2017

UNIVERSITY *of the*
WESTERN CAPE

A thesis submitted in fulfillment of the degree of Master of Science
Faculty of Natural Sciences
University of the Western Cape

Author   : Mihlali Mlaza

Supervisor   : Prof. David Pugh

Date   : December 2017

Investigation of the role of the ubiquitin-like DWNN domain in targeting Retinoblastoma Binding Protein 6 to nuclear speckles

Cancer Signalling Laboratory, Biotechnology Department, University of the Western Cape, Robert Sobukwe Road, Bellville 7535, South Africa.

GENERAL PLAGIARISM DECLARATION

Name: MIHLALI MLAZA

Student number: 3312906

1. I hereby declare that I know what plagiarism entails, namely to use another's work and to present it as my own without attributing the sources in the correct way. (Refer to University Calendar part 1 for definition)

2. I know that plagiarism is a punishable offence because it constitutes theft.

3. I understand the plagiarism policy of the Faculty of Natural Science of the University of the Western Cape.

4. I know what the consequences will be if I plagiarize in any of the assignments for my course.

5. I declare therefore that all work presented by me for every aspect of my course, will be my own, and where I have made use of another's work, I will attribute the source in the correct way.

20 December 2017

Signature   Date

ii

ACKNOWLEDGEMENTS

I would like to thank Prof. David Pugh for allowing me the opportunity to complete my project in his laboratory. I would like to voice my most sincere gratitude for the invaluable support he has provided towards my growth as a scientist over the course of the last three years, I could not have asked for a better supervisor.

I would like to thank Dr. Andrew Faro for always pointing me in the right direction every time I had an idea that seemed slightly out of reach or scientifically nonsensical. A number of people have provided resources, moral support and when I needed it the most, constructive criticism at UWC. Amongst them, I would like to give special thanks to Prof. Mervin Meyer for allowing me the use of his cell culture laboratory, Ms. Andronica Ramaila, Ms. Tephney Mahomedy and Ms. Zaida Parsons for providing some of the pCMV-UWC DNA constructs used in this study.

I also extend my greatest and most sincere gratitude to Mrs. Susan Cooper and her team at the Confocal and Light Microscope Imaging Facility at the University of Cape Town for all the time they put in helping me operate the inverted fluorescent microscope and for all the invaluable advice on how to better represent my data.

My family, both from home and the ones I made on campus has been the firm and unwavering support from which I drew most of the energy I needed to continue my research. I would love to give a very special thanks to each and every last one of you.

Thank you to the National Research Foundation of South Africa (NRF) for providing the necessary financial backing to pursue this project.

LIST OF ABBREVIATIONS

bp	Base pairs
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
EDTA	Ethylene diamine tetra acetic acid
EST	Expression Tag Sequence
HA	Hemagglutinin
HGNC	HUGO Gene Nomenclature Committee
KDa	Kilo Daltons
mRNA	Messenger RNA
ORF	Open Reading Frame
P2P-R	Proliferation Potential Protein-Related
PACT	p53 Associated Cellular protein-Testis derived
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate Buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein database
Rb /pRb	Retinoblastoma protein
RBBP6	Retinoblastoma binding protein 6
RBQ-1	Rb Binding Q protein-1
SDS	Sodium dodecyl sulphate
SR	Serine/Arginine
SUMO	Small ubiquitin-like modifier
TEMED	*N, N, N', N' – tetramethylethylenediamine*
Ub	Ubiquitin
UBL	Ubiquitin-like
UBP	Ubiquitin-domain protein

LIST OF FIGURES

LIST OF TABLES

UNIVERSITY *of the*
WESTERN CAPE

ABSTRACT

Retinoblastoma Binding Protein 6 (RBBP6) is a 200 KDa protein shown to play a role in 3'-polyadenylation of mRNA transcripts, as well as to function as an E3 ligase catalysing ubiquitination of cancer-associated proteins. RBBP6 has been previously reported to localise to nuclear speckles, which are thought to play a role in mRNA splicing, presumably as a result of its RS domain, which is known to target mRNA splicing factors to nuclear speckles. However recent studies in our laboratory have shown that isoform 3 of RBBP6, consisting mainly of the DWNN domain, also localises to speckles in resting cells, but more strongly in cells subjected to various stresses, suggesting that the DWNN domain may be the speckle-targeting domain.

In order to test this hypothesis, we designed mammalian expression constructs encoding wild type RBBP6, RBBP6-ΔRS, which lacks the RS domain, and RBBP6-ΔDWNN, which lacks the N-terminal ubiquitin-like DWNN domain. All the constructs were fitted with an N-terminal HA tag to facilitate detection using immunofluorescence microscopy.

Our results confirmed that RBBP6 localises exclusively to nuclear speckles, co-localising with the splicing factor SC35, which is known to localise in nuclear speckles. Surprisingly, omitting the RS domain did not abolish the targeting to nuclear speckles, indicating that the RS domain is not required for targeting RBBP6 to nuclear speckles. Even more surprisingly, omitting the DWNN domain made the protein completely cytoplasmic, indicating that the DWNN domain is necessary for localisation in nuclear speckles. However expression of a truncation including the DWNN domain but excluding much of the C-terminus of the protein showed that the DWNN domain is not sufficient for speckle-localisation on its own.
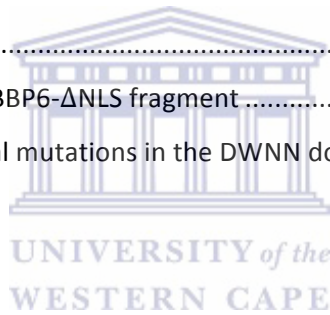
We concluded that the DWNN domain, with the help of the additional nuclear localising sequences on RBBP6, facilitates the translocation of the protein into nuclear speckles.

Keywords: RBBP6, DWNN domain, protein localisation, Site-directed mutagenesis, nuclear speckles

TABLE OF CONTENTS

1   INTRODUCTION

1.1   *Retinoblastoma Binding Protein 6*

The gene that encodes the Retinoblastoma binding protein 6 (RBBP6) was cloned over a decade ago and its protein product was discovered to interact with the tumour suppressor proteins, the retinoblastoma protein (pRb/Rb) and p53 (*Sakai et al., 1995; Simons et al., 1997*). RBBP6 is a 200 kDa multi-domain and multifunctional nuclear protein with involvement in a number of vital biological processes including the induction of apoptosis, mRNA processing, cell cycling regulation and protein turnover through the ubiquitin-proteasome pathway and independently through its conserved N-terminal ubiquitin-like domain (*Sakai et al., 1995; Simons et al., 1997; Pugh et al., 2006; Chibi et al., 2008; Mbita et al., 2012; Pretorious et al., 2011; Di Giammartino et al., 2014*).

Pretorius and co-workers (2011) showed that two promoters are responsible for the transcriptional regulation of RBBP6, namely promoter 0 (P0) and promoter 1 (P1). Both promoters showed, albeit to varying degrees, increased activity following camptothecin-induced apoptosis. Computational analysis suggested that both P0 and P1 transcriptional start sites are located within CpG islands, suggesting the need for additional gene regulation, and that P1, unlike P0 also contained a functional TATA-box. Further analysis of the promoter content indicated the presence of six transcription factor binding sites common to both the promoters and a few additional ones that are unique for each of the two promoters. Functional analysis conducted for all the transcription factors through bioinformatics tools associated P0 more firmly with apoptosis, while P1 was mainly associated with cell cycle regulation and the ubiquitin-mediated proteolysis pathway. In addition, both the promoters were shown to play a role in the antigen-processing and presentation pathway (*Pretorius et al., 2011*).

Initially, RBBP6 was cloned and sequenced during three separate studies, and the sequences entered into sequence data bases under different names. Sakai and co-workers (1995) were able to clone a truncated form of human RBBP6, also called Rb Binding Q protein-1 (RBQ-1) using the Rb protein as a probe. This protein was isolated from a human small cell lung

1

carcinoma (H69c) library as a 140 kDa protein. RBQ-1 was shown to selectively bind to underphosphorylated Rb and its binding to Rb was disrupted by the binding of the E1A protein from Adenovirus. A region of 34 amino acids in the middle of the RBQ-1 sequence was identified to be alternatively spliced. Around the same time, Simons and his co-workers (1997) identified PACT (p53 Associated Cellular protein-Testis derived) through a novel approach for identifying proteins that interact with the tumour suppressor protein p53. Through this approach, cDNA encoding a novel nuclear protein was isolated. The cDNA encoded a 250 kDa cellular protein, containing a putative 34 amino acid region that is alternatively spliced (Figure 1a, boxed area; aa 446-479), a serine/arginine (SR) rich region (Figure 1a, single underlined, aa 475-566), and a highly basic lysine rich C-terminus immediately adjacent to the stop codon (Figure 1a, double underline; aa 1520-1577). This PACT protein showed specific binding to p53 by outcompeting DNA binding of p53-sequence specific transcriptional targets. Sequence alignment and comparison of PACT and RBQ-1 showed that PACT predicted a much longer version of the RBQ-1 protein (1583 amino acids amino acids, compared to the 948 amino acids of RBQ-1), despite the RBQ-1 sequence containing 182bp upstream to the most 5'-sequence of PACT, coding for an additional 30 amino acids up to a putative initiating methionine (*Sakai et al., 1995*). The sequence alignment also revealed that the RBQ-1 open reading frame (ORF) encodes a stop codon (TAA) corresponding to position 919 (amino acids) of PACT. This stop codon was also found on PACT and on an expression sequence tag (EST) fragment that was used, but it was out of frame because of two extra nucleotides. The authors were led to postulate that the possible reason for the short sequence of RBQ-1 was due to a mutational event in the small cell lung carcinoma (Figure 1c, *Sakai et al., 1995*). Simons and co-workers (1997) also postulated a possible function for PACT in cellular pre-mRNA splicing processing based on three features of PACT: firstly was the fact that anti-PACT antibodies localise PACT to nuclear speckles, where a number of known pre-mRNA splicing components reside (*Spector et al., 1993; Simons et al., 1997*). The second implication stemmed from the apparent interaction of PACT with the Sm antigens in nuclear extracts as indicated by co-immunoprecipitation, suggesting that PACT is associated with these splicing factors *in vivo*. The third implication was from the fact that the PACT cDNA encodes an SR region that shares biochemical properties with known SR proteins, most of which are involved in pre-mRNA processing (*Simons et al., 1997*). The RBQ-1/PACT gene was later identified to be localised to

2

chromosome 16p11.2-p12.2 through a combination of polymerase chain reaction (PCR) and fluorescence in situ hybridization (FISH) technique (*Sakai et al., 1995*).

P2P-R (Proliferation Potential Protein-Related) was later identified by Witte and Scott (1997) through library screening with two monoclonal antibodies specific for core heterogeneous nuclear ribonucleoproteins (hnRNPs), suggesting that P2P-R cDNA derived peptides contain at least two hnRNP-related epitopes. The cDNA for P2P-R contains an ORF with the potential to encode for a 1404 amino acids protein with a predicted molecular weight of 156.9 kDa. P2P-R was later established to be an alternatively spliced form of PACT, and that P2P-R appears to be the most predominantly expressed product in multiple murine cell lines (*Scott et al., 2003*). Witte and Scott (1997) also showed that a P2P-R fusion protein consisting of the P2P-R cDNA encoding the hnRNP association is able to bind single-stranded DNA, as do hnRNPs. In addition, they also showed that P2P-R expression is highly reduced during terminal differentiation, and because terminal differentiation is associated with Rb functional regulation, they further presented evidence that the P2P-R cDNA product contains an Rb binding domain that binds to the pocket domain of Rb (*Witte and Scott, 1997*). Other important findings relating to P2P-R have since been published, these include:

I. The increased P2P-R immunoreactivity levels observed in late G2/M cells, when compared to G0 cells appears to occur without any observable change in P2P-R mRNA levels throughout these different cell cycle stages (*Gao et al., 2002*).

II. Gao and co-workers (2002) also showed that P2P-R primarily localises in the nucleoli of interphase murine and human cells and at the periphery of chromosomes in mitotic cells lacking nucleoli. This observation is further collaborated by the fact that p53 can also be localised to nucleoli of interphase cells along with factors that regulate p53 metabolism.

III. Scott and his co-workers (2003) suggested that the possible reason for the markedly increased levels of P2P-R levels during mitosis might be due to post translational modifications when they showed that the SR region of P2P-R can be phosphorylated by the mitotic cdc2 kinase. The SR region of P2P-R is also phosphorylated by SRPK1a.

3

Sequence alignments with ESTs showed that P2P-R is a truncated version of the full length RBBP6 protein (GenBank accession number NP_008841). The published P2P-R sequence lacks the 5'-most N-terminal DWNN domain and Zinc finger domain found on full length RBBP6 (Figure 1). Analysis of the RBBP6 locus suggested that it may give rise to three key transcripts of 6.1, 6.0 and 1.1 Kb generated through the specific combination of alternative splicing and alternative poly-adenylation. These transcripts encode proteins of 1792, 1758 and 118 amino acids, which have been identified from one another as RBBP6 isoforms 1, 2 and 3 respectively (Genbank:NP_008841, Genbank:NP_061173, Genbank:NP_116015; *Pugh et al., 2006*). Isoform 1 is transcribed by the 6.1 Kb transcript, with alternative splicing of the same transcript giving rise to isoform 2. The 1.1 Kb transcript encodes isoform 3 which is also referred to as DWNN-13. The coding region of the RBBP6 gene comprises of 18 exons; with isoform 1 containing all 18 exons, contrary to isoform 2 which has 17 exons (with exon 16 removed on account of alternative splicing) and isoform 3 which comprises of only the first 3 exons (*Pugh et al., 2006*). These three exons encoding DWNN are also present in transcripts 1 and 2, and as a result, all the RBBP6 isoforms consist of a conserved N-terminal domain, DWNN (*Mbita et al., 2012*). The DWNN domain, corresponding to residues 1-81 of the human RBBP6 protein was initially identified through genetic screening that was focused on identifying novel components of the antigen processing and presentation pathway through the major histocompatibility class I (MHC class I) molecules (*Mbita et al., 2012; Pugh et al., 2006*).

4

Figure 1: Predicted amino acid sequence of the PACT cDNA and comparison with human sequences. (a) Predicted amino acid sequence of PACT. The 34 amino acid putative alternative spliced region is boxed. The serine/arginine (SR) region is single-underlined and overlaps with the boxed region by 5 amino acids. The lysine rich region is double-underlined. **(b)** Schematic diagram showing the alignment of the human RBQ-1 and murine PACT. M indicates the putative initiating methionine. **(c)** Alignment of DNA sequences from PACT, RBQ-1 and a human EST fragment. The numbers on the right indicate the position in the PACT DNA sequence. The gaps only present in the RBQ-1 sequence are boxed. These two gaps result in a frameshift in RBQ-1, resulting in the TAA in the stippled box to encode for a stop codon. (Image adapted from *Simons et al., 1997*).

5

1.2   *DWNN*

The gene encoding the DWNN domain was initially isolated from retrovirally-mutagenized Chinese hamster ovary 22 (CHO22) cells using promoter-trapping technology. This gene was isolated due to its apparent role in the cytotoxic T-lymphocyte killing pathway and was later discovered to be homologous to the human cDNA 21c4 (Genbank: T25012) (*A.E. George, D. Phil. Thesis, Oxford university, 1995*). Further analysis of the cDNA database indicated that the gene knocked out by the retrovirus integration also encoded the 200 kDa RBBP6 protein. The corresponding gene was discovered to be located on human chromosome 16p12.2, upstream of the previously identified RBBP6/PACT/P2P-R gene. Complete sequencing of the DWNN gene showed it to be a 1.1 Kb cDNA clone that encodes a 118-150 residue protein composed of the DWNN domain and a short C-terminal tail (isoform 3). Analysis of the cDNA indicated that the sequence coded for the previously unidentified N-terminus of the RBBP6 protein, which was termed DWNN (*Pugh et al. 2006*). Sequence alignment searches using the BLAST engine against all available sequence data indicated that the DWNN domain is present only at the N-terminus of the RBBP6 family of proteins. All of the identified eukaryotic RBBP6 homologues to date are comprised of the DWNN domain, a CCHC Zinc finger domain and RING finger domain, which are found as single copy genes (Figure 2). A protein consisting only of the DWNN, Zinc knuckle and RING finger is the most predominantly expressed RBBP6 isoform in lower eukaryotes. The RBBP6 homologues in vertebrates and insects are longer and consist of additional domains, including the SR domain, an Rb and p53-binding domains. However, no homologous sequences have been identified in prokaryotes (*Pugh et al., 2006*). Comparing the representative DWNN structure against the entire protein database (PDB) indicated that DWNN is homologous to human ubiquitin (PDB: 1UBI), with 18% sequence homology between the two proteins, including the C-terminal di-glycine motif known for its functional role in the covalent attachment of ubiquitin molecules to specific substrates, suggesting that DWNN may function as a ubiquitin-like modifier in a process that our laboratory has coined 'DWNNylation' (*Pugh et al., 2006*). This discovery places the RBBP6 proteins under the superfamily of ubiquitin-like domains; which can be sub-divided into the ubiquitin-domain proteins (UDP's) and ubiquitin-like proteins (UBL's). To date, it appears that DWNN is the first example of an

6

ubiquitin-like domain that is alternatively expressed as both a UBL and as a UDP (*Pugh et al., 2006*).
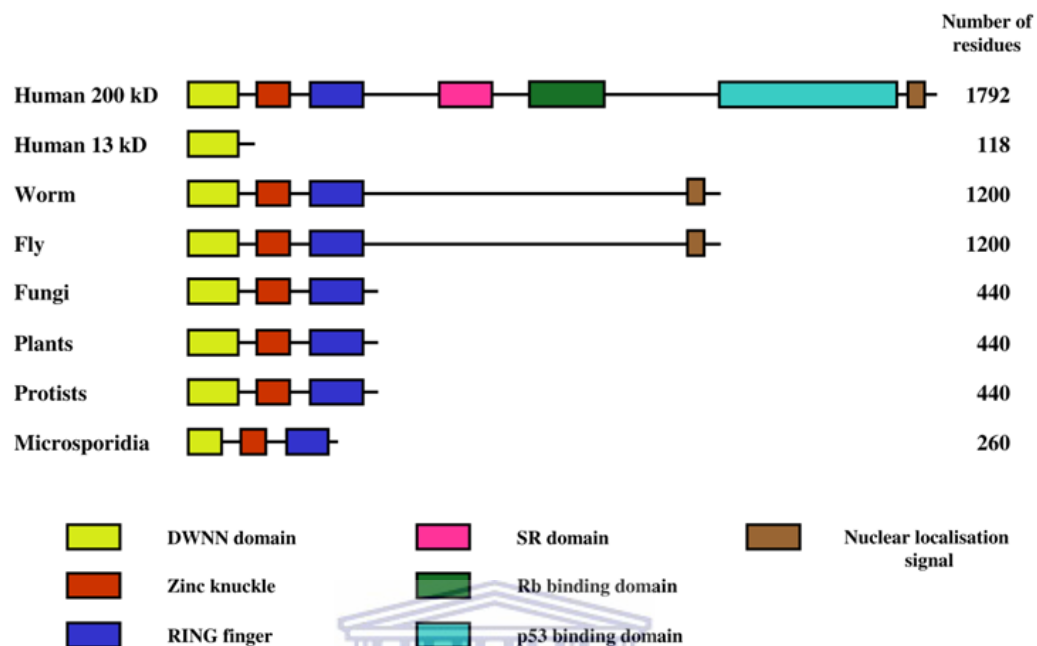
Figure 2: Domain organization of the RBBP6 family of proteins. RBBP6 homologues identified in all complete eukaryotic genomes to date contain the DWNN domain, a zinc knuckle and a RING finger. In higher eukaryotes the protein contains a long C-terminal end consisting of an SR domain, a p53 and Rb binding domains in human and mouse. The two human isoforms shown are isoform 1, which is approximately 200 kDa in length, and isoform 3, which is approximately 13 kDa. Isoform 3 consisting of just the DWNN domain and poorly conserved C-terminal tail has also been identified in higher eukaryotes (Image adapted from *Pugh et al., 2006*).

Additional publications have since been added under the HGNC-approved name RBBP6, indicating a more complex and sometimes overlapping role in RBBP6 functionality. These findings include the following:

I. Chibi and co-workers (2008) have reported that the RING finger domain of RBBP6 confers E3-ligase activity to RBBP6, and that both the full length protein and the isolated RBBP6 RING domain are able to ubiquitinate the Y-box binding protein 1(YB-1) causing its degradation in a proteasome dependent manner. Other findings concerning the RBBP6 RING have reported that RBBP6 may function in an E4-like manner to enhance Hdm2-mediated p53 ubiquitination (*Li et al., 2007*). E3-ligase activity of RBBP6 also plays a role in maintaining genomic stability by ubiquitinating the transcriptional repressor, zBTB38 (*Miotto et al., 2014*).

7

II. Di Giammartino and co-workers (2014) suggested that RBBP6 may function to link pre-mRNA 3'-end processing to different nuclear events through its indispensable role in regulating the human polyadenylation machinery. The study revealed a novel mechanism to regulating 3'-end processing, in which RBBP6 isoform 3 competes with RBBP6 isoform 1 to control cleavage efficiency. The experiments were conducted with both full length RBBP6 and a truncated version consisting of the three most N-terminal domains of RBBP6; the three N-terminal domains are necessary and sufficient for RBBP6 function in cleavage, supporting the idea that 3'-end processing is the primary role of RBBP6 (*Di Giammartino et al., 2014*).

III. These findings correspond with the findings reported by Lee and Moore (2014) on the RBBP6 homologue, MpeI, found in *Saccharomyces cerevisiae* (*S. cerevisiae*). MpeI is comprised of the three most N-terminal domains of RBBP6 known to be conserved in all eukaryotes and has also been shown to be indispensable in mRNA 3'-end processing (*Lee and Moore, 2014*).

IV. Mbita and co-workers (2012) reported on the ability of RBBP6 isoforms to activate different cellular processes, with the over-expression of isoform 3 leading to increased G2/M cell cycle arrest. In contrast, over-expression of RBBP6 isoform 1 in MCF-7 cells caused a decrease in cell viability compared to over-expression of isoform 3 (*Mbita et al., 2012*). The results also showed that over-expression of isoform 3 induced high expression of p53, suggesting that isoform 3 growth inhibitory effects are most likely to occur through a p53-dependent manner (*Mbita et al., 2012*).

V. RBBP6 was shown to be highly expressed in a number of malignant cell types. The expression shows a correlation between RBBP6 protein expression with cell differentiation of cell carcinomas and depth of tumour invasion (*Yoshitake et al., 2004; Motadi et al., 2011; Chen et al., 2013; Morisaki et al., 2014*). Motadi and co-workers (2011) also reported what seems to be a p53-dependent regulation of RBBP6, where transfection of lung cancer cell lines (A549, MRC-5 and Lsq) with siRNA targeting p53 resulted in a markedly increased expression of RBBP6 mRNA in all the cell lines being investigated. In contrast, p53 mRNA showed no observable alterations following treatment with siRNA targeting RBBP6, suggesting that RBBP6

knockdown has no effect on p53 mRNA expression in lung cancer cells (*Motadi et al., 2011*).

VI. Different research groups have suggested that a possible RBBP6 targeting inhibitor with therapeutic potential may be devised, that utilizes RBBP6 expression as a biomarker for subtypes of human lung cancers (*Pugh et al., 2006; Motadi et al., 2011*).

1.3   *Ubiquitin-like proteins*

The eukaryotic ubiquitin family comprises of almost 20 proteins that partake in the posttranslational modification of numerous macromolecules. The ubiquitin-like proteins (UBLs) that form part of this family characteristically adopt the β-grasp fold that is structurally well-documented from ubiquitin (Ub). Despite the shared structural similarities, UBLs regulate a wide range of cellular processes, including proteolysis, nuclear transport, translation, antiviral pathways, and autophagy. The functional diversity of UBL pathways in maintaining cellular homeostasis and physiology continues to expand as more UBL substrates are being discovered and their functional roles elucidated. A number of these UBLs are discussed below, each showing its own unique functionality in contributing to the maintenance of cellular homeostasis and often contributing to larger complex protein structures (*Van der veen and Ploegh, 2012*).
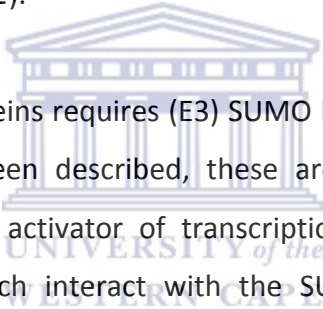
1.3.1   *SUMO-1*

The small ubiquitin-related modifier, SUMO-1 forms part of the growing UBLs involved in posttranslational protein modification. Members of the SUMO gene family appear to be conserved across the eukaryotic kingdom (*Choudhury and Li, 1997; Hanania et al, 1999; Melchior, 2000*). Four different SUMO isoforms, denoted SUMO-1, SUMO-2, SUMO-3 and SUMO-4 have been identified in mammals. SUMO-2 and SUMO-3 are highly similar, sharing almost 86% sequence homology and are therefore often referred as SUMO-2/3 (*Dohmen, 2004*). SUMO-1 appears to be the most predominantly conjugated isoform under physiological conditions, while SUMO-2/3 is preferentially conjugated to proteins under

9

stress conditions such as increased temperature (*Saitoh and Hinchey, 2000*). The recently identified fourth isoform, SUMO-4, is encoded by a sequence that lies within an intron of the human TAB2 gene. The expression of this gene is strongest in kidney cells. SUMO-2, SUMO-3, and SUMO-4 have been shown to contain a SUMO attachment consensus site, which is absent from SUMO-1. Supporting this observation, in contrast to SUMO-1, SUMO-2/3 and SUMO-4 have been shown to form SUMO chains *in vitro* and *in vivo* (*Tatham et al, 2001; Bohren et al, 2004*). SUMO-1 is composed of 101 amino acid residues and has been shown to share 18% sequence homology with ubiquitin, along with structural features, such as the β-grasp fold of the ubiquitin family, and the conserved C-terminal di-glycine motif required for isopeptide bond formation (*Bayer et al, 1998*). The three dimensional structure of human SUMO-1 was determined by nuclear magnetic resonance (NMR) and was found to match the structure of ubiquitin (*Bayer, 1998*). SUMO-1 adopts the characteristic fold of ubiquitin containing five β-strands and two α-helices (amino acids 22-97). However, in contrast to ubiquitin, SUMO-1 also has a long and flexible N-terminal domain (amino acids 1-21) which protrudes from the ubiquitin-like core. The two proteins also have very different surface charge distribution, suggesting that they interact with specific enzymes and substrates. Additionally, ubiquitin Lys48, which is indispensable for the generation of ubiquitin polymers, is replaced in SUMO-1 by Gln69 corresponding to the same position, which may provide an explanation of why SUMO-1 has not been reported to form polymers (*Bayer et al, 1998*). Bylebyl and co-workers reported on a lysine residue within the N-terminal domain of SUMO-1 which is implicated in the formation of poly-SUMO chains in yeast. Surprisingly, the entire extension including this lysine residue can be deleted without severe consequences for the yeast, indicating that, in contrast to ubiquitin, chain formation is not important for SUMO-1 function in *S. cerevisiae* (*Bylebyl et al, 2003*).

The conjugation of SUMO to its substrates, referred to as sumoylation, is carried out in an analogous fashion to ubiquitination. Similar to ubiquitin, the C-terminus of mature SUMO requires an activation step for isopeptide bond formation with the substrate lysine residues. This task is accomplished by the SUMO-activating enzyme (E1). The E1 in both yeast and humans is a heterodimer composed of the AoS1 and Uba2 proteins with sequence identities similar to the N-and-C-terminal parts, respectively, of ubiquitin-activating enzymes (*Dohmen et al, 1995; Johnson et al, 1997; Azuma et al, 2001*). Both the subunits for the heterodimeric

E1 are required for SUMO activation, despite the fact that the Uba2 subunit of the dimer contains the active site cysteine indispensable for SUMO activation. The activated SUMO is transferred from the Uba2 subunit of the SUMO-activating enzyme via a transesterification reaction to a monomeric SUMO-conjugating enzyme (E2), Ubc9 (*Johnson and Hochstrasser, 1997*). The structure of Ubc9 contains small insertions, which play critical roles in its interaction with SUMO, the SUMO-specific E1, and with SUMO substrates containing a sumoylation consensus sequence (as was observed with the sumoylation of RanGap1), ψKxE, where ψ is a hydrophobic residue, K is the SUMO acceptor site and E is an acidic residue (*Okuma et al, 1999; Melchoir et al, 2003*). A number of van der Waals interactions between Ubc9 residues and residues found in both the hydrophobic and the acidic parts of the consensus sequence stabilize the interaction between SUMO-E2 and substrate, which may provide an explanation as to why RanGap can be sumoylated in the absence of a SUMO ligase (*Bernier-Villamor et al, 2002*).

Sumoylation of other target proteins requires (E3) SUMO ligase activity. To date, only three types of SUMO ligases have been described, these are the PIAS (protein inhibitor of activated signal transducer and activator of transcription) family, RanBP2 (Ran binding protein 2) and PC2, all of which interact with the SUMO-conjugating enzyme, Ubc9, enhancing the potency of sumoylation both *in vivo* and *in vitro*. The PIAS family of SUMO ligases contain a RING-finger-like structure and are capable of binding directly to Ubc9 and to some SUMO target proteins to promote their sumoylation. RanBP2 SUMO ligase contains a 30 kDa domain housing its E3 catalytic activity, and was itself identified as one of the first SUMO targets (*Saitoh et al., 1998*). This E3 ligase does not display any obvious sequence similarities to PIAS-type SUMO ligases or ubiquitin ligases. RanBP2 is localised at the cytoplasmic filaments of the nuclear pore complex (NPC), where it interacts with sumoylated RanGAP and the GTPase Ran. PC2 forms part of the polycomb group (PcG) protein family. PcG proteins play a role in gene silencing (*Wang et al., 2016*).

Similar to ubiquitination, sumoylation is a reversible process. De-sumoylation is carried out by a set of ubiquitin-like specific proteases (ULPs). These ULPs play important roles at two significant steps in the SUMO cycle. Firstly, they cleave the precursor molecule to expose the C-terminal di-glycine motif required for SUMO conjugation. Secondly, the ULPs function

as isopeptidases to release SUMO from the conjugates so it can be used in subsequent conjugation cycles. Two ULPs, Ulp1 and Ulp2/Smt4, were identified in yeast (*Li and Hochstrasser, 1999; 2000*). Both Ulp1 and Ulp2 share sequence similarity in a region of 200 amino acids at the C-terminus, called the ULP domain, which contains a Cys-His-Asn catalytic triad (*Muller et al., 2000; Kim et al., 2002; Melchior et al., 2003*). Ulp1 was shown to be essential for cell viability in yeast, unlike Ulp2 which is not essential for cell viability, but is required for maintaining proper cellular growth without any temperature associated hindrances, chromosomal and cell-cycle progression defects (*Li and Hochstrasser, 1999, 2000; Schwienhorst et al., 2000*). Seven mammalian ULPs have since been identified, with amino acid residues spanning from 238 to 1112 amino acids (*Yeh et al., 2000*). The C-terminal conserved region found in yeast is also found in human ULPs. Mammalian ULPs show a varied localization pattern including the nucleus, nuclear bodies, the nuclear pore complex and the cytoplasm (*Melchior et al., 2003*).

Unlike ubiquitin, SUMO does not target proteins for degradation, but appears to be involved in the modulation of protein-protein interactions. Independent studies demonstrate an essential function of SUMO in protein trafficking between the cytosol and nucleus, such as the sumoylation of RanGap1 (*Mahajan et al., 1997*). In mammals, aside from modifying RanGap1, a substantial fraction of SUMO conjugates are associated with microscopically detectable sub-nuclear structures known as PML (promyelocytic leukaemia protein) nuclear bodies (*Sternsdorf et al., 1997; Duprez et al., 1999*). The integrity of PML bodies appears to be of great significance to maintaining normal cell growth and development, since their disruption is associated with human diseases such as acute promyelocytic leukaemia (APL) and spinocerebellar ataxia type 1 (SCA1). Sumoylation is also implicated in the regulation of transcription factors such as the activation of the transcriptional activity of p53 that occur post-sumoylation (*Gostissa et al., 1999; Rodriguez et al., 1999; Muller et al., 2000*). Sumoylation can even serve as an inhibitor of ubiquitin-mediated protein degradation, as was discovered in studies done on the mammalian inhibitor of the inflammatory response; IκBα (*Desterro et al., 1998*). Sumoylated IκBα is resistant to TNFα-induced ubiquitin-dependent degradation.

1.3.2   NEDD-8

In 1992, Kumar and co-workers identified 10 genes that showed developmental downregulation of their expression in the mouse brain and were named NEDD1-10 (neural precursor cell expressed, developmentally downregulated 1-10; *Kumar et al., 1992*). NEDD-8 was subsequently elucidated to be a UBL, showing the most similarity to ubiquitin at both the sequence and secondary structure levels amongst all the known UBLs (*Kamitani et al., 1997; Whitby et al 1998*). NEDD-8 is an 81 amino acid polypeptide displaying 60% sequence identity and 80% homology to ubiquitin. The protein is highly conserved in mammalian species, displaying only a single amino acid difference between the human and mouse NEDD-8. Furthermore, its C-terminus contains the Leu-Arg-Gly-Gly residues required for conjugation to specific target proteins. The Lys48 residue responsible for multi-ubiquitin chain formation is also conserved in NEDD-8 (*Kamitani et al., 1997*). Taking into account the similarities between NEDD-8 and ubiquitin, Kumar and co-workers (1992) hypothesized that NEDD-8 could also be conjugated to other proteins in an enzymatic cascade analogous to the enzymatic cascade system observed for ubiquitin conjugation. NEDD-8 was later reported to be activated and transferred to other proteins in a process analogous to ubiquitination; a process referred to as "Neddylation". Furthermore, NEDD-8-tagged proteins appear to predominantly localise in the nucleus, unlike ubiquitin tagged proteins which display a more even distribution between the nucleus and cytoplasm.

Similar to ubiquitin, NEDD-8 is initially synthesized as an inactive precursor molecule that needs to be processed at the conserved carboxyl-terminal Gly76 residue by the hydrolase activity of deneddylating enzymes, exposing the di-glycine motif required for neddylation. Two NEDD-8 processing enzymes have been reported; NEDP1 (also referred to as DEN1 or SENP8) and UCHL3 (*Wada et al., 1998; Mendoza et al., 2003*). NEDD-8 conjugation proceeds in a three-step enzymatic process similar to ubiquitination, using E1, E2 and E3 enzymes specific for neddylation. The NEDD-8 E1, NAE (NEDD-8 activating enzyme), is composed of the NAE1-UBA3 heterodimer, with NAE1 and UBA3 displaying homology to the amino-and-carboxyl-terminal regions, respectively, of the ubiquitin E1 UBE1. Despite the similarities between the two UBLs and their E1s, specificity appears to be maintained for each UBL and its corresponding activating enzymes through repulsion forces occurring between Arg190 on

13

UBA3 and Arg72 on ubiquitin. The Arg72 residue on ubiquitin, corresponding to Ala72 on NEDD-8, has been shown to be crucial for E1 specificity (*Burch and Hass, 1994; Souphron et al., 2008*). Despite these regulatory features, however, NEDD-8 activation by the ubiquitin E1 UBE1 is possible and has been reported under situations where the NEDD-8: Ubiquitin balance in cells is disturbed either by overexpression of NEDD-8 or following physiological depletion of ubiquitin (*Hjerpe et al., 2012; Leidecker et al., 2012*).

There are two known NEDD-8 E2s; UBE2M and UBE2F (*Liakopoulus et al., 1998; Osaka et al., 1998; Gong and Yeh, 1999; Huang et al., 2009*), which preferentially couple with the E3 RBX1 (also known as ROC1) or RBX2 (also known as ROC2 or RNF7), respectively. RBX1 and RBX2 are RING E3s which simultaneously bind their substrate, NEDD-8 and the E2 to catalyse neddylation (*Scott et al., 2010*). Interestingly, RBX1 and RBX2 can also function as the ubiquitin E3s for cullin-RING complexes (*Kamura et al., 1999; Petroski and Deshaies, 2005*). An additional group of proteins, Dcn1 (defective in cullin neddylation 1) in budding yeast and DCUN1D1-5 (defective in cullin neddylation 1, domain-containing 1-5) in humans have been designated as functional cofactors for RBX1 and RBX2, potentiating their neddylation activity (*Kurz et al, 2005 & 2008; Kim et al, 2008, Meyer-Schaller et al, 2009; Scott et al., 2010*).

As is the case with most post-translational modifications (PTMs), neddylation is a reversible process, with the CSN complex being the predominant de-neddylase. The CSN is a conserved complex of eight proteins (CSN1-8), named after a group of COP (constitutive photomorphogenesis) mutants that displayed constitutive photomorphogenesis, pigment seed coats and premature death in plants (*Wei and Deng, 1992*). CSN5 contains a JAMM motif, which was later shown to be indispensable for the de-neddylating activity of the CSN and associated the CSN to the regulation of the cullin-RING ligases (CRLs) (*Cope et al., 2002*). The CSN is the only isopeptidase that efficiently cleaves NEDD-8 from cullins substrates *in vivo*, although a number of other proteins have been reported to possess NEDD-8 isopeptidase activity *in vitro* (*Rabut and Peter, 2008*).

One of the most studied neddylation substrates are the cullins, of which there are eight encoded by the human genome (CUL1, 2 ,3, 4A, 4B, 5, 7and PARC), and they all share an

evolutionary conserved cullin homology domain and a carboxyl-terminus neddylation site. The APC2 subunit of the anaphase promoting complex/cyclosome (APC/C) also contains cullin homology domains, but lacks the carboxyl-terminal NEDD-8 consensus sequence (IVRIMKMR) and is therefore functionally distinct from the other cullins. Cullins are molecular scaffolds for the CRLs, which are E3 ubiquitin ligases responsible for up to 20% of all ubiquitin-mediated proteasomal degradation (*Soucy et al., 2009; Sarikas et al., 2011; Lydeard et al., 2013*). A number of non-cullin substrates for NEDD-8 have been reported in literature, including p53 (*Xirodimas et al., 2004*), the ribosomal protein L11 (*Sundqvist et al., 2009*), and histone H4 (*Ma et al., 2013*), adding to the pool of data representing NEDD-8 as a key protein in a number of regulatory cellular processes.

1.3.3 SAP18

SAP18 was initially identified as one of two novel proteins complexed with the mammalian transcriptional repressor Sin3 (mSin3) and histone deacetylases (HDAC) 1 and 2 (*Zhang et al., 1997*). This study was targeted at elucidating the significance of mSin3 in regulating gene expression, and to accomplish this, a Sin3-containing complex was immune-purified from human cells. The purified complex showed that mSin3 was found associated with HDAC1 and HDAC2, as well as with the histone tethering polypeptides, RbAp48 and RbAp46, and two novel mSin3-associated polypeptides, SAP18 and SAP30. In light of the discovery of these mSin3-associated polypeptides, two peptide sequences derived from the SAP18 polypeptide were obtained and used to isolate a full length cDNA clone encoding full length SAP18. The nucleotide sequence encoding SAP18 produced an open reading frame encoding a polypeptide of 153 amino acids with an estimated molecular mass of 17.56 kDa. A search of the GenBank database identified potential SAP18 homologs in both *C. elegans* and in the yeast genomes, but at this point, the sequence of SAP18 still gave no insight to its functional role in the mSin3-HDAC complex. However, like mSin3, it was abundantly clear that SAP18 was important enough to be ubiquitously expressed across all tissue types analysed (*Zhang et al., 1997*). The authors attempted to specify the role played by SAP18 in mSin3-mediated transcriptional repression. To this end, GST-pull downs and transient transfection assays indicated that SAP18 interacts with both mSin3 and HDAC1 both *in vitro* and *in vivo*. Since

15

there exist an interaction between mSin3 and SAP18, the authors suggested that tethering SAP18 to a promoter would effectively constitute a transcriptional repressing complex through the recruitment of mSin3-containig complex. Zhang and co-workers were able to show that tethering HDAC1 or SAP18 to a promoter by making use of DNA binding domains resulted in transcriptional inhibition and hence, they were able to show that SAP18 forms an integral component of the mSin3-repressing complex and can enhance mSin3-mediated transcriptional repression (*Zhang et al., 1997*).

Following the discovery of the interaction between SAP18 and mSin3, a number of studies reported that SAP18 has been identified to bind a number of transcription factors (*Espinas et al., 2000; Zhu & Hanes, 2000; Cheng & Bishop, 2002*). While the functional role of SAP18 was still not clearly defined, it was clear that SAP18 functioned as a protein-protein adapter that bridges the transcription factors. A study reported by McCallum and co-workers (2006) provided the nuclear magnetic resonance (NMR)-based structure of SAP18, and went further to speculate on the basis of the SAP18 structure and structure-function relationships based on structurally related proteins, and their functional roles as protein adapter molecules within larger molecular assemblies. An automated search of the Protein Data Bank (PDB) structure repository using a SAP18 fragment spanning amino acid residues 15-149 revealed close structural similarities to ubiquitin-like proteins and other related members of the large β-grasp fold class. The SAP18 β-grasp fold consists mostly of an anti-parallel five-stranded β-sheet curled around a central α-helix. Ubiquitin and SAP18 share high levels of structurally conserved features despite the given 11% sequence identity between the two proteins.

To solidify the theory that the SAP18 ubiquitin-like β-grasp fold is a protein motif involved in protein-protein interactions and the assembly of multiprotein complexes, SAP18 was identified as an integral component of the apoptosis-and splicing-associated protein (ASAP) complex, along with RNPS1 and Acinus (*McCallum et al., 2006*). In contrast to SAP18, both RNPS1 and Acinus display characteristic hallmarks of splicing regulatory proteins such as the RNA-binding motifs and the presence of RS and related domains, all of which are known to regulate alternative splicing by modulating spliceosome assembly and splice site choice. Singh and co-workers (2010) elucidated the specific role played by SAP18 in the ASAP

complex by making use of tethered function assays, mutational studies, and indirect immunofluorescence analysis. The collective results from these studies indicated that the ubiquitin-like fold of SAP18 was required for efficient splicing regulatory activity of the entire ASAP complex. The results indicated that the ubiquitin-like fold of SAP18 provided an interaction surface required for splicing modulation via assembly of a nuclear speckle-localised splicing regulatory protein complex containing RNPS1 and Acinus (*Singh et al., 2010*).

1.4   *Nuclear speckles*

Nuclear speckles are domains located in the interchromatin nuclear space of mammalian cells. These nuclear clusters are well known as the enrichment sites for pre-mRNA splicing factors, from small nuclear ribonucleo-proteins (snRNPs) to spliceosome subunits and other auxiliary non-snRNP protein splicing factors such as SR proteins (*Spector and Lamond, 2011*). Based on proteomic analysis, speckles contain a diverse composition of proteins apart from pre-mRNA splicing factors, including 3'-end RNA processing factors (*Krause et al., 1994*), transcription factors (*Larsson et al., 1995*), eukaryotic translation initiation factor eIF4E (*Dostie et al., 2000*), eIF4I II, a protein that plays a role in the inhibition of translation (*Li et al., 1999*), and structural proteins (*Jagatheesan et al., 1999; Sharma et al., 2010*). Based on their apparent reconfiguration upon inhibiting the transcriptional process, becoming more enlarged and rounded due to the accumulation of splicing factors; speckles are suspected to function more as storage/assembly/modification compartments that supply splicing factors to active transcription sites and not the direct sites of transcription themselves (*Spector and Lamond, 2011*).

Additional studies conducted with electron microscopy identified two morphologically distinct nuclear structures that constitute nuclear speckles: interchromatin granule clusters and perichromatin fibrils. The nuclear localization of splicing factors is dynamic, as shown by the fact that nuclear speckles can reorganize themselves when transcription and /or splicing activity is inhibited by either the use of chemical inhibitors or as a result of heat shock (*Spector et al., 1983; O'Keefe et al., 1994*). The most intense speckled regions are the ones adjacent to active transcriptional sites, as indicated by fluorescence labelling of nascent

17

RNAs (*Wansink et al., 1993*) and by in situ hybridization (*Huang and Spector, 1991; Xing et al., 1993, 1995*). The observation that splicing does not appear to occur at the majority of sites where splicing factors are most concentrated (*Huang and Spector, 1996*) provided the basis for the notion that splicing factors shuttle between interchromatin granule clusters (serving as storage and/or assembly and modification compartments) and perichromatin fibrils (sites of transcription and splicing) (*Jiménez-Garcia and Spector, 1993*).

The first detailed description of these nuclear bodies was published by Ramón y Cajal (1910), in which he referred to these nuclear bodies as "grumos hialinas" (directly translating to "translucent clumps"). Work later conducted by Swanson Beck brought about the term "speckles" and showed that these speckles were not randomly distributed, but rather occurred in localised "clouds" (*Beck, 1961*). By utilizing cytochemical analysis, Swift also observed that these speckles contained RNA (*Swift, 1959*). The initial connection between pre-mRNA splicing and nuclear speckles or interchromatin granule clusters came to light during the examination of the distribution pattern of snRNPs using anti-splicing factor-specific antibodies, displaying a speckled distribution of snRNPs in cell nuclei (*Spector et al., 1983*). It later became clear that much of the punctuate localization attributed to splicing factors as viewed under immunofluorescence microscopy correspond to the presence of these factors in nuclear speckles of variable sizes and irregular shape.
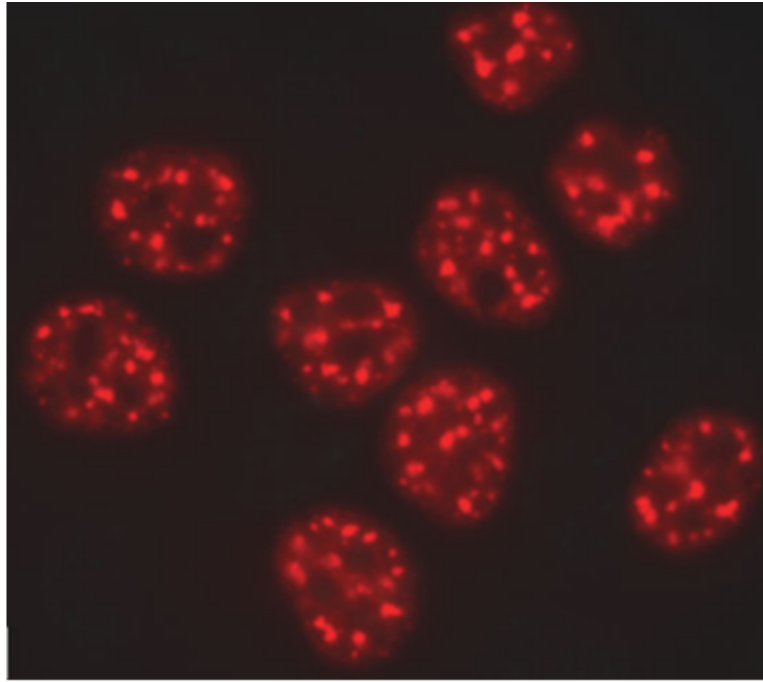
Figure 3: Speckles form in the interchromatin space of mammalian cells. Hela cells displaying splicing factors localised in a speckled pattern along with a diffusely distributed pattern across the nucleoplasm (Image adapted from *Spector and Lamond, 2011*).

Some of the speckle components have been shown to have a speckle targeting motif. The arginine/serine-rich domain (RS domain) of some SR pre-mRNA splicing factors has been shown to be necessary and sufficient for the targeting of these factors to nuclear speckles (*Li et al., 1991; Hedley et al., 1995; Caceres et al., 1997*). In addition, the forkhead-associated domain in NIPP1 (*Jagiello et al., 2000*) and the threonine-proline repeats of SF3b1 (*Eilbracht and Schmidt-Zachmann, 2001*) have also been implicated in targeting these specific proteins to nuclear speckles. Despite the focus of speckle-related studies being conducted on mammalian cells, structures similar to nuclear speckles have been identified during transcription in both *Drosophila melanogaster* (*D. melanogaster*) embryos (*Segalat and Lepesant, 1992*) and in amphibian oocyte nucleus (*Gall et al., 1999*), but not in yeast (*Potashkin et al., 1990*).

Importantly, not all nuclear proteins that are represented with the speckling-like pattern by immunofluorescence microscopy localise to interchromatin granule clusters (IGCs). Nayler and co-workers (*Nayler et al., 2000*) showed that the ER repeat protein YT521-B localises in a speckle-like distribution that corresponds to YT bodies. The protein PSPC1 also localises in punctate interchromatin structures referred to as "paraspeckles" (*Fox et al., 2002; Fox and*

19

Lamond, 2010). For this particular reason, it becomes necessary to conduct double labelled immunofluorescence using anti-speckling control antibodies such as anti-SC-35 to confirm the localization of any novel factors to nuclear speckles.

1.5   *SR proteins*

Pre-mRNA splicing occurs within a macromolecular complex, referred to as the spliceosome. The process is catalysed by several proteins and ribonucleoprotein components, with the major constituents of the spliceosome being the small nuclear ribonucleoprotein particles (snRNPs) U1, U2, and U4/6-U5; the polypeptides that associate with hnRNA to constitute hnRNP particles (hnRNP proteins); and a large complex of non-snRNP splicing factors, including the SR family of proteins (*Caceres et al., 1997*). These SR proteins constitute a family of structurally and functionally related non-snRNP splicing factors with indispensable roles in constitutive splicing and a well-documented regulatory role in alternative splicing by influencing the selection of 5'splice sites (*Ge and Manley, 1990; Krainer et al., 1990*). Another group of related RS domain-containing proteins, often referred to as SR protein-related polypeptides or SR-like proteins, some of which are known to contain RNA recognition motifs (RRMs), also play a role in splicing regulation (*Fu, 1995; Graveley, 2000*). SR family and SR-like proteins also play significant roles in the recognition of exonic splicing enhancers, promoting the activation of otherwise inefficient upstream 3' splice sites (*Cazalla et al., 2002*). SR proteins are characterized by a domain rich in alternating serine (S) and arginine (R) residues (RS domain), one or two RNA Recognition motifs (RRMs) and subcellular localization to nuclear speckles (*Hedley et al., 1995*). Studies have shown that the RS domain is essential for a number of different protein functions, these include mediating protein-protein (*Manley and Tracke, 1996*) and RNA-protein interactions (*Graveley, 2004*), to influence nuclear import (*Lai et al., 2001*) and to play a role in directing proteins such as SC35 and Transformer to nuclear speckles (*Li and Bingham, 1991*). RS domains from SR proteins, non-SR proteins and artificial domains have been reported to activate splicing (*Philipps et al., 2003*).

On the other hand, the RS domain does not appear to function as a nuclear import and localization signal for all RS domain proteins, as SF2/ASF and SRp40 can localise to nuclear

speckles even without their respective RS domains (*Cáceres et al., 1997*). With regards to non-shuttling SR proteins such as SC35, it appears that the RS domain imparts a localised nuclear retention signal as was shown with heterokaryon assays conducted with chimeric proteins in which the SC35 RS domain was fused to the shuttling protein, hnRNP A1 (*Cazalla et al., 2002*). The RS domain's ability to influence sub-nuclear localization appears to be tied to the rest of the protein domains, more specifically, the RRMs. In the context of proteins with two RRMs, such as SF2/ASF, the RS domain is not required for localization into nuclear speckles, whereas the RS domain appears to be indispensable for localization into nuclear speckles for proteins with a single RRM, such as SC35 (*Cáceres et al., 1997*). The subcellular localization of SR proteins is most likely dependant on the phosphorylation state of their RS domains, which most likely influences both nuclear import and export (*Sanford and Bruzik, 2001; Huang et al., 2004*). SR trasportins (TRN-SR) have also been identified in both mammals (*Kataoka et al., 1999; Lai et al., 2001*) and Drosophila that mediate SR-protein nuclear import (*Allemand, 2002*). More evidence has been presented, linking RS domain phosphorylation with mRNA export (*Gilbert and Guthrie, 2004*) and RNA binding specificity (*Tacke et al., 1997*).
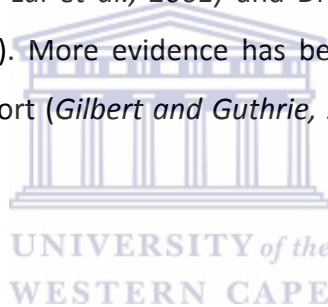
1.6   *Nuclear translocation*

The sequestering of genetic entities within the nucleus in eukaryotic cells offers a powerful mechanism for the regulation of gene expression and other cellular process, such as signal transduction and cell cycle progression through the selective translocation of proteins through the nuclear membrane (*Hodel et al., 2001; Lange et al., 2007*). The mechanism of regulated protein translocation across the nuclear membrane requires the identification of specific sequences that mark a macromolecule complex for nuclear import or export. This protein transport is mainly mediated by transport factors in the karyopherin-β superfamily, which are also referred to as importins or exportins depending on which directional pathway they function in (*Hodel et al., 2001*). The direction of nuclear-cytoplasmic translocation is regulated by the identified nuclear targeting or nuclear export sequences within the cargo proteins. Nuclear localisation signals (NLSs) direct proteins into the nucleus, whereas nuclear export signals (NESs) directing export of proteins from the nucleus to the cytoplasm (*Xu et al., 2012*).

1.6.1 Nuclear import

The best characterized pathway of nucleocytoplasmic transport is the classical nuclear import pathway *(Dingwall and Laskey, 1991)*. In this pathway, importin α identifies and binds cargo protein in the cytoplasm, connecting it to the β-karyopherin, importinβ. Importin β then mediates interaction of the complex with the nuclear pore as it translocates into the nucleus. Once the import complex reaches the nucleus, it is separated by RanGTP and importin α redirected into the cytoplasm for another round of import *(Lange et al., 2007)*.

1.6.1.1 Nuclear localisation signals

The first step of nuclear import occurs when a importin recognizes its cargo amongst other cellular proteins. Proteins intended for transport into the nucleus contain amino acid targeting sequences called nuclear localisation signals (NLSs). The best understood transport signal for nuclear import is the classical NLS (cNLS), which consists of either one (monopartite) or two (bipartite) stretches of basic amino acids *(Lange et al., 2007; Nie et al., 2007)*. Monopartite cNLSs are exemplified by the SV40 large T antigen NLS (^{126}PKKKRRV^{132}) and bipartite cNLSs are exemplified by the nucleoplasmin NLS [^{155}KRPAATKKAGQAKKKK^{170}] *(Lange et al., 2007)*. Consecutive residues from the N-terminal lysine of the monopartite NLS are referred to as P1, P2, etc. Structural and thermodynamic studies have identified many of the crucial requirements for a cNLS *(Fontes et al., 2000; Hodel et al., 2001)*. These studies show that a monopartite cNLS requires a lysine in the P1 position, followed by basic residues in positions P2 and P4 to yield a general consensus sequence of K(K/R)X(K/R). Hodel and co-workers presented data to establish boundaries for the binding capacity of a cNLS to importin α, where a functional cNLS has a binding constant of $\approx$10 nM. Non-functional near-cNLS sequences either binds importin α too weakly to be effectively to be translocated into the nucleus or too tightly for the complex to be efficiently dissociated by RanGTP in the nucleus *Hodel et al., 2001)*.

1.6.2 Nuclear export

The importin/karyopherin superfamily facilitates both nuclear import and export by interacting specifically with either a nuclear localisation signal or nuclear export signal (NES) contained in cargo proteins. The conventional nuclear export pathway is mediated by an

evolutionary conserved CRM1/exportin protein that belongs to the importin β family. The CRM1-RanGTP complex binds directly to the NES and directs the export of the protein complex from the nucleus *(Kosugi et al., 2008)*. The cargo is released from the complex following hydrolysis of RanGTP to RanGDP in the cytoplasm *(Ossareh-Nazari et al., 2001; Fornerod and Ohno, 2002)*.

1.6.2.1   *Nuclear export signals*

Nuclear export signal (NES), sometimes referred to as leucine-rich NESs due to the high number of leucine residues in the first identified proteins *(Fischer et al., 1995; Wen et al., 1995)* typically contain large hydrophobic conserved residues separated by a variable number of amino acids, represented by the consensus sequence L-X(2,3)-[LIVFM]-X(2,3)-L-X-[LI], where X(2,3) represents any two or three amino acids *(Bogerd et al., 1996)*. Despite the general sequence, a large portion of the proteome contains this consensus sequence, only 36% of experimentally defined NESs follows this consensus *(la Cour et al., 2003; 2004)*.
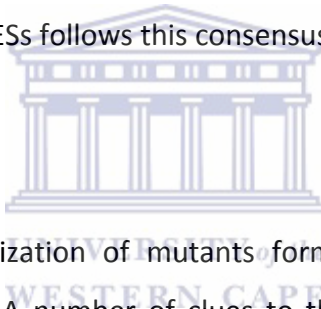
1.7   *Mutagenesis*

The construction and characterization of mutants forms the basis for any studies on structure-function relationships. A number of clues to the way in which macromolecules function can be ascertained from the knowledge of their three-dimensional structure, but proof that a specific mechanism has been elucidated necessitate the analysis of mutants that have nucleotide or amino acid changes at specific residues. Classically, mutants are generated by treating the test organism with mutagens. This method of generating mutants has been extremely successful, but is prone to a number of disadvantages. First, any gene in the organism is susceptible to the mutagens and the rate at which mutations occur in the target gene may be very low. To overcome this, selection strategies had to be developed. Second, even after mutants with the desired phenotype have been selected, there is no certainty that the mutation occurs in the target gene. Third, prior to the advancements in gene-cloning and sequencing techniques, there was no way of pinpointing where exactly in the gene the mutation had occurred. With the advancements in molecular biology, mutagenesis techniques have also been improved upon. Instead of the standard approach, it is now possible to target specific nucleotide bases within a cloned DNA sequence. This

technique is known as site-directed mutagenesis. The significance of site-directed mutagenesis extends way beyond gene structure-function relationships, allowing for mutant proteins with novel properties to be constructed (*Primrose & Twyman, 2006*).

The first method for site-directed mutagenesis to be developed was the single-primer method (*Gillam et al., 1980; Zoller & Smith 1983*). This technique involves *in vitro* DNA synthesis carried out with a chemically synthesized oligonucleotide that contains a base mismatch with the template sequence. This technique requires that the DNA to be mutated be readily available in its single-stranded form, and the gene of interest would be cloned into M13-based vectors to facilitate this condition. The synthetic oligonucleotide primes DNA synthesis and is itself incorporated into the resulting heteroduplex molecule. After transformation into the host *E. coli*, this heteroduplex subsequently gives rise to homoduplexes whose sequences are either that of the initial wild-type DNA or that of newly synthesized mutant DNA. The rate at which the mutated clones arise, in correlation to wild-type clones, may be low. For screening purposes, a number of selection strategies for picking out mutants over wild-type clones have been developed each with their own set of stringencies to follow (*Wallace et al., 1981; Traboni et al., 1983*). It is wise to check the sequence of the mutant directly by DNA sequencing, in order to confirm that the procedure has not introduced an incorrect or additional mutation.

The proficiency of the single-primer method is dependent on a number of factors. The double-stranded heteroduplex molecules produced will co-exist with any single-stranded template DNA that remained uncopied and any partially double-stranded DNA molecules. The presence of these DNA entities drastically reduces the population of the mutant progeny. These can be removed by agarose gel electrophoresis or by sucrose gradient centrifugation, but the process is time consuming. Following transformation into the host *E.coli*, *in vivo* DNA synthesis and segregation takes place, producing a mixed population of mutant and wild-type progeny. The newly synthesized mutants have to be purified from wild-type molecules, and this process is made difficult by the cell's mismatch repair system. The mismatch repair system of *E. coli* favours the repair of non-methylated DNA, and hence the non-methylated *in vitro*-generated DNA strands are preferentially repaired at the position corresponding to the mismatch, effectively eliminating a mutation (*Kramer et al.*

1984). Mutant host strains carrying the *mutH*, *mutL*, or *mutS* mutations can be used to overcome the problem associated with the mismatch repair system of *E. coli*. Various strategies have been developed with the expressed goal of supressing the growth of non-mutants, and hence improve the selection process for mutant progeny (*Kramer, 1984; Kunkel, 1985; Sayers & Eckstein, 1991*). Another disadvantage attributed to all the primer extension methods is the specific requirement that template DNA is single-stranded, instead of the more stable and easy to prepare double-stranded DNA templates used in PCR-based mutagenesis techniques.

The initial stages in the development of the PCR method for DNA amplification displayed its potential for mutagenesis. Single bases mismatched between the synthetic primer and the template DNA become incorporated into the template sequence following amplification cycles. Higuchi and co-workers (1988) reported on a variation of the standard method which allows for a mutation in a PCR-produced DNA fragment to be introduced anywhere along its length. This method requires four primers and three PCRs (a pair of PCRs to amplify the overlapping fragments and another PCR to join the two fragments). Commercial suppliers for mutagenesis reagents have developed much simpler methods making use of the two fundamental features of PCR-based mutagenesis. First, the technique is not limited to single base changes, but can also be used to make insertions and deletions, provided that the right set of primers are selected. Second, to avoid extraneous mutations being introduced into the amplified DNA product, a range of high fidelity thermostable polymerases, along with high template concentration, and fewer cycles can be used. Two commercial mutagenesis kits exist: the Excite and GeneTailor methods (Life Technologies, Carlsbad, California, USA).

In the Excite method, both the strands of the vector consisting of the gene of interest are amplified using a primer set with only one of the primers carrying the desired mutation. This inevitably produces a population of linear duplexes containing the mutated gene that is contaminated with low levels of the initial circular template DNA. If the template DNA was derived from an *E. coli* strain with an intact restriction modification system, then it will be methylated and thus be susceptible to restriction by the *DpnI* endonuclease. The newly synthesized linear DNA will be resistant to *DpnI* cleavage and following circularization by blunt-end ligation can be recovered by transformation into *E. coli*. Any hybrids consisting of

the *in vitro*-generated DNA and the methylated template DNA will also be sensitive to the endonuclease activity of *DpnI*.

In the GeneTailor method, the target DNA is methylated *in vitro* prior to the mutagenesis step using overlapping primers. This amplification process also produces linear amplicons that carry the desired mutation but in this case there is no cleavage with endonucleases, instead the DNA is transformed directly into *E. coli*. The host cell repair system circularizes the linear DNA while the *Mcr*BC endonuclease cleaves the methylated template DNA leaving behind only the unmethylated, mutated PCR-product.

The techniques described above are all specific for the construction of defined mutations within a gene and thus are of particular use in determining structure-function relationships. However, techniques exist for the researcher with the expressed objective of creating a mutant library at random and then selecting for mutants with altered and/or improved properties (*Primrose & Twyman, 2006*).
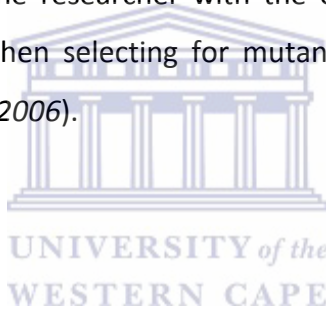
1.8 *Aims of the project*

Previous investigations of the localisation of RBBP6 isoforms carried out in our laboratory used Green Fluorescent Protein attached to the N-terminus, and mCherry attached to the C-terminus. The results, however, were not consistent with results obtained using endogenous RBBP6 detected using protein-specific antibodies, suggesting that the presence of large proteins attached to either termini altered the localisation of the target protein, thereby making them poor models for the behaviour of endogenous proteins. Hence, a different strategy was adopted in which the RBBP6 proteins were fused to an N-terminal HA-immuno-tag enabling for detection using anti-HA antibodies.

Fluorescence microscopy studies using endogenous RBBP6 had shown that whereas isoform 1 is entirely nuclear and confined to speckle-like bodies, isoform 3 is predominantly cytoplasmic, but nevertheless shows evidence of weak speckle-like bodies. Since isoform 3 consists of little more than the ubiquitin-like DWNN domain, the different localisation of the two isoforms is intriguing. Hence we decided to investigate whether the DWNN domain plays

26

a role in determining the localisation of the various isoforms of RBBP6. In particular, whether the DWNN plays a role in targeting the protein to nuclear speckles.

1.8.1   Objectives of the project

- The first objective was to make HA-tagged coding for wild type isoforms and test whether their localisation is consistent with that of endogenous RBBP6. Also to confirm whether the speckle-like bodies in which the protein localises are indeed nuclear speckles by carrying out a co-localisation study with SC35, a splicing factor which is commonly used as a marker for nuclear speckles.

- The next objective was to make a number of constructs encoding truncations or mutations of wild type RBBP6 and investigate their localisation in order to determine which domains are responsible for localisation to the nucleus or nuclear speckles.

- A final objective was to investigate the stress-dependent behaviour of the HA-RBBP6 deletion constructs used in this study in order to test conclusions derived from endogenous studies indicating that isoform 3 accumulates in nuclear speckles following a number of different cellular stresses.
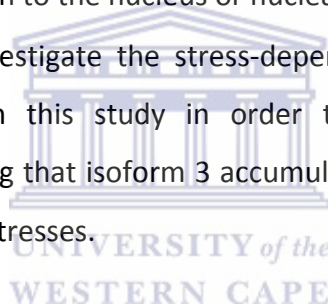
2   MATERIALS AND METHODS

2.1 *Antibodies Used*

2.1.1 *Primary antibodies*

Anti-HA:

Commercial mouse monoclonal raised against the HA-antigen (YPYDVPDYA), Cat. No: A01244-100 (Genscript Inc, Piscataway, NJ 08854, USA).

Commercial goat polyclonal raised against the HA-antigen (YPYDVPDYA), Cat. No: A00168-40 (Genscript Inc, Piscataway, NJ 08854, USA).

Anti-SC35:

Commercial mouse monoclonal raised against a phospho-epitope of the non-snRNP factor SC35. Cat. No: AB11826 (Abcam Plc, Cambridge, MA, USA).
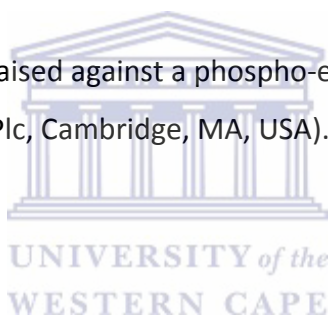
2.1.2 *Secondary antibodies*

Donkey anti-Goat

Alexa Flour 488-conjugated AffiniPure Donkey anti-goat antibody, Cat No: 705-545-003 (Jackson ImmunoResearch Laboratories, West Grove, PA, USA 19390).

Donkey anti-Goat

Cy3-conjugated AffiniPure Donkey anti-goat antibody, Cat No: 705-165-147 (Jackson ImmunoResearch Laboratories, West Grove, PA, USA 19390).

Donkey anti-Mouse

Cy3-conjugated AffiniPure Donkey anti-mouse antibody, Cat No: 715-165-150 (Jackson ImmunoResearch Laboratories, West Grove, PA, USA 19390).

2.2   Bacterial Culture

2.2.1   Preparation of competent E. coli cells for transformation

The *Escherichia coli (E. coli)* XL10 GOLD strain used in this study was streaked out on a nutrient agar plate and incubated overnight at 37 ℃. A single colony was picked and used to inoculate 50 ml of TYM broth and grown with vigorous shaking at 37 ℃ overnight. The overnight culture was then scaled up to 500 ml of TYM and grown under the same conditions until the optical density (OD) at 550 nm was in the range 0.4-0.6. The cells were rapidly chilled in ice water, with constant swirling, transferred to a 500 ml tube and harvested by centrifugation at 6000 g for 10 minutes using a pre-cooled Beckman JA-10 rotor. After discarding the supernatant the cells were re-suspended in 100 ml ice cold transformation buffer 1 (Tfb1) and allowed to incubate at 0 ℃ for 30 minutes. The cells were harvested by centrifugation at 6000 g for 8 minutes. The supernatant was discarded and the pellet was re-suspended in 30 ml ice cold transformation buffer 2 (Tfb2). The cells were frozen in liquid nitrogen as 50 μl aliquots and these were stored at -80 ℃.
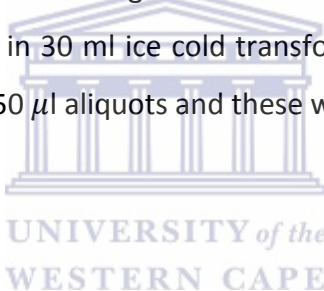
2.2.2   Bacterial transformations

DNA transformations were carried out with 50 μl competent cells (prepared as described in Section 2.1.1). The cells were thawed on ice and added to 50-100 ng of plasmid DNA solution, gently mixed and incubated on ice for 20 minutes. The cells were heat shocked at 42 ℃ for 90 seconds and placed on ice for 5 minutes. Following the incubation on ice, 500 μl of pre-warmed LB broth from Merk (Kenilworth, New Jersey, USA) was added and the mixture incubated with shaking at 37 ℃ for an hour to allow for the expression of the antibiotic resistance marker. The transformed cells were then plated on nutrient agar plates or scaled up to the required volume containing the ampicillin at a final concentration of 0.1 mg/ml and incubated overnight at 37 ℃, shaking for LB broth cultures.

29

2.3 Cloning

2.3.1 Vector specifications

The pCMV-UWC was created previously by replacing the MCS of pCMV with the MCS of pEGFP-C1, to facilitate subcloning of pEGFP-based constructs directly into pCMV, thereby replacing the GFP tag with an HA tag (see Section 6.1). The pCMV-UWC vector contains an ampicillin selection marker and an HA epitope tag coding sequence to facilitate detection of the target protein by immunofluorescence or western blotting. Strong, constitutive gene expression is driven by the CMV promoter from cytomegalovirus.
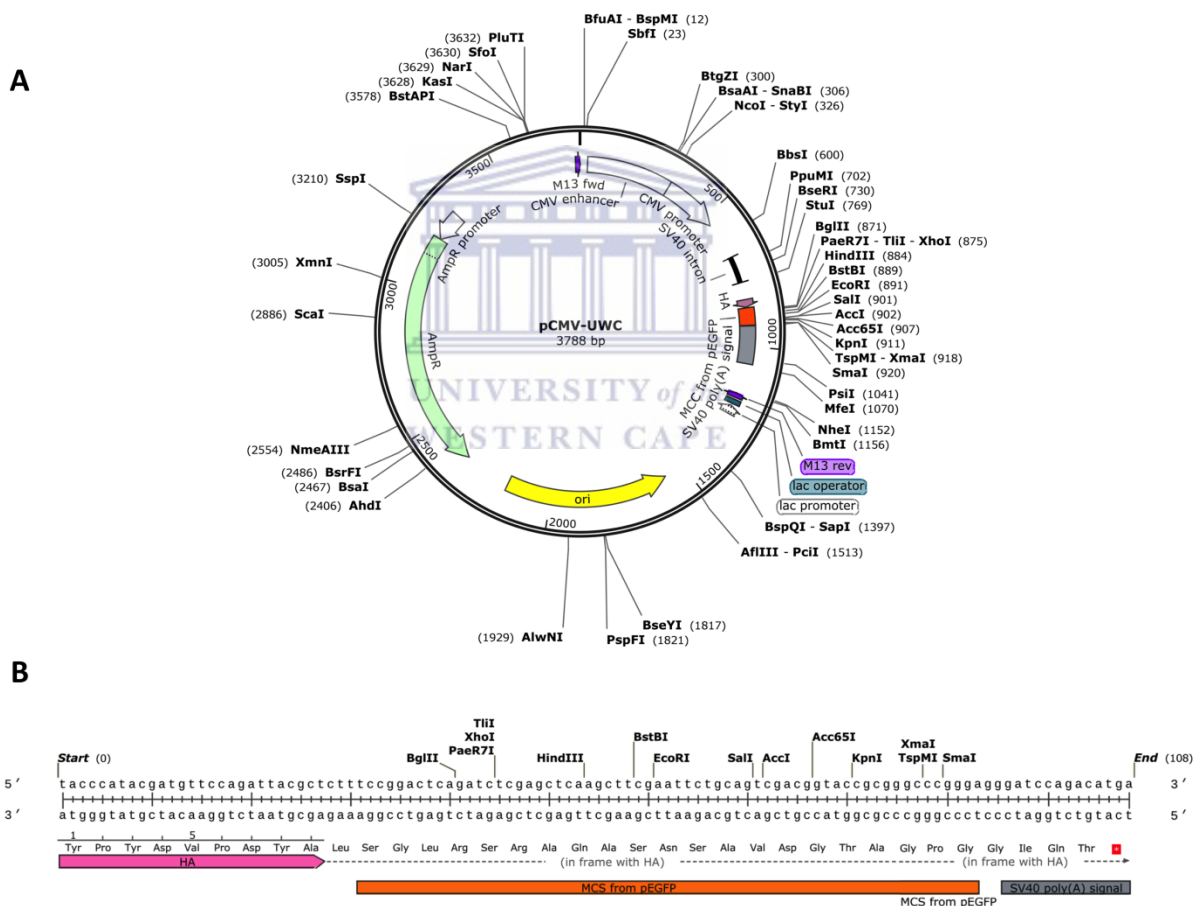
Figure 4: Plasmid map (A) and MCS (B) of the pCMV-UWC mammalian expression vector. pCMV-UWC was previously generated in our laboratory by inserting the multiple cloning sequence (MCS) of the pEGFP-C1 vector into the pCMV-C1 mammalian expression vector, allowing straight-forward sub-cloning of constructs from pEGFP-C1 into pCMV. pCMV-UWC incorporates an N-terminal HA-immuno-tag into proteins expressed from cDNA sequences cloned into the MCS. Figure generated using SnapGene Viewer (GSL Biotech LLC).

30

2.3.2 RBBP6 deletion constructs

The aim of the project was to investigate the hypothesis that the ubiquitin-like DWNN domain of Retinoblastoma Binding Protein 6 (RBBP6) is responsible for targeting it to speckle-bodies within the nucleus of mammalian cells. In order to check this hypothesis, we designed a series of truncations, mutations and deletions (Figure 5), each attempting to elucidate a specific role played by the domains constituting full length RBBP6 in its localisation phenotype. All these constructs were cloned into a pCMV-UWC vector. This cloning strategy incorporates an HA-tag on the N-terminus of the expressed proteins to facilitate the detection using immunofluorescence microscopy.
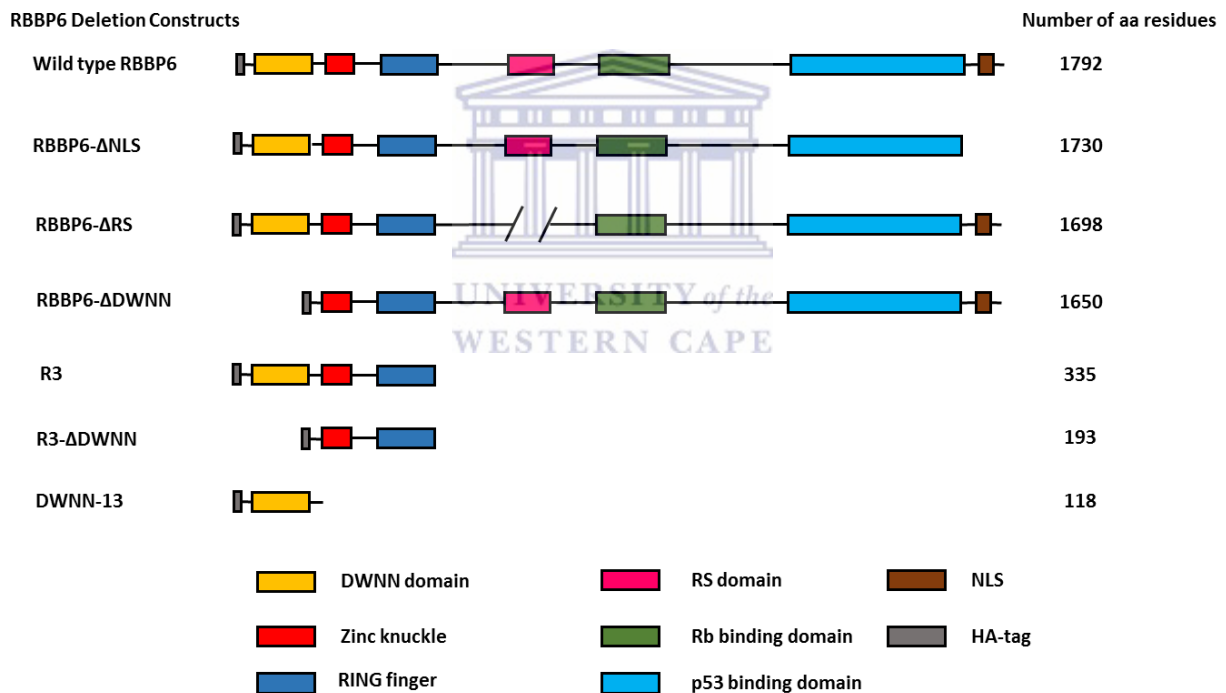
Figure 5: Domain organization of the RBBP6 deletion constructs used in this study. The different constructs used in this study were design through a series of gene manipulation techniques, allowing for specific domains to either be modified or removed entirely. The schematic diagram shows the domain organization of the different constructs. The HA tag YPYDVPDYA is included at the N-terminus of all constructs.

2.3.3 PCR amplifications

Generally, the PCR reactions were made up into a final volume of 50 µl with nuclease free water, unless stated otherwise. The primers were designed to incorporate specific restriction cut sites based on the DNA sequence of interest and the multiple cloning cassette of the vector to be utilized.

A standard PCR reaction was composed of 50 ng template DNA, 1X PCR reaction buffer, 1 µl of 50 ng primer set, 1 µl of 10 mM dNTPs, and 1U of the DNA polymerase.

The amplification reaction was carried out under the following parameters:

95℃ for 5 min

95℃ for 1 min

T_m- 5℃ for 1 min   25 cycles

72 ℃ for 1 min/1 Kb

72 ℃ for 1 min/1 Kb

Table 1. Sequence-specific oligonucleotides for the different cloning strategies used in this work. The different primers were designed to incorporate an XhoI (cyan) or KpnI (green) restriction site on the forward and reverse primers, respectively. Extra bases (yellow) and stop codons (red) were added were required to avoid any frame shifts and to dictate the length of the DNA sequence to be amplified. The specific bases mutated for the respective mutagenesis cloning are indicated in blue. The melting temperatures (Tm) for the overlapping (OL) and total primer (TP) are displayed below the sequences. No Tm were added for RBBP6-ΔNLS primers since the primers were not used for actual PCR cycles.

Construct name	Forward primer	Reverse primer
RBBP6–ΔDWNN deletion	5'–GAGGCGCTCGAGCTGACCCAATCAATTACATGAAGAA–3' $T_m = 62\ °C$	5'–CCTTTCATATTAGGATCTTTCAC–3' $T_m = 62\ °C$
DWNN–GG deletion	5'–GAGGCGCTCGAGCTTCCTGTGTGCATTATAAATTTT–3' $T_m = 62\ °C$	5'–GAGGCGGGTACCTCATTAACCTCCAATAGGAATTCTTCTAA–3' $T_m = 62\ °C$
DWNN–ΔNES deletion	5'–GAGGCGCTCGAGCTTCCTGTGTGCATTATAAATTTT–3' $T_m = 62\ °C$	5'–GAGGCGGGTACCTCATTATGCTTTTGTAGTTGCCATCGC–3' $T_m = 62\ °C$
DWNN–TY deletion	5'–GAGGCGCTCGAGCTTCCTGTGTGCATTATAAATTTT–3' $T_m = 62\ °C$	5'–GAGGCGGGTACCTCATTAATATGTCTTGCTTGTAGATTTAAC–3' $T_m = 62\ °C$
RBBPS–ΔNLS deletion	5'–TCAGCAGAAAGTCAGGACAGCTCTGTCACTGTGTGAGGTAC–3'	5'–CTCACACAGTGACAGAGCTGTCCTGACTTTCTGC–3'
D60A–L63A mutant	5'–GATGCTAATGCTGCGATTCCTAAGAATTCTTCTGTAATTGTTAGAAGAATTCCTATTGGAGG–3' $T_{m(TP)} = 79\ °C$ $T_{m(OL)} = 60\ °C$	5'–GGAATCGCAGCATTAGCATCAGTATATTCTTCTTTCGTCTGCGC–3' $T_{m(TP)} = 79\ °C$ $T_{m(OL)} = 60\ °C$
L29P–I33P mutant	5'–TCCCTCTGCGACCCTAAGAAGCAGCCTATGGGGAGAGAG–3' $T_{m(TP)} = 81\ °C$ $T_{m(OL)} = 66\ °C$	5'–CTTCTTAGGGTCGCAGAGGGAGATGTGGAGCCCATCAAA–3' $T_{m(TP)} = 81\ °C$ $T_{m(OL)} = 66\ °C$
V70P–V72P mutant	5'–TCTTCTCCTATTCCTAGAAGAATTCCTATTGGAGGTGTTAAATCTACAAGC–3' $T_{m(TP)} = 78\ °C$ $T_{m(OL)} = 58\ °C$	5'–TCTTCTAGGAATAGCAGAAGAATTCTTAGGAATCAGAGCATTATCATCAGTATATTCTTCTTTCG–3' $T_{m(TP)} = 78\ °C$ $T_{m(OL)} = 58\ °C$
G78A–G79A mutant	5'–ATTCCTATTGCTGCCGTTAAATCTACAAGCAAGACATATGTTATAAGTCGAACTGAACC–3' $T_{m(TP)} = 78\ °C$ $T_{m(OL)} = 58\ °C$	5'–TTTAACGGCAGCAATAGGAATTCTTCTAACAATTACAGAAGAATTCTTAGGAATCAGAGC–3' $T_{m(TP)} = 78\ °C$ $T_{m(OL)} = 58\ °C$

Mutagenesis PCR reactions were carried out using the Q5 High-Fidelity 2X Master Mix from New England Biolabs (Ipswich, Massachusetts, USA). The reaction was composed of 1X Q5 master mix, 50 - 100 ng of template DNA, and 1.25 ul of 125 ng primer set. The amplification reaction was carried out under the following parameters:

Table 2. Cycling parameters for the QuikChange Site-Directed Mutagenesis Method

Segment	Cycles	Temperature	Time
1	1	95℃	30 sec
2	12-18	95℃	30 sec
		55℃	1 minute
		68℃	1 minute/Kb of plasmid length

*Table adapted from the QuikChange Site-Directed Mutagenesis Kit Instruction Manual (Qiagen N.V., Hilden, Germany).

The *in vitro* synthesized DNA from the mutagenesis PCR was enriched for by treating the entire reaction volume (save for a 10 µl transformation control) from the thermo-cycler with DpnI endonuclease, which targets the methylated template DNA. The restriction digest was carried out with 1 µl DpnI added directly into the rest of the PCR reaction volume (≈40 µl). Following digestion, the DNA was transformed into and recovered from *E. coli* as described in Section 2.1.2. The experimental set-up is depicted in Figure 6 below.

1. Template DNA (wild-type)

Mutagenesis PCR
using mutagenesis primers,
control is the same but excluding primers

2. Wild-type + Amplicons (mutants)

DpnI Treatment
to destroy methylated template
leaving mutants

3. Mutant Progeny

Transformation into and recovery from *E. coli*
colonies grown on ampicillin plates

A

Wild type
+
DpnI

B

Wild-type + Mutants
+
DpnI
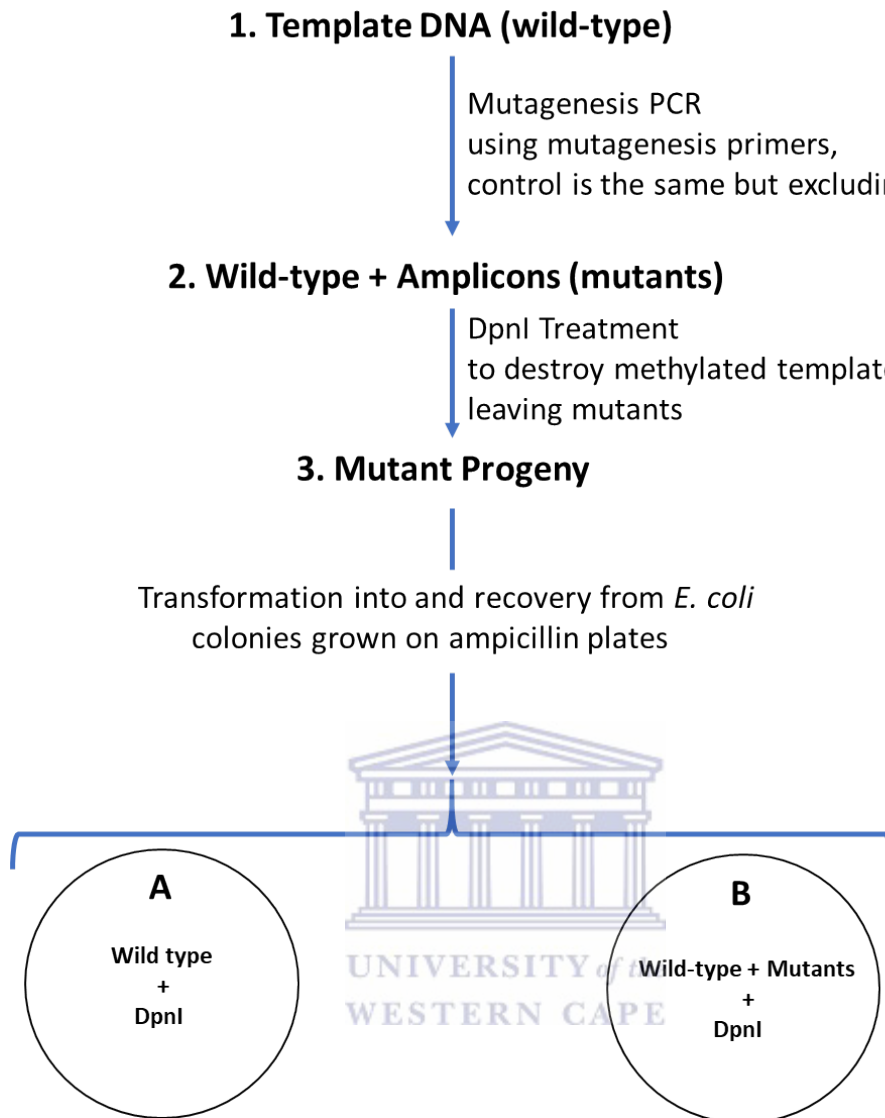
Figure 6: Schematic representation of the mutagenesis protocol. The mutagenesis PCR was carried out using the Q5-high fidelity master mix, a high template concentration and fewer cycles to reduce any chances of introducing unwanted mutations into the cloned sequence. The mutant progeny were selected for by treatment with the DpnI endonuclease which targets the methylated template DNA. (A) Wild type DNA was also digested with DpnI to indicate the proficiency of the enzyme. (B) Experimental plate where the parental DNA was digested with DpnI, allowing for the selection of the mutant progeny.

2.3.4 Restriction digest and DNA ligation

Generally, 1 ug of DNA was digested according to the manufacturers' specifications (restriction enzymes from Thermo Scientific, Waltham, MA USA) and the digested DNA fragments were subjected to gel electrophoresis for restriction digest confirmation and subsequent purification using the GeneJet gel extraction kit (Thermo Scientific, Waltham, MA USA). The purified DNA was quantified at absorbance 260 nm using a Nano Drop

35

spectrometer (Thermo Scientific, Waltham, MA USA) and ligated at room temperature for a period ranging from 1 to 4 hours.

The ligation reactions were carried out with 20 ng of vector DNA with the appropriate insert to vector ratio, calculated using the equation below:

$$ng \ of \ insert \ required = \frac{(ng \ of \ vector)(Kb \ size \ of \ insert)}{(Kb \ size \ of \ vector)} \ x \ insert \ vector \ ratio$$

Following ligation, *E. coli* XL10 Gold cells were transformed with 10 µl of the ligation reaction mixture, an appropriate volume spread plated out with the appropriate selection marker and grown overnight at 37 ℃ to obtain single colonies for screening. Transformations were carried out as described in Section 2.1.2.

2.3.5 Colony screening

Colonies were selected for screening from the experimental plates. No colonies were observed from the vector only control plate. Cultures were made by sub-culturing colonies in 25 ml LB-Ampicillin media and grown overnight at 37 ℃, shaking. This was followed by DNA extraction using the GeneJet plasmid mini-prep kit (Thermo Scientific, Waltham, MA USA) and quantified by Nano Drop spectrometry (Thermo Scientific). The purified DNA was digested with the respective restriction enzymes to produce the initial DNA fragments, which were confirmed by DNA gel electrophoresis, and thus providing the preliminary confirmation of successful cloning. The preliminary results for each of the different DNA constructs were confirmed through direct sequencing by Inqaba Biotechnical Industries (Hatfield, South Africa). Each of the cloned DNA constructs were given appropriate names based on the protein domains and/or the mutated amino acid residues each DNA construct encoded.

2.4 *Large scale DNA preparation for transfections*

Escherichia coli XL10 Gold cells were transformed with the respective pCMV-UWC DNA constructs. The transformation cultures were incubated overnight shaking at 37 ℃ in 25

ml LB-Ampicillin media. The overnight cultures were scaled up to sterile 250 ml LB-Ampicillin growth media flasks. The flasks were incubated overnight under the same conditions and the cells harvested by centrifugation at 6000 g for 20 min. Large-scale DNA extractions were carried out to obtain high-quality, endotoxin-free plasmid DNA for transfection using the GeneJet Maxi-prep kit from Thermo Scientific (Waltham, MA USA). The DNA extraction was carried out according to the manufacture's protocol for each of the constructs used in this study.

2.5 *Cell culture and transfections*

A549 cells were grown in Dulbecco's Modified Eagle's Medium supplemented with 10 % (v/v) fetal bovine serum and 10 mg/ml of penicillin and streptomycin each. Once confluent, the adherent cells were trypsinized and grown on coverslips in 12-well plates at a concentration of 1 x10^{5}, for 18-24 hours prior to transfecting. Transfections were carried out with 0.5 - 1.1 μg of plasmid DNA per well of 60 − 75 % confluent cells in the presence of Xtreme-GENE HP DNA transfection reagent (Roche Life Science, Randburg 2125, South Africa) in a 3:1 transfection reagent to DNA ratio. The proteins were transiently expressed for 24 to 48 hours before processing for detection using immunofluorescence microscopy.

2.6 *Immunofluorescence microscopy*

The cells were heat-shocked at 42 ℃ for 10 minutes and permeabilized by adding ice cold methanol. If no heat shocking was required, the cells were directly treated with methanol and fixed with 4 % formaldehyde made in 1XPBS, the incubation periods were 5 minutes long at 4 ℃ and room temperature, respectively. The cells were incubated with 1 % BSA-PBST blocking solution for 30 minutes at room temperature. The cells were washed three times with PBST for 5 minutes and incubated with either an anti-HA primary antibody on its own for direct protein localisation studies or coupled with an anti-SC35 antibody for co-localisation studies.

The HA and SC35 antibodies were diluted in 1 % BSA-PBST solution to working concentrations of 1:100 and 1:1000, respectively. The cells were incubated with the

37

antibodies overnight at 4 ℃. The cells were washed with PBST and subsequently incubated with either a Cy3-conjugated secondary antibody on its own or coupled with an Alexa 488 secondary antibody depending on whether it's a direct or co-localisation assay. The secondary antibody dilutions were made up to a 1:1000 dilution in 1% BSA-PBST solution and incubations were carried out for an hour at room temperature shielded from direct light.

The cells were incubated with the nuclei staining solution Hoechst (blue), at a dilution of 1:1000 (v/v) to differentiate it from the protein localisation pattern detected by an anti-HA primary antibody and Cy3-conjugated (red) secondary antibody. The cells were washed as described earlier and the coverslips mounted with Mowiol onto microscope glass slides to be viewed once they have completely dried. All the transfection results were analysed through fluorescence microscopy using a ZEISS Axioplan 2 imaging system for the standard or the ZEISS Axiovert 200M for the inverted fluorescent microscope, both coupled with an AxioCam HRm camera and the data was recorded for analysis of the different protein localisation phenotypes.
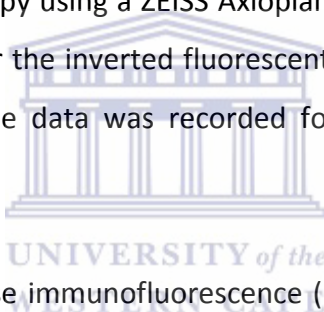
Controls were included in these immunofluorescence (IF) experiments to ensure antibody specificity and cell integrity. These controls included non-transfected cells treated with primary and secondary antibodies to confirm antibody specificity. Cells treated with just a nuclear stain were also included as a means to check for any contamination, particularly mycoplasma contamination and to check for any form of cellular degradation which may in turn affect the experiments and consequently any results obtained.

3   RESULTS AND DISCUSSION

3.1   *Isoform 1 of RBBP6 localises entirely within the nucleus in speckle-like bodies*

pCMV-UWC-RBBP6, a mammalian expression vector generated by inserting RBBP6-isoform 1 between the XhoI and KpnI sites of pCMV-UWC, was already on hand in the laboratory. pCMV-UWC, the backbone vector used to make all of the expression constructs used in this thesis, was generated previously by inserting the MCS of pEGFP-C1 into pCMV-C1, as decribed in more depth in Chapter 2. pCMV-UWC was used for historical reasons because the first expression construct for RBBP6 made in our laboratory was made using pEGFP-C1, and, considering the size of isoform 1, when the decision was made to move the construct into pCMV it was considered easier to adapt the MCS of pCMV-C1 than to re-amplify isoform 1. Hence pCMV-C1 was adapted so that isoform 1 could be sub-cloned directly from pEGFP-C1 into the same sites in pCMV-UWC. The construct expresses a protein of 1792 amino acid residues, preceded by an HA tag, under control of the *CMV* constitutive promoter.
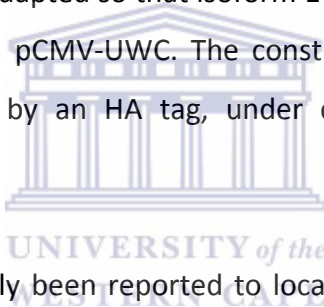
Endogenous RBBP6 has previously been reported to localise in nuclear speckles, punctate nuclear bodies most closely associated with mRNA splicing. To investigate whether HA-RBBP6 (HA-isoform 1) also localises to nuclear speckles, pCMV-UWC-RBBP6 was used to transfect A549 lung cancer cells and to investigate the localisation of HA-RBBP6 using immunofluorescence microscopy. Detection of HA-RBBP6 was carried out by staining the cells with mouse anti-HA primary antibodies (Cat No: A01244-100, Genscript Inc, Piscataway, NJ 08854, USA) followed by donkey anti-mouse secondary antibodies conjugated to Cy3 (Cat No: 715-165-150, Jackson ImmunoResearch Laboratories, West Grove, PA, USA 19390). A detailed description of the primary and secondary antibodies used in this work can be found in Section 2.1.1 and 2.1.2, respectively.

As can be seen from Figure 7 panels A-C, HA-isoform 1 is entirely nuclear, with clear evidence of punctate bodies consistent with nuclear speckles (see Figure 3). Isoform 1 is not entirely confined to the speckle-like bodies, but there appears to be a smooth distribution of low intensity signal across the nucleus.

In order to check that the signal observed in Figure 7 comes from RBBP6 and not from other proteins, the same investigation was carried out on untransfected cells (Figure 7, panel D and E). No signal was observed when both primary and secondary antibodies were added to unstransfected cells (panel D), which shows that the signal is due to the presence of HA-isoform 1. The very weak signal in panel E must therefore corresponds to auto-fluorescence (red light emitted by the cells themselves) and confirms that the amount of auto-fluorescence can be ignored in comparison with the amount of signal emanating from HA-RBBP6.
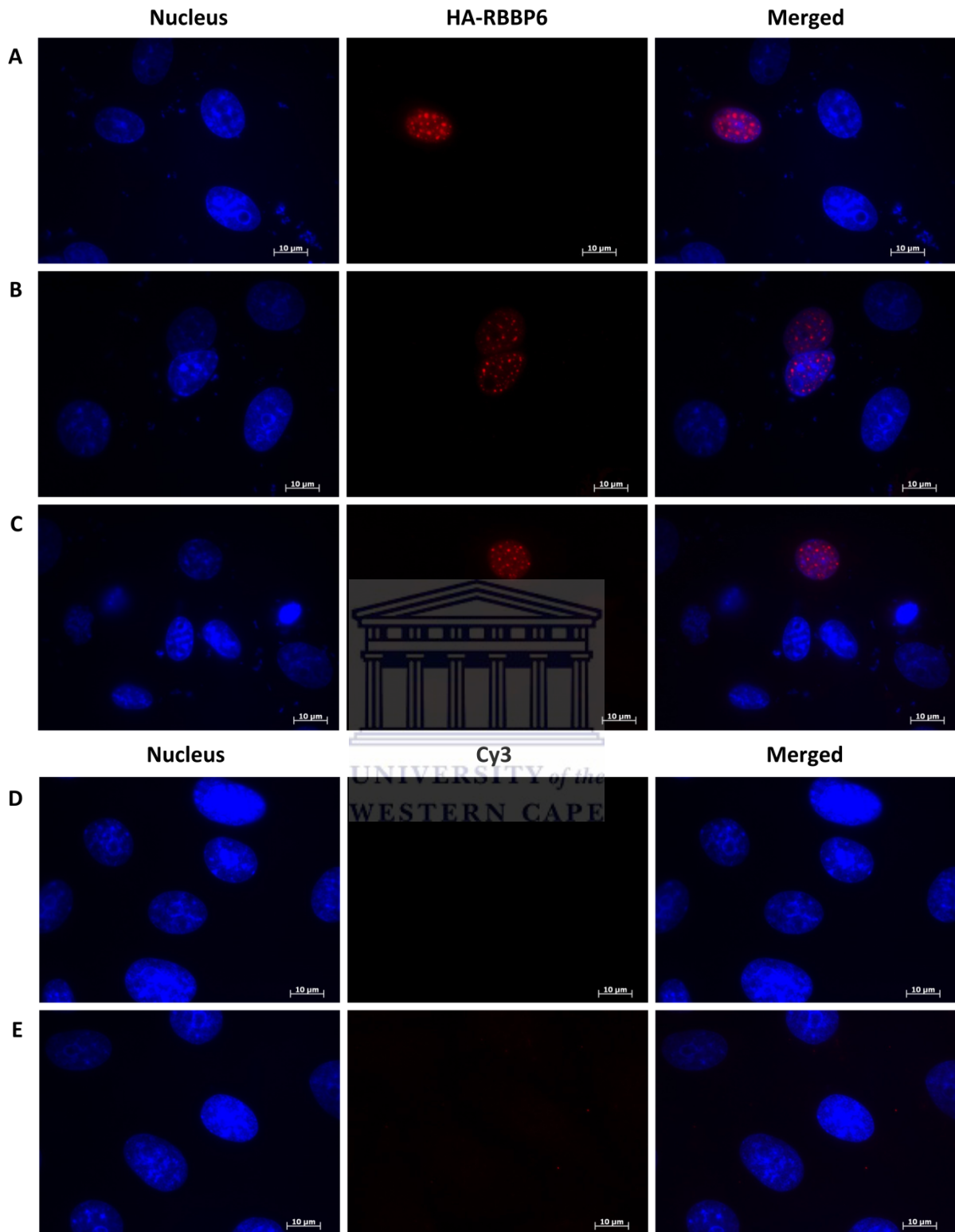
Figure 7: HA-RBBP6 localisation, antibody specificity and background controls. Cells transfected with pCMV-UWC-RBBP6 display a punctate nuclear localisation pattern (A-C). No signal was observed from untransfected cells in the presence of both primary and secondary antibodies (D) and with neither primary nor secondary antibodies (E). The nucleus was stained with Hoechst to differentiate it from the protein signal.

3.2 *HA-RBBP6 isoform 1 co-localises with SC35 in nuclear speckles*

To investigate if the punctate nuclear bodies in which HA-RBBP6 preferentially localises are indeed splicing speckles, co-localisation assays were carried out between RBBP6 and SC35, a splicing factor known to localise to these speckles. SC35 was detected using mouse anti-SC35 primary antibodies (Cat. No: AB11826, Abcam Plc, Cambridge, MA, USA) and Cy3-conjugated anti-mouse secondary antibodies (Cat No: 715-165-150, Jackson ImmunoResearch Laboratories, West Grove, PA, USA 19390). HA-isoform 1 was detected using goat anti-HA primary antibodies and Alexa488-conjugated anti-goat secondary antibodies (Figure 8).

As can be seen from Figure 8, the pattern of speckles produced by HA-RBBP6 (green) is highly similar to that produced by SC35 (red). This is confirmed in the panel on the right in which the red and green panels are superimposed; the red and green speckles combine to produce yellow spots, implying that the speckles superimpose exactly on one another.

Like the HA-tag on RBBP6, the anti-SC35 antibody has been shown to be specific for SC35 in numerous published studies, so we can be confident that the signal seen in the red and green channels emanate from HA-RBBP6 and SC35 respectively and not from any other proteins in the cell. Nevertheless, in co-localisation studies addition care must be taken that signal from one protein is not observed in the channel corresponding to the other protein. This can happen in two ways: (i) light from the green fluorophore may be allowed into the red channel and *vice versa*, and (ii) in addition to binding to its corresponding primary antibody, the green-conjugated secondary may also bind to the primary antibody corresponding to the red-conjugated secondary, and *vice versa*. The first problem is resolved by choosing the fluorophores to have non-overlapping emission spectra and by setting the filters on each channel so that red light is not allowed into the green channel, and *vice versa*. The second problem is resolved by choosing secondary antibodies that only recognize one of the two primaries.

Controls showing that the above conditions were fulfilled should have been performed. These would be carried out by repeating the above co-localisation experiment, but leaving out the

42

primary corresponding to the green secondary and confirming that no signal is observed in the green channel. That would confirm that signal is only present in the green channel when the corresponding primary antibody is present. The same procedure would be carried out for the red channel.

Unfortunately these controls were not carried out. Nevertheless, in this particular case, the fluorophores Cy3 and Alexa488 have non-overlapping spectra and the red and green channels are equipped with the standard filters. Furthermore mouse and goat are sufficiently distant in evolutionary terms that anti-goat secondary is unlikely to recognize mouse anti-SC35 primary and anti-mouse secondary is unlikely to recognize goat anti-HA primary.
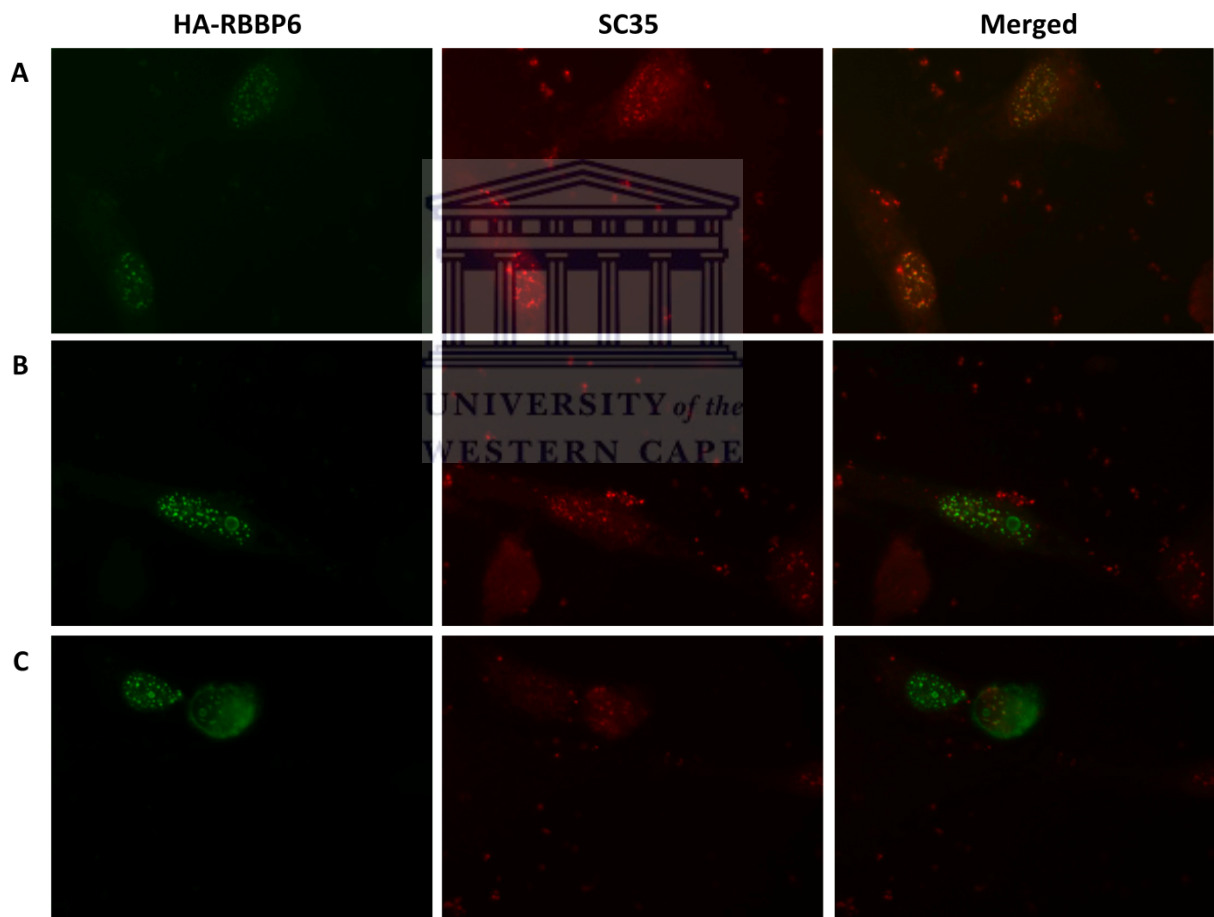
Figure 8: HA-RBBP6 isoform 1 co-localises with SC35 in nuclear speckles. A549 cells over expressing HA-RBBP6 were probed with both anti-HA and anti-SC35 specific antibodies to compare the localisation of both these proteins. In the green channel (first column), over expressed HA-RBBP6 probed with anti-HA primary and Alexa-488-conjugated secondary antibodies display a punctate nuclear localisation phenotype. In the red channel (second column), endogenous SC35 probed with anti-SC35 primary and Cy3-conjugated secondary antibodies also display a nuclear speckling pattern. The third column indicates regions of co-localisation between HA-RBBP6 and SC35, where regions of co-localisation are represented by the yellow colour obtained from an overlay of the two speckling proteins. Panels A-C show the same thing, however panel A shows a much clearer signal which could be a result of cell-dependent expression of the proteins.

43

3.3 *The ubiquitin-like DWNN domain is the nuclear localisation motif for isoform 1*

To investigate the role played by the DWNN domain in the intra-cellular localisation of RBBP6, a truncated form of the protein, spanning residues 143-1792, and therefore lacking the DWNN domain, was constructed from pCMV-UWC-RBBP6.

Large genes such as RBBP6 tend to contain few unique restriction sites, complicating the process of restriction enzyme-based cloning. Fortunately RBBP6 contains a unique AvrII site at nucleotide 555 (codon 186), between the DWNN domain and the Zinc finger domain. The XhoI insertion site immediately precedes the beginning of the RBBP6 insert. The cloning strategy involved excising the XhoI/AvrII fragment and replacing it with a short "bridging" fragment generated by amplifying nucleotides 426-626 (codons 143-209) using a forward primer incorporating an XhoI site at the 5'-end and a reverse primer designed to anneal on the C-terminal side of the AvrII site. No restriction enzyme site was required in the reverse primer due to the presence of the AvrII site in the amplicon at codon 186.

The nucleotide sequences of the required primers can be found in Table 1, first row. The XhoI site is shown in green. Complementary regions are underlined. The two bases indicated in yellow were added to bring the end of the XhoI site back into frame before the start of codon 143 (5'-GAC). The bridging fragment was successfully amplified, yielding a fragment consistent with the expected size of 214 bp (Figure 9A, lane 4), which was then digested with XhoI and AvrII and cloned into pCMV-UWC-RBBP6 following excision of the XhoI/AvrII fragment.

The cloning was judged to be successful due to larger numbers of colonies on plates transformed with vector + ligase + insert as compared to those transformed with vector + ligase only. DNA was extracted from putative clones and digested with XhoI and AvrII, which released the expected fragment of 214 bp rather than the fragment of 568 bp if the DWNN domain had not been excised (Figure 9B, lanes 2-6). Despite the very faint 214 bp in panel B, sequencing of the cloned fragment showed it to be identical to the expected sequence. The construct was named pCMV-UWC-RBBP6-ΔDWNN.
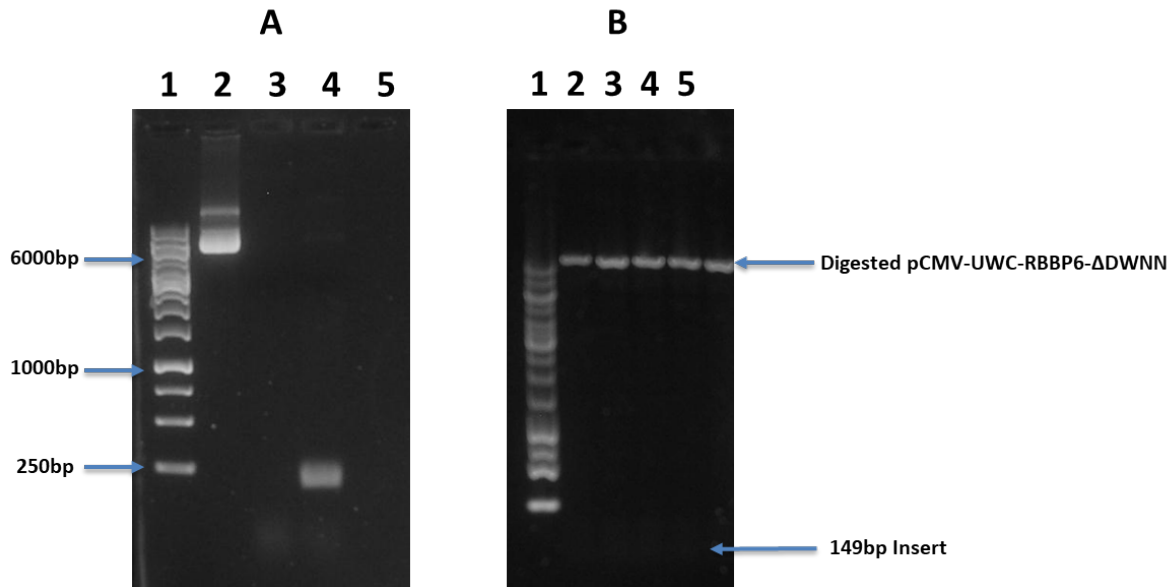
Figure 9: Cloning of the HA-RBBP6-ΔDWNN construct. (A) PCR amplification of the HA-RBBP6-ΔDWNN bridging fragment**.** The 214 bp fragment in lane 4 was amplified from full length RBBP6, incorporating an XhoI restriction site followed immediately by codon 143 of RBBP6. Lanes 2 and 3 contain the template DNA and the negative (no-template) control, respectively. (B) Restriction digest screen of putative RBBP6-ΔDWNN constructs. Digestion with XhoI and AvrII released fragments corresponding to the amplicon in panel A. The faint bands on the bottom of lanes 2-6 correspond to the 149 bp insert and the high molecular weight bands in the same lanes indicate the vector from the digested pCMV-UWC-RBBP6-ΔDWNN construct.

High quality, endotoxin-free pCMV-UWC-RBBP6-ΔDWNN plasmid DNA was prepared and transfected into cultured A549 mammalian cells as described in Sections 2.4 and 2.5. The cellular localisation of full-length HA-RBBP6 and HA-RBBP6-ΔDWNN were compared through the use of immunofluorescence microscopy. As reported above, HA-RBBP6 was found to be entirely nuclear and localise in nuclear speckles (Figure 7, panels A-C). In contrast, HA-RBBP6-ΔDWNN is almost entirely cytoplasmic (Figure 10), with almost no signal in the nucleus. Since these images were generated using a wide-field rather than a confocal microscope, any signal apparently coming from the nucleus may in fact originate from the cytoplasm above and below the nucleus along the line of sight, in which case the nuclei may in fact be darker than they already appear. These results strongly suggest that the DWNN domain of RBBP6 contains the speckle-directing motif or at least a nuclear localisation signal.

45
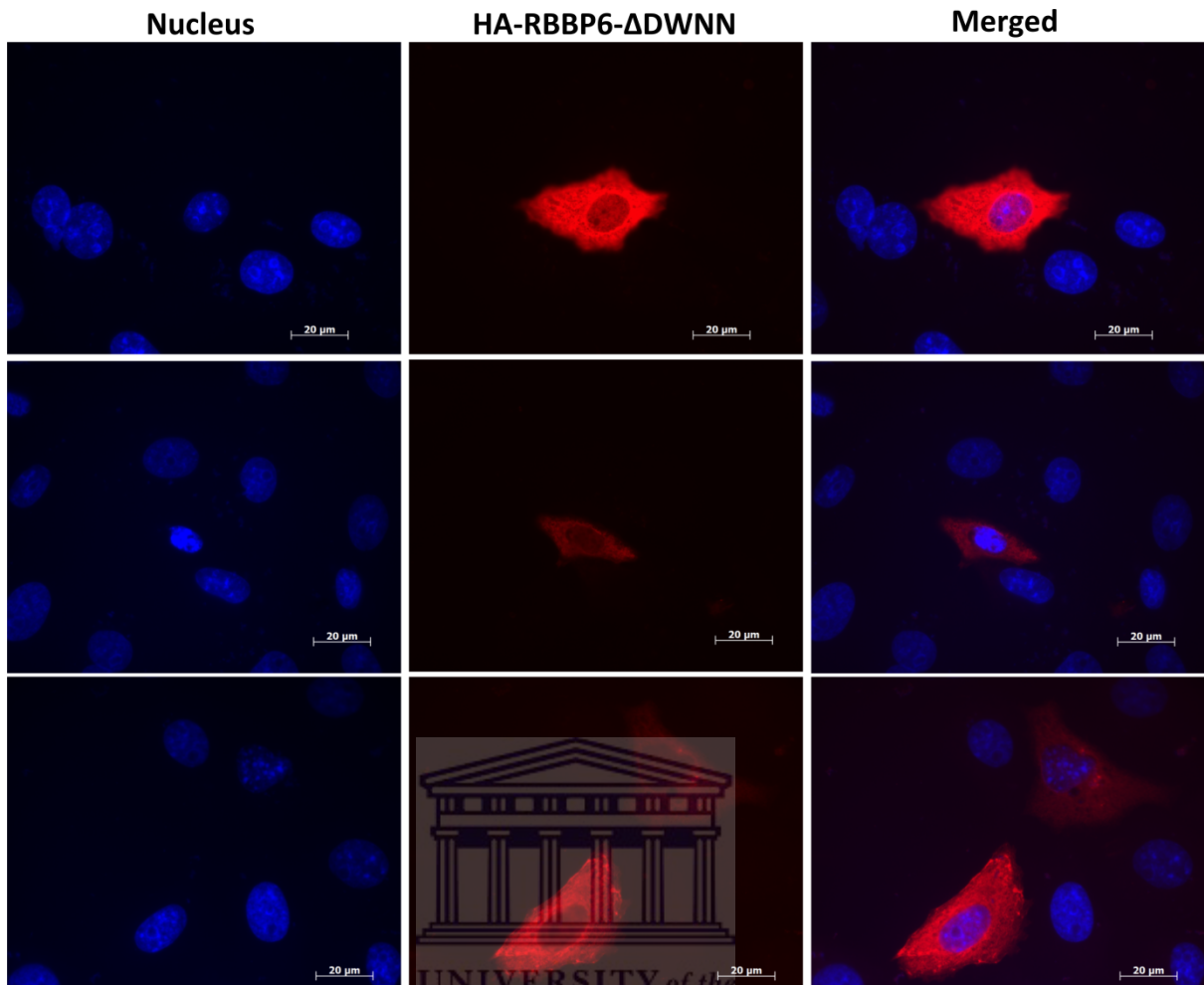
Figure 10: RBBP6-ΔDWNN localises to the cytoplasm of transfected cells. Cells transfected with pCMV-UWC-ΔDWNN display an entirely cytoplasmic localisation phenotype. The nucleus was counter-stained with Hoechst (blue) to differentiate it from the protein localisation pattern detected by an anti-HA primary antibody and Cy3-conjugated secondary antibody (red).

3.4   *The DWNN domain is not sufficient for targeting RBBP6 to nuclear speckles*

RBBP6 orthologues are generally smaller in lower eukaryotes than in higher eukaryotes (see Figure 2, *Pugh et al., 2006*), nevertheless all eukaryotes contain at least a transcript containing the DWNN domain, zinc finger and RING finger domains. It was therefore of interest to determine whether the corresponding truncation of human RBBP6, which we have termed R3, would present a similar localisation phenotype to the one observed for full length RBBP6, since both contain the DWNN domain. The corresponding fragment is the predominant form expressed in yeast, which has been shown to be essential for 3'-end processing of mRNA constructs (*Lee and Moore, 2014*). A similar human fragment, denoted RBBP6-N, has been shown to be functional in 3'-end splicing in humans (*Di Giammartino et*

46

al., 2014). Hence the intra-cellular localisation of this fragment of RBBP6 may have implications for the role of RBBP6 in mRNA processing, both in humans and yeast.

A mammalian expression plasmid coding for HA-tagged R3, denoted pCMV-UWC-R3, had previously been constructed by a co-worker in the laboratory, Ms Andronica Ramaila. It contained residues 1-335 of RBBP6 isoform 1 cloned between the BglII and SalI sites of pCMV-UWC. The localisation phenotype of the pCMV-UWC-R3 construct was investigated using immunofluorescence microscopy; it showed a very different distribution to HA-RBBP6 (isoform 1). Instead of being entirely nuclear and confined to nuclear speckles, HA-R3 was uniformly distributed across both the cytoplasm and the nucleus (Figure 12, panels A and B). The nuclear localisation, rather than punctate, was smoothly distributed across the nucleus.

The localisation phenotype of HA-R3 suggested that the DWNN domain of RBBP6 may be required to translocate the protein to the nucleus, but plays no further role in targeting the protein to nuclear speckles. These results hint at another possible domain within residues 336-1792 that is solely responsible for the speckling pattern observed for full length RBBP6, most likely the RS-domain.
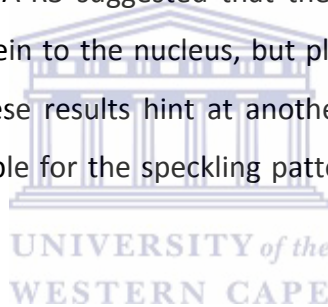
To test whether the DWNN domain is also required for nuclear localisation of R3, the DWNN domain was removed from R3. pCMV-UWC-R3-ΔDWNN was created through restriction enzyme-base manipulation of the previously described pCMV-UWC-RBBP6-ΔDWNN and pCMV-UWC-R3 constructs, as described below.

pCMV-UWC-RBBP6-ΔDWNN encodes residues 143-1792 of RBBP6, inserted between the XhoI and KpnI sites of pCMV-UWC. pCMV-UWC-R3 encodes residues 1-335 of RBBP6, inserted between the BglII and SalI sites of pCMV-UWC. The unique AvrII site referenced above is located in codon 186 of RBBP6. The strategy was to excise the AvrII/KpnI fragment from pCMV-UWC-R3, corresponding to residues 187-335 of RBBP6, and insert it into the same sites in pCMV-UWC-RBBP6-ΔDWNN.

Restriction digest of pCMV-UWC-RBBP6-ΔDWNN with AvrII and KpnI generated bands of 4828 bp, corresponding to residues 187-1792 of RBBP6, and 3881 bp, corresponding to residues

143-186 of RBBP6 cloned into pCMV-UWC. The bands can be seen in lane 4 of Figure 11A; marked with a red box, the lower band, corresponding to the 3881 fragment was excised from the gel and purified using the GeneJet gel extraction kit from Thermo Scientific (Waltham, MA USA). Digestion of pCMV-UWC-R3 with AvrII and KpnI generated bands of 466 bp, corresponding to residues 187-335, and 4309 bp, corresponding to residues 1-186 of RBBP6 cloned into pCMV-UWC. Both bands can be seen in lane 8 of Figure 11A; the 466 bp band marked with a red box was excised and purified.

The 466 bp band from lane 8 of Figure 11A and the 3881 bp band from lane 4 of Figure 11A were ligated together and transformed into competent *E. coli* cells. Plasmid DNA was extracted from putative positive clones and subjected to restriction digestion with AvrII and KpnI, which released the expected fragment of 466 bp, (Figure 11B, lanes 3-8). DNA sequencing was carried out which confirmed that the constructs were correct. The new DNA construct was named pCMV-UWC-R3-ΔDWNN.

Large-scale endotoxin-free extraction of pCMV-UWC-R3-ΔDWNN was carried out and the plasmid transfected into A549 cells for immunofluorescence microscopy studies. In contrast to HA-R3, which is found strongly and uniformly concentrated in the nucleus in addition to uniform distribution in the cytoplasm (Figure 12, panels A and B), HA-R3-ΔDWNN is predominantly cytoplasmic with little evidence of any protein in the nucleus (Figure 12, panels C and D). Hence the localisation of HA-R3-ΔDWNN is similar to that of HA-RBBP6-ΔDWNN.
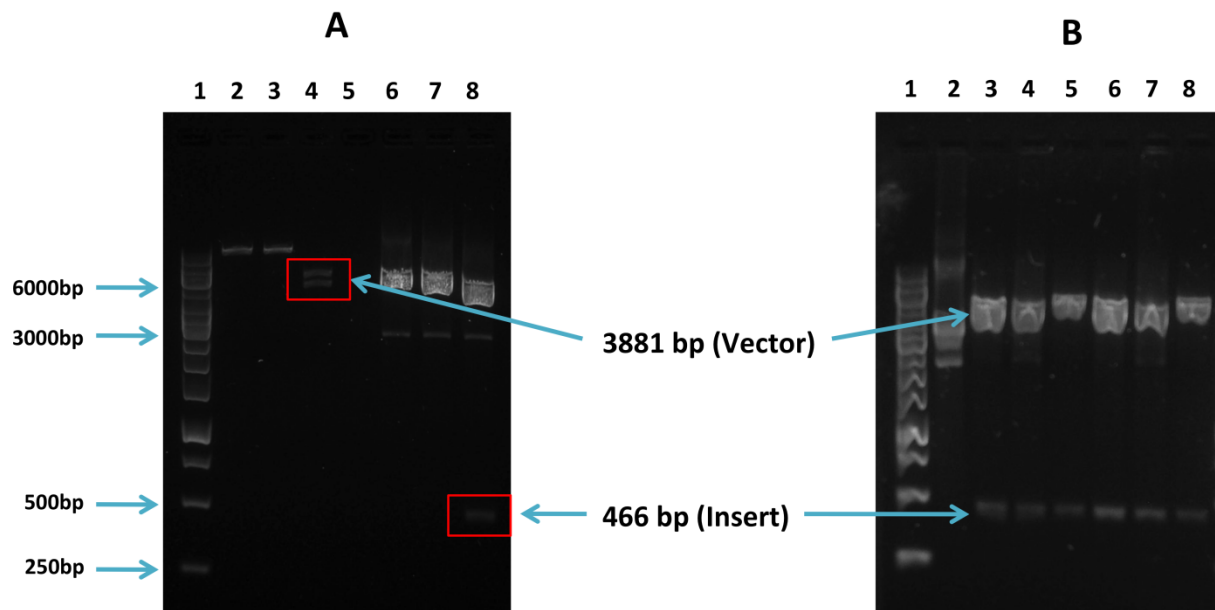
Figure 11: Construction of the pCMV-UWC-R3-ΔDWNN expression construct. (A) Restriction digest of pCMV-UWC-RBBP6-ΔDWNN (lanes 2-4) and pCMV-UWC-R3 (lanes 6-8) constructs were carried out with AvrII and KpnI to release the vector and the insert DNA, respectively. Lanes 2 and 6 indicate the single digests with AvrII, and lanes 3 and 7 indicate the single digests with KpnI. The double digests of the HA-RBBP6-ΔDWNN and R3 constructs are indicated in lanes 4 and 8, respectively. Only the smaller DNA fragments, marked with a red box from both lanes 4 (3881bp vector DNA) and 8 (466bp insert DNA) were gel purified and used for ligations. (B) Positive pCMV-UWC-R3-ΔDWNN clones were screened by restriction digest using AvrII and KpnI to release the ligated DNA fragments. Lane 2 represents the undigested pCMV-UWC-R3-ΔDWNN construct and the bands on the bottom of lanes 3 to 8 correspond to the 466bp insert released from the vector (the 3881bp bands in the same lanes).

49
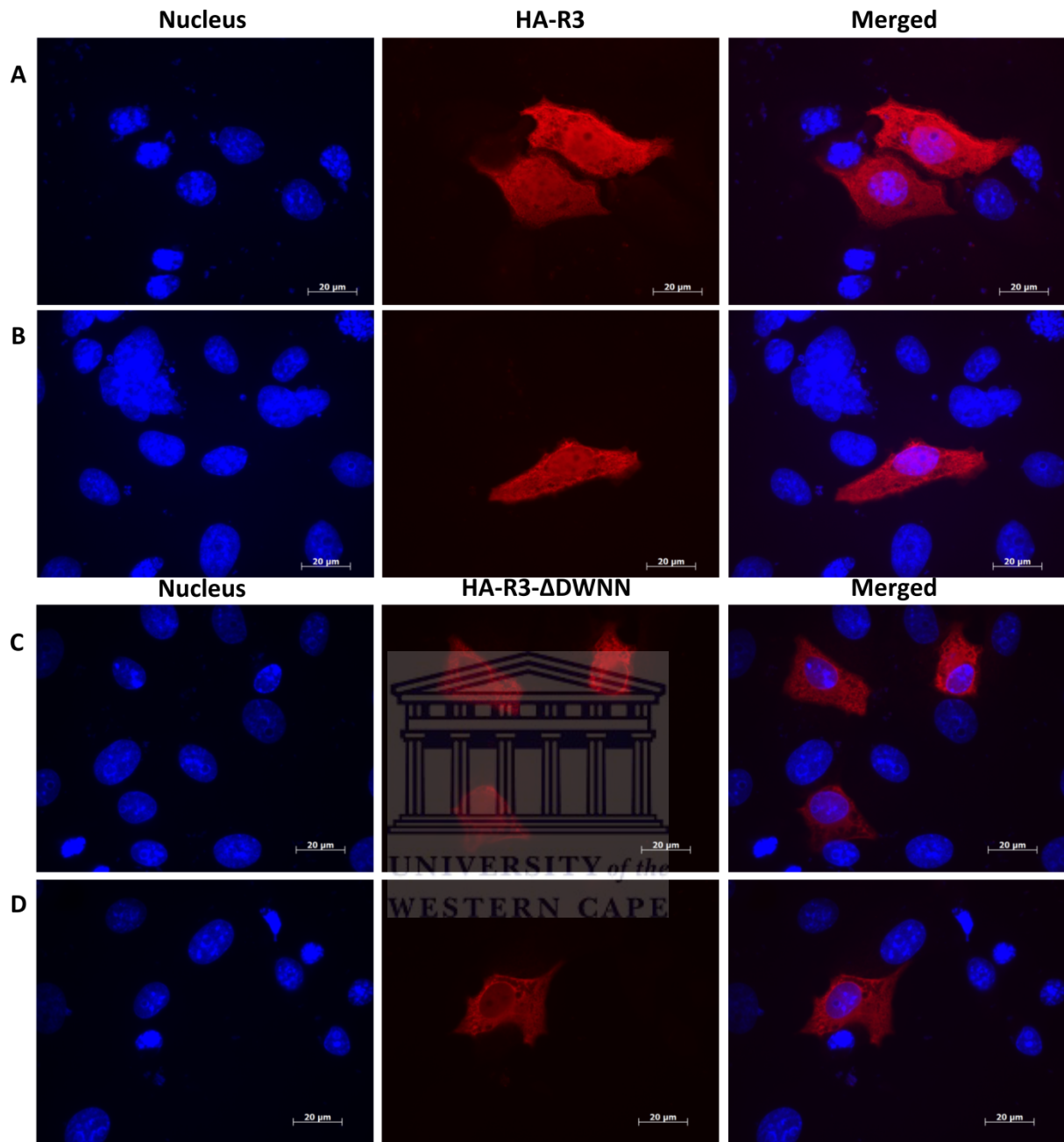
Figure 12: The DWNN domain functions as a nuclear targeting signal for RBBP6 proteins. HA-R3 (panels A and B) is evenly distributed across the cytoplasm and nucleus of transfected cells, in contrast to the localisation of HA-R3-ΔDWNN (panels C and D), which shows a predominantly cytoplasmic localisation phenotype. The cells were transfected with either the pCMV-UWC-R3 or pCMV-UWC-R3-ΔDWNN DNA construct and processed as described in Materials and Methods.

3.5   *The RS-domain is required for nuclear localisation of RBBP6 but not for targeting to*
       nuclear speckles

The absence of both HA-RBBP6-ΔDWNN and HA-R3-ΔDWNN from the nucleus suggests that the DWNN domain is required for targeting RBBP6 to the nucleus. However the presence of HA-R3 in the cytoplasm shows that the DWNN domain is not sufficient for nuclear localisation. And the lack of discrete nuclear speckles in HA-R3 shows that the DWNN domain is not sufficient for targeting to nuclear speckles.

Isoform 1 of RBBP6 contains a serine-arginine-rich region which is commonly referred to as an RS domain. Similar regions are found in many splicing factors and splicing-associated proteins and have been shown to be responsible for targeting some of them to nuclear speckles (*Li and Bingham, 1991; Cáceres et al., 1997; Cazalla et al., 2002*). Since the RS domain is present in HA-RBBP6, which does speckle, but absent in HA-R3, which doesn't speckle, a reasonable hypothesis is that the RS domain may be responsible for targeting the protein to nuclear speckles. In order to test this hypothesis, mutagenesis was used to delete the RS domain, yielding pCMV-UWC-RBBP6-ΔRS, as described below.

The RS domain spans residues 769-788 of RBBP6 and is flanked by unique Sal I and Hind III sites as shown in Figure 13. A cDNA sequence corresponding to the amino acid residues shown in yellow (omitting the RS domain shown in grey) and including the flanking SalI and HindIII sites, was synthesized by Genscript Inc. (Piscataway, NJ 08854, USA) and supplied in a pUC57 vector. The deletion was designed so that it preserved the frame of RBBP6 on the C-terminal side of the deletion.

The pUC57 construct from GenScript was digested with SalI and HindIII, yielding fragments of 1835 bp and 875 bp (Figure 14A, lane 4), the 875 bp band (lane 4, red box) corresponding to the synthesized fragment of RBBP6, was excised and purified. pCMV-UWC-HA-RBBP6 construct was digested with SalI and HindIII, yielding fragments of 7936 bp and 1199 bp (Figure 14A, lane 8). The 7936 bp band (lane 8, red box) corresponding to the bulk of HA-RBBP6, was excised and purified. The 875 bp and 7936 bp fragments were ligated to create the pCMV-UWC-RBBP6-ΔRS construct; as shown in panel B, release of an 875 bp fragment (rather than an 1199 bp fragment) on digestion with SalI/HindII confirmed that the RS

51

domain had been successfully removed. Sequencing carried out by Inqaba Biotech using a custom-synthesised primer (RBBP6-Int-3001: 5'-CCTCCTCAATTGCAATTACC-3') gave high quality coverage across the complete insert and showed it to be correct.

Large quantities of endotoxin-free pCMV-UWC-RBBP6-ΔRS were prepared and transfected into A549 cells. Removal of the RS domain had a significant effect on the localisation of HA-RBBP6, as can be seen in Figure 15. HA-RBBP6-ΔRS is significantly more cytoplasmic than HA-RBBP6, suggesting that the RS domain plays a role in nuclear localisation. Hence both the DWNN domain and the RS domain are required for complete nuclear localisation. Whether they are sufficient to produce total nuclear localisation will be addressed in the next section.

Due to its reported role in targeting other proteins to nuclear speckles, it was expected that HA-RBBP6-ΔRS would show no speckling and may therefore resemble HA-R3 most closely. Surprisingly it is clear from Figure 15 that HA-RBBP6-ΔRS does show clear evidence of nuclear speckling, although whether they are exactly the same as those of HA-RBBP6 is difficult to judge from our data. Nevertheless we can conclude that the RS domain is not required for targeting of RBBP6 to nuclear speckles.
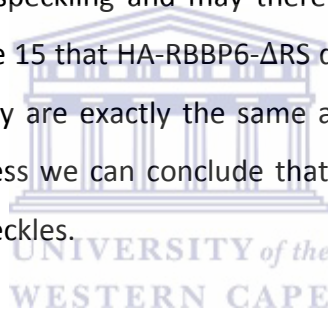
```
RBBP6 SalI–HindIII fragment
gtcgaccaggctgggaacattccaacaaacttggctatctggtttctccaccacaacaaat
     R  P  G  W  E  H  S  N  K  L  G  Y  L  V  S  P  P  Q  Q  I
agaagaggggagaggagctgctacagaagtataaaccgtgggcgacaccacagcgaaaga
  R  R  G  E  R  S  C  Y  R  S  I  N  R  G  R  H  H  S  E  R
tcacagaggactcaaggcccgtcactaccagcaactccagtctttgtacctgttccacca
  S  Q  R  T  Q  G  P  S  L  P  A  T  P  V  F  V  P  V  P  P
cctcctttgtatccgcctcctccccatacacttcctctccctccgggtgttcctcctcca
  P  P  L  Y  P  P  P  P  H  T  L  P  L  P  P  G  V  P  P  P
cagttttctcctcagtttcctcctggccagccaccacccgctgggtatagtgtccctcct
  Q  F  S  P  Q  F  P  P  G  Q  P  P  P  A  G  Y  S  V  P  P
ccagggtttcctccagctcctgccaatttatcaacaccttgggtatcatcaggagtgcag
  P  G  F  P  P  A  P  A  N  L  S  T  P  W  V  S  S  G  V  Q
acagctcattcaaataccatcccaacaacacaagcaccacctttgtccagggaagaattc
  T  A  H  S  N  T  I  P  T  T  Q  A  P  P  L  S  R  E  E  F
tatagagagcagcgacgactaaaagaagaggaaaagaaaaagtccaagctagatgagttt
  Y  R  E  Q  R  R  L  K  E  E  E  K  K  K  S  K  L  D  E  F
acaaatgattttgctaaggaattgatggaatacaaaaagattcaaaaggagcgtaggcgc
  T  N  D  F  A  K  E  L  M  E  Y  K  K  I  Q  K  E  R  R  R
tcattttccaggtctaaatctccctatagtggttcttcgtattcaagaagttcatatact
  S  F  S  R  S  K  S  P  Y  S  G  S  S  Y  S  R  S  S  Y  T
tattctaaatcaagatctggttcaacacgttcacgctcttattctcgatcattcagccgc
  Y  S  K  S  R  S  G  S  T  R  S  R  S  Y  S  R  S  F  S  R
tcacattctcgttcctattcacggtcacctccatacccccagaagaggcagaggcaagagc
  S  H  S  R  S  Y  S  R  S  P  P  Y  P  R  R  G  R  G  K  S
cgcaattaccgttcacggtctagatctcatggatatcatcgatctaggtcaaggtcaccc
  R  N  Y  R  S  R  S  R  S  H  G  Y  H  R  S  R  S  R  S  P
ccttacagacgctatcattcacgatcaagatctcctcaagcgtttaggggacagtctcct
  P  Y  R  R  Y  H  S  R  S  R  S  P  Q  A  F  R  G  Q  S  P
aataaacgtaatgtacctcaaggggaaacagaacgtgaatatttttaatagatacagagaa
  N  K  R  N  V  P  Q  G  E  T  E  R  E  Y  F  N  R  Y  R  E
gttccaccaccatatgacatgaaagcatattatgggagagagtgttgactttagagaccca
  V  P  P  P  Y  D  M  K  A  Y  Y  G  R  S  V  D  F  R  D  P
tttgaaaagaacgctaccgagaatgggagagaaaatatagagagtggtatgaaaaatat
  F  E  K  E  R  Y  R  E  W  E  R  K  Y  R  E  W  Y  E  K  Y
tataaaggttatgctgctggagcacagcctagaccctcagcaaatagagagaactttct
  Y  K  G  Y  A  A  G  A  Q  P  R  P  S  A  N  R  E  N  F  S
ccagagagattttgccacttaacatcaggaattctcccttcacaagaggccgcagagaa
  P  E  R  F  L  P  L  N  I  R  N  S  P  F  T  R  G  R  R  E
gactatgttggtgggcaaagtcatagaagtcgaaacataggtagcaactatccagaaaag
  D  Y  V  G  G  Q  S  H  R  S  R  N  I  G  S  N  Y  P  E  K
ctt
  L
```

Figure 13: DNA sequence mapping the RS domain on RBBP6. The RS domain spans residues 769-788 of RBBP6 flanked by unique SalI and HindIII cut sites. cDNA corresponding to the sequence highlighted in yellow was synthesized by GenScript Inc., omitting the bulk of the repetitive serine-arginine residues (shown in grey) that constitute the RS domain of RBBP6. This RBBP6-ΔRS cDNA was supplied in a pUC57 vector for sub-cloning into a pCMV-UWC vector as a SalI/HindIII fragment.
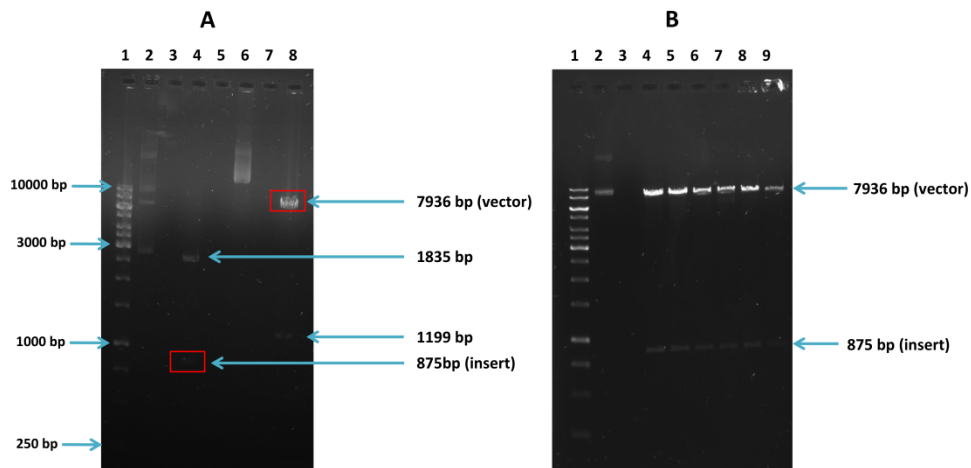
Figure 14: Construction of the pCMV-UWC-RBBP6-ΔRS DNA. (A) Restriction digest of the pUC57 construct from GenScript containing the ΔRS fragment (lane 4) and the pCMV-UWC-RBBP6 (lane 8) construct with SalI and HindIII to release the insert (875 bp) and the vector DNA (7936 bp), respectively. Lanes 2 and 6 indicate the undigested pUC57 and pCMV-UWC-RBBP6, respectively. (B) Positive pCMV-UWC-RBBP6-ΔRS clones were screened by restriction digest using SalI and HindIII to release the initially cloned DNA fragments. Lane 2 represents the undigested pCMV-UWC-RBBP6-ΔRS construct and the bands on the bottom of lanes 4 to 8 correspond to the 875 bp insert released from the vector.
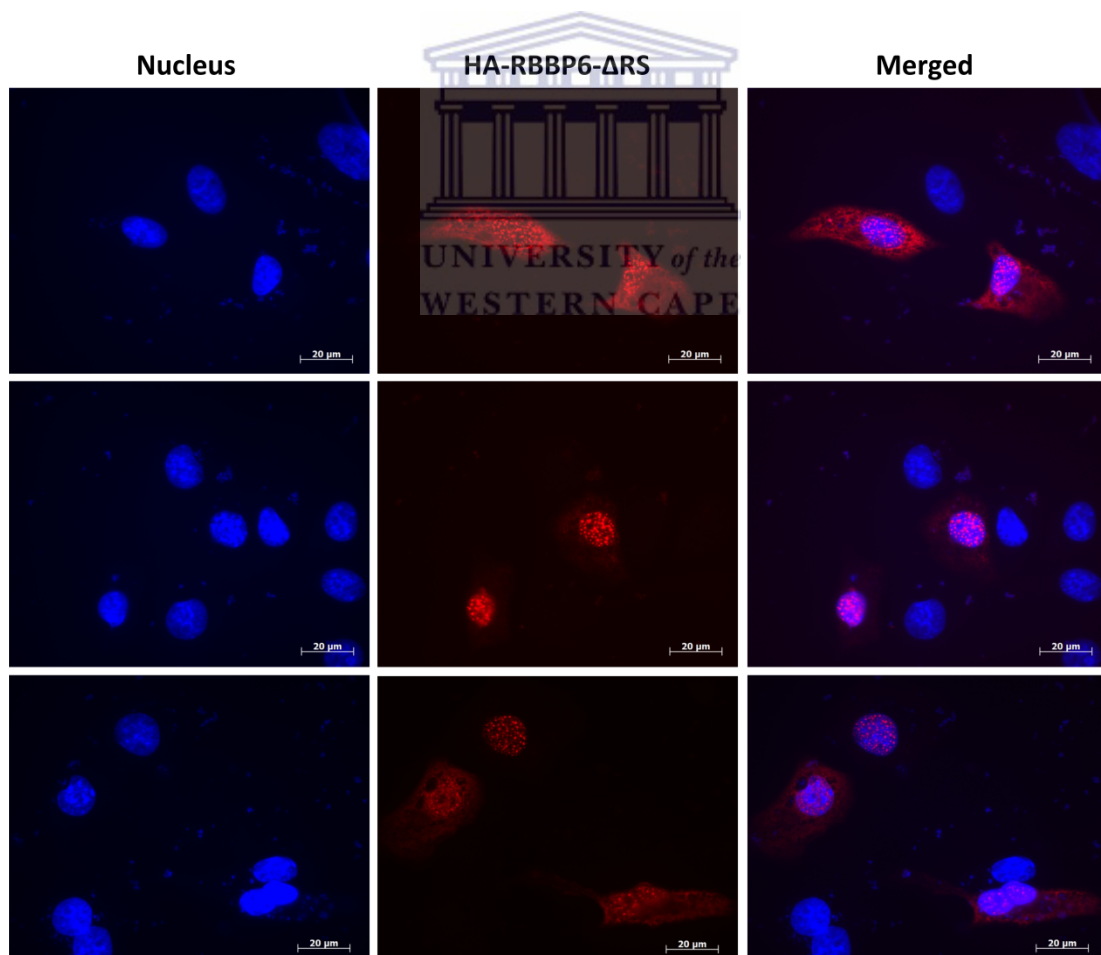
Figure 15: The RS domain of RBBP6 is not required for localisation to nuclear speckles. Cells transfected with the pCMV-UWC-RBBP6-ΔRS DNA construct displays a mixed localisation phenotype that appears to be cell dependent. The protein shows a diffuse localisation into the nucleus and cytoplasm of transfected cells, along with the distinct speckling phenotype associated with full length RBBP6.

54

3.6 *The most C-terminal end of RBBP6 encodes a weak nuclear localisation sequence*

The extreme C-terminus of RBBP6 has been previously characterised as a nuclear localisation sequence (NLS) based on the high density of lysine residues (*Simons et al., 1997*). We decided to test whether this region does in fact play the role of a NLS, by deleting residues 1727 - 1788 to produce HA-RBBP6-ΔNLS.

To do so we took advantage of an unique BlpI site (5'-GC|TNAGC-3') site spanning codons 1719-1721, just upstream of the start of the putative NLS at 1727, as well as the KpnI cloning site at the C-terminus of the protein. Overlapping complementary oligonucleotides were designed which, when annealed, would produce the correct overhangs for ligation into HA-RBBP6 following removal of the BlpI/KpnI fragment. The oligonucleotides are presented in Table 1, row five. As an additional step to facilitate suppression of wild type background, the KpnI site was designed so that the reconstituted KpnI site was replaced with 5'-CGTACC-3' (instead of 5'-GGTACC-3') so that it could no longer be digested by KpnI. Hence HA-RBBP6-ΔNLS would be linearized by BlpI but not by KpnI (see Figure 16A, lanes 5, 7 and 8), whereas HA-RBBP6 would be linearized separately by both enzymes (see Figure 16A, lane 3). Undigested HA-RBBP6-ΔNLS (lane 2) was also added to indicate how the banding profile should look in a case, where, for some reason, the BlpI enzyme did not cut to completion. Positive clones were sequenced using the custom-designed M13-R primer (Inqaba Biotechnical Industries, Hatfield, South Africa) and found to be correct.

Endotoxin-free preparations of HA-RBBP6-ΔNLS were transfected into A549 cells, as described previously. Surprisingly it showed very little cytoplasmic localisation, as can be seen in Figure 16B. We conclude from this that the so-called nuclear localisation sequence at the extreme C-terminus of RBBP6, unlike the DWNN and RS domains, plays a minimal role in localizing HA-RBBP6 to the nucleus. The DWNN domain functions as the most influential NLS, followed by the RS domain and lastly the C-terminal 'NLS' (see Figure 17). Figure 17 displays the relative strengths of the NLS on RBBP6, this judgment of 'relative' strengths is based on the immunofluorescence results, and so in needs to be taken into account that it's more a qualitative and suggestive reasoning, rather than definitive conclusions.

55
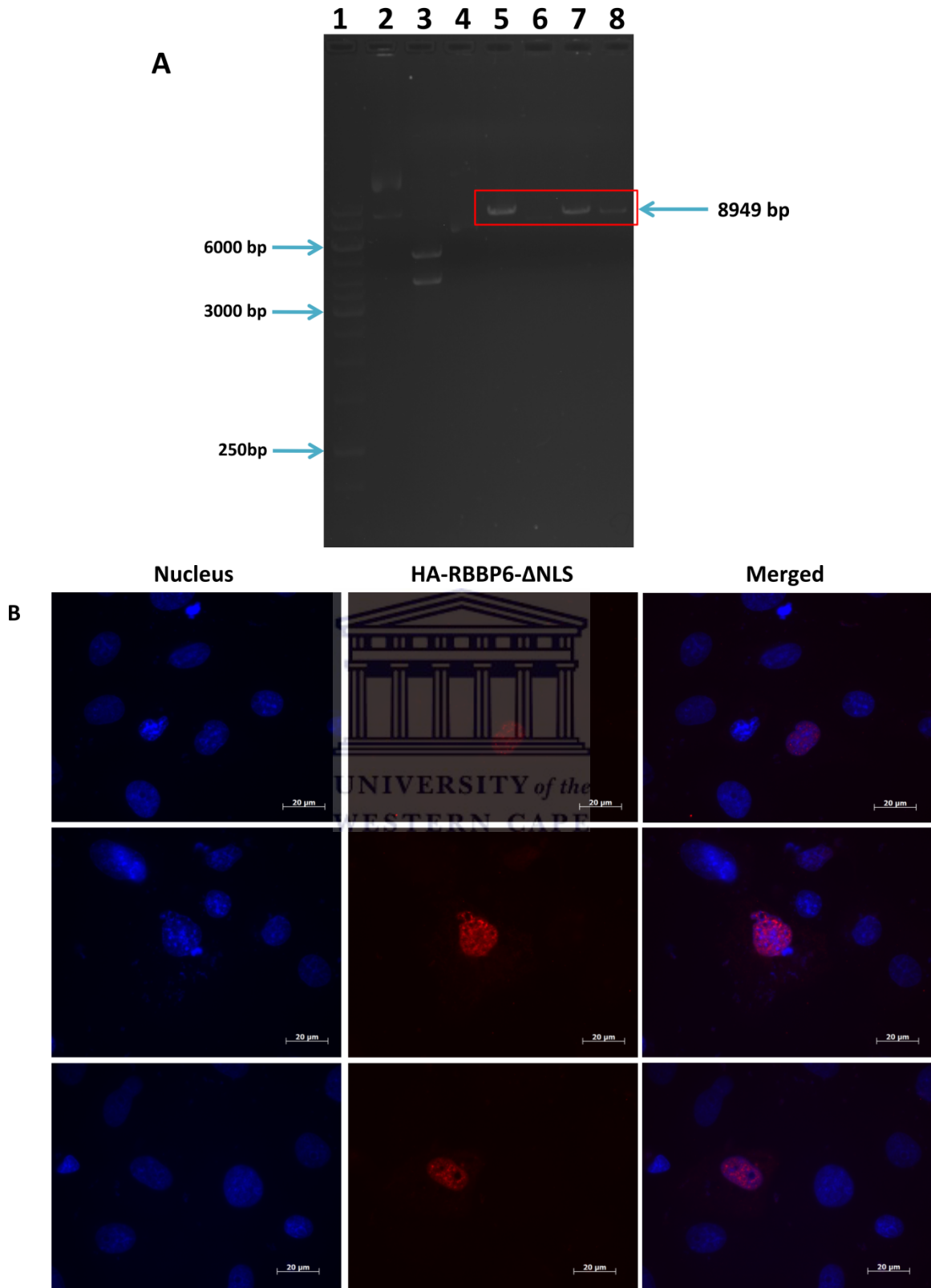
Figure 16: Cloning and cellular localisation of pCMV-UWC-RBBB6-ΔNLS. (A)Restriction enzyme screening of the positive RBBP6-ΔNLS clones. The cloning was screened using BlpI and KpnI. Positive clones can only be linearized by BlpI since the KpnI site was destroyed. Hence, lanes 5, 7 and 8 contain RBBP6-ΔNLS. Lanes 2 and 3 show the undigested RBBP6-ΔNLS DNA and digested wild type RBBP6, respectively. (B) Shows the localisation phenotype of HA-RBBP6-ΔNLS. The protein still localises to nuclear speckles, although the speckles are less distinct than those seen for wild type RBBP6. HA-RBBP6-ΔNLS also shows a very weak cytoplasmic diffusion.

56
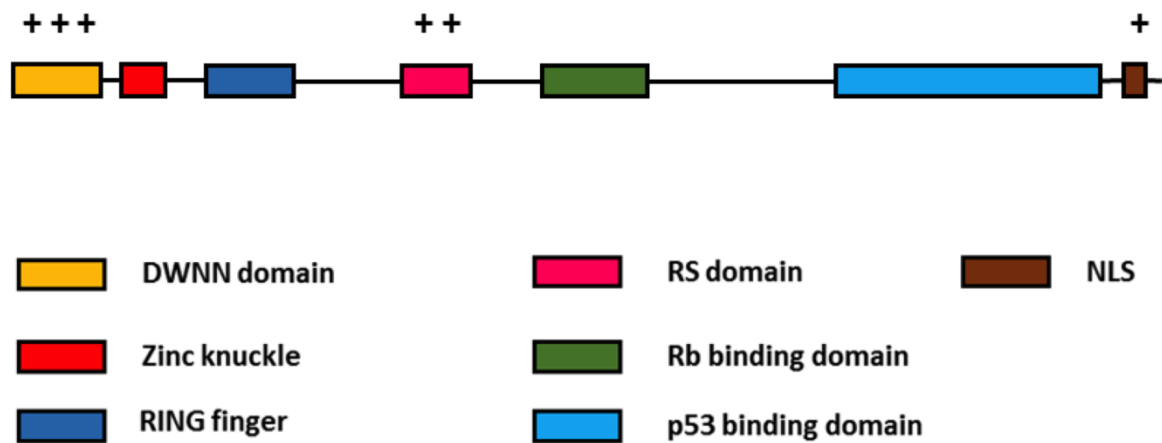
Figure 17: Schematic representation of wild type RBBP6 isoform 1, the distribution of the NLS and their relative strengths. RBBP6 constitutes at least three functional NLS, all functioning at varying degrees. The DWNN is the most influential NLS, followed by the RS domain and lastly the C-terminal 'NLS'. The (+) indicates the relative strengths of each of the NLS found along the length of RBBP6.

3.7   *Attempts to identify a nuclear export signal in the C-terminal tail of isoform 3*

A construct expressing HA-DWNN13 (isoform 3) was previously made by Ms Andronica Ramaila in our laboratory. When transfected into A549 cells the protein was found to have a primarily cytoplasmic localisation with evidence of nuclear speckles (Figure 18, panels A-C). This is consistent with previous results showing that endogenous isoform 3 is primarily cytoplasmic with evidence of nuclear speckles (A. V. Szmyd-Potapczuk, UWC PhD thesis, 2017). As in the endogenous case, the nuclear speckling appears to become more intense following stress (Figure 18, panels D-F). Since our previous results suggest that DWNN domain plays the role of a strong nuclear targeting signal, this suggests that the tail region [residue 80-118 (see Figure 20 showing a schematic representation of the DWNN-13 deletion constructs)] may contain a cytoplasmic localisation signal, also called a nuclear export signal (NES), sufficiently strong to override the DWNN domain.
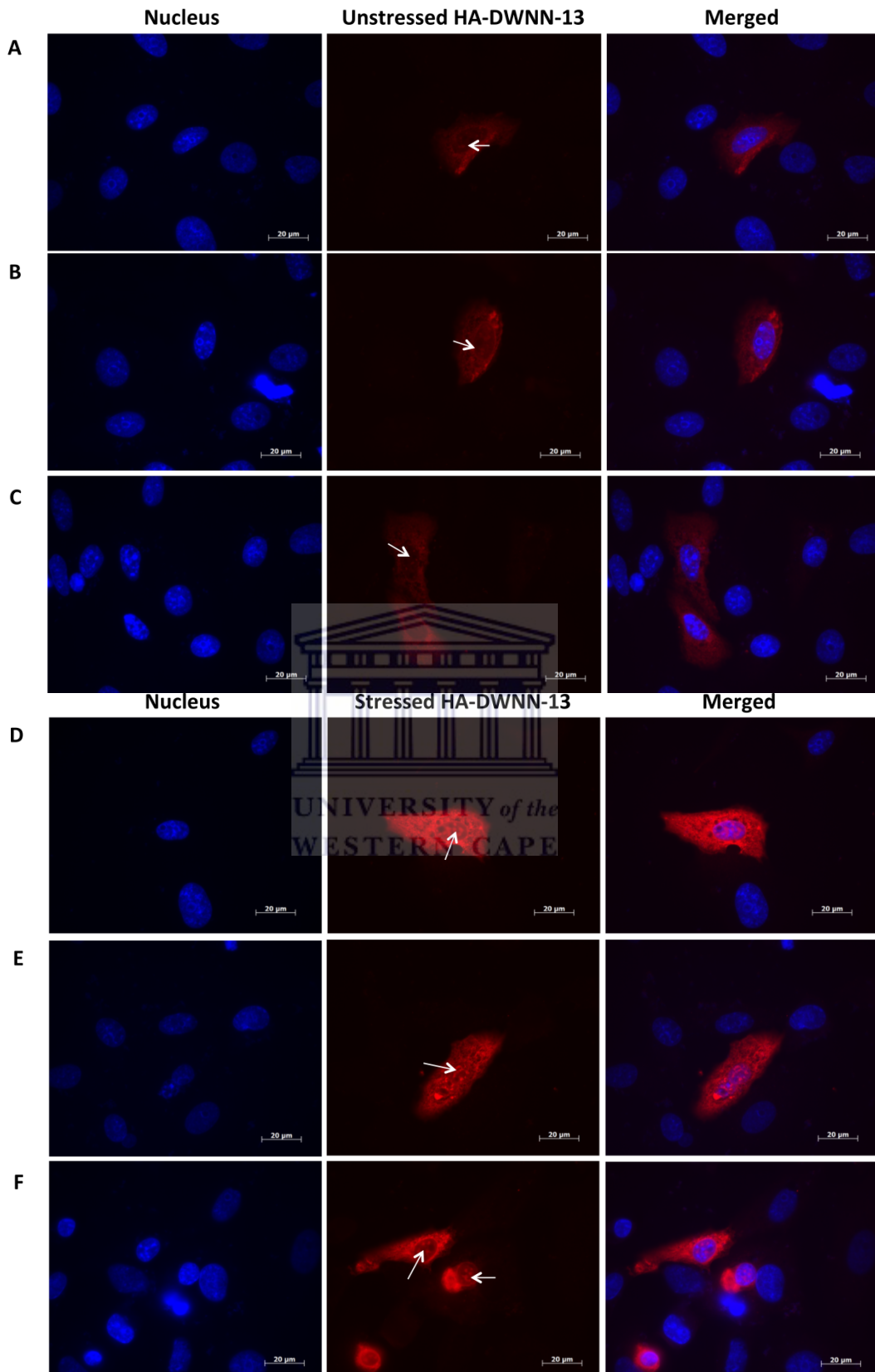
Figure 18: DWNN13 is predominantly cytoplasmic and the accumulation to nuclear speckles appears to be stress-dependent. Cells transfected with pCMV-UWC-DWNN13 display a predominantly cytoplasmic localisation phenotype under physiological and stressful conditions. Under stress, HA-DWNN13 also shows an accumulation of the protein in nuclear speckles.

58

In order to investigate the possibility of there being a NES in the C-terminal tail of isoform 3, we designed two truncations of HA-DWNN13. The first spanned residues 1-79, ending in Gly-Gly, and was therefore named HA-DWNN-GG. The second truncation spanned residues 1-77, ending in Pro-Ile, and was therefore named HA-DWNN-PI (See Figure 20). pCMV-UWC-DWNN-PI had previously been cloned by Ms Tephney Mahomedy. pCMV-DWNN-GG was constructed as part of this work, by amplification using the primers shown in Table 1, second row (construction shown in Figure 19). Sequencing was carried out using a custom-designed primer (M13-Int: 5'-GTTGCCTTTACTTCTAGG-3'), which anneals approximately 70 bp upstream of the MCS in the pCMV-UWC plasmid. This was necessary because the M13 forward primer anneals 830 bp upstream of the beginning of the MCS, which is too far for standard sequencing reads.

The localisation of both HA-DWNN-PI and HA-DWNN-GG was found to be similar to that of HA-R3, with similar intensity of signal inside and outside the nucleus (Figure 21). Like HA-R3, the nuclear signal was uniform, with no evidence of speckling. Unlike HA-DWNN13, neither HA-DWNN-PI nor HA-DWNN-GG showed any evidence of speckling, either before or after induction of stress. The localisation of HA-DWNN-PI and HA-DWNN-GG can therefore be characterised as being more similar to that of HA-R3 than to HA-DWNN13.
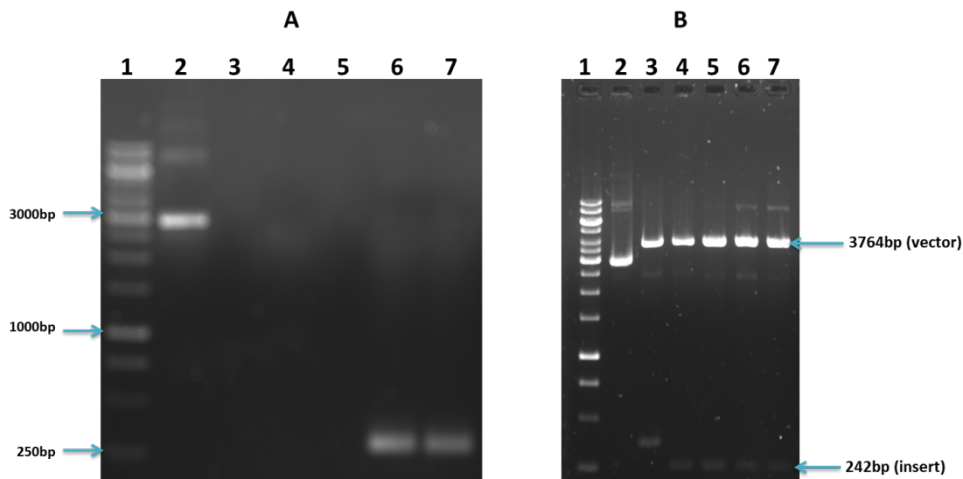
Figure 19: Cloning of the pCMV-UWC-DWNN-GG DNA. (A) PCR Amplification of DWNN-GG from the pCMV-UWC-DWNN13 construct**.** The 234 bp fragment in lanes 6 and 7 was amplified from wild type DWNN-13, incorporating an XhoI and KpnI restriction sites on the forward and reverse primers, respectively. Lane 2 represents the template DNA and lane 4 represents the negative control. (B) Restriction digest screening for the positively cloned DWNN-GG construct. Positive clones were verified by restriction digest using XhoI and KpnI to release the ligated DNA fragments. The bands on the bottom of lanes 4-8 correspond to the 242 bp insert and the high molecular weight bands in the same lanes indicate the vector from the digested pCMV-UWC-DWNN-GG construct. Lane 2 represents the undigested clone DNA.
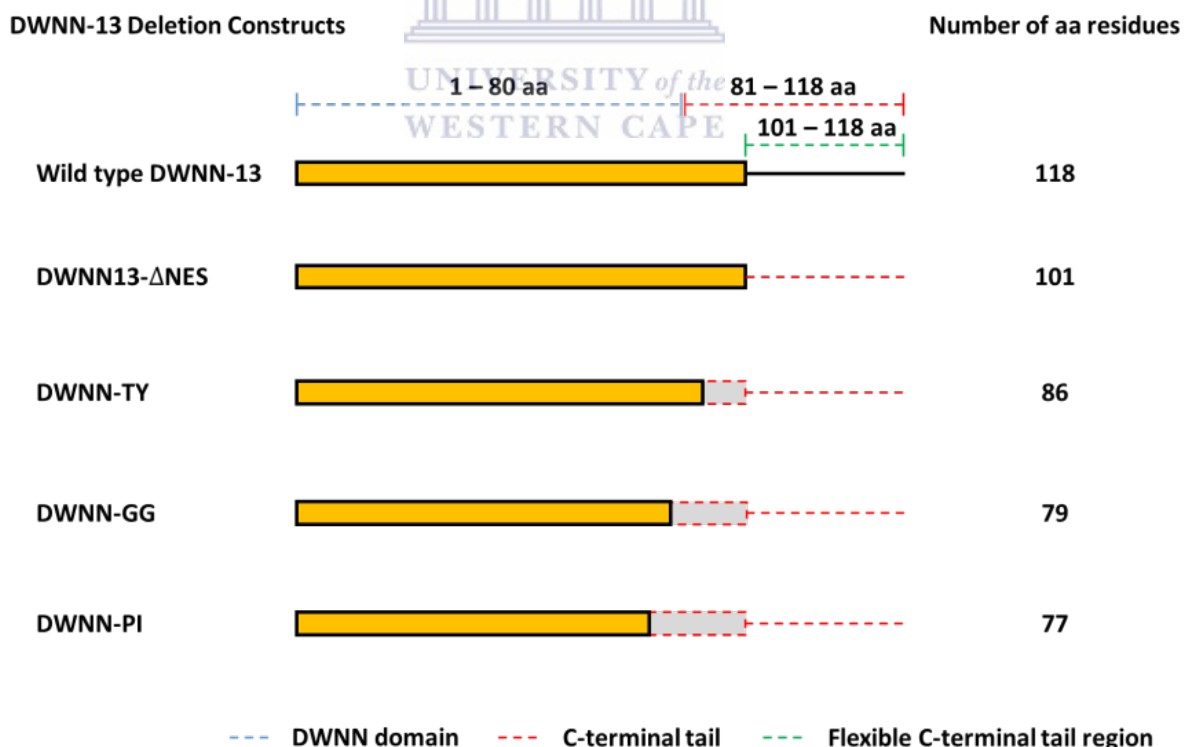
Figure 20: Schematic representation of the DWNN-13 deletion constructs used in this study. The different DWNN-13 constructs used in this study were designed using PCR-based gene manipulation techniques, allowing for specific fragments of DWNN-13 to be removed. The forward primer used was the same for all these constructs.
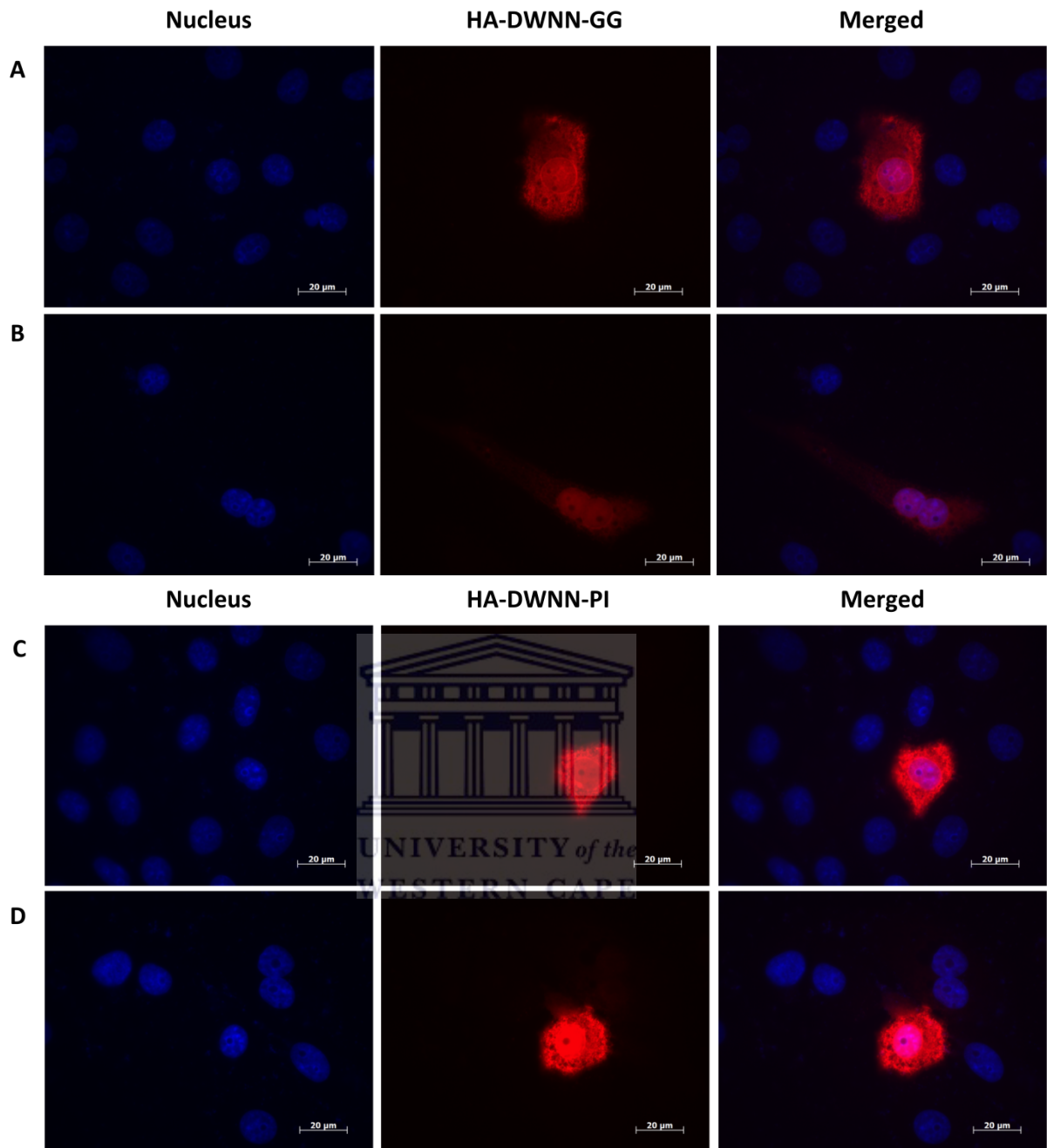
Figure 21: Both DWNN-GG and DWNN-PI are evenly distributed across the cytoplasm and nucleus of transfected cells. Cells transfected with pCMV-UWC vectors containing DWNN-GG (panel A and B) and DWNN-PI (panel C and D) truncations of HA-DWNN13 displayed an even localisation pattern across the nucleus and cytoplasm.

61

The above results lend support to our hypothesis that the DWNN domain plays a role in targeting a significant fraction of cellular RBBP6 to the nucleus. Since HA-DWNN13 is almost completely cytoplasmic, we concluded that somewhere in the C-terminal tail (residues 80-118) there must be a signal driving export of HA-DWNN-13 to the cytoplasm; i.e. a nuclear export signal (NES).

The last 17 residues of HA-DWNN13 [residues 102-118 – VCKNTISHFFYTLLLPL (see Figure 20)] are not found in any of the other isoforms of RBBP6. NES sequences typically contain a number of lysines; the signature (canonical) sequence is LxxxLxxLxL (*Nie et al., 2007*). According to the NetNES 1.1 Server the sequence ISHFFYTL forms a NES; according to the LocNES Server the sequence VCKNTISHFFYTLLL forms a NES. Hence we decided to truncate isoform 3 immediately following residue 101. A reverse primer was designed incorporating translational stop codons followed by a KpnI site, as shown in Table 1, row three. The forward primer was the same as for HA-DWNN-GG, and incorporated an XhoI site. The cDNA was amplified using HA-DWNN13 as template and cloned into the XhoI-KpnI sites of pCMV-UWC to create pCMV-UWC-DWNN13-ΔNES (data not shown). The construct was sequenced using the M13-Int primer and found to be correct.

Large quantities of endotoxin free pCMV-UWC-DWNN13-ΔNES DNA were isolated and used for DNA transfections as previously described for the other pCMV constructs. Surprisingly, when investigated using immunofluorescence microscopy, the localisation of the new construct (shown in Figure 22) was more similar to that of HA-DWNN13 (predominantly cytoplasmic, see Figure 18) than HA-DWNN-PI or HA-DWNN-GG. This suggests that the last 17 amino acids do not act as a nuclear export signal. However there is little evidence for nuclear speckling, which possibly suggests that the driver of speckling is in the extreme C-terminus of isoform 3. There was also no evidence of increased accumulation in nuclear speckles following heat shock.
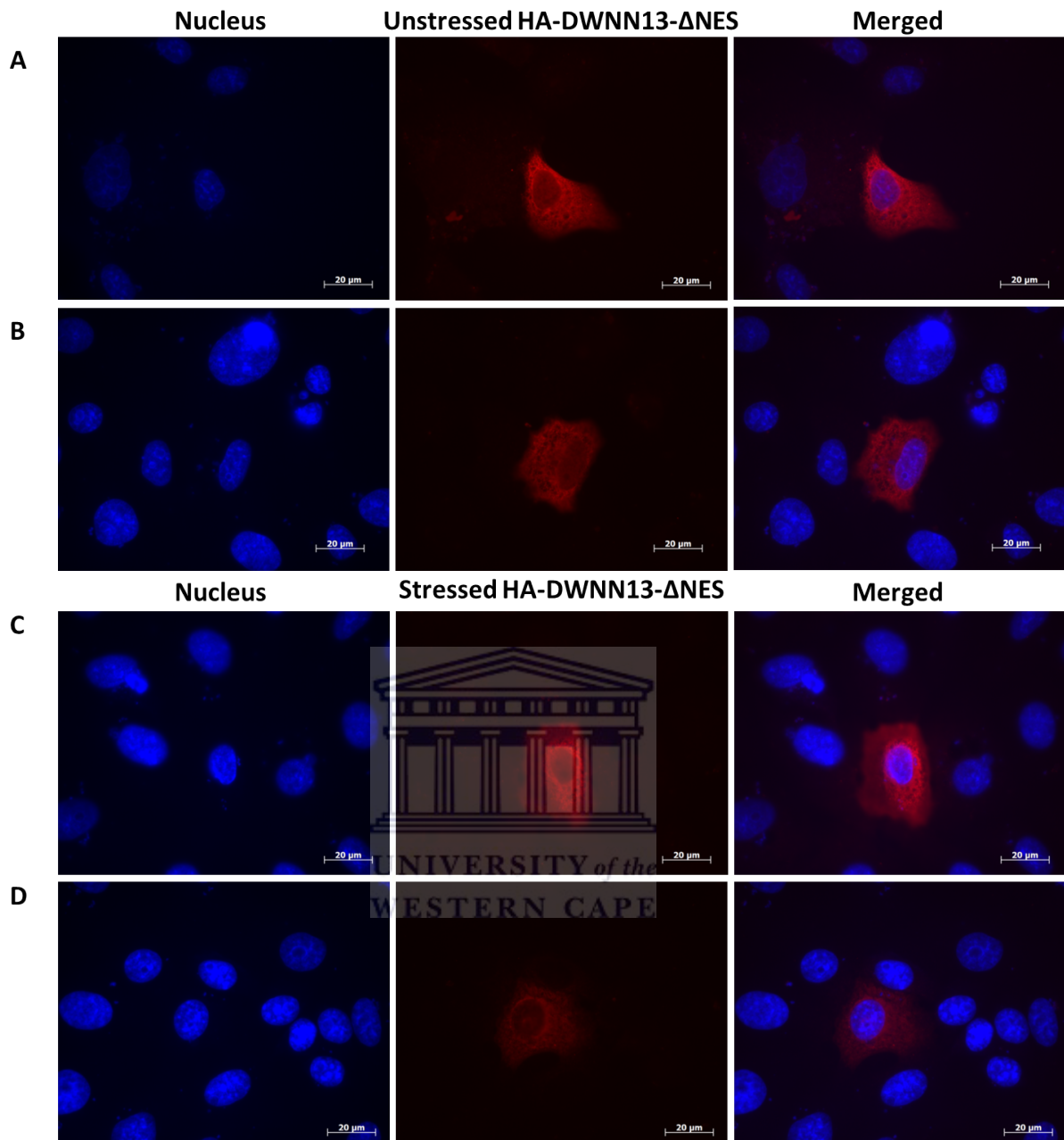
Figure 22: HA-DWNN13-ΔNES is predominantly cytoplasmic and does not accumulate in nuclear speckles. Cells transfected with pCMV-UWC-DWNN13-ΔNES also display a predominantly cytoplasmic localisation phenotype, with no evidence of nuclear speckling, either before or after stress.

So far, we have shown that truncating isoform 3 following Ala101 (i.e. DWNN-ΔNES) does not change the phenotype (mainly cytoplasmic), whilst truncating it following Ile77 or Gly79 produces a uniform cytoplasmic and nuclear phenotype distinct to that observed for wild type isoform 3. This suggests that the nuclear export signal is somewhere between residues 80 and 100. Of this region, the region 80-86 (KSTSKTY) is particularly rich in phosphorylatible amino acids and positively charged residues. Hence we constructed HA-DWNN-TY by

63

truncating isoform 3 immediately after Tyr86 (see Figure 20).

HA-DWNN-TY was constructed by amplifying residues 1-86 of HA-DWNN13. A reverse primer was designed incorporating stop codons followed by a KpnI site, as shown in Table 1, row four. The forward primer was the same as for HA-DWNN-GG, and incorporated an XhoI site (see Figure 23). The construct was validated by direct DNA sequencing and found to be correct.

Unfortunately time did not allow for investigation of the localisation of HA-DWNN-TY.
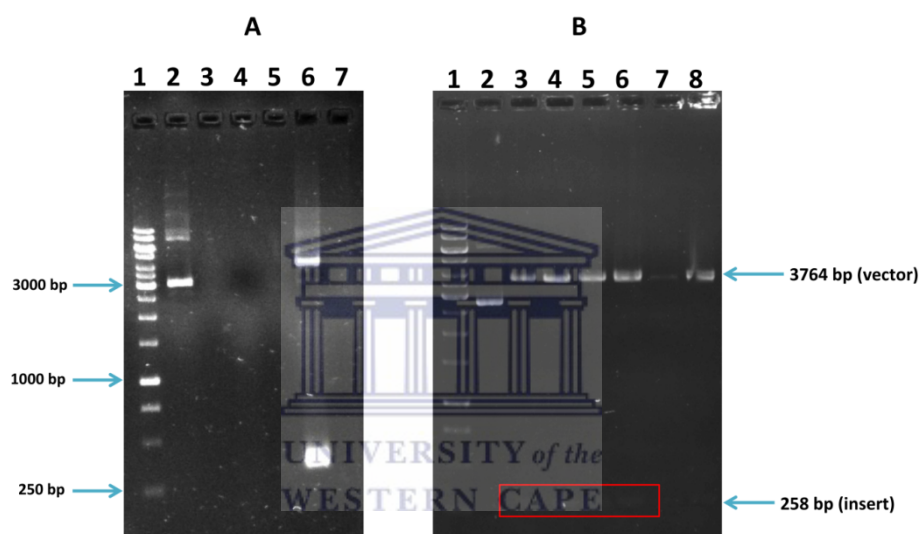
Figure 23: Cloning of pCMV-UWC-DWNN-TY. (A) PCR Amplification of DWNN-TY from the pCMV-UWC-DWNN13 construct**.** The 258 bp fragment in lane 6 was amplified from wild type DWNN13, incorporating an XhoI and KpnI restriction sites on the forward and reverse primers, respectively. Lane 2 represents the template DNA used for amplification, while lane 4 represents the negative PCR control. (B) Restriction digest screening with XhoI and KpnI to release the DWNN-TY construct. Positive clones were verified by restriction digest using Xho I and Kpn I to release the ligated DNA fragments, indicated in red at the bottom of lanes 3-7. The high molecular weight bands in the same lanes indicate the vector from the digested pCMV-UWC- DWNN13-TY construct. Lane 2 is the undigested DNA.

3.8   *Attempts to disrupt the nuclear localisation potential of the DWNN domain*

One of the conclusions of this work is that the DWNN domain acts as a nuclear localisation signal. This is most likely to occur as a result of a protein-protein interaction between the DWNN domain and some unidentified protein that transports it to the nucleus. If so, then it should be possible to find single amino acid mutations on the surface of the DWNN domain that will abolish the interaction, and therefore abolish the nuclear localisation. In other words, it should be possible to find a single amino acid mutation in R3 that has the same localisation phenotype as HA-R3-ΔDWNN.
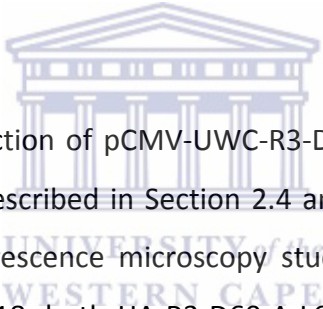
SAP18 is a ubiquitin-like protein which targets other proteins to nuclear speckles as a result of a protein-protein interaction (*Singh et al., 2010*). Based on the analogy with SAP18 we initially targeted residues D60 and L63 of RBBP6, which correspond to residues D118 and T121 of SAP18 (see alignment on Figure 24). Overlapping primers were designed based on the method of Liu and Naismith (see Table 1, row six), which is an extension of the QuikChange protocol (*Liu and Naismith, 2008*). pCMV-UWC-R3 (residues 1-335 of full length RBBP6) was mutated rather than pCMV-UWC-RBBP6 (1792 residues) in an attempt to avoid unwanted mutations being introduced into the DNA sequence, as it is more likely to occur with a much larger construct such as full length RBBP6. The parameters used for the mutagenesis PCR are shown in Table 2 found in Section 2.3.3.

Sequencing of putative pCMV-UWC-R3-D60A-L63A mutants was carried out using the custom-designed primer described above (M13-Int: 5'-GTTGCCTTTACTTCTAGG-3') which gave high quality coverage across the DWNN domain and showed it to be correct, including the desired mutations. The D60A-L63A mutations were then transferred to full length RBBP6 through restriction enzyme-base manipulation as described below.

pCMV-UWC-R3-D60A-L63A encodes residues 1-335 of RBBP6 cloned between the BglII and SalI sites of pCMV-UWC. The unique AvrII site referenced earlier is located in codon 186 of RBBP6. The strategy was to remove the AvrII/KpnI fragment from pCMV-UWC-R3-D60A-L63A, corresponding to residues 187-335 (end) of R3, and replace it with the AvrII/KpnI fragment excised from pCMV-UWC-RBBP6, corresponding to residues 187-1792 (end). In simple language this corresponded to inserting the back end of RBBP6 into a mutated R3 to

produce a mutated RBBP6.

Restriction digest of pCMV-UWC-RBBP6 with AvrII and KpnI generated bands of 4828 bp, corresponding to residues 187-1792 of RBBP6, and 4307 bp, corresponding to residues 1-186 of RBBP6 cloned into pCMV-UWC. The 4828 bp band from RBBP6 was excised and purified, to serve as the "insert". Restriction digest of pCMV-UWC-R3-D60A-L63A with the same enzymes yielded a 4309 bp band, corresponding to the mutant R3 cloned into pCMV-UWC, and a 149 bp fragment corresponding to residues 187-335 of R3. The 4309 bp band was excised and purified to serve as the "vector". Finally, the "insert" and "vector" were ligated together and transformed into competent *E. coli* cells. DNA sequencing was carried out using the M13 reverse primer to confirm that the C-terminus of RBBP6 had been correctly inserted, and using the M13-Int custom primer to show that the D60A-L63A mutations were still present in the longer construct. The construct was found to be correct and was named pCMV-UWC-RBBP6-D60A-L63A.

Large-scale endotoxin-free extraction of pCMV-UWC-R3-D60A-L63A and pCMV-UWC-RBBP6-D60A-L63A was carried out as described in Section 2.4 and the constructs were transfected into A549 cells for immunofluorescence microscopy studies. In contrast to the mutations made in the model protein, SAP18, both HA-R3-D60-A-L63A and RBBP6-D60A-L63A showed no observable difference compared to the wild type proteins (intense signal inside the nucleus and throughout the cytoplasm for R3 (Figure 25, panel A and B) and a completely nuclear and speckling phenotype for RBBP6 (Figure 25, panel C and D)). From these results, we concluded that D60 and L63 do not play a significant role in the function of the DWNN domain as a nuclear localising signal.

```
A




B
Query = 2HDE (SAP18)
SBJCT = 2C7H (DWNN)

Query  gshmrvtqeeikkepekpidrektsPLLLRVFTTNngrhhrmdefsrgnvPSSELQIYTW   60
ident
Sbjct  ------------------------MSCVHYKFSS---------------KLNYDTVTFD   20


Query  -MDATLKELTSLVKEVYPeARKKGTHFNFAIVFtdvkrpgyRVKEIGstmsgrkgtDDSM  119
ident       |  |                                  |         ||
Sbjct  gLHISLCDLKKQIMGREK-LKAADCDLQITNAQ--------TKEEYT---------DDNA   62


Query  TLqsqkfQIGDYLDIAITPPnrAPPPsgr  148
ident     |
Sbjct  LI-----PKNSSVIVRRIPI--GGVK---   81

DALI Server at http://ekhidna2.biocenter.helsinki.fi/dali/
```

Figure 24: Alignment of the three dimensional structures of the DWNN and SAP18 proteins. Both the proteins adopt a ubiquitin-like "β-grasp" fold consisting of a five-stranded anti-parallel β-sheet curled around a central α-helix. (A) Superposition of the three dimensional structures of DWNN (blue) and SAP18 (green), showing that D118 and T121 of SAP18 superimpose almost exactly on D60 and L63 of DWNN. (B) Structural alignment of the primary sequences of DWNN (subject) and SAP18 (query). The amino acid residues indicated in green are D118 and T121 from SAP18, the ones indicated in pink are the structurally equivalent residues from DWNN, D60 and L63. The image was generated using the PyMOL molecular graphics system (Delano Scientific LLC, San Carlos, CA, USA).
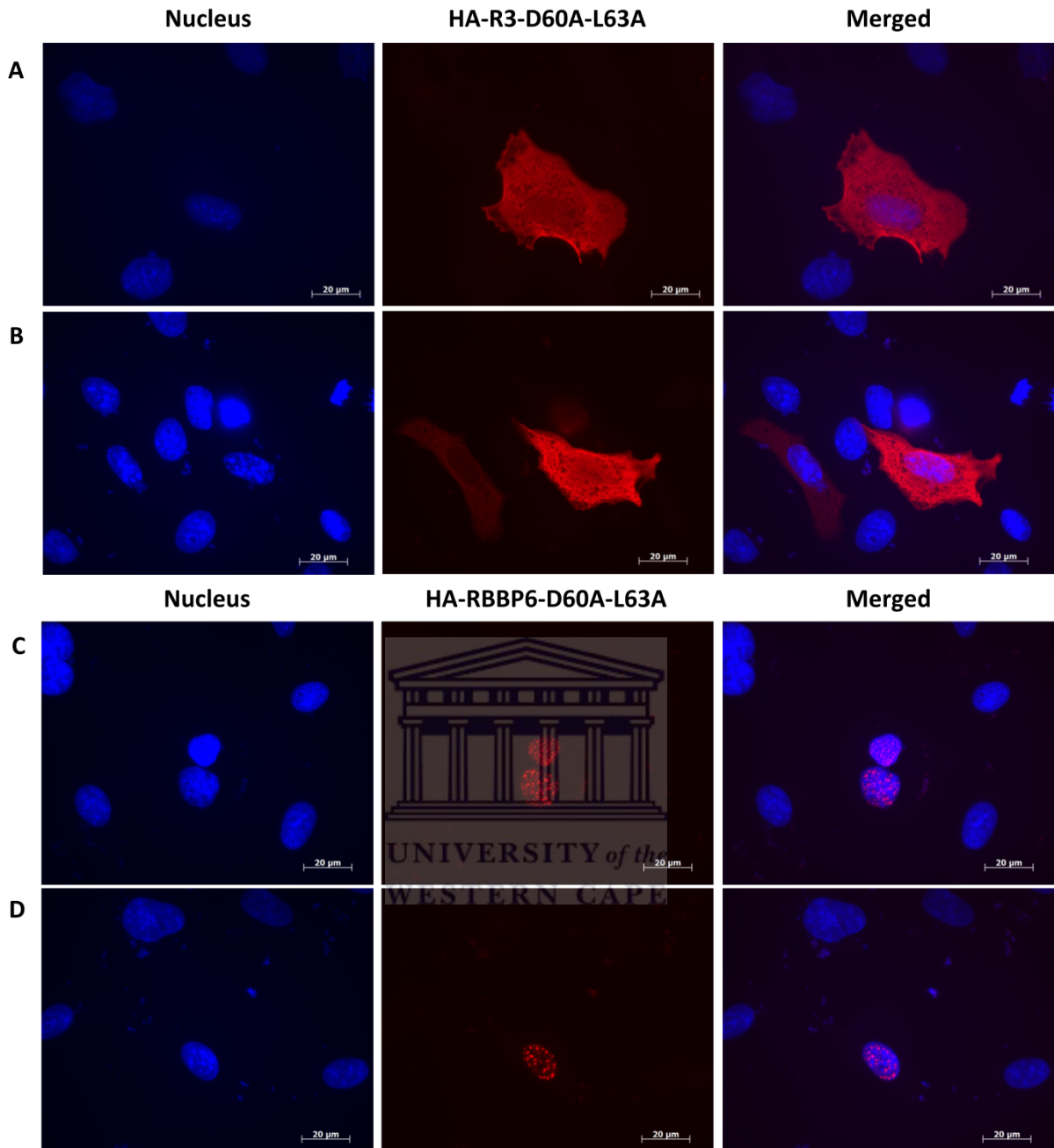
Figure 25: HA-R3-D60A-L63A and HA-RBBP6-D60A-L63A mutants show the same localisation phenotypes as their respective wild type counter parts. Cells transfected with the pCMV-UWC-D60A-L63A mutant constructs showed no distinction to the wild type HA-R3 and HA-RBBP6 proteins. HA-R3-D60A-L63A (top two rows) displays an even distribution across the nucleus and cytoplasm, whilst HA-RBBP6-D60A-L63A (bottom two rows) displays a sub-nuclear localisation into nuclear speckles.

68

Mutating D60 and L63 was something of a gamble that did not pay off. A better approach would have been to conduct a comprehensive alanine scan, replacing all surface residues on the DWNN domain with alanine residues. However, time only allowed for a smaller number of mutations to be performed, and the chances were good that whichever surface residues we mutated would not be the correct ones. Hence we decided to take a more conservative approach which involved trying to unfold the DWNN domain. Unfolding the entire DWNN domain is likely to abolish any biologically-significant protein-protein interaction and has the advantage that it is more likely to succeed than mutation of individual surface residues. Hence, with reference to the published structure of the DWNN domain (*Pugh et al., 2006*), we designed two double mutations that were highly likely to disrupt the folded structure.

The first mutation involved replacing Leu29 and Ile33 with prolines. Leu29 and Ile33 are hydrophobic residues on the inside of the main α-helix which extend into the core between the α-helix and the 5-stranded β-sheet which makes up the "β-grasp" fold of ubiquitin (shown in red in Figure 26). Prolines are much smaller than leucines and isoleucines, which is likely to disrupt the packing of the core, thereby unfolding the domain. In addition, the twisted backbones of prolines are likely to distort the α-helix, thereby disrupting any interaction involving the outside face of the α-helix.

The second involved replacing Val70 and Val72 with prolines. Val70 and Val72 are situated at the center of the 5-stranded β-sheet and their hydrophobic side-chains extend into the core of the domain between the α-helix and the β-sheet (shown in blue in Figure 26). Here too, the twisted backbones of the prolines are likely to distort the β-sheet, in addition to their more widespread effect on the packing of the core. Both mutants were generated using overlapping-primer mutagenesis, which is described in Chapter 2. The two sets of primers are shown in Table 1, rows seven (L29P-I33P) and eight (V70P-V72P). Mutagenesis was carried out and significantly more colonies were obtained on the experimental plate as compared to the control plate. DNA sequencing was carried out using the M13-Int custom primer which gave high quality coverage across the DWNN domain and confirmed the presence of the expected mutations and no others. The constructs were named pCMV-UWC-R3-L29P-I33P and pCMV-UWC-R3-V70P-V72P respectively.
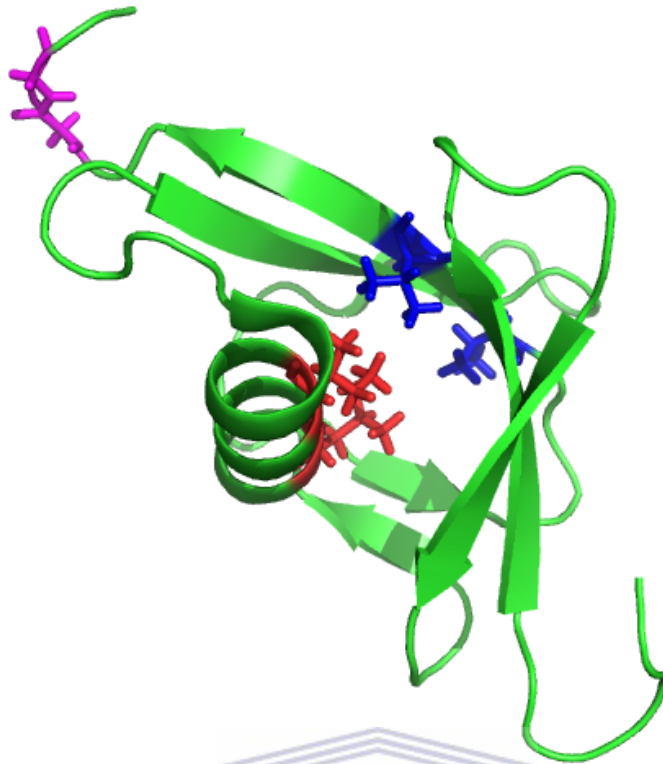
Figure 26: Unfolding mutations made to the DWNN domain to disrupt its nuclear targeting. The unfolding mutations were made on both the β-sheets and α-helix of the DWNN domain. Residues Leu29 and Ile33 (red) on the inside face of the α-helix and residues Val70 and Val72 (blue) located on the inside face of the β-sheet were all simultaneously mutated to proline in the hopes of destabilizing the protein structure, since these residues point into the core between the β-sheet and α-helix. The C-terminal di-glycine motif is shown in purple. The image was generated using the PyMOL molecular graphics system (Delano Scientific LLC, San Carlos, CA, USA).

As expected, immunofluorescence microscopy showed that both mutants are localised entirely in the cytoplasm (Figure 27). Hence their phenotypes were closer to that of R3-ΔDWNN than wild type R3 (compare with Figure 12, panels C and D). In fact, the amount of signal in the nucleus in Figure 27 appears to be even less than in Figure 12 panels C and D, although the difference is not likely to be significant. It is likely that R3-ΔDWNN, R3-L29P-I33P and R3-V70P-V72P are all entirely cytoplasmic and that the faint signal appearing to come from the nucleus is due to the wide-field microscope used in the study, which detects some signal coming from the cytoplasm above and below the plane of the image. Repeating the study using a confocal microscope is therefore expected to produce entirely dark nuclei. Unfortunately time did not allow for confocal analysis to be performed.
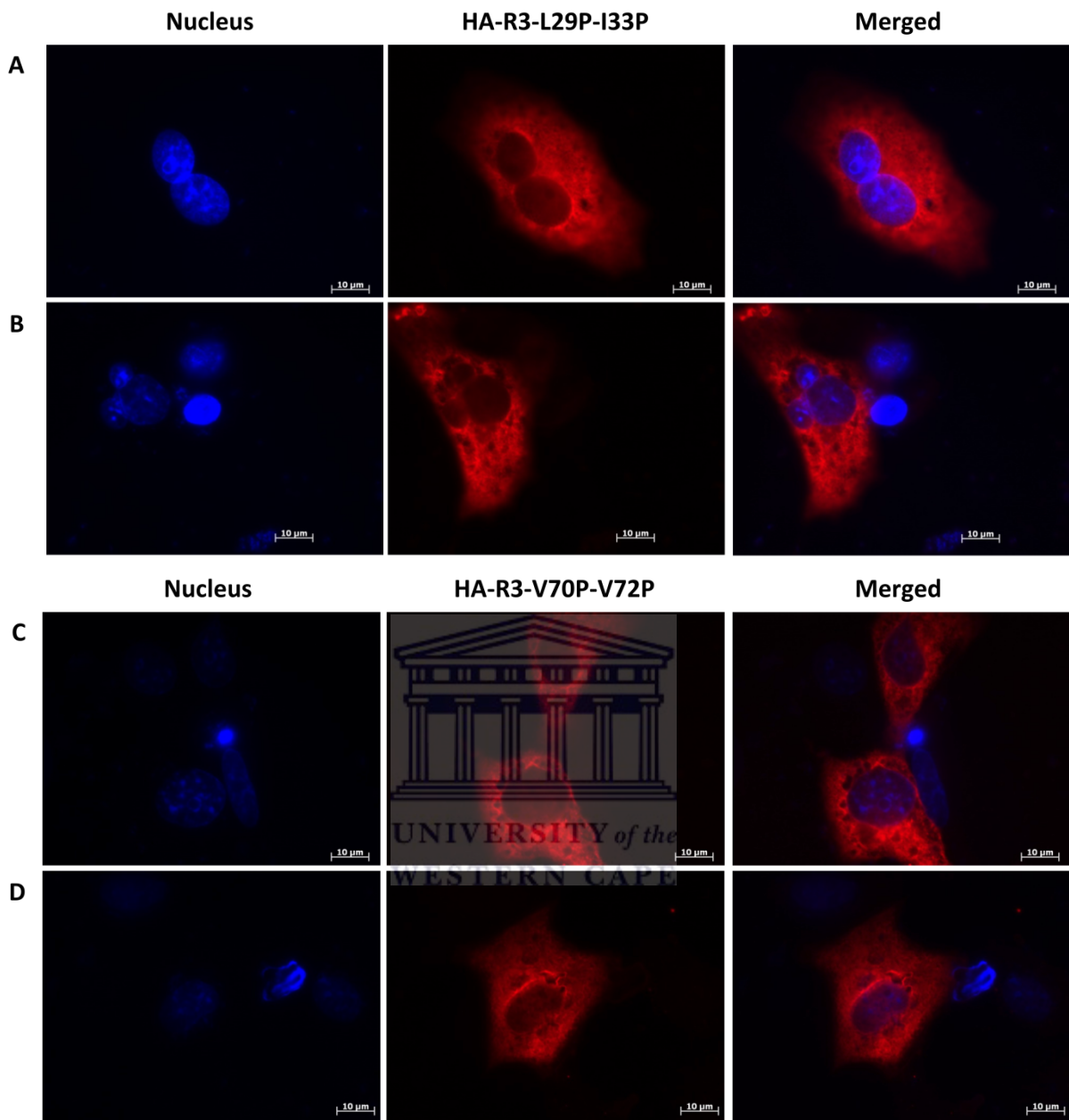
Figure 27: The folded state of DWNN domain is required for nuclear targeting. Cells transfected with the unfolding mutations, L29P-I33P (panels A and B) and V70P-V72P (panels C and D), showed a predominantly cytoplasmic localization phenotype more similar to HA-R3-ΔDWNN, in contrast to the even distribution of wild type R3.

We conclude from the above results that the DWNN domain must be folded in order for it to act as a nuclear localisation signal. The results also show that the interaction with the putative transport protein involves the DWNN domain, and not the region corresponding to residues 80-143, which was also missing from the HA-RBBP6-ΔDWNN and HA-R3-ΔDWNN constructs. This study has therefore laid the groundwork for further mutagenesis studies aimed at identifying surface residues of the DWNN domain which abolish the interaction and therefore targeting to the nucleus.

Since R3 is already partly cytoplasmic, the change produced by the mutants, although convincing, is nevertheless not as impressive as it would be expected to be in full length RBBP6, for which the wild type RBBP6 is exclusively nuclear. Hence the two unfolding mutations were transferred to full length RBBP6 using the same methodology described for the D60A-L63A mutant; the AvrII/KpnI fragment, containing the C-terminal end of RBBP6, was excised from pCMV-UWC-RBBP6 and inserted into the same sites in the R3 mutants. Sequencing was carried out using the M13-R primer to check that the C-terminal end of RBBP6 was correctly inserted, followed by the M13-Int primer to confirm the presence of the mutated residues. Both constructs, named pCMV-UWC-RBBP6-L29P-I33P and pCMV-UWC-RBBP6-V70P-V72P respectively, were found to be correct. Unfortunately time did not permit investigation of the localisations of these mutants using immunofluorescence microscopy.

4 CONCLUSIONS AND OUTLOOK

In this study immunofluorescence microscopy was used to characterize the role played by the DWNN domain in localising HA-RBBP6 isoforms 1 and 3 into the nucleus of mammalian cells and, more specifically, into nuclear speckles. In an attempt to achieve this goal, a number of regions of RBBP6, both wild type and mutated, were exogenously expressed in mammalian cells and their localisations investigated using antibodies detecting the HA-immuno-tag fused to the N-termini of the proteins (see Table 3 below for details of the HA-constructs). As in the case of endogenous RBBP6, HA-isoform 1 was found to be exclusively nuclear and confined to nuclear speckles, showing that the presence of the HA-tag does not change the localisation of the protein. The identity of the speckles was verified beyond doubt by showing that they co-localise with SC35, a splicing factor conventionally used as a marker of nuclear speckles. The presence of RBBP6, a protein that has recently been shown to play a central role in mRNA processing, in bodies more closely associated with mRNA splicing is intriguing and warrants further investigation.

An outstanding feature of RBBP6 is the ubiquitin-like DWNN domain found at the N-terminus of all isoforms. The results presented here show that the DWNN domain is required for nuclear localisation, since none of the proteins used in this study localise to the nucleus without the presence of the DWNN domain. Mutants designed to unfold the domain show that the DWNN domain must be folded in order to act as a nuclear targeting signal.

However the DWNN domain is not suffcent for complete nuclear localisation, since HA-R3, which contains residues 1-335 as well as the DWNN domain, is found in the cytoplasm as well as the nucleus. Furthermore isoform 3, which contains the DWNN domain followed by a 40-residue C-terminal tail, is predominantly cytoplasmic, although there is also evidence that it accumulates in nuclear speckles. Studies using endogenous RBBP6 provide strong evidence that the accumulation in nuclear speckles increases following a range of cellular stresses, including heat shock, UV-radiation and treatment with cisplatin (AV Szmyd-Potapczuk, UWC PhD thesis, 2017). Our results provide some evidence for nuclear speckling

73

of HA-isoform 3 (here called HA-DWNN13) and its induction by heat shock, but more data will be required to establish this beyond doubt.

An extremely lysine-rich region at the extreme C-terminus of isoform 1 was identified as a putative nuclear localisation signal when the protein was first reported. By explicitly removing this region we have shown that the putative NLS in fact plays a minimal role in nuclear localisation.

One of the aims of the work was to attempt to locate the region(s) of the protein responsible for targeting it to nuclear speckes. The presence of an arginine-serine-rich domain (RS domain) in RBBP6 suggests that this may be responsibe for targeting the protein to nuclear speckes. Although not the only domain reported to target protens to nuclear speckles, RS domains have been shown to be required for speckle-localisation in a number of mRNA splicing factors (*Li and Bingham, 1991*). In order to test this hypothesis the RS domain was excised from RBBP6; suprisingly this did not prevent the protein from localising strongly in nuclear speckles. In addition, in the absence of the RS domain the protein was more cytoplasmic than wild type isoform 1, suggesting that the RS domain plays a weak role in nuclear localisation.

Although our results have led us closer to an understanding of the reasons for the localisation of isoform 1, they have not led us closer to understanding the mechanism of targeting of isoform 1 to muclear speckles. Since isoform 3 also exhibits nuclear speckling and is much smaller than isoform 1, it may simplify the problem of narrowing down the region of interest. Isoforms 3 also exhibits predominant cytoplasmic localisation, despite the presence of the DWNN domain which we have already established acts as a nuclear localisation factor. A possible explanation is that a cytoplasmic localisation sequence (also called a nuclear export signal [NES]) is present in the C-terminal tail of isoform 3. The last 17 amino acids of isoform 3 are found only in this isoform. The same residues contain a leucine-rich sequence predicted by two online servers to function as a NES. However when those residues were removed the resulting protein remained predominantly cytoplasmic, similar to wild type isoform 3 (DWNN13), which suggested that the last 17 residues were not responsisble for targeting isoform 3 to the cytoplasm. However our data provide

74

evidence that the nuclear speckling exhibited by isoform 3 may no longer be present. If so, it would suggest that C-terminus plays the role of the elusive nuclear speckling sequence. However this possibility needs to be investigated further before it can be stated with certainty.

If the nuclear export signal is not in the last 17 residues of isoform 3, it is presumably somewhere within residues 80 – 101. As exected, truncation immediately following the Gly-Gly motif at the C-terminus of the DWNN domain (between residues 79 and 80) yielded a different phenotype: the protein was either uniformly distributed between the nucleus and the cytoplasm, or else more intensely in the nucleus. In both cases the nucleus distribution was very uniform, with no evidence of speckling. The same phenotype was observed when the protein was truncated immediately before the Gly-Gly motif (between residues 77 and 78). This observation supports a model in which the DWNN domain, by itself, acts as a nuclear localisation signal.

Table 3. RBBP6 deletion constructs, amino acid residues and the cellular localization of the resultant proteins.

pCMV-UWC Construct	Amino Acid Residues	Cellular Localisation
RBBP6 Isoform 1	1792	Completely nuclear – speckling
RBBP6-D60A-L63A	1792	Completely nuclear – speckling
RBBP6-ΔNLS	1730	Predominantly nuclear, poorly defined speckles with a weak cytoplasmic diffusion
RBBP6-ΔRS	1698	Predominantly nuclear, well defined speckles with cytoplasmic diffusion
RBBP6-ΔDWNN	1650	Completely cytoplasmic
R3	335	Evenly distributed across the nucleus and cytoplasm
R3-D60A-L63A	335	Evenly distributed across the nucleus and cytoplasm
R3-G78A-G79A	335	Evenly distributed across the nucleus and cytoplasm
R3-L29P-I33P	335	Predominantly cytoplasmic
R3-V70P-V72P	335	Predominantly cytoplasmic
R3-ΔDWNN	193	Completely cytoplasmic
RBBP6 Isoform 3/DWNN1	118	Predominantly cytoplasmic – speckling
DWNN13-ΔNES	100	Predominantly cytoplasmic – no speckling
DWNN-GG	78	Evenly distributed across the nucleus and cytoplasm
DWNN-PI	76	Evenly distributed across the nucleus and cytoplasm

4.1 *Future work*

Time did not permit the number of small investigations to be included in this thesis. These include transfer of the unfolding mutations L29P-I33P and V70P-V72P from R3 into isoform 1. Since isoform 1 is more strongly nuclear than R3, we expect the effects of unfolding of the DWNN domain on isoform 1 to provide more convincing evidence of the nuclear-targeting role of the DWNN domain than was provided by R3.

Also uncompleted was the investigation of the localisation of the HA-DWNN-TY construct. If this exhibits a localisation similar to that of wild type isoform 3 (HA-DWNN13), it will provide strong evidence that the short region spanning residue 80-85 is the driver of cytoplasmic localisation. If so, considering the high density of phosphorylatable residues, this would suggest a phosphorylation-dependent mechanism for regulation of the localisation of isoform 3. Combined with previous reports that the nuclear localisation of isoform 3 is induced by stress, this would be of great interest, especially in the context of the model proposed by Di Giammartino and co-workers (2014), that isoform 3 competitively regulates the polyadenylation activity of isoform 1.
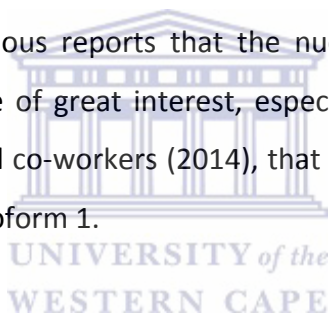
An important direction for future work will involve identification of surface residues involved in nuclear localisation of the DWNN domain, and hence isoform 1. In this work we have shown that the DWNN needs to be folded in order to target RBBP6 to the nucleus. This implies that an unidentified protein binds somewhere on the surface of the DWNN domain. The exact locations should not be difficult to identify by mutagenesis of surface residues, which can be identified on the basis of the published 3-dimensional structure of the domain.

One candidate for the unidentified protein is CstF-64, a core component of the 3'-polyadenylation machinery. Di Gammartino and co-workers (2014) have shown that the DWNN domain binds to CstF-64, thereby modulating the 3'-polyadenylation of mRNA transcripts. The first objective would be to determine whether mutations that abolish nuclear localisation of RBBP6 interfere with the binding of the DWNN to CstF-64 *in vitro*. The effect of such mutations on mRNA polyadenylation would also be of great interest. Such investigations would involve over-expression of exogenous wild type and mutant RBBP6 and

comparison of their effects on polyadenylation. The effect on mRNA splicing may also be worth investigating, considering the number of factors linking RBBP6 to splicing as well as 3'-end processing.

Future work should also focus on understanding the mechanism and significance of the accumulation of RBBP6 in nuclear speckles. There is disagreement in the literature as to whether nuclear speckles are sites of active mRNA processing or whether they represent storage locations for mRNA processing factors when not in use. The apparent exclusive localisation of isoform 1 in nuclear speckles suggests that it carries out its function within nuclear speckles. The role of the tail of isoform 3 in targeting it to nuclear speckles will also be worth investigating.

REFERENCES

1. Azuma, Y., Tan, SH., Cavenagh, MM., Ainsztein, AM., Saitoh H., Dasso, M. (2001), Expression and regulation of the mammalian SUMO-1 E1 enzyme, FASEB J, 15: 1825–1827

2. Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R., Becker, J. (1998), Structure determination of the small ubiquitin-related modifier SUMO-1, J Mol. Biol. 280: 275–286

3. Beck, J, S. (1961), Variations in the morphological patterns of "autoimmune" nuclear fluorescence, Lancet 1: 1203– 1205

4. Bernier-Villamor, V., Sampson, DA., Matunis, MJ., Lima, CD. (2002), Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 andRanGAP1, Cell 108: 345–356

5. Bogerd, HP., Fridell, RA., Benson, RE., Hua, J., Cullen, BR. (1996), Protein sequence requirements for function of the human T. cell leukemia virus type I Rex nuclear export signal delineated by a novel in vivo randomization-selection assay, Mol. Cell. Biol. 16: 4207-4214

6. Bohren, KM., Nadkarni, V., Song, KH., HGabbay, J., Owerbach, D. (2004), A M55V polymorphism in a novel SUMO gene (SUMO-4) differentially activates heat shock transcription factors and is associated with susceptibility to type I diabetes mellitus, J Biol. Chem. 29

7. Burch, TJ and Haas, AL. (1994), Site-directed mutagenesis of ubiquitin, Differential roles for arginine in the interaction with ubiquitin-activating enzyme, Biochemistry 33: 7300–7308

8. Bylebyl, GR., Belichenko, I., Johnson, ES. (2003), The SUMO isopeptidase Ulp2 prevents accumulation of SUMO chains in yeast, J Biol. Chem. 278: 44112–44120

9. Caceres, JF., Misteli, T., Screaton, GR., Spector, DL., Krainer, AR. (1997), Role of the modular domains of SR proteins in sub- nuclear localization and alternative splicing specificity, J Cell Biol. 138: 225–238

10. Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, AR., Caceres, JF. (2002), Nuclear export and retention signals in the RS domain of SR proteins, Mol. Cell Biol. 22:6871–6882, [PubMed: 12215544]

11. Chen, J et al. (2013), Overexpression of RBBP6, alone or combined with mutant TP53, is predictive of poor prognosis in colon cancer, PLoS ONE 8(6):e66524

12. Cheng, SY and Bishop JM. (2002), Suppressor of fused represses Gli-mediated transcription by recruiting the SAP18- mSin3 corepressor complex, Proc. Natl. Acad. Sci. USA. 99: 5442-5447

13. Chibi, M., Meyer, M., Skep, A., Rees, DJG., Moolman-Smook, JC., Pugh, DJR. (2008), RBBP6 interacts with multifunctional protein YB-1 through its RING finger domain, leading to ubiquitination and proteosomal degradation of YB-1, J Mol. Biol. vol 384: 908-16

14. Choudhury, BK and Li, SS. (1997), Identification and characterization of the SMT3 cDNA and gene from nematode *Caenorhabitis elegans*, Biochem. Biophy. Res. Commun. 234: 788-791

15. Cope, GA., Suh, GSB., Aravind, L., Schwarz, SE., Zipursky, SL., Koonin, EV., Deshaies, RJ. (2002), Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of NEDD-8 from Cul1, Science 298: 608–611

16. Delano WL., "The PyMOL Molecular Graphics System," Delano Scientific LLC, San Carlos, CA, USA, http://www,pymol,org)

17. Desterro, JM., Rodriguez, MS., Hay, RT. (1998), SUMO-1 Modification of IκBα Inhibits NF-κB Activation, Mol. Cell 2: 233–239

18. Di Giammartino, DC., Li, W., Ogami, K., Yashinskie, JJ., Hoque, M., Tian, B and Manley JL. (2014), RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3'UTRs, GENES & DEVELOPMENT 28: 2248-2260

19. Dingwall, C and Laskey, RA. (1991), Nuclear targeting sequences-a consensus?, Trends Biochem. Sci. 16: 478-481

20. Dohmen, RJ. (2004), SUMO protein modification, Elsevier: Biochimica et Biophysica Acta 1695: 113–131

21. Dohmen, RJ., Stappen, R., McGrath, JP., Forrova, H., Kolarov, J., Goffeau, A., Varshavsky, A. (1995), An essential yeast gene encoding a homolog of ubiquitin-activating enzyme, J Biol. Chem. 270: 18099–18109

22. Dostie, J., Lejbkowicz, F., Sonenberg, N. (2000), Nuclear eukary- otic initiation factor 4E (eIF4E) colocalizes with splicing factors in speckles, J Cell Biol. 148: 239–247

23. Duprez, E., Saurin, AJ., Desterro, JM., Lallemand-Breitenbach, V., Howe, K., Boddy, MN., Solomon, E., de The, H., Hay, RT., Freemont, PS. (1999), SUMO-1 modification of the acute promyelocytic leukaemia protein PML: implications for nuclear localisation, J Cell Sci. 112: 381–393

24. Eilbracht, J and Schmidt-Zachmann, MS. (2001), Identification of a sequence element directing a protein to nuclear speckles, Proc. Natl. Acad. Sci. 98: 3849–3854

25. Espinas, ML., Canudas, S., Fanti, L., Pimpinelli, S., Casanova, J., Azorin, F. (2000), The GAGA factor of Drosophila interacts with SAP18, a Sin3-associated polypeptide, EMBO Rep, 1: 253- 259

26. Fischer, U., Huber, J., Boelens, WC., Mattaj, IW., Luhrmann, R. (1995), The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs, Cell 82:475-483

27. Fontes, MR., Teh, T., Kobe, B. (2000), Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha, J. Mol. Biol. 297: 1183-1194

28. Fornerod, M and Ohno, M. (2002), Exportin-mediated nuclear export of proteins and ribonucleoproteins, Results Probl. cell Differ. 35: 67-91

29. Fox, AH and Lamond, AI. (2010), Paraspeckles, Cold Spring Harb. Perspect Biol. doi:10,1101/cshperspect,a000687

30. Fox, AH., Lam, YW., Leung, AK., Lyon, CE., Andersen, J., Mann, M., Lamond, AI. (2002), Paraspeckles: A novel nuclear domain, Curr. Biol. 12: 13–25

31. Gall, JG., Bellini, M., Wu, Z., Murphy, C. (1999), Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes, Mol. Biol. Cell 10: 4385–4402

32. Gao, S., Witte, MM., Scott, RE. (2002), P2P-R protein localizes to the nucleolus of interphase cells and the periphery of chromosomes in mitotic cells, which show maximum P2P-R immunoreactivity, J Cell Physiol. Vol. 191: 45-154

33. Gillam, S., Astell, CR., Smith, M. (1980), Site-specific mutagenesis using oligodeoxyribonucleotides: isolation of a phenotypically silent ФX174 mutant, with a specific nucleotide deletion, at very high efficiency, *Gene* 12: 129–37

34. Gong, L and Yeh, ETH. (1999), Identification of the activating and conjugating enzymes of the NEDD-8 conjugation pathway, J Biol. Chem. 274: 12 036–12 042

35. Gostissa, M., Hengstermann, A., Fogal, V., Sandy, P., Schwarz, SE., Scheffiner, M., Del Sal, G. (1999), Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1, EMBO Journal 18: 6462-6471

36. Hanania, U., Furman-Matarasso, N., Ron, M., Avni, A. (1998), Isolation of a novel SUMO protein from tomato that suppresses EIX-induced cell death, Plant J. 19: 533–541

37. Hedley, ML., Amrein, H., Maniatis, T. (1995), An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine rich splicing factor, Proc. Natl. Acad. Sci. 92: 11524–11528

38. Higuchi, R., Krummel, B., Saiki, RK. (1988), A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions, Nucl. Acids Res. 16: 7351–67

39. Hjerpe, R., Thomas, Y., Chen, J., Zemla, A., Curran, S., Shpiro, N., Dick, LR., Kurz, T. (2012), Changes in the ratio of free NEDD-8 to ubiquitin triggers NEDDylation by ubiquitin enzymes, Biochem. J. 441: 927–936

40. Hodel, MR., Corbett, AH., Hodel, AE. (2001), Dissection of a nuclear localization signal, J Biol. Chem. 267: 1317-1325

41. Huang, DT et al. (2009), E2-RING expansion of the NEDD-8 cascade confers specificity to cullin modification, Mol. Cell 33: 483–495

42. Huang, S and Spector, DL. (1991), Nascent pre-mRNA transcripts are associated with nuclear regions enriched in splicing factors, Genes Dev. 5:2288– 2302

43. Huang, S and Spector, DL. (1996), Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription, J Cell. Biol. 133(4):719-32

44. Jagatheesan, G., Thanumalayan, S., Muralikrishna, B., Rangaraj, N., Karande, AA., Parnaik, VK. (1999), Colocalization of intranuclear lamin foci with RNA splicing factors, J Cell Sci. 112: 4651–4661

45. Jagiello, I., Van Eynde, A., Vulsteke, V., Beullens, M., Boudrez, A., Keppens, S., Stalmans, W., Bollen, M. (2000), Nuclear and subnuclear targeting sequences of the protein phospha- tase-1 regulator NIPP1, J Cell Sci. 21: 3761–3768

46. Jiménez-García, LF and Spector, DL. (1993), In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism, Cell, 73: 47–59

47. Johnson, ES., Schwienhorst, I., Dohmen, RJ., Blobel, G. (1997), The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer, EMBO J. 16: 5509–5519

48. Johnson, PR and Hochstrasser, M. (1997), SUMO-1: ubiquitin gains weight, Trends in Cell Biology 7: 408-413

49. Kamitani, T., Kito, K., Nguyen, HP., Yeh, ETH. (1997), Characterization of NEDD-8, a developmentallydown-regulated ubiquitin-like protein, J Biol. Chem. 272: 28 557–28 562

50. Kamura, T., Conrad, MN., Yan, Q., Conaway, RC., Conaway, JW. (1999), The Rbx1 subunit of SCF and VHL E3 ubiquitin ligase activates Rub1 modification of cullins Cdc53 and Cul2, Genes Dev. 13: 2928–2933

51. Kim, AY et al. (2008), SCCRO (DCUN1D1) is an essential component of the E3 complex for neddylation, J Biol. Chem. 283: 33 211–33 220

52. Kim, KI., Baek, SH., Chung, CH. (2002), Versatile protein tag, SUMO: its enzymology and biological function, Journal of Cellular Physiology 191: 257-268

53. Kosugi, S., Hasebe, M., Tomita, M., Yanagama, H. (2008), Nuclear export signal consensus sequences defined using a localization-based yeast selection system, Traffic, 9: 2053-2062

54. Kramer, B., Kramer, W., Fritz, HJ. (1984), Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E, coli*. Cell, 38: 879, 1984

55. Krause, S., Fakan, S., Weis, K., Wahle, E. (1994), Immunodetection ofPoly(A) BindingProtein II in theCellNucleus, Exp. Cell Res. 214: 75–82

56. Kumar, S., Tomooka, Y., Noda, M. (1992), Identification of a set of genes with developmentally down- regulated expression in the mouse brain, Biochem. Biophys. Res. 185: 1155–1161

57. Kunkel, TA. (1985), Rapid and efficient site-specific mutagenesis without phenotypic selection, Proc. Natl. Acad. Sci. USA: Vol. 82: 488-492, January 1985 Genetics

58. Kurz, T., Chou, YC., Willems, AR., Meyer-Schaller, N., Hecht, ML., Tyers, M., Peter, M., Sicheri, F. (2008), Dcn1 functions as a scaffold-type E3 ligase for cullin neddylation, Mol. Cell 29: 23–35

59. Kurz, T., Ozlu, N., Rudolf, F., O'Rourke, SM., Luke, B., Hofmann, K., Hyman, AA., Bowerman, B., Peter, M. (2005), The conserved protein DCN-1/Dcn1p is required for cullin neddylation in *C. elegans* and *S. cerevisiae*, Nature 435: 1257–1261

60. La Cour, T., Gupta, R., Rapacki, K., Poulsen, FM., Brunak, S. (2003), NESbase version 1.0: a database of nuclear exports signals, Nucleic Acids Res. 31: 393-396

61. La Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K., Brunak, S. (2004), Analysis and prediction of leucine-rich nuclear export signals, Protein Eng. Des swl. 17: 527-536

62. Lang, A., Mills, RE., Lange, CJ., Stewart, M., Devine, SE., Corbette, AH. (2007), Classical nuclear localization signals: definition, function, and interaction with importin α, J Biol. Chem. 282 (8): 5101-5105

63. Larsson, SH., Charlieu, JP., Miyagawa, K., Engelkamp, D., Ras-soulzadegan, M., Ross, A., Cuzin, F., van Heyningen, V., Hastie, ND. (1995), Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing, Cell 81: 391–401

64. Lee, SD and Moore, CL. (2014), Efficient mRNA polyadenylation requires a ubiquitin-like domain, a zinc knuckle, and a RING finger domain, all contained in the Mpe1 protein, Mol. Cell. Bio.

65. Leidecker, O., Matic, I., Mahata, B., Pion, E., Xirodimas, DP. (2012), The ubiquitin E1 enzyme Ube1 mediates NEDD-8 activation under diverse stress conditions, Cell Cycle 11: 1142–1150

66. Li, H and Bingham, PM. (1991), Arginine/Serine-rich domains of the su(wa) and tra RNA processing regulators target proteins to a subnuclear compartment implicated in splicing, Cell 67: 335–342

67. Li, L et al. (2007), PACT is a negative regulator of p53 and essential for cell growth and embryonic development, Proc. Natl. Acad. Sci. USA, vol. 104: 7951-7956

68. Li, Q., Imataka, H., Morino, S., Rogers, GW., Jr, Richter-Cook, NJ., Merrick, WC., Sonenberg, N. (1999), Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII, Mol. Cell Biol. 19: 7336–7346

69. Li, SJ and Hochstrasser, M. (1999), A new protease required for cell-cycle progression in yeast, Nature 398: 246–251

70. Li, SJ and Hochstrasser, M. (2000), The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein, Mol. Cell Biol. 20: 2367–2377

71. Liakopoulos, D., Doenges, G., Matuschewski, K., Jentsch, S. (1998), A novel protein modification pathway related to the ubiquitin system, EMBO J. 17: 2208–2214

72. Liu, H and Naismith, JH. (2008), An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotech. 8:91

73. Lydeard, JR., Schulman, BA., Harper, JW. (2013), Building and remodelling Cullin-RING E3 ubiquitin ligases, EMBO Rep. 14: 1050–1061

74. Ma, T., Chen, Y., Zhang, F., Yang, C., Wang, S., Yu, X. (2013), RNF111-dependent neddylation activates DNA damage-induced ubiquitination, Mol. Cell 49: 897–907

75. Mahajan, R., Delphin, C., Guan, T., Gerace, L., Melchior, F. (1997), A small ubiquitin-related polypeptide involved in targeting RanGap1 to nuclear pore complex RanBP2, Cell 88: 97-101

76. Mbita, Z., Meyer, M., Skepu, A., Hosie, M., Rees, J., Dlamini, Z. (2012), De-regulation of the RBBP6 isoform 3/DWNN in human cancers, Molecular and Cellular Biochemistry 362: 249–262

77. McCallum, SA., Bazan, JF., Merchant, M., Yin, J., Pan, B., de Sauvage, FJ., Fairbrother, WJ. (2006), Structure of SAP18: A ubiquitin fold in histone deacetylase complex assembly, Biochemistry 45: 11974– 11982

78. Melchior, F. (2000), SUMO-nonclassical ubiquitin, Annu. Rev. Cell Dev. Biol. 16: 591–626

79. Melchior, F., Schergaut, M., Pichler, A. (2003), SUMO: ligases, iso- peptidases and nuclear pores, Trends Biochem. Sci. 28: 612–618

80. Mendoza, HM., Shen, LN., Botting, C., Lewis, A., Chen, J., Ink, B., Hay, RT. (2003), NEDP1, a highly conserved cysteine protease that deNEDDylates cullins, J Biol. Chem. 278: 25 637–25 643

81. Meyer-Schaller, N et al. (2009), The human Dcn1-like protein DCNL3 promotes Cul3 neddylation at membranes, Proc. Natl. Acad. Sci. USA 106, 12: 365–12 370

82. Miotto, B et al. (2014), The RBBP6/ZBTB38/MCM10 axis regulates DNA replication and common fragile site stability, Cell Reports 7: 575–587

83. Morisaki, T et al. (2014), Comparative proteomics analysis of gastro cancer stem cells, PLoS ONE 9(11): e110736

84. Motadi, LR., Bhoolab, KD., Dlamini, Z. (2011), Expression and function of retinoblastoma binding protein 6 (RBBP6) in human lung cancer, ELSEVIER, Immunobiology 216: 1065–1073

85. Muller, S., Berger, M., Lehembre, F., Seeler, JS., Haupt, Y., Dejean, A. (2000), c-Jun and p53 activity is modulated by SUMO-1 modification, Journal of Biological Chemistry 275: 13321-13329

86. Muller, S., Hoege, C., Pyrowolakis, G., Jentsch, S. (2001), SUMO, ubiquitin's mysterious cousin, Nature Reviews Molecular Cell Biology 2: 202-210

87. Nayler, O., Hartmann, AM., Stamm, S. (2000), The ER repeat protein YT521-B localizes to a novel subnuclear compartment, J Cell Biol. 150: 949–962

88. Nie, L et al., (2007), Regulation of p53 nuclear export through sequential changes in conformation and ubiquitination, The Journal of biological chemistry 282: 14616-14625

89. O'Keefe, RT., Mayeda, A., Sadowski, CL., Krainer, AR., Spector, DL. (1994), Disruption of pre-mRNA splicing in vivo results in reorganization of splicing factors, J Cell Biol. 124: 249–260

90. Okuma, T., Honda, R., Ichikawa, G., Tsumagari, N., Yasuda, H. (1999), In vitro SUMO-1 modification requires two enzymatic steps, E1 and E2, Biochemical and Biophysical Research Communications 254: 693-698

91. Osaka, F., Kawasaki, H., Aida, N., Saeki, M., Chiba, T., Kawashima, S., Tanaka, K., Kato, S. (1998), A new NEDD-8-ligating system for cullin-4A, Genes Dev. 12: 2263–2268

92. Petroski, MD and Deshaies, RJ. (2005), Function and regulation of cullin-RING ubiquitin ligases, Nat. Rev. Mol. Cell Biol. 6: 9–20

93. Potashkin, JA., Derby, RJ., Spector, DL. (1990), Differential dis- tribution of factors involved in pre-mRNA processing in the yeast cell nucleus, Mol. Cell Biochem. 10: 3524–3534

94. Pretorious, A., Kaur, M., Wamalwaz, M., February, MF., Essack, M, Bajic, VB., and Rees, JG. (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

95. Primrose, SB and Twyman, RM. (2006), Principles of gene manipulation and genomics (7th edition), American journal of human genetics, Chapter 8: 141 – 147

96. Pugh, DJR., Ab, E., Faro, A., Lutya, PT., Hoffmann, E., Rees, DJG. (2006), DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways, BMC Structural Biology, 6:1

97. Rabut, G and Peter, M. (2008), Function and regulation of protein neddylation, 'Protein modifications: beyond the usual suspects' review series, EMBO Rep. 9: 969–976

98. Ramo´n y Cajal, S. (1910), El nucleo de las celulaspiramidales del cerebrohumano y de algunosmamiferos, Trab Lab Invest Biol. 8: 27–62

99. Rodriguez, MS., Desterro, JM., Lain, S., Midgeley, CA., Lane, DP., Hay, RT. (1999), SUMO-1 modification activates the transcriptional response of p53, EMBO Journal 18: 6455-6461

100. Saitoh, H and Hinchey, J. (2000), Functional heterogeneity of small ubiquitin- related protein modifiers SUMO-1 versus SUMO-2/3, J Biol. Chem. 275: 6252–6258

101. Saitoh, H., Sparrow, DB, Shiomi, T., Pu, RT., Nishimoto, T., Mohun, TJ., Dasso, M. (1998), Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2, Curr. Biol. 8: 121–124

102. Sakai, Y., Saijo, M., Coelho, K., Kishino, T., Niikawa, N., and Taya, Y. (1995), cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product, Genomics, vol. 30: 98-101

103. Sarikas, A., Hartmann, T., Pan, Z, Q. (2011), The cullin protein family, Genome Biol. 12: 220

104. Sayers, J, R and Eckstein, F. (1991), A single-strand specific endonuclease activity copurifies with overexpressed T5 D15 exonuclease, Nucleic Acids Res. 19 (15): 4127-4132

105. Schwienhorst, I., Johnson, ES., Dohmen, RJ. (2000), SUMO conjugation and deconjugation, Molecular Gene and Genetics 263: 771-786

106. Scott, DC., Monda, JK., Grace, CRR., Duda, DM., Kriwacki, RW., Kurz, T., Schulman, BA. (2010), A dual E3 mechanism for Rub1 ligation to Cdc53, Mol, Cell 39: 784–796

107. Scott, RE., Giannakouros, T., Gao, S., Peidis, P. (2003), Functional potential of P2P-R: a role in the cell cycle and cell differentiation related to its interactions with proteins that bind to matrix associated regions of DNA, J Cell Biochem. Vol. 90: 6-12

108. Segalat, L and Lepesant, JA. (1992), Spatial distribution of the Sm antigen in Drosophila early embryos, Biol. Cell 75: 181– 185

109. Sharma, A., Takata, H., Shibahara, K., Bubulya, A., Bubulya, PA. (2010), Son is essential for nuclear speckle organization and cell cycle progression, Mol. Biol. Cell 21: 650–663

110. Simons, A., Melamed-Bessudo, C., Wolkowicz, R., Sperling, J., Sperling, R., Eisenbach, L., and Rotter, V. (1997), PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb, Oncogene, vol. 14: 145-155

111. Singh, N et al. (2010), Human SAP18 mediates assembly of a splicing regulatory multiprotein complex via its ubiquitin-like fold, RNA 2010 16: 2442-2454

112. Soucy, TA et al. (2009), An inhibitor of NEDD-8- activating enzyme as a new approach to treat cancer, Nature 458: 732–736

113. Souphron, J., Waddell, MB., Paydar, A., Tokgo¨z-Gromley, Z., Roussel, MF., Schulman, BA. (2008), Structural dissection of a gating mechanism preventing misactivation of ubiquitin by NEDD-80s E1, Biochemistry 47: 8961–8969

114. Spector, DL and Lamond, AI. (2011), Nuclear Speckles, Cold Spring Harb. Perspect. Biol. 3: a000646

115. Spector, DL., O'Keefe, RT., Jime´nez-Garcı´a, LF. (1993), Dynamics of transcription and pre-mRNA splicing within the mammalian cell nucleus, Cold Spring Harb. Symp. Quant Biol. 58: 799–805

116. Spector, DL., Schrier, WH., Busch, H. (1983), Immunoelectron microscopic localization of snRNPs, Biol. Cell 49: 1–10

117. Sternsdorf, T., Jensen, K., Will, H. (1997), Evidence for covalent modification of the nuclear dot-associated proteins PML and Sp100 by PIC1/ SUMO-1, J Cell Biol. 139: 1621–1634

118. Sundqvist, A., Liu, G., Mirsaliotis, A., Xirodimas, D, P. (2009), Regulation of nucleolar signalling to p53 through NEDDylation of L11, EMBO Rep. 10: 1132–1139

119. Swift, H. (1959), Studies on nuclear fine structure, Brookhaven Symp. Biol. 12: 134–152

120. Tatham, MH., Jaffray, E., Vaughan, OA., Desterro, JM., Botting, CH., Naismith, JH., Hay, RT. (2001), Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9, J Biol. Chem. 276: 35368–35374

121. Traboni, C., Cortese, R., Cilibert, G., Cesarini, G. (1983), A general method to select M13 clones carrying base pair substitution mutants constructed *in vitro,* Nucl. Acids Res. 11: 4229–39

122. Van der veen, AG and Ploegh, HL. (2012), Ubiquitin-Like Proteins, Annu. Rev. Biochem. 81: 325-57

123. Wada, H., Kito, K., Caskey, LS., Yeh, ET., Kamitani, T. (1998), Cleavage of the C-terminus of NEDD-8 by UCH-L3, Biochem. Biophys. Res. Commun. 251: 688–692

124. Wallace, RB., Schold, M., Johnson, MJ., Dembek, P., Itakura, K. (1981), Oligonucleotide directed mutagenesis of the human β-globin gene: a general method for producing specific point mutations in cloned DNA, Nucl. Acids Res. 9: 3647–56

125. Wang, W., Qin, J., Voruganti, S., Zhou, J., Zhang, R. (2016), Polycomb Group (PcG) proteins and human cancers: Multifaceted functions and therapeutic implications, Med. Res. Rev. 35 (6): 1220-1267

126. Wansink, DG., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R., de Jong, L. (1993), Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus, J Cell Biol. 122: 283–293

127. Wei, N and Deng, XW. (1992), COP9: a new genetic locus involved in light-regulated development and gene expression in arabidopsis, Plant Cell 4: 1507–1518

128. Wen, W., Meinkoth, JL., Tasien, RY., Taylor, SS. (1995), Identification of a signal for rapid export of proteins from the nucleus, Cell 82: 463-473

129. Whitby, FG., Xia, G., Pickart, CM., Hill, CP. (1998), Crystal structure of the human ubiquitin-like protein NEDD-8 and interactions with ubiquitin pathway enzymes, J Biol. Chem. 273: 34 983–34 991

130. Witte, MM and Scott, RE. (1997), The proliferation potential protein-related (P2P-R) gene with domains encoding heterogeneous nuclear ribonucleoprotein association and Rb1 binding shows repressed expression during terminal differentiation, Proc. Natl. Acad. Sci. USA, vol. 94: 1212–1217

131. Xing, Y., Johnson, CV., Dobner, PR., Lawrence, JB. (1993), Higher level or- ganization of individual gene transcription and RNA splicing, Science (Wash, DC), 259: 1326–1330

132. Xing, Y., Johnson, CV., Moen, PT., McNeil, J., Lawrence, JB. (1995), Non- random gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC35-domains, J Cell Biol. 131: 1635–1647

133. Xirodimas, DP., Saville, MK., Bourdon, JC., Hay, RT., Lane, DP. (2004), Mdm2-mediated NEDD-8 conjugation of p53 inhibits its transcriptional activity, Cell 118: 83–97

134. Xu, D., Farmer, A., Collett, G., Grishin, NV., Chook, Y. (2012), Sequence and structural analysis of nuclear export signals in the NESdb database, MBoC. 23: 3677-3693

135. Yeh, ETH., Gong, L., Kamitani, T. (2000), Ubiquitin-like proteins: new wines in new bottles, Gene 248: 1-14

136. Yoshitake, Y., Nakatsura, T., Monji, M., Senju, S., Matsuyoshi, H., Tsukamoto, H. (2004), Proliferation potential-related protein, an ideal oesophageal cancer antigen for immunotherapy, identified using complementary DNA microarray analysis, Clin. Cancer Res. Vol. 10: 6437-6448

137. Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., Reinberg, D. (1997), Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex, Cell 89: 357-364

138. Zhu, W and Hanes, SD. (2000), Identification of Drosophila bicoid-interacting proteins using a custom two-hybrid selection, Gene 245: 329-339

139. Zoller, MJ and Smith, M. (1983), Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, Methods Enzy. mol. 100: 468–500

APPENDIX

A1. Strategy for synthesizing the RBBP6-ΔNLS fragment

The most C-terminal sequence of RBBP6 with BlpI cut site

agcagaagccacagtccttctggaagccagacccgaagccacagtagcagtgcca`gctca`

  S   R   S   H   S   P   S   G   S   Q   T   R   S   H   S   S   S   A   S   S

`gc`agaaagtcaggacagcaagaagaagaagaaaaagaaggaaaagaaaaaacacaagaaa

  A   E   S   Q   D   S   **K**   **K**   **K**   **K**   **K**   **K**   **E**   **K**   **K**   **K**   **H**   **K**   **K**

cataaaaagcataagaagcataagaaacatgcaggcactgaagtggaattggaaaaaagc

  H   **K**   **K**   **H**   **K**   **K**   **H**   **K**   **K**   **H**   **A**   **G**   **T**   **E**   **V**   **E**   **L**   **E**   **K**   **S**

caaaaacacaaacacaagaaaaagaagtcaaagaagaacaaagataaagagaaggagaag

  Q   **K**   **H**   **K**   **H**   **K**   **K**   **K**   **S**   **K**   **N**   **K**   **D**   **K**   **E**   **K**   **E**   **K**

gagaaagatgaccaaaaagtgaaatctgtcactgtg

  E   **K**   **D**   **D**   **Q**   **K**   **V**   **K**   S   V   T   V

- The sequence highlighted in cyan is the area to be removed on the final RBBP6-ΔNLS construct.

The desired sequence across the deletion, including the re-formed Bpu and KpnI sites

  Bpu1102I   KpnI

5'–GCTCAGCAGAAAGTCAGGACAGCTCTGTCACTGTGTGAGGTACC–3'

3'–CGAGTCGTCTTTCAGTCCTGTCGAGACAGTGACACACTCCATGG–5'

Synthesize the following 2 fragments, anneal and add to RBBP6 digest

5'–TCAGCAGAAAGTCAGGACAGCTCTGTCACTGTGTGAGGTAC–3'

5'–CTCACACAGTGACAGAGCTGTCCTGACTTTCTGC–3'

Final sequence:

```
tcagcagaaagtcaggacagctctgtcactgtgtgaggtac
```

`S  A  E  S  Q  D  S  S  V  T  V  –  G`

<u>Suppressing background</u>

The uncut construct is 9135 bp, compared to the BpuI/KpnI digested construct which will produce a band of 8908 bp. This would make separating the doubly cut from the uncut RBBP6 difficult due to the large sizes of these plasmids. This may lead to singly cut vector re-ligating when ligase added. This may lead to re-ligated parental transforming in large numbers. This background should be apparent in the vector-only control.

Following transformation, it will be difficult to distinguish between positives and negatives due to the similar sizes of the plasmids. To facilitate post-transformation screening, we synthesized our oligos in such a way that the "reconstituted" KpnI site was not reconstituted. We changed the base pair marked in red from G to C (it can be changed to anyone of the three bases), hence the sequence still retains the same 4-base sticky end, but the KpnI site is destroyed (CGTACC instead of GGTACC). Hence, if minipreps cut with KpnI, they are parentals, but if they do not cut they are possibly positives.

```
 Bpu1102I                                          KpnI

5'–GCTCAGCAGAAAGTCAGGACAGCTCTGTCACTGTGTGACGTACC–3'

3'–CGAGTCGTCTTTCAGTCCTGTCGAGACAGTGACACACTGCATGG–5'
```

A2. Strategy to synthesize additional mutations in the DWNN domain

The complementary regions (underlined - also called the overlapping regions) are at the 5'-ends of both primers and the overhanging parts are at the 3'-ends. The Tm's of the complete primers are in the region of 59 - 80 °C; the Tm of the overlapping regions range from 47 - 60 °C. So by setting the extension temperature at 5 °C less that the Tm of the total primers (i.e. at 54 - 75 °C), the primers will anneal to the template and not to each other.

Creating the L29P-I33P mutation

Wild type (Leu29 and Ile33)

```
Forward: 5'-TCCCTCTGCGACTTAAAGAAGCAGATTATGGGGAGAGAG-3' (Tm = 78 °C)
Reverse: 5'-CTTCTTTAAGTCGCAGAGGGAGATGTGGAGCCCATCAAA-3' (Tm = 81 °C)
Overlap region
Tm = 60 °C
```

Mutants (Replace Leu29 and Ile33 with Proline (CTT))

```
Forward: 5'-TCCCTCTGCGACCCTAAGAAGCAGCCTATGGGGAGAGAG-3' (Tm = 81 °C)
Reverse: 5'-CTTCTTAGGGTCGCAGAGGGAGATGTGGAGCCCATCAAA-3' (Tm = 81 °C)
Overlap region
Tm = 66 °C
```

Creating the V70P-V72P unfolding mutation

Wild type V70V72 primers

```
Forward: 5'-TCTTCTGTAATTGTTAGAAGAATTCCTATTGGAGGTGTTAAATCTACAAGC
Reverse: 5'-TCTTCTAACAATTACAGAAGAATTCTTAGGAATCAGAGCATTATCATCAGT

         ATATTCTTCTTTCG-3'
```

```
Tm(overlap) = 47 °C; 21 bp
Tm(3'-overhang_for) = 59 °C; 30 bp
Tm(3'-overhang_rev) = 62 °C; 44 bp
```

90

Mutant V70V72 primers (CCT codes for Pro)

Forward: 5'-TCTTCT<mark>CCT</mark>ATT<mark>CCT</mark>AGAAGAATTCCTATTGGAGGTGTTAAATCTACAAGC

Reverse: 5'-TCTTCT<mark>AGG</mark>AAT<mark>AGG</mark>AGAAGAATTCTTAGGAATCAGAGCATTATCATCAGT

ATATTCTTCTTTCG

Creating the G78A-G79A mutation

Wild type GG primers

Forward: 5'-ATTCCTATT<mark>GGAGGT</mark>GTTAAATCTACAAGCAAGACATATGTTATAAGT

CGAACTGAACC

Reverse: 5'-TTTAAC<mark>ACCTCC</mark>AATAGGAATTCTTCTAACAATTACAGAAGAATTCTT

AGGAATCAGAGC

Tm(overlap) = 51; 21 bp
Tm(3'-overhang_for) = 62; 38 bp
Tm(3'-overhand_rev) = 60; 39 bp

Mutant GG primers (GCT and GCC both code for Ala)

Forward: 5'-ATTCCTATT<mark>GCTGCC</mark>GTTAAATCTACAAGCAAGACATATGTTATAAGT

CGAACTGAACC

Reverse: 5'-TTTAAC<mark>GGCAGC</mark>AATAGGAATTCTTCTAACAATTACAGAAGAATTCTT

AGGAATCAGAGC

Tm(overlap) = 52; 21 bp
Tm(3'-overhang_for) = 62; 38 bp
Tm(3'-overhang_rev) = 60; 39 bp

Table 4: Chemicals, Kits and Suppliers

Chemicals	Suppliers

40 % 37:1 acrylamide: bis-acrylamide	Promega
Acetic acid	Merck
Agarose	Separation
Ammonium persulphate	Merck
Ampicillin	Roche
Antibodies	Santa Cruz
Bovine Serum Albumin	Roche
Dulbecco's Modified Eagle's Medium –F12	White Science (LONZA)
Ethanol	Merck
Ethylene Diamine Tetra Acetic acid (EDTA)	Merck
Formaldehyde	Merck
GeneJet Gel extraction kit	Thermo Scientific
GeneJet maxi-prep kit	Thermo Scientific
GeneJet plasmid mini-prep kit	Thermo Scientific
Hoechst	Life Technologies
Isopropanol	Merck
Luria Broth	Merck
MG132	Sigma
Mowiol	Sigma
Nutrient Agar	Merck
PCR Reagents	Thermo Scientific
Qiagen Maxi-prep kit	Qiagen
Restriction Enzymes	Thermo Scientific
Sodium Chloride	Merck
Sodium Dodecyl Sulfate	Merck
T4 DNA Ligase	Thermo Scientific
Tris [hydroxymethyl] aminomethane	BDH
Trypsine-EDTA	Life Technologies

Tween-20	Merck
TEMED	Merck
Tryptophan Blue	Invitrogen
X-tremeGENE HP DNA Transfection Reagent	Roche

Table 5: General Stock Solutions and recipes

Stock solutions	Recipe
10X PBS (1 L)	Dissolved 80 g NaCl, 2 g KCl, 14.4 g $Na_2HPO_4.2H_2O$ and 4.8 g KH_2PO_4 in 500 ml distilled water. Adjusted the pH to 7.4 using NaOH and stored the stock at room temperature. The stock was diluted to 1X for application in experiments.
10X Running Buffer	Dissolved 30.2 g Tris, 144.1 g Glycine, and 20 g SDS in distilled water to the required final volume. The stock was stored at room temperature and diluted to 1X concentration for running SDS-PAGE Gels.
10X Transfer Buffer (1 L)	Dissolved 29 g Glycine and 58 g Tris to the required final volume. The 10X stock was stored at room temperature, diluted ten-fold by adding 20% isopropanol and the required amount of distilled water to the working stock. The working stock was stored at -20 ℃.
2X SDS PAGE sample buffer (100ml)	Mixed 10 ml 1.5 M Tris (pH 6.8), 6 ml 20% SDS, 30 ml glycerol and 1mg/ml bromophenol blue. The volume was adjusted to the required volume with distilled water and stored at room temperature. 200 mM β-mercaptoethanol was added immediately before use.
50X TAE (1 L)	Dissolved 242 g Tris in 500 ml distilled water. Added 100ml 0.5 M EDTA (pH 8.0) and 57.1 ml Glacial acetic acid. The volume was adjusted to 1 liter and the stock was stored at room temperature. The stock was diluted to 1X for running agarose gels.
Ampicillin	Ampicillin was dissolved in distilled water to a final concentration of 100 mg/ml. The solution was filter sterilized and stored at -20 ℃.

MG132	MG132 was dissolved in DMSO to a final stock concentration of 50 mM. The stock was stored at -20 ℃.
Tfb1	Mixed 30 mM Potassium acetate, 50 mM MnCl2, 0.1 M KCl, 10 mM CaCl2 and 15 % glycerol. The volume was adjusted to the required final volume with distilled water. The buffer was filter sterilized and stored at 4 ℃ until required.
Tfb2	Mixed 9 mM Na-MOPS, 50 mM CaCl2, 10 mM KCL and 15 % glycerol. The volume was adjusted to the required final volume with distilled water. The buffer was filter sterilized and stored at 4 ℃ until required.
TYM broth	Dissolved 2 % peptone, 0.5 % yeast extract, 0.1 M NaCl, 0.2 % glucose and 10 mM MgCl2. The volume was adjusted to the required final volume with distilled water. The broth was autoclaved and stored at room temperature until required.