UNIVERSITY OF ZULULAND

Recombinant expression, purification and structural characterization of *Saccharomyces cerevisiae* RING finger domain using Nuclear Magnetic Resonance (NMR) spectroscopy

By

Makhosazana Siduduzile MATHENJWA

(2000951523)

BSc (Biochemistry and Microbiology); BSc Hons (Biochemistry)

Dissertation submitted in fulfilment of the requirement for the degree **MASTERS OF SCIENCE (MSc) IN BIOCHEMISTRY**

Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand, KwaDlangezwa, KwaZulu Natal

Supervisor: Prof Abidemi Paul Kappo

April 2018

Recombinant expression, purification and structural characterization of *Saccharomyces cerevisiae* RING finger domain using Nuclear Magnetic Resonance (NMR) spectroscopy

Makhosazana Siduduzile Mathenjwa

(200951523)

A dissertation submitted in fulfilment of the requirements for the degree of Master of Science in the Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, KwaDlangezwa, South Africa

Supervisor: **Prof Abidemi Paul Kappo**

April 2018

Abstract

RING finger domain is a domain that is incorporated with other domains to form a large protein known as RBBP6. This is a 250kDa protein that stands for Retinoblastoma binding domain protein 6. It is classified according to different homologues namely RBQ1, PACT, SNAMA, Mpe-1 and P2P-R depending on the organisms in which they are found. P2P-R (a RING finger containing protein) was found to be highly up-regulated in the oesophageal cells compared to normal epithelial cells. This was a pre-indication that P2P-R might serve as an antigen against cancer or be used in the diagnosis of cancer. *Saccharomyces cerevisiae* RING finger domain is a protein that is represented by the richness of cysteine sequence motif that binds to two zinc atoms. It has a cross-brace topology consisting of 70 amino acids and its sequence is defined as Cys1-Xaa2-Cys2-Xaa9–39-Cys3-Xaa1–3-His4-Xaa2-3-Cys/His5-Xaa2-Cys6-Xaa4–48-Cys7-Xaa2-Cys8 (where Xaa represents any amino acid residue). The RING finger domain is essential in proteins involved in a range of diverse cellular processes such as apoptosis, viral infections, oncogenesis and ubiquitination.

The study presented in this dissertation reports on the recombinant expression and partial purification *S. cerevisiae* RING finger protein, as well as its biophysical and biochemical characterization. Structural predictions using the ProtParam tool on ExPASy provided information regarding the high structural stability due to the abundance of leucine in the protein. The protein has a limited ability for light absorption which is indicative of the absence of two important aromatic amino acids; tryptophan and tyrosine. More so, Swiss-Model showed that the 3D homology model structure is composed of two α-helices contributing to the stability of the structure,

two β-sheets that prevents proteolysis and one 3_{10} helix that maintains the helix-coiling transitions. The local quality estimate revealed the structure is rigid and stable. The non-redundant set of the Protein Data Bank analysis showed the structural prediction of the RING finger protein is accurate to a larger extent. Validation of the predicted model by Ramachandran plot showed that 88% of the amino acids residues are in the favoured region meaning that the modelled structure is an acceptable representation of the *S. cerevisiae* RING finger protein.

More so, interacting partners for *S. cerevisiae* RING finger protein were also ascertained and these protein partners were shown to be involved in cleavage and polyadenylation, however, zinc fingers from YTH1 showed similarity with RING finger domain as promoters of DNA binding. Phylogenetics and sequence alignment showed the ancestral proteins are involved in ubiquitination, which is one of the major pathways of RING finger domains. Lastly, the biochemical characterization using a 1D proton NMR spectroscopy showed that *S. cerevisiae* RING finger protein is well folded and can be further worked upon. These results provide the baseline information for the structural determination of the *S. cerevisiae* RING finger domain for future studies.

Keywords: RING finger domain, RBBP6, ubiquitination, zinc topology, cancer

DECLARATION

I declare that ***"Recombinant expression, purification and structural characterization of S. cerevisiae RING finger domain using Nuclear Magnetic Resonances (NMR) spectroscopy"*** is my own research and has not been submitted for examination in any other university for degree. All the work written has been indicated or acknowledged by complete references.

Makhosazana Siduduzile Mathenjwa April 2018

Signed……

ACKNOWLEDGEMENTS

I would like to thank the engineer of this work, my Lord Almighty for being my pillar of strength, His protection and granting me this opportunity, my family BT Mathenjwa (Daddy), GZ Mathenjwa (Mummy), Sihle (Sister), Thando (Brother) and Khetha (Nephew) for supporting me in all sides of my life.

To Prof. Kappo my supervisor for seeing a potential in me and trusting me with this project. To Prof. Andy and Dr Mosa for the support they gave me towards the success of this work. My friends Ntombikhona “Appear” and Nathi “David” thank you so much for everything and for being my stress reliever.

I would also like to express my gratitude to my friends from church (TWCI) and University of Zululand for praying with me and sharing a smile with me. Lastly, I like to thank my colleagues; it has been great being with you guys until we meet again.

“*Isandla sedlula ikhanda”*

TABLE OF CONTENT

ABBREVIATIONS

Amp	Ampicillin
APS	Ammonium persulphate
^{113}Cd, Cd^{2+}	Cadmium ion
Cv	Column volume
Cys, C	Cysteine
Da	Dalton
DNA	Deoxyribonucleic acid
dH_2O	Deionised water
DTT	Dithiothreitol
DWNN	Domain With No Name
E_1	Ubiquitin-activating enzyme
E_2	Ubiquitin-conjugating enzyme
E_3	Ubiquitin protein ligase
EDTA	Ethylene diamine tetra acetic acid
EEA_1	Early endosome antigen1
Fab1p	Formation of aploid and binu
Fig	Figure
FTIR	Fourier Transformation Infrared spectroscopy
FYVE	Fab1p, YOTB, Vac1p, EEA1
GST	Glutathione S-transferase
Hdm2	Human double minute 2
HECT	Homologous to E_6-AP carboxyl terminus
His, H	Histidine

HSQC	Heteronuclear Single Quantum Coherence
Hr	Hour
Ile,I	Isoleucine
IPTG	Isopropyl β-D-thiogalactoside
kDa	Kilodalton
Leu, L	Leucine
LB	Luria Bertani broth
LIM	Lin11/ Isl-1/ Mec-3
Lys	Lysine
Mdm2	Mouse double minute 2
Met, M	Methionine
Min	Minutes
mRNA	Messenger RNA
Nm	Nanometer
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear overhauser enhancement spectroscopy
N-terminus	Amino terminus
p44	Interferon-induced protein 44
Pab1p	Poly(A)-binding protein 1
P2P-R	Proliferation potential protein-related
PACT	p53-associated cellular protein-testes
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffer saline
PINCH	Particularly Interesting New Cysteine-Histidine protein
PI(3)P	Phosphatidylinositol-3 phosphate

PHD	Plant homeodomain
Phe, F	Phenylanine
PMSF	Phenymethylsulphonyl fluoride
Ppm	parts per million
PRb	Retinoblastoma gene product
Rad6	Radiation protein 6
RBBP6	Retinoblastoma binding protein 6
RING	Really Interesting New Gene
RBQ-1	Retinoblastoma-binding Q-protein 1
S	Seconds
SDS	Sodium dodecyl sulphate
SR	Serine/ arginine
TEMED	N',N',N'-tetramethylethylenediamine
Tris	2-amino-2-hydroxymethyprone-1,3-diol
UFD2	Ubiquitin fusion degradation protein 2
UV	Ultraviolet
YB-1	Y-box binding protein 1
Zn^{2+}	Zinc ion

LIST OF TABLES

LIST OF FIGURES

LIST OF RESEARCH OUTPUTS

CONFERENCE PROCEEDINGS

Makhosazana Siduduzile Mathenjwa, David JR Pugh and Abidemi Paul Kappo. Recombinant expression, purification and structural characterisation of *Saccharomyces cerevisae* RING finger domain using heteronuclear Nuclear Magnetic Resonance (NMR) spectroscopy. South African Society of Biochemistry and Molecular biology (SASBMB) East London International Conference Centre, 10th July to the 14th July 2016.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Retinoblastoma-**b**inding **p**rotein 6 (RBBP6) is a multifunctional 250 kDa protein present in all eukaryotes except prokaryotes and is located on chromosome 16p12. In humans, RBBP6 is also known as RBQ-1 and is found in human carcinoma cells (Sakai *et al.*, 1995), PACT (p53 associated cellular protein) (Simons *et al.*, 1997) and P2P-R (proliferation potential related protein) (Scott and Gao, 2002); the latter two both originating from testis. In yeast, the protein is known as mpe-1 (Vo *et al.*, 2001) and in *Drosophila melanogaster*; it is termed SNAMA (Ntwasa, 2005). The RBQ-1 and PACT homologues were cloned based on their capability in binding to p53 and Rb as observed in both human and mouse cells. P2P-R was cloned after being identified by two antibodies that are specific for heterogeneous nuclear ribonucleoproteins (hnRNPs) and was further defined by its characteristic interaction in p53-Hdm2, thereby enhancing its degradation (Saijo *et al*., 1995; Yoshitake *et al.*, 2004; Zhou *et al.*, 1999).

The size of RBBP6 evolves from its multi-domain characteristic nature which consists of the Domain With No Name (DWNN), zinc finger domain, RING finger domain, Retinoblastoma (Rb)-binding domain, a serine/arginine-rich (SR) region and the p53-binding domain. Functionally, RBBP6 is known to be involved in mRNA metabolism (Shi *et al.*, 2009), cell cycle progression (Scott *et al*, 2003), apoptosis and tumorigenesis (Li *et al.*, 2007; Yoshitake *et al.*, 2004). Regulation of p53 is critically controlled to maintain normal cell growth and is regulated by ubiquitin-mediated Hdm2 which is also responsible for the maintenance of proteasome function i.e. the degradation of misfolded proteins

(Franklin, 2000; Terasaki *et al.*, 2005; Zitting *et al.*, 2002). RBBP6 is known for its interactive behaviour with p53 and Rb. This interaction has been seen to suppress interaction between p53 and DNA (Saijo *et al.*, 1995; Simons *et al.*, 1997; Zhou *et al.*, 1999) and also functions in the restriction of adenovirus E1A protein binding to Rb (Sakai, 1997). It has also been reported that stable over-expression of PACT limits the progression of mitosis at prometaphase and encourages camptothecin-induced apoptosis (Gao and Scott, 2002; Gao *et al.*, 2003). Li and co-workers (2007) showed that the interruption of PACT leads to premature death of embryos before embryonic day 7.5 (E7.5) followed by abnormally enhanced apoptosis. Hence, PACT reduces risk in premature embryonic lethality and prolongs life span to E11. 5.

Kappo and co-workers (2012) showed that the role played by RBBP6 is facilitated by its RING finger–like domain. RING (**R**eally **I**nteresting **N**ew **G**ene) finger domain is rich in cysteine with a few cysteine residues conserved within the protein structure (Aravind *et al.,* 2003). It also coordinates two zinc ions in a cross-brace manner; where one zinc ion is coordinated by the first and third Cys/His pairs and the other zinc ion is coordinated by the second and fourth Cys/His pairs. RING fingers has the ability to fold independently but also with the help of zinc ions. It is also considered a small domain due to its length of approximately 70 amino acids.

The classification of RING finger domain depends on the coordination pattern of the cysteine/histidine residues to zinc ions. The most common class of RING fingers is C3HC4, even though C3HHC3, C2H2C4, and C4C4 RING fingers have also been observed. The primary sequence of the RING finger domain found in RBBP6 was noticeably saturated with eight conserved cysteine residues. This led Kappo and co-workers (2012) to propose that this protein may belong to the C4C4 RING family with the two zinc ions adopting a characteristic pattern which can either be octahedral in aqueous solution (Bock *et al.*,

1999), or pentahedral as found in ‘catalytic’ zinc-sites; rarely adopting hexa-coordination. Lastly, it has been shown that the tetrahedral coordinated pattern is found in both “structural and catalytic” zinc-sites, which is also the most stable form as compared to other zinc polyhedral coordinations (Dudev *et al.*, 2003). Tetrahedral coordination is also the most prevalent coordination pattern found in RING domains (Dudev and Lim, 2003).

RING finger domain was originally discovered by Freemont and co-workers (1991) and was found to have the canonical consensus sequence: Cys-X_2-Cys-$X_{(9-39)}$-Cys-$X_{(1-3)}$-His-$X_{(2-3)}$-Cys-X_2-Cys-$X_{(4-48)}$-Cys-X_2-Cys, where X represent the any amino acids. A three-dimensional structure of the RING finger domain shows that the cysteine and histidine residues are conserved at the core of the domain, where they assist in overall structure maintenance by binding to two zinc ions. The other semi-conserved residues are either implicated in forming a hydrophobic core of the RING domain or as recruiting agencies of other proteins. Also, unlike the zinc finger domain, the RING finger domain binding sites coordination by zinc are interleaved, giving rise to a rigid and globular scaffold for protein-protein interactions (Barlow *et al.*, 1994; Borden *et al.*, 1995; Kappo *et al.*, 2012).

However, there exist proteins with similar zinc configuration as seen in RING domains but their structures are entirely different from that of the RING finger domain class. These proteins are termed RING finger-like domains and include the PHD and LIM domains which have the same cysteine/histidine residue coordination but a different folding pattern and these proteins are not involved in ubiquitination (Aravind *et al.*, 2003; Borden and Freemont, 1996; Scheel and Hofman, 2003). Other RING variants include B-box domains (Tao *et al.,* 2008), U-box domains (Aravid and Koonin, 2000) and SUMO (Vander Kooi *et al.*, 2006).

The major function of the RING finger domain is the recruitment of ubiquitin-conjugating enzyme E2 complex releasing ubiquitin to attach to the substrate by binding to E3 (substrate-binding domain). Almost all of the 300 RING proteins that are found in humans function as ligation enzymes. It has also been shown that the RING domain rarely possesses intrinsic E3 activity; however two RING domains join together to form a heterodimer that mediates E3 activity. Various functions of the RING finger domain include differentiation, multimerization and cell growth (Kappo *et al.*, 2012). It has also been implicated in other cellular processes such as development, oncogenesis, apoptosis and viral replication (Dominguez *et al.*, 2004).

A genome-wide bioinformatic analysis by Kappo and co-workers (2012), using a RING finger domain derived from human RBBP6 as a searching bait, showed that the zinc ion coordination pattern was the same in all organisms except in three: *Saccharomyces cerevisae* and *Pichia pastoris*, where the first zinc binding-site was missing; and in *Aspergillus niger* in which there was substitution of the second coordinating signature cysteine residue for aspartic acid. Therefore, the aim of this study was to recombinantly express and purify *Saccharomyces cerevisae* RING finger domain to homogeneity, thereafter characterize the protein biophysically as a basis for sequential assignment of the protein backbone using NMR. Resonance assignment is the first step in solving protein structure by NMR; hence, this study will then be used to determine the structural effect of the putative previously-mentioned loss of the zinc binding site on the *S. cerevisiae* RING domain protein as a step towards structural determination, as well function (stability) of the protein.

CHAPTER 2

LITERATURE REVIEW

2.1 LITERATURE REVIEW

Retinoblastoma binding protein 6 has different homologues which include RBQ-1, PACT, P2P-R, mpe-1 and SNAMA.

2.1.1 RBQ-1 (RB-binding Q-protein 1)

RBQ-1 is the RBBP6 homologue that was identified from the human small cell carcinoma (H69c) library as a new protein of 140kDa. This protein has been shown to bind to the Rb gene product and also has a high homologue to the p53 associated cellular protein testis derived (PACT) protein (Sakai *et al.*, 1995). This implies that RBQ-1 has a role in cell cycle regulation. Two other proteins have been identified to be related to RBQ-1; RBQ-2 and RBQ-3 are proteins that bind the under-phosphorylated form of pRb. Though all these proteins were discovered in the same study, RBQ-1 and RBQ-3 show no striking similarities (Saijo *et al.*, 1995). RQB-1 is a short protein composed of 948 amino acids compared to PACT derived from mice. However, the overall homology between the two proteins is 94%, which means that these proteins are highly conserved (Simons *et al.*, 1997). RBQ-1 is localised on chromosome 16p12.2 and has a sequence length of 3011bp with an open reading frame of 948 amino acids.

2.1.2 PACT (p53 Associated Cell Testis)

PACT is a 250kDa protein derived from mice and is also known as a p53-associated cellular testis-protein with a conserved RING finger domain located at the N-terminal. This protein was originally isolated using a p53 probe in an expression library of mouse testes (Simons *et al.*, 1997). The sequence analysis of PACT showed that the cDNA is composed of 1577bp coding for 1583 amino acids and contains a 437bp 3' non-coding region with a polyA tail. Witte and Scott (1997) showed that PACT interacts both *in vitro* and *in vivo* with p53 and Rb (Simons *et al.*, 1999). It has also been suggested to be the primary target of

immunotherapy because of its high up-regulation in oesophageal cancer (Yoshitake *et al.*, 2004).

Li and co-workers (2007) showed the interaction of endogenous PACT with Hd2m enhances ubiquitination and cellular degradation of p53 mediated by Hd2m as a result of the high affinity between p53 and Hd2m. PACT has been shown to suppress p53 but when endogenous PACT activity gets blocked, p53 accumulates *in vivo* causing a p53-dependent suppression both in apoptosis and cell growth. Hence, PACT is implicated in the functioning of p53 via negative regulation by proteolysis that is dependent on ubiquitin showing the risks displayed by maturation of PACT.

2.1.3 P2P-R (Proliferation Potential Protein-Related)

P2P-R cDNA full length was first cloned and characterised in the early 1990s (Witte and Scott, 1997). It was shown that P2P-R protein has a pI of >9.0 and is composed of a highly basic nuclear protein with a size of 35-40kDa. It binds RNA and associates with hnRNP (Heterogeneous Nuclear Ribonucleoproteins). Antibodies that are prepared against the hnRNP core have been shown to recognise the P2P proteins. These antibodies include FA12 monoclonal antibodies, which detect the core of hnRNPs and AC88 monoclonal antibodies, thus demonstrating the similarity between the epitope of Hsp90 and P2P-R (Witte and Scott, 1997). The other monoclonal antibody that is able to recognise P2P-R epitope is C130, which shows that P2P-R cDNA is composed of three distinct binding sites similar to what is seen in hnRNP core proteins. Witte and Scott (1997) further showed the levels of expression of P2P-R mRNA in multiple tissues of mice as being more expressed in the testis, moderate in the skeletal muscles, liver, lungs and heart as well as lower but detectable in the spleen, kidneys and brain.

Peidis and co-workers (2010), in search of the co-variant that had high co-expression of the P2P-R in the three tissues of the BXD mouse, found a gene product Pum2, which was identified to have a positively strong mutual expression with P2P-R. This gene product is part of the Puf family; that contains highly conserved RNA binding proteins which are known as translation regulators specific for a particular gene i.e. the specific interaction of elements found in the targeted mRNAs is located at the 3' end of the untranslated region (UTR). Futhermore, the 3' UTR of P2P-R has been shown to have a single perfect consensus and dual (almost perfect consensus) where Pum2 can bind. Thus, Pum2 surely binds P2P2-R mRNA to regulate its expression, as shown by a pull-down assay and the reverse transcription followed by RT-PCR (Peidis *et al.*, 2010). The expression of the P2P-R protein is shown to be repressed by terminal differentiation of human normal keratinocytes (when the keratinocytes irreversibly lose their proliferative potential) and aging of human cells (Minoo *et al.*, 1989).

P2P-R has various functions that include mitosis, mRNA processing and translation. Recent studies have also shown that it functions as the ubiquitination and degradation factor of p53 and YB-1 proteins (Peidis *et al.*, 2010). The gene product of P2P-R has been shown as highly significant in the different pathological and biological processes, including controlled cell growth, differentiation, carcinogenesis and above all, as a tumor suppressor (Witte and Scott, 1997). Peidis and co-workers (2010), using rat HXB/BXH recombinant strains in an adipocyte genetic reference panel from the eye together with mice BXD recombinant inbred strains, suggested P2P-R as also present in the eye and adipose tissue of a wide genetically-defined transcription network.

2.1.4 Mpe-1p

Mpe-1p is a protein composed of 441 amino acids with a size of 49.5kDa. This protein was identified by Vo *et al.* (2001) and is implicated in the 3' cleavage and polyadenylation of mRNA in eukaryotic cells. These are important steps in the synthesis of active mRNA, which is also implicated in the cleavage of transcription and mRNA export from the nucleus (Huang *et al.,* 2000). The poly(A) tail and poly(A) binding protein (Pab1p) are stimulators of mRNA translation and during degradation, are dependent on deadenylation, these two factors control the stability of mRNA by deadenylation. It is through this action that the poly(A) tail and poly(A) binding protein (Pab1p) control the expression of the gene within the system (Vo *et al.*, 2001). The *cis*-acting sequences recognised in the transcript of the mRNA maintain the specificity of these processes (cleavage and polyadenylation). More so, Mpe1 is needed for specific cleavage of the 3' end in mRNA *in vivo*. There are three multiprotein complexes involved in mRNA cleavage and polyadenylation: the cleavage factor I (CFI), cleavage and polyadenylation factor (CPF) and the Pab1p (Vo *et al.*, 2001). The CPF recognises the polyadenylation signal (AAUAAA sequence), while the CSF recognises the "CA sequence". After these factors have been recruited, the CPF then cleaves the CA sequence. The polyadenylate polymerase enzyme is responsible for the addition of adenine which recruits the Pab1p to join adenine and mRNA through the CA sequence. Mpe1p has also been shown to be an integral subunit of the CFI, however it is not responsible for its stability. Vo and co-workers (2001) further showed Mpe1 is important for the normal growth and development of the cell which is common in most genes that are part of mRNA 3' end processing. Searching for the known protein motif in Mpe1p, a zinc knuckle was revealed (CX_2-CX_4-HX_4-C) between amino acids 182 and 195, and is responsible for protein-protein interactions as well as interactions with single-stranded nucleic acids.

2.1.5 SNAMA (something that sticks like glue…)

SNAMA is a 139kDa protein composed of 1231 amino acid residues. It has the same structure as the protein N-terminal conserved domain that folds like ubiquitin called the DWNN (Domain With No Name), a zinc finger and RING-finger-like motif. The SNAMA protein has homologues that have been reported in yeast, rats and humans. The SNAMA homologue identified in yeast is associated with 3'-end mRNA processing, while those found in mammals interact with Rb and p53 (Ntwasa *et al.*, 2005). It was also found that PACT and RBBP6 are the shorter transcripts of SNAMA found in rats and humans respectively (Ntwasa *et al.*, 2005).

SNAMA is composed of four domains which include the conserved DWNN domain, C2CH zinc finger motif, cysteine-rich RING finger like motif and a lysine-rich region. Ntwasa and co-workers (2005) show that in the RING-like domain of SNAMA, histidine4 (His4: which is characteristic of RING fingers), is replaced by a serine. The sequence of SNAMA has about 22% similarity to ubiquitin; however, the structural topology is identical. According to the predicted structure of SNAMA, it is shown that the second α-helix present in ubiquitin is absent; however, this reappears when SNAMA is modelled against ubiquitin, although upon energy minimization, it disappears.

The presence of Lysine at position 29 in SNAMA makes it responsible for the formation of a poly-ubiquitin chain and recruiting damaged proteins to 26S proteasome for degradation. It is also suggested that SNAMA plays a significant role in apoptosis, either by acting as an apoptosis suppressor or in the negative regulation of the apoptosis activator. Lysine 48 and 63 (Lys48 and Lys63) are missing in SNAMA; these amino acids are significant for the recruitment of modified proteins so that they can be degraded by the 26S proteasome.

Furthermore, these amino acids are also important in the regulation of cellular functions that include DNA repair and intracellular cell sorting (Ntwasa *et al.*, 2005).

Ntwasa and co-workers (2005) also reported that SNAMA is expressed during the development of the *Drosophila melanogaster* embryo as a single transcript of 4.6Kb with the gene located at the right arm of chromosome 2 at position 60B9 in the genome of *Drosophila*. However, it has also been observed that in adulthood, the concentration of this protein is vastly reduced.

2.2 RBBP6 domains

RBBP6 homologues include the DWNN (Domain With No Name), zinc finger domain, RING finger domain, serine-arginine domain, Retinoblastoma (Rb)-binding domain and p53 binding domain. However the truncated RING domain, zinc knuckle and DWNN have been observed in vertebrates (Pugh *et al.*, 2006).

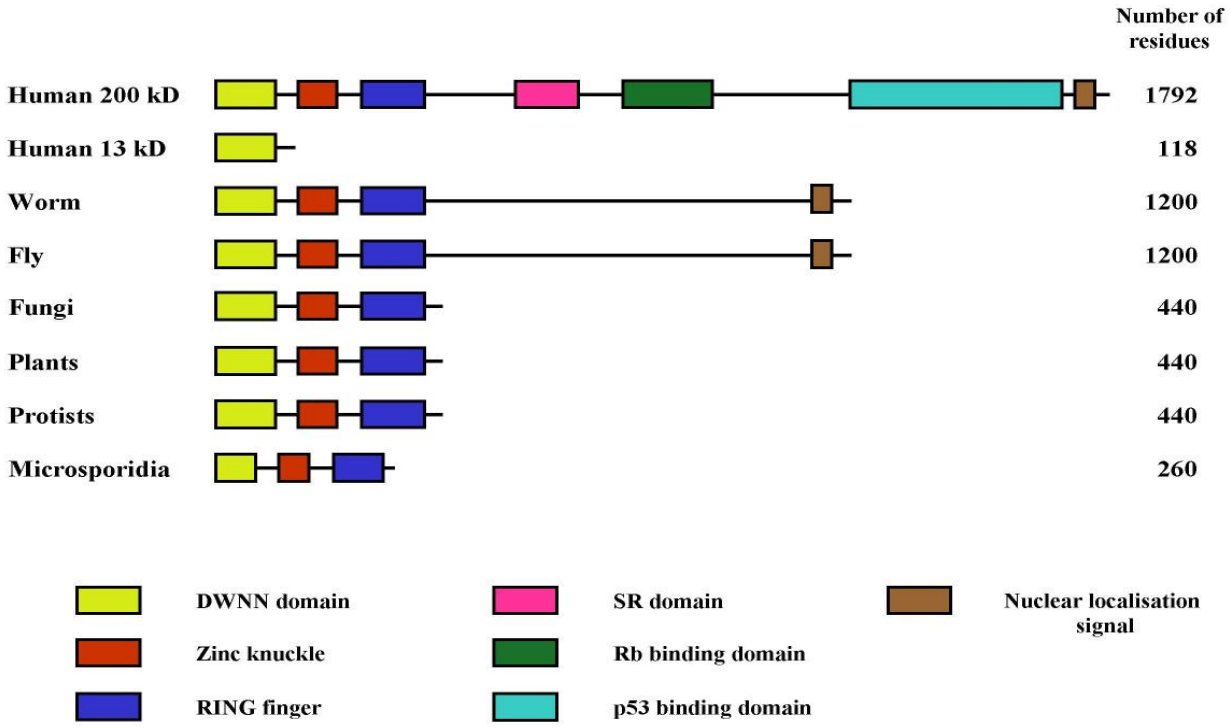


Figure 2.1: **RBBP6 structural domains found in different organisms**. DWNN (Domain With No Name), Zinc knuckle, RING (Really Interesting New Gene), SR (domain serine/arginine-rich region), Retinoblastoma (Rb)-binding domain, and the p53-binding domain. Taken from Pugh *et al.*, 2006.

2.2.1 DWNN (Domain With No Name)

DWNN is one of the domains that make up the RBBP6 protein and is the first domain in the RBBP6 domain structure. RBBP6 is expressed as a singular protein domain containing the DWNN at the N-terminal end between 118-150 amino acid residues in vertebrates. It has been reported that the DWNN, which is at the N-terminal of RBBP6, adopts the ubiquitin-fold, which suggests that it also plays a role as an ubiquitin-like modifier (Pugh *et al.*, 2006).

DWNN is present in fungi, plants, microsporidia, protozoa and also in *Encephalitozoon cuniculi*, in a shorter version in the single cell parasite (Pugh *et al.*, 2006). Looking at the sequence database for DWNN homologues alignment for a variety of eukaryotes, no similarities was found. This finding confirms the novelty of the DWNN protein motif, and that it arose after eukaryotes emerged (Pugh *et al.,* 2006). Expression of RBBP6 in bacteria has been observed to be sensitive to proteolysis from amino acid residues 82-118 while residues 1-18 DWNN is more stable (Pugh *et al.,* 2006). DWNN can both be classified as an ubiquitin domain protein (UDP) as well as an ubiquitin-like domain (UBL) and is the first discovered ubiquitin-like domain (Pugh *et al.*, 2006). Additionally, ubiquitin-like proteins contains the GG motif that is recognised by the cleavage of protease between the di-glycine which occurs during the conjugation process. Therefore, the presence of this motif in the human and mouse DWNN may imply that the DWNN plays a role in conjugation called ‘DWNNylation’ (Pugh *et al.*, 2006). However, the GG motif is absent in lower organisms, showing that DWNN is not a covalent modifier in lower organisms (Pugh *et al.*, 2006).

2.2.2 ZINC finger domain (Zinc Knuckle)

Zinc is a metal ion found abundantly in living organisms where it functions as a co-factor for metabolic enzymes and transcription factors (Dudev and Lim, 2003). The zinc-binding sites in a protein are divided into two: namely the catalytic zinc site and structural zinc site; the

latter being the best studied site out of the two. The most studied ‘structural’ zinc-proteins are those of the zinc-finger family that are responsible for the regulation of gene and nucleic acid binding (Dudev and Lim, 2003). Zinc-finger domains, also known as the zinc knuckle, are small protein domains that use zinc to stabilize their structural domains (Krishna *et al.*, 2002), forming a core domain which is set apart from solvent by cysteine and histidine. The stability of these proteins is through the binding of zinc metal ion. This binding of zinc assists in heat fluctuation and conformation stability but is not necessarily involved in the functioning of these proteins (Krishna *et al.*, 2002).

The ‘structural’ zinc-sites are either partially or fully buried (at the core) where they are surrounded by hydrogen bonds from the second layer. The zinc preferable ligands are cysteine and histidine because of their soft nature; however, zinc can also bind or be coordinated with aspartate and glutamate side chains. There are different ways in which zinc can be coordinated in different metabolic enzymes and transcription factors, including octahedral coordination which is normally in aqueous solutions, pentahedral and hexahedral coordination in zinc-catalytic sites and lastly, the tetrahedral coordination which is more abundant in ‘structural’ zinc binding sites (Dudev and Lim, 2003). In the environment of the protein, zinc may either be octahedrally or tetrahedrally coordinated depending on the type of protein ligand it binds to and the solvent accessibility to the site that binds the metal.

The tetrahedral coordination is found to be the more stable as compared to other coordinations and its structures are presented by shorter distances of metal ligands that coordinate 5- or 6-coordinates in all structures of the inner-sphere. The zinc tetra-coordinated state is favoured because of inaccessibility of the solvent and low dielectric medium, which is proven by the stability in the tetrahedral zinc hydrate compared to the

octahedral in the gaseous phase. Secondly, the charge in the ligands is transferred to the metal cation and lastly, the availability of an unoccupied orbital of zinc that goes to the highest occupied orbital and its ligands are able to accept the charge (Dudev and Lim, 2003).

There are two reasons why the zinc-coordination is more stable than the other metals. Firstly, zinc-binding sites do not affect the “coordination strain” to the catalytic functions of zinc enzymes. Secondly, the shorter metal-ligand bond lengths of the tetrahedral zinc-binding sites are preferred for stabilizing a protein fold. Zinc-binding sites are more specific compared to other metal cofactors; this allow the protein to successfully exclude zinc from a pool of other naturally occurring metal co-factors in the cell fluid (Dudev and Lim, 2003). The displacement of zinc from the zinc-finger core by other metal ions has been confirmed by Hartwig and co-workers (2002). More so, the other metal ions cannot maintain the proper shape of the protein which in turn disturbs the process of DNA-binding (Dudev and Lim, 2003). It has also been reported that metallothioneins are essential for repairing damaged zinc-binding sites by removing toxic metals from the binding site and bringing in the zinc co-factor to the same binding site (Dudev and Lim, 2003).

Zinc-finger domains are divided into eight folds which are separated into different families (Dudev and Lim, 2003). The different fold groups include metallothioneins, zinc-binding loops, TAZ2 domain like Zinc2/Cys6, zinc ribbon, Gag knuckle and C2H2-like finger. Among the eight fold groups, the classical C2H2-like finger is the best studied fold group. This group domains are presented by the β-hairpin and α-helix that forms the ββα-unit situated at the left hand. The C2H2-like finger is composed of two families which are inhibitors of apoptosis protein (IAP) domains and C2H2 fingers (Krishna *et al.*, 2002). The C2H2 finger family was originally discovered in transcription factor IIIA of *Xenopus laevis*

and it has also been found in other varieties of transcription factors including those that recognize specific sequences of DNA. This family is composed of a repeated 28-30 amino acid sequence with cysteine and histidine residues (Krishna *et al.*, 2002). It is suggested that the major function of the C2C2 finger family is to regulate transcription; however, it has also been recently shown that members of this class may play a role in mediating protein-protein interactions (Krishna *et al.*, 2002).

Conversely, the IAP domain is coordinated by zinc ions in a C2HC pattern. It has also been reported that it inhibits caspases for the regulation of apoptosis (Krishna *et al.*, 2002). The structural domain of IAP is like the C2H2 motif because the knuckle is the provider of the first two ligands whilst the broken α-helix has the other two ligands at its C-terminal. Although, there are similarities in structures of both C2H2 and IAP, there is no evidence of homology; therefore, they are two distinct families (Krishna *et al.*, 2002). It has also been reported that zinc fingers having a similar folding pattern, also have the same ligands structurally. However, these similarities do not mean these zinc fingers are evolutionarily-related (Krishna *et al.*, 2002).

2.2.3 Serine/Arginine domain

The serine/Arginine domain is commonly found in the SR protein (Graveley *et al.,* 1998). These are conserved family types of protein composed of repeats of serine and arginine amino acids, thus the origin of this protein name (Long and Caceres, 2009). The serine/arginine domain functions as RNA splicing factors, which contain the RNA binding domain and the serine/arginine rich domain required for protein-protein interactions (Graveley *et al.,* 1998). The serine/arginine proteins are factors represented by the occurrence of the RNA recognition motif and the SR domain and are also involved in alternative splicing decisions of RNA (Wenhog *et al.*, 1998). During the interaction of the SR

protein to form the spliceosome, the ASF/SF2 and SC35, which are also the members of serine/arginine family, interact with U1-70k and ASF/SF2 RS (Wenhog *et al.*, 1998). U1-70k is a human protein rich in serine/arginine domains and is the first to assemble with the pre-mRNA employing RNA and protein constituents to recognise the 5' splice site during splicing. It is also noted that mutational analysis implicated the importance of the sequence (Arg-Arg-Arg-Ser-Arg-Ser-Arg-Asp) as compared to other animal species homologues (Wenhog *et al.*, 1998).

2.2.4 Rb-binding domain (Retinoblastoma binding domain)

Rb-binding domain is a domain found in pRB, a tumour suppressor protein encoded by the RB gene (Saijo *et al.*, 1995). pRB is composed of 928 amino acids. The inactivation or mutation of RB gene is seen in various cancer forms. However, the re-introduction of the wild type RB gene has shown to reduce the size of tumour cells and reverse their transformed phenotype (Saijo *et al.*, 1995). pRB also plays an important role in apoptosis and regulation of cell cycle progression (Hsieh *et al.*, 1999). In addition, it has been postulated that Mdm2 anti-apoptotic activity, enhanced by the presence of p53 can be prevented by pRB-Mdm2 complex (Hiesh *et al.,* 1999). One of the functions of Mdm2 is to target p53 for degradation. However, it has also been observed that in the presence of pRB, p53 is protected from being degraded which may imply that Mdm2 degradation is inhibited. pRB is also implicated in the maintenance of DNA replication by forming a complex with transcriptional factor E2F (Sherr and McCormick, 2002). pRB has two sites referred to as A/B where viral oncoproteins such as adenovirus E1A protein, the human papilloma virus E7 protein and SV40 large T antigen bind. These viral oncoproteins inactivate the growth inhibitory activity of pRB; however, SV40 selectively binds under-phosphorylated pRB as it is the one responsible for anti-apoptotic activity (Sherr and McCormick, 2002).

2.2.5 p53 binding domain

The p53 protein is a DNA-binding nuclear phosphoprotein that can either be in the form of homotetramer or a complex of tetramers. This is one of the domains that make up the RBBP6, which is a known tumor suppressor protein. Its gene, *TP53* is located on chromosome 17p13.1 (Freed-Pastor and Prives, 2012). p53 is subdivided into three parts which include the amino-terminal domain with the activation domain, the central core with sequence-specific DNA-binding domain and the multifunctional carboxyl-terminal domain (Ko and Prives, 1996). The amino terminal of p53 has the ability to bind the TATA box-binding protein (TBP) so as to be able to bind DNA. Apart from this being one of p53 unique features, the protein also has the ability to associate with the Mdm2 gene. It has also been observed that mutation of the p53 is common in human malignancies (Bellamy, 1996) and that mutated residues inactivate the transcriptional function of p53 and also disturbs its interaction with TBP and Mdm2 (Ko and Prives, 1996). The main role of p53 is to recognise damaged DNA and repair it, regulate the cell cycle and act as the initial signalling factor for apoptosis (Bellamy, 1996).

2.3 RING (Really Interesting New Gene) finger domain

RING finger protein is a protein that is represented by the richness of cysteine sequence motifs that bind two zinc atoms (Dominguez, 2004). It consists of ~70 amino acids that are covalently attached to each other (Callis *et al.*, 2005). The canonical structure of the RING is shown to have four cysteine residues bound to one zinc atom, while three other cysteines and one histidine residue are bound to each other (Dominguez, 2004). This is a tetrahedral type of coordination, also known as “cross brace” topology (Krishna *et al.,* 2002). The “cross brace” of the RING finger domain gives it the globular conformation which is characterised by a central α-helix and variable-length loops separated by several small β-strands (Chasapis *et al.*, 2010). The RING finger was initially discovered from the human

protein product RING1 (Really Interesting New Gene 1) located on chromosome 6 proximal to the histocompatibility region (Chasapis *et al.*, 2010). The RING finger sequence is defined as Cys_1-Xaa_2-Cys_2-$Xaa_{9–39}$-Cys_3-$Xaa_{1–3}$-His_4-Xaa_{2-3}-Cys/His_5-Xaa_2-Cys_6-$Xaa_{4–48}$-Cys_7-Xaa_2-Cys_8 (where Xaa represents any amino acid residue) (Chasapis *et al.*, 2010). The 3D structure of the domain revealed conserved cysteines and histidines are buried in the core and this assists in the maintenance of the structure through its zinc binding properties (Deshaies and Joazeiro, 2009).

The RING finger domain is known to function as protein-protein and protein-DNA promoters (Chasapis *et al.,* 2010). They are also involved in several cellular processes such as viral function, apoptosis, oncogenesis and development (Dominguez, 2004). The other major function of the RING finger domain is that it is involved in ubiquitination, where it recruits targeted proteins and transfers them to 26S proteasome for degradation (Deshaies and Joazeiro, 2009).

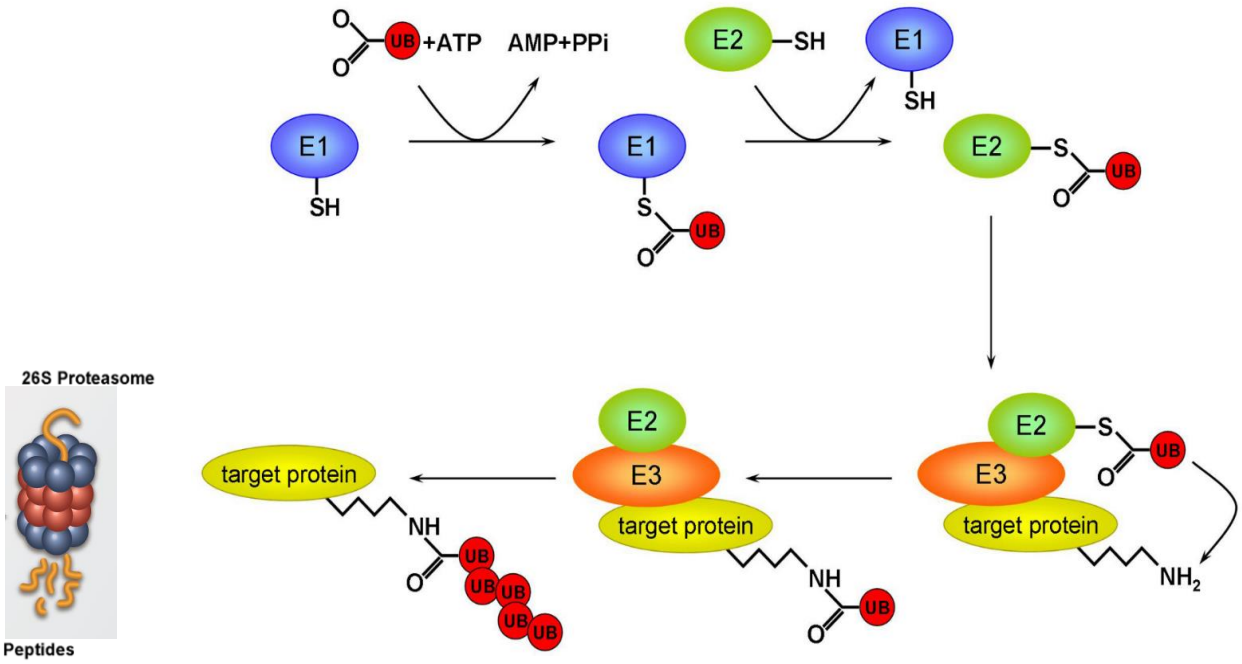


Figure 2.2: The ubiquitination pathway. UB (ubiquitin), E1 (Activating enzyme), E2 (Transfer/Conjugating enzyme), E3 (Ligase enzyme). Adapted from Panasenko, 2014.

Ubiquitination is the irreversible tagging of intracellular proteins by ubiquitin for degradation in a multi-enzymatic complex. This process is achieved by a cascade of three enzymes referred to as ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin-ligating (E3) enzyme (Callis *et al.*, 2005). First, the E1 enzyme forms a thiol ester with the carbonyl terminal group of a small ubiquitin protein at position Gly76. Activated ubiquitin is then transferred in a thioester linkage from the active-site cysteine of E1 to the active-site of E2 ubiquitin-conjugating enzyme. The E2~Ub thioester is transferred to E3 ubiquitin ligase, which causes the ubiquitin transfer from E2~Ub to a lysine residue of a substrate. Lastly, ubiquitin is transferred from E2 to the target protein promoting the ubiquitination of the substrate. The polyubiquitination substrates are recognised by the 26S proteasome where they are then degraded by proteolysis (Deshaies and Joazeiro, 2009; Domingues, 2009). The biological effect of ubiquitination, whether in degradation or signalling, depends on the ubiquitin receptors that binds and interprets the ubiquitin signals (Deshaies and Joazeiro, 2009; Domingues, 2009). Studies have shown that ubiquitination is not only seen in diseases such as cancer, but also in Parkinson's disease, which is due to the mutation of the RING domains This is because mutation of the RING finger domain prevents an efficient ubiquitination and degradation process (Dominguez, 2004).

The RING finger domain is not the only protein/domain with an atypical 'cross-brace' topology, other RING-finger like domains such as PHD, FYVE and U-box have shown similar coordinating topology (Kappo *et al.,* 2012; Kappo, 2009). Other RING-like domains such as LIM also coordinate two zinc ions but not in a cross-brace motif, rather, in a tandem topology (Dominguez, 2004). The similarities in the structural topology of these domains do not necessitate similarity in functions. For example, the secondary structure elements of PHD differ from that of RING domains (Kappo, 2009). The same is seen when looking at the structure of the FYVE domain in comparison to other domains; there may be similarities

in topology but their structures are very different due to the spaces between the coordinating residues (Dudev and Lim, 2003; Kappo, 2009). The U-box structure is different to that of the RING-finger domain in that U-boxes do not bind zinc ions and the signature zinc-coordinating residues are lacking (Aravind and Koonin, 2000). Depending on the combination of cysteine and histidine, length and sequence homology, the RING finger domain can be classified into canonical and modified RING finger domains, yielding the following subclasses: C3HC4, C3HHC3, C2H2C4 and C4C4 (Lian *et al.*, 2009; Kappo, 2009).

2.3.1 C3HC4 classical RING finger domain

The C3HC4 classical RING finger domain is also known as the RING-HC canonical subclass of the RING finger domain (Lian *et al.*, 2009). This domain differs from other domains in that histidine is at position 4 while the other domains have it at position 5 (Lian *et al.*, 2009). The C3HC4-type RING finger genes are abundant in the plant kingdom which contributes to the vast biological functions in the physiological processes of plant life. This is confirmed by studies of C3HC4 in *Arabidopsis* in which the genes were found to possess different functions in cellular processing and signalling transduction (Lian *et al.*, 2009). C3HC4 acts as an E3 ubiquitin ligase which plays a role in the downstream ubiquitination and degradation of targeted photomorphogenesis-promoting transcription factors. OsRHC24 and OsRHC1 are C3HC4 genes that are related to disease resistance mechanism (Lian *et al.*, 2009).

Figure 2.3: A typical C3HC4 RING-type domain. It is composed of two grey spheres representing zinc ions that is used for it stability. Taken from Kellenberger *et al*, 2005.

2.3.2 C3HHC3 RING finger domain

This is also known as RING-H2. The expression of this protein is induced by *N*-acetychitoologosaccharides in rice. RING-H2 is provided by EL5 which is structurally related to the *Arabidopsis ATL* protein family. This protein is characterised by a transmembrane domain (I), a basic domain (II), a conserved domain (III) and a RING-H2 finger domain (IV) which is followed by the C-terminal region, composed of highly diverse amino acid sequences (Katoh *et al.*, 2003). The solution structure of the *EL5* RING-H2 is stabilized by the tetrahedral coordination of zinc ions (Katoh *et al.*, 2003). The difference between RING-H2 and other RINGs is the replacement of Cys4 by Histidine (Kappo *et al.*, 2009). Katoh and co-workers (2003) determined the structure of the RING-H2 type of RING finger *EL5* by heteronuclear NMR and showed it is composed of β-strands and α-helix in a ββα fold (β1, Ala^{147}-Phe^{149}; β2, Gly^{156}-His^{188}; α1, Cys^{161}-Leu^{166}). It was also shown to have a long flexible loop on each side of the ββα structure (N-loop, Val^{133}-Glu^{146}; C-loop, Gly^{167}-Val^{180}) (Katoh *et al.*, 2003). Ubiquitination studies done *in vitro* have also shown that RING-H2 finger domain has the ability to bind E2 (Katoh *et al.*, 2003).

2.3.3 C2H2C4 RING domain

This domain is an example of the Hdm2 (Human domain minute 2), which is the human homologue of Mdm2. Hdm2 is an ubiquitin protein ligase responsible for the suppression of the tumor suppressor p53 transcription activity and is simultaneously an inducer of its degradation (Kappo, 2009). The solution structure of this domain has shown a novel cross-brace zinc binding topology and a symmetrical dimer. Kostic and co-workers (2006) showed Hdm2 as the only protein that has C2H2C4 zinc coordination. C2H2C4 is composed of the ββαβ fold, a small hydrophobic core and two zinc ions. The two zinc ions in the C2H2C4 domain structure are important for the maintenance and stability of the structure (Kostic *et al.,* 2006). As earlier mentioned, the C2H2C4 is the novel rearrangement of zinc-binding

sites that Hdm2 possesses in which the tight binding of the two provide E3 ubiquitin ligase activity, thereby creating the isopeptide bond between a lysine of the target protein and a C-terminal glycine of ubiquitin (Metzger *et al.*, 2012).

2.3.4 C4C4-type RING finger domain

The C4C4-type RING finger domain is a modified type of RING finger domain that is also known as RING-C2 (Yu *et al.*, 2011). This C4C4 domain was identified from *Arabidopsis thaliana* along with other domain types (Yu *et al*, 2011). Yu and co-workers (2011) showed that the analysed amino acid sequence of VpRFP1, a protein from grapevines is composed of a nuclear localization signal at the N-terminal and the RING finger at the C-terminal. The RING finger on the C-terminal is seen to be from the novel RING finger C4C4 with the consensus sequence of Cys-X_2-Cys-X_{13}-Cys-X_1-Cys-X_4-Cys-X_2-Cys-X_{10}-Cys-X_2-Cys, where the X represents any amino acid (Yu *et al.*, 2011). The C4C4 is also very similar to the C3HC4, with only one cysteine residue replacing histidine. The C4C4-type RING domain structure has a βββα fold and the chemical nature of this domain is said to be identical according to Houben and co-workers (2005). The C4C4-type RING domain has significantly different dynamic properties, raising the assumption that this difference might be related to the stability of the domain and perhaps the functionality of protein-protein interactions that they are involved in (Houben *et al.*, 2005).

Figure 2.4: The C4C4 RING finger domain. The two grey spheres represent zinc ions binding. Taken from Kumeta *et al*., 2011.

2.4 Bioinformatics

Bioinformatics has become a useful tool for information relating to genome sequences. It is a multidisciplinary field that utilizes different methods, including computational technology, to keep biological information for structural studies using either nucleotides or amino acids (Pawelkowicz *et al.*, 2015). This has now become a solution and even serves as a guide to novel biological discoveries and in designing new experiments. On the contrary however, the use of bioinformatics technology has limitations due to the data obtained not always being reliable. It can also be time consuming depending on the data one may be searching for (Wanichthabarak, 2014). Building blocks of proteins known as amino acids play a vital role in structural prediction. Knowing the structure of a protein allows one to have an understanding of the protein amino acid sequences (Abdullah *et al.*, 2015). Therefore, secondary structure predictions make use of amino acid sequences. However, the use of these methods may not be as accurate as those obtained using X-ray crystallography and NMR spectroscopy (Rokde and Kshirsagar, 2013).

2.5 Biophysical characterization of proteins

Once pure protein is obtained, biophysical characterisation helps in ascertaining the stability of the protein, whether the protein is folded or not and the intensity of light the protein is able to absorb or emit (optical properties). These results eventually provide a suitable platform for downstream structural genomics and structural determination.

2.5.1 Ultra-violet (UV) spectroscopy for proteins

In two different types of cells, either eukaryotes or prokaryotes, there are important compounds found within their proteins that are responsible for the emission of light (Rosselló-Mora and Amann, 2001). These compounds may either be an enzymatic co-

factor or amino acids. The structures of these compounds are similar when they all possess an aromatic ring which is responsible for light fluorescence or for light absorption.

There are three types of amino acids that regulate the emission and absorption of light in a protein: they are tryptophan, tyrosine and phenylalanine. Although they have common traits of an aromatic ring, they absorb and emit radiation at different wavelengths resulting in different quantum yields (McLaren and Shugar, 2014). Out of the three amino acids mentioned, tryptophan is the one that fluorescence the most because of the intensity of the solvent polarity. When solvent polarity levels decrease, the tryptophan spectrum is allowed to move towards the shorter wavelength which in turn results in a high intensity of fluorescence (Demchenko, 2013). According to the intensity of the fluorescence scale, tyrosine comes in as the second amino acid which signals as much as tryptophan even though the scale unit is smaller than that of tryptophan. This is because tyrosine is commonly found in most proteins. More so, it is noted that neighbouring tryptophan residues contribute to tyrosine fluorescence intensity by transferring energy resonances through ionization of the hydroxyl group. Lastly, in absence of the two amino acids mentioned above, phenylalanine gets a chance to show its fluorescence properties. Its significance is only recognised in the absence of the other two amino acids aforementioned as its transmission is overshadowed (McLaren and Shugar, 2014).

2.5.2 NanoDrop

For NMR studies, a protein concentration between 0.1 and 2.5mM is needed (Wider and Dreier, 2006). Therefore, to accurately determine protein concentration, a NanoDrop spectrophometry is used (O'Neill *et al.,* 2011). The NanoDrop is a spectrophotometer that gives estimated concentration values of desired proteins (Desjardins *et al.*, 2009). When eventually compared to conventional spectrophotometry used in previous years, the

NanoDrop spectrometry is a more feasible method requiring only 0.5-2µl of the sample instead of 1-2ml needed by cuvettes or capillaries (Desjardins *et al.*, 2010). This is advantageous as it saves more protein sample for structural studies. Only purified protein samples with aromatic amino acids (tryptophan, tyrosine and phenylalanine) qualify for concentration estimation. Cysteine-enriched proteins are also easily determined by NanoDrop because of their disulphide bonds (Desjardins *et al.*, 2009).

2.5.3 Protein analysis by mass spectrometry

Mass spectrometry is a technique used to determine the mass of small stable fragments of biomolecules and can also be used to identify the folded regions of the protein after the generation of shortened constructs to eliminate the unfolded regions of the protein (Pugh, 2005). It gives an insight into the molecular structure of the protein but not the molecular scaffold or the 3D structure of the protein (Seger *et al.*, 2013). Due to size limitations, the mass spectrometry is recognised as the smallest scale in the world (Whitehouse *et al.*, 1985) as it measures the mass/charge ratio of a molecule (Fenn *et al.*, 1990). Mass spectrometry is also used as a tool to determine the purity of a protein and to ascertain the incorporation of stable isotopes into the protein of interest. The limitation of this technique is that it can only read one label of a protein at a time hence, for a double-labelled protein, decoupling is required. Decoupling refers to the removal of protons from the proteins so that the energy released cannot be conserved (Nedergaard *et al.*, 2005).

2.5.4 Fourier Transform Infrared (FTIR) spectroscopy

Fourier Transform Infrared Spectroscopy is used to determine the identity of protein secondary elements, which is representative of the secondary structure of the protein (Kong and Yu, 2007). This is an easy-to-use experimental technique under different conditions that requires lesser sample volume preparations leading to limited chances of sample

damage (Abidi *et al.,* 2014). Unlike the conventional Infrared (IR) experimental technique that scans the spectrum at a time, FTIR is a modern spectroscopy that is able to rapidly perform multiple scans in a minute (Kong and Yu, 2007). The success of the scans however, depends on the wavelength and the powerful strength of the protein to absorb the IR radiation rays (Abidi *et al.*, 2014).

2.5.5 Nuclear Magnetic Resonance (NMR) spectroscopy

Macromolecules such as DNA, RNA and proteins are essential components of the protein body hence; it is of utmost importance to ascertain their functions in various physiological states. The study of the biological structure of these macromolecules allows for easier determination of their functions. Two of the techniques used to solve the structure of proteins are X-ray crystallography and heteronuclear NMR spectroscopy. X-ray crystallography was established by Kendrew in 1958, while NMR was started by Ernst and co-workers in 1970 and was developed for biological studies by Wuthrich and his group in 1980 (Pugh, 2005).

Besides the determination of protein structure by NMR, the technique also provides biological information pertaining to protein dynamics such as the rate of folding, stability and the occurrence of hydrogen bonds. The study of the protein structure using heteronuclear NMR becomes easier when the proteins are incorporated with magnetic isotopes such as ^{15}N and ^{13}C to allow visualization of resonances (Pugh, 2005).

The accuracy of the heteronuclear NMR spectroscopy is dependent on the size and concentration of the protein. Proteins of 30kDa and above are cumbersome to analyze because of the insensitivity of the NMR technique, a high concentration of protein is therefore needed (approximately 0.5mM and above). The disadvantages with the analysis of large-sized proteins include overlapping of spectra, sensitivity and suppression of water.

Overlapping resonances seen in bigger proteins are as a result of resonances being crammed in the same region of the spectrum due to the high number of amino acids present. This phenomenon is known as spectral overlap and is usually observed at the aliphatic region. Hence, to overcome this limitation, larger proteins are expressed as smaller domains with the assumption that these domains may be able to fold and function independently (Pugh, 2005).

The sensitivity of NMR spectroscopy depends on the concentration of the protein sample. Now, NMR has the tendency to be less sensitive and this requires the enrichment of the sample with isotopes such as ^{15}N and ^{13}C (Pugh, 2005). The other challenge faced when using NMR is to suppress water to avoid 'swamping' of the protein spectrum. Therefore, a means of water suppression must be applied. The initial technique for water suppression is to add deuterium oxide (D_2O) due to its identical formula to H_2O. However, it was found that some deuterons may swap with the backbone resonances of the protein resulting in the loss of amide protons, which play a role in sequential assignment of protein spectra. Other methods used for water suppression include selective pre-saturation of water resonances and the use of pulse field gradients (Pugh, 2005).

The resonances collected from an 2D-HSQC are further used to assign the backbone of the protein using two triple resonance experiments HNCACB and CBCA(CO)NH. These experiments are used to assign each peak from the ^{15}N-HSQC spectrum to each residue of the isotopically-enriched protein and also for the assignment of side-chain carbon and proton resonances (Pugh, 2005). However, these experiments are most effective for proteins that are less than 30kDa due to the fact that larger proteins, because of size, may cause resonances to overlap or even in some instances, lead to missing cross-peaks (Pugh, 2005).

2.6 Problem Statement

A genome-wide bioinformatics analysis by Kappo and co-workers (2012), using the human RING finger domain from RBBP6 as a searching bait, showed the zinc ion coordination pattern was the same in all but three organisms: *Saccharomyces cerevisae* and *Pichia pastoris,* where the first binding-site of zinc has been lost completely or weakly bound; and in *Aspergillus niger* where the second coordinating signature cysteine residue has been substituted by an aspartic acid (Figure 2.5). However, it is unclear whether the coordinating zinc ion was lost or was weakly bound. Structural characterization of the loss of the zinc first binding site will add to our understanding of the biological role played by the RING finger domains either in ubiquitination, multimerization or in protein quality control. Hence, since the structure of a protein reveals the function of that protein, it is imperative to solve the structure of the *S. cerevisae* RING finger domain. However, biophysical characterization of this protein will provide the biological basis of high throughput expression and purification strategy to obtain appropriate amount of labelled *S. cerevisiae* RING finger domain protein. It is also most important to ascertain the optical properties of the protein before embarking on the acquisition of triple resonance coherent data for structural determination. This is because NMR or whatever methods employed, are all based on the light-absorbing properties of the protein.

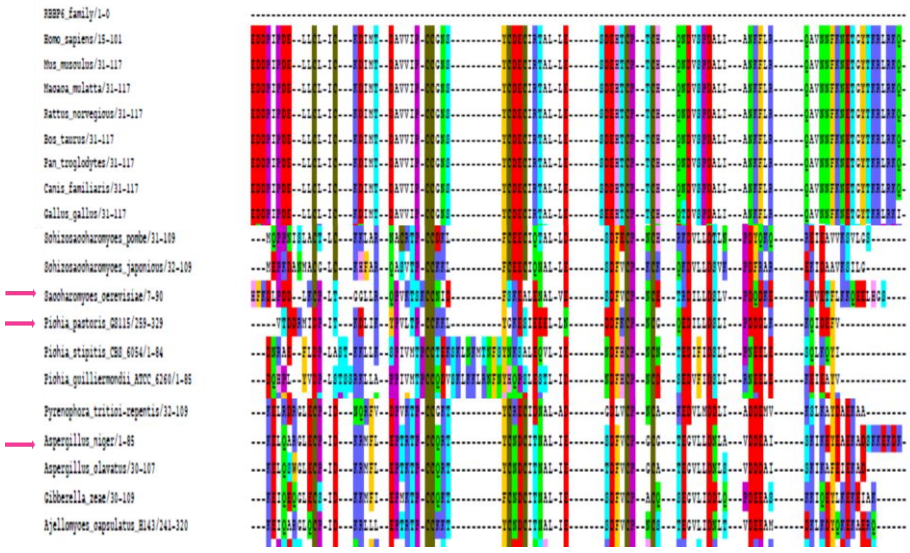


Figure 2.4: Sequential alignment showing the amino acid sequences of RING finger domain in various organisms. The pink arrows shows organisms (*Pichia pastoris, Aspergillus niger* and *Saccharomyces cerevisiae*) in which different kind of RING sequence was discovered compared to other organisms. Taken from Kappo *et al.*, 2012.

2.7 Aim of the study

The aim of the study was to express, purify and structurally characterize the *Saccharomyces cerevisae* RING finger domain in order to elucidate the biological significance of the protein. This aim was achieved via the following objectives:

2.8 Research objectives

- Overexpress soluble samples of *S. cerevisiae* RING finger in *E. coli.*
- Investigate the folding and oligomeric state of the *S. cerevisiae* RING domain (whether monomeric or dimeric in solution).
- Determine the purity of the *S. cerevisiae* RING protein by mass spectrometry.
- Ascertain biophysical properties of the *S. cerevisiae* RING finger as a guide towards structure determination.
- Bioinformatically-determine other proteins that interact with the *S. cerevisiae* RING protein and predict its 3D structure using computational techniques.

CHAPTER 3

MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 General stock solutions, buffers and media

30% Glycerol: 30ml of glycerol was mixed with 70ml of dH_2O to make up a 100ml of 30% glycerol solution.

YT Broth: 1.6g of Tryptone, 1.0g of Yeast and 0.5g of NaCl was added together and dissolved in 100ml of dH_2O.

YT Agar: 1.6g of tryptone, 1.0g of yeast, 0.5g of NaCl and 1.5g of Nutrient Agar was added together and dissolved in 100ml.

Luria Broth: 10.0g Tryptone powder, 5.0g Yeast extract, 5.0g NaCl and 2.0g Glucose were mixed in 1000ml of dH_2O.

0.1M Calcium Chloride: 1.1099g was dissolved in 100ml of dH_2O.

0.1M Magnesuim Chloride: 2.033g was dissolved in 100ml of dH_2O.

PBS: 137mM NaCl, 27mM KCl, 4.3mM $Na_2HPO_4.7H_2O$ and 1.4 mM KH_2PO_4 were dissolved in 1 litre of deionized water and autoclaved.

Ampicillin: In 10ml of deionized water, 1.0g of ampicillin was dissolved, filter-sterilized and stored at 4°C until needed.

Agorose gel: 2.0g of Agarose powder was dissolved in 200ml of TAE buffer, boiled and allowed to cool to 55 ºC before adding Ethidium bromide.

10X TAE buffer: 48.4g of Tris, 20ml of 0.5M EDTA and 11.42ml of glacial acetic acid were mixed together and made up to 1000ml with deionized water.

IPTG: 1.0g of IPTG was dissolved in 4.2ml of dH_2O and filter-sterilized and then stored at -20°C.

10% SDS: 10.0g of SDS was made to solution by dissolving it in 100ml of deionized water.

SDS Running buffer (pH 8.3): 15.0g of Tris, 72.0g of glycine and 0.5g of SDS were made up to 1000ml with deionized water. The pH was checked but not adjusted.

2X SDS Sample buffer: 4.0ml of 10% SDS, 2.5ml of 0.5M Tris and 1.5ml Glycerol were added to 10ml of deionized water and stored at 4°C.

Destaining Buffer: 100ml of glacial acetic acid and 300ml of ethanol were mixed together and topped up with 600ml of deionized water (dH_2O) to make up a 1000ml solution.

Staining Buffer: 0.25g of Coomassie Blue R250, 40% of methanol and 10% Acetic acid was made up to a litre solution with deionized water.

Zinc sulphate: 3.22g of zinc sulphate was dissolved in 200ml of dH_2O, covered with foil and stored in 4°C.

Ethidium bromide: 0.2g of ethidium bromide was dissolved to 20ml of dH_2O and stored at 4°C covered with foil.

30% Bis-acrylamide solution: 150.0g of Acrylamide and N,N-Methylene-bisacrylamide were mixed together with 300ml of dH_2O. The solution was lightly heated at 37°C until a true solution was formed; thereafter it was topped up with deionized water to make 1000ml, filter-sterilized and stored at 4°C covered in foil.

1.5M Tris (pH 8.8): 36.3g of Tris powder was dissolved in 200ml of dH_2O and pH adjusted to 8.8.

0.5M Tris (pH 6.8): 12.11g of Tris powder was dissolved in 200ml of dH_2O and pH adjusted to 6.8.

Lysis buffer: 100µg/ml Lysozyme, 1% Triton X-100, 1mM PMSF, 50µM $ZnSO_4$ and 1mM DTT were made up to 20ml with 1X PBS buffer.

PMSF stock solution: 1.74g of PMSF was dissolved in 100ml of ethanol and stored at -80°C.

DTT stock solution: 3.9g of DTT was dissolved in sodium acetate, filter-sterilized and stored at -80°C.

Lysozyme stock solution: 1g of Lysozyme powder was dissolved in 10ml of deionized water, filter-sterilized and stored at -80°C.

10% Ammonium sulphate: 0.1g of $(NH_4)_2SO4$ was dissolved in 900µl of dH_2O, filter-sterilized and stored at -80°C.

Equilibration/wash buffer: 50mM Tris containing 150mM NaCl and made up to 500ml with deionized water.The solution was thereafter kept at 4ºC until needed.

Elution buffer: 50mM Tris containing 150mM NaCl pH 8 was used to dissolve 15mM glutathione. The solution was kept at 4ºC.

70% Ethanol: 70ml of ethanol was dissolved in 30ml of deionized water and kept at room temperature.

Separating buffer: 36.3g of Tris was dissolved in 200ml of deionized water and adjusted to pH 6.8, giving a solution with a final concentration of 1.5M.

Stacking buffer: 12.1g of Tris was dissolved in 200ml of deionized water and adjusted to pH 8, giving a solution with a final concentration of 0.5M.

NMR buffer: A 100mM Sodium phosphate buffer at pH 6 containing 50mM NaCl, 5mM DTT and 0.2% of NaN_3.

3.1.2 Bacterial strains

a. Escherichia coli BL21 (DE3) pLysS: B F-*omp T gal dcm Ion hsd*S_B(*r*$_B^-$*m*$_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K\text{-}12}$(λ^S) pLysS[*T7p20 ori*$_{p15A}$](Cm^R).

b. Escherichia coli XL 1 Blue cells: endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15 Amy Cm^R] hsdR17(r_K^- m_K^+).

3.2 METHODOLOGY

3.2.1 Preparation of *E. coli BL21(DE3) pLysS* competent cells

Five test tubes were prepared with 5ml of YT broth for culturing *E. coli* BL21 cells and incubated overnight in a shaking incubator at 37°C. The following day, one of the cultured test tubes, with the best growth judging by the turbidity of the culture, was chosen. The cells were resuspended in 50ml of YT broth without antibiotics in a concave flask for further incubation at 37°C. The absorbance of the cultured cells was periodically checked at optical density (OD_{600}) until it was between 0.3 and 0.6. When the desired absorbance was reached, the cells were harvested by centrifugation at 5000 Xg at 4°C. The supernatant was discarded and the pellet was resuspended in 0.1M $MgCl_2$, lightly agitated and kept on ice for an additional 30 minutes. After this incubation period, the cells were once again harvested by centrifugation. The supernatant was discarded and 0.1M $CaCl_2$ was added. The cells were chilled on ice for about 20 minutes and were centrifuged at 5000 Xg at 4°C. The pellet was resuspended in 0.1M $MgCl_2$; and the cells were incubated for 4hrs at 4°C with occasional swirling. Thereafter, the suspension was again centrifuged and the supernatant decanted. Lastly, 0.1M $MgCl_2$ containing 30% glycerol solution was added to the pellet to make a suspension, which was later aliquoted in 300µl into sterile Eppendorf tubes and stored at -80°C for future use.

3.2.2 Transformation of *E. coli BL21(DE3)* pLysS competent cells using pGEX-6P-2~*S. cerevisiae* RING plasmid

Two eppendorf tubes were used to transfer 100µl of competent cells. The first tube was inoculated with only cells and labelled as the control, while the other one was inoculated with 2µl of the plasmid (pGEX-6P-2~*S. cerevisiae* RING). Both tubes were incubated on ice for 20 minutes. After this period, the cells were heat-shocked on a heating block at 42°C for 45 seconds and immediately incubated back on ice for 10 minutes. A 900µl of pre-warmed YT broth without ampicillin was added to each tube and both tubes were incubated on a shaking incubator at 37°C for 1-2 hours. After this period, 100µl of cells were transferred to ampicillin-containing YT agar plates and spread evenly using a sterile hockey-stick glass rod. The remaining cells were centrifuged, supernatant removed; pellet re-suspended in media and spread as earlier described (serving as the concentration sample). Both spread-plates were then incubated at 37°C overnight.

3.2.3 Small-scale expression screening of *E. coli BL21(DE3)pLysS* cells transformed with *S. cerevisiae* pGEX-6P-2-RING domain

Fresh colonies were picked from the overnight spread-plates mentioned in Section 3.3.2 and were used to inoculate five 5ml cultures of LB broth containing 100µg/ml ampicillin, which was then incubated for 4 hours at 37°C with vigorous shaking. From each 5ml cultured cells, 1ml was each transferred to two separate eppendorf tubes to create two set of cultures. One set of the cultures were induced with 0.5mM IPTG and labelled induced cultures, while the other set was labelled un-induced cultures. Both sets of cell cultures were incubated for an additional 2hrs at 37°C in a shaking incubator. Thereafter, the cells were harvested by centrifugation at 13200 xg for 5 minutes. The supernatants were discarded and the pellets used to run a 16% SDS-PAGE gel by adding 10µl of 2x sample buffer to the sample preparation. Based on the protein expression profile shown on the

SDS-PAGE gel, the remaining 3ml of the best recombinant protein expression culture was used to set-up an overnight culture at 37°C for a large-scale recombinant protein expression.

3.2.4 Large-scale expression of recombinant *S. cerevisiae* RING domain

A 100µl of the best recombinant protein expressing clone was inoculated into 100ml of LB with a concentration of 1000µg/ml of ampicillin and incubated at 37°C in a shaking incubator overnight. The overnight culture was scaled-up to 4L flask with the same concentration of ampicillin and incubated for a further 1-2hrs in a shaking incubator at 37°C. The optical density OD_{600} was periodically measured during the incubation period until it reached the absorbance of 0.4-0.6. Filter-sterilized $ZnSO_4$ (100µM) was added prior to induction and recombinant expression of *S. cerevisiae* RING domain was induced by the addition of IPTG to a final concentration of 0.5mM with gentle shaking at 25°C for 16 hours. The overnight culture was then harvested by centrifugation at 6000 xg for 10 minutes at 4°C; cell pellets were stored at -70 °C until required for use.

3.2.5 Extraction and affinity purification of *S. cerevisiae* RING finger protein

The cell pellets previously stored at -80°C were allowed to thaw on ice. After the cell pellets thawed, they were re-suspended in a protein lysis buffer by vigorous vortexing. To break the cell wall, the cell pellets were sonicated for 10 minutes at 2 minutes intervals for each eppendorf tube. The sonicated cell mixture was then centrifuged at 10000 xg for 30 minutes at 4°C to harvest the total bacterial cell lysate, clear of cell debris. The clear bacterial lysate was collected into fresh 50ml Greiner tubes. Following this, an affinity purification column was prepared using an overnight swollen glutathione-agarose bead at 4°C, which was poured into a plastic column, allowing buffer to flow-through. Thereafter, the packed affinity column was equilibrated with 3 column volumes of (3CV) PBS at pH 7.4. After the

equilibration buffer had drained off, the cleared total bacterial lysate, containing the GST~*S. cerevisiae* RING domain was poured onto the column and allowed to flow by gravitational force. The flow-through was collected and stored on ice. Thereafter, the affinity column was washed with 5CV of PBS at pH 7.4. The last three drops of the wash buffer before the column dry up were collected and used to ascertain the purity of the GST-column before eluting the GST-tag *S. cerevisiae* RING protein. A 3CV protein elution buffer (see Section 3.1) containing 15mM Glutathione was used to elute the fusion *S. cerevisiae* GST-RING domain proteins off the GST-agarose affinity column. The different fractions from the affinity purification process were then electrophoresed on a 16% SDS-PAGE gel.

3.2.6 SDS-PAGE

A short glass was placed on a spacer plate, while using the green casting frame to hold the plates on a flat surface by clamping the plates. A freshly-prepared separating gel was added between the two glasses to about 5cm before it gets to the top. Thereafter, 10% isopropanol was added to help level the top of the gel, while it is allowed to solidify for some time. In the time the separating gel is solidifying, a stacking gel was prepared, which was later added to the top of the separating gel and a comb inserted into the stacking gel to create wells. The gel was allowed to solidify for about 45 minutes. After the gel had solidified, the comb was removed and the gel was placed in an electrophoresis tank with containing 1X SDS-PAGE running buffer, which was allowed to cover the entire gel cast system before the protein samples were loaded. The first well of the SDS-PAGE gel was loaded 2µl of a pre-mixed protein ladder while the other wells were used to load pre-boiled protein fractions. The protein samples were prepared by mixing 20µl of protein samples and 10µl of 2X sample buffer together, followed by boiling the mixture at 95°C before 20µl of the boiled mixture were loaded into the wells. The gel was then allowed to run for 1-2 hours at

120V. After the dye front of the gel was about to be released into the running buffer, the process was stopped, the gels were carefully taken out of the tank and removed from the caster before being transferred to container. Appropriate volume of Coomasie brilliant blue solution was then poured onto the gel and allowed to stain for 60 minutes or overnight. After the gel has been properly stained, a destaining solution was added unto the gel in the container and allowed to shake for as long as possible until the gel becomes very clear and the protein bands can be properly seen.

3.2.7 Biochemical characterization of *S. cerevisiae* RING domain protein

Prior to subjecting the protein the protein to biophysical characterization methods, the concentration of the recombinant *S. cerevisiae* RING domain protein was ascertained using NanoDrop ND2000 (Thermo Scientific) spectrophotometer. This is achieved by measuring 1µl of the protein sample and pipetting it onto the pedestal of a NanoDrop spectrophotometer. Prior to this, the pedestal of the machine was cleaned with deionized water, followed by blanking the machine with the buffer in which the protein was held.

3.2.8 Nuclear Magnetic Resonances Spectroscopy (NMR)

NMR spectroscopy is a tool that allows the analysis of macromolecular structures and functions. Before a protein could be structurally analysed by NMR, there is need to ensure that the protein is appropriately folded into a typical protein 3D space. 1D proton spectrum and 2D HSQC spectra can be used to confirm the ‘foldedness’ of the protein. A 1D proton spectrum indicates whether the protein was folded or not by the number of resonances that are shifted up-stream of 8.5ppm (HN proton shifts) and down-stream between +1 to -1ppm (methyl regions). However, the 1D spectrum could not be used to ascertain folding for proteins with sizes greater than 30kDa. A large dispersal of amide backbone shift and chemical shift of the protons define the folded state of a protein. The methyl group

resonances had higher intensities because they had three equivalent protons to contribute whereas the amide group only had one proton. Conversely, unfolded or denatured proteins show very poor dispersion of chemical shifts.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

This section of the study shows computational biology results depicting predictions of the physico-chemical parameters and other structural characterizations as a guide towards recombinant expression, purification and biophysical characterization of the *Saccharomyces cerevisiae* RING finger protein.

4.1.1 Protein Parameter Predictions

Preliminary computational, biophysical and biochemical studies of the *S. cerevisiae* RING finger domain first entailed acquiring the physico-chemical properties of the protein (Table 4.1) by using its sequence via the ProtParam tool of ExPASy web server. Figure 4.1A showed the primary sequence of the *S. cerevisiae* RING finger domain as composed of 84 amino acids making up a molecular weight of 9505.78 Da. Thirty-four of these amino acids are positively charged while 38 are negatively charged, thus making the protein negatively charged. In literature it is found that lysine, arginine and histidine individually comprise of 7.6, 9.7 and 10.8 isoelectric point respectively which are weak to strong basic amino acids that require a basic solution to avoid them forming conjugated acids (Damodaran, 2008). Aspartic acid and glutamic acid are acidic in nature with an isoelectric point of 2.8 and 3.2 respectively and like basic residues, they also need an acidic solution to avoid conjugated bases when they lose their protons (Mellon *et al.*, 2000). All these isoelectric points contribute to the average 6.68 theoretical pI of all the amino acids found in the *S. cerevisiae* RING finger domain. The molecular formula of the protein is $Carbon_{144}$, $Hydrogen_{665}$, $Nitrogen_{109}$, $Oxygen_{136}$ and $Sulphur_{5}$, hence the 1329 atomic composition. The extinction coefficient was found to be 250, assuming that all cysteine residues formed simply because this amino acid comprises of disulphide bonds that have the ability to absorb light. This protein was also found to have an estimated half-life of 3.5 hours in mammalian reticulocytes *in vitro*, 10 minutes in yeast *in vivo* and greater than 10 hours in *E. coli in vivo*.

The instability index was above 39, which is the limit for protein stability, since it was predicted to be at 41.98, which is moderately stable (partially unstable). The aliphatic index was found to be 84.64 and this reflected the volume of aliphatic side chains (Mbah *et al.*, 2012) which comprised of leucine (14.3%), valine (6.0%), alanine (2.4%) and isoleucine (2.4%) (Figure 4.1B).

Table 4.1: Physico-chemical properties of *S. cerevisiae* RING finger domain

ProtParam Parameters	Values
Number of amino acids	84
Molecular weight	9505.78
Theoretical pI	6.68
Atomic composition	Carbon 414 Hydrogen 665 Nitrogen 109 Oxygen 136 sulphur 5
Formula	$C_{414}H_{665}N_{109}O_{136}S_5$
Total number of atoms	1329
Extinction coefficient	250 Abs 0.1% (=1 g/l) 0.026, assuming all pairs of Cys residues form cystines 0 Abs 0.1% (=1 g/l) 0.000, assuming all Cys residues are reduced
Number of positively charged residues	34
Number of negatively charged residues	38
Estimated half-life	3.5 hours (mammalian reticulocytes, *in vitro*). 10 min (yeast, *in vivo*). >10 hours (*Escherichia coli*, *in vivo*).
Instability index	41.98
Aliphatic index	84.64
Grand average of hydropathicity (GRAVY)	-0.567

A

```
    10         20         30         40         50         60

HFKDLPDDLK CPLTGGLLRQ PVKTSKCCNI DFSKEALENA LVESDFVCPN CETRDILLDS

    70         80
LVPDQDKEKE VETFLKKQEE LHGS
```

B

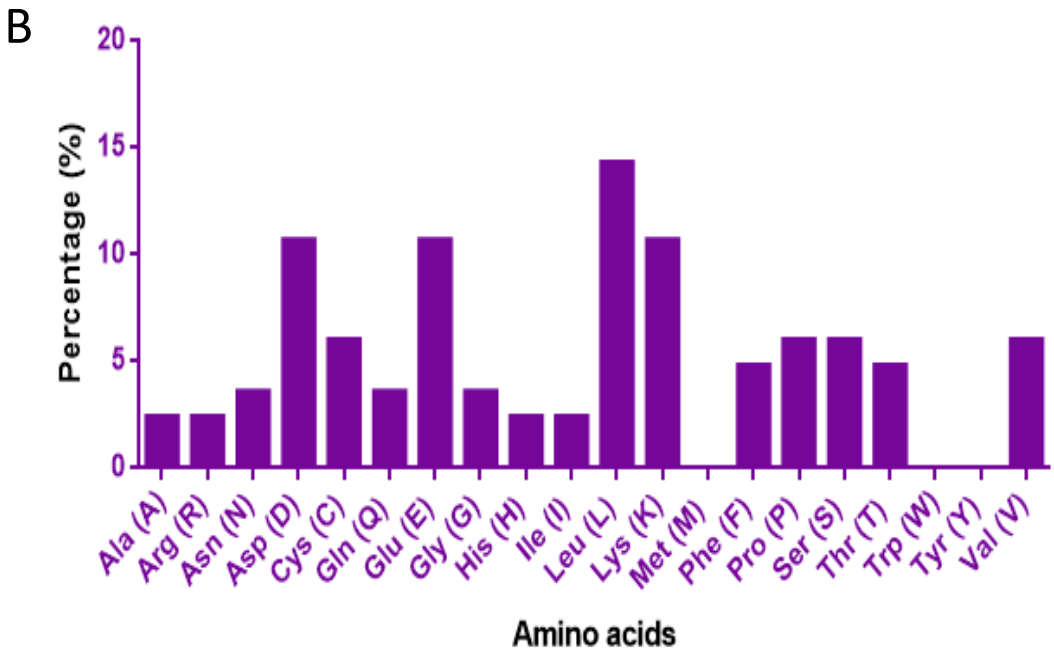


Figure 4.1A: Primary sequence for *S. cerevisiae* RING finger domain. The table shows the amino acids that make up the protein while **Figure 4.1B** Illustrates a bar graph demonstrating the different percentage yields from each individual amino acid found in the *S. cerevisiae* RING finger domain. It shows that leucine is the most abundant in the protein.

4.1.2 *S. cerevisae* RING finger domain structural predictions

The secondary structure of the protein was predicted using PSIPRED and was shown to have two α-helices, one 3_{10} helix and no β-sheets (Figure 4.2). In contrast, the cartoon structure (Figure 4.3) showed two anti-parallel β-sheets, two α-helices and one 3_{10} helix. Figure 4.4 shows the binding sites of the protein which are highlighted in red and comprise of five cysteine residues. The amino acids that contribute to the formation of β-sheets present in the *S. cerevisiae* RING finger domain are phenylalanine, isoleucine, valine and threonine, while the α-helix is composed of methionine, alanine, leucine, glutamic acid and lysine (Fioramonte *et al.*, 2012). It is known that α-helices are responsible for the stability of the protein and most of the amino acids involved are hydrophobic in nature, buried at the core of the protein; the reason of this being that α-helix N-terminals (N_1, N_2 and N_3) and C-terminals (C_1, C_2 and C_3) form hydrogen bonds which have a high impact on helical structure stability (Diod *et al.*, 2001).

The 3_{10} helix, also known as the A_{10} helix, is composed of 3-4 amino acids, most of which are aspartate, having a spacing of i and i+3 in the hydrogen bonds between the residues compared to the i+4 found in α-helices (Berk *et al.*, 2000). Under normal circumstances, 3_{10} helices are found at the C-terminals or N-terminals of the protein; however, in the *S. cerevisae* RING finger domain they were found between the two β-sheets. Only 3_{10} helices have hydrogen bonds needed for stability even though the strength is the same as that of the α-helix. Pal and co-workers (2002) postulated that this is because in most cases the 3_{10} helix is thought to be an extension of the β-sheets, having i and i+3 spacing to decide whether the residues will be β-sheet or 3_{10} helix. The role of 3_{10} helix hydrogen bonds are to regulate helix-coiling transitions which form part of the intermediates responsible for the folding and unfolding of the α-helix (Roger *et al.*, 2003).

The β-sheet is composed of 3-10 amino acids which make up 1-2 strands (Liu *et al.,* 2003). β-sheets can either be parallel or anti-parallel, but in this study it was observed that the protein contained antiparallel β-sheets. According to Perczel *et al.* (2005), proteins with an abundance of β-sheets tend to be resistant to proteolysis and are very stable (Perczel *et al.*, 2005). β-sheets originates from the polypeptide strand, joined with another polypeptide by hydrogen bonds. Therefore, the more the polypeptide extends, the more stable the structure becomes (Perczel *et al.*, 2005).

Lastly, the binding sites observed in the molecular surface protein structure (Figure 4.3) showed that one zinc ion is coordinated by four cysteine residues and the other zinc ion is coordinated by one cysteine residue where, according to the amino acids sequence (Figure 4.1A), two histidines are involved in the binding sites and missing residues are assumed to bind weakly.

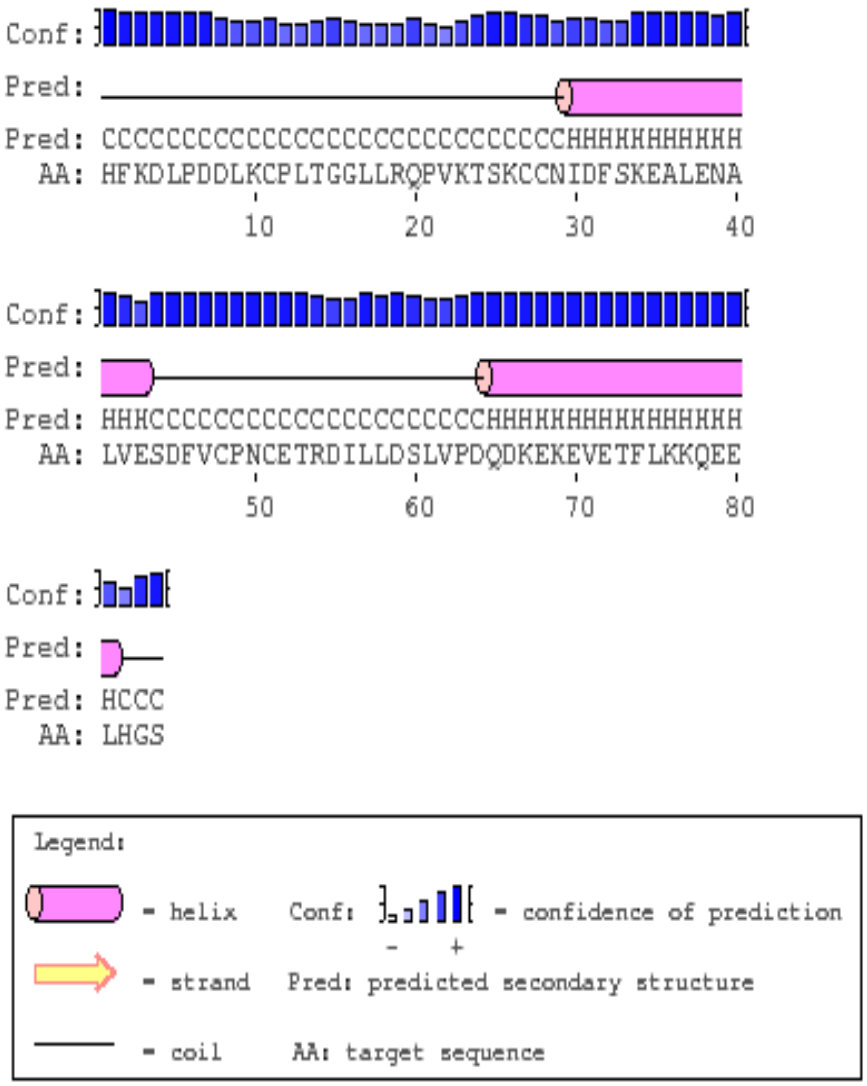


Figure 4.2: Secondary structure predictions demonstrating secondary structure elements and their configuration within the protein. The protein was predicted to contain 2 α-helices and no β-sheets.

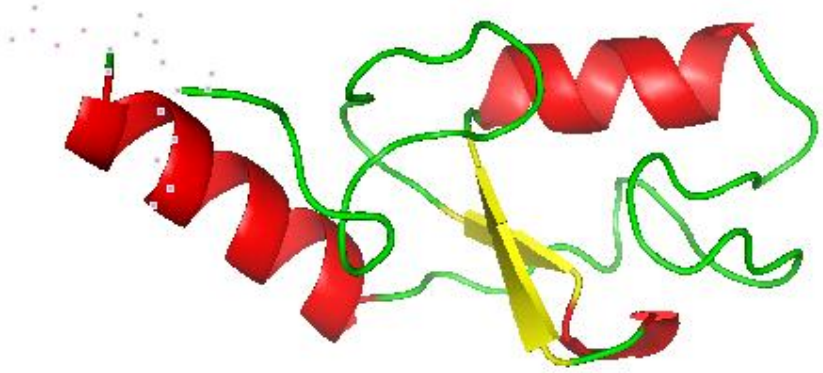


Figure 4.3: Cartoon structure of the *S. cerevisiae* RING domain displaying 2 α-helices (red colour), 2 β-sheets (yellow colour) and an A_{10} helix (short helix in red).

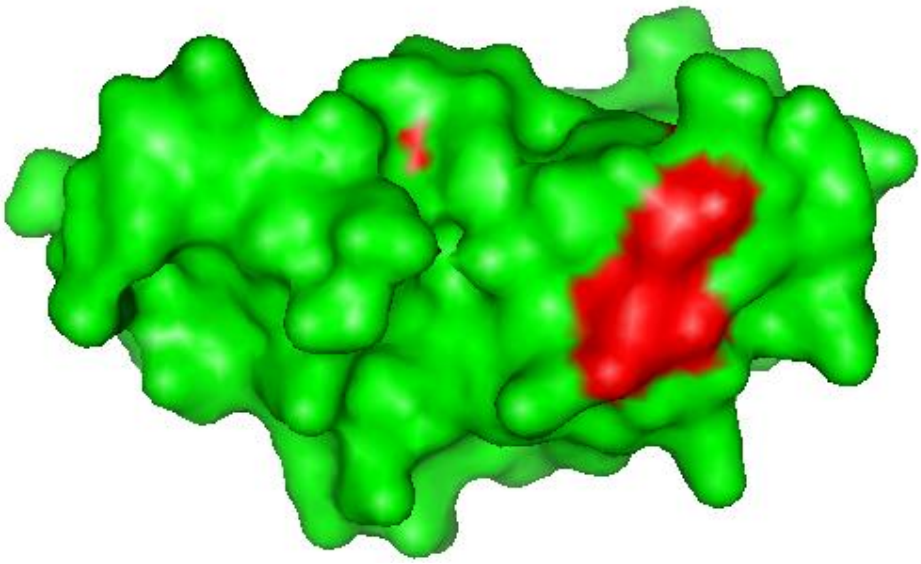


Figure 4.4: Molecular surface protein representation of binding pockets found in the *S. cerevisiae* RING finger domain.

4.1.3 Local quality estimate, Ramachandran plot and comparison of non-redundant set of PDB structures

The main purpose of the local estimation is to help identify potentially incorrect regions. The major role for estimating the local quality of the protein is to identify incorrectly predicted regions early so that they may be reported and corrected before being published (Benkert *et al.,* 2009). The essentiality of protein quality estimation is vital due to the fact that during secondary structural predictions many models are harvested and only the best convincing candidate is selected. In figure 4.5, the peaks observed show the quality of the protein secondary structure. Peaks above 0.8 and near to 1 meaning the desired protein is stable and rigid. This may be due to the abundance of leucine in α-helices and valine in β-sheets. Furthermore, the structure was compared to other models in the non-redundant set of the Protein Data Bank (PDB) where the quality standard of the protein is measured in z-score (Read *et al.*, 2011). Z-score represents a normal score normalised to mean 0 while standard deviation 1 indicates how many standard deviations the model QMEAN score differs from the expected values of experimental structures. The higher z-score is an indicative of a less accurate protein structure. Therefore, the position of the star (figure 4.6) shows that the protein model was accurate but still had a few errors (Read *et al.*, 2011). Lastly, a Ramachandran plot was constructed (a tool used to determine the quality of the model based on the phi and psi torsion angles present in the amino acids the protein is composed of). This showed that 88% (66) of the amino acids were present in the favoured region, meaning that the modelled structure was an acceptable representation of the protein and biologically informative. Residues in the allowed region were 8% (6) which is above the expected 2%, while residues in the outlier region were 4% (3).

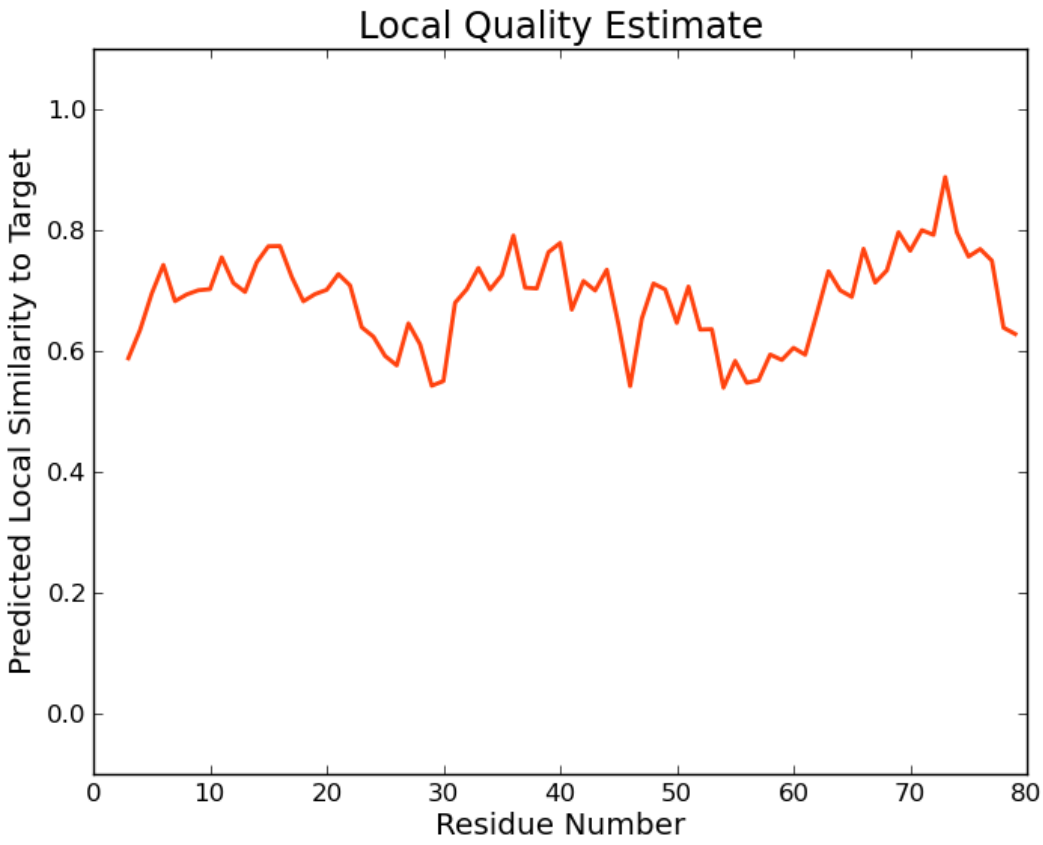


Figure 4.5: Local quality estimation which shows the rigidity of the protein. Peaks above 0.8 and nearest to one 1 indicate the rigidity of the protein.

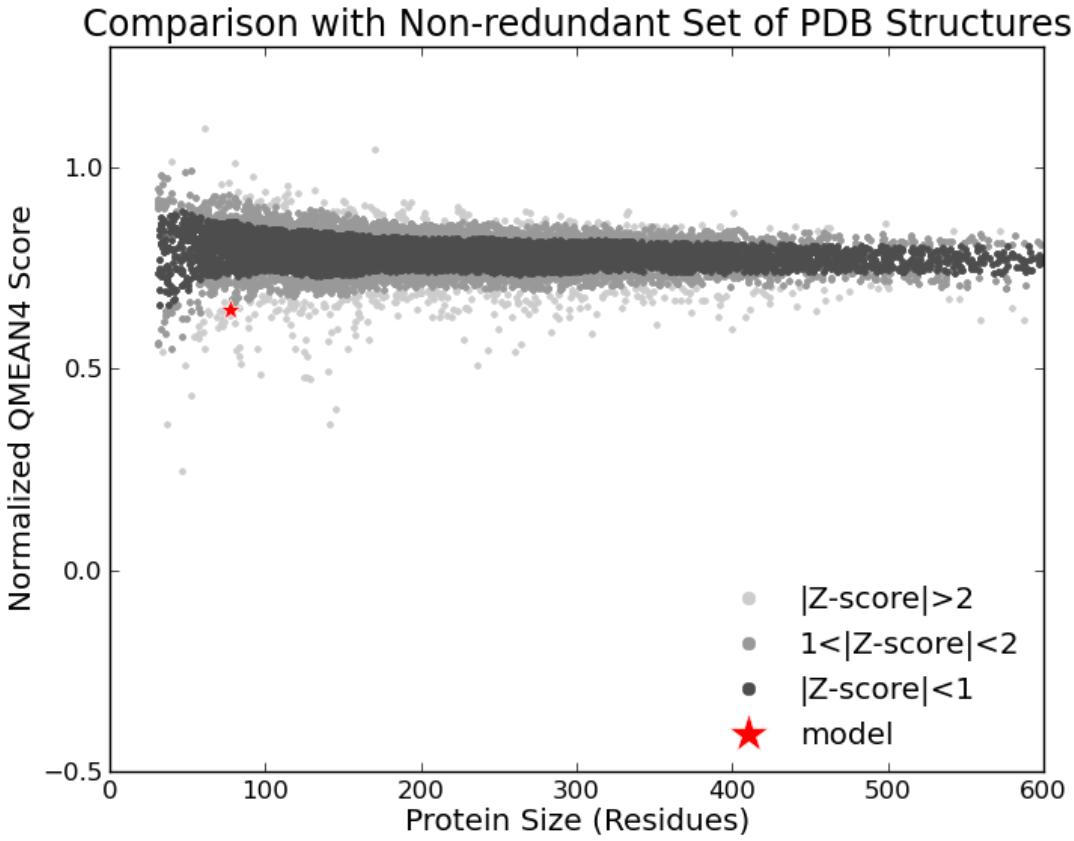
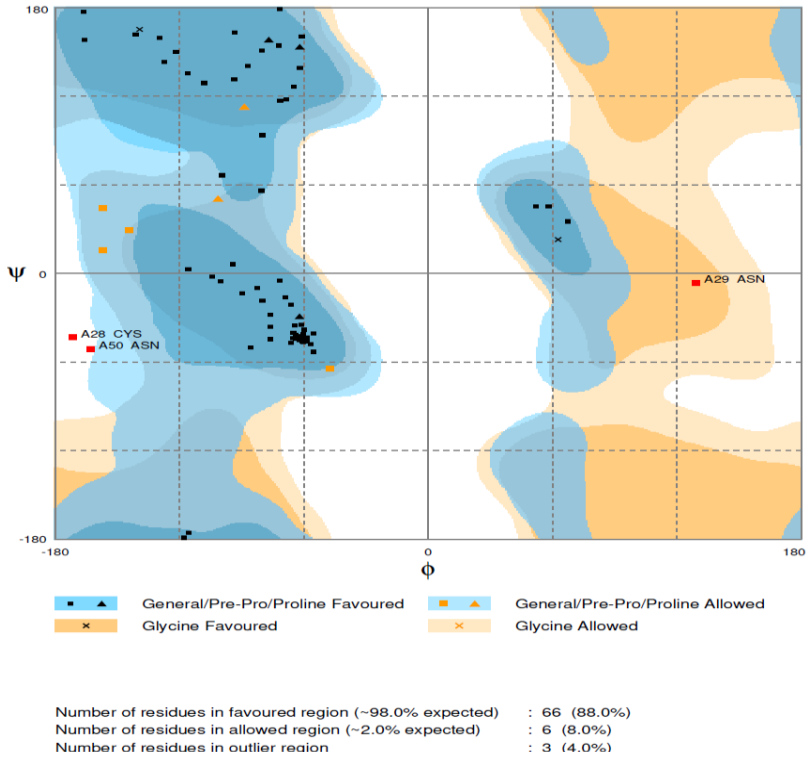


Figure 4.6: Representation of the non-redundant set of PDB structures which confirm whether the model predictions were accurate or not. The red star in |Z-score|>2 zone show that the predictions were not 100% accurate.

A



B

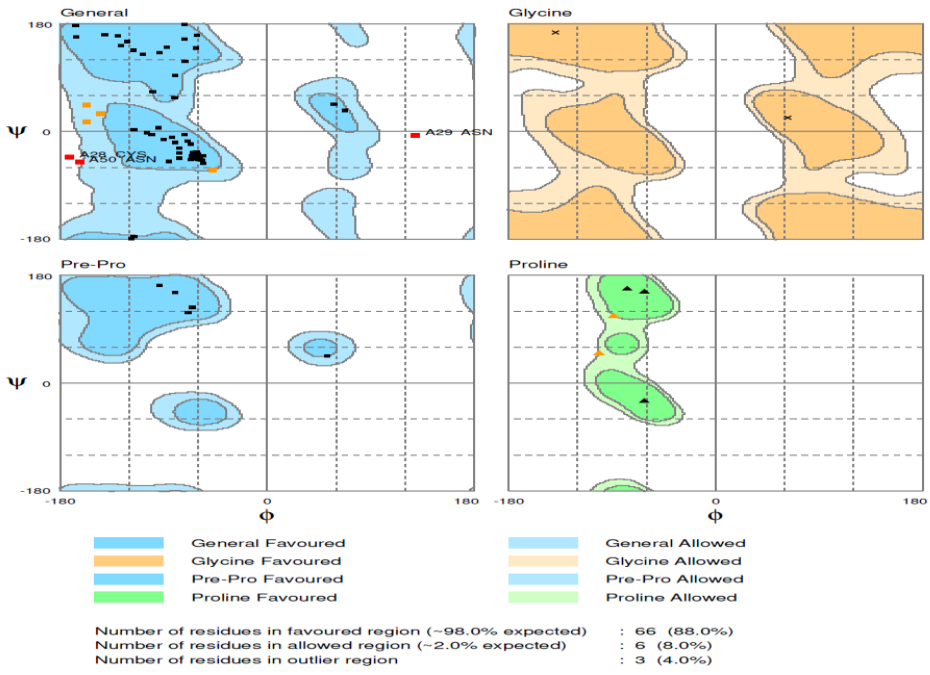


Figure 4.7: The Ramachandran plot (A and B) represent phi-psi torsion angles for all residues in the structure (except those at the chain termini). Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other side chain types and about 80% of the amino acids are at the favoured region meaning it a good prediction.

4.1.4 Analysis of protein-protein interactions

In drugs, vaccines or biosensor discovery, it is very important to know the interacting partners related to the protein of interest. *S. cerevisiae* RING finger domain was used as bait to identify a maximum of ten interaction partners and thereafter trace any biological significance of their interaction (see Figure 4.8). Most of the proteins found to putatively interact with the *S. cerevisiae* RING finger domain (CFT2, CFT1, YSH1, RNA15, SWD2, Ref2 and GLC7) seemed to all have the same function of being involved in polyadenylation and pre-mRNA/snomRNA cleavage. CFT2 (Cleavage Factor 2) is a subunit factor responsible for the cleavage of mRNA 3'-end from yeast by binding to RNA (depending on the availability of RNA). This is a similar role observed in cleavage and polyadenylation specificity factor (CPSF) in mammals (Zhao *et al.*, 1997).

Cleavage Factor 1 (CFT1) is also involved in cleavage and polyadenylation; however, unlike CFT2, CFT1 plays a role in making sure the integrity of the genome is maintained to prevent hindrances during its replication (Zhao *et al.*, 1997). It has been shown that in the presence of RNA polymerase, which prevents stability in the genome during cell cycle, CFT1 acts as an inhibitor to avoid DNA damage (Gaillard and Aguilera, 2014). The process of polyadenylation and mRNA cleavage involves many subunits playing different roles.

Yeast seventy-three homology, commonly known as YSH1, is a subunit that functions as a carrier of a highly conserved β-lactamase found in metal-dependent hydrolysis enzymes (Garas *et al.*, 2008). The β-lactamase protein plays a crucial role in ascertaining the formation of the 3'-end in endonucleolytic cleavage of pre-mRNA and also acts as a splicing factor (Garas *et al.*, 2008).

Another subunit which plays the same role as the aforementioned subunit is RNA15 (poly(A) mRNA metabolism), but this function occurs only in the presence of Hrp1 and

RNA14 that assists the protein to interact with an A-rich element (Gross and More, 2001). Unlike the other subunits that are more focus on the mRNA, SWD2 (Set1C, WD40 repeat protein) is known to be involved in snoRNA termination by being a component of adenine phosphoribosyltransferase (APT). Its absence is related to APT recruiting hindrances and termination of snoRNA (Soares and Buratowski, 2012). Furthermore, Ref2 (RNA end formation 2), an intermediate of APT, is responsible in assuring permanently secured stability within the complex that functions in the maturation of mRNA 3' end before polyadenylation (Ferrer-Dalmau, 2010). Ref2 was also observed to have a close relation with GCL7 (Glycogen deficient), as they are both involved in cation homeostasis, even though Ref2 is more sensitive and is known to specifically bind non-covalently to chromatin (Ferre-Dalmau, 2010). GLC7 is a gene that encodes a catalytic subunit of protein phosphatase (PPases) that regulates biological processes such as glucose and glycogen metabolism, sporulation, chromosome segregation, meiosis, mRNA transport, transcription and amino acid biosynthesis (Logan *et al.*, 2008).

Conversely, the other proteins (namely SSU72, PTI1 and YTH1) that were harvested by the STRING protein interaction software related to the RING finger domain were found to be up-regulated in cancer aside the roles they play in cleavage and polyadenylation. SSU72, which stands for suppressor of sua7-1 clone 2, is a gene that was discovered during suppressors of sua7-1 (cold-sensitive mutation in yeast TFIIB) screening studies, hence its name (Sun *et al.*, 1996). It is famously known as a cleavage and polyadenylation factor (CPF) component that regulates transcription *in vitro* (Krishnamurthy *et al.*, 2004). It has been found to be highly expressed in embryogenesis, specifically in the nervous system and intestines; however, in adult and human it is up-regulated in multiple tumour cell lines (St-Pierre *et al.,* 2005). In addition, prostate tumour inducing gene 1 (PTI1) is an essential gene encoding a predicted 425 amino acid protein conserved in eukaryotes. It is also

known as an intermediate needed for the maturity of snoRNA that is implicated in pre-mRNA 3' end processing complex (Dheur *et al.*, 2003). It is a truncated form of EE1A1 (elongation factor 1α1) protein that has been observed in prostate carcinoma patient-derived blood samples (Rehman *et al.*, 2012). Lastly, YTH1 (Yeast THirty kDa Homolog) is a 24kDa subunit of CPF containing C3H zinc finger type hence it is involved in cleavage and polyadenylation. It is composed of five types of zinc fingers (ZF1-5) conserved at the core with a motif of cysteine and histidine residues. Proteins with similar motifs are known to be responsible for RNA binding and stability in various organisms and also involved in facilitating protein-protein interactions (Tacahashi *et al.*, 2003). Further studies concerning YTH1 revealed that ZF2 absence affects the *in vitro* cleavage of yth1-4 while lack of ZF4 prevents the interaction of YTH1 with Fip1 (protein responsible for polyadenylation). However, when YTH1 interacts with Fip1 it prevents YTH1 binding to RNA. Moreover, it has been shown that ZF1 participates in double stranded RNA binding and also functions as a metal-sensing domain. ZF2-4 is involved in DNA binding (Tacahashi *et al.*, 2003). The domain involved for the interacting abilities of YTH1 to Fip1 is ZF4 while ZF5 is more responsible for the strength of the interaction (Tacahashi *et al.*, 2003).

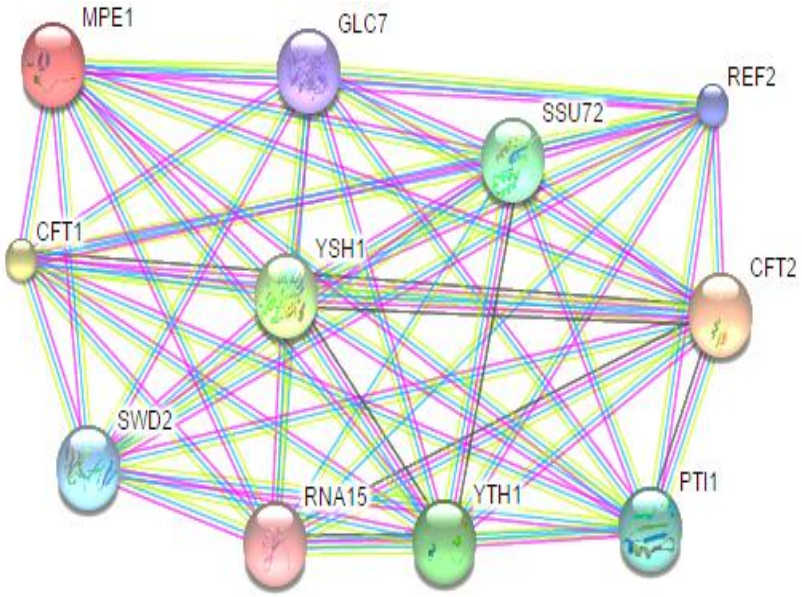


Figure 4.8: STRING diagram showing different types of proteins that interacts with the *S.cerevisiae* RING finger protein. *S.cerevisiae* RING finger is represented by Mpe1, a homology of RBBP6. The proteins with structures indicate the characterized protein whereas the uncharacterised proteins are the one with no structures.

4.1.5 Phylogenetic tree and Sequence Alignment

The evolutionary status of a desired protein may be depicted pictorially by constructing a phylogenetic tree. Amongst all the proteins represented in the phylogenetic tree (Figure 4.9), XP_003680012.1, XP_003959084.1 and XP_001644384.1 have not been characterised but are known to play a huge role in ubiquitin-protein transferase activities, zinc ion binding and pre-mRNA binding. They are also located in the cytosol. More so, looking at the other characterised proteins (SCU81809.1, SCW02477.1, CEP63933.1, SCV03645.1 and SCV05680.1), it was observed that they are involved in molecular functions such as binding nucleic acids and zinc ions, and also in biological functions like ubiquitination (by transferring ubiquitin). However the construction of the phylogenetic tree would have been impossible without going through a process called multiple sequence alignment (figure 4.8), which viewed conserved residues within the selected proteins; these are highlighted in grey. The least conserved residues and dotted lines seen in the figure represent deletions or possible mutations that may have taken place during evolutionary processes.

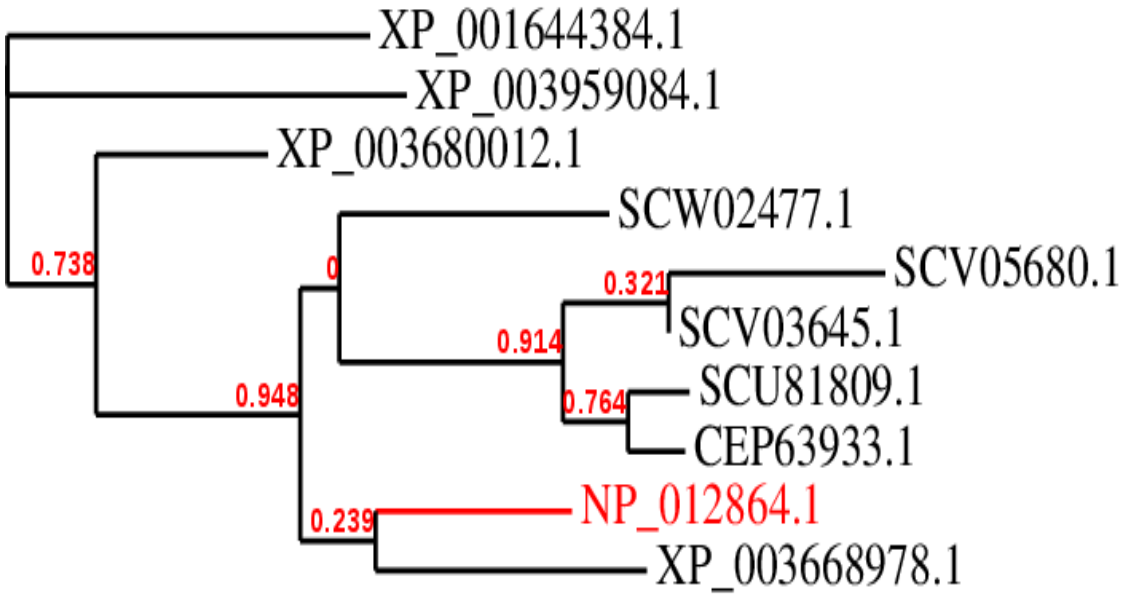


Figure 4.9: Phylogenetic tree showing the evolution of the *Saccharomyces cerevisiae* RING finger domain and homologous proteins. Accession number in red indicates the protein of interest while the numbers on top of the braches represent their length size which is proportional to the number of substitutions per site.

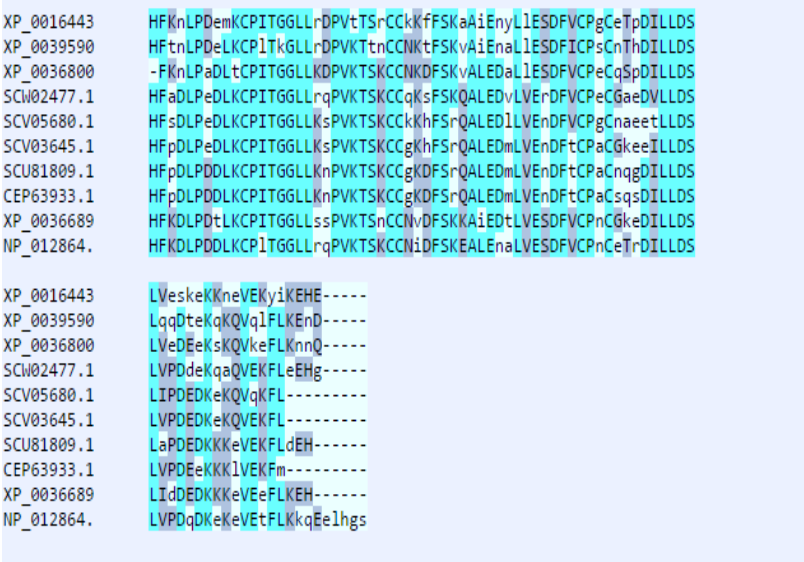


Figure 4.10: Sequential alignment showing the conserved residues in different proteins that were shown to be related to the *S. cerevisiae* RING finger domain. The blue represents highly conserved residues within the proteins while grey represents the lowest conserved residues.

4.2 Bacterial transformation, expression and purification of *S. cerevisiae* RING finger domain

After performing the above bioinformatics studies, which assisted in predicting certain characteristics of the protein of interest, it was imperative to conduct bench top (wet lab) studies to confirm the results provided by the *in silico* studies. The first step was to prepare competent cells from *E. coli* BL21 cells. The *S. cerevisiae* RING finger domain plasmid was then incorporated into the cells through a process called transformation. The results in figure 4.11 showed that the bacterial transformation was a success due to the single colonies present on the test plate and no growth on the control plate. Thereafter, individual colonies were picked and grown, with the use of 0.5mM IPTG and ampicillin, through a process known as expression screening. Figure 4.12 shows that colony 4, on the SDS-PAGE gel depicted an intense protein band as compared to the others, confirming that this particular colony was capable of producing more of the desired *S. cerevisiae* RING finger protein. Glycerol stock made from this colony was then used to produce more of the protein in 6L of LB broth (large-scale protein expression) as shown in figure 4.13. Having produced more protein, purification of the protein in a GST-column then followed as shown in figure 4.14. The SDS-PAGE gel shows that the *S. cerevisiae* RING finger domain protein eluted in the first elution (Lane E1).

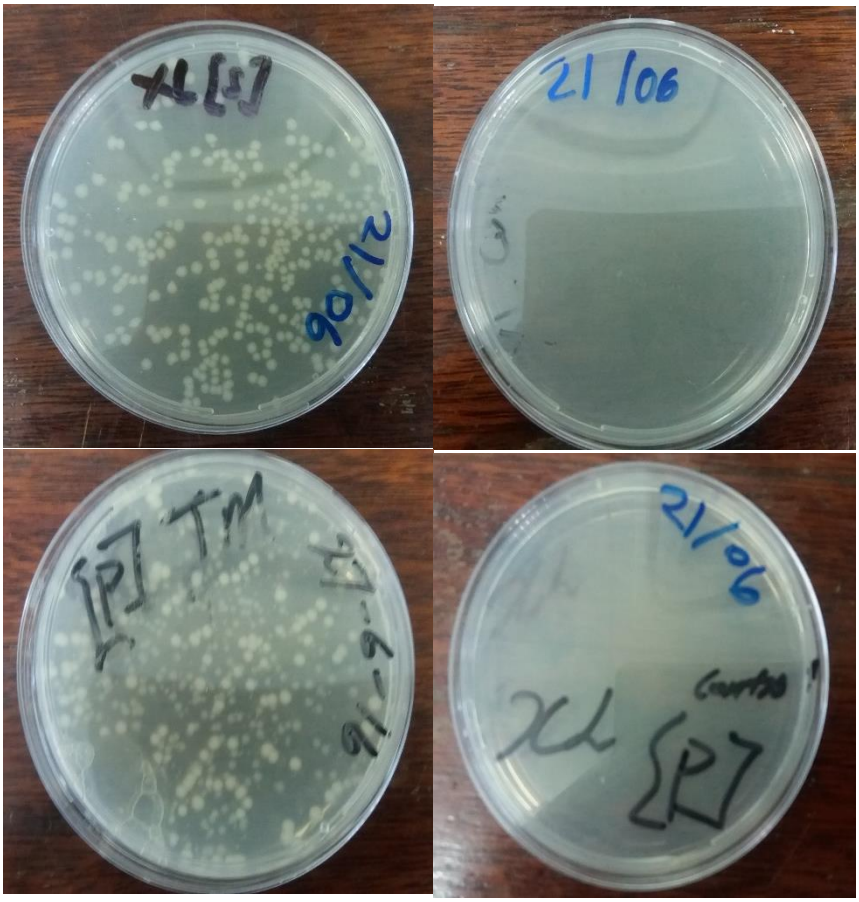


Figure 4.11: Transformation plates of the pGEX-6P-2-*S. cerevisiae* RING plasmid in *E. coli BL21* cells. The petri dishes with colonies represent transformed cells from both the supernatant and pellet with their controls beside them represented by petri dishes with no growth.

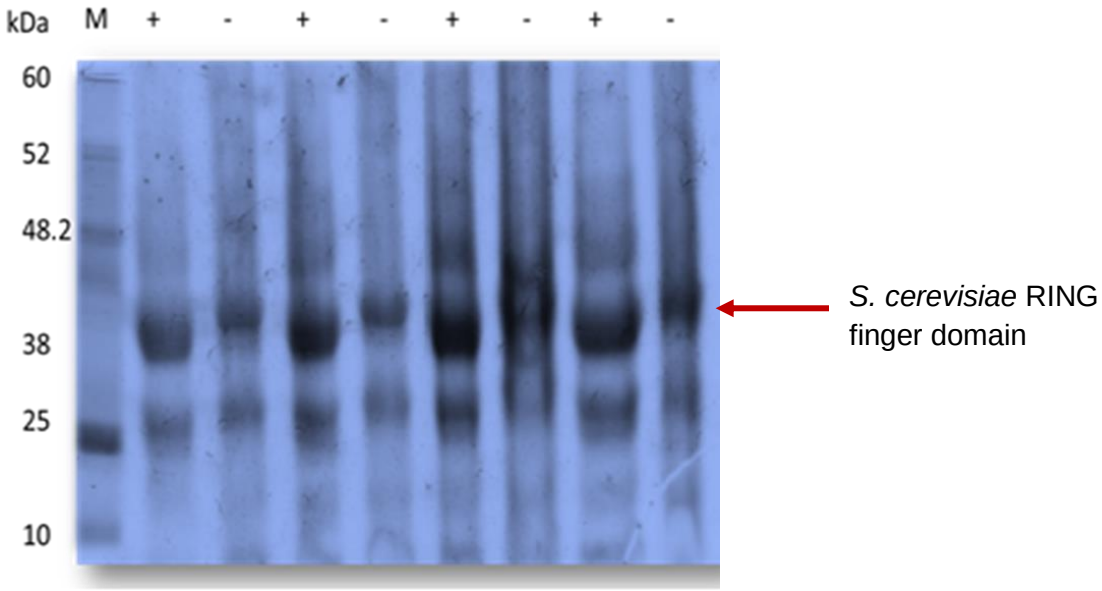


Figure 4.12: Expression screening where **M** stands for marker, **+** for samples induced with IPTG, **–** for samples uninduced and an arrow showing the protein of interest.

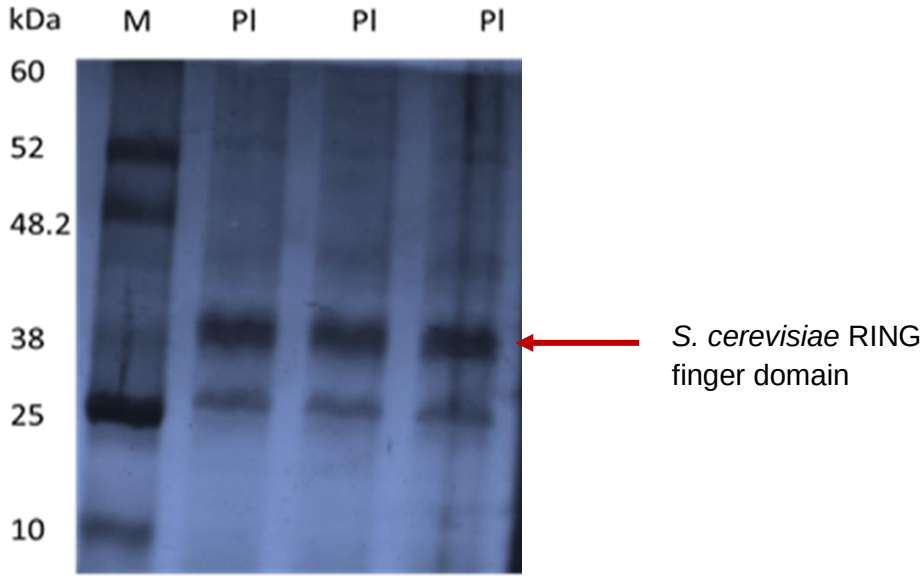


Figure 4.13: Large-scale expression where **M** stands for marker, **Pl** for protein lysate and the red arrow showing the protein of interest.

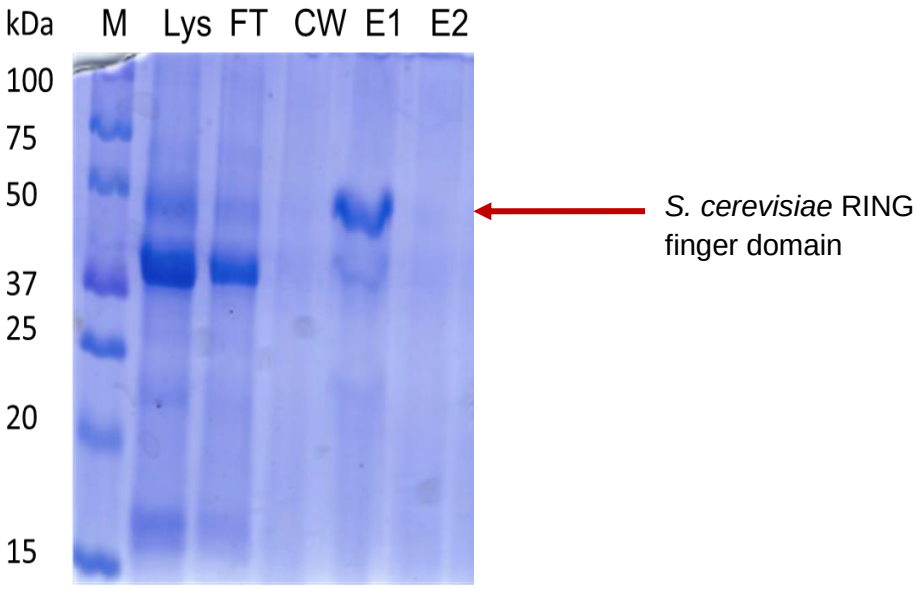


Figure 4.14: The *S. cerevisae* RING fusion protein SDS-PAGE purification. The **M** stands for marker, **Lys** for protein lysate, **FT** for flow-through, **CW** for clean wash, **E** for elution 1-2.

4.3 Biophysical characterization of *S. cerevisiae* RING finger domain

Protein absorbs light at a 280nm wavelength in any spectrophotometer used for analysis. This possibility is through aromatic amino acids such as tryptophan, tyrosine and phenylalanine, based on their strength of absorbance. Cysteine residues also absorb light by forming disulphide bonds with each other. In this study, UV spectrophotometer analysis was used to measure the protein ability to absorb light. As shown in figure 4.13, the protein was shown to have a high absorbance ability judging by the height of the peaks. NanoDrop ND2000 spectrophotometer was used to determine the concentration of the protein which was measured to be 5.44mg/ml (figure 4.14).The importance of this technique was to ensure that protein NMR studies were conducted at a suitable concentration. Thereafter, FTIR was done to measure the *S. cerevisiae* RING finger protein light intensity, its ability to absorb infrared radiation as well as the wavelength. In figure 4.15, an amide I is present between 1600-1700 cm^{-1} which is the most sensitive part of the carbonyl (c=o) group. The aliphatic stretch was observed between a 2500-3000 cm^{-1} wavelength and lastly, the single bonds of the amine group N-H and O-H stretch were observed between a 3000-3500 cm^{-1} wavelength in the FTIR spectrophotometer.

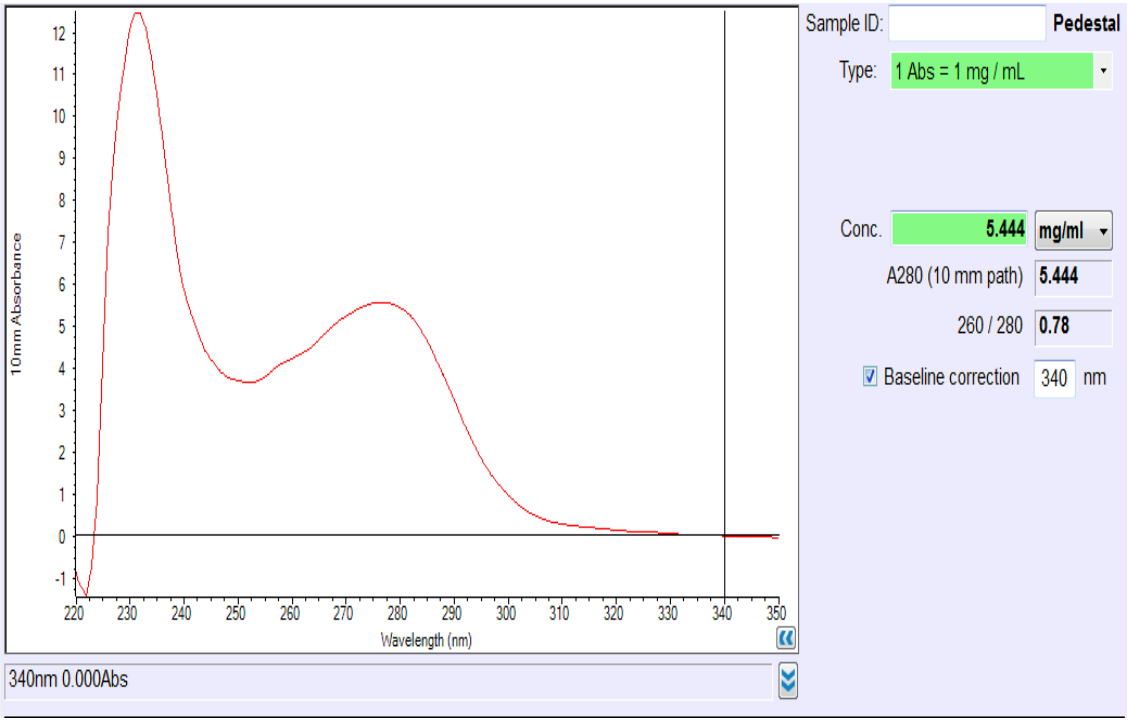


Figure 4.17: NanoDrop results showing protein concentration which was found to be 5.4mg/ml.

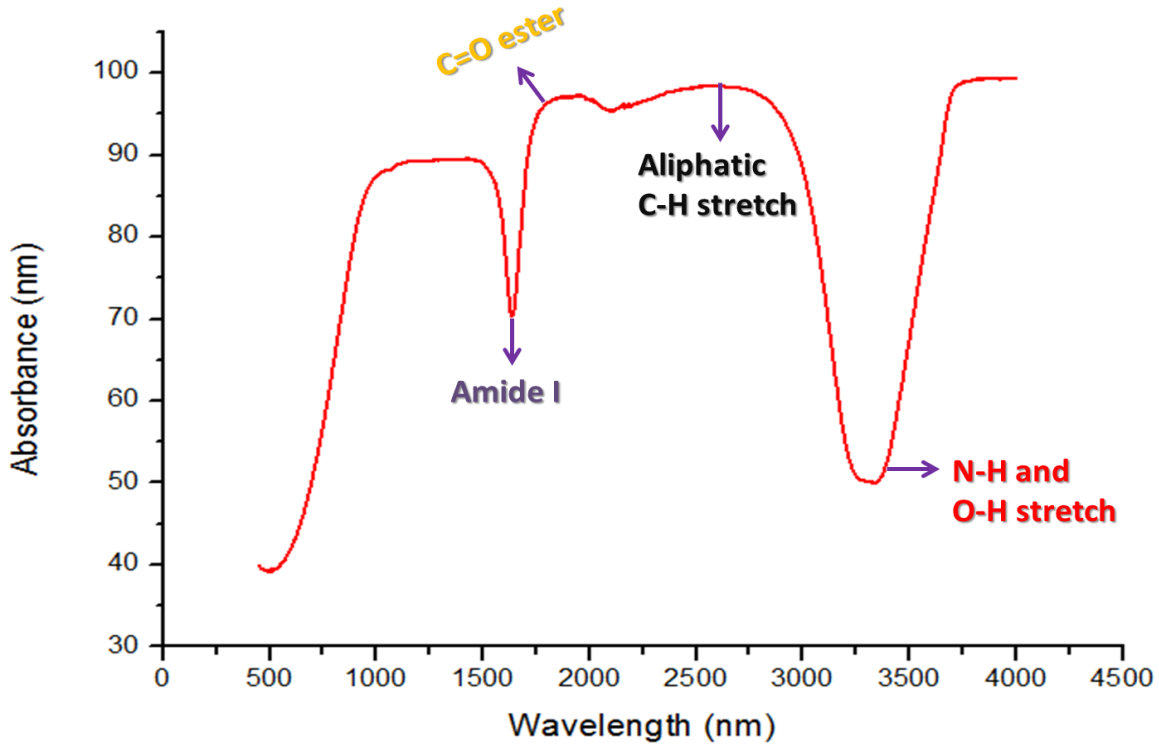


Figure 4.16: FTIR spectrum showing the secondary elements of the protein to elucidate molecules that absorb light. The protein absorption ability specifically corresponds to bonds present in the molecules.

4.4 Mass spectrometry and NMR spectroscopic analysis of *S. cerevisiae* RING finger domain

It was imperative to perform mass spectroscopy analysis for purity determination and mass per charge ratio of the *S. cerevisiae* protein. This information was necessary to initiate NMR studies. As shown in figure 4.18, mass spectrometry was done and revealed that the protein was well purified. However, the mass spectrogram showed different peaks, this was evidence that the protein was undergoing proteolysis due to the fact that these peaks represent various species and shortened fragments of the protein. For the mass per charge ratio identification, a method known as *de novo* sequencing was used where residues were assigned from fragmented peptide masses of the protein. Thereafter, 1D NMR was done to determine whether the protein was folded or not; this knowledge was essential for structural studies. For this analysis, NMR spectroscopy was used which revealed that the protein was well folded according to results shown in figure 4.19. This was verified by the dispersion of chemical shifts more specifically in the methyl region between 1.0 to -1.0 ppm and the amine region at 8.5 ppm (Scheich *et al*, 2004).

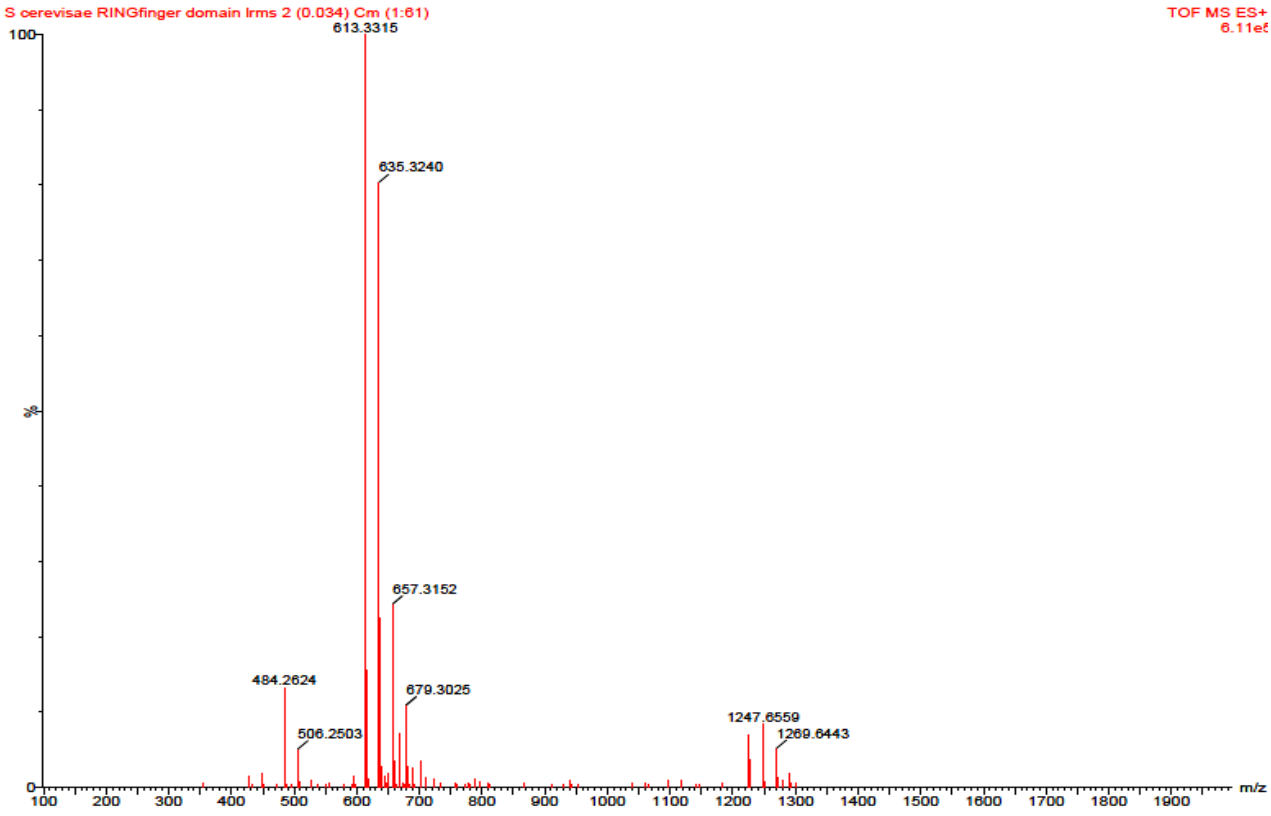


Figure 4.18: TOF-Mass spectrometry which shows that the protein was well purified, confirmed by the dispersion of peaks. The different peaks represent different species due to the protein undergoing proteolysis.

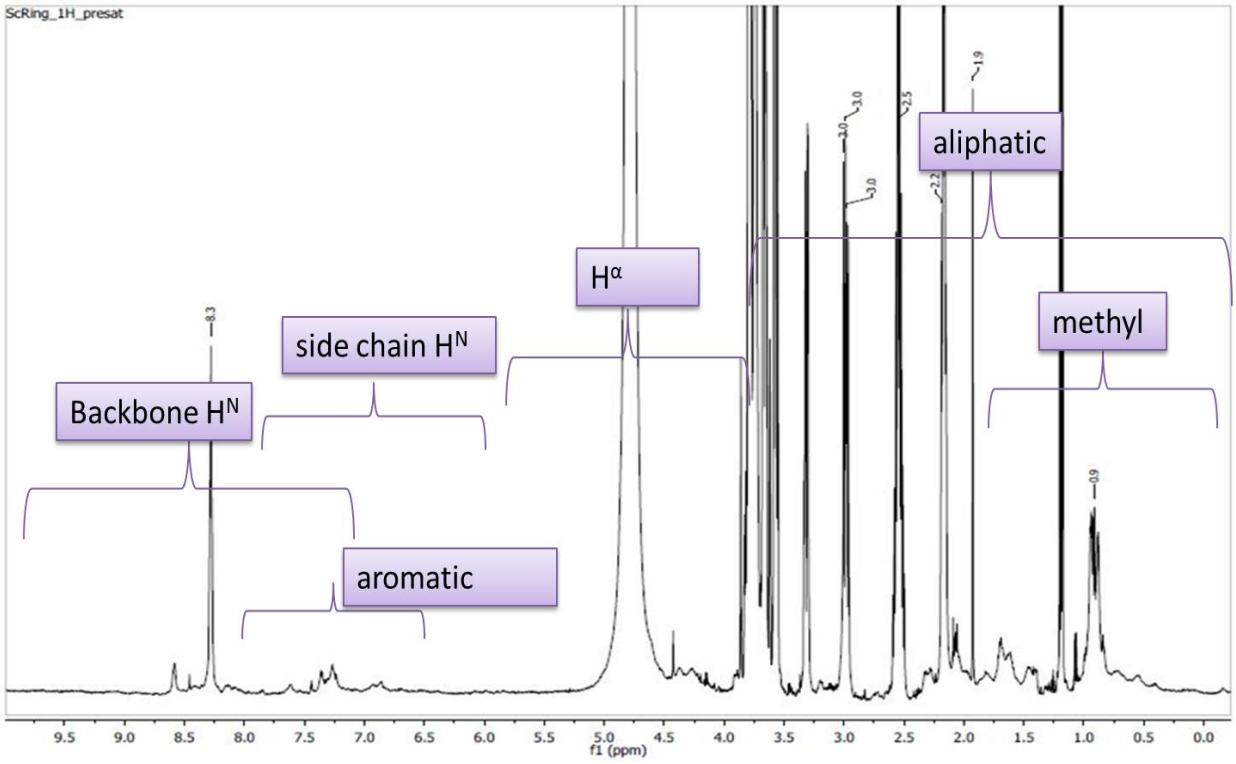


Figure 4.19: 1D H NMR for *S. cerevisiae* RING finger domain which shows a well-folded protein. This is confirmed by the appearance of amide protons (H^N) in the 6.5-9.4ppm region, α-protons (H^α) in the 4.5-6.5ppm region, and methyl protons in the 0-3.5 ppm region.

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSION

5.1 General discussion

The results obtained from the secondary structure predictions correlate with already known information regarding the properties of the *S. cerevisiae* RING finger domain. The theoretical pI of the protein (pH of 6.68) whose value was obtained prior to the bench work, made it feasible for protein purification and preparation of appropriate buffers (Damodaran, 2008). The presence of cysteine disulphide bonds situated at the hydrophobic core contributed greatly to the stability of the protein. The secondary structure predictions revealed the secondary elements present in the *S. cerevisiae* RING finger domain. These include two α-helix, two β-sheets and one 3_{10} helix. The helices are responsible for the stability of the protein due to their C- and N-terminals which form hydrophobic bonds. The β-sheets play a role in the rigidity of the protein by preventing unfolding in its backbone while the 3_{10} helix regulates the folding and unfolding of the α-helix to ensure that it always remains intact for optimum function (Diod *et al.*, 2001). Thereafter, a Ramachandran plot was done to confirm the accuracy of the predictions. The plot showed that the predictions were good since about 88% of the total number of residues appeared in the favoured region and a meagre 8% were present in the allowed region (Abdullah *et al.*, 2015).

STRING protein-protein interaction was done to investigate the identity of interacting partners of the *S. cerevisiae* RING finger domain and examine any relationships or similar functions between them and the *S. cerevisiae* RING finger domain. CFT2, CFT1, YSH1, RNA15, SWD2, Ref2 and GLC7 (Ferrer-Dalmau, 2010; Gross and Moore, 2001; Soares and Burataloski, 2012; Tacahashi *et al.,* 2003; Zhao *et al.*, 1997) were among those proteins discovered as interacting partners and exhibited similar functions to those associated with *S. cerevisiae* RING domain such as involvement in polyadenylation and mRNA/snoRNA cleavage. These functions were also observed in Mpe1, a homologue of

RBBP6 that contains a RING finger domain; this may possible mean that the aforementioned proteins have a RING finger domain in their structures. Chromatin binding is also synonymous with RING finger domain function and is an interacting promoter of protein-protein and protein-DNA. In contrast, other proteins obtained from the STRING server namely SSU72 (Krishnamurthy *et al.*, 2004), PTI1 (Dhaur *et al.*, 2003) and YTH1 (Tacahashi *et al.*, 2003), are known to be up-regulated in cancer, in addition to the roles they play in cleavage and polyadenylation. Looking specifically at YTH1, it was observed that out of all the zinc fingers found in its composition, ZF4 shows more similar function to that of the RING finger domain by acting as an interacting promoter and having its location at the C-terminal of YTH1, which is the same location of the RING finger in RBBP6 and its homologues.

The multiple sequence alignment and the phylogenetic tree revealed the evolutionary and ancestral relationship of different proteins to that of the *S. cerevisiae* RING finger domain. Both were able to show how certain residues in the different protein sequences evolved and had similar biological and molecular functions such as being involved in zinc ion binding, ubiquitin-protein transferase and binding pre-mRNA. The RING finger domain is an independent self-replicating and self-expressing protein found incorporated with other domains to form a large protein known as RBBP6. This protein has been noted in most studies for its possible role as an anticancer therapeutic antidote according to Yoshitake and co-workers (2004), where they discovered PACT (a homologue of RBBP6), which was highly up-regulated in oesophageal cancer. Therefore, this study entails both the biochemical and biophysical characterisation of whether the first missing hallmark residue in the *S. cerevisae* RING finger domain had any biological significance in protein function which requires structural determination by NMR spectroscopy.

The first objective was to recombinantly express the *S. cerevisae* RING finger domain in *E. coli* BL21 cells. The reason this microorganism was chosen as a host was because of its advantageous properties which include a fast growing rate that has the potential to yield a high density of culture cells, ready availability, cheap reagents to make growth media and lastly, its easy acceptance of foreign DNA for a fast transformation process (Rosano and Ceccarelli, 2014). Having successfully transformed the plasmid onto LB-Ampicillin plates, colonies were randomly picked for expression screening to determine which one of these colonies would produce more of the desired protein. This was determined by the size of the band after running an SDS-PAGE gel. The colony that produced more of the protein as compared to other colonies was used for large-scale expression, followed by successful extraction to release the recombinant protein. The lysate recovered was run on an SDS-PAGE gel, which confirmed the presence and size of *S. cerevisiae* RING finger protein.

However, the protein was not pure since the lysate still contained other proteins apart from the protein of interest. Hence, the protein was purified using a GST column to recover the GST-tagged fusion protein. According to Harper and Speicher (2011), the GST-tag has solubility enhancers which maintain the solubility of the desired protein, making it easier to work with. Affinity purification showed about 99% purified *S. cerevisiae* RING finger protein. However, because the GST-tag was still attached to the protein, further purification was done by cutting off the tag to recover a tag free protein for biophysical characterisation (Rosano and Ceccarelli, 2014).

It was important to determine how much light and at what intensity the protein was able to yield radiation rays. Therefore, UV absorption was done and it was discovered that the protein of interest does not absorb much light since it lacks tryptophan and tyrosine in its primary sequence (McLaren and Shugar, 2014). However, the little light the protein is able to absorb may be due to the presence of phenylalanine (an aromatic residue), whose

percentage was very small. It was also assumed that the cysteine produced forms disulphide bonds that are able to absorb light. NanoDrop was then done to estimate the concentration of the protein which was found to be 5.44mg/ml, which was adequate to perform structural studies (Desjardins *et al.*, 2009). FTIR revealed the secondary structural elements within the *S. cerevisiae* RING finger domain. In this technique, it was observed that Amide I, which is the most sensitive part of the protein showed the presence of the carbonyl group, while Amide II showed the presence of the amine group (Coates, 2000). These two functional groups play a huge role in protein structural integrity and stability (Abidi *et al.*, 2014).

The main aim of this study was to perform structural characterisation using NMR spectroscopy. Prior to that however, mass spectrometry to determine the purity state of the protein and the mass per charge ratio, was done (Pugh, 2005). The results showed that the protein was well purified due to the clear dispersion of the peaks. Unfortunately, the results also revealed that the peaks observed did not correspond to the molecular weight of the protein which may be due to proteolysis. The 1D NMR studies revealed that the protein of interest was folded, judging by the dispersion of chemical shifts, specifically in the methyl region between -1 and +1 ppm and also by the peaks observed in the amine region around 8.5ppm (Scheich *et al.*, 2004).

5.2 Conclusion

In conclusion, this study has shown that the *Saccharomyces cerevisae* RING finger domain can be recombinantly expressed in bacteria, purified adequately using affinity purification methods and biophysical parameters showed the optical properties of the protein. Above all, NMR analysis of the protein showed it is very well folded and can be amenable to structural determination by NMR spectroscopy.

5.3 Future work

Future studies will entail double labeling the *S. cerevisiae* RING finger domain in order to acquire triple resonance data in the quest to perform both backbone and side-chain resonances assignment in order to solve the solution structure of the protein. The biological significance of the loss of the first zinc binding site bioinformatically observed in the *S. cerevisae* RING finger domain would thereafter be determined using exchange experiment with cadmium.

REFERENCES

Abidi N, Cabrales L and Haigler CH (2014). Changes in the cell wall and cellulose content of developing cotton fibers investigated by FTIR spectroscopy. *Carbohydrate Polymers* **100**: 9-16.

Ahmed S, Palermo C, Wan S and Walworth NC (2004). A novel protein with similarities to Rb Binding protein 2 compensates for loss of Chk1 function and affects histone modification in fission yeast. *Molecular and Cellular biology* **24** (9): 3660-3669.

Aravind L and Koonin EV (2000).The U-box is a modified RING finger- a common domain in ubiquitination. *Current biology* **10** (4): 132-134.

Aravind L, Iyer LM and Koonin EV (2003). Scores of RINGS but no PHDs in ubiquitin signalling. *Cell Cycle* **2**: 123–26.

Armen R, Alonso DO and Daggett V (2003). The role of α-, 310-, and π-helix in helix→ coil transitions. *Protein Science* **12** (6): 1145-1157.

Bach I (2000).The LIM domain: regulation by association. *Mechanism of Development* **91**: 5-17.

Baker AL, Allis CD and Wang GG (2008). PHD fingers in human diseases: Disorder arising from misinterpreting epigenetic marks. *Mutation Research* **647**: 3-12.

Barlow PN, Luisi B, Milner A, Elliott A and Everett R (1994). Structure of the C3HC4 domain by ^{1}H-nuclear Magnetic Resonance Spectroscopy. Journal of Molecular Biology **237**: 201-211.

Bellamy COC (1996). p53 and apoptosis. *British Medical Bulletin* **53** (3): 522-538.

Berk A, Zipursky S and Lodish H (2000). Section 3.1 Hierarchical Structure of Proteins. *Molecular of Cell Biology 4th edition*.

Bienz M (2006). The PHD finger, a nuclear protein-interaction domain. *Trends in Biochemical Sciences* **31** (1): 35-40.

Bieri M, Kwan AH, Mobli M, King GF, Mackay JP and Gooley PR (2011). Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. *FEBS Journal* **278:** 704-715.

Billeter M, Braun W and Wu¨ thrich K (1982). Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Computation of sterically allowed proton–proton distances and statistical analysis of proton–proton distances in single crystal protein conformations. *Journal of Molecular Biology* **155**: 321–346.

Bock CW, Katz AK, Markham GD and Glusker JP (1999). Journal of American Chemical Society **77**: 2777.

Borden KL and Freemont PS (1996). The RING finger domain: a recent example of a sequence-structure family. *Current Opinion Structural Biology* **6**: 395–401.

Borden KLB, Boddy MN, Lally J, O'Reilly NJ, Martin S, Howe K, Solomon E and Freemont PS (1995). The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO Journal* **14** (7): 1532-1541.

Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-cycle binding. *Analytical Biochemistry* **72**: 248-254.

Callis J, Stone SL, Hauksdottir H, Troy A, Herschleb J and Kraft E (2005). Functional analysis of the RING-type ubiquitin ligase family of *Arabidopsis*. *Plant physiology* **137**: 13-30.

Cannon JF (2010). Function of protein phosphatase-1, Glc7, in Saccharomyces cerevisiae. *Advances in Applied Microbiology* **73**: 27-59.

Cao W and Garcia-Blanco M (1998). Nucleic acids, protein synthesis, and molecular genetics. *Journal of Biological Chemistry* **273**: 20629-20635.

Chasapis CT, Loutsidou AK, Orkoula MG and Spyroulias GA (2010). Zinc binding properties of engineered RING finger domain of Arkadia E3 ubiquitin ligase. *Bioinorganic Chemistry and Applications.* doi:10.1155/2010/323152.

Chumakov PM (2000). Function of the p53 Gene: Choice between life and death. *Biochemistry (Moscow)* **65** (1): 28-40. *Translated from Biokhimiya* **65** (1): 34-47.

Coates J (2000). Interpretation of infrared spectra, a practical approach. *Encyclopedia of analytical chemistry*.

Correa DHA and Ramos CHI (2009). The use of circular dichroisim spectroscopy to study protein folding, form and function. *African Journal of Biochemistry Research* **5** (5): 164-173.

Coscoy L and Ganem D (2003). PHD domains and E3 ubiquitin ligases: viruses make the connection. *Trends in Cell Biology* **13** (1): 7-12.

Damodaran S (2008). Amino acids, peptides, and proteins. *Chemical Rubber Company*. **4**: 217-329

Dawid IG, Breen JJ and Toyama R (1998). LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *TIG* **14** (4).

Demchenko AP (2013). *Ultraviolet spectroscopy of proteins*. Springer Science & Business Media.

Deshaies RJ and Joazeiro CAP (2009). RING domain E3 ubiquitin ligases. *Annual Revised Biochemistry* **78**: 399-434.

Desjardins P and Conklin D (2010). NanoDrop microvolume quantitation of nucleic acids. *Journal of Visualized Experiments* **45**: e2565, DOI: 10.3791/2565.

Dheur S, Voisinet-Hakil F, Minet M, Schmitter JM, Lacroute F, Wyers F and Minvielle-Sebastia L (2003). Pti1p and Ref2p found in association with the mRNA 3′ end formation complex direct snoRNA maturation. *The EMBO journal* **22** (11). 2831-2840.

Doig AJ, Andrew CD, Cochran DA, Hughes E, Penel S, Sun JK, Stapley BJ, Clarke DT and Jones GR (2001). Structure, stability and folding of the α-helix. *Biochemical Society Symposia* **68**: 95-110.

Domingueuz C (PhD thesis) 2004. NMR-based docking of protein-protein complexes. **3**: 338-351.

Dominguez C, Folkers GE and Boelens R (2004). Biological introduction: RING domain proteins. *Handbook of Metalloproteins* **3**: 338-351.

Dudev T and Lim C (2003). Metal binding and selectivity in zinc proteins. *Journal of the Chinese Chemical Society* **50**: 1093-1102.

Dudev T and Lim C (2003). Principles governing Mg, Ca, and Zn binding and selectivity in proteins. *Chemical Reviews* **103**: 773-788.

Dudev T, Lin Y, Dudev M and Lim C (2003). First-second shell interactions in metal binding sites in proteins: A PDB survey and DFT/CDM calculations. *Journal of American Chemical Society* **125** (10): 3168-3180.

Ernst RR, Bodenhausen G and Wokaun A (1988). Principles of Nuclear Magnetic Resonance in One and Two Dimensions. *The International Series of Monographs on Chemistry* **14**: Clarendon Press, Oxford.

Fenn JB, Mann M, Meng CK, Wong SF and Whitehouse CM (1990). Electrospray ionization - principles and practice. *Mass Spectrometry Reviews* **9**: 37.

Ferrer-Dalmau J, González A, Platara M, Navarrete C, Martínez JL, Barreto L, Ramos J, Ariño J and Casamayor A (2010). Ref2, a regulatory subunit of the yeast protein phosphatase 1, is a novel component of cation homoeostasis. *Biochemical Journal* **426** (3): 355-364.

Fioramonte M, dos Santos AM, McIlwain S, Noble WS, Franchini KG and Gozzo FC (2012). Analysis of secondary structure in proteins by chemical cross-linking coupled to MS. *Proteomics* **12** (17): 2746-2752.

Franklin WA (2000). Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia. *Chest* **117**: 80S–89S.web.

Freed-Pastor WA and Prives C (2012). Mutant p53: one name, many proteins. *Genes and Development* **26**: 1268-1286.

Freemont PS, Hanson IM and Trowsdale J (1991). A novel cysteine-rich sequence motif. *Cell Biology* **64**: 483–84.

Gaillard H and Aguilera A (2014). Cleavage factor I links transcription termination to DNA damage response and genome integrity maintenance in Saccharomyces cerevisiae. *PLoS Genetics*. **10** (3). 1004203.

Garas M, Dichtl B and Keller W (2008). The role of the putative 3′ end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *Publication of RNA society* **14** (12): 2671-2684.

Graveley BR and Maniatis T (1998). Arginine/Serine-rich domain of SR proteins can function as activators of pre-mRNA splicing. *Molecular and Cellular Biology* **1**: 765-771.

Greenfield NJ (2006). Using circular dichroisim spectra to estimate protein secondary structure. *Nature protocol* **1** (6): 2876-2860.

Gross S and Moore CL (2001). Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3′-end formation. *Molecular and Cellular biology* **21** (23): 8045-8055.

Hatakeyama S and Nakayama KI (2003). U-box proteins as a new family of ubiquitin ligases. *Biochemical and Biophysical Research Communications* **302**: 635-645.

Houben K, Wasielewski E, Dominguez C, Kellenberger E, Atkinson AR, Timmers HT, Kieffer B and Boelens R (2005). Dynamics and Metal exchange properties of C4C4 RING domains from CNOT4 and the p44 subunit of TFIIH. *Journal Molecular Biology* **349**: 621-637.

Hsieh JK, Chan FSG, O'Connor DJ, Mittnacht S, Zhong S and Lu X (1999). RB regulates the stability and the apoptotic function of p53 via MDM2. *Molecular cell*. 181-193.

Huang A, de Jong RN, Folkers GE and Boelens Rolf (2010). NMR characterization of foldness for the production of E3 RING domains. *Journal of Structural Biology* **172**: 120-127.

Kappo MA (PhD thesis) 2009. Solution structure of the RING finger domain from the human splicing-associated protein RBBP6 using heteronuclear Nuclear Magnetic Resonance (NMR) Spectroscopy.

Kappo MA, AB E, Hassem F, Atkinson RA, Faro A, Muleya V, Mulaudzi T, Poole JO, McKenzie JM, Chibi M, Moolman-Smook JC, DJG Rees and DJR Pugh (2012). Solution structure of RING finger-like domain of Retinoblastoma-binding protein-6 (RBBP6) Suggests it functions as a U-box. *Journal of Biological Chemistry* **287**: 7146-7158.

Katoh S, Hong C, Tsunoda Y, Murata K, Takai R, Minami E, Yamazaki T and Katoh E (2003). High precision NMR structure and the function of the RING-H2 finger domain of EL5, a rice protein whose expression is increased upon exposure to pathogen-derived oligosaccharides. *Journal of Biological Chemistry* **278**: 15341-15348.

Kellenberger E, Dominguez C, Friborug S, Wasielewski E, Moras D, Poterszman A, Boelens R and Kieffer B (2005). Solution structure of the carboxy-terminal of human TFIIH P_{44} subunit. *The journal of Biological Chemistry,* (PMID: 15790571).

Kendrew JC (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**: 662–666.

Ko LJ and Prives C (1996). p53: puzzle and paradigm. *Genes and Development* **10**: 1054-1072.

Kostic M, Matt T, Matinez-Yamout MA, Dyson HJ and Wright PE (2006). Solution structure of the Hdm2 C2H2C4 RING, a domain critical for ubiquitination. *Journal Molecular Biological* **363**: 433-450.

Krishna SS, Majumdar I and Grishin NV (2003). Structural classification of zinc fingers. *Nucleic Acids Research* **31** (2): 532-550.

Krishnamurthy S, He X, Reyes-Reyes M, Moore C and Hampsey M (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. *Molecular cell Biology* **14** (3): 387-394.

Kruger NN (2009). *The Bradford method for protein quantitation*. The protein protocol Handbook. John M. Walker, School of life, University of Hertfordshine, Humara press, 3rd Edition: 17-24.

Kumeta H, Kabashigawa Y and Inagaki F (2004). Solution structure of Helix-RING domain of Cbl-b in the Tyr363 phosphorylated form. *Proceedings of the National Academy of Sciences,* (PMID: 22158902).

Kwan AH, Mobli M, Gooley PR, King GF and Mackay JP (2011). Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS Journal* **278**: 687-703.

Li L, Deng B, Xing G, Teng Y, Tian C, Cheng X, Yin X, Yang J, Gao X, Zhu Y, Sun Q, Zhang L, Yang X, and He F (2007). PACT is a negative regulator of p53 and essential for cell growth and embryonic development. *National Academy of Science U.S.A* **104**: 7951–7956.

Lin MZ, McKeown MR, Ng HL, Aguilera TA, Shaner NC, Campbell RE, Adams SR, Gross LA, Ma W, Alber T and Tsien RY (2009). Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chemistry & biology* **16** (11): 1169-1179.

Liu Y, Carbonell J, Klein-Seetharaman J and Gopalakrishnan V (2003). Prediction of parallel and anti-parallel beta-sheets using Conditional Random Fields.

Logan MR, Nguyen T, Szapiel N, Knockleby J, Por H, Zadworny M, Neszt M, Harrison P, Bussey H, Mandato CA and Vogel J (2008).Genetic interaction network of the Saccharomyces cerevisiae type 1 phosphatase Glc7. *BioMed Central genomics* **9** (1): 336.

Ma K, Xiao J, Li X, Zhang Q and Lian X (2009). Sequence and expression analysis of the C3HC4-type RING finger gene family in rice. *Genetics* **444**: 33-45.

Masanori T, Evrad JL, Steinmetz A and Dawid IB (1995). Classification of LIM proteins. *Trends in Genetics* **11** (11): 43-432.

Mbah AN, Kamga HL, Awofolu OR and Isokpehi RD (2012).Drug target exploitable structural features of adenylyl cyclase activity in Schistosoma mansoni. *Drug target Insights*.41-59.

McLaren AD and Shugar D (2014). *Photochemistry of Proteins and Nucleic Acids: International Series of Monographs on Pure and Applied Biology: Modern Trends in Physiological Sciences* **22**.

Mellon FA, Self R, Startin JR and Belton PS (2000). Amino acids, peptides and proteins.

Metzger MB, Hristova VA and Weissman AM (2012). HECT and RING finger families of E3 ubiquitin ligases at a glance. *Journal of Cellular Sciences* **125** (3): 531-7.

Minoo P, Sullivan W, Solomon LR, Martin TE, Toft DO and Scott RE (1989). Loss of proliferative potential during terminal differentiation coincides with the decreased abundance of a subset of hnRNP proteins. *Journal of Cellular Sciences* **109**: 1937-1946.

Montgomerie S, Sundararaj S, Gallin WJ and Wishart DS (2006). Improving the accuracy of protein secondary structure prediction using structural alignment. *BioMed Central bioinformatics* ***7*** (1): 301.

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

Mulaudzi T (Master's thesis) 2007. An investigation of the zinc binding characteristic of the RING finger domain from the human RBBP6 protein using heteronuclear NMR spectroscopy.

Ndabambi N (Master's thesis) 2009. Recombinant expression of the pRb- and p53-interacting domains from the human RBBP6 protein for *in vivo binding* studies.

Nedergaard J, Ricquier D and Kozak LP (2005). Uncoupling proteins: current status and therapeutic prospects. *EMBO R*ep **6** (10): 917-921.

Ntwasa M, Rakgotho M and Mather A (2005). SNAMA, a novel protein with DWNN domain and a RING finger-like motif: A possible role in apoptosis. *Biochimica et Biophysica Acta* **1727**: 169-176.

Pawełkowicz ME, Skarzyńska A, Posyniak K, Ziąbska K, Pląder W and Przybecki Z (2015). Advantages and disadvantages in usage of bioinformatics programs in promoter region analysis. *International Society for Optics and Photonics.*96621L-96621L

Peidis P, Giannakouros T, Burow ME, Williams RW and Scott RE (2010). Systems genetics analyses predict a transcription role for P2P-R: Molecular confirmation that P2P-R is a transcriptional co-repressor. *BMC System Biology* **4**: 14.

Perczel A, Gáspári Z and Csizmadia IG (2005). Structure and stability of β-pleated sheets. *Journal of Computational Chemistry* **26** (11): 1155-1168.

Pretorius A (PhD thesis) 2007. Functional analysis of the mouse RBBP6 gene using interference RNA.

Proudfoot NJ (2011). Ending the message: poly(A) signal then and now. *Genes and Development* **25**: 1770-1782.

Pugh DJR (2005). Biomolecular NMR spectroscopy in South Africa: the first five years. *South African Journal of Science* **101**: 421-429.

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

Rehman I, Evans CA, Glen A, Cross SS, Eaton CL, Down J, Pesce G, Phillips JT, Ow SY, Thalmann GN and Wright PC (2012). Correction: iTRAQ Identification of Candidate Serum Biomarkers Associated with Metastatic Progression of Human Prostate Cancer. *PloS one* **7** (5).

Rokde CN and Kshirsagar M (2013). Bioinformatics: Protein structure prediction. *In Computing, Communications and Networking Technologies*.1-5

Rosselló-Mora R and Amann R (2001). The species concept for prokaryotes. *FEMS microbiology reviews* **25** (1): 39-67.

Saijo M, Sakai Y, Kishino T, Nikana N, Mastuura Y, Morino K, Tamai K and Taya Y (1995). Molecular cloning of a human protein that binds to the Retinoblastoma protein and chromosomal mapping. *Genomics* **27**: 511-519.

Sakai F (1997). Headache-free cerebrovascular reactivity. *Cephalagia An International Journal of Headache* **17** (3): 144.

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* **30**: 98-101.

Sandra H and David W S (2011). Purification of proteins fused to glutathione S-tranferase. *Methods in Molecular Biology* **681**: 259-280.doi:10.1007/9/978-1-60761-913-0_14.

Scheel H and Hofmann K (2003). No evidence for PHD fingers as ubiquitin ligases. *Trends Cell Biology* **13**: 285–87.

Scott RB, Tzen CY, Witte MM, Blatti S and Wang H (1993). Regulation of differentiation, proliferation and cancer suppressor activity. *International Journal Development Biology* **37**: 67-74.

Scott RE and Gao S (2002) P2P-R deficiency modifies nocodazol induced mitotic arrest and UV-induced apoptosis. *Anticancer Revised* **22**: 3837–3842.

Scott RE, Giannakouros T, Gao S and Peidis P (2003). Functional potential of P2P-R.Arole in the cell cycle and cell differentiation related to its interactions with proteins that bind to matrix-associated regions of DNA? *Journal of Cellular Biochemistry* **90**: 6–12.

Sellers WR, Rodgers JW and Kaelin Jr. WG (1995). Potent transrepression domain in the retinoblastoma protein induces a cell cycle arrest when bound to E2F sites. *Protocol National Academy of Science USA* **92**: 11544-11548.

Sherr CJ and McCormick F (2002). The RB and p35 pathways in cancer. *Cancer cell.* **2**: 103-112.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd Frank J and Manley JL (2009) Molecular architecture of the human premRNA 3_-processing complex. *Molecular Cell* **33**: 365–376.

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* **14**: 145-155.

Soares LM and Buratowski S (2012). Yeast Swd2 is essential because of antagonism between Set1 histone methyltransferase complex and APT (associated with Pta1) termination factor. *Journal of Biological Chemistry* **287** (19): 15219-15231.

Stenmark H and Aasland R (1999). FYVE-finger proteins-effectors of an inositol lipid. *Journal of Science* **112**: 4175-4183.

Stenmark H, Aasland R and Driscoll PC (2002). The phosphatidylinositol 3-phosphate-binding FYVE finger. *FEBS Letters* **513**: 77-84.

St-Pierre B., Liu X., Lan-Chau T.K., Zhu X., Ryan O., Jiang Z. and Zacksenhaus E. (2005). Conserved and specific functions of mammalian ssu72. *Nucleic acids research* **33** (2): 464-477.

Sun ZW and Hampsey M (1996). Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection *in vivo. Molecular Cell Biology* **16**:1557–1566.

Tacahashi Y, Helmling S and Moore CL (2003). Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor. *Nucleic acids research* **31** (6): 1744-1752.

Tao H, Simmons BN, Singireddy S, Jakkidi M and Short KM (2008). Structure of the MID1 tandem B-boxes reveals an interaction reminiscent of intermolecular ring heterodimers. *Biochemistry* **47** : 2450–57.

Terasaki Y, Fuduka Y, Suga M, Ikeguchi N and Takeya M (2005). Epimorphin expression in interstitial pneumonia. *Respiratory Research* **6:** 6.

Vander Kooi CW, Ohi MD, Rosenberg JA, Oldham ML and Newcomer ME (2006). The Prp19 U-box crystal structure suggests a common dimeric architecture for a class of oligomeric E3 ubiquitin ligases. *Biochemistry* **45**: 121–30.

Vo LTA, Minet M, Schmitter JM, Lacroute F and Wyers F (2001). Mpe1, a Zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation. *Molecular and Cellular biology* **21** (24): 8346-8356.

Wanichthanarak K (2014). Applied Bioinformatics in Saccharomyces cerevisiae*: Data storage, integration and analysis*. Chalmers University of Technology.

Whitehouse CM, Dreyer RN, Yanashita M and Fenn JB (1985). Electrospray interface for liquid chromatographs and mass spectrometers. *Annual Chemistry* **57**: 675-679.

Whitmore L and Wallace BA (2007). Protein secondary analyses form circular dichroism spectroscopy: Methods and reference database. *Wiley International Science* **89**: 392-400.

Witte MM and Scott RE (1997). The proliferation potential protein-related (P2P-R) gene with domains encoding heterogeneous nuclear ribonucleoproteins association and Rb1 binding

shows repressed expression during terminal differentiation. *Proceedings National Academy of Science USA* **94**: 1212-1217.

Wüthrich K (1982). Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *Journal Molecular. Molecular cell Biology* **155** (3): 311–319.

Yoshitake Y, Nakatsura T and Monji M (2004). Proliferation potential-related protein, an ideal oesophageal cancer antigen for immunotherapy, identified using complementary DNA microarray analysis. *Clinical Cancer Revised,* **10**: 6437–6448.

Yu Y, Xu W, Wang S, Xu Y, Li H, Wang Y and Li S (2011).VpRFP1, a novel C4C4-type RING finger protein gene from Chinese wild Vitis pseudoreticulata, functions as a transcriptional activator in defence response of grapevine. *Journal of Experimental Botany* **62** (15): 5671-5682.

Zhao J, Kessler MM and Moore CL (1997). Cleavage Factor II of Saccharomyces cerevisiaeContains Homologues to Subunits of the Mammalian Cleavage/Polyadenylation Specificity Factor and Exhibits Sequence-specific, ATP-dependent Interaction with Precursor RNA. *Journal of Biological Chemistry* **272** (16):10831-10838.

Zhou P, Yao Y, Soh JW and Weinstein IB (1999). Overexpression of p21Cip1 or p27Kip1 in the promyelocytic leukaemia cell line HL60 accelerates its lineage specific differentiation. *Anticancer Revised* **19**: 4935–4945.

Zitting A, Husgafvel-Pursianen K and Rantanen J (2002). Environmental tobacco smoke—a major preventable cause of impaired health at work. *Scandal Journal Work Environmental Health* **28**: 3–6.

APPENDIX I

GENERAL CHEMICALS AND ENZYMES

Chemicals	Company
Glutathione	Merck
Ammonium chloride (NH_4Cl)	Merck
Sodium hydrogen phosphate (Na_2HPO_4)	Merck
Magnesium sulphate ($MgSO_4$)	Merck
Sodium azide (NaN_3)	Merck
Isopropanol	Merck
Tryptone	Merck
Yeast	Merck
Glycerol	Merck
Sodium chloride (NaCl)	Merck
Calcium chloride ($CaCl_2$)	Merck
Magnesium chloride ($MgCl_2$)	Merck
Potassium chloride (KCl)	Merck
Sodium phosphate ($NaHPO_4$)	Merck
Potassium dihydrogen phosphate (KH_2PO_4)	Merck
Tris[hydroxymethyl]amino ethane (Tris)	Merck
Glucose	Merck
Glycine	Merck
Coomasie Brillant Blue R-250	Merck

Ampicillin	Calbiochem
Phenyl Methyl Sulphonyl Fluoride (PMSF)	Calbiochem
Agarose gel powder	Calbiochem
Glutathione agarose beads	Sigma
Ethidium bromide	Sigma
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Sigma
iso-octylphenoxypolyethoxyethanol (Triton X-100)	Sigma
BamHI	Thermo Scientific
HindIII	Thermo Scientific
DNA marker	Thermo Scientific
Ethylene diamine tetra Acetic Acid (EDTA)	Associated chemical enterprises
Glacial acetic acid	Associated chemical enterprises
Zinc sulphate ($ZnSO_4$)	Associated chemical enterprises
Ethanol	Associated chemical enterprises
Isopropylβ-D-thiogalactopyranoside (IPTG)	Melford
Sodium Dodecyl Sulphate (SDS) powder	Promega
Bis-acrylamide	Bio-Rad
Ammonium sulphate (APS)	Bio-rad
Protein marker	Bio-rad
Lysozyme	Amersco
Dithiothreitol (DTT)	Amersco
Methanol	Merck

APPENDIX II

PHYLOGENIC TREE PROTEIN MULTIPLE SEQUENCE ALIGNMENT

Cleavage polyadenylation factor subunit MPE1 [Saccharomyces cerevisiae S288c]

Sequence ID: NP_012864.1Length: 441Number of Matches: 1
HFKDLPDDLKCPLTGGLLRQPVKTSKCCNIDFSKEALENALVESDFVCPNCETRDILLDSLVPDQDKEKEVETFLKKQEELHGS

Hypothetical protein NDAI_0C00740 [Naumovozyma dairenensis CBS 421]

Sequence ID: XP_003668978.1Length: 462Number of Matches: 1

HFKDLPDTLKCPITGGLLSSPVKTSNCCNVDFSKKAIEDTLVESDFVCPNCGKEDILLDSLIDDEDKKKEVEEFLKEH

LAFE_0F07294g1_1 [Lachancea fermentati]

Sequence ID: SCW02477.1Length: 397Number of Matches: 1

HFPDLPDDLKCPITGGLLKNPVKTSKCCGKDFSRQALEDMLVENDFTCPACNQGDILLDSLAPDEDKKKEVEKFLDEH

LALA0S09e05754g1_1 [Lachancea lanzarotensis]

Sequence ID: CEP63933.1Length: 396Number of Matches: 1

HFADLPEDLKCPITGGLLRQPVKTSKCCQKSFSKQALEDVLVERDFVCPECGAEDVLLDSLVPDDEKQAQVEKFLEEHG

LAME_0H12090g1_1 [Lachancea meyersii CBS 8951]

Sequence ID: SCV03645.1Length: 396Number of Matches: 1

HFPDLPDDLKCPITGGLLKNPVKTSKCCGKDFSRQALEDMLVENDFTCPACSQSDILLDSLVPDEEKKKLVEKFM

LANO_0H12816g1_1 [Lachancea nothofagi CBS 11611]

Sequence ID: SCV05680.1Length: 395Number of Matches: 1

HFPDLPEDLKCPITGGLLKSPVKTSKCCGKHFSRQALEDMLVENDFTCPACGKEEILLDSLVPDEDKEKQVEKFL

hypothetical protein TDEL_0B06720 [Torulaspora delbrueckii]

Sequence ID: XP_003680012.1Length: 421Number of Matches: 1

HFSDLPEDLKCPITGGLLKSPVKTSKCCKKHFSRQALEDLLVENDFVCPGCNAEETLLDSLIPDEDKEKQVQKFL

hypothetical protein KAFR_0I01680 [Kazachstania africana CBS 2517]

Sequence ID: XP_003959084.1Length: 398Number of Matches: 1

FKNLPADLTCPITGGLLKDPVKTSKCCNKDFSKVALEDALLESDFVCPECQSPDILLDSLVEDEEKSKQVKEFLKNNQ

hypothetical protein Kpol_1064p6 [Vanderwaltozyma polyspora DSM 70294]

Sequence ID: XP_001644384.1Length: 461Number of Matches: 1

HFTNLPDELKCPLTKGLLRDPVKTTNCCNKTFSKVAIENALLESDFICPSCNTHDILLDSLQQDTEKQKQVQLFLKEND

LAQU0S01e08636g1_1 [Lachancea quebecensis]

Sequence ID: CUS20528.1Length: 396Number of Matches: 1
HFKNLPDEMKCPITGGLLRDPVTTSRCCKKFFSKAAIENYLLESDFVCPGCETPDILLDSLVESKEKKNEVEKYIKEHE

APPENDIX III

AMINO ACIDS AND THEIR ABBREVIATIONS

Amino acid	3-letter abbreviation	1-letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V