Protein Interactions with *Drosophila* p53

Umar-Faruq Cajee

Supervisor: Dr. Monde Ntwasa

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, in fulfillment of the requirements for the degree Master of Science.

July, 2014

School of Molecular & Cell Biology

University of the Witwatersrand

Johannesburg

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

Umar-Faruq Cajee

Johannesburg, 1 July 2014

Abstract

Drosophila melanogaster, a key model organism, has cognates of over 70% of human disease genes. This has created opportunities in the development of treatments for life threatening illnesses like cancer. Mutations on the p53 tumour suppressor protein, which is an activator of apoptosis, are common in many cancers. In mammals, p53 interacts with the Retinoblastoma Binding Protein 6 (RBBP6) which enhances the activity of MDM2, the prototypical negative regulator of p53, that is absent in invertebrates. In the absence of MDM2 the *Drosophila* RBBP6 homolog, SNAMA, through its DWNN Catalytic Module (DCM), is suspected to play an important role in the regulation of p53, probably via the ubiquitin proteasome pathway. Through bioinformatics analyses, and experimental analysis of transcripts, this study has shown the existence of two isoforms of SNAMA named here SNAMA A and SNAMA B for the long and short isoforms, respectively. SNAMA B appears to be expressed after genotoxic stress (DNA damage) in adults as well as during embryonic development. Recombinant protein expression in bacterial and yeast systems as well as HIS-tag chromatography and Western blot analyses were used to investigate interactions with Dmp53. Due to poor expression of recombinant Dmp53 protein in both prokaryotic and eukaryotic systems and unreliable commercial antibodies, it was impossible to complete interaction studies. Overall, these studies show that the SNAMA isoforms may play important roles during development and in response to DNA damage.

Quotation

“Alice: You're quite right, Mr. Hatter. I do live in a topsy-turvy world. It seems like I have to do something wrong first, in order to learn from what not to do. And then, by not doing what I'm not supposed to do, perhaps I'll be right. But I'd rather be right the first time, wouldn't you?”

Lewis Caroll –Alice in Wonderland

Acknowledgements

Bismillahir Rahmanir Rahim
(In the name of Allah, most gracious most merciful)

My sincere gratitude goes out to my supervisor Dr. Monde Ntwasa, for the opportunity to perform this research and all that goes with it.

A special thank you to the FlyLAB members Pfariso, Ntando for giving me hope. To the maybe-FlyLAB member Hlogzas for deserting us in the Drakensberg and hopefully soon to be ex-FlyLAB member Brent, for well, being Brent. To Zameer, for appearing out of nowhere. And lastly Rodney, and Shüné, for the friendship and especially Dr. Poopiepants, for the help, recipes, reagents, lab-work, putting up with me, advice, medication, etc. etc.

I thank the "curse of the FlyLAB" for granting me a few years off my sentence, presumably for good behaviour.

To MCB staff and colleagues, a thank you for all the conversations in the corridors, laughs, support and advice.

I acknowledge the National Research Foundation for a percentage of funding towards this degree, and thank them for the Xbox, TV, and Car.

Dear RaeesaRocks, THANK YOU! I LOVE YOU! P.S. Still another thesis to come, Insha Allah.

Dear Family, I would not be where I am today without you all, thank you for all the prayers and a lifetime of support and encouragement.

Research Output

Publications:

Umar-Faruq Cajee, Rodney Hull and Monde Ntwasa. (2012). Modification by Ubiquitin-Like Proteins: Significance in Apoptosis and Autophagy Pathways. *Int. J. Mol. Sci.* 13, no. 9: 11804-11831.

Conference Proceedings:

Umar-Faruq Cajee, Brent Oosthuysen, Lehlogonolo Mokgohloa and Monde Ntwasa. (2012). The characterization of *Drosophila melanogaster* RBBP6 protein, SNAMA. Poster presented at: SASBMB/FASBMB Congress 2012, 29 January – 1 February, Champagne Sports Resort, Drakensberg, Kwazulu Natal, South Africa.

Pfariso Maumela, Ntando Mnisi, Umar-Faruq Cajee, Rodney Hull and Monde Ntwasa. (2012). Identification and characterisation of immune related molecules in *Euoniticellus intermedius*. Poster presented at: EMBO Global Exchange Lecture Course Innate Immunity: Evolution and advances in clinical medicine, 2 – 5 September 2012, Stay City, Johannesburg, South Africa.

Table of Contents

Declaration i

Abstract ii

Quotation iii

Acknowledgements iv

Research Output v

Table of Contents vi

List of Figures xi

List of Tables xiii

List of Abbreviations xiv

1. Introduction 1

1.1 *Drosophila* as a model organism 2

1.2 History and Role of p53 and Dmp53 3

1.3 Apoptosis 6

1.4 Apoptosis in Humans 7

1.5 Apoptosis in *Drosophila* 8

1.6 Ubiquitin and Ubiquitination 11

1.7 RBBP6 and SNAMA 13

1.8 Overview, Aim and Objectives 14

2. Materials and methods 16

2.1 Materials 16

2.1.1 Flies 16

2.1.2 Bacterial Cells 16

2.1.3 Bacterial plasmids 17

2.1.4 Yeast 17

2.1.5 Media 17

2.1.6 Selection/antibiotic resistance 19

2.1.7 Buffers and other solutions 20

2.1.8 Primers for PCR 25

2.1.9 Antibodies 26

2.1.10 Restriction Enzymes 27

2.1.11 Molecular weight Markers 28

2.1.12 General reagents and kits 28

2.1.13 Initial clone 29

2.2 Methods 30

2.2.1 Bioinformatics 30

2.2.2 Agarose gel electrophoresis 30

2.2.3 Gel purification of DNA 30

2.2.4 Plasmid DNA purification by ethanol acetate method 31

2.2.5 Restriction digestion of DNA 31

2.2.6 Preparation of plasmid DNA for ligation (dephosphorylation) 32

2.2.7 Ligation of insert into linearised plasmid 32

2.2.8 Transformation of chemically competent cells 32

2.2.9 Transfection of yeast 34

2.2.10 Protein expression 34

2.2.11 Protein extraction 35

2.2.12 Yeast protein expression 36

2.2.13 Yeast protein extraction from pelleted cells 36

2.2.14 HIS nickel affinity chromatography 36

2.2.15 Protein concentration 37

2.2.16 SDS PAGE 38

2.2.17 Western blotting 38

2.2.18 Extraction of genomic DNA 39

2.2.19 Polymerase chain reaction 40

2.2.20 Southern blotting 40

2.2.21 Camptothecin treatment of flies 41

2.2.22 Extraction of RNA 41

2.2.23 cDNA synthesis RT PCR SNAMA Oligo(dT) & specific primers 41

3. Results 42

3.1 Gene Structure and isoforms of Dmp53 42

3.2 Heterologous expression of Dmp53 in *E. coli* 47

3.3 Protein expression and purification in Yeast 50

3.4 Gene structure and isoforms of SNAMA 61

3.5 Differential expression of SNAMA during development and in response to DNA damage 75

4. Discussion 78

4.1 SNAMA B, a shorter SNAMA isoform comprises an independent DWNN Catalytic module and is activated by DNA damage. 79

4.2 SNAMA A and B are probably differentially expressed during embryonic development and induced by DNA damage 81

4.3 Dmp53 interactors 82

5. Conclusion 86

5.1 Future work 86

References 87

Appendix A, Dmp53 in pET41a sequence 92

Appendix B, pGEM-T Easy 94

Appendix C, pET41a 95

Appendix D, pTG1895 96

Appendix E, Generuler 1kb plus 97

Appendix F, PageRuler Protein ladder 98

List of Figures

Figure 1: Schema of *Drosophila* p53 isoforms. 4

Figure 2: Superimposed crystal structures of human and *Drosophila* p53. 5

Figure 3: A diagrammatic overview of the Apoptotic mechanism in humans. 9

Figure 4: A comparison between the apoptotic mechanisms of *Drosophila*, and humans 10

Figure 5: The ubiquitin pathway of RING class E3 ligases which interact directly with E2 and the fates of proteins ubiquitinated by the pathway, ... 12

Figure 6: Schematic overview of *Drosophila* p53 gene structure 43

Figure 7: Agarose gel of biotin probe and Southern blot thereof. 45

Figure 8: Digestion of genomic DNA and Southern blot thereof 46

Figure 9: Western blot analysis of Dmp53 in soluble and insoluble fractions of expressed protein. 48

Figure 10: Purified soluble protein fraction from heterologous expression in *E. coli*, 6 hours after induction. 49

Figure 11: Confirmation of plasmid preparation by restriction digest 51

Figure 12: Dmp53 insert was successfully amplified by polymerase chain reaction, ... 52

Figure 13: Colony PCR analysis of transformed clones to detect clones with Dmp53 insert 54

Figure 14: Crude extracts from Protein expressions of transformed and untransformed yeast concentrated by stirred cell, and Western blot thereof. 55

Figure 15: SDS PAGE to visualise Dmp53. 57

Figure 16: Yeast expression with the inclusion of protease inhibitors 59

Figure 17: Non-specific binding of Dmp53 antibodies during western blot analysis 60

Figure 18: A schematic overview of the gene structure of the SNAMA A and B isoforms. 61

Figure 19: SNAMA B sequence with translation. 65

Figure 20: SNAMA A sequence with translation. 69

Figure 21: Aligned protein sequences of SNAMA A and SNAMA B. 73

Figure 22: A schematic overview of the possible domain structure of SNAMA isoforms. 74

Figure 23: SNAMA A was detected after Reverse Transcriptase PCR. .. 76

Figure 24: SNAMA B was detected after Reverse Transcriptase PCR. .. 77

List of Tables

Table 1: Bacterial Cell Strains used 16

Table 2: Bacterial Plasmid DNA used 17

Table 3: Composition of media for flies, bacteria and yeast maintenance 17

Table 4: Antibiotic resistance and selection of bacterial and yeast, cells and plasmids. 19

Table 5: Composition of Solutions and buffers used in the study 20

Table 6: Primer sequences used in Polymerase Chain reaction 25

Table 7: Primary Antibodies used in Western blot analysis 26

Table 8: Secondary Antibodies used in Western blot analysis 26

Table 9: Restriction enzymes used for DNA digestion 27

Table 10: Molecular Weight markers used in DNA and Protein Electrophoresis 28

Table 11: General reagents and kits used 28

Table 12: A Comparison of the *SNAMA* A and B gene structure. 62

List of Abbreviations

&	:	and
µg	:	micrograms
µg/ml	:	micrograms per millilitre
µJ	:	micro Joules
µl	:	microlitres
3D	:	Three Dimensional
aa	:	amino acids
ATP	:	Adenosine Triphosphate
auto	:	automatic
bp	:	base pairs
C. elegans	:	*Caenorhabditis elegans*
cDNA	:	complementary DNA
Cmpt.	:	Camptothecin
Cys	:	Cysteine
DCM	:	DWNN Catalytic Module
Dmp53	:	*Drosophila melanogaster* p53
DNA	:	Deoxyribonucleic acid
DWNN	:	Domain With No Name
E. coli	:	*Escherichia coli*
EtBr	:	Ethidium Bromide
g	:	relative centrifugal force
Gly	:	Glycine
H_2O	:	Dihydrogen monoxide, (water)
HIS	:	Histidine
hr	:	hour/hours
IgG	:	Immunoglobulin G
IPTG	:	Isopropyl β-D-1-thiogalactopyranoside

kb	:	kilobases / kilobase pairs
kDa	:	Kilo Daltons
KoAC	:	Potassium Acetate
LB	:	Luria-Bertani Broth
Lys	:	Lysine
MDM2	:	Mouse Double Minute 2
mins	:	minutes
ml	:	millilitres
mM	:	millimolar
mm	:	millimetres
mRNA	:	messenger RNA
MW	:	Molecular Weight
MWM	:	Molecular Weight Marker
NCBI	:	National Centre for Biotechnology Information
nm	:	nanometres
no.	:	number
^{o}C	:	degrees Celsius
OD	:	Optical Density
ORF	:	Open reading frame
PBS	:	Phosphate Buffered Saline
PCD	:	Programmed Cell Death
PCR	:	Polymerase Chain Reaction
PVDF	:	Polyvinylidene fluoride
Rb	:	Retinoblastoma
RBBP6	:	Retinoblastoma binding protein 6
RING	:	Really Interesting New Gene
RNA	:	Ribonucleic Acid
S. cerevisiae	:	*Saccharomyces cerevisiae*
SD	:	Synthetic minimal medium

SDS	:	Sodium dodecyl sulphate
SDS PAGE	:	Sodium dodecyl sulphate Polyacrylamide gel electrophoresis
SNAMA	:	Something that sticks like glue
SSC	:	Saline Sodium Citrate
STC	:	Signal Transduction Cascade
TCA	:	Trichloroacetic acid
TM	:	melting temperature
TNFR	:	tumour necrosis factor receptor
trans	:	transformed
Ub	:	Ubiquitin
UBL	:	Ubiquitin like
untrans	:	untransformed
UTR	:	untranslated region
YPG	:	Yeast extract peptone glucose

1. Introduction

Cancer is a disease that refers to an accumulation of cells or a certain degree of uncontrolled growth of cells, which have transformed from normal to an abnormal or diseased state. This collection of cells or tumour may remain localised or spread to other parts of the body in a process known as metastasis (Hanahan and Weinberg, 2000; Vidal and Cagan, 2006). To date cancer remains one of the leading causes of death worldwide. In a study conducted on cancer records from 2008, by (Ferlay et al., 2010), it was found that the most common cancers were lung, breast and colorectal cancers, while lung, stomach and liver cancers accounted for most deaths. Furthermore, over half of cancer cases and deaths occurred in less developed regions in Africa, Central and South America and Asia (Ferlay et al., 2010). Causes of cancer are varied, and include heredity and lifestyle stresses.

Over a hundred different human cancer types exist (Selivanova, 2004). These complex and diverse diseases are a result of what appears to be an endless number of combinations of genetic and epigenetic changes (Selivanova, 2004; Vidal and Cagan, 2006). It would seem that these diverse diseases would therefore require different types of therapies. However, recent advances in understanding the molecular biology of tumours has enabled the design of therapies directed against a variety of tumours that target a finite number of specific pathways, amongst these the p53 pathway in particular (Selivanova, 2004). Amongst genes involved in tumour development, the tumour suppressor protein p53 stands out. The inactivation of p53 function, occurs in virtually all human tumour varieties. Mutations in the *p53* gene, causing inactivation of p53 protein (an important DNA damage detector which

activates apoptosis) is common in more than half of human cancers (Farnebo et al., 2010; Hanahan and Weinberg, 2000). p53 is also found in a number of lower eukaryotes, including *Drosophila* in which cancer is not naturally prevalent (Dichtel-Danjoy et al., 2012)

1.1 *Drosophila* as a model organism

Model organisms have become a key in the study of human disease, where treatments and causes can be researched feasibly and ethically. Organisms, such as the common fruit fly – *Drosophila melanogaster*, are extensively studied due to advantages they provide (Daines et al., 2010). Conservation of genetic material as well as pathways, both metabolic and developmental, allow for the elucidation of biological workings in other organisms. *Drosophila* has been adopted as a key model organism because of its complete genomic sequence being available, as well as the similarity that exists between it and other systems. Ease of breeding, small size, low cost of maintenance and the ability to manipulate the fly are other factors making *Drosophila* more amenable for experimentation. In 2001, Reiter et al., found that 77 percent of human disease genes have cognates in *Drosophila*. This includes the tumour suppressor p53. *Drosophila melanogaster* p53 (Dmp53) shares similar functions with multiple mammalian p53 family members. This makes Dmp53 an attractive model system to study apoptotic pathways involving p53 activation (Fan et al., 2010; Herzog et al., 2012).

1.2 History and Role of p53 and Dmp53

The p53 protein, discovered in 1979, is a well-known tumour suppressor gene product (Farnebo et al., 2010; Machado-Silva et al., 2010). To date p53 studies have been somewhat limited to mammalian systems and thus more is known about the human homologue than that of the fly. On activation, p53 induces apoptosis, the repair of DNA, or cell cycle arrest, by different mechanisms (Schuler and Green, 2005). The inactivation of the *p53* gene serves as a marker for cancer as it is the most common event in human cancers (Bourdon, 2005; Farnebo et al., 2010). Furthermore, p53 is mutated in 50 percent of all cancers and mutations in the gene increase the risk of cancer development (Bourdon, 2005; Murray-Zmijewski et al., 2006).

There is significant similarity between human p53 and its paralogues (p63 and p73). These paralogues can sometimes mimic the action of p53 by binding p53 targets or inducing cell cycle arrest and apoptosis (Bourdon, 2005; Khoury and Bourdon, 2009). *Drosophila* p53, however, has no paralogues. Instead three theoretical isoforms identified by Bourdon and colleagues (2005) exist, as confirmed by a recent review of the Dmp53 gene structure, through its 2 internal promoters (Marcel et al., 2011). These isoforms are described as the Dmp53 long isoform of 495 amino acid sequence, the originally published Dmp53 of 385 amino acid length and lastly Dmp53n (110 amino acids). It was shown that the originally discovered Dmp53 was a truncated form of the long isoform generated by transcription initiated at an internal promoter

region (Bourdon, 2005). The three isoforms are diagrammatically represented in Figure 1.

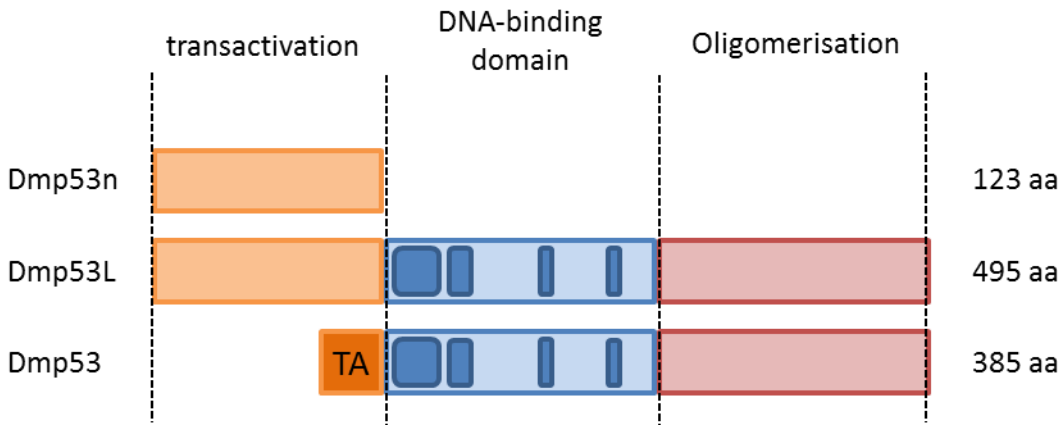


Figure 1: Schema of *Drosophila* p53 isoforms. The figure shows the 3 domains of p53 and the 3 isoforms formed by variation in these. Dmp53L is the long isoform, and Dmp53 was the originally published isoform. Dmp53n is the shortest isoform, but only contains the transactivation domain (adapted from Bourdon, 2005).

The schema identifies 3 major regions in Dmp53. The transactivation domain and oligomerisation domain are responsible for binding regulatory proteins and post translational modifications. These domains flank a central DNA binding core domain (Dichtel-Danjoy et al., 2012). Furthermore, on comparison to other p53 family members this core domain is conserved and shows a high degree of similarity to the DNA binding domains from p53 proteins from other species. The core domain functions to recognise and bind DNA as well as co-ordinate a zinc ion, through the presence of vital amino acid residues

(Brody, 2000). The similarity of the core region extends to the three dimensional structure of the protein, which suggests similarity in function. Aligned 3D structures of p53 and Dmp53 are shown in Figure 2.

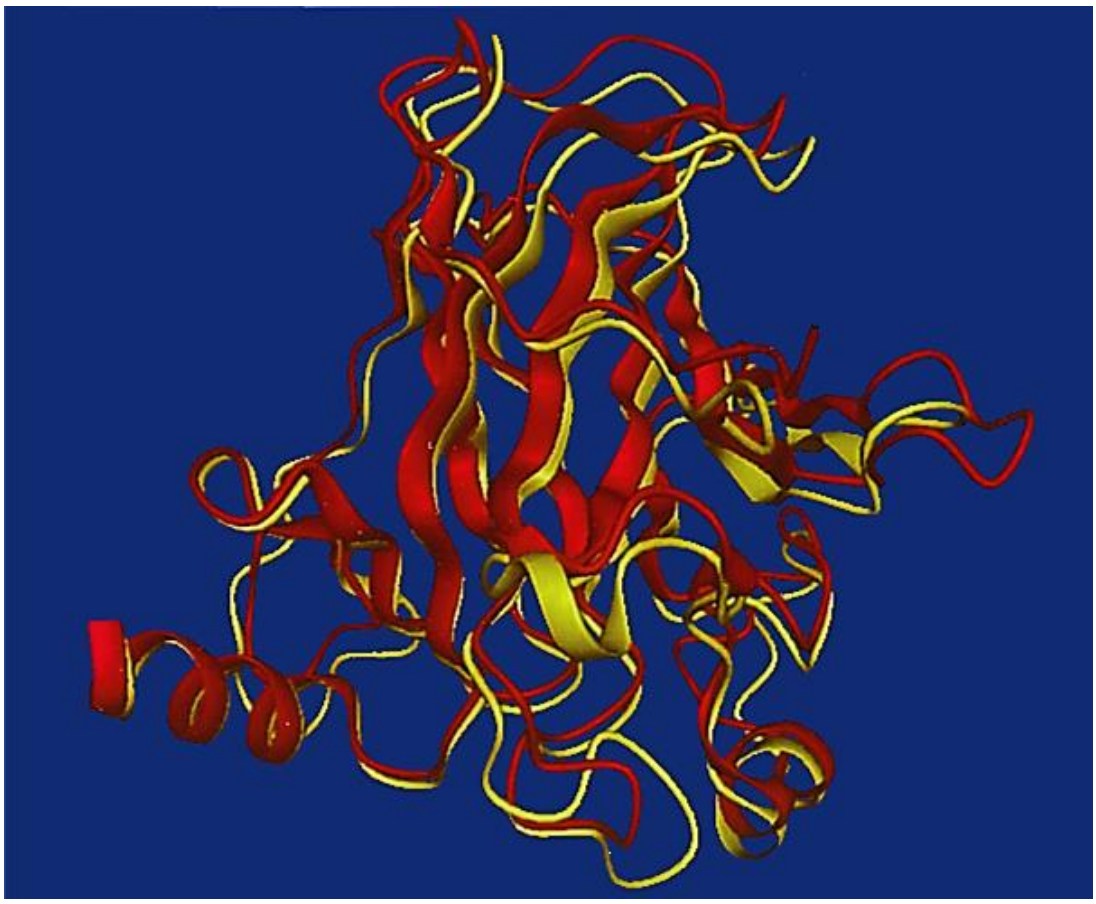


Figure 2: Superimposed crystal structures of human and *Drosophila* p53. Human p53 is shown in yellow and Dmp53 in red, the structures were predicted by Jin et al., (2000) using 'MODELLER'.

p53 is toxic in high quantities and may cause the death of the organism if not regulated. Kim et al., (2011), explain that p53 regulation involves

protein stabilisation, interactions, modification, and localization; furthermore, this stabilisation is important in its activation. As such, p53 in humans normally exists in a latent form and in low quantities (Brooks and Gu, 2006). This stringent regulation is largely a result of the ubiquitin proteasome pathway, and the E3 ligase Mouse Double Minute 2 (MDM2). While p53 inactivation in cancers is often through direct mutation of the p53 gene, its function is also sometimes affected through an increased level of MDM2, a downregulator of p53 (Michael and Oren, 2002). MDM2 binds the transactivation domain of human p53 and regulates p53 transport to the cytoplasm and its degradation by ubiquitination. This occurs in the absence of external factors (Brooks and Gu, 2006; Jin et al., 2000).

An MDM2 homologue is yet to be discovered in *Drosophila melanogaster*. As such, the regulation of p53 in invertebrates is not clearly understood, and its pathways are not completely known. In *Drosophila*, Dmp53 is required for apoptosis. Peculiarly, Dmp53 does not block the cell cycle at the G1 phase (Ollmann et al., 2000). Ollmann et al., (2000), explain that this reveals an ancestral proapoptotic function for Dmp53, induced by DNA damage. This makes *Drosophila* a model system to study and elucidate programmed cell death pathways.

1.3 Apoptosis

When the p53 tumour suppressor is activated, it induces apoptosis. Apoptosis, or programmed cell death (PCD), is a highly regulated biological process in multicellular organisms. The mechanism is

present throughout the organism's development from embryo to adult, cell homeostasis, and the body's defense mechanisms. During apoptosis, cells commit suicide in an orderly fashion when signaled to do so. Apoptosis generally occurs when cells are under stress; infected by viruses, damaged beyond repair, or have undergone DNA damage. In contrast, necrosis, or uncontrolled cell death, results in cell lysis, immune responses and other unfavourable conditions. An underlying characteristic of cancer is an impaired apoptotic mechanism; damaged cells, which have impaired or irregular function, continue dividing and tumours often develop. According to Hanahan and Weinberg (2000), the rate of tumour progression is dependent not only on cell proliferation rate but also that of cell attrition, which is predominantly as a result of PCD.

1.4 Apoptosis in Humans

Comparison of the apoptotic pathways of humans, *Drosophila* and the nematode *C. elegans* (another model organism) reveals that the fly provides a system with intermediate complexity. The fly system, as described by Richardson and Kumar (2002), has less redundancy thereby allowing simpler analysis of function than these systems. Two main activation routes divide the human apoptotic pathway into two component pathways; namely the intrinsic and extrinsic pathway. Although activated by different triggers, both pathways activate a common signal transduction cascade (STC) of cysteine proteases, otherwise known as caspases, which results in cell death.

The human extrinsic pathway is activated primarily by ligand-stimulated activation of receptors, such as FAS ligand binding to the FAS receptor, thus activating the STC. The intrinsic pathway is associated with a buildup of both death promoting and death inhibiting molecules on the mitochondrial outer membrane. Cell death ensues when the presence of death promoters, such as Bax, outweighs the presence of death inhibitors, such as Bcl-2. This causes the release of cytochrome C from the mitochondria, which in turn binds Apaf-1, activates the STC (Richardson and Kumar, 2002).

1.5 Apoptosis in *Drosophila*

The *Drosophila* apoptotic mechanism is remarkably similar to the mammalian mechanism. Many of the proteins involved in the human apoptotic pathway have homologues, which are present in that of the fly. It is therefore thought that the apoptotic mechanism in *Drosophila* has in essence been maintained through evolution and is a conserved form of the pathway. The human pathway's classical tumour necrosis factor receptor (TNFR) family of death receptors which signal apoptosis in the extrinsic mammalian pathway, are represented in the fly by Eiger and Wengen (Kauppila et al., 2003). Eiger and Wengen share a Fadd homologue, dFadd, and caspase 8 homologue, Dredd (Richardson and Kumar, 2002). In mammals, cytotoxic agents and stress activate the intrinsic pathway by the activation of p53. Dmp53, induces expression of pro-apoptotic, Reaper (Rpr) (Brodsky et al., 2000). The PCD inducers in *Drosophila* are Rpr, Grim and Hid, with Smac/Diablo being possible functional homologues in mammals. These inducers, may act upon DARK (APAF-1 in mammals) directly, or cause the release of

cytochrome C from the mitochondrion, which acts upon DARK via Debcl (Bcl2 in mammals). DARK initiates the *Drosophila* STC via the initiator caspases. An overview of the apoptotic pathway in humans is shown in Figure 3. Figure 4 compares the apoptotic mechanisms of humans and *Drosophila.*

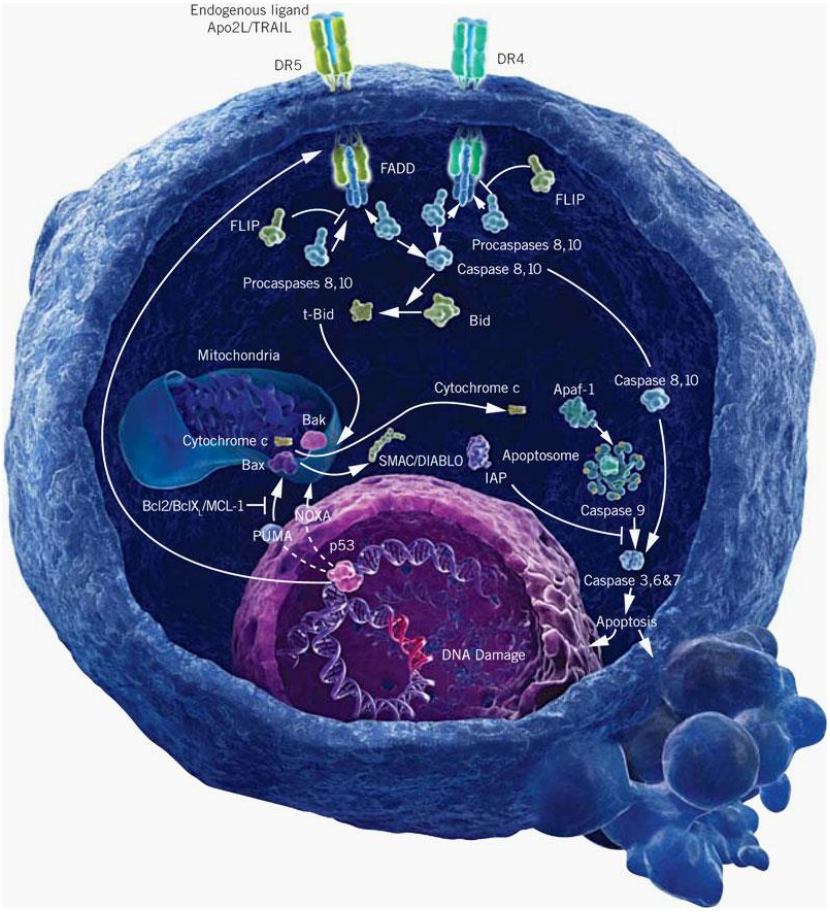


Figure 3: A diagrammatic overview of the Apoptotic mechanism in humans. Initiation of apoptosis may be via an extrinsic pathway, where an external ligand binds its receptor and activates a signal transduction cascade of caspases, or via an intrinsic pathway where cytotoxic molecules or stresses cause cytochrome c release from the mitochondria, which activates the same cascade through Apaf-1 (Genentech, 2014).

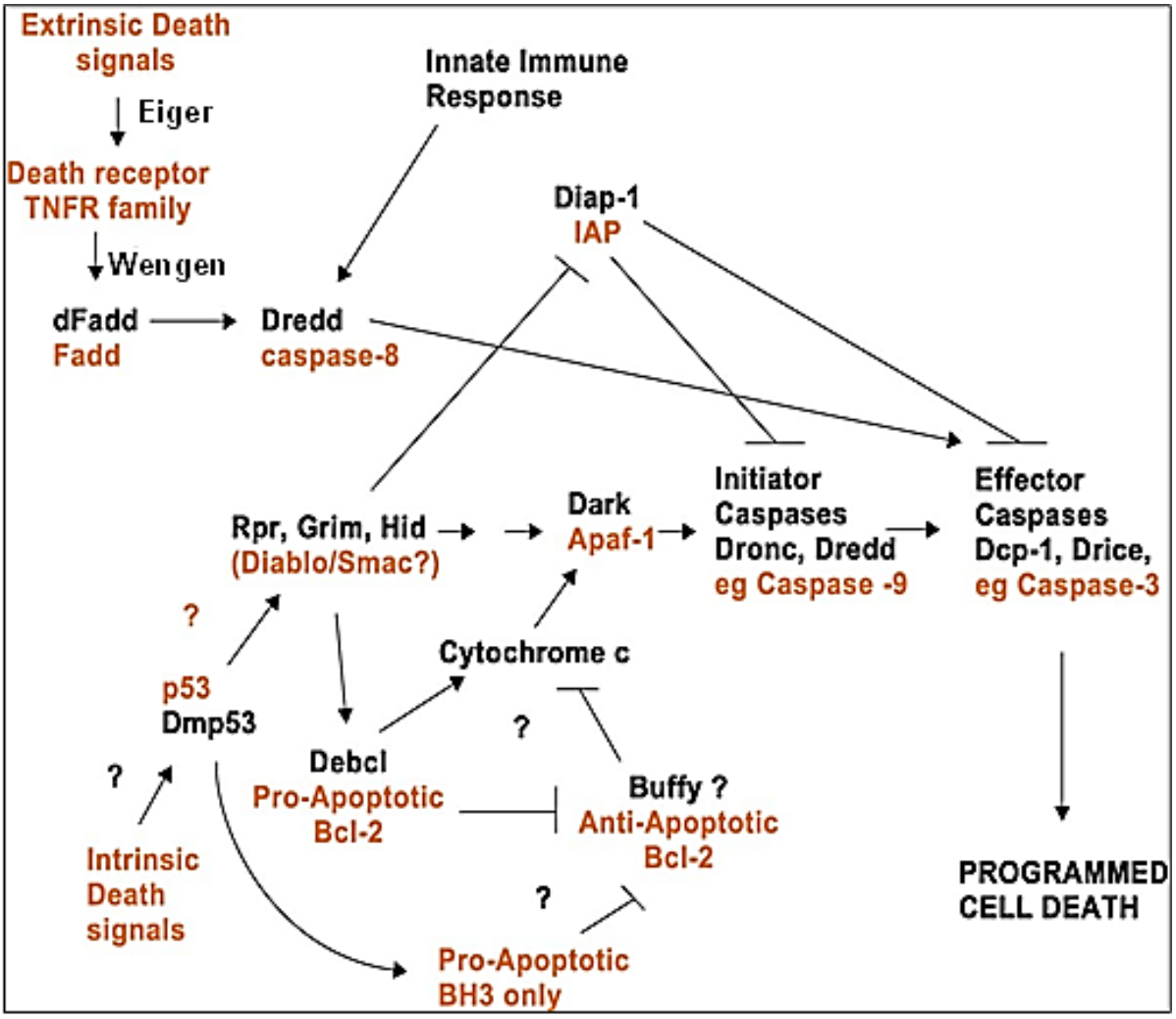


Figure 4: A comparison between the apoptotic mechanisms of *Drosophila*, and humans. The *Drosophila* mechanism is shown in black and that of humans is shown in red. Question marks are shown where uncharacterised steps occur (adapted from Richardson and Kumar, 2002).

1.6 Ubiquitin and Ubiquitination

The ubiquitin-proteasome pathway plays a major role in p53 regulation through MDM2. The ubiquitin pathway as described previously by Hershko and Ciechanover (1998), begins with the activation of ubiquitin, by the E1 enzyme Ubiquitin Activating Enzyme. The reaction requires ATP and conjugates ubiquitin at its Gly-76 to a Cys residue on the enzyme. The second step in the pathway involves the transfer of ubiquitin to an E2 enzyme, Ubiquitin Conjugating Enzyme. The ubiquitin once again is specifically transferred to a Cys sulphydryl group on the enzyme. The final step in the pathway is the transfer of the activated ubiquitin from the E2 enzyme to a Lys ε amino group on the target protein. This is carried out by an E3 enzyme, ubiquitin-protein ligase. Each E3 is specific to a specific set of proteins. The poly-ubiquitinated protein, depending on the length of the ubiquitin side-chain, undergoes a specific fate. Generally, targeted proteins with 4 or more linked ubiquitin molecules are degraded by hydrolysis in the Proteasome. Figure 5 gives an overview of the ubiquitin proteasome pathway and the fates of target protein depending on where and how many ubiquitin molecules are linked.

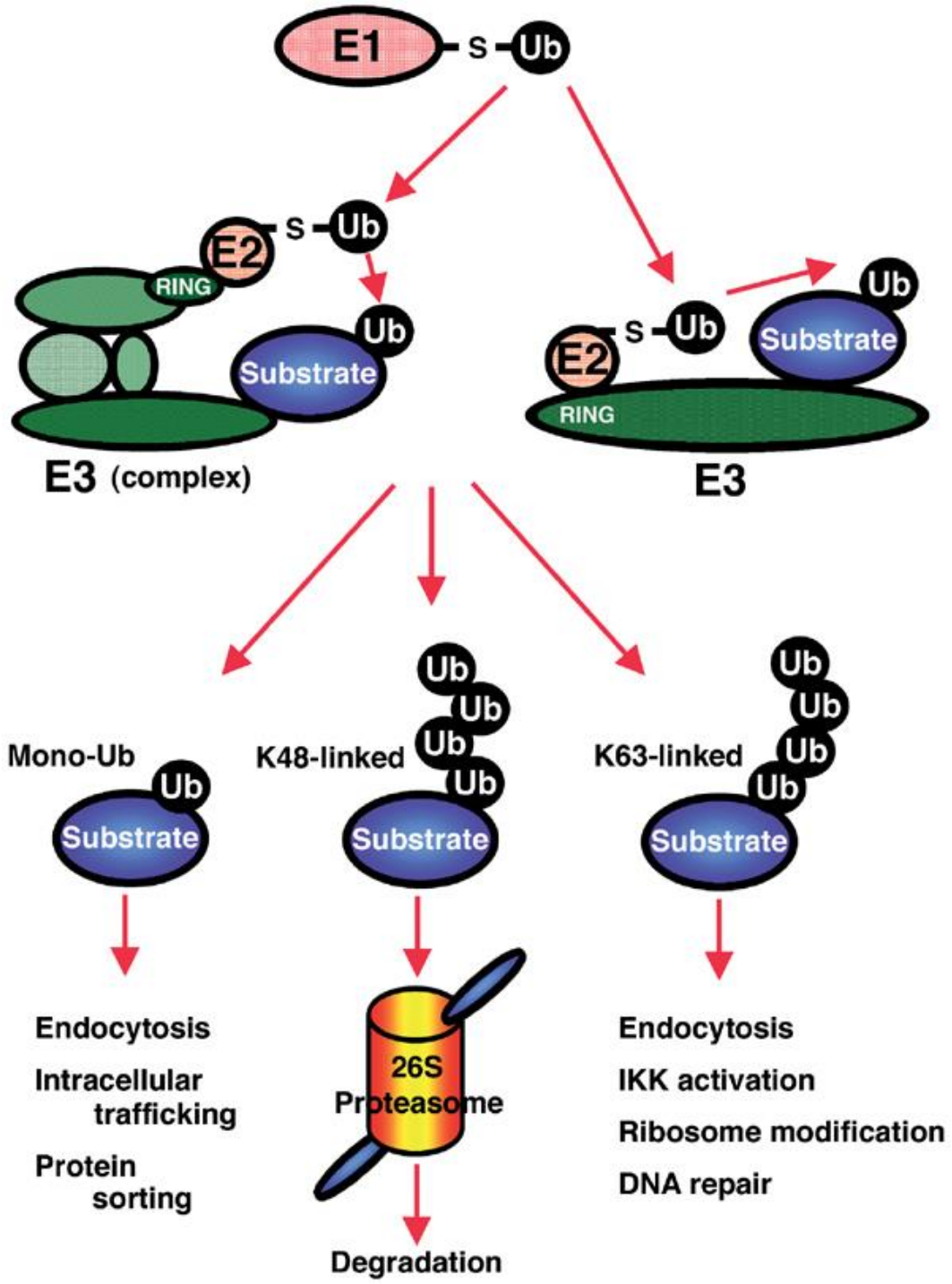


Figure 5: The ubiquitin pathway of RING class E3 ligases which interact directly with E2 and the fates of proteins ubiquitinated by the pathway, as described by Ohta and Fukuda (2004).

The E3 ligases are the most important in the pathway. They bind a target protein specifically and are divided into 3 classes. Of the three classes, the largest class is that of the RING (Really Interesting New Gene) family. MDM2, a negative regulator of p53 mentioned previously, is perhaps the most well-known E3 enzyme in the p53 apoptotic pathway. MDM2 is not present in invertebrates but other commonalities between systems have been identified, for example RBBP6.

1.7 RBBP6 and SNAMA

Retinoblastoma binding protein 6 (RBBP6) is a homo sapiens E3 ligase, characterised by the presence of a RING finger domain (Mather et al., 2005; Pugh et al., 2006). This nuclear protein is also one of the few known regulators of p53. The RBBP6 human homolog has Retinoblastoma protein (Rb) and p53 protein binding sites at its C terminal end, implicating it in regulating the cell cycle and apoptosis (Mbita et al., 2011; Simons et al., 1997). It has been identified as a down-regulator of p53, which functions by blocking the binding of p53 to DNA (Pugh et al., 2006). RBBP6 homologs all share a N terminal region, called the DWNN (Domain With No Name) catalytic module (Antunes, 2008). This module contains a DWNN domain, a Zinc knuckle and RING finger. DWNN has approximately 28% identity with ubiquitin, but has an almost exactly matching 3 dimensional structure to it (Cajee et al., 2012). The homologue of this protein in *Drosophila* is the recently characterised SNAMA (Mather et al., 2005). SNAMA, like RBBP6, is an E3 ligase and catalyses the addition of ubiquitin side chains to a substrate due to the presence of the DCM (Antunes, 2008). Research to date has shown that only SNAMA's homologs in mice (P2P-R) and humans (RBBP6) are known to bind p53 and Retinoblastoma

protein (Rb) (Simons et al., 1997). SNAMA has recently been shown to exhibit E3 ligase activity (Antunes, 2008) and contains a long uncharacterised C terminal extension (Mather et al., 2005) which may have a p53 binding site. We therefore suspect that it plays a role in transcriptional regulation and cell cycle control as suggested by Mather *et al*., in a 2005 study.

1.8 Overview, Aim and Objectives

Little is known about the regulation of p53 in invertebrates. When activated, p53 induces apoptosis, the repair of DNA, or cell cycle arrest, by different mechanisms. The inactivation the p53 gene, is the most common event in human cancer and these mutations increase the risk of cancer development. p53 has a high degree of toxicity in high quantities, and must therefore be regulated in order to prevent the death of the organism.

Recent studies have found that *Drosophila* have a third p53 isoform which has not yet been characterised completely. In addition, its interactions and regulation is yet to be determined (Dichtel-Danjoy et al., 2012).

By studying this protein in this model organism, we can draw comparison to the human system and thereby increase our knowledge of the human homologues of these molecules. New knowledge of p53 regulation and interactions will lead to new drugs being developed, and the identification of other key role players.

In an attempt to gain more insight into the regulation of p53 in *Drosophila*, we ask what other regulatory proteins exist, and specifically do SNAMA and Dmp53 interact? Furthermore, what is the

role of SNAMA in *Drosophila* p53 regulation? We attempt to answer this by determining whether p53 can act as a substrate for SNAMA.

The first stage of this research involved the isolation of the *Drosophila* p53 protein. This protein was to be expressed in vitro and used in various studies to determine what other proteins interact with it, as well as any specific interactions with the SNAMA protein and ubiquitin. By identifying role players in p53 pathways, one can attempt to regulate these role players and thereby regulate downstream processes such as cell death and the cell cycle.

The main aim of this study was to express and purify *Drosophila* p53, which would be used in protein interaction studies to determine whether SNAMA or any other proteins are involved in Dmp53 regulation.

The main aim was to be achieved by reaching the following objectives:

1. To heterologously express the long isoform of *Drosophila* p53 in *E. coli,* for solubility studies.
2. To purify the p53 protein expressed by the *E. coli.*
3. To heterologously express the long isoform of *Drosophila* p53 in *Saccharomyces cerevisiae* (Tgy 48-1).
4. To purify the p53 expressed by the *S. cerevisiae* Yeast (Tgy 48-1).
5. To perform immunoprecipitation studies using purified Dmp53, and anti Dmp53 purified antibodies.

Furthermore, an attempt was made to gain further insight into both the Dmp53 and SNAMA genes through bioinformatic analyses.

2. Materials and methods

2.1 Materials

2.1.1 Flies

Drosophila melanogaster Oregon R, adult flies and fly embryos were utilised where required throughout the study. Flies are bred in the laboratory, under 12hr day night cycles at a constant 25°C.

2.1.2 Bacterial Cells

Table 1: Bacterial Cell Strains used

	Supplier #catalogue no.
Epicurian Coli BL21 (DE3) pLysS Competent Cells	Stratagene #200132
E. cloni 10G Chemically Competent Cells DUOs	Lucigen #60107
One shot Top10 Chemically Competent *E. coli*	Invitrogen # C4040-10

2.1.3 Bacterial plasmids

Table 2: Bacterial Plasmid DNA used

	Supplier #catalogue no.
Pet-41a (+) (Appendix C)	Novagen #70556-3
pGEM-T Easy Vector System 1 (Appendix B)	Promega #A1360

2.1.4 Yeast

Saccharomyces cerevisiae (strain Tgy 48-1) as well as yeast expression plasmid pTG1895 (Appendix D) were kindly provided by Jean Marc Reichhart, University of Strasbourg, France.

2.1.5 Media

Table 3: Composition of media for flies, bacteria and yeast maintenance

	Recipe
Apple Juice agar	1.25% sucrose2% agar 0.025% streptomycin 0.25% methyl paraben 25% (v/v) Apple juice
Cornmeal Molasses agar	1% sucrose 0.75% agar 5% molasses

	7.5% cornmeal 2.5% yeast 0.05% methyl paraben
Luria-Bertani Broth (LB)	1% Tryptone 0.5% Yeast extract 1% NaCl
LB Agar	LB+ 1.5% Agar
YPG	2% Peptone 1% Yeast extract (pH5.8) 0.004% adenine sulphate 2%Glucose
YPG agar	YPG + 1.5% Agar
Synthetic Minimal (SD) media	0.67% Yeast Nitrogen Base 18.22% D-Sorbitol (pH 5.8) 10% 10x Dropout Solution 2% Glucose
SD agar	SD + 1.5% Agar
10 x Dropout Solution	300mg/l L-Isoleucine 1500mg/l L-Valine 200mg/l L-Adenine Hemisulfate salt 200mg/l L-Arginine HCl 200mg/l L-Hisitidine HCl monohydrate 300mg/l L-Lysine HCl

	200mg/l L-Methionine 500mg/l L-Phenylalanine 2000mg/l L-Threonine 200mg/l L-Tryptophan 300mg/l L-Tyrosine 1000mg/l L-Glutamic acid 1000mg/l L-Aspartic acid 4000mg/l L-Serine

2.1.6 Selection/antibiotic resistance

Table 4: Antibiotic resistance and selection of bacterial and yeast, cells and plasmids.

		Concentration in media
Epicurian Coli BL21 (DE3) pLysS	Chloramphenicol (Cm)	100µg/ml (stock 34mg/ml in ethanol)
E. cloni 10G DUOs	Streptomycin (Sm)	50µg/ml
One shot Top10	Ampicillin (Amp)	50 µg/ml
Pet41a	Kanamycin (Kan)	30 µg/ml
pGEM-T Easy	Ampicillin (Amp)	50 µg/ml
Yeast Tgy 48-1/pTG1895	Ampicillin (Amp) & SD -leucine –uracil	50 µg/ml

2.1.7 Buffers and other solutions

Table 5: Composition of Solutions and buffers used in the study

Birnboim 1 (resuspension buffer)	50mM Tris.HCl pH8.0 10mM EDTA 100µg/ml RNAse A
Birnboim 2 (Lysis Buffer)	200mM NaOH 1% w/v SDS
Birnboim 3 (Neutralization Buffer)	3M Potassium Acetate pH5.5
Ethanol Acetate	24 Parts Absolute Ethanol 1 part Sodium Acetate pH5.2
Bacterial Lysis Buffer	50mM Tris-HCl pH7.5 2mM EDTA 200mM NaCl 2mM DTT 0.1% Triton-X 100 0.1% PMSF 200 µg/ml Lysozyme 3mM MgCl2

Bacterial Denaturation Buffer	8M Urea 50mM Sodium Phosphate buffer pH 7.5 0.1 M NaCl 1mM PMSF 2mM DTT
Yeast Lysis Buffer	50mM Tris HCl pH8 1% DMSO 200mM NaCl 1mM EDTA 1mM PMSF 1 µg/ml Leupeptin
Transfection Mix (per Transfection)	240µl PEG3500 w/v 18 µl Lithium Acetate 2.0M 47µl H_2O
Genomic DNA grinding buffer	5% Sucrose 80mM $MgCl_2$ 100mM Tris pH 8.5 0.5% SDS

	50mM EDTA
Southern blotting: nucleic acid transfer buffer (20x SSC)	8.823% TriSodium Citrate 17.532% NaCl pH 7-8
Southern blotting: Depurination buffer	11ml HCl 989ml dDH_2O
Southern blotting: Denaturation buffer	8.766% NaCl 2% (w/v) NaOH
Southern blotting: Neutralization buffer	8.766% NaCl 6.05% Tris pH 7.5
Southern blotting: Hybridization Solution	5X SSC 1% Blocking Solution 0.1% N Lauryl Sarcosine 0.02% SDS 1.5 % Roche DIG Easy Hyb
Southern blotting: Low stringency wash buffers	2X SSC 0.1% SDS
Southern blotting: Medium stringency wash buffers	1X SSC 0.1% SDS

Southern blotting: High stringency wash buffers	0.1% SSC 0.1% SDS
Camptothecin stock	50 mM In DMSO
Coomassie stain	0.1% Coomassie Blue R-250 40% ethanol 10% glacial acetic acid 50% ddH20
Coomassie destain	40% ethanol 10% glacial acetic acid 50% ddH20
SDS PAGE running buffer 5x	25mM Tris Base 192mM Glycine 0.1% SDS Distilled water to 3L pH8.3
SDS PAGE sample buffer	62.5mM Tris-HCl pH6.8 10% Glycerol 2% SDS 5% B-Mercaptoethanol

	0.05% Bromophenol Blue 50% ddH20
Towbin	25mM Tris Base 192mM Glycine 0.1% SDS 20% Methanol pH8.3
PBS 10x	1.4M NaCl 27mM KCl 0.1M Na2HPO4 18mM KH2PO4 pH to 7.4 with HCl
PbsT	1X PBS + 0.1% Tween 20
HIS nickel chromatography native conditions equilibration buffer	50mM Sodium Phosphate pH8.0 0.3M NaCl 10mM Imidazole
HIS nickel chromatography native conditions wash buffer	50mM Sodium Phosphate pH8.0 0.3M NaCl 10mM Imidazole

HIS nickel chromatography native conditions elution buffer	50mM Sodium Phosphate pH8.0 0.3M NaCl 250mM Imidazole
HIS nickel chromatography denaturing conditions equilibration buffer	0.1M Sodium Phosphate pH 8.0 8M Urea
HIS nickel chromatography denaturing conditions wash buffer	0.1M Sodium Phosphate pH 6.3 8M Urea
HIS nickel chromatography denaturing conditions elution buffer	0.1M Sodium Phosphate pH 4.5-6.0 8M Urea

2.1.8 Primers for PCR

Table 6: Primer sequences used in Polymerase Chain reaction

Dmp53			
p53 for:	5′	AGGAGATATACAAGCTTCCCCTATAC	3′
Sal1p53rev:	5′	GTCGACTTCATGGCAGCTCGT	3′
pET41histag:	5′	TAGTGAAGCTTGTCATCACCAT	3′
Biotin probe			
p53.biotin.fwd:	5′	ATGTACTCGATTCCGCTGAACAAGC	3′
SNAMA A&B			
ABfwdexon3SNAMA	5′	TGAGGAATTCGGTGAACG	3′
BrevUTRSNAMA	5′	ATCCACATGGTAGCGGGAT	3′
ArevEXON7SNAMA	5′	CAACGGATCGTCAATGG	3′

2.1.9 Antibodies

Table 7: Primary Antibodies used in Western blot analysis

		Antibody target	Supplier #catalogue no.
Anti-*Drosophila* p53 (dD-21)	goat polyclonal IgG	internal epitope	Santa Cruz #sc-17577
Anti-*Drosophila* p53 (dA-17)	goat polyclonal IgG	C-terminal	Santa Cruz #sc-17578
Anti-*Drosophila* p53 (d-200)	rabbit polyclonal IgG	186-385	Santa Cruz #sc-25767

Table 8: Secondary Antibodies used in Western blot analysis

	Supplier #catalogue no.
Anti-Goat IgG (whole molecule)–Peroxidase antibody produced in rabbit	Sigma Aldrich #A8919
Anti-Rabbit IgG (whole molecule)–Peroxidase antibody produced in goat	Sigma Aldrich #A0545

2.1.10 Restriction Enzymes

Table 9: Restriction enzymes used for DNA digestion

	Supplier #catalogue no.
*Hind*III	NEB: #R0104 Roche: #10656321001 Fermentas: #ER0505
*Sal*I	NEB: #R0138 Roche: #10567663001 Fermentas: #ER0645
*Bam*HI	NEB: #R0136 Roche: #10656275001 Fermentas: #ER0055
*Eco*RI	NEB: #R0101 Roche: #11175084001 Fermentas: #ER0275
*Xho*I	NEB: #R0146 Roche: #10899194001 Fermentas: #ER0695

2.1.11 Molecular weight Markers

Table 10: Molecular Weight markers used in DNA and Protein Electrophoresis

		Supplier #catalogue no.
DNA	Gene ruler 1kb plus	Fermentas #SM1331
Protein	Pageruler Prestained Protein ladder	Fermentas #SM0671

2.1.12 General reagents and kits

Table 11: General reagents and kits used

	Supplier #catalogue no.
FastAP Thermosensitive Alkaline Phosphatase	Fermentas #EF0651
t4 DNA ligase	Fermentas #EL0015
HIS select nickel affinity gel	Sigma #p6611
Super block	Pierce Thermo Scientific #37545B
Chemiluminescent Nucleic Acid Detection Module kit	Pierce Thermo Scientific #89880
glass beads unwashed 425-600um	Sigma Aldrich #g9268
unitek glass beads 5mm	Merck Millipore #104017
DNA clean & concentrator-5 kit	Zymo Research #d4003
Nuclease free water	Qiagen #129115
SuperSignal® West Pico chemiluminescent substrate	Pierce Thermo Scientific #34080

TRI reagent (TRIzol)	Sigma Aldrich #T9424
RevertAid Reverse Transcriptase	Fermentas #EP0441
RevertAid First Strand cDNA synthesis kit	Fermentas #K1622
PCR Master Mix	Fermentas #K0171
DIG Easy Hyb	Roche #11603558001
Taq DNA Polymerase with Standard Buffer	NEB: #M0273S Fermentas: #EP0402
10mM dNTP mix	Fermentas: #R0192
Oligo(dT)18 Primer (0.5µg/µl)	Fermentas #S0131
25mM $MgCl_2$	Thermo Scientific #R0971
IPTG	Sigma Aldrich #i5502
PALL Biotrace PVDF	#66543
Kodak Biomax film	#8194540),
Axim fixer	#ffc-fixruhs10bb
Axim developer	#fdc-devru0010bb

2.1.13 Initial clone

A clone of *E. coli* BL21 (DE3) pLysS transformed with an N terminal truncated Dmp53 insert in pET41a plasmid was kindly provided by Brent Oosthuysen. The sequence of this insert appears in Appendix A.

2.2 Methods

2.2.1 Bioinformatics

A sequence analysis of the gene, mRNA and protein sequences of p53 and SNAMA was undertaken using online tools and databases (flybase, NCBI, Clustal W2) and the DNAman (Lynnon Biosoft v4.03) software package.

2.2.2 Agarose gel electrophoresis

Following PCR or restriction digestion, DNA was analysed using agarose gel electrophoresis. DNA was separated on a 1% Agarose gel with Fermentas Generuler 1kb plus (Appendix E) as a molecular weight marker. Ethidium bromide (EtBr) was used as a visualising agent and Gels were visualised on a calibrated Bio-Rad Gel Doc XR. For DNA purification from gels, EtBr was replaced with sybr gold and visualised on a Dark reader system from Clare.

2.2.3 Gel purification of DNA

For ligation into plasmids, DNA inserts of interest were excised from a 1% agarose gel, and purified using the Zymo DNA Clean and Concentrator kit, as per manufacturer's instructions. DNA was resuspended in nuclease free water.

2.2.4 Plasmid DNA purification by ethanol acetate method

Preparation of plasmid DNA, pTG1895 and pET41a containing Dmp53 insert, was performed based on the method of Birnboim (1983). The cells were lysed and DNA, proteins and cell debris was pelleted by centrifugation. Thereafter pellets were resuspended and DNA separated out into the supernatant fraction. The DNA was precipitated by ethanol acetate precipitation. Isopropanol was used to further precipitate DNA. Contaminating salts were removed from the DNA pellet by washing with 70% ethanol. The pellet was air-dried and the DNA was resuspended in nuclease free water.

2.2.5 Restriction digestion of DNA

Restriction digestion of DNA was performed according to manufacturer's instructions. Digestions were incubated overnight at 37^{o}C.

a. p53 and pTG1895 were hydrolysed with *Hind*III and *Sal*I for cloning
b. p53 in pET41a was hydrolysed with *Hind*III and *Bam*HI for linearization.
c. Cloned pGem-T Easy plasmid DNA was hydrolysed with *Eco*RI to screen for p53 insert.
d. Genomic DNA was hydrolysed with *Eco*RI, *Sal*I and *Bam*HI, *Sal*I and *Xho*I, for Southern blotting.

2.2.6 Preparation of plasmid DNA for ligation (dephosphorylation)

Following restriction enzyme digestion of pTG1895 Plasmid DNA, the resulting linear DNA was subsequently dephosphorylated using alkaline phosphatase (Fermentas #EF0651) according to the manufacturer's instructions.

2.2.7 Ligation of insert into linearised plasmid

p53 fragment (from the initial clone supplied by Brent Oosthuysen) was cloned into pTG1895 as well as into pGEM-T Easy vector using T4 DNA ligase (Fermentas #EL0015), according to manufacturer's instructions.

2.2.8 Transformation of chemically competent cells

The following transformations of chemically competent cells were performed.

- E. cloni Duos cells were transformed with intact pTG1895. For large-scale preparation of pTG1895 plasmid, the recombinant plasmid was used to transform E. cloni Duos cells (Lucigen #60107) as per manufacturer's instructions. 1µl cells were heat shocked at 42°C for 45 seconds and cold snapped on ice for 2 mins. Clones were screened for insert using Ampicillin/Streptomycin resistance.

- E.cloni Duos cells were transformed with p53 in pGEM-T Easy. For gel purification of p53 insert, p53 was amplified by PCR from the initial clone, and transformed into pGEM-T Easy vector. This ligation reaction was used to transform *E.cloni* duos cells (Lucigen #60107) as per manufacturer's instructions. Clones were screened for insert using Blue/White screening and Ampicillin/Streptomycin resistance.

- Top10 cells were transformed with p53 in pTG1895. p53 was ligated into pTG1895 plasmid and the ligation reaction was used to transform top10 competent cells (Invitrogen #C4040-10) as per manufacturer's instructions. Clones were screened for insert using Ampicillin resistance.

Resulting colonies were screened through restriction digest analysis. Clones that contained the recombinant plasmid were used to inoculate large-scale cultures. Recombinant plasmid DNA was then isolated from these cultures.

2.2.9 Transfection of yeast

From a 2ml overnight preculture of untransfected yeast, a 100ml culture was inoculated and grown at 30°C to an OD of 0.5. The yeast cells were pelleted, the supernatant was discarded and the pellet resuspended in the residual fluid. This was transferred to a smaller microcentrifuge tube and the cells were pelleted, the supernatant was discarded and the cells were washed in water (H_2O). The cells were pelleted again and then resuspended in water to a ratio of 10ml of 0.5ml culture: 50µl water. Each of these aliquots was then transfected. For transfection, 100µg of salmon sperm DNA was added to 50µl of yeast cells from above, and mixed by pipetting. To this 10µl of plasmid DNA (p53 in pTG1895) was added. 305µl of transfection mix (Lithium Acetate, PEG3500) was added, and the tube mixed by vortexing, before incubation at 42°C for 40mins. After incubation cells were pelleted at 3000g for 1min, the supernatant was discarded and the cells washed with water. A further round of washing was performed before the cells were resuspended in 100µl water and plated on synthetic minimal medium containing Ampicillin and standard dropout solution without Leucine and Uracil. Plates were incubated at 30°C.

2.2.10 Protein expression

Protein expression in bacteria was performed in cultures containing Chloramphenicol (100µg/ml) and Kanamycin (30µg/ml) (inoculated from a starter culture of the initial clone). The cultures were grown to an OD_{600} of 0.6, upon which expression was triggered using IPTG added to a final concentration of 0.2mM. Cultures were incubated on a

shaking platform incubator at 37°C, for a specified time (0, 2, 4, 6 hours). Following this, cells were spun down at 6000g for 15 mins at 4-10°C. All supernatants were retained for analysis.

For small culture expressions 2ml cultures were used. The cultures were grown to an OD_{600} of 0.6, upon which expression was triggered by adding IPTG to a concentration of 0.1mM. Cultures were incubated on a rotating mixer, in a walk in 37°C incubator room. 0, 2, 4 and 6 hour expressions (after induction) were pelleted by centrifugation at 6000g for 5 mins. All supernatants were retained for analysis.

2.2.11 Protein extraction

Pelleted bacterial cells were resuspended in lysis buffer (20ml for large cultures and 500µl for small cultures). Cells were allowed to lyse for 30mins on a shaking incubator at 0°C. Large cultures were then subjected to sonication. Following lysis, cells were pelleted by centrifugation, supernatants were retained as the soluble protein fraction. Cell pellets were resuspended in lysis buffer for protein extraction from inclusion bodies. This was achieved by repeated washes, pelleting and resuspension in lysis buffer. Inclusion bodies were then resuspended in denaturation buffer (Table 5).

2.2.12 Yeast protein expression

Protein expression of p53 in yeast was performed at room temperature on a shaking incubator. Large (500ml or 1litre) and small (100ml) cultures were inoculated with a single yeast colony containing the p53 insert in a pTG1895 plasmid. The culture media used was Synthetic minimal (SD) media. Cultures were incubated for 3-7days, before cells were pelleted by centrifugation in preparation for protein extraction. The supernatant media was kept for the purification of any excreted protein.

2.2.13 Yeast protein extraction from pelleted cells

After protein expression, pelleted cells were resuspended in yeast lysis buffer, or water. The suspension was frozen at -30°C, thawed and cells were pelleted by centrifugation. Cells were then ground using acid washed glass beads on a vortex for multiple rounds of 3 minutes, and were placed on ice for 5 mins between vortexing. Cells were again pelleted and frozen, before repeating the process, (2 more rounds). The supernatant resulting from this extraction was carefully decanted, and purified by chromatography.

2.2.14 HIS nickel affinity chromatography

Proteins expressed in inclusion bodies (bacterial expressions) were purified by chromatography using a Sigma HIS select nickel affinity gel (#p6611), the purification was performed according to manufacturer's instructions under denaturing conditions.

Yeast protein expressions were purified by chromatography using a Sigma HIS select nickel affinity gel (#p6611), the purification was performed according to manufacturer's instructions under native conditions for large scale purification. For cultures > 200ml a 3ml chromatography column was used, smaller cultures were purified using a 1ml column. All purification took place at room temperature.

2.2.15 Protein concentration

To concentrate proteins expressed by yeast, so as to make them visible by coomassie staining 3 methods were attempted.

- Freeze drying: Protein samples were freeze dried overnight in a Virtis SRC 10 freeze drier. Proteins were then resuspended in minimal volume of buffer before SDS PAGE.

- Stirred cell: Protein fractions from HIS purification were concentrated using a Millipore stirred ultrafiltration cell, according to manufacturer's instructions. Following concentration, proteins were visualised by SDS PAGE.

- Trichloroacetic acid (TCA) precipitation: Protein fractions from HIS purification were concentrated by TCA precipitation. An equal volume of 20% TCA was added to the sample. Samples were then incubated for 30 mins on ice before being pelleted (18000g, 4°C, 15mins). After decanting the supernatant, 300µl of cold acetone was then added to the pellets before further centrifugation for 5mins. The supernatant was removed and the

pellets air dried. Samples were resuspended in SDS PAGE sample buffer and heated to 65°C for 3mins before loading on an SDS gel.

2.2.16 SDS PAGE

The Laemmli method (Laemmli, 1970) of Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis was used to establish the purity of expressed protein as well as to separate protein in preparation for transfer to Polyvinylidene fluoride (PVDF) membrane for Western blotting. For visualisation of proteins, gels were stained with Coomassie Blue overnight and then destained for 2 to 3 hours using Coomassie Destain. Pageruler prestained protein ladder (Appendix F) was used as a molecular weight marker. Gels were photographed on a Bio-Rad GS800 Calibrated Densitometer. For blotting, gels were not stained but instead protein was transferred to a PVDF membrane by electroblotting.

2.2.17 Western blotting

Following separation by SDS PAGE, protein samples were transferred in a Hoefer mini VE blotting module, to PVDF using Towbin buffer. Blots were blocked overnight in blocking buffer (SuperBlock™ Dry Blend Blocking buffer (Pierce cat #37545) to prevent non-specific binding, and then probed with Anti Dmp53 for 1 hour. A 1:5000 dilution of primary antibody was used. The primary antibodies manufactured by Santa Cruz are polyclonal antibodies raised against

Dmp53 (Table: 7). Before incubation with secondary antibody the blot was washed twice with PBS tween (0.1 %) for 10 mins per wash. A 1:10000 dilution of secondary antibody in PBS tween was used to detect primary antibody. Incubation in secondary antibody was for 1 hour. Secondary antibodies were manufactured by Sigma Aldrich and are IgG horse radish peroxidase conjugates (Table 8). Membranes were then washed 6 times ten minutes per wash with PBS tween, before detection. Detection was performed using the SuperSignal® West Pico chemiluminescent substrate (Pierce Cat# 37077) as per manufacturers' instructions. Membranes were exposed to Kodak Biomax film, then fixed and developed using Axim fixer and. Exposed film was photographed on a Bio-Rad GS800 Calibrated Densitometer. All incubations and washes were performed with agitation from a platform shaker at slow speed.

2.2.18 Extraction of genomic DNA

Genomic DNA was prepared using a protocol based on the method described by Bender et al., (1983). Flies were immobilised using Carbon dioxide, collected and homogenised in 100µl grinding buffer. After a 30 minute incubation at 75°C, 35 µl of 8 M KOAc was added and incubated on ice for 30 minutes to pellet the cellular debris and protein. The DNA was then precipitated with isopropanol and washed with 70% ethanol before being resuspended in nuclease free water.

2.2.19 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify DNA sequences for cloning, to make biotin labelled DNA probes for Southern blotting, and to detect the presence of specific DNA sequences. Conditions for PCR were according to enzyme manufacturer's instructions. Annealing temperature was optimised to ±5°C of primer TM (as specified by the primer manufacturer). For colony PCR DNA template was omitted and a 0.5mm bacterial clone used.

2.2.20 Southern blotting

After digestion of genomic DNA from adult flies, hydrolysed DNA was run on 1% agarose gel and visualised on a Bio-Rad Geldoc XR. DNA on the gels was then depurinated for 10 mins, denatured for 30 mins and neutralized for 30 mins by soaking the gel in the respective buffer (Table 5) with gentle agitation. DNA was then transferred to a nylon membrane by capillary transfer overnight, using Saline Sodium Citrate (SSC) as a transfer buffer. Membranes were then crosslinked in Stratagene UV crosslinker using the auto crosslink feature (1200000 µJ). Membranes were then prehybridised in preheated hybridisation solution for 30mins at 65°C. Probe (biotin labelled DNA) was then added to the solution and the membrane was hybridised overnight at 65°C. After hybridisation, membranes were washed in a series of buffers of increasing stringency. Membranes were rinsed briefly in, then washed twice (5 mins each) in low stringency buffer. Then washed twice (10 mins each) in medium stringency buffer, and washed four times (5 mins each) in high stringency buffer. Hybridisation and

washing was carried out in tubes in an Amersham life science Hybridisation oven/shaker. The chemiluminescent nucleic acid detection module kit was then used to detect DNA bound to probe, according to the manufacturer's instructions, before exposing the membrane to film and developed as per Western blotting above.

2.2.21 Camptothecin treatment of flies

Adult flies were exposed to 1mM Camptothecin (Sigma #C9911) treated baker's yeast for 48 hours alongside their usual cornmeal feed.

2.2.22 Extraction of RNA

The RNA extraction procedure was based on that of Chomczynski (1993), using the TRIzol LS reagent. Larvae and adults (both camptothecin treated and untreated) were both homogenized in 500µl of TRIzol using an eppendorf pestle homogenizer. RNA was then recovered as per the manufacturer's instructions. The RNA pellet was then dissolved at 65°C in 50µl nuclease free water. Samples were snap cooled on ice and the concentration was determined on a Nanodrop spectrophotometer at 260 nm.

2.2.23 cDNA synthesis RT PCR SNAMA Oligo(dT) & specific primers

First strand cDNA synthesis was performed using Fermentas Reverse transcriptase, according to the manufacturer's protocol. Synthesis used Oligo(dT) primers in one instance and specific SNAMA reverse primers ("BrevSNAMAUtr", "ArevExon7SNAMA") in another. cDNA was then amplified by PCR.

3. Results

3.1 Gene Structure and isoforms of Dmp53

Dmp53 is a well-known tumour suppressor protein, whose function is affected by numerous cancers. The research used *Drosophila* as a model organism to gain valuable insight into Dmp53 and its interactors. The project aims to clone and overexpress Dmp53 protein and look for possible interacting proteins. In humans, MDM2 marks Dmp53 for degradation by ubiquitination, however, in *Drosophila* a MDM2 homolog is yet to be discovered. RBBP6 is a known interactor in other species and it is thought that SNAMA, the *Drosophila* RBBP6 homolog would thus also interact. Furthermore, SNAMA is suspected of being a Ubiquitin like protein (UBL) and perhaps to play this role in the absence of MDM2. In mammals the SNAMA homologue, RBBP6 has been shown to enhance the activity of MDM2.

Dmp53 is encoded by the *p53* gene found on chromosome 3R of *Drosophila*, its eleven exons form 3 major domains, namely Transactivation, DNA binding and Oligomerisation domains. At least 2 isoforms, a full length Dmp53 and an N-truncated Dmp53 which lacks the transactivation domain exist (Dichtel-Danjoy et al., 2012). A third isoform which contains only the transactivation domain is also identified by Bourdon, (2005).

A schematic overview of the Dmp53 gene, showing exon positions and sizes is presented below (Figure 6).

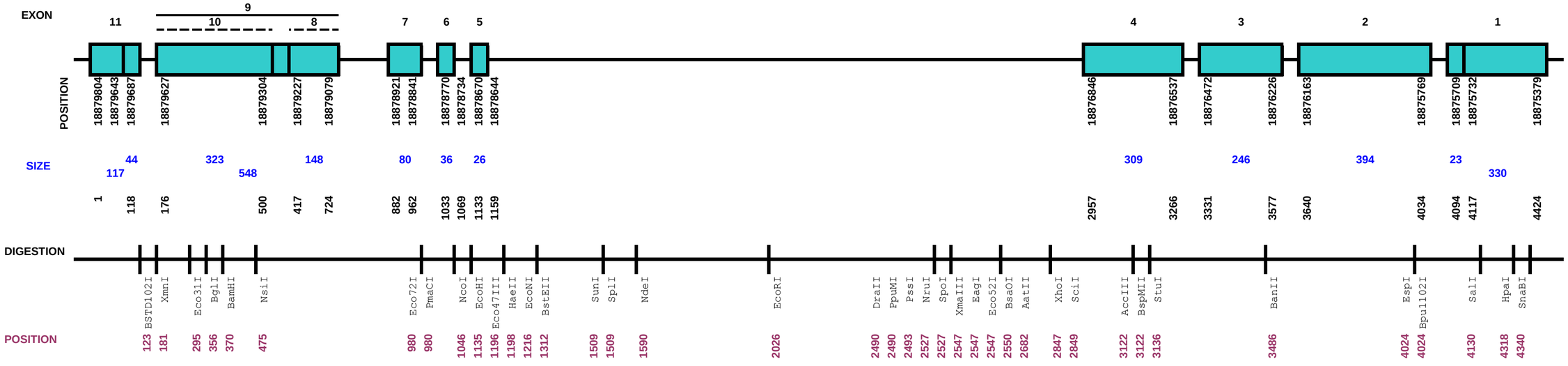


Figure 6: Schematic overview of *Drosophila* p53 gene structure. The figure shows the gene to scale, with the positions and numbering of the specific exons. The sizes of the expected exons are shown with their positions relative to the gene. An analysis of restriction digestion by enzyme which have single cut sites on the gene was carried out using DNAman software and the positions of these cut sites relative to the gene start are also shown.

Southern blotting was performed to establish the gene copy number of Dmp53, however, the results were inconclusive. Genomic DNA was extracted from adult flies and DNA was hydrolysed using *Eco*RI and combinations *Sal*I, *Bam*HI and *Xho*I , which were found to cut the strand once in the Dmp53 gene region. Agarose gel electrophoresis shows successful digestion of genomic DNA (Figure 8). Upon Southern blot analysis, only PCR generated Dmp53 shows binding with the probe (Figure 8). The protocol was tested by Southern blot analysis of varying dilutions of biotin labelled Dmp53 probe, which were generated by PCR (Figure 7).

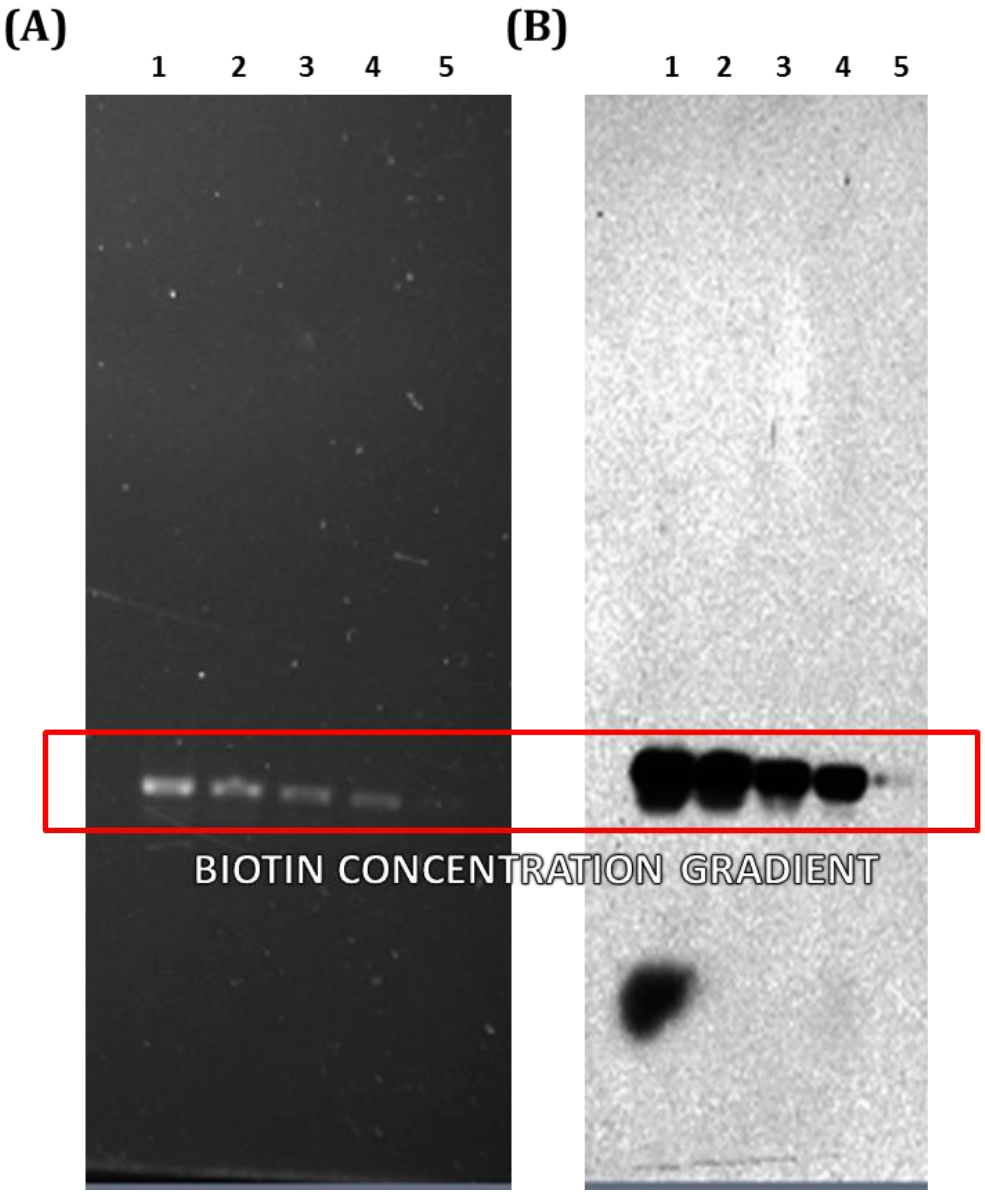


Figure 7: Agarose gel of biotin probe and Southern blot thereof. 100, 80, 60, 40, and 20% dilutions of the biotin probe in water were analysed on an agarose gel (A) then transferred to a membrane for Southern blotting. The membrane was probed with biotin labelled Dmp53 DNA and detected using the chemiluminesence detection module kit. The blot (B) confirms that the protocol was successful.

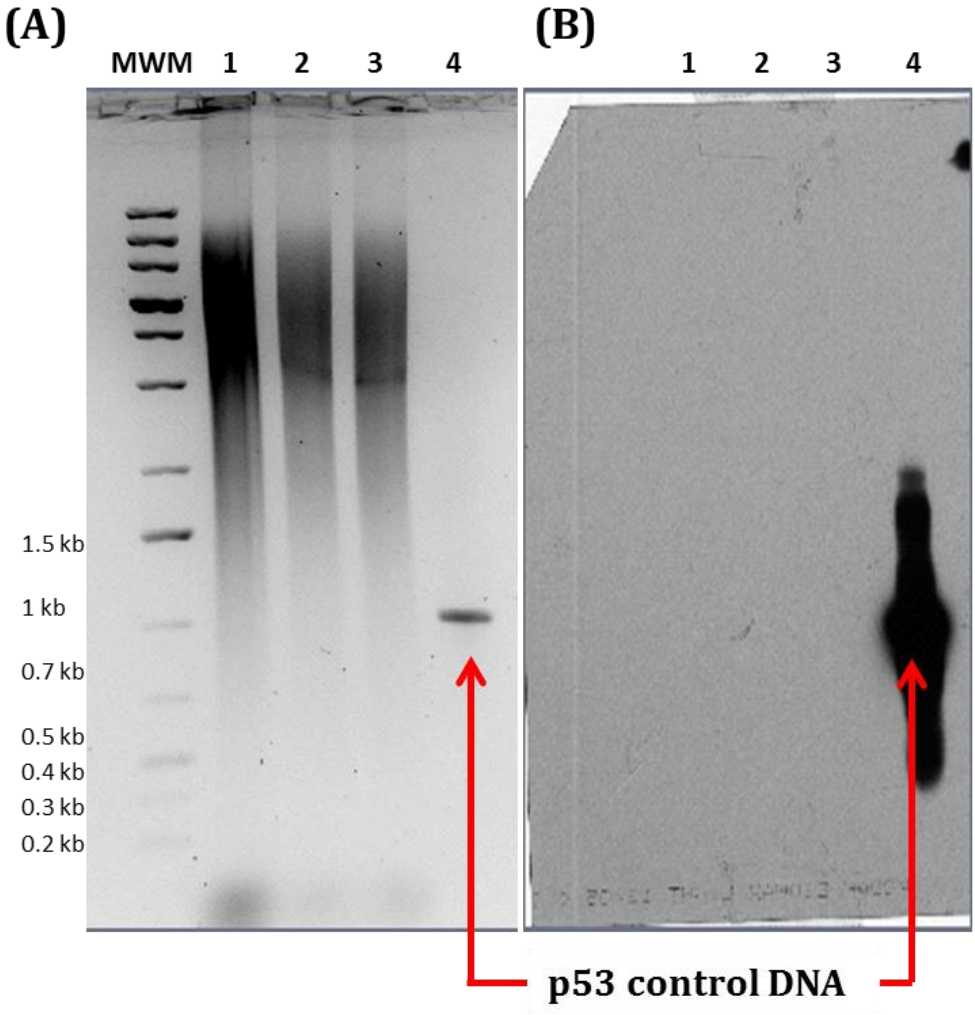


Figure 8: Digestion of genomic DNA and Southern blot thereof. (A) Lanes 1 to 3 show the presence and digestion of genomic DNA in preparation for Southern blotting. Lane 1 = *Eco*RI digestion, 2 = *Sal*I and *Bam*HI digestion, 3 = *Sal*I and *Xho*I digestion. Lane 4 shows a Dmp53 control. The Southern blot alongside (B) shows only, probe binding to the control insert DNA.

3.2 Heterologous expression of Dmp53 in *E. coli*

E. coli BL21 (DE3) pLys S, was transformed with, a clone of Dmp53 insert in pET41a plasmid. The transformed cells were grown to an optical density of 0.6 in LB media and protein expression was induced using IPTG. 2, 4 and 6 hour inductions were performed and an uninduced sample kept. After induction cells were lysed chemically and proteins re-suspended. Proteins in inclusion bodies were re-suspended in denaturation buffer. SDS PAGE confirmed that heterologous expression had taken place. Subsequently, Western blot analysis was used to probe for Dmp53 (Figure 9). After induction for 6 hours a fraction of the supernatant (soluble proteins) was purified by chromatography to determine if Dmp53 was present (Figure 10). Both of these results clearly show that Dmp53 is expressed as an insoluble inclusion body, irrespective of the time period of protein expression after induction. Specifically, Figure 9 shows Dmp53 detected only in pellet fractions in 2, 4 and 6 hr expressions and Figure 10 shows no eluted soluble protein 6 hours after induction. It is notable that the cross reacting proteins detected may be foreign to *E. coli,* as they also appear in pellet fractions. The lack of an untransformed *E. coli* control leaves this question unanswered.

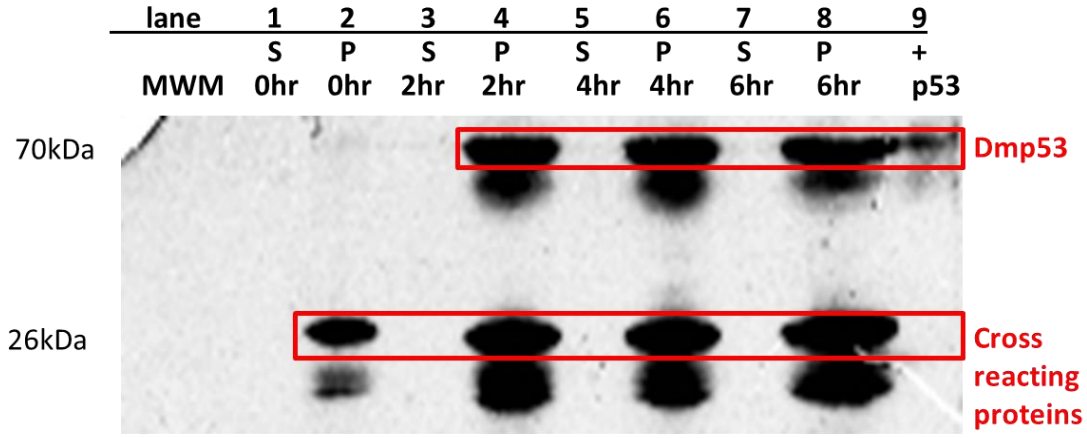


Figure 9: Western blot analysis of Dmp53 in soluble and insoluble fractions of expressed protein. After heterologous expression of Dmp53 in *E. coli,* proteins were extracted and inclusion bodies re-suspended. Crude extracts were probed for Dmp53 using anti Dmp53 rabbit primary antibody, and anti-rabbit peroxidase secondary antibody. Odd numbered lanes contain soluble protein fractions. Even numbered lanes contain re-suspended pellet fractions (inclusion bodies). Lane 1 and 2 contain uninduced samples. Lanes 3-4; 5-6; and 7-8; contain 2, 4 and 6 hour induction samples respectively. Lane 9 contains previously prepared purified heterologously expressed Dmp53 as a control. The Figure shows that Dmp53 is expressed as an insoluble inclusion body, irrespective of the time period of protein expression after induction.

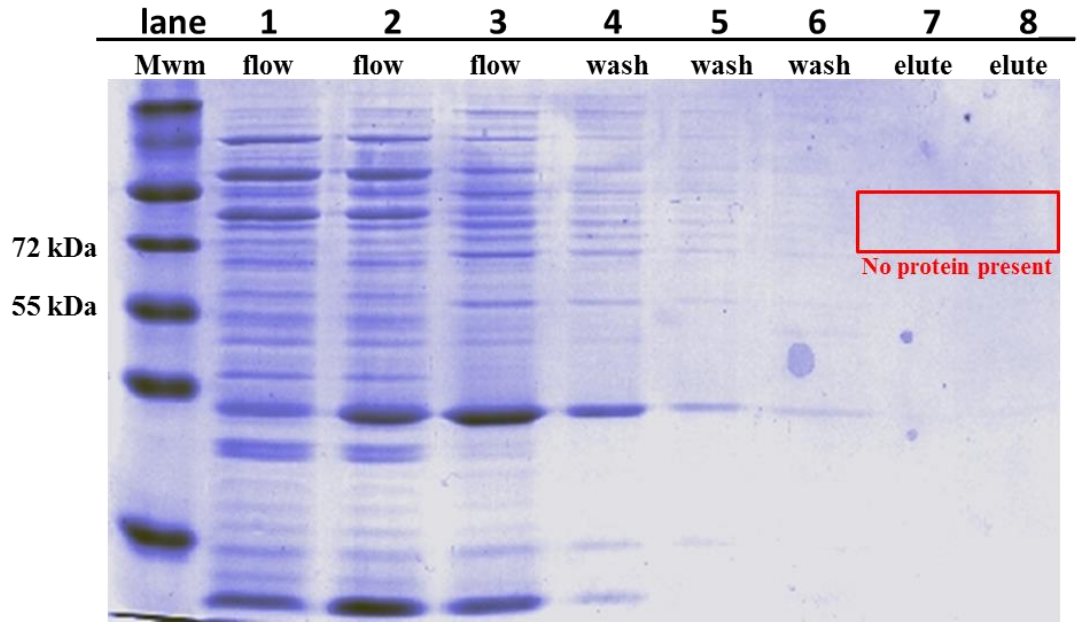


Figure 10: Purified soluble protein fraction from heterologous expression in *E. coli*, 6 hours after induction. The crude protein extract was purified by HIS chromatography. No Dmp53 is present in the elution fractions. This once again shows that Dmp53 is expressed as insoluble inclusion bodies in *E. coli*.

3.3 Protein expression and purification in Yeast

The yeast plasmid pTG1895, was prepared on a large scale. Subsequently some plasmid was restricted with *Eco*RI to determine if preparation was successful. The presence of two bands (at positions 7000bp and 2800bp) confirms that pTG1895 has the expected size and was linearised as compared to a typical 3 band migration pattern of undigested plasmid (Figure 11). The remainder of the plasmid was linearised and dephosphorylated in preparation for cloning. Dmp53 in pET41a was also prepared. PCR was performed on the Dmp53 in pET41a with primers designed to amplify the Dmp53 insert. These primers change the restriction sites to those of pTG1895 so as to clone this amplified insert into pTG1895 in the correct frame and orientation. Figure 12, shows the successful amplification of Dmp53 insert (lane1), which is slightly larger in size when compared to the original fragment (lane2). This is due to the inclusion of a HIS tag and some upstream sequence from the pET41a vector.

The amplified insert was ligated into the TA cloning site of pGEM-T Easy plasmid and subsequently *E. coli* was transformed with the plasmid. Bacterial clones were screened for successful transformants, using blue white screening and then using restriction digestion analysis (with *Hind*III and *Sal*I, which form the cloning site in pTG1895).

Plasmid DNA from successfully transformed clones was purified on a large scale in preparation for DNA recovery from an agarose gel. This was subsequently ligated into pTG1895, and bacteria transformed.

Clones were screened for successful transformants and plasmid DNA was isolated and purified on a large scale.

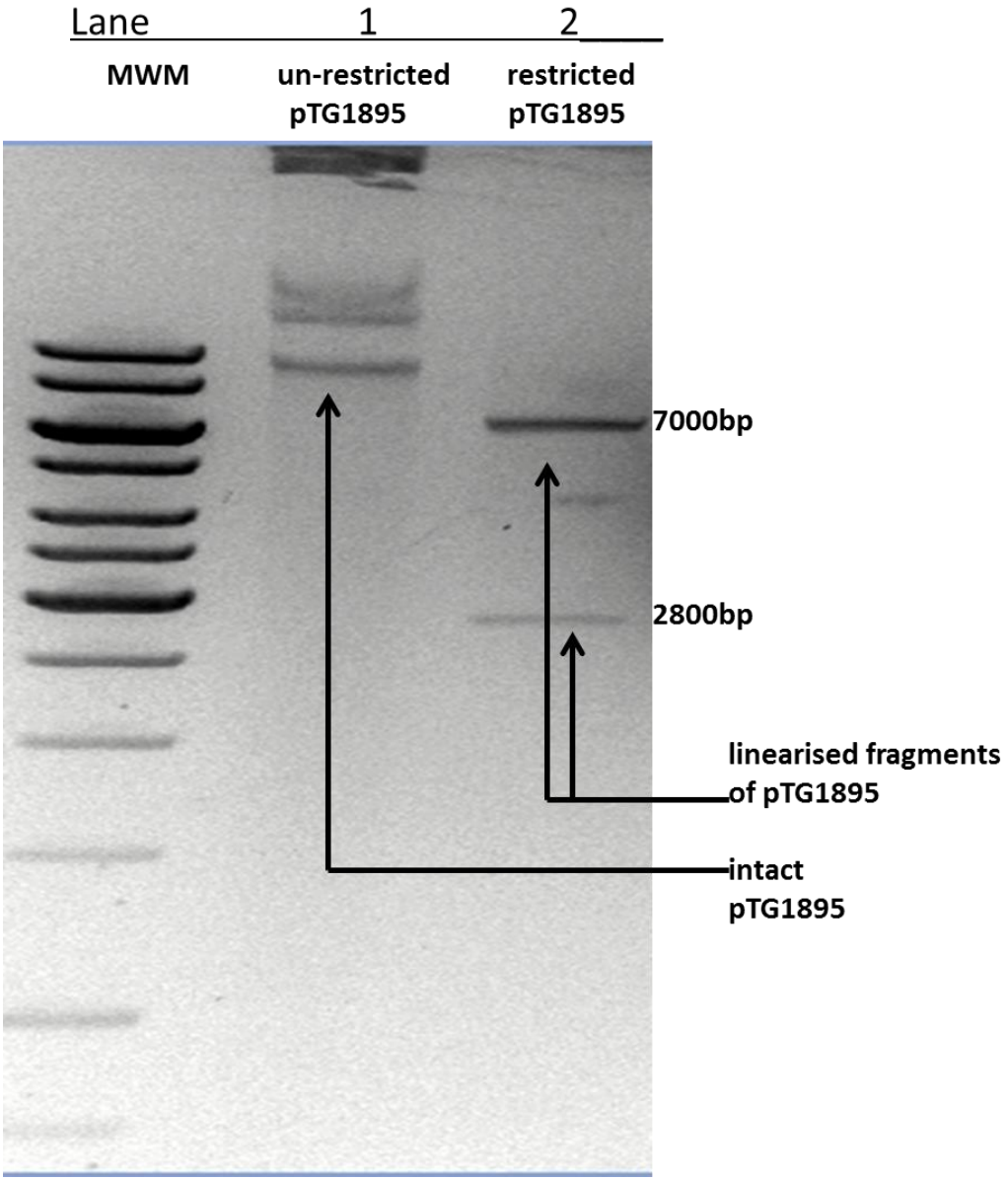


Figure 11: Confirmation of plasmid preparation by restriction digest. Lane two contains plasmid hydrolysed with *Eco*RI. Two bands of the expected size are present, confirming that pTG1895 has been successfully prepared.

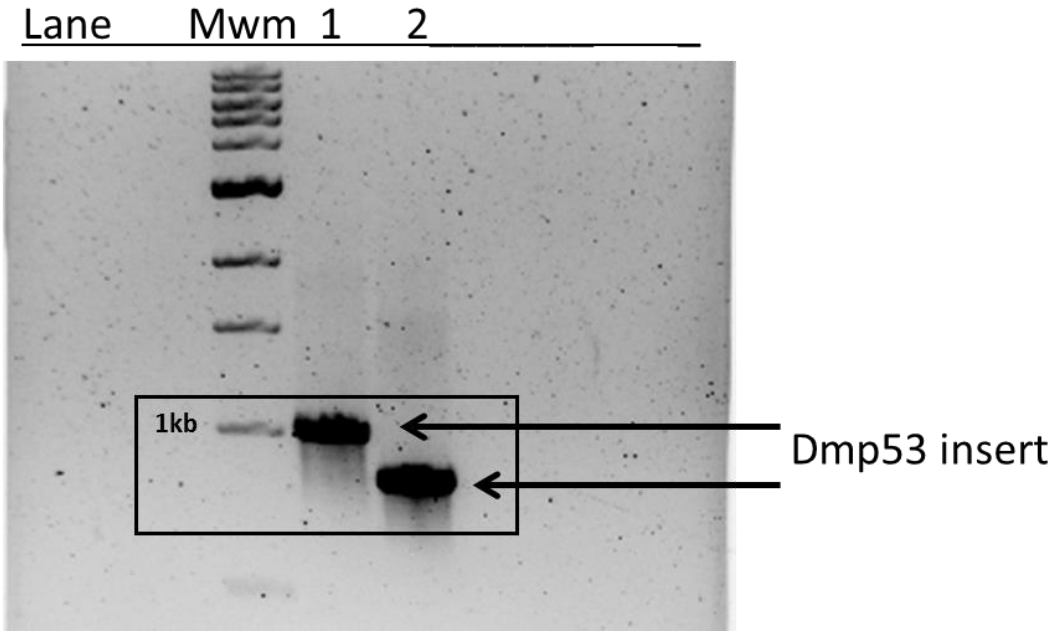


Figure 12: Dmp53 insert was successfully amplified by polymerase chain reaction, shown in lane 1. Lane 2 shows a control PCR of Dmp53 using different primers and template. The insert was gel purified, hydrolysed and ligated into pGem-T Easy for large scale preparation before being ligated into yeast plasmid pTG1895 in further reactions.

Yeast (Tgy 48-1) was then cultured and made competent. Following this the competent yeast was immediately transfected with the purified plasmid containing the Dmp53 insert. The transfection followed a modified Lithium Acetate method. Colonies which grew on selective media, were then screened by colony PCR, to confirm the presence of the Dmp53 insert. Clones that have taken up Dmp53 show a band at 1000bp as confirmed by a positive control (Figure 13).

Having confirmed the presence of Dmp53 insert in clones 1 and 4, these clones were cultured in SD media, for the expression of Dmp53 protein. After growing to a suitable OD, the yeast was pelleted and re-suspended. Protein extracts were purified using HIS-nickel affinity chromatography. The cloned Dmp53 insert (1020bp) gives rise to a 340 aa sequence, which equates to a protein approximately 37.5 kDa in size, however, Dmp53 could not initially be visualized by SDS PAGE, where gels appeared blank or showed very faint staining in lanes of interest. To determine if protein expression was indeed occurring, transformed and untransformed yeast were cultured. Crude protein extracts from these cultures were concentrated using a Millipore stirred cell, SDS PAGE showed that some protein was being expressed and secreted by the yeast, however, Western blot analysis could not positively identify Dmp53 due to non-specific binding (Figure 14).

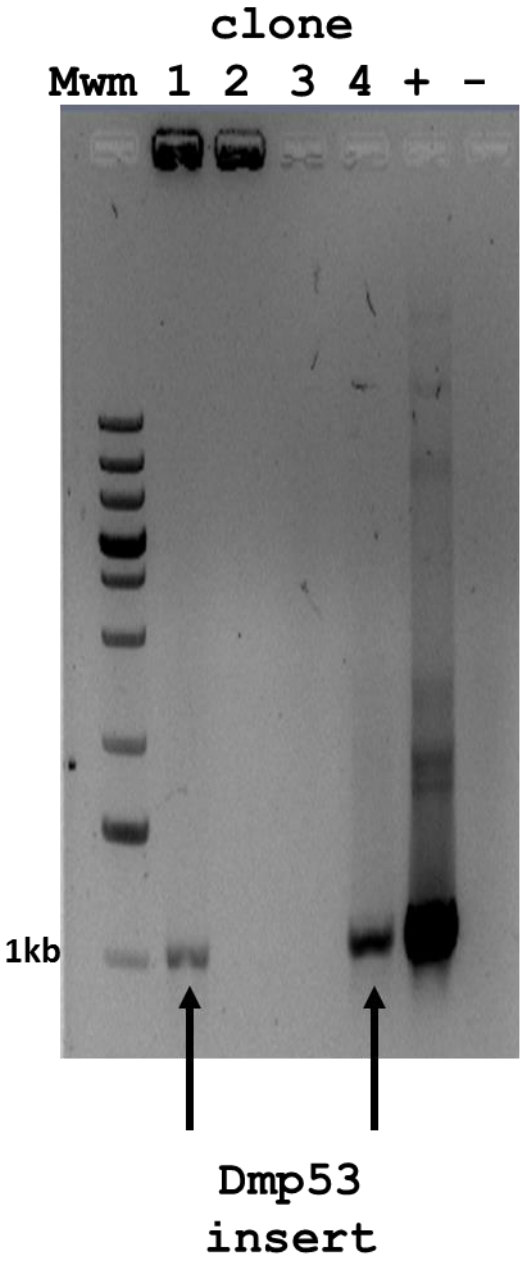


Figure 13: Colony PCR analysis of transformed clones to detect clones with Dmp53 insert, clone 1 and 4 show the presence of insert, correlating to the positive control (+), a PCR amplification of Dmp53 from bacterial plasmid. The negative control (-), pTG1895 without insert, shows no amplification as expected.

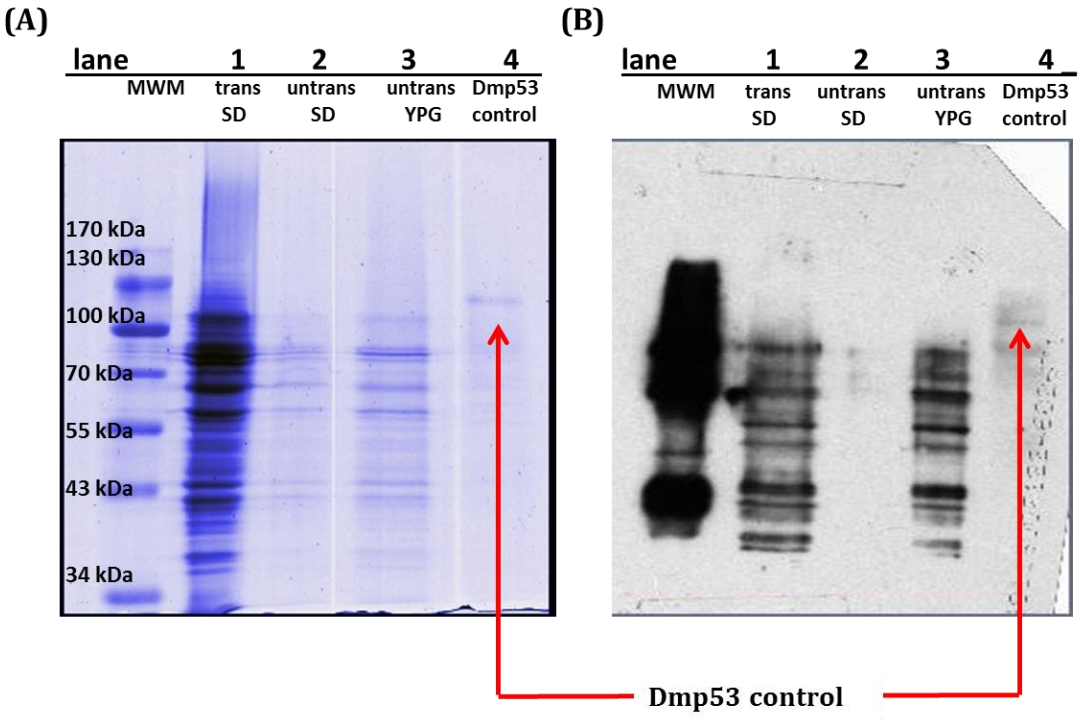


Figure 14: Crude extracts from Protein expressions of transformed and untransformed yeast concentrated by stirred cell, and Western blot thereof. SDS analysis after coomassie blue staining, shows that transformed yeast cultured in S Synthetic minimal (SD) media is expressing protein, lane 2. Untransformed yeast cultured in SD media (lane3) and YPG media (lane4) also express some protein, but far less as compared to Transformed yeast. Lanes 1 contained the molecular weight marker. Lane 5 contained Dmp53 purified from bacterial protein expression as a control. Western blot analysis shows non-specific binding of the Dmp53 antibody. This experiment cannot positively confirm that Dmp53 is being expressed and that it is being secreted.

Upon experimentation, it was found that the yeast was expressing Dmp53 in very low concentrations. SDS PAGE gels of proteins from expression cultures, some in excess of 1 litre, showed faint bands, or none whatsoever, after staining. Expressed proteins only became visible on gels after TCA precipitation (Figure 15b). Also the yeast expression system was retaining the protein in the cytosol as opposed to exporting it, as per the experimental design, analyses by SDS PAGE showed expressed Dmp53 in cytosolic fractions (Figure 15b) and not in the media (Figure 15a).

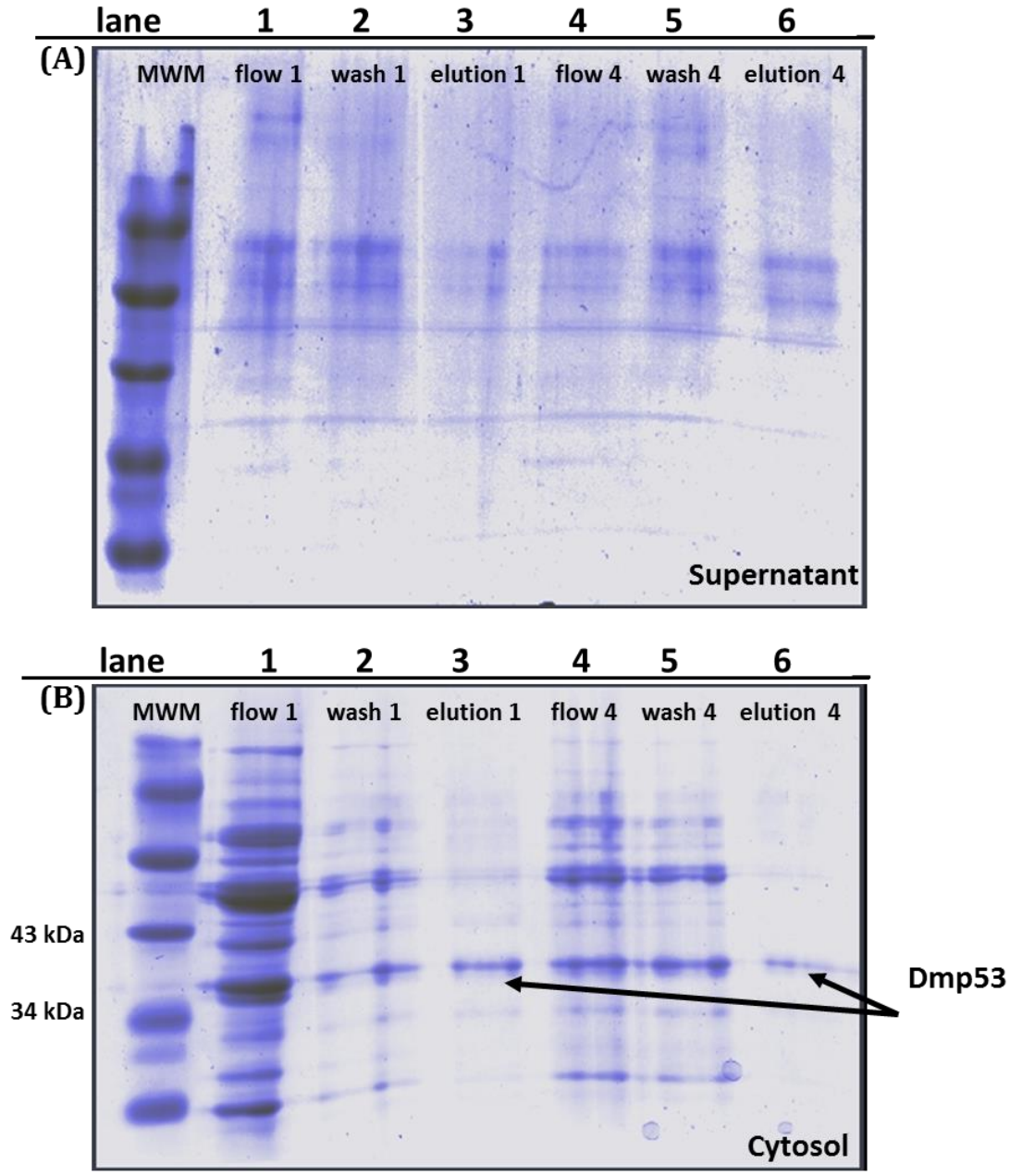


Figure 15: SDS PAGE to visualise Dmp53. Cultures of clones 1 and 4 were grown. Supernatant (A) and mechanically lysed yeast pellet (B), were HIS purified, and then concentrated by TCA precipitation. After coomassie blue staining, no Dmp53 is visible in the supernatant (A), proving retention of the protein. (B) shows that both clones 1 and 4 have retained Dmp53 in their cytosol.

TCA precipitation, causes denaturation of proteins, therefore proteins were expressed and purified again. In an attempt to increase protein concentration of the expression, a protease inhibitor cocktail was included in the culture media. Proteins were expressed in 50ml cultures grown at room temperature for five days. After expression, the media was kept aside and protein extracted from the cells by mechanical lysis using acid washed glass beads. Both the supernatant media and the cell extract were then purified by HIS chromatography. Figure 16 shows a successful purification and a suspected Dmp53 eluting in the cell fraction, again confirming the retention instead of excretion of the protein. The size of the protein is approximately 37kDa, which equates to an insert size of roughly 1020aa, equal to the cloned Dmp53 insert. This is also the same size protein that became visible after TCA precipitation, as shown in Figure 15b.

Once again Western blot analysis of the gel could not positively confirm the identity of Dmp53, due to non-specific binding, as seen in Figure 17. The Dmp53 antibody binds to most proteins present including the molecular weight marker.

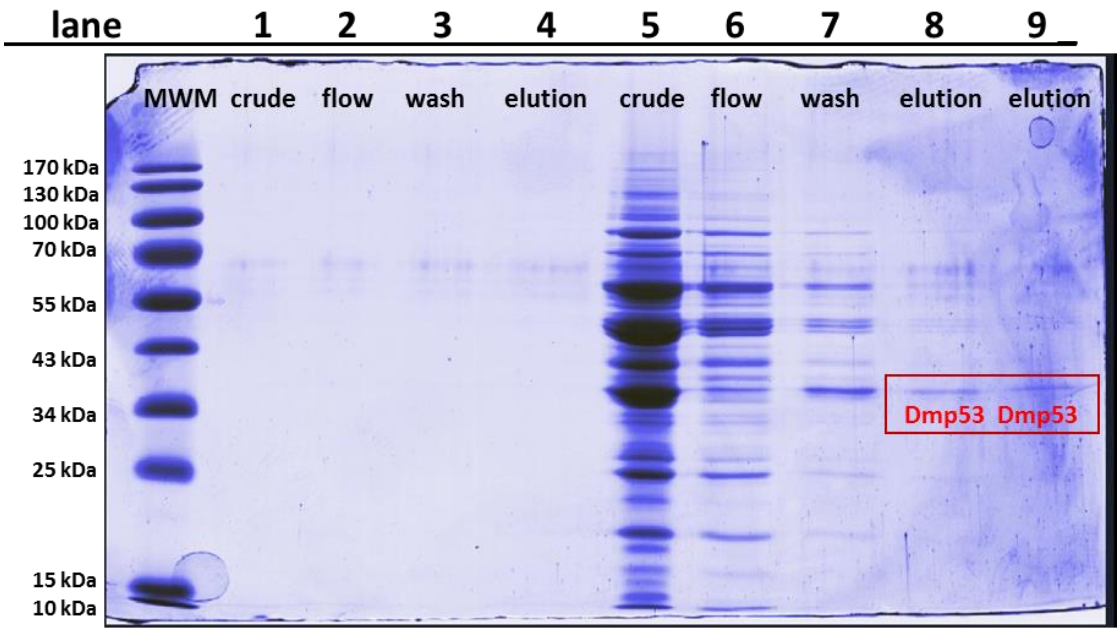


Figure 16: Yeast expression with the inclusion of protease inhibitors. Dmp53 is visible in elution fractions of the coomassie blue stained SDS PAGE gel. Lane 1-4 contain the media supernatant after purification by HIS chromatography, lanes 5-9 contain proteins extracted from the cell pellet and purified by HIS chromatography (crude extract, flow-through, wash and elution fractions respectively).

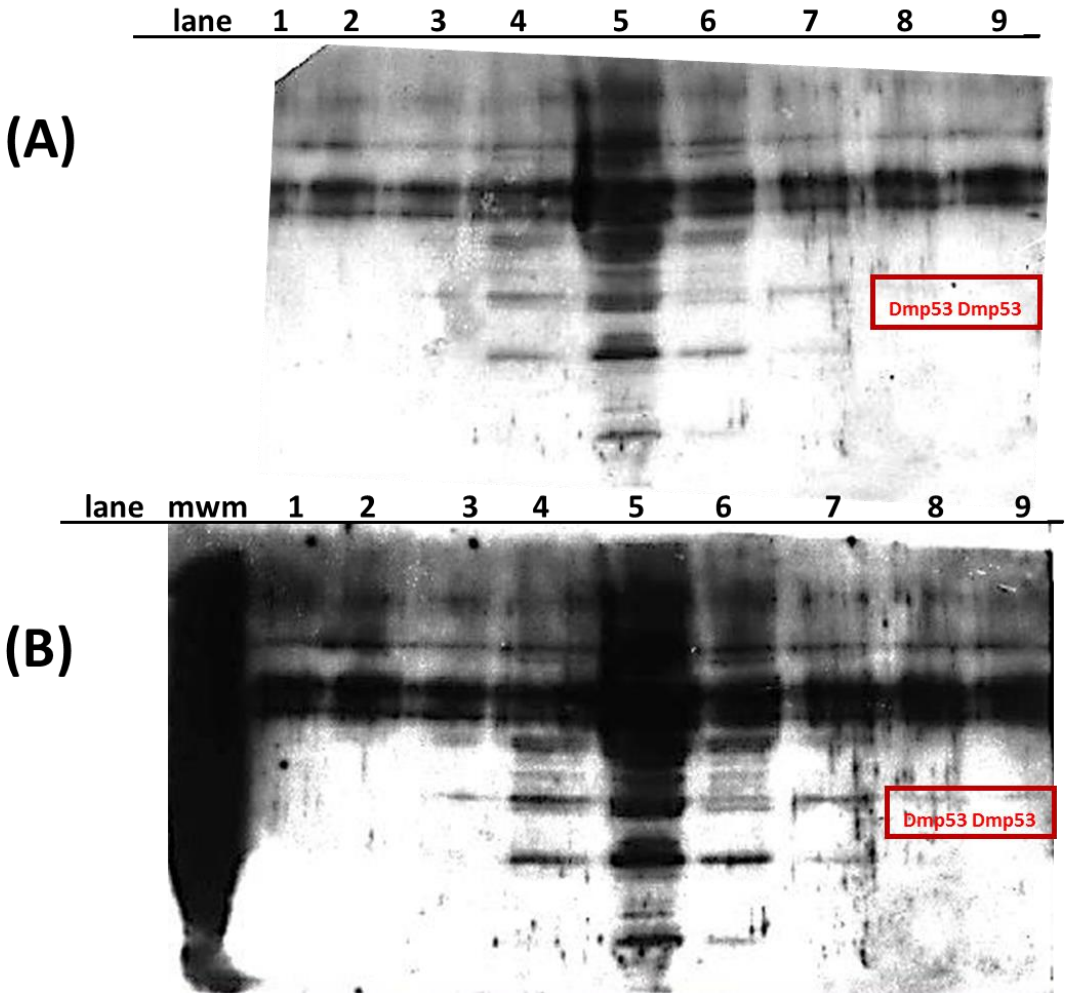


Figure 17: Non-specific binding of Dmp53 antibodies during western blot analysis on successive exposures of membrane to film. The above exposure times were 5 seconds, and 10 seconds respectively.

3.4 Gene structure and isoforms of SNAMA

A recent update of the flybase genome database theoretically identified a second *SNAMA* transcript, referred to as *SNAMA B*. The *SNAMA* transcripts have 5′ UTRs which start at different sites and share exons 1 through 4. There is a minor difference in exon 5 where *SNAMA B*, the shorter transcript, stops and *SNAMA A* continues for a few base pairs. *SNAMA A* has a further 3 exons after this (Table 12). The *SNAMA* gene structure is depicted in Figure 18.

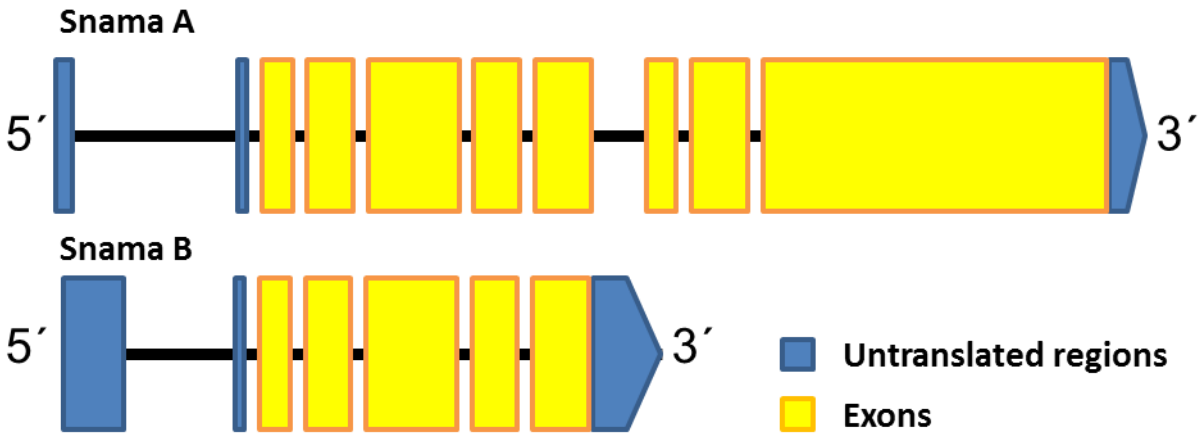


Figure 18: A schematic overview of the gene structure of the SNAMA A and B isoforms. Untranslated gene regions are shown in blue while exons are shown in yellow. SNAMA A and B share 5 exons. SNAMA A, the full length SNAMA isoform has a further 3 exons (adapted from flybase (St Pierre et al., 2014)).

Table 12: A Comparison of the *SNAMA* A and B gene structure. After bioinformatic analysis of the *SNAMA* A and B gene structure and sequence results were tabulated to highlight differences and commonalities. Differences found are shown in red, and common introns and exons are shown in yellow.

	RA		RB	
5′UTR	**4129-4225**		**4197-4634**	5′ UTR
-				-
5′UTR	5055-5073		5055-5073	5′UTR
Exon	5074-5239	165	5074-5239	Exon
Intron				Intron
Exon	5299-5533	234	5299-5533	Exon
Intron				Intron
Exon	5588-6066	478	5588-6066	Exon
Intron				Intron
Exon	6127-6401	274	6127-6401	Exon
Intron				Intron
Exon	**6465-6790**	**325/329**	**6465-6794**	Exon
Intron			6795-7035	3′UTR
Exon	7079-7228	209		
Intron				
Exon	7289-7582	293		
Intron				
Exon	7645-9415	1770		
3′UTR	9416-9537			
	Differences			
	Common introns and exons			

After bioinformatic analysis of the *SNAMA* gene region and translation, we hypothesize that transcript B arises from alternative splicing. The *SNAMA* A&B transcripts occur in shifted open reading frames through alternative splicing introducing a stop codon before exon 6 (Figure 19). The transcripts also differ in their 5´ UTR start sites (Figure 20). Figure 20 shows the aligned amino acid sequences of SNAMA A and SNAMA B highlighting conserved functional domains.

```
1      TCGGAGTTTTTATGTGCTTGCAGAGTAGTGTAAGTAATTGCTATTTTTGTGCCAGCTGCT

61     TGCCGGTAGTAAACAACGCGCGTGAACTGTGATGCCTTGCAATCCTCAGCGCAATTCAAG

121    TGCGGTTTAAGTCCGAACGACCCAAGAAACCCGAAGTGTACGTACATATGTATGTTTTCG

181    CGTGTGTATTTTTATGCACATGCAATTCTATGCTGCTGCCATGCAAAGGTGCAAAAAATG

241    CGAAATGCAAATATCGTTCACAGAATGCCCATTGATCTGACCGGATTTTAGGTGGCACAT

301    ACGCATATGCATATAGCATGTTCCCGGAATTGTGCAACTTCGCATACATACATATGTAAT

361    GTCTTATGTTGGCATAATTATCCGCTGTTCAAGGGAAGTGCATGGTGCAAATCGGGAATC

421    TATATTCGGCATACTTTGATTTCTTTCATATGCAACTATGTCGGTACACTATAAATTTAA
1                                                  M  S  V  H  Y  K  F  K

481    GAGTACACTCAACTTTGATACAATTACTTTTGATGGACTTCACATTTCTGTCGGGGACTT
9        S  T  L  N  F  D  T  I  T  F  D  G  L  H  I  S  V  G  D  L

541    AAAAAGGGAGATTGTGCAGCAGAAGCGACTGGGCAAAATCATCGACTTTGATCTCCAAAT
29       K  R  E  I  V  Q  Q  K  R  L  G  K  I  I  D  F  D  L  Q  I

601    AACAAATGCGCAGAGTAAAGAAGAATACAAGGACGATGGGTTCCTTATTCCCAAAAACAC
49       T  N  A  Q  S  K  E  E  Y  K  D  D  G  F  L  I  P  K  N  T

661    AACGCTGATCATATCGCGCATCCCCATCGCCCATCCCACAAAAAAGGGCTGGGAGCCACC
69       T  L  I  I  S  R  I  P  I  A  H  P  T  K  K  G  W  E  P  P

721    AGCAGCAGAAAATGCCTTTTCGGCGGCGCCTGCCAAGCAGGACAACTTCAACATGGACCT
89       A  A  E  N  A  F  S  A  A  P  A  K  Q  D  N  F  N  M  D  L

781    GTCCAAAATGCAAGGCACGGAGGAGGACAAAATCCAGGCCATGATGATGCAGAGCACAGT
109      S  K  M  Q  G  T  E  E  D  K  I  Q  A  M  M  M  Q  S  T  V

841    CGACTATGATCCTAAGACGTACCATCGTATTAAAGGACAATCGCAAGTGGGAGAAGTTCC
129      D  Y  D  P  K  T  Y  H  R  I  K  G  Q  S  Q  V  G  E  V  P

901    CGCATCCTACCGATGCAACAAATGCAAGAAAAGCGGACACTGGATCAAAAACTGTCCCTT
149      A  S  Y  R  C  N  K  C  K  K  S  G  H  W  I  K  N  C  P  F

961    TGTGGGGGGAAAGGACCAGCAAGAGGTCAAACGGAATACTGGTATTCCGCGGTCTTTCCG
169      V  G  G  K  D  Q  Q  E  V  K  R  N  T  G  I  P  R  S  F  R

1021   CGACAAGCCAGATGCGGCTGAGAACGAATCAGCCGATTTTGTGCTGCCTGCTGTACAAAA
189      D  K  P  D  A  A  E  N  E  S  A  D  F  V  L  P  A  V  Q  N

1081   CCAAGAGATACCGGAGGATCTGATATGCGGCATATGCCGAGATATATTCGTCGATGCTGT
209      Q  E  I  P  E  D  L  I  C  G  I  C  R  D  I  F  V  D  A  V

1141   CATGATACCCTGCTGCGGAAGTTCCTTTTGTGACGACTGTGTGCGAACCTCCTTATTGGA
229      M  I  P  C  C  G  S  S  F  C  D  D  C  V  R  T  S  L  L  E


1201   GTCAGAGGATAGTGAGTGCCCCGATTGCAAGGAGAAGAACTGTTCGCCTGGCTCCCTGAT
249      S  E  D  S  E  C  P  D  C  K  E  K  N  C  S  P  G  S  L  I
```

```
1261  ACCTAATCGGTTCTTGAGGAATTCGGTGAACGCCTTTAAAAATGAGACTGGGTATAACAA
269     P  N  R  F  L  R  N  S  V  N  A  F  K  N  E  T  G  Y  N  K

1321  AAGCGCGGCTAAGCCAGCTGCAGTAAAAAATGAGGAAAAACCTCCTGTTGAAAAAGAAGT
289     S  A  A  K  P  A  A  V  K  N  E  E  K  P  P  V  E  K  E  V

1381  GGAGAAAAAGCCAGTCGCGGAGGTGGAACCCGAAGAGACTGAGGTGAAACCTGAAAAGCA
309     E  K  K  P  V  A  E  V  E  P  E  E  T  E  V  K  P  E  K  Q

1441  AAAAGAATCCGAAACCAATGGCAGTAATCCGCCAAAATCGGAATCTCCAGAGCCTCCCGC
329     K  E  S  E  T  N  G  S  N  P  P  K  S  E  S  P  E  P  P  A

1501  AACCACAGAACCATCACAGAAGGAGAAAGATAAATATGATTCAGACTACGAGGATAACAT
349     T  T  E  P  S  Q  K  E  K  D  K  Y  D  S  D  Y  E  D  N  I

1561  TACCATAAAAATGCCCCAGCCTGCAGCTGATTCTACAACAGTGCCCAGCAAAAGATCCCC
369     T  I  K  M  P  Q  P  A  A  D  S  T  T  V  P  S  K  R  S  P

1621  CAGTTATTCCCACAGAAGTGAATCCTCTCATCGACGGGACAGGTCGGATTATGTTTCCGA
389     S  Y  S  H  R  S  E  S  S  H  R  R  D  R  S  D  Y  V  S  D

1681  TCACGATCACAAGCACCAACGTCCATCAAAATCGGAGTCTGTTAACAAGGATCGCAGTCT
409     H  D  H  K  H  Q  R  P  S  K  S  E  S  V  N  K  D  R  S  L

1741  CCTGCCCTTGCCCATTGGCACCCTGCCTAGCTACCAGGGCCACATGATGGCCGAATCAGA
429     L  P  L  P  I  G  T  L  P  S  Y  Q  G  H  M  M  A  E  S  E

1801  AGAAGCTCGTCGATCGAGTGCCTATAAGCCCCCTTATATGCAAATGCAGCGGGGCCCACC
449     E  A  R  R  S  S  A  Y  K  P  P  Y  M  Q  M  Q  R  G  P  P

1861  TCCTATGCACATGATGAGTCACCACATGCCAGCCTACAACAACGGGTTTAACAACATGGG
469     P  M  H  M  M  S  H  H  M  P  A  Y  N  N  G  F  N  N  M  G

1921  ACAGAGGCCTCCCCTCAGGTAGGTTTCTTTTCCACCCGTCCGGATCCCGCTACCATGTGG
489     Q  R  P  P  L  R  *

1981  ATCGAGTCAAGTGTCGAGGCCACTGGAACTCCGAGGTGTTGACGAGTCCATACGCAGCCG

2041  GTCCTGCCTGATCTCAATATTTTTAAGTTCCAACAGAGACTATACACACACACTAGTTAA

2101  GAGTAATTTAAGCTCTGGGCTCTGCGCCTTTATGCTTTGTTGAATATTGGTTATTCCTAA

2161  TTAAAACAGTATTGCTTACAATC
```

Figure 19: SNAMA B sequence with translation. The ORF appears in translation frame 2. Exons shared with SNAMA A (Figure 20) are highlighted in green. The stop codon is highlighted in red. This is the alternative splice site where the addition of further exons gives rise to SNAMA A.

```
1     GCATTTCACATCTCTGGGGCTTTGGCGTCACGTTCGCATCTCTAGTAAATTGGAAAAAAA

61    TAAAATCGTCGGAGTTTTTATGTGCTTGCAGAGTAGTATTTCTTTCATATGCAACTATGT
1                                                               M  S

121   CGGTACACTATAAATTTAAGAGTACACTCAACTTTGATACAATTACTTTTGATGGACTTC
3        V  H  Y  K  F  K  S  T  L  N  F  D  T  I  T  F  D  G  L  H

181   ACATTTCTGTCGGGGACTTAAAAAGGGAGATTGTGCAGCAGAAGCGACTGGGCAAAATCA
23       I  S  V  G  D  L  K  R  E  I  V  Q  Q  K  R  L  G  K  I  I

241   TCGACTTTGATCTCCAAATAACAAATGCGCAGAGTAAAGAAGAATACAAGGACGATGGGT
43       D  F  D  L  Q  I  T  N  A  Q  S  K  E  E  Y  K  D  D  G  F

301   TCCTTATTCCCAAAAACACAACGCTGATCATATCGCGCATCCCCATCGCCCATCCCACAA
63       L  I  P  K  N  T  T  L  I  I  S  R  I  P  I  A  H  P  T  K

361   AAAAGGGCTGGGAGCCACCAGCAGCAGAAAATGCCTTTTCGGCGGCGCCTGCCAAGCAGG
83       K  G  W  E  P  P  A  A  E  N  A  F  S  A  A  P  A  K  Q  D

421   ACAACTTCAACATGGACCTGTCCAAAATGCAAGGCACGGAGGAGGACAAAATCCAGGCCA
103      N  F  N  M  D  L  S  K  M  Q  G  T  E  E  D  K  I  Q  A  M

481   TGATGATGCAGAGCACAGTCGACTATGATCCTAAGACGTACCATCGTATTAAAGGACAAT
123      M  M  Q  S  T  V  D  Y  D  P  K  T  Y  H  R  I  K  G  Q  S

541   CGCAAGTGGGAGAAGTTCCCGCATCCTACCGATGCAACAAATGCAAGAAAAGCGGACACT
143      Q  V  G  E  V  P  A  S  Y  R  C  N  K  C  K  K  S  G  H  W

601   GGATCAAAAACTGTCCCTTTGTGGGGGGAAAGGACCAGCAAGAGGTCAAACGGAATACTG
163      I  K  N  C  P  F  V  G  G  K  D  Q  Q  E  V  K  R  N  T  G

661   GTATTCCGCGGTCTTTCCGCGACAAGCCAGATGCGGCTGAGAACGAATCAGCCGATTTTG
183      I  P  R  S  F  R  D  K  P  D  A  A  E  N  E  S  A  D  F  V

721   TGCTGCCTGCTGTACAAAACCAAGAGATACCGGAGGATCTGATATGCGGCATATGCCGAG
203      L  P  A  V  Q  N  Q  E  I  P  E  D  L  I  C  G  I  C  R  D

781   ATATATTCGTCGATGCTGTCATGATACCCTGCTGCGGAAGTTCCTTTTGTGACGACTGTG
223      I  F  V  D  A  V  M  I  P  C  C  G  S  S  F  C  D  D  C  V

841   TGCGAACCTCCTTATTGGAGTCAGAGGATAGTGAGTGCCCCGATTGCAAGGAGAAGAACT
243      R  T  S  L  L  E  S  E  D  S  E  C  P  D  C  K  E  K  N  C

901   GTTCGCCTGGCTCCCTGATACCTAATCGGTTCTTGAGGAATTCGGTGAACGCCTTTAAAA
263      S  P  G  S  L  I  P  N  R  F  L  R  N  S  V  N  A  F  K  N

961   ATGAGACTGGGTATAACAAAAGCGCGGCTAAGCCAGCTGCAGTAAAAAATGAGGAAAAAC
283      E  T  G  Y  N  K  S  A  A  K  P  A  A  V  K  N  E  E  K  P

1021  CTCCTGTTGAAAAAGAAGTGGAGAAAAAGCCAGTCGCGGAGGTGGAACCCGAAGAGACTG
303      P  V  E  K  E  V  E  K  K  P  V  A  E  V  E  P  E  E  T  E

1081  AGGTGAAACCTGAAAAGCAAAAAGAATCCGAAACCAATGGCAGTAATCCGCCAAAATCGG
323      V  K  P  E  K  Q  K  E  S  E  T  N  G  S  N  P  P  K  S  E

1141  AATCTCCAGAGCCTCCCGCAACCACAGAACCATCACAGAAGGAGAAAGATAAATATGATT
343      S  P  E  P  P  A  T  T  E  P  S  Q  K  E  K  D  K  Y  D  S
```

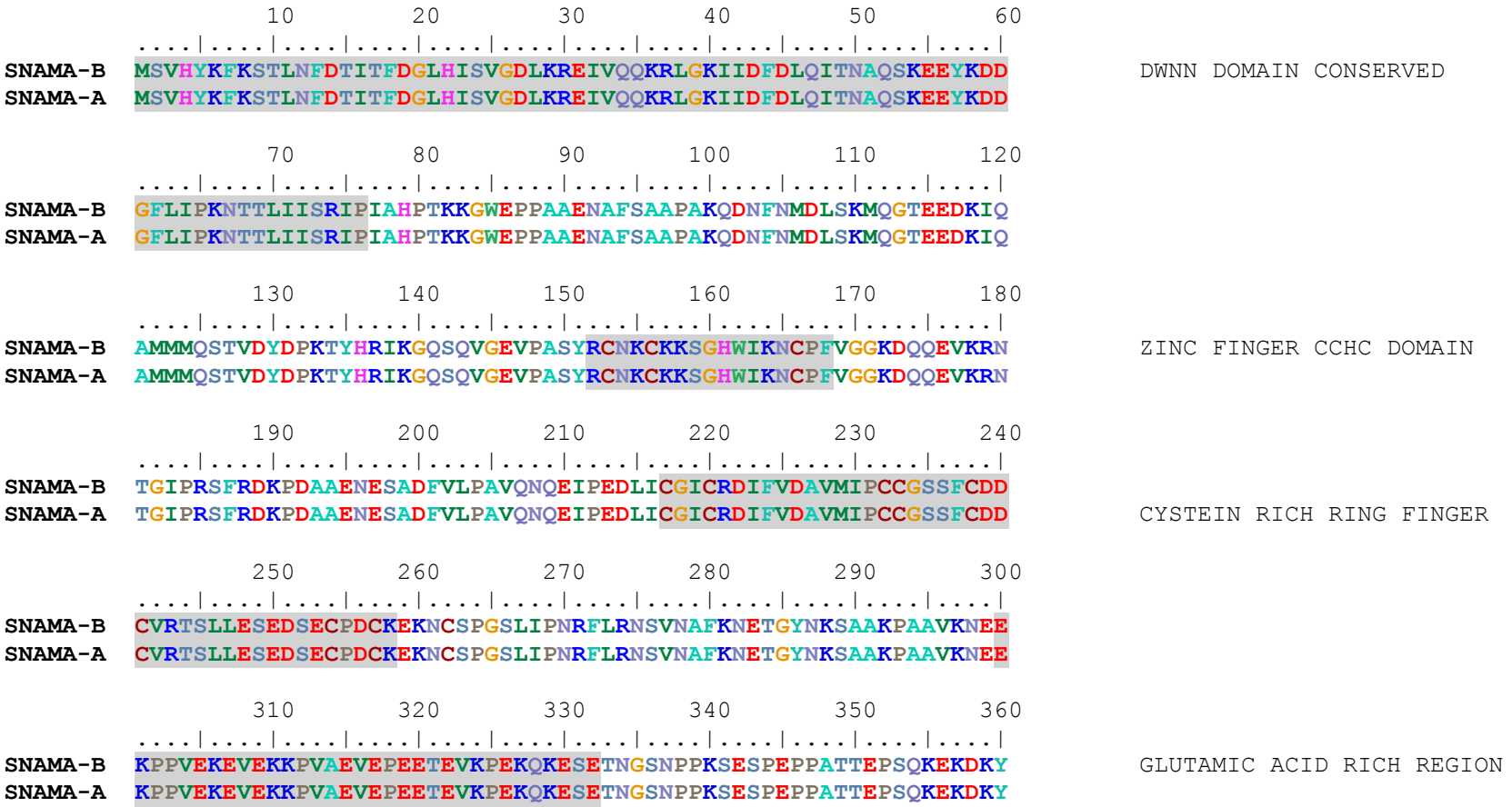
```
1201 CAGACTACGAGGATAACATTACCATAAAAATGCCCCAGCCTGCAGCTGATTCTACAACAG
363     D  Y  E  D  N  I  T  I  K  M  P  Q  P  A  A  D  S  T  T  V

1261 TGCCCAGCAAAAGATCCCCCAGTTATTCCCACAGAAGTGAATCCTCTCATCGACGGGACA
383     P  S  K  R  S  P  S  Y  S  H  R  S  E  S  S  H  R  R  D  R

1321 GGTCGGATTATGTTTCCGATCACGATCACAAGCACCAACGTCCATCAAAATCGGAGTCTG
403     S  D  Y  V  S  D  H  D  H  K  H  Q  R  P  S  K  S  E  S  V

1381 TTAACAAGGATCGCAGTCTCCTGCCCTTGCCCATTGGCACCCTGCCTAGCTACCAGGGCC
423     N  K  D  R  S  L  L  P  L  P  I  G  T  L  P  S  Y  Q  G  H

1441 ACATGATGGCCGAATCAGAAGAAGCTCGTCGATCGAGTGCCTATAAGCCCCCTTATATGC
443     M  M  A  E  S  E  E  A  R  R  S  S  A  Y  K  P  P  Y  M  Q

1501 AAATGCAGCGGGGCCCACCTCCTATGCACATGATGAGTCACCACATGCCAGCCTACAACA
463     M  Q  R  G  P  P  P  M  H  M  M  S  H  H  M  P  A  Y  N  N

1561 ACGGGTTTAACAACATGGGACAGAGGCCTCCCCTCAGCTATGTGCCGTATCAAAACCAAT
483     G  F  N  N  M  G  Q  R  P  P  L  S  Y  V  P  Y  Q  N  Q  S

1621 CCGTACACCCAATGCGTGCGCCGTACGGATCTGCAGGCGGAGGTATGAATATGAATATGT
503     V  H  P  M  R  A  P  Y  G  S  A  G  G  G  M  N  M  N  M  S

1681 CACAACCATTTCAGTCCCCAAATTTAGCCTCGATATACCAAGGGGTGGCAGCGAAGGTCG
523     Q  P  F  Q  S  P  N  L  A  S  I  Y  Q  G  V  A  A  K  V  G

1741 GTTCCGGTCCCATTGACGATCCGTTGGAGGCCTTCAATCGCATCATGAAGGAGAAGGAGC
543     S  G  P  I  D  D  P  L  E  A  F  N  R  I  M  K  E  K  E  R

1801 GGAAGAAGGTGGACCGCTTTCGAAGCTCTGACCGCCACAGGTCAAGGTCCCCGGATAGAC
563     K  K  V  D  R  F  R  S  S  D  R  H  R  S  R  S  P  D  R  Q

1861 AAAGGCACCGCTTTAAGTCTCCCATGTACGAAAAGGACAACTCCAGGGATAATCTCAAGG
583     R  H  R  F  K  S  P  M  Y  E  K  D  N  S  R  D  N  L  K  D

1921 ACAAAAGACCGCGATCCCGGGAAAGGAAGCGAGAACATAGCTACGAACGGCATATACGCC
603     K  R  P  R  S  R  E  R  K  R  E  H  S  Y  E  R  H  I  R  H

1981 ACCCTCGTTCTAGTCGCCAGCCGAATGATGGCTCTAAGTCCCCAGGTGGCAGAATCAAAA
623     P  R  S  S  R  Q  P  N  D  G  S  K  S  P  G  G  R  I  K  R

2041 GATCTGGACATCGTCGCTCTGCATCTCCAAAGCCGGGCTACAAGAGTGATTACAGAGACA
643     S  G  H  R  R  S  A  S  P  K  P  G  Y  K  S  D  Y  R  D  K

2101 AGCCGTACAACAAGCCTAGTGCTCCCAAAACGGAGGCAGTTGAGCCTCCTCCCCCCGGAT
663     P  Y  N  K  P  S  A  P  K  T  E  A  V  E  P  P  P  P  G  F

2161 TCGAGCCGTTGCAGCTGACGGATGAAGACGGCTACAGGAACAAGCACCCGACCAGTTCGG
683     E  P  L  Q  L  T  D  E  D  G  Y  R  N  K  H  P  T  S  S  E

2221 AAGCATCACAAAGCAGCAAGGGTGATAGCAGCAAGAAGAGAGGGGAAAACAGGCACGAAG
703     A  S  Q  S  S  K  G  D  S  S  K  K  R  G  E  N  R  H  E  E

2281 AGGCGCCACGAAAGAGGCACAGGTCTCGCAGCATTAGCAAGGAACCGAAGCCGAATGACA
723     A  P  R  K  R  H  R  S  R  S  I  S  K  E  P  K  P  N  D  S
```

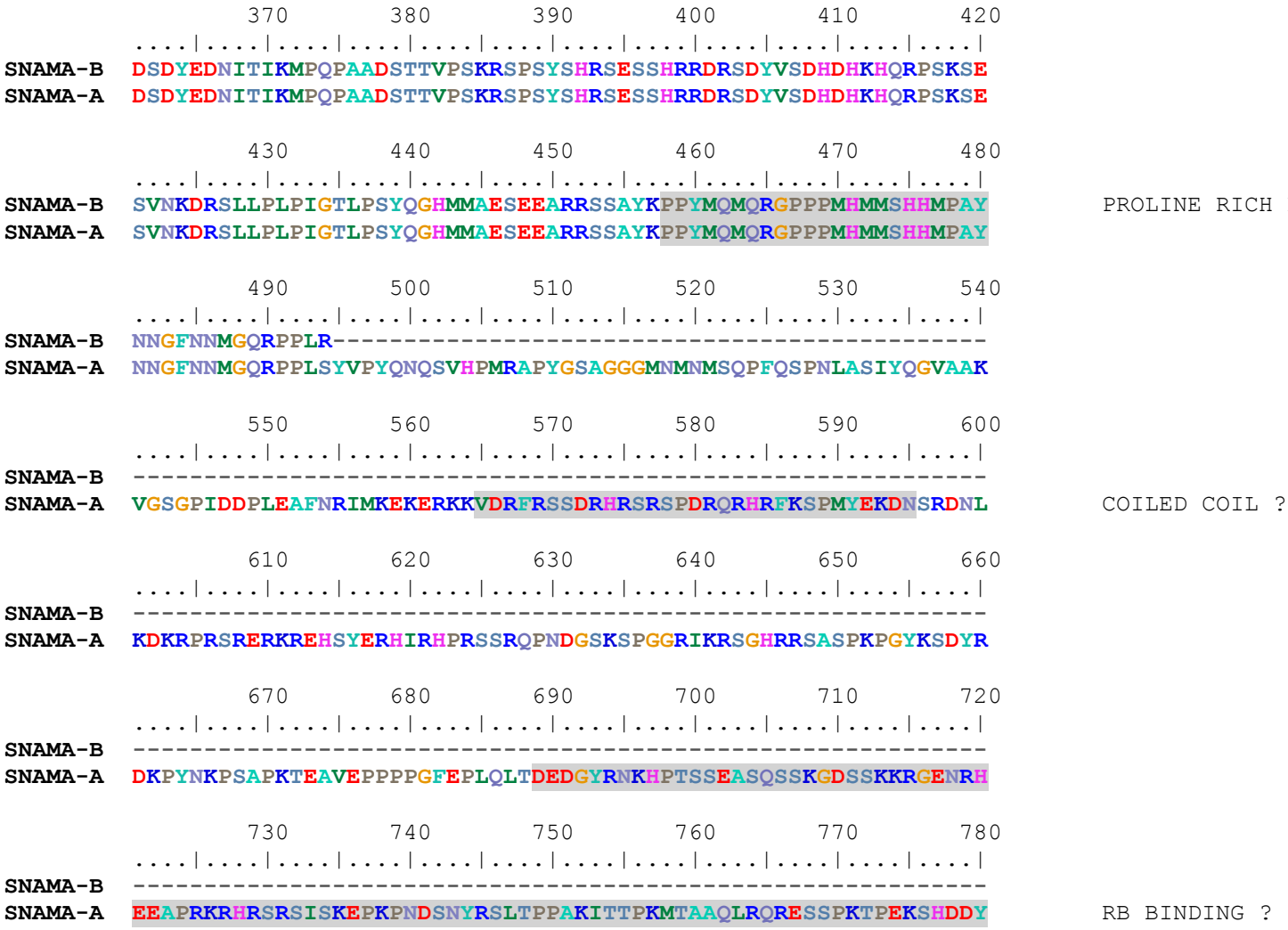
```
2341 GCAACTACAGGAGCCTGACTCCACCAGCAAAGATCACCACACCGAAAATGACTGCTGCCC
743    N  Y  R  S  L  T  P  P  A  K  I  T  T  P  K  M  T  A  A  Q

2401 AGTTGAGGCAACGCGAAAGTTCACCGAAGACGCCGGAAAAGAGTCACGACGATTATCTGA
763    L  R  Q  R  E  S  S  P  K  T  P  E  K  S  H  D  D  Y  L  T

2461 CCGCGAAGGCCAGAATTATGGCCTCCCAGCCCGTCATCAACGACACGGAAATGGAGACCA
783    A  K  A  R  I  M  A  S  Q  P  V  I  N  D  T  E  M  E  T  N

2521 ATGTGGGCAAGGAGAACAAGGCCAAGAGTCCGTTGTCAAAAGATCGCAAGAAGAAGAAGA
803    V  G  K  E  N  K  A  K  S  P  L  S  K  D  R  K  K  K  K  K

2581 AGGACAAGGACAAGGCTGAGCGCAAGAAAAACAAGAAGGACAAGCGCGCTAAGAAGGAGA
823    D  K  D  K  A  E  R  K  K  N  K  K  D  K  R  A  K  K  E  K

2641 AAGGGGATCGCCAGAAGAAGAGCTCCTCAGTTAATCGATCTGACTCGGATATTAACAACA
843    G  D  R  Q  K  K  S  S  S  V  N  R  S  D  S  D  I  N  N  S

2701 GCTCACTAATGAACGAGTCAAATTATAAAGTATTGTCTCCCAGGGCTCAAAGTCCCAGCA
863    S  L  M  N  E  S  N  Y  K  V  L  S  P  R  A  Q  S  P  S  I

2761 TTGAGATCAATGCGGCTCAACTTTCCCCTACTCACAACGCTACTGAAAACGTTAATCCGA
883    E  I  N  A  A  Q  L  S  P  T  H  N  A  T  E  N  V  N  P  K

2821 AGAGTCATTCCATCCTTACTGTGGGTGCTGCTAGCGACGATAATCTTGGCCCAAGAAGCA
903    S  H  S  I  L  T  V  G  A  A  S  D  D  N  L  G  P  R  S  K

2881 AACTCAGCGAGGCTAATTCTGTCAATCTATCCAAATGGGAAATCGACGAGAATATCTTAG
923    L  S  E  A  N  S  V  N  L  S  K  W  E  I  D  E  N  I  L  G

2941 GTTTGGAGGATTCCTCCAAAAAAGCTGCCGGGGCCTCCGACGATCCGTCGGAAATAACTT
943    L  E  D  S  S  K  K  A  A  G  A  S  D  D  P  S  E  I  T  S

3001 CAGACGTCCTGCGAAAGGCTGAGAACGCAATATTTGCAAAGGCTATTAATGCCATCAGGC
963    D  V  L  R  K  A  E  N  A  I  F  A  K  A  I  N  A  I  R  P

3061 CTATGGAGTTTCAAGTTATTATCAATTCCAAGGACAACAGCAAGGACCGCTCCGTAGTTC
983    M  E  F  Q  V  I  I  N  S  K  D  N  S  K  D  R  S  V  V  R

3121 GAAGTGACAAGGATCGCTCCTCCTCACCCAGGCGTAACAACAGCAGCAGGTCGGTAAAGG
1003   S  D  K  D  R  S  S  S  P  R  R  N  N  S  S  R  S  V  K  D

3181 ATAGGCTGGGCACCAAGATTTCCAATGATAGAAGCCGTTCGCGAGACAAGTCGAAGGGCA
1023   R  L  G  T  K  I  S  N  D  R  S  R  S  R  D  K  S  K  G  R

3241 GGCGCCGGGCCGCAAGGAGCTCCGACGACGATGCGAACCGCGGCAGGTCGGATCGTCATG
1043   R  R  A  A  R  S  S  D  D  D  A  N  R  G  R  S  D  R  H  G

3301 GCAGCCGGAAGAGGGACAACAGATCCCGCGACAGGGCGGCGCCTTCAGAGAAGAGGCAGG
1063   S  R  K  R  D  N  R  S  R  D  R  A  A  P  S  E  K  R  Q  E

3361 AGCGTTCGTACAAGCGAAGCTCGCCGGAGGACGACAAGCTGAGGCGCCAGAACAAGGAAC
1083   R  S  Y  K  R  S  S  P  E  D  D  K  L  R  R  Q  N  K  E  Q

3421 AGTCCGAATCCAAGCACGGAAAGCATGATCAAAACAATAGCGACGACTCGGATCGGAGGG
1103   S  E  S  K  H  G  K  H  D  Q  N  N  S  D  D  S  D  R  R  A
```

```
3481 CGGCCAAAAACACCAAGTCCAGCGACAGCCGAGTGGTCTCCTCTGTAACAGCCGTGGTTG
1123    A  K  N  T  K  S  S  D  S  R  V  V  S  S  V  T  A  V  V  A

3541 CTCCTCCCAAACCCTGTCGTCCAGACAACCCGTTCCGCAAGTTCGTCGACACCAGTTCGT
1143    P  P  K  P  C  R  P  D  N  P  F  R  K  F  V  D  T  S  S  S


3601 CGAGCAGCTTAGTTGTAAAATATGATAACACGATACAGAAGGAGGGCGCGTCCTCGGACA
1163    S  S  L  V  V  K  Y  D  N  T  I  Q  K  E  G  A  S  S  D  N

3661 ACGGCATGGAGCACAGGAAGCAGAGGGATAAGAAGCTGAAGAAACATTCAAAATATTCGT
1183    G  M  E  H  R  K  Q  R  D  K  K  L  K  K  H  S  K  Y  S  S

3721 CAACCGATTCGTTGAAGAGCGAGAAGCGCAAGGATCCGAAGAGCAAAAAGAAGAGCAAGA
1203    T  D  S  L  K  S  E  K  R  K  D  P  K  S  K  K  K  S  K  I

3781 TTTTGAAGAAGAAGAAAAAATCAAAGAAGTAGGTTACGGTAGGCTACGAGATAAGAATGA
1223    L  K  K  K  K  K  S  K  K  *

3841 TATAAATATTGAAGATTAATGTGTACAAATCAAAGATTTTTAATGTATGTATTTATCATG

3901 CAACTATAAGTATACAAATAAAACAGAACTACTC
```

Figure 20: SNAMA A sequence with translation. The ORF appears in translation frame 3. Exons shared with SNAMA B (Figure 19) are highlighted in yellow. The alternative splice site is highlighted in magenta, and the stop codon in red.

On alignment of protein sequences of SNAMA A and B with Human RBBP6 and DWNN, we can map the presence of possible functional domains in the isoforms. Although aligning the sequences does not show high conservation of residues across species, certain key residues are present with some variation. Figure 21 highlights the probable locations of functional domains on SNAMA A and B. Figure 22 presents the possible domain structure of these isoforms in a schematic overview.


10 20 30 40 50 60
SNAMA-B MSVHYKFKSTLNFDTITFDGLHISVGDLKREIVQQKRLGKIIDFDLQITNAQSKEEYKDD
SNAMA-A MSVHYKFKSTLNFDTITFDGLHISVGDLKREIVQQKRLGKIIDFDLQITNAQSKEEYKDD
DWNN DOMAIN CONSERVED
70 80 90 100 110 120
SNAMA-B GFLIPKNTTLIISRIPIAHPTKKGWEPPAAENAFSAAPAKQDNFNMDLSKMQGTEEDKIQ
SNAMA-A GFLIPKNTTLIISRIPIAHPTKKGWEPPAAENAFSAAPAKQDNFNMDLSKMQGTEEDKIQ
130 140 150 160 170 180
SNAMA-B AMMMQSTVDYDPKTYHRIKGQSQVGEVPASYRCNKCKKSGHWIKNCPFVGGKDQQEVKRN
SNAMA-A AMMMQSTVDYDPKTYHRIKGQSQVGEVPASYRCNKCKKSGHWIKNCPFVGGKDQQEVKRN
ZINC FINGER CCHC DOMAIN
190 200 210 220 230 240
SNAMA-B TGIPRSFRDKPDAAENESADFVLPAVQNQEIPEDLICGICRDIFVDAVMIPCCGSSFCDD
SNAMA-A TGIPRSFRDKPDAAENESADFVLPAVQNQEIPEDLICGICRDIFVDAVMIPCCGSSFCDD
CYSTEIN RICH RING FINGER
250 260 270 280 290 300
SNAMA-B CVRTSLLESEDSECPDCKEKNCSPGSLIPNRFLRNSVNAFKNETGYNKSAAKPAAVKNEE
SNAMA-A CVRTSLLESEDSECPDCKEKNCSPGSLIPNRFLRNSVNAFKNETGYNKSAAKPAAVKNEE
310 320 330 340 350 360
SNAMA-B KPPVEKEVEKKPVAEVEPEETEVKPEKQKESETNGSNPPKSESPEPPATTEPSQKEKDKY
SNAMA-A KPPVEKEVEKKPVAEVEPEETEVKPEKQKESETNGSNPPKSESPEPPATTEPSQKEKDKY
GLUTAMIC ACID RICH REGION


370 380 390 400 410 420
SNAMA-B DSDYEDNITIKMPQPAADSTTVPSKRSPSYSHRSESSHRRDRSDYVSDHDHKHQRPSKSE
SNAMA-A DSDYEDNITIKMPQPAADSTTVPSKRSPSYSHRSESSHRRDRSDYVSDHDHKHQRPSKSE
430 440 450 460 470 480
SNAMA-B SVNKDRSLLPLPIGTLPSYQGHMMAESEEARRSSAYKPPYMQMQRGPPPMHMMSHHMPAY
SNAMA-A SVNKDRSLLPLPIGTLPSYQGHMMAESEEARRSSAYKPPYMQMQRGPPPMHMMSHHMPAY
PROLINE RICH REGION
490 500 510 520 530 540
SNAMA-B NNGFNNMGQRPPLR----------
SNAMA-A NNGFNNMGQRPPLSYVPYQNQSVHPMRAPYGSAGGGMNMNMSQPFQSPNLASIYQGVAAK
550 560 570 580 590 600
SNAMA-B ----------
SNAMA-A VGSGPIDDPLEAFNRIMKEKERKKVDRFRSSDRHRSRSPDRQRHRFKSPMYEKDNSRDNL
COILED COIL ?
610 620 630 640 650 660
SNAMA-B ----------
SNAMA-A KDKRPRSRERKREHSYERHIRHPRSSRQPNDGSKSPGGRIKRSGHRRSASPKPGYKSDYR
670 680 690 700 710 720
SNAMA-B ----------
SNAMA-A DKPYNKPSAPKTEAVEPPPPGFEPLQLTDEDGYRNKHPTSSEASQSSKGDSSKKRGENRH
730 740 750 760 770 780
SNAMA-B ----------
SNAMA-A EEAPRKRHRSRSISKEPKPNDSNYRSLTPPAKITTPKMTAAQLRQRESSPKTPEKSHDDY
RB BINDING ?

```
                790       800       810       820       830       840
           ....|....|....|....|....|....|....|....|....|....|....|....|
SNAMA-B    ------------------------------------------------------------
SNAMA-A    LTAKARIMASQPVINDTEMETNVGKENKAKSPLSKDRKKKKKDKDKAERKKNKKDKRAKK
                850       860       870       880       890       900
           ....|....|....|....|....|....|....|....|....|....|....|....|
SNAMA-B    ------------------------------------------------------------
SNAMA-A    EKGDRQKKSSSVNRSDSDINNSSLMNESNYKVLSPRAQSPSIEINAAQLSPTHNATENVN

                910       920       930       940       950       960
           ....|....|....|....|....|....|....|....|....|....|....|....|
SNAMA-B    ------------------------------------------------------------
SNAMA-A    PKSHSILTVGAASDDNLGPRSKLSEANSVNLSKWEIDENILGLEDSSKKAAGASDDPSEI

                970       980       990       1000      1010      1020
           ....|....|....|....|....|....|....|....|....|....|....|....|
SNAMA-B    ------------------------------------------------------------
SNAMA-A    TSDVLRKAENAIFAKAINAIRPMEFQVIINSKDNSKDRSVVRSDKDRSSSPRRNNSSRSV

                1030      1040      1050      1060      1070      1080
           ....|....|....|....|....|....|....|....|....|....|....|....|
SNAMA-B    ------------------------------------------------------------
SNAMA-A    KDRLGTKISNDRSRSRDKSKGRRRAARSSDDDANRGRSDRHGSRKRDNRSRDRAAPSEKR

                1090      1100      1110      1120      1130      1140
           ....|....|....|....|....|....|....|....|....|....|....|....|
SNAMA-B    ------------------------------------------------------------
SNAMA-A    QERSYKRSSPEDDKLRRQNKEQSESKHGKHDQNNSDDSDRRAAKNTKSSDSRVVSSVTAV        p53 BINDING ?
```

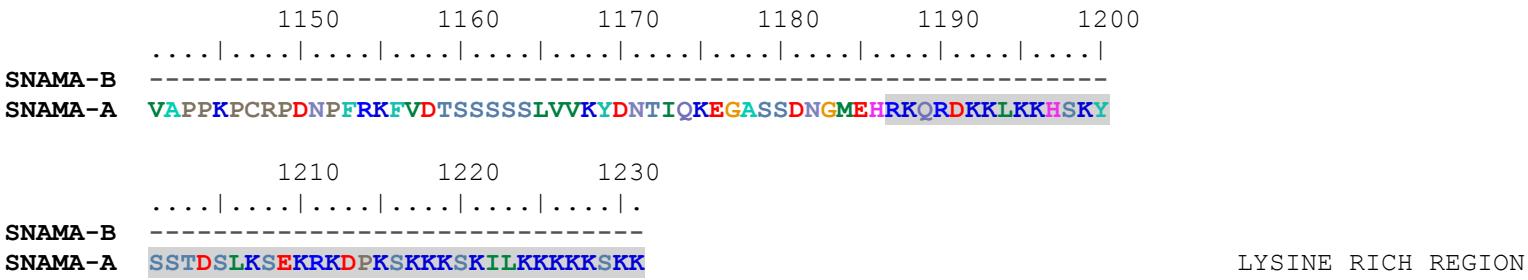


Figure 21: Aligned protein sequences of SNAMA A and SNAMA B. The transcripts have identical sequences and the DWNN Catalytic module (DCM) functional domains are conserved in both transcripts. The possible locations of all functional domains are highlighted in grey with their labeling alongside the alignment.

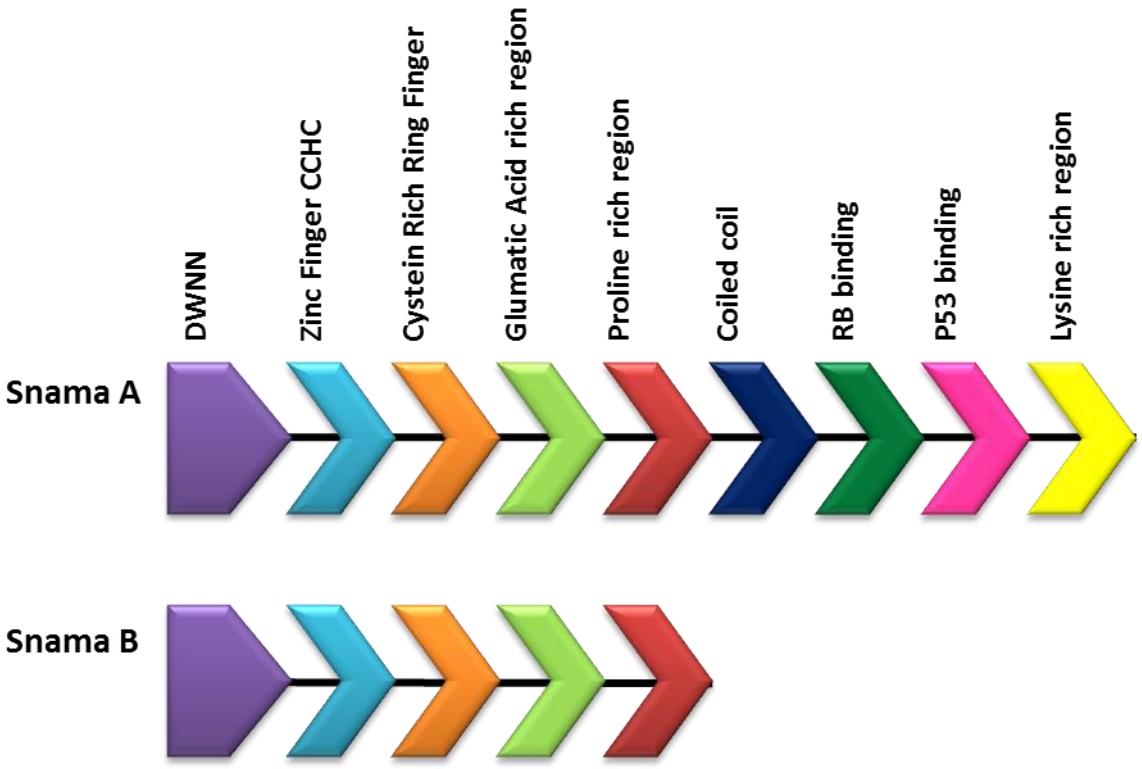


Figure 22: A schematic overview of the possible domain structure of SNAMA isoforms. On comparison of the protein coding regions of the two isoforms we find that: SNAMA B shares the DCM (DWNN, Zinc Knuckle, RING finger), with SNAMA A. Furthermore, SNAMA B lacks downstream p53 and Rb binding domains as well as, coiled coil and lysine rich regions.

3.5 Differential expression of SNAMA during development and in response to DNA damage

We have shown experimentally, the presence of the SNAMA B transcript by cDNA synthesis targeted at the gene, and amplification by subsequent PCR, shown in Figure 23 and 24. Specific primers were designed targeting the amplification of fragments of SNAMA A and B from wild type adult flies, Camptothecin treated adult flies as well as wild type fly embryos. Camptothecin causes DNA damage through the inhibition of DNA Topoisomerase 1. cDNA was synthesised using Oligo(dT) primers in one instance and SNAMA specific primers in another. cDNA synthesis using specific primers presents a limitation in that primers are not always able to bind the tail of the product cDNA making PCR amplification difficult.

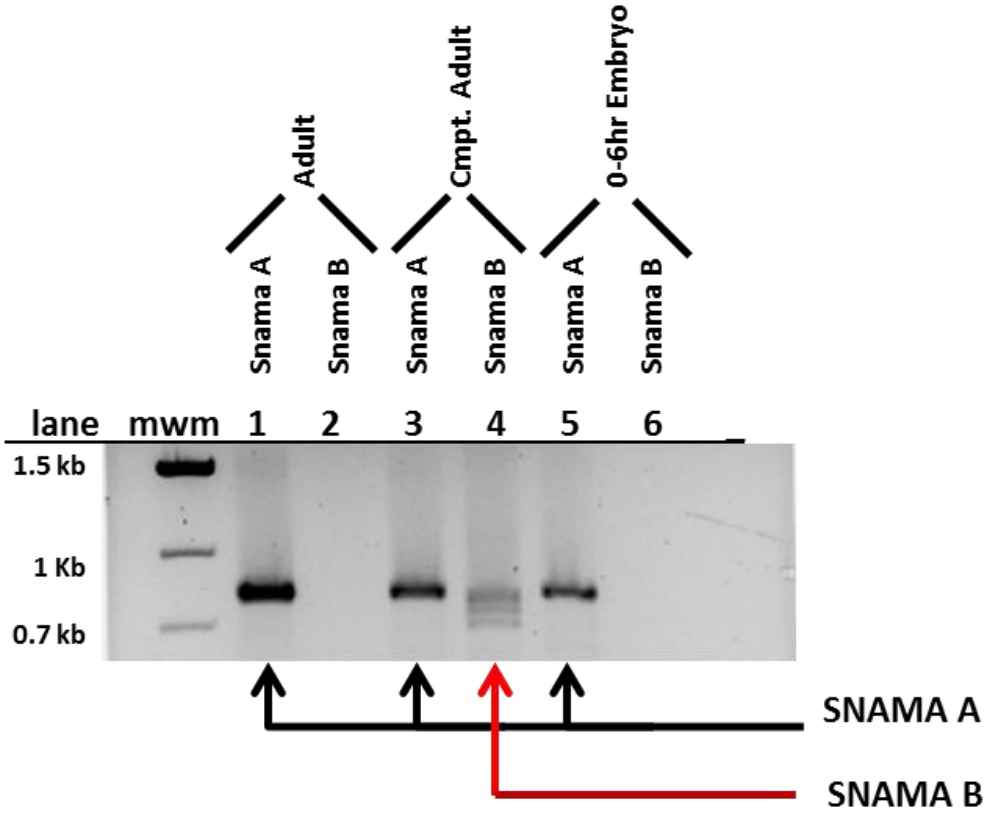


Figure 23: SNAMA A was detected after Reverse Transcriptase PCR. cDNA synthesised using specific reverse primers from wildtype adult flies, camptothecin treated adult flies and 0-6hr fly embryos RNA, was used as a template for DNA amplification of SNAMA A and B using targeted primers. DNA was visualized on a 1% Agarose gel containing EtBr. Lanes 1, 3, 5 show SNAMA A from adults, camptothecin treated adults and embryos respectively, while lanes 2, 4 and 6 contain reactions targeted at amplifying SNAMA B. SNAMA A is present in all three instances, whereas SNAMA B is present after DNA damage.

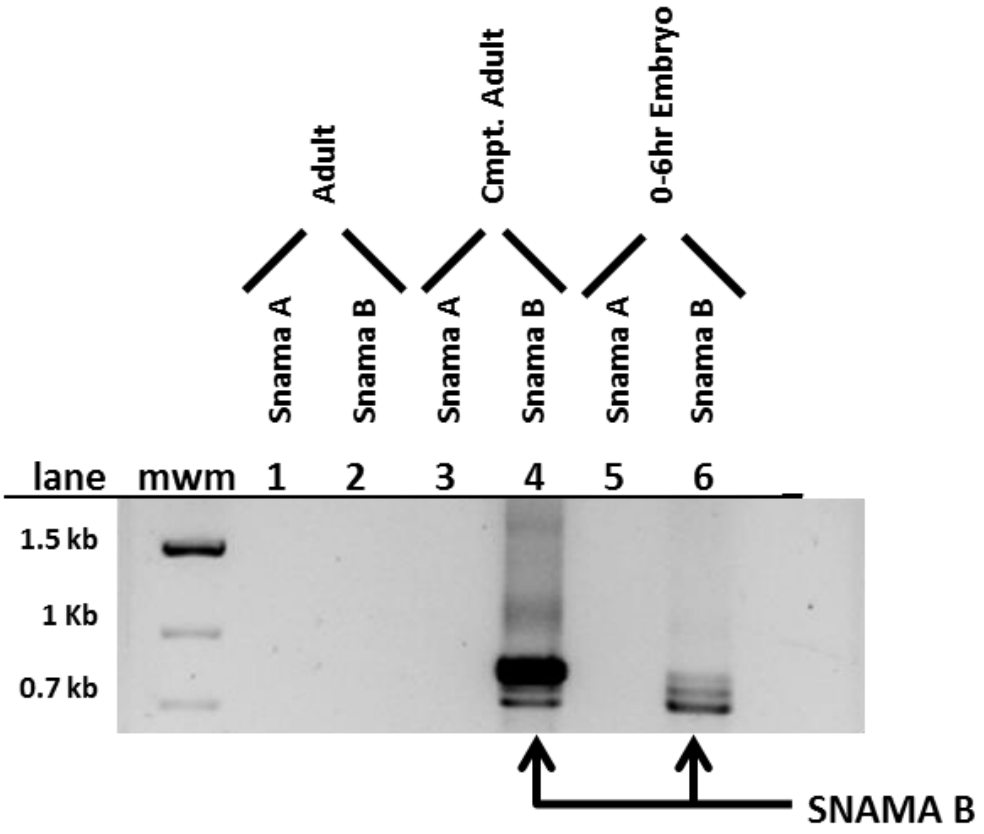


Figure 24: SNAMA B was detected after Reverse Transcriptase PCR. Using RNA from wildtype adult flies, camptothecin treated adult flies and 0-6hr fly embryo's cDNA synthesis was performed with Oligo(dT) primers. DNA was amplified by PCR using specific forward and reverse primers. DNA was visualized on a 1% Agarose gel containing EtBr. Lanes 4 and 6 shows SNAMA B from camptothecin treated adults and untreated embryo's respectively. Lanes 1, 3 and 5 contain reactions targeted at amplifying SNAMA A. Lane 2 shows that SNAMA B was not amplified from wild type adult flies, but was present after DNA damage and during embryonic development.

4. Discussion

This study has revealed more information on expression patterns as well as the existence of two isoforms of SNAMA (a possible interacting partner of Dmp53). *SNAMA* gene studies have revealed two alternative splice sites for *SNAMA*. Furthermore, PCR studies have shown, experimentally, the existence of the transcripts that are formed by alternative splicing. The *SNAMA B* transcript is stimulated by DNA damage and also appears in developing fly embryos.

The study aimed at identifying interacting partners for *Drosophila* p53, using heterologous expression of proteins and chromatography, followed by immunoprecipitation studies to identify interacting proteins. Small amounts of soluble Dmp53 were expressed in yeast expression experiments and the protein was found only in inclusion bodies in prokaryotic systems. Refolding these proteins proved insurmountable leading to the employment of the eukaryotic system. The unavailability of high quality commercial antibodies prevented protein interactions studies. Consequently, interaction studies were not concluded in this study. In this section, the key discoveries about the expression of the SNAMA gene and its possible involvement in genotoxic stress are discussed. Furthermore, insights are made about the implications of these transcription developments with regard to the interaction of SNAMA with other proteins.

4.1 SNAMA B, a shorter SNAMA isoform comprises an independent DWNN Catalytic module and is activated by DNA damage.

A recent review of the flybase *Drosophila* genome has shown the possible existence of a second SNAMA isoform. We have studied the sequences of the proposed transcripts and have concluded that *SNAMA* is alternatively spliced to produce two transcripts. The 1231 amino acid SNAMA isoform yielding a predicted protein of 139 kDa is encoded from a 3.9 kb transcript. SNAMA B has a 494 amino acid isoform, giving rise to a predicted 55 kDa protein and encoded by a 2.1 kb transcript.

The transcripts differ in their 5′ untranslated regions, with *SNAMA* A (the longer transcript) having a shorter UTR. Both transcripts encode the DCM, consisting the DWNN, Zink knuckle and Cysteine Ring finger as well as glutamic acid and proline rich regions. These transcripts should therefore share identical functions and be involved in the same processes and pathways related to the DCM. The presence of a proline rich region in both isoforms also indicates possible nucleic acid interactions and interactions with other proteins. Proline rich regions have been shown to bind a variety of motifs (Kay et al., 2000). Alternative splicing introduces a stop codon in exon 5, producing transcript B. The majority of SNAMA B is thus made up of the DWNN Catalytic Module. This suggests that SNAMA B functions and exists in a similar fashion to the mammalian independent DWNN isoform of RBBP6, which also lacks downstream binding domains and has been suggested to play a role in regulation of the cell cycle, and to be down regulated in cancers (Mbita et al., 2011).

Possible downstream functional domains which only occur in the full length SNAMA A transcript include the RB and p53 binding domains and coiled-coil and lysine rich regions. We expect that SNAMA A may interact with other proteins and perhaps with nucleic acids due to the presence of these domains. According to Ntwasa (2008), the presence of the coiled-coil domain further suggests interactions with other proteins and possible dimerization of the isoform. Proteins with coiled-coil domains are often implicated in the control of oligomerization and interact with various structures including the Golgi apparatus, cytoskeleton, nuclear matrix centromere and centrosomes and chromatin (Ntwasa, 2008). Furthermore, both SNAMA A and B may have other interactions with Dmp53 owing to the presence of the DCM. These are most probable through E3 ligase activity in apoptosis and ubiquitin pathways. Hull and Ntwasa (2010), also suggest that SNAMA isoforms may be integral in DNA damage response. Interestingly Hull (2012), has shown a 55kDa protein detected through Western analyses with anti-SNAMA antibodies after methyl pyruvate treatment but not after Camptothecin treatment. It is possible that this protein is SNAMA B and was amplified here through PCR after treatment with either camptothecin or methyl pyruvate. Since PCR is a more sensitive detection method, the Western analysis may have failed to detect small amounts of protein. This suggests that *SNAMA B* is present as a transcript after DNA damage but may not be expressed in high levels after only camptothecin treatment and that expression may be enhanced by methyl pyruvate treatment.

4.2 SNAMA A and B are probably differentially expressed during embryonic development and induced by DNA damage

SNAMA is crucial for embryonic development, having a probable role in suppressing apoptosis(Hull and Ntwasa, 2010). Furthermore, knockout studies have shown that a lack of SNAMA is lethal (Mather et al., 2005). Similarly Dmp53 is expressed throughout development (Jin et al., 2000) and programmed cell death is known to have an important role during development (Bangs and White, 2000). Studies into the effect of camptothecin induced apoptosis on human RBBP6 has shown an upregulation of RBBP6 promoters (Pretorius et al., 2011). During early embryonic development, SNAMA expression is considerably higher. SNAMA A is present in early stages of embryonic development and expression is reduced in late stages.

The study was also able to show the existence of the SNAMA B transcript experimentally. SNAMA B did not seem to be present in normal adult flies but was amplified from cDNA with poly A tails using specific primers in wild type 0-6 hour embryos, as well as adults with camptothecin induced DNA damage. PCR using cDNA amplified from specific primers as a template was successful in amplifying SNAMA A in adult flies; camptothecin treated adult flies, as well as 0-6 hour fly embryos. SNAMA B using this specifically amplified cDNA, was present in adult flies that had undergone camptothecin treatment. This suggests that SNAMA B is upregulated after DNA damage.

This further suggests that SNAMA, like its counterparts (RBBP6), is involved in apoptosis and interacts with Dmp53. Upon DNA damage, p53 triggers apoptosis through the ubiquitin proteasome system in which it is suspected that SNAMA is an active ubiquitin like modifier, due to the presence of a DCM (Cajee et al., 2012).

4.3 Dmp53 interactors

In order to study Dmp53 interactions with other proteins, a recombinant Dmp53 needed to be expressed, and purified. This protein would then be used in immune precipitation studies. Initial expression of Dmp53 was carried out using a bacterial expression system (*E. coli*, transformed with Dmp53 DNA in pET41a plasmid), Human p53 is routinely cloned expressed in a similar way as described by Midgley et al., (1992). It was found that Dmp53 in the bacterial expression system was expressed as insoluble inclusion bodies regardless of time period of induction. The protein can sometimes be solubilised and refolded, however, methods are still cumbersome and expensive and often result in poor recovery of protein (Singh and Panda, 2005). Also, these techniques may render the protein unusable in further studies, if antibody epitopes are disrupted. By changing the expression system, the study then attempted to express already soluble protein.

A yeast expression system was used in order to express soluble Dmp53, after bacterial expression had expressed the protein as insoluble inclusion bodies. Yeast expression systems offer a step up from prokaryotic/bacterial expression systems, for example post

translational modifications and secretory expression. Yeast grow on simple media to high cell densities, and have advanced genetics amongst eukaryotes, allowing easy manipulation similar to manipulation of *E. coli* (Romanos et al., 1992). The eukaryotic yeast subcellular organisation can carry out the necessary post translational modifications (including disulphide bonds, glycosylation, proteolytic modification and other modifications) needed for complex proteins, through its multi-component apparatus (Madzak et al., 2004).

Yeast was transformed with a Dmp53 DNA fused with a HIS tag in pTG1895. The pTG1895 plasmid used in yeast protein expression, included the factor alpha pre-pro cassette in its sequence. Upon expression, the cassette signals the secretion of the protein by the cell, to the culture media, making extraction and purification of expressed protein simpler. Upon experimentation we were not able to retrieve Dmp53 from the secreted into the supernatant but were able to detect and purify soluble Dmp53 protein expressed and retained in the cytosol. The difficulties in expressing secreted protein could be due to the choice of Yeast strain, where *S. cerevisiae* is becoming less popular having found to be limiting in many factors for example, low product yield, low secretion capacity as well as poor plasmid stability amongst others (Madzak et al., 2004). Another limiting factor may also be that yeast cells exhibit a high codon bias (Akashi, 2003).

We were able to express soluble Dmp53 protein in yeast; however, the level of expression was low. Proteins from crude extracts were not visible on stained SDS PAGE gels. Concentrating the extracts by TCA precipitation or stirred cell filtration proved effective in making proteins visible. Later the quantity of expression increased slightly with the inclusion of protease inhibitors, however, protein yield from large cultures, was not enough for further biological studies. The use of a yeast strain with impaired proteolytic function could further enhance peptide recovery (Thill et al., 1986) .

The study demonstrated the heterologous expression of a suspected soluble Dmp53 protein in yeast. The protein size after electrophoresis matched the expected size from calculations. Thus, the protein was purified out of crude extract through HIS nickel select chromatography. However, a benchmark for positive identification of a protein is immunoblotting. Unfortunately, due to non-specific antigen antibody binding it could not be confirmed that the expressed protein was Dmp53. The study had used/tested three different anti-Dmp53 antibodies, having different epitopes. It was found that these antibodies bind non-specifically. Furthermore, there was great variation in the antibody properties per batch. All these were polyclonal antibodies, as monoclonal Dmp53 antibodies were not readily available.

Immunoprecipitation studies use specific antibody antigen interactions to pull down a protein of interest together with possible interacting partners. The availability of an A grade specific Dmp53 antibody is thus

essential to further study Dmp53. We were unable to perform immunoprecipitation studies, as all readily available commercial antibodies bind non-specifically (as shown in Western blot analyses), and thus could not reliably be used.

Protein interactions, between Dmp53 and SNAMA, can only be confirmed with further research after Dmp53, as well as SNAMA, can be efficiently expressed and purified in their soluble states. In addition, specific monoclonal antibodies directed towards both Dmp53 and SNAMA, must be developed to ensure accurate results. We can thus only speculate that SNAMA interacts with Dmp53 in one or more systems, where both proteins are present, involved or upregulated. Likely interations may occur in the ubiquitin system due to the DCM of SNAMA and Dmp53's role in apoptosis, after DNA damage, and during fly embryonic development.

5. Conclusion

i. SNAMA is alternatively spliced to produce two isoforms.
ii. The expression of alternate splice forms of SNAMA is dependent on DNA damage and developmental stages.
iii. SNAMA domain structure and its upregulation after DNA damage suggests possible interactions with Dmp53, and/or the Ubiquitin system and other proteins.
iv. Overexpression of truncated Dmp53 in a bacterial overexpression system resulted in an insoluble protein that could not be purified or refolded.
v. The yeast expression system resulted in native full length Dmp53 but it was only expressed in small amounts.

5.1 Future work

i. Effective overexpression of native Dmp53 will require the development of an alternative overexpression system. Ideally, this would be yeast based and the vector used will be effective in attaching an export signal to the Dmp53 protein resulting in its export from the cell.
ii. In order for this research to proceed, an antibody, which binds Dmp53 specifically, would need to be developed.

References

Akashi, H., (2003). Translational selection and yeast proteome evolution. *Genetics* 164, 1291–1303.

Antunes, R.J., (2008). Characterisation of the DWNN Domain and RING finger-like motif within the DCM of the Drosophila Melanogaster SNAMA Protein. (Masters Dissertation). University of the Witwatersrand, Johannesburg, South Africa.

Bangs, P., White, K., (2000). Regulation and execution of apoptosis during Drosophila development. *Dev. Dyn.* 218, 68–79.

Bender, W., Spierer, P., Hogness, D.S., (1983). Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in Drosophila melanogaster. *J. Mol. Biol.* 168.

Birnboim, H.C., (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol.* 100, 243–255.

Bourdon, J.-C., (2005). p53 isoforms can regulate p53 transcriptional activity. *Genes Dev.* 19, 2122–2137.

Brodsky, M.H., Nordstrom, W., Tsang, G., Kwan, E., Rubin, G.M., Abrams, J.M., (2000). < i> Drosophila</i> p53 Binds a Damage Response Element at the< i> reaper</i> Locus. *Cell* 101, 103–113.

Brody, T.B., (2000). p53 [WWW Document]. *Interact. Fly*. URL http://www.sdbonline.org/sites/fly/hjmuller/p53-1.htm (accessed 2.20.14).

Brooks, C.L., Gu, W., (2006). p53 Ubiquitination: Mdm2 and Beyond. *Mol. Cell* 21, 307–315.

Cajee, U.-F., Hull, R., Ntwasa, M., (2012). Modification by Ubiquitin-Like Proteins: Significance in Apoptosis and Autophagy Pathways. *Int. J. Mol. Sci.* 13, 11804–11831.

Chomczynski, P., (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* 15, 532–4, 536–537.

Daines, B., Wang, H., Wang, L., Li, Y., Han, Y., Emmert, D., Gelbart, W., Wang, X., Li, W., Gibbs, R., Chen, R., (2010). The Drosophila melanogaster transcriptome by paired-end RNA sequencing. *Genome Res.* 21, 315–324.

Dichtel-Danjoy, M.L., Ma, D., Dourlen, P., Chatelain, G., Napoletano, F., Robin, M., Corbet, M., Levet, C., Hafsi, H., Hainaut, P., (2012). Drosophila p53 isoforms differentially regulate apoptosis and apoptosis-induced proliferation. *Cell Death Differ.* 20, 108–116.

Fan, Y., Lee, T.V., Xu, D., Chen, Z., Lamblin, A.-F., Steller, H., Bergmann, A., (2010). Dual roles of Drosophila p53 in cell death and cell differentiation. *Cell Death Differ.* 17, 912–921.

Farnebo, M., Bykov, V.J.N., Wiman, K.G., (2010). The p53 tumor suppressor: A master regulator of diverse cellular processes and therapeutic target in cancer. *Biochem. Biophys. Res. Commun.* 396, 85–89.

Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., Parkin, D.M., (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* 127, 2893–2917.

Genentech, (2014). Apoptosis [WWW Document]. *Biooncology*. URL http://www.biooncology.com/biological-pathways/apoptosis (accessed 2.20.14).

Hanahan, D., Weinberg, R.A., (2000). The hallmarks of cancer. *cell* 100, 57–70.

Hershko, A., Ciechanover, A., (1998). THE UBIQUITIN SYSTEM. *Annu Rev Biochem* 67, 425–479.

Herzog, G., Joerger, A.C., Shmueli, M.D., Fersht, A.R., Gazit, E., Segal, D., (2012). Evaluating Drosophila p53 as a Model System for Studying Cancer Mutations. *J. Biol. Chem.* 287, 44330–44337.

Hull, R., (2012). Stress response to genotoxic agents and to infection (Thesis). University of the Witwatersrand, Johannesburg.

Hull, R., Ntwasa, M.M., (2010). Glycolytic flux occurs in Drosophila melanogaster recovering from camptothecin treatment: *Anticancer. Drugs* 21, 945–957.

Jin, S., Martinek, S., Joo, W.S., Wortman, J.R., Mirkovic, N., Sali, A., Yandell, M.D., Pavletich, N.P., Young, M.W., Levine, A.J., (2000). Identification and characterization of a p53 homologue in Drosophila melanogaster. *Proc. Natl. Acad. Sci.* 97, 7301–7306.

Kauppila, S., Maaty, W.S., Chen, P., Tomar, R.S., Eby, M.T., Chapo, J., Chew, S., Rathore, N., Zachariah, S., Sinha, S.K., (2003). Eiger and its receptor, Wengen, comprise a TNF-like system in Drosophila. *Oncogene* 22, 4860–4867.

Kay, B.K., Williamson, M.P., Sudol, M., (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* 14, 231–241.

Khoury, M.P., Bourdon, J.-C., (2009). The Isoforms of the p53 Protein. *Cold Spring Harb. Perspect. Biol.* 2, a000927–a000927.

Kim, H., Lee, J.M., Lee, G., Bhin, J., Oh, S.K., Kim, K., Pyo, K.E., Lee, J.S., Yim, H.Y., Kim, K.I., Hwang, D., Chung, J., Baek, S.H., (2011). DNA Damage-Induced RORα Is Crucial for p53 Stabilization and Increased Apoptosis. *Mol. Cell* 44, 797–810.

Laemmli, U.K., (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.

Machado-Silva, A., Perrier, S., Bourdon, J.-C., (2010). p53 family members in cancer diagnosis and treatment. *Semin. Cancer Biol.* 20, 57–62.

Madzak, C., Gaillardin, C., Beckerich, J.-M., (2004). Heterologous protein expression and secretion in the non-conventional yeast Yarrowia lipolytica: a review. *J. Biotechnol.* 109, 63–81.

Marcel, V., Dichtel-Danjoy, M.-L., Sagne, C., Hafsi, H., Ma, D., Ortiz-Cuaran, S., Olivier, M., Hall, J., Mollereau, B., Hainaut, P., (2011). Biological functions of p53 isoforms through evolution: lessons from animal and cellular models. *Cell Death Differ.* 18, 1815–1824.

Mather, A., Rakgotho, M., Ntwasa, M., (2005). SNAMA, a novel protein with a DWNN domain and a RING finger-like motif: A possible role in apoptosis. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* 1727, 169–176.

Mbita, Z., Meyer, M., Skepu, A., Hosie, M., Rees, J., Dlamini, Z., (2011). De-regulation of the RBBP6 isoform 3/DWNN in human cancers. *Mol. Cell. Biochem.* 362, 249–262.

Michael, D., Oren, M., (2002). The p53 and Mdm2 families in cancer. *Curr. Opin. Genet. Dev.* 12, 53–59.

Midgley, C.A., Fisher, C.J., Bartek, J., Vojtesek, B., Lane, D., Barnes, D.M., (1992). Analysis of p53 expression in human tumours: an antibody raised against human p53 expressed in Escherichia coli. *J. Cell Sci.* 101, 183–189.

Murray-Zmijewski, F., Lane, D.P., Bourdon, J.C., (2006). p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. *Cell Death Differ.* 13, 962–972.

Ntwasa, M., (2008). The retinoblastoma binding protein 6 is a potential target for therapeutic drugs. *Biotechnol Mol Biol Rev* 3, 24–31.

Ohta, T., Fukuda, M., (2004). Ubiquitin and breast cancer. *Oncogene* 23, 2079–2088.

Ollmann, M., Young, L.M., Di Como, C.J., Karim, F., Belvin, M., Robertson, S., Whittaker, K., Demsky, M., Fisher, W.W., Buchman, A., (2000). < i> Drosophila p53</i> Is a Structural and Functional Homolog of the Tumor Suppressor< i> p53</i>. *Cell* 101, 91–101.

Pretorius, A., Kaur, M., Wamalwa, M., February, M.F., Essack, M., Bajic, V.B., Rees, D.J.G., (2011). Functional analysis and characterization of the human RBBP6 promoters based on a combination of molecular biology and in silico approaches provide additional evidence for RBBP6 role in apoptosis. *GSTF Int. J. Bioinforma. Biotechnol.* 1.

Pugh, D.J., Eiso, A.B., Faro, A., Lutya, P.T., Hoffmann, E., Rees, D.J.G., (2006). DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. *BMC Struct. Biol.* 6, 1.

Reiter, L.T., Potocki, L., Chien, S., Gribskov, M., Bier, E., (2001). A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. *Genome Res.* 11, 1114–1125.

Richardson, H., Kumar, S., (2002). Death to flies:< i> Drosophila</i> as a model system to study programmed cell death. *J. Immunol. Methods* 265, 21–38.

Romanos, M.A., Scorer, C.A., Clare, J.J., (1992). Foreign gene expression in yeast: a review. *Yeast* 8, 423–488.

Schuler, M., Green, D.R., (2005). Transcription, apoptosis and p53: catch-22. *Trends Genet.* 21, 182–187.

Selivanova, G., (2004). p53: fighting cancer. *Curr. Cancer Drug Targets* 4, 385–402.

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.

Singh, S.M., Panda, A.K., (2005). Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* 99, 303–310.

St Pierre, S.E., Ponting, L., Stefancsik, R., McQuilton, P., FlyBase Consortium, (2014). FlyBase 102--advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* 42, D780–D788.

Thill, G.P., Harpold, M.M., Tschopp, J.F., (1986). Expression and secretion of polypeptides from saccharomyces cerevisiae. Google Patents.

Vidal, M., Cagan, R.L., (2006). Drosophila models for cancer research. *Curr. Opin. Genet. Dev.* 16, 10–16.

Appendix A, Dmp53 in pET41a sequence

Sequence of Dmp53 insert in pET41a with primer locations and restriction site

```
1      TTAACTTTAA GAAGGAGATA TACATATGTC CCCTATACTA GGTTATTGGA AAATTAAGGG
       AATTGAAATT CTTCCTCTAT ATGTATACAG GGGATATGAT CCAATAACCT TTTAATTCCC

61     CCTTGTGCAA CCCACTCGAC TTCTTTTGGA ATATCTTGAA GAAAAATATG AAGAGCATTT
       GGAACACGTT GGGTGAGCTG AAGAAAACCT TATAGAACTT CTTTTTATAC TTCTCGTAAA

121    GTATGAGCGC GATGAAGGTG ATAAATGGCG AAACAAAAAG TTTGAATTGG GTTTGGAGTT
       CATACTCGCG CTACTTCCAC TATTTACCGC TTTGTTTTTC AAACTTAACC CAAACCTCAA

181    TCCCAATCTT CCTTATTATA TTGATGGTGA TGTTAAATTA ACACAGTCTA TGGCCATCAT
       AGGGTTAGAA GGAATAATAT AACTACCACT ACAATTTAAT TGTGTCAGAT ACCGGTAGTA

241    ACGTTATATA GCTGACAAGC ACAACATGTT GGGTGGTTGT CCAAAAGAGC GTGCAGAGAT
       TGCAATATAT CGACTGTTCG TGTTGTACAA CCCACCAACA GGTTTTCTCG CACGTCTCTA

301    TTCAATGCTT GAAGGAGCGG TTTTGGATAT TAGATACGGT GTTTCGAGAA TTGCATATAG
       AAGTTACGAA CTTCCTCGCC AAAACCTATA ATCTATGCCA CAAAGCTCTT AACGTATATC

361    TAAAGACTTT GAAACTCTCA AAGTTGATTT TCTTAGCAAG CTACCTGAAA TGCTGAAAAT
       ATTTCTGAAA CTTTGAGAGT TTCAACTAAA AGAATCGTTC GATGGACTTT ACGACTTTTA

421    GTTCGAAGAT CGTTTATGTC ATAAAACATA TTTAAATGGT GATCATGTAA CCCATCCTGA
       CAAGCTTCTA GCAAATACAG TATTTTGTAT AAATTTACCA CTAGTACATT GGGTAGGACT

481    CTTCATGTTG TATGACGCTC TTGATGTTGT TTTATACATG GACCCAATGT GCCTGGATGC
       GAAGTACAAC ATACTGCGAG AACTACAACA AAATATGTAC CTGGGTTACA CGGACCTACG

541    GTTCCCAAAA TTAGTTTGTT TTAAAAAACG TATTGAAGCT ATCCCACAAA TTGATAAGTA
       CAAGGGTTTT AATCAAACAA AATTTTTTGC ATAACTTCGA TAGGGTGTTT AACTATTCAT

601    CTTGAAATCC AGCAAGTATA TAGCATGGCC TTTGCAGGGC TGGCAAGCCA CGTTTGGTGG
       GAACTTTAGG TCGTTCATAT ATCGTACCGG AAACGTCCCG ACCGTTCGGT GCAAACCACC

   Forward primer with HindIII cut site 5'TAGTGAA GCTTGTCATC ACCAT 3'
661    TGGCGACCAT CCTCCAAAAT CGGATGGTTC AACTAGTGGT TCTGGTCATC ACCATCACCA
       ACCGCTGGTA GGAGGTTTTA GCCTACCAAG TTGATCACCA AGACCAGTAG TGGTAGTGGT

721    TCACTCCGCG GGTCTGGTGC CACGCGGTAG TACTGCAATT GGTATGAAAG AAACCGCTGC
       AGTGAGGCGC CCAGACCACG GTGCGCCATC ATGACGTTAA CCATACTTTC TTTGGCGACG

781    TGCTAAATTC GAACGCCAGC ACATGGACAG CCCAGATCTG GGTACCGGTG GTGGCTCCGG
       ACGATTTAAG CTTGCGGTCG TGTACCTGTC GGGTCTAGAC CCATGGCCAC CACCGAGGCC

841    TGATGACGAC GACAAGAGTC CCATGGGATA TCGGGGATCC ATGTACTCGA TTCCGCTGAA
       ACTACTGCTG CTGTTCTCAG GGTACCCTAT AGCCCCTAGG TACATGAGCT AARGCGACTT

901    CAAGCTCTAC ATCCGGATGA ACAAGGCCTT CAACGTGGAC GTTCAGTTTA AGTCTAAGAT
       GTTCGAGATG TAGGCCTACT TGTTCCGGAA GTTGCACCTG CAAGTCAAAT TCAGATTCTA
```

961 GCCCATCCAG CCACTTAATT TGCGTGTGTT CCTTTGCTTC TCCAATGATG TCAGTGCTCC
CGGGTAGGTC GGTGAATTAA ACGCACACAA GGAAACGAAG AGGTTACTAC AGTCACGAGG

1021 CGTGGTCCGC TGTCAAAATC ACCTTAGCGT TGAGCCTTTG ACGGCCAATA ACGCAAAAAT
GCACCAGGCG ACAGTTTTAG TGGAATCGCA ACTCGGAAAC TGCCGGTTAT TGCGTTTTTA

1081 GCGCGAGAGC TTGCTGCGCA GCGAGAATCC CAACAGTGTA TATTGTGGAA ATGCTCAGGG
CGCGCTCTCG AACGACGCGT CGCTCTTAGG GTTGTCACAT ATAACACCTT TACGAGTCCC

1141 CAAAGGAATT TCCGAGCGTT TTTCCGTTGT AGTCCCCCTG AACATGAGTC GGTCTGTAAC
GTTTCCTTAA AGGCTCGCAA AAAGGCAACA TCAGGGGGAC TTGTACTCAG CCAGACATTG

1201 CCGCAGTGGG CTCACGCGCC AGACCCTGGC CTTCAAGTTC GTCTGCCAAA ACTCGTGTAT
GGCGTCACCC GAGTGCGCGG TCTGGGACCG GAAGTTCAAG CAGACGGTTT TGAGCACATA

1261 CGGGCGAAAA GAAACTTCCT TAGTCTTCTG CCTGGAGAAA GCATGCGGCG ATATCGTGGG
GCCCGCTTTT CTTTGAAGGA ATCAGAAGAC GGACCTCTTT CGTACGCCGC TATAGCACCC

1321 ACAGCATGTT ATACATGTTA AAATATGTAC GTGCCCCAAG CGGGATCGCA TCCAAGACGA
TGTCGTACAA TATGTACAAT TTTATACATG CACGGGGTTC GCCCTAGCGT AGGTTCTGCT

1381 ACGCCAGCTC AATAGCAAGA AGCGCAAGTC CGTGCCGGAA GCCGCCGAAG AAGATGAGCC
TGCGGTCGAG TTATCGTTCT TCGCGTTCAG GCACGGCCTT CGGCGGCTTC TTCTACTCGG

1441 GTCCAAGGTG CGTCGGTGCA TTGCTATAAA GACGGAGGAC ACGGAGAGCA ATGATAGCCG
CAGGTTCCAC GCAGCCACGT AACGATATTT CTGCCTCCTG TGCCTCTCGT TACTATCGGC

1501 AGACTGCGAC GACTCCGCCG CAGAGTGGAA CGTGTCGCGG ACACCGGATG GCGATTACCG
TCTGACGCTG CTGAGGCGGC GTCTCACCTT GCACAGCGCC TGTGGCCTAC CGCTAATGGC

1561 TCTGGCTATT ACGTGCCCCA ATAAGGAATG GCTGCTGCAG AGCATCGAGG GCATGATTAA
AGACCGATAA TGCACGGGGT TATTCCTTAC CGACGACGTC TCGTAGCTCC CGTACTAATT

1621 AGGAGGCGGC GGCTGAAGTC CTGCGCAATC CCAACCAAGA GAATCTACGT CGCCATGCCA
TCCTCCGCCG CCGACTTCAG GACGCGTTAG GGTTGGTTCT CTTAGATGCA GCGGTACGGT

1681 ACAATTGCTG AGCCTTAAGA AACGTGCCTA CGAGCTGCCA TGAAAGCTTG CGGCCGCACT
TGTTAACGAC TCGGAATTCT TTGCACGGAT GCTCGAC**GGT ACTT**TCGAAC **GCCG**GCGTGA
Forward primer with *Sal*I cut site 3′ **GGT ACTT**CAGCTG **GCCG** 5′

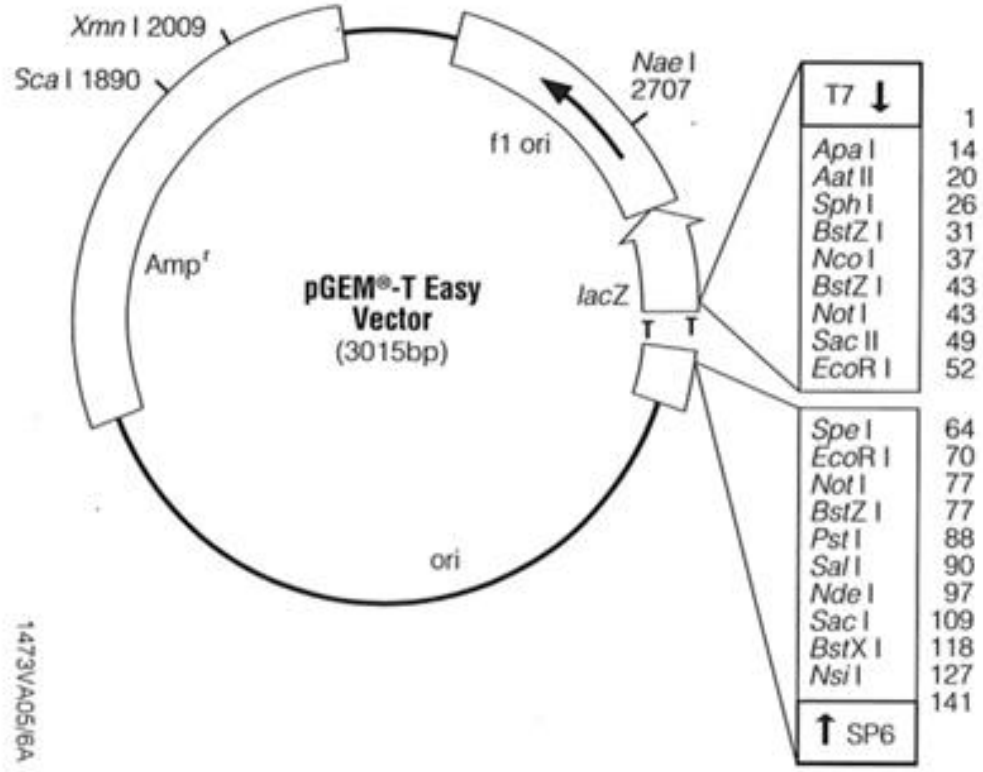
1741 CGAGCACCAC CACCACCACC ACCACCACTA ATTGATTAAT ACCTAGGCTG CTAAACAAAG
GCTCGTGGTG GTGGTGGTGG TGGTGGTGAT TAACTAATTA TGGATCCGAC GATTTGTTTC

1801 CCCGAAAGGA AGCTGAGTTG GCTGCTGCCA CCGCTGAGCA ATAACTAGCA TAACCCCTTG
GGGCTTTCCT TCGACTCAAC CGACGACGGT GGCGACTCGT TATTGATCGT ATTGGGGAAC

1861 GGGCCTCTAA ACGGGTCTTG AGGGGTTTTT TGCTGAAAGG AGGAACTATA TCCGGAT
CCCGGAGATT TGCCCAGAAC TCCCCAAAAA ACGACTTTCC TCCTTGATAT AGGCCTA

Key: Primer Sequences, Pet41a Sequence, Dmp53 insert, Pet41a HIS tag

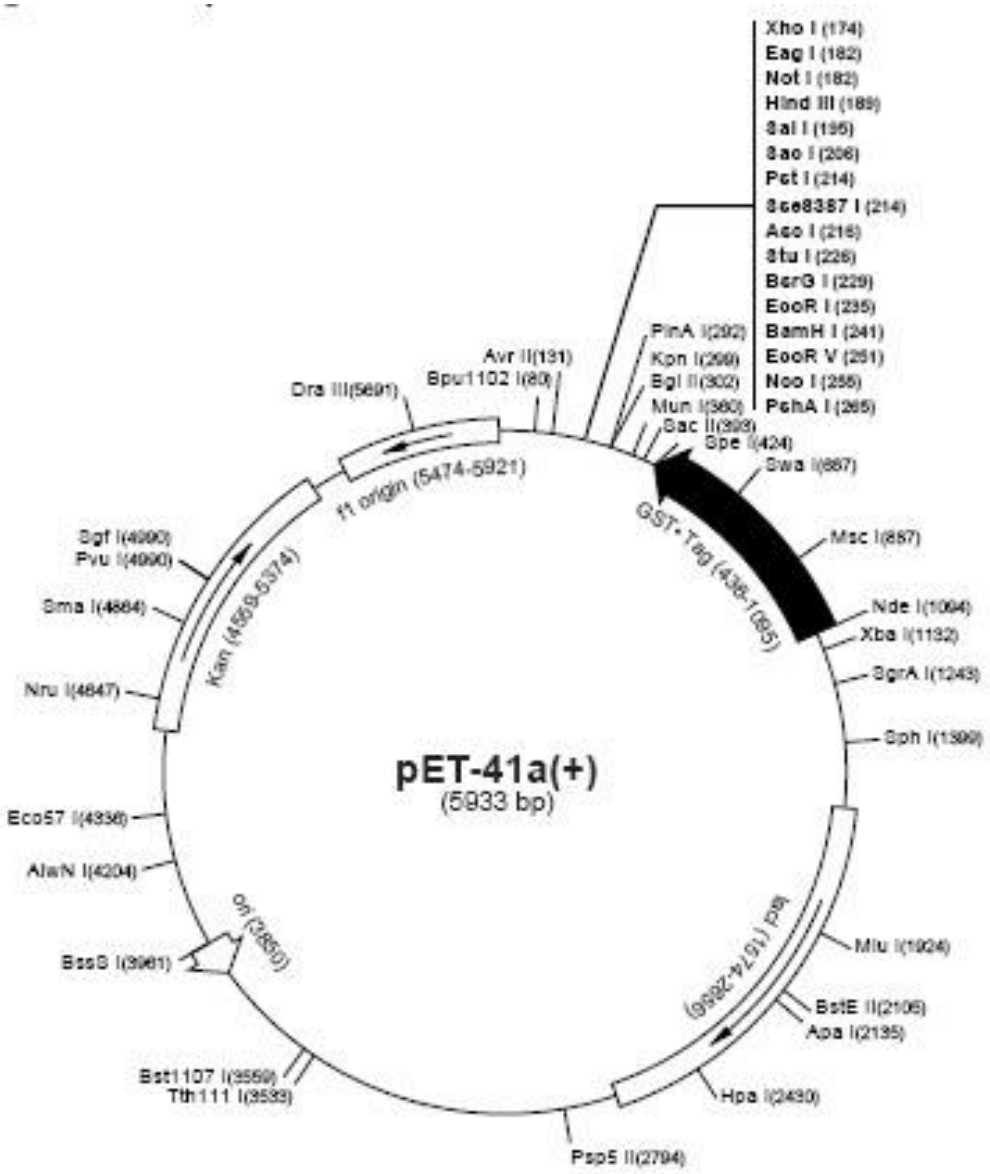
Appendix B, pGEM-T Easy



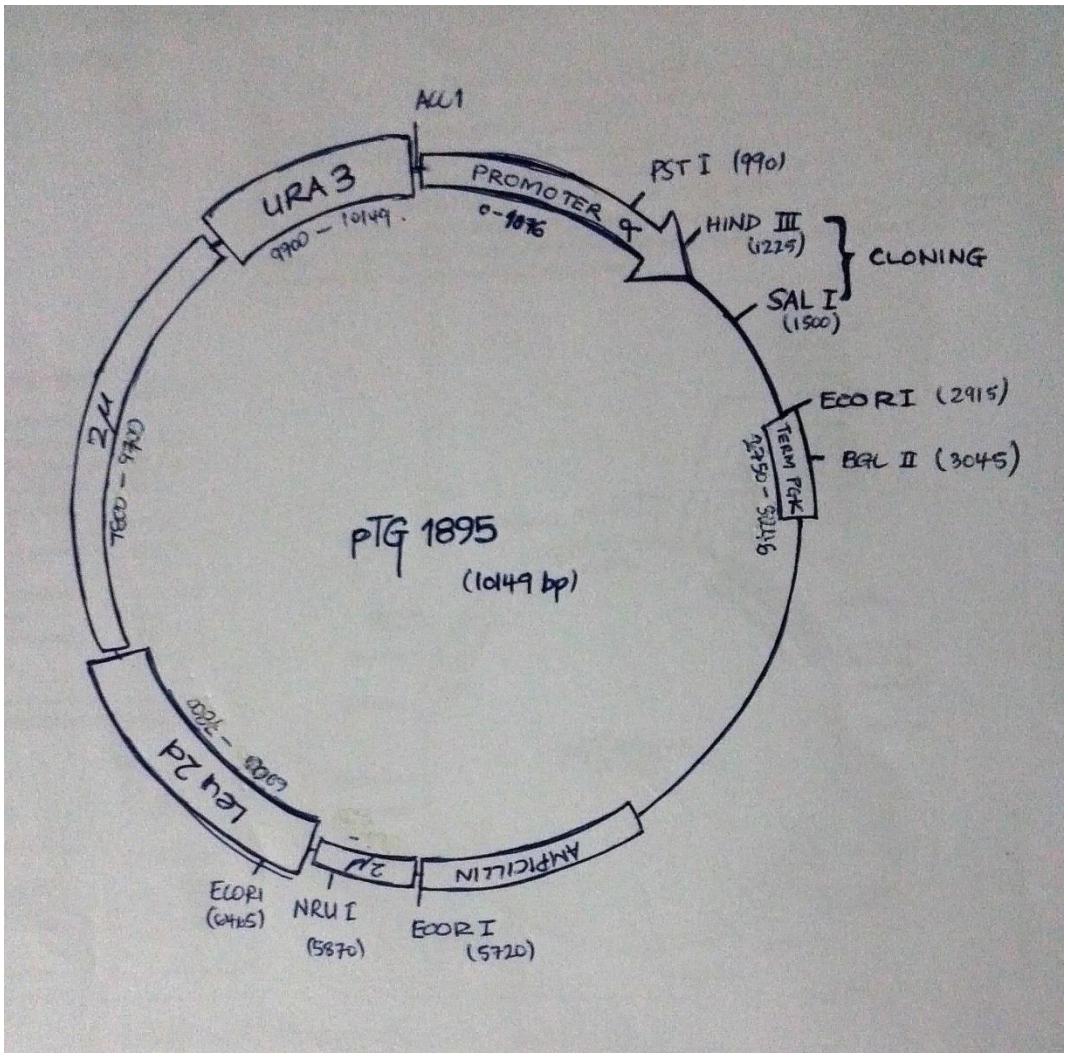
pGEM-T Easy Vector Sequence reference points:

T7 RNA Polymerase transcription initiation site	1
Multiple doning region	10-128
SP6 RNA Polymerase promoter (-17 to +13)	139-158
SP6 RNA Polymerase transcription initiation site	141
pUC/M13 Reverse Sequencing Primer binding site	176-197
lacZ start codon	180
lac operator	200-216
β-lactamase coding region	1337-2197
phage f1 region	2380-2835
lac operon sequences	2836-2996, 166-395
pUC/M13 Forward sequencing Primer binding site	2949-2972
T7 RNA Polymerase promoter (-17 - +13)	2999-3

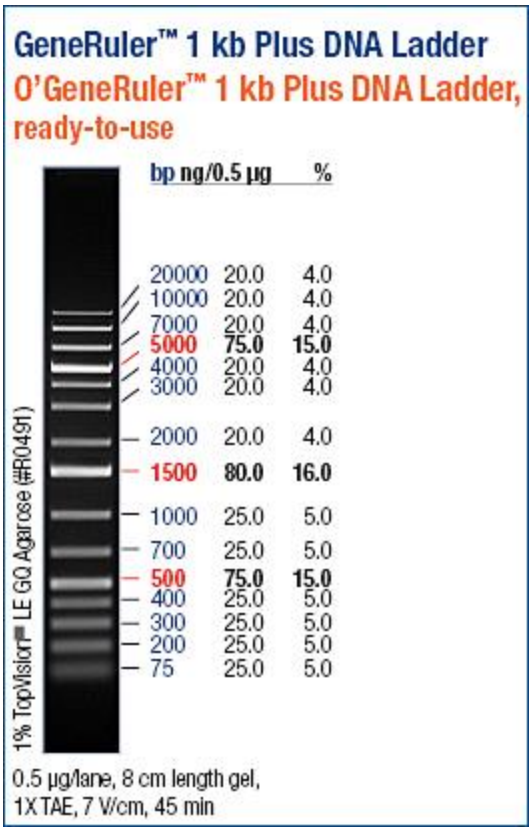
Appendix C, pET41a



Appendix D, pTG1895



Appendix E, Generuler 1kb plus



Appendix F, PageRuler Protein ladder