signaling and cancer: structural and functional consequences of mutations in smads. *Molecular Medicine Today*, 4(6):257–262.

Hayashi, H. *et al.* (1997). The MAD-related protein Smad7 associates with the TGFβ receptor and functions as an antagonist of TGFβ signaling. *Cell*, 89(7):1165–1173.

Hegde, A. N. (2004). Ubiquitin-proteasome-mediated local protein degradation and synaptic plasticity. *Progress in Neurobiology*, 73(5):311–357.

Heldin, C.-H., Landström, M., and Moustakas, A. (2009). Mechanism of TGF-β signaling to growth arrest, apoptosis, and epithelial–mesenchymal transition. *Current Opinion in Cell Biology*, 21(2):166–176.

Hershko, A. (1983). Ubiquitin: roles in protein modification and breakdown. *Cell*, 34(1):11–12.

Hershko, A. and Ciechanover, A. (1998). The Ubiquitin System. *Annual Review of Biochemistry*, 67(1):425–479. PMID: 9759494.

Hershko, A., Ciechanover, A., Heller, H., Haas, A. L., and Rose, I. A. (1980). Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proceedings of the National Academy of Sciences*, 77(4):1783–1786.

Hershko, A., Ciechanover, A., and Varshavsky, A. (2000). The ubiquitin system. *Nature Medicine*, 6(10):1073–1081.

Hongbao, M. and Shen, C. (2007). Review of stem cell studies. *Nature and Science*, 5:45–65.

Hu, L., McArthur, C., and Jaffe, R. (2010). Ovarian cancer stem-like side-population cells are tumourigenic and chemoresistant. *British Journal of Cancer*, 102(8):1276–1283.

Huang, L. *et al.* (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science*, 286(5443):1321–1326.

Huang, X. and Dixit, V. M. (2016). Drugging the undruggables: exploring the ubiquitin system for drug development. *Cell Research*, 26(4):484–498.

Huibregtse, J. M., Scheffner, M., Beaudenon, S., and Howley, P. M. (1995). A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proceedings of the National Academy of Sciences*, 92(7):2563–2567.

Hurley, J. H., Lee, S., and Prag, G. (2006). Ubiquitin-binding domains. *Biochemical Journal*, 399(3):361–372.

Iconomou, M. and Saunders, D. N. (2016). Systematic approaches to identify E3 ligase substrates. *Biochemical Journal*, 473(22):4083–4101.

Ilsley, J. L., Sudol, M., and Winder, S. J. (2002). The WW domain: linking cell signaling to the membrane cytoskeleton. *Cellular Signaling*, 14(3):183–189.

Ingham, R. J. *et al.* (2005). WW Domains Provide a Platform for the Assembly of Multiprotein Networks. *Molecular and Cellular Biology*, 25(16):7092–7106.

Ingham, R. J., Gish, G., and Pawson, T. (2004). The Nedd4 family of E3 ubiquitin ligases: functional diversity within a common modular architecture. *Oncogene*, 23(11):1972–1984.

Inoue, Y. and Imamura, T. (2008). Regulation of TGF-β family signaling by E3 ubiquitin ligases. *Cancer Science*, 99(11):2107–2112.

Iwai, H., Züger, S., Jin, J., and Tam, P.-H. (2006). Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Letters*, 580(7):1853–1858.

Izzi, L. and Attisano, L. (2004). Regulation of the TGFβ signaling pathway by ubiquitin-mediated degradation. *Oncogene*, 23(11):2071–2078.

Jackson, P. K. *et al.* (2000). The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. *Trends in Cell Biology*, 10(10):429–439.

Jacobsen, N. E. (2007). *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. Wiley.

Jauch, R., Choo, S. H., Ng, C. K. L., and Kolatkar, P. R. (2011). Crystal structure of the dimeric Oct6 (POU3f1) POU domain bound to palindromic MORE DNA. *Proteins: Structure, Function, and Bioinformatics*, 79(2):674–677.

Jiang, J. *et al.* (2015). Characterization of substrate binding of the ww domains in human wwp2 protein. *FEBS Letters*, 589(15):1935–1942.

Jones, D. P., Buchan, D., Cozzetto, D., and Bryson, K. (2018). The PSIPRED Protein Sequence Analysis Workbench. `http://bioinf.cs.ucl.ac.uk/psipred/`. [Online; Accessed September 2018].

Kanelis, V., Rotin, D., and Forman-Kay, J. D. (2001). Solution structure of a Nedd4 WW domain–ENaC peptide complex. *Nature Structural Biology*, 8(5):407–412.

Kato, Y., Miyakawa, T., Kurita, J.-i., and Tanokura, M. (2006). Structure of FBP11 WW1-PL ligand complex reveals the mechanism of proline-rich ligand recognition by group II/III WW domains. *Journal of Biological Chemistry*, 281(52):40321–40329.

Kavsak, P. *et al.* (2000). Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGFβ receptor for degradation. *Molecular Cell*, 6(6):1365–1375.

Keeler, J. (2011). *Understanding NMR spectroscopy*. Wiley, 2 edition.

Keepers, J. W. and James, T. L. (1984). A theoretical study of distance determinations from NMR. Two-dimensional nuclear overhauser effect spectra. *Journal of Magnetic Resonance (1969)*, 57(3):404–426.

Kim, H. C. and Huibregtse, J. M. (2009). Polyubiquitination by HECT E3s and the determinants of chain type specificity. *Molecular and Cellular Biology*, 29(12):3307–3318.

Kim, J. B. *et al.* (2009). Oct4-induced pluripotency in adult neural stem cells. *Cell*, 136(3):411–419.

Kitisin, K. *et al.* (2007). TGF-beta signaling in development. *Sci STKE*, 399:cm1.

Klemm, J. D. and Pabo, C. O. (1996). Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes & Development*, 10(1):27–36.

Komander, D. (2009). The emerging complexity of protein ubiquitination. *Biochemical Society Transactions*, 37(5):937–953.

Kozlowski, L. P. and Bujnicki, J. M. (2012). Metadisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics*, 13(1):111.

Kretzschmar, M., Liu, F., Hata, A., Doody, J., and Massague, J. (1997). The TGF-beta family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. *Genes & Development*, 11(8):984–995.

Kuliopulos, A., Walsh, C. T., *et al.* (1994). Production, purification, and cleavage of tandem repeats of recombinant peptides. *Journal of the American Chemical Society*, 116(11):4599–4607.

Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F., and Mackay, J. P. (2011). Macromolecular NMR spectroscopy for the non-spectroscopist. *The FEBS Journal*, 278(5):687–703.

Kwon, Y. *et al.* (2013). The Hippo signaling pathway interactome. *Science*, 342:737–740.

Larkin, M. A. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948.

Lecker, S. H., Goldberg, A. L., and Mitch, W. E. (2006). Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *Journal of the American Society of Nephrology*, 7(17):1807–1819.

Lee, E. Y. and Muller, W. J. (2010). Oncogenes and Tumor Suppressor Genes. *Cold Spring Harbor Perspectives in Biology*, 2(10):a003236.

Lee, M. J., Lee, B.-H., Hanna, J., King, R. W., and Finley, D. (2011). Trimming of ubiquitin chains by proteasome-associated deubiquitinating enzymes. *Molecular & Cellular Proteomics*, 10(5):R110–003871.

Levitt, M. (2013). *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. Wiley.

Li, J. *et al.* (2015). An overview of predictors for intrinsically disordered proteins over 2010–2014. *International Journal of Molecular Sciences*, 16(10):23446–23462.

Li, S.-J. and Hochstrasser, M. (1999). A new protease required for cell-cycle progression in yeast. *Nature*, 398(6724):246.

Li, W. *et al.* (2008). Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PloS One*, 3(1):e1487.

Liao, B. and Jin, Y. (2010). Wwp2 mediates Oct4 ubiquitination and its own auto-ubiquitination in a dosage-dependent manner. *Cell Research*, 20(3):332–344.

Liggett, A., Crawford, L., Walker, B., Morris, T., and Irvine, A. (2010). Methods for measuring proteasome activity: current limitations and future developments. *Leukemia Research*, 34(11):1403–1409.

Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003). ARIA: Automated NOE assignment and NMR structure calculation. *Bioinformatics*, 19(2):315–316.

Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods in Enzymology*, 339:71–90.

Lorick, K. L. *et al.* (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proceedings of the National Academy of Sciences*, 96(20):11364–11369.

Lowe, J. *et al.* (1995). Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. *Science*, 268(5210):533–539.

Lu, P.-J., Zhou, X. Z., Shen, M., and Lu, K. P. (1999). Function of WW domains as phosphoserine-or phosphothreonine-binding modules. *Science*, 283(5406):1325–1328.

Macias, M. J., Martin-Malpartida, P., and Massagué, J. (2015). Structural determinants of Smad function in TGF-β signaling. *Trends in Biochemical Sciences*, 40(6):296–308.

Maddika, S. *et al.* (2011). WWP2 is an E3 ubiquitin ligase for PTEN. *Nature Cell Biology*, 13(6):728–733.

Malakhov, M. P. *et al.* (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of Structural and Functional Genomics*, 5(1-2):75–86.

Malvern Panalytical (2018). Isothermal titration calorimetry (ITC). `https://www.malvernpanalytical.com/en/products/technology/microcalorimetry/isothermal-titration-calorimetry`. [Online; Accessed September 2018].

Marblestone, J. G. *et al.* (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Science*, 15(1):182–189.

Martinez-Noel, G., Müller, U., and Harbers, K. (2001). Identification of molecular determinants required for interaction of ubiquitin-conjugating enzymes and RING finger proteins. *European Journal of Biochemistry*, 268(22):5912–5919.

Massagué, J. (1998). TGF-β signal transduction. *Annual Review of Biochemistry*, pages 753–791.

Massagué, J. and Chen, Y.-G. (2000). Controlling TGF-β signaling. *Genes & Development*, 14(6):627–644.

Mehra, A., Attisano, L., and Wrana, J. L. (2000). Characterization of SMAD phosphorylation and SMAD/receptor interaction. *Methods in Molecular Biology*, 142:67–78.

Metzger, M. B., Pruneda, J. N., Klevit, R. E., and Weissman, A. M. (2014). RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochimica et Biophysica Acta*, 1843(1):47–60.

Nakao, A. *et al.* (1997). TGF-β receptor-mediated signaling through Smad2, Smad3 and Smad4. *The EMBO Journal*, 16(17):5353–5362.

Nalepa, G., Rolfe, M., and Harper, J. W. (2006). Drug discovery in the ubiquitin–proteasome system. *Nature Reviews Drug Discovery*, 5(7):596–613.

NESG Wiki (2011). NMR determined Rotational correlation time. `http://www.nmr2.buffalo.edu/nesg.wiki/NMR_determined_Rotational_correlation_time`. [Online; Accessed August 2018].

Nguyen, L. K., Dobrzynski, M., Fey, D., and Kholodenko, B. N. (2014). Polyubiquitin chain assembly and organization determine the dynamics of protein activation and degradation. *Frontiers in physiology*, 5:4.

Nichols, J. *et al.* (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, 95(3):379–391.

Nilges, M. (2016). Aria Webportal. `http://enmr.chemie.uni-frankfurt.de/portal/aria.html`. [Online; Accessed August 2018].

Niwa, H., Miyazaki, J. I., and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, 24(4):372–376.

Ogunjimi, A. A. *et al.* (2005). Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain. *Molecular Cell*, 19(3):297–308.

Okamoto, K. *et al.* (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell*, 60(3):461–472.

Olsen, S. K., Capili, A. D., Lu, X., Tan, D. S., and Lima, C. D. (2010). Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature*, 463(7283):906–912.

Olsen, S. K. and Lima, C. D. (2013). Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. *Molecular cell*, 49(5):884–896.

Özkan, E., Yu, H., and Deisenhofer, J. (2005). Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proceedings of the National Academy of Sciences*, 102(52):18890–18895.

Pan, G. J., Chang, Z. Y., Schöler, H. R., and Duanqing, P. (2002). Stem cell pluripotency and transcription factor Oct4. *Cell Research*, 12(5-6):321–329.

Panwalkar, V. *et al.* (2016). The Nedd4–1 WW domain recognizes the PY motif peptide through coupled folding and binding equilibria. *Biochemistry*, 55(4):659–674.

Passmore, L. A. and Barford, D. (2004). Getting into position: the catalytic mechanisms of protein ubiquitylation. *Biochemical Journal*, 379(3):513–525.

Petrovich, M., Jonsson, A. L., Ferguson, N., Daggett, V., and Fersht, A. R. (2006). ϕ-analysis at the experimental limits: mechanism of β-hairpin formation. *Journal of Molecular Biology*, 360(4):865–881.

Pettersen, E. F. *et al.* (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612.

Pickart, C. (1997). Targeting of substrates to the 26S proteasome. *The FASEB Journal*, 11(13):1055–1066.

Pickart, C. M. and Eddins, M. J. (2004). Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1695(1-3):55–72.

Pires, J. R. *et al.* (2001). Solution structures of the YAP65 WW domain and the variant L30 K in complex with the peptides GTPPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *Journal of Molecular Biology*, 314(5):1147–1156.

Pirozzi, G. *et al.* (1997). Identification of novel human WW domain-containing proteins by cloning of ligand targets. *Journal of Biological Chemistry*, 272(23):14611–14616.

Popovic, D., Vucic, D., and Dikic, I. (2014). Ubiquitination in disease pathogenesis and treatment. *Nature Medicine*, 20(11):1242–1253.

Qi, S., O'Hayre, M., Gutkind, J. S., and Hurley, J. H. (2014). Structural and biochemical basis for ubiquitin ligase recruitment by arrestin-related domain-containing protein-3 (ARRDC3). *Journal of Biological Chemistry*, 289(8):4743–4752.

Qin, Y. *et al.* (2016). Silencing of WWP2 inhibits adhesion, invasion, and migration in liver cancer cells. *Tumor Biology*, 37(5):6787–6799.

Quan, T., He, T., Kang, S., Voorhees, J. J., and Fisher, G. J. (2002). Ultraviolet irradiation alters transforming growth factor β/Smad pathway in human skin in vivo. *Journal of Investigative Dermatology*, 119(2):499–506.

Rechsteiner, M. (1998). The 26 S proteasome. In *Ubiquitin and the Biology of the Cell*, pages 147–189. Springer.

Rehm, T., Huber, R., and Holak, T. A. (2002). Application of NMR in Structural Proteomics: Screening for Proteins Amenable to Structural Analysis. *Structure*, 10(12):1613–1618.

Reményi, A. *et al.* (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes & Development*, 17(16):2048–2059.

Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature*, 414(6859):105–111.

Reynolds, R. J. and Schecker, J. A. (1995). Radiation, cell cycle, and cancer. *Los Alamos Science*, 23:51–89.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E., and Nilges, M. (2006). ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3):381–382.

Riling, C. *et al.* (2015). Itch WW Domains Inhibit its E3 ubiquitin Ligase Ativity by Blocking E2-E3 Trans-thiolation. *Journal of Biological Chemistry*, 290(39):23875–23887.

Rizo, J. and Südhof, T. C. (1998). C2-domains, structure and function of a universal Ca2+-binding domain. *Journal of Biological Chemistry*, 273(26):15879–15882.

Rock, K. L. *et al.* (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell*, 78(5):761–771.

Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirie, F., Rigby, P. W., and Staudt, L. M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature*, 345(6277):686–692.

Rotin, D. and Kumar, S. (2009). Physiological functions of the HECT family of ubiquitin ligases. *Nature Reviews Molecular Cell Biology*, 10(6):398–409.

Rule, G. S. and Hitchens, T. K. (2006). *Fundamentals of Protein NMR Spectroscopy*, volume 5. Springer Science & Business Media.

Sarkar, S. K. (1996). *NMR Spectroscopy and Its Application to Biomedical Research*. Elsevier.

Savage, C. *et al.* (1996). Caenorhabditis elegans genes sma-2, sma-3, and sma-4 define a conserved family of transforming growth factor beta pathway components. *Proceedings of the National Academy of Sciences*, 93(2):790–794.

Schägger, H. (2006). Tricine-SDS-page. *Nature Protocols*, 1(1):16–22.

Scheffner, M., Nuber, U., and Huibregtse, J. M. (1995). Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature*, 373(6509):81–83.

Scheffner, M. and Staub, O. (2007). HECT E3s and human disease. *BMC Biochemistry*, 8(Suppl 1)(1):S6.

Schimke, R. T. (1964). The importance of both synthesis and degradation in the control of arginase levels in rat liver. *Journal of Biological Chemistry*, 239(11):3808–3817.

Schimke, R. T. and Doyle, D. (1970). Control of Enzyme Levels in Animal Tissues. *Annual Review of Biochemistry*, 39(1):929–976.

Schlesinger, D. H., Goldstein, G., and Niall, H. D. (1975). Complete amino acid sequence of ubiquitin, an adenylate cyclase stimulating polypeptide probably universal in living cells. *Biochemistry*, 14(10):2214–2218.

Schöler, H. R. (1991). Octamania: the POU factors in murine development. *Trends in Genetics*, 7(10):323–329.

Sekelsky, J. J., Newfeld, S. J., Raftery, L. A., Chartoff, E. H., and Gelbart, W. M. (1995). Genetic characterization and cloning of mothers against dpp, a gene required for decapentaplegic function in Drosophila melanogaster. *Genetics*, 139(3):1347–1358.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TTALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, 44(4):213–223.

Shi, D. and Grossman, S. R. (2010). Ubiquitin becomes ubiquitous in cancer: emerging roles of ubiquitin ligases and deubiquitinases in tumorigenesis and as therapeutic targets. *Cancer Biology & Therapy*, 10(8):737–747.

Shi, Y. and Massagué, J. (2003). Mechanisms of TGF-β signaling from cell membrane to the nucleus. *Cell*, 113(6):685–700.

Smith, M. C., Furman, T., Ingolia, T. t., and Pidgeon, C. (1988). Chelating peptide-immobilized metal ion affinity chromatography. a new concept in affinity chromatography for recombinant proteins. *Journal of Biological Chemistry*, 263(15):7211–7215.

Solter, D. (2000). Mammalian cloning: advances and limitations. *Nature Reviews Genetics*, 1(3):199–207.

Soond, S. M. and Chantry, A. (2011). Selective targeting of activating and inhibitory Smads by distinct WWP2 ubiquitin ligase isoforms differentially modulates TGFβ signaling and EMT. *Oncogene*, 30(21):2451–2462.

Soond, S. M. *et al.* (2013). Novel WWP2 ubiquitin ligase isoforms as potential prognostic markers and molecular targets in cancer. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1832(12):2127–2135.

Spratt, D. E., Walden, H., and Shaw, G. S. (2014). RBR E3 ubiquitin ligases: new structures, new insights, new questions. *Biochemical Journal*, 458(3):421–437.

Staub, O. and Rotin, D. (1996). Ww domains. *Structure*, 4(5):495–499.

Sterneckert, J., Höing, S., and Schöler, H. R. (2012). Concise review: Oct4 and more: the reprogramming expressway. *Stem Cells*, 30(1):15–21.

Stewart, M. D., Ritterhoff, T., Klevit, R. E., and Brzovic, P. S. (2016). E2 enzymes: more than just middle men. *Cell Research*, 26(4):423–440.

Stolfi, C., Marafini, I., De Simone, V., Pallone, F., and Monteleone, G. (2013). The dual role of Smad7 in the control of cancer growth and metastasis. *International Journal of Molecular Sciences*, 14(12):23774–23790.

Sturm, R. A., Das, G., and Herr, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes & Development*, 2(12a):1582–1599.

Sudol, M. (1996a). Structure and function of the WW domain. *Progress in Biophysics and Molecular Biology*, 65(1-2):113–132.

Sudol, M. (1996b). The WW module competes with the SH3 domain? *Trends in Biochemical Sciences*, 21(5):161–163.

Sudol, M., Chen, H. I., Bougeret, C., Einbond, A., and Bork, P. (1995). Characterization of a novel protein-binding module – the WW domain. *FEBS Letters*, 369(1):67–71.

Sudol, M. and Hunter, T. (2000). NeW wrinkles for an old domain. *Cell*, 103(7):1001–1004.

Sudol, M., Recinos, C. C., Abraczinskas, J., Humbert, J., and Farooq, A. (2005). WW or WOW: the WW domains in a union of bliss. *IUBMB Life*, 57(12):773–778.

Sudol, M., Sliwa, K., and Russo, T. (2001). Functions of WW domains in the nucleus. *FEBS Letters*, 490(3):190–195.

Tang, L.-Y. *et al.* (2011). Ablation of Smurf2 reveals an inhibition in TGF-β signaling through multiple mono-ubiquitination of Smad3. *The EMBO Journal*, 30(23):4777–4789.

Tanksley, J. P., Chen, X., and Coffey, R. J. (2013). NEDD4l is downregulated in colorectal cancer and inhibits canonical WNT signaling. *PloS one*, 8(11):e81514.

Ten Dijke, P. and Hill, C. S. (2004). New insights into TGF-β-Smad signaling. *Trends in Biochemical Sciences*, 29(5):265–273.

Teng, R. W., Wang, D. Z., Wu, Y. S., Lu, Y., Zheng, Q. T., and Yang, C. R. (2005). NMR assignments and single-crystal X-ray diffraction analysis of deoxyloganic acid. *Magnetic Resonance in Chemistry*, 43(1):92–96.

Tenno, T. *et al.* (2004). Structural basis for distinct roles of Lys63-and Lys48-linked polyubiquitin chains. *Genes to Cells*, 9(10):865–875.

Thomas, P. D., Basus, V. J., and James, T. L. (1991). Protein solution structure determination using distances from two-dimensional nuclear Overhauser effect experiments: Effect of approximations on the accuracy of derived structures. *Proceedings of the National Academy of Sciences*, 88(4):1237–1241.

Thrower, J. S., Hoffman, L., Rechsteiner, M., and Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. *The EMBO Journal*, 19(1):94–102.

Umadevi, N., Kumar, S., and Narayana, N. (2005). Crystallization and preliminary X-ray diffraction studies of the WW4 domain of the Nedd4-2 ubiquitin–protein ligase. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 61(12):1084–1086.

van Laar, J. M. and Huizinga, T. W. (2005). Tweaking microtubules to treat scleroderma. *PLoS Medicine*, 2(12):1230–1231.

van Wijk, S. J. and Timmers, H. M. (2010). The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. *The FASEB Journal*, 24(4):981–993.

Verdecia, M. A. *et al.* (2003). Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. *Molecular Cell*, 11(1):249–259.

Voorhees, P. M., Dees, E. C., O'Neil, B., and Orlowski, R. Z. (2003). The proteasome as a target for cancer therapy. *Clinical Cancer Research*, 9(17):6316–6325.

Vranken, W. F. *et al.* (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*, 59(4):687–696.

Wahl, L. (2016). *Decoding the structure WWP2 in the TGFβ signaling pathway.* PhD thesis, University of East Anglia.

Walz, J. *et al.* (1998). 26S proteasome structure revealed by three-dimensional electron microscopy. *Journal of Structural Biology*, 121(1):19–29.

Wang, X. *et al.* (2007). NEDD4-1 is a proto-oncogenic ubiquitin ligase for PTEN. *Cell*, 128(1):129–139.

Wassenaar, T. A. *et al.* (2012). WeNMR: Structural Biology on the Grid. *Journal of Grid Computing*, 10(4):743–767.

Wenzel, D. M., Stoll, K. E., and Klevit, R. E. (2011). E2s: structurally economical and functionally replete. *Biochemical Journal*, 433(1):31–42.

Whitman, M. (1998). Smads and early developmental signaling by the TGFβ superfamily. *Genes & Development*, 12(16):2445–2462.

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73:1–16.

Wilton, D. J., Tunnicliffe, R. B., Kamatari, Y. O., Akasaka, K., and Williamson, M. P. (2008). Pressure-induced changes in the solution structure of the GB1 domain of protein G. *Proteins: Structure, Function, and Bioinformatics*, 71(3):1432–1440.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995). ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR*, 5(1):67–81.

Wrana, J. L. and Attisano, L. (2000). The Smad pathway. *Cytokine & Growth Factor Reviews*, 11(1):5–13.

Wrana, J. L., Attisano, L., Wieser, R., Ventura, F., and Massagué, J. (1994). Mechanism of activation of the TGF-β receptor. *Nature*, 370(6488):341.

Wu, G. and Schöler, H. R. (2014). Role of Oct4 in the early embryo development. *Cell Regeneration*, 3(1):7.

Xu, H. *et al.* (2009). WWP2 promotes degradation of transcription factor OCT4 in human embryonic stem cells. *Cell Research*, 19(5):561–573.

Xu, H. M. *et al.* (2004). Wwp2, an E3 ubiquitin ligase that targets transcription factor Oct-4 for ubiquitination. *Journal of Biological Chemistry*, 279(22):23495–23503.

Xu, S.-Q. *et al.* (2016). Inhibition of WWP2 suppresses proliferation, and induces G1 cell cycle arrest and apoptosis in liver cancer cells. *Molecular Medicine Reports*, 13(3):2261–2266.

Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1993). Two-Dimensional NMR Experiments for Correlating ^{13}Cβ and ^{1}hδ, ϵ Chemical Shifts of Aromatic Residues in ^{13}c-labeled Proteins via Scalar Couplings. *Journal of the American Chemical Society*, 115(23):11054–11055.

Yan, X. and Chen, Y.-G. (2011). Smad7: not only a regulator, but also a cross-talk mediator of TGF-β signaling. *Biochemical Journal*, 434(1):1–10.

Yan, X., Liu, Z., and Chen, Y. (2009). Regulation of TGF-β signaling by Smad7. *Acta Biochimica et Biophysica Sinica*, 41(4):263–272.

Yang, H. P. *et al.* (2015). PTEN expression in benign human endometrial tissue and cancer in relation to endometrial cancer risk factors. *Cancer Causes & Control*, 26(12):1729–1736.

Ye, Y. and Rape, M. (2009). Building ubiquitin chains: E2 enzymes at work. *Nature Reviews Molecular Cell Biology*, 10(11):755–764.

Zarrinpar, A. and Lim, W. A. (2000). Converging on proline: the mechanism of WW domain peptide recognition. *Nature Structural and Molecular Biology*, 7(8):611–613.

Zhang, W. *et al.* (2016). System-wide modulation of HECT E3 ligases with selective ubiquitin variant probes. *Molecular Cell*, 62(1):121–136.

Zhou, P. and Wagner, G. (2010). Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. *Journal of Biomolecular NMR*, 46(1):23–31.

Zhu, H., Kavsak, P., Abdollah, S., Wrana, J. L., and Thomsen, G. H. (1999). A SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation. *Nature*, 400(6745):687–693.

Zou, W. *et al.* (2011). The E3 ubiquitin ligase Wwp2 regulates craniofacial development through mono-ubiquitylation of Goosecoid. *Nature Cell Biology*, 13(1):59–65.

Zuo, X. *et al.* (2005). Enhanced expression and purification of membrane proteins by SUMO fusion in Escherichia coli. *Journal of Structural and Functional Genomics*, 6(2-3):103–111.

Appendix A

GB1-WW3 NMR Resonance Assignment

Table A.1: The NMR assignments for the GB1-WW3 construct. Previous work by Lloyd Wahl [Wahl, 2016] led to many residues being successfully assigned. Assignments in bold show residues which were assigned or corrected during this project. Where assignments do not exist, a dash is used.

3	H	4.60	-	-	56.43	31.37
4	M	4.64	8.43	120.66	54.67	34.95
5	Q	4.93	8.03	119.90	55.84	30.77
6	Y	5.35	9.24	125.04	57.25	43.04
7	K	5.20	9.13	122.70	55.32	36.08
8	L	4.95	8.57	126.45	52.58	42.82
9	I	4.32	9.11	126.41	60.31	38.15
10	L	4.50	8.71	125.70	54.50	42.25
11	N	5.26	8.87	125.78	51.21	38.18
12	G	4.47	7.97	110.04	44.81	-
13	K	4.06	9.32	121.38	59.27	32.64
14	T	4.08	8.85	108.95	62.08	69.77
15	L	4.47	7.36	125.22	55.23	43.56
16	K	5.14	8.15	124.10	53.94	34.77
17	G	4.32	8.45	109.68	45.24	-
18	E	5.60	8.42	118.88	54.74	34.00

NUM	RES	HA (ppm)	H (ppm)	HN (ppm)	Cα (ppm)	Cβ (ppm)
19	T	4.72	8.75	115.90	60.63	69.47
20	T	5.80	8.05	111.95	59.90	73.39
21	T	4.71	8.99	114.87	62.32	69.97
22	E	5.45	8.08	124.57	55.04	31.00
23	A	4.88	9.19	124.88	51.07	23.82
24	V	4.04	8.46	115.68	63.17	32.00
25	D	4.41	7.30	114.82	52.62	42.11
26	A	3.50	8.41	121.55	54.74	17.43
27	A	3.97	8.27	122.71	54.85	17.92
28	**T**	**3.69**	**8.36**	**117.10**	**66.82**	**67.85**
29	A	3.10	6.95	123.19	54.98	17.12
30	E	2.76	7.96	116.31	60.08	28.92
31	K	3.73	7.06	116.92	59.90	32.41
32	V	3.64	7.43	120.67	66.14	31.81
33	F	4.80	8.46	120.78	56.44	37.48
34	K	4.17	9.16	123.04	59.94	31.74
35	Q	4.06	7.49	119.73	58.75	28.28
36	Y	4.27	8.29	121.17	62.01	38.61
37	A	3.83	9.21	122.70	56.38	17.94
38	N	4.46	8.27	117.69	57.07	38.90
39	D	4.37	8.97	121.59	57.01	40.04
40	N	4.61	7.41	115.47	53.80	40.07
41	G	3.94	7.82	108.32	46.98	-
42	V	4.16	8.14	120.88	62.06	33.53
43	D	4.87	8.49	127.45	52.38	43.34
44	G	4.23	7.96	107.62	45.60	-
45	E	4.70	8.05	120.89	55.53	31.60
46	W	5.38	9.36	128.40	57.82	30.46
47	T	4.85	9.31	115.01	60.50	72.32
48	Y	5.00	8.61	120.92	56.63	41.83
49	D	4.62	7.72	128.57	51.88	42.93
50	D	4.04	8.53	124.91	56.38	41.97
51	A	4.12	8.30	120.07	55.05	18.36

NUM	RES	HA (ppm)	H (ppm)	HN (ppm)	Cα (ppm)	Cβ (ppm)
52	T	4.40	7.00	103.42	60.36	70.16
53	K	4.14	7.85	123.43	57.11	29.71
54	T	5.53	7.40	111.21	62.13	72.10
55	Y	5.64	10.30	131.14	58.13	42.52
56	T	5.21	9.12	117.03	61.58	71.05
57	V	4.49	8.25	123.55	57.99	32.38
58	T	4.73	8.38	123.46	61.02	70.59
59	E	4.55	8.14	129.02	56.33	32.02
60	G	4.12	8.79	112.72	45.23	-
61	S	4.14	8.29	115.33	58.01	64.26
62	D	-	8.44	123.59	52.32	41.22
63	P	4.39	-	-	64.12	32.31
64	L	4.29	8.44	119.07	55.21	42.14
65	**G**	**4.20**	**7.85**	**107.60**	**44.63**	-
66	P	4.50	-	-	62.53	32.19
67	**L**	**4.16**	**8.70**	**122.45**	**53.63**	**41.15**
68	P	-	-	-	-	29.60
69	P	4.42	-	-	64.21	32.01
70	G	4.20	8.85	112.28	45.30	-
71	W	5.55	7.64	117.88	56.88	32.17
72	E	4.71	9.46	121.44	55.07	34.66
73	K	4.51	8.84	126.15	55.77	34.59
74	R	4.42	8.70	127.14	54.09	33.59
75	Q	5.11	8.16	116.37	54.84	31.40
76	**D**	**4.94**	**8.89**	**122.27**	**53.61**	**43.08**
77	N	4.44	-	-	54.29	37.97
78	G	4.21	8.65	104.74	45.50	-
79	R	4.71	7.89	121.61	55.25	32.28
80	V	4.65	8.53	125.50	62.08	33.30
81	Y	4.80	8.49	122.32	55.80	39.08
82	Y	5.37	8.92	116.48	57.02	41.83
83	V	4.87	9.23	121.01	59.78	35.50
84	N	4.22	8.40	124.85	50.67	37.35

NUM	RES	HA (ppm)	H (ppm)	HN (ppm)	Cα (ppm)	Cβ (ppm)
85	H	4.26	8.70	123.75	58.68	30.69
86	N	4.54	8.43	116.84	55.64	38.35
87	T	4.32	7.27	106.87	61.11	69.26
88	R	3.83	7.93	119.34	57.27	26.81
89	**T**	**4.61**	**7.51**	**112.47**	**60.73**	**71.23**
90	T	5.35	8.42	117.79	60.79	71.33
91	Q	4.92	9.40	119.06	54.56	32.26
92	W	4.96	8.63	120.23	58.26	30.18
93	E	4.19	8.16	118.19	57.37	29.82
94	D	4.89	8.30	124.14	50.93	41.38
95	P	3.98	-	-	63.06	31.11
96	R	3.98	8.52	118.51	57.35	30.11
97	T	4.32	7.51	109.59	61.35	69.58
98	Q	-	7.30	126.95	57.34	30.62

Appendix B

GB1-WW2 NMR Resonance Assignment

Table B.1: The NMR assignments for the GB1-WW2 construct. Where assignments do not exist, a dash is used.

NUM	**RES**	**HN (ppm)**	**N (ppm)**	**Cα (ppm)**	**Cβ (ppm)**
1	G	8.54	110.51	-	-
2	S	-	-	-	-
3	H	8.34	120.90	-	-
4	M	8.40	120.52	54.46	33.68
5	Q	7.52	118.86	55.90	30.66
6	Y	9.13	124.34	57.23	43.02
7	K	9.10	122.70	55.33	36.02
8	L	8.56	126.35	52.68	42.74
9	I	9.10	126.41	60.27	38.32
10	L	8.71	125.68	54.43	42.29
11	N	8.86	125.50	51.25	38.30
12	G	7.97	110.02	44.72	-
13	K	9.27	121.34	59.28	32.63
14	T	8.85	108.95	62.09	69.68
15	L	7.35	125.05	55.30	43.65
16	K	8.15	124.13	54.09	34.73

17	G	8.45	109.81	45.27	-
18	E	8.43	118.94	54.80	33.93
19	T	8.76	115.93	60.77	69.49
20	T	8.07	112.22	59.81	73.24
21	T	8.96	114.86	62.25	69.95
22	E	8.06	124.71	55.13	31.16
23	A	9.17	125.59	51.20	23.43
24	V	8.34	114.20	63.50	32.00
25	D	7.31	114.87	52.71	42.25
26	A	8.47	121.77	54.86	17.48
27	A	8.29	122.71	54.93	17.89
28	T	8.35	117.24	67.01	67.78
29	A	6.99	123.22	55.10	17.17
30	E	7.95	116.50	60.04	28.50
31	K	7.07	117.01	59.80	32.32
32	V	7.46	120.62	66.11	31.78
33	F	8.47	120.75	56.56	37.42
34	K	9.16	123.00	59.87	31.64
35	Q	7.51	119.81	58.77	28.21
36	Y	8.28	121.17	62.06	38.67
37	A	9.21	122.73	56.43	17.93
38	N	8.28	117.78	56.96	38.90
39	D	8.97	121.54	57.01	40.05
40	N	7.42	115.44	53.94	40.06
41	G	7.83	108.42	47.01	-
42	V	8.15	120.86	62.13	33.24
43	D	8.53	127.59	52.58	43.31
44	G	7.97	107.73	45.61	-
45	E	8.04	120.62	55.58	31.57
46	W	9.36	128.34	57.92	30.51
47	T	9.30	115.00	60.62	72.22
48	Y	8.62	120.94	56.70	41.95
49	D	7.73	128.61	52.00	42.93
50	D	8.47	124.80	56.45	41.84

51	A	8.25	120.02	55.14	18.48
52	T	6.99	103.38	60.48	70.17
53	K	7.82	123.60	57.44	29.94
54	T	7.36	111.05	62.13	71.96
55	Y	10.29	3 131.154	58.19	42.18
56	T	9.11	117.09	61.53	70.97
57	V	8.28	123.69	58.16	32.38
58	T	8.39	123.37	60.95	70.51
59	E	8.18	128.75	56.60	31.77
60	G	8.84	112.87	45.48	-
61	S	8.34	115.66	55.90	-
62	E	8.50	122.65	56.04	30.61
63	R	8.51	124.88	54.07	29.90
64	P	-	-	-	-
65	L	8.49	122.89	52.78	41.44
66	P	-	-	-	-
67	P	-	-	-	-
68	G	8.92	112.26	45.69	-
69	W	7.63	118.28	56.80	31.89
70	E	9.35	121.80	-	-
71	K	-	-	-	-
72	R	8.70	127.54	-	-
73	T	-	-	-	-
74	D	9.53	129.86	-	-
75	P	-	-	-	-
76	R	8.17	115.26	55.25	30.23
77	G	8.33	108.70	45.45	-
78	R	8.40	121.88	55.45	31.12
79	F	8.47	122.23	56.34	39.90
80	Y	8.67	117.56	55.84	38.66
81	Y	-	-	-	-
82	V	-	-	-	-
83	D	-	-	-	-
84	H	-	-	-	-

85	N	-	-	-	-
86	T	6.66	103.37	-	-
87	R	7.65	118.87	-	-
88	T	-	-	-	-
89	T	8.33	115.93	60.76	71.48
90	T	8.09	119.27	-	-
91	W	-	-	-	-
92	Q	8.31	120.88	54.72	30.18
93	R	8.46	127.45	53.94	30.10
94	P	-	-	-	-
95	T	7.94	113.13	61.00	70.26
96	A	8.21	125.42	52.46	19.59
97	E	8.40	119.92	56.72	30.33
98	Y	8.13	121.45	57.93	38.86
99	V	7.99	123.91	62.34	32.98
100	R	7.90	130.26	57.61	31.56

Appendix C

GB1-WW2-OCT4 NMR Resonance Assignment

Table C.1: The NMR assignments for the GB1-WW2 construct in the presence of OCT4 peptide. Where assignments do not exist, a dash is used.

3	H	-	-	56.09	31.12
4	M	8.45	120.68	54.77	34.80
5	Q	8.03	120.00	55.78	30.85
6	Y	9.22	124.86	57.25	43.06
7	K	9.13	122.71	55.36	36.00
8	L	8.57	126.43	52.67	42.71
9	I	9.10	126.38	60.27	38.22
10	L	8.72	125.75	54.52	42.30
11	N	8.86	125.47	51.27	38.24
12	G	7.96	110.08	44.75	-
13	K	9.27	121.32	59.25	32.63
14	T	8.84	108.90	62.17	69.65
15	L	7.36	124.99	55.27	43.66
16	K	8.15	124.09	54.03	34.83
17	G	8.45	109.76	45.18	-
18	E	8.42	118.84	54.78	33.93
19	T	8.76	115.91	60.68	69.46

NUM	**RES**	**HN (ppm)**	**N (ppm)**	**Cα (ppm)**	**Cβ (ppm)**
20	T	8.06	112.05	59.97	73.41
21	T	8.98	114.82	62.31	69.98
22	E	8.08	124.61	55.17	31.19
23	A	9.16	124.60	51.16	23.68
24	V	8.45	115.49	63.18	31.93
25	D	7.29	114.81	52.77	42.12
26	A	8.42	121.59	54.71	17.45
27	A	8.28	122.76	54.88	17.92
28	T	8.36	117.11	66.85	67.77
29	A	6.96	123.21	55.09	17.17
30	E	7.94	116.30	6-	28.80
31	K	7.07	116.93	59.90	32.32
32	V	7.45	120.56	66.19	31.91
33	F	8.47	120.72	56.41	37.44
34	K	9.15	122.93	59.96	31.71
35	Q	7.51	119.81	58.76	28.24
36	Y	8.28	121.13	62.00	38.65
37	A	9.20	122.72	56.31	17.95
38	N	8.28	117.73	57.07	38.88
39	D	8.96	121.55	57.13	40.07
40	N	7.41	115.37	53.77	40.08
41	G	7.82	108.37	46.99	-
42	V	8.15	120.89	62.13	33.28
43	D	8.52	127.58	52.53	43.27
44	G	7.98	107.75	45.59	-
45	E	8.02	120.61	55.68	31.65
46	W	9.36	128.36	57.84	30.56
47	T	9.30	114.84	60.52	72.31
48	Y	8.61	120.86	56.63	41.93
49	D	7.71	128.56	51.81	42.90
50	D	8.52	124.89	56.29	42.05
51	A	8.30	120.08	55.17	18.37
52	T	7.00	103.38	60.43	70.16

NUM	RES	HN (ppm)	N (ppm)	Cα (ppm)	Cβ (ppm)
53	K	7.84	123.61	57.00	29.73
54	T	7.39	111.15	62.21	72.00
55	Y	10.30	131.14	58.14	42.48
56	T	9.10	116.99	61.62	71.05
57	V	8.27	123.59	58.06	32.37
58	T	8.38	123.47	60.96	70.54
59	E	8.17	128.90	56.53	31.79
60	G	8.84	112.90	45.39	-
61	S	8.31	115.64	58.32	64.06
62	E	8.49	122.62	56.01	30.64
63	R	8.53	124.94	54.05	29.89
64	P	-	-	62.92	32.24
65	L	8.48	122.35	52.97	41.22
66	P	-	-	-	-
67	P	-	-	64.08	31.83
68	G	9.07	112.63	45.98	-
69	W	7.70	117.41	56.07	32.41
70	E	9.40	119.94	55.11	35.42
71	K	8.91	126.13	55.46	34.97
72	R	8.63	127.07	54.49	34.11
73	T	8.77	114.92	60.71	70.61
74	D	9.65	129.67	53.10	40.70
75	P	-	-	65.42	32.40
76	R	8.14	115.03	54.86	30.13
77	G	8.28	108.59	45.39	-
78	R	8.56	122.43	55.82	31.36
79	F	8.58	123.64	56.76	39.59
80	Y	9.03	117.39	55.75	40.42
81	Y	9.22	116.77	57.61	40.89
82	V	9.34	123.09	60.88	34.22
83	D	8.34	124.29	51.17	39.36
84	H	8.95	123.24	58.99	30.31
85	N	8.35	115.07	56.04	38.51

NUM	**RES**	**HN (ppm)**	**N (ppm)**	**Cα (ppm)**	**Cβ (ppm)**
86	T	6.43	103.41	60.66	70.52
87	R	7.66	118.94	56.56	25.96
88	T	7.21	107.67	59.97	73.25
89	T	8.06	109.14	58.74	72.25
90	T	9.13	113.61	59.65	69.23
91	W	8.89	125.58	58.37	31.72
92	Q	8.56	119.37	55.36	30.04
93	R	8.41	128.89	53.95	30.05
94	P	-	-	62.32	31.34
95	T	7.90	113.49	61.00	70.21
96	A	8.21	125.60	52.45	19.50
97	E	8.36	119.88	56.51	30.37
98	Y	8.08	121.20	57.75	38.88
99	V	8.03	123.76	62.34	33.00
100	R	7.90	130.27	57.63	31.55

Appendix D

GB1-WW2-Smad7 NMR Resonance Assignment

Table D.1: The NMR assignments for the GB1-WW2 construct in the presence of Smad7 peptide. Where assignments do not exist, a dash is used.

4	M	8.40	120.55	54.49	33.67
5	Q	7.52	118.86	55.87	30.77
6	Y	9.14	124.34	57.33	42.95
7	K	9.11	122.68	55.35	36.00
8	L	8.56	126.37	52.72	42.58
9	I	9.10	126.40	60.21	38.16
10	L	8.72	125.70	54.33	42.31
11	N	8.86	125.53	51.28	38.24
12	G	7.96	110.04	44.75	-
13	K	9.28	121.34	59.27	32.67
14	T	8.85	108.94	62.16	69.61
15	L	7.36	125.05	55.29	43.61
16	K	8.16	124.11	53.98	34.72
17	G	8.45	109.80	45.14	-
18	E	8.43	118.87	54.86	33.95
19	T	8.75	115.89	60.57	69.39
20	T	8.06	112.13	59.95	73.30

NUM	RES	HN (ppm)	N (ppm)	Cα (ppm)	Cβ (ppm)
21	T	8.96	114.85	62.28	69.88
22	E	8.06	124.70	55.17	31.33
23	A	9.17	125.59	51.20	23.46
24	V	8.34	114.16	63.50	31.98
25	D	7.31	114.86	52.74	42.24
26	A	8.47	121.77	54.87	17.50
27	A	8.30	122.70	55.01	18.04
28	T	8.35	117.21	66.72	67.83
29	A	6.99	123.21	55.23	17.15
30	E	7.94	116.31	60.01	28.78
31	K	7.07	116.94	59.91	32.33
32	V	7.46	120.61	66.16	31.87
33	F	8.47	120.71	56.28	37.56
34	K	9.16	123.02	59.97	31.68
35	Q	7.51	119.80	58.70	28.30
36	Y	8.29	121.14	62.04	38.64
37	A	9.21	122.71	56.25	18.04
38	N	8.28	117.73	57.18	38.85
39	D	8.97	121.58	57.20	40.06
40	N	7.42	115.38	53.73	40.08
41	G	7.83	108.41	47.00	-
42	V	8.15	120.88	62.12	33.29
43	D	8.53	127.60	52.60	43.27
44	G	7.98	107.72	45.55	-
45	E	8.02	120.62	55.66	31.63
46	W	9.36	128.35	57.84	30.53
47	T	9.31	114.96	60.45	72.24
48	Y	8.62	120.90	56.64	41.93
49	D	7.73	128.60	51.77	42.78
50	D	8.48	124.82	56.25	41.81
51	A	8.26	120.03	55.10	18.45
52	T	6.99	103.35	60.34	70.08
53	K	7.82	123.60	57.52	29.95

NUM	RES	HN (ppm)	N (ppm)	Cα (ppm)	Cβ (ppm)
54	T	7.36	111.02	62.17	71.84
55	Y	10.29	131.17	58.11	42.45
56	T	9.11	117.03	61.72	71.00
57	V	8.27	123.64	58.02	32.40
58	T	8.38	123.46	60.98	70.47
59	E	8.17	128.93	56.44	31.83
60	G	8.85	112.95	45.35	-
61	S	8.32	115.62	58.28	64.06
62	E	8.48	122.58	55.95	30.65
63	R	8.54	125.07	54.02	29.81
64	P	-	-	-	-
65	L	8.49	122.31	53.00	41.19
66	P	-	-	-	-
67	P	-	-	-	-
68	G	9.08	112.60	45.94	-
69	W	7.70	117.28	56.16	32.43
70	E	9.45	121.55	55.04	35.38
71	K	8.98	127.33	55.39	34.98
72	R	8.59	126.89	55.20	34.59
73	T	8.20	112.39	61.00	70.50
74	D	9.67	130.08	53.17	40.74
75	P	-	-	-	-
76	R	8.12	115.07	55.22	30.39
77	G	8.27	108.63	45.30	-
78	R	8.62	123.16	55.80	31.31
79	F	8.59	124.56	56.67	39.48
80	Y	9.02	117.13	55.71	40.18
81	Y	9.19	116.76	57.74	40.74
82	V	9.44	123.60	60.89	33.85
83	D	8.36	124.58	51.22	39.23
84	H	8.95	123.22	59.05	30.19
85	N	8.35	115.20	55.95	38.46
86	T	6.41	103.35	60.83	70.50

NUM	**RES**	**HN (ppm)**	**N (ppm)**	**Cα (ppm)**	**Cβ (ppm)**
87	R	7.65	119.03	56.73	25.70
88	T	7.17	107.54	59.97	73.30
89	T	8.09	108.84	58.94	72.56
90	T	9.10	113.54	59.73	69.15
91	W	8.91	125.92	58.40	31.71
92	Q	8.59	119.34	55.41	29.85
93	R	8.42	128.99	53.79	30.07
94	P	-	-	-	-
95	T	7.89	113.43	61.01	70.14
96	A	8.22	125.59	52.38	19.51
97	E	8.37	119.90	56.48	30.42
98	Y	8.08	121.16	57.72	38.85
99	V	8.04	123.76	61.47	33.00
100	R	7.90	130.28	57.62	31.53