

**EXPRESSION AND PURIFICATION OF THE NOVEL
PROTEIN DOMAIN DWNN**

PORTIA THANDOKAZI LUTYA

**A thesis submitted in partial fulfilment of the requirements for the degree of
Magister Scientiae in the department of Biochemistry, University of the Western
Cape**

Supervisor: Dr David Pugh

December 2002

DECLARATION

I declare that *Expression and purification of the novel protein domain DWNN* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full Name : **Portia Thandokazi Lutya** Date : **December 20, 2002**

Signed.....



ABSTRACT

DWNN is a novel protein domain first identified in knock-out experiments on CHO cells as conferring resistance to cytotoxic T-lymphocyte killing. It is found in all eukaryotic genomes so far investigated, with a conserved consensus region of 78 amino acids. The human form is expressed both as an independent domain of 118 amino acids, and as part of a larger protein which also includes a zinc finger, a ring finger and a domain known from previous studies to bind both of the tumour suppressor proteins, p53 and Rb. Structure prediction studies suggest that the domain adopts the same 3-dimensional fold as the protein ubiquitin, which is interesting in the light of the recently discovered role of ubiquitin-like proteins and Ring fingers in the regulation of tumour suppressor proteins.



This work describes the cloning, expression and purification of the DWNN domain for structural studies using NMR. Both the 118 amino acid form and a shortened 82 amino acid form corresponding to the consensus region were successfully expressed as fusions with the protein GST. Two different vector systems were used, allowing removal of GST using either thrombin or 3C protease. The C-terminal region of the 118 amino acid protein was found to be susceptible to proteolysis, confirming that this region forms an unstructured tail. Samples of the shortened form were successfully prepared at concentrations and purity suitable for NMR analysis.

Dedicated to the memory of

Mr Zalisile ‘Tamnci’ Gladstone Joyi

And

Mr Luthando ‘Wankie’ Anthony Majikela



ABBREVIATIONS.

$^{\circ}\text{C}$	degrees Celsius
1D	one-dimensional
3D	three-dimensional
APS	ammonium per sulphate
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
C-terminal	carboxyl-terminal
cDNA	complementary deoxyribonucleic acid
CHO	Chinese hamster ovary
CTL	cytotoxic T lymphocyte
CV	column volume
dH ₂ O	distilled water
DNA	deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
Fig	Figure
GST	glutathione-S-transferase
hr	hour

IPTG	isopropyl β -D-thiogalactopyranoside
kDa	kiloDalton
kb	kilobase pairs
LB	Luria broth
MALDI-TOF	Matrix Assisted Laser Desorption/Ionization-Time Of Flight
MCS	Multiple Cloning Site
MHC	Major Histocompatibility Complex
min	minute
MW	Molecular Weight
NCBI	National Centre for Biotechnology Information
NMR	Nuclear Magnetic Resonance spectroscopy
N-terminal	amino terminal
OAc	acetate
o/n	overnight
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEG	polyethylene glycol
PMSF	phenylmethylsulphonyl fluoride
Rb	Retinoblastoma
RNA	ribonucleic acid
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate



SDS-PAGE	SDS-polyacrylamide gel electrophoresis
TBS	Tris-buffered saline
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	tris (hydroxymethyl)aminoethane
Tris-Cl	Tris (containing HCl)



TABLE OF CONTENTS.

	Page
ABBREVIATIONS	i
CHAPTER 1.	
Introduction: The DWNN family and its putative function	
1.1 Introduction	1
1.2 Variable domain organization of DWNN	2
1.3 DWNN is highly conserved across evolution	2
1.4 DWNN and the immune system	4
1.4.1 Components of the immune system	5
1.4.2 CTL Pathways	6
1.4.3 Identification of DWNN	8
1.5 DWNN and Apoptosis	9
1.5.1 Apoptosis	9
1.5.2 Apoptosis resistance screens and DWNN	11
1.6 DWNN and Cancer	11
1.6.1 Tumour suppressor genes	11
1.6.2 Interactions between DWNN and tumour suppressors	13
1.7 Ubiquitin and DWNN	14
1.7.1 Ubiquitin	14
1.7.2 Ubiquitin-like proteins	15
1.7.3 Tertiary structure prediction of DWNN	16

1.8	Protein structure determination by NMR	17
	1.8.1 1-D NMR	18
1.9	MALDI-TOF mass spectrometry	20
 CHAPTER 2		
	Materials and methods	
2.1	Materials and suppliers	21
2.2	Solutions	22
2.3	Bacterial cultures	24
	2.3.1 Bacterial strains used	24
	2.3.2 Selection	24
	2.3.3 Storage of bacterial strains	24
2.4	pGEX expression vectors	24
2.5	Preparation of <i>E. coli</i> competent cells for transformation	25
2.6	Transformation of <i>E. coli</i>	26
2.7	Large-scale preparation of plasmid DNA (Maxiprep)	26
	2.7.1 PEG precipitation	27
	2.7.2 Double CsCl/ethidium bromide fractionation	27
2.8	Small scale preparation of plamid DNA	28
2.9	Restriction enzyme digests	30
	2.9.1 Shrimp alkaline phosphatase treatment of restriction digested DNA	30
2.10	Ligations of insert to vector	30

2.11	Agarose gel electrophoresis of DNA	31
2.11.1	Gel preparation and electrophoresis	31
2.11.2	Sample preparation.	31
2.11.3	Detection of DNA	31
2.12	Amplification of DNA by the polymerase chain reaction (PCR)	32
2.13	<u>Colony PCR used in Section 3.3</u>	32
2.14	Sequencing of DNA	33
2.15	SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of proteins	34
2.16	Expression screening of transformants	35
2.17	Recombinant production of 3C protease	35
2.18	Expression and purification of recombinant DWNN	36
2.18.1	Expression and purification of GST-long-form DWNN (Section 3.4)	36
2.18.2	Expression and purification of GST-long-form DWNN (Section 4.2)	37
2.18.3	Expression and purification of GST-short-form DWNN	38
2.17	Lyophilisation of a protein sample	38

CHAPTER 3

Recombinant expression long-form human DWNN domain in *E.coli*

3.1	Introduction	39
-----	--------------	----

3.2	PCR amplification of the coding region for human long-form DWNN	39
3.3	Insertion of long-form DWNN PCR product into pGEX-4T-3 vector	40
3.4	Expression and purification of GST-long-form DWNN fusion protein	42
3.5	The cleavage of the fusion protein by thrombin and purification of DWNN	43
3.6	Mass spectrometric analysis of long-form DWNN	44
3.7	Optimisation of thrombin cleavage	45

CHAPTER 4
Recombination expression of the long-form human DWNN using pGEX-6P-2 vector



4.1	Insertion of the coding region for long-form DWNN into the pGEX-6P-2 vector	48
4.2	Large-scale expression and purification of long-form DWNN	49
4.3	Sequencing of the pGEX-6P-2 - long-form DWNN construct	50
4.4	Mass spectrometric analysis of long-form DWNN	51
4.5	1D proton NMR spectrum of purified long-form DWNN	52

CHAPTER 5
Recombinant expression of the human short-form DWNN domain

5.1	Insertion of the coding region for short form DWNN into the	54
-----	---	----

pGEX-6P expression vector	
5.1.1 PCR amplification of the coding region for human short-form DWNN	54
5.1.2 Insertion of short-form DWNN PCR product into pGEX-6P-2 vector	55
5.1.3 Expression and purification of GST-short form DWNN fusion protein	56
5.2 Sequencing of the pGEX-6P-2 –short-form DWNN construct	56
5.3 Reverse-phase purification of short-form DWNN	57
5.4 Short-form DWNN MALDI-TOF mass spectrometry results	58
5.5 1D NMR results of purified short-form DWNN	59
CHAPTER 6	
Discussion and Conclusion	
6.1 DWNN expression constructs	60
6.1.1 pGEX-4T-3 - long-form DWNN construct	60
6.1.2 pGEX-6P-2 – long-form DWNN construct	61
6.1.3 pGEX-6P-2 - Short-form DWNN construct	63
6.2 Conclusion and future prospects	64
REFERENCES	65
APPENDICES	
Appendix A Full Length cDNA sequence of Human DWNN, with translation of open reading frame	A
Appendix B Expected pGEX (4T-3) – DWNN (long-form) construct	C



	(Multiple Cloning Site)	
Appendix C	Expected pGEX 6P-2 – DWNN (long-form) construct (Multiple Cloning Site)	D
Appendix D	Expected pGEX 6P-2 - DWNN (short-form) construct (Multiple Cloning Site)	E
Appendix E	pGEX 5' Sequencing Primer	F



CHAPTER 1

Introduction: The DWNN family and its putative function

1.1 Introduction

DWNN is a novel gene originally identified in Chinese hamster ovary (CHO) cells. CHO cells are functionally haploid, meaning that they are a diploid cell line that appear to transcribe only one of the two alleles at a large proportion of loci. Retroviral promoter-trap mutagenesis was used to mutagenise the cells by insertion into the genome; the mutagenised cells were then screened for susceptibility to CTL killing using the lactate-dehydrogenase (LDH) assay. Inverse PCR and RACE PCR were used to identify the gene into which the retrovirus was inserted.



Using these methods, a 101bp fragment of hamster genomic DNA was identified. BLAST searches results using the fragment yielded a large number of close homologues in a wide range of eukaryotic organisms. EST analysis shows that in humans the domain is expressed both as a single domain of 13kDa and forms part of a 200kDa protein which also includes a zinc finger, a RING finger and a PACT domain (unpublished data by Zodwa Dlamini). The PACT domain (p53 associated cellular protein-testes derived) is also known as the RBBP6 (Retinoblastoma binding protein 6) and interacts with tumour suppressor proteins, retinoblastoma (Rb) and p53 (Simons *et al.*, 1997). This domain is found as a single copy per genome analysed. This domain was named DWNN (Domain With No Name). The consensus region of

DWNN is 78 amino acids long. Tertiary structure predictions show that DWNN shares the same fold as the protein ubiquitin. Ubiquitin is believed to be involved in apoptosis (Macleod, 2000). The findings that the full-length 200kDa protein may interact with p53 and Rb raises the possibility that DWNN forms part of essential cellular regulation, which may have important implications for the development of novel anti-cancer therapies. The work in this study involves generation of recombinant 13kDa DWNN, its expression in *E.coli* and purification for structure determination.

1.2 Variable domain organization of DWNN

DWNN is conserved across eukaryotic organisms, from humans to plants to algae. In worms and flies DWNN is found as a 200kDa domain and in algae and ascidians a 13kDa DWNN has been identified but it is not yet known if it interacts with other domains. The full-length 200kDa protein is composed of the DWNN domain, a zinc finger, a RING finger and a PACT domain. In organisms such as fungi and plants DWNN is associated with the zinc finger and the RING finger and no PACT domain as shown in Fig. 1.1.

1.3 DWNN is highly conserved across evolution

Homologues of DWNN have been found in all eukaryotic organisms for which sequence data exist in the database at the NCBI. An alignment of these sequences was produced using the Clustal X sequence alignment tool as shown in Fig. 1.2. A colour scheme has been used to highlight all the conserved amino acids. The

conserved negatively charged amino acids with polar side chains are shown in purple (D and E). All the conserved amino acids with non-polar side chains and Cysteine are highlighted in blue (M, I, V, L, F and A). The positively charged amino acids with polar side chains, K and R are shown in red. Proline is the only cyclic amino acid and is highlighted in lime. S, T, N and Q have uncharged polar side chains and are shown in green. H and Y, which have polar, aromatic side chains, are highlighted in powder blue. Glycine (G), an amino acid that has only hydrogen as its side chain is shown in orange in the diagram. This colour scheme makes it easier to identify all the conserved amino acids.

From the sequence alignment generated by Clustal X, a phylogenetic tree was produced as shown in Fig. 1.3. An alignment is produced to represent the evolutionary relationship between species, it is necessary for aligning inference to reflect the fact that species are related through an evolutionary tree (Thorne and Kishino, 1992). Interesting results were observed from the DWNN phylogenetic tree in Fig. 1.3. As expected mouse and human DWNN's are closely related to each other. DWNN sequences from plants (Arabidopsis, tomato, medicago) are closely related as they originate from the same branch.

The high level of conservation of the DWNN domain may be an indication that it plays an important role in cellular regulation of these organisms. DWNN is also found in organisms with minimal genomes such as Encephalitozoa, which has a very simple genome as compared to the human genome (Vivarès and Méténier, 2000). The

fact that this domain is found in an organism with such a simple genome suggests that DWNN may play a very crucial role in eukaryotic cells.

1.4 DWNN and the immune system

DWNN was first identified in a screen for cells resistant to killing by cytotoxic T cells (George, 1995). An implication that can be drawn from this is that DWNN plays a role in the killing of cells by cytotoxic T-cells. In order to place this finding into context, a brief review of the mammalian immune system is given below.

The mammalian immune system is comprised of T-lymphocytes, B-lymphocytes, natural killer (NK) cells, macrophages and antigen-presenting cells (APCs) and their various subclasses. Most of these components originate in the bone marrow. The responses of the immune system can be divided into two major classes: antibody responses and cell-mediated responses. Antibody responses involve the production of antibodies that circulate in the bloodstream and permeate other body fluids, where they bind specifically to foreign antigens that induce them.

Cell mediated responses involve the production of specialised cells that react with foreign antigens on the surface of other host cells thereby getting rid of the infected cell. In different situations, the reacting cell secretes chemical signals that activate macrophages to destroy the invading microorganisms (Alberts *et al.*, 2000).

1.4.1 Components of the immune system

B lymphocytes

B-cells produce cell membrane receptors that are also known as antibodies and have single antigen specificity. These lymphocytes are selected in the bone marrow and when matured they leave for secondary lymphoid organs such as the spleen, lymph nodes and the gut associated lymphoid tissue. When B cells are activated by antigen, they undergo a second round of selection in the follicles of secondary lymphoid organs after which they mature into plasma cells that produce and secrete antigen-specific antibodies. These antibodies then re-circulate to the bone marrow. This re-circulation ensures that the appropriate B lymphocytes come into contact with antigens (Krammer, 2000).



T lymphocytes

Pre-T lymphocytes move from the bone marrow into the thymus. T cells are selected in the thymus on the basis of the affinities of their T-cell antigen receptors (TCRs) for self-major histocompatibility (MHC) antigens. Two classes of the MHC antigen are present and are known as MHC class I antigen and MHC class II antigen. MHC I antigen sample peptides are from foreign proteins and MHC II antigens sample peptides are from self-proteins. Three key steps involved in antigen processing in association with MHC class I have been identified, cytosolic peptide generation with the help of the proteasome, peptide transport into the endoplasmic reticulum and peptide assembly with class I molecules (van Endert, 1999).

T-cells with high affinity for self-MHC molecules are eliminated to prevent autoimmunity. T-cells that interact with MHC class II molecules develop into cells that possess CD 4 antigens on their surface and T-cells with high affinity for MHC class I possess CD 8 antigens on their surface. Only mature T-cells with functional TRC leave thymus for secondary lymphoid organs. Mature CD 4⁺ T-cells develop into T-helper 1 cells that secrete cytokines, which regulate cellular immune response and T-helper 2 cells that secrete cytokines, which regulate antibody response. Mature CD 8⁺ T-cells become cytotoxic killer cells in the presence of infectious agent and APCs present antigenic peptides to T-cells. Antigen-MHC complexes stimulate APS surface molecules and cytokines to drive T cell into clonal expansion. These T-cells then send a message to the T-cells or B-cells to regulate their responses. The immune response then undergoes a down-phase. Most lymphocytes are eliminated by apoptosis and the few cells that survive will join the pool of memory cells (Krammer, 2000).

1.4.2 CTL pathways

Cytotoxic lymphocytes include activated CD4⁺ and CD8⁺ lymphocytes, natural killer (NK) cells, and lymphokine activated killer (LAK) cells (Thomas *et al.*, 2000). These cells play a vital role in the functioning of host immune system's response against viral infections and tumours (McMichael, 1992). The target cells for CTLs display viral or tumour peptides that are bound to MHC-1 glycoprotein on their plasma membrane (Townsend and Bodmer, 1989). CTLs morphological changes and motility play an important role in their functioning (Waters *et al.*, 1996). Activated CTLs are

highly motile and must move through all tissues of the organism (Hahn *et al.*, 1994). Cytotoxic T lymphocytes induce apoptosis in two distinct pathways to lyse target cells (Atkinson *et al.*, 1998 and Beresford *et al.*, 1998). These two pathways are the Fas mediated pathway and perforin pathway (Barry *et al.*, 2000).

The Fas pathway involves binding of Fas death receptor on the surface of target cells (Peter and Krammer, 1998). The interaction of Fas receptor/ligand plays a role in the recruitment of caspase 8 through the adapter molecule FADD (Fas-associated death domain) also known as MORT1 (Boldin *et al.*, 1996). Caspase 8 gets activated autocatalytically and then activates downstream caspases (Barry *et al.*, 2000), after-which apoptosis of target cells resumes.



The Perforin pathway involves exocytosis of cytolytic granules from CTLs (Barry *et al.*, 2000). These granules are made up of a variety of proteins required for apoptosis of the target cell (Griffiths and Argon, 1995). One of these proteins is a pore-forming protein known as perforin, which is thought to facilitate the entry of other proteins involved in this pathway and a family of serine proteases known as granzymes (Barry *et al.*, 2000). Studies have shown that the substrates for these proteases are the caspases; for example, the substrate for granzyme B was found to be caspase 3 (Darmon *et al.*, 1995). Other caspases have been shown to be granzyme substrates *in vitro* suggesting that granzyme B induces apoptosis of the target cell by triggering the activation of caspases (Barry *et al.*, 2000). However, only caspase 3 and 8 have been shown to be substrates for granzyme *in vivo* (Atkinson *et al.*, 1998; Yang *et al.*, 1998

and Medema *et al.*, 1997). These studies prove the importance of caspase 8 in both the Fas-mediated cell death and perforin pathway.

1.4.3 Identification of DWNN

CTL killing plays a role in the removal of viral particles, damaged cells and cancerous cells. These CHO cells were sensitised to lysis by HA-specific K^K restricted CTL, by the transfection of LHA and MHC class I K^K. As mentioned previously, these cells are functionally haploid and thus creating suitable cell lines for mutagenesis as integration event may lead to loss of gene function.

Insertional mutagenesis was performed by using a defective promoter-trap retrovirus. The retrovirus is inserted downstream of an active promoter of the host gene thereby disrupting the expression of that gene. The promoter-trap retrovirus is characterised by the presence of a hygromycin resistance gene which is expressed once the retrovirus integrates downstream of an active cellular promoter. If the disrupted gene is involved in the CTL pathway then the cells should be resistant to CTL killing.

The mutagenised CHO cells were subjected to HA-specific CTL lysis. More than 100 CTL resistant cell lines were generated using promoter-trap mutagenesis and 3 cell lines were shown to be totally resistant to CTL killing (George, 1995). Inverse PCR and rapid amplification of cDNA ends (RACE) PCR were utilised to identify the sequence of the inactivated gene from a number of the CTL resistant cell lines. One

of the cells yielded a fragment that when BLASTed against the genomic database, it was found to be novel (Pretorius, unpublished data 1998)

1.5 DWNN and Apoptosis

A study showing that there might be a relationship between DWNN and apoptosis was conducted and is discussed in this section. Since DWNN was initially identified in a screen for cells resistant to CTL killing and CTL killing has two pathways of which one leads to apoptosis, this raised the possibility that DWNN plays a role in apoptosis.

1.5.1 Apoptosis

Apoptosis is a fundamental process, also known as programmed cell death that is essential for cellular development and tissue homeostasis. Changes in this process may lead to many pathological conditions including viral infection, autoimmune disorder and cancer. The human body is composed of $\sim 10^{14}$ cells, each of which has the ability to commit suicide by apoptosis (Nicholson, 2000), thus this fact shows the importance of apoptosis in development. During development of organisms, apoptosis is induced in supernumerary, misplaced or damaged cells during the construction, maintenance and repair of tissues (Meier *et al.*, 2000). When this process is compromised inappropriate cell death and disease pathogenesis can result. Apoptosis is characterised by condensation and fragmentation of the chromatin network, compaction of cytoplasmic organelles, dilation of the endoplasmic reticulum, a decrease in cell volume and changes in the plasma membrane resulting

in the recognition and phagocytosis of apoptotic cells (Cohen, 1997; Arends and Wyllie, 1991). The apoptotic cells are then removed by phagocytosis.

It has been shown that cells dying from apoptosis undergo morphological features that are different from cells that die through pathology and necrosis. These morphological changes include cell shrinkage, the blebbing of plasma membranes and nuclear condensation (Sakahira *et al.*, 1998). This led to the hypothesis that a common conserved endogenous cell death program is responsible for apoptosis. These common morphological changes were caused by cysteine proteases that are activated specifically in apoptotic cells. These proteases are homologous to one another and form part of a large family known as the caspases (Alnemri *et al.*, 1996). The caspases are well conserved throughout species ranging from humans to nematodes. Since caspases lead to a visible change that characterise apoptosis, they are thought to be the central executioners of the apoptotic pathway (Hengartner, 2000). Some caspases are cleaved and activated by other caspases. It is thought that the initiator caspases get activated when they are isolated and combined by bridging or regulatory factors. Caspase-mediated cell death is suppressed by Bcl-2, the origin of a family of proteins that suppress apoptosis. (Chau *et al.*, 2000).

There are many pathways involved in apoptosis but only the Fas pathway is common to both apoptosis and CTL killing since DWNN has been shown to be involved in CTL killing and apoptosis (see below). A possible scenario is that DWNN might play a role in the Fas pathway. Fas pathway is discussed in section 1.4.2.

1.5.2 Apoptosis resistance screens and DWNN

Apoptosis was induced in the mutagenised cell lines generated as described in section 1.4.3 using staurosporine. The cells did not die, indicating that they were resistant to staurosporine (unpublished data). These results suggested that DWNN could play an important role in apoptosis. If so, DWNN may be an important anti-cancer therapy candidate because cancer frequently develops when apoptosis fails to occur.

1.6 DWNN and Cancer

The full protein of which the DWNN domain forms the N-terminal part has previously been shown to interact with both p53 and Rb (Simons *et al.*, 1997). Since p53 and Rb are both known to play vital roles in the control of cancer, this suggests a possible role for DWNN in the regulation of cancer. In order to place these hypotheses into context, a brief review of tumour suppressor genes is presented below.

1.6.1 Tumour suppressor genes

Tumour suppressor genes are required to keep cell growth under control. They are responsible for example, for controlling the cell cycle, DNA replication and cell division. When these genes fail, uncontrolled cell growth will occur which is a characteristic of cancerous cells. Tumours develop when tumour suppressor genes do not function properly. The tumour suppressor genes that are of importance in this study are p53 and Rb.

p53

p53 induces apoptosis and it is known to prevent genomic instability as well as being involved in growth arrest (Macleod, 2000). It has been found that p53 does not function properly in approximately half of the tumours (Vogelstein *et al.*, 2000). It is also known that p53 has a short half-life in cultured cells, and that it accumulates after cellular exposure to DNA-damaging agents (Hupp *et al.*, 2000), which suggests that there has to be a p53 regulator which is responsible for p53 degradation.

Mdm2

p53 interacts with murine double minute clone 2 (Mdm2), which binds the N-terminus of p53 and is one of the most important p53 regulators. The interaction of Mdm2 with p53 results in the inhibition of the ability of p53 to act as a transcription factor and also targets p53 for degradation through ubiquitin-dependent proteolysis (Zhang *et al.*, 1998). A negative feedback loop is thus formed because Mdm2 causes p53 to activate a component of its own degradation (Haupt *et al.*, 1995 and Chao *et al.*, 2000). Mdm2 has been shown to act as an E3-ubiquitin ligase and this activity plays a role in the targeting of p53 for ubiquitin-dependent degradation. E3 ligases are essential in ubiquitin conjugation, which is a process that leads to the covalent modification of proteins with ubiquitin and their subsequent degradation by the proteasome. Mutations in the RING finger that is present in Mdm2 and most E3 ligases results in the stabilisation of Mdm2 and loss of the ability to degrade p53 as it can also ubiquitinate itself. The E3 activity of Mdm2 is dependent on zinc and multiple metal co-ordinating ring residues (Lorick *et al* 1999). The ability of Mdm2

to act as ubiquitin-ligase might serve to target p53 directly to proteasome but ubiquitination of p53 may have other effects on the regulation of p53 stability and activation (Macleod, 2000; Kubbutat and Vousden, 1998).

Rb

Rb interacts with Mdm2 and thereby promotes p53 activity by blocking its Mdm2-mediated degradation. Unlike most tumour suppressor genes, Rb acts to prevent apoptosis through induction of the pro-apoptotic Bax gene and also through an uncharacterised mechanism requiring Apaf¹ and caspase 9. Rb is a critical negative regulator of the cell cycle and by linking loss of Rb to the induction of apoptosis, the cell prevents deregulated proliferation and tumour formation (Macleod, 2000). Rb suppresses cell proliferation and is mutated in various types of human cancer (Weinberg, 1992).



1.6.2 Interactions between DWNN and tumour suppressors

DWNN domain forms part of a large protein containing the PACT domain, which is known to interact with Rb and p53. The significance of the interaction of the PACT domain and the Rb is not yet fully understood (Sakai *et al.*, 1995). It also contains a RING finger domain, which raises the possibility that it may play a role in the ubiquitin-dependent regulation of Rb and DWNN.

1.7 Ubiquitin and DWNN

Modelling studies predict that DWNN is likely to adopt the same 3D fold as ubiquitin (Pugh and Jonas, unpublished). In the context of recent findings linking ubiquitin and ubiquitin-like proteins to the regulation of both apoptosis and cancer, a possible role for DWNN in these processes is suggested.

1.7.1 Ubiquitin

Ubiquitin is a 76 amino acid protein found in all eukaryotic cells either independent or bound covalently to cellular proteins and highly conserved in evolution (Ciechanover, 1994 and Yeh *et al.*, 2000). Ubiquitin tags protein for degradation by the proteasome or through the endocytic route. Conjugation of ubiquitin to other proteins includes the formation of an isopeptide bond between the C-terminal glycine residue of ubiquitin and the ϵ -amino group of a lysine residue of an acceptor protein. It is thought that all known functions of ubiquitin are carried out through this reaction.

Ubiquitin binds to its substrates through an enzymatic cascade, involving ubiquitin-activating (E1) enzyme, ubiquitin-conjugating (E2) enzyme and in some cases ubiquitin-protein ligases (E3) (Tyers and Jorgensen 2000; Hershko and Ciechanover 1998). The E1 enzyme hydrolyses ATP and through an E1-bound ubiquitinyl adenylate intermediate, forms a high-energy thioester between a cysteine of its active site and the C-terminus of ubiquitin. Ubiquitin is then moved on to E2 enzyme to

form thioester-linked complexes in the similar manner as with E1 enzyme. At the end, ubiquitin is covalently bound to the substrate protein by the E2 or by the E3 enzyme, which may possess substrate-binding properties (Liakopoulos *et al.*, 1998 and Scheffner *et al.*, 1995). The principal regulatory step of substrate recognition is carried out via the E3 ubiquitin ligases. The recognition of target proteins as substrates appears to involve protein-protein interactions with specific E3s but the role of E2 enzyme in substrate recognition has not yet been established (Kumar *et al.*, 1997).

1.7.2 Ubiquitin-like proteins

Ubiquitin-like proteins include proteins that function in a similar manner as ubiquitin and those that possess ubiquitin domains but are unrelated in sequence with ubiquitin (Jentsch and Pyrowolakis, 2000). The list of these proteins is expanding but only the well-characterized proteins will be mentioned in this section.

SUMO-1 (small ubiquitin-related modifier)/sentrin-1 is a small ubiquitin-like protein found in higher eukaryotic cells, which determines the sub-cellular localisation of proteins. SUMO has recently been shown to interact and activate p53 although the relevance of this has not yet been discovered (Lohrum and Vousden, 2000 and Hupp *et al.*, 2000). The NMR structure of SUMO-1/sentrin-1 suggests that it contains a ubiquitin-homology domain and a very flexible N-terminus that protrudes from an ubiquitin-like core (Bayer *et al.*, 1998). This protein is only 18% identical in sequence with ubiquitin. Although many SUMO-1/sentrin-1 targets have been

identified, the consequences of SUMO/sentrin-1 modification are not yet known. SUMO modification does not appear to lead to protein degradation in the same manner as the ubiquitin pathway but has been shown in some instances to be linked to the ubiquitin system (Liakopoulos *et al.*, 1998). The sentrinisation pathway deviates from the ubiquitination pathway at an early stage (Yeh *et al.*, 2000). NEDD8 (Neural precursor cell-Expressed Developmentally Down regulated protein 8) is a protein that shares 60% identity and 80% similarity with ubiquitin and was firstly found to be a novel mRNA highly enriched in foetal mouse brain (Kumar *et al.*, 1992). The major substrate for this protein is Cdc53, which is a yeast cullin (Yeh *et al.*, 2000).



1.7.3 Tertiary structure prediction of DWNN

Tertiary structure prediction studies were performed using Modeller, a computer program that models the 3D structures of proteins. Tertiary structure prediction studies showed that DWNN's fold is comparable to that of ubiquitin, see Fig. 1.4. The similarity of amino acid sequences of two different proteins is not related to the function of those proteins but if they have a similar fold they are more likely to perform similar functions.

These results reveal that both DWNN and ubiquitin possess five β -sheets and one α -helix each that fold in the similar manner on both proteins. Three-dimensional studies are required in this study to confirm these findings. It has been found that

ubiquitin ligases, which bind ubiquitin to proteins for degradation have RING fingers (Tyers and Jorgensen, 2000). Human 200kDa DWNN protein also has a RING finger and led to the hypothesis that DWNN might be one of or related to ubiquitin ligases but more investigation is required to confirm this.

1.8 Protein structure determination by NMR

Proteins play an important role in cells, as the morphology, function and activities of the cell depend on the proteins they express. The key to understanding how different proteins function lies in an understanding of the molecular structure. The overall aim of this research is the determination of the structure of DWNN domains; this thesis describes the preparation of samples of human DWNN suitable for structural analysis by NMR, as well as preliminary NMR analysis.



Nuclear Magnetic Resonance spectroscopy (NMR) was introduced as a second structure determination method after X-ray crystallography, which required crystallisation of proteins prior to collection of data. Some proteins do not crystallise well and even when they do crystals may diffract poorly or there may be difficulties in solving the phase problem (Clare and Gronenborn, 1991). For NMR studies, the protein in question should be soluble and must be able to be concentrated to at least 1 mM in a volume of 0.5ml. The upper molecular weight limit of a protein to be studied using NMR is ~ 50 -60kDa for the applicability of current solution NMR technology.

1.8.1 1-D NMR

This NMR technique is based on the fact that some nuclear isotopes, including ^1H , ^{13}C and ^{15}N , have magnetic properties. When a molecule containing such isotopes is placed in a magnetic field of strength B_0 , the nuclei precess about the direction of the field, giving rise to a fluctuating magnetic field. The fluctuating field is observed, amplified and Fourier transformed to give a 1-D spectrum, as shown in Fig. 1.5. Each peak (or resonance) in this spectrum corresponds to a particular proton in the protein. The amide protons, which are on the left hand side of the spectrum in Fig. 1.5, precess at a different frequency from the aromatic protons, near the centre of the spectrum, and the aliphatic protons at the right hand side, due to the different chemical environment of the protein backbone, the rings or the hydrocarbon side chains. The difference in precession frequency between protons in different chemical environments is known as “chemical shift”, and is represented by the Greek symbol δ . Chemical shift is a consequence of protein folding and a broad spread of δ is a good indicator that a protein is well folded; the addition of chemical denaturants such as urea and guanidinium causes the peaks to collapse in towards the middle of the spectrum.

δ is proportional to the magnetic field B_0 , which means that spectra from spectrometers with different magnetic field strengths cannot be superimposed. To get

around this problem, 1D spectra are plotted in dimensionless units defined relative to an arbitrarily chosen reference frequency δ_0 , as follows:

$$(\delta_0 - \delta) / (\delta_0 \times 10^6)$$

Because of the inclusion of the factor of 10^6 , these units are commonly referred to as “parts per million”, or “ppm”. Since δ_0 and δ are both proportional to the strength of the magnetic field, chemical shifts expressed in terms of ppm are independent of the magnetic field strength, and therefore the same for all spectrometers. Proton resonances typically fall in the range 0 – 10ppm.

Water is used as a solvent for most biological molecules such as proteins. The chemical shift of water protons lies in the centre of the 1D spectrum of the proteins at 4.7ppm. Given the molarity of the water protons is 110M and the maximum molarity of a protein is ~ 10mM, the water resonance would ordinarily be expected to swamp protein resonances. Sophisticated “water-suppression” techniques therefore have to be used to reduce the magnitude of the water resonance relative to the protein resonances. Alternatively H_2O can be replaced as solvent by D_2O , deuterium oxide, which is chemically identical to water, but has the advantage that the deuterons do not have a magnetic signal in the same range as proteins.

1.9 MALDI-TOF mass spectrometry

Matrix Assisted Laser Desorption/Ionization – Time Of Flight (MALDI-TOF) mass spectrometry was used in this study to confirm the sizes of recombinant protein. This

technology is appropriate for bio-molecules due to the gentle ionisation method used. A laser is used to ionize the sample, which has been pre-absorbed out of a readily vaporizable matrix (van Baar, 2000). Once the sample molecule has been vaporised and ionised they are electrostatically transferred into a time-of-flight (TOF) spectrometer where they are separated from the matrix ions and analysed, based on their mass-to-charge (m/z) ratios (Lennon and MALDI basics). In this study, MALDI-TOF was used to determine the molecular weight of DWNN.



CHAPTER 2

Materials and Methods

2.1 Materials and Suppliers

40% 37.5:1 acrylamide:bis-acrylamide	Promega
Ampicillin	Roche
Boric acid	Saarchem UNIV AR
Calcium chloride	BDH anala R
Cesium chloride	Roche
Chloroform	BDH anala R
Coomasie Brilliant Blue R 250	Sigma
DTT (Dithiothreitol)	Roche
EDTA	Saarchem UNIV AR
Ethanol	BDH anala R
Ethidium bromide	Sigma
Glucose	Saarchem UNIV AR
Glycine	Saarchem UNIV AR
IPTG	Promega
Mops	Roche
PBS (Phosphate-buffered saline) tablets	Life Technologies



PEG 8000	Sigma
PMSF	Sigma
Sodium hydroxide	Saarchem UNIV AR
Sodium dodecyl sulfate	Roche
TEMED	Promega
Triton X-100	Riedel-De Haën Ag Seelze-Hannover

All other chemicals (powder compounds) were from MERCK and the rest of organic compounds were from Saarchem UNIV AR

2.2 Solutions

1xTBE	89mM Tris, 89mM Boric acid, 3mM EDTA pH 8
2x SDS-PAGE Sample Buffer	0.125M Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol
APS	10% ammonium persulphate
SDS-PAGE Destaining Solution	50% methanol, 10% acetic acid, 40% dH ₂ O
SDS-PAGE Drying Solution	5% glycerol, 30% ethanol, 65% dH ₂ O
GTE	50mM Tris-HCl pH 8, 10mM EDTA, 1% glucose
Initiator	10% ammonium persulphate

L-agar	0.1% meat extract, 0.5% peptone, 0.2% yeast extract, 0.8M NaCl, 1.5% agar
LB	1% bacto-tryptone, 0.5% yeast extract, 10g l ⁻¹ NaCl
NaOH/SDS	200mM NaOH, 1% SDS
PBS	Phosphate-buffered saline
Phenol/chloroform	1:1 phenol: chloroform (v/v)
PMSF Stock	10mM PMSF, dissolved in isopropanol
SDS-PAGE Running Gel Buffer	1.5M Tris-HCl pH 8.8
Salt-saturated propan-2-ol	3M NaCl dissolved in propan-2-ol
SDS-PAGE Stacking Gel Buffer	0.5M Tris-HCl pH 6.8
Coomassie Stain	0.125% Coomassie stock, 50% methanol, 10% acetic acid, 39.875% dH ₂ O
Commassie Stain Stock	1% Coomassie Blue R-250
SDS-PAGE Tank Buffer	0.025M Tris-HCl pH 8.3, 0.192M glycine, 0.1% SDS
TBS	50mM Tris-HCl pH 8.5, 150mM NaCl
TE	10mM Tris-HCl pH 7.4, 1mM EDTA
Tfb 1	30mM KOAc, 50mM MnCl ₂ , 100mM KCl, 10mM CaCl ₂ , 15% glycerol
Tfb 2	10mM Na-MOPS, pH 7, 75mM CaCl ₂ , 10mM KCl, 15% glycerol
TTBS	TBS with 0.1% (v/v) Triton X-100

pGEX vector is an expression vector that forms the basis of the GST (glutathione-S-transferase) expression system. There are three major types of pGEX plasmid (T, P and X) which differ in the protease cleavage site (Thrombin, PreScission or Factor Xa) they possess upstream the poly-linker region. The genes are expressed as fusion proteins with the 26kDa GST protein from *Schistosoma japonicum*. The pGEX plasmids produce high-level of intracellular expression driven by a *tac* promoter and induced with IPTG. The expression vectors pGEX-4T-3 and pGEX-6P-2 were used in this study. pGEX-4T-3 contains a thrombin recognition site and pGEX-6P-2 contains the PreScission recognition site.

2.5 Preparation of *E. coli* competent cells for transformation

- a) A single colony was taken from a freshly streaked nutrient agar plate and transferred to 20ml of TYM and grown with vigorous shaking at 37⁰C to an OD₅₅₀ = 0.2. The cells were then transferred to 100ml TYM and grown to OD₅₅₀ = 0.2. The culture was transferred to 400ml TYM and grown under the same conditions to OD₅₅₀ = 0.5.
- b) The cells were rapidly chilled at 0⁰C and pelleted by centrifugation at 3000rpm for 10min at 4⁰C.
- c) The bacterial pellet was re-suspended in 250ml of cold Tfb 1 and incubated at 4⁰C for 30min. The cells were recovered by centrifugation at 3000rpm for 10min at 4⁰C. The pellet was gently re-suspended in 50ml Tfb 2, divided into 500µl aliquots and frozen in liquid nitrogen before long-term storage at -70⁰C. This method yielded

transformation efficiencies of $1 - 5 \times 10^7$ transformants per μg super-coiled DNA.

2.6 Transformation of *E. coli*

a) Competent *E. coli* strains were transformed with super-coiled DNA according to the following protocol:

b) Frozen competent cells were thawed on ice

c) $100\mu\text{l}$ of cells were added to 1-10ng of DNA and incubated on ice for 30 min

d) The mixture was heat shocked at 37°C for 5min. 1ml of LB was added to the transformation mix and the cells incubated for a further 30min, to allow expression of the antibiotic resistance markers. Cells were plated on appropriate antibiotic-containing plates and incubated overnight at 37°C .



2.7 Large-scale preparation of plasmid DNA

a) A 500ml – 1000ml culture of plasmid-containing *E. coli* in the presence of $100\mu\text{l.ml}^{-1}$ ampicillin was grown overnight with shaking at 37°C .

b) The bacteria were pelleted by centrifugation at 5000rpm for 10min at 4°C .

c) The pellet was re-suspended in 4ml GTE and incubated on ice for 5min. The cells were lysed by the addition of 8ml NaOH/SDS with gentle swirling and incubated on ice for 5min. 6ml of pre-chilled 3:5M KOAc was added and mixed gently to neutralise the alkali, and the mixture was incubated on ice for 5min.

d) The precipitate of cell debris, chromosomal DNA and SDS was removed by centrifugation at 10 000rpm for 15min at 4°C . The supernatant was filtered through

glass wool to remove particulate material and the nucleic acids precipitated by the addition of 1 volume propan-2-ol, followed by incubation at -20°C for 30min.

e) The precipitate was pelleted by centrifugation at 10 000rpm 15min at 4°C . Plasmid DNA was separated from RNA by either polyethylene glycol (PEG) precipitation or double CsCl/ethidium bromide fractionation.

2.7.1 PEG precipitation

- (i) The pellet was re-suspended in 5ml TE. $100\mu\text{g/ml}$ of RNase was added and incubated at 37°C for 60min.
- (ii) The RNase was removed by a single extraction with 1:1 phenol: chloroform. The supernatant was transferred to a clean tube, 0.3M NaOAc was added and the nucleic acids precipitated by the addition of 2.5 volumes of ethanol and incubated at -20°C for 30min.
- (iii) The pellet was collected by centrifugation at 10 000rpm for 10min at 4°C and re-suspended in 2ml 2.5M NaCl by vortexing. 2ml 13% polyethylene glycol 8,000 was added to precipitate the DNA. The mixture was vortexed thoroughly and incubated on ice for 30min.
- (iv) The precipitate was pelleted at 10 000rpm for 10min at 4°C . The pellet was washed with 70% ethanol and re-suspended in $100\mu\text{l}$ TE.

2.7.2 Double CsCl/ethidium bromide fractionation

This method was used to produce high quality plasmid DNA

- (i) The propan-2-ol pellet was re-suspended in 4.5ml TE, 5.75g CsCl and 400 μ l of 10mg/ml ethidium bromide was added and the mixture was vortexed.
- (ii) 1.61g/ml density of the mixture was required. TE and saturated CsCl were used to decrease and increase the density respectively until the required density was achieved. The mixture was centrifuged at 12 000g for 10min at 20⁰C.
- (iii) The solution was transferred to quickseal ultracentrifuge tubes and was centrifuged at 55 000rpm for 16hrs at 20⁰C. The plasmid DNA was recovered using a syringe and extracted 3 times with salt-saturated propan-2-ol to remove ethidium bromide.
- (iv) 2 volumes of water were added and the solution thoroughly mixed to dissolve salts, 1 volume propan-2-ol was added and the solution was mixed and incubated on ice for 10min to precipitate DNA.
- (v) The mixture was centrifuged at 10 000rpm for 15min and the pellet was washed with 70% ethanol. The pellet was re-suspended in 100 μ l TE

DNA purified by both methods was found to be free of contaminating chromosomal DNA and RNA when examined by agarose gel electrophoresis. Plasmid DNA was stored at 4⁰C (short term) or -20⁰C (long term).

2.8 Small scale preparation of plasmid DNA.

In order to analyse large numbers of transformants, plasmid DNA was isolated from overnight cultures of *E. coli* picked from single colonies. This method yielded DNA of sufficient quantity and purity to perform restriction analysis.

- a) 1.5ml saturated overnight cultures of plasmid-containing *E. coli* were grown in LB containing $100\mu\text{g ml}^{-1}$ ampicillin.
- b) The bacteria were pelleted in a microfuge at 6000rpm for 10min and re-suspended in 200 μl GTE after which the mixture was incubated at room temperature for 5min. 400 μl NaOH/SDS was added to lyse the cells and incubated at room temperature for 5min. The mixture was neutralised by the addition of 300 μl 3:5 KOAc, mixed gently and incubated at room temperature for 5min.
- c)  The precipitate was pelleted at 13 000rpm for 15min. 800 μl of the supernatant was added to 600 μl propan-2-ol in a fresh tube and incubated at -20°C for 30min.
- d) The precipitate of nucleic acid was pelleted at 13 000rpm for 10min, washed with 70% ethanol and re-suspended in 500 μl TE.
- e) 100 $\mu\text{g/ml}$ RNase was added and incubated at 37°C for 1hr and RNase was removed by single extraction with 1:1 phenol: chloroform.
- f) The DNA was recovered by ethanol precipitation in which 1/10 volume of 3M NaOAc and 2.5 volumes of ethanol were added to precipitate the DNA and the pellet was once again washed with 70% ethanol and re-suspended in 50 μl TE.

2.9 Restriction enzyme digests

Restriction enzymes were used according to the manufacturers' instructions. Where multiple digests were to be performed, the buffer conditions were selected to be compatible with both enzymes. Where this was not possible, after the completion of the first digest, the enzyme was removed by phenol:chloroform extraction and ethanol precipitation. The buffer conditions were adjusted accordingly and the second digest performed. All restriction enzymes used in this study were purchased from Roche. The PCR product was cut out of 1% agarose gel, recovered using BIO 101 gene-clean kit (Southern Cross Biotechnology) and re-suspended in 10ml of 1 x TE before restriction digest.

2.9.1 Shrimp alkaline phosphatase treatment of restriction digested DNA

200µg/µl of DNA was digested with *Bam* HI (10U/µl) and *Xho* I and included shrimp alkaline phosphatase and 0.5µl BSA in the digestion mixture. The total volume of the digestion was 50µl. Restriction enzymes and BSA used in this experiment were from New England Biolabs. The digestion mixture was incubated at 37⁰C for 3hrs

2.10 Ligations of insert to vector

Ligations were carried out in 300mM Tris-HCl (pH 7.8), 100mM MgCl₂, 100mM DTT and 10mM ATP, using 3U DNA ligase and less than 10ng DNA (i.e. vector and insert) at room temperature for 16hrs. T4 DNA ligase kit from Promega was used in this study. Ligation reactions were then transformed into competent *E. coli* MC 1061 cells and 100µl of the ligation reaction was plated on LB plates with 100µg/ml ampicillin, in duplicate.

2.11 Agarose gel electrophoresis of DNA

2.11.1 Gel preparation and electrophoresis

Gels were prepared by boiling the appropriate mass of agarose in 1xTBE, cooling to 50°C, adding ethidium bromide to 1µg ml⁻¹ and then setting this in the chosen gel former. TBE was used in the preparation and electrophoresis of all gels except when gene cleaning was to be conducted.

2.11.2 Sample preparation

Approximately 1µg DNA in 5 – 20µl was mixed with 0.5 volume of DNA loading buffer and loaded into the gel.

2.11.3 Detection of DNA

Ethidium bromide intercalates DNA, and in this state fluoresces when illuminated by ultraviolet (UV) light. For analytical purposes, DNA was visualised by illuminating the gel with short wave UV light on a trans-illuminator. When DNA was to be

recovered from the gel, a hand-held long wavelength lamp was used to avoid damage of the DNA.

2.12 Amplification of DNA by the polymerase chain reaction (PCR)

The two different PCR kits used in this study are discussed below.

Takara Taq PCR

1U of Takara *Taq* polymerase per PCR reaction was used, 2.5mM MgCl₂, 10ng of template, 50µM dNTPs and 25pmol of each primer were used in each reaction. A standard buffer of 20mM Tris-HCl pH 8, 100mM KCl, 0.1mM DTT, 0.5% Tween 20, 0.5% Nonidet P-40, 50% Glycerol was used. The reaction mix was incubated at 95⁰C, 2min followed by 94⁰C, 30sec; 64⁰C, 30sec; 72⁰C, 1min for 35 cycles followed by an incubation at 72⁰C for 10min.



Mastermix PCR

A master mix from Promega was used in PCR reactions. PCR Master Mix^(a,b,c) includes Nuclease-Free Water and PCR Master Mix. PCR Master Mix is a premixed, ready-to-use solution containing *Taq* DNA Polymerase, dNTPs, MgCl₂ and reaction buffers at optimal concentrations. PCR reactions included addition of 10ng DWNN encoding cDNA, 1µl of 10pmol forward and reverse primer, an extra 2mM MgCl₂ and 18µl nuclease free dH₂O from Promega mastermix.

2.13 Colony PCR

The Colony PCR technique was used to screen for possible clones and carried out using a protocol modified from Gussow and Clackson 1989. The colonies to be screened were re-suspended in 10 μ l of dH₂O. 11 colonies from ligation plate 1 and 1 colony from ligation plate 2 were selected using sterile toothpicks and re-suspended in 10 μ l of dH₂O. 1 μ l of the re-suspended colony from each tube was used in each colony PCR reaction. The colony PCR was set-up in exactly the same way as the PCR technique presented in Section 3.2., except that the re-suspended colonies were used instead of template DNA. A negative control that contained dH₂O instead of DNA was prepared to give an indication of contamination occurrence, which would lead to the presence of false positives PCR bands. A positive control was also prepared containing a DWNN cDNA and was expected to generate long-form DWNN PCR product.



2.14 Sequencing of DNA

Sequencing was done using the ABI PRISM 310 Genetic Analyser (Applied Biosystem). The samples to be sequenced were prepared as follows:

a) PCR Chain-termination based cyclic reactions

The reaction included 4 μ l of terminator ready reaction mixture, 4 μ l of 2.5 x sequencing buffer, 100ng/ μ l of DNA to be sequenced, 4 μ l of primer DNA and 4 μ l of dH₂O. 0.2ml tubes were used for cycling reaction. The cycling conditions used were:

96⁰C for 10sec }
}

50⁰C for 5sec 25 cycles

60⁰C for 4min

b) Precipitation of DNA

The cycled samples were transferred into 1.5ml microcentrifuge tubes. 16µl of dH₂O and 64µl of non-denatured 95% ethanol were added into each PCR sample. The samples were vortexed briefly. The tubes were then incubated at room temperature for 15min. The tubes were then centrifuged at room temperature for 20min at 13 000rpm and the orientation of the tubes were marked in order to know where the pellets are. The supernatant was carefully removed using a new pipette tip. 250µl of 70% ethanol was added and the tubes were vortexed briefly, this wash step was done twice. The tubes were centrifuged for 10min at 13 000rpm, again marking the orientations of the tubes. The supernatant was carefully removed. The samples were dried for 15-30min at 37⁰C.

The pellets were re-suspended in template suppressor reagent and incubated at 96⁰C for 5min. The samples were then loaded in the sequencing machine and sequenced.

2.15 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of proteins

Proteins were separated by denaturing them in discontinuous SDS-polyacrylamide gel electrophoresis, under reducing conditions. All gels were constructed from a 40% commercial Bis-acrylamide solution. The resolving/separating gel consisted of 12%

acrylamide, 0.375M Tris-Cl pH 8.8, 0.1% SDS, 0.1 APS, 0.04% TEMED. The stacking gel was made on top of the resolving gel and consisted of 5% acrylamide, 0.125M Tris-Cl pH 6.8, 0.1% SDS, 0.1% APS, 0.05% TEMED.

The 2x protein loading buffer was added at 50% of the final sample volume. Samples were boiled for 5min and vortexed thoroughly before loading. Gels were run at a constant 200V for 45min to 60min. Gels were stained in the staining solution and de-staining was achieved by multiple washes with destaining solution. Gels were incubated in the drying solution for 10min. Gels were dried onto a filter paper under vacuum for 2hr at 80⁰C.

2.16 Expression screening of transformants



This was prepared by inoculation of *E. coli* transformant colonies into 0.5ml LB with 100µg/ml ampicillin and 0.3% glucose to improve growth as this experiment was set up in 1.5ml centrifuge tubes. The tubes were incubated shaking at 37⁰C for 4 hrs. 3µl of 0.1M IPTG was added to the tubes and were incubated for further 2hrs. The cells were centrifuged and the supernatant was discarded from each tube. 0.5ml PBS was used to re-suspend the cells. 15µl of cell suspension and 5µl gel dye were added and heated at 80⁰C for 5min and were analysed on SDS/PAGE.

2.17 Recombinant production of 3C protease

pGEX-3C protease construct was transformed into *E.coli* BL 21 Gold competent cells and plated onto ampicillin LB plates. One colony was inoculated into 100ml LB with 100µg/µl ampicillin and incubated at 37⁰C o/n. The o/n culture was transferred into 900ml ampicillin LB and the culture was grown at 37⁰C until OD₅₅₀ reached 0.6. The final concentration of 0.3mM IPTG was added and the culture was induced at 25⁰C o/n. The cells were centrifuged at 5000rpm for 10min at 4⁰C and the supernatant was discarded. The pellet was re-suspended in 10ml PBS and lysed using a freeze-thaw method. The cell extract was clarified at 4⁰C for 30min at 10 000rpm. The GST-3C protease fusion was purified on glutathione 4B column. The GST column was prepared by washing with 5 column volumes (CV's) of dH₂O and equilibrated with 5 CV's of PBS. The protein sample was loaded on the column (the flow through was kept). The column was washed with 5 CV's of PBS to remove unbound proteins. The bound GST-3C protease fusion protein was eluted with 15mM glutathione in 50mM Tris-Cl pH 8. The expression and purification was analysed on 12% SDS gel, shown in Fig. 4.2 and was successful. 100µl aliquots of 3C protease in PBS were prepared and kept at -20⁰C.

2.18 Expression and purification of recombinant DWNN

2.18.1 Expression and purification of GST-long-form DWNN (Section 3.4)

A single colony from the transformation plate was inoculated in 100ml LB with 100µg/ml ampicillin and incubated at 37⁰C o/n. The o/n culture was transferred into 900ml LB with ampicillin and the culture was grown at 37⁰C until OD₆₀₀ reached 0.6.

IPTG was added to a final concentration of 0.3mM in a litre culture to induce expression of the fusion protein and the culture was incubated at room temperature o/n. The culture was centrifuged at 5 000rpm for 10min at 4⁰C and the pellet was re-suspended in 10ml TTBS containing 2mM PMSF to inhibit bacterial proteases. Working on ice, the re-suspended pellet was sonicated using 5 bursts of 30sec each, to lyse the cells. The sonicated sample was then centrifuged at 10 000rpm for 30min at 4⁰C to precipitate cell membranes and insoluble proteins. 100µl of the clarified extract was kept for analysis on SDS-PAGE gel.

The clarified extract was purified on a glutathione sepharose 4B column using a BIORAD Econo low-pressure pump. After loading, the column was washed with 5CVs of TTBS to remove unbound proteins. The bound fusion protein was eluted from the column with 15mM glutathione in TBS and collected in 1ml fractions using BIORAD fraction collector.

2.18.2 Expression and purification of GST-long-form DWNN (Section 4.2)

5ml o/n cultures of positive transformants were prepared and transferred into 45ml LB containing 100µg/ml ampicillin. Cultures were incubated at 37⁰C until OD₅₅₀ reached 0.5 and were then induced with 0.3mM IPTG for 4hrs at 25⁰C. The induced cells were centrifuged at 5 000rpm for 10min after which pellets were re-suspended in 2ml PBS and sonicated in bursts of 30 seconds five times.

Two BIORAD polyprep^R columns containing 1ml of Glutathione sepharose 4B beads each were prepared, washed with 5CV's of dH₂O and equilibrated with 5CV's of

PBS. Each sample was loaded onto each column, and washed with 4CV's of PBS to remove non-specifically bound proteins. The fusion protein was eluted with 10mM glutathione in 50mM Tris pH 8 and 2ml fractions were collected and analysed on 12% SDS-PAGE gel.

2.18.3 Expression and purification of GST-short-form DWNN

A single colony was selected and inoculated into 200ml LB containing 100µg/µl ampicillin and incubated at 37⁰C o/n. The o/n culture was transferred into 1800ml of LB with ampicillin and incubated at 37⁰C until OD₅₅₀ reached 0.6. The cells were then induced with 0.3mM IPTG at 25⁰C o/n. The 2000ml culture was centrifuged at 5 000rpm for 10min at 4⁰C to pellet the cells and the pellet was re-suspended in 20ml of PBS. The cells were lysed using the freeze-thaw method to release the proteins. Freeze-thaw method includes incubation of sample at -70⁰C for 5min and then at 37⁰C for 5min (repeating this cycle three times). The crude extract was centrifuged at 10 000rpm for 30min to clarify the extract for purification on a GSH column. The glutathione sepharose 4B column was packed and equilibrated with 5CV's of PBS. The clarified extract was loaded onto the column and the column was washed with 5CV of PBS. The bound fusion protein was eluted in a volume of 10ml using 15mM glutathione in 50mM Tris-Cl pH 8. 200µl of the eluate was kept at 4⁰C for analysis.

2.19 Lyophilisation of a protein sample

Protein sample (~ 1ml) was transferred into a 15ml tube, used liquid Nitrogen to freeze the sample, uncapped the tube and freeze-dried o/n.

CHAPTER 3

Recombinant expression of the long-form human DWNN domain in *E.coli*.

3.1 Introduction

This chapter and the next two chapters describe the recombinant expression of two forms of the human DWNN domain. Amplification of the DWNN gene from DWNN cDNA using PCR, cloning of the PCR product into pGEX vectors, expression in *E. coli* and subsequently purification of recombinant proteins are presented in detail. The aim of this study was to express high levels of the DWNN domain in order to produce concentrated samples for Nuclear Magnetic Resonance studies. Human DWNN cDNA was used as a template for amplification of all human DWNN fragments.



Long-form DWNN consists of the consensus DWNN sequence (81 amino acids) and a 37 amino acid tail (see Fig. 3.1). Full-length DWNN is also termed “long-form DWNN” in this presentation. cDNA corresponding to the human DWNN domain was ordered from ATCC (Rockville, USA) and its sequence is shown in Appendix A.

3.2 PCR amplification of the coding region for human long-form DWNN

Forward and reverse primers for the amplification of the coding region for the human DWNN domain were designed and synthesised, as shown in Fig. 3.2 and Fig. 3.3 respectively. The forward primer covered the eukaryotic transcription initiation sequences (AUG) and a *Bam* HI restriction site for cloning was introduced. Six bases

before a *Bam* HI site were incorporated to allow efficient restriction digest of the PCR product. The complementary region of the primer consisted of 24 base pairs.

The reverse primer was designed such that it included *Xho* I restriction site for cloning and two stop codons (UAA and UGA) to ensure that the translation of the fusion protein ends in the required region. Six bases were also introduced in this primer prior to *Xho* I site. The coding region for long-form DWNN was successfully amplified by PCR using Takara *Taq* polymerase as described in Section 2.12 and the results are shown in Fig. 3.4.

3.3 Insertion of long-form DWNN PCR product into pGEX-4T-3 vector

The PCR product of the long-form DWNN was cloned into the pGEX-4T-3 expression vector as described in Sections 2.10. The ligation results show that 11 colonies were present in one plate and 13 colonies in another plate, as two plates were prepared for the transformation of the ligation reaction. 12 colonies from the ligation plates were used to screen for the presence of transformants using colony PCR following the protocol in Section 2.13).

The engineered primers were used in colony PCR and thus the expected product would be equal to the size of PCR product (384bp). The results of the colony PCR are shown in Fig. 3.5. The absence of a negative control PCR band in Lane 2 indicated that there was no contamination and presence of the PCR band at 380bp in Lane 3 showed that the PCR worked. The presence of one positive transformant in Lane 11 indicated the possibility that the expression construct was generated successfully.

A plasmid DNA from a colony that contained a positive transformant was extracted using the CsCl₂ gradient large-scale DNA purification method and digested with *Bam* HI and *Xho* I to verify the presence of the insert from the transformant. The results were visualised on 1% agarose gel (see Fig. 3.6). The presence of a 370 bp band of insert and that of the vector at 4.9kb indicated that, DWNN-pGEX-4T-3 construct was successfully generated.

The pGEX-4T-3-long-form DWNN construct was sequenced using the engineered forward and reverse primers (Fig. 3.2 and Fig. 3.3 respectively) on the ABI PRISM 310 Genetic Analyser (Applied Biosystems), to determine if any mutations had occurred that would lead to the truncation of the expressed fusion protein. The sequence alignment of the sequenced DWNN (using forward primer) (Query) against the expected pGEX-6P-2 – long-form DWNN construct (APPENDIX B) is shown in Fig. 3.7, indicating that the alignment starts from base 35 of the DWNN sequence as the first few bases of the DWNN sequence form part of the primer-binding region. The *Xho* I site is visible at base 357-362 of the Query sequence. Fig. 3.8 shows the alignment of the DWNN sequence (produced by the reverse primer) with the expected pGEX-6P-2 – long-form DWNN construct. The sequence alignment shows that the entire sequence of the construct generated has no mutations because the first 35 bases of DWNN sequence that were absent from the sequence generated by the forward primer are included in the sequence generated by the reverse primer. The third base of the query sequence mismatched with the base of the expected sequence

because the base-caller cannot read accurately at the beginning or at the end of the sequence.

3.4 Expression and purification of GST-long-form DWNN fusion protein

Human DWNN - pGEX-4T-3 construct was transformed into competent *E.coli*. BL 21 (expression host strain) cells. pGEX-4T-3 vector is a GST gene fusion vector that has a thrombin recognition site for cleaving the target protein from the fusion protein. GST-fusion protein could be purified from the cell lysate by affinity chromatography using glutathione sepharose 4B column. The size of GST subunit is 26kDa and the size of the DWNN is 13kDa and thus the expected subunit size of the fusion protein is 39kDa. The long-form human DWNN was expressed and purified as described in Section 2.18.1. A_{280} of the elution fractions were determined to investigate fractions that contained the fusion protein. The graph of fraction number v/s A_{280} was plotted and is shown in Fig. 3.9. It can be concluded from the results that most of the fusion protein was eluted on fractions 14 – 21. These eight fractions were pooled and were analysed on SDS/PAGE gel, as shown in Fig. 3.10. The results show that the long-form DWNN is highly expressed in *E. coli* as seen in Lane 2 and was purified successfully as shown in Lane 3. The purified protein in Lane 2 appears diffuse, which could suggest possible heterogeneity in the sample.

3.5 The cleavage of the fusion protein by thrombin and purification of DWNN

pGEX-4T-3 expression vector incorporates a thrombin cleavage site between the GST and the target protein. Thrombin (Sigma) from the human plasma was used in this study. Half of the pooled purified fusion protein (4ml) was digested with 10 units of thrombin at 37⁰C o/n and the results were visualised on 12% SDS/PAGE gel (see Fig. 3.11). The expected subunit size of long-form DWNN was 13kDa and that of GST was 26kDa but the results reveal that the digested sample had three bands of sizes, 39kDa, 26kDa and 13kDa. These three bands corresponded to the fusion protein, GST and long-form DWNN respectively. The presence of the fusion protein in the digestion sample indicated that the fusion protein was not digested to completion.



The partially cleaved, DWNN fusion sample was dialysed against TBS at 4⁰C o/n to remove glutathione from the sample which prevents binding of GST to the GST column. After dialysis the sample was loaded on the glutathione agarose column, where DWNN was collected as the flow through. GST and GST-fusion bound on the column were eluted using 15mM glutathione in TBS. The samples were analysed on 12% SDS/PAGE gel, illustrated in Fig. 3.12, and the results show that the long-form DWNN domain was purified successfully from the GST and the fusion protein..

3.6 Mass spectrometric analysis of long-form DWNN

Fig. 3.13 shows MALDI-TOF analysis of a sample of undigested fusion protein. The expected size of GST-long-form DWNN fusion protein was 39kDa. The peak at 37kDa corresponds to the fusion protein. The peak at approximately 18kDa is a common occurrence in MALDI-TOF mass spectrometry; it is the result of the fusion molecules having acquired two units of charge during the ionisation phase rather than one, and which consequently appear with a mass/charge ratio of $37.5/2$ which is ~ 18.8 kDa.

The presence of smaller peaks at ~ 10 kDa (corresponding to DWNN) and 26kDa (corresponding to GST) is an indication that cleavage of the fusion protein took place between GST and DWNN in the absence of thrombin. Whether it happened *in vivo* or *in vitro* is not clear. Furthermore the presence of the DWNN fragments in a GST-affinity purified sample means either that proteolysis took place after elution of the fusion protein from the glutathione column, or else that the cleaved DWNN domain co-elutes with either the GST or the fusion protein.

Purified DWNN was digested from fusion protein by thrombin and isolated to prepare for MALDI-TOF mass spectrometry. The expected sequence is shown in Fig. 3.1.

There are two extra amino acids from the GST that are included in the DWNN after cleavage with thrombin and these two amino acids are G and S. The molecular

weight of the protein was found using the on-line server which forms part of the ExPASy suite of proteomics tools (URL: www.expasy.org) and found to be 13 535.69Da.

DWNN was subjected to MALDI-TOF mass spectrometry analysis using the facility at the Department of Molecular and Cell Biology, UCT. The results are presented in Fig. 3.14. These results show that long-form DWNN was shorter than expected with a MW of approximately ~10kDa. This suggested that DWNN had been proteolysed and was suspected that the proteolysis had taken place in the hydrophobic tail region, which is expected to be un-structured and therefore exposed to proteolysis. It was also suspected that thrombin played part in the unspecific cleavage of DWNN.



3.7 Optimisation of thrombin cleavage

Optimisation of conditions for thrombin cleavage was required as initial efficiency was approximately 60%. The potential factors considered were the components of the cleavage buffer and the possibility that the steric hindrance between the GST and the DWNN domain was suspected on preventing the thrombin molecule from gaining entry to the cleavage site. To test the buffer conditions; four 100 μ l thrombin digests were prepared, each containing the purified fusion protein and 10 units of thrombin in TBS, with CaCl₂ to a final concentration of 2.5mM. In addition, 0.5mM DTT was added to digest 2, 0.01% SDS was added to digest 3 and 0.5mM DTT and 0.01% SDS were added to digest 4. In all these digests, the amount of fusion protein in TBS

added was ~100 μ l. The digests were incubated o/n at room temperature and then analysed on a 12% SDS/PAGE gel, as shown in Fig 3.15.

It has been observed in Fig. 3.15, that Lane 3 (digest 1) contains three main bands in the region ~ 13kDa, which are clearly alternative forms of the DWNN domain. The question arose as to whether these were N- or C- terminal digestions and whether these were the result of bacterial proteases, either pre or post lysis, non-specific cleavage by thrombin or C-terminal truncations during transcription. The tightness of the GST band and the diffuse nature of the uncut fusion in lane 2 suggest that there were variant C-terminal forms. This came as no surprise since this was expected because the last 37 amino acids of the protein were known to be un-structured and thus vulnerable to bacterial proteases. It was suspected that the largest band of DWNN (13kDa) corresponded to the complete DWNN domain, and the smallest to a complete removal of the 37 amino acid tail.

Lane 4 shows a similar pattern as Lane 3, from which it was deduced that DTT had little effect on the digestion. DTT might be expected to lead to increased proteolysis of proteins stabilised by disulphide bonds, which are reduced by DTT; the absence of any effect comes as no surprise since DWNN was not expected to be stabilised by disulphide bridges. Lanes 5 and 6 show a number of new bands, indicating that the unfolding effect of the SDS had exposed new sites to proteolysis by thrombin. The band below that of GST represents proteolysis at a site in the GST sequence,

indicating that 0.01% SDS was sufficient to unfold GST enough for thrombin to cleave it internally. The three DWNN bands in lane 3 and 4 had been proteolysed down to a ~ 9kDa band. This is consistent with a model of a partially un-structured C-terminal tail, which becomes totally un-structured in the presence of SDS, allowing complete removal of the tail. The absence of a band at ~39kDa (fusion protein) in Lane 3-6, showed that thrombin digestion was efficient. This is presumably due to the addition of 2.5mM CaCl_2 that play a role in the activity of thrombin.



CHAPTER 4

Recombinant expression of the long-form human DWNN using pGEX-6P-2 vector

4.1 Insertion of the coding region for long-form DWNN into the pGEX-6P-2 vector

The pGEX-4T-3 expression vector was replaced by pGEX-6P-2 expression vector for expression of long-form DWNN, due to the problems experienced with thrombin cleavage of the fusion protein. The pGEX-6P-2 vector contains a recognition site for the 3C protease at the immediate N-terminal to the target protein for removal of the GST domain. 3C protease has a higher site-specificity than thrombin (Walker *et al.*, 1994).



The DWNN construct produced in Section 3.2 was digested with *Bam* HI and *Xho* I (from New England Biolabs) and inserted into pGEX-6P-2 expression vector, which had been digested with the same enzymes, using the protocol described in Section 2.9.1 Shrimp alkaline phosphatase (SAP) was included in the digestion of the vector in order to remove phosphate groups at the end of singly digested vector fragments to prevent re-circularisation of vector, which leads to high background in transformation plates. The digestion sample was loaded on a QIAprep spin column from QIAGEN to purify the digested vector DNA. The ligation reactions were prepared as shown in Table 2.2.

The ligation samples were transformed into *E.coli* MC1061. The results were recorded based on the number of colonies on the LB plates. In sample 1 plate (background control in which no vector was added in the ligation mixture) only 5 colonies were observed and the colonies in this reaction resulted either from undigested vector or re-circularisation of singly cut vector. The number of colonies in ligation reaction samples 2, 3 and 4 were 11, 4 and 8 respectively thus showing that the number of colonies in the ligation reaction samples were greater than the number of colonies in sample 1, indicating that there was a high probability that some of the colonies in samples 2-4 contained inserts.

8 colonies from sample 4 of the ligation reaction were screened for protein expression as described in Section 2.16. A positive control transformant containing the long-form DWNN - pGEX-4T-3 construct and a negative control transformant pGEX vector were included in the experiment. The clarified extract samples were visualised on a 12% SDS/PAGE gel. The 39kDa band visible in Lane 2 (positive control) corresponds to the fusion protein. The 26kDa band in Lane 3 (negative control) corresponds to GST subunit. A band at 39kDa corresponding to the fusion protein is visible in lanes 6, 7 and 11, indicating that the desired construct were present in these transformants. In Lane 4, a band corresponding to GST (26kDa) is visible, indicating, that the ligation was unsuccessful in this transformant.

4.2 Large-scale expression and purification of long-form DWNN

Plasmid DNA was extracted from the *E. coli* transformant colonies resulting from transformations above. The plasmid DNA was then transformed into competent *E. coli* BL21 expression host cells. Large-scale expression and purification of long-form DWNN was prepared according to a protocol in Section 2.18.2. 500µl of the eluted protein was digested o/n at 4⁰C with 20µl of non-commercial 3C protease (its expression and purification is described in Section 2.17 and shown in Fig. 4.2). 9mM EDTA was added for 3C protease activity because this protease is inhibited by Zn ions at 10mM and thus EDTA chelates those ions ensuring that 3C protease activity is optimal. Digests were visualised on a 12% SDS/PAGE gel, which is shown in Fig. 4.3. It is evident from Fig. 4.3, that the fusion protein was successfully purified (Lanes 2 and 4) and cleaved to completion by 3C protease (Lanes 3 and 5) through the presence of a 26kDa protein band that represents GST and double bands smaller than 14kDa representing DWNN. The cleaved DWNN was in the form of a doublet, one of approximately 13kDa and one of approximately 10kDa. These results led to the conclusion that proteolysis of the tail region of DWNN occurred.

4.3 Sequencing of the pGEX-6P-2 - long-form DWNN construct

To determine if the expression construct had been correctly assembled, it was sequenced using the original PCR primers shown in Fig. 3.2 and 3.3. An alignment of the sequence generated using the forward primer (Query), with the expected pGEX-6P-2 – long-form DWNN construct sequence (Sbjct) (see Appendix C). Bases 1-51 (corresponding to DWNN sequence) were not present in the sequenced construct

because they formed part of the primer-binding region. The rest of the sequence matched the expected sequence exactly; indicating that the construct contained no mutations, see Fig. 4.4. The sequencing covered the *Xho* I site which is in base 329-334 of the Query sequence, and the two engineered stop codons. The corresponding alignment of the sequence generated using the reverse primer, which is shown in Fig. 4.5, shows that the entire sequence (1-364) is correct.

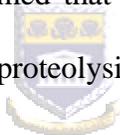
4.4 Mass spectrometric analysis of long-form DWNN

The cleaved fusion protein was dialysed against PBS to remove free glutathione and loaded onto a glutathione agarose column. GST was retained on the column, allowing DWNN to flow through. Purified DWNN fractions were pooled and analysed by MALDI-TOF mass spectrometry.



Figures 4.6 and 4.7 show mass spectrograms of two different samples, both prepared as described above. Both samples show evidence of proteolysis: in Fig. 4.6 the predominant species has a molecular weight of approximately 11kDa, while in Fig. 4.7 there are two clusters of approximately 10.5kDa and 12kDa respectively. A magnified picture of the DWNN region from Fig. 4.6 is shown in Fig. 4.8. The pattern of discrete peaks separated by approximately 100kDa is consistent with a sample in which any two molecules differ from each other by an integral number of amino acids.

To determine what this would correspond to in the context of DWNN, the molecular weight of a number of C-terminal truncations of DWNN was predicted using the ExPASy MW server (see Fig. 3.1). In each case the additional amino acids GPLGS added by the pGEX-6P-2 vector were included in the calculation. Truncation after K81 and S82 would give rise to fragments with molecular weights of 9.5kDa and 9.617kDa respectively, from which it was concluded that proteolysis was taking place in the tail region. This is perfectly consistent with the model of this region as an unstructured tail. The absence of any peaks of molecules smaller than 9.6kDa suggests that the consensus region of DWNN is tightly folded and therefore resistant to proteolysis. The fact that proteolysis was still occurring despite the use of 3C protease rather than thrombin confirmed that thrombin was not responsible for the degradation; the most likely cause of proteolysis was due to bacterial proteases, either *in vivo* or *in vitro*.



4.5 1D proton NMR spectrum of purified long-form DWNN

The purified DWNN was lyophilised o/n, re-suspended in 600µl D₂O and inserted into a 5mm NMR tube using a Pasteur pipette. The 1D proton spectrum collected on the 600MHz spectrometer at Stellenbosch University is shown in Fig. 4.9. The aliphatic peaks to the right of the water resonance (4.7ppm) are an indication that protein is present in the sample as it is shown in Fig. 1.5 by the 1-D proton NMR spectrum of ubiquitin. The fact that some of the methyl groups are shifted as far as 0ppm is an indication that the protein is well folded. The amide region to the left of

the water is suppressed due to exchange of the amide protons and replacement with deuterons, which have no resonance in this part of the spectrum. The large peak in the vicinity of 3.6ppm corresponds to the Tris buffer.



CHAPTER 5

Recombinant expression of the human short-form DWNN domain

5.1 Insertion of the coding region for short-form DWNN into the pGEX-6P-2 expression vector

On the evidence of the mass spectrometric analysis shown in Fig. 4.8, it was concluded that the tail region (see Fig. 3.1) is un-structured and therefore sensitive to proteolysis. It was therefore decided to shorten the construct by removing the tail region after S82. The coding region was re-amplified from the cDNA using the same forward primer shown in Fig. 3.2; the design of the new reverse primer is described below.



5.1.1 PCR amplification of the coding region for human short-form DWNN

As before, the new reverse primer incorporates an *Xho* I site for insertion into the pGEX-6P-2 multiple cloning cassette, as well as a double stop codon to further decrease the possibility of read through. The complementary region corresponds to bases 413 to 433 of the cDNA sequence (Appendix A), see reverse primer sequence in Fig. 5.1. PCR reactions were prepared using Master Mix as described in Section 2.12. The results are presented in Fig. 5.2. The successfully amplified product is shown in Lane 3; the size as estimated from the gel is consistent with the expected size of 273bp. The absence of any product in the negative control (Lane 2) is evidence that the band in Lane 3 does not result from contamination.

5.1.2 Insertion of short-form DWNN PCR product into pGEX-6P-2 vector

The PCR product was gene-cleaned and digested with restriction enzymes as described in Section 2.9. The digested short-form DWNN PCR product was ligated with digested pGEX-6P-2 vector as shown in Table 2.3 and described in Section 2.10. From the transformation results of the ligation reactions in sample 1 (representing background control), 21 colonies were observed. The colony number in the background control plate 1 is reasonable compared to the ligation plates of samples 3, 4 and 5, which had 119, 107 and 96 number of colonies respectively. 7 colonies were present in sample 2 (insert control) ligation plate and the presence of colonies in this ligation plate is an indication of the template carry-over from the gene-cleaning procedure.



8 colonies were selected from the transformation plates for the preparation of small-scale DNA extractions and DNA samples were digested with *Bam* HI and *Xho* I to confirm the presence of an insert. The results were analysed on 1% agarose gel electrophoresis (see Fig. 5.4). The presence of bands in the region of 250bp in Lanes 2, 5, 7 and 8 confirmed that the target gene had been successfully cloned into the corresponding colonies. A large-scale DNA extraction using a CsCl₂ gradient was performed on one of the colonies, and the presence of the insert was verified by *Bam* HI and *Xho* I digestion, as shown in Fig. 5.5.

5.1.3 Expression and purification of GST-short form DWNN fusion protein

The short-form DWNN construct was transformed into *E.coli* BL 21 competent cells and plated on LB plates with ampicillin (100µg/µl). The GST-short-form DWNN fusion protein was expressed and purified as described in Section 2.18.3. The fusion protein was then digested with 3C protease to release the short-form DWNN as described previously and the results are shown in Fig. 5.6. From Lane 3 it is clear that the fusion protein had been cleaved to completion, resulting in a band of 26kDa, which corresponds to GST, and one of ~ 10kDa, which corresponds to DWNN.

5.2 Sequencing of the pGEX-6P-2 - short-form DWNN construct

To determine if the pGEX-6P-2 construct contained the correctly shortened form of DWNN, DNA was sequenced using the pGEX 5' (forward) primer (Appendix E) on the ABI PRISM 310 Genetic Analyser (Applied Biosystems). The full sequence alignment of the sequence generated using the pGEX 5' primer (Query) against the expected pGEX-6P-2 – DWNN construct (Sbjct) (see Appendix D) is shown in Fig. 5.7. The sequencing data confirmed that the construct contained no mutations and was in the correct reading frame. The pGEX 5' primer-binding site is not close to the MCS of the pGEX vector and that is why the full sequence of the construct was attained during sequencing.

5.3 Reverse-phase purification of short-form DWNN

Separation of the DWNN domain from GST after cleavage had previously been achieved passively by removal of the GST using a glutathione agarose column following the protocol in Section 4.2 . Various methods were attempted to purify the DWNN domain more directly, including anion and cation exchange chromatography at a range of different pH's, hydrophobic interaction chromatography and reverse-phase chromatography. All except reverse-phase were unsuccessful, with DWNN eluting from the column without being retained. DWNN was successfully retained by an anionic exchange media (POROS 20HQ from PerSeptive Biosystems) at pH 11. However it was not clear whether this was due to specific ion-exchange interactions or simply that the protein was becoming denatured, and so this method was not pursued.



Reverse-phase chromatography is a technique that separates molecules according to hydrophobicity (Boyer, 1992). Hydrophobic proteins are retained on a hydrophobic solid phase in the presence of a polar mobile phase, but are selectively eluted as the polarity of the mobile phase is reduced. In this case the mobile phase was water, and the polarity was reduced using an increasing gradient of acetonitrile (ACN). The reverse phase media was 20RS POROS, which has similar characteristics to C18 media and is thus highly hydrophobic. The medium was packed into a 1.6 ml column and the mobile phase was pumped at a rate of 15ml/min using a BioCAD Sprint Perfusion Chromatography System (PerSeptive Biosystems). The protein sample was

loaded in 100% Buffer A (5% ACN, 0.1% TFA, pH 2) and then eluted with a gradient of 0 – 60% Buffer B (95% ACN, 0.087% TFA, pH 2).

The chromatogram is shown in Fig. 5.8. Three major peaks can be identified; the first corresponds to the protein that eluted from the column without being retained, while the second and the third were retained by the matrix and successively eluted. 20µl samples from each peak were analysed on a 12% SDS/PAGE gel, which is shown in Fig. 5.9. The gel shows that the second peak corresponds to DWNN, and the third to GST and 3C protease. Surprisingly, DWNN appeared as a doublet of bands in the vicinity of ~ 14kDa, with additional bands running between 21 and 31kDa (marked “A”). The additional bands were interpreted as being due to multimerisation, possibly due to disulphide bridges involving Cysteine residues (short-form DWNN contains three Cysteine residues). This is possible because the buffer contained no DTT. With the addition of 10mM DTT the high molecular weight bands were no longer observed.

5.4 Short-form DWNN MALDI-TOF mass spectrometry results

The purified DWNN domain was analysed by MALDI-TOF mass spectrometry to confirm its size. The expected molecular weight of short-form DWNN, as determined by ExPASy MW server, is 9.6kDa. The results are shown in Fig. 5.10. The two major peaks correspond to 9.8 and 10.1kDa respectively. While these are a little larger than expected, it was confirmed that the un-structured tail region has been successfully

removed through sequencing. Possible explanations for the slightly larger size of the molecule include cleavage by 3C upstream its expected cleavage site, or the addition of small-molecule adducts to the protein.

5.5 1D NMR results of purified short-form DWNN

Short-form DWNN sample was dialysed against 50mM Tris-Cl, pH 8 and then concentrated to the final volume of 500 μ l with concentration of \sim 0.5mM. 1D NMR data of purified short-form DWNN was collected using the 600MHz Varian spectrometer in the Chemistry Department of Stellenbosch University. The results are shown in Fig. 5.11. The ratio of signal to noise was low due to the low concentration of DWNN protein and the fact that the probe used was not optimised for protein work. However the amide peaks are visible to the left of water, aliphatics to the right and characteristic methyls in the vicinity of 0ppm. The spectrum confirms that the short- form DWNN domain is folded.

CHAPTER 6

Discussion and conclusion

6.1 DWNN expression constructs

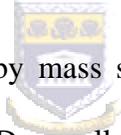
The coding regions for the two forms of DWNN domains have been inserted into two different types of pGEX vectors, pGEX-6P-2 and pGEX-4T-3, to produce expression constructs. Both vectors are components of the GST expression system and differ only in the type of protease recognition site they incorporate.

6.1.1 pGEX-4T-3 - long-form DWNN construct

The first expression construct was generated by cloning the full-length DWNN domain into pGEX-4T-3 vector, which has a thrombin recognition site. pGEX-4T-3 expression vector was selected because after thrombin cleavage of the fusion protein, no extra amino acids (from the GST) were included in the target protein (see Appendix B). The full-length DWNN domain, which is also designated “long-form DWNN”, is 118 amino acids long with a MW of approximately 13kDa. The amino acid sequence is shown in Fig. 3.1. The construct was successfully prepared. Expression and purification of the fusion protein was also successful but thrombin did not digest the fusion protein efficiently as it is evident from Fig. 3.11.

A number of experiments were done in order to improve the efficiency with which thrombin cleaved the fusion protein. These included the addition of denaturation agents in order to expose the thrombin cleavage site of the fusion protein. 0.01% SDS

and 0.5mM DTT were added as components of the cleavage buffer. It was observed that DTT had little effect on digestion while 0.01% SDS increased the efficiency of digestion to almost 100%. DWNN appears in a number of bands, the smallest of which is approximately 9kDa, which is the size of the domain after the removal of the tail region. These findings led to the understanding that these denaturation agents, especially SDS, led to the total unstructuring of the C-terminal tail of DWNN thus allowing proteolysis and subsequently the removal of the un-structured tail. The absence of any major bands less than 9kDa in size suggested that the consensus region of DWNN is resistant to proteolysis by thrombin even in the presence of 0.01%SDS.



These conclusions were confirmed by mass spectrometry, which showed that the fusion protein was approximately 2kDa smaller than expected, and cleaved DWNN domain was smaller than the expected by the same amount. The conclusion that was drawn was that the protein, DWNN domain was being non-specifically proteolysed in the C-terminal tail.

6.1.2 pGEX-6P-2 – long-form DWNN construct

Due to the problems encountered with thrombin cleavage, the pGEX-4T-3 vector was replaced with the pGEX-6P-2 expression vector, which contains a 3C protease recognition site. 3C protease has a longer recognition sequence and therefore a higher site-specificity than thrombin (Walker *et al.*, 1994).

Long-form DWNN coding region was successfully inserted into pGEX-6P-2 and the fusion protein expressed at high levels. 3C protease digested the fusion protein to completion.

The long-form DWNN was analysed by MALDI-TOF mass spectrometry, which confirmed that the expressed long-form DWNN was still subject to proteolysis. This was shown by the presence of peaks at the range of 10 – 11kDa in the spectrogram when the expected size of long-form DWNN was 13kDa. The absence of peaks smaller than 9.6kDa (see Fig. 4.6 – 4.8), which is the expected size of DWNN excluding the C-terminal, hydrophobic tail, suggested that the consensus region of DWNN was tightly folded and thus resistant to proteolysis. The fact that proteolysis was still taking place despite the use of 3C protease rather than thrombin suggested that thrombin had not been the main cause of proteolysis of the pGEX-4T-3 construct.

1D proton NMR data of purified DWNN was collected on the 600MHz spectrometer at Stellenbosch University. The data (see Fig. 4.9) showed that the protein was well folded, and that the buffers used were therefore suitable for large-scale purification for structure determination.

6.1.3 pGEX-6P-2 - Short-form DWNN construct

Because of the problems encountered with proteolysis of the C-terminal tail region, it was decided to remove it. The coding region for a shortened form of the domain was amplified using PCR and inserted into the pGEX-6P-2 vector.

The shortened domain was successfully expressed, purified and cleaved to completion with 3C protease. It was successfully separated from GST and 3C protease using two different methods: (i) reverse-phase chromatography and (ii) glutathione affinity chromatography to remove the GST and 3C protease, which is also a fusion with GST. Efforts to develop an ion exchange purification protocol were unsuccessful. Both purification methods yielded highly purified protein, as can be seen from Fig 5.8 and 5.9.



It was noticed that under certain conditions the purified DWNN was at a higher molecular weight on SDS-PAGE gels than expected (see Fig. 5.9). The higher molecular weight bands were interpreted as being due to disulphide bridges involving three cysteine residues present in short-form DWNN. This was possible because the buffer contained no DTT; addition of 10 mM DTT led to the removal of the high MW bands.

The purified short-form DWNN domain was analysed by MALDI-TOF mass spectrometry to confirm its size. The expected MW of short-form DWNN, as determined by ExPASy MW server, was 9.6kDa. The results are shown in Fig. 5.10,

and two major peaks corresponding to 9.8 and 10.1kDa were visible. Although these were a little larger than expected, they confirm that the un-structured region was removed. The sequencing data in Fig. 5.7 also confirmed the latter. Possible explanations for the slightly larger size of the molecule include cleavage by 3C protease upstream of its cleavage site, or the addition of small-molecule adducts to the protein.

1D proton NMR spectrometry in Fig. 5.13 confirmed that the short-form DWNN domain was folded.

6.2 Conclusion and future prospects

The final result of the work described in this thesis was the generation of the reagents and protocols necessary for the large-scale recombinant production of the DWNN domain for structural analysis. The DWNN domain was shown to be soluble and correctly folded under the buffer conditions employed. As expected, the C-terminal hydrophobic tail forming part of the full-length protein was unstable in the bacterial expression system employed, and was therefore removed.

Since the work described in this thesis was completed, the expression construct has been taken over by a co-worker and used to produce concentrated samples of pure short-form DWNN. These have been used to generate high quality NMR data from which the structure of the domain will shortly be generated.

REFERENCES

- Alberts, B., Bray, D., Lewis, J., Raff, m., Roberts, K., Watson, J.D. (2000)
Molecular Biology of the cell, 3rd Edition.
- Alnemri, E. S., Livingston, D.J., Nicholson, D.W., Salvesen G., Thornberry, N.A.
Wong, W.W., Yuan, J. (1996). Human ICE/CED-3 protease nomenclature.
Cell **87**: 171.
- Arends, M.J., Wyllie, A.H. (1991). Apoptosis: mechanisms and roles in
pathology. *Int. Rev. Exp. Pathol.* **32**: 223-254.
- Atkinson, E.A., Barry, M., Darmon, A.J., Shostak, I., Turner, P.C., Moyer, R.W.,
Bleackley, R.C. (1998). Cytotoxic T lymphocyte-assisted suicide Caspase 3
activation is primarily the result of the direct action of granzyme B. *J. Biol.
Chem.* **273**: 21261-21266.
- Barry, M., Heibein, J.A., Pinkoski, M.J., Lee, S.F., Moyer, R.W., Green, D.R.,
Bleackley, R. C. (2000). Granzyme B short-circuits the need for caspase 8
activity during granule-mediated cytotoxic T-lymphocyte killing by directly
cleaving Bid. *Mol. and Cell. Biol.* **20**: 3781-3794.
- 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. *Mol. Biol.* **280**: 275-286.
- Beresford, P.J., Jaju, M., Friedman, R.S., Yoon, M.J., Lieberman, J. (1998). A
role for heat shock protein 27 in CTL-mediated cell death. *J. Immunol.*

161: 161-167.

Boldin, M.P., Goncharov, Y.V., Wallach, D. (1996). Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell* **85:** 803-815.

Boyer, R. F., (1992). Modern Experimental biochemistry 2nd edition, pg. 96

Cavanagh, J., Fairbrother, W. J., Palmer II, A.G., Skelton, N.J (1996) Protein NMR Spectroscopy, *Principles and Practice*, Academic Press, San Diego, p 534

Chao, C., Saito, S., Kang, J., Anderson, C.W., Appella, E., Xu, Y. (2000). p53 transcriptional activity is essential for p53-dependent apoptosis following DNA damage. *EMBO J.* **19:** 4967-4975.

Chau, B.N., Cheng, E.H., Kerr, D.A., Hardwick, J.M. (2000). Aven, a novel inhibitor of caspase activation, binds Bcl-xL and Apaf-1. *Mol. Cell* **6:** 31-40.

Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell* **79:**13-21.

Clore, G.M., Gronenborn, A.M. (1991). Structures of large r proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252:** 1390-1399.

Cohen, G.M. (1997). Caspases: the executioners of apoptosis. *Biochem. J.* **326:** 1-16.

Darmon, A.J., Nicholson, D.W., Bleackley, R.C. (1995). Activation of the

- apoptotic protease CPP32 by cytotoxic T-cell-derived granzyme B. *Nature* **377**: 446-448.
- George, A.E. (1995). A new method for isolating genes involved in the processing and presentation of antigens to cytotoxic T-cells. *D. Phil. Thesis, Oxford*.
- Griffiths, G.M., Argon, Y. (1995). Structure and biogenesis of lytic granules. *Curr. Top. Microbiol. Immunol.* **198**:39-58.
- Guo, M., Hay, B.A. (1999). Cell proliferation and apoptosis. *Curr. Opin. Cell Biol.* **11**: 745-752.
- Gussow, D., and T. Clackson. (1989). Direct clone characterization from plaques and colonies by the polymerase chain reaction. *Nucleic Acids Res.*, **17**: 4000.
- Hahn, K., DeBiasio R., Tishon A., Lewicki H., Gairin J.E., LaRocca G., Taylor D.L., Oldstone, M. (1994). Antigen presentation and cytotoxic T lymphocyte killing studied in individual, living cells. *Virology* **201**: 330-340.
- Haupt, Y., Maya, R., Kazaz, A. and Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature* **387**: 296-299.
- Hengartner, M.O. (2000). The biochemistry of apoptosis. *Nature* **407**: 770-776.
- Hershko, A., Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* **67**: 425-479.
- Hupp, T.R., Lane, D.P., Ball, K.L. (2000). Strategies for manipulating the p53 pathway in the treatment of human cancer. *Biochem. J.* **352**: 1-17.
- Jentsch, S., Pyrowolakis, G. (2000). Ubiquitin and its kin: how close are the family ties? *Trends Cell Biol.* **10**: 335-342.

- Krammer, P.H. (2000). CD95's deadly mission in the immune system. *Nature* **407**: 789-795.
- Kubbutat, M.H.G. and Vousden, K.H. (1998). Keeping an old friend under control: regulation of p53 stability. *Mol. Med. Today* **4**: 250-256.
- Kumar, S., Tomooka, Y., Noda, M. (1992). Identification of a set of genes with developmentally down-regulated expression in the mouse brain. *Biochem.Biophys.Res. Commun.* **185**: 1155-1161.
- Kumar, S., Kao, W. H., Howley, P.M. (1997). Physical interaction between specific E2 and Hect E3 enzymes determines functional cooperativity. *J. Biol. Chem.* **272**: 13548-13554.
- Lennon, J.J.,  www.abrf.org/ABRFNews/1997/June1997/jun97lennon.html
- Liakopoulos, D., Doenges, G., Matuschewski, K., Jentsch, S. (1998). A novel protein modification pathway related to the ubiquitin system. *EMBO J.* **17**: 2208-2214.
- Lohrum, M.A.E., Vousden, K.H. (2000). Regulation and function of the p53-related proteins: same family, different rules. *Trends Cell Biol.*, **10**: 197-202.
- Lorick, K.L., Jensen, J.P., Fang, S., Ong, A.M. Hatakeyama, S., Weissman, A.M., (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl. Acad. Sci. USA* **96**: 11364-11369.
- Macleod, K. (2000). Tumor suppressor genes. *Curr. Opin. Genet. Dev.* **10**: 81-93.
- MALDI Basics, www.srsmaldi.com/Maldi/Maldi.html

- McMichael, A. (1992). Cytotoxic T lymphocytes: specificity, surveillance, and escape. *Adv. Cancer Res.* **59**: 227-244.
- Medema, J.P., Toes, R.E.M., Scaffidi, C., Zheng, T.S., Flavell, R.A., Melief, C.J.M., Peter, M.E., Offringa, R., Krammer, P.H. (1997). Cleavage of FLICE (caspase-8) by granzyme B during cytotoxic T lymphocyte-induced apoptosis. *Eur. J. Immunol.* **27**: 3492-3498.
- Meier, P., Finch, A., Evan, G. (2000). Apoptosis in development. *Nature* **407**: 796-801.
- Nicholson, D. W. (2000). From bench to clinic with apoptosis-based therapeutic agents. *Nature* **407**: 810-816.
- Peter, M.E., Krammer, P.H. (1998). Mechanisms of CD95 (APO-1/Fas)-mediated apoptosis. *Curr. Opin. Immunol.* **10**: 545-551.
- Sakahira H., Enari M., Nagata S. (1998). Cleavage of CAD inhibitor in CAD activation and DNA degradation during apoptosis. *Nature* **391**: 96-99.
- Sakai, Y., Saijo, M., Coelho, K., Kishno, T., Niikawa, N., Taya, Y. (1995). cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. *Genomics* **30**: 98-101.
- Savill, J., Fadok, V. (2000). Corpse clearance defines the meaning of cell death. *Nature* **407**: 784-788.
- Scheffner, M., Nuber, U., Huibregtse, J.M. (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**: 81-83.

- Sherr, C. J., Weber, J.D. (2000). The ARF/p53 pathway. *Curr. Opin. Genet. Dev.*, **10**: 94-99.
- 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.
- Thomas, D. A., Du, C., Xu, M., Wang, X., Ley, T. J. (2000). DFF45/ICAD can be directly processed by granzyme B during the induction of apoptosis. *Immunity* **12**: 621-632.
- Thompson, J. D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25** : 4876-4882.
- Thorne J. L., Kishino H. (1992). Freeing phylogenies from artifacts of alignment. *Mol. Biol. Evol.* **9**:1148-62.
- Townsend, A., Bodmer. H. (1989). Antigen recognition by class I-restricted T lymphocytes. *Annu. Rev. Immunol.* **7**: 601-624.
- Tyers, M, Jorgensen, P. (2000). Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr. Opin. Genet. Dev.* **10**: 54-64.
- van Baar, B.L.M. (2000). Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. *FEMS Microbiol. Rev.* **24**: 193-219.

- van Endert, P.M. (1999). Genes regulating MHC class I processing of antigen. *Curr. Opin. Immunol.* **11**; 82-88.
- van Parijs, L., Abbas, A.K. (1996). Role of Fas-mediated cell death in the regulation of immune responses. *Curr. Opin. Immunol.* **8**: 355-361.
- Vivarès, C. P., Méténier, G. (2000). Towards the minimal eukaryotic parasitic genome. *Curr. Opin. Microbiol.* **3**: 463-467.
- Vogelstein, B, Lane, D, Levine, A.J. (2000). Surfing the p53 network. *Nature* **408**: 307-310.
- Walker, P.A., Leong, L.E., Ng, P.W., Tan, S.H., Waller, S., Murphy, D., Porter, A.G. (1994). Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology* **12**: 601-605.
- Waters, J.B., Oldstone, M.B.A., Hahn, K.M. (1996). Changes in the cytoplasmic structure of CTLs during target cell recognition and killing. *J. Immunol.* **157**: 3396-3403.
- Weinberg, R.A. (1992). The retinoblastoma gene and gene product. *Cancer Surv.* **12**: 43-57.
- Wüthrich, K. (1989). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* **243**: 45-50.
- Yang, X., Stennicke, H.R., Wang, B., Green, D.R., Janicke, R.U., Srinivasan, A., Seth, P., Salvesen, G.S., Froelich, C.J. (1998). Granzyme B mimics apical caspases. Description of a unified pathway for trans-activation of executioner caspase-3 and -7. *J Biol. Chem.* **273**; 34278-34283.
- Yeh E.T.H., Gong L., Kamitani. (2000). Ubiquitin-like proteins: new wines in

new bottles. *Gene* **248**: 1-14.

Zhang, Y., Xiong, Y., Yarbrough, W.G. (1998). ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumor suppression pathways. *Cell* **92**: 725-734.



APPENDIX A

Full Length cDNA sequence of Human DWNN, with translation of open reading frame

```
CGGGGGTCTCTGGATTATTGTTCTGACGAACCCCTGCTTGTGGTTGGGGGGTATTTAAT
1 -----+-----+-----+-----+-----+-----+-----+ 60

CTGAGGCCTTAGGGTCCTTCGGTGTCTTTGAGTGTTTTGTGTGTACATATTTTGCTCTTA
61 -----+-----+-----+-----+-----+-----+-----+ 120

AAGTTTATAAATATACGTATATTGAGAGTGTCCACGTCTCCTCGCTGAACCTTAGGAATC
121 -----+-----+-----+-----+-----+-----+-----+ 180

      M S C V H Y K F S S K L N Y D T V
CCTTGGCACCATGTCCTGTGTGCATTATAAAATTTTCTCTAAACTCAACTATGATACCGT
181 -----+-----+-----+-----+-----+-----+-----+ 240

      T F D G L H I S L C D L K K Q I M G R E
CACCTTTGATGGGCTCCACATCTCCCTCTGCGACTTAAAGAAGCAGATTATGGGGAGAGA
241 -----+-----+-----+-----+-----+-----+-----+ 300

      K L K A A D C D L Q I T N A Q T K E E Y
GAAGCTGAAAGCTGCCGACTGCGACCTGCAGATCACCAATGCGCAGACGAAAGAAGAATA
301 -----+-----+-----+-----+-----+-----+-----+ 360

      T D D N A L I P K N S S V I V R R I P I
TACTGATGATAATGCTCTGATTCCCTAAGAATTCTTCTGTAATTGTTAGAAGAATTCCTAT
361 -----+-----+-----+-----+-----+-----+-----+ 420

      G G V K S T S K T Y V I S R T E P A M A
TGGAGGTGTTAAATCTACAAGCAAGACATATGTTATAAGTCGAACTGAACCAGCGATGGC
421 -----+-----+-----+-----+-----+-----+-----+ 480

      T T K A V C K N T I S H F F Y T L L L P
AACTACAAAAGCAGTATGTAAAAACACAATCTCACACTTTTTCTACACATTGCTTTTACC
481 -----+-----+-----+-----+-----+-----+-----+ 540

L *
TTTATAATGTAGCAGTGAAGTAAATCATTTTAGAACTTAATATCCAACCTGATCATAGTAC
```

541 -----+-----+-----+-----+-----+-----+ 600

 ATATTGTAAATAAAATGTATTTTGGATGACAGCTCAGTTGAATATGGATAATATGTGGCAT
 601 -----+-----+-----+-----+-----+-----+ 660

 CACTTGCACACTTATTTTGTAGAAATGGGTAATTTGTGCCCGTAACACTGTTTCATATTA
 661 -----+-----+-----+-----+-----+-----+ 720

 AATATGATAGCATTATCCCTGTATGACACTGTGTTGTACAGTTAATGTATGATCCTTTTT
 721 -----+-----+-----+-----+-----+-----+ 780

 AGATCGTTTAGGTTTTACACTAAGGAACATGATGACATGTTCTACATTTGTCTGTCTATA
 781 -----+-----+-----+-----+-----+-----+ 840

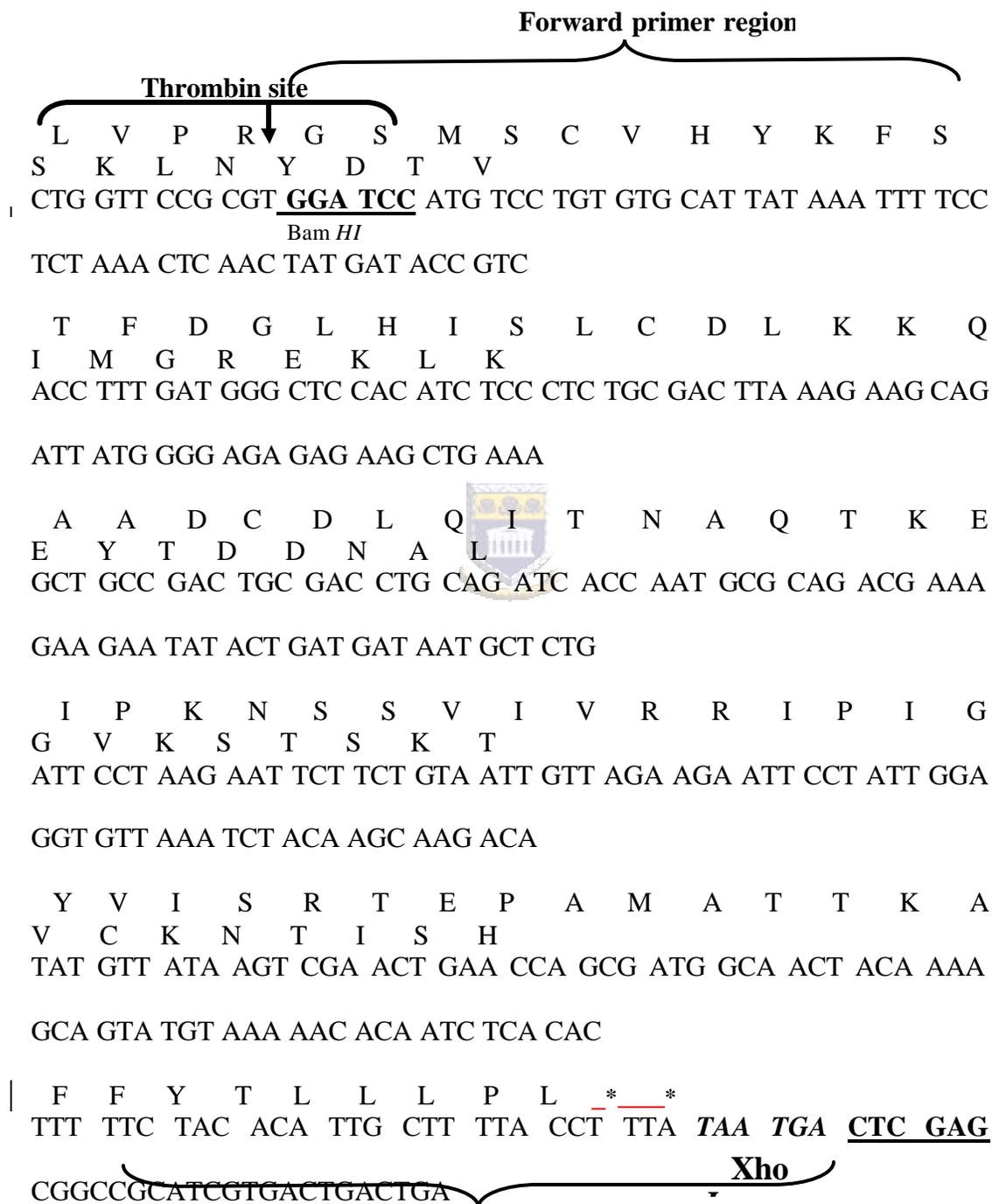
 GTTAGTATTTTGTATGTATGTACAGGCTGTTGTGTGCTTTTTGTTTCTTGAATAAAAAA
 841 -----+-----+-----+-----+-----+-----+ 900

 TGTTTGGAGTGTAATAAAAAAAAAAAAAAAAAA
 901 -----+-----+-----+-----+-----+-----+ 929



Appendix B

Expected pGEX (4T-3) – DWNN (long-form) construct (Multiple Cloning Site)



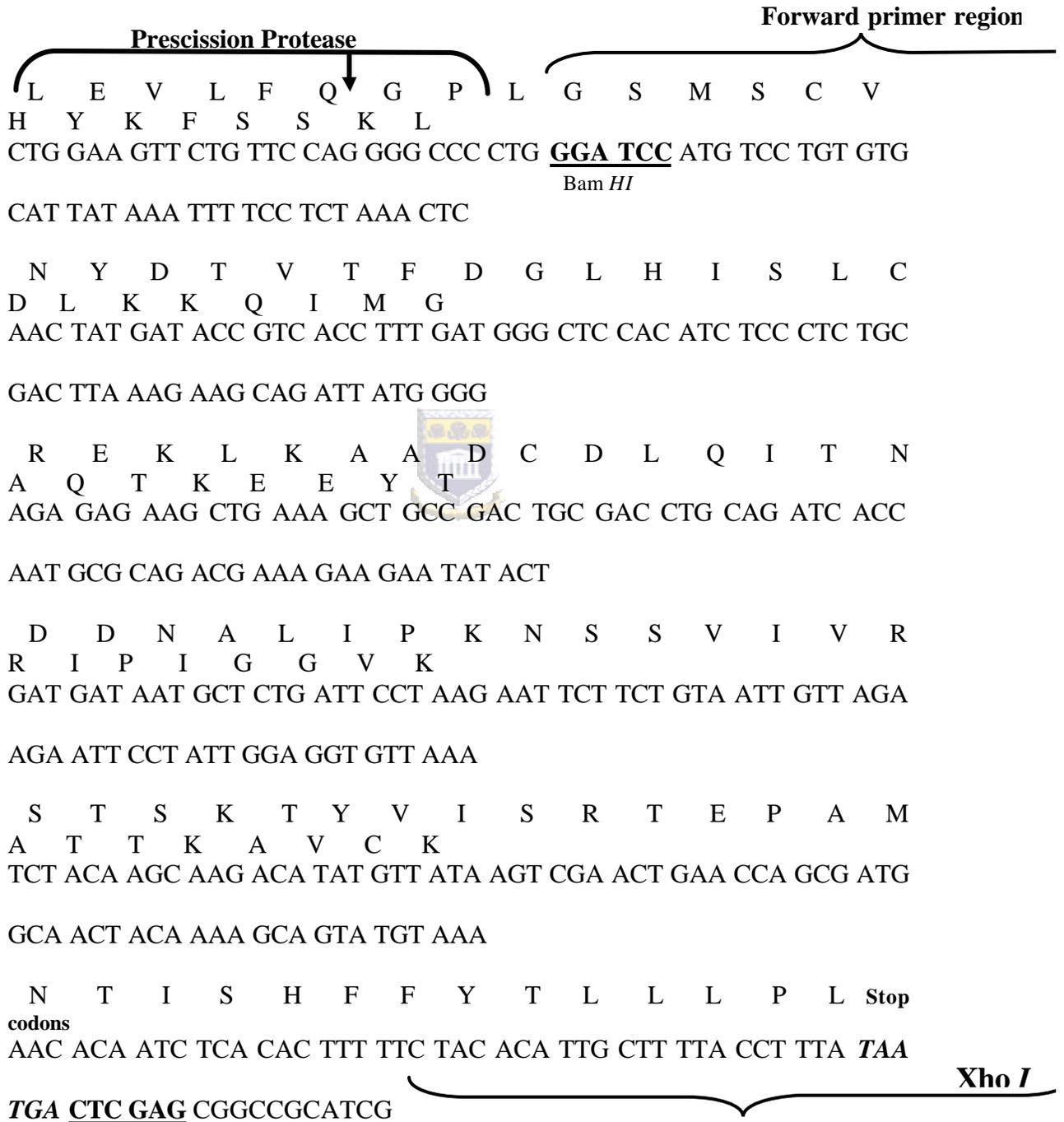
*** Stop Codon**

Reverse primer region



APPENDIX C

Expected pGEX 6P-2 – DWNN (long -form) construct (Multiple Cloning Site)

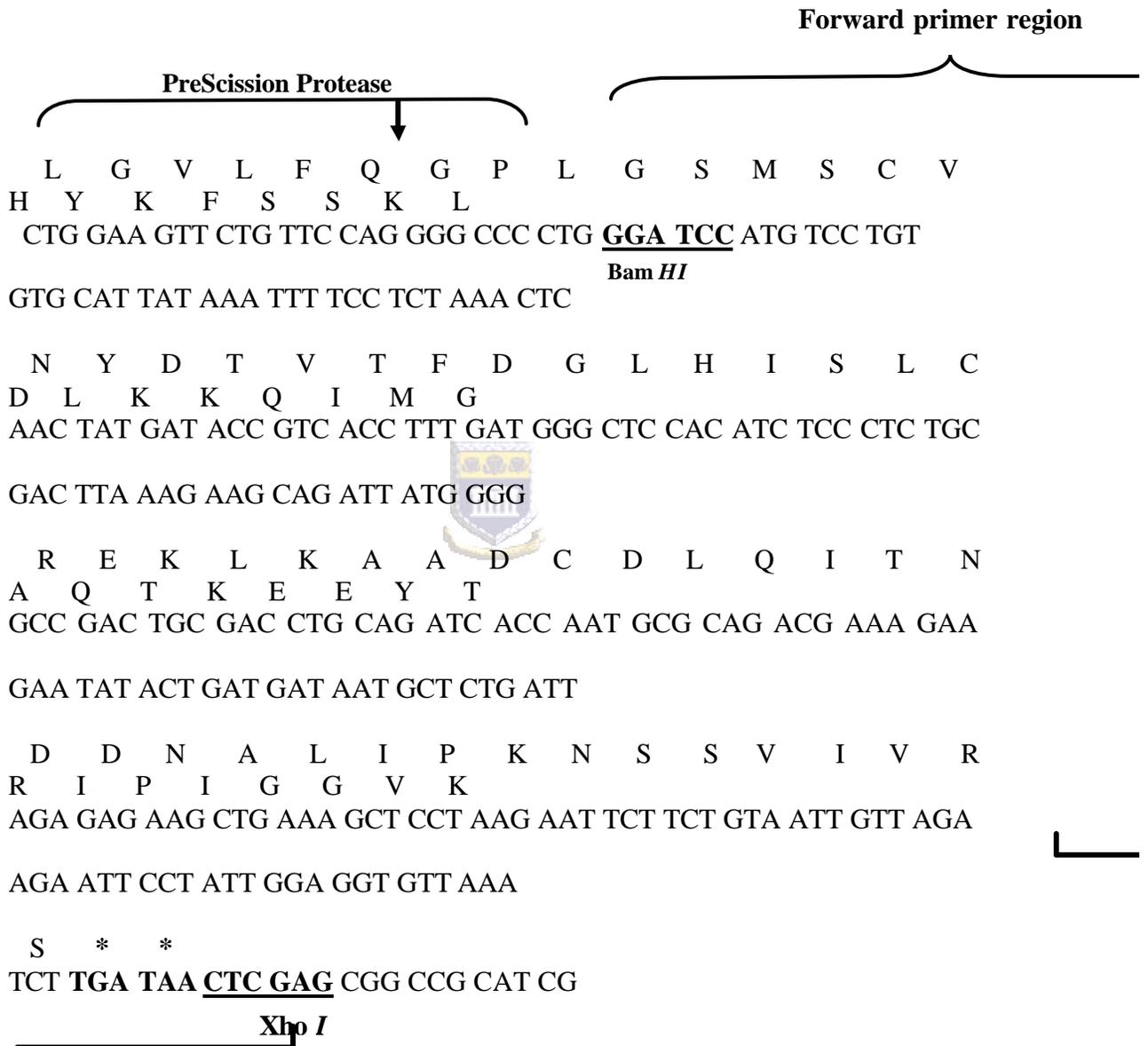


Reverse primer region



APPENDIX D

Expected pGEX 6P-2 - DWNN (short-form) construct (Multiple Cloning Site)



* Stop Codon

APPENDIX E

pGEX 5' Sequencing Primer:

5'-

d[GGGCTGGCAAGCCACGTTTGGT

G]-3'



pGEX 5' Sequencing Primer Binding Site is situated at position 869-891 of all pGEX vectors.