

**Recombinant expression and full backbone assignment of the human
DWNN domain using heteronuclear NMR.**

Andrew Faro

A thesis submitted in fulfilment of the degree of MSc in the Faculty of Natural Sciences,

University of the Western Cape



Supervisor: Dr. DJR Pugh

June 2004

**Recombinant expression and full backbone assignment of the human
DWNN domain using heteronuclear NMR.**

Andrew Faro

Keywords:

DWNN

Apoptosis

Protein

Recombinant

NMR

Nuclear

Magnetic

Resonance

Heteronuclear

Structural



Abstract

Recombinant expression and full backbone assignment of the human DWNN domain using heteronuclear NMR.

Andrew Faro

MSc thesis, Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape.

The cellular levels of a number of proteins have been found to be regulated by the ubiquitin-proteasome pathway. In this pathway, proteins are covalently tagged (“ubiquitinated”) by ubiquitin, which acts as a signal for degradation by the proteasome. A number of key cellular processes, including cell-cycle progression, transcription and DNA repair, are regulated in this way. In recent years a number of cellular proteins resembling ubiquitin in structure or function, the so-called ubiquitin-like proteins, have been identified. Ubiquitin-like proteins can be divided into two classes-the so-called “ubiquitin-like modifiers”, which consist of a single domain that structurally resembles ubiquitin, and “ubiquitin-domain” proteins, which are multi-domain proteins, which include domains that resemble ubiquitin.

This thesis describes the recombinant expression, purification and full backbone assignment of the human DWNN domain, a novel ubiquitin-like domain. The DWNN domain occurs at the N-terminus of RBBP6, a protein that has been shown to interact with p53 and Rb as well as to be involved in mRNA processing and apoptosis. A bacterial expression system was used to overexpress the DWNN domain as a GST fusion

protein. The domain was labelled with ^{15}N and ^{13}C to perform triple-resonance heteronuclear NMR experiments, from which full backbone assignments were obtained.

Although full structure determination of the DWNN domain falls outside the scope of this thesis, the backbone assignments formed the basis for the subsequent structure determination, which confirmed that the DWNN domain is indeed a novel ubiquitin-like domain. The RBBP6 protein may therefore represent a novel E3 ubiquitin ligase that plays a role in regulating the cellular levels of p53 and Rb.



Declaration

I declare that “*Recombinant expression and full backbone assignment of the human DWNN domain using heteronuclear NMR*” is my own work that has not been submitted for any degree or examination at any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Andrew Faro

June 2004



Signed:.....

Acknowledgements

I would like to express my sincere gratitude to Prof. Jasper Rees and Dr. David Pugh for allowing me the opportunity to do this project in their laboratory. Special thanks go to Dr. Pugh for his invaluable input and guidance during the last 3 years. Without him, this thesis would not have been possible.

Thanks also to Dr. Eberhard Hoffmann in Darmstadt, Germany, as well as Prof. Ian Campbell and his laboratory in Oxford, UK for their assistance with the NMR work. So many people have provided moral support and constructive input at UWC. Amongst them, I would like to give special thanks to Mervin Meyer, Portia Lutya, Amanda Skepu and Raymond Daniels.



My family has been considerably patient with me during this phase of my life and I would like to give a very special thanks to each and everyone of them.


Thank you to the National Research Foundation of South Africa and the Royal Society of England for providing the necessary financial support.

List of tables and charts

- Table 1.1:** P2P-R protein domains and associated binding factors.
- Table 2.1:** NMR experiments performed on the DWNN domain.
- Table 3.1:** Concentration and absorbance of lysozyme at 620 nm.
- Table 4.1:** C_{β} chemical shift values for DWNN domain cysteine residues.
- Table 5.1:** Summary of the secondary structural motifs formed by the DWNN backbone, as determined using the methods as indicated.
- Chart 3.1:** Lysozyme standard curve.



List of figures


- Fig. 1.1:** Multiple sequence alignment of DWNN domain amino acid sequences.
- Fig. 1.2:** The tertiary structure of ubiquitin as determined by X-ray crystallography (Vijay-Kumar et al, 1987).
- Fig. 1.3:** Domain organization of human RBBP6/DWNN-containing protein.
- Fig. 1.4:** The ubiquitin-proteasome pathway (Hershko and Ciechanover 1998).
- Fig. 3.1:** Screening colonies transformed with pGEX-6P2-DWNN for the expression of GST-DWNN fusion protein.
- Fig. 3.2:** Expression and purification of the GST-DWNN fusion protein.
- Fig. 3.3:** PreScission™ protease cleavage of the GST-DWNN fusion protein.
- Fig. 3.4:** Removal of the GST domain using a GSTPrep™ FF 16/10 column.
- Fig. 3.5:** Removal of the GST domain.
- Fig. 3.6:** Chromatogram showing  gel filtration chromatography of the DWNN domain.
- Fig. 3.7:** Silver stained gel of purified DWNN.
- Fig. 4.1:** 1D ¹H spectrum of the DWNN domain at pH 6.0, 25⁰ C, recorded at 600 MHz.
- Fig. 4.2:** Through bond transfer between the backbone nitrogen and the attached proton in a ¹⁵N-HSQC experiment.
- Fig. 4.3:** ¹⁵N-HSQC spectrum of the DWNN domain at pH 6.0, 25⁰ C, recorded at 600 MHz.
- Fig. 4.4:** Coherence transfer in CBCA(CO)NH and CBCANH experiments for a pair of consecutive residues.

- Fig. 4.5:** Strips showing backbone sequential connectivities of residues 29 to 33 of the DWNN domain.
- Fig. 4.6:** (A) Average C_α and C_β chemical shifts for all 20 commonly occurring residues.
(B) Average ^{13}C chemical shifts for aromatic ring carbons (Cavanagh et al, 1996).
- Fig. 4.7:** 3D ^{15}N -HSQC-NOESY strips depicting backbone H^{N} - H^{N} NOE's.
- Fig. 4.8:** Coherence transfer pathway in an HNCO experiment.
- Fig. 4.9:** Projection of the HNCO spectrum of the DWNN domain onto the H^{N} -N plane.
- Fig. 4.10:** HMQC-J spectrum of the DWNN domain at pH 6.0, 25⁰ C, recorded at 500 MHz.
- Fig. 4.11:** Diagram depicting the 3-bond coupling $^3J_{\text{H}^{\text{N}}-\text{H}^\alpha}$.
- Fig. 4.12:** The Karplus Curve, showing the relationship between $^3J_{\text{H}^{\text{N}}-\text{H}^\alpha}$ coupling constants, the backbone ϕ angle and polypeptide secondary structure.
- Fig. 4.13:** TALOS output for the DWNN domain
- Fig. 5.1:** Jpred predictions of the secondary structural motifs formed by the DWNN backbone
- Fig. 5.2:** Chemical shift index (CSI) of the DWNN domain

Abbreviations

Amp	Ampicillin
APF-1	ATP-dependent proteolysis factor 1
ATP	Adenosine triphosphate
BAG1	Bcl2-associated athanogene-1
BLAST	Basic local alignment search tool
cDNA	Complementary DNA
CHO	Chinese hamster ovary
CP	Core particle
CPF	Cleavage and polyadenylation factor
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DUB	Deubiquitinating
DWNN	Domain with no name
EDTA	Ethylene diaminetetra acetic acid
EST	Expressed sequence tag
GST	Gluthathione S-transferase
GTP	Guanosine triphosphate
HMQC	Heteronuclear multiple quantum coherence
hnRNP	Heterogeneous nuclear ribonucleoprotein
Hsp	Heat shock protein
HSQC	Heteronuclear single quantum coherence
I κ B	Inhibitor of NF- κ B
IPTG	Isopropyl β -D-thiogalactopyranoside




ISG15	Interferon-stimulated gene 15
JAMM	Jab1/Pad1/MPN-domain metallo-enzyme
kD	Kilodalton
LB	Lurea broth
MAR	Matrix associated region
mRNA	Messenger RNA
MHz	Megahertz
NER	Nucleotide excision repair
NF- κ B	Nuclear factor κ B
nm	Nanometer
NMR	Nuclear magnetic resonance
Otubain	OTU-domain ubiquitin-aldehyde-binding protein
P2P-R	Proliferation potential-related
Pab1p	poly(A)-binding protein 
PACT	p53 associated cellular protein-testis derived
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIM	Proteasome-interacting motif
PML	Promyelocytic leukaemia
PMSF	Phenylmethanesulphonyl fluoride
Rb	Retinoblastoma
RBBP6	Retinoblastoma binding protein 6
RING	Really interesting new gene

RNA	Ribonucleic acid
RP	Regulatory particle
Rpt1-6	Regulatory particle triple-A protein 1-6
Rpn	Regulatory particle non-ATPase
SDS	Sodium dodecyl sulphate
SR	Serine/arginine
SUMO	Small ubiquitin-like modifier
TEMED	<i>N, N, N', N'</i> - tetramethylethylenediamine
UBA	Ubiquitin-associated
UBL	Ubiquitin-like
UBP	Ubiquitin-specific processing protease
UCH	Ubiquitin C-terminal hydrolase
UDP	Ubiquitin-domain protein
ULP	Ubiquitin-like specific protease



Table of contents

	Page
Title.....	i
Keywords.....	ii
Abstract.....	iii
Declaration.....	v
Acknowledgements.....	vi
List of tables and charts.....	vii
List of figures.....	viii
Abbreviations.....	x
	
Chapter 1	1
1.1 The DWNN domain.....	1
1.2 The RBBP6 protein.....	2
1.3 Ubiquitin-mediated proteolysis.....	7
1.3.1 Ubiquitin.....	7
1.3.2 Ubiquitin-mediated proteolysis.....	7
1.4 Ubiquitin-like proteins.....	12
1.4.1 SUMO-1.....	13
1.4.2 Ubiquitin domain proteins.....	17
1.5 Aims.....	19

Chapter 2: Materials and methods	20
2.1 General stock solutions and buffers.....	20
2.2 Materials and methods	22
2.2.1 Preparation of competent <i>E.coli</i> BL21 (DE) pLysS cells.....	22
2.2.2 Small-scale preparation of plasmid DNA	23
2.2.3 Expression of recombinant DWNN protein.....	24
2.2.3.1 Construction of the pGEX-6p-2-DWNN plasmid.....	24
2.2.3.2 Transformation of <i>E.coli</i> BL21 (DE) pLysS with pGEX-6p-2-DWNN.....	25
2.2.3.3 Screening of <i>E.coli</i> BL21 (DE) pLysS transformed with pGEX-6p-2-DWNN.....	25
2.2.3.4 Large scale expression of recombinant DWNN protein.....	26
2.2.3.5 Double labelling of recombinant protein.....	27
2.2.4 Recombinant protein extraction and purification	27
2.2.4.1 Cell lysis and protein extraction.....	27
2.2.4.2 Purification of recombinant protein.....	28
2.2.4.3 PreScission Protease cleavage of recombinant protein and recovery of DWNN domain.....	28
2.2.4.5 Removal of the GST domain.....	29
2.3 SDS-PAGE	29
2.4 Bradford assay	30
2.5 NMR analysis and structure determination.....	31
2.5.1 Sample preparation	31
2.5.2 NMR experiments and data processing.....	31

Chapter 3: Expression and purification of the DWNN domain..... 33

3.1 Introduction.....	33
3.2 Expression and purification of the DWNN domain.....	34
3.2.1 Screening of <i>E.coli</i> BL21 (DE) pLysS transformed with pGEX-6P-2-DWNN for recombinant protein expression	34
3.2.2 Large scale expression and purification of recombinant DWNN protein.....	35
3.2.3 Removal of the GST domain	36
3.2.4 Bradford assay	37
3.2.5 Lyophilization.....	38
3.3 Summary	38



Chapter 4: Full backbone assignment of the human DWNN domain

using triple-resonance nuclear magnetic resonance..... 39

4.1 Introduction.....	39
4.2 Assessment of DWNN sample integrity and spectral quality using homonuclear 1D and ¹⁵ N-HSQC experiments spectra	40
4.3 Assignment of backbone resonances	42
4.3.1 Sequential assignment of C _α and C _β resonances using CBCA(CO)NH and CBCANH experiments	42
4.3.2 Identification of H ^N -H ^N NOE cross-peaks	45
4.3.3 Extraction of ¹³ CO chemical shifts	46

4.4 Extraction of secondary structure information from backbone chemical shifts	47
4.4.1 Extraction of $^3J_{H^N-H^\alpha}$ coupling constants	47
4.4.2 Prediction of DWNN domain backbone torsion angle restraints using TALOS	49
4.5 Summary	50

Chapter 5: Discussion and conclusion 52

5.1 Introduction.....	52
5.2 Recombinant expression and purification of the DWNN domain.....	52
5.3 Full backbone assignment of the DWNN domain	53
5.4 Extraction of secondary structure information from backbone chemical shifts	53
5.5 Backbone geometry of the DWNN domain.....	54
5.6 Conclusion	56



References..... 57

Appendix 1: DWNN domain chemical shifts.....72

Appendix 2: HMQC-J splittings and secondary structure motif.....83

Appendix 3: TALOS phi and psi angle predictions.....86

Chapter 1

1.1 The DWNN domain

The DWNN domain gene was isolated from retrovirally mutagenised Chinese hamster ovary 22 (CHO22) cells based on their resistance to cytotoxic T-lymphocyte killing (A.E. George, D.Phil. thesis, Oxford University, 1995). Regions flanking the retroviral integration site were amplified using inverse PCR and sequenced. Comparison of these sequences with the Genbank database showed 100% identity with human cDNA clone 21C4 (Genbank accession number T25012). Subsequent analysis of the cDNA database showed that the gene knocked out by the retrovirus integration coded for the 250 kD RBBP6 protein (GenBank accession number NP_008841). The retroviral integration site was into the N-terminal 80 residues of the protein. Database searches using BLAST (Altschul *et al.*, 1990) showed that this region of the protein had homologues in all eukaryotes for which sequence data was available (see Fig. 1.1), but had no significant homology to any previously characterised protein domains. We termed this new domain “DWNN”, the **D**omain **W**ith **N**o **N**ame.

Fold recognition using servers such as 3D-PSSM (Kelly *et al.*, 2000) and BIOINBGU (Fischer, 2000) predicted that the tertiary structure of the DWNN domain strongly resembles that of ubiquitin (Jonas and Pugh, unpublished data). Ubiquitin is a 76 residue globular protein that folds into a β -grasp structure consisting of a long α -helix packing against a five-stranded β -sheet (Vijay-Kumar *et al.*, 1987) as shown in Fig. 1.2. The four C-terminal residues, Leu-Arg-Gly-Gly, form a functionally important but unstructured

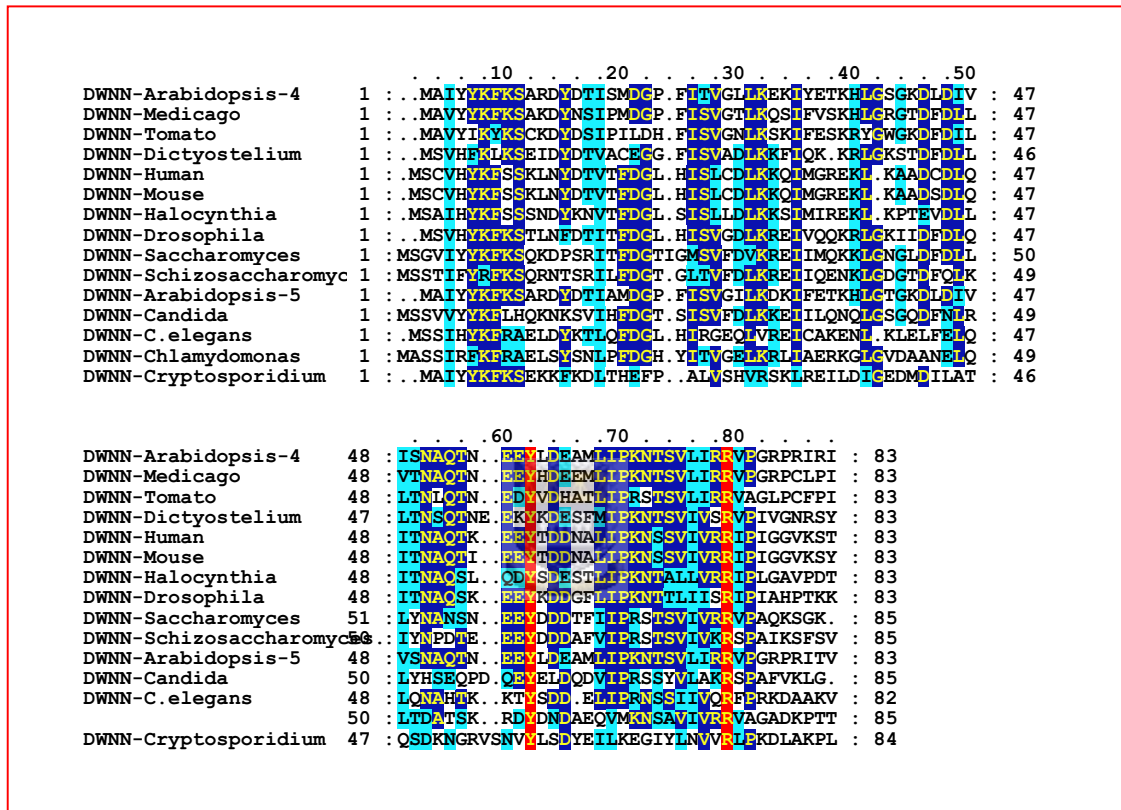


Fig. 1.1: Multiple sequence alignment of DWNN domain amino acid sequences. Highlighted residues indicate conserved or semi-conserved amino acids. Y57 and R74 (human sequence) are absolutely conserved since they occur in all eukaryotic sequences. This indicates that these residues might be important for the tertiary structure formation or function of the DWNN domain.

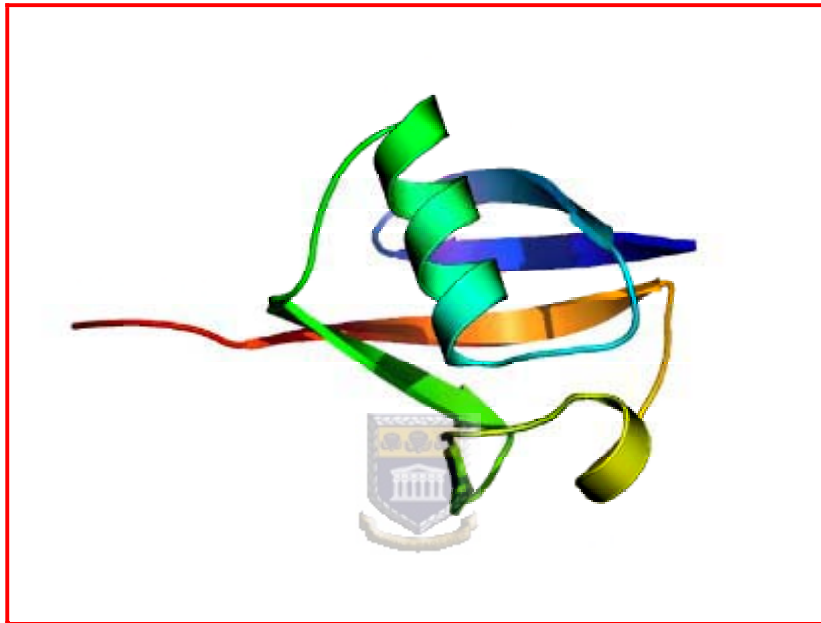


Fig. 1.2: The tertiary structure of ubiquitin as determined by X-ray crystallography (Vijay-Kumar et al, 1987). Ubiquitin folds into a β -grasp structure, consisting of a long α -helix packing against a five-stranded β -sheet. This figure was generated using PyMOL (Warren L. Delano “The PyMOL Molecular Graphics System.” Delano Scientific LLC, San Carlos, CA, USA. <http://www.pymol.org>).

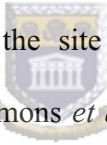
tail that extends from the compact structure. The C-terminal glycine is activated during the ubiquitin-conjugation cascade to covalently tag target proteins. Interestingly, for the majority of eukaryotic DWNN domain sequences the C-terminal di-glycine motif present in ubiquitin and other ubiquitin-like proteins is absent. Only the human and mouse DWNN domains contain a potential C-terminal di-glycine motif, namely Gly78 and Gly79. The finding that human DWNN is expressed as both a 250 kD DWNN-containing fusion protein and as a freestanding DWNN domain-only protein of 118 residues may be significant in this regard. It raises the possibility that the freestanding DWNN domain in humans may be conjugated to other proteins in the same way as other ubiquitin-like proteins such as SUMO-1 (see Section 1.4).



1.2 The RBBP6 protein

A database search for DWNN related sequences revealed that the 250 kD DWNN domain-containing protein showed significant homology to RBQ-1, also called retinoblastoma binding protein 6 (RBBP6). RBBP6 was first isolated from a human small cell lung carcinoma expression library using pRB as a probe (Saijo *et al.*, 1995). RBBP6 is a 140 kD protein that binds hypophosphorylated pRB but not phosphorylated pRB (Sakai *et al.*, 1995). The binding interaction was disrupted by adenovirus E1A protein, suggesting that RBBP6 possibly binds to the pRB pocket domain. A region of 34 amino acids in the middle of the RBBP6 sequence was found to be alternatively spliced. Simons and his co-workers (Simons *et al.*, 1997) identified a mouse homologue of RBBP6, which they called PACT, for **P**53 **A**ssociated **C**ellular protein-**T**estes derived, on the basis of its ability to bind p53. Analysis of the cDNA sequences for PACT and RBBP6 subsequently

revealed that RBBP6 is a truncated form of PACT. RBBP6 is truncated because of a stop codon in its open reading frame. This stop codon also occurs in PACT, but is out of frame due to two extra nucleotides present in the PACT coding sequence. The authors (Simons *et al.*, 1997) postulated that the stop codon in RBBP6 may be due to a mutational event in the small lung cell carcinoma, since the extra two nucleotides present in PACT occurred in all normal human sequences examined. The C-terminal part of PACT is able to bind to wild type p53 and interferes with the ability of p53 to bind DNA (Simons *et al.*, 1997), while the N-terminal region of PACT binds pRB (Sakai *et al.*, 1995). Simons and co-workers also showed that PACT, p53 and pRB co-precipitate in immunocomplexes, suggesting that PACT can bind p53 and pRB simultaneously.

PACT localizes to nuclear speckles,  the site of pre-mRNA splicing, and also co-precipitates with Sm splicing factor (Simons *et al.*, 1997). In addition, PACT has an N-terminal serine/arginine (SR) rich domain, shared by many pre-mRNA splicing factors. These observations pointed to a possible role for PACT in pre-mRNA splicing. Witte and Scott (Witte and Scott, 1997) reported that P2P-R (proliferation potential-related protein), a murine protein highly homologous to PACT, associates with heterogeneous nuclear ribonucleoprotein (hnRNP) particles. It was later established that P2P-R is the alternatively spliced form of PACT, lacking the 34 amino acid exon described by Saijo and his colleagues (Saijo *et al.*, 1995), and that P2P-R appears to be the dominant product expressed in multiple murine cell lines (Scott *et al.*, 2003). A P2P-R fusion protein derived from a region of the P2P-R cDNA coding for hnRNP association is able to bind single-stranded DNA (Witte and Scott, 1997). Witte and Scott also showed that P2P-R

expression is markedly repressed during terminal differentiation (Witte and Scott, 1997). In addition, they confirmed that P2P-R binds pRB by precipitating pRB from cellular extracts using GST-P2P-R fusion protein. P2P-R binding to pRB was blocked by E1a protein confirming that P2P-R binds to the pRB pocket domain (Witte and Scott, 1997).

Vo and co-workers (Vo *et al.*, 2001) identified a protein in *Saccharomyces cerevisiae*, which is highly homologous to PACT/P2P-R, which they named Mpe1p. Mpe1p is essential for cell viability and contains a zinc knuckle, a motif that has been implicated in protein-nucleic acid interactions (Berg and Shi, 1996; Laity *et al.*, 2001). The zinc knuckle motif found in Mpe1p is conserved in the human homologue, PACT/P2P-R. Vo and his colleagues demonstrated that Mpe1p is an integral subunit of the *S. cerevisiae* cleavage and polyadenylation factor (CPF) and that Mpe1p is essential for mRNA 3'-end cleavage and polyadenylation. CPF and the poly(A)-binding protein, Pab1p, are responsible for the specific cleavage and polyadenylation of pre-mRNA.

Other important findings concerning P2P-R have since been published. These include the following:

- (i) Perkins and his co-workers (Perkins *et al.*, 1999) identified P2P-R as one of many proteins that contribute to genome stability.
- (ii) P2P-R localises primarily to the nucleoli of interphase murine and human cell lines, consistent with earlier reports that proposed a role for PACT/P2P-R in RNA metabolism (Vo *et al.*, 2001; Gao *et al.*, 2002).

(iii) P2P-R immunoreactivity increases more than ten fold in mitotic cells compared to G₀ cells, without an increase in P2P-R mRNA expression. In these mitotic cells, P2P-R localises to the periphery of chromosomes (Gao *et al.*, 2002).

(iv) Transfection and overexpression of near-full length P2P-R restricts cell cycle progression at prometaphase and promotes mitotic apoptosis in Saos2 cells (Gao and Scott, 2002). Since Saos2 cells lack p53 and have a non-functional pRB (Masuda *et al.*, 1987; Shew *et al.*, 1989; Stott *et al.*, 1998), P2P-R promotes apoptosis in a p53- and pRB-independent manner.

(v) A possible pro-apoptotic region exists within P2P-R from amino acid 1156-1314 (Gao and Scott, 2003). Overexpression of this region in MCF-7 cells promotes camptothecin-induced apoptosis. The potential pro-apoptotic region overlaps with the region of P2P-R that is responsible for p53 and single stranded DNA binding (amino acids 1204-1314, see Table 1.1).



(vi) The SR region of P2P-R can be phosphorylated by the mitotic cdc2 kinase (Scott *et al.*, 2003), which might explain the earlier observation by Gao and co-workers (Gao *et al.*, 2002) that P2P-R immunoreactivity increases significantly during mitosis. P2P-R is also phosphorylated by SRPK1a in the SR region.

(vii) The SR region of P2P-R can bind two factors that associate with matrix associated regions (MARs) of DNA, namely SAF-B and nucleolin (Scott *et al.*, 2003). SAF-B is a MARs binding factor (Renz and Fackelmayer, 1996) which has been reported to couple transcription and pre-RNA splicing and localises to nuclear speckles (Nayler *et al.*, 1998). MARs regions of DNA and associated factors are believed to regulate transcription and other nuclear functions (Glazko *et al.*, 2003).

Table 1.1, adapted from Scott *et al.* (Scott *et al.*, 2003), summarizes the domain organisation and currently known binding factors of the P2P-R protein.

P2P-R protein domains	P2P-R binding factors					
	p53	RB	ssDNA	SAF-B	Nucleolin	SRPK1a
1. Ring-type zinc finger (57-107aa)						
2. Proline rich (362-411aa)						
3. SR-like (460-540aa)				+	+	+
4. RB binding (735-908aa)		+				
5. p53/ssDNA binding (1204-1314aa)	+		+			
6. Lysine rich (1497-1550aa)						

Table 1.1: P2P-R protein domains and associated binding factors.



Examination of expressed sequence tags (ESTs) showed that P2P-R is a truncated form of the full-length RBBP6 protein. The full-length RBBP6 protein (GenBank accession number NP_008841) contains the DWNN domain and a zinc finger domain N-terminal to the published P2P-R sequence. Fig. 1.3 shows the domain organisation of the full-length human RBBP6 protein. Vo and his colleagues (Vo *et al.*, 2001) reported that Mpe1 contains a highly conserved N-terminal domain. Our data confirms that the DWNN domain is highly conserved, occurring in all eukaryotic sequences examined. The DWNN domain occurs once in all eukaryotic genomes, except for the *Arabidopsis thaliana* genome, which encodes two copies of the DWNN domain. Fig. 1.1 shows a multiple alignment of DWNN amino acid sequences from various eukaryotic organisms.

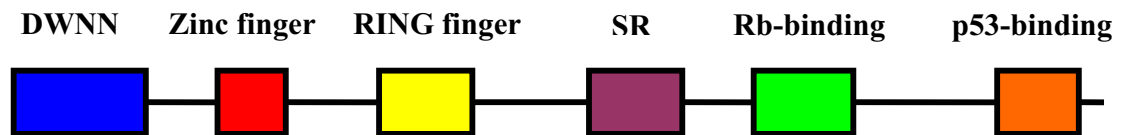


Fig. 1.3: Domain organization of human RBBP6/DWNN-containing protein. Full-length RBBP6 is a 1782 amino acid protein, with the DWNN domain occurring at the N-terminus. The zinc finger has been implicated in protein-nucleic acid interactions and the RING finger domain is found in a number of E3 ubiquitin ligases. SR domains occur in many mRNA splicing factors.

1.3 Ubiquitin-mediated proteolysis

1.3.1 Ubiquitin

Ubiquitin (Fig. 1.3) was first identified as the heat-stable ATP-dependent proteolysis factor 1 (APF-1) in rabbit reticulocytes through amino acid analysis, electrophoretic mobility, and enzymatic assays (Wilkinson *et al.*, 1980). APF-1/ubiquitin is covalently linked to reticulocyte proteins during ATP-dependent proteolysis. Ubiquitin is a highly conserved 76 residue polypeptide found in all eukaryotic cells and tissues. The yeast ubiquitin sequence differs from human ubiquitin by only 3 amino acids (Hershko and Ciechanover, 1998). Ubiquitin is coded as polyubiquitin genes (Wiborg *et al.*, 1985) or as ribosomal fusion proteins (Finley *et al.*, 1989). The polyubiquitin chains and ribosomal ubiquitin fusion proteins are cleaved by deubiquitinating enzymes (DUBs) to free monoubiquitin molecules.



1.3.2 Ubiquitin-mediated proteolysis

Fig. 1.4, adapted from Hershko and Ciechanover (Hershko and Ciechanover, 1998), illustrates the sequential steps in ubiquitin-mediated proteolysis. In step 1, an E1 enzyme activates the C-terminal glycine residue of ubiquitin in a reaction that requires ATP (Haas and Siepmann, 1997). This leads to the formation of an intermediate ubiquitin adenylate and the release of pyrophosphate. In step 2, ubiquitin is transferred to a cysteine residue of the E1 enzyme to form a high-energy thioester bond. The activated ubiquitin is then transferred to a cysteine residue in the active site of an E2 enzyme. In step 3 of the ubiquitin-conjugation cascade, catalysed by an E3 enzyme, ubiquitin is linked via its C-

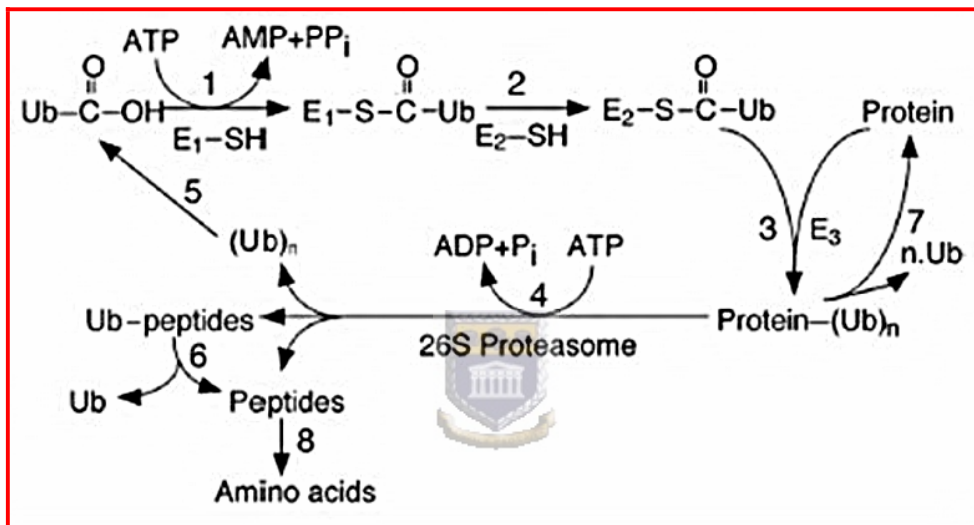


Fig. 1.4: The ubiquitin-proteasome pathway (Hershko and Ciechanover 1998).

terminus to a ϵ -amino group of a lysine residue in the substrate protein to form an amide isopeptide linkage.

The polyubiquitin tagged substrate protein is degraded by the proteasome, which is a 2.5 megadalton multi-protein complex. The proteasome consists of two large complexes, the 19S regulatory particle (RP) and the 20S core particle (CP). The 20S CP consists of four stacked heptameric rings, two outer rings formed by 7 α -subunits each and two inner rings formed by 7 β -subunits each. The 3.4Å crystal structure of the proteasome from the archeobacterium *Thermoplasma acidophilum* has revealed that the catalytically active sites of the CP are in the β -subunits, within a central chamber of the 20S particle (Löwe, *et al.* 1995). All of the β -subunits in archeobacteria are identical, thus giving 14 active sites (Glickman and Ciechanover, 2002). These active sites have a chymotrypsin-like activity, cleaving only after hydrophobic residues (Hochstrasser *et al.*, 1999).

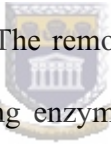
The crystal structure of the yeast 20S proteasome was solved at 2.4Å resolution by Groll and co-workers (Groll *et al.*, 1997). It revealed that the yeast CP is also made up of four stacked rings, containing 7 subunits units each, as in the case of the *Thermoplasma acidophilum* CP. However, the subunits are not identical and the catalytic activity is located in only some of the β -subunits (Voges *et al.*, 1999). Eukaryotic proteasomes also display, in addition to the chymotrypsin-like activity of *T. acidophilum* proteasomes, trypsin-like activity (cleavage after basic residues) and peptidylglutamyl peptide hydrolysing activity (cleavage after acidic residues). Three β -subunits β 1/PRE3, β 2/pup1 and β 5/PRE2, carry the protease active sites. The active sites are N-terminal threonine

residues, generated by cleavage of the β 1, β 2 and β 5 pro-proteins during proteasome maturation. The threonine 1 side chains of these β subunits are used as nucleophiles in a catalytic attack on carbonyl carbons of substrate proteins (Löwe *et al.*, 1995, Hershko and Ciechanover, 1998). The α -subunits play a role in stabilizing the structure of the two β -subunit rings (Hershko and Ciechanover, 1998). In addition, the N-terminal domains of the α -subunits α 1/C7, α 2/Y7, α 3/Y13, α 6/PRE5 and α 7/C1 extend towards each other in the space at the ends of the catalytic inner chamber, forming a network of interacting side chains that obstructs access to the chamber. This suggests that the α -subunits may control access to the catalytic inner chamber of the proteasome (Groll *et al.*, 1997, Groll *et al.*, 2000).

The 19S regulatory particle consists of two subcomplexes, called the lid and the base respectively. The RP is formed by 18 different subunits. Six of these subunits are ATPases, designated Rpt1-6 (regulatory particle triple-A protein 1-6). The ATPase subunits are members of the AAA-ATPase family, which is a diverse group of proteins whose members include translocators, transporters and proteases (Beyer, 1997; Patel and Latterich, 1998; Ogura and Wilkinson, 2001). Some of the Rpt subunits interact directly with the α -ring of the 20S CP and is responsible for opening the catalytic chamber and translocating substrates into the CP (Kohler *et al.*, 2001; Glickman and Ciechanover, 2002). The non-ATPase subunits of the RP are designated Rpn1-12. The Rpt subunits, together with Rpn1 and Rpn2, form the base of the RP, while the remaining Rpn subunits form the lid (Glickman *et al.*, 1998). Rpn10 resides at the interface of the base and lid

subcomplexes, stabilizing the interaction between the two subcomplexes (Glickman and Ciechanover, 2002).

Rpn10 and its human homologue S5a can bind polyubiquitin chains (Deveraux *et al.*, 1994, Nocker *et al.*, 1996, Kominami *et al.*, 1997, Wilkinson *et al.*, 2000). Two additional 19S base subunits, Rpn1 and Rpn2, have been shown to bind the ubiquitin-like (Ubl) domains of Rad23 and Dsk2 (Saeki *et al.*, 2002). These yeast proteins have N-terminal Ubl domains, as well as C-terminal ubiquitin-associated (UBA) domains. UBA domains can bind ubiquitin, polyubiquitin chains and ubiquitinated cellular proteins (Bertolaet *et al.*, 2001; Wilkinson *et al.*, 2001).

Ubiquitination is a reversible process.  The removal of polyubiquitin chains from target proteins is catalyzed by deubiquitinating enzymes (DUBs). DUBs belong to a diverse family of enzymes and play a role in the generation of ubiquitin, recycling of ubiquitin and editing of polyubiquitin chains (Glickman and Ciechanover, 2002). Kim and co-workers (Kim *et al.*, 2003) have grouped deubiquitinating enzymes into four distinct families: (i) ubiquitin C-terminal hydrolase (UCH) family, (ii) ubiquitin-specific processing protease (UBP) family, (iii) OTU-domain ubiquitin-aldehyde-binding protein (Otubain) family and (iv) Jab1/Pad1/MPN-domain metallo-enzyme (JAMM) family. UCHs are small enzymes of 20-30 kD and are classified as cysteine proteases. Members of the UCH family remove short peptides and small molecules from the carboxyl terminus of ubiquitin (Glickman and Ciechanover, 2002; Kim *et al.*, 2003). UBP family members are also cysteine proteases and vary in size from 50-300 kD. UBPs can

hydrolyse the isopeptide bonds that link ubiquitin moieties in polyubiquitin chains as well as those between ubiquitin and target proteins to generate free ubiquitin. The Otubain family of DUBs, recently discovered by Balakirev and his colleagues (Balakirev *et al.*, 2003) and Evans and co-workers (Evans *et al.*, 2003) belongs to the ovarian tumour superfamily of proteins, which are cysteine proteases (King and Storto, 1988). Otubain family members can cleave isopeptide linkages in polyubiquitin chains (Kim *et al.*, 2003). Eytan and co-workers (Eytan *et al.*, 1993) identified a deubiquitinating activity that is associated with the 26S proteasome. Verma and co-workers (Verma *et al.*, 2002) and Yao and Cohen (Yao and Cohen, 2002) suggested that Rpn11, a subunit of the 19S regulatory particle of the proteasome, is responsible for this DUB activity. Rpn11 contains a conserved motif, EX_nHXHX₁₀D, dubbed the JAMM motif, which coordinates metal ions in the active sites of hydrolytic enzymes (Hochstrasser, 2002). Rpn11 appears to be responsible for the release of polyubiquitin chains from target proteins, prior to translocation and degradation of the target by the proteasome.

Many cellular proteins are degraded by the ubiquitin-proteasome pathway. The ever-growing list of ubiquitin substrates include cell-cycle regulators such as cyclins and cyclin-dependent kinases, as well as transcription factors, tumour suppressors and oncoproteins like NF- κ B, p53, c-Jun and E2F-1 (Hershko and Ciechanover, 1998). Ubiquitin-mediated proteolysis serves to regulate the levels of these proteins to ensure that cells progress through the cell cycle. Ubiquitination has also been implicated in the degradation of cell surface proteins such as the growth hormone receptor and the platelet-derived growth factor receptor, as well as transporters such the uracil and maltose

permeases in yeast (Hershko and Ciechanover, 1998; Glickman and Ciechanover, 2002). In these cases ubiquitination serves to internalize substrate proteins, which are then degraded by lysosomes or vacuoles in yeast. These examples illustrate the important and varied roles that ubiquitination of target proteins plays in ensuring cell stability and smooth progression through the cell cycle.

1.4 Ubiquitin-like proteins

Ubiquitin-like proteins (UBLs) are a diverse group of proteins, some of which bear very little sequence similarity to ubiquitin. Some of these proteins can be covalently attached to other proteins in a manner that resembles ubiquitin conjugation. UBLs are grouped into two classes. Class 1 UBLs exist as free-standing molecules of 76-103 amino acids, called ubiquitin-like modifiers, which can be post-translationally conjugated to other proteins (Jentsch and Pyrowolakis, 2000; Glickman and Ciechanover, 2002). Members of this class of UBLs include SUMO/Sentrin/Smt3, NEDD8/Rub1, HUB1 and ISG15/UCRP. ISG15 (interferon-stimulated gene 15), the first UBL identified, contains two tandem UBL domains (Haas *et al.*, 1987). Analogous to ubiquitin, most of the members of this class have a di-glycine C-terminal motif, which is necessary for recognition and activation of the UBL (Glickman and Ciechanover, 2002). The second class of UBLs are fusion proteins that bear domains resembling ubiquitin but show no further resemblance to each other. Proteins of this class are referred to as ubiquitin-domain proteins (UDPs). UDPs are not conjugated to other proteins (Jentsch and Pyrowolakis, 2000). UDP members include Rad23, Dsk2, PLIC-1, Parkin and BAG1.

1.4.1 SUMO-1

SUMO (small ubiquitin-like modifier), the most well-studied UBL, occurs in all eukaryotic kingdoms. Members of the SUMO gene family have been reported in protozoa, metazoa, plants and fungi, as well as *Drosophila melanogaster* and *Caenorhabditis elegans* (Choudbury and Li, 1997; Katami *et al.*, 1998; Huang *et al.*, 1998; Hanania *et al.*, 1999; Melchior, 2000) The human genome encodes three functional SUMO proteins, designated SUMO-1/2/3 (Su and Li, 2002). In addition, the human genome also contains eight SUMO-1 pseudogenes and 23 SUMO-2 pseudogenes, but no SUMO-3 pseudogenes. SUMO-1 shares 44% sequence identity with SUMO-2 and SUMO-3 proteins while SUMO-2 and SUMO-3 are 86% identical. Human SUMO-1 is a 101 residue protein that shares 18% sequence identity with ubiquitin.



The three dimensional structure of SUMO-1 has been solved by NMR and found to resemble ubiquitin closely (Bayer *et al.*, 1998; Jin *et al.*, 2001). SUMO-1 folds into a conformation that contains five β -strands and two α -helices (amino acids 22-97). Unlike ubiquitin, SUMO-1 has a long and highly flexible N-terminus (amino acids 1-21) which extends from the ubiquitin-like core. Lys48 of ubiquitin, which is required for polyubiquitin chain formation, is replaced in SUMO-1 by Gln69, which may provide an explanation for the observation that SUMO-1 does not form polymers like ubiquitin (Bayer *et al.*, 1998).

“Sumoylation”, the conjugation of SUMO to target proteins, proceeds in a similar fashion to ubiquitination. The E1 for SUMO-1 activation in yeast and humans has been identified

as the heterodimeric protein, AOS1/UBA2 (Johnson *et al.*, 1997; Gong *et al.*, 1999; Desterro *et al.*, 1999; Okuma *et al.*, 1999). AOS1 bears sequence similarity to the amino-terminal region of the ubiquitin E1, UBA1, while UBA2 shows similarity to the C-terminal region of UBA1 (Yeh *et al.*, 2000; Muller *et al.*, 2001). Even though the UBA2 subunit of the dimer carries the active site cysteine required for SUMO activation, both subunits are required for the E1 activity. In addition to AOS1/UBA2, SUMO conjugation requires the E2 enzyme Ubc9 (Johnson and Hochstrasser, 1997). Ubc9 is specific for SUMO and does not act on ubiquitin (Desterro *et al.*, 1997). In an *in vitro* system, AOS1/UBA2 and Ubc9 was sufficient for sumoylation of RanGap1, a known target of the SUMO conjugation system (Okuma *et al.*, 1999). Thus no E3 was required for sumoylation of RanGap1. A minimal sumoylation motif, ψ KxE/D, has been identified in RanGap1, where ψ is a bulky aliphatic residue and K is the SUMO acceptor site (Melchior *et al.*, 2003). In RanGap1 the sumoylation motif forms part of an exposed loop that reaches into the catalytic cleft of Ubc9 (Bernier-Villamore *et al.*, 2002). Residues flanking this motif in RanGap1 stabilize the interaction with SUMO, while additional residues at the C-terminus of RanGap1 also interact with Ubc9.

Other SUMO targets require SUMO E3 ligases for conjugation. Currently, three types of E3's have been identified for SUMO (Melchior, 2003). These are the PIAS (protein inhibitor of activated signal transducer and activator of transcription) family, RanBP2 and Pc2. PIAS family members contain a RING-finger-like structure and can bind directly to Ubc9 and some SUMO target proteins to stimulate their sumoylation. RanBP2 is a nuclear pore complex protein, which contains a 30 kD domain with E3 catalytic activity,

even though the domain bears no relation to RING fingers or other E3 ligases. Pc2 is a Polycomb group member. Polycomb group proteins are involved in gene silencing.

Like ubiquitination, sumoylation is reversible. De-sumoylation is catalysed by ULPs (ubiquitin-like specific proteases). Two ULPs, Ulp1 and Ulp2/Smt4, were identified in yeast (Li and Hochstrasser, 1999; Schwienhorst *et al.*, 2000). Ulp1 and Ulp2 share sequence similarity in a region of 200 amino acids at the C-terminus, called the ULP domain, which contains a Cys-His-Asn catalytic triad (Muller *et al.*, 2001; Kim *et al.*, 2002; Melchior *et al.*, 2003). Ulp1 is essential for cell viability in yeast and can generate the mature form of Smt3 as well as cleave the isopeptide linkage between Smt3 and its targets (Li and Hochstrasser, 1999). Unlike Ulp1, Ulp2 is not essential for cell viability, but Ulp2 deficient mutants accumulate Smt3 conjugates and show defects such as slow and temperature-sensitive growth, chromosome instability, sporulation defects and defects in cell-cycle progression (Li and Hochstrasser, 2000; Schwienhorst *et al.*, 2000). Seven mammalian ULPs have been identified, ranging in size from 238 to 1112 amino acids (Yeh *et al.*, 2000). Human ULPs share the C-terminal conserved region of 200 residues found in yeast. Mammalian ULPs show a diverse cellular localization including the nucleolus, nucleoplasm, nuclear bodies, the nuclear envelope, the nuclear pore complex, the cytoplasm and cytoplasmic vesicles (Melchior *et al.*, 2003).

RanGap1 was the first cellular target identified for sumoylation (Matunis *et al.*, 1996; Mahajan *et al.*, 1997). RanGap1 is a regulator of the Ras-related GTPase Ran, which controls transportation between the cytoplasm and the nucleus (Melchior *et al.*, 1993;

Gorlich and Kutay, 1999). RanGap1 associates with RanBP2 at the cytoplasmic side of the nuclear pore complex. Sumoylation of RanGap1 occurs between Gly-97 of SUMO-1 and Lys-526 of RanGap1, which induces a conformational change in RanGap1 that allows it to associate with and target RanBP2 to the nuclear pore complex (Mahajan *et al.*, 1997; Jentsch and Pyrowolakis, 2000; Muller *et al.*, 2001; Kim *et al.*, 2002; Glickman and Ciechanover, 2002). RanBP2 has recently been identified as an E3 ligase for SUMO (Pichler *et al.*, 2002).

Most of the known SUMO targets are nuclear proteins. Sumoylation of the tumour suppressor, p53, activates the transcriptional activity of p53 (Gostissa *et al.*, 1999; Rodriguez *et al.*, 1999; Muller *et al.*, 2000). PML (promyelocytic leukaemia protein) undergoes sumoylation at three different lysine residues (Muller *et al.*, 2001). Sumoylated PML is targeted to nuclear bodies called PML bodies. Sumoylation of PML appears to regulate the assembly and stability of these bodies, since PML mutants lacking SUMO- modification sites fail to target to nuclear bodies (Sternsdorf *et al.*, 1997; Muller *et al.*, 1998; Kamitani *et al.*, 1998; Jackson, 2001). NF- κ B, a transcription factor that activates genes involved in numerous cellular processes including cell adhesion, cell growth, inflammatory responses and immune function, is also regulated by sumoylation (Desterro *et al.*, 1998). NF- κ B is kept in inactive complexes in the cytoplasm by association with inhibitory proteins called I κ Bs. SUMO-1 exerts its effects on NF- κ B transcriptional activity by sumoylating I κ B α . Polyubiquitination of I κ B α at Lys21 or Lys22 results in the degradation of I κ B α by the proteasome, releasing NF- κ B, which then translocates to the nucleus to activate transcription. SUMO-1 works antagonistically to

ubiquitin through conjugation of I κ B α at Lys21, thereby preventing ubiquitin-mediated degradation of I κ B α ; thus preventing NF- κ B transcriptional activity. Sumoylation of p53, PML, RanGap1 and I κ B α illustrates the role that SUMO and other UBLs plays in diverse cellular processes.

1.4.2 Ubiquitin domain proteins

Ubiquitin domain proteins are fusion proteins containing domains which resemble ubiquitin. Unlike ubiquitin and free-standing UBLs like SUMO, UDPs are not conjugated to target proteins. Rather it appears that the ubiquitin-like domains of these proteins provide a functional link with the proteasome (Glickman and Ciechanover, 2002). Rad23 is an example of the UDP class of UBLs. Rad23 was originally identified in yeast as a nucleotide excision repair (NER) protein that associates with Rad4 to form a dimer that preferentially binds to damaged DNA (Gudzer *et al.*, 1998; Jansen *et al.*, 1998). The factors constituting the NER pathway has been identified and includes the transcription factor complex TFIIH, xeroderma pigmentosum group proteins and replication protein A (Gudzer *et al.*, 1995b; de Laat *et al.*, 1999). Rad23 promotes the formation of the NER complex (Gudzer *et al.*, 1995a). Watkins and co-workers (Watkins *et al.*, 1993) reported that Rad23 contain an N-terminal ubiquitin-like domain and Schaubert and colleagues (Schauber *et al.*, 1998) subsequently showed that Rad23 interacts with the proteasome through its UBL domain. Two 19S base subunits, Rpn1 and Rpn2, have been shown to bind the ubiquitin-like domains of Rad23 and Dsk2, another yeast UDP (Saeki *et al.*, 2002). Recently, a potential proteasome-interacting motif (PIM) was identified in the UBL domain of Rad23, parkin and other UDPs (Upadhya and Hegde, 2003).

Phylogenetic analysis revealed that PIM-containing UDPs are evolutionarily more closely related to ubiquitin than UDPs lacking the PIM motif. Furthermore, a missense mutation associated with Parkinson's disease occurs in a PIM residue in parkin, suggesting that the PIM motif is functionally significant. In addition to the UBL domain, Rad23 also contains two C-terminal ubiquitin associated (UBA) domains that can bind polyubiquitin chains (Wilkinson *et al.*, 2001). The UBA domains of Rad23 have been shown to inhibit multi-ubiquitin chain assembly (Ortolan *et al.*, 2000; Chen *et al.*, 2001). Chen and Madura (Chen and Madura, 2002) also reported that Rad23 promotes the targeting of proteolytic substrates to the proteasome; both the UBL domain and the UBA domains were required for this function. Sweder and Madura (Sweder and Madura, 2002) proposed a model in which Rad23 regulates the activity of the nucleotide excision repair complex. They propose that in the absence of DNA damage some NER factors, specifically Rad4, since it has been identified as a target of the proteasome, are degraded by the proteasome. The degradation of these factors is mediated by Rad23 and inhibited in the presence of DNA damage. Thus Rad23 not only forms an integral part of the NER complex but is also involved in the regulation of the DNA damage response.

Another UDP, BAG1 (Bcl2-associated athanogene-1), has been shown to modulate the molecular chaperone function of Hsp70. Members of the Hsp70 family are involved in protein folding, translocation and degradation (Bukau and Horwich, 1998). Bag1 has been identified as a regulator of cell signalling molecules like Raf1, Bcl2, Siah1A and growth and steroid receptors (Nollen *et al.*, 2000). At least four different isoforms of BAG1 have been identified in humans (Yang *et al.*, 1998; Takayama *et al.*, 1998). All

BAG1 isoforms have a UBL domain near their N-termini (Takayama *et al.*, 1995). Nollen and co-workers (Nollen *et al.*, 2000) showed that BAG1 associates with the ATPase domain of Hsp70 and inhibits Hsp70 chaperone activity. In this regard BAG1 opposes Hip, an Hsp70 cofactor that also binds to the ATPase domain of Hsp70 and promotes the protein folding activity of the chaperone. Lüders and co-workers (Lüders *et al.*, 2000) showed that BAG1 associates with the proteasome in an ATP dependent manner and that the association of BAG1 with the proteasome was dependent on its N-terminal UBL domain. Hartmann-Petersen and co-workers (Hartmann-Petersen *et al.*, 2003) proposed that Hsp70 chaperones could associate with the proteasome via the UBL of BAG1 to degrade misfolded proteins.



1.5 Aims

The objectives of this study were two-fold. Firstly, to heterologously express the DWNN domain in a bacterial expression system to produce samples suitable for heteronuclear NMR analysis, for which ^{15}N - and ^{13}C -labeled samples of more than 0.5 mM concentration is required. Secondly, to fully assign the backbone resonances using triple resonance NMR, which is an essential prerequisite for structure determination. The full structure determination is, however beyond the scope of this thesis.

Chapter 2: Materials and methods

All chemicals were from Promega, Sigma, Roche, BDH and Amersham, unless otherwise stated.

2.1 General stock solutions and buffers

10X DNase I buffer: 0.5 M Tris-Cl pH 7.5, 0.1 M MnCl_2 , 0.5 mg/ml BSA. This buffer was filter sterilised before storage at -20°C . The buffer was diluted 10 times with deionised water before use.

2X SDS gel sample buffer: 4% SDS, 0.125 M Tris pH 6.8, 15% glycerol and 1 mg/ml bromophenol blue. This buffer was stored at room temperature. 200 mM DTT was added to the buffer immediately prior to use.

5x SDS electrophoresis buffer: 25 mM Tris, 0.1% SDS and 250 mM glycine, pH 8.3.

Bradford assay dye stock: 100 mg Coomassie brilliant blue G 250, 8.5% phosphoric acid and 4.75% ethanol. The solution was kept at 4°C . The solution was diluted 5 times with deionised water before use.

Cell lysis buffer: 10 $\mu\text{g/ml}$ DNase I, 100 $\mu\text{g/ml}$ lysozyme, 1x CompleteTM EDTA free protease inhibitor cocktail in PBS.

Coomassie staining solution: 0.25 g Coomassie Blue R-250, 45% methanol, 10% acetic acid.

Destaining solution: 45% methanol, 10% acetic acid.

DNase free RNase buffer: 0.1 M sodium acetate anhydrous, 0.1 mM EDTA, pH 4.8.

This buffer was stored at -20°C.

DTT: A 1 M stock solution was prepared in 0.01 M sodium acetate, pH 5.2. The solution was sterilised by filtration and stored in 1 ml aliquots at -20°C.

IPTG: A 0.1 M stock solution was prepared in deionised water. The solution was sterilised by filtration and stored at 4°C.

Luria agar: 10 g/l tryptone powder, 5 g/l yeast extract, 5 g/l NaCl and 14 g/l bacteriological agar.

Luria broth: 10 g/l tryptone powder, 5 g/l yeast extract, 5 g/l NaCl and 2 g/l glucose

Minimal media: 12.8 g/l Na₂HPO₄·7H₂O, 3 g/l KH₂PO₄, 0.5 g/l NaCl and 1 g/l NH₄Cl.

The pH of this solution was adjusted to 7.4 with NaOH and the solution was autoclaved.

The solution was then supplemented with 2 ml of filter-sterilised 1 M MgSO₄, 0.1 ml of filter-sterilised 1 M CaCl₂ and 10 ml of filter-sterilised 20% glucose. The solution was stored at room temperature.

NMR buffer: 100 mM sodium phosphate buffer pH 6.0, 150 mM NaCl, 1 mM DTT.

PBS: 137 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄ and 1.5 mM KH₂PO₄. The pH was adjusted to 7.4 with HCl and the solution sterilised by autoclaving.

PMSF: A 10 mM stock solution was prepared in isopropanol. The solution was stored at -20°C.

PreScissionTM Protease cleavage buffer: 50 mM Tris-HCl, pH 7.0, 150 mM NaCl, 1 mM DTT, 1 mM EDTA.

Protein elution buffer: 10 mM glutathione and 50 mM Tris, pH 8.0

Separating buffer: 1.5 M Tris, pH 8.8. This solution was stored at 4°C.

Stacking buffer: 0.5 M Tris pH 6.8. This solution was stored at 4°C.

2.2 Materials and methods

2.2.1 Preparation of competent E.coli BL21 (DE) pLysS cells

Escherichia coli BL21 (DE) pLysS cells were streaked out on a Luria agar plate. The cells were incubated at 37°C overnight. A single colony was picked and used to inoculate 20 ml TYM broth. The culture was incubated at 37 °C until it reached an optical density at 550 nm of 0.2. The cells were then transferred to 100 ml of fresh TYM broth and incubated at 37 °C until the optical density at 550 nm again reached 0.2. 400 ml of fresh TYM broth was added to the flask and the culture incubated until the optical density at 550 nm reached 0.5.



The cells were rapidly cooled by gently swirling the flask in ice water and then transferred to two 250 ml polypropylene tubes and centrifuged at 6000 rpm for 10 minutes. The media was discarded and the pellets resuspended in 100 ml of transformation buffer 1 and kept on ice for 30 minutes. This was followed by centrifugation at 3000 rpm for 8 minutes at 4°C. The cell pellet was then gently resuspended in 30 ml of transformation buffer 2. 0.3 ml aliquots were transferred into microfuge tubes and snap-frozen using liquid nitrogen. Competent cells were stored at –80°C.

2.2.2 Small-scale preparation of plasmid DNA

E. coli MC106 cells transformed with the pGEX-6P-2-DWNN plasmid (see Section 2.2.3.1) were streaked out on Luria agar plates containing 100 µg/ml of ampicillin. Single colonies were picked and used to inoculate 10 ml of Luria broth containing 100 µg/ml of ampicillin and grown overnight at 37⁰C. The overnight cultures were centrifuged at 9000 rpm for 10 minutes. The supernatant was discarded and the tubes placed upside down on paper towel to drain any residual supernatant, where after the pellets were resuspended in 200 µl ice-cold GTE. The suspensions were transferred to 1.5 ml microfuge tubes and incubated for 5 minutes at room temperature, before 400 µl of lysis solution (0.2N NaOH/1% SDS) was added. The samples were mixed with the lysis solution by gentle finger tapping and incubated at room temperature for 5 minutes. Following this incubation, 300 µl of neutralising solution (potassium acetate) was added to the samples and the samples were mixed with the neutralising solution by inverting the tubes several times. The samples were then centrifuged at 13 200 rpm for 15 minutes. 800 µl of the supernatant was transferred to new 1.5 microfuge tubes and 600 µl isopropanol was added to precipitate nucleic acids. The samples were incubated for 30 minutes at -20⁰C to facilitate precipitation. The samples were then centrifuged at 13 200 rpm for 10 minutes. The supernatant was discarded and the pellets washed by adding 500 µl of 70% ethanol, vortexing briefly and centrifuging at 13 200 rpm for 5 minutes. The supernatant was removed and the pellets dried at 37⁰C. The dried pellets were resuspended in 500 µl of 1 x TE.

To remove contaminating RNA, the samples were treated with RNase A. 5 µl of 20 mg/ml RNase A was added to 500 µl of the DNA samples. The samples were incubated in a waterbath at 37⁰C for 1 hour. The DNA was then extracted by phenol: chloroform (1:1) treatment. 800 µl of phenol: chloroform was added to the samples. The samples were then vortexed and centrifuged at 13 200 rpm for 10 minutes. 400 µl of the top aqueous phase was removed and transferred to 2 ml microfuge tubes.

To precipitate the DNA, 200 µl of 7.5 M ammonium acetate and 1 ml of absolute ethanol were added to the samples. The samples were then incubated at -20⁰C for 1 hour. The DNA was pelleted by centrifugation at 10 000 rpm for 10 minutes. The supernatant was removed and the pellets washed with 70% ethanol. 500 µl of 70% ethanol was added to the pellets, followed by centrifuging at 13 200 rpm for 5 minutes. The ethanol was removed and the pellet dried at 37⁰C. The dried pellets were resuspended in 50 µl of 1 x TE. The DNA was stored at 4⁰C until required.

2.2.3 Expression of recombinant DWNN protein

2.2.3.1 Construction of the pGEX-6P-2-DWNN plasmid

The pGEX-6P-2-DWNN plasmid was constructed by the cloning of base pairs 1 to 243 from the human DWNN domain nucleotide sequence into the pGEX-6P-2 expression vector (PT. Lutya, MSc thesis, University of the Western Cape, 2003). After removal of the N-terminal GST fusion tag using 3C protease, the expected sequence of the protein was:

GPLGS^MSCVHYKFSSKLN^YDTVTFDGLHISLCDLKKQIMG
REKLKAADC^DLQITNAQTKEEY^TDDNALIPKNSSVIVRRIP
IGGVK

Methionine-6 (highlighted in red) represents the initiation methionine of the native DWNN domain. Residues 1 to 5 (GPLGS) originate from the recognition footprint of the 3C protease in the linker region between the GST tag and the DWNN domain. All sequence numbers in this thesis are recorded relative to the above sequence.

2.2.3.2 Transformation of *E.coli* BL21 (DE) pLysS with pGEX-6P-2-DWNN

Aliquots of competent *E.coli* BL21 (DE) pLysS cells were thawed on ice and 100 µl of cells added to 2 µl of pGEX-6P-2-DWNN DNA in a pre-cooled eppendorf tube. The transformation mixture was kept on ice for 30 minutes and then placed in a 37⁰C waterbath for 5 minutes. The mixture was placed on ice for 2 minutes before 0.9 ml of pre-warmed LB was added. The mixture was then placed on a shaker at 37⁰C and incubated for 1 hour. 200 µl of the transformation mixture was plated on a LB plate that containing 100 µg/ml ampicillin. The plated cells were incubated overnight at 37⁰C.

2.2.3.3 Screening of *E.coli* BL21 (DE) pLysS transformed with pGEX-6P-2-DWNN for recombinant protein expression

E.coli BL21 (DE) pLysS cells transformed with pGEX-6P-2-DWNN were plated on an LB-amp plate and incubated overnight at 37⁰C (see Section 2.2.4). Single colonies was picked and used to inoculate 5 ml of LB that contained 100 µg/ml ampicillin. The

cultures were grown at 37⁰C for 4 hours with shaking. 1 ml of each culture was transferred to 1.5 ml microfuge tubes and placed on ice. These would serve as uninduced samples. A further 1 ml of each culture was transferred to 1.5 ml microfuge tubes and isopropyl β -D-thiogalactopyranoside (IPTG) was added to each tube to a final concentration of 0.3 mM. These cultures were grown further at 37⁰C for 2 hours.

Both induced and uninduced samples were pelleted by centrifugation at 13 200 rpm in a bench top centrifuge. The supernatant was discarded and the pellets resuspended in 0.1 ml phosphate buffered saline (PBS). 20 μ l of each cell suspension was taken and 20 μ l of 2x sodium dodecyl sulphate (SDS) gel sample buffer added to it. The samples were boiled at 95⁰C for 5 minutes. SDS polyacrylamide gel electrophoresis (PAGE) was performed on uninduced and induced samples. The remaining 3 ml of strongly expressing colonies was used for large-scale recombinant protein expression and to make glycerol stocks.

2.2.3.4 Large scale expression of recombinant DWNN protein

E.coli BL21 (DE) pLysS competent cells transformed with pGEX-6P-2-DWNN were grown overnight on LB-amp plates at 37⁰C. Transformants were screened for recombinant protein expression (Section 2.2.3.2). 10 μ l of cultures showing good expression of recombinant protein in the expression screen were used to inoculate 100 ml of LB containing 100 μ g/ml ampicillin and 50 μ g/ml chloramphenicol. This culture was grown overnight at 37⁰C on a rotary shaker. The culture was scaled up to 1 litre by the addition of 900 ml of LB medium containing 100 μ g/ml ampicillin and grown at 37⁰C

until the culture reached an optical density at 550 nm of 0.5. Recombinant protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM. Induction was carried out overnight at 30⁰C.

2.2.3.5 Double labelling of recombinant protein

For heteronuclear (¹⁵N and ¹³C) labelling of recombinant protein, fusion protein was expressed in minimal medium (12.8 g/l Na₂HPO₄-7H₂O, 3 g/l KH₂PO₄, 0.5 g/l NaCl and 2 ml 1M MgSO₄, 0.1 ml 1M CaCl₂) containing 1g/l ¹⁵N-H₄Cl (Spectra Gas) and 2g/l ¹³C-glucose (Spectra Gas) as the sole nitrogen and carbon sources respectively. The recombinant protein was either unlabelled (natural abundance), singly labelled (¹⁵N only) or doubly labelled (¹⁵N and ¹³C). The expression of labelled protein was performed as described in Section 2.2.3.3



2.2.4 Recombinant protein extraction and purification

2.2.4.1 Cell lysis and protein extraction

After overnight induction of recombinant protein expression, cultures were transferred to 250 ml polypropylene tubes and pelleted by centrifugation at 6000 rpm at 4⁰C. The supernatant was discarded and the pellet stored at -70⁰C until required.

Cell pellets were resuspended in 20 ml lysis buffer (PBS supplemented with 100 µg/ml lysozyme, 20 µg/ml DNase I, 1x CompleteTM EDTA-free protease inhibitor cocktail (Roche), 1 mM EDTA (ethylenediaminetetraacetate) and 0.5% Triton-X 100). Cells were lysed by repeated freeze-thawing at -70⁰C and 37⁰C and cell debris was pelleted by

centrifugation at 10 000 rpm at 4⁰C. The lysate was transferred to a 50 ml tube and stored at 4⁰C.

2.2.4.2 Purification of recombinant protein

Recombinant protein was purified by affinity chromatography on a glutathione agarose (Sigma) column, prepared according to the manufacturer's instructions. Prior to loading the lysate, the column was equilibrated with 5 column volumes of PBS. The lysate was then passed over the column and the flow through collected. The column was then washed with 5 column volumes of PBS. Recombinant protein was eluted from the column with 15 ml of elution buffer (50 mM Tris, 10 mM glutathione). The column was then washed with 5 column volumes of PBS, followed by 3 column volumes of 1 M NaCl. 1 mM PMSF, 1 mM EDTA, 0.02% sodium azide and 1 mM dithiothreitol (DTT) were added to the eluted protein before storage at 4⁰C.

2.2.4.3 PreScission Protease cleavage of recombinant protein and recovery of the DWNN domain

The recombinant protein was cleaved with PreScission[™] Protease (Amersham Pharmacia Biotech) to release the DWNN domain. 12 units (6 µl) of protease were added to 15 ml of recombinant protein in a dialysis bag (molecular weight cut off (MWCO) 3500 Da). The dialysis bag was placed in cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA and 1 mM DTT, pH 7.0). Cleavage was done overnight at 4⁰C. Following cleavage, the DWNN domain was recovered using a glutathione agarose column as described in Section 2.2.4.2.

2.2.4.4 Removal of the GST domain

The DWNN domain was purified further on a BioCad Sprint Perfusion Chromatography system (Perkin Elmer) using a GSTPrep™ FF 16/10 column (Amersham Pharmacia Biotech). The column was equilibrated with 50 mM Tris. The protein was loaded and wash buffer (50 mM Tris) run through at a flow rate of 5 ml per minute. The DWNN domain was not retained by the column and remained in the flow through. The GST protein was eluted with elution buffer (50 mM Tris, 15 mM glutathione).

The DWNN domain was concentrated to a final volume of about 1ml using YM-3 Centriprep concentrators (MWCO 3500 Da, Millipore) and loaded onto a Sephacryl S100 gel filtration column (Amersham Pharmacia Biotech) which had been equilibrated with NMR buffer (100 mM sodium phosphate buffer pH 6.0, 150 mM NaCl, 1 mM DTT). Fractions were collected and subjected to SDS-PAGE analysis. DWNN containing fractions were pooled and concentrated to a final volume of 600 µl using Centriprep concentrators. Concentrated protein was lyophilised and stored at 4⁰C.

2.3 SDS-PAGE

Protein samples were separated on SDS-PAGE gels according to Laemmli's method (Laemmli, 1970). 16% SDS-PAGE separating gels were prepared from a 40% bis-acrylamide: acrylamide stock as follows: 1.286 ml distilled H₂O, 1.6 ml 40% bis-acrylamide: acrylamide, 1.05 ml 1.5 M Tris pH 8.8, 42 µl 10% SDS, 20 µl ammonium persulphate (APS) and 2 µl TEMED (N, N, N', N'-tetramethylethylenediamine). 4% stacking gels, made from 1.268 ml distilled H₂O, 0.2 ml 40% bis-acrylamide: acrylamide,

0.5 ml 0.5 M Tris pH 6.8, 10 μ l APS and 2 μ l TEMED, was poured on top of the separating gels to a depth of about 1.5 cm.

Protein samples were prepared by adding an equal volume of 2x SDS sample buffer to the samples. The samples were vortexed briefly and then boiled at 95⁰C for 5 minutes. 20 μ l of each sample was loaded on the SDS-PAGE gels, unless otherwise stated. The samples were electrophoresed at 150V. Gels were incubated in Coomassie staining solution (90 ml 50% methanol, 10 ml glacial acetic acid and 0.25g Brilliant Blue R-250 (Sigma)) for about an hour on an orbital shaker. The stained gels were destained overnight in destaining solution (90 ml 10% methanol, 10 ml glacial acetic acid). Alternatively, SDS-PAGE gels were silver stained using Blum's method (Blum *et al.*, 1987).



2.4 Bradford assay

Protein concentrations were determined by Bradford assay (Bradford, 1976), modified for microtiter plates. Bradford dye stock (0.01% Coomassie brilliant blue G250, 8.5% phosphoric acid and 4.75% ethanol) was diluted five times with deionised water and 0.05N NaOH added. Serial dilutions, in the range 6.25 μ g/ml to 0.1 mg/ml, were made of lysozyme, which was used as a standard protein. Serial dilutions were also made of DWNN domain protein preparations. 20 μ l aliquots of the protein standards, as well as diluted DWNN preparations were added to individual wells of 96 well plates and 180 μ l of Bradford assay reagent were added to the samples. 100 mM phosphate buffer, containing 150 mM NaCl and 1 mM DTT was used as a blank. After the addition of Bradford assay

reagent, the samples were incubated at room temperature for 5 minutes. Absorbance was measured at 620 nm. Microsoft Excel (Microsoft Corporation) was used to draw standard curves and calculate DWNN domain sample concentrations.

2.5 NMR analysis and structure determination

2.5.1 Sample preparation

Purified DWNN domain preparations were concentrated and lyophilised as described in Section 2.2.4.4. Lyophilised samples were dissolved in 600 μ l distilled water. 10% deuterium oxide (D_2O) was added to these samples to act as a lock signal.



2.5.2 NMR experiments and data processing

^{15}N -HSQC-NOESY and ^{15}N -HMQC-J experiments were carried out at Oxford University on hybrid Oxford/General Electric 600 MHz and 500 MHz spectrometers (Oxford Instruments) respectively. All other NMR experiments were performed using 600 MHz Varian Inova spectrometers (Varian, Inc.) at the University of Stellenbosch and at the Varian Applications Laboratory in Darmstadt, Germany. 5 mm NMR tubes were used for all experiments. Spectra were processed using NMRPipe (Delaglio *et al.*, 1995), while NMRView (Johnson and Blevins, 1994) was used to visualise and analyse spectra. Table 1 lists the NMR experiments performed. All NMR experiments were performed at 25 $^{\circ}C$, using 100 mM sodium phosphate buffer pH 6.0, 150 mM NaCl and 1 mM DTT as the NMR sample buffer.

Experiment	Spectrometer field strength	Purpose
1D ^1H homonuclear	600MHz	Evaluation of sample integrity and state of folding
^{15}N -HSQC	600 MHz	Evaluation of sample integrity and state of folding; ^1H and ^{15}N chemical shifts
CBCA(CO)NH	600 MHz	Sequential assignment of backbone resonances; $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts
CBCANH	600 MHz	Sequential assignment of backbone resonances; $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts
HNCO	600 MHz	Carbonyl carbon chemical shifts
^{15}N -HSQC-NOESY	600 MHz	$\text{H}^{\text{N}}\text{-H}^{\text{N}}$ NOE's
HMQC-J	500 MHz	$^3\text{J}_{\text{H}^{\text{N}}-\text{H}^{\alpha}}$ coupling constants

Table 2.1: NMR experiments performed on the DWNN domain

Chapter 3: Expression and purification of the DWNN domain

3.1 Introduction

Nuclear magnetic resonance (NMR) has been described as “the unrivalled method for answering biological questions at the atomic level in the solution state” (Primrose, W.U, 1993). It has allowed the study of biological substances (proteins, DNA, RNA) and interactions between these molecules to be studied under conditions that closely resemble the physiological state. Conditions such as pH, temperature and salt concentrations can be adjusted to resemble the physiological state. These considerations have made NMR the method of choice for the study of biomolecular substances in solution. However, the use of NMR for the study of biomolecular substances is hampered by its intrinsically low sensitivity. The intensity of the NMR signal is proportional to the amount of material in the spectrometer’s sensitive volume, which is typically between 400µl and 600ul. In order to increase the signal-to-noise ratio, the signal from the sample is usually averaged many times over the course of an experiment, since the signal-to-noise ratio is proportional to the square root of the number of times the signal is averaged. This means that weaker samples will take longer to give the same signal-to-noise ratio as stronger samples. For structural studies, this implies that sample concentrations in the order of 1-5 mM are required, which for the DWNN domain corresponds to 5-25 mg of protein.

Therefore, the aim of this section of the work was to recombinantly express and purify the DWNN domain in sufficient quantities to allow the determination of the structure of the DWNN domain using heteronuclear NMR. The domain was expressed as a C-

terminal Glutathione S-transferase (GST) fusion protein using the pGEX gene fusion system (Amersham Biosciences). The system allows for the inducible, high-level expression of gene or gene fragments as fusions with *Schistosoma japonicum* GST (Smith and Johnson, 1988), in *E.coli*. GST fusion proteins are affinity purified from bacterial lysates using immobilised glutathione. The protein of interest can be separated from GST using a site-specific protease whose recognition sequence lies in the linker region between the GST domain and the protein of interest. PreScissionTM Protease (Amersham Pharmacia Biotech), a fusion of GST and human rhinovirus 3C protease, was specifically engineered for cleavage of fusion proteins expressed from pGEX-6P fusion constructs. The protease recognises the amino acid sequence Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro and cleaves between the Gln and Gly residues. The protease is removed from cleavage reactions using immobilised glutathione, since it is itself a GST fusion protein.



3.2 Expression and purification of the DWNN domain

3.2.1 Screening of *E.coli* BL21 (DE) pLysS transformed with pGEX-6P-2-DWNN for recombinant protein expression

E.coli BL21 (DE) pLysS cells were transformed with pGEX-6P-2-DWNN as described in Section 2.2.3.1 and transformed cells were screened for recombinant protein expression as described in Section 2.2.3.2. Fig. 3.1 shows the total bacterial lysates for four different colonies. Lanes **B**, **D**, **F** and **H** shows the uninduced lysates, while lanes **C**, **E**, **G** and **I** show the corresponding induced lysates. A protein of approximately 36.2 kD

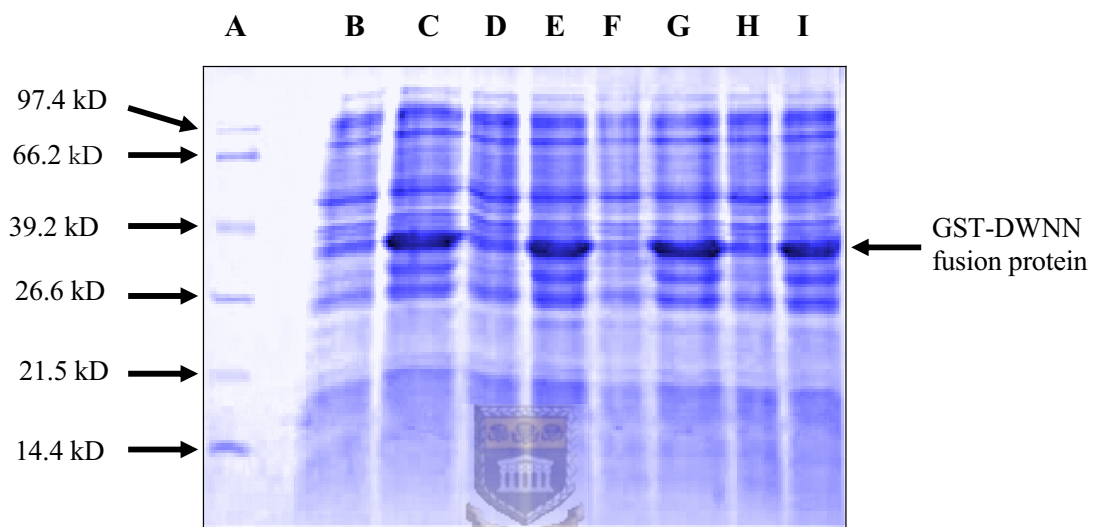


Fig. 3.1: Screening colonies transformed with pGEX-6P2-DWNN for the expression of GST-DWNN fusion protein. Lane **A** shows the molecular weight marker; lanes **B**, **D**, **F**, and **H** show the total cell lysate for the uninduced samples. Lanes **C**, **E**, **G** and **I** show the total cell lysate for the induced samples. The presence of a strong band in the vicinity of 36 kD in the induced but not in the uninduced samples confirms that GST-DWNN is being expressed.

corresponding to the expected size of the GST-DWNN fusion protein (GST = 26.6 kD, DWNN ~ 9.6 kD) can be seen in the induced sample lanes.

10 µl of culture **E** was used for large-scale protein expression. The remainder of this culture was used to make glycerol stocks that were stored at -80°C.

3.2.2 Large scale expression and purification of recombinant DWNN protein

10 µl of culture **E** was used to inoculate 1 litre of either minimal media or Luria broth for large-scale expression of recombinant DWNN protein. Cell lysis and purification of the GST-DWNN fusion protein was carried out as described in Sections 2.2.3.3 and 2.2.3.4.

Fig. 3.2 shows SDS-PAGE analysis of the expression and affinity purification of the fusion protein. The fusion protein was eluted in 15 ml of elution buffer. 20 µl was loaded in each sample lane. Lane **E**, which represents the column elute, shows a band of approximately 36.2 kD, the expected size of the GST-DWNN fusion. Fig. 3.2 shows that fusion protein yield was very good at the induction conditions used (overnight induction at 30 °C, using 0.5 mM IPTG).

Cleavage of recombinant GST-DWNN, using recombinant 3C protease, was performed as described in Section 2.2.4.3. The cleaved protein was analysed by SDS-PAGE as shown in Fig. 3.3. 20 µl was loaded in each sample lane. Lane **C** indicates that almost complete cleavage was achieved. Three bands, corresponding to GST (26.6 kD), the DWNN domain (9.6 kD) and a very small amount uncut fusion protein (36.2 kD) are visible in lane **C**.

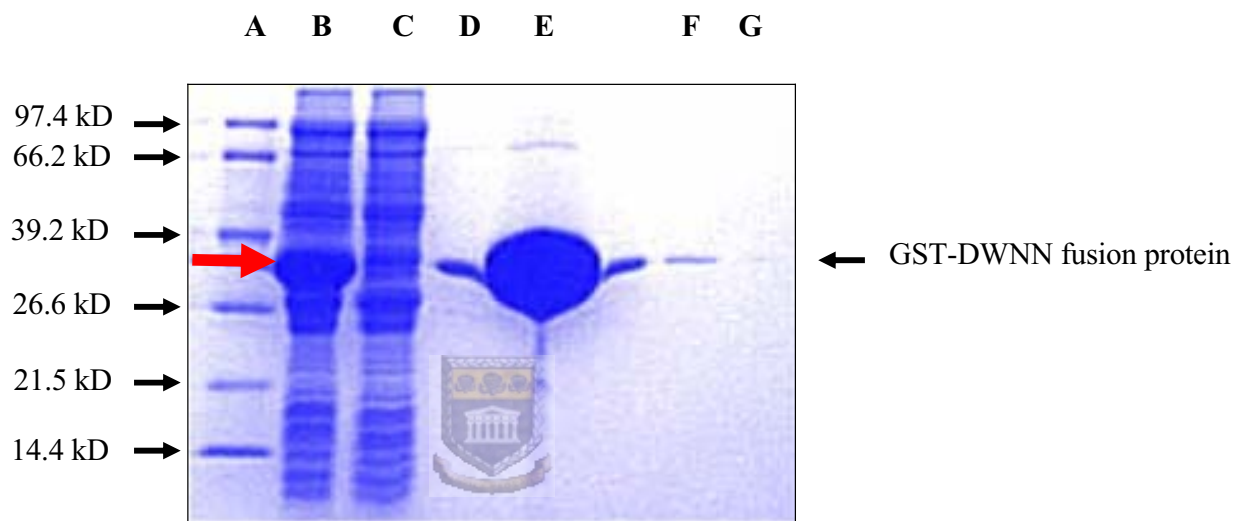


Fig. 3.2: Expression and purification of the GST-DWNN fusion protein. Lane **A** shows the molecular weight marker. Lane **B**, the cell lysate, shows a protein of approximately 36.2 kD (indicated by the red arrow), which corresponds to the GST-DWNN fusion protein. Lane **E** shows the GST-DWNN fusion protein that was eluted from the glutathione agarose column.

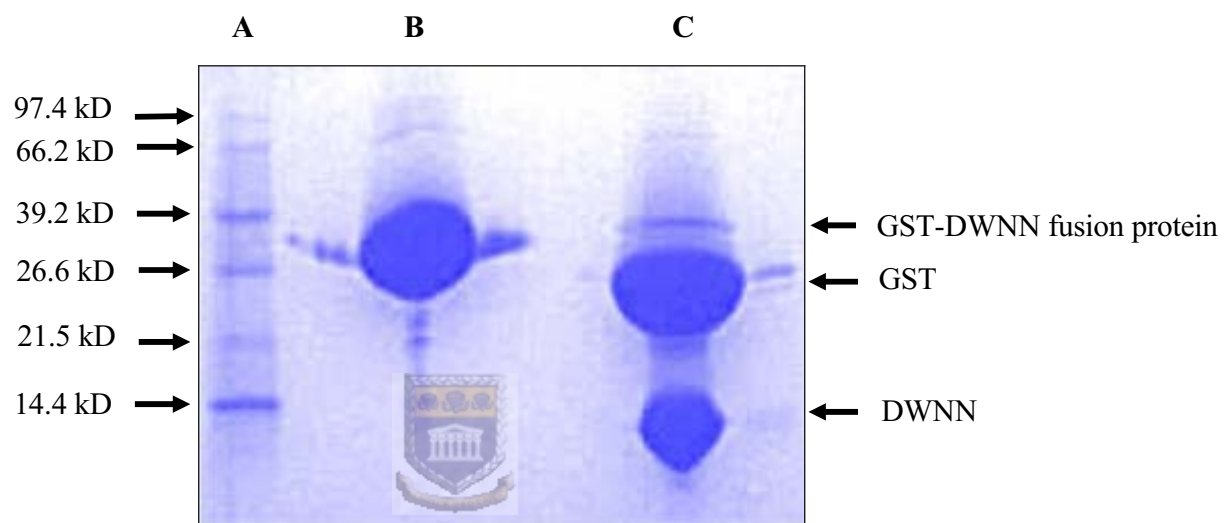


Fig. 3.3: PreScission™ protease cleavage of the GST-DWNN fusion protein. Lane **A** shows the molecular weight marker, while lanes **B** and **C** show the uncleaved and cleaved GST-DWNN fusion proteins, respectively.

3.2.3 Removal of the GST domain

GST was removed from the cleaved sample using a GSTPrepTM FF 16/10 column (Amersham Pharmacia Biotech), running on a BioCAD Sprint Perfusion Chromatography system (Perkin Elmer), as described in Section 2.2.4.4. The GSTPrepTM FF 16/10 column has a glutathione-linked Sepharose 4B matrix, which has a high affinity for GST. Fig. 3.4 is a chromatogram depicting the purification. Protein purity was analysed by SDS-PAGE, as shown in Fig. 3.5. A fraction of the cleaved product is shown in lane **B**. Lane **C** shows that the DWNN domain occurs in the column flow through, while the GST (lane **D**) remained bound to the column and occurs in the eluent. As a final purification step, in particular to remove any residual GST, the DWNN domain was subjected to gel filtration chromatography, as described in Section 2.2.4.4. The resulting chromatogram is depicted in Fig. 3.6.



DWNN peak fractions were pooled and concentrated to a final volume of 600 μ l. Fig. 3.7 shows a silver-stained gel of the purified DWNN domain. Lane **A** shows the molecular weight marker. Lane **B** is a 1.5 μ l fraction of the DWNN peak in Fig. 3.7, while lane **C** is a 0.5 μ l fraction of the concentrated protein. Lane **C** shows a single band of approximately 9.6 kD corresponding to the expected size of the DWNN domain. No other bands are visible in this lane. Since silver staining is a very sensitive staining method (as little as 1 ng of protein can be detected), this indicates that the DWNN domain was highly purified.

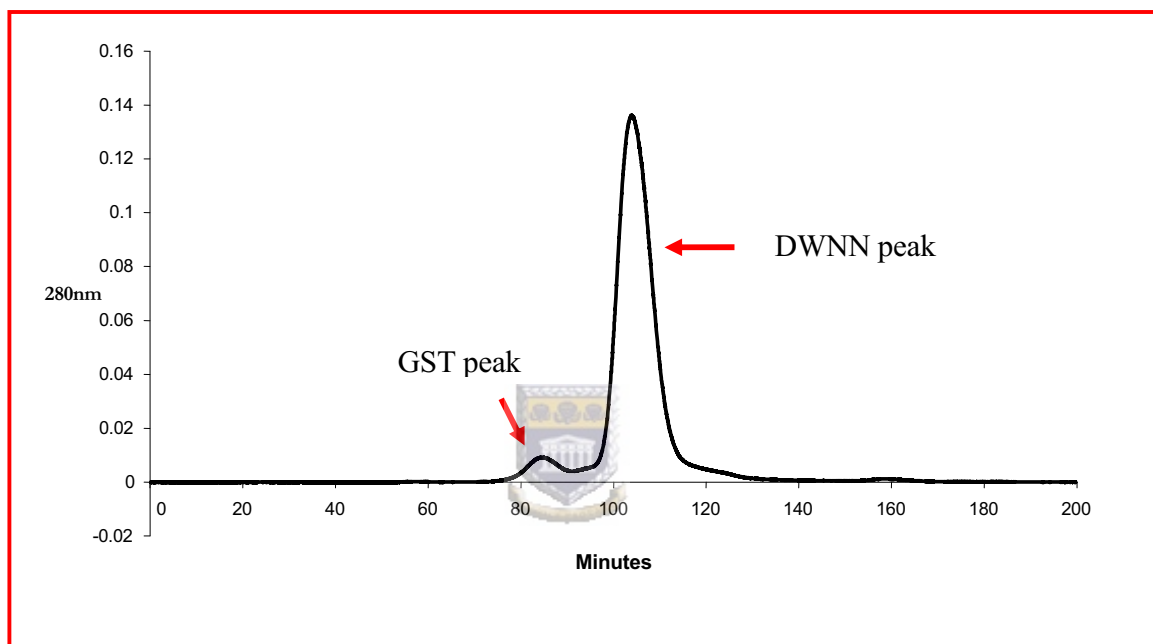


Fig. 3.6: Chromatogram showing gel filtration chromatography of the DWNN domain. The chromatogram shows two well-resolved peaks; a major peak, corresponding to the DWNN domain, and a smaller peak, corresponding to GST.

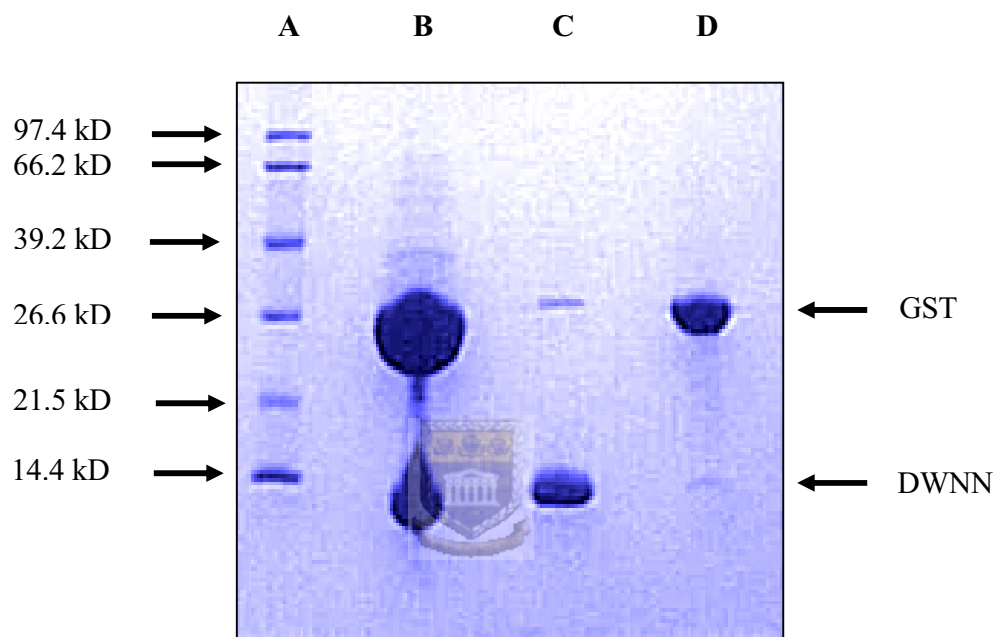


Fig. 3.5: Removal of the GST domain. Lane **A** shows the molecular weight marker and lane **B** shows a fraction of the PreScission™ protease cleavage products. Lane **C** shows a fraction from the DWNN peak in fig. 3.4, while lane **D** is a fraction from the GST peak. Lane **C** shows two bands, a 9.6 kD band corresponding to the DWNN domain and a 26.6 kD band corresponding to the GST domain.

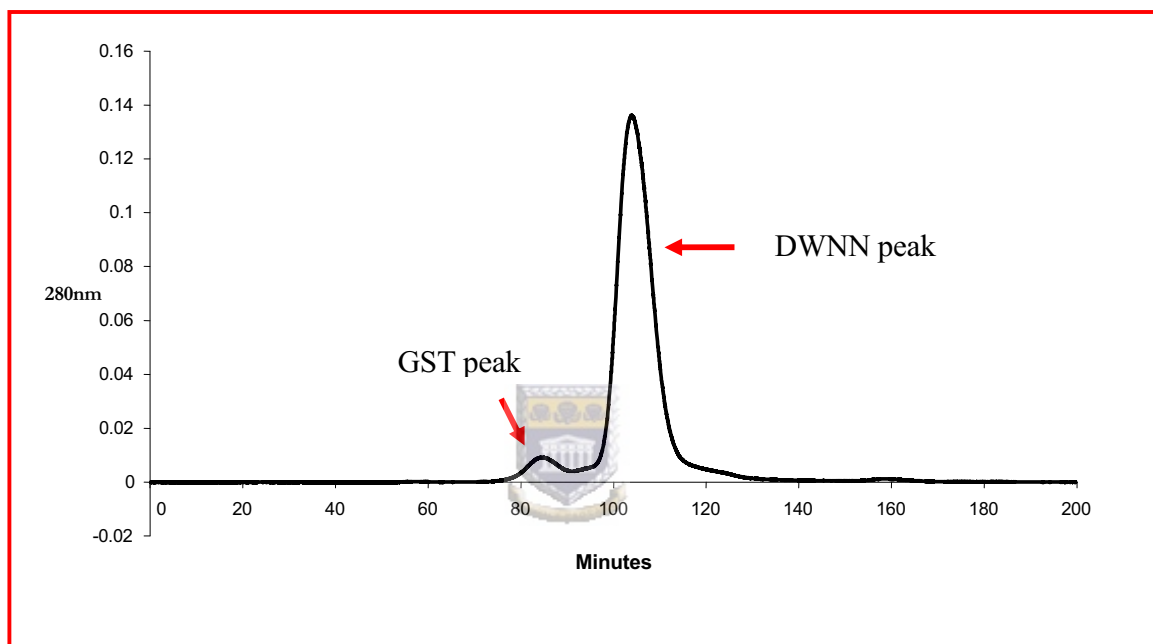


Fig. 3.6: Chromatogram showing gel filtration chromatography of the DWNN domain. The chromatogram shows two well-resolved peaks; a major peak, corresponding to the DWNN domain, and a smaller peak, corresponding to GST.

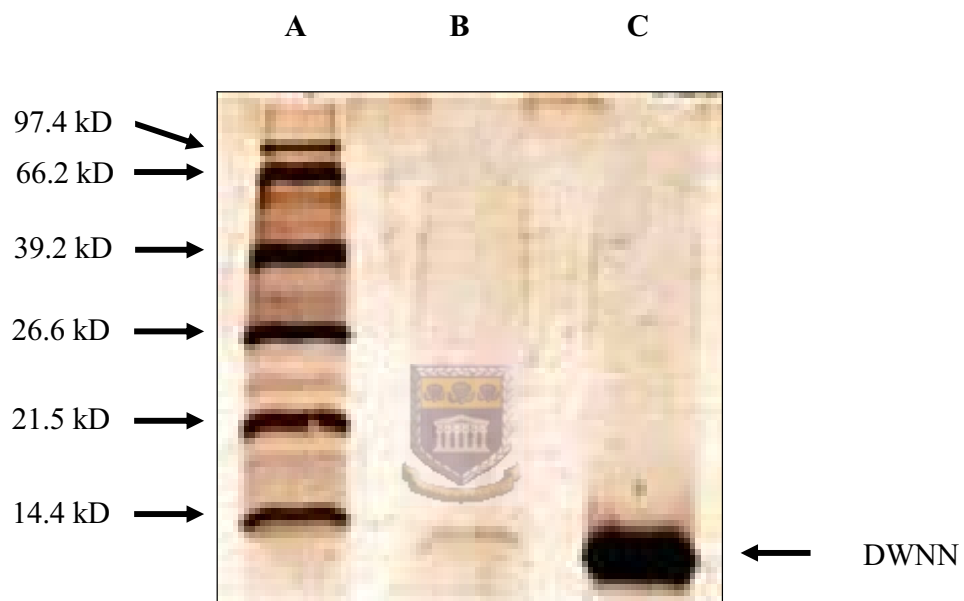


Fig. 3.7: Silver stained gel of purified DWNN. Lane **A** contains the molecular weight marker; lanes **B** and **C** show the purified DWNN domain before and after concentration, respectively.

3.2.4 Bradford assay

The concentration of the concentrated DWNN preparations (unlabeled and isotopically labelled) was determined by the Bradford assay as described in Section 2.4. Table 3.1 lists the concentrations and absorbance of the lysozyme standards and Chart 1 depicts the resulting standard curve.

Concentration in mg/ml	Absorbance at 620 nm
0.0125	0.012
0.025	0.055
0.05	0.1235
0.1	0.237
DWNN: 1/1000 dilution	0.045

Table 3.1: Concentration and absorbance of lysozyme at 620 nm.

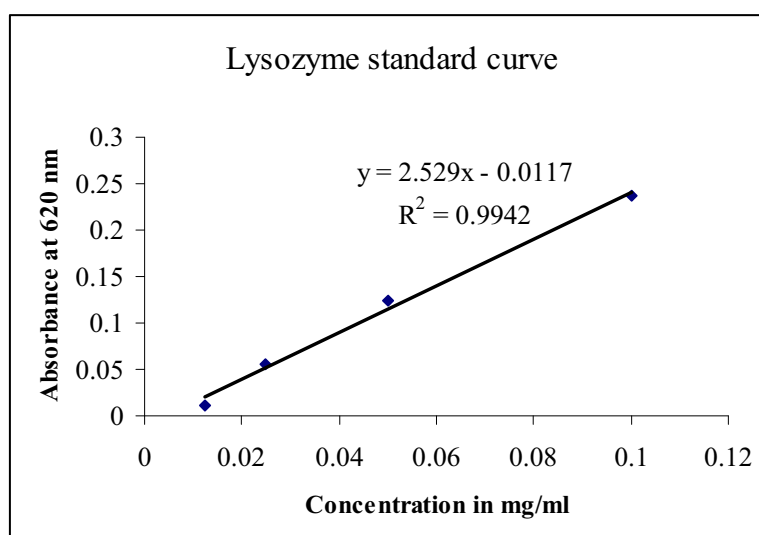


Chart 3.1: Lysozyme standard curve. The equation, $y = 2.529x - 0.0117$, was used to determine the concentration of DWNN domain preparations.

Use of the standard curve gave a final concentration of 22.42 mg/ml for the concentrated DWNN domain shown in Fig. 3.7. Concentrations of other samples ranged from 14 mg/ml to 22.42 mg/ml. For a protein the size of the DWNN domain, 9.6 kD, this corresponds to 1.45 mM and 2.33 mM, respectively. Typically, protein concentrations of 1 mM are sufficiently concentrated for NMR studies.

3.2.5 Lyophilization

Purified protein samples were snap-frozen in liquid nitrogen and lyophilised overnight on a Virtis Sentry (Virtis Company Inc.) freeze-drier. Lyophilised samples were stored successfully at 4⁰ C for periods of more than two months.



3.3 Summary

E.coli BL21 (DE) pLysS cells were successfully transformed with pGEX-6P-2-DWNN. Transformed cells were screened for recombinant protein expression and positive transformants showing good expression were used for large-scale protein expression. GST-DWNN fusion protein, which has a molecular weight of about 36.2 kD, was purified from the bacterial lysate and cleaved with recombinant 3C protease to yield a 26.6 kD GST domain and the 9.6 kD DWNN domain. The DWNN domain was purified to homogeneity using a combination of affinity chromatography and gel filtration chromatography. The purified DWNN domain was concentrated to 600 μ l, lyophilised and stored at 4⁰ C.

Chapter 4: Full backbone assignment of the human DWNN domain using triple-resonance nuclear magnetic resonance

4.1 Introduction

The first NMR structures of proteins were determined from unlabelled samples using 2D homonuclear TOCSY, COSY and NOESY spectra (Wütrich, 2003). This approach works well for smaller proteins. In the case of larger proteins, the larger number of resonances and the increased linewidths caused by longer molecular tumbling times lead to extensive overlap in proton spectra, making assignment of spectra very difficult. The development of heteronuclear experiments (reviewed by Sattler *et al.* 1999), using ^{15}N and ^{13}C -enriched samples produced in heterologous expression systems has helped to overcome these problems. In these experiments, proton-proton NOE cross-peaks are labelled with the chemical shift of the attached ^{15}N or ^{13}C resonance, allowing spectra to be separated into the heteronuclear direction and thereby greatly reducing overlap.

This chapter describes the methodology used to assign all ^{15}N , ^{13}C and ^1H resonances in the backbone of the DWNN domain using triple-resonance coherence transfer experiments, as well as the information that can be obtained about the overall fold of the backbone using these assignments. Backbone conformational information can be obtained using two independent methods: (i) estimation of ϕ dihedral angle restraints from the measurement of $\text{H}^\alpha\text{-H}^{\text{N}}$ coupling constants using the Karplus Relation (Eqn. 1), and (ii) knowledge-based

prediction of the ϕ and ψ angles from backbone chemical shifts using the TALOS algorithm.

4.2 Assessment of DWNN sample integrity and spectral quality using homonuclear 1D and ^{15}N -HSQC experiments spectra

NMR samples were prepared as described in Section 2.3.1. Samples were dissolved in H_2O , with the addition of 10% D_2O , to act as a lock signal. To assess whether the protein was folded, a 1D ^1H homonuclear experiment was performed (Fig. 4.1) at 25 $^\circ\text{C}$ using a 600 MHz Varian Inova spectrometer (Varian Inc.). Pre-saturation was used to suppress the water resonance. The spectrum was visualised and analysed using the NMRView software package (Johnson and Blevins, 1994).



The folded state of a protein can be gauged by the degree of resonance dispersion in a 1D spectrum (Cavanagh *et al.*, 1996, Woestenenk *et al.*, 2003, Scheich *et al.*, 2004). Unfolded or denatured proteins will exhibit chemical shifts similar to random coil values: H^{N} resonances will appear in the region 8.0-8.5 ppm, H^{α} around 4.1-4.4 ppm and methyl groups in the region of 0.8-1.1 ppm (Cavanagh *et al.*, 1996). In contrast, folded proteins will exhibit a range of chemical shifts due to the shielding effect of nearby aromatic rings and carbonyl groups. This shielding stems from induced magnetic fields in the electronic orbitals of these groups, resulting from the applied static magnetic field B_0 . The presence of H^{N} resonances in the range 8.5-10 ppm, H^{α} resonances between 5 and 6 ppm, and

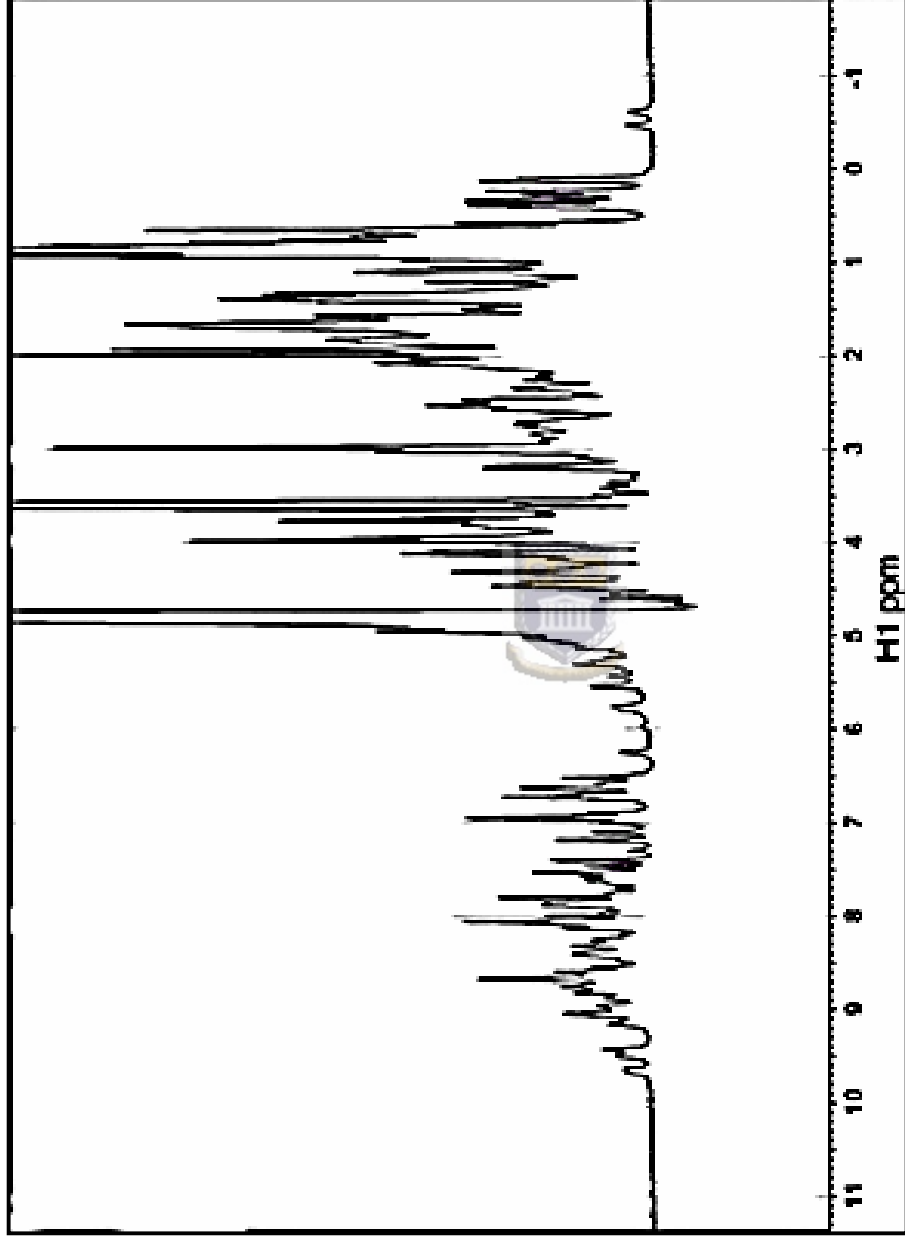


Fig. 4.1: 1D ¹H spectrum of the DWNN domain at pH 6.0, 25^o C, recorded at 600 MHz. The good dispersion of resonances (some H^N in the region 8.5- 10 ppm, some H^α in the region 5-6 ppm, methyl protons in the region 0-0.5 ppm and two resonances <0 ppm) is an indication that the protein is folded.

methyl resonances between -1 and 1 ppm clearly indicates that the DWNN domain is folded under the given experimental conditions.

The folded state of the DWNN domain was confirmed by performing a ^{15}N -heteronuclear single quantum coherence (^{15}N -HSQC) experiment. The ^{15}N -HSQC experiment transfers coherence between the backbone nitrogen and its attached proton, resulting in a single peak for each residue (excluding proline residues since they have no amide proton) (see Fig. 4.2). Coherence transfer between the side-chain nitrogen atoms of asparagine and glutamine residues and their attached pairs of protons results in pairs of peaks at the same ^{15}N chemical shift, which are typically found towards the top right hand corner of the ^{15}N -HSQC spectrum. The ^{15}N -HSQC spectrum of the DWNN domain, recorded at 600 MHz on a Varian Inova spectrometer, is shown in Fig. 4.3.



The high degree of dispersion in both the ^{15}N and ^1H dimensions is indicative of a well-folded protein (Rehm *et al.*, 2002). Resonances 11, 53, 56, 72, 74 and 75 are all strongly down-shifted in the ^1H dimension, while 33, 36, 43, 60 and 73 are up-shifted in the ^1H dimension, which is strong evidence of folding. The fact that the peaks are well-resolved in both dimensions greatly facilitated the assignment using triple-resonance spectra, as the individual strips in the strip plots contained only one residue in the majority of cases (see Fig. 4.5 and 4.7). This spectrum was used to extract amide ^1H and ^{15}N chemical shifts (labelled HN and N respectively in Appendix 1).

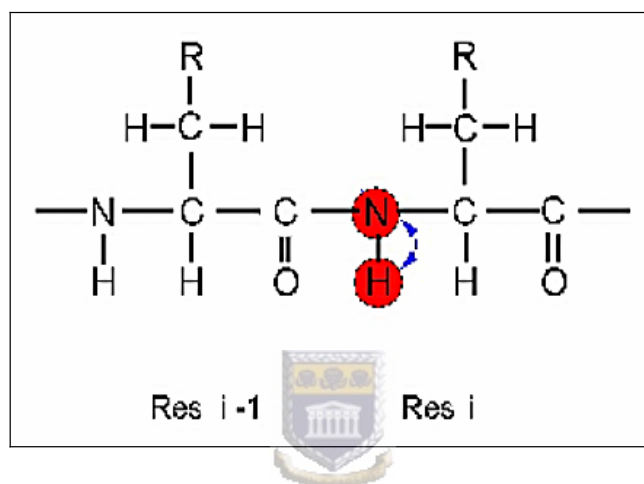


Fig. 4.2: Through bond transfer between the backbone nitrogen and the attached proton in a ^{15}N -HSQC experiment.

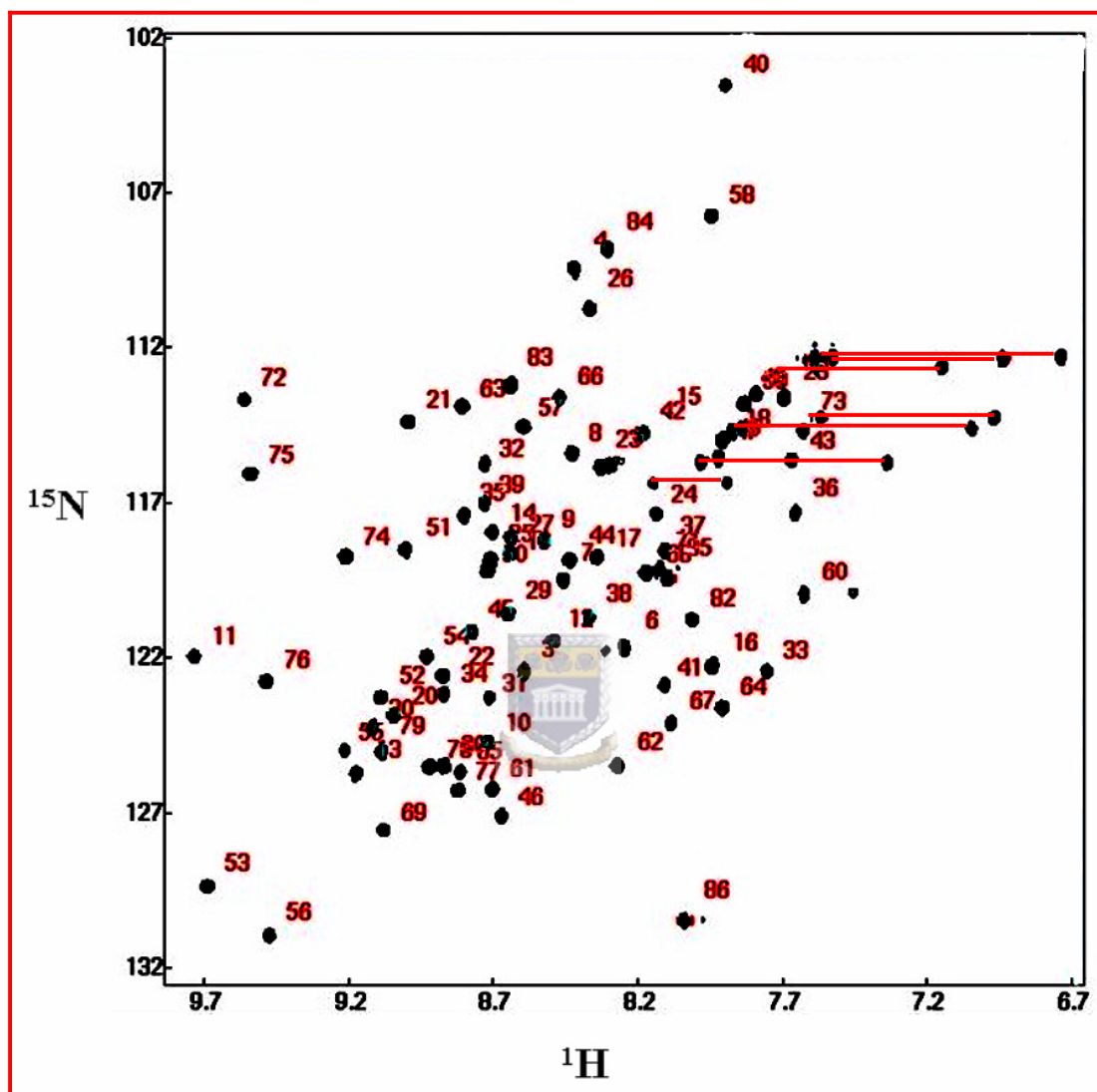


Fig. 4.3: ^{15}N -HSQC spectrum of the DWNN domain at pH 6.0, 25°C , recorded at 600 MHz. Each peak is labelled with the number of the corresponding residue, using a numbering scheme in which residue 6 corresponds to the initiation methionine (see Section 2.2.3.1). The seven pairs of resonances corresponding to NH_2 sidechains of asparagine and glutamine residues are joined by horizontal lines. These resonances could not be easily assigned to their corresponding residues using the experiments discussed in this thesis, and so are left unlabelled.

4.3 Assignment of backbone resonances

4.3.1 Sequential assignment of C_α and C_β resonances using CBCA(CO)NH and CBCANH experiments

Resonance assignment is a prerequisite for any structure determination exercise, i.e. each resonance must be assigned to a particular nucleus in the molecule being investigated. CBCA(CO)NH (Grzesiek and Bax, 1992) and CBCANH (Grzesiek and Bax, 1992) are both three dimensional spectra, in which the dimensions correspond to ^1H , ^{15}N and ^{13}C respectively. In both experiments, magnetic energy is first excited on the H^{N} proton, and then transferred first to the attached nitrogen and then to the C_α and C_β nuclei of the same residue (the “in-residue” C_α and C_β) and of the preceding residue (the “preceding” C_α and C_β). During the first evolution period, t_1 , the magnetic energy is allowed to be modulated by the precession frequencies of the C_α and C_β nuclei. It is then transferred back to the N nucleus, and allowed to be modulated by the ^{15}N precession frequency during the second evolution period, t_2 . Finally it is transferred back to the H^{N} nucleus, which modulates it with the proton frequency during the acquisition period, t_3 . The difference between the CBCA(CO)NH and the CBCANH experiments lies in the fact that in the CBCANH experiment the magnetization is transferred back from the C_α and C_β 's to the N without going via the carboxyl carbon (the “CO”) whereas in the CBCA(CO)NH the magnetization is forced to go via the “CO”, thus eliminating any signal from the in-residue C_α and C_β . The “CO” is placed in brackets in the name “CBCA(CO)NH” to indicate that it is used only as a conduit and its frequency is not allowed to label the signal.

When the data is Fourier transformed the result is a three-dimensional spectrum in which the ω_1 dimension corresponds to the ^{13}C chemical shifts, the ω_2 dimension to the ^{15}N shifts and the ω_3 dimension to the H^{N} shifts. In the CBCA(CO)NH experiment (Fig. 4.4) two peaks, with the chemical shifts of the C_α and C_β respectively of residue **i-1** can be observed at the $(\text{H}^{\text{N}}, \text{N})$ chemical shifts of residue **i**. In the CBCANH experiment (Fig. 4.4) four peaks, with the chemical shifts of the C_α and C_β of both residue **i-1** and residue **i** respectively can be observed at the $(\text{H}^{\text{N}}, \text{N})$ chemical shifts of residue **i**. Pairs of consecutive residues can thus be identified using these two experiments, making it possible to “walk along” the protein backbone. Having the CBCA(CO)NH in addition to the CBCANH makes it possible to distinguish easily between peaks belonging to the “in-residue” and to the preceding residue respectively.



Two dimensional “strip plots” are generated from three dimensional spectra by extracting tubes centred on each resonance in the ^{15}N -HSQC spectrum and extending across the full ^{13}C spectral width. The tubes are reduced to two dimensional strips by taking cross sections through the tubes, either in the ^1H or the ^{15}N dimensions. A double strip plot generated from the CBCA(CO)NH and CBCANH spectra corresponding to residues 29 to 33 can be seen in Fig. 4.5. In each pair of strips, the yellow strip corresponds to the CBCA(CO)NH and the blue strip to the CBCANH spectrum. In the CBCANH spectrum, C_α and C_β peaks can be distinguished by the fact that they have opposite signs: the C_α peaks are shown in red and C_β peaks in black.

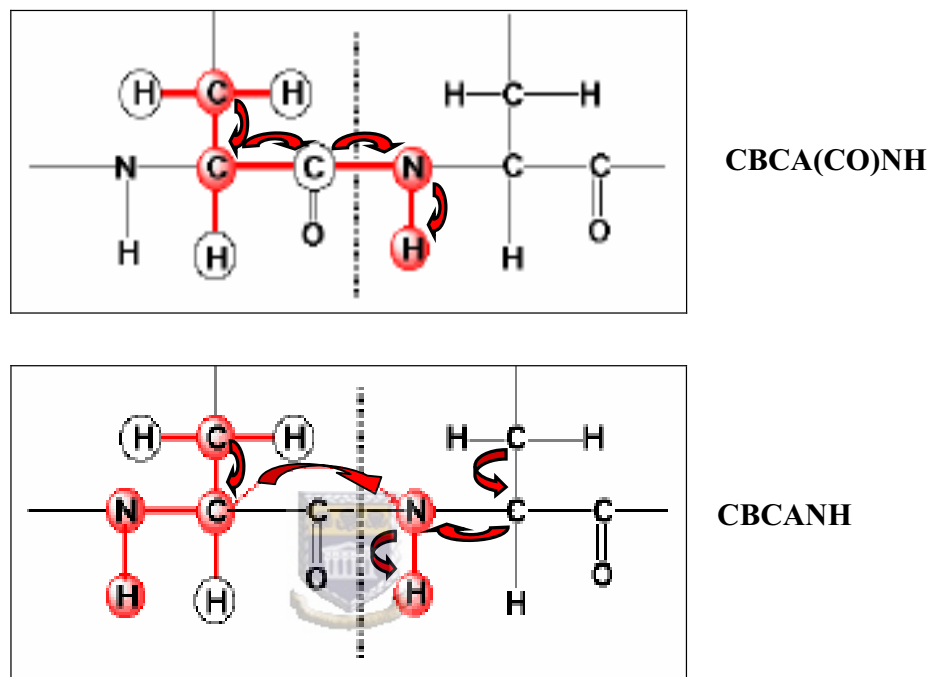


Fig. 4.4: Coherence transfer in CBCA(CO)NH and CBCANH experiments for a pair of consecutive residues. The red arrows indicate the magnetization transfer pathway. In the CBCA(CO)NH experiment, coherence is transferred only between the H-N group and the C_β and C_α of the preceding residue, whereas in the CBCANH experiment the coherence is transferred between the H-N group and the C_β and C_α of both the “preceding” residue and the “in-residue”.

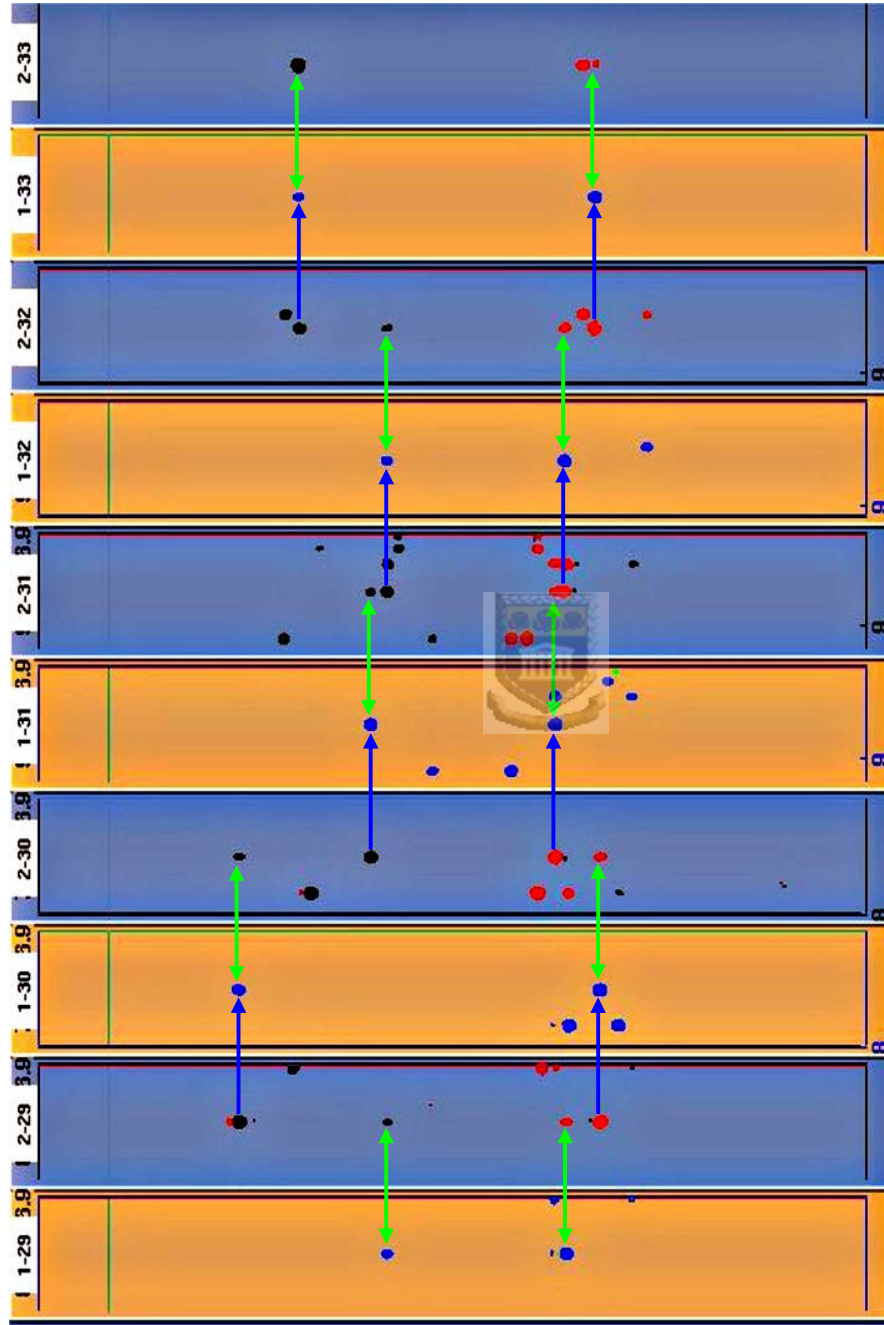
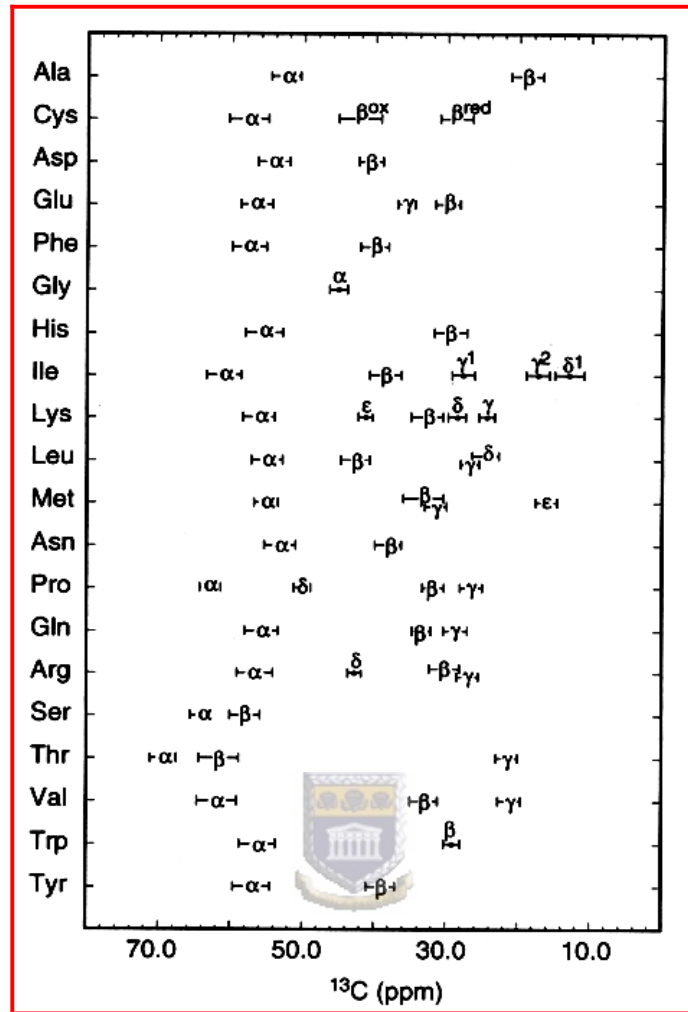


Fig. 4.5: Strips showing backbone sequential connectivities of residues 29 to 33 of the DWNN domain. Yellow strips are taken from a CBCA(CO)NH spectrum while blue strips are taken from a CBCANH spectrum. Blue arrows indicate transfer of magnetization from a single C_{α} or C_{β} resonance to adjacent H-N groups, allowing sequential relationships to be identified.

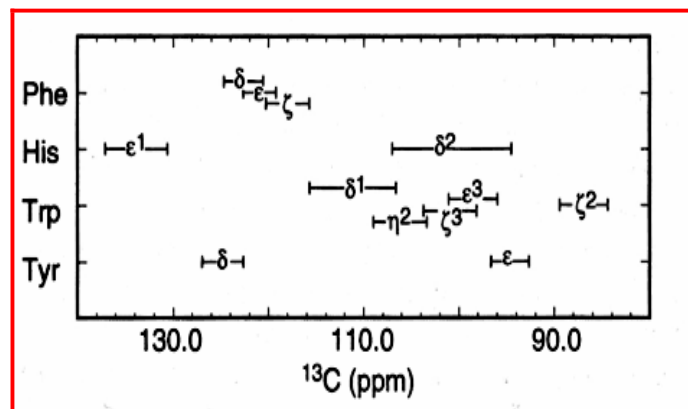
With a few exceptions, each yellow strip in Fig. 4.5 contains two peaks, corresponding to the C_α and C_β resonances of residue **i-1** respectively. Corresponding peaks are seen in the blue CBCANH strips. The correspondence is highlighted with the aid of double-headed green arrows. In addition to these peaks, CBCANH strips also contain C_α and C_β resonances corresponding to residue **i**. By matching these peaks to the **i-1** peaks in another pair of strips, strips corresponding to adjacent residues can be identified. These pairings are indicated by blue arrows in Fig. 4.5. The strips in Fig. 4.5 have been arranged in order of the sequence.

In order to link the sequentially assigned strips to stretches of the amino acid sequence, C_α and C_β shifts were compared with the shifts expected for particular amino acids. The expected amino acid specific ^{13}C chemical shifts are shown in Fig. 4.6 (Cavanagh *et al.*, 1996). Using these chemical shifts as reference, alanine residues were identified because of their high C_β shifts, while threonine and serine residues were identified due to low chemical shifts for both C_α and C_β . Glycine residues lack C_β 's and therefore has only one peak in the CBCA(CO)NH, corresponding to the C_α , making them very easily identifiable. This information, taken together with the known amino acid sequence of the DWNN domain, allowed the backbone sequential assignment to be completed.

The C_β of cysteine residues has a chemical shift in the vicinity of 30 ppm if the thiol group is reduced, but in the vicinity of 40 ppm if the thiol is oxidised, as would be the case if the cysteine was involved in a disulfide bond. Due to the presence of 1 mM DTT in all NMR samples, all cysteines were expected to be reduced. The C_β chemical shifts for the three



A



B

Fig. 4.6: (A) Average C_α and C_β chemical shifts for all 20 commonly occurring residues. These values were calculated using a database of 13 proteins. (B) Average ^{13}C chemical shifts for aromatic ring carbons (Cavanagh et al, 1996).

cysteine residues in DWNN are shown in Table 4.1; they confirm that none of the cysteines are involved in disulfide bonds.

Residue	¹³ C chemical shift
Cys8	32.22
Cys32	27.00
Cys49	30.69

Table 4.1: C_β chemical shift values for DWNN domain cysteine residues.



4.3.2 Identification of H^N-H^N NOE cross-peaks

Sequential connectivities were confirmed by identification and analysis of NOE cross-peaks between adjacent residues, in particular between H^N nuclei of adjacent residues, so-called “H^N-H^N” or “NN” NOE’s. A ¹⁵N-HSQC-NOESY spectrum (Fig. 4.7) recorded at 600 MHz was used for the identification of NOE’s. A ¹⁵N-HSQC-NOESY experiment is a three-dimensional experiment, combining the two dimensional proton-proton NOESY and ¹⁵N-HSQC experiments and allowing the NOESY to be separated into the ¹⁵N dimension. In the ¹⁵N-HSQC-NOESY experiment magnetic energy is first excited on all protons and modulated with their frequencies in t₁, and then allowed to transfer “through space” to

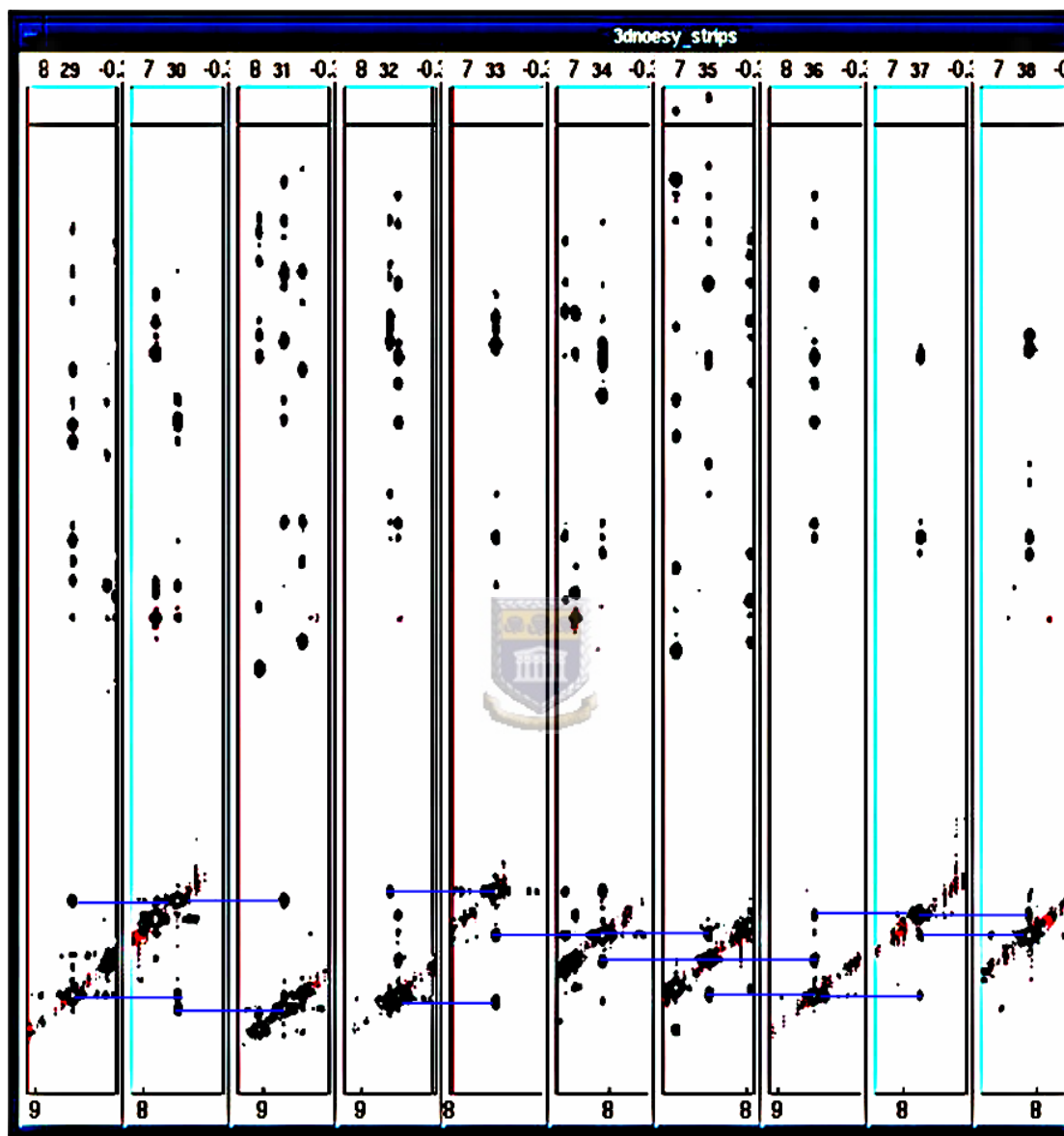


Fig. 4.7: 3D ^{15}N -HSQC-NOESY strips depicting backbone H^{N} - H^{N} NOE's. The strips shown correspond to residues 29 to 38, which form the central α -helix. The horizontal lines indicate the extensive i to $i+1$ and i to $i-1$ H^{N} - H^{N} NOE's expected in an α -helix.

other protons close in space. The magnetic energy is then transferred “through bond” to the attached nitrogen and modulated with the nitrogen frequency in t_2 . Finally the magnetization is transferred back to the attached proton which modulates it during the acquisition period, t_3 . The resulting spectrum has ^1H resonances in dimensions ω_1 , N resonances in ω_2 and H^{N} in ω_3 . Because of the transfer to the nitrogen, the only peaks observed in ω_3 are the H^{N} protons, i.e. the spectrum appears on the left of the water resonance.

Strips corresponding to residues 29 to 38 of the ^{15}N -HSQC-NOESY spectrum are shown in Fig. 4.7. Each strip corresponds to NOE's between a $\text{H}^{\text{N}}/\text{N}$ pair and another proton in the sample. The pairs of peaks indicated by blue lines correspond to “through-space” transfer between H^{N} protons corresponding to the two strips. The strips show that residues 29 to 38 make a series of backward and forward $\text{H}^{\text{N}}\text{-H}^{\text{N}}$ NOE's, indicating that this region of the DWNN domain forms an α -helix, since in an α -helix the H^{N} protons of adjacent residues are only 2.5Å apart.

4.3.3 Extraction of ^{13}C CO chemical shifts

Carbonyl carbon “CO” chemical shifts were extracted from an HNCO spectrum of the DWNN domain (Fig. 4.9). The HNCO experiment (Kay *et al.*, 1990) is the highest sensitivity heteronuclear experiment and correlates the amide ^1H and the ^{15}N chemical shifts of residue **i** with the carboxyl carbon chemical shift of residue **i-1**, by using the large

one-bond J_{N-CO} coupling. The coherence transfer pathway in the HNC0 experiment is depicted graphically in Fig. 4.8.

In the HNC0 experiment magnetic energy is first excited on the H^N proton, and then transferred first to the attached N nucleus and then to the “CO” in the preceding residue. During the first evolution period, t_1 , the magnetic energy is modulated by the precession frequency of the “CO” nuclei. It is then transferred back to the N nuclei during the second evolution period, t_2 , and allowed to be modulated by its precession frequency. Finally, the magnetic energy is transferred back to the H^N and modulated with the proton precession frequency during the acquisition period, t_3 . Fourier transformation of the data results in a three dimensional spectrum, with the ω_1 dimension corresponding to “CO” chemical shifts, ω_2 to the ^{15}N chemical shifts and the ω_3 dimension to the H^N chemical shifts. Since every amide group is covalently bonded to a single “CO”, only a single cross peak per residue is observed in the HNC0 spectrum (see Fig. 4.9). The “CO” chemical shifts are listed in Appendix 1.

4.4 Extraction of secondary structure information from backbone chemical shifts

4.4.1 Extraction of $^3J_{H^N-H^{\alpha}}$ coupling constants

$^3J_{H^N-H^{\alpha}}$ coupling constants were derived from a HMQC-J spectrum (Fig. 4.10) performed at 500 MHz on a hybrid General Electric/Oxford Instruments spectrometer. The pulse sequence for the HMQC-J experiment is adapted from the HMQC experiment, and gives

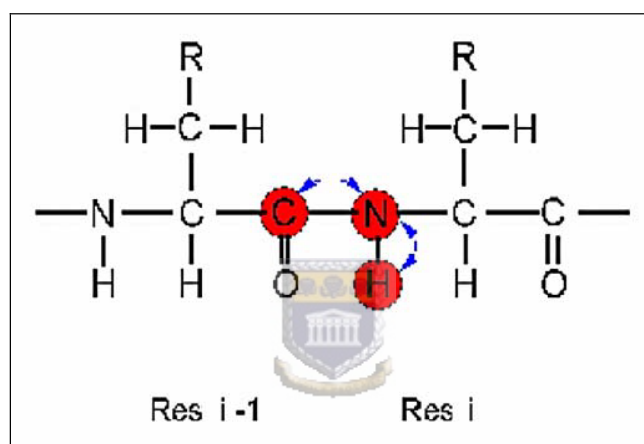


Fig. 4.8: Coherence transfer pathway in an HNCO experiment.

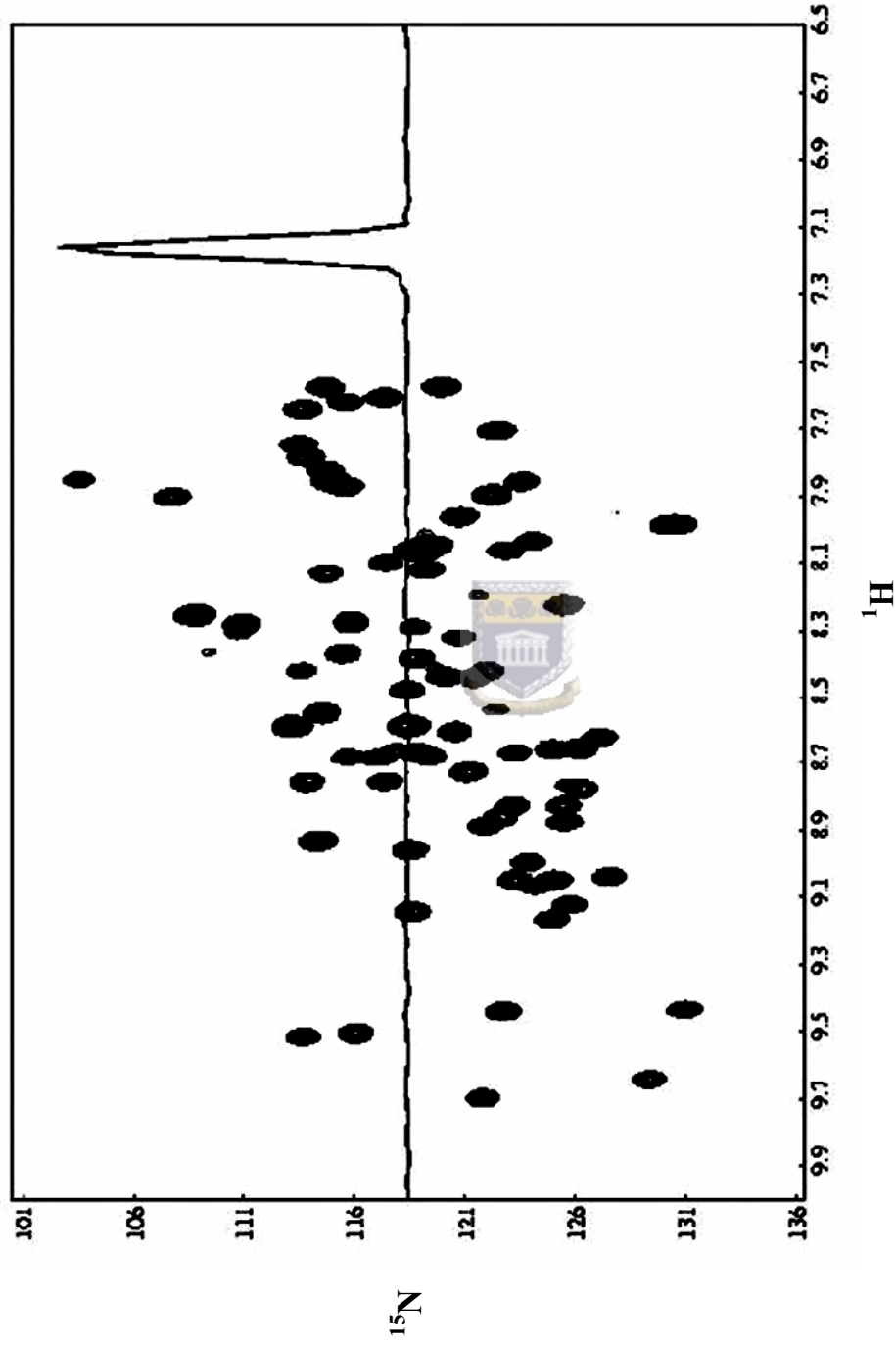


Fig. 4.9: Projection of the HNCO spectrum of the DWNN domain onto the ^1H - ^{15}N plane. The HNCO spectrum was recorded at the Varian Applications Laboratory in Darmstadt, Germany, at pH 6.0, 25 $^\circ$ C, at 600 MHz. The solid line represents a trace along the ^{13}C dimension (into the plane of the page) through one of the peaks, and contains only a single peak as expected. The high signal-to-noise ratio is evidence of the highly concentrated state of the protein sample.

a spectrum identical to the HSQC experiment apart from the fact that the 3-bond coupling, ${}^3J_{H^N-H^\alpha}$ (see Fig. 4.11), between the H^N and H^α of each residue shows up as a splitting of the relevant peak in the ${}^{15}N$ dimension (see Fig. 4.10). The splitting is measured in Hz rather than ppm, which indicates that its magnitude is independent of the field strength B_0 . The splitting is particularly obvious in the resonances on the extreme left of the spectrum, corresponding to residues 11, 74 and 76; all of which occur in β -sheets. Splittings were measured by fitting to the theoretically expected lineshapes (Kay and Bax, 1990) and are listed in Appendix 2.

Comparison of Fig. 4.3 and Fig. 4.10 shows that in the HMQC-J spectrum resonance 40 is aliased. “Aliasing” or “folding” refers to the production of “false” peaks which originate outside the spectral window used for a particular spectrum. In this case, the ${}^{15}N$ chemical shift for peak 40 is 105.939 (obtained from the ${}^{15}N$ -HSQC spectrum and listed in Appendix 1), which falls outside the spectral window of 131-107 ppm used for recording this HMQC-J experiment. The peak therefore appears the same distance above the bottom of the spectrum as the original peak would have been above the top. In addition it has a negative sign.

The splittings derived from the HMQC-J spectrum were used to estimate the φ angle restraints using the Karplus Relation (Eqn. 1), which gives the theoretical relationship between ${}^3J_{H^N-H^\alpha}$ (in Hz) and the backbone φ angle (Pardi *et al.*, 1984).

$${}^3J_{H^N-H^\alpha} = 6.4 \cos^2 (\varphi - 60^\circ) - 1.4 \cos (\varphi - 60^\circ) + 1.9 \quad [\text{Eqn.1}]$$

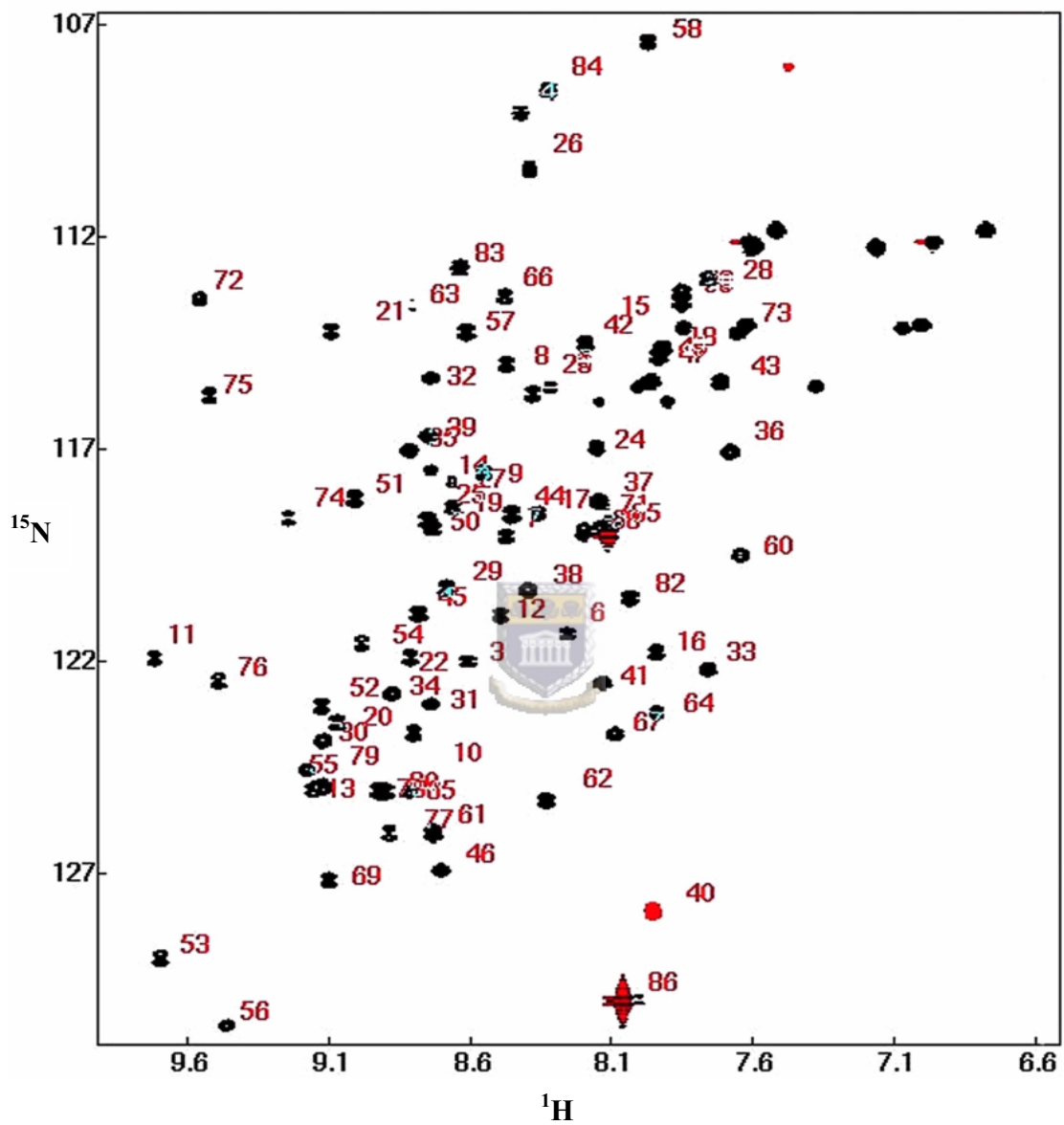


Fig. 4.10: HMQC-J spectrum of the DWNN domain at pH 6.0, 25⁰ C, recorded at 500 MHz. The peaks are split in the ^{15}N (vertical) direction by the $^3J_{\text{H-N}}^{\alpha}$ coupling.

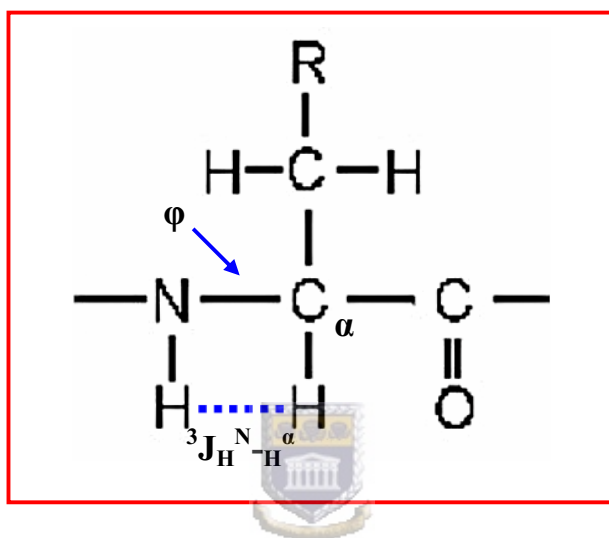


Fig. 4.11: Diagram depicting the 3-bond coupling ${}^3J_{\text{H}^{\text{N}}-\text{H}^{\alpha}}$. The blue dashed line represents the coupling between the H^N and H^α protons, while the arrow indicates the dihedral angle, ϕ . The magnitude of the coupling depends on ϕ in a manner given by the Karplus Relation (Eqn. 1). Figure provided by DJR Pugh.

A graph of 3J as a function of ϕ is shown in Fig. 4.12. The regions of ϕ space indicated by “ β ” and “ α ” correspond to the regions of the Ramachandran Plot associated with β -sheets and α -helices respectively.

From the curve, splittings ≥ 8 Hz correspond to β -sheet regions, while splittings ≤ 6 Hz correspond to α -helical regions of the backbone. Splittings in the region 6-8 Hz correspond to the time-averaging between α -helix type and β -sheet type conformations, as occurs in regions of poorly defined secondary structure. Splittings in this region therefore cannot be used to generate restraints on the ϕ angle.

Based on these criteria, residues were assigned to α -helical or β -sheet regions (see Appendix 2). For the generation of molecular structures using restrained molecular dynamics, ϕ angles of residues in α -helical regions were assigned a restraint of $-60^\circ \pm 20^\circ$; ϕ angles in β -sheet regions were assigned a restraint of $-120^\circ \pm 20^\circ$.

4.4.2 Prediction of DWNN domain backbone torsion angle restraints using TALOS

TALOS (Torsion Angle Likelihood Obtained from Shift and sequence similarity) is a protein backbone dihedral angle prediction program (Cornilescu *et al.*, 1999). The program is based on the observation that positive or negative chemical shifts relative to the average for random coil correlate very well with secondary structure. Wishart *et al.* (Wishart *et al.*, 1992) found that $^1H^\alpha$ chemical shifts are shifted to lower ppm values relative to random coil values in α -helix regions, while showing an upward shift in β -

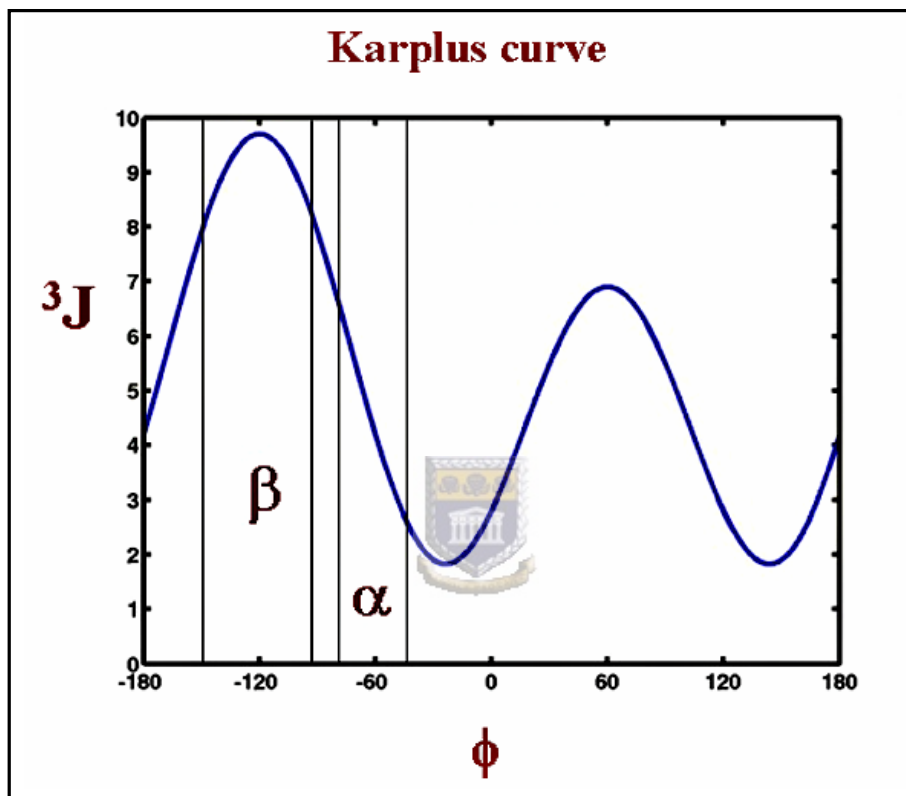


Fig. 4.12: The Karplus Curve, showing the relationship between $^3J_{H^N-H^\alpha}$ coupling constants, the backbone ϕ angle and polypeptide secondary structure. Figure provided by DJR Pugh.

sheets. Similar findings were made for $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{13}CO chemical shifts (Wishart and Sykes, 1994).

TALOS uses these observations, together with amino acid sequence similarities, to search for triplets of amino acids with the closest backbone chemical shifts to the protein under investigation in a database of 20 well-resolved protein X-ray crystal structures. The program uses as input the $^1\text{H}_\alpha$, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, ^{13}CO and ^{15}N chemical shifts for each residue of the target protein. It then uses the average and standard deviation for the 10 best matches to predict the ϕ and ψ angles for the central residue in a triplet of residues in the target protein.

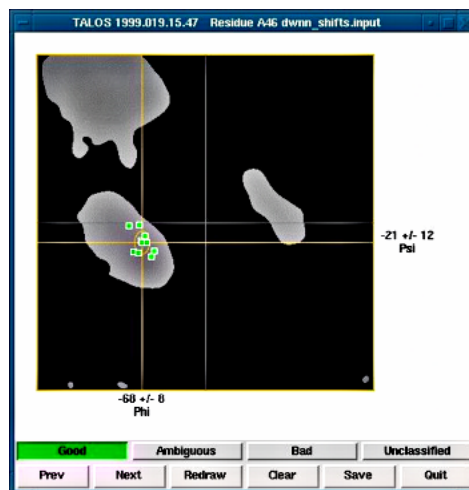
The TALOS output for the DWNN domain is depicted graphically in Fig. 4.13A. It shows good dihedral angle predictions for 72% of residues, while predictions for 28% of residues were ambiguous. The TALOS suite provides a graphical user interface for the inspection of the closest matches for each residue, which allows the user to control the matches used in the prediction. For the purpose of illustration, the ϕ and ψ angle predictions for Ala46 are shown in Fig. 4.13B and C. These are $-68^\circ \pm 8^\circ$ and $-21^\circ \pm 12^\circ$, respectively. The TALOS ϕ and ψ angle restraints for the DWNN domain are listed in Appendix 3.

4.5 Summary

Complete backbone chemical shift assignments of the DWNN domain were obtained using triple resonance NMR spectroscopy. Chemical shift data were extracted from



A



B

Residue A46, Triplet K45 A46 A47							
■	-69	-19	17.82	N24	A25	K26	bpti
■	-68	-21	17.87	N86	L87	D88	cvn
■	-82	-3	17.89	D35	L36	N37	cvn
■	-77	-31	18.07	S125	A126	K127	hca_I
■	-63	-21	19.50	Q164	S165	N166	alpha_LP
■	-72	-32	21.70	D56	A57	D58	calmodulin
■	-71	-2	22.02	D41	T42	S43	hca_I
■	-65	-14	23.50	T149	S150	A151	alpha_LP
■	-55	-30	24.23	N118	K119	D120	mbp
■	-58	-36	24.62	D134	S135	A136	cutinase
┘	-68	-21	20.72				Average

C

Fig. 4.13: TALOS output for the DWNN domain: (A) Sequence window; backbone chemical shifts for residues highlighted in green shows good matches to at least ten proteins in the TALOS database (B) The Ramachandran window for residue A46 shows that the chemical shift values for A46 match those of database proteins in which the residue falls within an α -helix (C) The prediction window for A46 lists the proteins in the database with closely matching chemical shifts and predicts the ϕ and ψ angles for A46 based on these matches. The predicted ϕ and ψ angles for A46 are -68° and -21° respectively.

^{15}N -HSQC, CBCA(CO)NH, CBCANH, ^{15}N -HSQC-NOESY and HNC0 spectra. A HMQC-J spectrum was used for the determination of $^3J_{\text{H}^{\text{N}}-\text{H}^{\alpha}}$ coupling constants, which were used for the estimation of backbone ϕ angle restraints. The TALOS software suite was used to predict ϕ and ψ angles of each residue based on the chemical shifts of the surrounding nuclei.



Chapter 5: Discussion and conclusion

5.1 Introduction

Molecular modelling of the DWNN domain indicated that the domain might assume an ubiquitin-like fold. In addition to the DWNN domain, the full-length RBBP6 protein also contains a RING finger domain, which is a motif found in a number of E3 ubiquitin ligases, as well as C-terminal domains that have been shown to interact with p53 and Rb. Confirmation that the DWNN domain assumes an ubiquitin-like fold might therefore point to a possible role for RBBP6 in regulating the levels of p53 or Rb. We therefore decided to determine the structure of the DWNN domain directly using NMR. The aims of this study were to express the DWNN domain using a bacterial expression system and to fully assign the backbone resonances as a prelude to full structure determination.

5.2 Recombinant expression and purification of the DWNN domain

The DWNN domain was expressed as a GST fusion protein using the pGEX-6P-2 expression vector (Section 2.2.3) and purified using a combination of affinity and gel filtration chromatography (Sections 2.2.4.2 and 2.2.4.4). To perform heteronuclear triple resonance NMR experiments the domain was double labelled with ^{13}C and ^{15}N by growing the bacterial cells on minimal media containing $^{15}\text{N}\text{-H}_4\text{Cl}$ and $^{13}\text{C}\text{-glucose}$ as the sole nitrogen and carbon sources respectively (Section 2.2.3.4). The expression and purification strategy employed in this study yielded highly purified samples of the

DWNN domain (see Fig. 3.7). The concentration of DWNN samples ranged from 1.45 mM to 2.33 mM (Section 3.1.4).

5.3 Full backbone assignment of the DWNN domain

DWNN domain backbone resonances were assigned using a combination of NMR spectra. Peaks in the ^{15}N -HSQC spectrum were randomly numbered and used to generate 2D strip plots from 3D-CBCA(CO)NH and CBCANH spectra. The strip plots were used to establish connectivities between the backbone resonances, as described in Section 4.3.1. Sequential connectivities were confirmed by the identification of NOE cross-peaks between sequential H^{N} groups, using a 3D ^{15}N -HSQC-NOESY spectrum (Section 4.3.2). The peaks in the ^{15}N -HSQC spectrum were then re-numbered according to residue numbers, and the backbone ^{15}N and H^{N} chemical shifts extracted (Section 4.2). C_α and C_β chemical shifts were extracted from the CBCA(CO)NH and CBCANH strip plots, while carboxyl carbon chemical shifts were extracted using an HNC(O) spectrum (Section 4.3.3).

5.4 Extraction of secondary structure information from backbone chemical shifts

Restraints on the ϕ dihedral angles were obtained by extracting $^3J_{\text{H}^{\text{N}}-\text{H}^\alpha}$ coupling constants from an HMQC-J spectrum (Section 4.4.1). The coupling constants were determined by fitting the peak splittings in the HMQC-J spectrum and the ϕ angles obtained using the Karplus Relation (Eqn. 1) to determine the backbone ϕ angle. Based on the Karplus Curve (Fig. 4.12), splittings of less than 6 Hz correspond to α -helical

regions, while splittings above 8 Hz correspond to β -sheet regions. No prediction could be made for splittings in the range of 6-8 Hz.

In addition to the HMQC-J splittings, the TALOS software suite was used to predict the backbone dihedral φ and ψ angles based on the backbone $^1\text{H}_\alpha$, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, ^{13}CO and ^{15}N chemical shifts (Section 4.4.2).

5.5 Backbone geometry of the DWNN domain

The DWNN domain backbone geometry was predicted by the Jpred secondary structure prediction server (Cuff *et al.*, 1998) to form secondary structural motifs as illustrated in Fig. 5.1. The server predicted that the domain forms four β -sheets and a central α -helix (see Table 5.1, column 2). Two alternate methods were used to validate the predicted secondary structure of the DWNN domain. Firstly, the Karplus relation was used to assign secondary structural motifs based on HMQC-J splittings. Secondly, the chemical shift index (CSI), based on backbone chemical shift data, was used to determine the secondary structural elements formed by the backbone. Table 5.1 summarizes the secondary structural motifs formed by the DWNN domain backbone as determined using these methods.

Based on the HMQC-J splittings (see Appendix 2), secondary structural motifs were assigned to regions of the DWNN domain backbone. The HMQC-J data suggests that the DWNN domain backbone forms six β -sheet regions, a central α -helix and a short α -

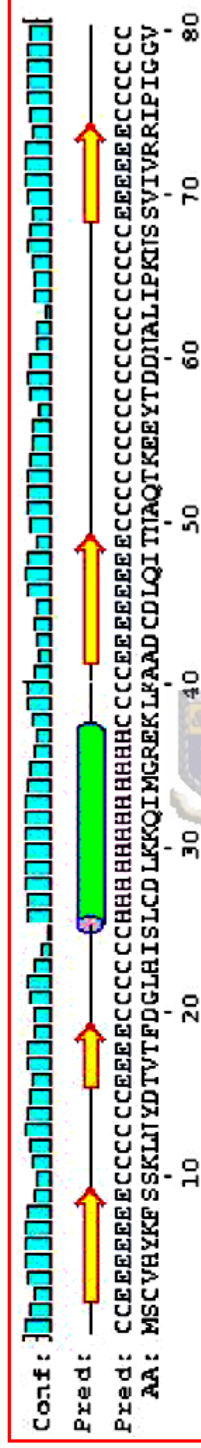



Fig. 5.1: Jpred prediction of the secondary structural motifs formed by the DWNN backbone. Two alternate methods are used to display the predicted secondary structural motifs: (i) a cartoon where the yellow arrows indicate β -sheets and the green barrel indicates an α -helix and (ii) one letter symbols where C indicates a coiled region, E refers to a β -sheet and H indicates an α -helix. The backbone forms four β -sheet regions and one central α -helix.

helical region, involving residues A46-D48. Table 5.1, column 3, lists these secondary structural elements of the DWNN domain.

In addition to the HMQC-J splittings, the CSI was also used to determine the secondary structural elements of the DWNN domain (see Fig. 5.2). The CSI was calculated using the random coil chemical shift database as compiled by Wishart and Sykes (1994). Wishart and co-workers (Wishart *et al.*, 1992) found that $^1\text{H}^\alpha$ chemical shifts are shifted to lower ppm values relative to random coil values in α -helical regions, while showing an upward shift in β -sheets. Similar findings were made for $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{13}CO chemical shifts (Wishart and Sykes, 1994). Based on the CSI, the DWNN domain backbone forms secondary structural elements as listed in Table 5.1, column 4.



Backbone region (aa)	Jpred prediction	HMQC-J splittings	CSI
6-14	β	β	β
19-25	β	β	β
27-30	-	β	β
31-42	α	α	α
46-48	-	α	-
49-54	β	β	β
57-59	-	β	-
74-77	β	β	β

Table 5.1: Summary of the secondary structural motifs formed by the DWNN backbone, as determined using the methods as indicated.

The two methods based on experimentally obtained data confirmed the secondary structural motifs formed by the DWNN domain, as predicted by Jpred. However, the

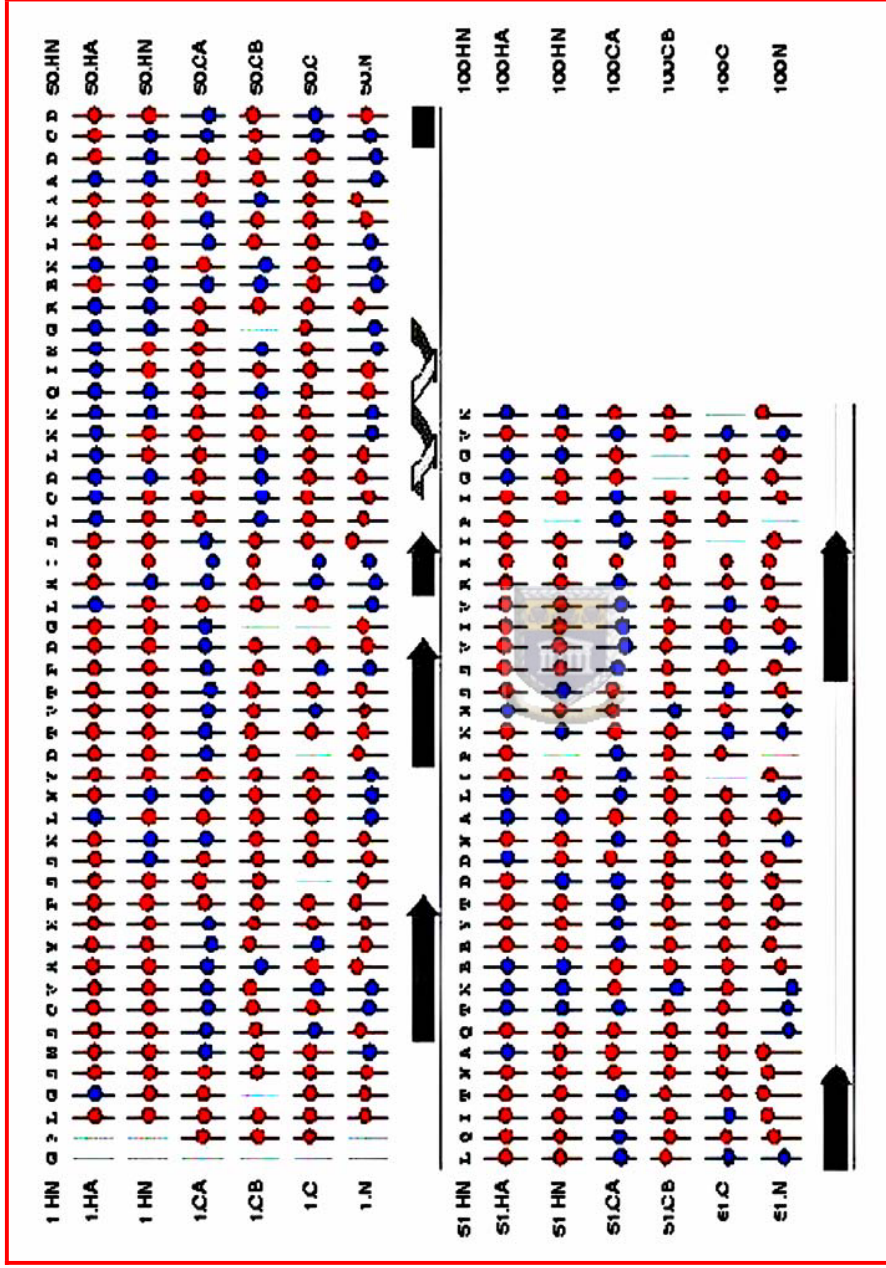


Fig. 5.2 Chemical shift index (CSI) of the DWN domain. The CSI was calculated using the method of Wishart et al (1994). The atoms used for the CSI calculation are indicated above. Positive chemical shift values relative to random coil values are indicated by red circles, while negative chemical shift values relative to random coil values are indicated by blue circles. Secondary structural elements are as indicated: arrows indicate β -sheet regions while the helix indicates an α -helical region of the DWN domain.

HMQC-J data and the CSI identified additional secondary structural elements that were not predicted by Jpred. These include a short β -sheet formed by residues L27-S30, identified from the HMQC-J data and the CSI. In addition, the HMQC-J data suggests that the DWNN backbone forms an additional short α -helix, involving residues A46-D48 and a short β -sheet region involving residues Q57-K59.

5.6 Conclusion


^{15}N - and ^{13}C -labeled samples of the DWNN domain were successfully expressed and used to fully assign all backbone resonances. This represents the first time isotopically enriched protein samples had been produced in South Africa, as well as the first time that a protein had been successfully assigned in South Africa using heteronuclear NMR.



The assignment formed the basis of the subsequent three dimensional structure determination of the DWNN domain (Faro A., Lutya PT., Rees DJG. and Pugh DJR., unpublished results). The structure confirms that the DWNN domain does indeed assume an ubiquitin-like fold. Consistent with this result, in much the same way that ubiquitin becomes conjugated to proteins such as p53, protein transduction experiments suggest that the DWNN domain may become conjugated to other proteins in the cell (Rees *et al.*, unpublished results). Future work will focus on identification of DWNN interaction partners and analysis of the binding interactions using heteronuclear NMR. The full set of assignments recorded in this thesis will greatly facilitate in the analysis of these interactions.

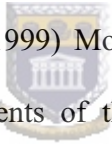
References

- Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ., (1990) Basic local alignment search tool. *The Journal of Molecular Biology* **215**, 403-410
- Balakirev MY., Tcherniuk SO., Jaquinod M., Chrobozek J. (2003) Otubains: a new family of cysteine proteases in the ubiquitin pathway. *EMBO Reports* **4**, 517-522
- Bayer P., Arndt A., Metzger S., Mahajan R., Melchior F., Jaenicke R., Becker J. (1998) Structure determination of the small ubiquitin-related modifier SUMO-1. *Journal of Molecular Biology* **280**, 275-286
- Bernier-Villamore V., Sampson DA., Matunis MJ., Lima CD. (2002) Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugation enzyme Ubc9 and RanGap1. *Cell* **108**, 345-356
- Bertolaet BL., Clarke DJ., Wolff M., Watson MH., Henze M., Divita G., Reed SI. (2001) UBA domains of DNA damage-inducible proteins interact with ubiquitin. *Nature Structural Biology* **8**, 417-422
- Beyer A. (1997) Sequence analysis of the AAA protein family. *Protein Science* **6**, 2043-2058
- Blum H., Beier H., Gross HJ. (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8**, 93-99.
- Bradford MM. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**, 248-254
- Bukau B., Horwich AL. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351-366

- Cavanagh J., Fairbrother WJ., Palmer AG., Skelton NJ. (1996) Protein NMR spectroscopy: principles and practice. Academic Press, Inc., San Diego.
- Chen L., Madura K. (2002) Rad23 promotes the targeting of proteolytic substrates to the proteasome. *Molecular and Cellular Biology* **22**, 4902-4913
- Chen L., Shinde U., Ortolan TG., Madura K. (2001) Ubiquitin-associated (UBA) domains in Rad23 bind ubiquitin and promotes inhibition of multi-ubiquitin chain assembly. *EMBO Reports* **2**, 933-938
- Choudbury BK., Li SS.-L. (1997) Identification and characterization of the SMT3 cDNA and gene from nematode *Caenorhabditis elegans*. *Biochemical and Biophysical Research Communications* **234**, 788-791
- Cornilescu G., Delaglio F., Bax A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* **13**, 289-302 
- Cuff JA., Clamp ME., Siddiqui AS., Finlay M. Barton G J. (1998) Jpred: A consensus secondary structure prediction server. *Bioinformatics* **14**, 892-893
- de Laat WL., Jaspers NGJ., Hoeimakers JHJ. (1999) Molecular mechanism of nucleotide excision repair. *Genes and Development* **13**, 768-785
- Delaglio F., Grzesiek S., Vuister GW., Zhu G., Pfeifer J., Bax A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293
- Delano WL. "The PyMOL Molecular Graphics System." Delano Scientific LLC, San Carlos, CA, USA. <http://www.pymol.org>)

- Desterro JM., Rodriguez MS., Kemp GD. Hay RT. (1999) Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *Journal of Biological Chemistry* **274**, 10618-10624
- Desterro JMP., Thomson J., Hay RT. (1997) Ubch9 conjugates SUMO but not ubiquitin. *FEBS Letters* **417**, 297-300
- Deveraux Q., Ustrell V., Pickart C., Rechsteiner M. (1994) A 26S protease subunit that binds ubiquitin conjugates. *The Journal of Biological Chemistry* **269**, 7059-7061
- Evans PC., Smith TS., Lai MJ., Williams MG., Burke DF., Heyninck K., Kreike MM., Beyart R., Blundell TL., Kilshaw PJ (2003) A novel type of deubiquitinating enzyme. *Journal of Biological Chemistry* **278**, 23180-23186
- Eytan E., Armon T., Heller H., Beck S. Hershko A. (1993) Ubiquitin C-terminal hydrolase activity associated with the 26S protease complex. *Journal of Biological Chemistry* **268**, 4668-4674
- Finley D., Bartel B., Varshavsky A. (1989) The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature* **338**, 394-401
- Fischer D. (2000) Hybrid fold recognition: combining sequence derived properties with evolutionary information. *Pacific Symposium on Biocomputing* **5**, 116-127
- Gao S. Scott RE. (2003) Stable overexpression of specific segments of the P2P-R protein in human MCF-7 cells promotes camptothecin-induced apoptosis. *Journal of Cellular Physiology* **197**, 445-452



- Gao S., Scott RE. (2002). P2P-R overexpression restricts mitotic progression at the prometaphase and promotes mitotic apoptosis. *Journal of Cellular Physiology* **193**, 199-207
- Gao S., Witte MM., Scott RE. (2002) P2P-R localizes to the nucleolus of interphase cells and the periphery of chromosomes in the mitotic cells which show maximum P2P-R immunoreactivity. *Journal of Cellular Physiology* **191**, 145-154
- Glazko GV., Koonin EV., Rogozin IB., Shabalina SA. (2003) A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends in Genetics* **19**, 119-124
- Glickman MH., Ciechanover A. (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiological Review* **82**, 373-428
- Gong L., Li B., Millas S., Yeh ET. (1999)  Molecular cloning and characterization of human AOS1 and UBA2 components of the sentrin-activating enzyme complex. *FEBS Letters* **448**, 185-189
- Gorlich D., Kutay U., (1999) Transport between the cell nucleus and the cytoplasm. *Annual Review of Cell Development and Biology* **15**, 607-660
- Gostissa M., Hengstermann A., Fogal V., Sandy P., Schwarz SE., Scheffner M., Del Sal G. (1999) Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1. *EMBO Journal* **18**, 6462-6471
- Groll M., Bajorek M., Kohler A., Moroder L., Rubin DM., Huber R., Glickman MH., Finley D. (2000) A gated channel into the proteasome core particle. *Nature Structural Biology* **7**, 1062-1067

- Groll M., Ditzel I., Löwe J., Stock D., Bochtler M., Bartunik HD., Huber R. (1997) Structure of the 26S proteasome from yeast at 2.4 Å resolution. *Nature* **386**, 463-471
- Grzesiek S. Bax A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance* **99**, 201-207
- Grzesiek S. Bax A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **114**, 6291-6293
- Gudzer SN., Bailly V., Sung P., Prakash L., Prakash S. (1995a) Yeast DNA repair protein RAD23 promotes complex formation between transcription factor TFIIH and DNA damage recognition factor RAD14. *Journal of Biological Chemistry* **270**, 8385-8388
- Gudzer SN., Habraken Y., Sung P., Prakash L., Prakash S. (1995b) Reconstitution of yeast nucleotide excision repair with purified Rad proteins, replication protein A, and transcription factor TFIIH. *Journal of Biological Chemistry* **270**, 12973-12976
- Gudzer SN., Sung P., Prakash S. (1998) Affinity of yeast nucleotide excision repair factor 2, consisting of the Rad4 and Rad23 proteins, for ultraviolet damaged DNA. *Journal of Biological Chemistry* **273**, 31541-31546
- Haas AL., Ahrens P., Bright PM., Ankel H. (1987) Interferon induces a 15-kilodalton protein exhibiting marked homology to ubiquitin. *Journal of Biological Chemistry* **262**, 11135-11323
- Haas AL., Siepmann TJ. (1997) Pathways of ubiquitin conjugation. *FASEB Journal* **11**, 1257-1268

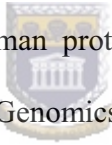
- Hanania U., Furman-Matasaro N., Ron M., Avni A. (1999) Isolation of a novel SUMO protein from tomato that suppresses EIX-induced cell death. *Plant Journal* **19**, 533-541
- Hartmann-Petersen R., Seeger M., Gordon C. (2003) Transferring substrates to the 26S proteasome. *Trends in Biochemical Sciences* **28**, 26-31
- Hershko A., Ciechanover A. (1998) The ubiquitin system. *Annual Review of Biochemistry* **67**, 425-479
- Hochstrasser M. (2002) New proteases in a ubiquitin stew. *Science* **298**, 549-552
- Hochstrasser M., Johnson PR., Arendt CS., Amerik AY., Swaminathan S., Swanson R., Li S., Laney J., Pals-Rylaarsdam R., Nowak J., Connerly PL (1999) The *Saccharomyces cerevisiae* ubiquitin-proteasome system. *Philosophical Transactions, Royal Society of London* **354**, 1513-1522
- Huang HW., Tsoi SCM., Sun YH., Li SS-L. (1998) Identification and characterization of the SMT3 cDNA and gene encoding ubiquitin-like protein from *Drosophila melanogaster*. *Biochemistry and Molecular Biology International* **46**, 775-785
- Jackson PK. (2001) A new RING for SUMO: wrestling transcriptional responses into nuclear bodies with the PIAS family of E3 SUMO ligases. *Genes and Development* **15**, 3053-3058
- Jansen LET., Verhage RA., Brouwer J. (1998) Preferential binding of yeast Rad4-Rad23 complex to damaged DNA. *Journal of Biological Chemistry* **273**, 33111-33114
- Jentsch S., Pyrowolakis G. (2000) Ubiquitin and its kin: how close are the family ties? *Trends in Cell Biology* **10**, 335-342

- Jin C., Shiyanova T., Shen Z., Liao X. (2001) Heteronuclear magnetic resonance assignments, structure and dynamics of SUMO-1, a human ubiquitin-like protein. *International Journal of Biological Macromolecules* **28**, 227-234
- Johnson BA, Blevins RA (1994) NMRView: a computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR* **4**:603-614
- Johnson ES., Schwienhorst I., Dohmen RJ., Blobel G. (1997) The ubiquitin-like protein Smt3 is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. *EMBO Journal* **16**, 5509-5519
- Johnson PR., Hochstrasser M. (1997) SUMO-1: ubiquitin gains weight. *Trends in Cell Biology* **7**, 408-413
- Kamitani T., Nguyen HP., Kito K., Fukuda-Kamitani T., Yeh ET. (1998) Covalent modification of PML by the sentrin family of ubiquitin-like proteins. *Journal of Biological Chemistry* **273**, 3117-3120
- Katami T., Kito K., Nguyen HP., Fukuda-Kamitani T., Yeh ET., (1998) Characterization of a second member of the sentrin family of ubiquitin-like proteins. *Journal of Biological Chemistry* **273**, 11349-11353
- Kay LE., Ikura M., Tschudin R., Bax A. (1990) Three dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance* **89**, 496-514
- Kay, L.E. and Bax, A. (1990) New methods for the measurement of NH-CH coupling constants in N-15-labeled proteins. *Journal of Magnetic Resonance* **86**: 110–126

- Kelly LA., MacCallum RM., Sternberg MJE. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *Journal of Molecular Biology* **299**, 499-520
- Kim JH., Park KC., Chung SS., Bang O., Chung CH. (2003) Deubiquitinating enzymes as cellular regulators. *Journal of Biochemistry* **134**, 9-18
- Kim KI., Baek SH., Chung CH. (2002) Versatile protein tag, SUMO: its enzymology and biological function. *Journal of Cellular Physiology* **191**, 257-268
- King RC., Storto PD. (1988) The role of the *otu* gene in *Drosophila* oogenesis. *Bioassays* **8**, 18-24
- Kohler A., Cascio P., Leggett DS., Woo KM., Goldberg AL., Finley D (2001) The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. *Molecular Cell* **7**, 1143-1152
- Laemmli UK. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685
- Li SJ., Hochstrasser M. (1999) A new protease required for cell-cycle progression in yeast. *Nature* **398**, 246-251
- Li SJ., Hochstrasser M. (2000) The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. *Molecular and Cellular Biology* **20**, 2367-2377
- Löwe J., Stock D., Jap B., Zwickl P., Baumeister W., Huber R. (1995) Crystal structure of the 20S proteasome from the archaeon *T.acidophilum* at 3.4 Å resolution. *Science* **268**, 533-539

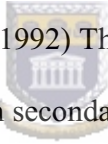
- Lüders J., Demand J. Höhfeld J. (2000) The ubiquitin-related BAG1 provides a link between the molecular chaperones Hsc70/Hsp70 and the proteasome. *Journal of Biological Chemistry* **275**, 4613-4617
- Mahajan R., Delphin C., Guan T., Gerace L. Melchior F. (1997) A small ubiquitin-related polypeptide involved in targeting RanGap1 to nuclear pore complex RanBP2. *Cell* **88**, 97-101
- Matunis MJ., Coutavas E., Blobel G. (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGap1 between the cytosol and the nuclear pore complex. *Journal of Cell Biology* **135**, 1457-1470
- Melchior F., Pascal B., Evans J., Gerace L. (1993) Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase RAN/TC4 as an essential transport factor. *Journal of Cell Biology* **123**, 1649-1659
- Melchior F., Schergaut M., Pichler A. (2003) SUMO: ligases, isopeptidases and nuclear pores. *Trends in Biochemical Sciences* **105**, 1-7
- Muller S., Berger M., Lehembre F., Seeler JS., Haupt Y., Dejean A. (2000) c-Jun and p53 activity is modulated by SUMO-1 modification. *Journal of Biological Chemistry* **275**,13321-13329
- Muller S., Hoege C., Pyrowolakis G. Jentsch S. (2001) SUMO, ubiquitin's mysterious cousin. *Nature Reviews Molecular Cell Biology* **2**, 202-210
- Muller S., Matunis MJ., Dejean A. (1998) Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *EMBO Journal* **17**, 61-70

- Nollen EAA., Brunsting JF., Song J., Kampinga HH., Morimoto RI. (2000) Bag1 functions in vivo as a negative regulator of Hsp70 chaperone activity. *Molecular and Cellular Biology* **20**, 1083-1088
- Ogura T., Wilkinson AJ. (2001) AAA+ superfamily ATPases: common structure-diverse function. *Genes and Cells* **6**, 574-597
- Okuma T., Honda R., Ichikawa G., Tsumagari N., Yasuda H. (1999) In vitro SUMO-1 modification requires two enzymatic steps, E1 and E2. *Biochemical and Biophysical Research Communications* **254**, 693-698
- Ortolan TG., Tongaonkar P., Lambertson D., Chen L., Schaubert C., Madura K. (2000) The DNA repair protein Rad23 is a negative regulator of multi-ubiquitin chain assembly. *Nature Cell Biology* **2**, 601-607
- Pardi A., Billeter M., Wütrich K. (1984) Calibration of the angular dependence of the amide-proton- C_{α} proton coupling constant $^3J_{H^N H_{\alpha}}$, in a globular protein. *Journal of Molecular Biology* **180**, 741-751
- Patel S., Latterich M. (1998) The AAA team: related ATPases with diverse functions. *Trends in Cell Biology* **8**, 65-71
- Perkins EL., Sterling JF., Hashem VI., Resnick MA. (1999) Yeast and human genes that affect the *Escherichia coli* SOS response. *Proceedings of the National Academy of Sciences* **96**, 2204-2209
- Peters JM. (1994) Proteasomes: protein degradation machines of the cell. *Trends in Biochemical Science* **19**, 377-382
- Pichler A., Gast A., Seeler JS., Dejean A., Melchior F. (2002) Nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell* **108**, 109-120

- Primrose WU. (1993) Sample preparation. In *NMR of macromolecules: a practical approach*. Edited by Roberts GCK, Oxford University Press Inc., New York, p7-34
- Rehm T., Huber R., Holak TA. (2002) Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. *Structure* **10**, 1613-1618
- Rodriguez MS., Desterro JM., Lain S., Midgeley CA., Lane DP., Hay RT. (1999) SUMO-1 modification activates the transcriptional response of p53. *EMBO Journal* **18**, 6455-6461
- Saeki Y., Sone T., Toh-e A., Yokosawa H. (2002) Identification of ubiquitin-like protein-binding subunits of the 26S proteasome. *Biochemical and Biophysical Research Communications* **296**, 813-819
- Saijo M., Sakai Y., Kishino T., Niikawa N., Matsuura Y., Morino K., Tamai K., Taya Y. (1995) Molecular cloning of a human protein which binds to the retinoblastoma protein and chromosomal mapping.  *Genomics* **27**, 511-519
- Sakai Y., Saijo M., Coelho K., Kishino T., Niikawa N., Taya Y. (1995) cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP60, that binds to the retinoblastoma gene product. *Genomics* **30**, 98-101
- Sattler M., Schleucher J., Griesinger C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**, 93-158
- Schauber C. Chen L. Tongaonkar P., Vega I., Lambertson D., Potts W., Madura K. (1998) Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature* **391**, 715-718

- Scheich C., Leitner D., Sievert V., Leidert M., Schlegel B., Simon B., Letunic I., Büsow K., Diehl A. (2004) Fast identification of folded human protein domains expressed in *E.coli* suitable for structural analysis. *BioMed Central Structural Biology* **4**, 4-12
- Schwienhorst I., Johnson ES., Dohmen RJ. (2000) SUMO conjugation and deconjugation. *Molecular Gene and Genetics* **263**, 771-786
- Scott RE., Giannakouros T., Gao S., Peidis P. (2003) Functional potential of P2P-R: a role in the cell cycle and cell differentiation related to its interactions with proteins that bind to matrix associated regions of DNA. *Journal of Cellular Biochemistry* **90**, 6-12
- Simons A., Melamed-Bessudo C., Wolkowicz R., Sperling J., Sperling R., Eisenbach L., Rotter V. (1997) PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb. *Oncogene* **14**, 145-155
- Smith DB., Johnson KS. (1988) Single step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**, 31-40
- Sternsdorf T., Jensen K., Will H. (1997) Evidence for covalent modification of the nuclear dot-associated proteins PML and Sp100 by PIC1/SUMO-1. *Journal of Cell Biology* **139**, 1621-1634
- Stott FJ., Bates S., James MC., McConnell BB., Starborg M., Brookes S., Palmero I., Ryan K., Hara E., Vousden KH., Peters G. (1998) The alternative product from the human CDKN2A locus, p14 (ARF), participates in a regulatory loop with p53 and MDM2. *EMBO Journal* **17**, 5001-5014
- Su H-L., Li SS-L. (2002) Molecular features of human ubiquitin-like SUMO genes and their encoded proteins. *Genes* **296**, 65-73

- Sweder K. Madura K. (2002) Regulation of repair by the 26S proteasome. *Journal of Biomedicine and Biotechnology* **2**, 94-105
- Takayama S., Krajewski S., Krajewski M., Kitada S., Zapata JM., Kochel K., Knee D., Scudiero D., Tudor G., Miller GJ., Miyashita T., Yamada M., Reed JC. (1998) Expression and location of Hsp70/Hsc-binding anti-apoptotic protein BAG-1 and its variants in normal tissues and tumor cell lines. *Cancer Research* **58**, 3116-3131
- Upadhy SC., Hegde AN. (2003) A potential proteasome-interacting motif within the ubiquitin-like domain of parkin and other proteins. *Trends in Biochemical Sciences* **28**, 280-283
- Verma R., Aravind L., Oania R., McDonald WH., Yates JR., Koonin EV., Deshaies RJ. (2002) Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**, 611-615
- Vijay-Kumar S., Bugg CE., Cook WJ. (1987) Structure of ubiquitin refined at 1.8 Å resolution. *Journal of Molecular Biology* **194**, 531-544.
- Vo LTA., Minet M. Schmitter J. Lacroute F., Wyers F. (2001) Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. *Molecular and Cellular Biology* **21**, 8346-8356
- Watkins JF., Sung P. Prakash L. Prakash S. (1993) The *Saccharomyces cerevisiae* DNA repair gene RAD23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Molecular and Cellular Biology* **13**, 7757-7765

- Wiborg O., Pedersen MS., Wind A., Bergland LE., Marcker KA., Vuust J. (1985) The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. *The EMBO Journal* **4**, 755-759
- Wilkinson CRM., Seger M., Hartmann-Petersen R., Stone M., Wallace M., Semple C., Gordon C. (2001) Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nature Cell Biology* **3**, 939-943
- Wilkinson KD., Urban MK, Haas AL (1980), Ubiquitin is the ATP-dependent proteolysis factor-1 of rabbit reticulocytes. *The Journal of Biological Chemistry* **255**, 7529-7532
- Wishart DS., Sykes BD. (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *Journal of Biomolecular NMR* **4**, 171-180
- Wishart DS., Sykes BD. Richards FM. (1992)  The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-1651
- Witte MM., Scott RE. (1997) The proliferation potential protein-related (P2P-R) gene with domains encoding heterogeneous nuclear ribonucleoprotein association and Rb1 binding shows repressed expression during terminal differentiation. *Proceedings of the National Academy of Sciences* **94**, 1212-1217
- Woestenenk EA., Hammarström M., Härd T., Berglund H. (2003) Screening methods to determine biophysical properties of proteins in structural genomics. *Analytical Biochemistry* **318**, 71-79
- Wütrich K. (2003) NMR studies of structure and function of biological macromolecules (Nobel Lecture). *Journal of Biomolecular NMR* **27**, 13-39

- Yang X., Chernenko G., Hao Y., Ding Z., Pater MM., Pater A., Tang SC. (1998) Human BAG-1/RAP46 protein is generated as four isoforms by alternative translation initiation and overexpressed in cancer cells. *Oncogene* **17**, 981-989
- Yao T., Cohen RE. (2002) A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**, 403-407
- Yeh ETH., Gong L., Kamitani T. (2000) Ubiquitin-like proteins: new wines in new bottles. *Gene* **248**, 1-14



Appendix 1: DWNN domain chemical shifts

Residue no.	Residue name	Atom name	Chemical shift
2	P	C _α	63.98
2	P	C _β	32.50
2	P	C	178.47
3	L	N	124.92
3	L	HN	8.59
3	L	C _α	55.82
3	L	H _α	4.38
3	L	C _β	42.24
3	L	C	179.42
4	G	N	111.87
4	G	HN	8.42
4	G	C _α	45.67
4	G	H _{α2}	4.00
4	G	H _{α1}	4.00
4	G	C	175.55
5	S	N	118.16
5	S	HN	8.30
5	S	C _α	58.24
5	S	H _α	4.50
5	S	C _β	64.02
5	S	C	175.84
6	M	N	124.04
6	M	HN	8.25
6	M	C _α	56.12
6	M	H _α	4.52
6	M	C _β	33.50
6	M	C	177.91
7	S	N	121.87
7	S	HN	8.46
7	S	C _α	57.34
7	S	H _α	4.49
7	S	C _β	64.77
7	S	C	173.12
8	C	N	117.85
8	C	HN	8.43
8	C	C _α	56.17
8	C	H _α	5.57
8	C	C _β	32.22
8	C	C	175.41
9	V	N	120.67
9	V	HN	8.53
9	V	C _α	61.57
9	V	H _α	4.40

Residue no.	Residue name	Atom name	Chemical shift
9	V	C _β	36.74
9	V	C	174.89
10	H	N	127.12
10	H	HN	8.72
10	H	C _α	54.08
10	H	H _α	5.79
10	H	C _β	30.81
10	H	C	176.20
11	Y	N	124.30
11	Y	HN	9.74
11	Y	C _α	55.07
11	Y	H _α	6.29
11	Y	C _β	44.11
11	Y	C	173.74
12	K	N	123.80
12	K	HN	8.49
12	K	C _α	54.34
12	K	H _α	4.84
12	K	C _β	35.29
12	K	C	177.27
13	F	N	128.18
13	F	HN	9.18
13	F	C _α	57.96
13	F	H _α	5.30
13	F	C _β	40.27
13	F	C	178.89
14	S	N	120.36
14	S	HN	8.71
14	S	C _α	60.85
14	S	H _α	4.72
14	S	C _β	63.04
15	S	N	116.60
15	S	HN	8.10
15	S	C _α	58.88
15	S	H _α	4.46
15	S	C _β	63.57
15	S	C	175.84
16	K	N	124.62
16	K	HN	7.94
16	K	C _α	55.80
16	K	C _β	34.08
16	K	C	177.62
17	L	N	121.17
17	L	HN	8.34
17	L	C _α	56.31
17	L	H _α	4.30
17	L	C _β	43.40

Residue no.	Residue name	Atom name	Chemical shift
17	L	C	178.67
18	N	N	117.28
18	N	HN	7.91
18	N	C _α	52.29
18	N	H _α	4.98
18	N	C _β	41.17
18	N	C	175.46
19	Y	N	120.99
19	Y	HN	8.64
19	Y	C _α	58.89
19	Y	H _α	4.53
19	Y	C _β	40.59
19	Y	C	177.62
20	D	N	126.29
20	D	HN	9.05
20	D	C _α	52.91
20	D	H _α	4.99
20	D	C _β	44.49
21	T	N	116.78
21	T	HN	9.00
21	T	C _α	62.58
21	T	H _α	5.34
21	T	C _β	72.69
21	T	C	176.36
22	V	N	124.92
22	V	HN	8.88
22	V	C _α	61.09
22	V	H _α	4.59
22	V	C _β	34.65
22	V	C	176.19
23	T	N	118.21
23	T	HN	8.33
23	T	C _α	59.82
23	T	H _α	5.48
23	T	C _β	71.55
23	T	C	175.70
24	F	N	119.80
24	F	HN	8.14
24	F	C _α	56.54
24	F	H _α	4.82
24	F	C _β	39.42
24	F	C	171.86
25	D	N	121.20
25	D	HN	8.71
25	D	C _α	53.74
25	D	H _α	5.05
25	D	C _β	42.81

Residue no.	Residue name	Atom name	Chemical shift
25	D	C	177.71
26	G	N	113.07
26	G	HN	8.37
26	G	C _α	44.51
26	G	H _{α2}	3.93
26	G	H _{α1}	4.62
27	L	N	120.46
27	L	HN	8.64
27	L	C _α	56.85
27	L	H _α	4.14
27	L	C _β	43.09
27	L	C	178.49
28	H	N	116.00
28	H	HN	7.70
28	H	C _α	54.24
28	H	H _α	5.16
28	H	C _β	35.17
28	H	C	173.81
29	I	N	122.92
29	I	HN	8.65
29	I	C _α	58.11
29	I	H _α	4.30
29	I	C _β	40.64
29	I	C	174.21
30	S	N	126.63
30	S	HN	9.12
30	S	C _α	57.52
30	S	H _α	5.07
30	S	C _β	64.83
30	S	C	177.08
31	L	N	125.90
31	L	HN	8.71
31	L	C _α	58.57
31	L	H _α	3.81
31	L	C _β	41.26
31	L	C	179.83
32	C	N	118.25
32	C	HN	8.73
32	C	C _α	61.85
32	C	H _α	4.03
32	C	C _β	27.00
32	C	C	178.41
33	D	N	124.87
33	D	HN	7.76
33	D	C _α	57.46
33	D	H _α	4.49
33	D	C _β	39.71

Residue no.	Residue name	Atom name	Chemical shift
33	D	C	179.93
34	L	N	125.68
34	L	HN	8.88
34	L	C _α	58.45
34	L	H _α	3.84
34	L	C _β	41.21
34	L	C	179.75
35	K	N	119.81
35	K	HN	8.80
35	K	C _α	61.38
35	K	H _α	3.54
35	K	C _β	32.78
35	K	C	179.42
36	K	N	119.73
36	K	HN	7.66
36	K	C _α	60.30
36	K	H _α	3.99
36	K	C _β	32.66
36	K	C	181.78
37	Q	N	120.92
37	Q	HN	8.11
37	Q	C _α	59.29
37	Q	H _α	4.14
37	Q	C _β	28.92
37	Q	C	180.66
38	I	N	123.15
38	I	HN	8.37
38	I	C _α	66.30
38	I	H _α	3.24
38	I	C _β	37.61
38	I	C	178.85
39	M	N	119.40
39	M	HN	8.73
39	M	C _α	60.13
39	M	H _α	3.84
39	M	C _β	31.53
39	M	C	180.06
40	G	N	105.94
40	G	HN	7.90
40	G	C _α	47.21
40	G	H _α 2	3.99
40	G	H _α 1	3.99
40	G	C	178.51
41	R	N	125.39
41	R	HN	8.11
41	R	C _α	59.21
41	R	H _α	4.17

Residue no.	Residue name	Atom name	Chemical shift
41	R	C _β	30.75
41	R	C	179.87
42	E	N	117.08
42	E	HN	8.18
42	E	C _α	55.13
42	E	H _α	4.35
42	E	C _β	29.19
42	E	C	176.05
43	K	N	118.06
43	K	HN	7.67
43	K	C _α	57.12
43	K	H _α	3.97
43	K	C _β	28.68
43	K	C	177.41
44	L	N	121.24
44	L	HN	8.44
44	L	C _α	53.34
44	L	H _α	4.64
44	L	C _β	44.58
44	L	C	177.96
45	K	N	123.49
45	K	HN	8.78
45	K	C _α	55.10
45	K	H _α	4.52
45	K	C _β	33.54
45	K	C	178.43
46	A	N	129.51
46	A	HN	8.67
46	A	C _α	54.19
46	A	H _α	4.37
46	A	C _β	18.25
46	A	C	179.62
47	A	N	117.91
47	A	HN	7.92
47	A	C _α	53.41
47	A	H _α	4.17
47	A	C _α	19.22
47	A	C	179.22
48	D	N	116.12
48	D	HN	7.83
48	D	C _α	55.12
48	D	C _β	43.08
48	D	C	178.06
49	C	N	117.09
49	C	HN	7.88
49	C	C _α	56.76
49	C	H _α	4.99

Residue no.	Residue name	Atom name	Chemical shift
49	C	C _β	30.69
49	C	C	173.41
50	D	N	121.60
50	D	HN	8.72
50	D	C _α	51.50
50	D	H _α	5.00
50	D	C _β	44.46
50	D	C	176.23
51	L	N	120.95
51	L	HN	9.01
51	L	C _α	53.22
51	L	H _α	5.35
51	L	C _β	45.57
51	L	C	176.23
52	Q	N	125.68
52	Q	HN	9.09
52	Q	C _α	54.77
52	Q	H _α	4.68
52	Q	C _β	31.16
52	Q	C	176.74
53	I	N	131.61
53	I	HN	9.70
53	I	C _α	61.03
53	I	H _α	4.98
53	I	C _β	38.33
53	I	C	175.45
54	T	N	124.32
54	T	HN	8.93
54	T	C _α	60.14
54	T	H _α	5.13
54	T	C _β	71.96
54	T	C	175.59
55	N	N	127.41
55	N	HN	9.22
55	N	C _α	55.33
55	N	H _α	4.81
55	N	C _β	40.59
55	N	C	177.95
56	A	N	133.36
56	A	HN	9.48
56	A	C _α	55.16
56	A	H _α	4.08
56	A	C _β	19.83
56	A	C	179.27
57	Q	N	117.12
57	Q	HN	8.60
57	Q	C _α	57.78

Residue no.	Residue name	Atom name	Chemical shift
57	Q	H _α	4.52
57	Q	C _β	30.28
57	Q	C	178.96
58	T	N	110.17
58	T	HN	7.95
58	T	C _α	61.85
58	T	H _α	4.36
58	T	C _β	70.25
58	T	C	177.49
59	K	N	115.86
59	K	HN	7.79
59	K	C _α	57.78
59	K	H _α	3.94
59	K	C _β	29.03
59	K	C	176.54
60	E	N	122.34
60	E	HN	7.63
60	E	C _α	56.89
60	E	H _α	4.05
60	E	C _β	30.59
60	E	C	176.16
61	E	N	128.62
61	E	HN	8.71
61	E	C _α	55.41
61	E	H _α	4.48
61	E	C _β	30.55
61	E	C	177.60
62	Y	N	127.88
62	Y	HN	8.27
62	Y	C _α	58.35
62	Y	H _α	4.75
62	Y	C _β	39.80
62	Y	C	177.20
63	T	N	116.22
63	T	HN	8.81
63	T	C _α	61.85
63	T	H _α	4.59
63	T	C _β	70.74
63	T	C	175.36
64	D	N	126.00
64	D	HN	7.91
64	D	C _α	53.65
64	D	H _α	4.79
64	D	C _β	42.93
64	D	C	177.49
65	D	N	128.05
65	D	HN	8.81

Residue no.	Residue name	Atom name	Chemical shift
65	D	C _α	57.50
65	D	H _α	4.43
65	D	C _β	41.99
65	D	C	177.45
66	N	N	115.98
66	N	HN	8.47
66	N	C _α	52.80
66	N	H _α	4.80
66	N	C _β	39.01
66	N	C	176.91
67	A	N	126.60
67	A	HN	8.09
67	A	C _α	53.17
67	A	H _α	4.16
67	A	C _β	19.61
67	A	C	177.37
68	L	N	121.62
68	L	HN	8.18
68	L	C _α	53.72
68	L	H _α	4.35
68	L	C _β	42.81
68	L	C	177.48
69	I	N	130.00
69	I	HN	9.08
69	I	C _α	58.86
69	I	H _α	4.36
69	I	C _β	38.19
70	P	C _α	62.75
70	P	C _β	33.24
70	P	C	179.45
71	K	N	121.42
71	K	HN	8.13
71	K	C _α	57.43
71	K	H _α	4.30
71	K	C _β	32.88
71	K	C	175.91
72	N	N	116.11
72	N	HN	9.56
72	N	C _α	55.71
72	N	H _α	4.36
72	N	C _β	37.00
72	N	C	176.25
73	S	N	117.02
73	S	HN	7.63
73	S	C _α	60.55
73	S	H _α	4.73
73	S	C _β	64.15

Residue no.	Residue name	Atom name	Chemical shift
73	S	C	173.12
74	S	N	121.15
74	S	HN	9.21
74	S	C _α	57.58
74	S	H _α	5.83
74	S	C _β	65.22
74	S	C	175.78
75	V	N	118.41
75	V	HN	9.54
75	V	C _α	58.33
75	V	H _α	5.59
75	V	C _β	35.19
75	V	C	174.97
76	I	N	125.18
76	I	HN	9.49
76	I	C _α	59.21
76	I	H _α	4.90
76	I	C _β	39.11
76	I	C	177.27
77	V	N	128.66
77	V	HN	8.82
77	V	C _α	60.27
77	V	H _α	5.01
77	V	C _β	34.24
77	V	C	175.42
78	R	N	127.99
78	R	HN	8.92
78	R	C _α	54.57
78	R	H _α	5.00
78	R	C _β	34.14
78	R	C	176.54
79	R	N	127.38
79	R	HN	9.09
79	R	C _α	56.52
79	R	H _α	4.75
79	R	C _β	31.56
79	R	C	176.73
80	I	N	127.91
80	I	HN	8.87
80	I	C _α	57.66
80	I	H _α	4.82
80	I	C _β	39.53
81	P	C _α	62.72
81	P	C _β	32.28
81	P	C	178.25
82	I	N	123.21
82	I	HN	8.02

Residue no.	Residue name	Atom name	Chemical shift
82	I	C _α	62.34
82	I	H _α	4.16
82	I	C _β	38.84
82	I	C	178.42
83	G	N	115.60
83	G	HN	8.64
83	G	C _α	45.36
83	G	H _{α2}	4.02
83	G	H _{α1}	4.02
83	G	C	176.04
84	G	N	111.22
84	G	HN	8.31
84	G	C _α	45.22
84	G	H _{α2}	4.02
84	G	H _{α1}	4.02
84	G	C	175.22
85	V	N	121.78
85	V	HN	8.10
85	V	C _α	62.70
85	V	H _α	4.14
85	V	C _β	32.91
85	V	C	176.86
86	K	N	132.81
86	K	HN	8.04
86	K	C _α	57.81
86	K	H _α	4.22
86	K	C _β	34.00

Appendix 2: HMQC-J splittings and secondary structure motif

Resonance no	Residue name	Splitting	Secondary structure assignment
1	G	Gly	-
2	P	Pro	-
3	L	6.491	-
4	G	Gly	-
5	S	7.465	-
6	M	6.688	-
7	S	8.156	β
8	C	8.702	β
9	V	8.458	β
10	H	9.246	β
11	Y	8.821	β
12	K	8.438	β
13	F	Not fit	-
14	S	-1.539	-
15	S	Not fit	-
16	K	7.882	-
17	L	6.767	-
18	N	6.828	-
19	Y	7.372	-
20	D	9.591	β
21	T	9.169	β
22	V	9.646	β
23	T	9.486	β
24	F	6.829	-
25	D	7.799	-
26	G	Gly	-
27	L	6.427	-
28	H	7.978	-
29	I	9.310	β
30	S	5.172	α
31	L	0.372	α
32	C	2.419	α
33	D	5.407	α
34	L	3.770	α
35	K	3.013	α
36	K	3.271	α
37	Q	4.882	α
38	I	4.624	α

Resonance no	Residue name	Splitting	Secondary structure assignment
39	M	3.002	α
40	G	Gly	-
41	R	4.588	α
42	E	7.657	-
43	K	7.085	-
44	L	8.068	β
45	K	8.333	β
46	A	2.994	α
47	A	4.254	α
48	D	Overlapped	-
49	C	9.017	β
50	D	9.062	β
51	L	9.357	β
52	Q	9.101	β
53	I	9.145	β
54	T	9.517	β
55	N	5.047	α
56	A	3.626	α
57	Q	9.031	β
58	T	7.725	-
59	K	Overlapped	-
60	E	4.671	α
61	E	7.368	-
62	Y	7.650	-
63	T	Not fit	-
64	D	7.129	-
65	D	2.772	α
66	N	8.543	β
67	A	4.438	α
68	L	7.951	-
69	I	8.249	β
70	P	Pro	-
71	K	Not fit	-
72	N	6.838	-
73	S	Not fit	-
74	S	9.239	β
75	V	9.668	β
76	I	9.347	β
77	V	9.732	β
78	R	Not fit	-
79	R	6.930	-
80	I	Not fit	-

Resonance no	Residue name	Splitting	Secondary structure assignment
81	P	Pro	-
82	I	7.030	-
83	G	Gly	-
84	G	Gly	-
85	V	Not fit	-
86	K	7.570	-



Appendix 3: TALOS phi and psi angle predictions

Residue no	Residue name	ϕ angle	ψ angle	Classification
1	G	9999	9999	None
2	P	-61	2	Warn
3	L	-91	-6	Warn
4	G	49	-22	Warn
5	S	-78	138	Warn
6	M	-68	139	Warn
7	S	-94	139	Good
8	C	-109	136	Good
9	V	-136	134	Good
10	H	-102	135	Good
11	Y	-122	149	Good
12	K	-110	122	Good
13	F	-94	132	Good
14	S	-77	-18	Good
15	S	-83	-20	Warn
16	K	-88	149	Good
17	L	-80	-32	Good
18	N	-95	154	Warn
19	Y	-80	136	Warn
20	D	-130	139	Good
21	T	-110	130	Good
22	V	-115	128	Good
23	T	-106	137	Good
24	F	-134	146	Good
25	D	-89	136	Warn
26	G	-165	-17	Warn
27	L	-80	-19	Warn
28	H	-123	144	Good
29	I	-110	118	Good
30	S	-92	140	Good
31	L	-65	-35	Good
32	C	-66	-37	Good
33	D	-66	-36	Good
34	L	-66	-40	Good

Residue no	Residue name	ϕ angle	ψ angle	Classification
35	K	-67	-36	Good
36	K	-67	-39	Good
37	Q	-65	-42	Good
38	I	-66	-43	Good
39	M	-63	-39	Good
40	G	-64	-40	Good
41	R	-66	-29	Good
42	E	-91	1	Good
43	K	52	29	Warn
44	L	-115	154	Warn
45	K	-88	130	Good
46	A	-68	-21	Good
47	A	-79	-27	Good
48	D	-99	-11	Warn
49	C	-104	147	Good
50	D	-117	145	Good
51	L	-113	122	Good
52	Q	-114	125	Good
53	I	-101	121	Good
54	T	-112	147	Good
55	N	-72	136	Good
56	A	-72	-32	Good
57	Q	-78	-20	Good
58	T	-75	-22	Good
59	K	-68	-20	Warn
60	E	-77	138	Good
61	E	-87	122	Good
62	Y	-95	134	Warn
63	T	-95	141	Warn
64	D	-76	123	Good
65	D	-69	-27	Good
66	N	-97	10	Good
67	A	-71	142	Warn
68	L	-89	125	Good
69	I	-98	112	Good
70	P	-74	145	Good
71	K	-72	-15	Good

Residue no	Residue name	ϕ angle	ψ angle	Classification
72	N	-73	-23	Good
73	S	-88	137	Warn
74	S	-108	135	Good
75	V	-123	147	Good
76	I	-117	120	Good
77	V	-113	132	Good
78	R	-108	133	Good
79	R	-91	126	Good
80	I	-123	132	Good
81	P	-63	140	Warn
82	I	-75	135	Warn
83	G	42	-21	Warn
84	G	-16	-142	Warn
85	V	-90	123	Warn
86	K	9999	9999	None



TALOS uses the following scheme to classify the predictions: (i) none, indicates that no matches to database entries were found and no prediction was made, (ii) good, indicates that at least nine out of ten predictions fall within the same populated region of the Ramachandran map and (iii) warn, indicates that less than nine out of ten predictions fall within the same populated region of the Ramachandran map; thus the prediction is ambiguous.