Acknowledgements

I am grateful to my Supervisors Prof. David P. Hornby and Dr Mark J Dickman for the opportunity to work in their labs and for their assiduous guidance and assistance throughout my PhD research.

I would like to thank Dr Qaiser I. Sheikh and Paul Brown for their help, especially during the early stages of my studies.

Many thanks to our collaborators from the Department of Protein-DNA Interactions, Institute of Biotechnology, Vilnius University, Lithuania.

I am very grateful to Abia State University, Uturu, Nigeria and TETFund for funding my study in the UK.

I thank my Fiancé, Ahuva, for always being there for me, championing my cause and for her perseverance. I also want to thank her parents for their kind support.

For their support and encouragement in all my endeavours I thank my Dad, Mr Martin Nwokeoji and my siblings.

For her enduring love and for teaching me the virtues of hard work and resilience I am eternally grateful to my Late Mother, Mrs Rosemary Nwokeoji, who unfortunately passed away while I was in the UK for this PhD.

Table of Contents

List of Figures

List of Tables

Abbreviations

AEX	Anion Exchange Chromatography
DNA (S)	Deoxyribonucleic acid (S)
Cas	CRISPR associated
Cascade	CRISPR associated complex for antiviral defence
dsDNA (S)	Double stranded ribonucleic acid (S)
dsRNA (S)	Double stranded ribonucleic acid (S)
EDTA	Ethylenediaminetetraacetic acid
ESI MS	Electrospray ionisation Mass spectrometry
ETD	Electron transfer dissociation
HFIP	1,1,1,3,3,3-hexafluoro-2-propanol
HILIC	Hydrophilic interaction chromatography
hnRNA	Heterogeneous ribonucleic acid
hnRNP	heterogeneous ribonucleoprotein
IPTG	Isopropyl-β-D-thiogalactopyranoside
LC	Liquid Chromatography
miRNA	micro ribonucleic acid
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
ncRNA	Non-coding ribonucleic acid
PCR	Polymerase chain reaction
pre-crRNA	Precursor CRISPR ribonucleic acid
pre-mRNA	precursor messenger ribonucleic acid
Pri-miRNA	Primary micro ribonucleic acid
Q-TOF	Quadruple time of flight
RF	Radio frequency
RISC	RNA-induced silencing complex
RNA (S)	Ribonucleic acid (s)

rRNA	Ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
siRNA	Small interfering ribonucleic acid
snRNA	small nuclear ribonucleic acid
ssRNA	Single stranded ribonucleic acid
TAP	Tandem affinity Purification
TEA	Triethyl amine
TEAA	Tetraethyl ammonium bicarbonate
TEMED	N,N,N',N,-tetramethylethane-1,2-diamine
TOF	Time of Flight
tRNA	Transfer ribose nucleic acid
UHR-TOF	Ultra-high resolution time-of-flight
UPLC	Ultra performance liquid chromatography
XIC	Extracted ion chromatogram

ABSTRACT

Protein-protein/protein-RNA interactions are assuming increasing significance with the recent discoveries of the diverse and important roles RNA-protein complexes play in biological systems. As a result there is an increasing demand for high throughput analytical approaches for the analysis of these complexes. In this study, affinity purification-mass spectrometry (AP-MS) approaches are used to analyse protein-RNA complexes. Tandem affinity purification (TAP)-MS was used to analyse complexes associated with the *Saccharomyces cerevisiae* spliceosome and RNA processing pathways providing insight into the coupling of the various steps of gene expression. These approaches were also used to exhaustively distinguish two similar but functionally distinct Lsm complexes providing further insight into their mRNA processing pathways. These studies are consistent with the model that splicing occur co-trancriptionally and that several steps of gene expression is coupled to transcription. Utilizing HPLC-MS approaches in conjunction with other molecular biology techniques, the recently discovered *Streptococcus thermophilus* CRISPR/Cas complex was also investigated. The Cas protein complexes were identified and their stoichiometry determined using semi-quantitative data. This provided insight into the roles of these Cas proteins in the biogenesis of the CRISPR/Cas complex. Furthermore, the architecture of the CRISPR RNAs (crRNA) associated with the complex was determined, providing further insight into this mechanism of the crRNA processing in the novel Type III CRISPR/Cas complex

Chapter 1

Introduction

Proteins are biological molecules consisting of one or more chains of amino acids whose roles in the cell include cellular catalysts (enzymes), transport molecules, biological sensors and structural units (building blocks of the cell)(Gutteridge and Thornton, 2005). Although, all proteins are ultimately composed of 20 standard amino acids, they differ from each other by the linear sequence of their amino acids and the folding of their amino acid chains into unique tertiary- or quaternary-dimensional structures. Protein sequence is to a large extent encoded by deoxyribonucleic acid (DNA) sequence but specifically dictated the messenger ribonucleic acid (mRNA) sequence of its gene. RNA, like proteins, are large biological molecules distinguished functionally by their sequence, but are composed of chains of four distinct nucleotide bases and are generally regarded as the secondary hereditary material of the cell (Gutteridge and Thornton, 2005; Higgs, 2000).

The term 'proteome', first coined in the mid-1990s, can be defined as the entire complement of proteins expressed in a cell, organism or tissue at any given time under defined conditions (Wilkins *et al*., 1996). Typically the proteome is much larger than the genome due to alternate gene splicing and post-translational modifications of proteins (Yates, 2000). In comparison to the relatively constant information retained in the genome, the proteome can be subject to great variations in response to environmental change. Proteomics is the study of all the proteins encoded within the genetic information and how these proteins perform the biological requirements of an organism. This enables a deeper understanding of cellular function to be developed; which proteins are present, their relative abundance, post-translational modifications, sub-cellular localisation, and how proteins interact with each other to perform their defined roles within a system (Wilkins *et al*., 1996).

The interest in protein interactions stems from the discovery that proteins/enzymes do not always catalyse biochemical pathways in the classical fashion predicted by

Beadle and Tatum (Beadle and Tatum, 1941); that is, one-gene/one-enzyme/one-function. In the model postulated by Tatum and Beadle only one protein is responsible for, and catalyses, one biochemical reaction. However, it is now known that many important cellular processes such as DNA replication and repair are carried out by large molecular machineries comprising several proteins (George et al., 2001; Labib et al., 2000; Wang et al., 2004). As the proteomics field has emerged, many researchers have put considerable effort into elucidating protein–protein interaction networks in different organisms, in order to better understand the interplay between proteins and to gain more insight into the potential for disease development in case of network disruptions.

Protein-protein interactions are a phenomenon that defines the binding of one protein to another usually to carry out a biological function. Early structural studies revealed that hydrophobicity is the major principle of protein-protein interactions (Chothia and Janin, 1975). Protein-protein interactions differ depending on the function that is to be performed and the number of molecules involved. Proteins may interact with another protein to transport it, for instance, from the nucleus to the cytoplasm or vice versa (Damelin and Silver, 2000). It may mediate signals from the surface to the inside of the cell which is the case in signal transduction, important in many biochemical processes and disease conditions such as cancer. Proteins may interact with one or more proteins to form a catalytic or structural unit (Finzel et al., 1985; Jones and Thornton, 1996; Keskin et al., 2008; Milligan, 1996). Protein interactions may be transient, as is the case when one of the interacting partners is to be structurally modified, or it can form stable protein complexes (Jones and Thornton, 1996; Keskin et al., 2008).

Protein-protein interactions are at the core of interactomics, a new field which interfaces bioinformatics with molecular biology to understand the nature and consequences of interactions among and between proteins, and other molecules (Keskin et al., 2008). Protein-RNA interaction is a recent innovation in the field which studies the interactions of proteins with RNAs. Several approaches have been used to study protein-protein interactions which will be discussed.

1.1 Principles of Protein-Protein Interactions

It is now widely recognized that the vast majority of cellular pathways are mediated by protein-protein interactions and that the recognition of the function of at least one binding partner in a protein-protein interaction network will facilitate the assignation of a pathway (Chothia and Janin, 1975; Ewing et al., 2007; Keskin et al., 2008). Conversely, the recognition of the binding partners of a protein opens up an avenue to identifying its function. Mapping cellular pathways and their complex connectivity is gaining momentum as researchers are increasingly gaining insight into the network of many protein-protein interactions (Ewing et al., 2007; Mann et al., 2001). Identification of protein-protein interactions is at the heart of proteomics, functional genomics and drug discovery. And since no two proteins can bind at the same site, understanding the way protein partners interact and insight into their mechanism of association will aid in deciphering the dynamic regulation of pathways (Keskin et al., 2008).

Protein-protein interactions fall under two broad categories: 1) interaction where protein partners bind with high affinity and 2) more transient interactions. The obligatory, stable or high affinity interactions form complexes while the transient interactions bind and uncouple continuously (Keskin et al., 2008). To understand what constitute functional protein interactions, the preferred mechanisms of these interactions have to be addressed and many chemical scientists are working towards this end. However, validating functional protein-protein interaction is a difficult task as any two proteins can interact from the chemical point of view. This is because protein-protein interactions are mainly driven by hydrophobic interactions but also by hydrophobic interactions and salt bridges (Xu et al., 1997) (Tsai and Nussinov, 1997). The major challenge becomes formulating a parameter for distinguishing functional interactions from the false ones. The two major considerations taken into account are the conditions and strength of interactions (Keskin et al., 2008).

A good appreciation of the mechanism of protein-protein interactions will require viewing proteins more as flexible structures than as the rigid molecules often depicted in crystal structures. Protein assumes various conformations depending on whether it is free or bound to a ligand (or another protein) and a protein in solution

may have a different conformation from that observed in crystal structure. Shape complementarity, allosteric effects, organisations chemical and physical contributions of components of the complex are also factors important in understanding and predicting protein interactions (Gunasekaran et al., 2004; James et al., 2003).

The amino acid residues that line the surfaces of protein are very important when considering protein-protein interaction as proteins interact through their surfaces. Cooperativity, a concept referring to the chemical and physical inputs of different components of complex towards thermodynamic stability, contribute to the stability, and hence the formation of protein-protein interaction. Giving that a wide spectrum of factors plays roles either directly or indirectly towards protein complex formation, the prediction of protein-protein interaction pieced together from the chemical and physical properties of components of a complex becomes a gruelling task (Keskin et al., 2008). It therefore becomes necessary when elucidating the functional structure of a protein in complex to employ a more holistic approach that captures the structural features of that protein when bound to its functional interacting partners. Crystal structures of the protein complex would have been a good, if not a more ideal, way of studying protein complexes. The problem is that large protein complexes are innately thermodynamic unstable at crystallization conditions, resulting in very few to be crystallized so far (Ke and Doudna, 2004; Keskin et al., 2008). Researchers are increasingly exploiting the powerful technology of tandem affinity purification-mass spectrometry to both identify the components of protein complexes and gain insight into the structure of the complex as a unit. A very practical approach towards gaining more complete structural information will be to complement the mass spectrometry data with crystal structural data of each component of the complex.

1.2 Protein-RNA Interactions

In addition to the importance of studying protein-protein interactions, RNA also plays an important role in complex structures and pathways such as the ribosome, spliceosome and RNAi (Fire et al., 1998; Kastner et al., 1990; Lührmann et al., 1990; Moore, 1998; Ramakrishnan and White, 1998; Thiede and von Janta-Lipinski, 1998).

A common phenomenon in these pathways is mediation of function by the interaction of proteins with RNA. RNA molecules are flexible structures that display secondary and tertiary features that are almost as diverse as their function (Jones et al., 2001). Although commonly single-stranded, some examples of RNA structures include hairpin loop, bulges, pseudoknots and short length of double helices. Proteins tend to bind to RNA where it forms complex secondary structures such as stem loops and bulges (Nagai, 1996). Also, non-Watson base-pairing can also occur in loop regions of RNA and such structures can be selectively recognized by proteins (Jones et al., 2001).

One of the most common RNA-binding motifs is the RNA recognition motif (RRM). Examples of these motifs are found in proteins such as the Sex-lethal (SXl), an RNA binding protein with two RRMs that control sex-differentiation (Crowder et al., 1999). The RNA binding protein, hnRNP A1, also utilizes two RRM motifs to complex with heterogenous RNA (hnRNA) and is thought to influence mRNA processing and export (Shamoo et al., 1994). SXl and hnRNP A1 have RRMs referred to as RRM1 and RRM2 and features the canonical RRM fold and flexible interdomain linker (see Figure 1.1) (Crowder et al., 1999; Shamoo et al., 1997). Whereas hnRNP A1 maintains an interdomain contact between its two RRMs (1HA1) (See Figure 1.1A), SXl (3SXL) does not (see Figure 1.1B)(Crowder et al., 1999)). The RRM is a small protein domain which has about 75-85 amino acids and forms a four-stranded β-sheet against two α-helices (Mattaj, 1993; Nagai, 1996). This recognition motif plays roles in various cellular pathways, such as mRNA/rRNA processing, splicing, translation regulation, RNA export and RNA stability (Haynes, 1992; Knight and Docherty, 1992; Lührmann et al., 1990; van Heugten et al., 1992). Two major ways in which proteins recognize and bind or interact with RNAs have been suggested: 1) by binding to the major groove and 2) by Beta-sheet binding. It is on this basis that two classes are distinguished: a) groove-binding protein-RNA complexes, where proteins position a secondary structure element such as the alpha-helix into an RNA groove, and b) Beta-sheet binding complexes where the protein creates a pocket with its beta-sheet elements that bind unpaired RNA bases (Draper, 1999) (see Figure 1.2).

Several line of evidence from analysis of many RNA binding proteins suggest that groove-binding proteins specifically bind double-stranded or single-stranded, single-

loop RNA elements and beta-sheet binders, single-stranded or single-stranded, single loops or single-stranded, multiple loops (as in U1A spliceosomal protein which has double RRM). However, this division is often complicated by the fact that sometimes different domains of a protein complex or even protein may exhibit both classes of RNA binding, for instance, glutaminyl tRNA synthetase where both classes of RNA-binding domains occur and which binds single stranded, multiple loops RNA structure. Several of these complications abound (see Figure 1.2 and 1.3) (Jones et al., 2001).

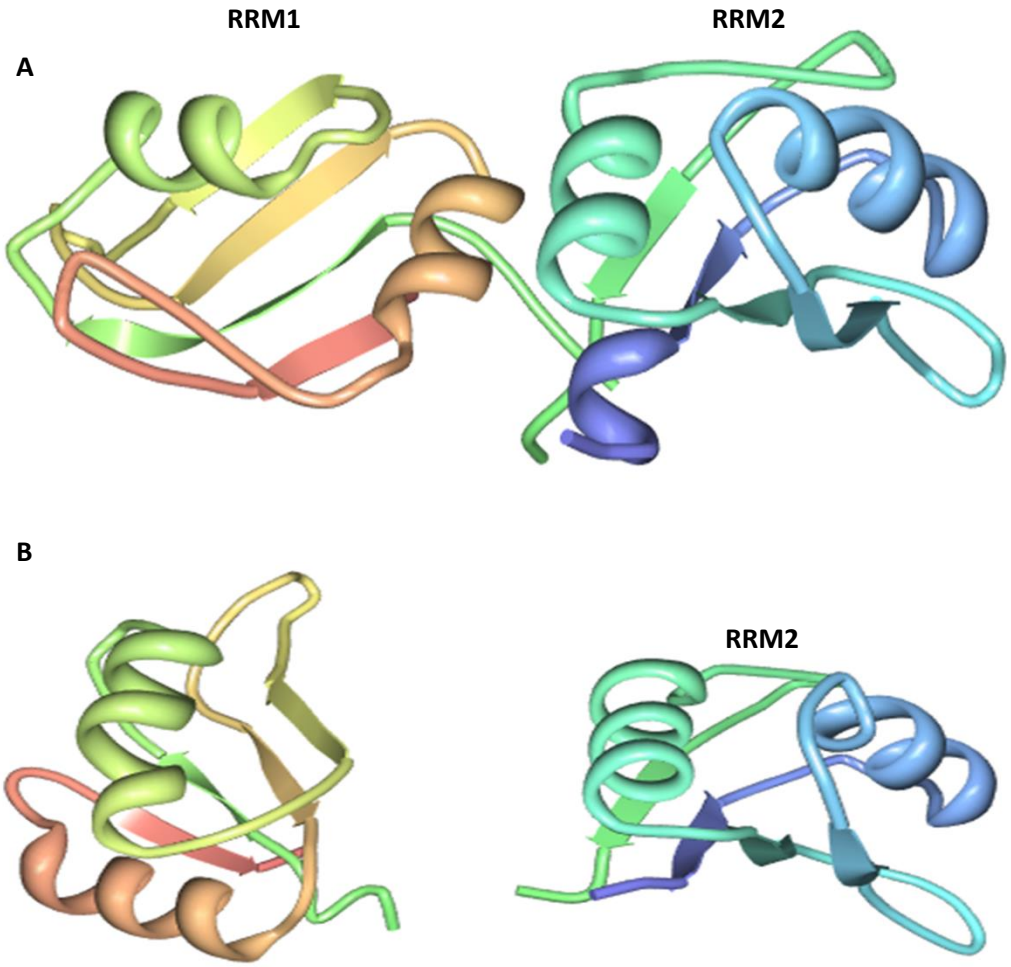


Figure 1.1 Relationship of RRM domains in SXL and hnRNP A1 Ribbon diagrams of hnRNP A1 and Sex-lethal RNA binding motifs (RRMS) using the secondary structure elements of RRM1 and RRM2 of both proteins. A) Crystal structure of hnRNP A1 RRM1 and 2 (1HA1) determined at 1.7 Å resolution reveals two RRMs that are independently folded but connected by a flexible linker. B) Crystal structure of SXL' RRM1 and 2 (3SXL) determined at 2.7 Å resolution displaying the same canonical fold and disordered linker observed in hnRNP A1 RRMs. SXL lacks interdomain contact between RRMs. Figure generated using PDB IDs: 1HA1 and 3SXL.

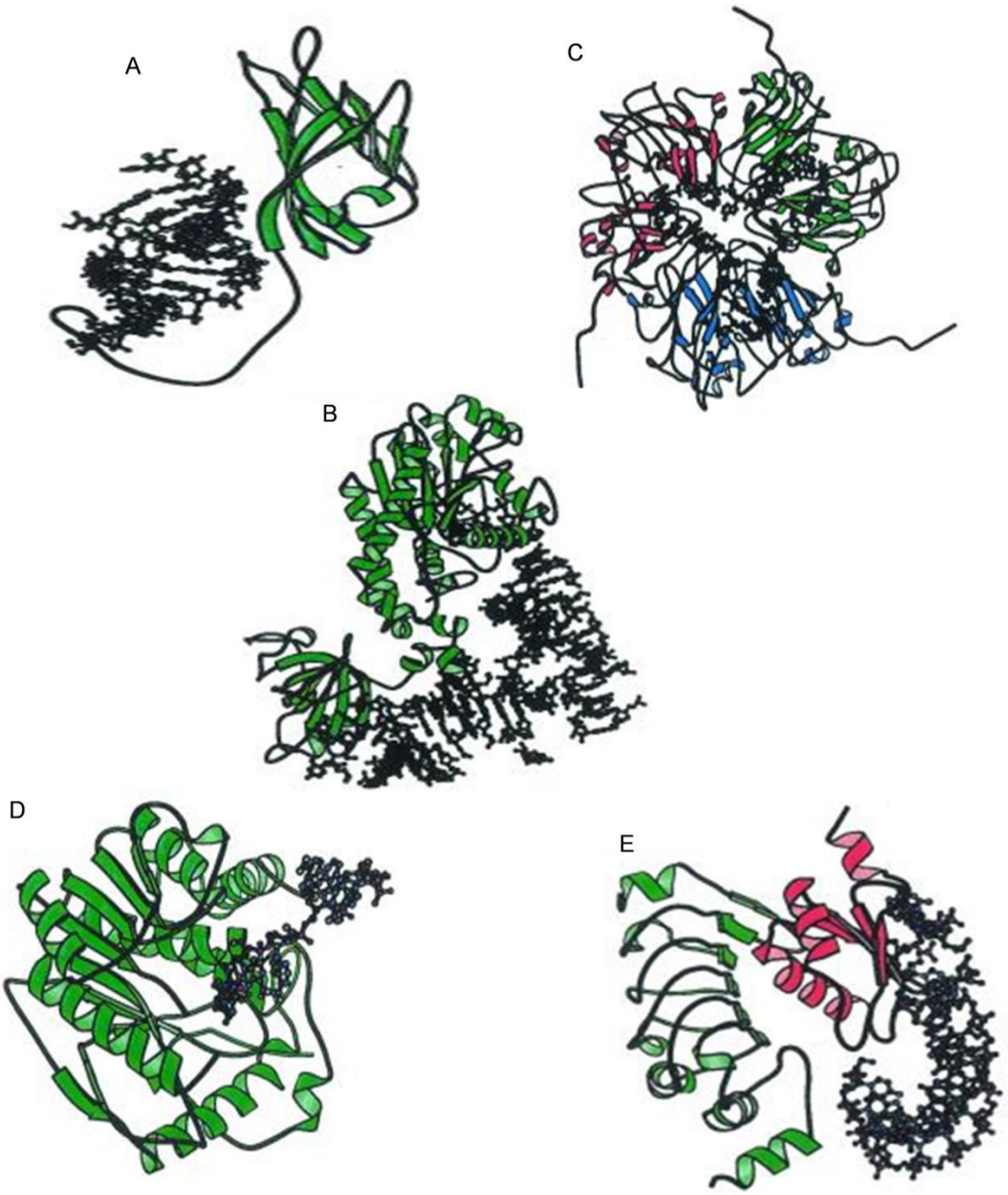


Figure 1.2 Diagrams depicting protein–RNA complexes Examples of complexes belonging to different families. The sizes of the proteins are not comparable between diagrams and each is viewed from an angle that best depicts both the protein and RNA. In each diagram the RNA molecule is shown in ball-and-stick format and the proteins in ribbon format. Different subunits of the same protein are differentiated by colour. (A) In coat protein from Satellite tobacco mosaic virus (1A34) beta-Sheets recognize double-stranded RNA structure; (B) In bean pod mottle virus (1BMV) β-sheet recognizes single-stranded RNA; (C) aspartyl tRNA synthetase (1ASY) utilizes groove binding + β-sheet to recognize single stranded, multiple loops RNA structures; (D) methyltransferase VP39 (1AV6) utilizes groove binding domains to recognize single-stranded RNA structure; (E) spliceosomal U2B″/U2A′ complex (1A9N) utilizes β-sheet to recognize and bind single-stranded, single loop RNA structures. Adapted from Jones et al., 2001.

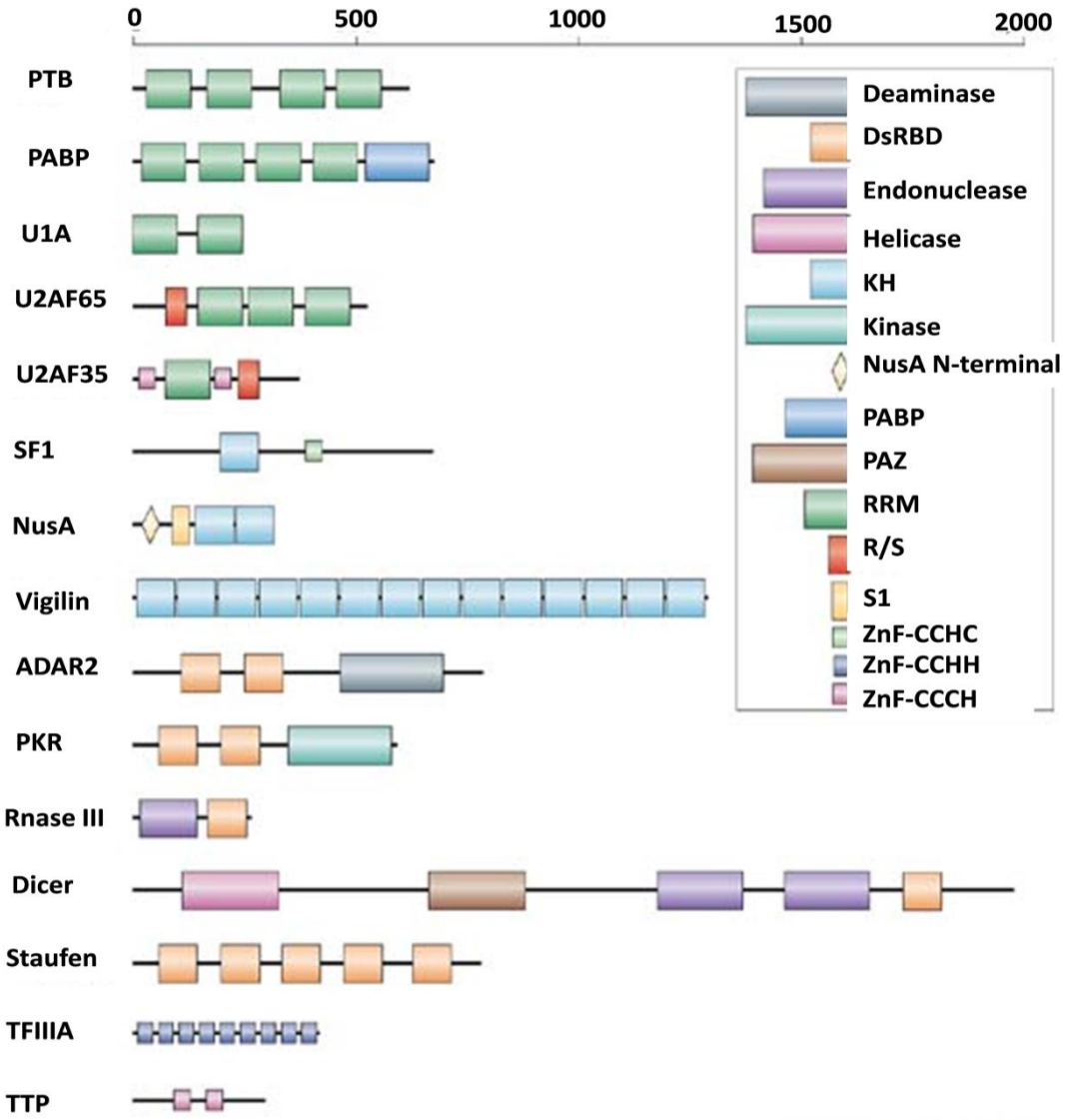


Figure 1.3 Representative examples from some of the most common RNA-binding protein families, as illustrated here demonstrate the variability in the number of copies (as many as 14 in vigilin) and arrangements that exist. This variability has direct functional implications. For example, Dicer and RNase III both contain an endonuclease catalytic domain followed by a double-stranded RNA-binding domain (dsRBD). So, both proteins recognize dsRNA, but Dicer has evolved to interact specifically with RNA species that are produced through the RNA interference pathway through additional domains that recognize the unique structural features of these RNAs. Different domains are represented as coloured boxes. These include the RNA-recognition motif (RRM; by far the most common RNA-binding protein module), the K-homology (KH) domain (which can bind both single-stranded RNA and DNA), the dsRBD (a sequence-independent dsRNA-binding module) and RNA-binding zinc-finger (ZnF) domains. Enzymatic domains and less common functional modules are also shown. PABP, poly(A)-binding protein; PTB, polypyrimidine-tract binding; R/S, Arg/Ser-rich domain; SF1, splicing factor-1; TTP, tristetraprolin; U2AF, U2 auxiliary factor. Taken from Lunde *et al.*, 2007

multiple copies are the higher affinity, specificity and versatility that arise from such modular design. Individual domains, in comparison, often bind short stretches of RNA with weak affinity (Lunde et al., 2007). By evolving an RNA interaction surface with multiple modules, a protein can achieve higher specificity and affinity for a particular target by combining multiple weak RNA-interacting motifs. This strategy regulates the formation of complexes and makes it easier to disassemble them when needed (Lunde et al., 2007). These multiple binding sites can evolve independently and is ideal for proteins that match in their specificity the poorly conserved sequence features that are observed in splicing and 3'-end processing sites of eukaryotic mRNA (Deka et al., 2005; Perez Canadillas and Varani, 2003; Sickmier et al., 2006). Another advantage that arises from a protein with multiple RNA-interacting domains is that such a protein can recognize a much longer stretches of nucleic acids than would be possible for a single domain (Lunde et al., 2007).

The specificity of RNA binding domains and their relative arrangement to each other in a protein is functionally important. In evolution, higher levels of conservation are often observed between domains occupying the same position in orthologous proteins, as against domains in the same protein but in a different position. For instance, this is observed in the splicing factor U2 auxiliary factor (U2AF) subunit 65 and in the poly(A)-binding protein (PABP) both of which contain RRM1, RRM3 and RRM4 domains. For instance, the RNA-recognition motif-1 (RRM1) of the yeast U2AF is more similar to its human ortholog proteins RRM1 (human U2AF RRM1) than it is to yeast U2AF RRM3 or RRM4. The same principle applies to PABP (Lunde et al., 2007). The linker between the two domains plays a major role in the ability of the protein to recognize RNA in a specific way. Long domains are generally disordered and predispose proteins towards recognizing diverse targets (See Figure 1.4a, right panel) while short linkers promote the specificity of the domains to bind contiguous stretch of nucleic acids (See Figure 1.4a, left side) (Lunde et al., 2007). During RNA binding, the linker domain generally becomes ordered and forms a short α-helix that positions the two binding domains relative to one another and sometimes contacts the RNA directly (Allain et al., 2000; Deo et al., 1999; Handa et al., 1999; Lunde et al., 2007; Perez-Canadillas, 2006).

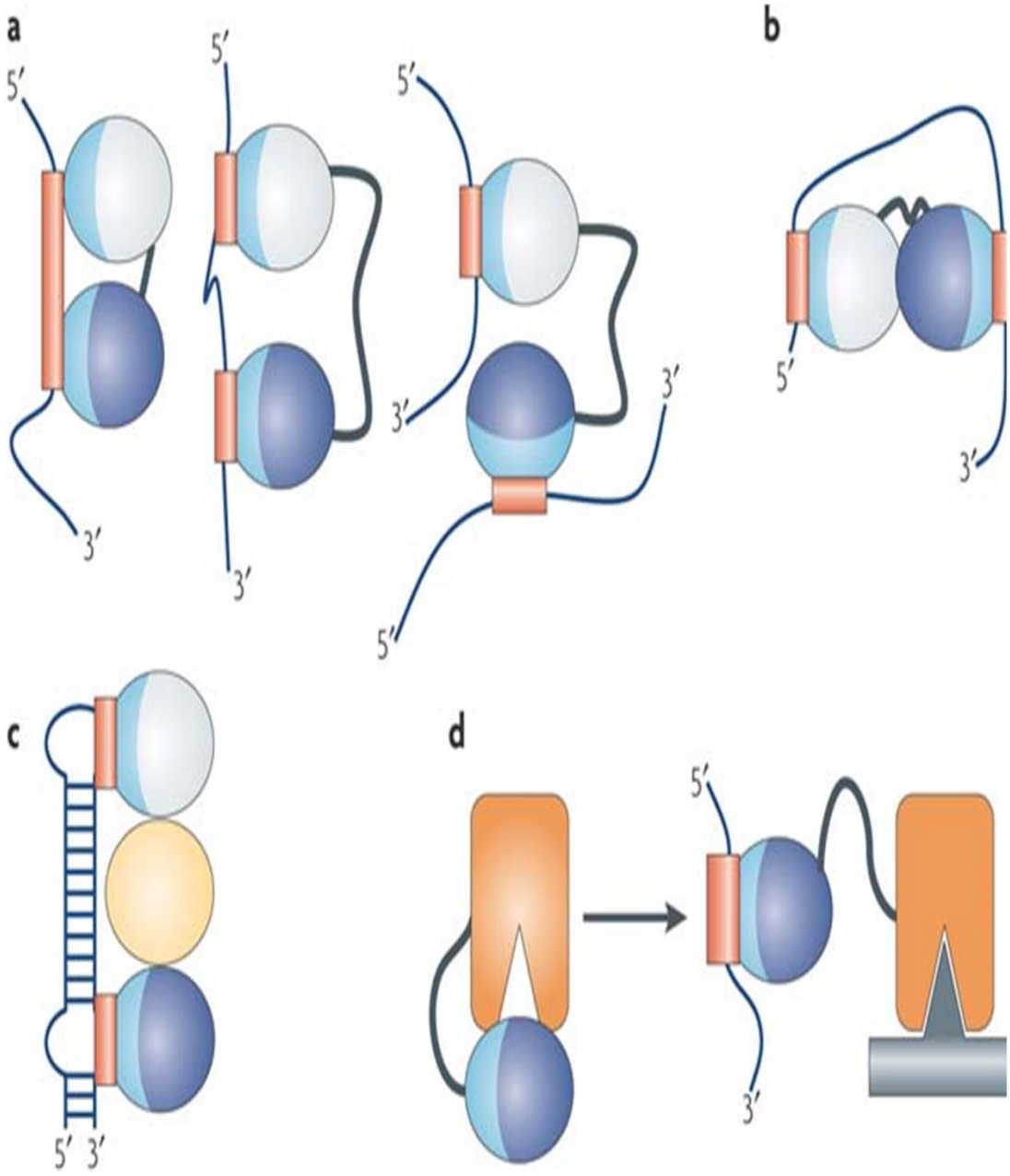


Figure 1.4 RNA-binding domains (RBDs) function in various ways a) They recognize RNA sequences with a specificity and affinity that would not be possible for a single domain or if multiple domains did not cooperate. Multiple domains combine to recognize a long RNA sequence (left), sequences separated by many nucleotides (centre), or RNAs that belong to different molecules altogether (right). b) RBDs can organize mRNAs topologically by interacting simultaneously with multiple RNA sequences. c) Alternatively, they can function as spacers to properly position other modules for recognition. D) They can combine with enzymatic domains to define the substrate specificity for catalysis or to regulate enzymatic activity. The RNA-binding modules are represented as ellipses with their RNA-binding surfaces coloured in light blue, and the corresponding binding sites in the RNA coloured in red; individual domains are coloured differently. Taken from Lunde et al., 2007.

In addition to broadening the scope of RNA recognition, multiple modules also allow RNA-binding proteins to simultaneously interact with other proteins and RNAs. Dimerization is the simplest example of this phenomenon and is observed in RNA interference. RNA interference is a mechanism that inhibits gene expression by using a protein complex harbouring an exonuclease activity and a specific strand of small double stranded RNA known as small interfering RNA (siRNA) as a guide to target and destroy specific mRNA (Fire et al., 1998). An example of dimerization is observed in tombus virus protein p19 which is expressed in response to viral infection in planta and can also suppress RNA interference (RNAi) pathways in human cells and Drosophila *in vitro* (Dunoyer et al., 2004; Lakatos et al., 2004). P19 function by specifically binding siRNA and preventing their loading to the RISC complex (Vargason et al., 2003). Although p19 binds the siRNA as a homodimer, it however, forms a dimer which allows it to measure the length of siRNA with great precision (Vargason et al., 2003; Xia et al., 2009; Ye et al., 2003). In addition, dimerization presents two recognition sites for RNA and therefore provides the cooperative interaction that confers higher affinity for the RNA (Lunde et al., 2007; Ramos et al., 2002).

There are other examples of RNA binding domains (RBD) that function by dimerizing or by forming protein-protein interactions. For instance, in the N-terminal RRM of U1A bound to an RNA-regulatory element in its own 3'untranslated region (3'UTR), two separate RRMs interact through their C-terminal helices to form a homodimer after binding to the RNA. Polyadenylation is a mechanism used in protecting mRNAs from 3' to 5' degradation by addition of adenosine monophosphate polymer (poly(A) tail) which is added by complex harbouring poly(A) polymerase (Balbo and Bohm, 2007). Dimerization creates an interface that inhibits polyadenylation by direct interaction with poly(A) polymerase (Varani et al., 2000). The specificity for RNA interaction can also be enhanced by formation of heterodimers resulting from interaction between an RBD and another protein. This is observed in the spliceosome, where U2A' is required for the binding of RRM of U2B'' to a stem-loop in U2 small nuclear RNA (U2 snRNA) (Price et al., 1998). In another example, the RRM of the spliceosomal cap-binding complex subunit, CBP80 can bind with high affinity to the 7-methylguanosine cap of mRNA only if it interacts with the RRM of

the CBP20 subunit (Calero et al., 2002; Mazza et al., 2002). In spite of recent advances in RNA-protein interactions, there are still few structures of proteins with RBD (Lunde et al., 2007). This is particularly the case in the large complex, spliceosome, where most of the spliceosomal proteins and snRNAs are known but few structures of protein-RNA complex exist. The difficulty in obtaining structural information from spliceosome is owed to its dynamic structure and complexity, making it difficult to crystallize (Lührmann et al., 1990; Will and Lührmann, 2011). It therefore becomes necessary to investigate such dynamic protein-RNA interactions using other strategies, such as Mass spectrometry.

1.3 Methods in the study of protein-protein interactions

Several methods exist to investigate protein-protein interactions. Most of these methodological approaches have undergone many refinements during their evolution and continue to be developed. Each of the methods has their strengths and weakness, especially, with regards to specificity and sensitivity. The specificity refers to the capability of a methodological approach to reveal functional interactions that occur *in vivo*. A highly specific method will show that most of its interactions are functional, reflect *in vivo* interactions and 'noisy' non-specific interactions will be minimal. The sensitivity refers to the ability of an approach to detect all protein-protein interactions possible for a given target *in vivo*. That means a perfectly sensitive method will detect all possible interactions for given protein target. A researcher's choice of any approach may also depend on the method's throughput capabilities and amenability to the specific study of interest. The methods used to study protein-protein interactions will be discussed briefly in the following sections.

1.4 Two-Hybrid studies

1.4.1 Yeast two-hybrid screen

The yeast two-hybrid (Y2H) technique allows detection of interacting proteins in living cells (Fields and Song, 1989). In this approach, two interacting proteins,

referred to as bait and prey, activate reporter genes that enable growth on specific media or a colour reaction (Bruckner et al., 2009; Fields and Song, 1989; Shaffer et al., 2012). Yeast two hybrid technique can now be adapted to high-throughput studies of protein interactions on a genome-wide scale, as shown in viruses like bacteriophage T7 (Bartel et al., 1996), *Saccharomyces cerevisiae* (Ito et al., 2001), *Drosophila melanogaster* (*Formstecher et al., 2005)*, *Caenorhabditis elegans* (Obrdlik et al., 2004)and humans(Rual et al., 2005). Two screening approaches can be distinguished: the matrix (or array) and the library approach. In the matrix approach, all possible combinations between full-length open reading frames (ORFs) are systematically examined by performing direct mating of a set of baits versus a set of preys expressed in different yeast mating types. The classical cDNA-library screen searches for pairwise interactions between proteins of interests (baits) and their interacting partners (preys) present in cDNA libraries or sub-pool of the libraries (Bruckner et al., 2009). However, the inherent disadvantage to this approach is the increase in the rate of wrongly identified proteins (called false positives) during screening of libraries. Another disadvantage is the need to sometimes verify interaction partners by colony PCR analysis and sequencing, making such screens expensive and time consuming (Bruckner et al., 2009). Also, all identified protein partners would need to be validated.

1.4.2 Bi-Molecular Fluorescent Complementation (BiFC) Studies

This method utilizes and is based on complementary fluorescent peptides attached to each of the allegedly interacting proteins which are introduced into live cells and can be visualized by microscopy (Kerppola, 2006, 2009). The interaction of the tagged proteins will bring within proximity the complementary fluorescent tags causing them to fold into their native three-dimensional structure and emitting fluorescence. This methodology is particularly useful for validating protein interactions that have been confirmed with other methods but can also be used for postulated protein interactions. The introduction of the dual expression recombinase based (DERB) vector system into BiFC technique has the potential to upgrade it to a high-throughput method (Voehringer et al., 2009). The main strength

of the technique is that proteins are expressed at native levels and visualized *in vivo*. However, it suffers from some of the drawbacks that mark two hybrid experiments. It is difficult to make conclusions based only on the BiFC experiment since observation of interaction or lack of it may be due to the structural modifications of interacting proteins resulting from the fluorescent tag attached to them.

1.5 Affinity Purification-MS approaches

Part of the challenges in expanding the field of proteomics is elucidating protein-protein interaction networks in different organisms in order to provide insight into the mechanism of interplay between proteins and disease development in case of network disruption (Gavin et al., 2002; Kuhner et al., 2009; Taylor et al., 2009). Determining such interactions in cell is a daunting task because of several reasons, among which are the intricate and complex networks formed by proteins (Albert and Albert, 2004; Gavin et al., 2006; Grigoriev, 2003; Johnson and Hummer, 2011) and the multiple states (for instance, phosphorylation, methylation) that exist for each protein. These states depend on cellular context of the process or cellular localization and confer different interaction potential and function (Hunter and Borts, 1997; Liao et al., 1999; Plowman et al., 1999). Another reason is that relevant interactions have a wide range of affinities (Costanzo et al., 2000; Goffeau et al., 1996). Also, proteins exist with a broad range of abundance (10^1 – 10^6 copies per cell) which varies depending on cellular context and impacts on complex formation and stoichiometry (Picotti et al., 2009). In addition, there are other proteins that interact with other molecules, such as DNA and RNA. The roles played by RNA both as a template for protein expression and regulatory role has increased interest in ribonuceoproteins. Advances in affinity purification mass spectrometry (AP-MS) studies have allowed the detailed, quantitative and rapid characterisation of macromolecules (Gavin et al., 2006; Gavin et al., 2002; Kuhner et al., 2009; Puig et al., 2001). AP-MS, generally, consists of rapidly purifying samples using affinity approaches and identifying them by mass spectrometry. In contrast to techniques, such as, yeast two-hybrid screen, AP-MS can be performed in a near physiological context and interactions can be monitored in any selected cell type, following

exposure to almost any cell type (Oeffinger, 2012; Puig et al., 2001). Protein interactions that depend on post-translational modifications can be identified and the PTM itself mapped by mass spectrometry (Annan and Zappacosta, 2005; Carr et al., 2005; Garcia et al., 2005; Witze et al., 2007).

Ideally, affinity purification allows the isolation of any macromolecule in its native states with its physiological environment intact. However, there are parameters that make it challenging to isolate a complex, RNP or any given macromolecule in its native state. Experimental conditions such as, sample preparations and purification strategies are examples of these parameters (Cristea et al., 2005; Oeffinger et al., 2007; Trinkle-Mulcahy et al., 2008). Experimental conditions that need to be taken into consideration, cell lysis, and preservation of complex, transient interactions and integrity of nucleic acid components of the complex. Also to be considered is the dynamics of complexes within a network, since in many instances complexes are not static structures but a mixture of dynamic intermediates. In some instances, the choice and placement of epitopes are important points to consider for a successful AP. Finally, the identification of components of complex, including distinguishing real interactors and contaminants as well as determining the stoichiometry is another challenge (Oeffinger, 2012).

The various affinity purification approaches will be briefly discussed in the following sections. Also, the advantages and disadvantages of each approach will be highlighted.

1.5.1 Co-immunoprecipitation (Co-IP)

Co-immunoprecipitation exploits the immuno-specific properties of antibodies to selectively bind proteins or antigens that they were raised against (Lee, 2007). Antibody specific to the target protein or bait is used to precipitate out of solution the bait and other proteins, if any, associating with the bait. The antibody is at some point in the procedure coupled to a solid phase. The target proteins and associating proteins can be immunological interrogated or identified by mass spectrometry. The Co-IP procedure is a one-step purification process and hence the presence of

background proteins as a result of non-specific interactions with either the antibody or bait reduces specificity of the method. Specific steps such as stringent pre-clearing of lysate and bead washes may reduce background proteins but may also eliminate legitimate interacting proteins (Isono and Schwechheimer, 2010)

1.5.2 Chemical-Cross linking

This approach utilizes chemical cross-linkers to 'fix' proteins that bind to or interact with each other before isolation/purification and identifying them (Tang and Bruce, 2009, 2010; Vasilescu et al., 2004; Zhang et al., 2009). Complex purification or isolation may exploit the Co-IP technique or any other purification method. Identification of interacting partners can be done by interrogation with antibodies/blotting or by mass spectrometry(Doneanu et al., 2004). It is particularly popular in chromatin immunoprecipitation (ChIP) assays (Carey et al., 2009; Herzberg et al., 2007). Common cross-linkers such as the non-cleavable N-hydroxysuccinide (NHS) ester cross-linker, bissulfosuccinimidyl suberate (BS3) and the imidoester cross-linker dimethyl dithiobispropionimidate (DTBP) are used. Strep-protein interaction experiment (SPINE) is a technique that uses the reversible cross-linker formaldehyde and affinity tag for *in vivo* analysis of protein-protein interaction (Herzberg et al., 2007).

1.6 Tandem affinity Purification-tandem mass spectrometry

Tandem affinity purification (TAP) is a high throughput approach that combines two affinity purification steps to isolate a homogenous protein complex. Two different affinity tags are attached in tandem to a bait protein by integrating the tags' DNA sequence at either the C- or N-terminal of the bait gene in the cellular genome of interest. This is achieved through PCR-mediated gene targeting that exploit homologous recombination. The method was developed by Puig et al.(Puig et al., 2001).

Tandem affinity purification (TAP)-mass spectrometry (MS) has vastly improved the sensitivity of protein identification and for about a decade now, provided the means of studying protein-protein interactions in biological systems. The advantage of this approach over 2-hybrid experiments and co-immunoprecipitation studies is its high throughput capabilities and less susceptibility to non-specific interactions. Another advantage is the expression of proteins under native conditions (under the regime of host endogenous promoters) therefore approximating the conditions of *in vivo* protein-protein interactions. The use of two consecutive affinity tags separated by TEV cleavage sequence and matrices affords double purification modules further improving the purification of complexes (see Figure 1.5a). This is another advantage TAP has over Co-IP which relies on a single purification module. Furthermore, antibodies used in Co-IPs may bind non-specifically to proteins and since elution is done with competitive ligand or by denaturation, target complexes can be easily contaminated thereby reducing the selectivity of the purification. TAP, on the other hand, utilizes TEV protease to cleave the TEV sequence sandwiched between the two consecutive affinity tags achieving gentle and overall, selective elution. Because of the potential structural impediments and the often arbitrary way in which antibodies are designed against target proteins, antibodies used in Co-IPs may indeed disrupt potentially important interactions by binding domains in the target protein important in protein-protein interactions, thereby decreasing the sensitivity. TAP has the advantage in this regard because its target peptide is the TAP tag localized at either terminal of the target protein. Lastly, TAP significantly improves the purification of protein-RNA complexes, reducing background nucleic acids. In spite of all these advantages, certain precautions are taken during preparation of extracts and mass spectrometry analysis.

Genome wide TAP Tagging in Yeast

In previous studies, about 491 protein complexes have been revealed by genome wide analysis of protein-protein interactions in yeast by TAP (CBP-protA.)-MS of which 257 were previously uncharacterised (Gavin et al., 2002). 4500 proteins have been successfully TAP-tagged from about 6500 attempted proteins (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/cellzome.html). From these large scale high-throughput studies, only 1,993 TAP-fusion proteins, 88% retrieved at least one

partner (Gavin et al., 2006; Gavin et al., 2002) Therefore, about 12% of the TAP-MS either failed or the baits do not interact with other proteins. The fact that about 4500 proteins were TAP-tagged from the attempted TAP-tagging of 6500 proteins means that approximately 2000 failed to be TAP-tagged. This means there are large numbers of TAP-tagged proteins whose interactions with other proteins have not been explored.

Although previous genome-wide studies based on the TAP-MS approach served as guide for exploration of protein-protein interaction, the scale of growth culture used was low and can potentially alter the result of the TAP-MS studies. In the high-throughput genome-wide study by Gavin et al., small-scale cell culture was used for TAP-MS; typically yeast strains were grown in 2 litres yeast cultures medium. Considering that most proteins are low abundant when expressed at physiological conditions and under the regime of natural promoters, sub-stoichiometric/transiently-interacting protein partner may not be identified because of the small scale cell culture. In addition, small-scale cell culture-based TAP-MS of low abundant protein baits can result in unsuccessful TAP-MS and poor quality data with protein partners or even the bait not identified. The scale of cell culture is therefore important for studying protein-protein interactions and can substantively alter the result of protein-protein interactions. To improve the sensitivity of the TAP-MS and quality of data, optimization and scaling up cell culture is crucial to enable the identification of transient/sub-stoichiometric interactions that are not possible in small-scale cell culture-based TAP-MS studies.

1.6.1 Extract Preparation

Because TAP requires that proteins are expressed at native conditions, significant amount of biomass are required for protein purification. Expression of the target protein is an important factor and can be determined empirically by comparing its expression level by Western blotting to that of a previously TAP tagged protein. A variety of cells, tissues and steps can be used for TAP. However, cells, tissues or steps that enrich the production of the target protein over other proteins are preferred (Oeffinger, 2012; Puig et al., 2001). Optimizing yield of target proteins without compromising the integrity of protein-protein

interaction is an important consideration during TAP. Therefore, extracts are prepared in optimized conditions, and if possible, simulate the intracellular conditions of the target protein. Therefore, pilot experiments and considerable exploration of the literature around the target protein are recommended for optimization of extract preparation procedure. There are, however, cases where it is impossible to meet these conditions, because interaction of the target protein with some structure may be difficult to break without destabilizing potentially functional interactions of the target protein (Puig et al., 2001)

During extract preparation, care must be taken not to introduce conditions that may destabilize protein-protein interactions as the process may be irreversible even if conditions are restored further downstream. Temperature of 4°C is recommended and gentle handling of samples is recommended (Puig et al., 2001) .

1.6.2 Conditions of Tandem affinity purifications

The choice of buffers (salt, detergent, etc.) is often adapted to conditions best suited to target proteins/complexes. The interaction of intracellular calmodulin with the calmodulin binding peptide (CBP) is a potential concern and theoretically possible, although this has not been observed. Yet if this is the case, the inclusion of EGTA during extract preparation should deter such interaction. Also there are cases where endogenous proteins of mammalian cells interact with calmodulin in calcium-dependent manner (Agell et al., 2002; Head, 1992). A replacement of the CBP affinity tag with other affinity tags such as, FLAG tag (Knuesel et al., 2003), can resolve this. Extensive washing in the two purification module should be included but at the same time being mindful that washing buffers do not destabilize functional but weak interactions.

1.6.3 Analyses of data: controls and verification

Performing TAP experiments requires the use of a number of controls. One control is the use of parent strain (wild type) that does not express TAP proteins. Although, this is an important control it is not sufficient to demonstrate the specificity of the TAP target protein-protein interaction. This is because there are interactions that may be a result of aggregation of non-specific proteins with target protein or binding of proteins that associate with unfolded proteins (Collins et al., 2007; Gavin et al., 2002; Krogan et al., 2006; Xu et al., 2010). Fortunately, TAP experiments from previous studies are often instructive; they have identified and catalogued a number of proteins that may be potential contaminants, which include abundant proteins (Translation factors, ribosomal proteins) and proteins known to associate with unfolded protein domains (such as proteasome, heat shock proteins). Usually these proteins are flagged as common contaminants pending further verification using an independent experimental procedure (Gavin et al., 2006; Gingras et al., 2007).

There is also the challenge of distinguishing non-specific interactions from weakly binding or transient interaction but functional interactions using TAP experiments. This is often overcome by including mild washing steps and again comparing the present data with data from previous TAP experiments which are often instructive in identifying potential non-specific interactions (Bouveret et al., 2000; Gingras et al., 2007; Oeffinger, 2012; Puig et al., 2001; Rigaut et al., 1999; Schlosshauer and Baker, 2004).

1.7 Tandem vs single step affinity purification approaches

Tandem MS approaches are not designed to monitor very transient or labile interactions (typically capturing interactions with K_d higher than the mid nM range (Schlosshauer and Baker, 2004)). To address these limitations and capture more transient interactions, shorter protocols with single step purification instead of two have been designed (Figure 1.5b). It was believed that single step approaches may lead to significantly higher background, however, several studies in yeast, mammalian and viral systems have shown that cryolysis, rapid sampling and use of low background resins such as magnetic beads over agarose/sepharose beads can significantly reduce background to a manageable level and yielding as clean a

sample as any TAP while preserving transient or weaker interactions (Cristea et al., 2005; Hubner et al., 2010; Oeffinger et al., 2007). Moreover, many of these protocols use less starting materials than any commonly used TAP since sample loss is also minimized by the simple step strategy (Oeffinger, 2012).

Another requirement for clean and efficient ssAP is an epitope tag with a high specificity and low K_d to its ligand. Several have been used so far which include, Protein A, HA, FLAG and GFP. Protein A has a remarkable high affinity for rabbit IgG making it ideal for rapid isolation of complexes. Despite the large size of the Protein A tag currently used in the literature (~15.3 kDa, containing 2 copies of z domains of protein A), it is innocuous to most (~95%) proteins and so far more than 300 proteins have been tagged with Protein A (not CBP-Prot A) in yeast (Archambault et al., 2004; Marelli et al., 2004; Oeffinger et al., 2007; Strambio-de-Castillia et al., 1999). Protein A is also readily removed from IgG using salts and denaturants, making the elution of complexes both straight-forward and economical. GFP, although widely applied to in vivo visualisation of proteins, can also be used in ssAP, and recently, it has been successfully applied in the purification of proteins. One advantage of this approach is that it can be used to visualise proteins in living cells and their interactions captured by ssAP procedure from the same culture. Given the wide use and availability of GFP tagged protein reagents for many organisms this tag will be an ideal tool for AP studies. The availability of reliable, high-affinity antibodies will be a problem, particularly ones that have not been pre-conjugated to a resin potentially at low densities thereby increasing background levels (Oeffinger et al., 2007; Rothbauer et al., 2008; Trinkle-Mulcahy et al., 2008). Comparison of the two affinity purification steps are illustrated in Figure 1.5

Also, affinity purification approaches enable quantitative analysis of protein complexes. One of the most common quantitative approaches is SILAC which involve metabolic labelling of proteins carried out in vivo. This results in the replacement of essential natural amino acids with amino acids labelled with heavy isotopes, during protein synthesis in the cell (Busk et al., 2011; Gouw et al., 2010). This leads to a difference in mass for tryptically digested peptides compared with the control which can be detected by MS (Gouw et al., 2010).

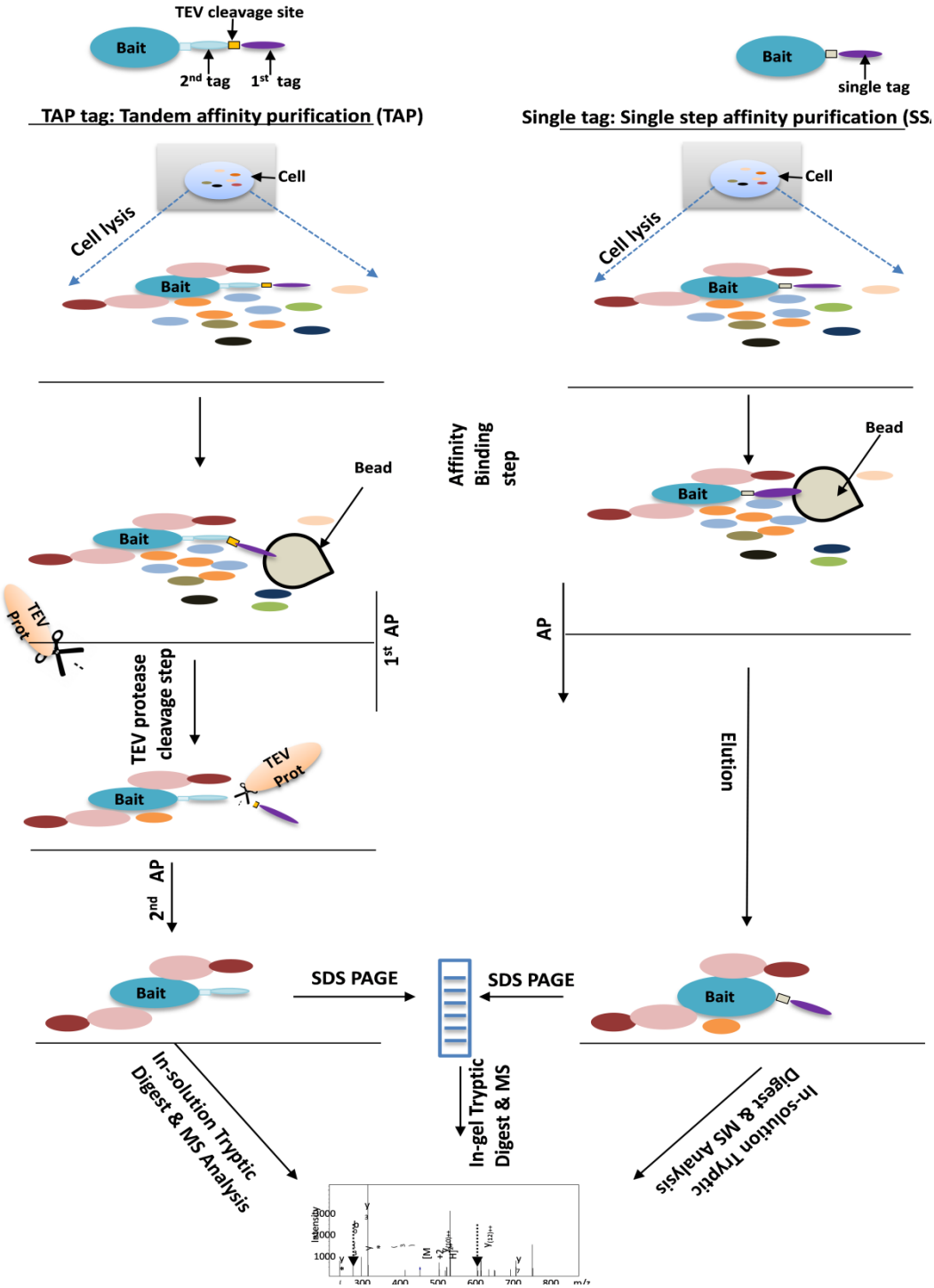


Figure 1.5 Diagramatic illustration of single step vs tandem affinity purification. Tandem affinity purification (TAP) and single step AP (SSAP) are illustrated on the left and right panels, respectively. Bait, bait-interacting and cellular proteins are illustrated with coloured spheres and affinity bead by a tear-drop shape. TAP involves two step affinity purifications: in the first AP, TEV protease cleaves the TEV cleavage site between the first and second tag to release bait-2nd-tag and specific interacting protein partners from affinity bead. In the second step both TEV protease and some other potential contaminants are eliminated and protein complex eluted. In SSAP, step up to affinity binding is same as TAP but only one affinity purification step is involved. Protein complexes derived from both steps are either analysed by SDS-PAGE in conjunction with in-gel tryptic or by in-solution tryptic digest MS analysis.

1.8 Approaches to purification and analysis of Protein-RNA Interactions

Research in proteomics and protein-DNA interactions have yielded insights and novel approaches to the study of protein-RNA interactions (Gavin et al., 2006; Jones et al., 2001; Lunde et al., 2007; Luscombe et al., 2001; Vogel et al., 2007). However, DNA-protein recognition has been shown to differ considerably from RNA-protein recognition. This will not be difficult to see when the diversity of RNA structures is taken into hand. The differences are further highlighted by the fact that greater percentage of RNA-protein interactions involves protein recognition of the RNA base edge and sugar rather than of phosphate backbone (Morozova et al., 2006).

Structural studies using NMR and crystallography have provided atomic resolution structures of RNAs and RNA- processing proteins (Kvaratskhelia and Grice, 2008). However, most biologically relevant protein-RNA complexes are not so amenable to NMR and crystallography. New approaches, prominently featuring mass spectrometry, promise to arm molecular biology with a tool for studying protein-RNA interactions. The advantages of mass spectrometry-oriented approaches include inspection of protein-RNA interaction in biologically relevant conditions as well as its capacity to maximize very limited protein-RNA samples (Kvaratskhelia and Grice, 2008).

Mass spectrometry is carried out in concert with other techniques. The strategy couples affinity pull down or Tandem affinity-tag (TAP-tag) techniques to HPLC-MS technology. In addition, other complementary techniques such as SDS-PAGE, and methods that map RNA-protein contacts by exploiting primary amine accessibility to (NHS)-biotin reagent (Kvaratskhelia and Grice, 2008), are used to understand different aspects of the Protein-RNA interaction. To facilitate proper identification, and because of the difficulties encountered in adapting direct MS-analysis of RNA, it is important that the components be isolated in relatively pure forms. Hence, separation of the protein and RNA subunits in the complex prior to the MS-analysis is very important. RNA separation from proteins can be achieved using phenol-chloroform extractions. RNAs in the complex can then be separated using PAGE and further digested in gel to prepare them for MS (Taoka et al., 2010). SDS-PAGE is used

to separate the proteins in the complex according to their masses. However, a novel and more convenient technique employing denaturing HPLC- can be used to separate proteins and RNAs in complexes. This approach has been demonstrated recently with bacterial CRISPR system and potentially useful in analysis of other protein-RNA complexes (Dickman and Hornby, 2006; Waghmare et al., 2009). This will also facilitate easy digestion of the components without resorting to the somewhat laborious process of in-gel digestion. Further separation of the RNA fragments and protein digests using liquid chromatography is coupled to MS for analysis.

1.9 Mass Spectrometry

The first mass spectrometer (MS) was developed close to a century ago as a result of the exploration of canal rays by J.J. Thompson, a study that was continued by his student, F.W. Aston (Budzikiewicz and Grigsby, 2006; Thomson, 1913). A mass spectrometer measures the mass to charge ratios of ions in the gaseous phase. The ioniser produces ions from the sample and the ions are passed to the analysers which accurately measure and separate ions according to their size. The detector records the mass of the ion emerging from the last analyser. The computer processes the data into a suitable mass spectrum and provides feedbacks and controls the mass spectrometer. There are different types of ionisers among which include, electrospray ionisation (ESI), fast atom/ion bombardment (FAB), electron ionisation (EI), electron spray ionisation, matrix assisted laser desorption (MALDI) and chemical ionisation (Fenn et al., 1989; Karas and Hillenkamp, 1988; Wait, 1993). However, significant advances in the application of MS are largely due to the soft ionisation technologies of ESI (Fenn et al., 1989) and MALDI(Karas and Hillenkamp, 1988), which allow the stable transfer of large, polar biomolecules to gaseous phase. John Bennett Fenn awarded the Nobel Prize in Chemistry in 2002 for the development of ESI for the analysis of biological macromolecules. Advances in analysers have also greatly contributed to the emerging applications of MS. Different analysers that exist today include quadrupole, quadrupole ion trap, time-of-flight (ToF), time-of-flight reflectron, quad-ToF and fourier transform ion

cyclotron resonance (FT-ICR) mass spectrometers. In the analysis of protein and RNA species in this study, ESI-MS utilizing Q-TOF and 3D ion trap instruments was used.

1.9.1 The ESI ion source

One of the features that make ESI attractive in biology is its soft ionisation capability. In the ESI system, ions are transferred from sample solution to the gaseous phase into the MS at atmospheric pressures. The metal coated glass capillary needle, maintained at voltages from 2000-4000 V), is used to spray the sample into the MS (Figure 1.6). The evaporated ions enter into the vacuum system of the mass spectrometer through series of sampling apertures that compartmentalize subsequent stages of the gas trajectory.

As the sample solution enters into the electric field set up between the capillary and a counter electrode positive or negative ions depending on the polarity of applied voltage, are concentrated towards the tip of the capillary. These charged ions are drawn out of the capillary surface tip into a shape known as 'Taylor cone'. This cone is further drawn out into a filament that subsequently form charged droplets. The droplets fly toward the counter electrode, which is opposite in polarity to their own. During this flight, a heated nebulizing or neutral carrier gas such as nitrogen causes the droplets to evaporate into the gaseous phase. The droplet shrinks in size and as a result increases the droplets surface charge density. The concentration of field densities causes the droplets to take on a tear-shape. At a point where surface tension is less than electrostatic repulsion due to like-charges, a Rayleigh limit is reached which causes the droplets to split into finer droplets. Depending on the size of the parent droplet, the droplet can either be finer droplets which repeat the process or discrete solvated surface ions. These ions are stripped of their solvent clusters by collision with surrounding gases, resulting in multi-charged ions (Lane, 2005)

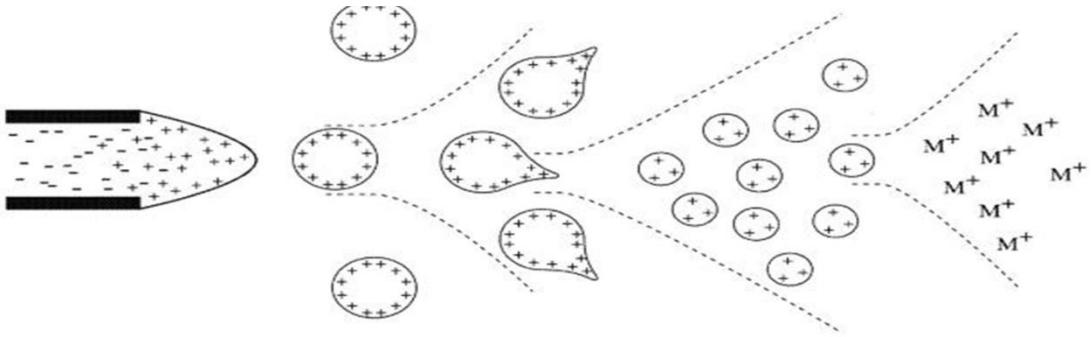


Figure 1.6 Mechanism of Electrospray ionisation Within an ESI source, a continuous stream of sample solution is passed through a stainless steel or quartz silica capillary

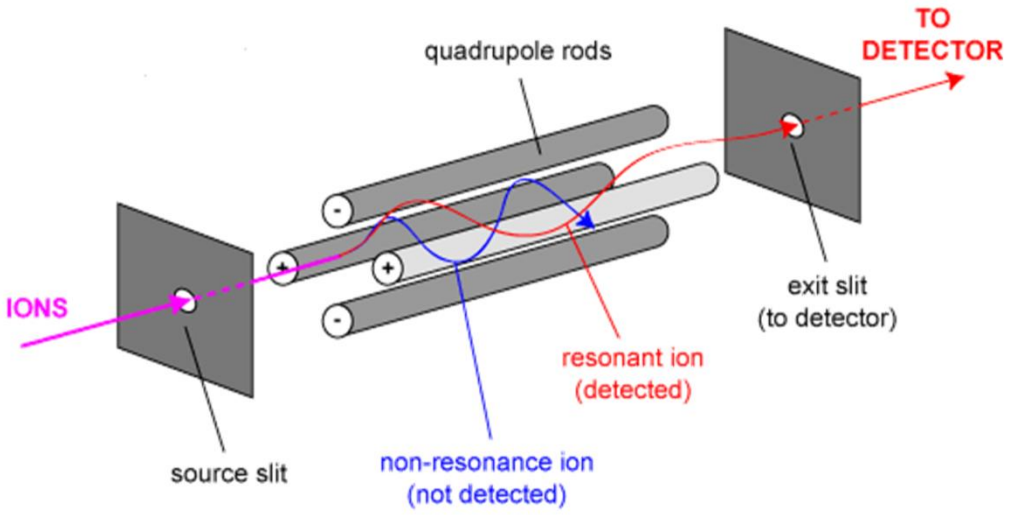


Figure 1.7 A Schematic diagram illustrating the mechanism of the Quadrupole mass filter. Resonant ion within the m/z range reaches the detector, the non-resonance ion are deflected and are not detected.

1.9.2 Mass analysers: the quadrupole and Ion trap mass analysers

The difference between quadrupole and ion trap systems is that the latter, first traps the ions then analyses them; both however, operate on the same basic principles, using radio frequency (RF) and direct current (DC) voltages for the separation of ions. In fact, ion traps are sometimes referred to as 'quadrupole ion trap'. Ion trap can be imagined as a quadrupole bent around itself to for a loop (Lane, 2005). Quadrupole mass analyser consists of four rods precisely placed parallel to each other and equally spaced around a central axis (see Figure 1.7). Direct current (DC)

and radio frequency (RF) voltage is applied to each pair of rods set opposite to each other, such that one rod in a pair is positive and the other negative. Depending on their mass/charge (m/z) ratio, ions streaming through the central axis between the poles are filtered. The oscillating RF voltage enables this filtration of ions by maintaining ions with a particular m/z ratio or within the scan range in a linear trajectory towards the detector and deflecting other ions so that they collide with the poles.

1.9.3 Time of Flight (TOF) analyser

An ion with charge, *z*, under an accelerating potential *V* in a TOF system is in principle given energy of *zV* which is equal to the kinetic energy of the ion, mv_2, Where m= mass and v=velocity. If all ions are accelerated with the same force, these ions having the same charge but different mass will be separated based on their different velocities. The TOF system has a long metal tube through which ions travel to the detector. Since ions with different mass travelling at different velocities (velocity (v) = distance (d)/time (t)) will take different amount of time to travel down the length of the metal tube, m/z ratio are measured based on the time it takes for the ions to travel the length of the tube which is field free. However, mass resolution can be affected by variations in flight times and factors that create a distribution in flight times among ions with the same m/z ratio will result in poor resolution (Lane, 2005). Reflectron, which consists of a series of electrostatic mirrors with the same charge, is often used to resolve this. Basically, an electrostatic field is applied at one end of the flight tube which retards the ions flying towards the end of the tube due to field effect. The reflectron then reflects the ions out of the mirror and back into the detector in opposite direction to the source. Ions with higher kinetic energy are retarded more as they penetrated deeper into the field and will not be readily reflected towards the detector. Hence, ions with lower kinetic energy are retarded less and reflected sooner toward the detector. Greater resolution can be achieved by interfacing quadrupole with TOF. It is the basis of the Q-TOF which consists of three quadrupoles and a TOF mass analyser. The first quadrupole serves to focus the

ions generated in the ESI; hence ions in the whole m/z spectrum are transmitted to the second quadrupole, which is really the first analyser in the system. This first analyser scans and effectively selects a narrow m/z range at any one time for transmission to the third quadrupole, which is the collision cell. The ions that enters into the third quadrupole or collision cell, known as the precursor ion, is bombarded with electrons and collision induced dissociation occurs resulting in the precursor ions breaking up into fragment ions. The resulting fragment species are used to give information on the identity of the precursor, by analysing their fragmentation patterns.

1.10 Mass Spectrometry analysis of proteins and RNA

1.10.1 Mass Spectrometry Characterization of Proteins

By analysing the fragmentation pattern of peptides in a mass spectrometer it is now possible to determine the sequence of a protein. In a tandem mass spectrometry, the first mass spectrometer's analyser scans and isolates one peptide species and dissociates it with a neutral gas such as nitrogen or argon. The fragmented or dissociated ions are then separated in the second analyser of the mass spectrometer to produce a tandem mass spectrum. Two major types of collision induced fragmentation used are the low-energy induced fragmentation and the high-energy induced fragmentation. In the former, a single collision is not sufficient to break a peptide bond while the latter can generate peptide bond cleavage in a single, high-energy impart. Most instruments today use multiple low-energy collision to generate peptide fragmentation (Mann et al., 2001). In a peptide backbone, there are several possible bonds that can be cleaved in a collision to generate fragment ions. The most common fragment ions are the *b*- and *y*-ions which are so called depending on whether the fragmentation charge is retained on the amino terminal (on the terminal carboxyl (COO^+) group) end of the peptide or on the carboxyl terminal (the terminal ammonium (NH_3^+) group) end of the peptide, respectively .

1.10.2 Identification of peptide and proteins

Peptide sequence and protein identity are determined from the precursor and product ion spectra generated during tandem MS experiment. Theoretical mass distributions of fragment ions can be predicted since peptide ions fragment in a predictable fashion when it undergoes collision induced dissociation (CID) (Hunt et al., 1986). In Table 1.1, the mass values of an example peptide is given. In Figure 1.8, the theoretical distribution of product masses which can be matched with actual product ion spectra of this peptide is shown.

Mass spectra can be interpreted de novo, to establish the sequence of a peptide without matching it to a theoretical fragment ion distribution. However, this is rarely done in proteomics workflow. Instead, experimentally derived ion spectra are probabilistically matched to entire databases of theoretical spectra. Database of protein sequences for a particular organism can be constructed from genomic data of the organism. Since proteolytic enzymes used in the generation of peptides act on specific amino acid sequences it is therefore possible to generate a database of theoretical peptides and peptide ions. Because it is also possible to predict ionisation and fragmentation of peptides during mass spectrometry, database of fragment ions for an entire sequenced organism can be created. MASCOT (Matrix Science) is one of the software used in matching acquired MS/MS data to such fragment ion databases (Perkins *et al.*, 1999). Ideally, a given spectra should match a single spectra. However, the quality of the spectra (b and y-ion coverage) may determine the number of peptide that is matched. MASCOT match possible matches according to how well the experimental data fits the predicted fragment ion. In the example shown in Table 1.1 and Figure 1.8, the spectra for ILVLALEDLK (Prp8 peptide sequence) also match peptide ILVKNLDRR.

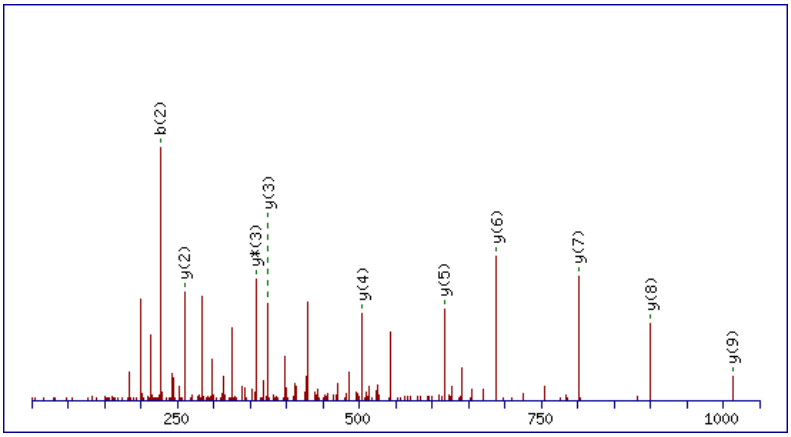


Figure 1.8 CID MS/MS spectra of the peptide ILVLALEDLK. The most abundant fragment ions are highlighted.

Table 1.1 Theoretical fragmentation pattern of peptide, ILVLALEDLK This peptide has a mass of 1125.7009 Da and an m/z (for a doubly charged species) of 563.857. The y-series highlighted in red corresponds to the y-series shown in Figure 1.8.

b-ions		y-ions		y-ions fragment mass	b-ions fragment mass
1	I	10			114.0913
2	IL	9	LVLALEDLK	1013.6241	227.1754
2	ILV	8	VLALEDLK	900.5401	326.2438
4	ILVL	7	LALEDLK	801.4716	439.3279
5	ILVLA	6	ALEDLK	688.3876	510.3650
6	ILVLAL	5	LEDLK	617.3505	623.4491
7	ILVLALE	4	EDLK	504.2664	752.4917
8	ILVLALED	3	DLK	375.2238	867.5186
9	ILVLALEDL	2	LK	260.1969	980.6027
10		1	K	147.112	

MASCOT searches these peptide sequences against a proteolytic peptide database to match the identified peptide to a specific protein. MASCOT uses various statistical parameters such as the score of the matched peptides, number of peptides

matched, etc., to assign scores for the identification of the protein which represents the confidence of protein identification based on matching experimental peptide ion spectra to the theoretical database.

1.11 Mass Spectrometric Characterisation of RNA

A number of methods are currently available for the identification of RNAs among including Northern blotting, microarrays and reverse transcription (RT)-sequencing based techniques. The former two techniques require prior knowledge of the nucleotide sequence under study. RT-sequencing based techniques, on the other hand, do not require prior knowledge of the whole sequence. Traditional laboratory techniques such as PCR, in conjunction with new "next generation" sequencing technologies can massively parallel sequence huge numbers of nucleic acids and provide quantitative information. However, these approaches are not able to provide valuable information on the original modifications [Geissing AMB et al., 2012]. In addition more recent direct RNA sequencing methods are also unable to detect such modifications [Ozsolak et al., 2009]. Mass spectrometry (MS)-based techniques complemented with conventional RNA analytic methods provides a sensitive and easier alternative for direct characterization of RNA. In addition to being able to measure spectra at the subfemtomole levels, the main advantages are its potential to highlight post-transcriptional modifications and contact between protein motifs and nucleic acids of RNA (Kvaratskhelia and Grice, 2008; Taoka et al., 2010). Intact masses of small RNAs (up to 60mer) can be accurately measured with mass spectrometry. Even larger RNA molecules can be mapped with the aid of sequence-specific nucleases such as RNAse A and RNAse T1. However, limited success has been recorded for de novo sequencing of RNA by mass spectrometry, mainly because of the difficulty in interpreting the complex fragmentation pattern of RNA during tandem mass spectrometry analysis(Schurch et al., 2002; Schurch et al., 2007).

1.12 Biology and Biochemistry of the Spliceosome

The spliceosome is one of the most widely studied complexes using interactomics approaches. A number of approaches have been used to study the spliceosome and

include, TAP-MS, Co-IP and yeast two-hybrid screening. Electron microscopy and NMR are among the structural tools that have been used to study the spliceosome (See Figure 1.9) (Lührmann and Stark, 2009). In the following sections, the biology of the spliceosome is discussed.

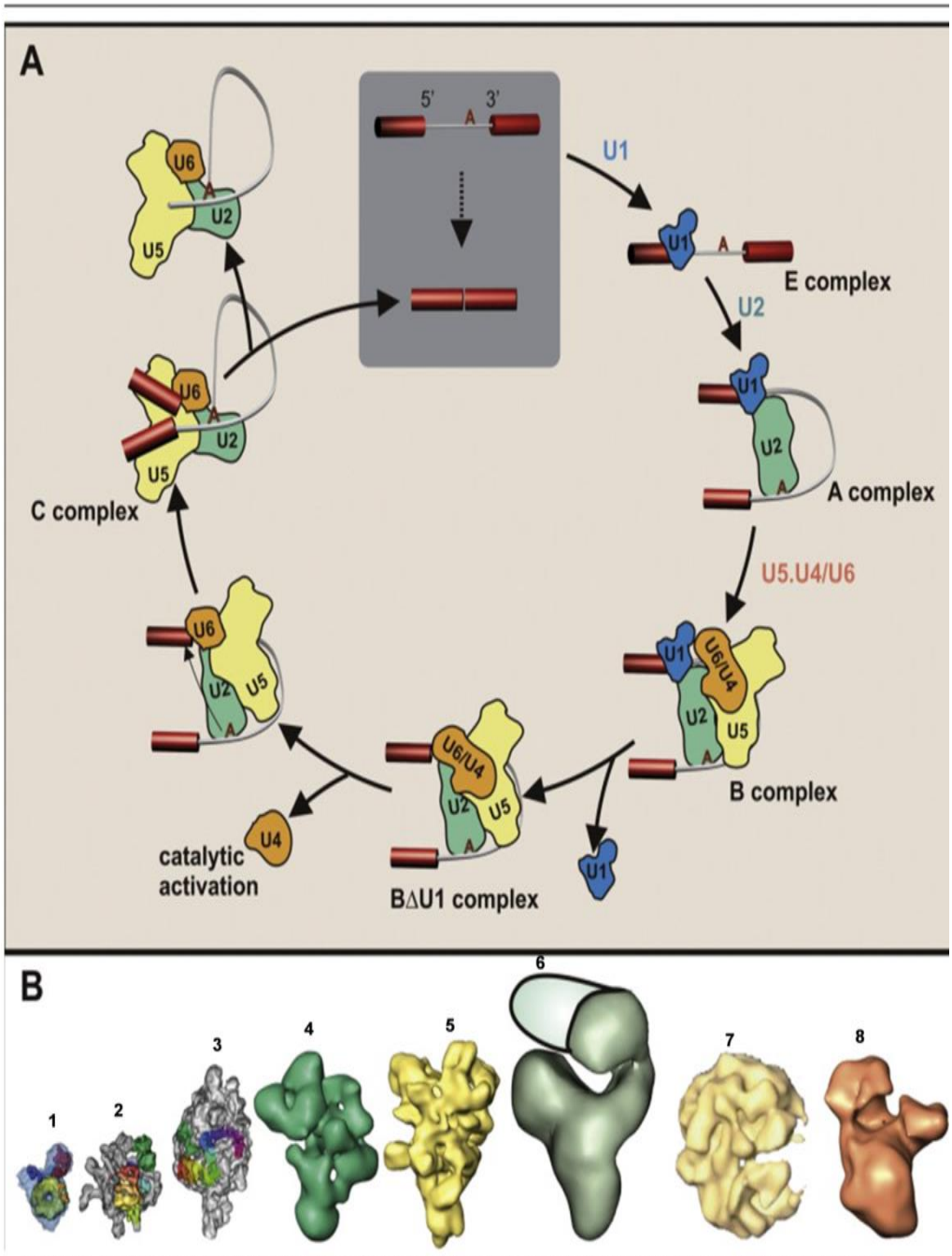


Figure 1.9 Schematic representations of pre-mRNA splicing and the spliceosome assembly pathway. Introns are excised from pre-mRNA by the spliceosome, which is assembled by the stepwise integration of U1, U2, and U4/U6.U5 snRNPs. **(B)** The currently available EM 3D structures of spliceosomes and spliceosomal components. From left to right: U1 snRNP [1], SF3b [2], U11/U12 di-snRNP [3], U5 snRNP [4], U4/U6.U5 tri-snRNP [5], BΔU1 spliceosome [6], 'native' spliceosome [7], and the C complex [8]. Taken from Lührmann and Stark, 2009.

1.12.1 Importance of the Spliceosome

The process of protein production in the cells of higher organisms is complex and involves many steps. The first step in this process is transcription which involves re-writing genetic information for protein from DNA into a working copy, the precursor messenger RNA (pre-mRNA). However, pre-mRNA contains regions that are not needed for the production of proteins called introns. These regions must be precisely excised and the remaining regions (called exons), which contain information for protein production joined together. This process is referred to as pre-mRNA splicing and represents the maturation process for eukaryotic messenger RNA (mRNA). Only mature mRNAs are transported from the cell nucleus into the cytoplasm and used by the ribosome as a template for protein production.

The presence of exons and introns is a great advantage for an organism, as different combinations of exons from a given pre-mRNA species can be chosen to be included in the mature mRNA. This strategy, referred to as alternative splicing, ensures that mRNAs corresponding too many different proteins can be made from a single gene. This is not only economical but represents an additional level of regulation of gene expression. This explains why humans are able to cope with only about 20,000 protein-encoding genes in their genomes. Understanding splicing at the molecular level is of great medical importance as aberrant splicing results in many human diseases.

The splicing of pre-mRNA takes place in two trans-esterification steps carried out by the macromolecular machinery, the spliceosome. There are over 100 proteins present in the spliceosome and about five small RNA molecules, the snRNAs, U1, U2, U4, U5 and U6. Many of the spliceosomal components are organised into smaller, stable sub-complexes whose association is highly dynamic. An example is the formation of small nuclear ribonucleoproteins (snRNPs), such as U1 and U2 snRNPs, and the U4/U6.U5 complexes from the stable binding of about 50 proteins and the snRNAs (Wahl et al., 2009; Will and Lührmann, 2011).

1.12.2 The Biology of the spliceosome

The splicing of pre-mRNA introns to form mature mRNA is a common phenomenon in almost all eukaryotes and is catalysed by a mega-dalton ribonucleoprotein (RNP) complex (reviewed by (Brow, 2002; Will and Lührmann, 2011). Although the isolation of a spliceosome complex with all five snRNPs in the past led to the hypothesis that pre-assembled forms of the spliceosome exist (Stevens et al., 2002), it is widely recognized that spliceosome is a highly dynamic structure, which assemble and dis-assemble in a step-wise manner (Fourmann et al., 2013; Will and Lührmann, 2011). The dynamic composition and conformation of the spliceosome confers splicing flexibility for the large and varied population of mRNAs that exist in the cell. It also ensures the high-fidelity intron-splicing that is the hallmark of spliceosome-mediated splicing. Two unique but coexisting spliceosomes can be distinguished in most eukaryotes: the U2-dependent spliceosome, which catalyzes the removal of U2-type introns, and the less abundant U12-dependent spliceosome, which is present in only the minority of eukaryotes and catalyses the removal of the rare U12-type class of introns (reviewed by Patel and Steitz, 2003).

The U2-dependent spliceosome are assembled in a stepwise manner from U1, U2, U5, and U4/U6 small nuclear ribonucleoprotein (snRNP) complexes and numerous non-snRNP proteins. Each snRNP consists of snRNA (two in the case of U4/U6), a common set of seven Sm proteins (B/B', D1, D2, D1, E, F, and G) and a variable number of particle-specific proteins. The snRNA U6, unlike the other spliceosomal snRNAs, does not associate with the Sm proteins; they associate specifically with a protein heteroheptamer complex consisting of seven proteins (Lsm2-8) out of the eight (Lsm1-8) proteins that are homologous to the Sm proteins (see Figure 1.10)(Bouveret et al., 2000; Tharun et al., 2000; Zhou et al., 2013).

Although, the spliceosome accomplishes intron splicing in two trans-esterification reactions, the process that leads up to the splicing reaction is more complex. The spliceosome snRNPs are assembled, splices off intron and disassemble in stages. These stages are classified as the earliest Complex E, Complex A, pre-activated Complex B, the activated complex B-act, the catalytic B* and the post-catalytic complex (Wahl et al., 2009; Will and Lührmann, 2011).

The spliceosome is a protein-rich complex and about 50 snRNP and 100 non-snRNP proteins have been identified in humans by mass spectrometry (Wahl et al., 2009; Will and Lührmann, 2011). However, the number of proteins is lower in yeast mainly because some metazoan spliceosomal proteins are involved in alternative splicing, a process that is apparently absent in yeast (Will and Lührmann, 2011). For instance, there are approximately 60 proteins and 50 proteins are identified in yeast Complex B and Complex C compared to approximately 110 proteins identified in metazoan B and C complexes (Fabrizio et al., 2009b; Will and Lührmann, 2011).

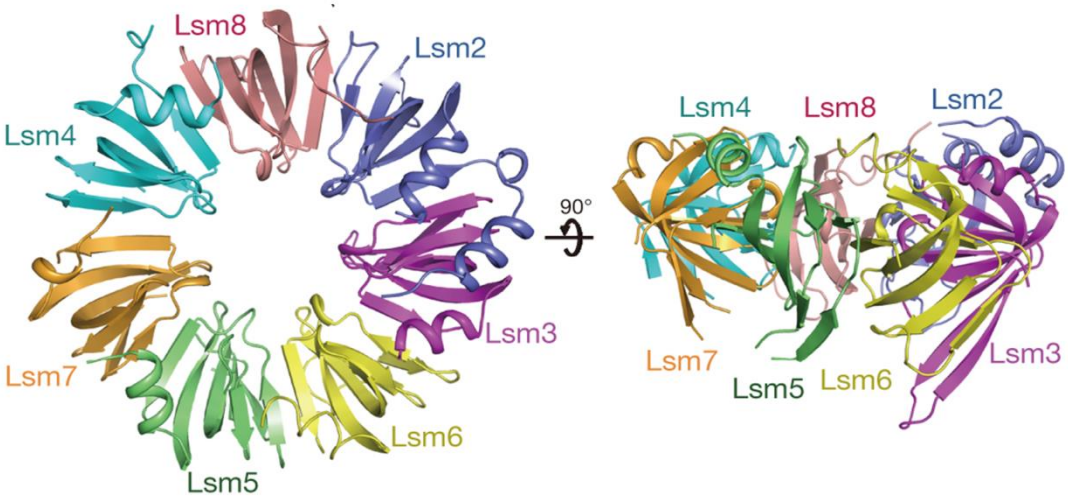


Figure 1.10 Structure of the Lsm2–8 heptameric complex. Overall structure of the Lsm2–8 complex in two perpendicular views. Taken from Zhou *et al.,* 2013.

1.12.3 Mechanism of the spliceosome assembly, splicing and disassembly

The mechanism of the U2-dependent splicing depends on information provided by short, conserved sequences at the 5' splice site (ss), 3' ss and the branch region (BR) of the pre-mRNA (Kolossova and Padgett, 1997; Will and Lührmann, 2005). The BR is

typically located 18-40 nucleotides upstream from the 3'ss. Introns are removed by two consecutive transesterification reactions. The earliest event of the assembly known as the complex A formation occurs in two broadly defined steps. First, U1 recognizes the intron's 5' end known as the 5'splice site. Second, the interaction and binding of the U2 snRNP with the branch site of the pre-mRNA (Kolossova and Padgett, 1997; Wahl et al., 2009; Will and Lührmann, 2005, 2011). According to the present model, a pre-formed U4/U5.U6 tri-snRNP complex then binds the complex A to form the penta-snRNP spliceosome known as complex B (Will and Lührmann, 2011). This is followed by a crucial spliceosomal remodelling caused by structural RNA rearrangement. This structural rearrangement leads to major compositional changes which includes U1 and U4 dissociation (Will and Lührmann, 2011). The structural and compositional changes are crucial because they lead up to the activated spliceosome, known as complex B-act. This complex is not yet catalytically active. The spliceosome is activated for catalysis by the association of the RNA dependent ATPase of the DEAH box family, Prp2p with concomitant hydrolysis of ATP. The catalytically active spliceosome, named Complex B*, catalyses the first trans-esterification reaction. The 2' OH group of the branch adenosine in the intron carries out a nucleophilic attack on the 5' ss, cleaves it, generates a 5' exon and ligates the 5' end of the intron to the branch adenosine to form a lariat structure containing the 3' exon (Will and Lührmann, 2005, 2011). This complex is known as complex C. The complex C is responsible for catalysing the second trans-esterification reaction. In this reaction, the intron is further cleaved but at the 3' ss in a tandem step that ligates the 3' exon to the 5' exon. Subsequent to the completion of the second splicing step, the spliceosome snRNPs disassemble and after further remodelling take part in another round of splicing. A summary of the yeast spliceosome assembly pathway is shown in Figure 1.9A. Note that terms used in human spliceosome are retained throughout this work for clarity and to avoid confusion.

The snRNA-pre-mRNA complex plays a crucial role in determining the spliceosome overall structure and ensuring that the pre-mRNAs partners are precisely aligned to maintain overall splicing accuracy (Nilsen, 1994). Although the spliceosomal snRNAs arguably play the most important roles in the splicing reaction, especially U6 and U2

form the splicing core catalytic site, it is however noted that the protein components do not merely play structural roles. The spliceosome proteins have been recognized to play both important recognition and catalytic functions (Pyle, 2008). Some of these proteins also play crucial roles in selecting the intron substrate and alternative splicing (Cáceres and Kornblihtt, 2002).

The spliceosome assembly pathway and the significant features of its catalytic chemistry appear has been intensively studied in metazoans and yeast for about two decades now. The high conservation between yeast and metazoan spliceosome is now recognized. The dynamic features of the spliceosome and the dramatic re-arrangement that takes place have been examined in a number of different studies (Fabrizio et al., 2009b; Fourmann et al., 2013; Tsai et al., 2005). However, the nature and extent of these changes is still not clear and have not been fully characterised. Lsm4 protein is one of the the eight proteins that binds U6 snRNA. According to the present model, almost all Lsm proteins dissociate from U6 as the spliceosome approaches the Complex B-act stage. Hence, baiting one of the Lsm proteins will pull down complexes from only U4/U6 to B-complex U4/U6.U5. Therefore, all complexes found in early spliceosomal complex E, A stages and later stages, complex B-act, B* and down-stream will be excluded if the model is accurate.

1.12.4 Compositional Dynamics of the Spliceosome during the Splicing Cycle

According to the present model, the spliceosome penta-snRNP complex observed in complex B has all the snRNP components and the Prp19/Nineteen complex (NTC) and NTC -related proteins. However, massive protein compositional changes are observed during the transition from B-complex to the activated spliceosomal B-act complex and it is suggested that approximately 35 proteins, including among others all U1 and U4/U6 proteins as well as some U5 proteins dissociate ((Fabrizio et al., 2009b). However, it is not clear whether these proteins are destabilized / released concomitantly or in discrete steps or how many discrete RNP modelling event exist during these massive compositional changes (Fabrizio et al., 2009b; Will and Lührmann, 2011). The U6 snRNA is thought to have lost most of its pre-activation binding partners and engages in novel base pairing with U2 with a concomitant RNA-

protein interactions taking place (Wahl et al., 2009). See Figure 1.11 for the compositional dynamics of the spliceosome.

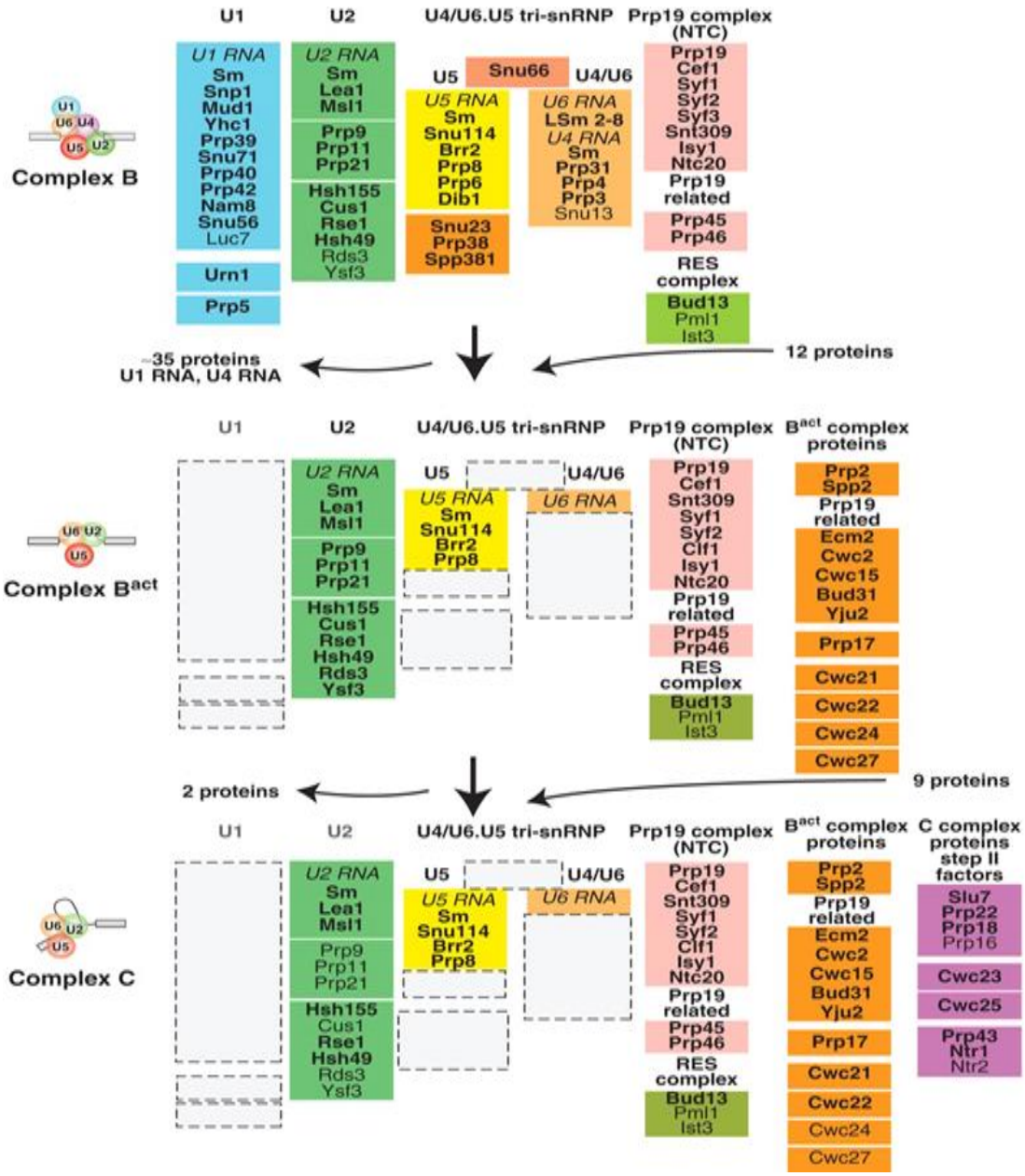


Figure 1.11 the compositional dynamics of spliceosome. This illustrates the compositional changes that take place during the splicing cycle. Taken from Will and Lührmann, 2011.

1.12.5 Current model for U4/U6 complex and the B-complex U4/U6.U5 complex

The U4/U6 complex and the spliceosome complex B has been extensively studied and characterised (Fabrizio et al., 2009a; Will and Lührmann, 2011). However, because the spliceosome is highly dynamic the presence of common proteins across the various splicing stages complicates the precise classification of these proteins. Recently, the various stages of the splicing cycle and some of the snRNP components participating during transition through these stages were characterised utilizing *in vitro* strategies (Fabrizio et al., 2009b) (See schematic, Figure 1.9 and electron micrograph, Figure 1.9B).

Aims of study

The splicing of pre-mRNA introns to form mature mRNA is a common phenomenon in almost all eukaryotes and is catalysed by the spliceosome, a mega-dalton ribonucleoprotein (RNP) complex. The dynamic features of the spliceosome and the dramatic re-arrangement that takes place have been examined in a number of different studies. However, the nature and extent of these changes is still not clear and have not been fully characterised.

It is proposed to use TAP-MS as an analytical tool to study spliceosome assembly, their links to other pathways and potential novel interactions. Although we have a great deal of information about factors involved in splicing, we do not however have profound structural information about the spliceosome. A crystal structure is the ideal, but this challenged in major part by the large size and dynamic nature of spliceosome, which undergoes major remodelling during its assembly. Another reason is the extreme difficulty in obtaining sufficient purified spliceosome complex to undertake these structural studies.

The advantage of TAP-MS strategy is that it has the potential of capturing several dynamic remodelling events during splicing. When targeted at different aspects of the splicing steps, TAP-MS has the potential of offering insight not only about the

structure but also about the remodelling events during splicing. This is because TAP-MS analyses of the target bait and co-purifying proteins/ribonucleoproteins will potentially yield insight into the spliceosomal structure and the sub-complex where they exist. It also has the potential of isolating and dissecting an event during splicing if the chosen bait participates in a single event during the splicing cycle. Taken together the information obtained in TAP-MS of different aspects of splicing will potentially offer important structural details and dynamics of the spliceosome.

Although, TAP-tagging has been performed on a genome-wide scale in yeast (Gavin et al., 2002) it was not successful for a wide number of low abundant spliceosome proteins such as Prp5 and Hsh155. In addition, many TAP MS studies have only been performed on a small scale often limiting the identification of interacting partners. These sub-stoichiometric/transient protein partners may be important in providing insight into the dynamics of protein interactions with its pathway as well as other inter-connected cellular pathways. Therefore, effective and efficient TAP strategy is crucial to obtain quality TAP-MS data. A number of reasons are responsible for this. First, as already highlight in section 1.6, many proteins are of low abundance in the cell. Therefore, interactions which are transient/sub stoichiometric are potentially not identified because of the small scale of cell culture used in genome-wide studies. Secondly, because of the length of time required to conduct TAP purification, many of the protein components are degraded during purification.

This study is therefore aimed at optimising and utilising large scale TAP-MS approaches using a number of proteins targeting different stages of assembly in the spliceosome pathway Moreover, it is proposed to focus and number of important proteins where TAP MS was not successful in small scale genome wide studies. It is proposed to study complexes associated with U6/U4.U5 by utilizing Lsm proteins as baits in order to provide insight into the assembly of the spliceosome B-complex. In addition, analysis of the U2 snRNP will be examined by TAP-MS of Hsh155-TAP and Prp5-TAP. Emerging evidence show that splicing is linked to other steps in genetic expression, although the dynamics of these connections is still unclear. The development of an effective TAP-MS approach that will identify sub-stoichiometric

interactions will offer further mechanistic insight into the assembly of the spliceosome and potentially improve our understanding of the interconnectivity of the steps in gene expression. Furthermore, it is proposed to examine the RNA components associated with these complexes using molecular and analytical tools.

The CRISPR/Cas system is a recently discovered prokaryotic anti-viral defense complex. The silencing of invading nucleic acids is performed by ribonucleoprotein complexes that utilise crRNAs as guides for targeting and degradation of foreign nucleic acids. This study aims at analysing the processing of crRNA and the role of the CRISPR associating proteins in a novel Type III CRISPR/Cas system in *Streptococcus thermophilus.*

Chapter II

Materials and Methods

2.1 Materials

2.1.1 Equipments and Columns

Equipment and Columns	Make
Automatic Pippettes (P10, P20, P200, P1000)	Gilson, Eppendorf
Centrifuge	Beckman
Heating Block	Techne
Water Bath	Fisherbrand, Grant, FALC
Horizontal slab gel	Bio-Rad
Vertical gel apparatus	Bio-RAD
Horizontal slab gel apparatus	Thistle Scientific
Power supplies	Bio-Rad
Spectrophotometer	Ultraspec 1100 Pro, Amersham Bioscience
Vibrator	Fisher Scientific
Microfuge	Eppendorf
pH meter	Mettler Toledo
Concentrator	Eppendorf
Real time PCR	Applied Biosciences
Orbital incubator	Bioline
Fermentor/Bioreactor BioFlo 4500®	New Brunswick Scientific
Sonicator	Soniprep, Fisherbrand
Laminar flow	Telstar
Constant Cell disruption system	Constant Systems Ltd
UV transilluminator	Syngene
Weighing balance	Kern
Autoclave	Priorclave

HPLC Agilent 1100	Agilent
HPLC Ultimate 3000	Dionex
amaZon ETD	Bruker Daltonics
HCT Ultra PTM Discovery system	Bruker Daltonics
maXis UHR ToF MS	Bruker Daltonics
Column oven HPLC	HITACHI L-7400
DNAsep column (4.6 mm i.d. x 50 mm)	Transgenomic
Monolith PS-DVB (200 µm i.d. x 50 mm)	Dionex
Proswift 4H (1 mm I.d x 250 mm)	Dionex
Proswift RP1S (4.6 mm i.d. x 50 mm)	Dionex
Proswift RP 10R (1 mm i.d. x 50 mm)	
PepMap C-18 RP capillary column (300 µm i.d. x 150 mm)	Dionex
Mini-BeadBeater	Biospec products

2.1.2 Chemicals and Reagents

Chemicals, Reagents, Kits and Samples	Make
Agarose	Bioline
Ammonium bicarbonate	Fluka/ Sigma
Ethylenediaminetetraacetic acid (EDTA)	Sigma
Ethylene glycol tetraacetic acid (EGTA)	Sigma
Ammonium acetate	Fluka
Potassium hydroxide	Fisher
Poly(ethylene glycol) analytical standard, for GPC, 200	Sigma
Sodium phosphate dibasic anhydride	Fisher
Sodium dodecyl sulphate (SDS)	Fisher/VWS
Acid-phenol Chloroform	Ambion
1,1,1,3,3,3,-Hexafluoro-2-propanol (HFIP)	Aldrich

Triethylammonium acetate (TEAA)	Fluka
Triethylamine (TEA)	Fisher
Acetonitrile	Fluka
Methanol	Fluka
Potassium phosphate dibasic	Acros/sigma
Potassium Chloride	Fisher
Tris(hydroxymethyl) aminomethane	AnalaR Normapur
Ethidium bromide	Sigma
Protein Markers	New England Biolabs
DNA ladders	New England Biolabs
ssRNA ladders	New England Biolabs
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Bioline
Oligonucleotides	MWG Biotech
CRISPR DNA	Eurofins MWG Operon
Taq polymerase	New England biolabs
RNAse Zap (wipes)	Ambion
Magnesium Chloride	Sigma
Magnesium Acetate	Sigma
Glycine	AnalaR Normapur
Boric acid	Sigma
Tween-20	Sigma
IgG SepharoseTM 6 Fast flow	GE Healthcare
Calmodulin sepharose 4B	GE Healthcare
Polyclonal Goat Anti-Rabbit Immuglobulins HRP	GE Healthcare
Peroxidase anti-peroxidase	Sigma-Aldrich
ECL kit reagent	GE Healthcare
Nitrocellulose membrane	Sigma-Aldrich
Milk blotto	Marvel
Isopropanol	Fluka
ECL™ Anti-rabbit IgG, HRP linked whole antibody (from donkey)	GE Healthcare
Peroxidase anti-peroxidase soluble complex	Sigma

antibody produced in rabbit (PAP)	
Plasmid miniprep kit	Sigma
Complete supplement (CSM) Dropout: -TRP	Formedium
Yeast Nitrogen Base w/o Amino Acids	Difco
Amersham Hyperfilm ECL	GE Healthcare
Trypsin, from porcine pancreas	Sigma
AcTEV™ protease	Invitrogen
Bacto® Peptone	SLS
Bacto® Yeast extract	SLS
D-glucose anhydrous	Fisher
HEPES potassium salt	Sigma
Benzamidine Hydrochloride	Sigma Aldrich
Phenylmethylsulfonyl fluoride (PMSF)	Sigma
Calcium Chloride	Analar Normapur
Brilliant Blue-R	Sigma-Aldrich
SYBR® Gold nucleic acid gel stain	Invitrogen
Dried Skim milk (Milk Blotto)	Marvel
Complete EDTA-free protease inhibitor tablet	Roche

2.1.3 List of yeast TAP tag strains and constructs used in this study

Accession No	Parent Strain	Strain genotype	TAP tag	Description	Source
Scoooo	*S. cerevisiae* Scoooo^{-}	*Mata ade2 arg4 leu2-3,112 trp1-289 ura3-52*		Wild type	EUROSCARF/Cellzome
SC1377	*S. cerevisiae*-Scoooo	*Mata ade2 arg4 leu2-3,112 trp1-289 ura3-52 hsh155-TAP::TRP1^{KI}.*	CBP-Protein A	Hsh155-TAP tagged	EUROSCARF/Cellzome
SC1170	*S. cerevisiae*-Scoooo	MATa ade2 arg4 leu2-3,112 trp1-289 ura3-52 Lsm1-TAP::*TRP1^{KI}.*	CBP-Protein A	Lsm1—TAP tagged	EUROSCARF/Cellzome

SWC0931	*S. cerevisiae*-Scoooo	*Mata ade2 arg4 leu2-3,112 trp1-289 ura3-52 Lsm4-TAP::TRP1^{KI}.*	CBP-Protein A	Lsm4-TAP tagged	EUROSCARF/Cellzome
Sc3758	*S. cerevisiae*-Scoooo	*Mata ade2 arg4 leu2-3,112 trp1-289 ura3-52 Lsm8-TAP::TRP1^{KI}.*	CBP-Protein A	Lsm8-TAP tagged	EUROSCARF/Cellzome
	S. cerevisiae Scoooo	*Mata ade2 arg4 leu2-3,112 trp1-289 ura3-52 Prp5-TAP::TRP1^{KI}.*	P54-Protein A	Prp5-TAP tagged	This study

Description of p54 tag

The p54 Tag is a 7.12 kDa peptide derived from the mammalian proteinp54nrb. It has a polyhistidine sequences and proline-rich region dispersed in the Nickel-binding domain. The p54-TeV-Prot.A tag was developed in our lab (Prof. D. P. Hornby's Lab.) by replacing the Calmodulin binding moiety (CBP) nucleotide sequence of the standard CBP-Tev-Prot.A tag in the pBS1479 plasmid (obtained from Euroscarf) with the p54nrb N terminal sequence containing polyhistidine and polyproline motifs (see Figure 2.2). This plasmid was identified as PAB101 and contains a tryptophan nutritional complement gene. The cassette containing p54-TEV-Prot.A and TRP nutritional complement gene was amplified and used as the integration cassette.

2.1.4 Details of the primers used to insert TAP tags into Prp5 gene

After designing primers, they were ordered from Eurofins MWG, HPLC purification grade and 0.01 µmol.

Fwd primer: GCT GGA GCA AAA AGT TAA AGA GGG GGT CGT AAA GGC TGC AAG CTT GTC TTT GAA GAG TAC TAA ATA CGG AGG ATC CAT GGA ACA GAG (87 bp)

Rev primer: GCT TTA AAT TAC CAC CCT AAA AAT TTC AGA ACT AAC TAC GAA AGT ATA TAG CAC CAC GAG TAA TAC GAC TCA CTA TAG GG (80 bp)

Rev complement of Rev. primer (80bp)
CCCTATAGTGAGTCGTATTACTCGTGGTGCTATATACTTTCGTAGTTAGTTCTGAAATTTTTAGGGTGGTAATTT AAAGC

Figure 2.1 Schematic overview of the Prp5 gene targeted for TAP tagging in this study. Cartoon illustrating the site of Prp5 gene targeted for TAP tagging using PCR-mediated homologous recombination. The site of recombination delineated by the forward and reverse primer construct is highlighted in bold red and light blue, respectively. The continuous diamonds represent intervening sequences. The continuous dash represents downstream sequences. Start and Stop codon are highlighted in light brown. The region highlighted in green is replaced by p54-Protein A TAP tag and –TRP complement gene during homologous recombination.

5’----------

GGAGGATCCATGGAACAGAGTAATAAAACTTTTAACTTGGAGAAGCAAAACCATACTCCAAGAAAGCATCATCAACATCACCACCAGCAGCAGCACCACCAGCAG
CAACAGCAGCAGCCGCCACCACCGCCAATACCTGCAAATGGGCAACAGGCCAGCAGCCAAAATGAAGGCTTGACTATTGACCTGAAGGAGAATTTGTATTTTCAG
GGTGAGCTCAAAACCGCGGCTCTTGCGCAACACGATGAAGCCGTGGACAACAAATTCAACAAAGAACAACAAAACGCGTTCTATGAGATCTTACATTTACCTAAC
TTAAACGAAGAACAACGAAACGCCTTCATCCAAAGTTTAAAAGATGACCCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCTAAAAAGCTAAATGATGCTCAGGCG
CCGAAAGTAGACAACAAATTCAACAAAGAACAACAAAACGCGTTCTATGAGATCTTACATTTACCTAACTTAAACGAAGAACAACGAAACGCCTTCATCCAAAGTT
TAAAAGATGACCCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCTAAAAAGCTAAATGGTGCTCAGGCGCCGAAAGTAGACGCGAATTCCGCGGGGAAGTCAACC
TGAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCGGAACGATCATTCACTATATATATATCAATTTATATATACGTATGTGT
AATTGAAGAAAGATACGTTTTTTCCTCTATTGAGAGGCCTGCTGGATGAATAGCTTTACCTTTTCTAAATCCTTGATACCATCAGTCTCTACTCCTCCACTTACATCG
ACACCGATCGCATTTGGTAACATATTAATGGCAACAGAAACGTTATCAGGATTCAATCCACCAGCGATAATGAATTTTATCTCGGGATGACTTGCAGACCAACTGG
AAATTGCACTCCAATTCAATTTCTCACCAGTGCCACCTTCACCAGAATCGAACAACGTCAGCACATTGTCTACGTGTTCATACAGGTCCAGTAGTAATTCACAATCCT
GTGGGAACTGGAACCTCTTAATGATTGGAATTGAAGATGGGATCAAAGATCTGTATTCTTTAATATCTTCATCTCCATGTAATTGTATCACATCTAGATTATATTCG
TGGTACAGTTGAAGGACATCATCAACGGACTGATTTCTAAACACCCCGACTAGTTTAGTACCTTTCACGTTCTCTTGTTGGTGAACTGCAGTTGAAATACCTTTCGC
AACAGATGAGTCAATGGTTCTTTTCCTACCGGGAACACAAATGATACCTAAGTAATCAGCACCATCATCCACAGCAGTCTTTGCAGCTTCAACGGTTTGCAAACCA
CACACTTTAACGAGCATCACAGTAGATTTCGCAAGAGGGTAGCTCGCTCAGAGTACCCAAGTAAATGATTAGTAAACTGATGTTTGATAGTTCAATTTTTCAATGA
AATAACCTTATATTAAAATTGATATTACTATTATACAAAAATAAAGAATAAAGGATTTGAGTTTATACATAAAATACCATTATTATTTGTTCAGTGAGAGATACCGG
GGTATATGGGATGTGTGTAGTGATACCATGCAATCATGTATCAAACATGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTA -3’----------
----

Figure 2.2 Schematic overview of the p54-Protein A TAP tag. The p54-Protein A sequence was cloned in a vector with both the TAP tag and nutritional complement as marker. A conjugate primer with plasmid binding and target gene binding sequence was designed. The green-highlighted sequence is the CBP moiety and the Protein A moiety is highlighted in red. The sequence highlighted in yellow is the minus Tryptophan (-TRP) nutritional complement CDS which is used as marked during recombinant selection. The sequences highlighted in blue represent the 5’ and 3’ plasmid binding sequence of the conjugate primer. The dash represents the target gene binding sequence of the conjugate primer.

ATGGAAAAGAGAAGATGGAAAAAGAATTTCATAGCCGTCTCAGCAGCCAACCGCTTTAAGAAAATC
TCATCCTCCGGGGCACTTGATTATGATATTCCAACTACTGCTAGCGAGAATTTGTATTTTCAGGGTGA
GCTCAAAACCGCGGCTCTTGCGCAACACGATGAAGCCGTGGACAACAAATTCAACAAAGAACAACA
AAACGCGTTCTATGAGATCTTACATTTACCTAACTTAAACGAAGAACAACGAAACGCCTTCATCCAAA
GTTTAAAAGATGACCCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCTAAAAAGCTAAATGATGCTCA
GGCGCCGAAAGTAGACAACAAATTCAACAAAGAACAACAAAACGCGTTCTATGAGATCTTACATTTA
CCTAACTTAAACGAAGAACAACGAAACGCCTTCATCCAAAGTTTAAAAGATGACCCAAGCCAAAGCG
CTAACCTTTTAGCAGAAGCTAAAAAGCTAAATGGTGCTCAGGCGCCGAAAGTAGACGCGAATTCCG
CGGGGAAGTCAACCTGA

Figure 2.3 Schematic overview of the CBP-Protein A TAP tag. The CBP-Protein A sequence is represented in colours. The red coloured sequence is the CBP moiety and the Protein A moiety is highlighted in blue.

2.1.5 Primers for Prp5 TAP tagging

Oligonucleotide	Sequence
Forward primer for PCR-mediated gene targeting	GCT GGA GCA AAA AGT TAA AGA GGG GGT CGT AAA GGC TGC AAG CTT GTC TTT GAA GAG TAC TAA ATA CGG AGG ATC CAT GGA ACA GAG (87 bp)
Forward primer for PCR-mediated gene targeting	5' GCT TTA AAT TAC CAC CCT AAA AAT TTC AGA ACT AAC TAC GAA AGT ATA TAG CAC CAC GAG TAA TAC GAC TCA CTA TAG GG 3'(80 bp Reverse complement 3' CCC TAT AGT GAG TCG TATTACTCGT GGTGCTAT ATACTTTCGTAGTTAGTTCTGAAATTTTTAGGGTGGTAA TTTAAAGC 5'
CBP-Forward primer (Internal primer)	TAG CCG TCT CAG CAG CCA AC (20 bp)
ProtA Rev-primer (Internal primer)	TCT ACT TTC GGC GCC TGA GC (20 bp)
Prp5 Forw-Primer (External primer)	CGA AGG CCA AGA TGA GAA AG (20 bp)
Prp5 Rev-Primer (External primer)	GAC CGC TGT CAA ATA TTC CG (20 bp)

2.1.6 Oligonucleotide sequences used in crRNA Analysis

Oligonucleotides	
BC6	AGCTA
BC7	AGCTAGC
BC8	AGCTAGCTA
E. coli CRISPR DNA	ATAAACCGACGGTATTGTTCAGATCCTGGCTTGCCAACAGGAGTTCCCCG CGCCAGCGGGG
E. coli crRNA	AUAAACCGACGGTAUUGUUCAGAUCCUGGCUUGCCAACAGGAGUUCC CCGCGCCAGCGGGG

2.1.7 Oligonucleotides sequences used in Northern Blot Analysis

Oligonucleotides	
U4 oligo	5'-AGACTATGTAGGGAATTTTTGGAATACC-3'
U5 large oligo	5'-GCCCTTTTTCTCAATGAGTAAGGAGGGCG-3'
U5 small oligo	5'-CCAGAACCATCCGGGTGTTGTCTCCATAG-3'
U6 Oligo	5'-GATCAGCAGTTCCCCTGCATAAGGATG-3'

2.2 Buffers and Gel matrices Composition

2.2.1 Tandem affinity purification Buffers

Lysis buffer – Buffer A	
K-Hepes (pH 7.9)	10 mM
KCl	10 mM
$MgCl_2$	1.5 mM
Benzamidine	2 mM
PMSF in DMSO	1 mM
EDTA-free Protease inhibitor cocktail tablet	Appropriate amount

IPP150	
Tris-base (pH 8.0)	10 mM
NaCl	150 mM
Tween-20	0.05%

TEV Cleavage Buffer (pH 8.0)	
IPP150	Required volume
EDTA	0.5 mM
DTT	1 mM

Calmodulin Binding Buffer (CBB) (pH 8.0)

Tris-base	10 mM
NaCl	150 mM
Magnesium Acetate	1 mM
Imidazole	1 mM
$CaCl_2$	2 mM
Tween-20	0.05%
DTT	10 mM

Calmodulin Elution Buffer (CEB) (pH 8.0)	
Tris-base	10 mM
NaCl	150 mM
Magnesium Acetate	1 mM
Imidazole	I mM
EGTA	2 mM
Tween-20	0.05%
DTT	10 mM

Denaturing Elution buffer (pH 8.0)	
Sodium dodecyl sulphate (SDS)	2%
Tris-base (pH 8.0)	125 mM
Dithiotheitol (DTT)	100 mM

2.2.2 SDS Polyacrylamide gel electrophoresis (PAGE)

Stock solutions were first prepared and working solution made by appropriate dilution of the stock solutions in water.

4% Stacking gel 6mls	
1.5 M Tris-base, 0.4% SDS (pH 6.8)	1.5 ml

ddH_2O	3.4 mls
30% Acrylamide	1 mls
10% Ammonium persulphate	100 µl
N,N,N',N'-tetramethylethene-1,2-diamine (TEMED)	20 µl

10% Resolving gel (10 ml)	
1.5 M Tris-base, 0.4% SDS (pH 8.8)	2.5 ml
ddH_2O	3.96 mls
30% Acrylamide	3.3 mls
10% Ammonium persulphate	200 µl
N,N,N',N'-tetramethylethene-1,2-diamine (TEMED)	40 µl

SDS gel loading buffer	
Tris-base (pH 6.8)	50 mM
DTT	100 mM
Bromophenol blue	0.10%
SDS	2%
Glycerol	10%

Tris-glycine-SDS (TGS) Running buffer	
Tris-base (pH 8.8)	25 mM
SDS	0.1 %
Glycine	250 mM
Phosphate Buffer Saline (PBS)	
NaCl	150 mM
NaH_2PO_4	50 mM

Ni-NTA Elution buffer

This elution is PBS containing 250mM Imidazole.

2.2.3 Denaturing Polyacrylamide gel electrophoresis

Stock solutions were first prepared and working solution made by appropriate dilution of the stock solutions in water.

TBE PAGE Running Buffer 10x(500 ml) (pH 8.0)	
Tris-base	54 g
Boric Acid	27.5 g
0.5 M EDTA	20 ml
ddH_2O	Make up to 500 ml

Denaturing PAGE GEL (100 ml)	
Urea	48g
10x TBE	10 ml
40% Acrylamide	15 ml
ddH_2O	Make up to 100 ml

2.2.4 Agarose gel electrophoresis

TAE gel Running Buffer 50x(500 ml)	
Tris-base	121 g
Acetic Acid (glacial)	28.55 ml (17.4N)
EDTA (pH 8.0) (0.5M)	50 ml
ddH_2O	Make up to 500 ml
1% Agarose gel (100 ml)	
Agarose	1g
1x TAE	Make up to 100 ml

2.2.5 LC ESI MS Buffers

LC ESI MS Solvent A for Protein Analyses	
Formic acid	0.1%
Solvent B	
Formic acid	0.1%
CAN	80%

LC ESI MS Buffer A For nucleic acid analyses	
1,1,1,3,3,3,-Hexafluoro-2-propanol	0.4 M
Triethylamine (TEA)	To pH 7.0
Triethylammonium acetate (TEAA)	0.1 mM

LC ESI MS Buffer B For nucleic acid analyses	
1,1,1,3,3,3,-Hexafluoro-2-propanol	0.4 M
Triethylamine (TEA)	To pH 7.0
Triethylammonium acetate (TEAA)	0.1 mM
Methanol	50%

2.2.6 Growth Media for Yeast growth

To select for recombinant TAP tagged yeast strains, -TRP media was prepared. To grow Yeast strains in enriched media, Yeast extract peptone dextrose (YPD) media was made. Both media were sterilized by autoclaving at 121°C for 20 minutes. For Large culture, such as 15L culture, YPD media was autoclaved in the Fermentor/Bioreactor at 121°C for 20minutes. In the preparation of SD-TRP media, it

was common practice to filter sterilize the Yeast Nitrogen Base w/o amino acid supplement and autoclave the other nutritional constituents; although the SD-TRP media are sometimes autoclaved with all the constituents included.

Selective minimal media lacking Tryptophan(SD-TRP) (1L)	
Complete supplement (CSM) Dropout: -TRP (Formedium)*	0.74 g
Yeast Nitrogen Base w/o amino acid**	6.7 g
Glucose	20g
dH20	Up to 1000

* Complete supplement (CSM) Dropout: -TRP is the trademark name and is supplied by Formedium. The formulation contains all essential amino acids except for Tryptophan. **The Yeast Nitrogen base w/o amino acid (supplier, Difco) formulation contains ammonium sulphate without amino acids.

YPD or YPAD media	
Bacto Yeast Extract	1% (W)
Bacto Peptone	2 % (W)
Glucose	2 %

2.3 Transformation by LiAC/ssDNA/PEG Method

A single colony of the yeast strain was inoculated with a sterile inoculation loop from a fresh YPAD plate into 10 ml of liquid medium (YPAD) and incubated overnight on a rotary shaker at 220 r.p.m. and 30 ˚C. After 16 h of growth, the yeast culture titre was determined by using a spectrophotometer. (10 µl of cells was pipetted into 1.0 ml of water in a cuvette, mixed thoroughly by inversion and the OD was measured at 600 nm.) The cell culture were added to pre-warmed (30 ˚C) 50 ml of YPAD in a pre-warmed (30 ˚C) culture flask to reach an OD of 0.6. The flask was placed in the shaking incubator at 30 ˚C and 200 rpm until the cell OD_{600} was at least, 1.0. A 1.0 ml

sample of carrier DNA was denatured in a boiling water bath for 5 min and chilled immediately in an ice box. The cells were then harvested by centrifugation at 3,000*g* for 5 min and the pellet was resuspended in 25 ml of sterile water and centrifuged at 3,000*g* for 5 min to pellet the cells. This wash was repeated with another 25 ml of sterile water by resuspending the cells and pelleting them again by centrifugation. The cells were resuspended in 1.0 ml of sterile water. The cell suspension was transferred to a 1.5 ml microcentrifuge tube, centrifuged for 30 sec. at 13,000*g* and the supernatant was discarded. The cells were resuspended in 1.0 ml of sterile water and 100 µl sample containing the cell pellet were pipetted into 1.5 ml microcentrifuge tubes, one for each transformation, centrifuged in a microcentrifuge at 13,000*g* for 30 sec. and the supernatant was removed. For a single plasmid in each transformation reaction, sufficient transformation mix was made up for the planned number of transformations, plus one extra using the recipe below and was mixed thoroughly on a vortex mixer.

Transformation Mix components	Volume (µl)
PEG 3350 (50% (w/v))	240
LiAc 1.0 M	36
Single-stranded carrier DNA (2.0 mg/ml)	50
Plasmid DNA plus sterile water	34
Total	360

360 µl of transformation mix was added to each transformation tube and the cells were resuspended by vortexing vigorously. The tubes were placed in a water bath at 42 ˚C and incubated for 40 min. The tubes were centrifuged at 13,000*g* for 30 sec. in a microcentrifuge and the supernatant was removed with a micropipettor. 1.0 ml of sterile water was pipetted into the transformation tube. The pellet was stirred with a sterile micropipette tip to uniformly resuspend the cell pellet. 200 µl of the cell suspension was plated onto the appropriate SC selection medium; the inoculum was spread with a glass rod, sterilised by an ethanol soak and passed through the flame of a Bunsen burner. The liquid was allowed to be absorbed into the medium by incubation at room temperature. The plates were incubated at 30 ˚C for 4 days.

2.4 Recombinant *Saccaromyces cerevisiae* growth and tandem affinity purification (TAP)

In this study, five different recombinant strains of *S. cerevisiae* in which each bears a TAP tag fused to the C-terminal of target proteins was constructed. These proteins include Hsh155-TAP, Prp5-TAP, Lsm1-TAP, Lsm4-TAP and Lsm8-TAP strains. To characterise various target ribonucleoprotein complexes (mRNA processing complexes), *S. cerevisiae* strains were grown at optimum conditions using yeast extract dextrose peptone media (YPD) and then purified using novel modified tandem affinity purification protocols adapted from the protocol developed by (Puig et al., 2001).

2.4.1 Growth of Lsm-TAP, Hsh155-TAP and Prp5-TAP yeast strains

A single colony of the recombinant *S. cerevisiae* strains were inoculated into 5 mls of SD-TRP media and incubated overnight (14 hours) at 30°C with continuous shaking at 200 rpm. The overnight culture was inoculated into a 150 ml Tryptophan-deficient minimal media, further incubated for 18 hours (harvested at exponential growth phase) using the same conditions. The 150 ml culture was then used as a seed culture in a 15 L YPD media culture grown in a fermentor at 30°C with continuous mixing at 200 rpm and with adequate air supply for a period ranging from 14 to 18 hours (but essentially at the exponential phase) depending on the experiment.

2.4.2 TAP procedure A: Purification of Lsm-TAP and Hsh155-TAP using less stringent TAP

1st purification module:

15 L yeast cell culture was grown on YPD medium to exponential growth phase and cells harvested and washed with distilled water. The pellet was placed on ice and resuspended in 1:2 (g/V) volumes of ice cold lysis buffer, buffer A (10 mM HEPES – potassium salt, pH = 7.9; 10 mM KCl; 1.5 mM $MgCl_2$). 1mM PMSF dissolved in DMSO

was added and the cell slurry stirred for a minute. Then 2 mM Benzamidine and EDTA-free protease inhibitor tablets were added (1 tablets: 100 ml buffer A) and cell slurry stirred for a minute. The cells were passed through cell disruption machine (Constant systems Ltd) at 4°C and lysate collected. The KCl concentration of the cell lysate was adjusted to 200 mM. The lysate was centrifuged at 9500 rpm and the supernatant collected. The supernatant was further subjected to centrifugation at 13500 rpm and supernatant collected. One volume of 3xIPP150 buffer (450 mM NaCl, 30 mM Tris-HCl pH 8.0, 0.15 % Tween-20 was added to two volumes of cell lysate, and the mixture was incubated with 1 ml pre-washed IgG-sepharose (GE Healthcare Life Sciences) beads at 4°C for 2 hours. The IgG beads were collected with the aid of a Vacuum suction filter device (Buchner filtering apparatus, with a filter) that was connected to a suction pressure (at a flow rate of approximately 10 mL per minute) as the supernatant flowed through. As the supernatant above the filter gets down to 5 mL the filtering apparatus was released from suction and allowed to flow under gravity. To wash the beads, 20 ml buffer A was immediately added to the bead collected on the filter as soon as the final remnant of lysate supernatant flowed through. It was again allowed to flow through until 5 mL of buffer A-bead slurry was left. The bead slurry was transferred to a Bio-rad mini-protein column; 10 ml IPP150 solution was added and allowed to flow through under gravity. 1 ml of TEV cleavage buffer (10 mM Tris-HCl, pH 8.0; 150 mM NaCl; 0.05% Tween-20; 0.5 mM EDTA; I mM DTT) was added and allowed to flow through. The IgG beads were incubated in 1.5ml of TEV cleavage buffer containing 100 units of TEV protease (Invitrogen) at room temperature for 1 hour. The IgG beads were removed from the suspension and the filtrate (IgG eluate) collected in 1.5 eppendorf tubes. About 300ul was reserved for analysis and the rest used for the second module of TAP.

2^{nd} purification module

In the second TAP module, calmodulin beads (GE Healthcare) were washed with Calmodulin binding buffer (CBB) (10 mM Tris-HCl, 150 mM NaCl, 2 mM imidazole, 2 mM $CaCl_2$, 2 mM Magnesium acetate 0.05 mM Tween-20 and 10 mM DTT). In order to ensure mild purification condition 7 ml of Calmodulin binding buffer (CBB) that does not contain DTT was added to the IgG eluate and then applied to the calmodulin beads column. The beads were subsequently washed with 1ml of CBB

without DTT and eluted with denaturing elution buffer (2% SDS, 100 mM DTT and 125 mM Tris-HCl) or native elution buffer (10 mM Tris-HCl, 150 mM NaCl, 2 mM imidazole, 2 mM EGTA, 2 mM Magnesium acetate 0.05 mM Tween-20 and 10 mM DTT).

2.4.3 TAP procedure B: Purification of Lsm-TAP and Hsh155-TAP using more stringent TAP

1st purification module

In this procedure, 15 L cells were grown exponential growth phase, harvested and cell pellets lysed and adjusted for KCl concentration using the same conditions as in TAP procedure A (see section 2.4.2).The cells were then first centrifuged at 13500 rpm for 10 minutes at 4°C and supernatant collected. The supernatant was further centrifuged at 24000 rpm for 20 minutes at 4°C. The supernatant was clarified by passing through sepharose 4-Cl (Pharmacia) and then incubated with IgG beads for 2 hours. IgG beads were extensively but mildly washed consecutively with 50 mls of buffer A, 50 mls of IPP150, and 20 mls of TEV cleavage buffer. The IgG beads were incubated in 2 mls of TEV cleavage buffer + 100 units of TEV protease for 1 hour at room temperature. The IgG eluates were collected and 300 µl reserved for analysis.

2nd purification module

7 mls of Calmodulin binding buffer (containing 10 mM DTT) was added to remaining IgG eluate (about 1.6 mls) and then applied to a column containing calmodulin beads pre-washed with CBB (containing 10 mM DTT). After allowing the mobile phase to flow through under gravity, the beads were washed with 20 mls of washing buffer (CBB + 10 mM DTT) and then eluted with either denaturing elution buffer or native elution buffer (see section 2.3.1 for buffer compositions).

2.4.4 Tandem affinity purification of Prp5-TAP

1st purification module

The Prp5-TAP yeast strain has a novel TAP tag containing Protein A and p54 (a novel affinity peptide that bind Ni-NTA) and the first purification module of TAP used here is the same as TAP procedure B (section 2.4.3). Hence, 300 ml IgG eluate was reserved for analysis and the rest used for the second purification module.

2nd purification module

In the second purification module, Ni-NTA resin was washed with 30 ml distilled water and 30 ml PBS (50 mM NaH_2PO_4 and 300 mM NaCl). The IgG eluate sample (1.5 ml) was loaded on to the column two times to ensure strong binding, followed by 50 ml wash with PBS. 5 fractions were eluted with the same buffer containing 250 mM imidazole.

2.5 Analysis of ribonucleoprotein complexes

2.5.1 Fractionation of the protein-RNA components.

300 µl of Phenol-Chloroform (ambion) was added to 300 µl purified sample (if eluted with native elution buffer (see section 2.2.1) add 80 µl 10% SDS) and Vortexed for about 30 seconds. It was heated for 5 minutes at 75 °C and vortexed for 1 minute (not necessary when using denaturing buffer). 300 µl of acid phenol-chloroform was added, vortexed for another minute and left to stand for 5 minutes at room temperature. It centrifuged at 13000 rpm for minutes and the aqueous phase (which contains RNAs) transferred to another 1.5 mL Eppendorf tube (about 200-250 µl was usually recovered) without disturbing the phenol/chloroform phase. It was vortexed for 30 seconds and 0.75 ml of ethanol added. This was mixed and centrifuged for 25 minutes at room temperature. The supernatant was discarded and pellet (RNAs) air-dried and resuspended with desired volume of HPLC grade water.

For the protein extraction, 0.9 mL of acetone was added to the phenol phase (about 300 – 350 µl), vortexed for 30 seconds and centrifuged for 25 minutes. Supernatant was discarded and pellet (proteins) air-dried and resuspended with distilled water.

EDTA-wash-Phenol-chloroform extraction of crRNA

To remove potential metal ions, appropriate volume of 0.5M EDTA was added to the Csm CRISPR complex sample. 40 µl of 10% SDS stock solution (4% in final volume) was added and the mixture vortexed for about 30 seconds. It was made up to 100 µl with HPLC grade water. It was heated for 5 minutes at 75°C and vortexed for 1 minute. 200 µl of HPLC grade water was added followed by 300 µl of acid phenol-chloroform. It was then vortex for 1 minute and left to stand for 10 minutes at room temperature. It was centrifuge for 5 minutes at 14,800 rpm (16,162 g). The aqueous phase was transferred to an eppendorf new tube, with care not to disturb the bottom phenol/chloroform phase. 1.2ml isopropanol or 900 µl of ethanol was added to the aqueous phase and vortexed for 30 seconds. It was spin centrifuged for 15-30 minutes at 14,800 rpm (16,162 g) at room temperature and the supernatant removed. The crRNA was re-suspended in RNAse free water.

2.5.2 Protein SDS-PAGE

The gel apparatus (Bio-Rad mini protean II system) was used for SDS PAGE analysis. 10% resolving gel (section 2.2.2) was introduced in between glass gel casting slides (70 x 100 x 1 mm) leaving approximately 0.7 cm gap between the top of the plate and the top of the gel. The resolving gel was overlaid with 100% isopropanol and allowed to polymerise for 15 – 20 minutes. Following polymerisation, isopropanol was removed and 4% stacking gel (section 2.2.2) was overlaid, a 1mm size comb inserted and allowed to polymerise for 15 – 20 minutes. Before use, the comb was removed and the gel cassette was assembled into a 1D gel apparatus filled with 1X Tris-glycine-SDS (TGS) buffer. Protein samples were prepared in appropriate volumes of gel loading buffer (section 2.2.2) and denatured at 95°C for 4 min. Gels were first run at 80 V for 10 minutes followed by 100 V until electrophoretic front

(dye front) had eluted to the bottom of the gel. Protein gels were stained in Coomassie blue stain (10% acetic acid, 0.45% Coomassie Brilliant Blue R-250, 50% methanol) and destained with destaining buffer (25% methanol, 5% acetic acid).

2.5.3 Western blot Analysis

5 ml of subclones cultures was transferred in Eppendorf tubes. The cells were spun down and supernatant discarded. 200 µl of glass beads were added followed by 200 µl of PBS buffer (150 mM NaCl, 50 mM Na_2P04) + Complete EDTA-free protease inhibitor tablet (Roche). Cells were lysed by vortexing in Bead beater and supernatant removed. 15µl SDS-PAGE loading buffer 2X was added to 15 µl of lysate (supernatant). The tubes were vortexed for 30 secs., boiled for 10 min., spun and 20 µl loaded on a mini 10% SDS-PAGE gel. Migration was done for 1 hour at 120 V. The gel was the transferred on a nitrocellulose membrane, blocked in Blotto/Tween-20 for 2 hours at 4 ˚C and washed 3 x 5 min. with PBS/Tween-20. 20 ml Blotto/Tween-20 + 10 µl of Peroxidase anti-peroxidase soluble complex antibody produced in rabbit (PAP) (GE Healthcare) was added. The membrane was incubated at room temperature for 1 h and rinsed 2x for 10 min. with PBS/Tween-20. Detection was done by chemiluminescence: ECL kit reagent (GE Healthcare) was applied on the membrane according to manufacturer's instruction, covered with a clean, cling film and signal detected using Chemiluminescence imaging system fitted with charge coupled device (CCD) camera (Biospectrum® Multispectral Imaging System) or X-ray film.

2.5.4 RNA denaturing polyacrylamide gel electrophoresis

Using a vertical gel apparatus was used for RNA denaturing PAGE. 10% 8M urea resolving gel (see section 2.2.2) was introduced in between glass gel casting slides (70 x 100 x 1 mm) and a 1mm size comb was inserted. The gel was allowed to polymerise for 15 minutes. Before use, the comb was removed and the gel cassette

assembled into a 1D gel apparatus that was filled with tank buffer (1X TBE). The RNA samples were prepared in RNA loading buffer NEB.

2.5.5 Northern blot analysis

Deoxyoligonucleotides of about 27 to 29 nt corresponding to regions within U4 snRNA, U6 snRNA, U5 snRNA (Large) and U5 snRNA (Small) were designed as PCR primers as shown in the section 2.1.7 table and synthesized by Eurofins MWG Operon. Following denaturing PAGE of Lsm1-TAP and Lsm8-TAP samples, the RNA was transferred to Nylon membrane by Wet transfer in electrophoresis running buffer (TBE buffer). Radiolabelled probes were prepared from the deoxyoligos using Hi Prime Labelling Kit (Roche Applied Sciences, cat. number 1 585 584) following manufacturer's instructions. Blocking, probing and visualization were performed following manufactures instructions.

2.5.6 Native Agarose gel electrophoresis

DNAs and RNAs were analysed in 1% agarose gel prepared by mixing 1 g agarose (electrophoresis grade) in 1X TAE buffer (described in section 2.2.4. The mixture was dissolved by heating in microven oven and then allowed to cool to approximately 60°C. For DNA gels, 0.5 µg/ml ethidium bromide was added at this point but not for RNA samples. It was poured into a horizontal gel apparatus (Thistle Scientific). It was allowed to cool and polymerise. Nucleic acids sample were prepared in an appropriate volume of gel loading buffer (NEB) and loaded into gel wells. Electrophoresis was run at 100 V for 1 hour. For RNA sample, the gels were stained with 0.5 µg/ml ethidium bromide for 10 minutes.

2.6 Preparation of protein samples for LC MS/MS

2.6.1 Protein Tryptic In-gel digestion

To produce peptide mixtures of protein complex components subsequent to LC MS/MS analysis, samples were first analysed on SDS PAGE and protein bands excised for tryptic in-gel digest which is described as follows. First, protein bands were excised from the SDS PAGE gel, cut into tiny bits to facilitate migration of trypsin into the gel in a subsequent step and placed in Eppendorf tubes. This was followed by addition of 400 µl of 200 mM Ammonium bicarbonate (ABC) in 40% acetonitrile (ACN) and incubated at 37°C for 15 to 30 minutes until the gel pieces are destained. The supernatant was removed and 400 µl of 100% ACN added to dehydrate the gel pieces. The acetonitrile (ACN) was removed and the gel pieces were further dried in Eppendorf concentrator for 10 minutes. 20 ng/µl trypsin working solution was prepared by dissolving 20 µg Trypsin from bovine pancreas (Sigma) 100 ml of 1 mM HCl and diluting with 50 mM ABC in 9% ACN to a final concentration of 20 ng/ul. 30 µl of the trypsin working solution was added to the dried gel pieces and incubated at 4°C for 10 minutes. 100 µl of 50 mM ABC was added and then incubated overnight at 37°C.The supernatant was collected into an Eppendorf tube and 100 µl of 100% ACN added to the gel pieces. After 15 mins, the supernatant was separated from the dried gel pieces and collected. This liquid was dried down by Eppendorf concentrator. The dried peptides were resuspended with appropriate volume of 0.1% trifluoroacetic acid (TFA) (usually 5 -10 µl) depending on the LC injection loop and the volume to be injected.

2.6.2 Protein in-solution digest

Protein sample in solution were prepared for LC MS/MS by digesting with tryspin and is described as follows. 5 µl of 10% SDS was added to protein samples in a solution volume of about 50 µl to 100 µl (solution could be water or very volatile buffer) and heated at 95°C for 2 minutes. This was allowed to cool for about 2 minutes and 400 ng of trypsin added. Incubation was overnight at 37°C. Subsequently the samples were dried down and pellets resuspended with appropriate volume of 0.1% TFA.

2.7 Protein Mass spectrometry Analysis

Mass spectrometry analysis of proteins were performed using an ion trap mass spectrometer (HCT Ultra PTM Discovery systems, Bruker Daltonics), Ultra-high resolution time of flight (UHR-ToF) mass spectrometer (maXis, Bruker Daltonics) and AmaZon ETD (Bruker Daltonics) coupled with an online nano liquid chromatography system (Ultimate 3000 RSLC nano liquid chromatography system, Dionex). Peptide separations were performed using a 5 mm x 300 µm trapping and 75 µm x 15 cm analytical PepMap C18 reverse-phase columns. Tryptic peptide elution was performed using a 55 minute linear gradient from 94% solvent A (0.1% (v/v) formic acid) to 40% solvent B (0.1% (v/v formic acid), 80% (v/v) acetonitrile) at a flow rate of 300 nL/min.

2.7.1 Protein ESI-TRAP and ESI-TRAP-QUOD analysis

MS1 profile scans (m/z 300-1800) were acquired in standard enhanced positive mode and were followed by 4 CID fragmentation experiments in ultrascan mode (m/z 100-1800). For fragmentation, the trap was loaded to a target value of 200,000 with a maximum accumulation time of 200 ms. The precursor isolation width was set to 4.0 and singly charged precursors were excluded. Mass spectra were acquired on-line using profile MS with automatic dependent MS/MS scans. Data files were processed into MGF format in order to identify the complex proteins in samples using MASCOT Server v. 2.2.01 (Matrix Science) The Swissprot yeast database was searched using the following parameters (analysis peptide tolerance = ±1.2 Da, MS/MS tolerance = ±0.6 Da, peptide charge 2+ and 3+. Tryptic enzyme specificity with up to two missed cleavages was applied to all searches. Oxidized methionine and carbamidomethylation were used as variable modifications for the tryptic digests. A peptide ion score of ≤ 27 as a cut-off as calculated by MASCOT was also used to filter. The false discovery rates were determined using the MASCOT decoy database option and were ≤2.5. Protein identifications were based on a minimum of two unique peptides.

2.7.2 Protein ESI-UHR TOF analysis

MS/MS analysis was performed using a maXis UHR TOF mass spectrometer (Bruker Daltonics) using an automated acquisition approach. MS and MS/MS scans (m/z 50-3000) were acquired in positive ion mode. Lock mass calibration was performed using HP 1221.990364. Line spectra data was then processed into peak list by Data analysis using the following settings. The sum peak finder algorithm was used for peak detection using a signal to noise (S/N) ratio of 10, a relative to base peak intensity of 0.1% and an absolute intensity threshold of 100. Spectra were deconvoluted and the peak lists exported as MASCOT Generic Files (MGF). and searched using MASCOT 2.2 server (Matrix Science) The Swissprot database was searched using the following parameters (analysis peptide tolerance = ±0.1 Da, MS/MS tolerance = ±0.1 Da, peptide charge 2+ and 3+. Tryptic enzyme specificity with up to two missed cleavages was applied to all searches. Oxidized methionine and carbamidomethylation were used as variable modifications for the tryptic digests. A peptide ion score of ≤ 27 as a cut-off as calculated by MASCOT was also used to filter. The false discovery rates were determined using the MASCOT decoy database option and were ≤2%. Protein identifications were based on a minimum of two unique peptides

2.7.3 Protein identifications

Protein identifications were based on a minimum of two unique peptides using a peptide ion score of ≤ 27 as a cut-off as calculated by MASCOT to filter. Protein identifications and interaction networks were identified with the aid of integrated software platform for Mass spectrometry, ProHits (downloaded athttp://prohitsms.com/Prohits_download/list.php). The TAP protein complexes data were further screened for potential contaminants such as abundant proteins (Translation factors, ribosomal proteins) and proteins known to associate with unfolded protein domains (such as proteasome, heat shock proteins).

2.8 Preparing RNA for Mapping by RNAse digest-LC MS/MSn

RNA digestion

Oligoribonucleotide fragment mixtures from RNA were generated by digesting~ 200 ng RNA sample with 20U RNAse T1 and RNAse A (Ambion) in water at 37°C for 30 minutes.

2.9 RNA Analysis by Chromatography and Mass spectrometry

2.9.1 crRNA Analysis by HPLC

The chromatographic analysis was performed using the following buffer conditions: A) 0.1 M triethylammonium acetate (TEAA) (pH 7.0) (Fluka); B) buffer A with 25% LC MS grade acetonitrile (v/v) (Fisher); C) buffer A with 25% EDTA. For this analysis, crRNA was obtained by phenol-chloroform extraction with an EDTA-wash step (see 2.5.1). Alternatively we obtained crRNA by injecting purified intact Csm-strep tagged complex at 75 °C starting with an isocratic EDTA wash at 100% buffer C for 6 minutes followed by a linear gradient starting with 5 minute run at a 15% buffer B and extending to 60% B in 12.5 min, followed by a linear extension to 100% B over 2 min at a flow rate of 1.0 ml/min. Analysis of the 3' terminus was performed by incubating the HPLC-purified or crRNA in a final concentration of 0.1 M HCl at 4 °C for 1 hour. The samples were concentrated to 10-20 µl on a vacuum concentrator (Eppendorf) prior to ESI-MS analysis.

Other Gradient conditions:

Gradient Condition 1: Starting at 15% buffer B the linear gradient was extended to 60% in 12.5 min, followed by a linear extension to 100% B over 2 minutes at a flow rate of 1.00 ml min^{-1} at 75°C.

Gradient Condition 2: Starting at 20% buffer B the linear gradient was extended to 22% Buffer B in 2 min, then from 22% to 52% buffer B over 15 min, then from 52% to 65% B for 2.5 min at a flow rate of 1.00 ml min^{-1} at 75°C.

2.9.2 RNA Mass Spectrometry Analysis

ESI-MS Analysis of crRNA on HCT Ultra PTM Ion Trap

ESI-MS was performed in negative mode using an ion trap mass spectrometer (HCT Ultra PTM Discovery systems, Bruker Daltonics) coupled with an online capillary liquid chromatography system (U3000 Dionex, UK). Intact RNA analysis was performed on a monolithic PS-DVB 200 µm i.d x 50 mm capillary column (Dionex UK) using the following gradient. Starting at 20% buffer B (see section 2.2.5), the linear gradient was extended to 90% buffer B over 15 minutes at a flow rate of 1.7 µl/min. The subsequent column washing extends over 3 minutes and column equilibration for 7 minutes. Analysis of RNA digests was performed on a monolithic PS-DVB 200 µm i.d x 50 mm capillary column (Dionex UK) using the following gradient. Starting at 5% buffer B (see section 2.2.5), the linear gradient was extended to 50% buffer B over 15 minutes at a flow rate of 1.7 µl/min. The subsequent column washing extends over 3 minutes and column equilibration for 7 minutes.

ESI-MS analysis of crRNA on Amazon Ion Trap

Electrospray Ionization Mass spectrometry was performed in negative mode using an Amazon Ion Trap mass spectrometer (Bruker Daltonics), coupled to an online capillary liquid chromatography system (Ultimate 3000, Dionex, UK). Intact RNA separations were performed using a monolithic (PS-DVB) capillary column (50 mm × 0.2 mm I.D., Dionex, UK). RNA analysis was performed at 50 °C with 20% buffer B, extending to 40% B in 5 min followed by a linear extension to 90% D over 8 min at a flow rate of 2 µl/min. For RNAse T1 digest analysis, 250 ng crRNA was digested with 20U RNAse T1 (Applied Biosystems). The reaction was incubated at 37oC for 30 minutes. The oligoribonucleotide mixture was separated on a PepMap C-18 RP capillary column (150 mm × 0.3 µm I.D., Dionex, UK) at 50 °C using gradient

conditions starting at 20% buffer C and extending to 35% D in 3 mins, followed by a linear extension to 60% D over 40 mins at a flow rate of 2 µl /min.

ESI-TRAP and ESI-TRAP QUOD conditions For RNA Analyses:

The electrospray ionisation-trap quadrupole (ESI-trap-QUOD) mass spectrometer (AmaZon, Bruker Daltonics) and ESI-TRAP (HCT Ultra PTM Discovery systems, Bruker Daltonics) was set to perform data acquisition in the negative mode with a selected mass range of 500−2500 m/z. The ionisation voltage of -2500 V was set to maintain capillary current between 30−50 nA. The temperature of nitrogen was set to 120°C at a flow rate of 4.0 L/h and N2 nebuliser gas pressure at 0.4 bar. Tandem mass spectrometry was performed with selection of charge state from -1 to -4 with the trap filled to a target of 100000 ions. Selected precursors that had been analysed >2 times were actively excluded from analysis for 1 min.

2.9.3 Nucleic acid Data analysis

Using mongo oligo mass calculator (http://mods.RNA.albany.edu/masspec/Mongo-Oligo) the molecular weight of the DNA, RNA and oligoribonucleotide fragments from RNAse A digests were calculated. Following data acquisition using ESI-trap QUOD MS (AmaZon, Bruker Daltonics) ,data was processed using Bruker Compass DataAnalysis (version 4.0 SP 2 (Build 234). For the charge / mass deconvolution the processing parameters were set to default except the following. a) Mass list > Sum peak: peak width (FWHM) 30 points. b) Charge deconvolution > Protein/ large molecules (low mass = 5000, high mass = 50000). c) Exclusion of the low intensity masses e.g. 500-700 m/z and 1300-2000 m/z.

Chapter III

Engineering a novel TAP tag for the analysis of Prp5 and its interacting proteins

Abstract

Prp5 is an RNA helicase of the DEAD-box family and plays a role in pre-spliceosome formation. It has been proposed that Prp5 bridges the U1 and U2 snRNPs to enable stable interaction of U2 snRNP with intron RNA. Prp5 was engineered with a novel TAP tag prior to tandem affinity purification in conjunction with mass spectrometry analysis to identify interacting partners. This work demonstrates that for the first time that a novel Protein A-p54 TAP tag can be used effectively to study protein-protein interactions in conjunction with TAP - MS studies. The results identified a number of U2 snRNP SF3b proteins consistent with the model that Prp5 interacts with the spliceosomal U2 snRNP through SF3b. Furthermore, a number of novel interacting proteins were identified including Npl3p in this study.

3.1 Introduction

Prp5 interactomics

Prp5 is an RNA helicase of the DEAD-box family and plays a role in pre-spliceosome formation (Abu Dayyeh et al., 2002a; de la Cruz et al., 1999; Xu et al., 2004). Further insight into the roles of Prp5 and spliceosome formation have been generated using a number of *in vitro* and *in vivo* studies indicating that Prp5 bridge the U1 and U2 snRNPs to enable stable interaction of U2 snRNP with intron RNA (Xu et al., 2004).

In an *in vitro* experiment by Kosowski et al, Prp5-TAP was used to pull-down splicing intermediates in conjunction with Western blot detection. Prp5-TAP was reported to pull down U1 and U2 snRNA fourfold to sevenfold above background levels (Kosowski et al., 2009). Prp5-GST was previously reported to pull down larger amounts of U2 snRNP than other snRNPs (Abu Dayyeh et al., 2002b). These studies suggest that Prp5p interacts more strongly with U2 snRNP than other snRNPs. More recently, a Prp5-TAP experiment performed by (Shao et al., 2012) in *S. pombe* using (CBP-Prot. A) U1A was reported to co-purify with Prp5p along with SF3b proteins. To further investigate the roles of Prp5p in spliceosome assembly and potential interactions with other pathways a novel TAP tag utilising the protein p54nrb was engineered to enable the TAP of Prp5 in *S. cerevisiae* in conjunction with LC MS analysis to identify the interacting proteins.

The Prp5-CBP-Prot A obtained from EUROSCARF yeast TAP fusion library (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/col_index.html) was analysed using PCR tag orientation experiments and western blot analysis. However, it was not possible to confirm the expression of the Prp5-CBP-Prot A. Therefore, an alternative Prp5 TAP construct was engineered prior to performing Prp5 TAP MS analysis.

In this Chapter it is proposed to engineer a novel *Saccharomyces cerevisiae* Prp5p with a p54-Protein A (TAP) tag, in order to study interactions of Prp5 within the spliceosome and other potential mRNA processing pathways. This TAP tag can be

used in conjunction with the high capacity properties of Ni-affinity resins to purify the low abundant Prp5, given that Prp5 is expressed under the regime of its endogenous promoter.

The aim of this study is to use affinity purification and mass spectrometry to study protein-protein interactions and protein-RNA interactions in the yeast spliceosome. It is proposed to use two TAP tags, a novel p54-TEV-Prot.A utilising the protein p54nrb from Hela cells and the standard CBP-Tev-Prot.A TAP tags in this study. The simpler TAP tag with poly (6X) His-Tev-Prot A (His6-TEV-Prot.A) is an alternative but has less affinity for Ni-NTA than the p54. Previous work in our lab has shown that p54 has a higher affinity for the Ni-NTA than the His6-TEV-Prot.A possibly because of the conformation and arrangement of Ni-binding domain of the p54 peptide. Therefore, its high-affinity makes p54 a more desirable choice especially for the purification of large protein complexes. In addition, the Ni-NTA provides relatively high yields of tagged proteins from inexpensive, high capacity resins and therefore reduces the cost of purification of protein complexes (Lichty et al., 2005). These advantages informed the choice of p54-Tev-Prot.A. These advantages informed the choice of p54-Tev-Prot.A

Although, TAP-tagging has been done on a more or less genomic-wide scale in yeast, Prp5-TAP was not successful. Often times interactions which are transient/sub stoichiometric are not identified because scale of cell culture. Secondly, because of the length of time required to conduct TAP purification, most of the protein components of complexes are degraded. In order to address these issues, it is therefore proposed in this study to scale up the cell culture and develop a more efficient and effective TAP procedure that will shorten the time of purification, effectively purifying complexes and identifying potential interactions which might be sub-stoichiometric. This study aims to provide further insight into *in vivo* protein-protein interactions of Prp5 in *S. cerevisiae* using the TAP-MS approach. Although the TAP-MS approach has been used to study Prp5 in *S. pombe*, this is the first time TAP-MS is used to study Prp5 in *S. cerevisiae*. This is also the first time a novel TAP tag, p54-Protein A, will be used to study Prp5 interactions.

3.2 Results and Discussion

3.2.1 Genetic engineering of *Saccharomyces cerevisiae* Prp5-TAP by PCR-mediated gene targeting

Tandem affinity purification, developed by (Puig et al., 2001), utilizes two sequential affinity tags to decrease non-specific interactions either with tagged protein or affinity chromatographic matrices (see Chapter 1). Since the development of the original CBP-Protein A TAP tag, alternative versions containing different affinity peptides and protease cleavage sites have been made (see Chapter 1.4.4). CBP-Prot A TAP is an effective affinity purification tag however, alternative affinity tags to replace CBP have been made to overcome potential problems associated with its properties (Lichty et al., 2005; Oeffinger, 2012). For instance, the elution of protein complexes from the CBP tag requires the use of EGTA or other chelating agents which presents a problem when isolating active complexes containing metal-binding domains. In this work, CBP tag is replaced with a 7.12 kDa peptide from the protein p54 which contains a proline-rich Ni-binding domain derived from the protein (see Figure 3.1). To engineer a p54-ProtA TAP tag, the IgG-binding subunit of protein A from *Staphylococcus aureus* (Prot A) and an N-terminal peptide from p54 (a nuclear protein) were used in tandem separated by a TEV protease cleavage site allowing for tandem affinity purification(see Figure 3.1). The p54-ProtA TAP tag was engineered because the Prp5-CBP-Prot A was not stably expressed and could not be confirmed by PCR or Western blot analysis.

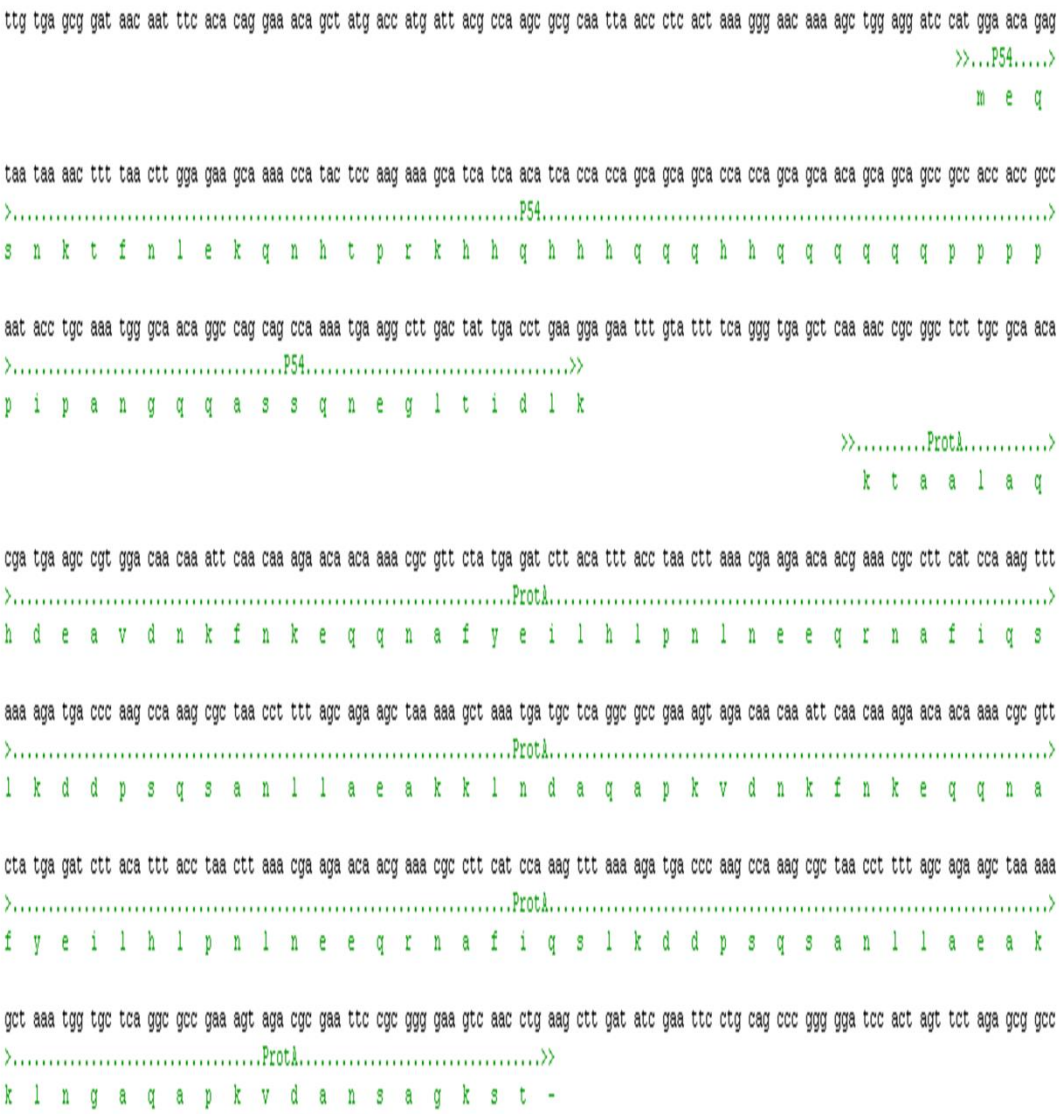


Figure 3.1 TAP tag peptide sequence. Two ORFs, ProtA and p54, are represented as fusion tags in tandem at the C-terminal region of Prp5 just before the stop codon. The order of the TAP tag is p54 → TEV protease cleavage site → ProtA.

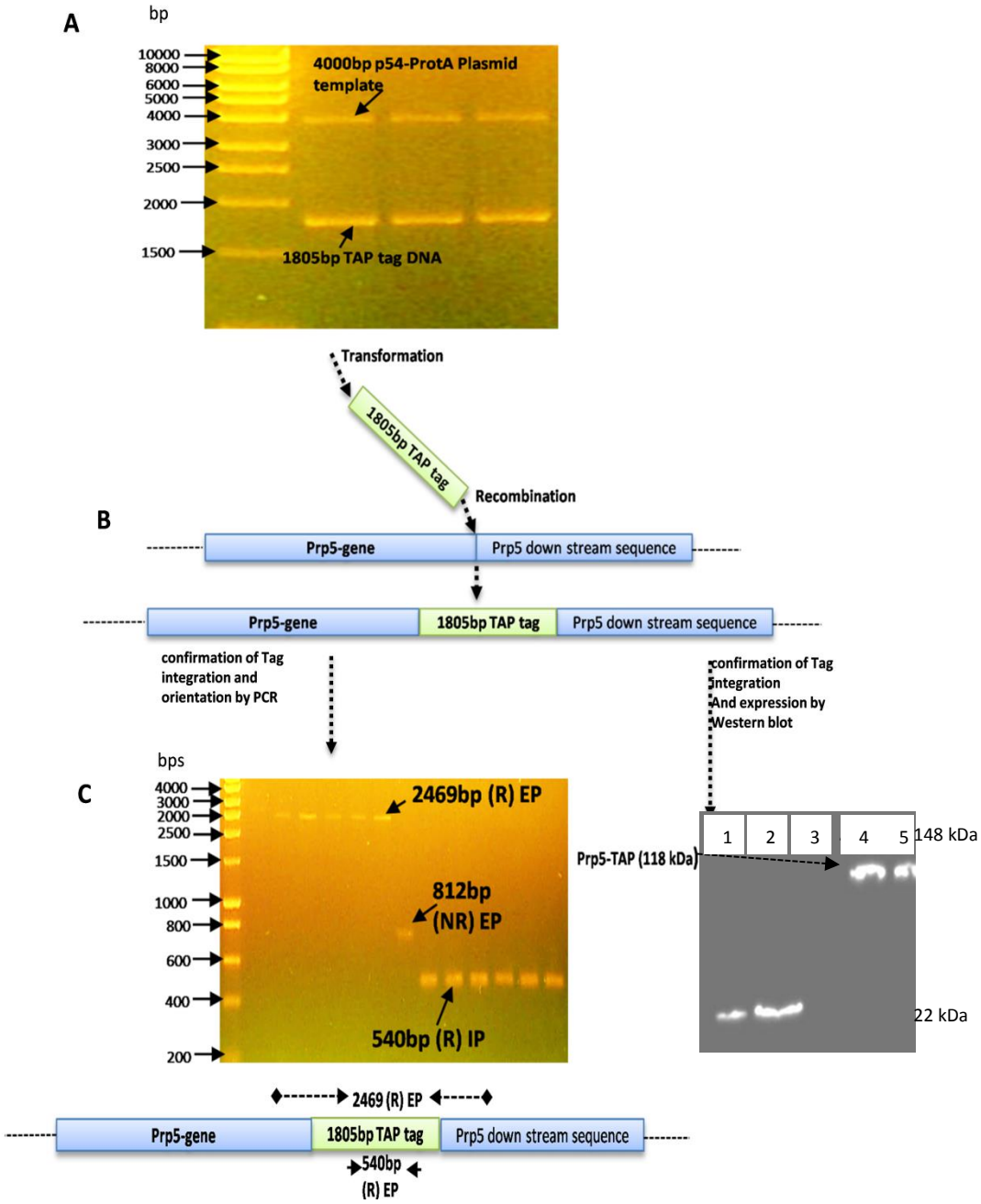


Figure 3.2 Genetic engineering of the Prp5-TAP strain. Integration of TAP tag into C-terminal the locus of the Prp5 gene was confirmed by checking TAP tag integration by PCR and expression of Prp5-TAP by Western blotting. A) Plasmid DNA containing the Protein-A-p54 gene and Trp- nutritional complement was amplified by PCR using flanking primers whose ends are designed to homologously recombine with Prp5 C-terminal locus. A PCR product corresponding to the expected 1805 bp was confirmed by agarose gel electrophoresis (annotated). B) The 1805 bp PCR product was used for transformation to integrate the TAP tag cassette at the Prp5 C-terminal locus. C) Following successful transformation and nutritional selection, tag integration and orientation was confirmed by PCR using external primers (EP) producing the expected 2469 bp PCR product (R) The –ve control without integration generates the 812 bp PCR product. PCR using an internal primer (IP) producing a PCR product of 540 bp was also used to confirm tag integration. D and ii) Western blot analysis targeting the Protein A epitope of the TAP tag was used to confirm Prp5-TAP tag expression. Lane1-2) Lsm4 as positive control 3) Wild type strain as negative control.

To perform analysis of Prp5p protein-protein interactions, *S. cerevisiae* strain was prepared in which a novel tandem affinity purification tag comprising of p54 (on the N-terminal end) and Protein A (on the C-terminal end) was inserted at the 3'-end of the endogenous yeast Prp5 gene via homologous recombination (see Figure 3.1 and 3.2). A section of the plasmid, Pab101, containing the p54-ProtA gene and Trp-complement was amplified with flanking primers and the resulting PCR product confirmed by agarose gel electrophoresis (see Figure 3.2A). The PCR-amplified product was used to transform a wild type strain of *Saccharomyces cerevisiae*, integrating the TAP tag at the 3'-end of endogenous Prp5 gene by homologous recombination using LiAC/ssDNA/PEG method (see Chapter 2.3). TAP tag integration and orientation was confirmed by PCR (see Figure 3.2C). Prp5-TAP tag expression was confirmed by Western blot analysis using anti-protein A antibody (see Figure 3.2D).

3.2.2 Tandem affinity Purification of Prp5

Single colony from Prp5-TAP and wild type YPD plate were inoculated into 5 mL YPD media and grown at 30°C overnight (for 14 hours). Their OD_{600} were compared and there was not much significant difference between the two strains suggesting that the Prp5-TAP strain was healthy. In two experiments, 10L Yeast cell culture were grown in YPD media to exponential growth phase at 30°C. Tandem affinity purification (TAP) procedure for these experiments was adapted from the basic principle developed by (Puig et al., 2001) (see Chapter 2.3.3). A workflow summarising the methodology is shown in Figure 3.3. In summary the cells were harvested and lysed at 4°C. The lysed cells were centrifuged at 13500 rpm, 4°C for 10 minutes, supernatant collected and further centrifuged at 24000 rpm for 20 minutes. Further clearing of the supernatant was performed using Sepharose-CL 4B.The first affinity purification was performed using IgG sepharose beads. Cleavage from the IgG beads was performed using TEV protease. The second purification step utilised Ni-NTA beads. After binding, the Ni-NTA beads were washed and the bound proteins were eluted using elution (250 mM imidazole in PBS). The fractions were analysed by SDS PAGE (see figure 3.4) and the proteins identified using in-gel

digestion in conjunction with reversed-phase chromatography interfaced to electrospray ionisation tandem mass spectrometry (LC ESI MS/MS) (see Chapter 2.7). Protein identifications were performed using MASCOT in conjunction with Prohits to further interrogate the protein interactions. Protein identifications were based on a minimum of at least two unique peptides and a MASCOT cut off score of 27 with typical false discovery rates were 1% (see Chapter 2.7.1). Proteins identified with only one unique peptide with an ion score above 27 and were manually verified using Bruker data analysis. Proteins identified in the mass spectrometry analysis were classified with the aid of an online database resource, DAVID Functional Annotation Table (http://david.abcc.ncifcrf.gov/) and Kegg's pathway.

Criteria for discussing identified proteins

MASCOT uses a probability based scoring algorithm which enables a simple rule to be used to judge whether a hit is significant or not (http://www.matrixscience.com/help/scoring_help.htm). Protein identifications were based on a minimum of at least two unique peptides and a MASCOT peptide ion score cut off of 27. Typical false discovery rates were 1% (see Chapter 2.7.1) .The false discovery rate (FDR) is an indication of the false positives in the data, with a high FDR indicating high frequency of false positives.

With the aid of the Prohits software, proteins that co-purified with the bait were screened against a control TAP done on strain harbouring a non-splicing protein bait- TAP tag. Using a heuristic approach, proteins which are frequently observed in TAP literature, such as ribosomal proteins, heat shock proteins and cytoskeletal proteins were eliminated from the list. Following protein identifications based on the above criteria, literature searches were used to provide further insight into the biological roles of the identified interacting partners.

Those proteins with no obvious connection to splicing or gene expression which had low MASCOT scores were not discussed but some with relatively significant scores, for instance, the CopII proteins were highlighted on the table (See Table 3.1). Each protein identification the protein MASCOT scores are shown which reflects the

probability that the observed peptide match is a random event (false positive). High MASCOT score is a low probability that the peptide match is a random and low MASCOT score, a high probability the peptide match is a random event (false positive) (http://www.matrixscience.com/help/scoring_help.htm.) This, therefore, reflect the confidence of the identification or probability of the identification). However, low MASCOT score could result from the length or nature of the protein amino acid sequence. For instance, a short protein length with relatively few trypsin cleavage sites will generate a low MASCOT score as only fewer peptides are generated. The paucity of tryptic-specific cleavage sequence even in a large protein (long protein sequence) can also mean fewer peptides, hence low MASCOT score.

3.2.3 Identification of Prp5 interacting proteins using mass spectrometry

Following TAP and SDS PAGE analysis MS spectrometry was used to identify over 34 potential interacting proteins (see Table 3.1). The results identified the SF3b proteins Hsh155 and Rse1p (Table 3.1).

Hsh155 has previously been reported to interact with Prp5 in a two-hybrid experiment (Wang et al., 2005) and U2 snRNP, but Prp5 has not been reported to co-purify or interact with Rse1p *in vivo*. For the first time we show evidence for interaction of Prp5p with Rse1p *in vivo*. It has been suggested that Prp5 mediates the interaction between U1 and U2 snRNP by serving as a bridge between the two snRNPs (Shao et al., 2012). It has been proposed that U1 snRNP interacts with the SR-like protein Rsd1p and the latter interacts with Prp5p. Subsequently Prp5p contacts SF3b of U2 snRNP thereby forming a bridge between U1 and U2 snRNP (Shao et al., 2012; Wang et al., 2005) . The co-purification of SF3b proteins, Hsh155p and Rse1p with Prp5p in this experiment is consistent with the model that Prp5 interacts with the spliceosomal U2 snRNP through SF3b. However, no components of the U1 snRNP were detected in the analysis suggesting that Prp5 may not interact directly with U1 snRNP. Rsd1p, on the other hand, has no homolog in *S. cerevisiae*, hence was not identified in this experiment. The non-detection of U1 snRNP and lack of Rsd1p homolog in *S. cerevisiae* raises questions about the conservation of the role

of Prp5p in bridging interactions between U1 and U2 – at least, in *S. cerevisiae.* First, if Prp5p bridges interaction with U1 and U2 Prp5p would be expected to co-purify with U1 snRNP. Secondly, it is likely that a homolog of Rsd1p or a complementary surrogate exist to perform this Rsd1p function if this mechanistic model is conserved.

To address these questions, if the interaction of Prp5p with U1 snRNP is indirect, as suggested (Shao et al., 2012), it may not co-purify with U1 snRNP or may do so at a very low level, especially if the interaction with Rsd1p-U1 snRNP is transient. However, the possibilities of the TAP tag (p54-Prot. A) interfering with the interaction of Prp5 with U1 snRNP has to be noted. In the Prp5-TAP experiment performed by (Shao et al., 2012) in *S. pombe* using (CBP-Prot. A) U1A was reported to co-purify with Prp5p along with SF3b proteins. The fact that the genetic background of *S. pombe* is different to *S. cerevisiae* makes it difficult to conclude that the variability in results may be due to merely differences in TAP tag used. In a previous *in vitro* study where Prp5-TAP was used to pull-down splicing intermediates using Northern blot analysis as the detection method, Prp5-TAP was reported to pull down U1 and U2 snRNA fourfold to sevenfold above background levels (Kosowski et al., 2009). Also, Prp5-GST was previously reported (Abu Dayyeh et al., 2002b) to pull down larger amounts of U2 snRNP compared to other snRNPs. These results suggest that Prp5p interacts more strongly with U2 snRNP than other snRNPs. There is no strong evidence to suggest that Prp5p interacts with other splicing complexes. However, the present study suggests that among the snRNPs, Prp5 interacts only with U2 snRNP SF3b. Again, the possibility of the TAP tag dislodging other snRNP cannot be excluded.

It is proposed that Prp5 acts as a U1-U2 bridging protein which is conserved in yeast but it utilizes a surrogate protein which performs a similar function as Rsd1p. Therefore, a protein that may not be a homolog of Rsd1p but which fits the profile was sought in the data. Interestingly, the SR-like protein, Npl3p, was identified (see figure 3.4 and table 3.1) which was not previously detected in previous TAP studies (*in vitro*). Previous results have shown that mutation of NPL3 reduces the occupancy of U1 and U2 snRNPs at genes, whose splicing is stimulated by Npl3p (Kress et al., 2008). The interaction of Npl3p with Prp5p will fit a model where Npl3p serves as a

surrogate Rsd1p since both proteins are SR-like proteins and they promote the recruitment of U1 and U2 to the pre-mRNA. However, further experiments are needed to confirm if the role of Npl3p and Prp5p in *S. cerevisiae* follows the mechanistic model suggested for *S. pombe* during the alleged U1 and U2 interaction. This is the first time it is demonstrated that Prp5p interacts with Npl3p.

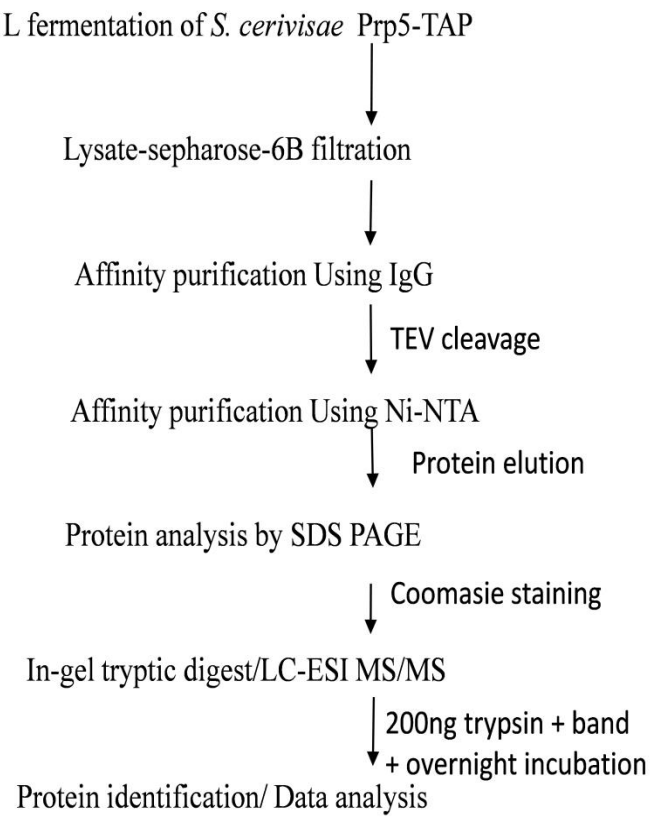


Figure 3.3 workflow summarising the methodology of TAP MS analysis

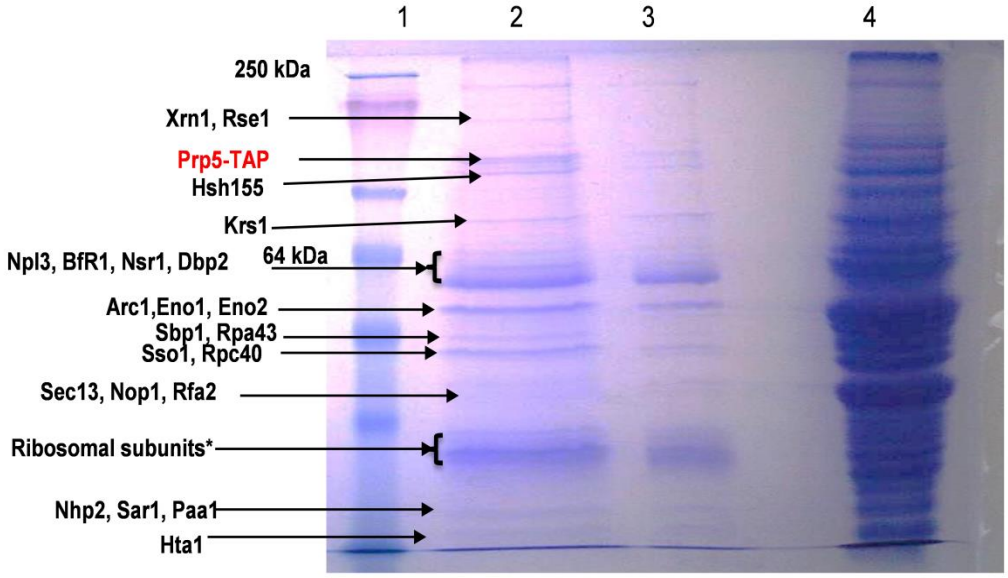


Figure 3.4 TAP analysis of Prp5-TAP. 10% SDS PAGE analysis of TAP of Prp5. Protein identifications were performed using LC MS analysis. Lane 1 is the marker; Lane 2 is the second purification step (p54-Ni-NTA); lane 3, first purification step (Protein A-IgG module) and Lane 4, the lysate (soluble fraction). A number of U2 snRNP associating proteins and novel proteins identified by MS are highlighted.

It would be interesting to investigate if Prp5p really interacts with U1 snRNP and to what extent. Another methodological approach may be needed to verify the interaction and especially if the interaction is transient since *in vitro* affinity based studies and TAP studies till date has not yielded sufficient evidence that indeed there is such an interaction. However, the data from this study is consistent with the model that Prp5p does interact with SF3b subcomponent of the U2 snRNP. It also demonstrates a novel interaction with Npl3p which raises the possibility that Npl3p may be a functional counterpart of Rsd1p in *Saccharomyces cerevisiae*. This argument is reinforced by the evidence that Npl3p co-immunoprecipitates with U1 snRNP and the branch point proteins (BBP) (Gavin et al., 2002). It has also been shown to co-purify with U2 snRNP (also see chapter 5), and mutation of Npl3p has been demonstrated to reduce the occupancy of U1 and U2 at genes whose splicing is stimulated by Npl3p, hence required for pre-spliceosome formation (Kress et al., 2008).

Protein	Mascot Score	Number of peptides	Prp5-TAP *S. pombe* (Shao et al. 2012)
Prp5	2144	132	✓
Hsh155/SF3b155	240	6	✓
Rse1	193	5	✓
Hsh49/SF3b49	-	-	✓
SF3b130	-	-	✓
Rsd1	-	-	✓
U1A/Mud1	-	-	✓
Snu71	-	-	✓
Snu145/Cus1	-	-	✓
Clu1	2048	57	-
Sbp1	752	31	-
Gcd11	851	27	-
Shm1	887	23	-
Npl3	648	22	-
Cdc48	801	19	-
Krs1	681	19	-
Sec13***	485	16	-
Bfr1	456	10	-
Gus1	379	9	-
Arc1	316	8	-
Kem1	288	9	-
Sui2	348	7	-
Rtn1	224	7	-
Sar1***	268	6	-
Nhp2	150	6	-
Fun12	246	5	-
Rpc40	182	4	-
Nop1	141	4	-
Tsr1	126	3	-
Ded81	95	2	-
Cdc33	87	2	-
Rpa43	75	2	-
Nsr1*	72	1	-
Sso1*	59	1	-
Rfa2*	53	1	-
Rtn2*	48	1	-
Scp160*	47	1	-
Sac1***	41	1	-
Dbp2*	30	1	-
Hrp1*	27	1	-

Table 3.1 Summary of interacting proteins identified in the Prp5-TAP. *S. pombe* and *S cerevisiae* Prp5-TAP are compared. A parent wild type of the recombinant Prp5-TAP was also used as a control. Protein identifications are only shown for those that were not identified in the –ve control. In addition common contaminants, such as proteasome, ribosomal, heat shock proteins and known cytoskeleton proteins were also excluded. * Those identified with single peptide at MASCOT ion score cut off, 27, were manually verified, and were not detected in the control. *** Protein components of CopII coat of vesicles.

3.2.4 Prp5 interacts with arginine methylated Npl3

Npl3p plays a role in the transport and localisation of mRNAs in the cytoplasm (Lee et al., 1996b) and mRNA translational repression (Windgassen et al., 2004). It has also been implicated in the co-transcriptional recruitment of U1 and U2 snRNP to the pre-mRNA (Kress et al., 2008). Arginine methylation of Npl3p is reported to facilitate its export directly to the cytoplasm by weakening contacts with nuclear proteins (McBride et al., 2005a). The diversity in the patterns of arginine methylation is enhanced by the ability of this residue to be methylated in three different ways on the guanidine group: monomethylated (MMA), symmetrically dimethylated (sDMA) and asymmetrically dimethylated (aDMA)(see Figure 3.5), each having potentially different functional roles (Bedford, 2007).

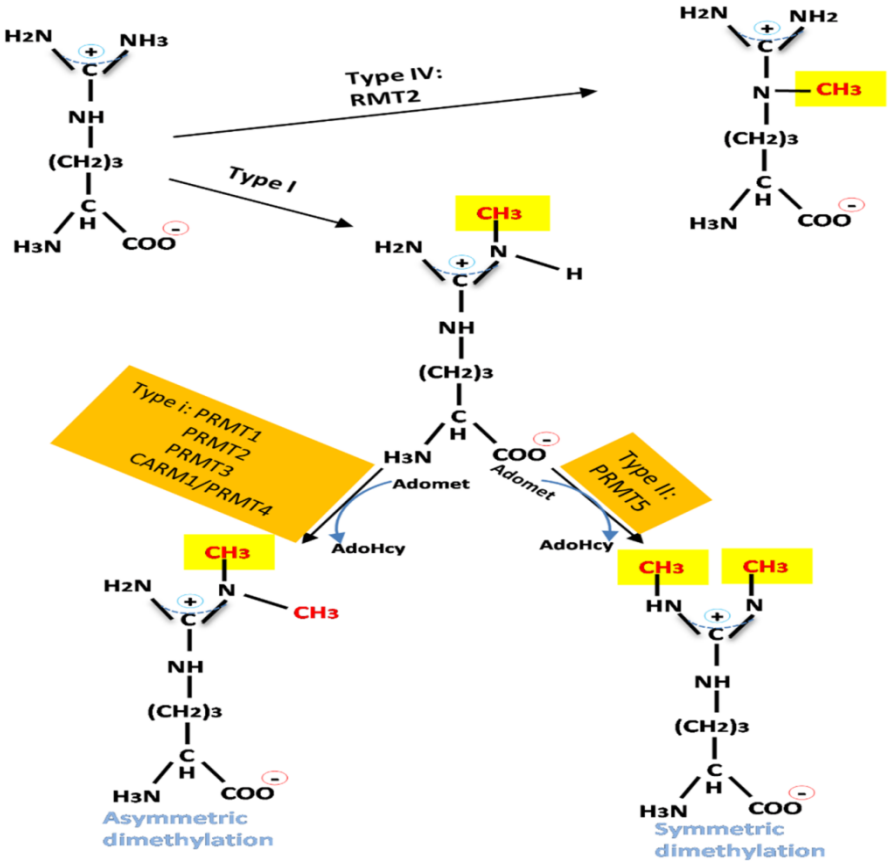


Figure 3.5 Schematic illustration of protein arginine methylation. The guanidine group can be monomethylated or asymmetrically dimethylated by Type 1 PRMT enzymes. It can also be symmetrically dimethylated by Type II PRMT enzymes.

Following identification of Npl3p further interrogation of the data was performed to identify sites of arginine methylation. The Npl3p RGG peptides, GSYGGSRGGYDGPR and GGYSRGGYGGPR, were identified with dimethyl group on the R7 and R5, respectively. The modified peptides were subsequently verified by manual inspection of the tandem MS (see Figure 3.6 and 3.7). The Npl3p RGG peptide, GSYGGSRGGYDGPR (monomethyl (R7)) was also identified with a low MASCOT score. The data was further analysed to characterise the type of arginine methylation present on Npl3p. The identification of the neutral loss of dimethylamine [M+nH-45]n+ shows that the dimethylated arginine on both R_5 and R_7 is asymmetric (aDMA). In addition, the MS/MS analysis also confirmed the neutral loss of monomethylamine $[M+nH-31]^{n+}$ consistent with monomethylation of R_7 (see Figure 3.6 and 3.7). This is the first time the type of arginine methylation on Npl3p has been characterised.

Previous studies have investigated the role of methylation of Npl3p (McBride et al., 2005a). All arginine sites were replaced with lysine on the basis that lysine is commonly employed to probe sites of methylation of arginine and to elucidate the significance of arginine methylation (Najbauer et al., 1993; Xu and Henry, 2004). They observed that replacement of arginine with lysine caused Npl3pto exit the nucleus. They reported that methylation of Npl3p influences its transportation out of the nucleus by weakening its association with nuclear proteins. But there may be potentially variable subsets of arginine methylation at RGG sites, many of which could influence Npl3p activity in variety of ways. Therefore the replacement of all RGG arginine does not address the question of whether variable subsets of methylations of RGG affect the activity and interaction of Npl3p with other proteins in different ways. These studies focussed on the interaction of methylated Npl3p with itself and Tho2. Methylation of Npl3p may facilitate the export of Npl3p as already shown by previous experiments (McBride et al., 2005a; Xu and Henry, 2004) but it does not necessarily suggest that association of Npl3p with all nuclear proteins are abolished by methylation. It is possible that methylation of all RGG sequences of Npl3p may weaken its interactions with certain nuclear proteins; however, it is possible that variable subsets of RGG methylated sites may affect Npl3p interactions with nuclear proteins differently.

As shown in this study, Prp5p co-purified with methylated Npl3p suggesting that some methylated Npl3p interacts with Prp5p.This shows that Npl3p associates with the nuclear Prp5p, as U2 snRNP proteins also co-purified. The identification of dimethylated arg^{363} and arg^{384} sites suggests that Prp5p interacts with Npl3p that is methylated at these sites. However, further experiments are needed to determine whether arginine methylation of Npl3 perturbs the interaction with Prp5 and if interactions of Prp5p with Npl3p is affected with variable levels of methylation at Npl3p RGG sites.

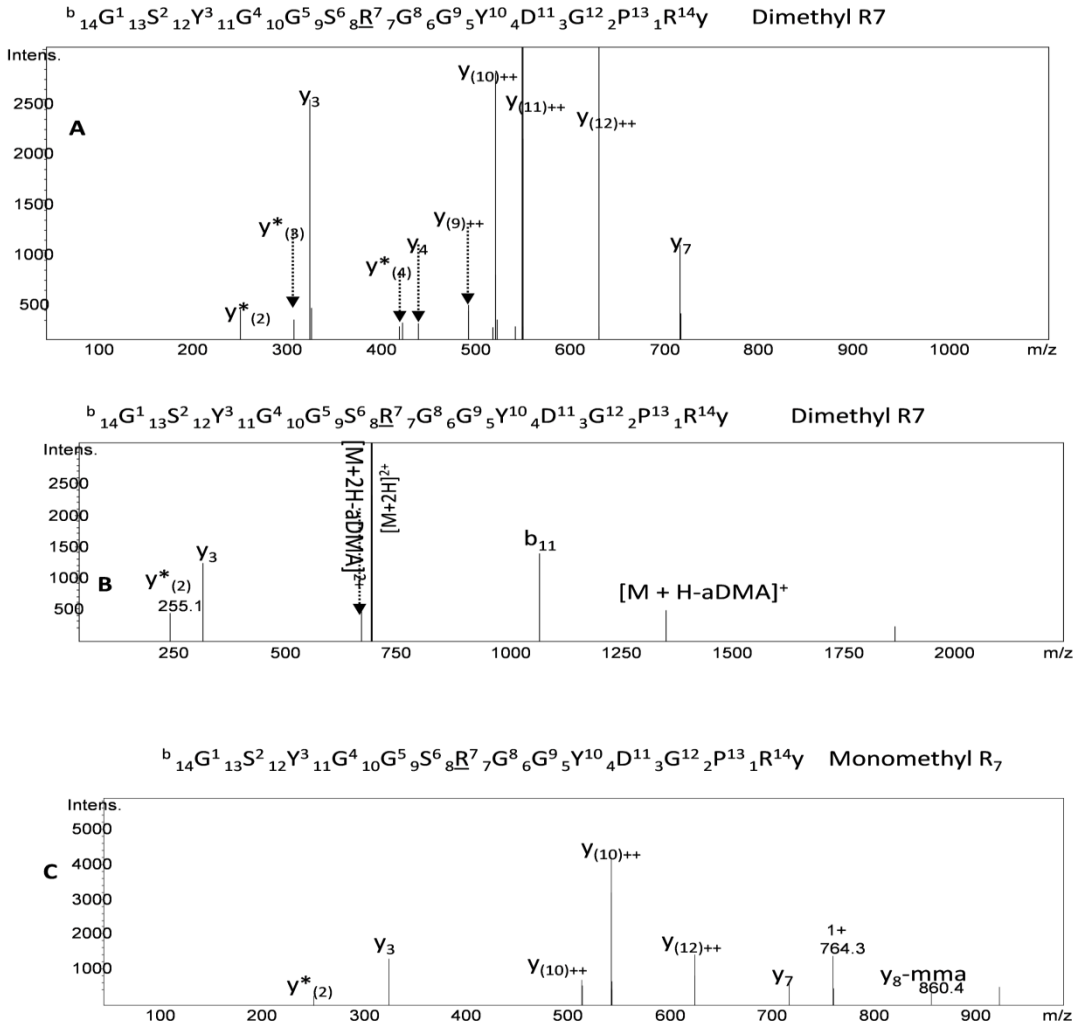
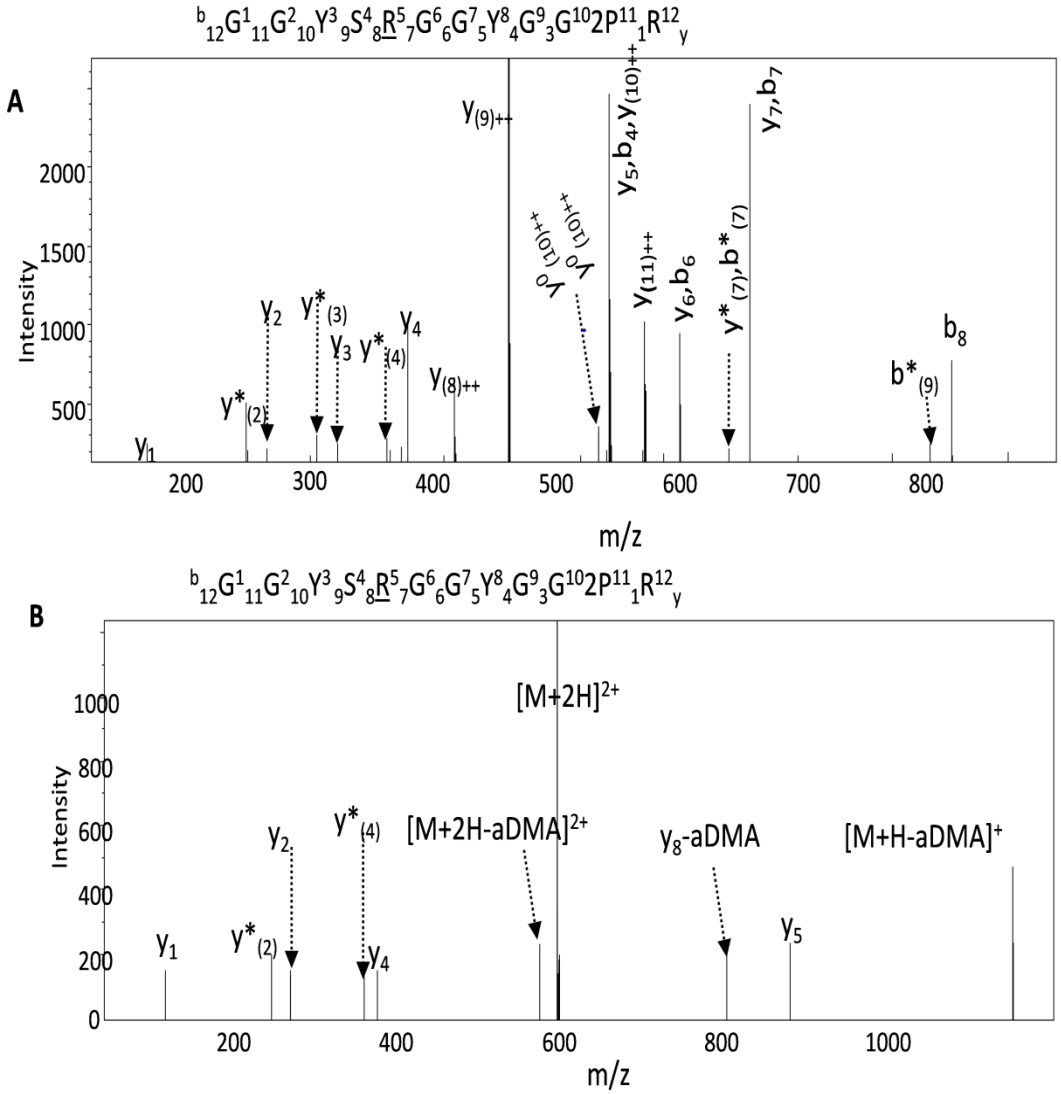


Figure 3.6 Tandem MS analysis of the methylated peptide GSYGGSRGGYDGPR. (A)CID MS/MS of GSYGGS(dimeR)GGYDGPR $(M + 3H)^{3+}$ B) CID MS/MS analysis of GSYGGS(dimeR)GGYDGPR $(M + 2H)^{2+}$ showing neutral loss of dimethylamine (DMA) C) CID MS/MS analysis of GSYGGS(monomethyl R) GGYDGPR $(M + 3H)^{3+}$. The prominent y and b ions are highlighted.



C

1 MSEAQETHVE QLPESVVDAP VEEQHQEPPQ APDAPQEPQV PQESAPQESA
51 PQEPPAPQEQ NDVPPPSNAP IYEGEESHSV QDYQEAHQHH QPPEPQPYYP
101 PPPPGEHMHG RPPMHHRQEG ELSNTRLFVR PFPLDVQESE LNEIFGPFGP
151 MKEVKILNGF AFVEFEEAES AAKAIEEVHG KSFANQPLEV VYSKLPAKRY
201 RITMKNLPEG CSWQDLKDLA RENSLETTFS SVNTRDFDGT GALEFPSEEI
251 LVEALERLNN IEFRGSVITV ERDDNPPPIR RSNRGGFRGR GGFRGGFRGG
301 FRGGFSRGGF GGPRGGFGGP RGGYGGYSRG GYGGYSRGGY GGSRGGYDSP
351 RGGYDSPRGGYSRGGYGGPR NDYGPPRGSY GGSRGGYDGP RGDYGPPRDA
401 YRTRDAPRER SPTR

Figure 3.7 Tandem MS analysis of the methylated peptide GGYS(dimeR)GGYGGPR (A)CID MS/MS of GGYS(dimeR)GGYGGPR $(M + 3H)^{3+}$. B) CID MS/MS analysis of GGYS(dimeR)GGYGGPR $(M + 2H)^{2+}$ showing neutral loss of dimethylamine (DMA). The prominent y and b ions are highlighted. C) Npl3p sequence with the methylated regions highlighted in red.

Observed Mass/charge (m/z)	Calculated Mass (Da)	Experimental mass (Da)	MASCOT Score	Expect value	Rank	Sequence
404.5350	1210.5832	1210.5843	27	0.07	1	R.GGYSRGGYGGPR.N + Dimethyl (R)
606.2990	1210.5834	1210.5843	(16)	0.78	1	R.GGYSRGGYGGPR.N + Dimethyl (R)
467.2170	1398.6292	1398.6277	(19)	0.29	1	R.GSYGGSRGGYDGPR.G + Methyl (R)
471.8890	1412.6452	1412.6433	25	0.086	1	R.GSYGGSRGGYDGPR.G + Dimethyl (R)

Table 3.2 Arginine methylated Peptides identified for Npl3 by mass spectrometry analysis. The methylated arginines are underlined and type of methylation indicated after the peptide sequence. Trypsin, used as the digestive enzyme, cleaves between R and subsequent amino acids. Details of the MS and tandem MS spectra analyses are shown in Figure 3.6 and 3.7.

3.2.5 Potential Interaction of Prp5 with Transcription, nucleolar and Translation factors

Transcription factors such as Rpc40, Rpa43 and Rfa2 were identified in the Prp5-TAP MS study (see Table 3.1). Rpc40p is a subunit of the RNA polymerase 1 and III, while Rpa43p is found in both RNA polymerase II and III of *S. cerevisiae*. ((Beckouet et al., 2008; Panov et al., 2006) and reviewed in Kuhn et al. (Kuhn et al., 2007) . Nucleolar proteins, such as Nhp1 and Nop1, also co-purified with Prp5-TAP in this experiment. Translation factors, such as the tRNA synthetases, Ksr1p and Gus1p co-purified with Prp5-TAP. Arc1 binds Gus1p and Mes1p, delivering tRNA to them also co-purified with Prp5-TAP. This evidence suggests a potential interaction of Prp5p with tRNA synthetases. Also factors implicated in translation, such as Cdc33p and Sbp3p, also co-purified with Prp5. Sbp3p has been implicated in repression of translation and

mRNA decapping. Dbp2, another protein involved in mRNA decay and repression of aberrant transcription (Bond et al., 2001; Cloutier et al., 2012) was also identified in the analysis. The data is consistent with a notion that Prp5p, in addition to its role in spliceosome, may potentially interact with RNA processing pathways, such as transcription, tRNA pathways and regulation of translation.

3.2.6 Prp5p co-purified with (Kem1p) Xrn1p

Kem1p is a 5' to 3' exoribonuclease associated with the cytoplasmic processing bodies (P-bodies) involved in mRNA decay (Sheth and Parker, 2003). It's role in other processes has been reported, such as telomere maintenance (Askree et al., 2004) and ribosomal rRNA processing (Geerlings et al., 2000). In this study, Kem1p co-purified with Prp5p. This is the first time an evidence for interaction of Prp5p with Kem1p has been reported. This evidence either suggests a potential role for Kem1p in splicing or the role of Prp5p in other mRNA processing pathways where Prp5p associates with Kem1p. Further studies will be required to bring more insight into the questions this association raises.

Conclusions

A novel TAP tag utilising the protein p54nrb was engineered to enable the TAP of Prp5. Following TAP using this novel strategy, mass spectrometry was utilised to identify a number of interacting partners. The co-purification of SF3b proteins, Hsh155p and Rse1p with Prp5p in this experiment is consistent with the model that Prp5 interacts with the spliceosomal U2 snRNP through SF3b. In addition, an SR-like protein, Npl3p, was also identified in this study which has not been previously identified as in interacting partner of Prp5. However, Npl3p has been implicated in the co-transcriptional recruitment of U1 and U2 snRNP to the pre-mRNA (Kress et al., 2008). Furthermore, a number of sites of arginine methylation of Npl3p have previously been identified and shown to affect the interaction with the transcription

elongation factor Tho2 and inhibit Npl3 self-association. The mass spectrometry analysis identified a number of sites of arginine methylation present on Npl3p from the TAP analysis. Furthermore MS studies were used to characterise the type of arginine methylation present on Npl3p. Asymmetric dimethylation (aDMA) at positions Arg^{363} and Arg^{384} were identified in this study. No proteins of the U1, U6 or U5 snRNP were identified in this study; therefore indicating that within the spliceosome Prp5 directly interacts only with U2 snRNP SF3b. The data also suggests that the Prp5 complex interacts with transcription factors and demonstrates potential novel interactions with other proteins involved in RNA processing, such as Kem1p (Xrn1p).

Chapter IV

TAP MS studies of the *S. cerevisiae* Lsm proteins –an insight into mRNA splicing/processing dynamics

Abstract

The Lsm proteins are Sm-Like proteins that are associated with spliceosomal U6 snRNA. However, they are also known to interact with RNA degradation pathways. In this Chapter, using standard and novel modified tandem affinity purification protocols, I analysed the protein-protein interactions of a number of Lsm proteins (1 and 8) using TAP MS approaches. This study aims to characterize Lsm1-7p and Lsm2-8p pathways and provide further insight into interactions with other novel cellular pathways. In addition, this study aims at providing further insight into the dynamic composition of the spliceosome at targeted stages of spliceosome assembly. In this Chapter I show that, in addition to co-purifying with U4/U5.U6 splicing factors, Lsm proteins participate in novel interactions with other mRNA processing pathways. Lsm1-TAP and Lsm8-TAP co-purified with mRNA degradation, TREX, transcription, H-ACA/CD Box and translation initiation factors. In addition, Lsm1-7p and Lsm2-8p co-purify with TREX, Transcription and H-ACA/CD Box proteins. Comparison across the two different complexes reveals that apart from the presence of Lsm proteins they differ in many respects with regards to snRNA and protein compositions and therefore are functionally distinct. Furthermore, this study supports the notion that the spliceosome undergoes remodelling during splicing and provides further evidence for coupling of splicing, mRNA transport, mRNA processing/degradation and translation to transcription.

4.1 Introduction

The spliceosome is a megadalton dynamic protein-RNA complex comprising of several proteins, snRNAs and pre-mRNA (see Chapter 1.5). Lsm proteins are involved in splicing and other RNA processing pathways (Mayes et al., 1999; Tharun et al., 2000). It is suggested that two evolutionary similar but distinct complexes are formed by Lsm proteins. A complex formed by Lsm2p to Lsm8p is suggested to be the heptameric complex of the spliceosome while a complex of Lsm1p to Lsm7p has been implicated in mRNA decapping/deadenylation complex (Chowdhury et al., 2007; Tharun et al., 2000). Apart from these two pathways/complexes, it is not clear if other pathways physically interact with Lsm1-7p and Lsm1-8p. Although splicing and mRNA transport are thought to be co-transcriptionally coupled, the mechanism is not well understood (Tilgner et al., 2012; Will and Lührmann, 2011).

The Lsm proteins are associated mainly with the U6 snRNP and the splicesome (Mayes et al., 1999). They bind specifically with the 3' terminal oligo(U) tract of U6 snRNA and are thought to function in pre-mRNA splicing by mediating U4/U6 snRNP formation(Achsel et al., 1999). However, the Lsm proteins are also known to play a role in mRNA degradation (Chowdhury et al., 2007; He and Parker, 2000; Tharun et al., 2000).

Lsm1 is a part of the complex, Lsm1-7p, known to accumulate in the cytoplasmic foci and implicated in yeast mRNA decapping and decay (Tharun et al., 2000). Messenger RNA degradation is mainly a deadenylation-mediated decapping process which leads to 5' to 3' mRNA decay in eukaryotes (Beelman and Parker, 1995; Jacobson and Peltz, 1996). The predominant decay mechanism, however, appears to be in 3' to 5' direction in mammalian cells (Chen et al., 2001; Mukherjee et al., 2002). The work of Tharun et al. showed that mutations in seven Lsm1-7 proteins leads to inhibition of mRNA degradation (Tharun et al., 2000). They also showed that Lsm1-7p co-immunoprecipitates with mRNA degradation proteins such as Pat1, a protein thought to activate decapping and the decapping enzyme, DCP1. Their study indicated that Lsm involved in mRNA degradation may be distinct from the U6-snRNP associated Lsm proteins and that they may form unique complexes, playing different roles in mRNA metabolism. Although, these studies show that Lsm1-7p

may be distinct from Lsm2-8p, no evidence has exhaustively shown that Lsm1-7p may not indeed interact with spliceosomal proteins, especially since Lsm1 has been observed in the nucleus (Tharun et al., 2000). This raises a question about biogenesis of Lsm1: is the Lsm1-7p complex a U6 snRNP complex that has merely undergone conformational changes to acquire new, non-splicing roles or is it just a distinct complex that forms independent of the spliceosome? Does Lsm1 interact with splicing factors? To provide further insight into these questions, this study elected to exhaustively probe any potential interaction between Lsm1-7p complex and the spliceosome/nucleus.

The immunofluorescence and Western blot techniques used previously to distinguish the two Lsm complexes were also limited to the number of antibodies and fluorescent-tagged proteins available. Therefore, there may be potential interactions that may not have been identified because of the limitation imposed by the immunological techniques. Mass spectrometry, on the other hand, has no such limitation as it can potentially identify all proteins co-purifying with given bait.

Although, Lsm1-TAP MS identified the Lsm complex protein partners, a few other previously reported proteins including Pat1p, Nam1p, Dcp2 and Xrn1p proteins, proteins such as Dcp1p and other potentially sub-stoichiometric protein where not identified. Therefore, the failure to identify potentially important interactions in Lsm1-TAP suggest that genome-wide scale can fail to identify important protein interactions that are potentially useful in gaining further insight into Lsm protein-protein interactions dynamics. Similarly, Lsm8-TAP identified 20 splicing proteins including core Lsm proteins and proteins associated with U4 snRNP. Proteins such as Snu13 and Snu114 known to associate with U6 snRNP Lsm proteins (Fabrizio et al., 2009a) were not identified.

In this Chapter the aim was to use TAP MS analysis of Lsm proteins in order to provide more insight into the dynamic composition of the spliceosome at targeted stages of spliceosome assembly. In this chapter, the study focuses on events taking place in the U4/U6 and B complex U4/U6.U5. It also aims at investigating how far down the stages the Lsm proteins go during the splicing cycle. Lastly, this work aims at characterising two similar but distinct Lsm complexes, Lsm1-7p and Lsm2-8p.

4.2 TAP MS Analysis of Lsm 1- and 8-TAP

15 L Yeast cell culture were grown using YPD media at the same temperature and general growth conditions for all Lsm-TAP experiments to exponential phase (see Chapter 2). In 3 experiments (for the sake of simplicity this will be referred to as experiment Lsm1-TAP, Lsm4 and Lsm8-TAP, respectively) the cells were harvested and purified using TAP procedure A (see Chapter 2.4.2) which is a less stringent but effective purification procedure. Lsm4 is a component of both the Lsm1-7 and Lsm2-8p complexes therefore Lsm4-TAP was used as a positive control (see Figure 4.1A and 4.2A). However, the focus of this Chapter will be on Lsm1-TAP and Lsm8-TAP. Tandem affinity purification (TAP) procedure for these three experiments is adapted from and follows the basic principle developed by (Puig et al., 2001) but is mild because it avoided the 10 mM DTT used in the binding and washing buffer of second purification step. This is to avoid losing potential weakly associated but functionally relevant protein components of the purified complex. All Lsm baits were obtained from the Euroscarf Cellzome yeast TAP library. The Lsm baits were tagged with calmodulin binding peptide (CBP) at the C-terminal. The expression of the fusion TAP proteins were confirmed by Western blot (see Figure 4.1, Chapter 2.5.3). A schematic illustration of the work flow is shown in Figure 4.1B. Following TAP, the complexes were analysed using SDS PAGE (see Figure 4.2). In-gel tryptic digests of the purified proteins were performed and the proteins were identified using reversed-phase chromatography interfaced to electrospray ionisation tandem mass spectrometry (LC ESI MS/MS) (see Chapter 2.6). All bait proteins were first identified by MASCOT and then manually analysed with the aid of Bruker data analysis software. All other proteins were identified using MASCOT and Prohits software (see Chapter 2.7). Protein identifications were based on a minimum of at least two unique peptides and a MASCOT peptide ion cut off score of 27. The false discovery rate for Lsm-TAP dataset was adjusted to 1%. With the aid of the Prohits software, proteins that co-purified with the bait were screened against a control TAP done on strain harbouring a non-splicing protein bait-TAP tag. Using a heuristic approach, proteins which are frequently observed in TAP literature, such as ribosomal proteins, heat shock proteins and cytoskeletal proteins were eliminated from the list. Following protein identifications based on the above criteria, literature searches

were used to provide further insight into the biological roles of the identified interacting partners.

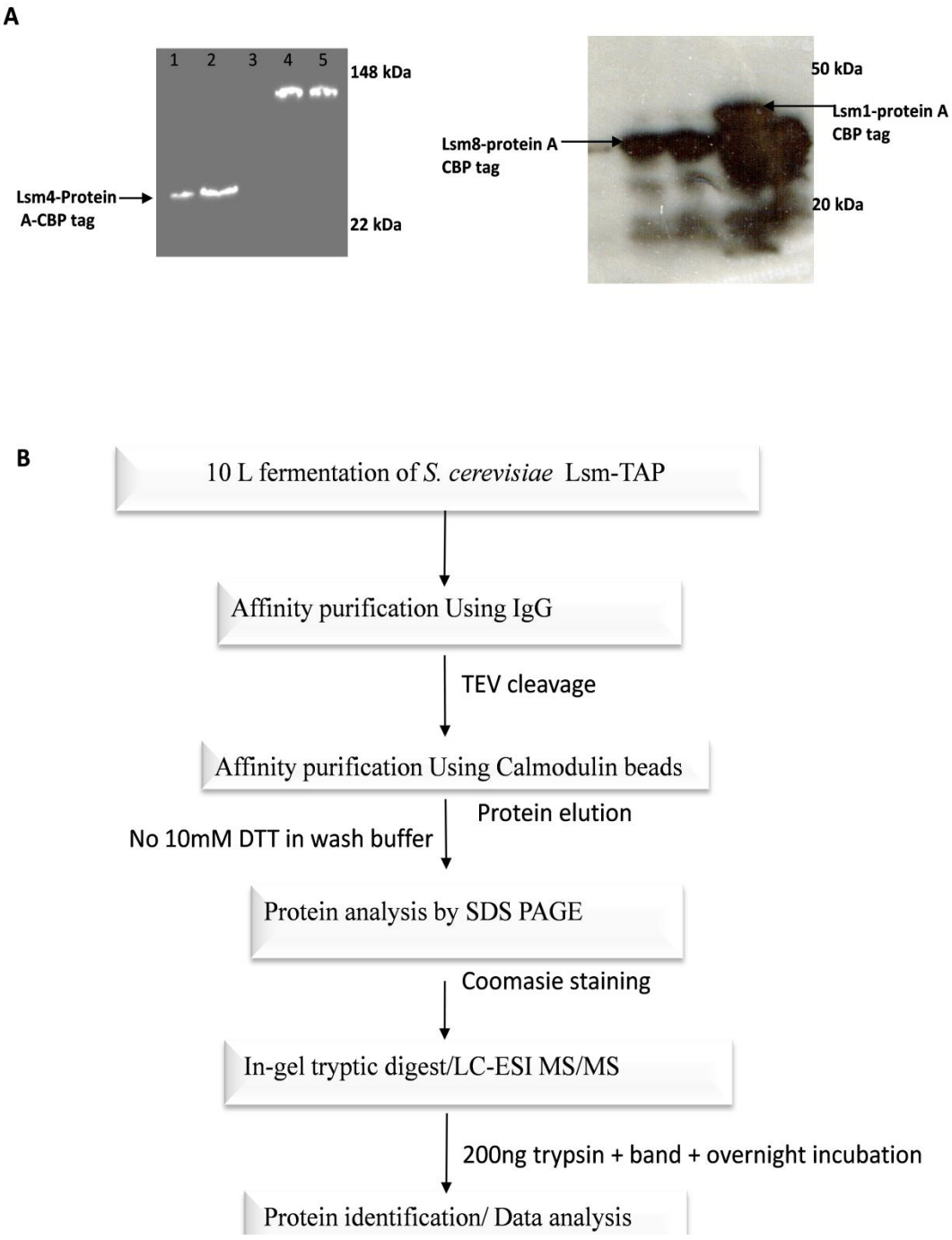


Figure 4.1 Verification of TAP tag expression and TAP work flow for Lsm-TAP A) Western blot verification of Lsm-TAP expression**.** The TAP on C-terminal of Lsm4 was confirmed by Western blot targeting the Protein A epitope. Lane 1-2: Lsm4-TAP, Lane 3: wild type strain as negative control, Lane 4-5: Prp5 as positive control. The other Lsm1 and 8 baits were confirmed in likewise manner. B) Workflow summarising Lsm-TAP-MS analysis**.** All Lsm-TAP experiments were performed under the same culture conditions. In the second purification module, 10 mM DTT was avoided in the wash buffer of the TAP procedure (see Chapter 2.4.3)

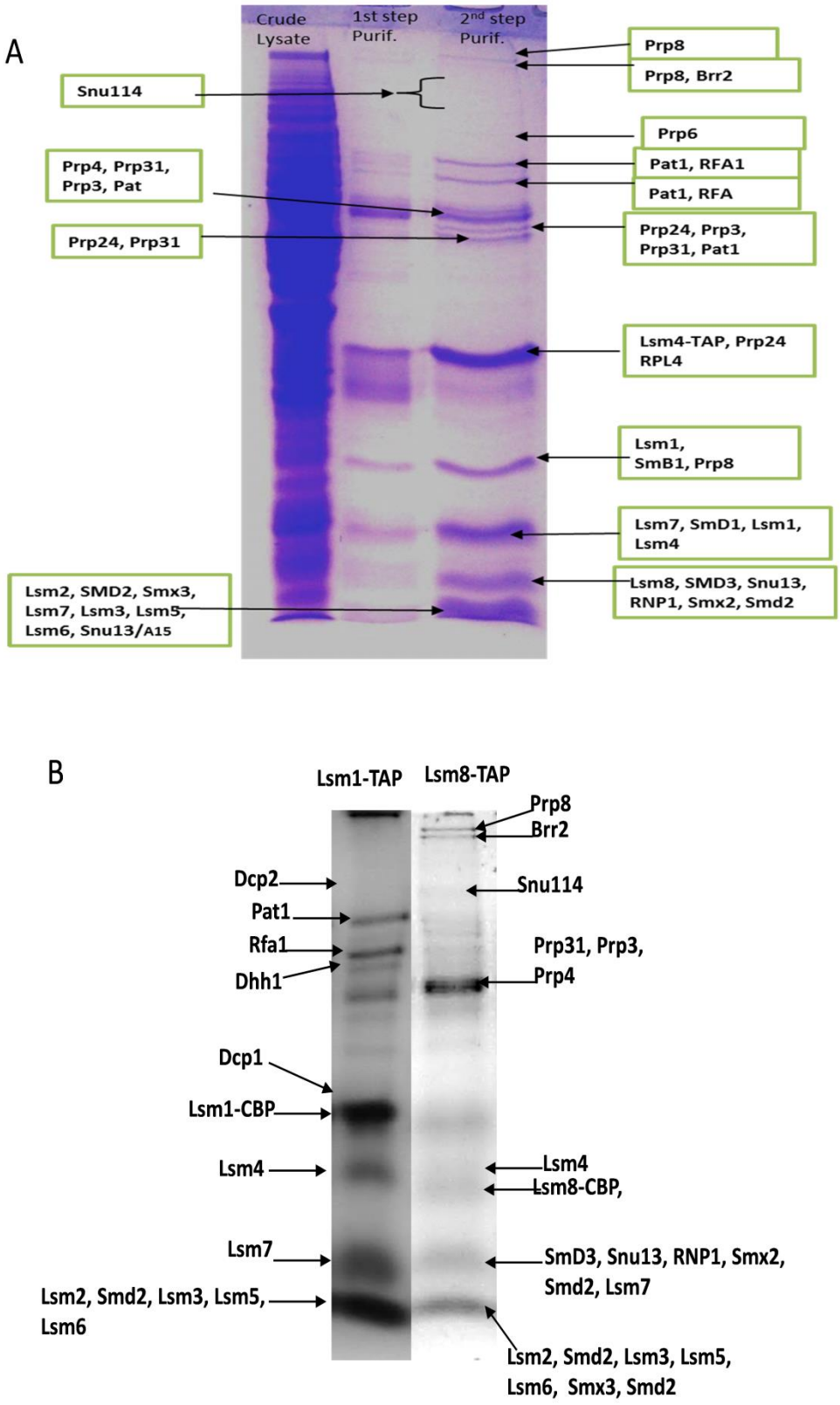


Figure 4.2 TAP MS analysis of Lsm-TAP. A) SDS-PAGE analysis of the proteins eluted from the Lsm4-TAP study. The proteins eluted from both the first (protein A module) and second (Calmodulin Binding Peptide module) purification steps are shown. Proteins identified from the mass spectrometry analysis are highlighted. B) SDS-PAGE analysis of the proteins eluted from the Lsm1/8-TAP study. Proteins identified from the mass spectrometry analysis are highlighted

4.3 Comparative analysis of Lsm TAP MS studies

Comparative analysis of the protein components identified in the TAP MS studies revealed a wide range of proteins that were common across all the Lsm proteins. However significant differences were observed in particular, comparing the Lsm1 and Lsm8 TAP MS studies. Following protein identifications the proteins were grouped according to classifications based on DAVID and KEGG databases (Huang da et al., 2009). Following functional classification a summary of the protein identifications and general functional classification across each of the Lsm TAP MS studies is discussed.

4.3.1 Comparative analysis of Lsm1 and Lsm8 TAP MS studies reveals distinct functional complexes

In the Lsm1-TAP experiment, Lsm1 to Lsm7 but no Lsm8 were identified (Figure 4.3, also see appendix 4.1 for details) which is consistent with the previous evidence that Lsm1-7 is distinct from the U6 snRNA-associating Lsm2-8p complex (Tharun et al., 2000). Also, Lsm1-TAP did not co-purify with Sm proteins and U5 snRNP proteins (see Table 4.2 and 4.3, respectively) which is consistent with the model that Lsm1-7p does not form a stable association with spliceosomal U4/U6 complexes and U5 (U4 and U5 has Sm as core snRNP proteins).

In contrast, the results from the Lsm8-TAP show that the Lsm2-8p complex co-purifies with most of the Sm proteins and U5 snRNP proteins (see Table 4.2 and 4.3, respectively). This provides further evidence that Lsm1-7p is a distinct complex from the Lsm2-8p and therefore, a distinct complex from the spliceosomal U6-associating Lsm complex. Further evidence demonstrating that the Lsm1-7p is a distinct complex from the U6 associating Lsm2-8p complex was shown by the absence of components of the U4/U6.U5 associating factors in the Lsm1-TAP. In contrast, a number of the components of the U4/U6.U5 were identified in the Lsm8-TAP (see Table 4.4 and 4.5).

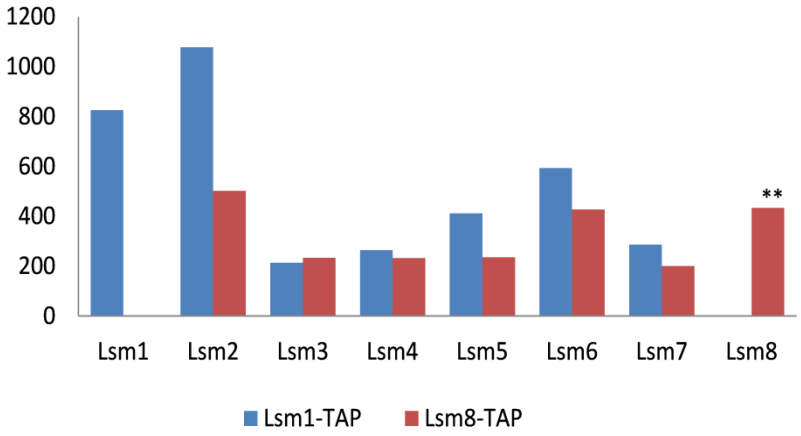


Figure 4.3 A column chart illustrating the Lsm proteins compositions of Lsm1-TAP and Lsm8-TAP bait proteins. The MASCOT scores of identified proteins were used and Lsm1-TAP and Lsm8-TAP are bait proteins. Lsm2- Lsm7 proteins co-purified with both bait proteins. Lsm1-TAP did not co-purify with Lsm8p and Lsm8-TAP did not co-purify with Lsm1p. **Lsm1 was detected with a low MASCOT score but because it was also detected in the blanks preceding Lsm8-TAP analysis it was deemed to be carryover.

	Lsm1-TAP	Lsm8-TAP	Lsm4-TAP
U4/U6 snRNP proteins	Absent*	Present	Present
Lsm2-7p complex	Present	Present	Present
Lsm1	Present	Absent	Present
Lsm8	Absent	Present	Present
Sm protein	Absent**	Present	Present
U4/U6.U5 proteins	Absent	Present	Present
U5 snRNP proteins	Absent	Present	Present
U2 snRNP	Absent	Absent	Absent
TREX Factors	Present	Present	Present
Deadenylation-Decapping complex and associating factors	Present	Absent	Present
H/ACA snoRNP complex	Present	Present	Present
Processome/Box C/D protein	Present	Present	Present
Translation initiation and regulation factors	Present	Present	Present
Chromatin-associating factors	Present	present	Present

Table 4.1 A summary of Lsm TAP-Mass spectrometry analysis. Following SDS PAGE, in-gel tryptic digest was performed in conjunction with HPLC-MSMS to identify proteins. In this summary table, Identification of proteins belonging to any group/type/complex is marked against the group name as 'present' and absent if not identified. Analysis shows that Lsm4-TAP co-purifies with all the group of proteins identified for Lsm1-TAP and Lsm8-TAP. Absent*: although, the majority of U4 proteins were absent in one TAP-mass spectrometry analysis, in another U4 proteins Prp3p and Prp4p were detected with a relatively lower MASCOT score. Absent**: All Sm proteins were absent except for Smd1p which was identified with a low MASCOT score of 31.

Sm proteins	Lsm1-TAP	Lsm8-TAP
SmB or SmB1	-	726
SmD1	31	250
SmD2	-	409
SmD3	-	347
SmE1 or SmX1	-	44
SmF or Smx3 or RuxF	-	83
SmG or Smx2 or RuxG	-	160

Table 4.2 Comparison of Sm proteins associating with Lsm1-TAP and Lsm8-TAP. The data shows that Sm proteins do not interact with Lsm 1-7p (Lsm1-TAP) complex but does with Lsm2-8p (Lsm8-TAP).

U5 snRNP proteins	Lsm1-TAP	Lsm8-TAP
Snu114 or 116K	-	2639
Prp8 or 220K	-	4362
Prp28	-	-
Prp6	-	2540
Brr2	-	3458
Lin1	-	-
Dib1	-	299

Table 4.3 Comparison of U5 proteins associating with Lsm1-TAP and Lsm8-TAP. The data shows that U5 proteins do not interact with Lsm 1-7p (Lsm1-TAP) complex however a number of the protein interact with the Lsm2-8p complex (Lsm8-TAP).

U4/U6 snRNP proteins	Lsm1-TAP	Lsm8-TAP
Prp31	-	1525
Prp3	-	1886
Prp4	-	1260
Prp24	-	528
Snu13	166	205

Table 4.4 Comparison of U4/U6 associating proteins co-purifying with Lsm1-TAP and Lsm8-TAP. The data shows that U4/U6 associating proteins do not interact with Lsm 1-7p (Lsm1-TAP) complex but does with Lsm2-8p (Lsm8-TAP).

U4/U6.U5 snRNP proteins	Lsm1-TAP	Lsm8-TAP
Snu66	-	978
Sad1	-	-
Spp381	-	53
Prp38	-	-
Snu23	-	56

Table 4.5 Comparison of U4/U6.U5 associating proteins co-purifying with Lsm1-TAP and Lsm8-TAP. The data shows that U4/U6.U5 associating proteins do not interact with Lsm 1-7p (Lsm1-TAP) complex in contrast to the Lsm2-8p (Lsm8-TAP).

4.3.2 Analysis of the U4, U6 or U5

Analysis of the protein compositions of Lsm1-TAP and Lsm8-TAP shows clear differences in purification of the Lsm1-7p and the Lsm2-8p complexes. To further analyse the complexes and support the evidence from the difference in protein components from the TAP MS studies, the RNA was extracted from the respective complexes and analysed using denaturing PAGE (see Figure 4.4A). Based on the predicted sizes of the RNAs bound to the Lsm complexes, the results indicate that U4, U6 or U5 snRNAs are bound to the Lsm8-TAP. In contrast the U4, U6 or U5 snRNAs were not observed in the Lsm1-7p complex. A single RNA species was observed in the Lsm1-7p complex which based on the size was not a predicted U4, U6 or U5 snRNA. To verify the snRNAs, Northern Blotting (see Chapter 2.5.5) was used in conjunction with probes specific for U4, U6 or U5 snRNAs (see Figure 4.4B/C). The Northern analysis shows that Lsm8-TAP co-purifies with U4, U6 or U5 snRNAs in contrast to Lsm1-TAP which does not. The results are consistent with the previous proteomic analysis where components of the U4/U6.U5 associating factors were identified in Lsm8-TAP but not Lsm1-TAP.

4.3.3 Mass Spectrometry analysis of Lsm-TAP is consistent with a role for Lsm2-8p in splicing

The Lsm8-TAP MS analysis identified almost all the splicing factors known to associate with the spliceosome U4/U5.U6 tri-snRNP including, Lsm2-8p, Prp3p, Prp4p, Prp31p, Prp24, Snu66p, Snu23p and Spp381p (see Figure 4.3 and Tables 4.1-4.5). This confirms that Lsm2-8p not only interacts with the spliceosome, but plays an important role in splicing. However, none of the factors associated with later stages of splicing (such as U2 snRNP, Prp43p, Prp22p) or the pre-spliceosome (such as U1 and U2 snRNP) co-purified with Lsm2-8p. The absence of the proteins such as, Prp43 and Prp22p (involved in later stages of splicing), is consistent with previous *in vitro* experiment which suggests that Lsm proteins assemble into the spliceosome and dissociate from the spliceosome at some point between B-complex and first splicing step ((Fabrizio et al., 2009a) and reviewed by Will and Lührmann, 2011). This data suggests that Lsm proteins dissociate from the spliceosome at some point between assembly of U4/U5.U6 and the first splicing step in vivo. However, the

Lsm1-TAP did not co-purify with splicing factors, further demonstrating that Lsm1-7p is a distinct complex from Lsm2-8p complex.

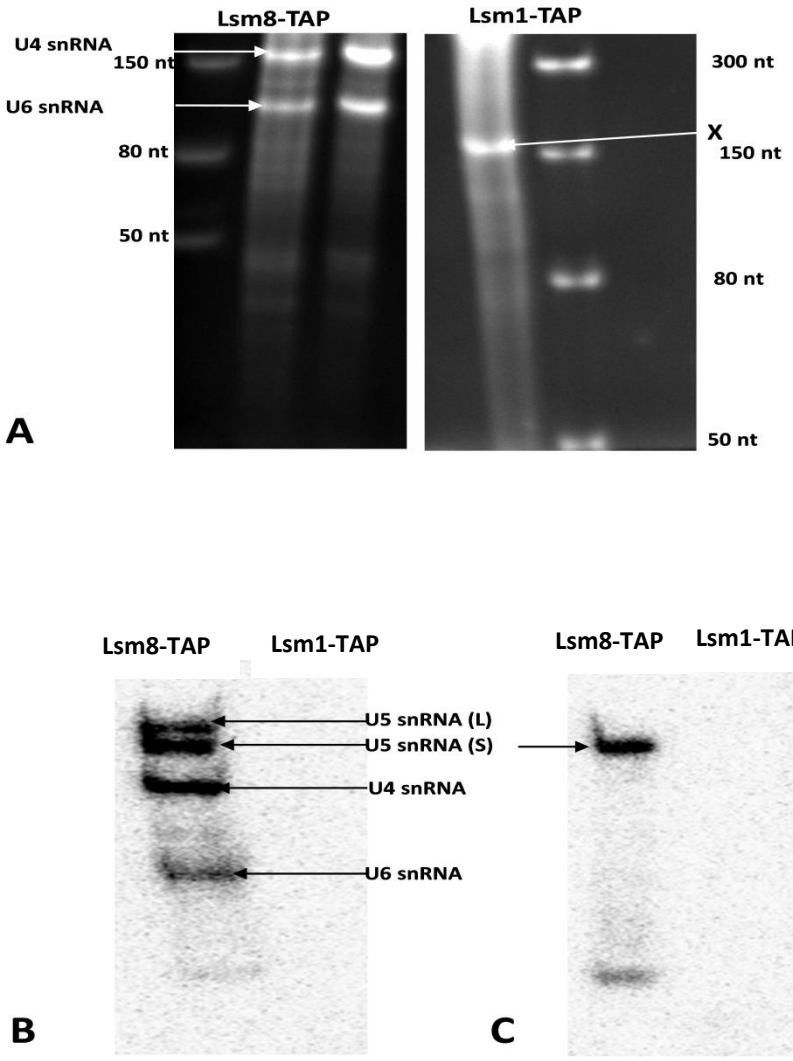


Figure 4.4 RNA binding in the Lsm1-TAP and Lsm8-TAP complexes. A) Denaturing PAGE analysis of the RNA extracted from the Lsm8-TAP (left panel) and Lsm1-TAP (right) panel. Molecular weights of ssRNA markers are highlighted. Based on predicted MWs the U4 and U6 snRNA is prominent in Lsm8-TAP and a single unknown RNA was identified in Lsm1-TAP **B)** Northern blot analysis of Lsm1-TAP and Lsm8-TAP**.** Results confirm **t**hat Lsm8-TAP co-purify with U4, U6, U5 (large) and U5(small) snRNAs. The Lsm1-TAP does not co-purify with any of the snRNAs probed for. **C)** Analyses using the U5 snRNA (S) probe. Lane I and II are for Lsm8-TAP and Lsm1-TAP, respectively.

4.3.4 A novel interaction between Lsm complex and Prp43p

In this study, analyses of Lsm1-TAP –MS data demonstrates a surprisingly novel interaction between Lsm1-7p and Prp43p (see Table 4.7). The analysis of the Lsm8-TAP-MS data shows that Lsm2-8p complex does not interact with Prp43p. The Dead-box RNA helicase Prp43p plays a role in both RNA I and II transcript metabolism as well as catalysing the removal of U2, U5, and U6 snRNPs from the post-splicing lariat-intron ribonucleoprotein complex (Arenas and Abelson, 1997). It is also required for efficient biogenesis of rRNAs (Leeds et al., 2006). Prp43p co-purification with Lsm1-7p complex suggest either that Lsm1-7p plays a role in spliceosome disassembly or that Prp43p plays a role in the recycling of Lsm complexes. However, further studies will be required to understand the role of Prp43p in Lsm1-7p mediated pathway/s or vice versa.

4.4 Roles of Lsm complexes in mRNA degradation pathways

Previous studies have shown that Lsm proteins participate in mRNA decay, specifically the Lsm1-7p complex (Tharun et al., 2000). The major pathway of mRNA decay is the deadenylation-dependent pathway. In this pathway, the poly(A) tail is removed by a deadenylase activity followed by either 5'→3' or 3'→5' decay (Garneau et al., 2007). The Lsm1-7p complex plays a role in this pathway by associating with the 3’ end of the mRNA transcript, inducing decapping by the DCP1-DCP2 complex (Chowdhury et al., 2007; Tharun et al., 2000). This leaves the mRNA susceptible to decay by the 5'→3' exoribonuclease Xrn1p. In an alternative pathway, the deadenylated mRNA can be degraded in the 3'→5' direction by the exosome, followed by the hydrolysis of the remaining cap structure by Dcps(Garneau et al., 2007). In *Saccharomyces cerevisiae*, de-adenylation-independent pathway requires recruitment of the decapping machinery. In this pathway, Rps28B interacts with enhancer of decapping Edc3p, to engage the decapping activity and then mRNA degradation by Xrn1p (Conti and Izaurralde, 2005; Garneau et al., 2007; Yamashita et al., 2005). Although, it is now increasingly evident that Lsm1-7p plays a role in deadenylation-decapping dependent mRNA degradation, it is not clear if it is involved in other mRNA degradation pathways. Furthermore, it is not clear if Lsm2-8p participate in mRNA degradation. In a previous study it was shown that Lsm2-8p

may be distinct from Lsm1-7p complex it did not, however, determine if Lsm2-8p also participates in mRNA degradation pathways. This is because the Western blot and immunofluorescence experiments used in previous studies to distinguish these complexes were limited to the number of antibodies and fluorescent peptide tags used (Tharun et al., 2000). Mass spectrometry on the hand has no such limitations.

4.4.1 Deadenylation-dependent and Decapping pathway of mRNA Degradation

In order to uncover other potential Lsm protein-protein interactions with mRNA degradation proteins which may not have been identified in previous studies, data from the Lsm1-TAP and Lsm8-TAP MS analyses were analysed and compared. The data shows that Lsm1-TAP co-purifies with deadenylase and decapping factors, including Dcp1, Dcp2, Edc1p, Pat1p, Dhh1p, deadenylation factor, Sup35p and the exoribonuclease, Kem1p (Xrn1p) (see Table 4.6). On the other hand, Lsm8-TAP did not co-purify with these deadenylase-decapping complex factors. However, Lsm8-TAP co-purified with the exoribonuclease, Kem1p and Eap1p (see Table 4.6). Eap1p is a factor that accelerates mRNA degradation by promoting decapping (Blewett and Goldstrohm, 2012).

Dcp1p in association with Dcp2p forms the decapping enzyme complex that removes the 5' Cap structure of mRNA prior to their degradation (Beelman and Parker, 1995, LaGrandeur and Parker, 1998, She et al., 2006) and this activity is enhanced by the RNA-binding protein, Edc3p, which activates mRNA decapping directly (Dunckley et al., 2001, Schwartz et al., 2003). The decapping activity is regulated by Dhh1p (Fischer and Weis, 2002). Pat1p is a deadenylation-dependent mRNA decapping factor and plays a role in mRNA decay (Bonnerot et al., 2000). Kem1p (Xrn1p) is a 5' to 3' exoribonuclease and a component of the cytoplasmic processing (P) bodies involved in mRNA decay (Sheth and Parker, 2003).

The association of the decapping factors, Dcp1-Dcp2/Dhh1p and the deadenylation-dependent mRNA decapping factors with Lsm1-TAP demonstrates that Lsm1-7p interacts with the Deadenylase-decapping complex. The absence of these proteins in Lsm8-TAP shows not only that Lsm1-8p is a distinct complex from Lsm1-7p but also

that Lsm2-8p is not involved in deadenylation-dependent mRNA decapping/degradation pathway. However, the co-purification of Lsm8-TAP with Kem1p suggests that it may play some role in mRNA decay (see Table 4.6). The co-purification of the decapping factor, Eap1p, with Lsm8-TAP but not Lsm1-TAP further supports the notion that Lsm8-TAP is involved in an mRNA decay pathway which is distinct from the mRNA decapping/decay pathway mediated from Lsm1-7p/ Dcp1-Dcp2p.

Other factors implicated in mRNA translational repression and mRNA degradation, Viz, Sbp1p and Ssd1p were detected in both Lsm1-TAP and Lsm8-TAP (see Table 4.6). Sbp1p plays a role in repression of translation, affects decapping in *S. cerevisiae* and is found in the cytoplasmic P bodies (Segal et al., 2006). Ssd1p plays a variety of roles in the cell, including, translational repression, definition of bound mRNA's destination by its nucleocytoplasmic shuttling activity and integration of cell fate with translational (Ohyama et al., 2010, Kurischko et al., 2011, Jansen et al., 2009).It is significant that Lsm1, Pat1 and Dhh1 localise in the nucleus (Tharun et al., 2000). The co-purification of Lsm1-7p with nucleocytoplasmic shuttling, Ssd1p is consistent with the idea that Lsm1-7p shuttles between the nucleus and the cytoplasm, playing a role in cytoplasmic mRNA degradation (in P-bodies) and the nucleus (Tharun et al., 2000).

In conclusion, the results demonstrate that Lsm1-7p and Lsm2-8p are distinct complexes and are involved in different mRNA degradation pathways. The results show that while Lsm1-7p is involved in deadenylation-decapping mediated mRNA decay, Lsm2-8p is not. Both complexes associated with other mRNA decay factors, such as Kem1p (Xrn1p), Sbp1p and Ssd1p. However, the results demonstrate a novel interaction between Lsm2-8p and the decapping factor, Edc1p. Conversely, the decapping enzyme complex, Dcp1-Dcp2p, interacts with Lsm1-7p but not with Lsm2-8p. This suggests that Lsm2-8p role in mRNA degradation may require the decapping activity of Eap1p in contrast with the Lsm1-7p complex that requires the decapping activity of Dcp1-Dcp2p enzyme complex. This is the first time that the Lsm2-8p is shown to interact with Kem1p (Xrn1p), Eap1p and the other mRNA decay factors suggesting a role for Lsm2-8p in mRNA degradation.

Deadenylase-Decapping complex and Cytoplasmic P-bodies proteins	Lsm1-TAP	Lsm8-TAP
Dcp1	281	-
DCP2	698	-
EDC3	207	-
Pat1	2379	-
Dhh1	2439	-
Sup35 or Erf3	516	-
SSD1	57	154
Sbp1 or SSBP1	924	865
EAP1	94	35
Xrn1 or Kem1	2409	332
XRN2 or RAT1	-	-
RNP1	-	-

Table 4.6 Comparison of Lsm-TAPs interacting with mRNA decay proteins.

Translation initiation and regulation factors	Lsm1-TAP	Lsm8-TAP
PAB1	1754	1921
Tif4631	1261	277
Tif4632	421	88
Tif45 /CDC33	164	252
Tif1	-	111
Lsm12	-	-
Tif31 or Clu1	2425	1666
Sup45	113	-
Sup35	516	-
Nip1 EIF3C	49	28
Fun12	1199	810
Pbp1	199	88
Pbp4	85	36

Table 4.8 Comparison of translation regulation factors associating with Lsm-TAP

Transcription-Export (TREX) complex and Poly(A)+ mRNA/RNA binding proteins	Lsm1-TAP	Lsm8-TAP
MEX67	270	-
YRA1	368	431
HRB1	49	57
GBP2	476	68
NOP3 or NPL3	665	334
Hpr1	34*	-
DBP2	1196	292
Sto1	115	-
Hrp1	311	101
Pub1	107	-
Sub2	56*	-
Bfr1	-	703
Sro9	802	148

Table 4.9 Comparison of TREX factors interacting with baited Lsm proteins. The bait proteins are Lsm1- and Lsm8-TAP. This suggests that Lsm proteins associate with TREX complexes but Lsm1 associate with more TREX components than Lsm8.
*Identified with single peptide hit.

Nonsense-mediated mRNA Decay (NMD) pathway	Lsm1-TAP	Lsm8-TAP
Nam7 or Upf1	110	-
UPF3	-	-
Dead-Box proteins	-	-
Ded1	2113	1326
DBP9	-	2830
DBP1	-	231
Dbp3	212	-
Prp43	202	-

Table 4.7 Comparison of Lsm-TAPs interacting with Nonsense-mediate mRNA decay and Dead-box proteins.

4.4.2 Role of Lsm proteins in Nonsense-mediated mRNA degradation

Nonsense mediated pathway is a surveillance and quality control mechanism that selectively degrades mRNAs harbouring premature termination (nonsense) codons (Leeds et al., 1992, Cui et al., 1995, Holbrook et al., 2004). If translated these mRNAs can produce truncated proteins with dominant-negative or deleterious effect in the cell (Holbrook et al., 2004, Chang et al., 2007). In Mammalian cells, UPF1 and UPF2 are part of a large complex of proteins deposited at the exon-exon junction during splicing (Chang et al., 2007). Although, the human Upf1p is a cytoplasmic protein involved in translational termination, evidence from previous studies shows that it shuttles between the nucleus and the cytoplasm suggesting that nuclear NMD is a possibility (Wilkinson and Shyu, 2002, Mendell et al., 2002).

In this study, proteins of the nonsense-mediated mRNA degradation (NMD) pathway such as Nam7p protein co-purified Lsm1-TAP but not Lsm8-TAP (see Table 4.7). The co-purification of NMD factors with Lsm1-TAP but not with Lsm8-TAP suggests that Lsm1-7p interacts with and plays a role in NMD. This is consistent with previous evidence suggesting that Lsm1-7p complex interacts with the NMD factor, Nam7p (UPF1) (Gavin et al., 2006). The result is consistent with a notion that Lsm1-7p interacts with NMD factors, targeting mRNA harbouring nonsense codons for degradation.

4.5 Interaction of Lsm proteins with mRNA capping factors and Translation regulation factors

Previous studies suggest that the decapping-activating Lsm1-7p complex interacts with deadenylated mRNAs when the elF4G (Tif4631) or elF4E (Cdc33) are no longer bound. (Tharun and Parker, 2001). In contrast, our data shows that Tif4631p (elF4G1), Tif4632p (elF4G2), Cdc33 elF4E, Pab1p, mRNA deadenylation-decay factor, Sup35p (erF3) and polypeptide release/ translation termination factor, Sup45p (Erf3) co-purifies with the Lsm1-TAP (see Table 4.8). This suggests that Lsm1-7p complex interacts with elF4G (Tif4631) and elF4E (Tif45). This will mean that the cytoplasmic

Lsm1-7p complex interacts with the translational initiation factors and possibly when they are bound to the mRNA. Whether the interaction between Lsm1-7p complex and these mRNA cap-binding proteins are in the cytoplasm or in the nucleus cannot be directly inferred from the data. However, it is of interest that nuclear cap-binding protein; Sto1p and TREX factor, Npl3p co-purified with Lsm1-7p complex (see Table 4.9). This is interesting because Sto1p has been reported to interact with Npl3p in the transport of mRNA to the nucleus and has been implicated in nuclear mRNA degradation (Das et al., 2003; Das et al., 2000a; Fortes et al., 1999; Gao et al., 2005). It is also interesting that Lsm8-TAP also co-purified with these translation initiation factors (see Table 4.8). This suggests that Lsm2-8p may play a role in the regulation of translation from the nucleus possibly by facilitating transcript/mRNA degradation in the nucleus.

4.6 Interaction of Lsm proteins with THO-Transcription-Export (TREX) mediated processes

The compartmentalization of the transcription and processing of RNAs and protein translation makes the export of various RNAs a crucial step in gene expression. Most RNAs, such as tRNAs, rRNAs, snRNAs and microRNAs, are transported via importing/Karyopherin-β family proteins while poly(A)+ mRNAs are exported by a non-importin –β type heterodimeric receptor Mex67-Mtr2 (Segref et al., 1997, Santos-Rosa et al., 1998, Katahira and Yoneda, 2009). To further study which of the Lsm proteins/complexes these TREX factors associate with, we analysed the mass spectrometry data from Lsm1-TAP and Lsm8-TAP. The results shows that both Lsm1-7p complex and Lsm2-8p co-purifies with the TREX factors, were found in the Lsm1-7p complex) (see Table 4.9). It is interesting that in this study, Lsm1-7p but not Lsm2-8p co-purified with the THO/TREX component, Hpr1p, and Sto1p, yeast ortholog of Cbp80. Hpr1p-THO/TREX is involved in the coupling of transcription elongation with mitotic recombination and with mRNA metabolism and export (Chavez et al., 2000; Jimeno et al., 2002), and Sto1p is implicated in nuclear mRNA degradation and transport of nuclear poly(A)+ mRNA to cytoplasm (Das et al., 2003; Das et al., 2000a; Gao et al., 2005). This is particularly significant since Lsm1-7p co-

purified with Npl3p in this experiment (see Table 4.9) which has been implicated in the transport of nuclear mRNA transport processes involving Sto1p (Das et al., 2003; Das et al., 2000a) and Hpr1p-Tho/TREX (Jimeno et al., 2002; Lee et al., 1996a). It is also interesting that Npl3p has been involved in a variety of roles such as acting as an alternative adaptor for mRNA export (Windgassen et al., 2004, Katahira and Yoneda, 2009), transcription, elongation and nuclear export of mRNAs (Strasser et al., 2002, Windgassen et al., 2004).

In addition, factors involved in transcription including RNA polymerase subunits, Rpc40p, Rpa49p, Rpa135p, and other transcription factors such as Hta1p, Hta2p co-purified with Lsm1-TAP and Lsm8-TAP (see Table 4.10). Other transcription associating factors such as RNA polymerase II degradation factor, Def1p and DNA replication and repair factors such as Rfa1p, Rfa2p, Rfa3p co-purified with Lsm1-TAP. This suggests that Lsm proteins form heterogeneous multicomplex with transcription/DNA repair proteins and TREX factors.

Since Lsm1-7p co-purified with all these factors implicated in mRNA processing in this study, the data is consistent with a model where Lsm1-7p forms a multi-complex with the THO/TREX to mediate transcription, nuclear mRNA transport and degradation. The co-purification of these factors with Lsm2-8p complex support a model where some TREX factors are recruited to the pre-mRNA during transcription and stays on during splicing thereby interacting with splicing factors, Lsm2-8p. The absence of Sto1p and Hpr1p from the Lsm8-TAP in this study (see Table 4.9) highlights some of the compositional difference between Lsm2-8p and Lsm1-7p. The data is consistent with the notion that Lsm1-7p is derived from Lsm2-8p by displacement of Lsm8 by Lsm1p and recruitment of more TREX factors such as Sto1p and Hpr1p required for export of mRNA to the nucleus after splicing.

4.7 Lsm proteins co-purifies with processosome proteins and Box C/D proteins

Analysis of the Lsm1-TAP and Lsm8-TAP shows that the C/D box proteins co-purify with Lsm1-7p and Lsm2-8p complexes (see Table 4.11). The fact that Snu13p, Nop1p

and Nop56p co-purified with the two bait proteins is significant. This is because Snu13p along with Nop56p, Nop58p and Nop1p are members of box C/D type sRNPs and are involved in 2'-O-ribose methylation of RNAs, including pre-rRNA, during their maturation (Lafontaine and Tollervey, 2000; Watkins et al., 2000; Zhou et al., 2002).

Furthermore, the co-purification of the nuclear-localizing sequence-binding/ nucleolar protein, Nsr1p with Lsm1-TAP provides further evidence that Lsm complex interacts with these nucleolar proteins. The co-purification of snoRNP protein with Lsm2-8p (Lsm8-TAP is the bait) suggests that snoRNP plays a role in the maturation of U6 snRNP or that Lsm2-8p play a role in snoRNP assembly. Previous work has demonstrated that U6 snRNA may be pseudouridylated in the nucleolus and that it localizes in the nucleolus (Lange and Gerbi, 2000; Sleeman and Lamond, 1999). In fact, U4, U5 and U6 snRNAs have all been observed in the nucleolus (Gerbi and Lange, 2002). The co-purification of U6 snRNP Lsm2-8p with Box C/D complex and Nsr1p (see table 4.11) provides further evidence for the transient localisation of U6 snRNP in the nucleolus (Gerbi and Lange, 2002).

It has been reported that some snoRNPs catalyse the 2'-O-ribose methylation of certain nucleotides of U6 snRNA in *Schizosaccharomyces pombe* and higher eukaryotes (Tycowski et al., 1998; Zhou et al., 2002). It is interesting that Nop1p and other components of the Box C/D particularly direct the 2'-O-ribose- methylation of rRNA. These proteins may also methylate snRNAs since they co-purified with splicing factors in this experiment. The notion that Box C/D may methylate snRNAs are corroborated by the evidence U4/U6 snRNP-associating protein, Snu13p, interacts with these Box C/D proteins, and along with the components of this snoRNP, form a protein core that pseudouridylates rRNAs (Dobbyn and O'Keefe, 2004). Previous studies suggest 2'O-ribose methylation and pseudouridylation of U6 snRNA is directed by nucleolar factors (Ganot et al., 1999). U2 snRNA is known to be either extensively pseudouridylated or 2-O-ribose methylated which can be RNA-guided or RNA-independent (Ma et al., 2005; Ma et al., 2003). The co-purification of the snoRNPs with protein-RNA components found in U4/U6.U5 suggests that it is required for snRNA modification, especially 2-O-ribose methylation of U6 snRNA and U2 snRNA.

The co-purification of Lsm1-TAP with C/D snoRNP proteins is particularly significant since Lsm1-7p complex does not interact with U4, U6 and U5 snRNAs. This, therefore, indicates that the Lsm1-7p complex is not involved in the processing of snRNAs. However, the co-purification of Lsm1-7p with C/D snoRNPs and Nsr1p would be consistent with Lsm1-7p complex playing a role in the processing and maturation of snoRNPs.

Transcription/Chromatin associating factors	Lsm1-TAP	Lsm8-TAP
Htb1, Htb2	265	220
Hho1	512	107
Hta1	44	57
Rfa2	43	-
RFA1	173	-
RFA3	133	-
TFG2	30	-
Rpc40	335	136
Rpa49	371	61
Rpa135	327	63
Def1	60	-

Table 4.10 Transcription/Chromatin associating factors co-purifying with baited Lsm proteins.

Processosome proteins and Box C/D proteins involved in snRNA and 18S rRNA processing	Lsm 1-TAP	Lsm8-TAP
Snu13+	166	205
Nop1	968	410
Nsr1	194	974
NOP56	983	94
Rrp8	-	-
Rrp9	-	-
UTP10	-	-
Kre33 UPF0202 protein	-	-
Nop2	340	97
NOP14	-	-
NOP12	-	-
Nop4	216	37
IPI1	-	-
Mrt4	189	92

Table 4.11 Comparison of Box C/D and processosome proteins co purifying with Lsm-TAPs

Yeast Box H/ACA snoRNP complex / 18S rRNA processing	Lsm1-TAP	Lsm8-TAP
NHP2	105	48
NOP10	143	-
CBF5	200	-
GAR1	176	69

Figure 4.12 Comparison of Box H/ACA proteins co-purifying with Lsm-TAPs

4.8 Yeast Box H/ACA and C/D type snoRNP complex

Nhp2p, Nop10p, Gar1p and Cbf5p are components of the H/ACA-type small nucleolar ribonucleoprotein (H/ACA snoRNP) complex involved in processing of diverse RNAs and RNA modifications such as isomerization of specific uridines into pseudouridine, a process referred to as pseudouridylation (Filipowicz and Pogacic, 2002; Henras et al., 2004; Weinstein and Steitz, 1999). Cbf5 is the putative psedouridine synthase of the H/ACA-type complex (Lafontaine et al., 1998; Lafontaine and Tollervey, 2000) and forms complex with Nop10p, Gar1p and Nhp1 (Henras et al., 2004). It has been suggested that the complex can be formed with or without Nhp1 (Henras et al., 2004). Both large and small rRNAs, as well as U2 snRNA, are pseudouridylated by Cbf5p (Ma et al., 2005).

Although there is evidence that many snoRNAs are encoded within pre-mRNA introns and are released by exonuclease digestion of the excised intron following splicing or during degradation of the unspliced pre-mRNA (Allmang et al., 1999; van Hoof et al., 2000; Villa et al., 1998), it is not clear if and how splicing and snoRNP synthesis and processing is coupled. There is also no report of an interaction between Lsm proteins and the snoRNPs.

The Lsm1-TAP co-purified with the entire yeast box H/ACA snoRNP proteins and some box C/D snoRNP while only two H/ACA box proteins were identified in Lsm8-TAP with relatively lower MASCOT scores (see Table 4.12). Since snoRNPs localised in the nucleolus it is reasonable to suggest that these snoRNPs associate with nuclear Lsm1-7p complex and Lsm2-8p. These results suggest that either Lsm proteins require some snoRNPs for certain roles in RNA processing such as splicing or that Lsm proteins are required for the processing of snoRNAs or both processes may take place in the cell. There is some evidence that Lsm proteins bind and are required for the processing of snoRNAs (FeRNAndez et al., 2004; Tomasevic and Peculis, 2002). Analysis of the Lsm1-TAP in yeast by denaturing RNA PAGE suggest that Lsm1-7p co-purify with, at least, one small RNA (see Figure 4.4). Further analysis by northern blot to see if this small RNA is one of the snRNAs showed that this is not the case (see Figure 4.4B and C). However, the Northern analysis identified U4, U5 and U6 snRNAs for the Lsm8-TAP (see Figure 4.4B and C). This shows that Lsm1-7p

co-purifies and enriches for a small RNA which may be a snoRNA, however further analysis is required to identify and characterize the RNA.

SnoRNPs have also been implicated in splicing and processing of snRNAs, for instance, in an experiment involving mutation of the H/ACA pseudouridine synthases, Cbf5p or Nhp2p deletion, it was shown the pseudouridylation of U2 snRNA can be catalysed by an RNA-guided or RNA independent mechanism (Ma et al., 2005). It has also been shown that all the U5 snRNA have modifications that include, 5'cap, pseudouridylation and methylation sites (Ganot et al., 1999; Ma et al., 2005). The co-purification of some Lsm2-8p complex with some H/ACA and C/D proteins will be consistent with the notion that snoRNP proteins are required in processing of snRNAs.

In conclusion, this study shows that snoRNP protein components associate with Lsm proteins. It supports the model that they are involved in the modification and processing of snRNAs and snoRNAs (FeRNAndez et al., 2004; Kufel et al., 2003; Kufel et al., 2002).

4.9 Lsm proteins interact with Npl3p

MS analysis of the Lsm1 and Lsm8-TAP also shows that Npl3p interacts with Lsm2-8p and Lsm1-7p complexes. The Sr-like protein, Npl3p, is implicated in the transport and localisation of mRNA in the cytoplasm ((Lee et al., 1996b) and regulation of translation. A previous study (McBride et al., 2005b) reported that methylation of Npl3p facilitates its nuclear exit to the cytoplasm by weakening nuclear protein-protein interactions. Further analysis of the MS data was used to verify sites of arginine methylation of Npl3p. The result shows that both Lsm1-TAP and the nuclear restricted Lsm8-TAP associates with methylated Npl3p. This study is also consistent with the findings in Chapter 3 (see section 3.2.3) where Prp5, involved in splicing, co-purified with methylated Npl3p.

The MS analysis identified a range of methylated peptides in Npl3p in both Lsm1-TAP and Lsm8-TAP (see Table 4.13). All tandem MS of the modified peptides was manually verified and are shown in Appendix 4.2. Furthermore, both Lsm1-7p and

Lsm2-8p co-purified Hrp1p (see Table 4.9). Hrp1p is involved in the TREX/hnRNP complex and its nuclear exit to the cytoplasm has been shown to be dependent on methylation of Npl3p (Xu and Henry, 2004). These results provide further evidence for the potential role of Lsm complexes in nuclear transport of mRNP. Furthermore the pattern of methylation at RGG sites of Npl3p, rather than its mere methylation, may be responsible for discrimination of mRNP nuclear retention signal from the mRNP export signal.

Calculated Mass (Da)	Experimental mass (Da)	MASCOT Score Lsm8-TAP	MASCOT Score Lsm1-TAP	Sequence
1210.5832	1210.5843	35		R.GGYSRGGYGGPR.N + Dimethyl (R)
1210.5834	1210.5843	(16)	49	R.GGYSRGGYGGPR.N + Dimethyl (R)
1398.6292	1398.6277	(16)	32	R.GSYGGSRGGYDGPR.G + Methyl (R)
1412.6452	1412.6433	23	13	R.GSYGGSRGGYDGPR.G + Dimethyl (R)

Table 4.13: Arginine methylated peptides identified for Npl3 by mass spectrometry analysis. The methylated arginines are underlined and type of methylation indicated after the peptide sequence. Details of tandem MS analyses of methylated Np3p peptides are shown in Appendix B1 to B4.

Conclusion

The Lsm-TAP MS studies show that the Lsm1-7p complex is distinct from Lsm2-8p complex because of mainly three reasons. First, Lsm2-8p co-purifies with U4/U6.U5

complexes but not Lsm1-7p complexes. Secondly, Lsm2-8p associates with splicing factors while Lsm1-7p does not. Thirdly, Lsm2-8p does not co-purify with proteins associated with cytoplasmic P-bodies such as Pat1/Dhh1p/Dcp1/2 proteins while data from Lsm1-TAP is consistent with the model that Lsm1-7p interacts with the cytoplasmic bodies and shuttles between the nucleus and the cytoplasm. However, both complexes interact with transcription factors, TREX factors, snoRNPs and mRNA degradation factors suggesting the complex plays a role in transcription, splicing and mRNA transport. One model proposed is that both complexes, although functionally distinct, may share the same evolutionary origin if not derived from the same parent particle. An alternative model proposed is that the Lsm1-7p complex may be derived from recycled Lsm2-8p, that is, by dissociating Lsm8 from the complex and incorporating Lsm1. Lsm proteins, however, do not interact with other splicing factors found in the early spliceosome or the core catalytic unit that facilitates the first and second splicing steps.

The study also provides insight into the role of Lsm1-7p and Lsm2-8p complexes in distinct mRNA decay pathways. It is of particular interest that Lsm1-TAP co-purifies with two important factors that mediate deadenylation-independent mRNA, namely, the decapping enhancing factor, Edc3p and the exoribonuclease, Kem1p (Xrn1p). Furthermore, Lsm1-7p co-purified with deadenylation-decapping complex, such as Dcp1-Dcp2p, Pat1p, Dhh1p and other proteins involved in mRNA degradation. This suggests that Lsm1-7p mediates both deadenylation-dependent and independent mRNA decay pathway Lsm8-TAP, other hand, co-purified with the decapping promoting factor, Eap1p and the exoribonuclease, Xrn1p suggesting. Altogether, the result suggests that while Lsm1-7p may be involved in deadenylation-dependent and deadenylation-independent pathways, Lsm2-8p complex may be involved in a distinct mRNA decay pathway utilizing Eap1p as a decapping enhancing factor. It is also interesting that Lsm1-7p co-purified with nonsense-mediate decay pathways (NMD) while Lsm2-8p did not, which suggests versatility of the Lsm1-7p complex in the mRNA degradation pathways.

The co-purification of Lsm proteins with TREX factors suggests that these proteins are present in the TREX/hnRNP complex and may play roles in transcription, mRNA splicing and export of mRNA. The co-purification of H/ACA and CD Box proteins

suggests that Lsm proteins may a play role in snoRNAs processing and vice versa. It is particularly significant that Lsm2-8p and Lsm1-7p both interact with the H/ACA and CD Box proteins. This study also revealed other novel interactions between Lsm proteins and other proteins for example, Npl3 and Nsr1p which will be interesting to investigate further.

Chapter V

Analysis of the Hsh155/U2 snRNP complex using TAP-MS studies

Abstract

Hsh155 is a U2 snRNP associated splicing factor which forms extensive associations with the branch site-3' splice site-3' exon region during pre-spliceosome formation. It is similar to the mammalian U2 snRNP-associated splicing factor, SAP155. In this Chapter, recombinant Hsh155-TAP *Saccharomyces cerevisiae* in conjunction with TAP-MS was used to identify novel interactions between U2 snRNP and Prp19-NTC/NTC-related proteins. The TAP- MS analysis identified the core U2 proteins, Rse1p, Cus1p, Hsh155p, and Prp9p. The U2 proteins, Prp21p, Prp11p, Msl1p, Rds3p and Ysf3p usually observed in the U2 mono-snRNP complex and early pre-spliceosome stage were not identified, indicating that the purified complex has undergone considerable remodelling. The analyses suggest that other NTC/NTC-related proteins identified were sub-stoichiometric interacting partners. The Hsh155-TAP experiments suggest that the NTC/NTC-related proteins associate with U2 snRNP complex. Also factors implicated in transcription, including Spt5p, mRNA export (TREX) and DNA repair proteins co-purified with Hsh155-TAP. Altogether, the results are consistent with the notion that splicing and mRNA export are coupled to transcription.

5.1 Introduction

The eukaryotic pre-mRNA is littered with numerous false splice sites that the spliceosome must distinguish quite early during the spliceosome assembly (Will and Lührmann, 2011). Proteins contribute directly to splicing by contacting several pre-mRNA species and indirectly by interacting with snRNA to promote splicing reactions (Staknis and Reed, 1994; Will and Lührmann, 2011). There are several weak protein/RNA binary interactions which alone are not sufficient to stably identify and carry out splicing. However, the existence of multiple weak binary interactions ensures that the spliceosome completes the major challenge of not only identifying the splice sites, but catalysing high precision splicing while maintaining the flexibility needed for the splicing machinery to regulate the splicing events (Will and Lührmann, 2011).

Hsh155 is a U2 snRNP associated splicing factor which forms extensive associations with the branch site-3' splice site-3' exon region during pre-spliceosome formation. It is similar to the mammalian U2 snRNP-associated splicing factor, SAP155 (McPheeters and Muhlenkamp, 2003; Pauling et al., 2000; Stevens et al., 2002). The U2 snRNP comprises of a heptameric Sm complex, U2A'-U2B', SF3a and SF3b sub-complexes. The SF3b sub-complex comprises of the Hsh155, Cus1p and Hsh49p (Pauling et al., 2000; Wells et al., 1996).

In order to discuss U2 snRNP and its role in the spliceosome it is important to briefly revisit the spliceosome and its mechanism of assembly. The spliceosome is a highly dynamic RNA-protein complex comprising five snRNPs, U1, U2, U4, U5 and U6 (Lerner et al., 1980; Nilsen, 1994; Raker et al., 1996). Each snRNP contains a unique snRNA and spliceosome associating factors. With the exception of U6 snRNP, these snRNPs contain a common set of Sm proteins (B/B', D3, D2, D1, E, F, and G)(Raker et al., 1996). U6 snRNP contains Lsm-proteins (i.e., Sm-related proteins) and the protein, Prp24p thought to associate with U6 to form an intermediate that facilitates association of U6 with U4 (Stevens et al., 2001). In the earliest stage of spliceosomal complex assembly, known as Commitment Complex 1 or early complex in yeast, the U1 snRNP associates with the pre-mRNA and U1 snRNA pairs with the 5' exon at the

positions -1 and -2 in both yeast and metazoan spliceosome (Newman et al., 1995; Wyatt et al., 1992). This is followed by U1 snRNP protein Prp40 contacting the branch point bridging protein (BBP) at branch point (BP) or site (BS) which in turn contacts Mud2 at the 3'-splice site effectively bridging the intron ends (Abovich and Rosbash, 1997) to form the commitment complex 1 (CC1) known in mammals as Complex E. The commitment complex incorporates U2 snRNP in an ATP-independent step; in fact, U2 snRNP may be required to form this complex but the stable association of the U2 to the pre-mRNA Branch Point (BP) to form the pre-spliceosome is ATP-dependent(Das et al., 2000b) . Recent evidence shows that the yeast Prp5 and Sub2/UAP56 play important roles in the association of U2 with the Branch Site (BS) to form the pre-spliceosome; it is thought that Sub2 displaces Mud2 to facilitate the addition of U2 snRNP (Kistler and Guthrie, 2001). Prp5 is thought to facilitate the smooth association and stable binding of U2 snRNA with the pre-mRNA Branch Site (BS) (Xu and Query, 2007).

There is evidence that U2 snRNA associates with the -6 site of the 5'exon in an ATP dependent manner during pre-spliceosome formation before the tri-snRNP U4/U6.U5 joins the complex (Newman et al., 1995). After the pre-spliceosome formation, the U4/U6.U5 complex associates with U1-5'exon. There is evidence that the U5 loop 1 displaces U1 at the exon side of the 5'ss (Newman, 1997) and the U6 snRNA replaces U1 at the intron side of the 5'ss (Sontheimer and Steitz, 1993). The interaction of U4/U6.U5 with U1 comes with conformational/structural changes cumulating into B* complex followed subsequently by dissociation of U4 snRNP from U6 snRNA and further compositional re-arrangement. The unwinding of U4 from U6 is catalysed by Brr2. During the transition from B to B-act the U6 base-pairs extensively with the U2 and further exchanges of proteins take place. The spliceosome assembly pathway has been extensively reviewed. Although the nature of these compositional exchanges/re-arrangement have been observed it is not known precisely how and in which spliceosome snRNP they take place when multiple snRNP is present at any one time in the splicing cycle. For instance, although it is known that Prp19-NTC/NTC-related proteins associate with the spliceosome at some point between the pre-spliceosome and Complex B-act, it is not precisely known when and if it associates directly with U6, U5 or U2 or all of them.

Proteomic analyses suggest that the Nineteen Complex (NTC) is evolutionarily conserved across the eukaryotes with both yeast and human protein complexes having similar compositions (Ajuh et al., 2000; Ohi et al., 2002). The *S. cerevisiae* NTC complex comprises of about eight proteins of which Prp19 is a principal component (Prp19, Snt309, Clf1, Lsy1, syf2, syf1, cef1, Ntc20) (Fabrizio et al., 2009b). Earlier immunopreicipitation studies suggest that NTC complex is recruited at B-act complex stage just as the U4 snRNA is released (Chan et al., 2003). More recent studies using pre-mRNA as bait have also shown that the NTC proteins are recruited to the pre-mRNA at the B-complex stage (Fabrizio et al., 2009b). Although, these studies show that NTC proteins associate with the spliceosome and pre-mRNA, they did not reveal which sub-complex of the spliceosome these group of proteins interact directly with. These proteins are well established splicing factors and are known from previous studies as Cef-associating factors or splicing factors which, along with Prp19, associate with the human 35S U5 snRNP (Ohi et al., 2002). It has been previously established that two of these proteins, Cww15/AD002 and Prp46/PRL1 are part of the human Prp19/CDC5L complex and are already integrated into the human B-complex (Deckert et al., 2006; Fabrizio et al., 2009b). The yeast proteins, Ecm2 and Bud31 whose human homologs are the Prp19-related proteins, RBM2 and G10, respectively are also members of NTC-related proteins (Makarova et al., 2004). In addition, splicing factors have been suggested to participate in transcription and splicing coupled to transcription (Conrad et al., 2000; Fong and Zhou, 2001) although its role in transcription remains unclear.

In this Chapter, it is proposed to investigate the interaction of Hsh155-TAP with U2 snRNP and other splicing factors when expressed under the regime of its endogenous promoter. This is to enable the identification of interacting partners that reflects its *in vivo* interaction. Hsh155p is a low abundant protein with a concentration of 521 molecules per cell (Ghaemmaghami et al., 2003) and therefore very challenging to purify. This also makes it difficult to explore its interaction with other proteins when expressed endogenously. Therefore, this study is aimed at developing a strategy for purifying endogenously expressed low abundant Hsh155p by performing a large scale TAP to provide further insight into U2 snRNP complex.

5.2 Result and Discussion

5.2.1 Analysis suggest four SF3b associating factors are stably associated

15 L yeast cultures were grown using the same conditions (temperature and nutrients) in a fermentor. The time of growth in three experiments was between 16 to 18 hours. Because of the difficulty in handling such large cultures, the time of growth, harvesting and lysing of cells were not synchronised. In this study, three tandem affinity purifications were performed. TAP purifications were essentially performed as described by Puig et al., 2001 (see Chapter 2.4). In addition TAP was performed in both the presence and absence of 10 mM DTT. The addition of 10 mM DTT has the potential of disrupting weakly associated proteins and mRNAs so that they do not co-purify with bait protein. Analysis of data and their discussion was based on criteria described in Chapter 3.

Following Hsh155-TAP, SDS PAGE was used to analyse the purified sample (see Figure 5.1). In gel tryptic digestion in conjunction with MS analysis was performed to identify the proteins (see Chapter 2.6-2.7) which are summarised in Table 5.1. The results show that Hsh155-TAP co-purify with components of U2 snRNP including, Rse1p, Hsh49p and Cus1p. The results demonstrate that Hsh155 is part of U2 snRNP complex consistent with previous observations (Pauling et al., 2000).

The roles of the four stably associated U2 snRNP proteins identified in this experiment have been investigated previously. Cus1p forms a complex with Hsh155p and Hsh49p, and is required for assembling U2 snRNP into the spliceosome (Pauling et al., 2000; Wells et al., 1996). Although the role of Rse1p is not clear, it is however known to be a component of the pre-spliceosome, and associates with U2 snRNA and contains two RNA recognition motifs (Caspary et al., 1999). All four proteins are components of the SF3b sub-complex of U2 snRNP. These results suggest that the four SF3b proteins are stably associated and their interactions are not disrupted during potential remodelling events. The co-purification of all four proteins in both the absence and presence of DTT suggests that all four proteins form a stable complex. The SDS PAGE analysis, although not directly quantitative, also suggests a stoichiometric association between the four proteins (see Figure 5.1).

These results demonstrate that all four SF3b proteins form a stable complex during spliceosome assembly *in vivo*, which has not previously been reported to date. The results suggest that the Hsh155 form a complex with all three proteins. This is also the first time these U2 snRNP associating factors have been identified using Hsh155-TAP-MS analysis.

5.2.2 SF3a proteins, Prp9 and Lea1 associate with the U2 snRNP sub-complex, SF3b.

The U2 snRNA is known to be bound by a heptameric Sm protein complex, U2A'-U2B' heterodimer (Scherly et al., 1990), SF3b and SF3a complexes. *In vitro* studies have shown that the U2 snRNP is assembled in a defined pathway. First, the Sm protein complex associates binds at the stem-loop IV of the U2 snRNA; at the same time U2A'-U2B" heterodimer binds to the U2 Sm site. The binding of Sm protein complex and U2A'-U2B" complex to the snRNA forms the 12S particle (Boelens et al., 1991; Scherly et al., 1990). Subsequently, a multimeric SF3b complex binds to the 12S particle to form a 15S which is then joined by SF3a to form the 17S particle (Behrens et al., 1993; Brosi et al., 1993; Kramer et al., 1999). The yeast SF3a is made up of Prp9, Prp11 and Prp21. Among all three analyses, only in one experiment was Prp9 detected. Lea1p is the yeast homolog of the human U2A' protein while Msl1p is the yeast U2B component of the U2 snRNP. However, only Lea1p was co-purified in a single experiment (see Table 5.2). Most of the SF3b sub-complex proteins were co-purified.

The absence of SF3a proteins, Prp11 and Prp21, SF3b14b and the U2B protein Msl1, and the presence of U2A (Lea1p), SmB1 and U2 snRNP-associating factors suggests that the purified U2 snRNP complex may have undergone remodelling where some of its components have disassembled (see Table 5.2). However, the data raises a number of other possibilities. One possibility is that the TAP tag may interfere with the association of SF3a protein components, such as Prp11 and Prp21 since they were not identified in any of the experiments. It may also be that Hsh155 and the SF3b proteins do not sufficiently interact with Prp11 and Prp21 in such a way as to

co-purify with them. The third possibility is that Prp11 and Prp21 are dislodged from the core U2 snRNP during remodelling events. The third possibility seems unlikely in the light of an *in vitro* experiment that used pre-mRNA as bait where Prp11 and Prp21 was shown to persist in the splicing complex or on the spliced pre-mRNA up to the B-act complex stage (Fabrizio et al., 2009b). Interestingly, a recent structural study shows that Prp21 wraps around Prp11 forming a compact complex and that Prp9 interacts via a bidentate mode to form the so called SF3a complex (Lin and Xu, 2012). This suggests a U2 snRNP that undergoes remodelling during splicing (See Figure 5.2).

In conclusion, these result suggest that the SF3b proteins interact with U2A'/Lea1 and the SF3a component, Prp9p. It also suggests that U2B''/Msl1p and the SF3a proteins, (Prp21 and Prp11p) do not form a stable complex with SF3b proteins.

5.2.3 The U5 snRNP component is present in purified U2 snRNP associating complex suggesting a novel interaction with U2 snRNP associating factors

In two experiments involving stringent (Hsh155-TAP + 10 mM DTT in Table 5.3) and mild (Hsh155-TAP No DTT in Table 5.3) purification conditions, the U5 snRNP component, Snu144p was identified in the MS analysis. Snu114p is a GTPase thought to bind directly to U5 snRNP and is suggested to play a role in the conformational changes of the spliceosome (Brenner and Guthrie, 2005; Dix et al., 1998; Fabrizio et al., 1997).. The co-purification of Snu114p under stringent and mild purification conditions suggests that Snu114p associate stably with U2 SF3b which has not previously been reported. It has been proposed that Snu114p is the only GTPase required for splicing and that GTP hydrolysis results in a rearrangement between Prp8p and the C- terminus of Snu114p which leads to the release of U1 and U4, thus activating the spliceosome for catalysis. This co-purification of Snu144p with U2 snRNP associating factors suggests that Snu114p may play a role in the assembly of U2 snRNP into the spliceosome.

The Hsh155 TAP MS experiments failed to detect U1 snRNP proteins, Sm and the majority of U4, U6 and U5 snRNP/associating proteins (see Table 5.1). The results suggest that the SF3b complex does not interact with most of the Sm and U1 and most of U4/U6.U5 complex proteins.

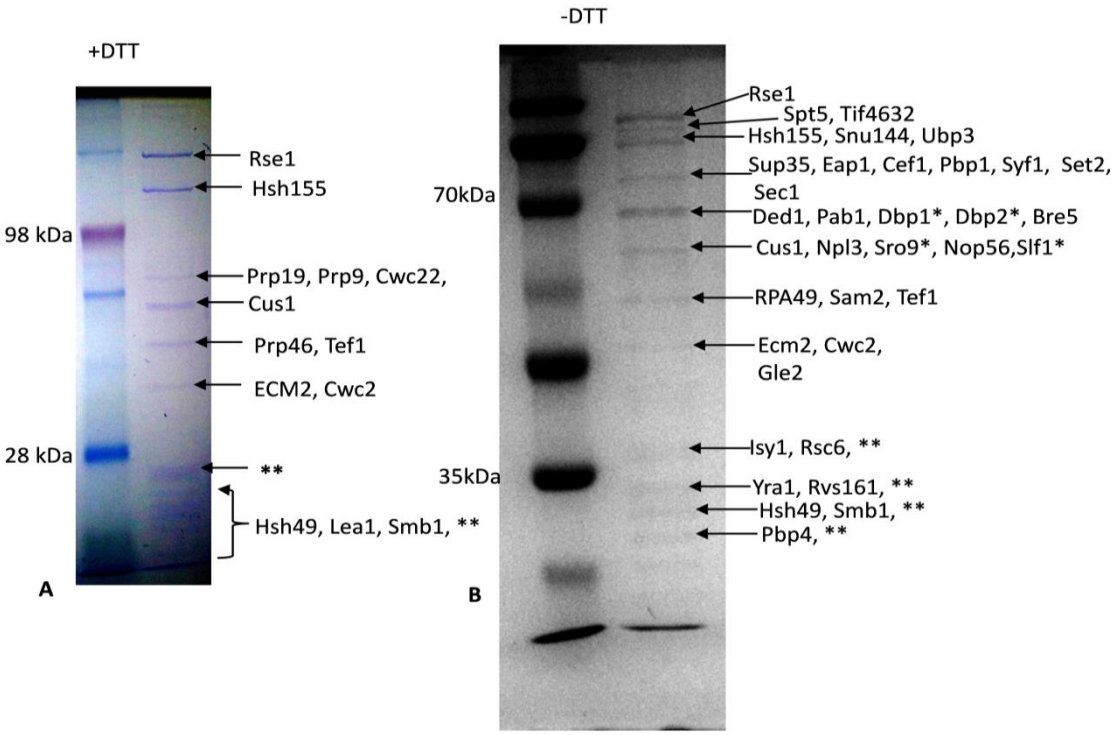


Figure 5.1 SDS PAGE analysis of Hsh155-TAP. A) 10% SDS PAGE analysis of Hsh155-TAP in the presence of DTT. Key proteins identified by MS are highlighted. B) 10% SDS PAGE analysis of Hsh155-TAP in the absence of DTT. Selected proteins identified by MS are highlighted.

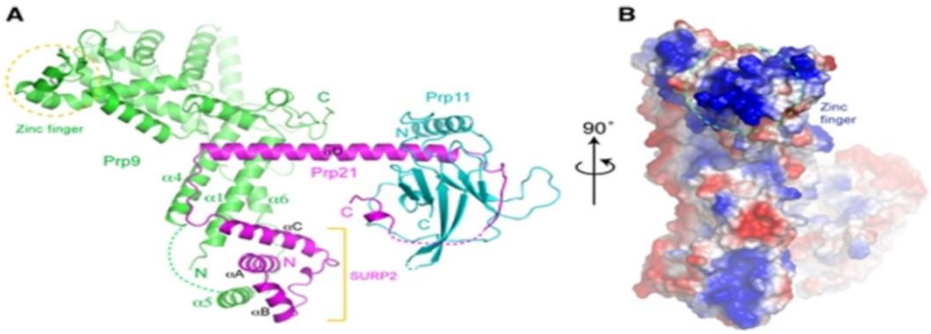


Figure 5.2 Structure of the SF3a core. (A) A ribbon representation of the core domain of the Prp9–Prp21–Prp11 complex. Prp9 is shown in green, Prp21 in magenta and Prp11 in cyan. Dashed lines denote disordered segments in the structure. A Cys2His2 U1C-type zinc finger (zinc atom shown as a sphere) of Prp9 and the SURP2 domains of Prp21 are indicated. (B) An orthogonal view of the heterotrimeric complex is shown in a surface representation with electrostatic potential distribution (blue, positively charged; red, negatively charged; white, neutral). The encircled area indicates the Prp9 region surrounding the zinc finger. Taken from (Lin and Xu, 2012).

The Nineteen-Complex/Prp19 is a non-snRNA protein complex associated with the spliceosome. Although previous studies suggest that NTC/NTC-related proteins stabilize U5 and is thought to be associated with the spliceosome before or after the dissociation of U4 from the U6 (Chan et al., 2003), no direct evidence exist for its interaction with U5 or U6 snRNPs. In fact, no study has established which of the snRNP proteins it specifically interacts with. The Hsh155-TAP MS analysis identified a number of NTC/NTC-related proteins (see Table 5.2). The results suggest that these proteins may associate with spliceosome and potentially interact with the U2 snRNP SF3b component of the spliceosome. Interestingly, none of the Lsm-TAP (see Chapter 4) co-purified with NTC/NTC-related proteins. Furthermore, Prp19, a prominent member of the NTC complex, binds the NTC proteins Cef1, Snt309 and the NTC-related protein, Cwc2 (Ohi et al., 2005) and all these proteins were detected in the purified Hsh155-TAP U2 complex.

Another interesting splicing factor identified is Cwc22. Cwc22 is a splicing factor known to be essential for the function of Prp2 during the first splicing step (Ohi et al., 2002). The detection of this protein in the Hsh155-TAP U2 snRNP in one Hsh155-TAP experiment (see Table 5.2) suggest that, although SF3b proteins may not be in direct contact or in complex with U2/U6 snRNA during the first catalytic step, they may indeed be present in the spliceosome and may be essential for the constitution of the Complex B-act required for the first catalytic step.

	Hsh155-TAP
U2 snRNP proteins	**Present**
NTC/Prp19 Complex	**Present**
NTC-related proteins	**Present**
Cwc22 *	**Present**
Snu114	**Present**
SmB or SmB1 or RSMB *	**Present**
Nucleolar complex proteins	**Present**
Translation Initiation and regulation factors	**Present**
Heat Shock Proteins	**Present**
Deadenylation-Decapping associating proteins	**Present**
Transcription-Export (TREX) Complex	**Present**
Chromatin associating proteins	**Present**
U4/U6.U5 snRNP complex proteins	**Absent**
U1 snRNP proteins	**Absent**

Table 5.1: Summary of spliceosome proteins identified in three Hsh155-TAP Proteins were identified using in gel tryptic digestion in conjunction with LC-ESI-MS .The table shows a summary of all proteins identified in three TAP preparations represented by their protein group/type/snRNP/ complex name.

U2 snRNP proteins	Hsh155-TAP - DTT	Hsh155-TAP +DTT
Rse1	1408	3080
Hsh155	1823	2047
Cus1	1806	1664
Hsh49	385	479
Prp9		378
Lea1 or RU2A		193
NTC/Prp19 Complex		
Prp19		910
Syf1	120	
Cef1	154	
Cwc23	60	70
Isy1	86	
Ntc20	40	
NTC-related proteins		
Prp46		189
Ecm2 or SLT11	86	124
Cwc2	60	70
Known Splicing factors		
Cwc22		193
Sm Proteins		
SmB or SmB1 or RSMB	85	46
Nucleolar complex proteins		
NOP56	77	
NOP14	37	
Nop53	77	
Imp3	76	
Translation Initiation and regulation factors		
PAB1	1620	
Tif4631	366	
Tif4632	220	
Rpg1 or Tif32	93	
Pbp1	186	
Pbp4	152	
Lsm12	82	
Deadenylation-Decapping associating proteins		
Sup35 or Erf3	1513	
EAP1	202	
SSD1	422	
Dead-Box protein family		
DBP1	362	
Ded1	1800	
Transcription-Export (TREX) Complex		
YRA1	164	
NOP3 or NPL3	182	
Gle2	41	
Chromatin associating proteins		
RPA49	130	
Set2	66	
Pol5	30	
Rad7	37	
Def1 Ubiquitination	308	
Ubp3 Ubiquitination	341	
Bre5	231	
Rsc6	44	
Spt5	1301	

Table 5.2 Hsh155-TAP MS analysis in the presence and absence of 10 mM DTT. Proteins were identified using in gel tryptic digestion in conjunction with LC-ESI-MS. Proteins identified in two TAP experiments where 10 mM DTT was used (Hsh155-TAP + 10 mM DTT) and proteins identified in one TAP experiments where DTT was excluded in one TAP experiment (Hsh155-TAP - DTT) are shown. The MASCOT scores of the proteins identified are highlighted.

5.4 Hsh155-TAP co-purifies with Transcription Factors, TREX factors and Translation factors

There is increasing evidence that splicing is coupled to transcription and that splicing factors function directly to promote transcriptional elongation (Conrad et al., 2000). In this study, we show that U2 snRNP co-purifies with a number of proteins involved in regulation of transcription. Moreover, U2 snRNP co-purified with important TREX factors suggesting that U2 forms a network with a multi-complex that is required for transcription, splicing and export of mRNAs.

Hsh155-TAP co-purified with Spt5p, Set2, RPA49, Bre1p, Ubp3p, Def1p and Rsc6 demonstrating a novel interaction between U2 snRNP and these proteins, which play roles in transcription or transcription-mediated processes. The result from this study suggests that U2 snRNP either interacts with these transcription protein complex or they interact with actively transcribed pre-mRNA at the same time as these proteins.

Interestingly, this is the first time the Spt5p and Set2p are shown to co-purify/interact with U2 snRNP. It is also of interest that mRNA capping complex co-purifies with U2 snRNP and transcription factors in this study since they have all been implicated in transcription initiation-elongation pathway. The relationship between Spt5p, transcription initiation-elongation and mRNA capping complex is now well-known. Briefly, the phosphorylation of the RNA pol II C-terminal domain leads to the recruitment of mRNA capping complex (Wen and Shatkin, 1999). The subsequent recruitment of the elongation factor Spt4/Spt5p to the elongating pol II facilitates the recruitment of the PAF complex. PAF complex provides the platform for the recruitment of H3K4 methylating Set1/COMPASS complex and the H2B ubiquitylation enzymes Rad6/Bre1. Further Pol II CTD phosphorylation, recruits the methyltransferase, Set2p, which catalyses H3k36 methylation. This methylation step promotes Rpd3s histone deacetylase complex, which is a crucial step to prevent cryptic transcription initiation at open reading frame (Liu et al., 2009). The co-purification of Spt5p, Set2p along and other transcription regulating factors with Hsh155-TAP (see Table 5.1) supports the model that Spt4/Spt5 plays a crucial role in transcription elongation and couples this role to promoting splicing.

Although, a role for Spt5p in splicing has been suggested, this study shows that it co-purifies with U2 snRNP SF3b complex. A previous study has shown that mutation in the genes of the transcription factor, Spt4p and Spt5p, leads to the accumulation of splicing factors (Lindstrom et al., 2003). The co-purification of Spt5p with U2 snRNP provides further evidence that some transcription factors play important roles in splicing, perhaps, by facilitating the recruitment of splicing factors. The data supports the model that Spt5p may play a role in the recruitment of U2 snRNP.

In a previous study it was shown that that mutations in components of the BUR complex lead to defective transcriptional phenotypes similar to SPT genes. Mutants of the BUR components, BUR1 and BUR2 were initially identified to increase transcription of SUC2 genes in mutants lacking its upstream activating sequence (UAS). Mutations that bypass the UAS requirements increase transcription from suc2Δuas exhibit BUR- phenotypes. Mutations in specific domains of the BUR1 and BUR2 components of the BUR complex have been found to exhibit SPT- phenotypes (Keogh et al., 2003; Malone et al., 1991; Murray et al., 2001; Prelich and Winston, 1993; Yao and Prelich, 2002). Mutations in Spt4, Spt5, Bur1 and Bur2 result in Spt- phenotypes which indicate their roles in regulating transcription events (Malone et al., 1991; Malone et al., 1993; Winston et al., 1984). The co-purification of splicing factors of the U2 snRNP complex and the transcription factor, Spt5p further confirms the role of splicing factors in transcription events and/or vice versa. Other interactions with transcription and DNA-repair associating proteins identified in this work will lend further credence to this argument. The findings in this work are consistent with the model that transcription is functional coupled to splicing. This thesis reports, for the first time, the physical interactions between splicing factors, U2 snRNP complex and transcription factors in yeast. Further studies are, however, required to determine the functional significance of such interactions, and the mechanism of such coupling between transcriptions and splicing in yeast.

It has been suggested in previous studies that transcription and splicing are functional coupled mechanisms. This model predicts that specific transcription factors and splicing factors should be able to interact physically and that these interactions would affect splicing, transcription, or both (McKay and Johnson, 2011).

Although such a relationship has been demonstrated in mammals, specifically between the transcriptional elongation factors P-TEFb and Tat-SF1 (Fong and Zhou, 2001), there is sparse evidence for such relationship in yeast. An attempt in a previous studies to examine this functional relationship between yeast P-TEFb homologs, BUR-proteins and U2 snRNP was unable to establish such interaction or any detectable effect of U2 snRNP on transcription (McKay and Johnson, 2011). However, a caveat was issued by the same study that their findings do not eliminate the possibility that splicing factors do not play a role in transcription perhaps via other genes and under conditions not tested in their studies (McKay and Johnson, 2011).

In contrast, the co-purification of splicing factors with transcription factors in the present study suggest that splicing factors form a complex with transcription factors, (see Table 5.2). The results showed that Def1 co-purified with Hsh155-TAP which demonstrates a novel interaction of Def1p with U2 snRNP complex. Def1p is known to enable the ubiquitylation and proteolysis of RNAPII during transcription coupled-DNA repair (Woudstra et al., 2002). It is interesting that previous *in vitro* studies have suggested transcription, in particular, RNAPII could affect splicing; it was further suggested that a similar interaction can take place *in vivo* (Fong and Zhou, 2001; McKay and Johnson, 2011). The co-purification of Def1p with U2 snRNP suggests that Def1p is recruited at some point during transcription for the degradation of RNAP II. It can be speculated that, ubiquitylation and degradation of RNAP II and other histone methylating complexes is essential for the recruitment of splicing factors and extension of pre-mRNA during transcription. Since Def1p has been implicated in transcription coupled-DNA repair (Woudstra et al., 2002), it may be that they are needed for the clearing of RNAPII stalled at lesions during transcription.

Other proteins that co-purified with Hsh155-TAP include the TREX factors Yra1, Npl3 and Gle2, which play important roles in the processing and nuclear export of mRNAs (see Table 5.2). Also factors, involved in translation initiation, such as Pab1p and the cytoplasmic mRNA cap binding proteins, Cdc33, Tif4631 and Tif4632 co-purified with Hsh155-TAP (Table 5.2). This supports the model that transcription, splicing and mRNA transport are coordinated by a network of highly dynamic multi-complex that

is recruited during transcription (Reed, 2003). It does appear as if most of this dynamic multi-complex is recruited to the pre-mRNA during transcription and undergoes remodelling with some proteins retained, while others are actively recruited and exiting the multi-complex during transcription, splicing and transport. The co-purification of most translation factors and proteins implicated in translation regulation and mRNA degradation will support a notion that RNA processing and regulation pathways are tightly linked.

Conclusions

The tandem affinity purification (TAP) tag was used to enable the TAP of Hsh155. Following TAP, mass spectrometry was used to identify a number of known and novel interacting partners. This study demonstrates that Hsh155p interacts and form a stable complex with the U2 snRNP SF3b complex components, including Cus1p, Rse1p and Hsh49. It also shows that U2 snRNP proteins, more specifically the SF3b sub-complex, interact with Prp19-NTC/NTC-related proteins. In addition, the results show that Hsh155-TAP complex associates with transcription/transcription-elongation factors, splicing factors, TREX factors and translation regulation factors suggesting a role for U2 snRNP in a RNA-processing multi-complex. This study demonstrates for the first time that the transcription elongation factors, Spt5p and Set2p, along with other transcription factors interact with splicing factors. This evidence supports the notion that pre-mRNA splicing and mRNA transport is tightly coupled to transcription.

Chapter VI

Studying crRNA processing in a novel Type III CRISPR/Cas system in *Streptococcus thermophilus*

Abstract

Bacteria and archaea integrate short fragments of foreign nucleic acid into the host chromosome at one end of a repetitive element known as CRISPR (clustered regularly interspaced short palindromic repeat). CRISPR loci are transcribed and the long primary transcript is processed into mature CRISPR RNAs (crRNAs) that each contains a sequence complementary to a previously encountered invading nucleic acid. The silencing of invading nucleic acids is performed by ribonucleoprotein complexes that utilise crRNAs as guides for targeting and degradation of foreign nucleic acids. In this Chapter, denaturing ion pair reversed phase chromatography in conjunction with ESI-MS was used to investigate the processing of crRNA and the role of the CRISPR associating proteins in a novel Type III CRISPR/Cas system in *Streptococcus thermophilus*. In this study, the architecture of the intermediate and mature crRNAs was determined. The results demonstrate that the pre crRNA is processed to a 72 nt crRNA which is further processed into smaller 40/41 nt crRNAs. Furthermore MS analysis revealed that the protein stoichiometry of the Cascade complex affects crRNA processing.

6.1 Introduction

Clusters of Regularly Interspaced Short palindromic Repeats (CRISPR)-mediated immunity is a defence mechanism evolved by prokaryotes to protect themselves against invading viruses and potential pernicious genetic elements. CRISPR complex undergoes three stages during this process. The first stage is adaptation where the host first encounters the invader and forms a memory of the infection by incorporating a piece of the foreign DNA into its genome. In the second stage, biogenesis, the CRISPR complex is formed which potentially mediates the third stage where the invader DNA is recognized and destroyed. The biogenesis is a crucial stage in the CRISPR-immunity pathway (See Figure 6.1) (Makarova et al., 2011).

The CRISPR system has been evolved in bacteria to combat invading viral and pernicious genetic agents in its environment and therefore is an important aspect of bacterial immune system (Barrangou et al., 2007; Brouns et al., 2008; Sapranauskas et al., 2011). The mechanism of defence by the CRISPR system utilizes a restriction strategy that recognizes and degrades viral DNA (Barrangou et al., 2007; Bickle and Kruger, 1993) and is mediated by a complex comprising of a number CRISPR associating proteins (Cas) and small RNA known as CRISPR RNA (crRNA) (Barrangou et al., 2007; Garneau et al., 2010; Horvath and Barrangou, 2010). The crRNA is a transcription product of a DNA from an invading virus/genetic material incorporated into the host bacterial genome during previous infection episode and may be described as the memory of infection (van der Oost et al., 2009). However, the crRNA requires the CRISPR associated complex (Cas) proteins for processing and efficient recognition and degradation of invading viral genetic material. A specific range of activities, including, endonuclease, exonuclease, helicase and DNA/RNA binding activity are associated with the Cas proteins (Carte et al., 2008; Haft et al., 2005; Jansen et al., 2002; Makarova et al., 2006). The crRNA is incorporated into the surveillance multi-subunit protein complex to form the CRISPR-associated complex for antiviral defence (Cascade) required for protection against invading viral agent (Brouns et al., 2008; Jore et al., 2011). In *E. coli*, there are about eight Cas genes which together with a CRISPR locus downstream make up its immune system (Brouns et al., 2008; Jore et al., 2011; Wiedenheft et al., 2011).

CRISPR-mediated immunity can be divided into three crucial stages: an adaptation, CRISPR biogenesis and Interference stage. The host first encounters a viral agent and incorporates a portion of the invading genetic material as a new spacer into the CRISPR locus in a step referred to as adaptation stage ((Barrangou et al., 2007; Horvath and Barrangou, 2010; van der Oost et al., 2009). In the second stage, the host cell transcribes the CRISPR region, which contains the leader, repeats and spacer sequences, into a precursor CRISPR RNA (pre-crRNA). The pre-crRNA is further cleaved into a shorter crRNA, which could be the 'maturated' or 'unmaturated' crRNA depending on the bacteria. For instance, in *E. coli* K-12, it was shown that pre-crRNA is processed in a stepwise manner to generate a small mature crRNA that is not further processed (Brouns et al., 2008). In *P. furiousus*, on the other hand, the pre-crRNA is first cleaved by Cas6 within the repeat at a point 8 nts upstream of the spacer to generate a precise 5' end. The repeat sequence at the 3'end is then specifically cleaved by an unidentified nuclease to produce approximately, 39 nt mature crRNA (Carte et al., 2008). In the interference stage which follows, the maturated Cas protein-crRNA complex potentially mediates the recognition of the target DNA to be degraded and requires the helicase activity of Cas3 (Barrangou et al., 2007; Brouns et al., 2008). There are, however, three types of CRISPR systems, namely, Types I, II and III and subtypes (see Figure 6.2) (Makarova et al., 2011). In addition to genes encoding proteins that probably form Cascade-like complexes with different compositions, the type I system contains a Cas3 gene which encodes a protein with separate helicase and DNase activities, and appears to target DNA during interference (Brouns et al., 2008; Makarova et al., 2011; Sinkunas et al., 2011). The type II contain in addition to the ubiquitous type I and II, a single protein, Cas9, which seem to be sufficient for processing crRNA and targeting DNA for cleavage (Kleanthous et al., 1999; Makarova et al., 2011). The type III CRISPR-Cas systems contain a polymerase and a repeat-associated mysterious protein (RAMP) that is probably involved in the processing of the spacer-repeat transcripts, analogous to Cascade.

There are three CRISPR systems in *Streptococcus thermophilus*, Type I, II and III (see Figure 6.2) (Makarova et al., 2006; Makarova et al., 2011). The type I system in *S thermophilus* is based on Cas3-mediated DNA cleavage and utilizes an effector

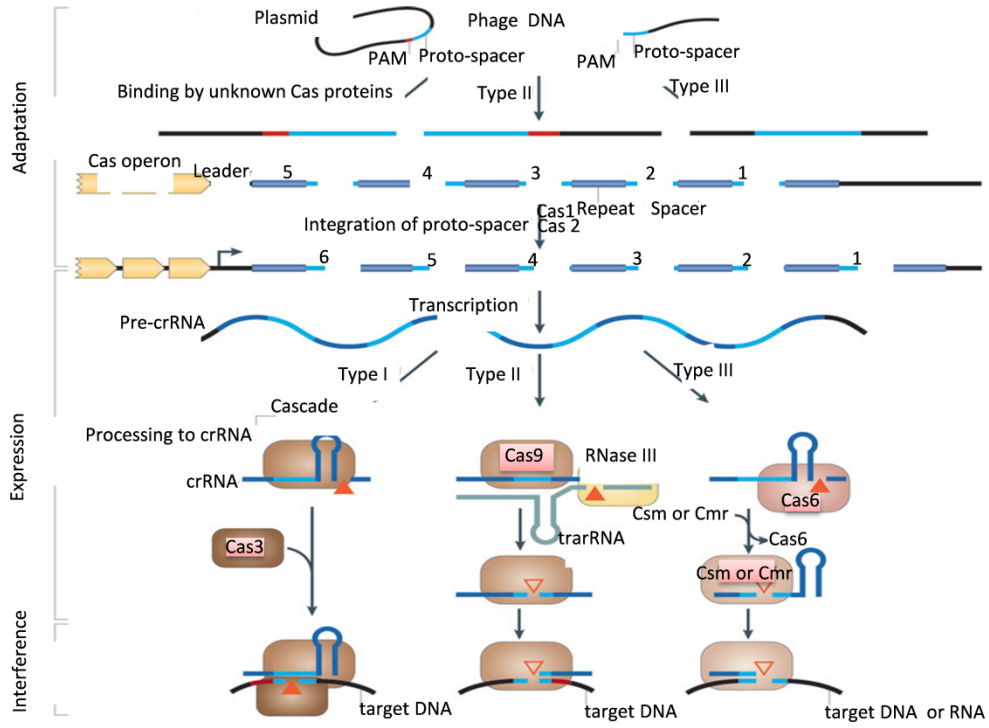


Figure 6.1 Stages of CRISPR Interference. CRISPR–Cas systems act in three stages: adaptation, expression and interference. In type I and type II CRISPR–Cas systems, but not in type III systems, the selection of proto-spacers in invading nucleic acid probably depends on a proto-spacer-adjacent motif (PAM). After initial recognition step, Cas1 and Cas2 incorporate the proto-spacers into the CRISPR locus to form spacers. During the expression stage, the CRISPR locus containing the spacers is expressed in pre-crRNA. The CRISPR-associated complex for antiviral defence (Cascade) complex binds pre-crRNA, which is cleaved by the Cas6e or Cas6f subunits, resulting in crRNAs with a typical 8-nucleotide repeat fragment on the 5' end and the remainder of the repeat fragment, which generally forms a hairpin structure, on the 3' flank. Type II systems use a trans-encoded small RNA (tracrRNA) that pairs with the repeat fragment of the pre-crRNA, followed by cleavage within the repeats by the housekeeping RNase III in the presence of Cas9 . In type III systems, Cas6 is responsible for the processing step, but the crRNAs seem to be transferred to a distinct Cas complex (called Csm in subtype III-A systems and Cmr in subtype III-B systems. During the interference step, the invading nucleic acid is cleaved. In type I systems, the crRNA guides the Cascade complex to targets that contain the complementary DNA, and the Cas3 subunit probably cleaves the invading DNA. The PAM may also plays an important part in target recognition in type I systems. In type II and type III systems, no Cas3 orthologue is involved. In type II systems, Cas9 loaded with crRNA may directly targets invading DNA, in a process that requires the PAM. The two subtypes of CRISPR–Cas type III systems target either DNA (subtype III-A systems) or RNA (subtype III-B systems. In type III systems, a chromosomal CRISPR locus and an invading DNA fragment are distinguished by either base pairing to the 5' repeat fragment of the mature crRNA (resulting in no interference) or no base pairing (resulting in interference). Filled triangles represent experimentally characterized nucleases, and unfilled triangles represent nucleases that have not yet been identified. Taken from Makarova et al., 2011.

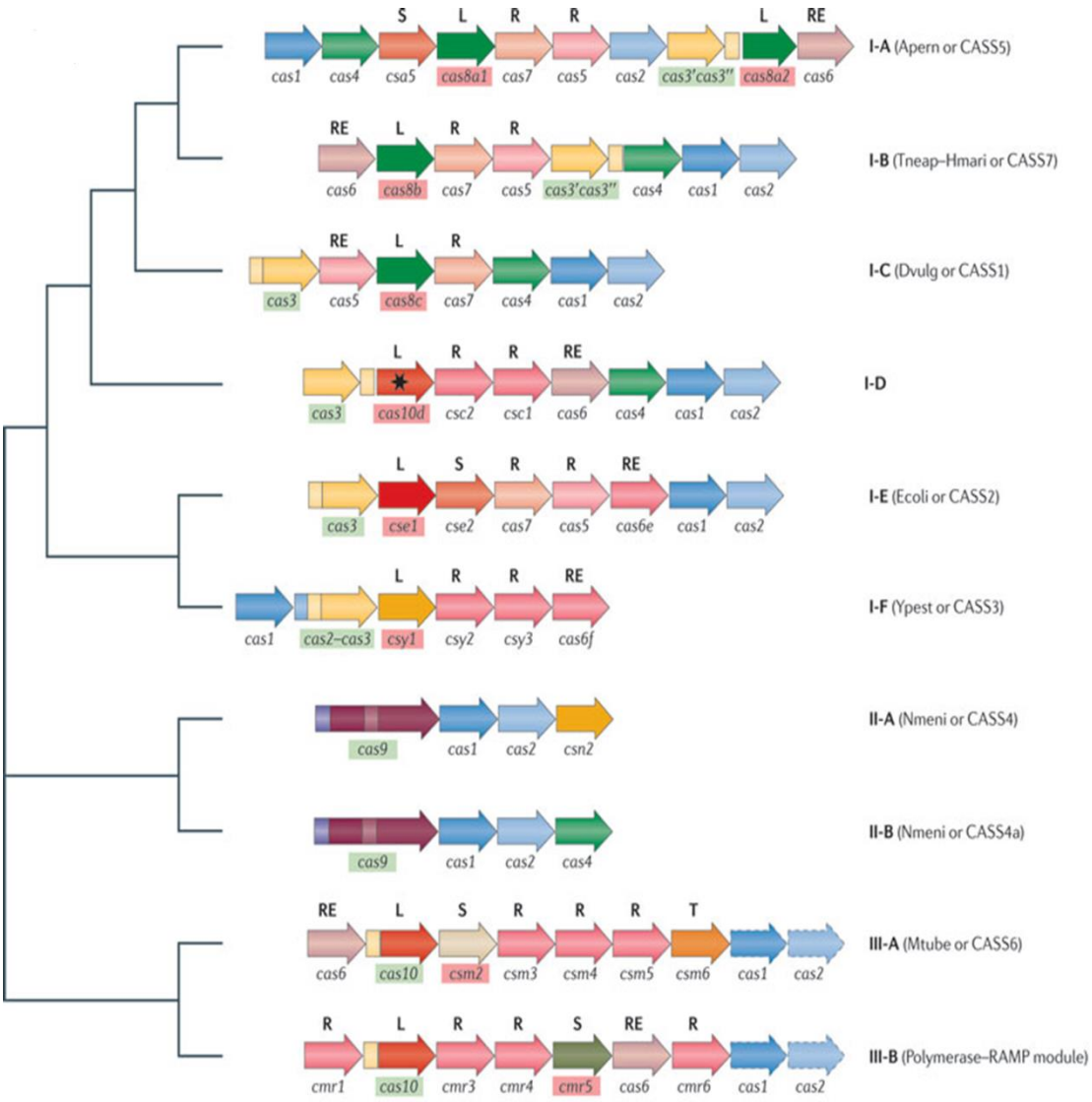


Figure 6.2 Classification of CRISPR/CAS systems. The typical, simplest operon architectures are shown for each type and subtype of CRISPR–Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) system; numerous variations exist. Orthologous genes are colour coded and identified by a family name. The signature genes for CRISPR–Cas types are shown within green boxes, and those for subtypes are shown within red boxes. The letters above the genes show major categories of Cas proteins: large CRISPR-associated complex for antiviral defence (Cascade) subunits (L), small Cascade subunits (S), repeat-associated mysterious protein (RAMP) Cascade subunits (R), RAMP family RNases involved in crRNA processing (RE) (note that only those in subtypes I-E, I-F and III-B systems have been characterized), and transcriptional regulators (T). The star indicates a predicted inactivated polymerase with an HD domain. For subtype I-A systems, the cas8a1 and cas8a2 genes are typically mutually exclusive but both can be considered signature genes for the subtype. For type III systems, the cas1 and cas2 genes in dashed boxes are not associated with all type III polymerase–RAMP modules. In addition to previously published data, this schematic shows Cas7 (COG1857) as a member of the RAMP superfamily. For each CRISPR–Cas subtype (except for the newly identified subtype I-D), the old names from are indicated in parentheses. Taken from (Makarova *et al*., 2011).

complex containing a 61 nt crRNA to bind *in vitro* a matching proto-spacer if a proto-spacer adjacent motif (PAM) is present (Sinkunas et al., 2013). In the type II system of S. thermophilus, the Cas9-crRNA complex has been shown to require an additional RNA molecule, trans-activating CRISPR RNA (tracrRNA) and in vitro, forms a ternary Cas9-crRNA-tracrRNA that cleaves DNA (Karvelis et al., 2013). Although in a recent study a role for a Type III protein, Csm3 was revealed in *Staphylococcus epidermidis* (Hatoum-Aslan et al., 2013), no studies have been done on the Type III system of *S. thermophilus*.

The aim of this chapter is to further analyse the Cascade-crRNA ribonucleoprotein complex from a type III CRISPR/Cas system in *Streptococcus thermophilus*. MS approaches will be used to both analyse the crRNA and the protein components in the RNP complex. The understanding of the architecture of the crRNA, including the cleavage site in the pre-crRNA, the sequence and characterization of the 5' and 3' termini, will provide further insight into the mechanism of crRNA interference. It also aims to provide further insight into the role of the CRISPR associated proteins in the crRNA biogenesis.

6.2 Results and Discussion

6.2.1 Proteomic analysis of the *Streptococcus thermophilus* Csm complexes

To study the processing of pre-crRNA by Cascade in the Type III CRISPR/Cas system in *Streptococcus thermophilus*, six Cascade proteins, Csm2, 3, 4, 5 and Cas 6 and 10 and *S. thermophilus* crRNA were overexpressed in *E. coli* in both the absence and presence of crRNA with Csm3 and Csm2 carrying the Strep-tactin affinity tag. A summary of the different complexes studied in this Chapter is shown in Table 6.1. The complexes were subsequently purified using affinity chromatography and gel exclusion chromatography in the laboratory of Prof. Virginijus Siksnys (Department of Protein-DNA Interactions, Institute of Biotechnology, Vilnius University, Lithuania). The purified complexes were analysed by SDS PAGE and proteins identified by both in solution digestion and in gel digestion and MS analysis (see Figure 6.3). A summary of the protein identifications is shown in Table 6.2

The MS analysis reveals a difference in the composition between the apoprotein complexes (-crRNA) and the ribonucleoprotein complexes. The results show that Cas6 is absent in the apoprotein complex while the ribonucleoprotein complex contains Cas6 (see Table 6.2) irrespective of the bait protein used. Cas6 co-migrates with Csm3 on the SDS PAGE making analysis via SDS PAGE difficult (see Figure 6.3). These results demonstrate that the Cascade complex is able to form in the absence of pre-crRNA, however pre-crRNA is required to recruit Cas6 to the Cascade complex. It is also consistent with the model that Cas6 is required for the processing of CRISPR RNA (Carte et al., 2008). This study demonstrates that CRISPR complex consisting of Csm1 (Cas10), 2, 3, 4 and 5 can form a complex in the absence of crRNA. This is consistent with a recent study where it was shown that crRNA is not required for the assembly of Cas10/Csm proteins (Hatoum-Aslan et al., 2013). However, the present study demonstrates that crRNA is required for the recruitment of Cas6.

Further comparative analysis of the Csm2+crRNA and Csm3+crRNA, using SDS-PAGE in conjunction with MS analysis was performed to determine the protein composition. The SDS PAGE analysis shows a potential reduction in the Csm5 protein

in the Csm3+crRNA complex (highlighted with a blue box in Figure 6.3). Further analysis of the MS data revealed an increased MASCOT score and sequence coverage consistent with SDS PAGE analysis (see Table 6.3) To obtain further semi-quantitative MS data, the ion intensities from a number of Csm5 peptides from the different complexes were compared (see Table 6.3). The results show that more abundant peptide intensities from Csm5 are observed for Csm2+crRNA compared to the Csm3+crRNA. In conclusion, these results suggest that the Csm5 forms a stable interaction with Csm2+crRNA but does not associate stably with Csm3+crRNA CRISPR complex.

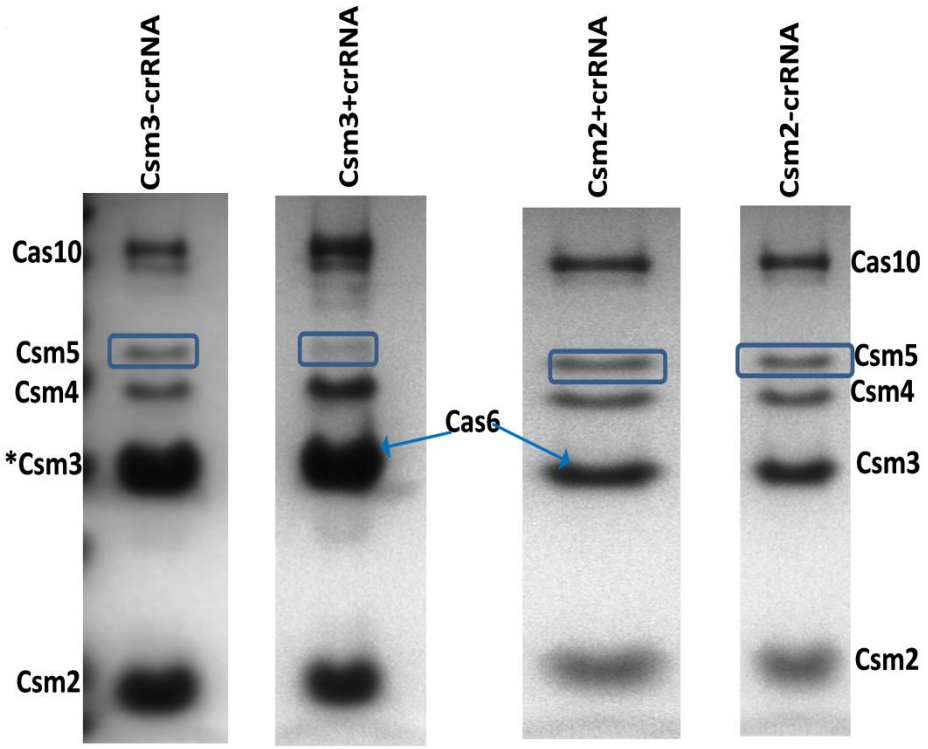


Figure 6.3 Characterisation of the Csm complexes purified from *S. thermophilus*. SDS PAGE analysis of the purified complexes in conjunction with MS identification was used to characterise each of the different Csm complexes. Tagged protein used as bait is indicated. Csm2-strep or Csm3-strep corresponds to the Csm-complex purified via the Csm2-strep-tactin or Csm3-strep-tactin complex. +crRNA or -crRNA-corresponds to expression of complex in the presence or absence of crRNA respectively. Blue arrows indicate highlight the identification of Cas6. Blue boxes highlight the Csm5 protein in the different Csm complexes.* indicates streptactin tag.

Table 6.1: **Composition of the different Csm complexes used in this study**

Sample complex	Description
Csm3+crRNA	Cascade proteins and *S. thermophilus* crRNA expressed, strep-tactin tag on Csm3
Csm2+crRNA	Cascade proteins and *S. thermophilus* crRNA expressed, strep-tactin tag on Csm2
Csm3-crRNA	Cascade proteins expressed but without *S. thermophilus* crRNA, strep-tactin tag on Csm2
Csm2-crRNA	Cascade proteins expressed but without *S. thermophilus* crRNA, strep-tactin tag on Csm2

6.3 Analysis of the crRNA processing the Type III *S. thermophilus* Cascade-crRNA complex.

6.3.1 Purification and analysis of crRNA from the Csm3 tagged complex

In order to further analyse the crRNA present in the Cascade-crRNA complexes, the RNA was purified from the Csm3+crRNA complex by chloroform-phenol extraction and analysed on 10% denaturing RNA PAGE (8M urea) (see Figure 6.4A). The size of the crRNA from the PAGE analysis was determined to be approximately 70 nt. For further verification and analysis using mass spectrometry, the crRNA was purified by injecting the intact Csm3+crRNA complex on the denaturing ion pair reversed phase chromatography (Dickman and Hornby, 2006; Waghmare et al., 2009) see Chapter 2.9.1. The results are shown in Figure 6.4B). A single peak was observed, collected and concentrated. The isolated crRNA reveals a single peak with a retention time consistent with a crRNA of approximately 70 nt length. The purified crRNA was further analysed using ESI-MS to obtain the accurate mass by separating on a 200 µm i.d. monolithic PS-DVB capillary column interfaced to ESI-MS (see Chapter 2.9.2). A molecular weight of 22998.51 Da was obtained (Figure. 6.4C) consistent with a theoretical 72 nt crRNA with a 5' OH and 3' P termini. To verify the 3' termini, acid treatment of the crRNA was performed with no change in mass observed using ESI-MS (data not shown). Using the accurate M_W of the crRNA (22998.51 Da) and the sequence of the pre-crRNA, the cleavage sites on the pre-crRNA could not be

unambiguously assigned. This is because a number of alternative cleavage sites are possible which corresponds to the intact mass observed (see Figure 6.5A). Therefore, it was necessary to map the sequence of the crRNA. As no other mass was detected in the Csm3+crRNA, the data suggests that the 72nt crRNA is not further processed in this complex.

6.3.2 Analyses of the 70 nt crRNA using RNAse mapping

In order to map the sequence of the crRNA, RNAse A/T digest was performed and the oligoribonucleotides analysed by tandem MS (see Chapter 2.8 and 2.9.2). The results are shown in Figure 6.6. A number of oligoribonucleotides were assigned to the spacer and its flanking repeat crRNA sequence. The assignment of the unique RNAse A digest fragment, GAGAGGGGp (see Figure 6.6B), in conjunction with the other assigned fragments from RNAse T1 and the intact mass (22998.51 Da) confirm that the cleavage point is within the repeat, 8 nt from the spacer (see Figure 6.5B). This corresponds to a cleavage point at the base of each hairpin loop of the crRNA sequence. The data is consistent with the generation of a 5' OH and 3' P termini. Further confirmation for the 3' P termini was provided by the ESI-MSMS of the terminal oligoribonucleotide GAGAGGGGp (see Figure 6.6C).

6.3.3 Purification and analysis of crRNA from the Csm2+crRNA complex

Analysis of the crRNA from the Csm2-strep complex was performed as previously described. The denaturing ion pair reverse phase chromatogram is shown in Figure 6.7B. The results show the presence of two major peaks with retention times corresponding to approximately 40 nt and 70 nt. These results are consistent with the denaturing RNA PAGE of the same complex (see Figure 6.7A). These results show that the 70 nt crRNA is further processed into smaller crRNAs in this complex. The presence of the lower abundant 70 nt crRNA suggests that it is a precursor of the crRNA of 40 nt approximate length. This suggests that the 70nt and 40 nt crRNAs, represent the unmature and mature crRNAs, respectively.

To further analyse the matured crRNAs it was necessary to determine the accurate sequence using ESI MS. Two intact crRNAs were found co-eluting and their intact mass molecular weights were determined as 12602.171 Da and 12907.6 (see Figure 6.7C). Although the accurate intact masses of the crRNAs were determined and the sequence of the pre-crRNA known, the cleavage sites on the pre-crRNA could not be unambiguously assigned. This is because a number of possible cleavage sites corresponding to the observed molecular weight also exist. However, the results predict 40 and 41 nt crRNAs retaining the 5' OH of the intermediate crRNA (72 nt) and having 3' OH terminus (see Figure 6.4). To verify the cleavage of the 72 nt crRNA at the predicted sites shown in Figure 6.5B, it was necessary to map the sequence of maturated crRNAs using RNAse mapping.

6.3.4 RNAse mapping of the maturated crRNA

Following purification of the crRNAs from the Csm2+crRNA complex by phenol-chloroform extraction (see Chapter 2.5.1), RNAse T1 digest was performed and the oligoribonucleotides were analysed by ESI-tandem MS. A number of oligoribonucleotides were assigned to the spacer and its flanking repeat crRNA sequence (see Figure 6.8A). Furthermore, these results enabled assignment of a unique oligoribonucleotide corresponding to the maturated crRNA termini, 5'OH-UUCACUUAUUC-3'OH (see Figure 6.7B). The combined accurate intact mass measurements and RNAse mapping confirm that the maturated crRNA is a 40 mer with a 5'OH and 3'OH termini. MS fragmentation of the unique oligoribonucleotide, UUCACUUAUUC, was poor and therefore could not be used to confirm through MS sequencing. The assignment of the unique fragment, UUCACUUAUUUCC (see Figure 6.8B) in conjunction the intact mass suggest that this is a 41 nt crRNA with 5'OH and 3'OH termini. It is also consistent with a 41nt crRNA that shares the same 5' terminus as the 40 nt crRNA with an extra C nucleotide at its 3'end (Figure 6.5C). It is interesting that the 3' terminal cleavage point is within the spacer sequence. The peaks for the two alternatively cleaved maturated crRNAs could not be resolved by the denaturing ion pair reverse phase chromatography apparently because of the

similarity in base sequence and length but the MS data identifies them and suggests that the 41 nt is much less abundant than the 40 nt crRNA.

Protein	Molecular Weight (kDa)	Csm3-crRNA Mascot Score	Csm2-crRNA Mascot score
Cas 10 or Csm1	86.89	2143	2257
Csm5	41	1587	1726
Csm4	33.7	865	1172
Csm3	24	1044	1632
Cas6	28		
Csm2	14.8	532	330

Table 6.2 Summary of proteins identified in Csm 2/3-crRNA complexes. Showing there is no appreciable difference between Csm3-crRNA and Csm2-crRNA. Both contain Csm5 and lack Cas6.

Protein/peptide	Ion mass	Csm2 Ion intensity	Csm3 ion intensity	Ion Score Csm2	Ion score Csm3
Csm5			-		-
WNNENAVNDFGR	718.340 (2+)	3.08 X 10^5	-	90	-
KGKEYDDLFNAIR	523.68 (3+)	1.64 X 10^5	-	81	-
IEFEITTTTDEAGR	791.94 (2+)	1.0 X 10^4	-	59	-
LISFLNDNR	546.39 (2+)	3.45 X 10^5	2.8 X 10^4	66	34
NHESFYEMGK	621.30 (2+)	1.70 X 10^5	1.31 X 10^4	40	32
DAFGNPYPGSSLK	733.3960 (2+)	4.09 X 10^4	2.5 X 10^4	50	30
Cas 10 (86 kDa)					
GDYAAIATR	469.33 (2+)	8.4 X 10^4	4.4 X 10^4	67	66
HNYKEDLFTK	647.87 (2+)	4.51 X 10^5	1.5 X 10^5	70	74
DSISLFSSDYTFK	755.36 (2+)	3.12 X 10^4	1.8 X 10^4	75	74
Cas 6 (28.4 kDa)					
LVFTFK	377.80 (2+)	3.86 X 10^4	1.95 X 10^4	23	27
LIFQSLMQK	554.3520 (2+)	4.61 X 10^4	2.62 X 10^4	50	42
RIDHPAQDLAVK	455.0 (3+)	5.77 X 10^4	2.5 X 10^4	43	36
SQGSYVIFPSMR	686.40 (2+)	2.94 X 10^4	1.45 X 10^4	57	70
Csm4 (33.7 kDa)					
HDQIDQSVDVK	642.40 (2+)	4.04 X 10^5	4.5 X 10^5	81	75
FELDIQNIPLELSDR	901.50 (2+)	2.43 X 10^4	3.453 X 10^3	100	78
DGNLYQVATTR	619.3580 (2+)	2.28 X 10^5	2.99 X 10^5	73	70
Csm2 (14.817 kDa)					
AQILEALK	443.30 (2+)	3.14 X 10^5	4.79 X 10^4	45	51
VQFVYQGR	534.40 (2+)	1.2 X 10^6	5.20 X 10^5	65	61
YMEALVAYFK	617.90 (2+)	-	1.49 X 10^5	-	74
Csm3 (24.541 kDa)					
LLELDYLGGSGSR	690.40 (2+)	3.4 X 10^6	1.73 X 10^6	113	110
VAEKPSDDSDILSR	766.40 (2+)	1.51 X 10^6	2.95 X 10^6	85	82
DAFLSNADELDSLGVR	861.50 (2+)	1.80 X 10^5	2.56 X 10^5	94	108

Table 6.3 Comparison of the ion intensities of the Csm2+crRNA and Csm3+crRNA complexes Following ESI/MSMS, the ion intensities of the Csm+crRNA complexes were compared. The results show significant differences in the intensities of Csm5 peptides in the two complexes, in contrast to the majority of the peptides from the Cas10/Csm proteins.

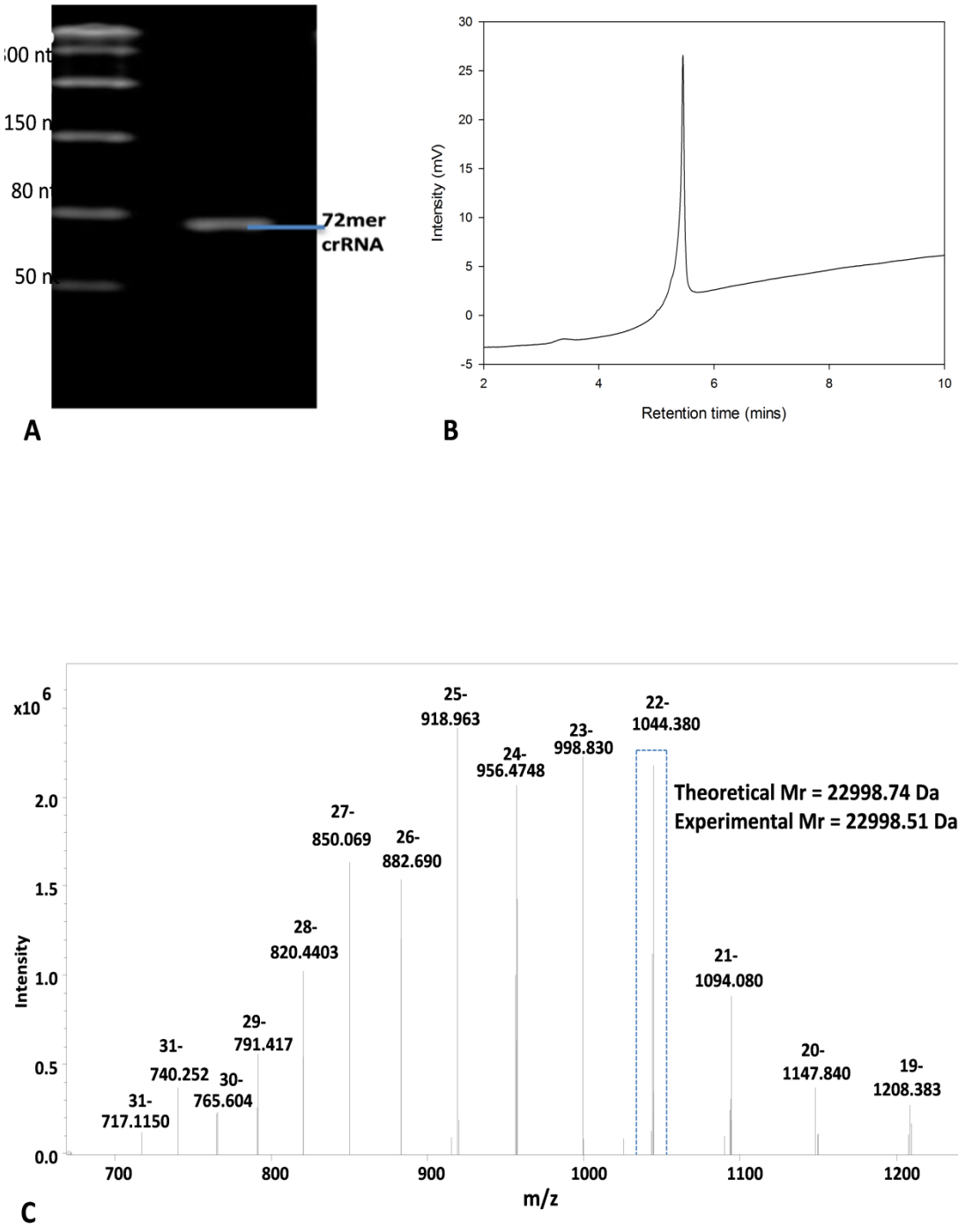


Figure 6.4 Analysis of crRNA processing in *S. thermophilus* Csm complex crRNA was purified from Csm3-strep+crRNA complex by phenol-chloroform extraction and analysed by (**A**) 8M urea denaturing PAGE. RNA of approximately 70 nts in length was observed (**B**) Denaturing IP RP HPLC analysis gradient condition 1 (see Chapter 2.9.1) was performed (**C**) LC ESI MS analysis the IP RP HPLC purified *S. thermophilus* unmaturated crRNA. Following deconvolution the intact mass was obtained = 22,998.51 Da consistent with a theoretical 72 nucleotide mature crRNA with a 5'-OH and 3'phosphate termini.

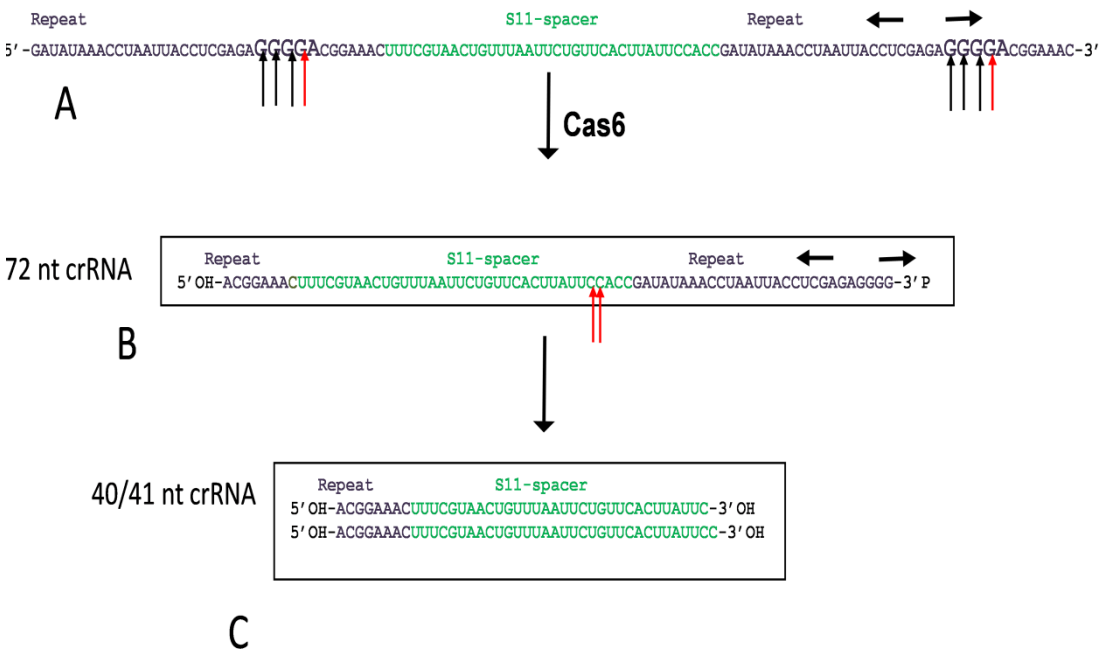
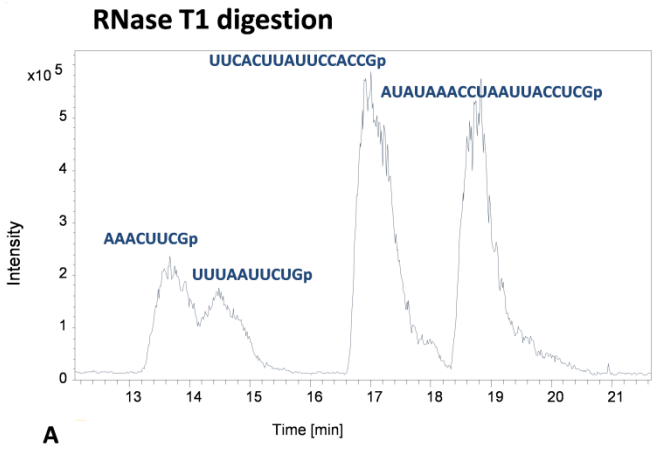
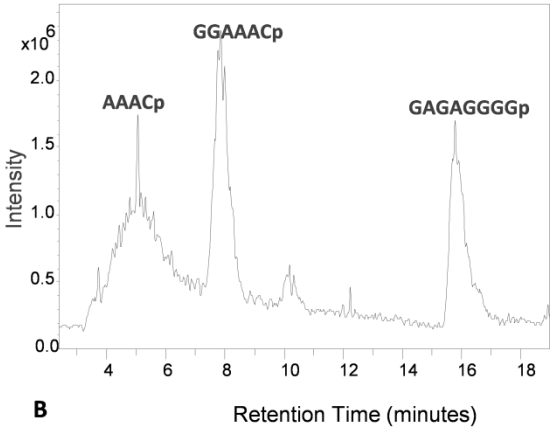


Figure 6.5 Analysis of crRNA processing in the Type III *S. thermophilus* Cascade-crRNA complex A) Pre-crRNA construct containing two 36 nt repeat sequences (purple) *S. thermophilus* and flanking a 36 nt spacer sequence (green). Possible cleavage sites on the repeat (pre-crRNA) that generate RNA product of 22998 Da is indicated by arrows. Red arrows indicate the Cas6 cleavage site revealed in this study. B) 72 nt unmaturated crRNA consisting of 36 nt spacer sequence (green) flanked by 8 nt 5' terminus and 28 nt 3' terminus. C) 40 nt and 41 nt mature crRNAs each consisting 8 nt 5' terminus followed by 32 nt and 33 nt spacer sequences, respectively.



Position	Theoretical mass (Da)	Experimental Mass (Da)	Sequence
A5:G13	2879.725	2879.529	AAACUUUCGp
U20:G29	3163.839	3163.663	UUUAAUUCUGp
U30:G45	5018.969	5018.918	UUCACUUAUUCCACCGp
A46:G65	6359.830	6359.347	AUAUAAACCUAAUUACCUCGp



Position	Theoreti-cal mass (Da)	Experi-mental Mass (Da)	Sequence
G3:C8	2001.244	2001.3	GGAAACp
A50:C53	1310.279	1310.3	AAACp
G65:G72	2747.306	2747.4	GAGAGGGGp

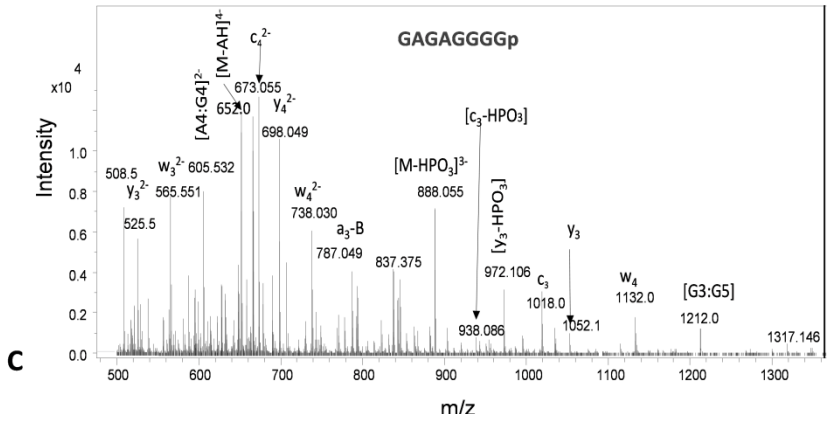


Figure 6.6 RNase mapping of *S. thermophilus* unmaturated 72 nt crRNA A) Base peak chromatogram of RNase T1 digest. RNase T1 cleaves single-stranded RNA 3' of guanine residues. Predominant oligonucleotide peaks to the unmaturated crRNA are highlighted. B) RNase A digest performed in a similar manner. Masses of each oligoribonucleotide are presented in the corresponding tables .C) ESI MS/MS analysis of the oligoribonucleotide GAGAGGGGp. Tandem MS was used to verify the oligoribonucleotide GAGAGGGGp. The predominant fragment ions are highlighted

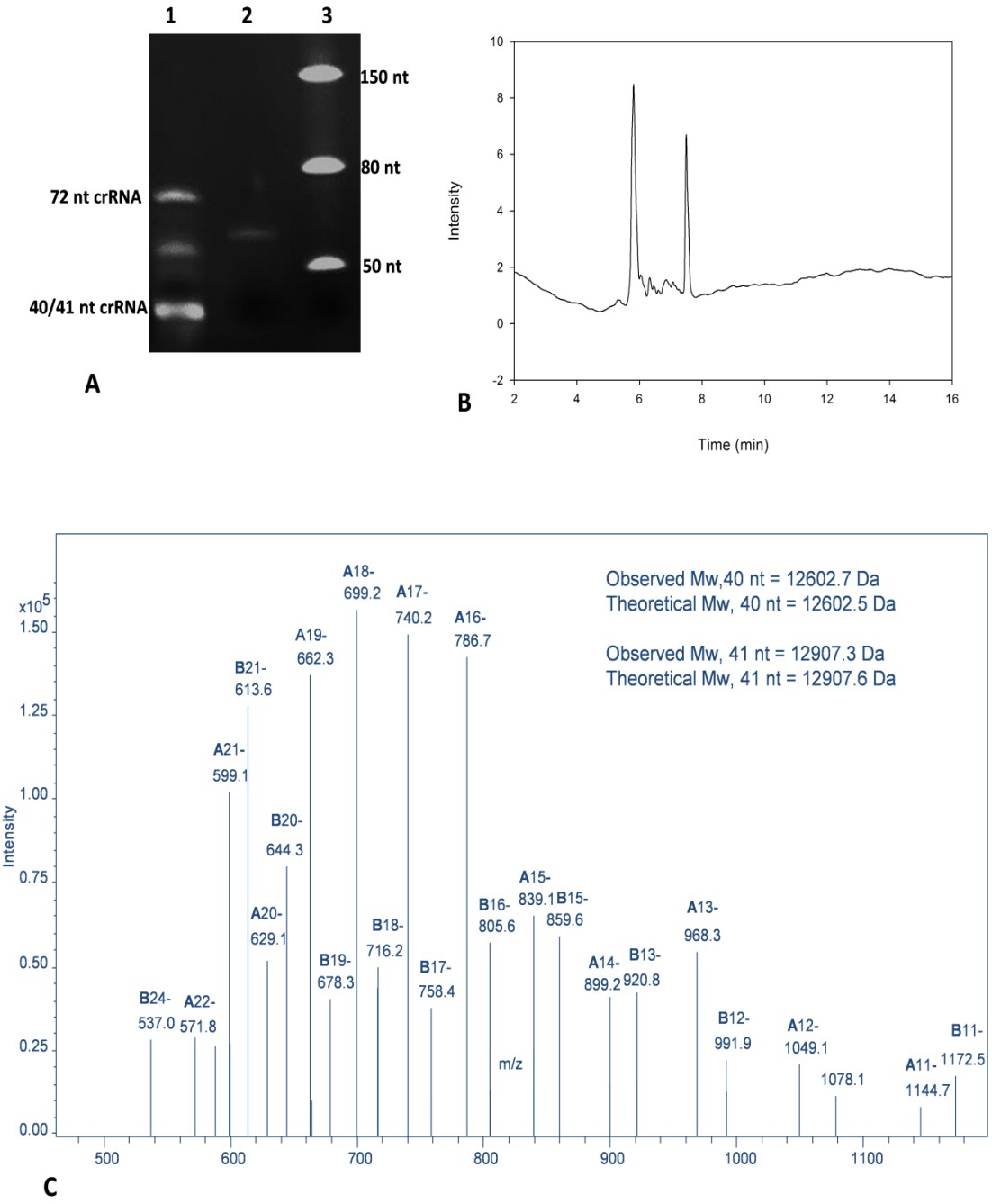
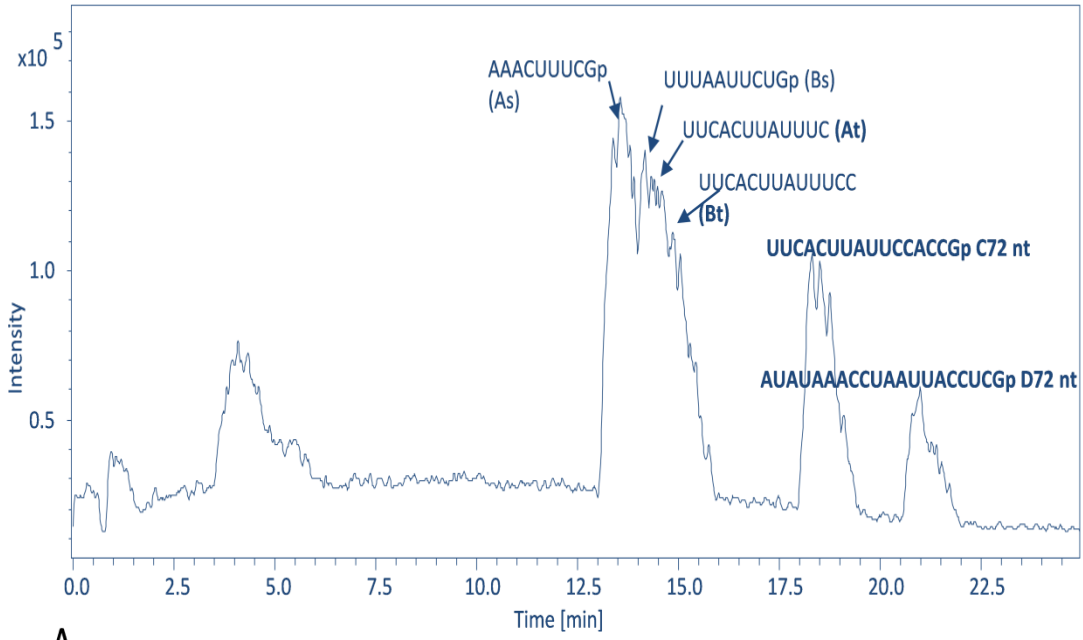


Figure 6.7 Analysis of the crRNA from the Csm2-strep complex crRNA was purified from Csm2-strep+crRNA complex by phenol-chloroform extraction and analysed by (**A**) 8M urea denaturing PAGE. Lane 1, Csm2-strep+crRNA complex; Lane 2, 61 nt RNA; lane 3, ssRNA ladder (NEB). (**B**) IP RP HPLC analysis of the Csm2-strep+crRNA complex. Intact Csm complex (10 ug) was analysed using gradient condition 2 (see Chapter 2.9.1). (C) LC ESI MS analysis showing the MS spectra of IP RP HPLC purified *S. thermophilus* unmaturated (intermediate) crRNA. Following deconvolution the intact masses were obtained = 12602.7 Da and 12907.3 Da.



A

Position	Theoretical mass	Experimental mass	Sequence and assignment
U39:C40	3349.019	3348.906	UUCACUUAUUC (40 nt 3' terminus)
U39:C41	3654.203	3654.400	UUCACUUAUUCC (41 nt 3'terminus)
A5:G13	2879.725	2879.526	AAACUUUCGp (spacer)
U20:G29	3163.839	3163.381	UUUAAUUCUGp (spacer)
U30:G45	5018.969	5018.478	UUCACUUAUUCCACCGp (72 nt)
A46:G65	6359.830	6359.333	AUAUAAACCUAAUUACCUCGp (72 nt)
U30:G45	5000.954	5001.082	UUCACUUAUUCCACCG>p (72 nt)*
A46:G65	6341.815	6342.009	AUAUAAACCUAAUUACCUCG>p (72 nt)*

B

Figure 6.8 RNase T1 digest analysis of maturated crRNAs: A) Chromatogram showing RNase T1 digest of total Csm2-strep+crRNA. Following ESI-MS, oligoribonucleotides were assigned which confirms the presence of 40 nt, 41 nt and the residual uncleaved 72 nt crRNAs: The oligos labelled, As and Bs, were assigned to spacer sequence of 40 nt, 41 nt and residual 72 nt crRNAs. Oligoribonucleotides (labelled C72 nt and D72 nt) were assigned to confirm the presence of residual uncleaved 72 nt crRNA. The termini of 40 nt (At) and 41 nt (Bt) where confirmed by assignment of the unique oligoribonucleotides. B) Table summarizing identified RNase T1 digest oligoribonucleotides. The theoretical and experimental masses are shown for the oligoribonucleotides identified with their corresponding regions in the crRNA. p= 3' phosphate, >p= 2'3' cyclic phosphate.

6.4 Csm3 specifies 3' nuclease cleavage point of Csm5

Recently, it was suggested that Csm3 is a ruler protein, in which multiple Csm3 molecules within the Cas10/Csm complex bind the crRNA with a 6-nucleotide periodicity to function as a ruler that measures the extent of crRNA maturation in *Staphylococcus epidermidis* Type III CRISPR/Cas system (Hatoum-Aslan et al., 2013). In this recent study by Hatoum et al., the tagging of Csm3 with C-terminal His-tag impacted on maturation of crRNA. This suggests a role for Csm3 in the maturation of crRNAs. The combined proteomic and crRNA analysis of the Csm2/Csm3 complexes suggests that the different protein composition in the two complexes reflects the difference in the processing of their pre-crRNAs. In the Csm2+crRNA complex, a 72 nt crRNA is generated which is further processed into the mature crRNAs (40/41 nts). In theCsm3+crRNA complex, the pre-crRNA is only processed into a 72 nt crRNA. The proteomic analysis revealed that Csm5 protein is more abundant in the Csm2+crRNA complex compared to the Csm3+crRNA complex suggesting thatCsm5 is required for the processing of the 'unmaturated' 72 nt crRNA into 'maturated' 40/41 nt crRNAs.

The data suggests that the tagging of Csm3 with Strep tag affected the recruitment of Csm5. In the Csm3+crRNA, Csm5 stably associates with the complex even though Csm3 is tagged and was in fact no different to composition of Csm2-RNA which also contains Csm5 (Figure 6.3 and Table 6.3). This suggests that tagging of Csm3 disrupts the binding of Csm5 only when the pre-crRNA is present. Therefore the dissociation of Csm5 is not only due the presence of C-terminal tag on Csm3 but a combinatorial effect of the tag and presence of crRNA. It has been shown that multiple copies of Csm3 protein binds in a periodic manner on the crRNA and potentially protects crRNA from cleavage by the 3' nuclease responsible for maturation (Hatoum-Aslan et al., 2013). It can be hypothesized that Csm5 binds and cleaves on a site close to one of the potential binding sites of Csm3 and that the presence of the tag may shield intermediate (72 nt) crRNA from binding and cleavage by Csm5. An alternative explanation could be that the presence of the tag resulted in addition of multiple copies of Csm3 to the complex thereby shielding potential site of Csm5 cleavage. Interestingly, Hatoum et al., also showed that overexpression/additional copies of

Csm3 resulted in a shift of crRNA size to the larger species (Hatoum-Aslan et al., 2013) and therefore suggesting that Csm3 protects crRNA from 3'end cleavage. However, these explanations do not adequately explain why Csm5 associate with Csm3-crRNA and not Csm3+crRNA. The presence of Csm5 in Csm2-crRNA and Csm3-crRNA (Table 6.2) suggests that crRNA is not required for the recruitment of Csm5. This is also in agreement with the observation by Hatoum et al. that the Cas10/Csm complex in *S. epidermidis* forms in the absence of crRNA. Csm3 is, on the other hand, is required for the formation of Cas10/Csm complex (Hatoum-Aslan et al., 2013). One hypothesis for the lack of association of Csm5 with Csm3+crRNA is that the presence of the tag and crRNA imposes an artificial configuration that excludes/destabilizes Csm5 from the complex and therefore prevents Csm5 from acting. Note that in Csm2+crRNA that Csm3 (which is not tagged) did not prevent the maturation of the intermediate (72 nt) crRNA into 40/41 nt crRNA. This study in conjunction with previous studies suggests that Csm3 as a ruler protein may indeed function to specify the cleavage point for the 3' nuclease by protecting regions from the 3' nuclease. This hypothesis is supported by the maturation of crRNA in the Csm2+crRNA (which contains untagged Csm3) in which the intermediate 72 nt crRNA was processed into 40/41 nt crRNAs.

Conclusions

In this Chapter, the crRNA processing in a Type III CRISPR/Cas complex in *Streptococcus thermophilus* was studied. Following overexpression and purification of the Cascade-crRNA complex, denaturing RNA chromatography in conjunction with electrospray ionization mass spectrometry (ESI-MS) was used to analyse the crRNAs. The results identified an unmature (intermediate 72 nt) crRNA and mature crRNAs (40/41 nt).The cleavage site in the pre crRNA was at the base of the hairpin loop resulting in the generation of 72 nt crRNA intermediate with a 8 nt 5' handle, 36 nt spacer and 28 nt 3' handle. The chemical nature of the 5' and 3' termini of the 72 nt unmaturated crRNA was determined to be 5'OH and 3'P. Analysis of the 40/41 nt

mature crRNA revealed that these crRNAs share the same 5'OH termini and are processed from the 3' end of the unmatured 72 nt crRNA Moreover, the results show that the maturated crRNAs have 3'OH termini. Therefore demonstrating that the catalytic mechanism of crRNA cleavage in processing the 72 nt crRNA from the pre-crRNA is distinct from that used in further processing the 72 nt into the mature 40/41 nt crRNAs. In addition these results also suggest that the protein composition of the Cascade-crRNA complex effects crRNA maturation which is affected by the association of Csm5 within the Cascade complex. Furthermore it is proposed that tagging of Csm3 in the presence of crRNA may adopt an artificial configuration that excludes Csm5 from the complex. In the future, it will be interesting to further investigate the mechanism of Csm5 action in the maturation of crRNAs in Type III CRISPR/Cas system in *S. thermophilus*. *In vitro* assays aimed at dissecting the function Csm5 in the presence of its substrate crRNA will also provide further insight into its mechanism of action.

Chapter VII

Final Discussion and Future work

In this Thesis, large scale tandem affinity purification in conjunction with mass spectrometry was used to investigate protein-protein/protein-RNA interactions in *Saccharomyces cerevisiae* and the biogenesis of Type III CRISPR/Cas complex in *Streptococcus thermophilus*. The advantage of using tandem affinity purification-mass spectrometry approach lays in its throughput capabilities and its ability to potentially harness structural information which cannot be obtained from genetic sequence or protein sequencing. Protein post-translational modifications are proving to be an important aspect of protein-protein interactions and biology of the cell in general. These modifications can be mapped using tandem affinity purification-MS (TAP-MS) approach. However, interference of affinity tags can be a potential hindrance in obtaining full information about the network of protein-protein/protein-RNA interactions. This is also compounded by transient interactions which might not be stable under standard affinity purification conditions. Tandem affinity purification has an advantage over other purification strategies because the purification conditions are usually milder. Specificity of interaction which can be a problem in affinity capture methodologies, such as Co-IPs, is usually averted in TAP by the use of TEV protease. TEV protease cleaves the baits off the affinity matrix, thereby ensuring the gentle elution of proteins specifically interacting with the bait protein and exclusion of non-specific proteins associating with the affinity matrix.

This chapter concludes on the successes achieved in the presented studies and discusses limitations of the approaches used and directions for future work.

7.1 Engineering a novel TAP tag for the analysis of Prp5 and its interacting proteins

In chapter three, a novel *Saccharomyces cerevisiae* Prp5-TAP strain was successfully engineered by fusing a novel TAP tag, p54-protA to the C-terminus of Prp5.

Following confirmation of TAP tag orientation and expression by PCR and Western blot, respectively, tandem affinity purification was performed. The proteins co-purifying with Prp5 were identified. The data suggests that Prp5 interacts with spliceosomal proteins and possibly proteins involved in transcription and mRNA transport.

In this Chapter, the results show that Prp5 interacts with U2 snRNP proteins, including Hsh155 and Rse1p, but does not interact with proteins found downstream of splicing. No proteins from the U4/U6.U5 complexes were identified in this study suggesting that Prp5p only interacts with U2 snRNP *in vivo*. Proteins of U1 snRNP were not identified. However, proteins involved in transcription and mRNA transport, such as Rpa43p and Npl3p were identified. Interestingly, the protein Kem1p, a 5' to 3' exonuclease, was also identified, suggesting a novel interaction with Prp5. Kem1p has been implicated in pathways such as, cytoplasmic bodies where the function in mRNA degradation, telomere maintenance and ribosomal RNA maturation (Askree et al., 2004; Geerlings et al., 2000; Sheth and Parker, 2003). This suggests that splicing factor, Prp5p, may also be involved in mRNA degradation pathways.

This study also revealed an interaction between Prp5p and methylated Npl3p. For the first time, the type of arginine methylation was characterised in Npl3p. Specifically, asymmetric dimethylation (aDMA) of arginine at positions Arg363 and Arg384 were identified in this study.

Although the analysis of RNA components of the purified Prp5 using RNA denaturing PAGE in this study revealed a band of about 200 nt, the northern analysis showed that it is not U2, U4, U5 and U6. In the future, the interaction of Prp5 with potential small RNAs will be further investigated using denaturing PAGE analysis/Northern blotting in conjunction with mass spectrometry with the aim to identify the RNA associated with Prp5.

7.2 Analysis of Lsm proteins and their Interactions with the Spliceosome and other coupled pathways

Chapter four presented the analysis of Lsm complexes in *Saccharomyces cerevisiae* using TAP-MS. The analysis of Lsm1-TAP and Lsm8-TAP led to an exhaustive distinction between two complexes formed by Lsm proteins. The presented studies showed that Lsm protein form two distinct complexes, namely, Lsm1-7p and Lsm2-8p. Lsm1-TAP co-purified with Lsm1-7p proteins while Lsm8-TAP co-purifies with Lsm2-8p proteins. This is consistent with earlier studies which demonstrated that Lsm proteins form two distinct complexes, Lsm1-7p and Lsm2-8p (Tharun et al., 2000).

The results also show that Lsm1-7p co-purifies with deadenylase-decapping complex in contrast to Lsm8-TAP, which suggests that Lsm1-7p participate in deadenylation-dependent-decapping mRNA decay while Lsm2-8p does not. Lsm1-TAP was also shown to co-purify with nonsense mediated decay pathway, while Lsm8-TAP does suggesting that Lsm1-7p also participate in NMD while Lsm2-8p does not. However, Lsm8-TAP co-purifies with 5'→3' exoribonuclease, Xrn1p (Kem1p) and the decapping promoting factor, Eap1p along. It is significant that Lsm1-7p did not co-purify with Eap1p but co-purified with a distinct decapping enhancer, Edc3p involved in deadenylation-independent-decapping mRNA decay. This suggests that Lsm1-7p and Lsm2-8p participate in two distinct deadenylation-independent-decapping mRNA degradation pathways utilizing, Edc3p and Eap1p, respectively. The result suggests that Lsm2-8p participate in deadenylation-independent decapping mRNA. This is consistent with previous observation that Lsm2-8p interacts with nuclear poly(A)mRNA and also promotes decapping of pre-mRNA apparently by deadenylation independent mechanism (reviewed by (Beggs, 2005; Kufel et al., 2004)). The result further highlights differences in the two Lsm complexes and suggests that their pathways in mRNA degradations are distinct. It also suggests the versatility of Lsm1-7p in mRNA degradation pathways, namely, deadenylase-dependent, deadenylase-independent-decapping and nonsense mediated decay pathways.

Furthermore, the results obtained in Chapter 4 show that Lsm2-8p complexes co-purify with U4/U5.U6 complex which is consistent with the model that the Lsm2-8p complex is associated with U6 snRNP and plays a role in the spliceosome. Lsm1-7p, on the hand, did not co-purify with these splicing factors. However, Lsm1-7p co-purify with Prp43p while Lsm2-8p complex does not. This is interesting because, Prp43p, catalyses the removal of U2, U5 and U6 snRNPs from the post-splicing lariat-intron ribonucleoprotein complex and processing of rRNA. This suggests that the role of Lsm1-7p may not be restricted to on deadenylation-decapping mRNA degradation but may participate in other RNA processing events.

Transcription factors/DNA repair proteins RNA polymerase subunits, Rpa49p, Rpa135p, Rpc40p, Def1p and repair proteins, including, Hta1p and 2p cap-binding/poly(A)mRNA binding proteins including Pab1p, Cdc33p (elF4E), Tif4631p (elF4G1), Tif4632p (elF4G2), THO/TREX factors and snoRNP factors also co-purified with the Lsm complexes. This suggests that Lsm complexes form a dynamic multi-protein complex with these factors during transcription, consistent with the proposed model that mRNA processing and export pathways are coupled to transcription(Masuda et al., 2005; Reed, 2003; Reed and Hurt, 2002). The formation of this dynamic multi-protein complex can be speculated to be an evolutionary mechanism for maximizing and coordinating the enzymatic resources of the cell during the different stages of RNA processing. Lsm complexes were also shown to co-purify with methylated Npl3p which is involved in mRNA nuclear export to the cytoplasm.

7.3 Analysis of Hsh155/U2 snRNP complex using TAP-Mass spectrometry approach

In Chapter 5, Hsh155-TAP was analysed using TAP in conjunction with mass spectrometry. Hsh155 is part of the U2 snRNP, specifically, the U2 SF3b sub-complex. The presented studies shows that Hsh155-TAP co-purifies with U2 SF3b proteins including Rse1p, Cus1p, Hsh155p and the SF3a proteins, including Lea1p and Prp9p which suggests Hsh155 co-purifies with U2 snRNP complexes.

The results in Chapter 5 show that Hsh155-TAP co-purifies with NTC/NTC-related proteins. This suggests that U2 snRNP interacts with NTC/NTC-related proteins during spliceosome assembly. Although, NTC/NTC-related proteins are known to play a role in splicing, this is the first time interaction with U2 snRNP is suggested by affinity purification-MS studies. NTC/NTC-related proteins are thought to specify the stable association of U5 and U6 snRNA with the pre-mRNA (Chan et al., 2003). Analysis of U4/U5.U6 proteins co-purifying with Lsm1, 2, and 4-TAPs in Chapter 4suggests that U5 or U6 complexes do not interact with the NTC/Prp19 complex. The co-purification of U2 snRNP with NTC/NTC-related proteins suggests a function for the U2 snRNP in facilitating association of U snRNAs with pre-mRNAs.

This study also examined the interaction between Hsh155-TAP and U1 snRNP but was unable to establish such association. This suggests that, at least, SF3b components if not U2 snRNP does not associate with U1 snRNP. Interaction between Hsh155-TAP and U4/U6.U5 snRNP was also examined and again such interaction was not established.

This study identified a number of potential novel interactions between U2 snRNP and transcription factors, which suggest coupling of the two RNA processing pathways. This is also consistent with the result and observation in Chapter 4 that splicing may be coupled to transcription in *S. cerevisiae*.

The identification of interaction between Spt5p and U2 snRNP in this Chapter provides evidence for the functional coupling of transcription and splicing in yeast. Although, previous *in vitro* studies suggested that coupling of splicing with transcription can take place in vivo, there is sparse evidence that such events occurs in yeast. This is the first time an *in vivo* interaction between splicing factor, U2 snRNP and transcription factors are demonstrated in yeast. The further identification of other splicing factors and chromatin associating factors such as, Def1p, Rpa49p and Set1p further strengthens the argument for coupling of transcription and splicing in yeast. This study also identifies interaction U2 snRNP with TREX factors suggesting that a highly-dynamic multi-complex involving transcription, splicing and mRNA export complexes exist for co-ordinated processing of RNAs. It is interesting that some TREX factors and transcription factors co-purified with U2 snRNP core protein,

Hsh155 as shown in Chapter 5. Since U2 snRNP is restricted to the nucleus, interaction with transcription factors and translation factors can only take place in the nucleus. Identification of the deadenylation and decay factor, Sup35p (Table 5.6) and Cap/poly(A) binding proteins suggests a role for U2 snRNP in nuclear mRNA decay (Table 5.7). The identification of nucleolar proteins (Table 5.7) will suggest that U2 is involved other RNA processing apart from mRNAs. In summary the data generated in this Chapter is consistent with the model that transcription, splicing and mRNA transport are functionally coupled mechanisms. It also suggests a role for U2 snRNP/splicing factors in nuclear mRNA turnover. However, further studies are needed to understand the functional nature of these interactions and coupling.

7.4 Overview of coupling of steps in gene expression

In all three studies involving Prp5-TAP, Lsm-TAP and Hsh155-TAP, splicing factors co-purified with transcription, capping/and polyadenylation factors, mRNA degradation, TREX and in some cases, snoRNP factors. These suggest that splicing factors form a multi-complex with factors involved in transcription, mRNA decay, mRNA export and RNA processing or that these splicing factors are present during these RNA processing events. The result is consistent with the model that transcription, splicing, mRNA export and event downstream processing events such as mRNA degradation are coupled events (Reed, 2003; Reed and Hurt, 2002). The results are also consistent with observations in previous studies. For instance, previous studies have shown that yeast TREX complex containing the THO transcription elongation complex, plays a role in coupling transcription to mRNA export by facilitating the recruitment of the mRNA export proteins Sub2 and Yra1 to nascent transcripts (Fischer et al., 2002; Jimeno et al., 2002; Maniatis and Reed, 2002; Reed, 2003). There is also evidence that 3'end processing machinery (polyadenylation machinery) of the transcript is physically linked to splicing (spliceosome).

These suggest that the various steps that constitute gene expression are coupled to each other via an extensive network of physical and functional cellular machineries that carry out each step of gene expression (Bentley, 2002; Hirose and Manley, 2000;

Maniatis and Reed, 2002; Reed, 2003). It has been suggested that this coupling may be a quality control mechanism which ensures that each step occur efficiently only if the proper contacts are made to the other steps in the pathway (Masuda et al., 2005).

In humans, splicing occurs co-transcriptionally and mRNA export is coupled to transcription indirectly through splicing in contrast to yeast where direct transcription-coupled mechanism is thought to occur (Masuda et al., 2005). Although, earlier studies suggests the possibility of direct coupling of mRNA export and splicing to transcriptionally in yeast it was not clear if splicing occurs co-transcriptionally and in what order this coupling steps take place (Görnemann et al., 2005; Lacadie et al., 2006; Reed, 2003; Reed and Hurt, 2002). In other words, it was not clear if splicing occurs during the process of transcription and if mRNA exports machinery is recruited to the transcript prior to or after co-transcriptional splicing. In this study, the co-purification of splicing factors with transcription factors (RNA Pol II), polyadenylation factors (including Pab1p) and mRNA export factors (TREX factors) suggest that splicing occurs co-transcriptionally in yeast, and that TREX factors are present during splicing event. This is consistent with a recent study where it was shown that some of the transcripts are spliced before the transcription termination in budding yeast (Carrillo Oesterreich et al., 2010).

This thesis supports a model which proposes that splicing in yeast occur co-transcriptionally and that mRNA export factors are recruited during co-transcriptional splicing event. It is also consistent with the model that the multi-steps of gene expression are physically and functionally linked through the formation of a multi-complex comprising of all the cellular machineries involved in each step of gene expression.

7.5 Studying CRISPR RNP Biogenesis

Chapter 6 examined the processing of CRISPR RNA and the roles of Cas proteins in CRISPR biogenesis in a type III CRISPR/Cas system in *Streptococcus thermophilus* using various analytical techniques including mass spectrometry. Although a number

of CRISPR/Cas systems have beencharacterised, there remains a considerable challenge in delineating mechanisms of CRISPR action owing to the diversity observed in a number of CRISPR families. In this Chapter the CRISPR RNA was shown to be important for the recruitment of Cas6 to the CRISPR/Cas complex. This study also, for the first time, suggests that the Cas/Csm protein, Csm5 is required for the 3'end processing of CRISPR RNA.

In this study, the architecture of the intermediate (unmaturated) and maturated CRISPR RNAs of *Streptococcus thermophilus* was determined. This study demonstrates that the intermediate CRISPR RNA is of 72 nt long and is cleaved at the base of the pre-crRNA hair-loop structure, 8-nt bases away from the 5'end of the spacer sequence (Figure 7.1A). Analysis of the maturated crRNAs reveals that the maturated crRNAs retains the 5'end from the intermediate crRNA while the 3'end is processed. It also reveals that the processing of the 3'end is somewhat degenerate, producing two maturated crRNAs of 40 and 41 nts in length having identical 5'OH and sequence but differs from each other by 1 nt from the 3'end (Figure 7.1B). The chemical nature of the 5' and 3'ends of the intermediate and maturated crRNAs were also determined. The 5' and 3'end of the intermediate crRNA was determined to be 5'OH and 3'P, respectively. However, the 5' and 3'end of the maturated crRNAs were determined to have OH at both the 5' and 3'ends. These results demonstrate that the catalytic mechanism of crRNA cleavage in processing the 72 nt crRNA from the pre-crRNA is distinct from that used in further processing the 72 nt into the mature 40/41 nt crRNAs.

7.6 Advantages, Limitations and Future work

It is clear that proteomic data generated from TAP-MS approaches provide valuable insight into the protein-protein interactions, especially in very dynamic complexes such as spliceosomes. However, this approach sacrifices the structural and functional details that are possible through molecular biology approaches. Large and dynamic complexes such as the spliceosome, on the other hand, are not amenable to structural techniques such as NMR and crystallization. And where such crystal

structures are obtained for a dynamic complex it only provides a snapshot which may represent an intermediate of such a complex. The combination of proteomics data obtained by TAP-MS with images produced by cryo-electron microscopy and AFM can be a powerful tool in the visualisation and production of a comprehensive protein-protein/protein-RNA interaction network for not only the dynamic spliceosome but also the mRNA processing/export network. Recently, a laboratory using negative strain cryo-EM looked at TAP-purified complexes from both pre-40S and pre-60S ribosomes to study the dynamics of structural changes during the transition between different late ribosomal maturation stages(Schafer et al., 2006). In the future, it will be interesting if this approach is employed in capturing snapshots of the different stages of the spliceosome assembly or mRNA processing events from transcription to mRNA export.

Another complementary approach for future work will be to integrate absolute/relative quantification into the TAP-MS strategy. There is an increasing demand for absolute protein abundance values for input into models. Recently, QconCAT technology is gaining grounds in the absolute multiplexed protein quantification (Brownridge et al., 2012). The QconCAT acts as a source of internal standards and enables parallel absolute/relative quantification of multiple proteins. QconCATs are created by concatenating peptide sequences taken from the target proteins into artificial proteins (see Figure 7.1) and typically applied in targeted proteomic workflows and so benefit from the greater sensitivity and wider dynamic ranges (Brownridge et al., 2012). It will be interesting if this quantitation workflow is integrated into TAP-MS analysis of splicing and mRNA processing complexes, such as the Lsm complexes. This approach will enable the determination of the relative stoichiometry of the protein complexes.

A second complex of interest that can be analysed using absolute/relative quantitative approach is the Type III CRISPR/Cas complex proteins of *S. thermophilus*. In this study, the stoichiometry of the CRISPR complex was inferred from semi-quantitative data. The complex of interest for quantitation will be the Type III Cas proteins. Three peptide sequences from each target protein generated from tryptic digest are chosen. An artificial gene is created by concatenating nucleotide sequences encoding these peptides. This gene will be expressed in *E. coli*

in the presence of heavy N (N^{15}) to produce a heavy labelled protein. This protein is digested with trypsin, mixed with sample protein complex and analysed by mass spectrometry. The mass spectra of the heavy and light isotope peptides from the internal control and sample protein complex, respectively are compared which will enable the determination of the complex stoichiometry (also see Figure 7.1).

Tandem affinity purification is not currently considered a powerful tool in studying transient or labile protein-protein/protein-nucleic acid interaction with lower Kd values of the lower nM range(Oeffinger, 2012; Schlosshauer and Baker, 2004), even though it is considerable valuable in proteomics. To address these limitations and capture more transient interactions, shorter protocols utilizing single step of purification instead of two could be adopted. In the presented studies, in order to capture more transient interactions and also remove carry-over TEV protease, TAP was performed without 10 mM DTT usually present in the standard TAP Calmodulin binding and washing buffers of the second step of purification. During the experiments in the presented studies it was observed that 10 mM DTT has the potential of dissociating labile or transient interacting partners. Using the simple step under suitable cryogenic conditions of sample preparation will be a simpler and effective way of capturing transient interactions in dynamic molecular machinery like the spliceosome.

There is a growing interest in the use of GFP tag as an affinity purification tag. This is because the tag has been used widely in the study of many proteins and so it will be easy to obtain an already tagged target protein of interest. Its desirability also stems from the fact that it can be visualized in cells, whether by fluorescence (e.g. GFP) or immunofluorescence. While this is not an essential requirement for successful AP, it provides the possibility to directly combine imaging with quantitative proteomics technology. By inserting a TEV cleavage sequence between the GFP tag and bait protein it is also possible to take advantage of the mild elution and reduced background contaminants afforded by TEV cleavage. It will be interesting if the use of GFP is explored in the study of the dynamic multicomplex mediating the mRNA processing pathways.

Finally, little is known of the mechanism of Interference of *S. thermophilus* Type III CRISPR/Cas complex. It will be interesting if this is investigated. A proposed approach to this study is using a combination of *in vitro* single-molecule techniques. For instance, it will be interesting to understand the mechanism of the R-loop formation. This can be done using magnetic tweezers to measure the R-loop formation as a function of torque. Also, the roles of CRISPR RNA (crRNA) sequence in R-loop initiation, formation and stability can be teased out using this approach.

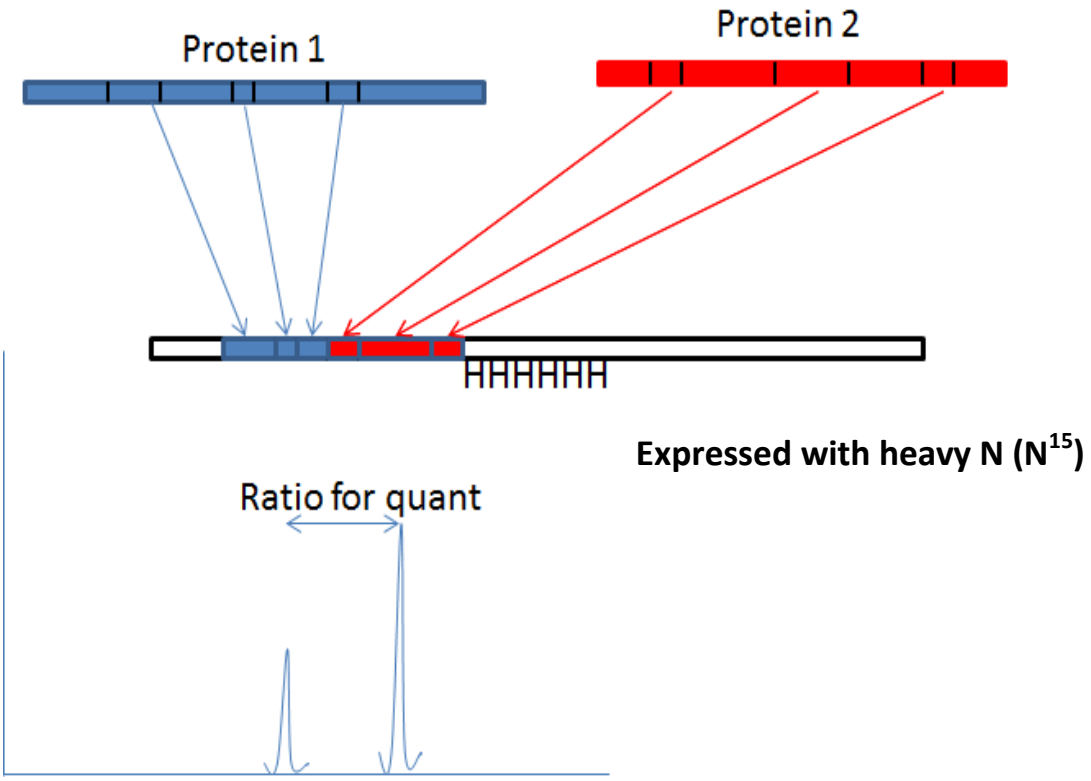


Figure 7.1 Diagrammatic illustration of the QconCAT design. Protein 1 (blue) and 2 (red) represent sample proteins to be quantified. Sample peptides from each protein are chosen and an artificial gene created by concatenating 'genes' encoding each sample peptide. The artificial gene is expressed in heavy labelled medium to produce an artificial protein/internal control, which contains sample peptides from sample proteins (3 blue and 3 reds, for protein 1 and 2, respectively). The peaks represent the spectral abundance of heavy-labelled and light versions of a sample peptide and can be used to determine the absolute abundance of desired sample protein.

BIBLIOGRAPHY

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403-412.

Abu Dayyeh, B.K., Quan, T.K., Castro, M., and Ruby, S.W. (2002a). Probing interactions between the U2 small nuclear ribonucleoprotein and the DEAD-box protein, Prp5. J Biol Chem *277*, 20221-20233.

Abu Dayyeh, B.K., Quan, T.K., Castro, M., and Ruby, S.W. (2002b). Probing interactions between the U2 small nuclear ribonucleoprotein and the DEAD-box protein, Prp5. J Biol Chem *277*, 20221-20233.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. EMBO J *18*, 5789-5802.

Agell, N., Bachs, O., Rocamora, N., and Villalonga, P. (2002). Modulation of the Ras/Raf/MEK/ERK pathway by Ca(2+), and calmodulin. Cell Signal *14*, 649-654.

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J.C., Mann, M., and Lamond, A.I. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. EMBO J *19*, 6569-6581.

Albert, I., and Albert, R. (2004). Conserved network motifs allow protein-protein interaction prediction. Bioinformatics *20*, 3346-3352.

Allain, F.H., Bouvet, P., Dieckmann, T., and Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. EMBO J *19*, 6870-6881.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J *18*, 5399-5410.

Annan, R.S., and Zappacosta, F. (2005). Protein posttranslational modifications: phosphorylation site analysis using mass spectrometry. Methods Biochem Anal *45*, 85-106.

Archambault, V., Chang, E.J., Drapkin, B.J., Cross, F.R., Chait, B.T., and Rout, M.P. (2004). Targeted proteomic study of the cyclin-Cdk module. Mol Cell *14*, 699-711.

Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proc Natl Acad Sci U S A *94*, 11798-11802.

Askree, S.H., Yehuda, T., Smolikov, S., Gurevich, R., Hawk, J., Coker, C., Krauskopf, A., Kupiec, M., and McEachern, M.J. (2004). A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. Proc Natl Acad Sci U S A *101*, 8658-8663.

Balbo, P.B., and Bohm, A. (2007). Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. Structure *15*, 1117-1131.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Bartel, P.L., Roecklein, J.A., SenGupta, D., and Fields, S. (1996). A protein linkage map of Escherichia coli bacteriophage T7. Nat Genet *12*, 72-77.

Beadle, G.W., and Tatum, E.L. (1941). Genetic Control of Biochemical Reactions in Neurospora. Proc Natl Acad Sci U S A *27*, 499-506.

Beckouet, F., Labarre-Mariotte, S., Albert, B., Imazawa, Y., Werner, M., Gadal, O., Nogi, Y., and Thuriaux, P. (2008). Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. Mol Cell Biol *28*, 1596-1605.

Bedford, M.T. (2007). Arginine methylation at a glance. Journal of cell science *120*, 4243-4246.

Beelman, C.A., and Parker, R. (1995). Degradation of mRNA in eukaryotes. Cell *81*, 179-183.

Beggs, J.D. (2005). Lsm proteins and RNA processing. Biochem Soc Trans *33*, 433-438.

Behrens, S.E., Tyc, K., Kastner, B., Reichelt, J., and Lührmann, R. (1993). Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol Cell Biol *13*, 307-319.

Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol *14*, 336-342.
Bickle, T.A., and Kruger, D.H. (1993). Biology of DNA restriction. Microbiol Rev *57*, 434-450.
Boelens, W., Scherly, D., Beijer, R.P., Jansen, E.J., Dathan, N.A., Mattaj, I.W., and van Venrooij, W.J. (1991). A weak interaction between the U2A' protein and U2 snRNA helps to stabilize their complex with the U2B" protein. Nucleic Acids Res *19*, 455-460.
Bond, A.T., Mangus, D.A., He, F., and Jacobson, A. (2001). Absence of Dbp2p alters both nonsense-mediated mRNA decay and rRNA processing. Mol Cell Biol *21*, 7366-7379.
Bork, P., Downing, A.K., Kieffer, B., and Campbell, I.D. (1996). Structure and distribution of modules in extracellular proteins. Q Rev Biophys *29*, 119-167.
Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Séraphin, B. (2000). A Sm-like protein complex that participates in mRNA degradation. EMBO J *19*, 1661-1671.
Brenner, T.J., and Guthrie, C. (2005). Genetic analysis reveals a role for the C terminus of the Saccharomyces cerevisiae GTPase Snu114 during spliceosome activation. Genetics *170*, 1063-1080.
Brosi, R., Groning, K., Behrens, S.E., Lührmann, R., and Kramer, A. (1993). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science *262*, 102-105.
Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.
Brow, D.A. (2002). Allosteric cascade of spliceosome activation. Annu Rev Genet *36*, 333-360.
Brownridge, P.J., Harman, V.M., Simpson, D.M., and Beynon, R.J. (2012). Absolute multiplexed protein quantification using QconCAT technology. Methods Mol Biol *893*, 267-293.
Bruckner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci *10*, 2763-2788.
Budzikiewicz, H., and Grigsby, R.D. (2006). Mass spectrometry and isotopes: a century of research and discussion. Mass Spectrom Rev *25*, 146-157.
Busk, O.L., Ndossi, D., Verhaegen, S., Connolly, L., Eriksen, G., Ropstad, E., and Sorlie, M. (2011). Relative quantification of the proteomic changes associated with the mycotoxin zearalenone in the H295R steroidogenesis model. Toxicon *58*, 533-542.
Cáceres, J.F., and Kornblihtt, A.R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet *18*, 186-193.
Calero, G., Wilson, K.F., Ly, T., Rios-Steiner, J.L., Clardy, J.C., and Cerione, R.A. (2002). Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. Nat Struct Biol *9*, 912-917.
Carey, M.F., Peterson, C.L., and Smale, S.T. (2009). Chromatin immunoprecipitation (ChIP). Cold Spring Harb Protoc *2009*, pdb prot5279.
Carr, S.A., Annan, R.S., and Huddleston, M.J. (2005). Mapping posttranslational modifications of proteins by MS-based selective detection: application to phosphoproteomics. Methods Enzymol *405*, 82-115.
Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol Cell *40*, 571-581.
Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev *22*, 3489-3496.
Caspary, F., Shevchenko, A., Wilm, M., and Séraphin, B. (1999). Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. EMBO J *18*, 3463-3474.
Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p-associated complex in spliceosome activation. Science *302*, 279-282.

Chavez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J *19*, 5824-5834.
Chen, C.Y., Gherzi, R., Ong, S.E., Chan, E.L., Raijmakers, R., Pruijn, G.J., Stoecklin, G., Moroni, C., Mann, M., and Karin, M. (2001). AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell *107*, 451-464.
Chothia, C., and Janin, J. (1975). Principles of protein-protein recognition. Nature *256*, 705-708.
Chowdhury, A., Mukhopadhyay, J., and Tharun, S. (2007). The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. RNA *13*, 998-1016.
Cloutier, S.C., Ma, W.K., Nguyen, L.T., and Tran, E.J. (2012). The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription. J Biol Chem *287*, 26155-26166.
Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S., and Krogan, N.J. (2007). Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol Cell Proteomics *6*, 439-450.
Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. (2000). A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics *154*, 557-571.
Conti, E., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Curr Opin Cell Biol *17*, 316-325.
Costanzo, M.C., Hogan, J.D., Cusick, M.E., Davis, B.P., Fancher, A.M., Hodges, P.E., Kondu, P., Lengieza, C., Lew-Smith, J.E., Lingner, C*., et al.* (2000). The yeast proteome database (YPD) and Caenorhabditis elegans proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. Nucleic Acids Res *28*, 73-76.
Cristea, I.M., Williams, R., Chait, B.T., and Rout, M.P. (2005). Fluorescent proteins as proteomic probes. Mol Cell Proteomics *4*, 1933-1941.
Crowder, S.M., Kanaar, R., Rio, D.C., and Alber, T. (1999). Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. Proc Natl Acad Sci U S A *96*, 4892-4897.
Damelin, M., and Silver, P.A. (2000). Mapping interactions between nuclear transport factors in living cells reveals pathways through the nuclear pore complex. Mol Cell *5*, 133-140.
Das, B., Butler, J.S., and Sherman, F. (2003). Degradation of normal mRNA in the nucleus of Saccharomyces cerevisiae. Mol Cell Biol *23*, 5502-5515.
Das, B., Guo, Z., Russo, P., Chartrand, P., and Sherman, F. (2000a). The role of nuclear cap binding protein Cbc1p of yeast in mRNA termination and degradation. Mol Cell Biol *20*, 2827-2838.
Das, R., Zhou, Z., and Reed, R. (2000b). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol Cell *5*, 779-787.
de la Cruz, J., Kressler, D., and Linder, P. (1999). Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends Biochem Sci *24*, 192-198.
Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol Cell Biol *26*, 5528-5543.
Deka, P., Rajan, P.K., Perez-Canadillas, J.M., and Varani, G. (2005). Protein and RNA dynamics play key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human Cstf-64 protein. J Mol Biol *347*, 719-733.

Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835-845.
Dickman, M.J., and Hornby, D.P. (2006). Enrichment and analysis of RNA centered on ion pair reverse phase methodology. RNA *12*, 691-696.
Dix, I., Russell, C.S., O'Keefe, R.T., Newman, A.J., and Beggs, J.D. (1998). Protein-RNA interactions in the U5 snRNP of Saccharomyces cerevisiae. RNA *4*, 1675-1686.
Doneanu, C.E., Gafken, P.R., Bennett, S.E., and Barofsky, D.F. (2004). Mass spectrometry of UV-cross-linked protein-nucleic acid complexes: identification of amino acid residues in the single-stranded DNA-binding domain of human replication protein A. Anal Chem *76*, 5667-5676.
Draper, D.E. (1999). Themes in RNA-protein recognition. J Mol Biol *293*, 255-270.
Dunoyer, P., Lecellier, C.H., Parizotto, E.A., Himber, C., and Voinnet, O. (2004). Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. Plant Cell *16*, 1235-1250.
Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., Li, M*., et al.* (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. Molecular systems biology *3*, 89.
Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009a). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell *36*, 593-608.
Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009b). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell *36*, 593-608.
Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W.S., and Lührmann, R. (1997). An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. EMBO J *16*, 4092-4106.
Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64-71.
FeRNAndez, C.F., Pannone, B.K., Chen, X., Fuchs, G., and Wolin, S.L. (2004). An Lsm2-Lsm7 complex in Saccharomyces cerevisiae associates with the small nucleolar RNA snR5. Mol Biol Cell *15*, 2842-2852.
Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.
Filipowicz, W., and Pogacic, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol *14*, 319-327.
Finzel, B.C., Weber, P.C., Hardman, K.D., and Salemme, F.R. (1985). Structure of ferricytochrome c' from Rhodospirillum molischianum at 1.67 A resolution. J Mol Biol *186*, 627-643.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.
Fischer, T., Strasser, K., Racz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., Lechner, J., and Hurt, E. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. EMBO J *21*, 5843-5852.
Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929-933.
Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C*., et al.* (2005). Protein interaction mapping: a Drosophila case study. Genome Res *15*, 376-384.
Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., and Mattaj, I.W. (1999). Genetic and physical interactions involving the yeast nuclear cap-binding complex. Mol Cell Biol *19*, 6543-6553.

Fourmann, J.B., Schmitzova, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.L., Fabrizio, P., and Lührmann, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. Genes Dev *27*, 413-428.
Ganot, P., Jady, B.E., Bortolin, M.L., Darzacq, X., and Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. Mol Cell Biol *19*, 6906-6917.
Gao, Q., Das, B., Sherman, F., and Maquat, L.E. (2005). Cap-binding protein 1-mediated and eukaryotic translation initiation factor 4E-mediated pioneer rounds of translation in yeast. Proc Natl Acad Sci U S A *102*, 4258-4263.
Garcia, B.A., Shabanowitz, J., and Hunt, D.F. (2005). Analysis of protein phosphorylation by mass spectrometry. Methods *35*, 256-264.
Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.
Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat Rev Mol Cell Biol *8*, 113-126.
Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B*., et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631-636.
Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M*., et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.
Geerlings, T.H., Vos, J.C., and Raué, H.A. (2000). The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5'-->3' exonucleases. RNA *6*, 1698-1703.
George, J.W., Stohr, B.A., Tomso, D.J., and Kreuzer, K.N. (2001). The tight linkage between DNA replication and double-strand break repair in bacteriophage T4. Proc Natl Acad Sci U S A *98*, 8290-8297.
Gerbi, S.A., and Lange, T.S. (2002). All small nuclear RNAs (snRNAs) of the [U4/U6.U5] Tri-snRNP localize to nucleoli; Identification of the nucleolar localization element of U6 snRNA. Mol Biol Cell *13*, 3123-3137.
Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.
Gingras, A.C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol *8*, 645-654.
Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M*., et al.* (1996). Life with 6000 genes. Science *274*, 546, 563-547.
Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell *19*, 53-63.
Gouw, J.W., Krijgsveld, J., and Heck, A.J. (2010). Quantitative proteomics by metabolic labeling of model organisms. Mol Cell Proteomics *9*, 11-24.
Grigoriev, A. (2003). On the number of protein-protein interactions in the yeast proteome. Nucleic Acids Res *31*, 4157-4161.
Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? Proteins *57*, 433-443.
Gutteridge, A., and Thornton, J.M. (2005). Understanding nature's catalytic toolkit. Trends Biochem Sci *30*, 622-629.
Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol *1*, e60.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. Nature *398*, 579-585.
Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., and Marraffini, L.A. (2013). A Ruler Protein in a Complex for Antiviral Defense Determines the Length of Small Interfering CRISPR RNAs. J Biol Chem *288*, 27888-27897.
Haynes, S.R. (1992). The RNP motif protein family. New Biol *4*, 421-429.
He, W., and Parker, R. (2000). Functions of Lsm proteins in mRNA degradation and splicing. Curr Opin Cell Biol *12*, 346-350.
Head, J.F. (1992). A better grip on calmodulin. Curr Biol *2*, 609-611.
Henras, A.K., Capeyrou, R., Henry, Y., and Caizergues-Ferrer, M. (2004). Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs. RNA *10*, 1704-1712.
Herzberg, C., Weidinger, L.A., Dorrbecker, B., Hubner, S., Stulke, J., and Commichau, F.M. (2007). SPINE: a method for the rapid detection and analysis of protein-protein interactions in vivo. Proteomics *7*, 4032-4035.
Higgs, P.G. (2000). RNA secondary structure: physical and computational aspects. Q Rev Biophys *33*, 199-253.
Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. Genes Dev *14*, 1415-1429.
Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167-170.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.
Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J Cell Biol *189*, 739-754.
Hunt, D.F., Yates, J.R., 3rd, Shabanowitz, J., Winston, S., and Hauer, C.R. (1986). Protein sequencing by tandem mass spectrometry. Proc Natl Acad Sci U S A *83*, 6233-6237.
Hunter, N., and Borts, R.H. (1997). Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. Genes Dev *11*, 1573-1582.
Isono, E., and Schwechheimer, C. (2010). Co-immunoprecipitation and protein blots. Methods Mol Biol *655*, 377-387.
Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A *98*, 4569-4574.
Jacobson, A., and Peltz, S.W. (1996). Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. Annu Rev Biochem *65*, 693-739.
James, L.C., Roversi, P., and Tawfik, D.S. (2003). Antibody multispecificity mediated by conformational diversity. Science *299*, 1362-1367.
Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565-1575.
Jimeno, S., Rondon, A.G., Luna, R., and Aguilera, A. (2002). The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. EMBO J *21*, 3526-3535.
Johnson, M.E., and Hummer, G. (2011). Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. Proc Natl Acad Sci U S A *108*, 603-608.
Jones, S., Daley, D.T., Luscombe, N.M., Berman, H.M., and Thornton, J.M. (2001). Protein-RNA interactions: a structural analysis. Nucleic Acids Res *29*, 943-954.
Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proc Natl Acad Sci U S A *93*, 13-20.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R.*, et al.* (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol *18*, 529-536.
Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Analytical chemistry *60*, 2299-2301.
Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., and Siksnys, V. (2013). crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. RNA Biol *10*, 841-851.
Kastner, B., Bach, M., and Lührmann, R. (1990). Electron microscopy of snRNPs U2, U4/6 and U5: evidence for a common structure-determining principle in the major UsnRNP family. Mol Biol Rep *14*, 171.
Ke, A., and Doudna, J.A. (2004). Crystallization of RNA and RNA-protein complexes. Methods *34*, 408-414.
Keogh, M.C., Podolny, V., and Buratowski, S. (2003). Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. Mol Cell Biol *23*, 7005-7018.
Kerppola, T.K. (2006). Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. Nat Protoc *1*, 1278-1286.
Kerppola, T.K. (2009). Visualization of molecular interactions using bimolecular fluorescence complementation analysis: characteristics of protein fragment complementation. Chem Soc Rev *38*, 2876-2886.
Keskin, O., Gursoy, A., Ma, B., and Nussinov, R. (2008). Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chem Rev *108*, 1225-1244.
Kistler, A.L., and Guthrie, C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. Genes Dev *15*, 42-49.
Kleanthous, C., Kuhlmann, U.C., Pommer, A.J., Ferguson, N., Radford, S.E., Moore, G.R., James, R., and Hemmings, A.M. (1999). Structural and mechanistic basis of immunity toward endonuclease colicins. Nat Struct Biol *6*, 243-252.
Knight, S.W., and Docherty, K. (1992). RNA-protein interactions in the 5' untranslated region of preproinsulin mRNA. J Mol Endocrinol *8*, 225-234.
Knuesel, M., Wan, Y., Xiao, Z., Holinger, E., Lowe, N., Wang, W., and Liu, X. (2003). Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. Mol Cell Proteomics *2*, 1225-1233.
Kolossova, I., and Padgett, R.A. (1997). U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. RNA *3*, 227-233.
Kosowski, T.R., Keys, H.R., Quan, T.K., and Ruby, S.W. (2009). DExD/H-box Prp5 protein is in the spliceosome during most of the splicing cycle. RNA *15*, 1345-1362.
Kramer, A., Gruter, P., Groning, K., and Kastner, B. (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. J Cell Biol *145*, 1355-1368.
Kress, T.L., Krogan, N.J., and Guthrie, C. (2008). A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. Mol Cell *32*, 727-734.
Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P.*, et al.* (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637-643.
Kufel, J., Allmang, C., Verdone, L., Beggs, J., and Tollervey, D. (2003). A complex pathway for 3' processing of the yeast U3 snoRNA. Nucleic Acids Res *31*, 6788-6797.
Kufel, J., Allmang, C., Verdone, L., Beggs, J.D., and Tollervey, D. (2002). Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. Mol Cell Biol *22*, 5248-5256.

Kufel, J., Bousquet-Antonelli, C., Beggs, J.D., and Tollervey, D. (2004). Nuclear pre-mRNA decapping and 5' degradation in yeast require the Lsm2-8p complex. Mol Cell Biol *24*, 9646-9657.
Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007). Functional architecture of RNA polymerase I. Cell *131*, 1260-1272.
Kuhner, S., van Noort, V., Betts, M.J., Leo-Macias, A., Batisse, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P.*, et al.* (2009). Proteome organization in a genome-reduced bacterium. Science *326*, 1235-1240.
Kvaratskhelia, M., and Grice, S.F. (2008). Structural analysis of protein-RNA interactions with mass spectrometry. Methods Mol Biol *488*, 213-219.
Labib, K., Tercero, J.A., and Diffley, J.F. (2000). Uninterrupted MCM2-7 function required for DNA replication fork progression. Science *288*, 1643-1647.
Lacadie, S.A., Tardiff, D.F., Kadener, S., and Rosbash, M. (2006). In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. Genes Dev *20*, 2055-2066.
Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev *12*, 527-537.
Lafontaine, D.L., and Tollervey, D. (2000). Synthesis and assembly of the box C+D small nucleolar RNPs. Mol Cell Biol *20*, 2650-2659.
Lakatos, L., Szittya, G., Silhavy, D., and Burgyan, J. (2004). Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. EMBO J *23*, 876-884.
Lane, C.S. (2005). Mass spectrometry-based proteomics in the life sciences. Cell Mol Life Sci *62*, 848-869.
Lange, T.S., and Gerbi, S.A. (2000). Transient nucleolar localization Of U6 small nuclear RNA in Xenopus Laevis oocytes. Mol Biol Cell *11*, 2419-2428.
Lee, C. (2007). Coimmunoprecipitation assay. Methods Mol Biol *362*, 401-406.
Lee, M.S., Henry, M., and Silver, P.A. (1996a). A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes Dev *10*, 1233-1246.
Lee, M.S., Henry, M., and Silver, P.A. (1996b). A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes & development *10*, 1233-1246.
Leeds, N.B., Small, E.C., Hiley, S.L., Hughes, T.R., and Staley, J.P. (2006). The splicing factor Prp43p, a DEAH box ATPase, functions in ribosome biogenesis. Mol Cell Biol *26*, 513-522.
Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980). Are snRNPs involved in splicing? Nature *283*, 220-224.
Liao, Y., Kariya, K., Hu, C.D., Shibatohge, M., Goshima, M., Okada, T., Watari, Y., Gao, X., Jin, T.G., Yamawaki-Kataoka, Y.*, et al.* (1999). RA-GEF, a novel Rap1A guanine nucleotide exchange factor containing a Ras/Rap1A-associating domain, is conserved between nematode and humans. J Biol Chem *274*, 37815-37820.
Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J., and Tan, S. (2005). Comparison of affinity tags for protein purification. Protein Expr Purif *41*, 98-105.
Lin, P.C., and Xu, R.M. (2012). Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J *31*, 1579-1590.
Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., 3rd, and Hartzog, G.A. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. Mol Cell Biol *23*, 1368-1378.
Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S. (2009). Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. Mol Cell Biol *29*, 4852-4863.

Lührmann, R., Kastner, B., and Bach, M. (1990). Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim Biophys Acta *1087*, 265-292.
Lührmann, R., and Stark, H. (2009). Structural mapping of spliceosomes by electron microscopy. Curr Opin Struct Biol *19*, 96-102.
Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol *8*, 479-490.
Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res *29*, 2860-2874.
Ma, X., Yang, C., Alexandrov, A., Grayhack, E.J., Behm-Ansmant, I., and Yu, Y.T. (2005). Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. EMBO J *24*, 2403-2413.
Ma, X., Zhao, X., and Yu, Y.T. (2003). Pseudouridylation (Psi) of U2 snRNA in S. cerevisiae is catalyzed by an RNA-independent mechanism. EMBO J *22*, 1889-1897.
Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct *1*, 7.
Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F*., et al.* (2011). Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol *9*, 467-477.
Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. EMBO J *23*, 2381-2391.
Malone, E.A., Clark, C.D., Chiang, A., and Winston, F. (1991). Mutations in SPT16/CDC68 suppress cis- and trans-acting mutations that affect promoter function in Saccharomyces cerevisiae. Mol Cell Biol *11*, 5710-5717.
Malone, E.A., Fassler, J.S., and Winston, F. (1993). Molecular and genetic characterization of SPT4, a gene important for transcription initiation in Saccharomyces cerevisiae. Mol Gen Genet *237*, 449-459.
Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.
Mann, M., Hendrickson, R.C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem *70*, 437-473.
Marelli, M., Smith, J.J., Jung, S., Yi, E., Nesvizhskii, A.I., Christmas, R.H., Saleem, R.A., Tam, Y.Y., Fagarasanu, A., Goodlett, D.R*., et al.* (2004). Quantitative mass spectrometry reveals a role for the GTPase Rho1p in actin organization on the peroxisome membrane. J Cell Biol *167*, 1099-1112.
Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev *19*, 1512-1517.
Mattaj, I.W. (1993). RNA recognition: a family matter? Cell *73*, 837-840.
Mayes, A.E., Verdone, L., Legrain, P., and Beggs, J.D. (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. EMBO J *18*, 4321-4331.
Mazza, C., Segref, A., Mattaj, I.W., and Cusack, S. (2002). Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. EMBO J *21*, 5548-5557.
McBride, A.E., Cook, J.T., Stemmler, E.A., Rutledge, K.L., McGrath, K.A., and Rubens, J.A. (2005a). Arginine methylation of yeast mRNA-binding protein Npl3 directly affects its function, nuclear export, and intranuclear protein interactions. J Biol Chem *280*, 30888-30898.
McBride, A.E., Cook, J.T., Stemmler, E.A., Rutledge, K.L., McGrath, K.A., and Rubens, J.A. (2005b). Arginine methylation of yeast mRNA-binding protein Npl3 directly affects its

function, nuclear export, and intranuclear protein interactions. J Biol Chem *280*, 30888-30898.
McKay, S.L., and Johnson, T.L. (2011). An investigation of a role for U2 snRNP spliceosomal components in regulating transcription. PLoS ONE *6*, e16077.
McPheeters, D.S., and Muhlenkamp, P. (2003). Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast. Mol Cell Biol *23*, 4174-4186.
Milligan, R.A. (1996). Protein-protein interactions in the rigor actomyosin complex. Proc Natl Acad Sci U S A *93*, 21-26.
Moore, P.B. (1998). The three-dimensional structure of the ribosome and its components. Annu Rev Biophys Biomol Struct *27*, 35-58.
Morozova, N., Allers, J., Myers, J., and Shamoo, Y. (2006). Protein-RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. Bioinformatics *22*, 2746-2752.
Mukherjee, D., Gao, M., O'Connor, J.P., Raijmakers, R., Pruijn, G., Lutz, C.S., and Wilusz, J. (2002). The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. EMBO J *21*, 165-174.
Murray, S., Udupa, R., Yao, S., Hartzog, G., and Prelich, G. (2001). Phosphorylation of the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase. Mol Cell Biol *21*, 4089-4096.
Nagai, K. (1996). RNA-protein complexes. Curr Opin Struct Biol *6*, 53-61.
Najbauer, J., Johnson, B.A., Young, A.L., and Aswad, D.W. (1993). Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase(s) modifying arginine in numerous proteins. J Biol Chem *268*, 10501-10509.
Newman, A.J. (1997). The role of U5 snRNP in pre-mRNA splicing. EMBO J *16*, 5797-5800.
Newman, A.J., Teigelkamp, S., and Beggs, J.D. (1995). snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. RNA *1*, 968-980.
Nilsen, T.W. (1994). RNA-RNA interactions in the spliceosome: unraveling the ties that bind. Cell *78*, 1-4.
Obrdlik, P., El-Bakkoury, M., Hamacher, T., Cappellaro, C., Vilarino, C., Fleischer, C., Ellerbrok, H., Kamuzinzi, R., Ledent, V., Blaudez, D*., et al.* (2004). K+ channel interactions detected by a genetic system optimized for systematic studies of membrane protein interactions. Proc Natl Acad Sci U S A *101*, 12242-12247.
Oeffinger, M. (2012). Two steps forward--one step back: advances in affinity purification mass spectrometry of macromolecular complexes. Proteomics *12*, 1591-1608.
Oeffinger, M., Wei, K.E., Rogers, R., DeGrasse, J.A., Chait, B.T., Aitchison, J.D., and Rout, M.P. (2007). Comprehensive analysis of diverse ribonucleoprotein complexes. Nat Methods *4*, 951-956.
Ohi, M.D., Link, A.J., Ren, L., Jennings, J.L., McDonald, W.H., and Gould, K.L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol Cell Biol *22*, 2011-2024.
Ohi, M.D., Vander Kooi, C.W., Rosenberg, J.A., Ren, L., Hirsch, J.P., Chazin, W.J., Walz, T., and Gould, K.L. (2005). Structural and functional analysis of essential pre-mRNA splicing factor Prp19p. Mol Cell Biol *25*, 451-460.
Panov, K.I., Panova, T.B., Gadal, O., Nishiyama, K., Saito, T., Russell, J., and Zomerdijk, J.C. (2006). RNA polymerase I-specific subunit CAST/hPAF49 has a role in the activation of transcription by upstream binding factor. Mol Cell Biol *26*, 5436-5448.
Pauling, M.H., McPheeters, D.S., and Ares, M., Jr. (2000). Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA. Mol Cell Biol *20*, 2176-2185.

Pawson, T. (2003). Organization of cell-regulatory systems through modular-protein-interaction domains. Philos Trans A Math Phys Eng Sci *361*, 1251-1262.
Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445-452.
Perez-Canadillas, J.M. (2006). Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. EMBO J *25*, 3167-3178.
Perez Canadillas, J.M., and Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. EMBO J *22*, 2821-2830.
Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551-3567.
Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B., and Aebersold, R. (2009). Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. Cell *138*, 795-806.
Plowman, G.D., Sudarsanam, S., Bingham, J., Whyte, D., and Hunter, T. (1999). The protein kinases of Caenorhabditis elegans: a model for signal transduction in multicellular organisms. Proc Natl Acad Sci U S A *96*, 13603-13610.
Prelich, G., and Winston, F. (1993). Mutations that suppress the deletion of an upstream activating sequence in yeast: involvement of a protein kinase and histone H3 in repressing transcription in vivo. Genetics *135*, 665-676.
Price, S.R., Evans, P.R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645-650.
Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.
Pyle, A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. Annu Rev Biophys *37*, 317-336.
Raker, V.A., Plessel, G., and Lührmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. EMBO J *15*, 2256-2269.
Ramakrishnan, V., and White, S.W. (1998). Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. Trends Biochem Sci *23*, 208-212.
Ramos, A., Hollingworth, D., Major, S.A., Adinolfi, S., Kelly, G., Muskett, F.W., and Pastore, A. (2002). Role of dimerization in KH/RNA complexes: the example of Nova KH3. Biochemistry *41*, 4193-4201.
Reed, R. (2003). Coupling transcription, splicing and mRNA export. Curr Opin Cell Biol *15*, 326-331.
Reed, R., and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. Cell *108*, 523-531.
Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol *17*, 1030-1032.
Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2008). A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics *7*, 282-289.
Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N*., et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.
Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res *39*, 9275-9282.
Schafer, T., Maco, B., Petfalski, E., Tollervey, D., Bottcher, B., Aebi, U., and Hurt, E. (2006). Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. Nature *441*, 651-655.

Scherly, D., Boelens, W., Dathan, N.A., van Venrooij, W.J., and Mattaj, I.W. (1990). Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B'' and their cognate RNAs. Nature *345*, 502-506.
Schlosshauer, M., and Baker, D. (2004). Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. Protein Sci *13*, 1660-1669.
Schurch, S., Bernal-Mendez, E., and Leumann, C.J. (2002). Electrospray tandem mass spectrometry of mixed-sequence RNA/DNA oligonucleotides. J Am Soc Mass Spectrom *13*, 936-945.
Schurch, S., Tromp, J.M., and Monn, S.T. (2007). Mass spectrometry of oligonucleotides. Nucleosides Nucleotides Nucleic Acids *26*, 1629-1633.
Shaffer, H.A., Rood, M.K., Kashlan, B., Chang, E.I., Doyle, D.F., and Azizi, B. (2012). BAPJ69-4A: a yeast two-hybrid strain for both positive and negative genetic selection. J Microbiol Methods *91*, 22-29.
Shamoo, Y., Abdul-Manan, N., Patten, A.M., Crawford, J.K., Pellegrini, M.C., and Williams, K.R. (1994). Both RNA-binding domains in heterogenous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. Biochemistry *33*, 8272-8281.
Shamoo, Y., Krueger, U., Rice, L.M., Williams, K.R., and Steitz, T.A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. Nat Struct Biol *4*, 215-222.
Shao, W., Kim, H.S., Cao, Y., Xu, Y.Z., and Query, C.C. (2012). A U1-U2 snRNP interaction network during intron definition. Mol Cell Biol *32*, 470-478.
Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. Science *300*, 805-808.
Sickmier, E.A., Frato, K.E., Shen, H., Paranawithana, S.R., Green, M.R., and Kielkopf, C.L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. Mol Cell *23*, 49-59.
Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J *30*, 1335-1342.
Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J *32*, 385-394.
Sleeman, J.E., and Lamond, A.I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. Curr Biol *9*, 1065-1074.
Sontheimer, E.J., and Steitz, J.A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science *262*, 1989-1996.
Staknis, D., and Reed, R. (1994). Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. Mol Cell Biol *14*, 2994-3005.
Stevens, S.W., Barta, I., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. RNA *7*, 1543-1553.
Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell *9*, 31-44.
Strambio-de-Castillia, C., Blobel, G., and Rout, M.P. (1999). Proteins connecting the nuclear pore complex with the nuclear interior. J Cell Biol *144*, 839-855.
Tang, X., and Bruce, J.E. (2009). Chemical cross-linking for protein-protein interaction studies. Methods Mol Biol *492*, 283-293.
Tang, X., and Bruce, J.E. (2010). A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies. Mol Biosyst *6*, 939-947.

Taoka, M., Ikumi, M., Nakayama, H., Masaki, S., Matsuda, R., Nobe, Y., Yamauchi, Y., Takeda, J., Takahashi, N., and Isobe, T. (2010). In-gel digestion for mass spectrometric characterization of RNA from fluorescently stained polyacrylamide gels. Anal Chem *82*, 7795-7803.
Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., and Wrana, J.L. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol *27*, 199-204.
Tharun, S., He, W., Mayes, A.E., Lennertz, P., Beggs, J.D., and Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. Nature *404*, 515-518.
Thiede, B., and von Janta-Lipinski, M. (1998). Noncovalent RNA-peptide complexes detected by matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom *12*, 1889-1894.
Thomson, J.J. (1913). On the Appearance of Helium and Neon in Vacuum Tubes. Science *37*, 360-364.
Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigo, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res *22*, 1616-1625.
Tomasevic, N., and Peculis, B.A. (2002). Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence. Mol Cell Biol *22*, 4101-4112.
Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., Boisvert, F.M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H*., et al.* (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol *183*, 223-239.
Tsai, C.J., and Nussinov, R. (1997). Hydrophobic folding units at protein-protein interfaces: implications to protein folding and to protein-protein association. Protein Sci *6*, 1426-1437.
Tsai, R.T., Fu, R.H., Yeh, F.L., Tseng, C.K., Lin, Y.C., Huang, Y.H., and Cheng, S.C. (2005). Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev *19*, 2991-3003.
Tycowski, K.T., You, Z.H., Graham, P.J., and Steitz, J.A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. Mol Cell *2*, 629-638.
van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., and Brouns, S.J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci *34*, 401-407.
van Heugten, H.A., Thomas, A.A., and Voorma, H.O. (1992). Interaction of protein synthesis initiation factors with the mRNA cap structure. Biochimie *74*, 463-475.
van Hoof, A., Lennertz, P., and Parker, R. (2000). Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. Mol Cell Biol *20*, 441-452.
Varani, L., Gunderson, S.I., Mattaj, I.W., Kay, L.E., Neuhaus, D., and Varani, G. (2000). The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. Nat Struct Biol *7*, 329-335.
Vargason, J.M., Szittya, G., Burgyan, J., and Hall, T.M. (2003). Size selective recognition of siRNA by an RNA silencing suppressor. Cell *115*, 799-811.
Vasilescu, J., Guo, X., and Kast, J. (2004). Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. Proteomics *4*, 3845-3854.
Villa, T., Ceradini, F., Presutti, C., and Bozzoni, I. (1998). Processing of the intron-encoded U18 small nucleolar RNA in the yeast Saccharomyces cerevisiae relies on both exo- and endonucleolytic activities. Mol Cell Biol *18*, 3376-3383.
Voehringer, D., Wu, D., Liang, H.E., and Locksley, R.M. (2009). Efficient generation of long-distance conditional alleles using recombineering and a dual selection strategy in replicate plates. BMC Biotechnol *9*, 69.

Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nat Protoc *2*, 1467-1478.
Waghmare, S.P., Pousinis, P., Hornby, D.P., and Dickman, M.J. (2009). Studying the mechanism of RNA separations using RNA chromatography and its application in the analysis of ribosomal RNA and RNA:RNA interactions. J Chromatogr A *1216*, 1377-1382.
Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.
Wait, R. (1993). Fast atom bombardment mass spectrometry of peptides. Methods Mol Biol *17*, 237-283.
Wang, Q., He, J., Lynn, B., and Rymond, B.C. (2005). Interactions of the yeast SF3b splicing factor. Mol Cell Biol *25*, 10745-10754.
Wang, X., Ira, G., Tercero, J.A., Holmes, A.M., Diffley, J.F., and Haber, J.E. (2004). Role of DNA replication proteins in double-strand break-induced recombination in Saccharomyces cerevisiae. Mol Cell Biol *24*, 6891-6899.
Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., and Lührmann, R. (2000). A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. Cell *103*, 457-466.
Weinstein, L.B., and Steitz, J.A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. Curr Opin Cell Biol *11*, 378-384.
Wells, S.E., Neville, M., Haynes, M., Wang, J., Igel, H., and Ares, M., Jr. (1996). CUS1, a suppressor of cold-sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145. Genes Dev *10*, 220-232.
Wen, Y., and Shatkin, A.J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev *13*, 1774-1779.
Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486-489.
Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., and Williams, K.L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev *13*, 19-50.
Will, C.L., and Lührmann, R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. Biol Chem *386*, 713-724.
Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*.
Windgassen, M., Sturm, D., Cajigas, I.J., Gonzalez, C.I., Seedorf, M., Bastians, H., and Krebber, H. (2004). Yeast shuttling SR proteins Npl3p, Gbp2p, and Hrb1p are part of the translating mRNPs, and Npl3p can function as a translational repressor. Mol Cell Biol *24*, 10479-10491.
Winston, F., Chaleff, D.T., Valent, B., and Fink, G.R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics *107*, 179-197.
Witze, E.S., Old, W.M., Resing, K.A., and Ahn, N.G. (2007). Mapping protein post-translational modifications with mass spectrometry. Nat Methods *4*, 798-806.
Woudstra, E.C., Gilbert, C., Fellows, J., Jansen, L., Brouwer, J., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2002). A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. Nature *415*, 929-933.
Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. (1992). Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. Genes Dev *6*, 2542-2553.
Xia, Z., Zhu, Z., Zhu, J., and Zhou, R. (2009). Recognition mechanism of siRNA by viral p19 suppressor of RNA silencing: a molecular dynamics study. Biophys J *96*, 1761-1769.

Xu, C., and Henry, M.F. (2004). Nuclear export of hnRNP Hrp1p and nuclear export of hnRNP Npl3p are linked and influenced by the methylation state of Npl3p. Mol Cell Biol *24*, 10742-10756.
Xu, D., Tsai, C.J., and Nussinov, R. (1997). Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng *10*, 999-1012.
Xu, X., Song, Y., Li, Y., Chang, J., Zhang, H., and An, L. (2010). The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification. Protein Expr Purif *72*, 149-156.
Xu, Y.Z., Newnham, C.M., Kameoka, S., Huang, T., Konarska, M.M., and Query, C.C. (2004). Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. EMBO J *23*, 376-385.
Xu, Y.Z., and Query, C.C. (2007). Competition between the ATPase prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. Mol Cell *28*, 838-849.
Yamashita, A., Chang, T.C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.Y., and Shyu, A.B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol *12*, 1054-1063.
Yao, S., and Prelich, G. (2002). Activation of the Bur1-Bur2 cyclin-dependent kinase complex by Cak1. Mol Cell Biol *22*, 6750-6758.
Yates, J.R., 3rd (2000). Mass spectrometry. From genomics to proteomics. Trends Genet *16*, 5-8.
Ye, K., Malinina, L., and Patel, D.J. (2003). Recognition of small interfering RNA by a viral suppressor of RNA silencing. Nature *426*, 874-878.
Zhang, H., Tang, X., Munske, G.R., Tolic, N., Anderson, G.A., and Bruce, J.E. (2009). Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. Mol Cell Proteomics *8*, 409-420.
Zhou, H., Chen, Y.Q., Du, Y.P., and Qu, L.H. (2002). The Schizosaccharomyces pombe mgU6-47 gene is required for 2'-O-methylation of U6 snRNA at A41. Nucleic Acids Res *30*, 894-902.
Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C., and Shi, Y. (2013). Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. Nature.