Investigation of Protein-protein Interactions within the Human Spliceosomal U4/U6.U5 tri-snRNP Particle

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von

Sunbin Liu

aus Fujian, V.R. China

Göttingen 2005

D7
Referent: Prof. Dr. Ralf Ficner
Korreferent: Prof. Dr. Hans-Joachim Fritz
Tag der mündlichen Prüfung: 28. April 2005

In the beginning God created the heavens and the earth.

[Genesis 1:1]

CONTENTS

1 Summary

The removal of introns from nuclear pre-messenger RNA (pre-mRNA) is catalysed by the spliceosome. One of the major building blocks of the human spliceosome is the 25S U4/U6.U5 tri-snRNP. The purified tri-snRNP from HeLa cells contains in addition to the U4, U5 and U6 snRNAs at least 29 distinct proteins. One set of seven Sm proteins is associated with both U4 and U5 snRNAs, while one set of LSm 2-8 proteins is bound to the 3' end of U6 snRNA. Proteins 15.5K, 20K, 60K, 61K, and 90K interact specifically with the U4/U6 snRNPs. U5 snRNA, in contrast, is associated with the 220K, 200K, 116K, 102K, 100K, 40K, and 15K U5-specific proteins. Three proteins, namely 110K, 65K, and 27K, more stably associated with the tri-snRNP than with free U4/U6- or U5-particle, represent the tri-snRNP specific group.

The tri-snRNP particle has to be assembled from U4, U6 and U5 snRNPs prior to association with the pre-spliceosome. First, U4 and U6 snRNPs interact through extensive base-pairing of their respective snRNAs to form the U4/U6 snRNPs. Subsequently, the U4/U6 snRNPs interacts with U5 snRNP to form the U4/U6.U5 tri-snRNP. The recruitment of tri-snRNP to the pre-spliceosome initiates the activation of the spliceosome, which undergoes a series of rearrangements to form the catalytic core. The proteins of the tri-snRNP play a central part in the tri-snRNP assembly and the activation of the spliceosome, and thus in the splicing of pre-mRNA. In this work, yeast two-hybrid techniques and *in vitro* binding assays were established to investigate the protein-protein interactions in these particles.

Two-hybrid data allowed me to define the regions through which proteins 220K, 200K, and 116K interact with each other. These regions include the N-terminal and C-terminal domains of protein 220K, the second helicase domain of 200K, and the C-terminal domain of 116K. Strikingly, most of the binding domains are also identified in the interactions of the yeast orthologue proteins Prp8p, Brr2p, and Snu114p.

Protein 102K interacts with several particle-specific proteins of both the U5 and the U4/U6 snRNP, and thus very probably acts as a bridging or scaffolding protein in the formation or structural stabilisation of the tri-snRNP. The interactions with U5-220K, 200K and 116K allow the 102K protein to bind stably to the U5-snRNP particle. The 102K protein is the only U5-specific protein identified in this work that interacts with U4/U6-specific proteins 61K and 90K. The former interaction has been proved to be critical for the formation of the tri-snRNP. One of the missense mutations (A194E) in the 61K protein responsible for the disorder retinitis pigmentosa has been shown a reduced interaction with the 102K in this study. The possible mechanisms of this are discussed. Mutational analysis showed that the TPR repeats of 102K protein are each dedicated to the interactions with specific partners. All repeats participate in the interaction with the U4/U6-61K, whereas only the first nine repeats interact with 110K, 200-4 fragment of 200K, and the 220-1 fragment of 220K.

The U4/U6-90K protein interacts with 60K within the 20K•60K•90K heterotrimer and contacts the U6 snRNA in the stem II region in the U4/U6-snRNP particle. In this study, it was shown that this protein interacts with human recycling factor U6-p110 and thus functions in the recycling of U4/U6 snRNPs. Mutational analysis showed that the C-terminal domain of 90K, comprising amino acids 417–683, is responsible for this binding. During the formation of tri-snRNP, the U5-102K protein interacts with 90K protein in the same domain, suggesting that U5-102K might function in the release of U6-p110. Protein 90K also interacts with U2-associated protein SPF30/SMNrp through its N-terminal region, and therefore functions in the recruitment of tri-snRNP into pre-spliceosome.

The tri-snRNP-specific 110K protein interacts with U4/U6-90K, U5-102K and U5-200K through the C-terminal region lacking an RS domain. Since the absence of 110K protein does not compromise the stability of the tri-snRNP, it appears reasonable that anchoring 110K to these proteins may be required to properly position 110K for its contribution in connecting the tri-snRNP to the pre-spliceosome.

The U5-52K protein interacts with the U5-102K and 15K proteins, suggesting that these interactions contribute to its integration into the U5 particle. Binding studies performed with 52K deletion mutants revealed that the N-terminal two-thirds of 52K interact with the 102K protein, whereas its C-terminal GYF-domain binds the 15K protein. The GYF domain has been characterised previously as a polyproline-targeting molecule. As the 15K protein lacks a proline-rich tract, these data indicate for the first time that a GYF-domain can also engage in specific protein-protein interactions in a polyproline-independent manner. The crystallography study of the 52K GYF domain in complex with 15K, a cooperative work with the laboratory of Prof. R. Ficner at the University of Göttingen, showed that the 15K protein makes contact with the distinctive surface of GYF domain, as does the proline-rich target. The data from this study and others demonstrated that the 52K protein is the only 20S U5-specific protein that is not integrated into the tri-snRNP.

On the basis of the data obtained in this work, I propose a model of the assembly of the U4/U6.U5 tri-snRNP.

2 Introduction

2.1 Pre-mRNA Splicing

Eukaryotic pre-mRNAs must undergo several posttranscriptional modifications before their export to the cytoplasm as functional mRNAs. Most pre-mRNAs contain intervening sequences (introns) that must be removed in order to place the coding sequences (exons) in a protein-reading frame. The mechanism of this critical processing event, known as pre-mRNA splicing, has been extensively studied (reviewed in Brow, 2002). It has long been known that intron removal and the ligation of exons occurs through two sequential transesterification reactions that are carried out by a multicomponent complex that is known as the spliceosome. Most introns have common consensus sequences near their 5' and 3' ends that are recognized by spliceosomal components and are required for spliceosome formation. The assembly of a spliceosome onto a pre-mRNA is an ordered process that involves five small nuclear ribonucleoprotein particles (snRNPs; U1, U2, U4, U5 and U6), as well as a large number of non-snRNP proteins. The spliceosome acts through a multitude of RNA-RNA, RNA-protein and protein-protein interactions to precisely excise each intron and join the exons in the correct order. However, the regulation of splicing is still not well understood, particularly in the context of the structure of the spliceosome.

2.2 The Chemistry of the Splicing Reaction

2.2.1 The two-step splicing reactions

Splicing comprises two transesterification reactions (figure 2.1; Ruby and Abelson, 1991; Sharp *et al*., 1987). In the first step, the phosphodiester bond at the 5’ splice site (SS) is cleaved by a nucleophilic attack by the 2’-hydroxyl group of the conserved adenosine (A) located at the intron branch point. This generates a 2’–5’ phosphodiester bond between the branch site and the 5’ end of the intron as well as a free 3’ hydroxyl group on the 5’ exon. The cleavage also results in two RNA molecules, i.e., a free exon 1 and a “lariat” intron-exon 2. In the second step, the free 3’ hydroxyl on the end of the 5’ exon attacks the

phosphodiester bond at the 3' splice site, resulting in the joining of the exons and release of the lariat intron. The basic chemistry of pre-mRNA splicing is similar to that of the group II introns, also called self-splicing introns, that are found in organelles. This group of introns catalyze their own removal and require no protein cofactors (Cech, 1990a, b; Michel and Jacquier, 1987). In contrast to these self-splicing introns, nuclear pre-mRNA splicing requires an input of energy and protein cofactors that are organized into a large ribonucleoprotein particle called the spliceosome.

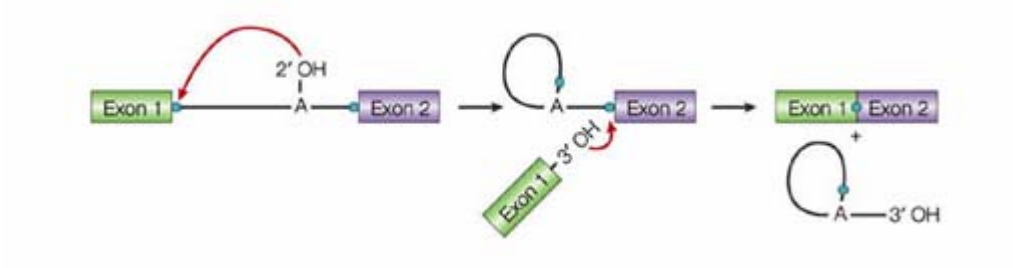


Figure 2.1. Pre-mRNA splicing occurs by two sequential transesterification reactions. A schematic pre-messenger RNA is shown on the left as a single intron (solid line) flanked by two exons. The first and second steps of splicing involve nucleophilic attacks (red arrows) on the terminal phosphodiester bonds (blue dots) by the 2' hydroxyl of the branch-point adenosine (A) and by the 3' hydroxyl of the upstream exon, respectively. The ligated exons and the lariat intron products are shown on the right (from Patel and Steitz, 2003).

2.2.2 The consensus sequence of the introns

Intron removal occurs in a process called nuclear pre-mRNA splicing after the gene has been transcribed. Conserved sequence elements within the primary RNA transcript provide the signals required for precise intron removal. These sequences include 5' and 3' splice sites (SS), a branch point sequence containing a strictly conserved adenosine residue, and a stretch of pyrimidines that is located between the branch point and the 3' splice site (figure 2.2).

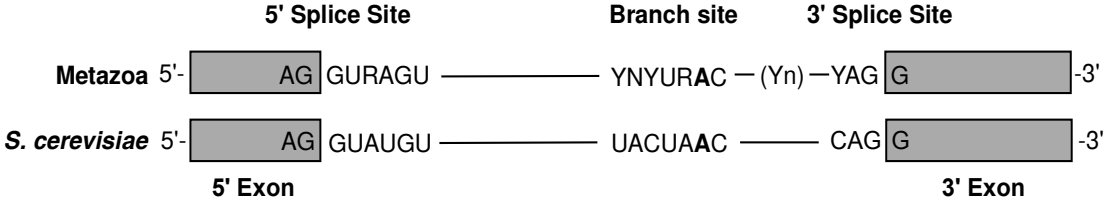


Figure 2.2. Conserved sequences of the introns in a typical metazoan or yeast pre-mRNA. The intron consensus sequences direct spliceosome assembly and define the splice sites and branch site (Y denotes pyrimidine; R denotes purine).

Interestingly, although the conserved sequence at the 5' and 3' splice sites and branch point are similar in all organisms, these signals are much more conserved in the yeast S*accharomyces cerevisiae* than in mammals. The 5' SS sequence signal in yeast is almost always GUAUGU while in humans only the first two positions (i.e., GU) are very highly conserved. Similarly, the branch point sequence found in yeast introns is almost always UACUAAC, but the sequence is very degenerate in mammalian introns. In both yeast and mammals, the 3' SS signal is rather short, consisting of a pyrimidine (U or C) followed by AG. The differences in signal sequence conservation probably reflect the relative complexity of the systems. As a simple, single-celled eukaryote, yeast has few introns and no alternative splicing. Their introns can therefore have very strong, clear-cut signals. Higher organisms that utilize alternative splicing, however, have a number of weaker signals that allow a choice between their usages under various conditions.

2.3 The Splicing Machinery

2.3.1 The spliceosome

Splicing is accomplished in a complex cellular machine called the spliceosome, which co-ordinates the removal of introns from pre-mRNAs. Since the discovery that a large complex was responsible for splicing, the list of spliceosome components has grown to include five small nuclear RNAs (snRNAs) and more than 180 proteins (Burge *et al*., 1999; Brow, 2002; Jurica and Moore, 2003). Each of the five snRNAs (U1, U2, U4, U5 and U6) is found in a complex with a number of proteins to form small nuclear ribonucleoprotein particles or snRNPs. Non-snRNP proteins are also required for splicing and are likely to interact with other splicing components only transiently. One particular class of proteins found in metazoans that are not present in yeast is the SR family of proteins. All SR proteins have a similar bipartite structure composed of two functional domains: an N-terminal RNA binding domain, comprising multiple RNA-recognition-motifs (RRMs), and a C-terminal arginine-serine-rich (RS) domain. Whereas the RRMs are sufficient for sequence-specific RNA binding, the RS domain is required for enhancement of splicing activities.

Until now, two distinct spliceosomes have been identified (reviewed by Padgett *et al*., 2002). The large majority of introns are called U2-dependent introns and are spliced by the major spliceosomal snRNPs U1, U2, U4, U5 and U6. Interestingly, a minor class of introns have been identified in some more advanced organisms including *Arabadopsis*, *Drosophila*, *Xenopus*, mice and humans. This rare class of introns are called U12-dependent and are spliced by a group of less abundant snRNPs including U11, U12, U4atac and U6atac in coordination with the standard U5 snRNP. This class of introns was first identified by their unusual intron termini, which contained an AT-AC sequence instead of the otherwise highly conserved GT-AG sequence (Burge *et al*., 1999). It was later found that the 5' splice site and branchpoint sequences in this class of introns were complementary to stretches of the U11 and U12 snRNAs, two low-abundance snRNAs with previously unknown function. The involvement of these two snRNAs in splicing was supported by evidence of interactions between U11, U12, and U5 (Hall and Padgett, 1996; Tarn and Steitz, 1996a). Two additional novel snRNA components, U4atac and U6atac, which appear functionally analogous to U4 and U6 in the standard U2-type spliceosome, were also identified (Tarn and Steitz, 1996b).

2.3.2 Spliceosomal snRNAs

The major spliceosome contains five snRNAs, U1, U2, U4, U5, and U6, which are packaged as ribonucleoprotein particles (snRNPs). The primary and secondary structure of the snRNAs is shown in figure 2.3. Spliceosomal snRNAs are uridine-rich RNA molecules. These snRNAs associate dynamically with each other and with pre-mRNA substrates in the spliceosome assembly and splicing catalysis. Among the five snRNAs and the snRNP particles they form, U1, U2, U4 and U5 are similar. The U1-U5 snRNAs are transcribed by RNA polymerase II and modified by a unique trimethylguanosine cap structure that is not found in the U6 snRNA. Similarly, only U1, U2, U4, and U5 snRNAs contain a conserved Sm site, a structural domain that allows binding to Sm core proteins, B/B', D1, D2, D3, E, F and G (Branlant *et al*., 1982; Mattaj *et al*., 1986). In contrast, U6 snRNA has no Sm site and consequently does not associate with the Sm proteins. Moreover, its biogenesis pathway differs in many respects from that of the other snRNAs, as it is transcribed by RNA polymerase III

(Dahlberg and Lund, 1991) and capped with γ-monomethyl triphosphate (Singh and Reddy, 1989). Finally, U6 does not leave the nucleus during snRNP biogenesis (Vankan *et al*., 1990; Terns *et al*., 1993; Boelens *et al*., 1995).

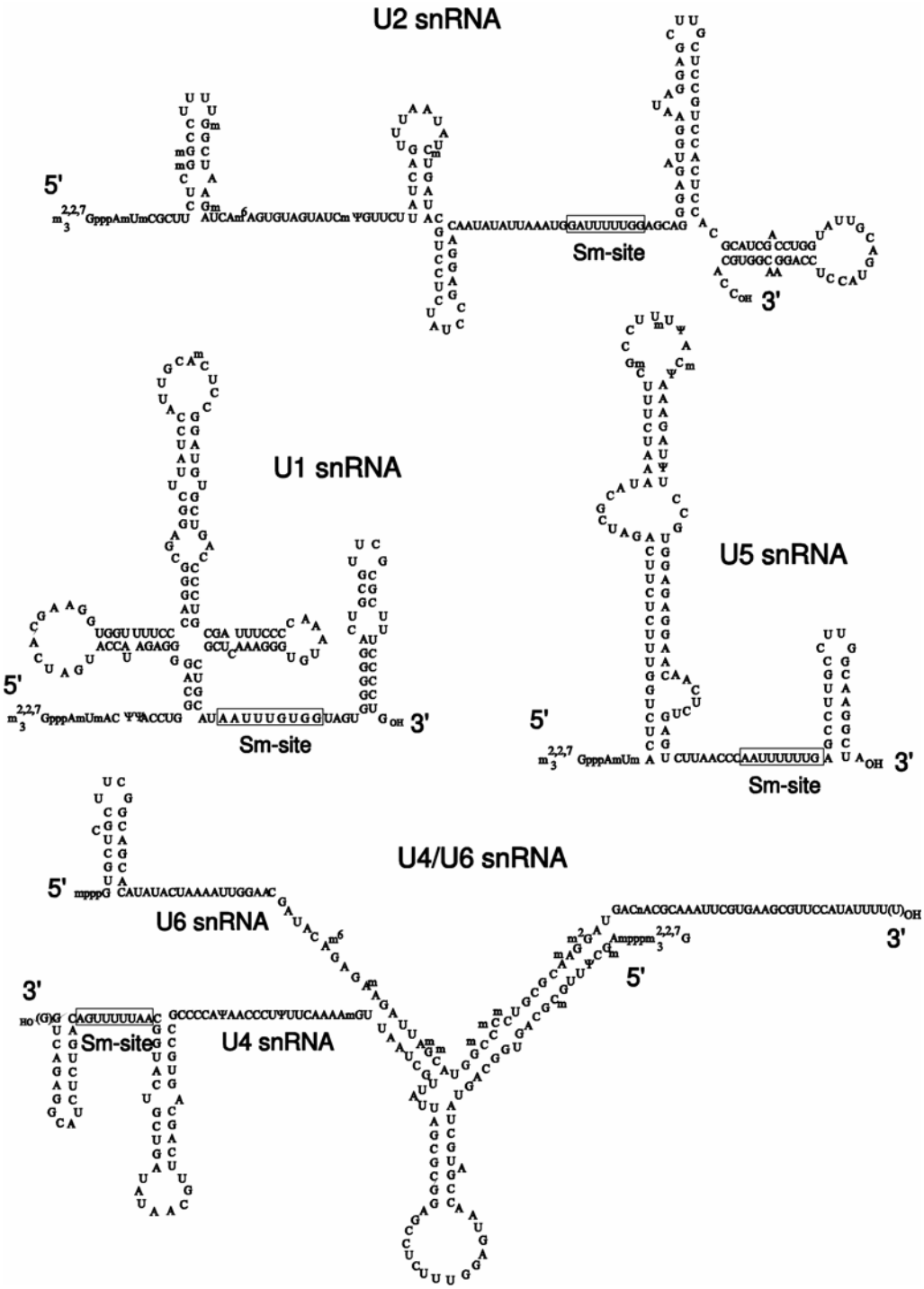


Figure 2.3. The primary and secondary structure of human U snRNAs U1, U2, U4, U5 and U6. The Sm site which binds Sm core proteins is indicated (provided by Dr. C. Will in the laboratory of Prof. Lührmann).

2.3.3 The spliceosomal snRNP proteins

The protein composition of the major spliceosomal snRNPs has been best characterized in HeLa cells and yeast (S*accharomyces cerevisiae*) and is summarized in table 2.1. Proteins associated with the U1, U2, U5, and U4/U6 snRNPs fall into two classes. The first class consists of the so-called common or Sm proteins, which are tightly associated with all snRNP particles. The second class is comprised of the particle-specific proteins, which associate with a particular snRNP particle or complex. These proteins exhibit a wide range of binding affinities, and their association with an snRNP particle is thus dependent on the ionic strength of the particle's environment.

2.3.3.1 The common proteins

The U1, U2, U4, and U5 snRNPs contain seven common Sm core proteins. The human Sm proteins were initially identified by their ability to cross-react with antisera from patients with the autoimmune disorder systemic lupus erythematosus and were named B/B', D1, D2, D3, E, F and G based on their relative mobilities during gel electrophoresis (Lerner and Steitz, 1979; van Venrooij, 1987). B and B' are encoded by alternatively spliced transcripts of the same gene. The homologues of the Sm proteins have been identified in *Saccharomyces cerevisiae* and named after their respective human counterparts. These proteins are known to initiate snRNP assembly in the cytoplasm by associating with the conserved Sm site in the U1, U2, U4, and U5 snRNAs upon their export from the nucleus (Branlant *et al*., 1982; Mattaj and De Robertis, 1985). The association of the Sm proteins then allows hypermethylation of the 5'-m^7G cap of these snRNAs to convert it to a 2,2,7-tri-methylguanosine (m_3G) which together with the Sm proteins provides a signal for import into the nucleus (Fischer and Lührmann, 1990; Hamm *et al*., 1990). After returning to the nucleus, the remainder of the snRNP specific proteins associate with the particle to complete the maturation process (Zieve and Sauterer, 1990). In contrast, U6 snRNP contains seven Sm-like (LSm) proteins, LSm2-8. LSm proteins bind directly to the 3'-end of U6 snRNA. All Sm and LSm proteins share a conserved Sm sequence motif consisting of two segments, Sm1 and Sm2, interrupted by a spacer region of variable length (Cooper *et al*., 1995; Hermann *et al*., 1995; Seraphin, 1995).

Homo sapiens							*Saccharomyces cerevisiae*					
Protein	**U1**	**U2**	**U5**	**U4/U6**	**U4/U6.U5**	**Sequence motif**	**U1**	**U2**	**U5**	**U4/U6**	**U4/U6.U5**	**Protein**
B/B'	◉	◉	◉	◉○	◉○		◉	◉	◉	◉○	◉○	B
D1	◉	◉	◉	◉○	◉○		◉	◉	◉	◉○	◉○	D1
D2	◉	◉	◉	◉○	◉○		◉	◉	◉	◉○	◉○	D2
D3	◉	◉	◉	◉○	◉○	Sm motif	◉	◉	◉	◉○	◉○	D3
E	◉	◉	◉	◉○	◉○		◉	◉	◉	◉○	◉○	E
F	◉	◉	◉	◉○	◉○		◉	◉	◉	◉○	◉○	F
G	◉	◉	◉	◉○	◉○		◉	◉	◉	◉○	◉○	G
70K	●					RRM	●					Snp1
A	●					2 RRMs	●					Mud1
C	●					Zn-finger	●					yU1-C
						3 RRMs	●					Nam8
						TPR repeats	●					Prp39
						WW domain	●					Prp40
						TPR repeats	●					Prp42
						-	●					Snu56
						-	●					Snu71
						Zn-finger	●					Luc7
A'		●				Leu-rich		●				Lea1
B''		●				2 RRMs		●				Msl1
SF3a120		●				2 SURP, UBQ		●				Prp21
SF3a66		●				Zn-finger		●				Prp11
SF3a60		●				Zn-finger		●				Prp9
SF3b155		●				HEAT repeats		●				Hsh155
SF3b145		●				SAP, Pro-rich		●				Cus1
SF3b130		●				CPSF A		●				Rse1
SF3b125		●				DEAD box						-
SF3b49		●				2 RRMs		●				Hsh49
SF3b14a		●				RRM						-
SF3b14b		●				Cys-rich		●				Rds3
SF3b10		●				-		●				Ysf3
hPrp5		●				DEAD box		●				Prp5
SR140		●				RRM, SWAP						-
CHERP		●				SWAP, G-patch						-
hPrp43		●				DEXH box		●				Prp43
SPF45		●				G-patch, RRM						-
SPF31		●				DnaJ domain						-
SPF30		●				Tudor						-
220K			●		●	-			●		●	Prp8
200K			●		●	2 DEXH-box			●		●	Brr2
116K			●		●	G-domain			●		●	Snu114
102K			●		●	TPR repeats				●	●	Prp6
100K			●		●	DEAD, RS			●			Prp28
52K			●			GYF domain			●			Snu40
40K			●		●	WD40 motif						-
15K			●		●	Thioredoxin fold			●			Dip1
90K				●	●	PWI, dsRNA bdg				●	●	Prp3
61K				●	●	Nop domain				●	●	Prp31
60K				●	●	WD40 motif				●	●	Prp4
20K				●	●	Cyclophilin fold						-
15.5K				●	●	RNA bdg. motif				●	●	Snu13
110K					●	RS domain					●	Snu66
65K					●	RS domain					●	Sad1
27K					●	RS domain						-
-						-					●	Prp38
-						Zn-finger					●	Snu23
-						PEST motif					●	Spp381

Table 2.1. Protein composition of the U snRNPs in human and yeast. The snRNP proteins consist of the so called common Sm (grey spots) or LSm (light grey spots) proteins, and particle-specific proteins (black spots).

2.3.3.2 The snRNP-specific proteins

The mammalian **12S U1 snRNP** consists of three particle-specific proteins, U1-70K, A and C, in addition to the common Sm proteins. The two largest proteins, U1-70K and U1-A, bind directly to U1 snRNA stem/loop I and II, respectively (Nagai *et al.*, 1994). The structure of the amino-terminal fragment of U1-A in complex with the loop of stem II has been solved by X-ray crystallography (Oubridge *et al*., 1994). The smaller U1-C protein is probably attached by protein-protein interactions with Sm proteins and U1-70K protein (Nelissen *et al*., 1994).

The **17S U2 snRNP** is a large particle containing two stably associated proteins: U2-A' and U2-B" (Lea1p and Msl1p/Yib9p in yeast) (Lührmann *et al.*, 1990; Tang *et al.*, 1996; Caspary and Séraphin, 1998), and two multisubunit complexes: SF3a and SF3b (Brosi *et al.*, 1993). SF3a is composed of three polypeptides (SF3a120, SF3a66, SF3a60, in human; Prp21p, Prp11p and Prp9p in yeast; Krämer, 1995), while human SF3b was reported to contain seven proteins (SF3b155, SF3b145, SF3b130, SF3b49, SF3b14a/p14, SF3b14b and SF3b10) (Gozani *et al.*, 1996; Das *et al.*, 1999; Will *et al.*, 2002). Yeast homologues of six of these factors have been characterized (Wells *et al.*, 1996; Igel *et al.*, 1998; Caspary *et al.*, 1999; Wang and Rymond, 2003; Dziembowski *et al*., 2004). Interestingly, human SF3a acts exclusively during the splicing of major introns (Will *et al.*, 1999), while SF3b is required for splicing of both types of introns by associating with U2 and U12 snRNPs, respectively (Das *et al.*, 1999; Will *et al*., 1999). Thus, both factors are essential for splicing. Furthermore, many SF3b subunits can be crosslinked to the branchpoint region of the pre-mRNA (Staknis and Reed, 1994; Gozani *et al.*, 1996; Query *et al.*, 1996; McPheeters and Muhlenkamp, 2003). Several additional human U2 associated proteins including SPF30/SMNrp, hPrp5p and hPrp43p have been identified recently (Meister *et al*., 2001; Will *et al*., 2002).

U4 and U6 snRNAs contain extensive complementary sequence and interact to form **U4/U6 snRNP**. In addition to the seven Sm proteins that bind the Sm site of the U4 snRNA, and the seven LSm proteins that are associated with U6 snRNA, 13S U4/U6 snRNP contains five particle-specific proteins, 90K, 60K,

20K, 61K, and 15.5K (the terminology is derived from their apparent molecular weights in electrophoresis; reviewed in Will and Lührmann, 2001). The proteins 20K, 60K, and 90K form a biochemically stable, heteromeric complex (Horowitz *et al*., 1997; Teigelkamp *et al.*, 1998). Protein 20K belongs to the cyclophilin family of peptidyl-prolyl isomerases, termed cyclophilin H (CypH) or SnuCyp-20 or USA-Cyp (Horowitz *et al.*, 1997; Teigelkamp *et al*., 1998). The crystal structure of CypH shows a typical cyclophilin fold (Reidt *et al*., 2000). Protein 61K shares a homologous central domain with the proteins Nop56 and Nop58, which are integral constituents of the box C/D snoRNPs (Gautier *et al*., 1997; Makarova *et al*., 2002). The smallest protein 15.5K possesses a novel RNA-binding domain and binds directly to U4 snRNA (Nottrott *et al*., 1999). This is the first U4/U6-specific protein identified as interacting directly with U4 snRNA, and the structure of protein 15.5K in complex with the 5' stem-loop of U4 snRNA has been determined by X-ray crystallography (Vidovic *et al*., 2000). Moreover, the binding of protein 15.5K to the U4 5' stem-loop is required for subsequent interaction of both 61K protein and the 20K•60K•90K protein complex with U4/U6 snRNA duplex (Nottrott *et al*., 2002). In this respect, protein 15.5K functions as a nucleation factor. Interestingly, 15.5K protein is also present in box C/D snoRNPs, providing a link between the pre-mRNA and pre-rRNA processing machineries (Walkins *et al*., 2000). Except for the 20K protein, orthologous proteins termed Snu13p (15.5K in human), Prp4p (60K), Prp3p (90K) and Prp31p (61K) are also associated with the yeast U4/U6 snRNP particle (Banroques and Abelson, 1989; Peterson-Bjørn *et al*., 1989; Anthony *et al*., 1997; Weidenhammer *et al.*, 1997; Gottschalk *et al.*, 1999; Stevens and Abelson, 1999). Like their human counterparts, Prp4p and Prp3p interact directly with each other (Wang *et al*., 1997; Gonzalez-Santos *et al*., 2002; Ayadi *et al*., 1998). Genetic and biochemical studies in yeast and HeLa cells have shown that all conserved U4/U6-specific proteins are essential for cell viability and required for pre-mRNA splicing (Lustig *et al*., 1986; Banroques and Abenlson, 1989; Peterson-Bjørn *et al*., 1989; Weidenhammer *et al*., 1997; Nottrott *et al*., 1999).

After each spliceosome cycle, the U4 and U6 snRNAs are released separately and are recycled to the functional U4/U6 snRNP. The p110 (SART3) protein,

the homologue of yeast Prp24p, has been identified in the mammalian system which is required for recycling of the U4/U6 snRNP from singular U4 and U6 snRNPs (Bell *et al*., 2002). Protein p110 associates only transiently with U6 and U4/U6 snRNPs during the recycling phase of the spliceosome cycle.

Interestingly, recent biochemical evidence indicated that these U4/U6 snRNP-specific proteins are also associated with the HeLa U4atac/U6atac snRNP (Schneider *et al*., 2002).

U5 snRNP possesses a complex protein composition, with a sedimentation coefficient of 20S (Bach *et al*., 1989). In addition to the seven common Sm proteins and the U5 snRNA, 20S U5 snRNP consists of a total of eight particle-specific proteins, referred to as 15K, 40K, 52K, 100K, 102K, 116K, 200K and 220K proteins (Bach *et al*., 1989; reviewed in Will *et al*., 1993). Of these, all except the 40K protein have identifiable homologues in yeast (Stevens *et al*., 2001). Proteins common to the U5 snRNPs from human and yeast (in parentheses) include 220K (Prp8p), 200K (Brr2p), 116K (Snu114p), 100K (Prp28p), 52K (Snu40p/Lin1p), and 15K (Dib1p). While protein 102K is stably associated with U5 snRNP in human, the yeast homologue Prp6p is present in the U4/U6 snRNP particle (Makarov *et al.*, 2000; Abovich *et al.*, 1990).

Most of the U5-specific proteins display significant domain features, and play central roles in the splicing machinery (reviewed in Will and Lührmann, 1997). For example, protein 200K (Brr2p) and 100K (Prp28p) contain the RNA helicase domain (Lauber *et al.*, 1996; Teigelkamp *et al*., 1997). Protein 116K (Snu114p) is the sole GTPase identified in the spliceosome to date and is related to translation elongation factor EF-2 (Fabrizio *et al*., 1997; Stevens *et al*., 2001; Jurica and Moore, 2003). Protein 220K is the most highly conserved splicing factor among species, displaying 62% identity between the human and yeast proteins throughout the entire amino acid sequence. However, it contains little in the way of recognizable sequence motifs, so that its domain structure is unclear and it is difficult to make testable predictions about the biochemical activities of the protein. The functions of these proteins are discussed in sections 2.4.2 and 2.4.4 in detail. The structural study using X-ray

crystallography has indicated that the small protein 15K adopts a thioredoxin fold, and may thus be involved in the extensive protein-protein interactions in the spliceosome (Reuter *et al*., 1999). Protein 102K (Prp6p) contains multiple 34-amino-acid TPR motifs. TPR domains provide a structural unit of two antiparallel α helices that assemble to a platform for specific protein-protein interactions (reviewed in Blatch and Lässle, 1999). The 52K protein had not been characterised when I started this work.

Although the composition of U5 snRNP is well defined, our knowledge of structural organization of the particle is still limited. Interestingly, several studies attempting to elucidate the protein-snRNA interactions in the U5 snRNP have indicated that the protein-protein interactions may contribute to the formation of U5-snRNP (Bach and Lührmann, 1991; Dix *et al*., 1998; Black and Pinto, 1989). First, the interaction of the U5-specific proteins with U5 snRNA was investigated by comparison of the differential accessibility towards nuclease and dimethylsulfate of defined regions of U5 snRNA in purified 20S and 10S U5 snRNPs (Bach and Lührmann, 1991). The results indicated that only the central part of stem/loop I of U5 snRNA contains binding sites for U5-specific proteins, suggesting that most of the U5-specific proteins may be bound to U5 snRNP via protein-protein interactions. These data support and extend the previous investigation of Black and Pinto, who performed comparative structural probing of naked U5 snRNA and U5 snRNP in HeLa nuclear extracts with chemical reagents. Consistently with this, photo-crosslinking demonstrated that only limiting proteins are in contact with U5 snRNA in the reconstituted yeast U5 snRNP (Dix *et al*., 1998). Among them, the Prp8p and Snu114p are significant candidates. Some additional data further support the observation that protein-protein interactions are the main building force in the U5 snRNP. For example, U5 protein 40K and three large proteins 116K, 200K and 220K can form a stable heteromeric complex in the absence of U5 snRNA (Achsel *et al*., 1998). Co-immunoprecipitation experiments showed that the U5-102K protein does not bind to either U5 snRNA or U5 core snRNP when this contains only U5 snRNA and Sm proteins; instead, it binds stably to the 20S U5 snRNP. This result suggested that one or more of the U5-specific proteins are required for the association of the protein 102K (Makarov *et al*., 2000).

The U4/U6 and U5 snRNPs must be pre-assembled *in vivo* as a **U4/U6.U5 tri-snRNP** particle in order to join the spliceosome. Within the tri-snRNP particle there is extensive base pairing between the U4 and U6 snRNAs, leading to the formation of two stem structures, and the U5 snRNP appears to be associated with the U4/U6 snRNP through protein-protein interactions. The tri-snRNP particle can be purified *in vitro* from HeLa nuclear extract and yeast cellular extract at low salt concentration (150 mM NaCl) and sediments at 25S in glycerol gradients (Behrens and Lührmann, 1991; Fabrizio *et al*., 1994; Gottschalk *et al*., 1999; Stevens and Abelson, 1999). The protein composition of the tri-snRNP is well established. In human cells, the tri-snRNP contains at least 15 specific proteins, and many of these are essential for splicing (reviewed in Will and Lührmann, 1997, 2001). The U4/U6-snRNP proteins 15.5K, 61K, the 20K·60K·90K complex and the U5-snRNP proteins including 220K, 200K, 116K, 102K, 100K, 40K and 15K are present in this particle. In addition, three other proteins (110K, 65K and 27K) are more stably associated with the 25S tri-snRNP particle, and are thus referred to as tri-snRNP-specific proteins. Interestingly, all three tri-snRNP-specific proteins contain an N-terminal arginine-serine-rich (RS) domain (Fetzer *et al*., 1997; Makarova *et al*., 2001). Of these, the two largest proteins are shown to be required for the recruitment of tri-snRNP to the pre-spliceosome (Makarova *et al*., 2001).

While some information has been accumulated on protein-protein and protein-RNA interactions within the individual U4/U6 or U5 snRNP, little is known about the interactions that bridge the U4/U6 and U5 snRNPs. There is no indication of RNA-RNA interactions between the two particles (Black and Pinto, 1989). Consistently with this, the tri-snRNP dissociates into U4/U6 and U5 snRNPs at monovalent salt concentrations between 300 and 400 mM, suggesting that tri-snRNP formation is mediated by protein-protein and/or protein-RNA interactions. However, apart from a crosslink observed between the yeast U5 protein Prp8p and U6 snRNA (Vidal *et al*., 1999), there is a paucity of information about interactions that potentially mediate the association of the U5 and U4/U6 snRNPs. In yeast, Prp6p appears to be important for the interaction of the U4/U6 and U5 snRNPs. Mutation of the *PRP6* gene inhibits tri-snRNP accumulation, while accumulation of the individual U4/U6 and U5 snRNPs is not

affected (Galisson and Legrain, 1993). This suggests that the Prp6p protein is not needed for the integrity and hence stability of the U4/U6 or U5 snRNPs, but only for tri-snRNP formation. Consistently with this, recent data from a mammalian system showed that U5-102K protein prepared by translation *in vitro* binds to purified 13S U4/U6 snRNP (Makarov *et al*., 2000). To substantiate the hypothesis that 102K protein bridges the U4/U6 and U5 snRNPs, it will be important in future experiments to elucidate in detail the interaction partners of 102K in the tri-snRNP particle.

2.3.4 U4/U6.U5 tri-snRNP proteins and retinitis pigmentosa

The importance of tri-snRNP in human disease was realised recently as a result of several studies (Vithana *et al*., 2001; McKie *et al*., 2001; Chakarova *et al*., 2002). In the year 2001, McKie *et al.* showed that mutations of several highly conserved residues in the C-terminal part of 220K/hPrp8p correlated with autosomal dominant retinitis pigmentosa (RP13), while Vithana *et al.* found that mutations in another splicing factor, 61K/hPrp31p, are linked to the autosomal dominant gene RP11. In the following year, mutations of the third gene encoding 90K/hPrp3p were implicated in RP18 (Chakarova *et al*., 2002). It was recently found that two missense mutations (A194E, A216P) in the 61K gene (*PRPF31*) may affect splicing by impeding the translocation of 61K protein into the nucleus. The splicing defect of A216P was further demonstrated by the finding that 61K with A216P failed to complement fully the functional deficiency in a temperature-sensitive, Prp31p-deficient yeast strain at the high restrictive temperature (higher growth rate, high demand for splicing) (Deery *et al*., 2002). It was argued that rod photoreceptors may have a high demand for splicing of important molecules such as opsin mRNA, and subtle defects in splicing due to loss of function of protein 61K may lead to a disease in this system (RP). However, the molecular mechanism of retinitis pigmentosa caused by mutations in these three tri-snRNP proteins is unclear at present.

2.4 The Dynamic Nature of the Spliceosome

2.4.1 The spliceosome cycle

Studies in both the yeast and mammalian systems have revealed a cycle of spliceosome assembly and disassembly on a pre-mRNA (figure 2.4) (Cheng and Abelson, 1987; Konarska and Sharp, 1986; Konarska and Sharp, 1987). The first step that commits a pre-mRNA to the splicing pathway is recognition of the 5' SS by the U1 snRNP, initiating the early (E) complex in mammals. This initial step does not require energy (ATP) for formation. U1 is the only component that can bind in the absence of ATP, and its interaction with the pre-mRNA is in general required for the other snRNPs to bind. After the association of the U1 snRNP with the 5' SS, the U2 snRNP recognizes and binds to the branchpoint to form complex A. Addition of the U2 snRNP is the first energy-dependent step in the splicing pathway, and binding of the U2 snRNP is mediated in part by the U1 snRNP as well as additional non-snRNP factors that bridge the two components. After U2 snRNP binding, the U4/U6.U5 tri-snRNP particle joins the A complex to form complex B. Complex B is structurally rearranged to form the catalytically activated spliceosome (B*), which subsequently catalyses the first transesterification reaction, generating complex C. After the second step of splicing, the mRNA is released, the post-spliceosomal complex dissociates, and the snRNPs are recycled for new rounds of splicing.

Thus, spliceosome assembly goes through many intermediate stages, the most stable of which (e.g., the A, B, C complexes) can be detected biochemically, for example, by native gel electrophoresis. Some additional stable intermediates, such as the immunoaffinity-purified activated spliceosome B* and pre-catalytic complex, which contains intact U4/U6.U5 tri-snRNP but lacks U1 snRNP (designated BΔU1) have been also reported recently (Makarov *et al*., 2002; Makarova *et al*., 2004).

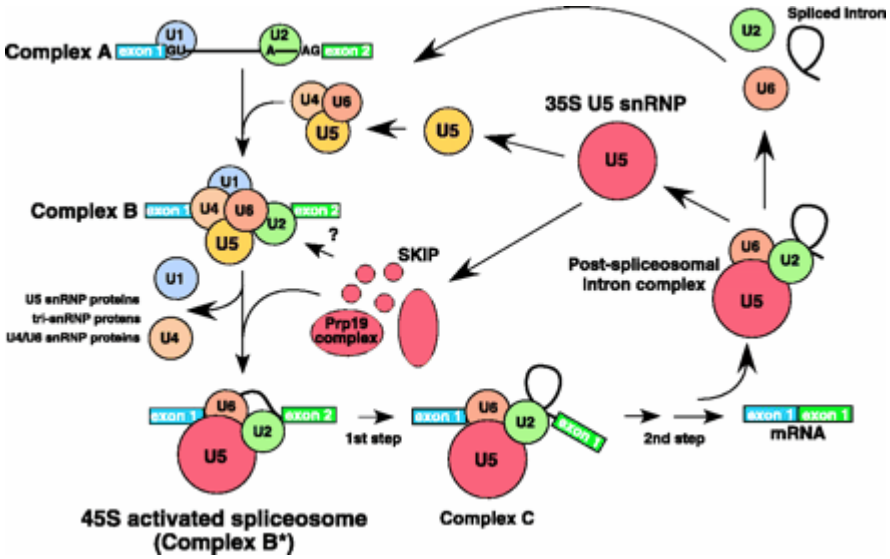


Figure 2.4. Schematic diagram of spliceosomal and snRNP remodelling events. In the early phase of spliceosome formation, U1 snRNA base pairs with the 5' splice site, and U2 snRNA interacts with the branch site to form the pre-spliceosomal A complexes. Spliceosome assembly is completed by the addition of the 25S U4/U6.U5 tri-snRNP particle, forming spliceosomal complexes B and C. Complexes with a red U5 snRNP contain the SKIP protein (from Makarov *et al*., 2002).

2.4.2 Dynamics of RNA-RNA interactions in the spliceosome

The most decisive step during the spliceosome maturation process is the conversion of complex B into the catalytically activated spliceosome B*. This occurs after the recruitment of U4/U6.U5 tri-snRNP into pre-spliceosome.

Once all five snRNPs are present, the complex undergoes a series of rearrangements. The interaction between the U1 snRNA and the 5' SS is disrupted, and the U1 snRNP particle is released from the complex. Similarly, the base pairing between the U4 and U6 snRNAs is also disrupted, allowing the release of the U4 snRNP and new base-pairing to form between the U2 and U6 snRNAs (Villa *et al*., 2002). The resulting RNA network forms the catalytic core of the spliceosome (figure 2.5). U2 and U6 snRNAs interact with the branch point and 5' splicing site respectively, and provide a structural basis for juxtaposing the branch site and 5' splicing site for the first catalytic step. U5 snRNA bears a highly conserved stem-loop that is implicated in aligning the exons for the second catalytic step (O'Keefe *et al*., 1996). The U6 snRNA is proposed to be the actual catalytic entity, indicating that the spliceosome may

be a true ribozyme. Most RNA enzymes require metal ions as cofactors for catalysis, and it was recently determined that specific bases in the U6 snRNA bind Mg^{2+} ions, a strong indication that this molecule may indeed be the catalytic agent (Collins and Guthrie, 2000; Yean *et al.*, 2000). A protein-free complex of the U2 and U6 snRNAs is able to bind a small RNA and activate the attack of a branch point adenosine on a catalytically important region of U6. This reaction is similar to the first step of splicing, further supporting the potential catalytic activity of the U2 and U6 snRNAs (Valadkhan and Manley, 2001).

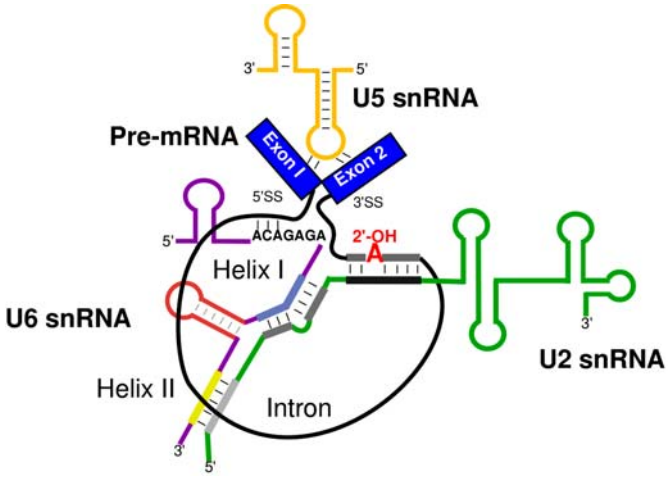


Figure 2.5. Model of the catalytic RNA core of the spliceosome. The U2 and U6 snRNAs are extensively base-paired with one another and with the branch site sequence and 5' splice site respectively. An invariant loop sequence in the U5 snRNA interacts initially with Exon 1, and then also with Exon 2 after the first catalytic step (modified from Dr. P. Fabrizio in the laboratory of Prof. Lührmann).

In addition to the U6 and U2 snRNA molecules, Prp8p (220K in human) has also been proposed to play a critical role at the catalytic core of the spliceosome. Prp8p is a component of the U5 snRNP in both mammalian and yeast systems and is the most highly conserved splicing factor among species. Prp8p is unique in making extensive contacts with U5 and U6 snRNAs and with the pre-mRNA substrate at the splice sites and intron branch site (Wyatt *et al.*, 1992; Teigelkamp *et al.*, 1995; Dix *et al.*, 1998; Vidal *et al.*, 1999). Prp8p probably plays a role in stabilizing the interactions between the U5 snRNA loop sequence and the exons (Teigelkamp *et al.*, 1995) and may also be responsible for juxtaposing the U5 loop with the rest of the catalytic core. Mutational studies

of Prp8p (Kuhn *et al*., 2002; Collins and Guthrie, 1999; Siatecka *et al*., 1999; Kuhn and Brow, 2000; Query and Konarska, 2004) also support the view that this protein is intimately involved in the functions of the catalytic core of the spliceosome.

2.4.3 Dynamic changes of protein components of U4/U6.U5 tri-snRNP during the spliceosome cycle

The restructuring that the U4/U6.U5 tri-snRNP undergoes during the spliceosomal cycle affects not only its RNA components but its proteins as well. This has become evident from proteomic analysis of purified spliceosomal complexes at defined functional stages. Mass spectrometry of a purified pre-catalytic spliceosomal B complex containing U2, U4/U6 (still base-paired) and U5 snRNAs but lacking U1 snRNA (therefore termed BΔU1) revealed that all of the U4/U6- and U5-specific proteins (with the single exception of U5-100K) are stably bound at this spliceosomal assembly stage (Makarov *et al*., 2004). In striking contrast, all of the U4/U6-specific proteins as well as the U5-15K protein are significantly destabilized upon transformation of complex B into the catalytically active spliceosome (complex B*) such that these proteins are no longer retained in complex B* when this is isolated under the same stringent conditions as complex BΔU1 (Makarova *et al*., 2002). Of the tri-snRNP-specific proteins, only 110K remains stably bound in complex B*. Recently, a novel 35S RNP particle containing U5 snRNA was identified in nuclear extracts from HeLa cells, and its protein composition was characterized by mass spectrometry. This particle contained – in addition to the established U5-specific proteins (220K, 200K, 116K and 40K) – a group of splicing factors (such as the human Prp19 protein complex) that are stably integrated into the spliceosome during the transformation of complex B into the catalytically activated B* spliceosome, indicating that the 35S U5 snRNP complex represents a post-spliceosomal disassembly intermediate of the splicing machinery (Makarov *et al*., 2002; Makarova *et al.*, 2004). Interestingly, several proteins which are stoichiometrically present in 20S U5 snRNPs – namely, the U5-specific 102K, 100K and 15K proteins – are completely absent in the 35S U5 RNP particle (Makarov *et al*., 2002). Thus, the protein composition of the U5 snRNP changes within the framework of the spliceosome, and the resulting U5 particle has to be

extensively reformed before it can build a U4/U6.U5 tri-snRNP and then re-enter the spliceosome. Very little is known about these remodelling steps or the proteins possibly involved.

U4/U6.U5 tri-snRNP	BΔU1 (Makarova *et al*., 2004)	45S activated spliceosome (B*) (Makarov *et al*., 2002) **	35S U5 snRNP (Makarov *et al*., 2002) **
Sm Proteins	Sm Proteins	Sm Proteins	Sm Proteins
B,D1,D2,D3,E,F,G	B,D1,D2,D3,E,F,G	B,D1,D2,D3,E,F,G	B,D1,D2,D3,E,F,G
LSm proteins	LSm proteins	LSm proteins	LSm proteins
LSm 2,3,4,5,6,7,8	LSm 3,4,7,8	LSm 2,3	-
U5 snRNP	**U5 snRNP**	**U5 snRNP**	**U5 snRNP**
220K	220K	220K	220K
200K	200K	200K	200K
116K	116K	116K	116K
40K	40K	40K	40K
102K	102K	102K	-
15K	15K	-	-
100K	-	-	-
tri-snRNP	**tri-snRNP**	**tri-snRNP**	**tri-snRNP**
110K	110K	110K	-
65K	-	-	-
27K	-	-	-
U4/U6 snRNP	**U4/U6 snRNP**	**U4/U6 snRNP**	**U4/U6 snRNP**
90K	90K	-	-
60K	60K	-	-
20K	20K	-	-
61K	61K	-	-
15.5K	15.5K	-	-

Table 2.2. The protein composition of the U4/U6.U5 tri-snRNP in the complex BΔU1, the 45S activated spliceosome (B*), and 35S U5 snRNP. ** A number of proteins (including CDC5/Prp19 complex) recruited in B* and associated with 35S U5 snRNP are not showed in this table.

2.4.4 The roles of tri-snRNP proteins in the activation of spliceosome

Although the splicing reactions themselves do not directly require energy input, remodelling of a network of RNA-RNA and RNA-protein interactions in the spliceosome consumes a number of ATP molecules. These rearrangements are catalysed by several ATP-dependent RNA helicases and may also involve the activity of a GTPase related to the translation elongation factor EF-2. The largest class of energy-requiring proteins involved in splicing belongs to the DEXD/H box family: these proteins are commonly known as RNA helicases but

in some cases may act as RNPases (ribonucleoproteinases), which disrupt RNA-protein complexes in the spliceosome (Schwer, 2001).

While the U4/U6 snRNP part of the tri-snRNP contributes a substantial component of the catalytic RNA network (described in section 2.4.2), the driving force of catalytic activation and the rearrangements of the spliceosome all appear to be U5 snRNP residents. U5 snRNP contains three NTPases: 100K (Prp28p in yeast), 200K (Brr2p) and 116K (Snu114p). The NTPases of the U5 snRNP are involved in the critical switch in which U1 is replaced by U6 at the 5´ splice site (figure 2.6). This is an important stage in spliceosome activation, which contributes to the fidelity of 5´ splice site recognition. The two unwinding events that disrupt base-pairing between U1 snRNA and the 5´ splice site, and between U4 and U6 snRNAs, allow the 5´ splice site, U6 and U2 catalytic core structure to form (described in section 2.4.2; Brow, 2002). It has been shown that Prp28p in yeast has a role in destabilizing the U1 snRNA interaction with the 5´ splice site. This DEXD/H box protein may unwind the helix formed between the 5´ splice site and U1 snRNA (Staley and Guthrie, 1999). Alternatively, it is a prime candidate to act as an RNPase. Under normal conditions, Prp28p is an essential protein in yeast, but if the U1C protein (a factor that stabilizes the interaction between U1 and the 5´ splice site) is mutated, then Prp28p becomes dispensable (Chen *et al*., 2001). This suggests that Prp28p participates in disrupting the U1C interaction, either by acting directly against the protein or by disrupting the double helix between the U1 snRNA and the 5´ splice site that forms its site of interaction.

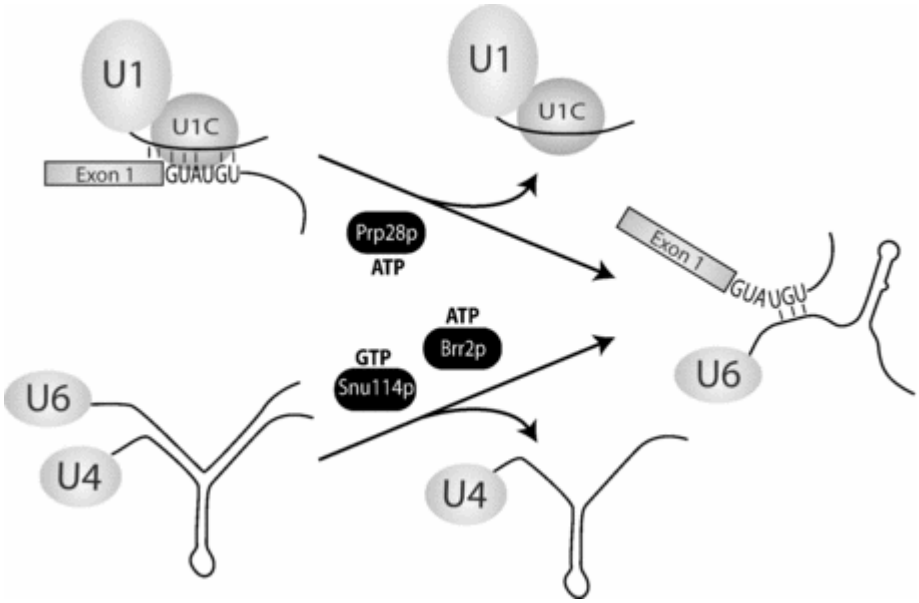


Figure 2.6. Model of the 5' splice site switch. During spliceosome assembly, the 5' splice site is initially recognized by the U1 snRNP. This interaction is later replaced by U6, which must base-pair with the 5' splice site before the first catalytic step. The unwinding of U4:U6 is also required for the spliceosome to become active. These events occur concurrently and allow the release of U1 and U4 snRNPs from the spliceosome. The roles that the U5 snRNP components, Prp28p, Brr2p and Snu114p, are hypothesized to perform during activation of the spliceosome are indicated: Prp28p displaces U1 (and the U1C protein, see text) from the 5' splice site presumably utilizing ATP, whereas Brr2p and Snu114p are involved in the unwinding of U4 and U6 snRNAs, a process requiring ATP and GTP. This results in free U6 paired with the 5' splice site. U6 is also capable of forming an intramolecular stem loop, a structure important for catalysis, after release from U4. The 5' splice site consensus from *Saccharomyces cerevisiae* is used with base-pairing to U1 and U6 snRNAs indicated. Adapted from Turner *et al*., 2004.

Disruption of the interaction between U1 snRNA and the 5´ splice site is accompanied by the release of U6 from U4 (figure 2.6). The U4/U6.U5 tri-snRNP contains the base-paired U4:U6 structure that prevents the formation of catalytic core structures until the correct time. *In vitro* work has shown that the 200K/hBrr2p protein is capable of unwinding RNA helices including a base-paired U4:U6 complex (Laggerbauer *et al*., 1998). Furthermore, it has been shown that a mutation in the ATPase domain of the yeast homologue Brr2p (*brr2-1*) inhibits the ATP-dependent disruption of U4/U6.U5 tri-snRNPs in yeast cell extracts (Raghunathan and Guthrie, 1998). These studies suggest that 200K protein is the motor that drives the release of U6 before formation of the U6:5´-splice-site and U6:U2 structures. Both Snu114p and Prp8p (a large U5 snRNP protein) are supposed to control the action of Brr2p (Bartels *et al*., 2002,

2003; Kuhn *et al*., 1999, 2002). Elegant studies using an Snu114p mutant that switched specificity from GTP to XTP allowed dissection of this regulatory role of Snu114p (Bartels *et al*., 2003). It was shown that stalled complexes would only unwind the U4:U6 helices when supplied with hydrolysable XTP, implying that Snu114p has a role either in unwinding U4:U6 or, more probably, in controlling the action of Brr2p. Prp8p has been implicated in several aspects of spliceosome remodelling and activation, in addition to putative cofactor activity in catalysis (described in section 2.4.2). These aspects include a central role in governing the activities of the Brr2p and Prp28p RNA-dependent ATPases (Kuhn *et al*., 1999, 2002). A mutation in *PRP8*, *prp8-201*, was identified as a suppressor of the U4-cs1 mutation. U4-cs1 extends the base pairing between the U4 and U6 snRNAs into the ACAGA box of the U6 snRNA, preventing its interaction with the 5' SS. This led to the proposal that Prp8p may trigger the unwinding of the U4 and U6 snRNAs only after correct recognition of the 5' SS by the ACAGA box of U6. Like many other *PRP8* mutations, this alteration in Prp8p maps to the C-terminal domain of the protein (Kuhn *et al*., 1999). Additional genetic interactions indicate that Prp8p may regulate Brr2p, the putative U4/U6 helicase, and Prp28p, the putative U1/5'-SS helicase, in order to control the rearrangements that they promote and thus to allow spliceosome activation (Kuhn and Brow, 2000; Kuhn *et al*., 2002).

Interestingly, experimental evidence also indicated a role for the 15.5K/Snu13p and Prp4p proteins in the transition of fully assembled spliceosome toward an active form (Ayadi *et al*., 1997; Nottrott *et al*., 1999).

2.5 The Aims of the Current Studies

The 25S U4/U6.U5 tri-snRNP, pre-assembled from U4/U6 and U5 snRNPs, is one of the major building blocks of the spliceosome. As mentioned in section 2.3.3, this particle contains – in addition to the U4, U6 and U5 snRNAs – ca. 30 distinct proteins which make up approximately 80% of the tri-snRNP's total mass. Most of the tri-snRNP proteins exhibit significant domain features, possibly providing hints of the potential functions of these proteins. For instance, the TPR repeats of U5-102K, the WD40 domain of U4/U6-60K and U5-40K, the

cyclophilin fold of U4/U6-20K, and the thioredoxin fold of U5-15K all predict a rich diversity of protein-protein interactions and protein modifications within the tri-snRNP particle (see section 2.3.3 for detailed description). The following evidence also suggests that the protein-protein interactions might be the major building force of the tri-snRNP. First, the stable heteromeric complexes 20K•60K•90K and 220K•200K•116K•40K could be isolated separately *in vitro* in the absence of snRNAs (see section 2.3.3.2). Second, protein U5-102K binds *in vitro* directly to the latter heterotetrameric complex (Makarov *et al*., 2000). However, this is probably only part of protein-protein interaction map in such a large complex. For example, little is known about the interactions that bridge the U4/U6 and U5 snRNPs although it has been suggested that the protein-protein interactions may dominate this connection (see section 2.3.3.2).

There is evidence that the DExH-box RNA helicase U5-200K and the GTPase U5-116K are driving forces behind the disruption of the U4/U6 snRNA helices (see section 2.4.4). However, the mechanism – that is, how these U5-specific proteins actually disrupt the U4/U6 RNA helices, i.e. by directly contacting the RNAs or more indirectly, by modulating the activity of U4/U6 snRNA binding proteins via a network of protein-protein interactions – is still not understood.

Therefore, the knowledge of protein-protein interactions appears to be important for understanding the mechanisms of the tri-snRNP assembly and its role in the activation of spliceosome and catalysis of splicing. A major part of my PhD work has been to identify the protein-protein interactions within the individual U4/U6 and U5 snRNPs, as well as protein contacts between these two snRNPs in the tri-snRNP particle, and protein domains involved. The combination of all the interaction data will allow me to draw up a comprehensive protein-protein interaction map, and thus aid in the understanding of the structure and functional mechanisms of the tri-snRNP complex and the dynamic nature of the spliceosome. In addition, the knowledge of protein-protein interaction will provide a basis for protein co-expression for the structure study by X-ray crystallography.

For this purpose, the yeast two-hybrid (Y2H) system and *in vitro* biochemical methods were first established. The Y2H system was developed by Fields and Song (1989) and is a sensitive method for identifying protein-protein interactions *in vivo*. Briefly, the two-hybrid system takes advantage of the properties of the *GAL4* transcription factor, which has separable domains for DNA binding (amino acids 1-147) and transcriptional activation (amino acids 768-881). The proteins being studied are expressed as hybrid constructs with a DNA binding domain or activation domain. If the two proteins interact with each other, both domains of *GAL4* transcription factor are brought into close physical proximity and allow trans-activation of the reporter genes, e.g. *HIS3* and *ADE2*. GST pull-down assays and co-immunoprecipitation assays are the two methods used for *in vitro* binding studies in this work. In the most case, one of the tested protein pair has to be expressed in *E. coli* and purified as a functional recombinant protein. Mutational analysis was used for identified the interacting domains.

The same methods are also applied in the following other studies. The interactions between tri-snRNP proteins and the U6-associated protein p110 and the U2-associated splicing factor SPF30/SMNrp are investigated in order to elucidate the functions of tri-snRNP proteins in the recycling of U4/U6 and the recruitment of tri-snRNP to the pre-spliceosome. Furthermore, I am interested in the role of tri-snRNP proteins 220K, 90K and 61K in retinitis pigmentosa, in the context of protein-protein interactions. To understand the function of protein 200K, the N-terminal 434-amino-acid fragment of 200K was used as a prey in a yeast two-hybrid screen to identify the interaction partners of 200K.

3 Materials and Methods

3.1 Materials

3.1.1 Chemicals

Acrylamide solution	Roth, Karlsruhe
Adenine sulfate	QBiogene, Heidelberg
Agarose	Invitrogen, USA
Ammonium persulfate (APS)	Merck, Darmstadt
Ampicillin	Sigma, Deisenhofen
Amplify™	Amersham Biosciences, Freiburg
Bovine serum albumin (BSA), acetyliert	Sigma, Deisenhofen
Bradford solution	Biorad, München
Bromophenol blue	Merck, Darmstadt
B-Per™ Bacterial Protein Extraction Reagent	Pierce, UK
Carrier DNA	Clontech, Heidelberg
Chloramphenicol	Boehringer, Mannheim
Coomassie Brillant Blau R250	Serva, Heidelberg
Dimethyl sulfate (DMS)	Fluka, Schweiz
Dimethyl sulfoxide (DMSO)	Sigma, Deisenhofen
Dithioerythrol (DTE)	Sigma, Deisenhofen
Dithiothreitol (DTT)	Roth, Karlsruhe
Sodium dodecyl sulfate (SDS)	Merck, Darmstadt
Acetic acid	Merck, Darmstadt
Ethanol	Merck, Darmstadt
Ethylendiamine tetra-acetic acid (EDTA)	Roth, Karlsruhe
Ethidium bromide solution	Roth, Karlsruhe
Glutathion (reduced)	Sigma, Deisenhofen
Glycin	Merck, Darmstadt
Glycerol	Merck, Darmstadt
Guanidinium hydrochloride	Fluka, Schweiz
Heparin	Sigma, Deisenhofen
HEPES	Calbiochem, USA
Imidazole	Merck, Darmstadt
Isopropyl-β-D-thiogalactoside (IPTG)	Sigma, Deisenhofen
Kanamycin	Boehringer, Mannheim
Lysozyme	Boehringer, Mannheim
Milk powder, instant	Cenovis GmbH, Radolfzell
Magnesium chloride	Merck, Darmstadt
Methanol	Merck, Darmstadt
NDSB-256[3-(1-Pyridino)-1-propan sulfate]	Calbiochem, USA
1 kb DNA ladder	Invitrogen, USA
Phenylmethylsulfonylfluoride (PMSF)	Boehringer, Mannheim
p-Nitrophenyl α-D-Galactopyranoside (PNP-α-Gal)	Sigma, Deisenhofen
Polyethylenglycol 3350 (PEG3350)	Sigma, Deisenhofen
Ponceau S	Serva, Heidelberg
Potassium chloride	Merck, Darmstadt
Precision protein standard marker	Biorad, München
Roti-Phenol/Chloroform	Roth, Karlsruhe
Silver nitrate	Merck, Darmstadt
Sodium chloride	Merck, Darmstadt
Sodium thiosulfate	Merck, Darmstadt
N, N, N', N'-Tetramethylethylendiamin (TEMED)	Sigma, Deisenhofen
3-Amino-1,2,4-Triazole (3-AT)	Sigma, Deisenhofen
tRNA, *E. coli*	Boehringer, Mannheim
Tris-(hydroxymethylen)aminomethan (TRIS)	Roth, Karlsruhe
Triton X-100	Sigma, Deisenhofen

Tween 20	Sigma, Deisenhofen
X-α-Gal (5-Bromo-4-Chloro-3-indolyl α -D-galactopyranoside)	Clontech, Heidelberg
X-Gal (5-Bromo-4-Chloro-3-indolyl β-D-galactopyranoside)	Clontech, Heidelberg
Xylene cyanol FF	Fluka, Schweiz
Urea	Merck, Darmstadt

3.1.2 Media (Qbiogene, Eschwege, Germany)

CSM-Leu, CSM-Trp, CSM-His, CSM-Ade, CSM-Ura, CSM-Leu-Trp, CSM-Leu-Trp-His, CSM-Leu-Trp-His-Ade, DOB, DOBA, NYZ^{+} Broth, SOC Broth, SOB Broth, LB Broth, LB Agar, YPD Broth, YPD Agar

3.1.3 Nucleotides, radionucleotides and amino acids

Nucleotide

Deoxynucleoside-5'-Triphosphate, 100 mM (dATP, dCTP, dGTP, dTTP)

Nucleoside-5'-Triphosphate, 100 mM (ATP, CTP, GTP, UTP)

Radionucleotide (Amersham Biosciences, Freiburg)

[α-^{32}P]-dATP	3000 Ci/mmol; 10 Ci/l
[γ^{32}P]-ATP	6000 Ci/mmol; 10 Ci/l
[α-^{32}P]-UTP	3000 Ci/mmol; 10 Ci/l
[γ-^{32}]-pCp	3000 Ci/mmol; 10 Ci/l

Amino acid (Amersham Biosciences, Freiburg)

L-[^{35}S] Methionine 1000 Ci/mmol

3.1.4 Antibodies

Anti-c-myc-peroxidase	Boehringer, Mannheim
Penta.His Antibody	Qiagen, Hilden
Monoclonal antibody, HA.11	Covance, USA
c-Myc monoclonal antibody	Clontech, USA
HA-tag polyclonal antibody	Clontech, USA
Polyclonal anti-110K rabbit antibody	Lührmann lab, α-pep 110
Polyclonal anti-60K rabbit antibody	Lührmann lab

3.1.5 Enzymes and inhibitors

AMV Reverse Transcriptase (15 U/µl)	Amersham Biosciences, Freiburg
DNase I	Roche, Mannheim
Pfu DNA Polymerase (2,5 U/µl)	Stratagene, Heidelberg
Proteinase Inhibitor Cocktail Complete™, EDTA-free	Roche, Mannheim

Proteinase K (10 mg/ml)	Sigma, Deisenhofen
Ready-To-Go T4 DNA ligase	Amersham Biosciences, Freiburg
Restriction Endonucleases	New England Biolabs, France
RNAsin (40 U/µl)	Promega, USA
SP6 RNA Polymerase (20 U/µl)	Promega, USA
Taq DNA Polymerase (5000 U/µl)	Promega, USA
T4 DNA Ligase (400 U/µl)	New England Biolabs, Frankfurt
T4 Polynucleotide Kinase (20 U/µl)	New England Biolabs, Frankfurt
T3 RNA Polymerase (17 U/µl)	Promega, USA
T7 RNA Polymerase (50 U/µl)	Promega, USA
Thrombin Protease	Amersham Biosciences, Freiburg

3.1.6 Oligonucleotide primers for PCR (Table 3.1)

Protein	Oligo	Description	Sequence (5'→3')
15.5K	HPV-140	full-length	CCTTCGAACATATGACTGAGGCTGATGTGAAT
	HPV-141		CCTTCGAACTCGAGTTAGACTAAGAGCCTTTCAAT
20K	HPV-142	full-length	CCTTCGAACATATGGCGGTGGCAAATTCAAGT
	HPV-143		CCTTCGAACTCGAGCTACATCTCCCCACACTGCGA
27K	HPV-148	full-length	CCTTCGAACATATGGGTCGCAGTCGCAGCCGC
	HPV-149		CCTTCGAACTCGAGTCATGCAATGAAATCCAAAGG
40K	HPV-82	full-length	CCCGGATCCCATATGATAGATCAGCAGAAGCGT
	HPV-83		CCCCTCGAGTCACTGAATCTCTCCCATATAC
52K (128/129)	HPV-128	full-length	CCTTCGAACATATGCCAAAGAGGAAAGTGACC
	HPV-129		CCTTCGAACTCGAGTCAGGTGTAGAGGTCAAAGTC
52N(128/243)	HPV-243	truncation	CCCCTCGAGTCACAACTCCTCAGCGAACATGTC
52C(242/129)	HPV-242		CCCCATATGGCGGAGGAGGAACTGGAGACC
60Kmut	HPV-144	full-length	CCTTCGAACATATGGCTTCCTCGCGAGCCTCT
	HPV-145		CCTTCGAACTCGAGCTATTCCAGCATCCAC AGC TT
60Kwt	HPV-324	full-length	CCCGGATCCATATGGCTTCCTCGCGAGCCTCTTCC
	HPV-325		CCCCTCGAGCTATTCAGCCATCCACAGCTTGAA
61K	HPV-345	full-length	CCCCATATGTCTCTGGCAGATGAGCTC
	HPV-346		CCCCTCGAGTCAGGTGGACATAAGGCCACTC
A194E	HPV-384	mutagenesis	GAGGCCTGCGACATGGAGCTGGAGCTGAACGCCTCC
	HPV-385		GCTTGGAGGCGTTCAGCTCCAGCTCCATGTCGCAGGCC
A216P	HPV-386	mutagenesis	GTGGAGTCCCGGATGTCCTTCATCCCACCCAACCTGTCC
	HPV-387		GATGGACAGGTGGGTGGGATGAAGGACATCCGGGACTCC
90K	HPV-146	full-length	CCATCGATCATATGGCACTGTCAAAGAGGGAG
	HPV-147		CCATCGATCTCGAGTCAATCAGTGGACTCTAACAC
90K_N194	LSB-1	truncation	GGTCTCCCATGGAATTCATGGCACTGTCAAAGAGGGAG

90K_N194	LSB-2	truncation	AAACTCGAGGTACCTTAAGTGGCAGCCTGGGAGGGCTG
90K_N309(1/8)	LSB-8	truncation	AAACTCGAGGTACCTTAGGTATTGGATTCCATGTCTTC
90K_N442(1/6)	LSB-6		AAACTCGAGGTACCTTACTTGGTAAGATATACTCC
90K_C267	LSB-5	truncation	GGTCTCCCATGGAATTCCTTGTTGAACATCCAGCCCAGC
	LSB-4		AAACTCGAGGTACCTCAATCAGTGGACTCTAACAC
90K_C381(7/4)	LSB-7	truncation	GGTCTCCCATGGAATTCGAAGACATGGAATCCAATACC
90K_C489(3/4)	LSB-3		GGTCTCCCATGGAATTCATGAATGATGCCATTGAG
P493S	HPV-359	mutagenesis	AAGCTGTTCAAGACCTCACGAAGGTAGAAGCCC
	HPV-360		GTGGGCTTCTACCTTCGTGAGGTCTTGAACAGC
T494M	HPV-347	mutagenesis	AAGCTGTTCAAGACCCCATGAAGGTAGAAGCCC
	HPV-348		GTGGGCTTCTACCTTCATGGGGTCTTGAACAGC
102K	HPV-235	full-length	CCCCCGCGGCATATGAACAAGAAGAAGAAACCGTTC
	HPV-237		CCCGTCGACTCAGAAGGTGTTCTTGATGCGGCCGGC
102K-NTD (235/272)	HPV-272	truncation	CCCGTCGACTCACTTCTTGATATCATTGATGTC
102K-TPR_M (273/278)	HPV-273	truncation	AAACCCCATATGGCGCGACTGCTCCTCAAGTC
	HPV-278		CCCGTCGACTTAGGACTCCCGAGTGCCATGG
102K-TPR_C (279/237)	HPV-279	truncation	AAACCCCATATGCTGGAAGCACTCCTGCAG
110K-RS	HPV-318	truncation	CCCGAATTCCATATGGGGTCGTCCAAGAAG CAT
	HPV-319		CCCCTCGAGTCAGCTAGTTTTGGAGCTGGCAGC
110K-ΔRS	HPV-320	truncation	CCCGAATTCCATATGTCAGGCGATGCCTCCTCACTC
	HPV-321		CCCCTCGAGTCACTTGGTGATGGTGTTCGC
116K (73/74)	HPV-73	full-length	CCCGGATCCCATATGGATACTGACTTGTATGATG
	HPV-74		CCCCTCGAGTCACATGGGATAATTGAGCAC
116-1(73/88)	HPV-88	truncation	CCCCTCGAGTCAGCCAAGCTCGTCTAGGGTCCG
116-2	HPV-76	truncation	CCCCATATGCGACACATTGTGGATGAGGTC
	HPV-89		CCCCTCGAGTCAAGGCTCAGCAATCATGGTGAT
116-3(78/74)	HPV-78	truncation	CCCCATATGCTTGATGGCCTGCGCAAG
200K	HPV-29	cDNA cloning from library	CAGAGGTCCTGTCTTAGAAGCTTT
	HPV-30		GTGCACTGACAACTCCAACTTGGG
	HPV-31		ATGTGTCCTCTGCGCCAGTTCCGG
	HPV-32		ATTGGGGTTCTGTGTCATGCGGCG
	HPV-33		GCCAAGCCTGTGTACCATGCTATC
	HPV-34		CCCCTCGAGTCAATCTGAATCACTGTCTGTCTC
	HPV-52		AGGGTCAACACGTAGAAAGGTGGC
200-1	HPV-63	truncation	CCGCGGCCGCATATGGCGGATGTAACCGCCCGTAGTC
	HPV-50		CCCTCATGCCTGGGCATACTTTGGCAGCTT
200-2	HPV-59	truncation	CCCATATGGATCTCGACCAGGGTGGA
	HPV-56		CCCTCACACTTGGGTGCCTTTGATGATGAC
200-3	HPV-36	truncation	CCCATATGCAGGTTTTAGTTTCCACCGCAACT
	HPV-46		AAATCAGTTGAAGTTCTGCACGTTCTTGCG

200-4	HPV-38	truncation	CCCATATGGACCTGCAGCCCTTGCCCGTGT
	HPV-44		AGCTCACATGCCTAGGTTCAGAGGCGCCAC
200-5	HPV-40	truncation	CCCATATGTGTCAGGGCTCCAAGAAGGAT
	HPV-34		CCCCTCGAGTCAATCTGAATCACTGTCTGTCTC
220K	HPV-3	cDNA cloning from library	CCCTTCGAAGCATATGGCCGGAGTGTTTCCTTATC
	HPV-4		CCCCTGAGACTGGTTCAACCGAGAC
	HPV-5		CCCACTAGTTATAAGCATGACACCAAGTTGCTC
	HPV-6		CCCCTCGAGACATTCCACTTATAGGAGGC
	HPV-7		ATATAAGATGAACTCTTCCTGTG
	HPV-8		CCCCTCGAGTTCGAAAGGCTTCGGCCTCGGGAGGCTG
220-1	HPV-3	truncation	CCCTTCGAAGCATATGGCCGGAGTGTTTCCTTATC
	HPV-66		CCCCTCGAGTCAGAAGGGCTCCACAAACTCCGG
220-2	HPV-25	truncation	CCCCATATGCTGAAGGACACACCCCTCTA
	HPV-26		CCCCTCGAGTCAGAATGGGATGGGTGAAAACC
220-3	HPV-67	truncation	CCCGGATCCCATATGCCCCCACTCTCCTATAAGCAT
	HPV-68		CCCCTCGAGTCAGTTGAGTCCAATCTTGATACG
220-4	HPV-27	truncation	CCCCATATGTCCAAGATGCCAAGTCGGTT
	HPV-28		CCCCTCGAGTCATGACTTTCGGGGATGGATTG
220-5	HPV-69	truncation	CCCGAATTCCATATGTATAAGATGAACTCTTCCTGT
	HPV-70		CCCCTCGAGTCAATCCTTGAGCTGCACCTCGAC
220-6	HPV-71	truncation	CCCGAATTCCATATGCTGATCTTGGCTGACTACGGC
	HPV-72		CCCCTCGAGTCAGGCATACAGGTCCTCCCG
220-6-ΔC	HPV-71	truncation	CCCGAATTCCATATGCTGATCTTGGCTGACTACGG
	HPV-326		CCCCTCGAGTCAGTTCGCCAGCTGTAGCTCATA
220-6-ΔN	HPV-327	truncation	AAACCCCATATGCGCCAGAACACAGACAAGGGCAAC
	HPV-72		CCCCTCGAGTCAGGCATACAGGTCCTCCCG
220-6-ΔN P2301T	HPV-331	mutagenesis	GAGCTACAGCTGGCGAACACCAAAGAGTTCTACC
	HPV-332		CCTCGTGGTAGAACTCTTTGGTGTTCGCCAGCTGTAGC
220-6-ΔN F2304L	HPV-333	mutagenesis	GCGAACCCCAAAGAGTTGTACCACGAGGTGCACAGG
	HPV-334		GCCTGTGCACCTCGTGGTACAACTCTTTGGGGTTCG
220-6-ΔN H2309P	HPV-335	mutagenesis	GAGTTCTACCACGAGGTGCCCAGGCCCTCTCACTTCC
	HPV-336		TTGAGGAAGTGAGAGGGCCTGGGCACCTCGTGGTAGAAC
220-6-ΔN H2309R	HPV-337	mutagenesis	GAGTTCTACCACGAGGTGCGCAGGCCCTCTCACTTCC
	HPV-338		TTGAGGAAGTGAGAGGGCCTGCGCACCTCGTGGTAGAAC
220-6-ΔN R2310G	HPV-339	mutagenesis	GAGTTCTACCACGAGGTGCACGGGCCCTCTCACTTCC
	HPV-340		TTGAGGAAGTGAGAGGGCCCGTGCACCTCGTGGTAGAAC
220-6-ΔN R2310K	HPV-341	mutagenesis	GAGTTCTACCACGAGGTGCACAAGCCCTCTCACTTCC
	HPV-342		TTGAGGAAGTGAGAGGGCTTGTGCACCTCGTGGTAGAAC
220-6-ΔN F2314L	HPV-343	mutagenesis	TGCACAGGCCCTCTCACTTACTCAACTTTGCTCTCCTGC
	HPV-344		AGGAGAGCAAAGTTGAGTAAGTGAGAGGGCCTGTGC

SPF30/SMNrp	LSB-17	full-length	CGCCGCCATATGTCAGAGGATTTAGCAAAGCAGC
	LSB-18'		GGAACTCGAGGTACCTTATTGAGGCATCAAATGCC
U6-p110	HPV-361	full-length	AAACCCGAATTCATGGCGACTGCGGCCGAAACC
	HPV-362		CCCCTCGAGTCACTTTCTCAGAAACAGCTTGG
pGEX-6P-1	pGEX-5'	modification	GGGCTGGCAAGCCACGTTTGGTG
	pGEX-3'		CCGGGAGCTGCATGTGTCAGAGG

3.1.7 Vectors and plasmids (Table 3.2)

Plasmid	Description	Source
pBluescript II KS	Cloning vector, T7 promoter, Amp^r	Stratagene
pGADT7	Y2H vector expresses proteins fused to GAL4 activation domain, *LEU2* marker, Amp^r	Clontech
pGBKT7	Y2H vector expresses proteins fused to GAL4 DNA binding domain, *TRP1* marker, Kan^r	Clontech
pGEX-6P-1	GST expression vector, Amp^r	AP Biotech
pET-28a-c	Expression vector, His-tag, T7 promoter, Kan^r	Novagen
pNoTA/T7	Cloning vector, T7 promoter, Amp^r	5'→3' Inc.
pGADT7-15K	Cloning sites *Nde*I→*Sal*I	HP Vornlocher (Lührmann lab)
pGBKT7-15K	Cloning sites *Nde*I→*Sal*I	HP Vornlocher (Lührmann lab)
pNoT-15.5K	Introduce cloning sites *Nde*I→*Xho*I by PCR, Amp^r	This study
pGADT7-15.5K	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-15.5K	Cloning sites *Nde*I→*Xho*I	This study
pNoT-20K	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-20K	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-20K	Cloning sites *Nde*I→*Xho*I	This study
pNoT-27K	Introduce cloning sites NdeI→XhoI by PCR	This study
pGADT7-27K	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-27K	Cloning sites *Nde*I→*Xho*I	This study
pNoT-40K	Introduce cloning sites NdeI→XhoI by PCR	This study
pGADT7-40K	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-40K	Cloning sites *Nde*I→*Xho*I	This study
pNoT-52K	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-52K	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-52K	Cloning sites *Nde*I→*Xho*I	This study
pET28a-52K	Cloning sites *Nde*I→*Xho*I	This study
pGADT7-52N	Cloning sites *Nde*I→*Xho*I, insert :1-255aa	This study
pGBKT7-52N	Cloning sites *Nde*I→*Xho*I, insert :1-255aa	This study
pET28a-52N	Cloning sites *Nde*I→*Xho*I, insert :1-255aa	This study
pGADT7-52C	Cloning sites *Nde*I→*Xho*I, insert :256-341aa	This study
pGBKT7-52C	Cloning sites *Nde*I→*Xho*I, insert :256-341aa	This study
pET28a-52C	Cloning sites *Nde*I→*Xho*I, insert :256-341aa	This study
pNoT-60K(mut)	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-60K(mut)	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-60K(mut)	Cloning sites *Nde*I→*Xho*I	This study
pNoT-60K(wt)	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study

pGADT7-60K(wt)	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-60K(wt)	Cloning sites *Nde*I→*Xho*I	This study
pGEX6P-60K(wt)	Cloning sites *Nde*I→*Xho*I	This study
pGADT7-61K	Cloning sites *Nde*I→*Xho*I	OV Makarova (Lührmann lab)
pGBKT7-61K	Cloning sites *Nde*I→*Xho*I	OV Makarova (Lührmann lab)
pGEX6P-61K	Cloning sites *BamH*I→*Xho*I	S. Nottrott (Lührmann lab)
pNoT-61K-A194E	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGBKT7-61K-A194E	Cloning sites *Nde*I→*Xho*I	This study
pNoT-61K-A216P	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-61K-A216P	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-61K-A216P	Cloning sites *Nde*I→*Xho*I	This study
pGADT7-65K	Cloning sites *EcoR*I→*Xho*I	OV Makarova (Lührmann lab)
pGBKT7-65K	Cloning sites *EcoR*I→*Xho*I	OV Makarova (Lührmann lab)
pNoT-90K	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-90K	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-90K	Cloning sites *Nde*I→*Xho*I	This study
pNoT-90K-P493S	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-90K-P493S	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-90K-P493S	Cloning sites *Nde*I→*Xho*I	This study
pNoT-90K-T494M	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-90K-T494M	Clonign sites *Nde*I→*Xho*I	This study
pGBKT7-90K-T494M	Cloning sites *Nde*I→*Xho*I	This study
pGADT7-90K-N194	Cloning sites *Nde*I→*Xho*I	This study
pGBKT7-90K-N194	Cloning sites *Nde*I→*Xho*I	This study
pGADT7-90K-N309	Cloning sites *EcoR*I→*Xho*I	This study
pGBKT7-90K-N309	Cloning sites *EcoR*I→*Xho*I	This study
pGADT7-90K-N442	Cloning sites *EcoR*I→*Xho*I	This study
pGBKT7-90K-N442	Cloning sites *EcoR*I→*Xho*I	This study
pGADT7-90K-C267	Cloning sites *EcoR*I→*Xho*I	This study
pGBKT7-90K-C267	Cloning sites *EcoR*I→*Xho*I	This study
pGADT7-90K-C381	Cloning sites *EcoR*I→*Xho*I	This study
pGBKT7-90K-C381	Cloning sites *EcoR*I→*Xho*I	This study
pGADT7-90K-C489	Cloning sites *EcoR*I→*Xho*I	This study
pGBKT7-90K-C489	Cloning sites *EcoR*I→*Xho*I	This study
pGADT7-100K	Cloning sites *Nde*I→*Sal*I	D. Ortlepp (Lührmann lab)
pGBKT7-100K	Cloning sites *Nde*I→*Sal*I	D. Ortlepp (Lührmann lab)
pNoT-102K	Cloning sites *Nde*I→*Sal*I	This study
pGADT7-102K	Cloning sites *Nde*I→*Sal*I	This study
pGBKT7-102K	Cloning sites *Nde*I→*Sal*I	This study
pGEX6P-102K	Cloning sites *Nde*I→*Sal*I	This study
pET28a-102K	Cloning sites *Nde*I→*Sal*I	This study
pGADT7-102K-NTD	Cloning sites *Nde*I→*Sal*I	This study
pGBKT7-102K-NTD	Cloning sites *Nde*I→*Sal*I	This study
pET28a-102K-NTD	Cloning sites *Nde*I→*Sal*I	This study

pGADT7-102K-TPR_M	Cloning sites *Nde*I→*Sal*I	This study
pGBKT7-102K-TPR_M	Cloning sites *Nde*I→*Sal*I	This study
pET28a-102K-TPR_M	Cloning sites *Nde*I→*Sal*I	This study
pGADT7-102K-TPR_C	Cloning sites *Nde*I→*Sal*I	This study
pGBKT7-102K-TPR_C	Cloning sites *Nde*I→*Sal*I	This study
pGADT7-110K	Cloning sites *Nde*I→*Xho*I	EM Makarov (Lührmann lab)
pGBKT7-110K	Cloning sites *Nde*I→*Xho*I	EM Makarov (Lührmann lab)
pET28a-110K	Makarov et al., 2001	EM Makarov (Lührmann lab)
pGADT7-110K_RS	Cloning sites *Nde*I→*Xho*I, 1-111 aa (insert)	This study
pGBKT7-110K_RS	Cloning sites *Nde*I→*Xho*I, 1-111 aa	This study
pGADT7-110K_ΔRS	Cloning sites *Nde*I→*Xho*I, 112-800 aa	This study
pGBKT7-110K_ΔRS	Cloning sites *Nde*I→*Xho*I, 112-800 aa	This study
pNoT-116K	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-116K	Cloning sites *Nde*I→*Xho*I, full-length 972 aa	This study
pGBKT7-116K	Cloning sites *Nde*I→*Xho*I, full-length 972 aa	This study
pNoT-116-1	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-116-1	Cloning sites *Nde*I→*Xho*I, 1-400 aa (insert)	This study
pGBKT7-116-1	Cloning sites *Nde*I→*Xho*I, 1-400 aa (insert)	This study
pNoT-116-2	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-116-2	Cloning sites *Nde*I→*Xho*I, 279-691 aa (insert)	This study
pGBKT7-116-2	Cloning sites *Nde*I→*Xho*I, 279-691 aa (insert)	This study
pNoT-116-3	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-116-3	Cloning sites *Nde*I→*Xho*I, 603-972 aa (insert)	This study
pGBKT7-116-3	Cloning sites *Nde*I→*Xho*I, 603-972 aa (insert)	This study
pBlue(KS)-200K	Cloning sites *Not*I→*Xho*I, full-length 2136 aa	HP Vornlocher (Lührmann lab)
pGADT7-200K	Cloning sites *Nde*I→*Xho*I, full-length 2136 aa	This study
pGBKT7-200K	Cloning sites *Nde*I→*Xho*I, full-length 2136 aa	This study
pNoT-200-1	Introduce cloning sites *Nde*I at 5' end by PCR	This study
pGADT7-200-1	Cloning sites *Nde*I→*BamH*I, 1-434 aa (insert)	This study
pGBKT7-200-1	Cloning sites *Nde*I→*BamH*I, 1-434 aa	This study
pET28c-200-1	Cloning sites *Nde*I→*BamH*I, 1-434 aa	This study
pNoT-200-2	Introduce cloning site *Nde*I at 5' end by PCR	This study
pGADT7-200-2	Cloning sites *Nde*I→*Eco*RI, 393-832 aa (insert)	This study
pGBKT7-200-2	Cloning sites *Nde*I→*Eco*RI, 393-832 aa	This study
pET28c-200-2	Cloning sites *Nde*I→*Eco*RI, 689-832 aa	This study
pNoT-200-3	Introduce cloning site *Nde*I at 5' end by PCR	This study
pGADT7-200-3	Cloning sites *Nde*I→*Eco*RI, 807-1449 aa (insert)	This study
pGBKT7-200-3	Cloning sites *Nde*I→*Eco*RI, 807-1449 aa	This study
pET28c-200-3	Cloning sites *Nde*I→*Eco*RI, 807-1449 aa	This study
pNoT-200-4	Introduce cloning site *Nde*I at 5' end by PCR	This study
pGADT7-200-4	Cloning sites *Nde*I→*Eco*RI, 1301-1816 aa (insert)	This study
pGBKT7-200-4	Cloning sites *Nde*I→*Eco*RI, 1301-1816 aa	This study
pET28c-200-4	Cloning sites *Nde*I→*Eco*RI, 1301-1816 aa	This study

pNoT-200-5	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-200-5	Cloning sites *Nde*I→*EcoR*I, 1705-2136 aa (insert)	This study
pGBKT7-200-5	Cloning sites *Nde*I→*EcoR*I, 1705-2136 aa	This study
pET28c-200-5	Cloning sites *Nde*I→*EcoR*I, 1705-2136 aa	This study
pNoT-220K-5'	Introduce cloning sites *BstB*I/*Nde*I at 5' end by PCR	HP Vornlocher (Lührmann lab)
pNoT-220K-Mi	Introduce cloning sites *Spe*I→*Xho*I by PCR	HP Vornlocher (Lührmann lab)
pNoT-220K-Mi	Introduce cloning sites *Xho*I/*BstB*I at 3' end by PCR	HP Vornlocher (Lührmann lab)
pBlue(KS)-220K	Cloning sites *Nde*I→*Xho*I, full-length 2335 aa	This study
pGADT7-220K	Cloning sites *Nde*I→*Xho*I, full-length 2335 aa	This study
pGBKT7-220K	Cloning sites *Nde*I→*Xho*I, full-length 2335 aa	This study
pNoT-220-1	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-220-1	Cloning sites *Nde*I→*Xho*I, 1-387 aa insert	This study
pGBKT7-220-1	Cloning sites *Nde*I→*Xho*I, 1-387 aa	This study
pET28c-220-1	Cloning sites *Nde*I→*Sst*I, 1-370 aa	HP Vornlocher (Lührmann lab)
pNoT-220-2	Introduce cloning sites *Nde*I→*Xho*I by PCR	HP Vornlocher (Lührmann lab)
pGADT7-220-2	Cloning sites *Nde*I→*Xho*I, 388-827 aa insert	This study
pGBKT7-220-2	Cloning sites *Nde*I→*Xho*I, 388-827 aa	This study
pET28c-220-2	Cloning sites *Nde*I→*Xho*I, 388-827 aa	HP Vornlocher (Lührmann lab)
pNoT-220-3	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-220-3	Cloning sites *Nde*I→*Xho*I, 828-1304 aa insert	This study
pGBKT7-220-3	Cloning sites *Nde*I→*Xho*I, 828-1304 aa	This study
pET28b-220-3	Cloning sites *Nde*I→*Sst*I, 1015-1274 aa	HP Vornlocher (Lührmann lab)
pNoT-220-4	Introduce cloning sites *Nde*I→*Xho*I by PCR	HP Vornlocher (Lührmann lab)
pGADT7-220-4	Cloning sites *Nde*I→*Xho*I, 1305-1619 aa insert	This study
pGBKT7-220-4	Cloning sites *Nde*I→*Xho*I, 1305-1619 aa	This study
pET28c-220-4	Cloning sites *Nde*I→*Xho*I, 1305-1619 aa	HP Vornlocher (Lührmann lab)
pNoT-220-5	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-220-5	Cloning sites *Nde*I→*Xho*I, 1620-1985 aa insert	This study
pGBKT7-220-5	Cloning sites *Nde*I→*Xho*I, 1620-1985 aa	This study
pET28c-220-5	Cloning sites *BamH*I→*Bgl*II, 1736-2006 aa	HP Vornlocher (Lührmann lab)
pNoT-220-6	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-220-6	Cloning sites *Nde*I→*Xho*I, 1986-2335 aa insert	This study
pGBKT7-220-6	Cloning sites *Nde*I→*Xho*I, 1986-2335 aa	This study
pET28c-220-6	Cloning sites *Nde*I→*Xho*I, 1986-2335 aa (?)	HP Vornlocher (Lührmann lab)
pNoT-220-6-ΔC	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-220-6-ΔC	Cloning sites *Nde*I→*Xho*I, 1986-2300 aa insert	This study
pGBKT7-220-6-ΔC	Cloning sites *Nde*I→*Xho*I, 1986-2300 aa	This study
pET28a-220-6-ΔC	Cloning sites *Nde*I→*Xho*I, 1986-2300 aa	This study
pNoT-220-6-ΔN	Introduce cloning sites *Nde*I→*Xho*I by PCR	This study
pGADT7-220-6-ΔN	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa insert	This study
pGBKT7-220-6-ΔN	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study

pET28a-220-6-ΔN	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study
pNoT-220-6-ΔN-P2301T	Introduce a point mutation by PCR	This study
pGADT7-220-6-ΔN-P2301T	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa insert	This study
pGBKT7-220-6-ΔN-P2301T	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study
pNoT-220-6-ΔN-F2304L	Introduce a point mutation by PCR	This study
pGADT7-220-6-ΔN-F2304L	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa insert	This study
pGBKT7-220-6-ΔN-F2304L	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study
pNoT-220-6-ΔN-H2309P	Introduce a point mutation by PCR	This study
pGADT7-220-6-ΔN-H2309P	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa insert	This study
pGBKT7-220-6-ΔN-H2309P	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study
pNoT-220-6-ΔN-H2309R	Introduce a point mutation by PCR	This study
pGADT7-220-6-ΔN-H2309R	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa insert	This study
pGBKT7-220-6-ΔN-H2309R	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study
pNoT-220-6-ΔN-R2310G	Introduce a point mutation by PCR	This study
pGADT7-220-6-ΔN-R2310G	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa insert	This study
pGBKT7-220-6-ΔN-R2310G	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study
pNoT-220-6-ΔN-R2310K	Introduce a point mutation by PCR	This study
pGADT7-220-6-ΔN-R2310K	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa insert	This study
pGBKT7-220-6-ΔN-R2310K	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study
pNoT-220-6-ΔN-F2314L	Introduce a point mutation by PCR	This study
pGADT7-220-6-ΔN-F2314L	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa insert	This study
pGBKT7-220-6-ΔN-F2314L	Cloning sites *Nde*I→*Xho*I, 2239-2335 aa	This study
pGADT7-SPF30/SMNrp	Cloning sites *Nde*I→*Xho*I, full-length 239 aa	This study
pGBKT7-SPF30/SMNrp	Cloning sites *Nde*I→*Xho*I, full-length 239 aa	This study
pGEX6P-SPF30/SMNrp	Cloning sites *Nde*I→*Xho*I, full-length 239 aa	This study
pNoT-U6-p110	Introduce cloning sites *EcoR*I→*Xho*I by PCR	This study
pGADT7-U6-p110	Cloning sites *EcoR*I→*Xho*I, 2-963 aa (insert)	This study
pGBKT7-U6-p110	Cloning sites *EcoR*I→*Xho*I, 2-963 aa	This study
pGADT7-p110-ΔN	537-963 aa insert; Medenbach et al., 2004	Bindereif lab
pGADT7-p110-ΔC	2-688 aa insert; Medenbach et al., 2004	Bindereif lab
pGADT7-p110-AF	2-350 aa insert; Medenbach et al., 2004	Bindereif lab

3.1.8 Bacterial strains

E. coli BL21(DE3)	Novagen, Darmstadt
E. coli DH5α	Invitrogen, USA
E. coli HMS174(DE3)	Novagen, Darmstadt
E. coli XL-1 blue	Stratagene, Heidelberg

3.1.9 Yeast strains (Clontech, USA)

AH109, Y187

3.1.10 Reaction sets (Kits)

ABI PRISM™ Dye Deoxy Cycle Sequencing Kit	Applied Biosystems, Weiterstadt
Amicon Centriplus Concentrator	Millipore, France
ECL Western Blot Detection Kit	Amersham Biosciences, Freiburg
GFX Purification Kit	Amersham Biosciences, Freiburg
Human Hela Marathon-Ready™ cDNA Library	Clontech, Heidelberg
Luminescent ß-galactosidase Detection Kit II	Clontech, Heidelberg
Matchmaker Yeast Two-hybrid System 3	Clontech, Heidelberg
Pretransformed Human HeLa cDNA Library	Clontech, Heidelberg
Prime-It® II Random Primer Labelling Kit	Stratagene, Heidelberg
Primer PCR Cloner™ Cloning System	5'→3' Inc., USA
pSTBlue-1 Perfertly Blunt™ Cloning Kit	Novagen, Darmstadt
QIAprep Spin Miniprep Kit	Qiagen, Hilden
QIAquick Gel Extraction Kit	Qiagen, Hilden
QIAquick PCR Purification Kit	Qiagen, Hilden
TNT® T7 Coupled Reticulocyte Lysate System	Promega, USA

3.1.11 Equipments

ABI PRISM™ 310 Genetic Analyzer	Applied Biosystems, Weiterstadt
Biofuge (pico/fresco)	Heraeus, Hanau
Electrophorese apparaturen	BiorRad, München
Gel documentation system	Biorad, München
Gradient Master Modell 106	BioComp Instruments, Canada
Heating block	Hybaid Biometra, UK
Head over tail Rotor 7637-01	Cole-Parmer, USA
Hybridisation oven	Hybaid, Biometra
Incubator BK-600	Heraeus, Hanau
Incubation shaker Multitron	Infors, Switzerland
Petrischalen 92/16 mit Nocken	Sarstedt, Nümbrecht
pH Meter	MettlerToledo,Switzerland
Phosphorimager Typhoon 8600	Molecular Dynamics

Replica plater	Qbiogene, Eschwege
SMART System	Pharmacia Biotech
Sonifier	Heinemann Labortechnik
Sorvall Rotor	Kendro, USA
Speed Vac Concentrator 5301	Eppendorf, Hamburg
Spectrophotometer Ultropsec 300 pro	Amersham Biosciences, Freiburg
SW60 Rotor	Beckman, USA
Scintillation counter LS 1701/TRI-CARB 2100TR	Beckman/Packard, USA
Ultra centrifuge	Sorvall/Beckman, USA
UV Lamp 254nm	Bachofer, Reutlingen
Thermal Cycler	Hybaid Omni Gene, UK
Vortex	Janke & Kunkel, Staufen i. Br.
X-ray film developer X-Omat 2000	Kodak, USA

3.1.12 Special materials

Amylose Resin	New England Biolabs, France
Cuvettes for Monolight 3010	Pharmingen, USA
Collodium Bags	Sartorius GmbH, Göttingen
Dialysis cassettes	Pierce, USA
Falcon tubes (5, 15, 50 ml)	Greiner, Kremsmünster
Glass Beads (425-600 microns)	Sigma, Deisenhofen
Glutathione Sepharose 4B	Amersham Biosciences, Freiburg
Nylon membrane Hybond-(N+)	Qiagen, Hilden
Pipettes	Eppendorf, Hamburg
Protein A Sepharose	Amersham Biosciences, Freiburg
Protein G Sepharose	Amersham Biosciences, Freiburg
Protran Nitrocellulose Membrane	Schleicher & Schuell, Dassel
Reaction tubes (0.5; 1.5; 2 ml)	Eppendorf, Hamburg
Sterile filter (0.2; 0.45 µm)	Millipore, France
Talon metal affinity resin	Clontech, Heidelberg
Vivaspin concentrators	Vivascience, Sartorius
X-ray film BioMax MR	Kodak, USA

3.2 Methods

3.2.1 Molecular Cloning

Two strategies were used for the plasmid construction and the cDNA cloning in this work. In one strategy, DNA was amplified and the desired restriction enzyme sites were introduced at the 5' end by PCR (section 3.2.1.1). The PCR products were subcloned into a shuttle cloning vector pNoTA/T7 which is provided in the PRIMER PCR CLONER cloning system (5'→3' Inc., Boulder, Colorado, USA). The correct insert in the pNoTA/T7 was subsequently transferred into the target vectors such as pGADT7 and pGBKT7. In the other strategy, PCR products were purified from agarose gel (section 3.2.1.2); the purified DNA was digested with appropriate enzymes and ligated with target vectors after purification (section 3.2.1.3). 1-2 µl ligation reaction was used for transformation (section 3.2.1.4). The positive colony was inoculated in LB medium and the plasmid DNA was prepared (section 3.2.1.5). 2 µl DNA sample from mini-preparation was digested with appropriate enzymes and checked on the agarose gel. The insert with the right size was verified by sequencing (3.2.1.6).

3.2.1.1 PCR amplification

Polymerase chain reaction (PCR) was mostly used as a tool for cDNA cloning and plasmid construction in this work. For this purpose, forward and reverse primers were designed to introduce compatible restriction enzyme sites and 3-6 additional bases were added at the 5' ends to allow efficient digestion by restriction enzymes (table 3.1). The following text describes a typical PCR reaction and a PCR cycling programme. The annealing temperature can be chosen on the basis of the melting temperature of the primers. This may work, but sometimes the results may not match up to expectations. Therefore, a simple procedure used many times in this work was to use an initial annealing temperature of 53°C (usually good for most primers with a length around 20 bp or more). If non-specific products resulted, this temperature was increased. If there were no products, the temperature was decreased.

PCR reaction mixture (50 µl)

1.0 µl	DNA sample (~20ng)
5.0 µl	5' oligo (10 pmol/µl)
5.0 µl	3' oligo (10 pmol/µl)
5.0 µl	10x cloned *Pfu* buffer
5.0 µl	DMSO
1.0 µl	dNTP (25 mM each)
2.0 µl	*Pfu* polymerase
26.0 µl	H_2O

PCR cycling programme

1 x	95°C 1'
30 x	95°C 30'', 53°C 1', 72°C 2'/kb
1 x	72°C 6'

hold temperature at 4°C

3.2.1.2 Agarose gel electrophoresis and DNA fragment isolation

For analysis of plasmid DNA, 0.8% agarose gel was usually used in this study. A 1-kb DNA ladder at a concentration of 0.05 mg/ml was loaded in one lane as a marker. Gel was run in 1 x TBE buffer and stained in 0.5 µg/ml ethidium bromide (EtBr). DNA was visualized under UV light.

10x TBE buffer (pH 8.3)

1 M	Tris base
0.83 M	Boric acid
10 mM	EDTA

5x DNA loading buffer

30%	glycerol (v/v)
0.25%	bromophenol blue (w/v)
025%	xylene cyanol FF (w/v)

QIAquick gel exaction kit (Qiagen, Hilden, Germany) was used for isolation of DNA fragments from agarose gel. The samples (PCR products or enzyme-digested DNA samples) were electrophoresed through 0.8% agarose gel and stained with EtBr. The band of interest was identified using a UV-illuminator and cut out of the gel. The gel slice containing the DNA fragment was placed in the prepared tube and weighed. Every 100 mg of gel was dissolved in 300 µl buffer QG at 50°C for 10 minutes. After the gel slice has dissolved completely, it was applied to the QIAquick column and spun down for 1 minute. The column was washed once with 0.75 ml wash buffer PE and the DNA sample was eluted with 30 µl elution buffer EB (10 mM Tris-HCl, pH 8.5).

3.2.1.3 Enzyme digestion and ligation

For the ligation reaction, the vector DNA and the insert DNA have to be prepared by digesting with appropriate restriction enzymes and then purifying with either GFX purification kit (Amersham Biosciences) or QIAquick PCR Purification Kit (Qiagen) according to the manufacture's protocol. The ligation reaction was established as follows. For the optimal ligation efficiency, two DNA molecules were added in the reaction in the range of 3:1 to 5:1 molar ratio of insert to vector. The reaction mixture was incubated at 16°C for 3 hours or more and was then incubated at 65°C for 15 minutes to inactivate the enzyme. The sample was spun down briefly and 2 µl was used for transformation.

20 µl ligation reaction mixture

2.0 µl	10x buffer for T4 DNA ligase
2.0 µl	linearised vector DNA
8.0 µl	insert DNA
7.5 µl	H_2O
0.5 µl	T4 DNA ligase (400 U/µl)

3.2.1.4 Bacterial transformation

Two methods were used for the bacterial transformation in this work.

3.2.1.4.1 Transformation of *E. coli* cells by electroporation

• Preparation of competent cells

One litre of culture was incubated until an OD_{600} of 0.4-0.6 was reached. The culture was centrifuged at 4000 x g for 15 minutes. The pellet was washed twice with 500 ml ice-cooled, sterile water, once with 200 ml ice-cooled, sterile 10% glycerol and once with 50 ml ice-cooled, sterile 10% glycerol. The last pellet was resuspended in 4 ml of ice-cooled, sterile 10% glycerol. 50 µl aliquots were prepared and were shock-frozen in liquid nitrogen and placed in the –70°C freezer for storage.

• Transformation

Before transformation, a 50 µl aliquot of competent cells was thawed on ice. The DNA to be transformed was then added to competent cells and the cell mixture was transferred to a pre-chilled electroporation cuvette (BioRad). After a brief incubation on ice, the cells were exposed to a voltage of 1.8 kV (for

cuvettes with 0.1 mm width) using the E. coli Pulser (BioRad). 950 µl of LB medium was added to the cuvette after electroporation. The cells were transferred to a 1.5-ml microcentrifuge tube, incubated at 37°C for 40-60 minutes, spread on LB plates containing appropriate antibiotics and incubated at 37°C overnight.

3.2.1.4.2 Transformation of *E. coli* cells by heat shock

• Preparation of competent cells (rubidium chloride method)

250 ml culture was incubated until an OD_{600} of 0.4-0.6 was reached. The culture was centrifuged at 4000 x g for 15 minutes at 4°C. The cell pellet was resuspended in 100 ml ice-cooled TFBI buffer and incubated on ice for 10 minutes. The resuspension was centrifuged at 4000 x g for 5 minutes at 4°C. The pellet was resuspended in 10 ml ice-cooled TFBII buffer and incubated on ice for at least 15 minutes. 100 µl aliquots were prepared and were shock-frozen in liquid nitrogen and placed in the –70°C freezer for storage.

TFBI
30 mM potassium acetate
100 mM rubidium chloride
10 mM calcium chloride
50 mM manganese chloride
15% v/v glycerol
Adjust to pH 5.8 with acetic acid and filter (0.20 µm, Millipore) to sterilize.

TFBII
10 mM MOPS
75 mM calcium chloride
10 mM rubidium chloride
15% v/v glycerol
Adjust pH to 6.5 with KOH and filter to sterilize.

• Transformation

Before transformation, a 100 µl aliquot of competent cells was thawed on ice. The DNA or ligation reaction to be transformed was then added to the competent cells and the cell mixture was incubated on ice for 20-30 minutes and then at 42°C for 90 seconds for heat shock. 900 µl of LB medium was added and incubated at 37°C for 45 minutes and then spread on the LB plates containing appropriate antibiotic and incubated at 37°C overnight.

3.2.1.5 Mini-preparation of plasmid DNA

QIAprep spin miniprep kit (Qiagen, Hilden, Germany) was used for this purpose. The purification of plasmid DNA was performed following the methods described by manufacturer. A single colony was inoculated into 5 ml LB liquid medium containing appropriate antibiotic and grown overnight at 37°C with shaking. 2 ml of overnight culture was transferred to a 2-ml microcentrifuge tube and centrifuged at 13,000 rpm at 4°C for 10 minutes. The cell pellet was resuspended in 250 µl buffer P1 and then 250 µl buffer P2 was added and mixed gently by inverting the tube 4-6 times. After 5 minutes incubation at room temperature, 350 µl buffer P3 was added and mixed immediately but gently by inverting the tube 4-6 times. The tube was centrifuged at 13,000 rpm for 10 minutes. The supernatant was applied to a QIAprep spin column by decanting. The column was spun down for 1 minute to remove the liquid and washed once with 0.75 ml buffer PB and subsequently with 0.75 ml buffer PE. The residual wash buffer was removed by additional centrifugation for 1 minute and air-dried for 1 minute. To elute DNA, 50 µl elution buffer EB was added to the centre of each column and the column was allowed to stand for 1 minute and centrifuged for 1 minute to collect the DNA sample. For checking the insert, 2 µl DNA sample was used for enzyme digestion and run on 0.8 % agarose gel.

3.2.1.6 DNA sequencing

To minimize potential misincorporation during PCR, I used high-fidelity *Pfu* DNA polymerase. In addition, the sequence accuracy of PCR products for each construct and cDNA was confirmed by sequencing. The sequencing reaction and PCR cycling programme were established as follows.

20 µl sequencing reaction

10.0 µl	H_2O
6.0 µl	BigDye (Applied Biosystems, Weiterstadt, Germany)
3.0 µl	DNA sample (~250 ng)
1.0 µl	sequencing oligo (10 pmol/µl)

PCR cycling programme

1 x	96°C 1’
25 x	96°C 30’’, 55°C 30’’, 60°C 4’

hold temperature at 4°C

Following the temperature cycling, the reaction was spun down briefly. To precipitate the sample, 15 µl 3 M NaAc pH 5.3, 65 µl H_2O and 300 µl 100% ethanol were added and mixed. The tube was centrifuged at 13,000 rpm at

15°C for 20 minutes. The pellet was washed once with 750 µl 70% ethanol, dried in air and resuspended in 25 µl of template suppression reagent (Applied Biosystems, Weiterstadt, Germany). The DNA sample was sequenced by ABI PRISM™ 310 Genetic Analyzer (Applied Biosystems).

3.2.1.7 PCR-based site directed mutagenesis

In vitro site-directed mutagenesis is an invaluable technique for studying protein structure-function relationships, gene expression and vector modification. Several methods have appeared in the literature, but many of these methods require single-stranded DNA as the template. The reason for this, historically, has been the need for separating the complementary strands to prevent reannealing. The use of PCR in site-directed mutagenesis accomplishes strand separation by using a denaturing step to separate the complementary strands and allowing efficient polymerization of the PCR primers. PCR-based site-directed methods thus allow site-specific mutations to be incorporated in virtually any double-stranded plasmid, eliminating the need for M13-based vectors or single-stranded DNA rescue.

In this study, this method was used to introduce single-point mutations into human genes *PRPF3* (U4/U6-90K) and *PRPF8* (U5-220K). *PfuTurbo* DNA polymerase from Stratagene (Heidelberg, Germany) was used in the PCR reaction. I followed basically the protocol provided by Stratagene. The mutagenic oligonucleotide primers were ordered from MWG Biotech (Ebersberg, Germany) and are listed in table 3.1. Primers HPV-331–344 were synthesised for U5-220K; HPV-347, 348, 359 and 360 for U4/U6-90K. The PCR reaction and cycling programme were established as follows.

35 µl PCR reaction

1.0 µl	DNA template (~16 ng)
3.5 µl	10x cloned *Pfu* buffer
2.0 µl	5’ oligo (10 pmol/µl)
2.0 µl	3’ oligo (10 pmol/µl)
1.0 µl	25 mM dNTP
24.5 µl	H_2O
1.0 µl	*PfuTurbo* DNA polymerase (2.5 U/µl)

PCR cycling programme

1 x	95°C 30’’
12 x	95°C 30’’, 55°C 1’, 68°C 2’/kb

Following the temperature cycling, the reaction mixture was cooled on ice for 2 minutes. To digest the methylated parental DNA template, 1 µl restriction endonuclease *Dpn*I was added to the reaction and incubated at 37°C for one hour. 1 µl of *Dpn*I-treated DNA was transformed into 50 µl XL-1 blue competent cells by electroporation as described in section 3.2.1.4.1. The cells were spread on the LB plates containing appropriate antibiotic and grown at 37°C overnight. Several single colonies were inoculated into 5 ml LB medium at 37°C overnight respectively. The protocol described in section 3.2.1.5 was used for preparing plasmid DNA. The mutation was confirmed by DNA sequencing (section 3.2.1.6).

3.2.1.8 cDNA cloning from library

To clone U4/U6-60K, two oligonucleotides HPV-324 and HPV-325 (see table 3.1) corresponding respectively to the 5' and 3' ends of the ORF (gi|2708304) were synthesized and used in PCR with human Marathon-Ready cDNA as template. The amplified cDNA was cloned into the PCR cloning vector pNoTA/T7 (5'→3' Inc.) to generate pNoT-60K. To construct pGADT7-60K and pGBKT7-60K, the *Nde*I–*Xho*I fragment of 60K was subcloned between the *Nde*I and *Xho*I sites of pGADT7 and the *Nde*I and *Sal*I sites of pGBKT7, respectively. Since *Nde*I cleaves in the 3'-terminal region of 60K, *Xho*I-digested pNoT-60K was partially digested with *Nde*I to isolate the intact ORF DNA.

To clone the U5-220K protein, three overlapping fragments comprising the 5', central and 3' regions of full-length 220K were amplified from Marathon-Ready cDNA by PCR using three pairs of oligonucleotides: HPV-3/HPV-4, HPV-5/HPV-6, and HPV-7/HPV-8 (see table 3.1). These were designed according to the sequence deposited in the GenBank data base (gi|3661609). The PCR products were cloned into the pNoTA/T7 vector. The three fragments were then released from pNoTA/T7 and jointly cloned into pBluescript-II-KS to generate the pBlue(KS)-220K full-length cDNA. To construct pGADT7-220K and pGBKT7-220K, 3.0-kb *Nde*I–*Nde*I and 3.9-kb *Nde*I–*Xho*I fragments of pBlue(KS)-220K were jointly cloned between the *Nde*I and *Xho*I sites of pGADT7 and the *Nde*I and *Sal*I sites of pGBKT7, respectively.

To clone the U5-200K gene, three overlapping fragments were PCR-amplified from Marathon-Ready cDNA by using three pairs of oligonucleotides: HPV-29/HPV-30, HPV-31/HPV-32 and HPV-33/HPV-34 (see table 3.1). These are complementary to the partial cDNA sequence of 200K (gi|20521659; Nagase *et al*., 1998). The three PCR products were cloned into pNoTA/T7 and subsequently cloned jointly into pBluescript-II-KS, resulting in a 4.8-kb 3'-region of the 200K cDNA. To obtain the 5'-end sequence, 5'-RACE PCR was performed using Marathon-Ready cDNA and the gene-specific primer HPV-52 (see table 3.1). The resulting 2.0-kb 5' RACE product was sequenced (it shows a 92-bp overlap with the sequence of the 3' region) and fused with the 3' sequence to yield full-length pBlue(KS)-200K cDNA. To construct pGADT7-200K and pGBKT7-200K, 1.0-kb *Nde*I–*Dra*III and 5.4-kb *Dra*III–*Xho*I fragments were jointly cloned between the *Nde*I and *Xho*I sites of pGADT7 and the *Nde*I and *Sal*I sites of pGBKT7, respectively. To clone various truncated forms of the 200K protein in pGADT7 and pGBKT7, fragments were generated and amplified by PCR by using pBlue(KS)-200K as a template. An *Nde*I site is at the same time introduced at the 5' ends. PCR products were cloned into pNoTA/T7 to generate the pNoT-200-(1 to 5) series. The *Nde*I–*BamH*I fragment of pNoT-200-1 was cloned between the corresponding sites of pGADT7 and pGBKT7. The *Nde*I–*Nde*I and *Nde*I–*EcoR*I fragments of pNoT-200-2 were jointly subcloned between the *Nde*I–*EcoR*I sites of pGADT7 and pGBKT7, respectively. The *Nde*I–*EcoR*I fragments of pNoT-200-(3 to 5) were subcloned between the *Nde*I–*EcoR*I sites of pGADT7 and pGBKT7, respectively.

To clone U2-SPF30/SMNrp, two oligonucleotides LSB-17 and LSB-18’ (see table 3.1) corresponding respectively to the 5' and 3' ends of the ORF (gi|54126076) were synthesized by introducing *Nde*I at 5’ end and *Xho*I at 3’ end and used in PCR with human Marathon-Ready cDNA as template. The PCR products were purified and digested with *Nde*I/*Xho*I and were subcloned between the *Nde*I and *Xho*I sites of pGADT7 and the *Nde*I and *Sal*I sites of pGBKT7, respectively.

3.2.1.9 Plasmid construction for two-hybrid analysis

To construct two-hybrid plasmids for the U4/U6-15.5K, U4/U6-20K, U4/U6.U5-27K, U5-40K, U5-52K, U4/U6-61K-A194E, U4/U6-61K-A216P, U4/U6-90K, and U5-116K proteins, the corresponding ORFs were amplified by PCR, at the same time including terminal *Nde*I and *Xho*I sites. PCR products were cloned into pNoTA/T7, and *Nde*I–*Xho*I fragments were transferred into pGADT7 linearised with *Nde*I/*Xho*I or into pGBKT7 linearised with *Nde*I/*Sal*I. Fragments of the 52K, 116K and 220K proteins were constructed in a similar manner. The cDNAs encoding full-length 100K and 102K were PCR-amplified, introducing terminal *Nde*I and *Sal*I sites, respectively. PCR products were cloned into pNoTA/T7 to generate pNoT-100K and pNoT-102K respectively. *Nde*I–*Sal*I fragments were transferred to pGADT7 cut with *Nde*I/*Xho*I or pGBKT7 cut with *Nde*I/*Sal*I. To construct pGADT7-15K and pGBKT7-15K, the *Nde*I–*Sal*I fragment from pXC35-15K (Reuter *et al*., 1999) was subcloned between the *Nde*I and *Xho*I sites of pGADT7 or between the *Nde*I and *Sal*I sites of pGBKT7. The *EcoR*I–*Xho*I fragment of pSK-65K and the *Nde*I–*Xho*I fragments of pET28a-61K and pSK-110K (Makarova *et al*., 2001, 2002) were processed in the same way to obtain pGADT7-65K and pGBKT-65K, pGADT7-61K and pGBKT-61K or pGADT7-110K and pGBKT-110K, respectively. To generate Y2H vectors expressing truncated forms of 102K, the fragments indicated in figure 4.13 were PCR-amplified, again introducing terminal *Nde*I and *Sal*I sites. PCR products were restriction-digested with *Nde*I and *Sal*I and introduced into pGADT7 or pGBKT7 linearised with *Nde*I and *Xho*I or *Nde*I and *Sal*I, respectively. The original constructs used for PCR cloning are showed in the table 3.3.

Table 3.3. List of the original constructs used in this study

Human protein	Original construct	mRNA Accession	Source*	Reference
U4/U6-90K	EST-90K (aa205466)	gi|2708306	S. Schieborr	Lauber *et al.*1997 Horowitz *et al.*, 1997
U4/U6-61K	pBISK(-)/zm27-61K	gi|18249846	Dr. O.V. Makarova	Makarova *et al.*, 2002
U4/U6-61K mut	pTriEX-61K-A194E		Dr. S. S. Bhattacharya	Vithana *et al.*, 2001 Deery *et al.*, 2002
U4/U6-61K mut	pTriEX-61K-A216P			
U4/U6-60K	pNoT-60K	gi|2708304	this study	Horowitz *et al.*, 1997
U4/U6-20K	pKS-20K	gi|3647229	Dr. D. Ingelfinger	Teigelkamp *et al.*, 1998
U4/U6-15.5K	pKS-15.5K	gi|6318598	Dr. S. Nottrott	Nottrott *et al.*, 1999
U5-220K	pKS-220K	gi|3661609	this study	Luo *et al.*, 1999
U5-200K	pKS-200K	gi|45861371	Dr. H.-P. Vornlocher	Lauber *et al.*, 1996
U5-116K	HA01047	gi|434758	Dr. P. Fabrizio	Fabrizio *et al.*, 1997 Nomura *et al.*, 1994
U5-102K	pKS-102K, pET21a-102K	gi|7658290	Dr. E.M. Makarov	Makarov *et al.*, 2000 Nishikimi *et al.*, 1999
U5-52K	EST262201	gi|33875510	Dr. O.V. Makarova	Laggerbauer *et al.*, 2005 Bialkowska *et al.*, 2002
U5-40K	pKS-40K	gi|3820593	Dr. T. Achsel	Achsel *et al.*, 1998
U5-15K	pXC35-15K	gi|6572635	Dr. K. Reuter	Reuter *et al.*, 1999
U4/U6.U5-110K	pSK(-)/110K	gi|13926067	Dr. O.V. Makarova	Makarova *et al.*, 2001
U4/U6.U5-65K	pSK-65K	gi|13926070	Dr. O.V. Makarova	Makarova *et al.*, 2001
U4/U6.U5-27K	RY-1	gi|431952	Dr. P. Fabrizio	Fetzer *et al.*, 1997
U6-p110	pFastBacHT-p110	gi|21327689	Dr. A. Bindereif	Medenbach *et al.*, 2004
U2-SPF30/SMNrp	pGADT7-SPF30/SMNrp	gi|54126076	this study	Meister *et al.*, 2001 Reppsilber *et al.*, 2001

* All these persons have been worked or are still working in the laboratory of Prof. Lührmann except Dr. Bindereif and Dr. Bhattacharya.

3.2.2 Yeast Two-hybrid System

The Matchmaker Two-Hybrid System 3 (Clontech, Palo Alto, USA) was used in this study. Two vectors pGADT7 and pGBKT7, two yeast strains AH109 and

Y187, and two plasmids pGADT7-T and pGBKT7-53 were provided in this system. pGADT7-T encodes a Gal4 activation domain/SV40 large T-antigen fusion protein, while pGBKT7-p53 encodes a Gal4 DNA binding domain/murine p53 fusion protein. Murine p53 and SV40 large T-antigen are known to interact in a yeast two-hybrid assay and therefore provide a positive control for interacting proteins. The map of the vectors pGADT7 and pGBKT7 is shown in the Appendix; the genotypes of the yeast strains are shown in table 3.4. AH109 utilizes two nutritional markers *HIS3* and *ADE2* to control the stringency of selection: the *ADE2* gene, under the control of the *GAL2* promoter, reduces the number of false positives, while the *HIS3* gene, under the control of the *GAL1* promoter, provides sensitive growth selection that helps to identify weak positive signals (section 3.2.2.1). β-galactosidase encoded by *lacZ* reporter can be detected using a variety of assays that differ in their relative levels of sensitivity (sections 3.2.2.2 and 3.2.2.3). I used yeast strain Y187 instead of AH109 for β-galactosidase assays, because Y187 exhibits a higher level of induced β-galactosidase activity than does AH109. X-α-Gal assay was used to detect inherent yeast *MEL1* gene expression level (section 3.2.2.4). The phenotypes of the yeast strains AH109 and Y187 were verified before they were used in a transformation or mating experiment.

The yeast co-transformation method is described in section 3.2.2.5. To check protein expression after transformation into yeast cells, I followed the methods given in sections 3.2.2.8 and 3.2.2.9.

Table 3.4 Yeast host strain genotype used in this study

Strain	Genotype	Reporters
AH109	***MATa,*** *trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2 : : $GAL1_{UAS}$-$GAL1_{TATA}$ -HIS3, $GAL2_{UAS}$-$GAL2_{TATA}$-ADE2 URA3 : : $MEL1_{UAS}$ -$MEL1_{TATA}$-lacZ MEL1*	*HIS3, ADE2, MEL1, lacZ*
Y187	***MATα,*** *ura3-52, his3-200, trp1, leu2, ade2-101, trp1-901, leu2-3, 112, gal4Δ, gal80Δ, met⁻, URA3 : : $GAL1_{UAS}$-$GAL1_{TATA}$-lacZ MEL1*	*MEL1, lacZ*

3.2.2.1 Yeast two-hybrid analysis of known proteins for interaction studies

All cDNAs encoding full-length or truncated proteins were cloned into pGBKT7 (a Gal4 DNA-binding domain "bait" vector) and pGADT7 (a Gal4 activation domain "prey" vector) as described in sections 3.2.1.8 and 3.2.1.9. Two-hybrid assays were carried out according to procedures described in the manufacturer's manual. To screen for protein-protein interactions, the bait and prey plasmids were co-transformed into the yeast strain AH109 by using the lithium acetate method (section 3.2.2.5). Co-transformation reaction mixtures were plated onto minimal synthetic dropout (SD) medium lacking leucine and tryptophan and incubated at 30 °C for 3-5 days. Co-transformants were then replicated on two different selective media: SD/-His for activation of the *GAL1-HIS3* reporter gene and SD/-His-Ade for the activation of both the *GAL1-HIS3* and the *GAL2-ADE2* reporter gene. In each assay, empty vectors pGADT7 and pGBKT7 were used as negative controls, and pGADT7-T/pGBKT7-53 as a positive control. At least two independent co-transformants were tested for each assay.

3.2.2.2 Colony-lift filter assay

The colony-lift filter assay is used to detect reporter *lacZ* gene activity. It measures the β-galactosidase activity encoded by *lacZ*. It is used primarily to screen the large number of cotransformants that survived the nutritional selection. It can also be used to assay for an interaction between two known proteins in a GAL4 two-hybrid system.

Plates were incubated at 30°C until colonies appear and grew to a size of 1-2 mm, and this usually took 4-5 days. For each plate of transformants assayed, a sterile Whatman #5 or VWR grade 410 filter was used. It was placed in 2.5-5.0 ml of Z buffer/X-gal solution in a clean 150 mm plate. Using clean forceps, a clean dry filter was placed over the surface of the plate of colonies and gently rubbed with the side of the forceps, which helped the colonies to cling to the filter. After the filter was evenly wetted, it was carefully lifted off the agar plate and transferred to a pool of liquid nitrogen (colonies facing up). Liquid nitrogen lyses the yeast cells, thus allowing detection of the reporter gene. The filters were submerged for 10 seconds after which they were completely thawed at

room temperature. The filter colony side facing up was then cautiously placed on the presoaked filter, with care not to trap any bubbles between the two filters. Filters were incubated at 30°C and checked periodically for the appearance of blue colonies.

Z buffer

$Na_2HPO_4.7H_2O$	16.1 g/L
$NaH_4PO_4.H_2O$	5.50 g/L
KCl	0.75 g/L
$MgSO_4.7H_2O$	0.246 g/L

Adjusted the pH to 7.0 and autoclaved

X-gal stock solution
20 mg /ml in DMF

Z buffer/ X-gal solution

100 ml	Z-buffer
0.27 ml	β-mercaptoethanol
1.67 ml	X-gal stock solution

3.2.2.3 Chemiluminescence β-galactosidase assay

The chemiluminescence β-galactosidase assay can be used to verify and quantify β-galactosidase activity in two-hybrid assays. In this work, I detected the potential effect of the point mutations of U5-220K on the interactions of this protein with U5-116K and U5-200K (see section 4.10.3) using the luminescence β-galactosidase detection kit II from Clontech (Palo Alto, USA).

As an example, the pGBKT7-116-3 and pGADT7-ΔN-P2301T were cotransformed into yeast Y187. The cells were plated on the medium SD/-Leu-Trp and incubated at 30°C until colonies appear and grew to size of 1-2 mm, and this usually took 4-5 days. The colonies were grown in 5 ml SD/-Leu-Trp liquid medium overnight at 30°C and the OD_{600} should be 0.5-1.0 in the next day. 2 ml of overnight culture was added to 8 ml of YPD medium and the ratio was adjusted so that the OD_{600} is approximately 0.2. The culture was incubated at 30°C for 3-5 hours with shaking (250 rpm) until the OD_{600} reaches 0.4-0.6. 1.5 ml of the culture was harvested and the pellet was washed once with 1.5 ml of Z buffer. The cells were resuspended in approximately 300 µl of Z buffer to bring the OD_{600} to ~2.5. 100 µl aliquot was transferred into a fresh tube. The tube was placed in liquid nitrogen until the cells were frozen and then moved to a 37°C water bath for 30-60 seconds to thaw the cells. This freeze/thaw cycle was

repeated twice. The tube was centrifuged for 5 minutes at maximum speed and the supernatant (i.e. yeast extract) was then transferred to a fresh tube and kept on ice. Three aliquots of 25 µl of the yeast extract were prepared and each was mixed with 200 µl of reaction buffer (4 µl reaction substrate+196 µl reaction buffer supplied in the kit). The mixture was incubated at room temperature (20-25°C) for 60 minutes. For measuring, the sample was transferred into a luminometer tube (Cuvettes for Monolight 3010, 12X75 mm) and measured in the Luminometer Monolight 3010 in the Department of Molecular Developmental Biology in MPIbpc. The light emission was recorded as 5-second integrals. The data from different samples including the negative and positive control groups were collected. The linear range of the assay was determined accordingly.

3.2.2.4 X-α-Gal assay

X-α-Gal is a chromogenic substrate for α-galactosidase, the product of the yeast *MEL1* gene in response to GAL4 activation. It can be used to detect GAL4-based two-hybrid interactions directly on the nutritional selection plates. As α-galactosidase accumulates in the medium, it hydrolyses X-α-Gal, causing yeast colonies to turn blue. In this work, the X-α-Gal assay was used in the yeast two-hybrid screening with pGBK-200K-1 (see section 4.11 for results).

The yeast dropout medium was poured onto the 15-cm plates and allowed to cool down and harden at room temperature. 250 µl of X- α -Gal stock solution (2 mg/ml in DMF) was then evenly spread onto a 15-cm plate. When plates dried, cells were plated and incubated at 30°C until blue colonies formed.

3.2.2.5 LiAc-mediated yeast co-transformation

Yeast colonies were grown in 15 ml YPDA/Kan medium overnight at 30°C to stationary phase (OD_{600}>1.5) and were used to inoculate 150 ml of YPD medium containing 0.003% adenine sulphate and 30µg/ml kanamycin to produce an OD_{600} of 0.2-0.3.The flask was incubated at 30°C for 3-5 hours with shaking (230-270 rpm) to get a mid-log phase measuring an OD_{600} of 0.4-0.6. This culture is sufficient for 30 transformations. The culture was harvested in sterile 50 ml centrifuge tube at 2000 rpm (Megafuge 1.0R) for 5 min at room

temperature (20°C). The medium was poured off and 15 ml of sterile distilled water was added to resuspend the cells. The cells were collected in one tube and centrifuged again to remove the water. 3.0 ml of PEG/LiAc solution was added to the pellet and the tube was vortexed to resuspend the cells. This allows permeabilization of the yeast cells, resulting in competent cells. 100 μl yeast competent cells were added to each 1.5-ml sterile microcentrifuge tubes containing 5 μl salmon sperm DNA (10mg/ml) and ~ 1.0 μg DNA plasmids from bait and prey and were mixed well by pipetting. The tubes were incubated at 42°C for 20-30 minutes for heat shock. The cells were chilled on ice for 1-2 minutes, and were centrifuged at 7000 rpm for 10 seconds to remove the supernatant. 50-100 ml of sterile dH_2O was added to the tubes and the pellet was resuspended by pipetting gently. The transformation mix was spread on an SD/ -Leu-Trp. The plates were incubated at 30°C until colonies appeared.

1 ml PEG/LiAc solution (prepare fresh just prior to use)
800 μl 50% PEG3350
100 μl 1 M LiAc pH7.5
100 μl 10 X TE pH7.5

3.2.2.6 Yeast two-hybrid screen with the N-terminal region of U5-200K

The fragment 200-1 of U5-200K (amino acids 1-434) was constructed as a bait fusion (pGBKT7-200-1) as described in section 3.2.1.8. Before it was used for screening library, the following tests were performed. (1) This fusion protein was transformed into AH109, and the protein expression was confirmed by western blotting (see sections 3.2.2.8 and 3.2.2.9 for the methods). (2) It was confirmed that this fusion protein had no toxicity on the host strain by comparing the growth rate in liquid culture (SD/-Trp) of cells transformed with the empty vector pGBKT7 and cells transformed with pGBKT7-200-1. (3) The AH109 strain transformed with pGBKT7-200-1 exhibited background growth on SD/-His medium, but the background was eliminated by adding 20 mM 3-AT; it was confirmed that there was no background growth on SD/-Ade-His medium.

The above experiments indicated that pGBKT7-200-1 was a suitable bait for the library screening. The pGBKT7-200-1 was therefore transformed into AH109 and grown on the medium SD/-Trp for 5 days. A single colony (2-3 mm in

diameter) was inoculated into 50 ml of SD/-Trp medium and incubated overnight (~23 hours) at 30°C; the OD_{600} was 1.6 on the following day. The cells were harvested by centrifugation at 2000 rpm (Megafuge 1.0R) for 5 min at 4°C. The cell pellet was resuspended in 5 ml SD/-Trp medium and kept in room temperature. Just before use, one frozen aliquot (~1.0 ml) of the pretransformed human HeLa cDNA library in Y187 (see below for description) was thawed in a room-temperature water bath and mixed gently. 10 µl was taken out and kept on ice for later titrating. The entire AH109:pGBKT7-200-1 culture and the ~1ml library culture were combined in a 2-L sterile flask and 45 ml of 1X YPD containing 0.003% adenine sulphate and 15µg/ml kanamycin were subsequently added. The flask was incubated overnight (~24 hours) at 30°C with gentle swirling (40 rpm). The mating mixture was transferred to a sterile 50-ml centrifuge tube and the cells were harvested by centrifugation at 1500 rpm for 10 minutes at 4°C. Meanwhile, the mating flask was rinsed with 1X YPD (2X 25 ml), and the rinses were used to resuspended the first pellet, and the cells were spun down again and resuspended in 10 ml of 0.5X YPD medium. This is the library mating mixture. 100 µl of the mixture was used for mating efficiency analysis and the rest was then spread on the 50 large (150-mm) plates of SD/-Leu-Trp-His-Ade at 200 µl per plate. The plates were incubated at 30°C. The positive clones were selected between 5 and 12 days after incubation. These clones were reselected on the more stringent plates of SD/-Leu-Trp-His-Ade+ 20 mM 3-AT+X-α-Gal. From this step, 126 positive clones were selected for preparing plasmid DNA samples (see section 3.2.2.7 for the method). The samples were then sequenced. The data were collected and analysed (see Results 4.11).

The pretransformed human HeLa cDNA library was purchased from Clontech (Palo Alto, USA). It was cloned into pGAD GH and amplified in E. coli strain DH10B. The plasmid DNA was isolated and used to transform yeast strain Y187.

3.2.2.7 Quick preparation of plasmid DNA from yeast

(Method from the Hahn laboratory)

The positive colony (2-3 mm in diameter) from the plate of SD/-Leu-Trp-His-Ade+20 mM 3-AT+X-α-Gal was picked and resuspended in 200 µl of lysis buffer.

An equal volume of glass beads (425-600 microns; Sigma) was added and the mixture was mixed on a vortexer at top speed for 1 minute. 200 µl of phenol/$CHCl_3$ (2:1) was then added for extraction. The sample was centrifuged at 13000 rpm at 4°C for 12 minutes. The supernatant was mixed with 2.5 volumes of 100% ethanol for precipitation of DNA at -20°C for one hour. The DNA pellet was washed once with 80% ethanol. DNA was then dried and resuspended in 100 µl TE buffer and used to transform *E. coli*.

Lysis buffer (10 ml)

100 mM	NaCl
10 mM	Tris-HCl pH8.0
1 mM	EDTA
0.1%	SDS

TE buffer pH8.0

10 mM	Tris-HCl pH 8.0
1 mM	EDTA pH 8.0

3.2.2.8 Preparation of yeast protein extract (Urea/SDS method)

For each transformed yeast strain assayed, a 5 ml overnight culture (using SD/-Leu or SD/-Trp medium) was used to inoculate 50 ml of YPD medium containing 0.003% adenine sulphate and 30µg/ml kanamycin to produce an OD_{600} 0.4-0.6. 15 OD_{600} units of each culture were taken (e.g. if OD_{600}=0.5, take 30 ml of the culture) for centrifugation at 4°C and the cells were washed once with cold water. The pellet can be stored at -80°C or go further for the next step. The pellet was resuspended in 200 µl pre-warmed cracking buffer (60°C) and the suspension was transferred to a 1.5-ml microcentrifuge tube containing 150 µl of glass beads. The mixtures were heated at 70°C for 10 minutes and then vortexed vigorously for 1 minute. Unbroken cells and debris were removed by centrifugation at 13000 rpm for 5 minutes at 4°C. The supernatant was transferred into a fresh tube, boiled for 3 minutes and loaded onto the protein gel.

Cracking buffer stock solution

8 M	Urea
5% (w/v)	SDS
40 mM	Tris-HCl pH 6.8
0.1 mM	EDTA
0.4 mg/ml	Bromophenol blue

Cracking buffer

1 ml	Cracking buffer stock solution
10 µl	β-mercaptoethanol
70 µl	protease inhibitor cocktail
50 µl	0.1 M PMSF

3.2.2.9 Western blotting

Proteins were separated by SDS-PAGE (a pre-stained MW standards from Bio-Rad were applied onto one lane as an indication of a successful transfer) and transferred to Hybond C or Hybond P membrane using a semi-dry transfer apparatus (Biometra). The gel, membrane and sheets of Whatman paper 3MM were soaked in transfer buffer, and the sandwich containing the gel and membrane was placed between sheets of 3MM paper. The transfer was performed for 45 minutes at 10 W. To decrease non-specific binding of antibodies, the membrane was blocked in TBS-Tween buffer containing 5% non-fat dried milk for two hours at room temperature or at 4°C overnight on a rocking platform. All further manipulations were carried out in TBS-Tween buffer containing 1% non-fat dried milk. After blocking, the membrane was washed 3 times, and then incubated with primary antibody against the protein of interest for one hour at room temperature on a rocking platform. The optimum dilution of the antibody was determined by titration and was typically in the range of 1:500 to 1:5000.

The membrane was washed three times for 15 minutes each and then incubated with either a secondary antibody or protein A, conjugated to horseradish peroxidase (HRP), in a manner described for the primary antibody. Protein A-HRP was used at 50 ng/ml, whereas the appropriate concentration for the secondary antibody was determined by titration. The membrane was washed three times and developed with enhanced chemiluminescence (ECL) detection reagents (Amersham). Equal volumes of Amersham detection reagents 1 and 2 (50-100 µl per 1 cm2 of the filter) were mixed and applied to the membrane. After incubation for 60-90 seconds at room temperature, the membrane was drained, wrapped in plastic film and exposed to X-ray film for 10 seconds to 5 minutes. The membrane could be re-probed with a different antibody after stripping it of the previous one. For this purpose, the membrane was incubated in stripping buffer at 50°C for 30 minutes.

10x Transfer buffer (1 liter)
30.3 g Trizma base (= 0.25 M)
144 g Glycine (= 1.92 M)
pH should be 8.3; do not adjust

1x Transfer buffer (1 liter)
100 ml 10x Transfer buffer
200 ml Methanol
700 ml H_2O

TBS-Tween
20 mM Tris-HCl pH 7.5
150 mM NaCl
1% (v/v) Tween-20

Stripping buffer
62.5 mM Tris-HCl pH 6.7
2% SDS (w/v)
100 mM β-mercaptoethanol

3.2.3 *In vitro* Biochemical Methods

3.2.3.1 Transcription and translation *in vitro*

The T7 promoter of pGADT7 and pGBKT7 vectors is positioned immediately upstream of the epitope tag (HA or c-Myc, respectively) and the insertion site of the cloned cDNA. All constructs could therefore be used as transcription templates in transcription/translation reactions *in vitro* to produce [^{35}S]-labelled HA- or c-Myc-tagged proteins, using [^{35}S]-methionine. The TNT T7 coupled reticulocyte lysate system (Promega, Madison, USA) was used. The procedure was followed according to the manufacture's instructions. Briefly, the reaction components were assembled in a 1.5-ml microcentrifuge tube and the reaction was incubated at 30°C for 90 minutes. This lysate was used for the subsequent co-immunoprecipitation experiments.

50 µl volume of reaction component
25 µl TNT rabbit reticulocyte lysate
2 µl TNT reaction buffer
1 µl TNT T7 RNA polymerase
1 µl amino acid mixture, minus methionine (1 mM)
2 µl [^{35}S]-methionine (1000 Ci/mmol, Amersham Pharmacia Biotech)
1 µl RNasin ribonuclease inhibitor (40 U/µl)
2 µl DNA template (~0.5 µg/µl)
16 µl nuclease-free H_2O

3.2.3.2 GST-102K pull-down assays (figures 4.9B, 4.11, 4.14B, 4.22A)

• Expression of U5-102K protein in *E. coli*

To express U5-102K protein in *E. coli* cells, pNoT-102K was digested with *Nde*I and *Sal*I, and the insert was subcloned into the same sites of pGEX-6P-1+, which has been modified from pGEX-6P-1 by insert the *Nde*I site to the multiple cloning site. The resulted pGEX-6P-102K plasmid was transformed into *E. coli* strain BL21(DE3) for expression. Several small expression tests were performed before the following step was started. 200 ml culture was induced at OD_{600} of 0.6 with 0.5 mM IPTG at 20°C for 3 hours. The culture was then harvested and resuspended in 10 ml ice-cold 1x PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) containing 1 x complete proteinase-Inhibitor cocktail (Roche). The proteins were released from cell by sonification for 4 x 15 seconds on ice. 20% Triton X-100 was added to a final concentration of 1% and incubated by mixing gently at 4°C for 30 minutes to aid in solubilisation of the fusion protein. The sample was centrifuged at 15,000 x g for 10 minutes at 4°C and the supernatant was added to 500 µl bed volume of glutathione–Sepharose, and incubation at 4°C for 20 minutes. The beads were washed several times with 1 x PBS buffer, and the next steps were followed below.

• GST pull-down assays

For GST pull-down assays, the GST-102K fusion protein bound to glutathione–Sepharose 4B beads was washed four times with binding buffer (20 mM HEPES pH 7.9, 100 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% Triton X-100, 1 mM DTT and protease inhibitor cocktail). For each reaction, 20 µl of beads (containing ~5 µg GST-102K protein) and 10 µl of [^{35}S]-labelled protein, prepared as described above, were incubated in 300 µl of binding buffer for 2 hours at 4 °C. Beads were washed five times with binding buffer, and bound proteins were eluted in SDS sample buffer and analysed by SDS-PAGE followed by autoradiography.

3.2.3.3 GST-61K pull-down assays (figure 4.10B)

This assay has been described by Makarova *et al*. (2002). Briefly, aliquots of the [^{35}S]-labelled protein 102K, prepared by translation *in vitro*, were incubated in 400 µl of TBST buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Triton X-100) with either a recombinant fusion protein GST-61K or recombinant GST alone. After incubation for 2 hours at 4°C, glutathione–Sepharose beads were added to the mixtures and incubated for another 2 hours. Beads were recovered by centrifugation, washed, and the bound material was eluted in SDS sample buffer and analysed by SDS-PAGE.

3.2.3.4 His pull-down assays (figure 4.14C)

This assay has been described by Laggerbauer *et al*. (2005). Full-length 52K protein and the truncated 52C protein were first subcloned into the pET28a vector and the recombinant His-tagged proteins were then expressed in *E. coli* BL21 and purified by affinity chromatography on Talon beads as described by the manufacturer (Qiagen). ~6 µg of purified His-tagged 52K or 52C protein was incubated with 10 µl of [^{35}S]-labelled HA-tagged 15K protein, which was produced from pGADT7-15K by translation *in vitro* as described in section 3.2.3.1, and 30 µl Talon beads in 250 µl IP buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM β-ME, 0.5% NP-40) at 4 °C for 2 h. In the control reaction, recombinant protein was omitted. Subsequently, beads were washed four times with 0.8 ml of IP buffer. Bound proteins were eluted in SDS sample buffer, analysed by SDS-PAGE, and visualised by autoradiography.

3.2.3.5 Co-immunoprecipitations

For the co-immunoprecipitations shown in figure 4.4B, 10 µg of purified His-90K was incubated with 20 µl of HA-tagged 60K protein prepared by translation *in vitro* for 1 hour at 4 °C. 1.6 µg of anti-pentahistidine antibody (mouse IgG1, Qiagen) and 25 µl protein G–Sepharose were incubated in 400 µl buffer IP (20 mM Tris-HCl pH 8, 150 mM NaCl, 0.05% NP-40). The two tubes were then combined for a further incubation for 1 hour at 4 °C until collection of beads. Precipitates were washed four times with IP buffer and precipitated proteins were analysed by SDS-PAGE followed by autoradiography.

For the co-immunoprecipitations shown in figure 4.9C, 10 µg of His-tagged 15K protein (Reuter *et al*., 1999) together with 1.6 µg anti-pentahistidine antibodies and 25 µl protein A–Sepharose were incubated in 400 µl IP buffer at 4 °C for 2 hours. His-tagged 15K was omitted in the control. Beads were spun down and resuspended in 300 µl IP buffer. 10 µl of c-Myc-102K protein prepared by translation *in vitro* were added to the sample and control reactions, and incubation was continued for 2 hours. Beads were washed five times with 0.8 ml of IP buffer. Bound proteins were eluted in SDS sample buffer and analysed by SDS-PAGE followed by autoradiography.

For the co-immunoprecipitations shown in figures 4.12B and 4.22B, the purified His-tagged 110K protein was dialysed against a buffer containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl and 0.05% NP-40. [^{35}S]-labelled full-length proteins and fragments were prepared by translation *in vitro* as described above. 100 µl of protein A–Sepharose beads were pre-incubated for 3 hours at 4 °C with 100 µl of affinity-purified anti-pep 110 antibody (Makarova *et al*., 2001) in a total volume of 1 ml of phosphate-buffered saline (PBS) pH 8.0, containing 0.5 mg/ml BSA and 50 µg/ml total yeast tRNA. Beads were then washed four times with 1 ml PBS and resuspended in 100 µl IP buffer, resulting in a final volume of 200 µl of α-pep-110 protein A–Sepharose beads. For each reaction, 30 µl of these beads was incubated for 2 hours at 4 °C in 250 µl IP buffer together with 10 µl of [^{35}S]-labelled proteins in the presence or absence of His-tagged 110K protein. Beads were then washed four times with IP buffer, and bound proteins were eluted with SDS sample buffer and analysed by SDS-PAGE followed by autoradiography. The cloning, expression and purification of the 110K protein have already been described (Makarova *et al*., 2001).

For the co-immunoprecipitation assays shown in figures 4.17 and 4.22C, the coding sequences of the 90K and 60K proteins were transcribed *in vitro* with T7 RNA polymerase from the pGBKT7 constructs and translated with [^{35}S]-labelled methionine by using a TNT coupled reticulocyte lysate system (Promega). A total of 15 µl of translation reaction mixture was incubated with 500 ng of baculovirus-expressed full-length p110 protein for 2 hours at 4°C in 250 µl of buffer IP (20 mM Tris-HCl pH7.4, 120 mm NaCl, 5 mM $MgCl_2$, 50 mM KCl,

0.05% NP-40, 1 mM PMSF, 1 mM DTT). Complexes formed were precipitated with p110 antibodies prebound to protein A–Sepharose, followed by seven washes with 0.75 ml each of buffer IP. The bound proteins were eluted by being heated in protein sample buffer, fractionated by SDS-PAGE, and visualized by autoradiography.

4 Results

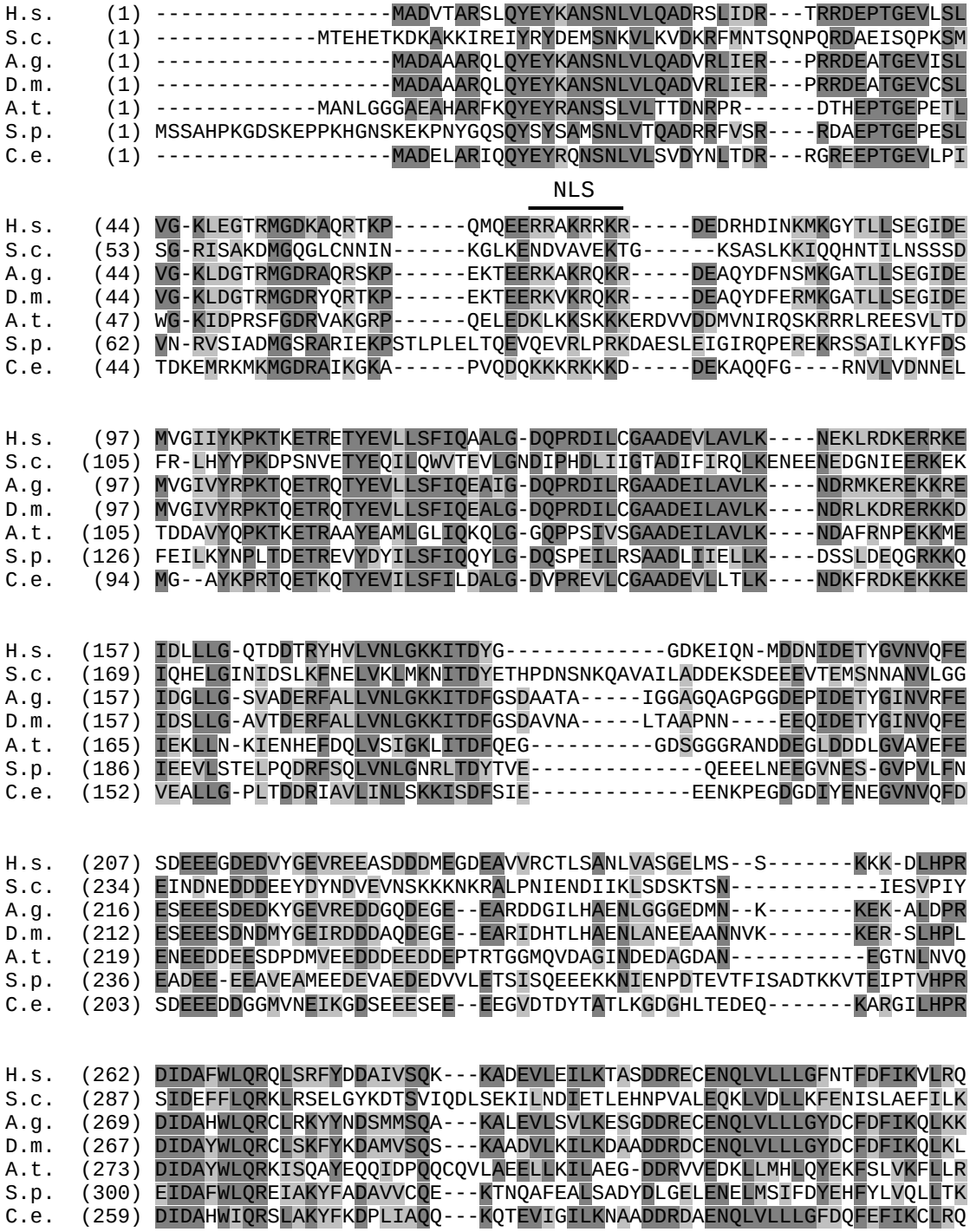
The yeast two-hybrid (Y2H) system and *in vitro* binding assays were used to detect protein-protein interaction in this study. The first part will present the cloning work which is necessary for the complete yeast two-hybrid screening (section 4.1). The following several parts are focused on the results of protein-protein interactions within the tri-snRNP particle (sections 4.2–4.7). Furthermore, the interactions of protein 90K with U6-p110 and U2-SPF30/SMNrp are described (sections 4.8, 4.9). The analysis of point mutations related to retinitis pigmentosa is shown in section 4.10. The final section concerns the Y2H screen with the N-terminal region of protein 200K.

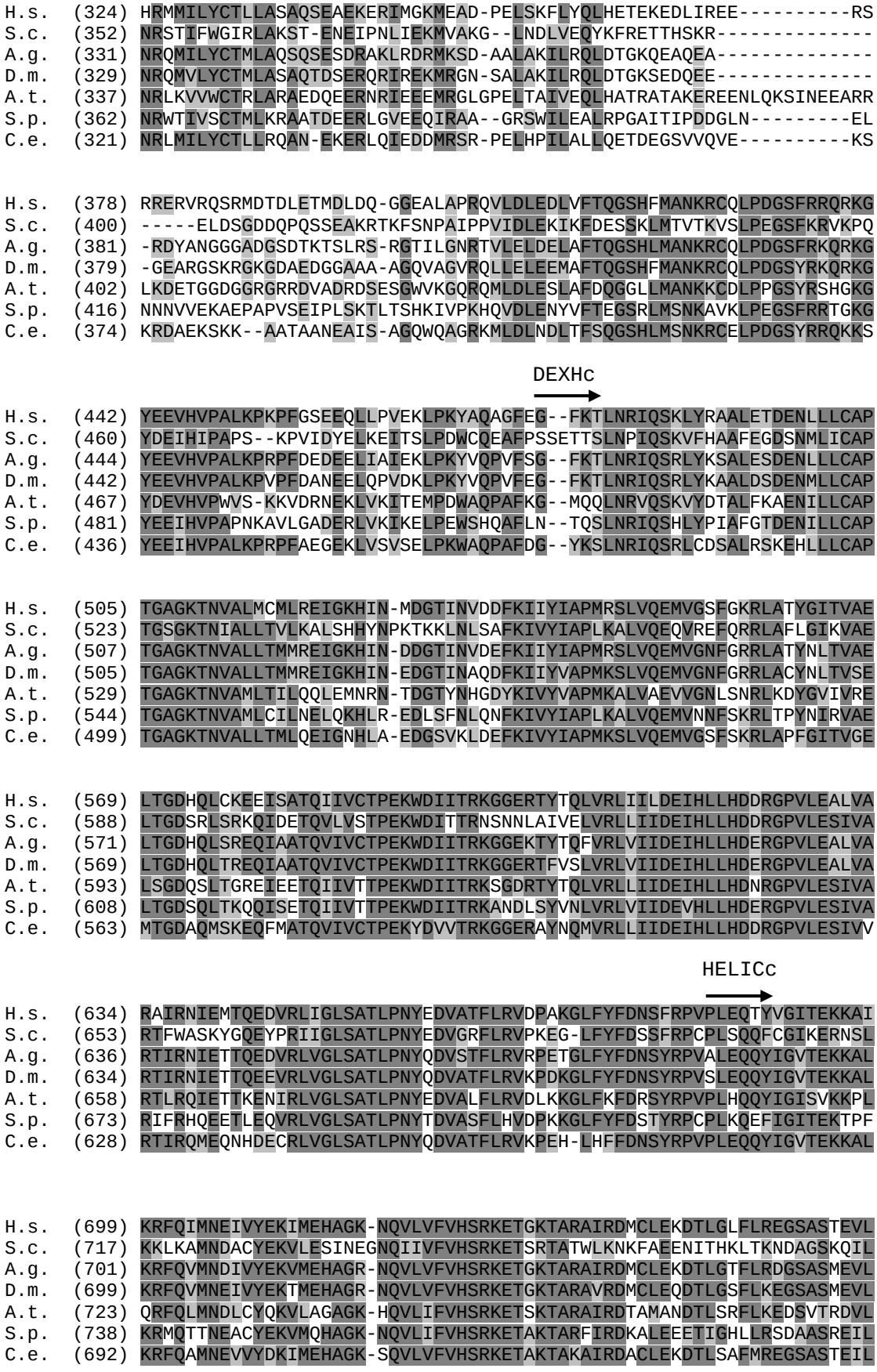
4.1 cDNA cloning

As shown in the table 3.3 in the Materials and Methods, cDNAs for the following proteins had previously been cloned in the laboratory of Prof. Lührmann: U4/U6-specific proteins 15.5K, 20K, 61K, and 90K; U5-specific proteins 15K, 40K, 52K, 100K, 102K and 116K; U4/U6.U5-specific proteins 27K, 65K and 110K. The U6-p110 (SART1) cDNA was generously donated by the laboratory of Prof. A. Bindereif (Medenbach *et al.*, 2004). The cDNAs for the proteins 60K, 200K, 220K, and SPF30/SMNrp were cloned directly from a human Marathon-Ready cDNA library (Clontech) in this study.

Since the sequence of the previously published U5-specific protein 200K (Lauber *et al*. 1996; GenBank accession no. gi|12643640) does not provide the entire N-terminal sequence, the complete 200K gene was cloned using 5'-RACE PCR techniques with a published partial cDNA sequence (Nagase *et al*., 1998) as described in Materials and Methods. The full-length cDNA encodes a protein that has 2136 amino acids and a calculated molecular mass of 244 517 Da. The sequence has been deposited with GenBank (accession no. gi|45861371). The protein not only contains two helicase domains; in addition, each helicase domain is followed by a SEC63 domain, a domain of as yet unknown function, possibly anchoring proteins to RNP complexes (shown schematically in the lower part of figure 4.1) (Ponting, 2000). Orthologues of human 200K protein display a high degree of conservation throughout the eukaryotic kingdom, suggesting that this DExH-box RNA helicase plays an

important role in all eukaryotes. The alignment of the N-terminal region (amino acids 1-870) including the first (N-terminal) helicase domain is shown in figure 4.1. A putative nuclear localization sequence, RRAKRRKR, is present at amino acids 67 to 74 of the human 200K protein. The C terminus of the second (C-terminal) helicase domain (HELICc domain) shows little conservation, which is consistent with its non-essential nature, in contrast to the first helicase domain, which contains lethal mutation sites (Kim and Rossi, 1999).




H.s. (324) HRMMILYCTLLASAQSEAEKERIMGKMEAD-PELSKFLYQLHETEKEDLIREE----------RS
S.c. (352) NRSTIFWGIRLAKST-ENEIPNLIEKMVAKG--LNDLVEQYKFRETTHSKR--------------
A.g. (331) NRQMILYCTMLAQSQSESDRAKLRDRMKSD-AALAKILRQLDTGKQEAQEA--------------
D.m. (329) NRQMVLYCTMLASAQTDSERQRIREKMRGN-SALAKILRQLDTGKSEDQEE--------------
A.t. (337) NRLKVVWCTRLARAEDQEERNRIEEEMRGLGPELTAIVEQLHATRATAKEREENLQKSINEEARR
S.p. (362) NRWTIVSCTMLKRAATDEERLGVEEQIRAA--GRSWILEALRPGAITIPDDGLN---------EL
C.e. (321) NRLMILYCTLLRQAN-EKERLQIEDDMRSR-PELHPILALLQETDEGSVVQVE----------KS
H.s. (378) RRERVRQSRMDTDLETMDLDQ-GGEALAPRQVLDLEDLVFTQGSHFMANKRCQLPDGSFRRQRKG
S.c. (400) -----ELDSGDDQPQSSEAKRTKFSNPAIPPVIDLEKIKFDESSKLMTVTKVSLPEGSFKRVKPQ
A.g. (381) -RDYANGGGADGSDTKTSLRS-RGTILGNRTVLELDELAFTQGSHLMANKRCQLPDGSFRKQRKG
D.m. (379) -GEARGSKRGKGDAEDGGAAA-AGQVAGVRQLLELEEMAFTQGSHFMANKRCQLPDGSYRKQRKG
A.t. (402) LKDETGGDGGRGRRDVADRDSESGWVKGQRQMLDLESLAFDQGGLLMANKKCDLPPGSYRSHGKG
S.p. (416) NNNVVEKAEPAPVSEIPLSKTLTSHKIVPKHQVDLENYVFTEGSRLMSNKAVKLPEGSFRRTGKG
C.e. (374) KRDAEKSKK--AATAANEAIS-AGQWQAGRKMLDLNDLTFSQGSHLMSNKRCELPDGSYRRQKKS
DEXHc
H.s. (442) YEEVHVPALKPKPFGSEEQLLPVEKLPKYAQAGFEG--FKTLNRIQSKLYRAALETDENLLLCAP
S.c. (460) YDEIHIPAPS--KPVIDYELKEITSLPDWCQEAFPSSETTSLNPIQSKVFHAAFEGDSNMLICAP
A.g. (444) YEEVHVPALKPRPFDEDEELIAIEKLPKYVQPVFSG--FKTLNRIQSRLYKSALESDENLLLCAP
D.m. (442) YEEVHVPALKPVPFDANEELQPVDKLPKYVQPVFEG--FKTLNRIQSRLYKAALDSDENMLLCAP
A.t. (467) YDEVHVPWVS-KKVDRNEKLVKITEMPDWAQPAFKG--MQQLNRVQSKVYDTALFKAENILLCAP
S.p. (481) YEEIHVPAPNKAVLGADERLVKIKELPEWSHQAFLN--TQSLNRIQSHLYPIAFGTDENILLCAP
C.e. (436) YEEIHVPALKPRPFAEGEKLVSVSELPKWAQPAFDG--YKSLNRIQSRLCDSALRSKEHLLLCAP
H.s. (505) TGAGKTNVALMCMLREIGKHIN-MDGTINVDDFKIIYIAPMRSLVQEMVGSFGKRLATYGITVAE
S.c. (523) TGSGKTNIALLTVLKALSHHYNPKTKKLNLSAFKIVYIAPLKALVQEQVREFQRRLAFLGIKVAE
A.g. (507) TGAGKTNVALLTMMREIGKHIN-DDGTINVDEFKIIYIAPMRSLVQEMVGNFGRRLATYNLTVAE
D.m. (505) TGAGKTNVALLTMMREIGKHIN-EDGTINAQDFKIIYVAPMKSLVQEMVGNFGRRLACYNLTVSE
A.t. (529) TGAGKTNVAMLTILQQLEMNRN-TDGTYNHGDYKIVYVAPMKALVAEVVGNLSNRLKDYGVIVRE
S.p. (544) TGAGKTNVAMLCILNELQKHLR-EDLSFNLQNFKIVYIAPLKALVQEMVNNFSKRLTPYNIRVAE
C.e. (499) TGAGKTNVALLTMLQEIGNHLA-EDGSVKLDEFKIVYIAPMKSLVQEMVGSFSKRLAPFGITVGE
H.s. (569) LTGDHQLCKEEISATQIIVCTPEKWDIITRKGGERTYTQLVRLIILDEIHLLHDDRGPVLEALVA
S.c. (588) LTGDSRLSRKQIDETQVLVSTPEKWDITTRNSNNLAIVELVRLLIIDEIHLLHDDRGPVLESIVA
A.g. (571) LTGDHQLSREQIAATQVIVCTPEKWDIITRKGGEKTYTQFVRLVIIDEIHLLHDERGPVLEALVA
D.m. (569) LTGDHQLTREQIAATQVIVCTPEKWDIITRKGGERTFVSLVRLVIIDEIHLLHDERGPVLEALVA
A.t. (593) LSGDQSLTGREIEETQIIVTTPEKWDIITRKSGDRTYTQLVRLLIIDEIHLLHDNRGPVLESIVA
S.p. (608) LTGDSQLTKQQISETQIIVTTPEKWDIITRKANDLSYVNLVRLVIIDEVHLLHDERGPVLESIVA
C.e. (563) MTGDAQMSKEQFMATQVIVCTPEKYDVVTRKGGERAYNQMVRLLIIDEIHLLHDDRGPVLESIVV
HELICc
H.s. (634) RAIRNIEMTQEDVRLIGLSATLPNYEDVATFLRVDPAKGLFYFDNSFRPVPLEQTYVGITEKKAI
S.c. (653) RTFWASKYGQEYPRIIGLSATLPNYEDVGRFLRVPKEG-LFYFDSSFRPCPLSQQFCGIKERNSL
A.g. (636) RTIRNIETTQEDVRLVGLSATLPNYQDVSTFLRVRPETGLFYFDNSYRPVALEQQYIGVTEKKAL
D.m. (634) RTIRNIETTQEEVRLVGLSATLPNYQDVATFLRVKPDKGLFYFDNSYRPVSLEQQYIGVTEKKAL
A.t. (658) RTLRQIETTKENIRLVGLSATLPNYEDVALFLRVDLKKGLFKFDRSYRPVPLHQQYIGISVKKPL
S.p. (673) RIFRHQEETLEQVRLVGLSATLPNYTDVASFLHVDPKKGLFYFDSTYRPCPLKQEFIGITEKTPF
C.e. (628) RTIRQMEQNHDECRLVGLSATLPNYQDVATFLRVKPEH-LHFFDNSYRPVPLEQQYIGVTEKKAL
H.s. (699) KRFQIMNEIVYEKIMEHAGK-NQVLVFVHSRKETGKTARAIRDMCLEKDTLGLFLREGSASTEVL
S.c. (717) KKLKAMNDACYEKVLESINEGNQIIVFVHSRKETSRTATWLKNKFAEENITHKLTKNDAGSKQIL
A.g. (701) KRFQVMNDIVYEKVMEHAGR-NQVLVFVHSRKETGKTARAIRDMCLEKDTLGTFLRDGSASMEVL
D.m. (699) KRFQVMNEIVYEKTMEHAGR-NQVLVFVHSRKETGKTARAVRDMCLEQDTLGSFLKEGSASMEVL
A.t. (723) QRFQLMNDLCYQKVLAGAGK-HQVLIFVHSRKETSKTARAIRDTAMANDTLSRFLKEDSVTRDVL
S.p. (738) KRMQTTNEACYEKVMQHAGK-NQVLIFVHSRKETAKTARFIRDKALEEETIGHLLRSDAASREIL
C.e. (692) KRFQAMNEVVYDKIMEHAGK-SQVLVFVHSRKETAKTAKAIRDACLEKDTLSAFMREGSASTEIL

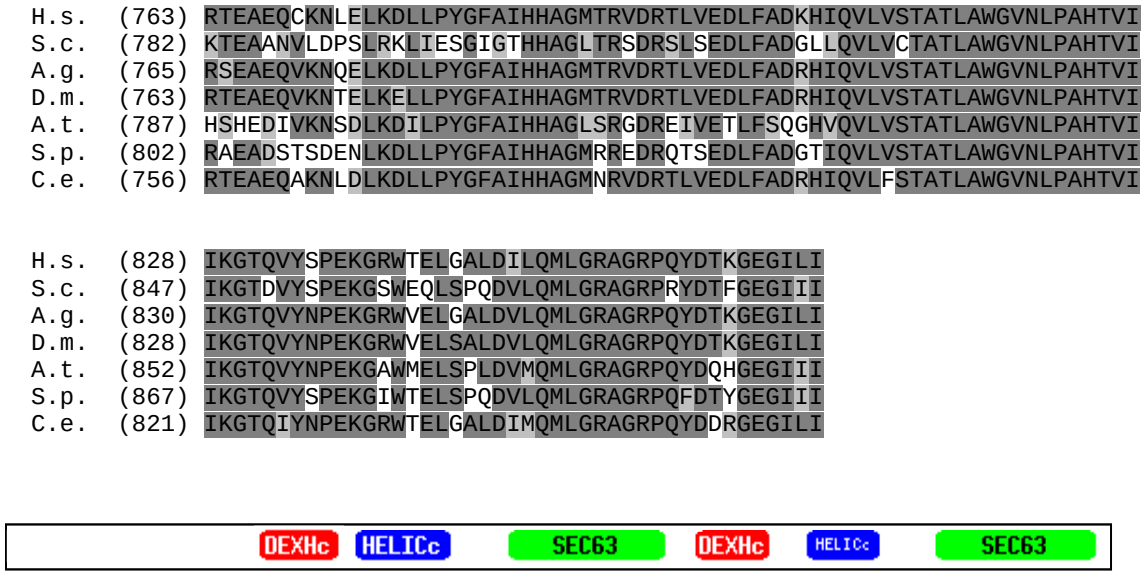
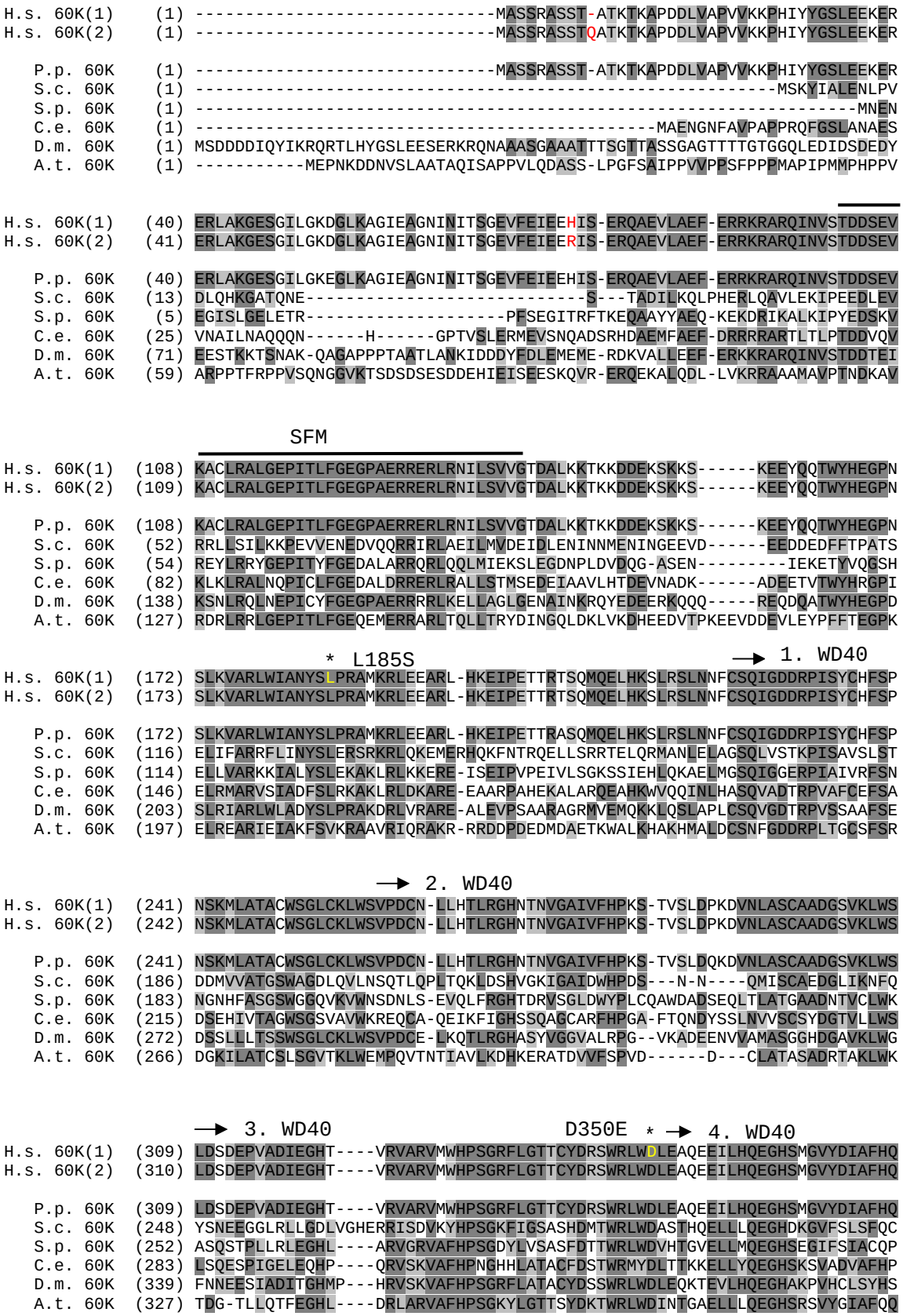


Figure 4.1. Alignment of the N-terminal region of putative hBrr2 orthologues from *Homo sapiens* (H.s., gi|45861372), *Saccharomyces cerevisiae* (S.c., gi|6321020), *Anopheles gambiae* (A.g., gi|31242349), *Drosophila melanogaster* (D.m., gi|28574898), *Arabidopsis thaliana* (A.t., gi|15218086), *Schizosaccharomyces pombe* (S.p., gi|19114258), *Caenorhabditis elegans* (C.e., gi|17537127). NLS, putative nuclear localization sequence. Identical and similar residues are indicated through dark grey or light grey boxes, respectively. The schematic drawing below depicts the domain structure of the protein: DEXHc, DEXH-box helicases; HELICc, helicase superfamily C-terminal domain; SEC63 domain of unknown function found in SEC63p. The domain structure was predicted from NCBI's Conserved Domain Database and Search Service, v2.02.

Because of inconsistencies between the various sequences deposited for the 60K protein in published GenBank databases (accession no. gi|2653735, Lauber *et al*., 1997; accession no. gi|2708304, Horowitz *et al*., 1997; accession no. gi|2853286, Wang *et al*., 1997), the cDNA was cloned again by PCR from the same cDNA library used for protein 200K. This new sequence has been deposited with GenBank (accession no. gi|45861373). Figure 4.2 shows the alignment of the new cloned sequence with Horowitz's human sequence (gi|2708304) and several homologue sequences from other organisms. Two human 60K sequences are nearly identical. Protein 60K contains two highly conserved domains, the SFM domain in the N-terminal region and the WD40 domain in the C-terminal region. However, yeast homologue proteins lack the SFM near the N-terminus.


SFM
* L185S
1. WD40
2. WD40
3. WD40
D350E * 4. WD40

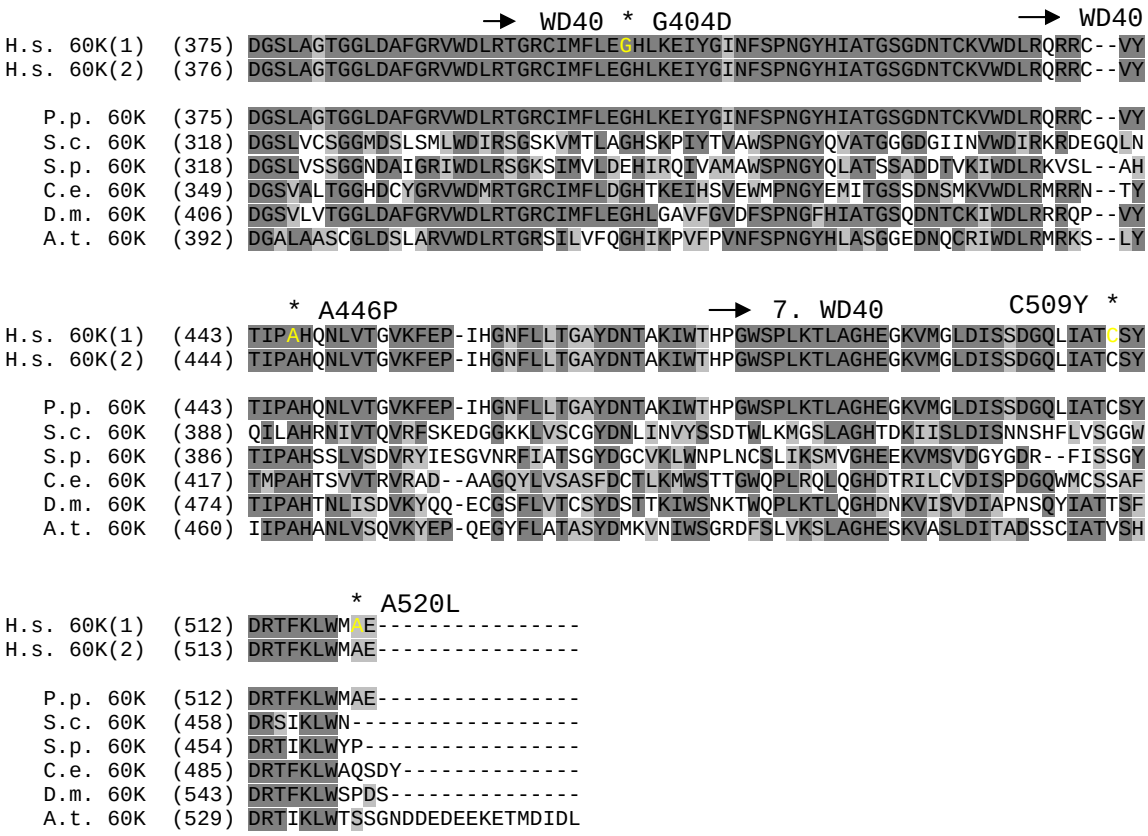


Figure 4.2. Alignment of putative 60K orthologues from *Homo sapiens* (H.s.(1), gi|45861373; H.s.(2), gi|2708304), *Pongo pygmaeus* (P.p., gi| 56403896), *Saccharomyces cerevisiae* (S.c., gi|6325435), *Schizosaccharomyces pombe* (S.p., gi|19113878), *Caenorhabditis elegans* (C.e., gi|17505895), *Drosophila melanogaster* (D.m., gi|21355245), *Arabidopsis thaliana* (A.t., gi|30794128). The variations in the two positions between the two human sequences are shown in red. The N-terminal SFM domain and the seven C-terminal WD40 repeats are indicated above the sequences. SFM, Splicing Factor Motif, present also in Prp18p. The residues shown in yellow and marked with an asterisk (*) above the residues were replaced in the mutant 60K clone. This clone was used for the subsequent binding study, as described below. Identical and similar residues are indicated by dark and light grey boxes, respectively.

The cloning of cDNA for the 220K protein is described in Materials and Methods. The entire sequence encoding the protein is nearly identical to the sequence published in the GenBank database (gi|3661609; Luo *et al.*, 1999), and the sequence is given in the Appendix.

4.2 Protein-protein interactions within U4/U6-snRNP particle

4.2.1 The interactions of proteins in the 20K•60K•90K RNA-free heterotrimer

Five U4/U6-specific proteins remain associated with the U4/U6 snRNA duplex under high-salt conditions, when the tri-snRNP dissociates into the U4/U6 and U5 snRNPs. These proteins are 90K, 61K, 60K, 20K and 15.5K. The stable binding defines them as U4/U6-specific proteins (reviewed by Will and Lührmann, 2001). Of these, 20K, 60K and 90K form a stable RNA-free heterotrimeric complex (Horowitz *et al*., 1997; Teigelkamp *et al*., 1998). The latter two proteins are also referred to as hPrp4 and hPrp3, since they are the orthologues of the *S. cerevisiae* Prp4 and Prp3 splicing factors, respectively. To investigate the binding interactions among these five proteins, each was subcloned into the pGADT7 and pGBKT7 vectors to generate bait and prey fusions respectively, and these plasmids were co-transformed into yeast and tested for interaction as described in Materials and Methods. Consistently with the previous co-immunoprecipitation results (Teigelkamp *et al*., 1998), the interaction between the 20K and 60K proteins was confirmed during Y2H analysis with each protein serving in turn as bait or prey (figure 4.3). Remarkably, this interaction remained the only one that could be detected among the five U4/U6 proteins by the two-hybrid method (figure 4.3).

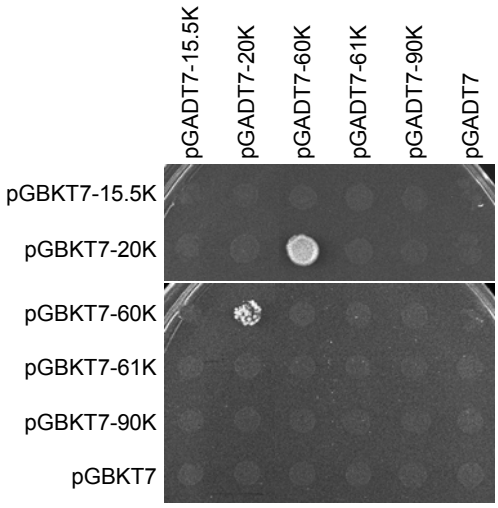


Figure 4.3. Y2H interactions between proteins 20K and 60K. Binding was selected on the minimal synthetic dropout medium lacking leucine, tryptophan, histidine, and adenine (hereafter termed SD/-Leu-Trp-His-Ade) at 30°C for 3-5 days. Gal4 DNA-binding domain fusion protein 20K (pGBKT7-20K) interacts with Gal4 activation domain fusion protein 60K (pGADT7-60K). The same result was observed in the reciprocal combination (i.e. pGADT7-20K/pGBKT7-60K). The empty vector pGADT7 or pGBKT7 served as a negative control.

Using *in vitro* binding assays, one could observe the interaction between 60K and 90K, in addition to the interaction between 20K and 60K (figure 4.4). In one experiment, all proteins were produced and labelled with [^{35}S]-methionine by *in vitro* translation as described in the Materials and Methods. As shown in figure 4.4A, [^{35}S]-labelled HA-tagged 20K (lane 5) and 90K protein (lane 7) can be co-precipitated with c-Myc-tagged 60K using anti-c-Myc antibody. However, the binding efficiency of 90K is much lower than that of 20K (compare lanes 5 and 7). The same results were obtained when these two proteins were added together in one reaction (lane 8). The binding of 60K to 90K was also observed when 90K was expressed as a recombinant His-tagged protein and 60K was translated *in vitro* as an HA-tagged protein. Figure 4.4B shows that His-tagged 90K protein specifically co-precipitated HA-60K protein that had been prepared by translation *in vitro*, and the level of HA-60K was significantly above the background level obtained in the absence of His-tagged 90K protein. Several conclusions can be drawn from above studies. First, proteins 60K and 90K are indeed associated *in vitro*, although this association was not observed in the Y2H system. This is in agreement with the recent result of another group, who observed binding between 60K and 90K by using co-immunoprecipitation and isothermal titration calorimetry (Gonzalez-Santos *et al*., 2002). Second, in the 20K•60K•90K complex, each interaction occurs independently (see figure 4.4A: lane 5 and 7). Third, in the complex, one interaction does not affect the other, implying that the proteins 20K and 90K bind to distinct domains of 60K (see figure 4.4A: lane 8). The observed Y2H interactions between 60K and 20K (above) and between 90K and tri-snRNP-110K, U6-p110, SPF30/SMNrp (figures 4.12, 4.17, 4.19) indicate that 60K and 90K are both properly expressed in the yeast cell. The failure of 60K and 90K to show an interaction in the Y2H system must therefore be attributed to another cause (see Discussion).

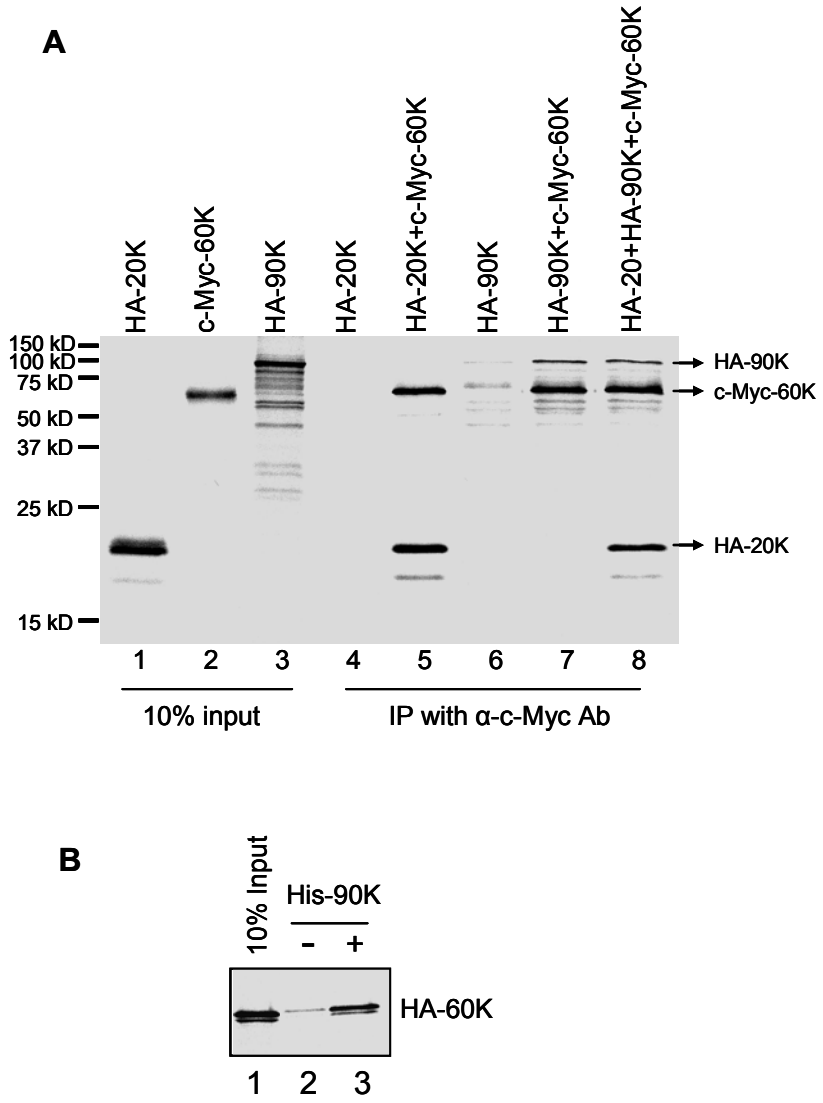


Figure 4.4. *In vitro* interactions between proteins 20K, 60K and 90K. (A) [^{35}S]-labelled fusion proteins HA-20K, c-Myc-60K and HA-90K were produced *in vitro* respectively and the protein positions on the SDS-PAGE are shown on the right of the panel. c-Myc-60K was incubated with HA-20K or/and 90K (lane 5, 7, 8), respectively, in 250 µl buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.05% NP-40) and precipitated by c-Myc Ab which immobilised to protein-A Sepharose. HA-20K (lane 4) or HA-90K (lane 6) alone was incubated with c-Myc Ab immobilised beads in the negative control reaction. 10% of each translates was loaded on gel as input (lane 1, 2, 3). Molecular marker is indicated on the left of panel. (B) *In vitro* interactions between 60K and 90K proteins. Purified recombinant His-tagged 90K was first incubated with *in vitro* translated [^{35}S]-labelled HA-tagged 60K protein and then with anti-pentahistidine antibodies and protein G-Sepharose to precipitate His-tagged 90K. The precipitated proteins were fractionated by SDS-PAGE and visualized by autoradiography. Mock precipitations without His-tagged 90K were included as a negative control. The input lane represents 10% of the 60K amount added to the reaction.

In the complex of 20K•60K•90K, it has been shown that human 20K protein binds to the SFM domain in the N-terminus of 60K (Horowitz *et al.*, 2002; Reidt *et al.*, 2003). In yeast, the C-terminal WD40 domain of Prp4p, which is conserved in the human homologue 60K protein, is required for the binding of protein Prp3p (Ayadi *et al*., 1998). One may ask whether the same WD40 domain in human 60K protein is also involved in the 90K binding. To obtain more information on the 90K binding sites, one 60K mutant clone, obtained during the cloning of wild-type 60K, was used for the binding assays. This clone contains six point mutations: L185S, D350E, G404D, A446P, C509Y, and A520L, which are indicated as asterisks (*) above the residues in the figure 4.2. The position of L185S is near to the upstream of the WD40 domain; the mutation D350E, G404D, A446P, C509Y is located in the third, fifth, sixth, and seventh WD40 repeats respectively; A520L is in the position of the second-last amino acid at the 3' end. Of these mutations, the residues L185, D350, G404, A446 are highly conserved among the putative 60K orthologues (see figure 4.2). The experimental data show that the mutant 60K, like its wild type, efficiently binds to 20K in the Y2H system (data not shown) and in the co-immunoprecipitation assay as well (figure 4.5, left panel). However, it failed to bind 90K protein in both systems (figure 4.5 and Y2H data not shown). As positive control, the wild-type 60K protein could specifically bind 90K as shown in the figure 4.5 (left panel). Again, this could be shown by another assay, where only wild-type 60K, but not mutant clone, was co-precipitated by recombinant His-tagged 90K protein (right panel of the figure 4.5). Taken together, these point mutants in the gene of 60K completely disrupt the interaction with 90K, while the interaction between 20K and 60K remains unaffected. Considering that most of these mutations are located in the WD40 domain and the rest close to this domain, one may propose that the WD40 domain is indeed responsible for the binding of 90K protein.

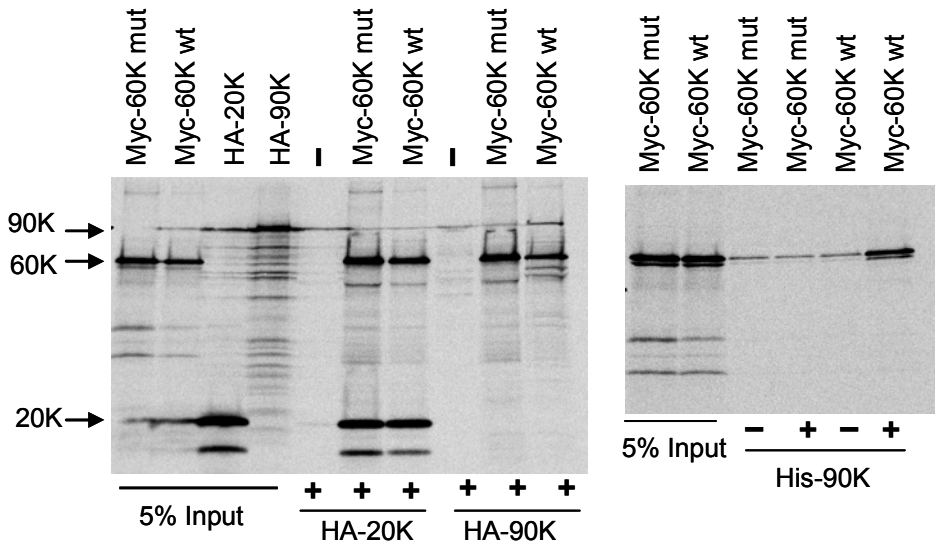


Figure 4.5. Mutational analysis of the *in vitro* interaction between proteins 20K, 60K and 90K. Left panel: [^{35}S]-labelled fusion proteins HA-20K, HA-90K and c-Myc-60K (wild-type and mutant clones) were produced *in vitro* respectively as mentioned in the Materials and Methods and the protein positions on the SDS-PAGE are shown on the left of the panel. c-Myc-60K wild-type or mutant fusion protein was incubated with HA-20K or 90K in 400 µl buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.05% NP-40) and precipitated by c-Myc Ab which immobilised to protein-G Sepharose. HA-20K or HA-90K alone was incubated with c-Myc Ab immobilised beads in the negative control reaction. 5% of each translates was loaded on gel as input. The precipitated proteins were fractionated by 12% SDS-PAGE and visualized by autoradiography. Right panel: Purified recombinant His-tagged 90K was first incubated with *in vitro* translated [^{35}S]-labelled HA-tagged 60K wild-type and mutant fusion proteins respectively and then with anti-pentahistidine antibodies and protein G-Sepharose to precipitate His-tagged 90K in 400 µl buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.05% NP-40). The precipitated proteins were fractionated by 12% SDS-PAGE and visualized by autoradiography. Mock precipitations without His-tagged 90K were included as a negative control. The input lane represents 5% of the 60K amount added to the reaction.

4.3 Protein-protein interactions within the U5-snRNP particle

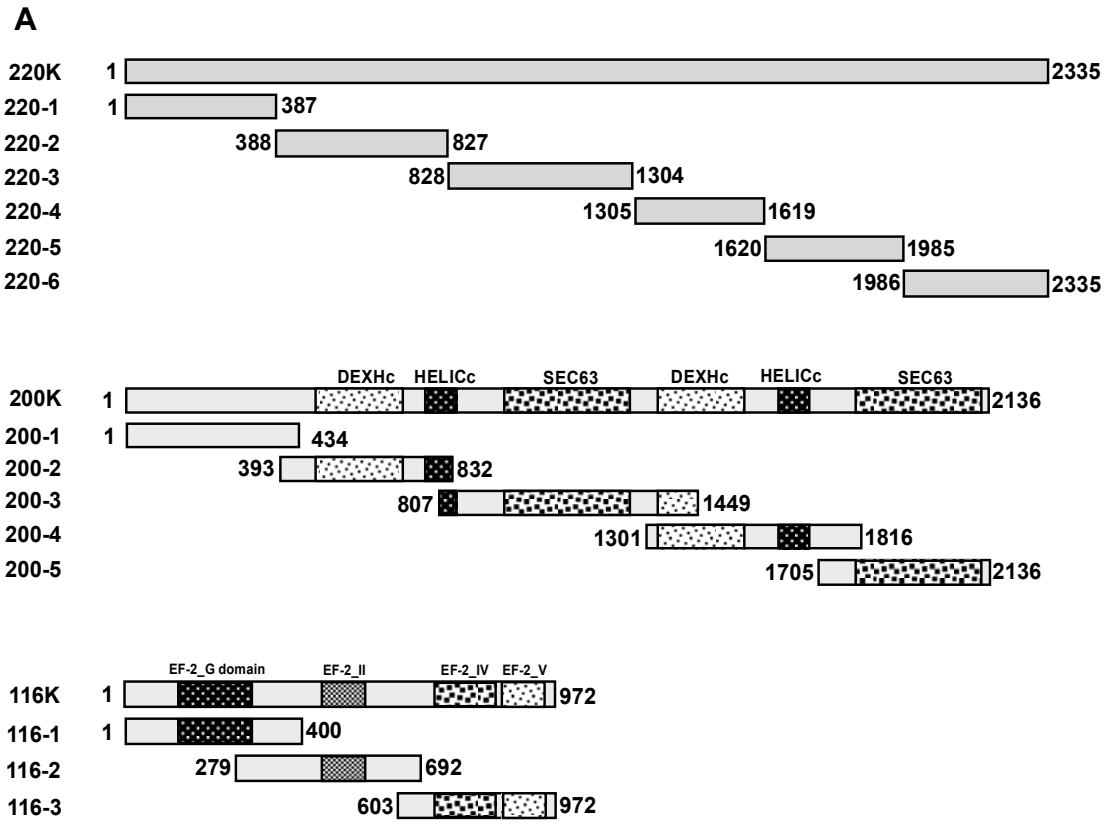
4.3.1 Interactions between U5 snRNP proteins in the RNA-free 220K•200K•116K•40K heterotetramer

The U5-specific proteins 220K, 200K, 116K and 40K, form an RNA-free protein complex which remains stable even in 0.2 M NaSCN. Dissociation studies identified the U5-220K as the hub of this complex (Achsel *et al.*, 1998). Y2H was applied to analyse the interactions between these proteins in more detail.

The use of full-length coding sequences of 220K, 200K, 116K and 40K in bait or prey constructs did not result in the appearance of any clear protein-protein interactions (figure 4.6B). Occasionally, two-hybrid interaction between two full-length proteins (especially for the large proteins) is just undetectable. Similar behaviour has been observed by others in the case of yeast Prp8p (van Nues and Beggs, 2001); full-length Prp8p as bait did not produce such obviously specific interactions as its fragments did. In a more detailed investigation, each of these proteins was truncated for bait and prey construction (figure 4.6A). For proteins 200K and 116K, the truncated proteins corresponded to recognisable, known protein domains. Thus, 200K was divided into five overlapping fragments on the basis of the structural domains deducible from the amino-acid sequence: fragments 2 and 4 cover the two helicase domains, while fragments 3 and 5 harbour the two SEC63 domains (see figures 4.1, 4.6A). Likewise, protein 116K was divided up by following the boundaries of the conserved GTPase structural features known from EF-2. The first fragment contains the GTP-binding domain of EF-2, while fragments 2 and 3 cover domain II and domains IV+V in the EF-2 nomenclature (figure 4.6A). In contrast to these two proteins, the sequence of 220K, although highly conserved in all eukaryotes, does not offer any clues as to the structural organization of the protein. 220K was therefore divided into six arbitrary fragments of roughly equal size (315-477 amino acids; figure 4.6 A).

Bait and prey constructs with full-length and truncated proteins were again tested in the Y2H assay. Binding between the three proteins 220K, 200K, 116K became clearly visible (figure 4.6B). The domain contacts deduced are shown schematically in figure 4.6C. The N-terminal fragment 220-1 (aa 1-387) and the C-terminal fragment 220-6 (aa 1986-2335) of 220K interact with each other, as well as with fragment 200-4 of 200K (aa 1301-1816). This fragment contains the C-terminal helicase domain of 200K and in turn interacts with the central and C-terminal fragments of the 116K protein (116-2: aa 279-692 and 116-3: aa 603-972). The central fragment of 116K (116-2) interacts with both terminal fragments of 220K (220-1 and 220-6), while the C terminal fragment (116-3) of 116K interacts with the C-terminal fragment of 220K (220-6), thus closing the circle of these three proteins. All fragments that interact with fragments of 116K

also interact with full-length 116K, indicating that protein 116K is properly expressed and correctly folded in yeast. It thus appears that contacts between specific conserved functional domains of the three large proteins make a major contribution to the creation of the heterotetrameric complex.



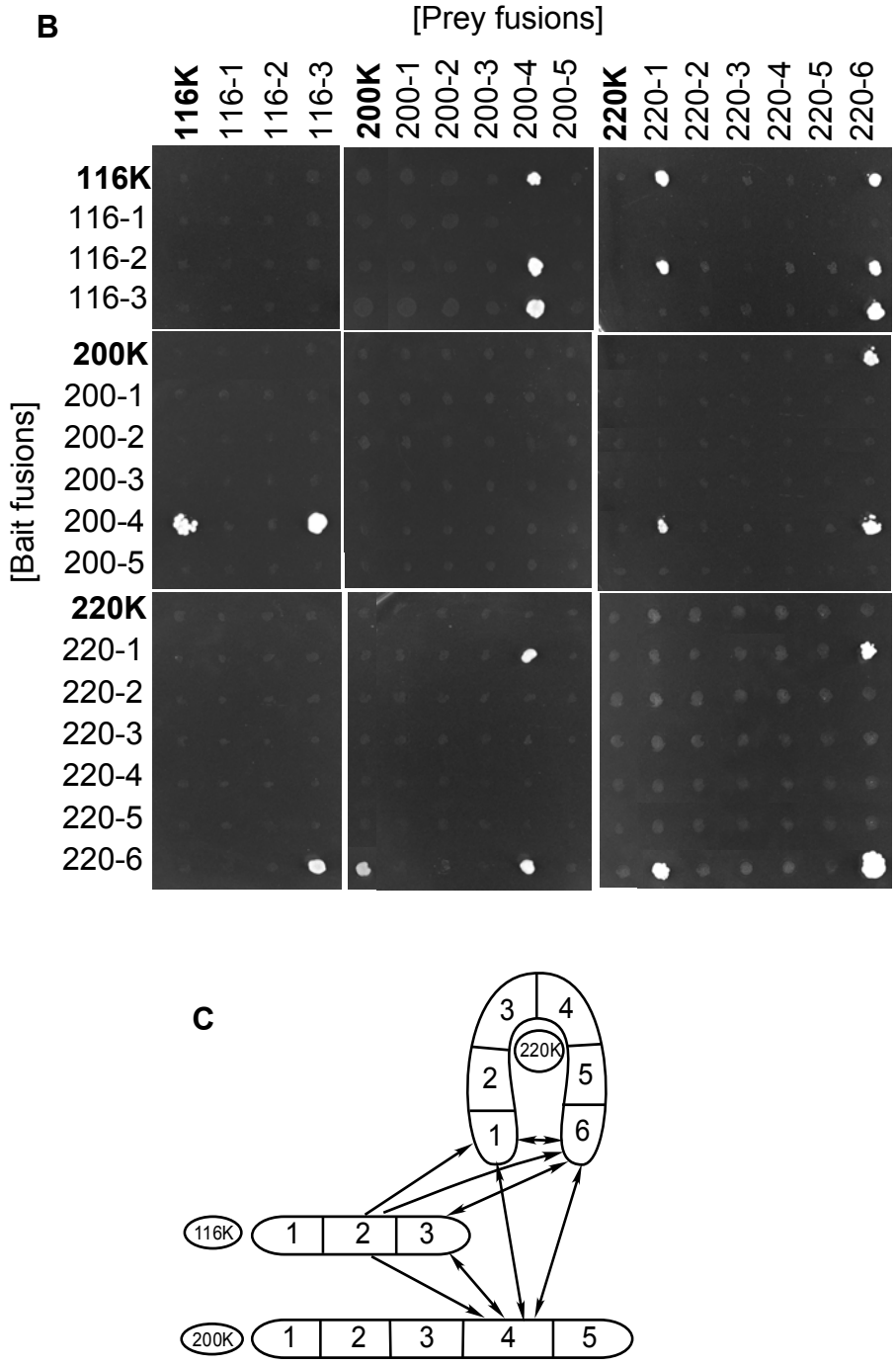


Figure 4.6. (A) Schematic representation of the 220K, 200K and 116K proteins and their fragments used in the yeast two-hybrid assays. DEXHc, DEXH-box helicases; HELICc, helicase superfamily C-terminal domain; SEC63, domain of unknown function in Sec63p and other proteins; EF-2_G, eukaryotic elongation factor 2 GTP-binding domain; EF-2_II (IV, V), eukaryotic elongation factor 2 domains II, IV and V. (B) Y2H interaction between the three large U5-specific proteins 220K, 200K and 116K. All the three proteins and truncated fragments were cloned in the both pGADT7 and pGBKT7 respectively, and then were co-transformed into AH109 cells in all combinations. The positive results were observed on the selective medium SD/-Leu-Trp-His-Ade at 30°C for 3-7 days. (C) Schematic presentation of the domain interactions between protein 220K, 200K and 116K found in Y2H system (B).

Binding partners of protein 40K could not be identified with any full-length or truncated proteins by Y2H analysis (data not shown). In a previous paper Achsel *et al.* have described binding interactions of protein 220K with protein 40K and 116K using far-western blot analysis (Achsel *et al*., 1998). In this study, the interaction was confirmed by the fact that [^{35}S]-labelled c-Myc-tagged protein 40K co-immunoprecipitated with [^{35}S]-labelled HA-tagged 220K, using protein A–Sepharose and anti-HA antibodies (figure 4.7).

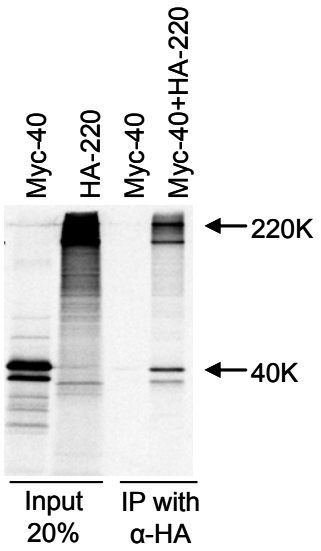


Figure 4.7. *In vitro* interactions between 40K and 220K proteins. *In vitro* translated [^{35}S]-labelled HA-tagged 220K protein was incubated with c-Myc-tagged 40K protein, and then precipitated with anti-HA antibodies and protein A–Sepharose. The precipitated proteins were fractionated by SDS-PAGE and visualized by autoradiography. Mock precipitation without HA-tagged 220K was included as a negative control. The input lane represents 20% of the proteins amount added to the reaction.

4.3.1.1 Overexpression of the protein fragments from 220K and 200K

In order to obtain recombinant proteins for biochemical study, all 220K and 200K fragments used in the Y2H were subcloned into the pET28a-c expression vectors. The plasmids were introduced into BL21(DE3) and the cell cultures were induced in 1 mM IPTG at 37°C for 3 hours. Most proteins were found to be expressed but insoluble. As an example, figure 4.8 shows the fractions of some fragments of 220K protein eluted under denaturing conditions with urea. Compared with fragments 220-1 and 220-4, the yields of 220-5 and 220-6 are much higher. The denatured protein 220-5 was used to produce antibodies in rabbits. The resulted anti-200-5 antibodies were found to be very specific and exhibited no cross reactions (data not shown).

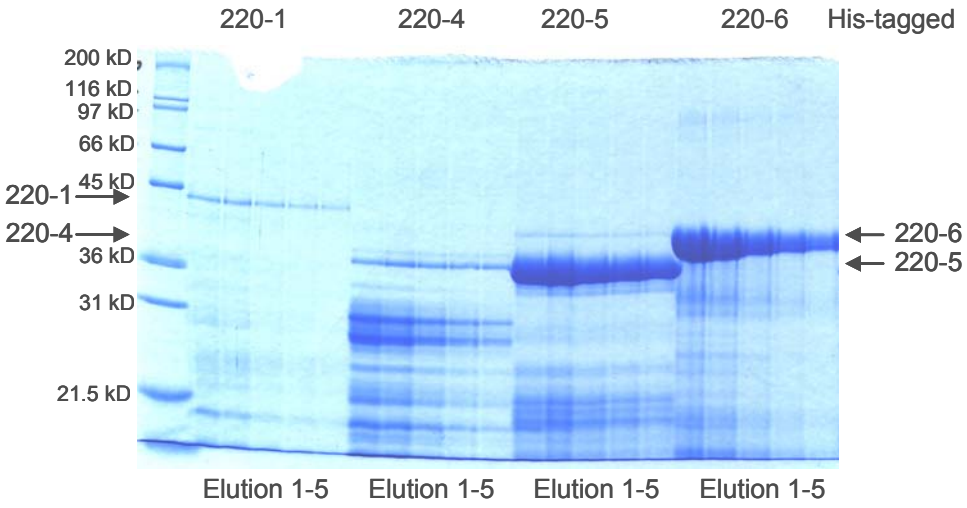


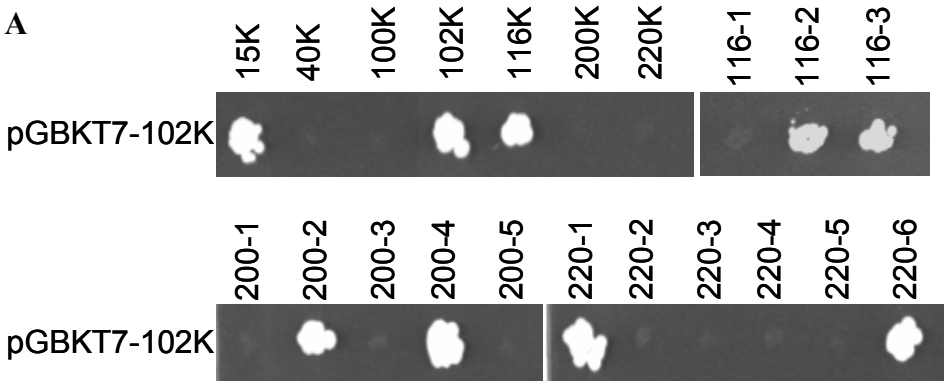
Figure 4.8. Expression of protein 220K fragments in *E. coli*. Five different inclusion body fractions of each protein were run on a 13% SDS-PAGE gel and stained with Coomassie blue.

4.3.2 Multiple interactions of 102K with other U5 snRNP proteins within the tri-snRNP particle

Previous studies had indicated that 102K is stably associated with U5 snRNP via protein-protein interactions (Makarov *et al*., 2000). However, little is known about its interacting partner. By the yeast two-hybrid method, protein 102K was tested against the full-length U5 snRNP proteins including 15K, 40K, 100K, 102K, 116K, 200K, and 220K, and subsequently the fragments from the latter three proteins, which are described in section 4.3.1. Figure 4.9A shows the outcome when protein 102K was used as a bait fusion. Multiple interactions were identified. First, protein 102K interacts with the full-length proteins 15K, 116K, and 102K itself. In the next experiment, protein 102K was demonstrated to interact with distinct domains of each of the three large proteins, which include 116K fragments 2 and 3, 200K fragments 2 and 4, and 220K fragments 1 and 6. The interaction between 102K and 15K were able to be observed in the reverse combinations, where 102K was constructed as a prey fusion (data not shown).

To confirm the results of Y2H analysis, an *in vitro* GST pull-down assay was performed. Proteins 15K and 102K, protein 116K and its mutants, protein 200K fragments 2 and 4, protein 220K fragments 1 and 6 – all produced by *in vitro*

translation – were incubated with immobilized GST-102K or GST. GST-102K (lane 3), but not GST (lane 2), significantly binds [^{35}S]-labelled c-Myc-102K, c-Myc-200-4 (aa 1301–1816), and c-Myc-220-1 (aa 1–387), but not protein 116K or its mutants, or fragments 200-2 and 220-6 (lower panel of figure 4.9B and data not shown). Protein 102K and 200-4 show higher binding efficiency than 220-1 does. Fragments 116-2 and 200-2 (the degraded band) display signals only slightly above background. The Coomassie blue staining gel shows that nearly equal amounts of recombinant proteins in each reaction were eluted from glutathione-Sepharose beads (figure 4.9B, upper panel). The fact that GST-102K fails to interact with the 15K protein can be attributed to the N-terminal GST tag possibly blocking the binding site for 15K. Therefore, to confirm the interaction between 102K and 15K proteins, a co-immunoprecipitation experiment with His-tagged 15K was performed. Figure 4.9C shows that [^{35}S]-labelled c-Myc-102K is now specifically precipitated with His-tagged 15K. The expression and purification of His-tagged 15K protein was described previously (Reuter and Ficner, 1999). In summary, the binding assay *in vitro* independently confirmed the interactions of protein 102K with proteins 15K, 102K, 220-1, and 200-4. However, there are no significant *in vitro* interactions of protein 102K with protein 116K or fragments 200-2 and 220-6.



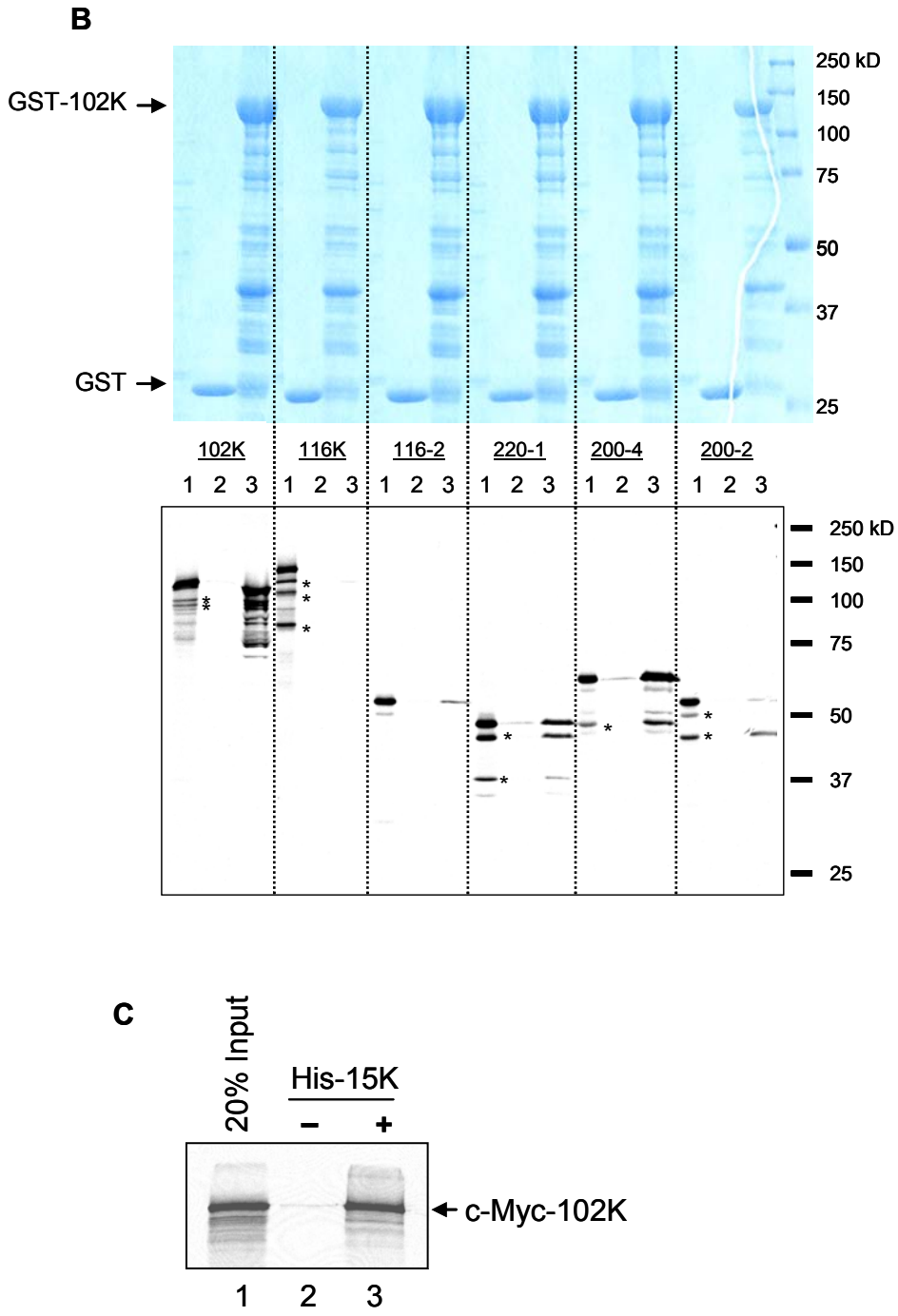


Figure 4.9. Interactions of protein 102K with other U5-specific proteins. (A) Y2H interaction assay of 102K. Protein 102K was cloned into pGBKT7 vector as a bait fusion, while the testing partners of U5-specific proteins as indicated in the figure were cloned into pGADT7 as prey fusions. The positive results were selected on the SD/-Leu-Trp-His-Ade medium at 30°C for 6 days. (B) *In vitro* interactions of GST-102K. 10 µl of [^{35}S]-labelled samples of interest were incubated with GST, or GST-102K immobilized on glutathione-Sepharose beads. Bound proteins were eluted in SDS sample buffer and analyzed with SDS-PAGE. The input lane (lane 1) contains 10% of the total amount of *in vitro* translated protein added to each reaction. The gel was first subjected to Coomassie blue staining to show that equivalent amount of expressed proteins were eluted from beads (upper panel). Subsequently, the gel was dried and

exposed to X-ray film (lower panel). Arrowheads indicate the position of GST or GST-102K protein. Asterisks (*) indicate the degraded form of the translated proteins. The protein size markers are shown on the right of the panel. (C) The interaction of His-15K with 102K protein prepared by translation *in vitro*. Purified His-tagged 15K protein was first incubated with anti-pentahistidine antibodies and protein A-Sepharose, and subsequently with [^{35}S]-labelled c-Myc-102K protein prepared *in vitro* (lane 3). Bound proteins were again visualized by autoradiography following SDS-PAGE. A mock precipitation without His-tagged 15K is included as a negative control (lane 2). The input lane represents 20% of the c-Myc-102K products added to the reaction (lane 1).

4.4 Protein-protein interactions between U4/U6 and U5 snRNP particles

4.4.1 U5-102K binds specifically to U4/U6-61K in the Y2H system and *in vitro*

While some information has been accumulated on protein-protein and protein-RNA interactions within the individual U4/U6 or U5 snRNP, little is known about the interactions that connect the U4/U6 and U5 snRNPs. Several lines of evidence have suggested that protein-protein interaction may contribute to the formation of the tri-snRNP. Therefore, the interactions between the U4/U6 and U5 snRNP particles were tested in the yeast two-hybrid system. Figure 4.10A shows the results from the combinations of all the prey fusions of the U4/U6-specific proteins with bait fusions of the U5-specific proteins. The protein fragments from the three large proteins 116K, 200K and 220K were included in the experiment. Surprisingly, only one interaction – between pGADT7-61K and pGBKT7-102K – was identified. The same results were obtained in the reciprocal two-hybrid screen, where U4/U6-specific proteins were constructed as bait fusions but U5-specific proteins as prey fusions (data not shown). The interaction between the 61K and 102K proteins was independently confirmed using a biochemical method (figure 4.10B). In the GST pull-down assay *in vitro* translated, [^{35}S]-labelled 102K protein binds strongly to a purified recombinant GST-tagged 61K fusion protein. The expression and purification of GST-61K fusion protein was described previously (Makarova *et al*., 2002; Nottrott *et al*., 2002). This interaction is specific, as only a low level of 102K protein was precipitated with beads containing GST alone. In summary, these results indicated that the binding between 102K and 61K is specific and stable which

may contribute to the formation of tri-snRNP from the U4/U6 and U5 snRNPs. This result was published recently in the EMBO Journal (Makarova *et al*., 2002).

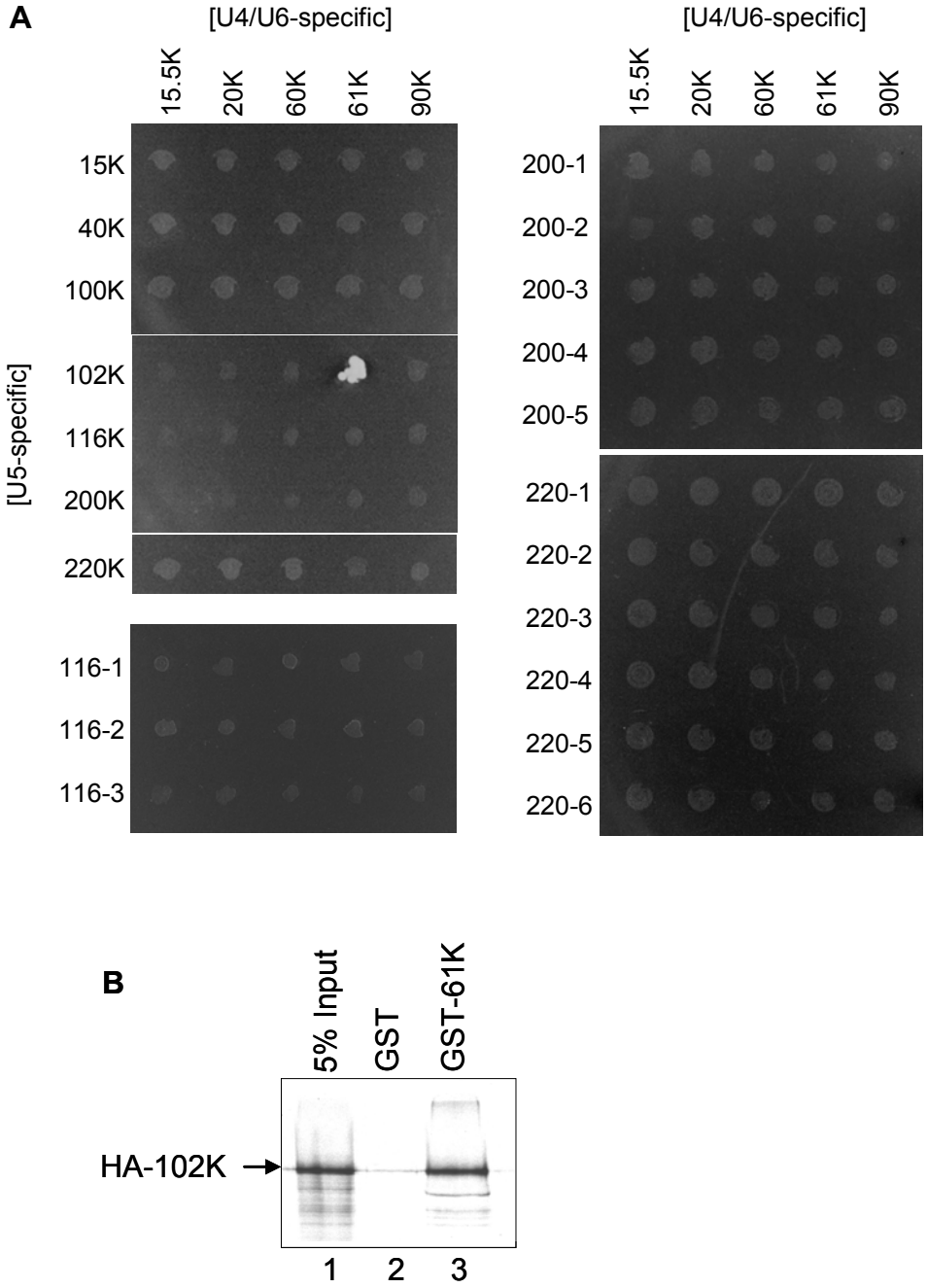


Figure 4.10. Protein 102K interacts specifically with 61K. (A) Y2H interaction between U4/U6 and U5 snRNP particles. All the U4/U6-specific proteins were cloned into pGADT7 as prey fusions, while the U5-specific proteins were cloned into pGBKT7 as bait fusions. The positive colony in the combination of pGADT7-61K and pGBKT7-102K was observed after growth on the selective medium of SD/-Leu-Trp-His-Ade at 30°C for 5 days. (B) *In vitro* interaction between protein 102K and 61K. [^{35}S]-labelled, *in vitro* translated protein 102K was incubated with GST (lane 2) or GST-61K fusion protein(lane 3) in the presence of glutathione–Sepharose. The precipitated proteins were

fractionated by SDS-PAGE and visualized by fluorography. The input lane represents 5% of the 102K amount added to the reaction.

4.4.2 The U5-102K binds specifically to the C-terminal region of U4/U6-90K in *in vitro* binding assay

As pointed out in section 4.2.1, an interaction between the U4/U6-specific proteins 60K and 90K could not be detected using the two-hybrid-system, but was clearly detectable in the *in vitro* binding assay. This suggested that other interactions of both proteins may have eluded detection in the Y2H system. The GST-102K fusion protein was therefore used in binding assays with 60K and 90K prepared by translation *in vitro*, and with 61K as a positive control (its interaction with 102K was described above; Makarova *et al*., 2002). While 90K bound very strongly to GST-102K, 60K did not (figure 4.11A).

To delimit further the boundaries of the interacting domain in the 90K, the protein was dissected into four fragments according to the design by Gonzalez-Santos *et al*. (2002), with two additional fragments respectively comprising the N- and C-terminal halves of the protein (see figure 4.21 for diagram). All six fragments were subcloned in pGBKT7, and subsequently were used for yeast two-hybrid testing with 102K, or used as templates to generate [^{35}S]-labelled c-Myc-tagged proteins for GST pull-down experiments respectively (see section 3.2.3.2 for methods). The full-length 90K protein was tested as well. Consistently, no binding was identified in the Y2H assay (figure 4.11B, upper panel). In contrast, the *in vitro* binding assay shows that the C-terminal fragments (C267, C381, and C489) strongly bind GST-102K, while the N-terminal fragments do not result in signals above background (figure 4.11B, lower panel). It therefore appears that a C-terminal region of the 90K, comprising amino acids 417–683, suffices to establish binding to 102K.

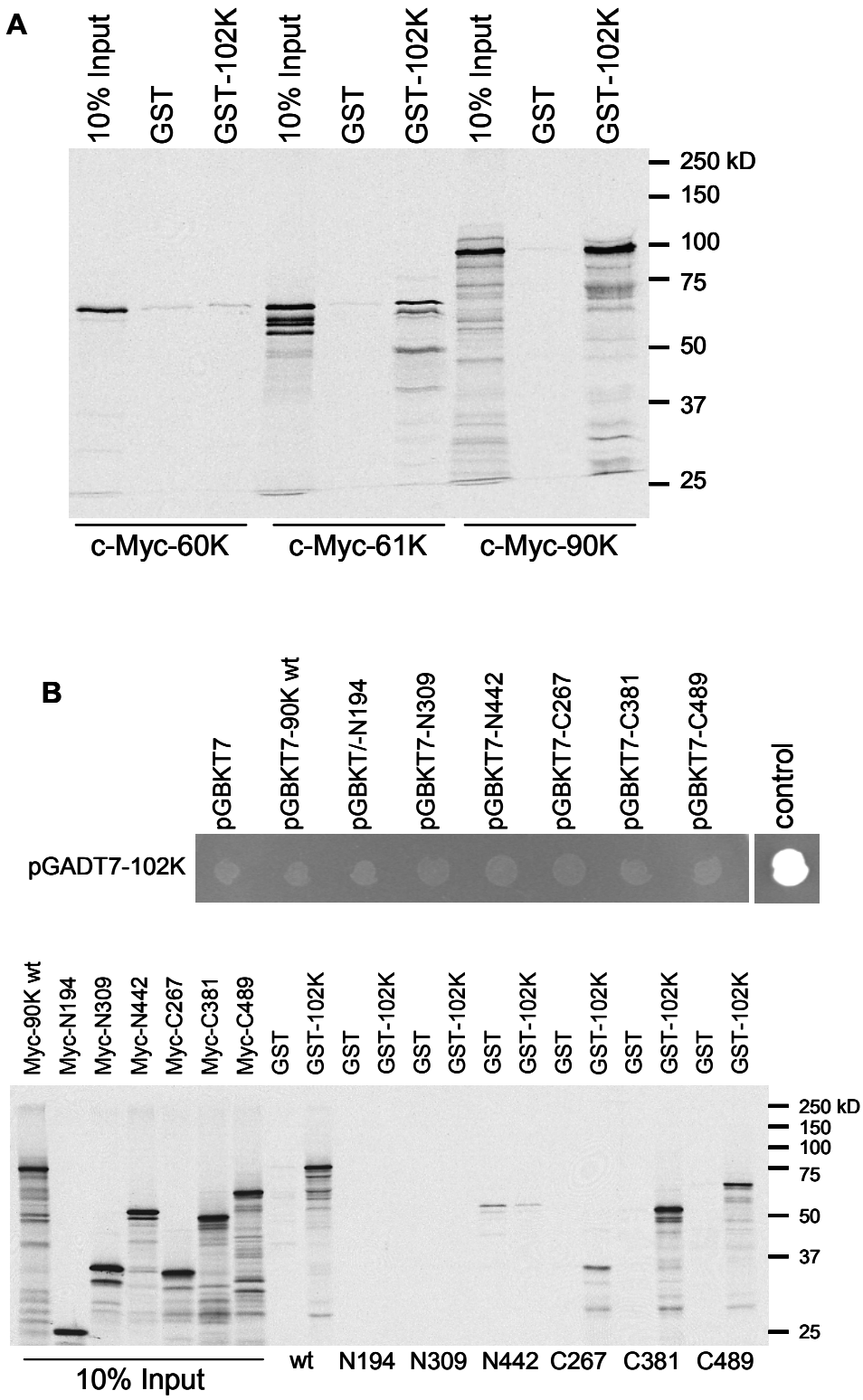


Figure 4.11. (A) *In vitro* interaction between proteins 102K and 90K. 10 µl of [^{35}S]-labelled samples of interest were incubated with GST, or GST-102K immobilized on glutathione–Sepharose beads. Bound proteins were eluted in SDS sample buffer and analyzed with SDS-PAGE and visualized by autoradiography. The input lane contains 10% of the total amount of *in vitro* translated protein added to each reaction. The protein size markers are shown on the right of the panel. (B) Upper panel: Y2H analysis to search for the 102K binding domain in 90K. pGADT7-102K was tested against six

90K fragments constructed in pGBKT7; the combination of pGADT7-T/pGBKT7-p53 served as a positive control. The positive interaction was observed by growing on the selective medium of SD/-Leu-Trp-His at 30°C for 3 days. Lower panel: Co-immunoprecipitation assay to search for the 102K binding domain in 90K. The same procedure as described in the legend for (A) was used.

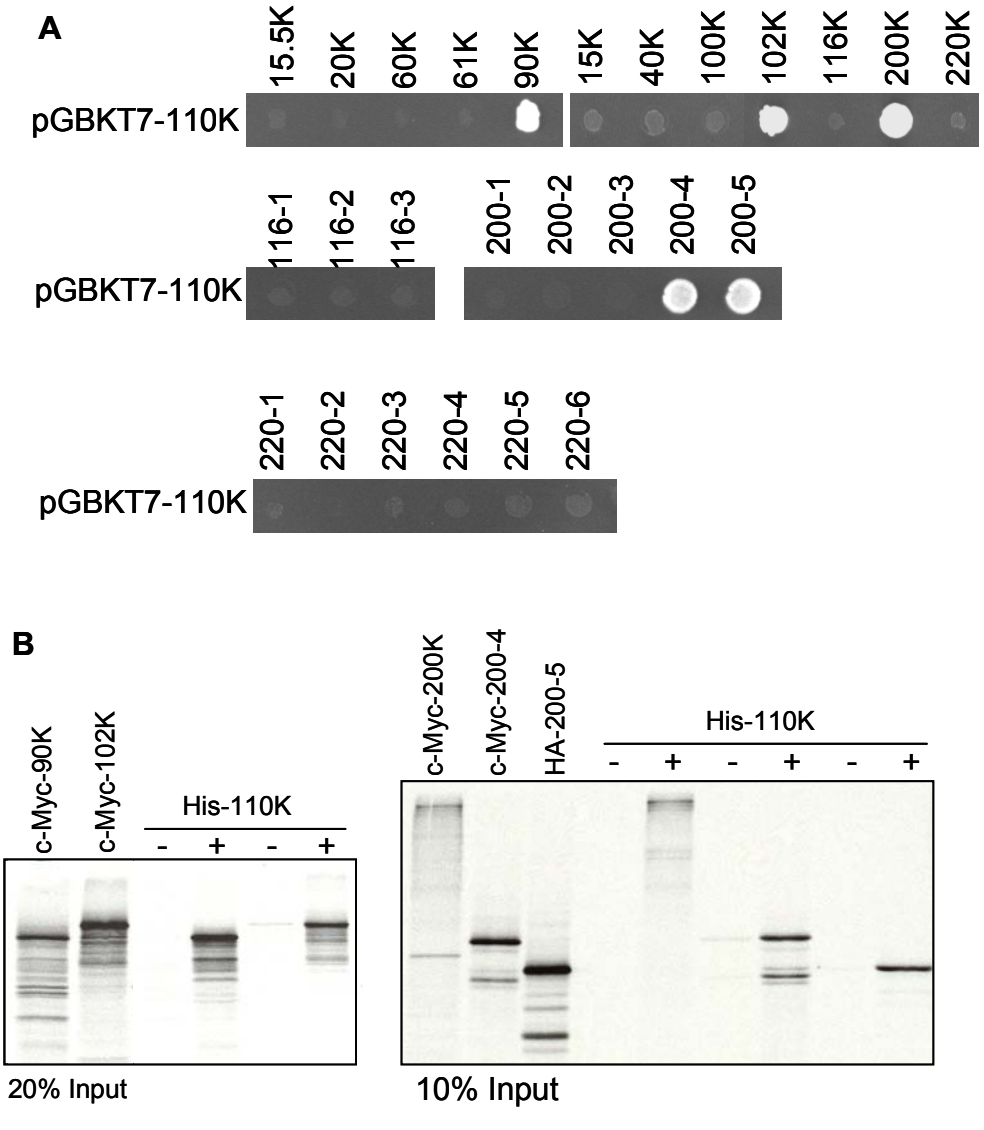
4.5 The interactions of the tri-snRNP-specific protein 110K

Three proteins, the 110K, 65K and 27K proteins, show specific stable association with the tri-snRNP (Makarova *et al*., 2001; Fetzer *et al*., 1997), but less so or not at all with its separate snRNP components. I was interested in identifying the binding partners of these three proteins. In Y2H experiments, full-length 110K was found to bind to the U4/U6-specific protein 90K and also to the full-length U5-specific proteins 102K and 200K (figure 4.12A). In the case of protein 200K, the interaction with 110K appears to depend on the C-terminal fragments 200-4 (aa 1301–1816) and 200-5 (aa 1705–2136) (figure 4.12A). Interactions were again confirmed by co-immunoprecipitations: in the presence of His-tagged 110K recombinant protein, the anti-pep-110 antibody co-precipitated [^{35}S]-labelled full-length c-Myc-90K, c-Myc-102K and c-Myc-200K protein as well as the c-Myc-200-4 and HA-200-5 fragments of 200K (figure 4.12B). Yeast two-hybrid interactions involving the U4.U6/U5-27K or the 65K protein could not be detected (data not shown). The expression and purification of His-tagged 110K recombinant protein has been described by Makarova *et al*. (2001).

The interacting domain of 90K was further investigated in the Y2H system. Like its full-length 90K protein, the two larger C-terminal fragments (C381 and C489) interact with 110K, while the three N-terminal fragments and the short C-terminal fragment do not (figure 4.12C). In summary, for the binding of protein 110K, a larger C-terminal region of 90K, comprising amino acids 303–683, is required, comparing to the binding for protein 102K (see section 4.4.2) and p110 (section 4.8), which need only amino acids 417-683.

The human 110K protein comprises 800 amino acids residues and contains an RS domain in the N-terminal region of about 120 residues, which is absent in the yeast homologue protein Snu66p. The remaining part of the 110K sequence does not contain known motifs that would suggest additional potential functions

of this protein (Makarova *et a*l., 2001). To evaluate the domain function, the full-length 110K was split into two segments, RS and ΔRS. RS contains the RS domain (1-111 amino acids), while ΔRS contains the C-terminal region lacking RS domain (112-800 amino acids). Two segments were subcloned into pGBKT7 for testing interaction with known partners 90K, 102K and 200K C-terminal fragments of 200-4 and 200-5 in Y2H system. The results clearly showed that the RS domain of 110K is not required for the binding, while the C-terminal region lacking RS domain is sufficient for the binding with protein 90K, 102K and 200K fragments (figure 4.12D). Consistently with this, yeast homologue protein Snu66p, which lacks RS domain in nature, is able to bind to the Brr2p (200K in human) and Prp6p (102K in human) (van Nues and Beggs, 2001). The potential role of the RS domain in the human 110K protein is discussed in the Discussion.



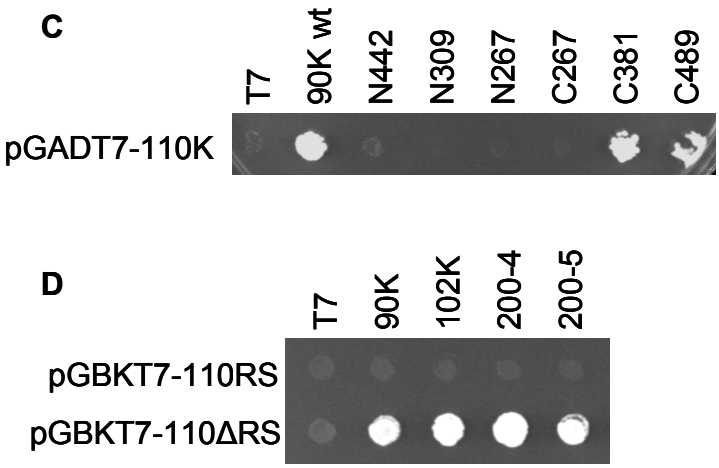


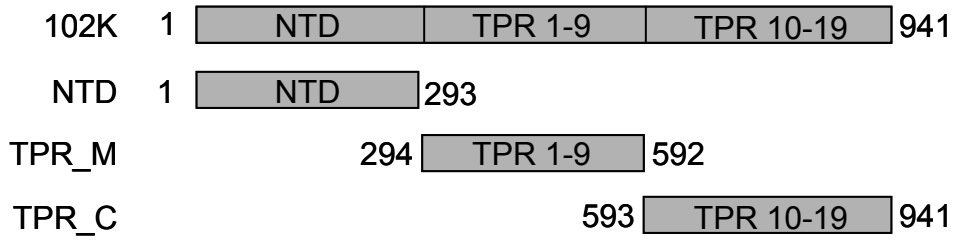
Figure 4.12. Analysis of the interactions of 110K protein. (A) Y2H investigation of the interaction between 110K and U5- and U4/U6-specific proteins. Protein 110K was cloned into pGBKT7 as a bait fusion, while all the U4/U6- and U5-specific proteins were cloned into pGADT7 as prey fusions. The positive interaction was observed by growing on the selective medium of SD/-Leu-Trp-His-Ade at 30°C for 3-6 days. (B) Analysis of interactions between 110K protein and *in vitro*-prepared fusion proteins of 90K and 102K (left panel), and the full length 200K protein as well as fragments 200-4, 200-5 (right panel). Purified His-tagged 110K was incubated with anti-pep-110 antibodies bound to protein A–Sepharose and [^{35}S]-labelled proteins produced by *in vitro* translation. The precipitated proteins were fractionated by SDS-PAGE and visualized by autoradiography. Mock precipitations without His-tagged 110K protein were included as a negative control. (C) Y2H analysis to search for the 110K binding domain in 90K. pGADT7-110K was tested against six 90K fragments constructed in pGBKT7; the empty vector and the full-length 90K served as controls. The positive interaction was observed by growing on the selective medium of SD/-Leu-Trp-His at 30°C for 3 days. (D) Analysis of protein 110K domain structure in the Y2H system. Protein 110K mutants were constructed in pGBKT7 and were tested for interaction with other proteins. The positive interaction was observed by growing on the selective medium of SD/-Leu-Trp-His-Ade at 30°C for 3 days.

4.6 The interaction domains of U5-102K protein

The 102K protein comprises an N-terminal domain followed by 19 repeats of TPR (tetratricopeptide) motif, as shown in Figure 4.13A (Makarov *et al*., 2000). Comparison with other TPR proteins revealed that the motifs in the protein 102K belong to a TPR protein subfamily which appears restricted to RNA processing proteins such as Prp39p, Prp42p, Rna14p, and Clf1p (Chung *et al*., 1999; McLean and Rymond, 1998). The TPR repeats in this subfamily were also called HAT (half a TPR) repeats (Preker and Keller, 1998) or cl-TPR (crooked neck-like TPR) repeats (Chung *et al*., 1999). The N-terminal domain (NTD) is rich in charged amino acids and conserved in the homologues of 102K

protein. However, this region contains no known sequence motifs. To obtain a more detailed view of which particular regions interact with the various proteins, protein 102K was dissected into three fragments (see figure 4.13A). NTD (amino acids 1 to 293) lacks the entire C-terminal TPR domain, while TPR_M (amino acids 294 to 592) and TPR_C (amino acids 593 to 941) contain first 9 TPR repeats and second 10 TPR repeats respectively. All protein-protein interactions of 102K, as identified by both Y2H and by *in vitro* GST pull-down experiments (see figures 4.9, 4.10, 4.12), were further examined with the 102K fragments used either as bait (figure 4.13B) or as prey (figure 4.13C). Figure 4.13D schematically summarises the interactions observed from figure 4.13B and C. The N-terminal domain was found to bind to 15K protein, while the central TPR_M fragment bound to 110K, the 200-4 fragment of 200K and the 220-1 fragment of 220K. The 61K protein interacts with both TPR_M and TPR_C fragments. In summary, it appears that distinct regions of 102K are dedicated to interactions with different proteins.

A



B

Gal4-BD hybrid	Gal4-AD hybrid: pGADT7	15K	61K	110K	200-4	220-1
pGBKT7						
NTD		ND	ND	ND	ND	ND
TPR_M						
TPR_C						

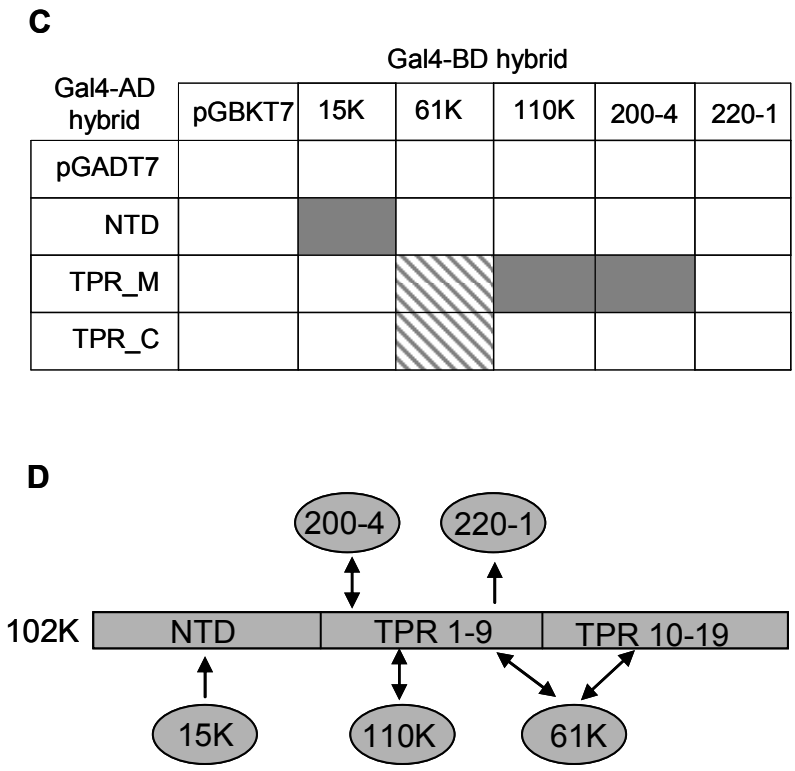


Figure 4.13. Y2H analysis of 102K domains involved in the protein-protein interactions. (A) Schematic representation of the protein fragments derived from full-length 102K used in two-hybrid assays. (B) Yeast strain AH109 was transformed with pGBKT7 carrying 102K fragments and pGADT7 carrying protein 15K, 61K, 110K and fragments 200-4 from 200K and 220-1 from 220K. Empty vectors were included as a negative control. Combinations labelled ND were not determined because of self-activation of the bait. Grey squares represent interactions selected on SD/Leu-Trp-His-Ade, hatched squares those selected on SD/Leu-Trp-His. White squares represent no interactions. (C) Interactions of inversed bait-prey pairs. (D) Schematic drawing of the observed bait→prey interactions involving 102K fragments.

4.7 Specific binding of the U5-52K protein to the U5 snRNP 102K and 15K proteins

The 52K protein is readily visible in immunoaffinity-purified 20–25S snRNPs (Behrens and Lührmann, 1991). The gene encoded for this protein was subsequently cloned in the laboratory of Prof. Lührmann. Database searches revealed that the sequence of the 52K protein is identical to that of a previously described protein, CD2BP2, which interacts with CD2, a surface marker of T cells, which facilitates activation of T lymphocytes on binding to its ligand CD58 (Laggerbauer *et al*., 2005; Nishizawa *et al.*, 1998). The protein 52K contains in the C-terminal region a GYF motif which has been shown to has the

ability to bind a proline-rich tandem repeat in the cytoplasmic tail of CD2 (Freund *et al*., 2002; Freund *et al*., 1999), and all of the residues that play a direct part in this binding (Tyr285, Trp287, Pro298, Phe299, Met304, Trp307, Tyr312 and Phe313) are well conserved among the homologue proteins. In *S. cerevisiae* the Snu40p/Lin1p can be clearly identified as the homologue of 52K, because it shares the other characteristic sequence motifs and also associates with the yeast U5 snRNP (Stevens *et al*., 2001).

To investigate whether the 52K protein is associated with both 20S U5 snRNP and 25S tri-snRNP, HeLa nuclear extract was loaded onto a glycerol gradient and the distribution of the 52K protein was analysed by SDS-PAGE followed by western blotting. For determination of the sedimentation coefficients, RNA was isolated from aliquots of each fraction, separated by PAGE and visualised by silver staining. The results showed that almost all of the 52K protein is present in a 20S complex, presumably the U5 snRNP, and the 52K protein appears to be excluded from the tri-snRNP. To test the latter conclusion, the 25S tri-snRNP were purified from a mixture of spliceosomal snRNPs by immunoaffinity chromatography and fractionated by glycerol gradient centrifugation; the protein and RNA content of each fraction was analysed by PAGE. The 52K protein was also absent in this snRNP particle. These results demonstrate clearly that the 52K protein is not a component of the 25S U4/U6.U5 tri-snRNP, and indicates that it dissociates from the U5 snRNP during the formation of the tri-snRNP particle (Laggerbauer *et al*., 2005).

Since the 52K protein is associated with the U5 snRNP in such a specific manner, it is of particular interest to examine how the protein is integrated into this particle. In this study, yeast two-hybrid analysis was first used to study the interactions of the 52K protein with other proteins. As bait, I tested all the U5 snRNP-specific proteins (proteins 15K, 40K, 100K, 102K, 116K, 200K and 220K), the tri-snRNP-specific proteins (27K, 65K and 110K), and the U4/U6-snRNP-specific proteins (15.5K, 20K, 60K, 61K and 90K). Of all these proteins, only the U5 snRNP proteins 15K and 102K were found to bind to the 52K protein (figure 4.14A, left column, and data not shown). In contrast, neither of the tri-snRNP or U4/U6 snRNP proteins interacted with the 52K protein (data

not shown). The 52K protein contains a known protein-protein interaction motif, namely the GYF domain in the protein's C-terminal 86 residues (Freund *et al*., 1999; Freund *et al*., 2002). Therefore, I next asked whether this C-terminal fragment is necessary, or possibly even sufficient, for the interaction with protein 15K or 102K. For this purpose, the full-length 52K was split into two fragments: 52N (amino acids 1–255) lacks the GYF motif, while 52C (amino acids 256–341) contains only this domain. As shown in figure 4.14A (right column), protein 15K binds to 52C, but not with 52N. In contrast, protein 102K binds to 52N, but not 52C. The negative controls with empty prey (pGADT7) or bait (pGBKT7) vector showed no cell growth on the selective medium, confirming that none of the constructs used here induced the reporter gene on its own.

To confirm the interactions observed in the two-hybrid system, I performed pull-down assays (see sections 3.2.3.2 and 3.2.3.4 for methods). As shown in figure 4.14B, the GST-102K fusion protein, but not GST alone, efficiently precipitated the radio-labelled 52K protein from reaction mixtures in which it had been generated by translation *in vitro*. This assay also confirms that protein 102K interacts with the N-terminal two-thirds of the 52K protein, and not with the GYF domain (figure 4.14B, lower panels). Similarly, I used the His-tagged 52K protein to precipitate protein 15K prepared by translation *in vitro*. As expected, a significant amount of protein 15K was precipitated in the presence of the His-tagged 52K protein, but not on the beads alone (figure 4.14C, lanes 2 and 3). As predicted form the two-hybrid results, the C-terminal GYF domain also precipitates protein 15K, and the efficiency of this precipitation is even higher than for the full-length protein (lane 4).

In conclusion, the data from both the yeast two-hybrid system and the binding assays *in vitro* demonstrate clearly that the 52K protein associates with the U5 snRNP particle by binding to proteins 102K and 15K. Interestingly, two different domains of the 52K protein are involved in these interactions, suggesting that they can occur simultaneously.

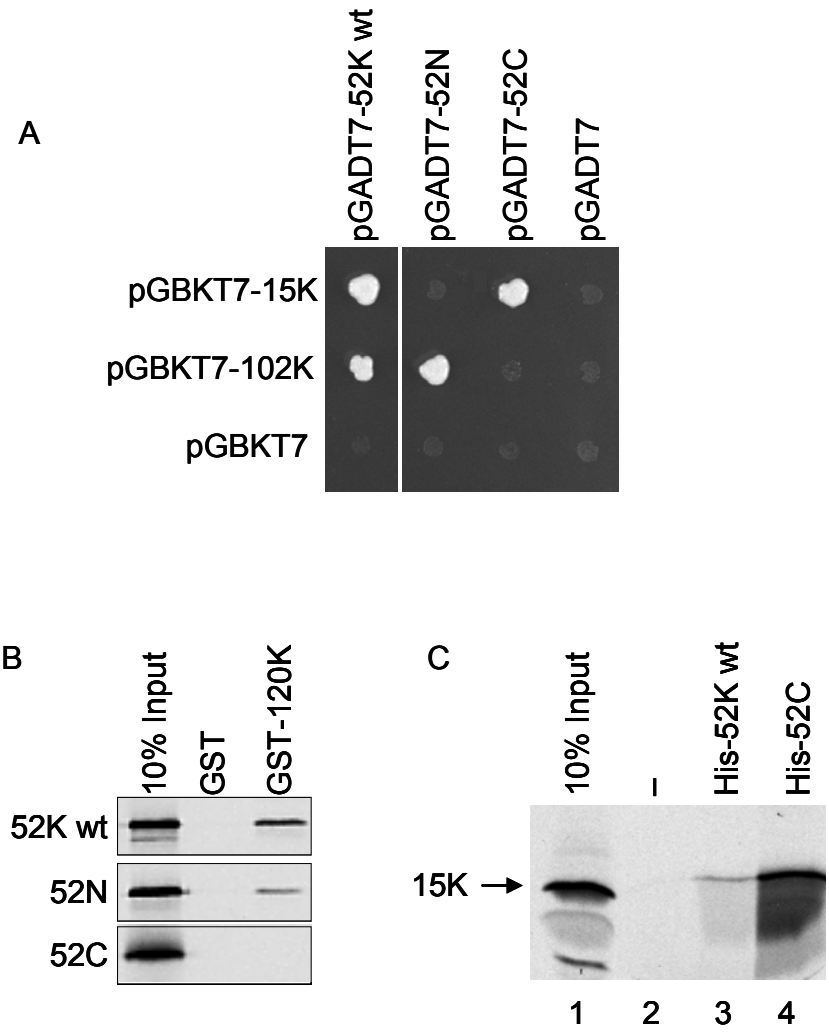


Figure 4.14. Association of protein 52K with U5 snRNP through the U5-specific proteins 15K and 102K. **(**A) Two-hybrid analysis. Plasmid pairs were co-transformed into yeast host AH109 as indicated in the figure, and the interactions of the full-length (left panel) or truncated 52K protein (right panel) with 15K and 102K proteins were detected by growth on selective medium SD/-Leu-Trp-His-Ade at 30 °C for three days. The empty vector pGADT7 or pGBKT7 served as a negative control. (B) GST pull-down assay. *In vitro* [^{35}S]-labelled, c-Myc-tagged 52K constructs were incubated with 102K protein fused to a GST tag and precipitated by glutathione-Sepharose. 10% of the translation products were loaded as input. (C) His pull-down assay. *In vitro* [^{35}S]-labelled HA-tagged 15K protein was incubated with His-tagged fusion protein 52K wild type or C-terminal fragment and precipitated with Talon beads. 10% of the translation products were loaded as input.

4.7.1 Overexpression of U5-52K proteins and X-ray crystallography study of 52K GYF domain in complex with U5-15K

As described in the Materials and Methods, protein 52K wild-type and two deletion mutations (52N and 52C) were subcloned in pET28a expressing vector. The plasmids were transformed into BL21(DE3) and the cell culture was

induced in 1 mM IPTG at 37°C for 3 hours. The expressed proteins were bound to Ni-NTA beads. After washing, the proteins were eluted in the elution buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 100 mM Imidazole). Figure 4.15 shows the first and second elution fractions of the proteins. Protein 52K wt (lane 1, 4) and 52C fragment (lane 3, 6) express high amount of proteins and are soluble, whereas the 52N does not (lane 2, 5).

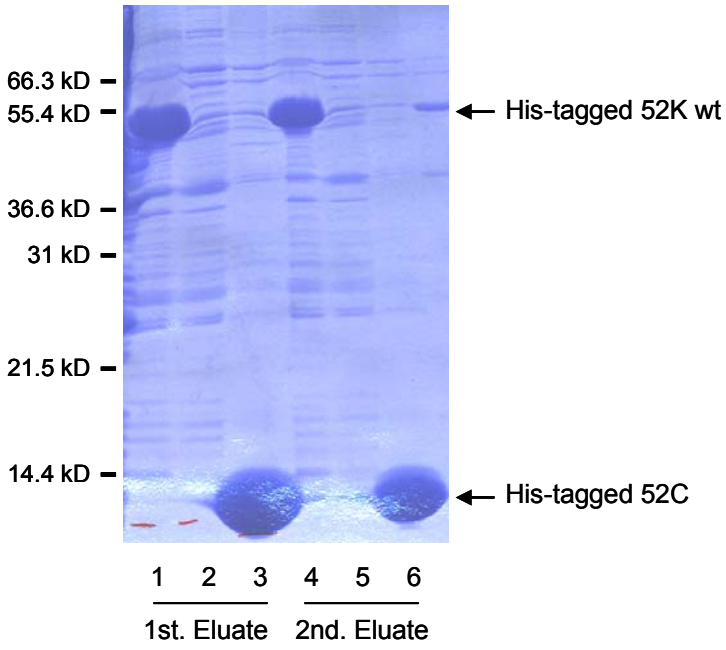


Figure 4.15. Protein 52K wt and deletion mutants were expressed in *E. coli*. The soluble fraction of each protein was run on a 15% SDS-PAGE gel and stained with Coomassie blue.

A crystallography study of the 52K GYF domain in the complex with U5-15K in the laboratory of Prof. R. Ficner (Nielsen *et al.*, in preparation) shows that the distinct residues of GYF domain are used for the binding of 15K comparing to the binding of CD2 protein. In the functional complex the GYF-domain and the U5-15K protein are bound mainly by polar interactions, including a salt bridge between Glu111, OE1 (U5-15K) and Gln328, NE2 (U5-52K), hydrogen bonds between side chains of Asp114, OD2, Glu117, OE2 and Ser132, OG (U5-15K) and side chains of Tyr330, OH, Lys321, NZ and Arg334, NH2 (U5-52K) and between main chains Ser132, O (U5-15K) and side chains Arg334, NE (U5-52K). Finally there is a hydrogen bond between the side chain of Lys125, NZ (U5-15K) and the C-terminal oxygen of Thr341 (U5-52K). CD1 and CD2 from

residue Leu339 (U5-52K) make hydrophobic contacts with Val130, CG2 and Gly122, CA (U5-15K).

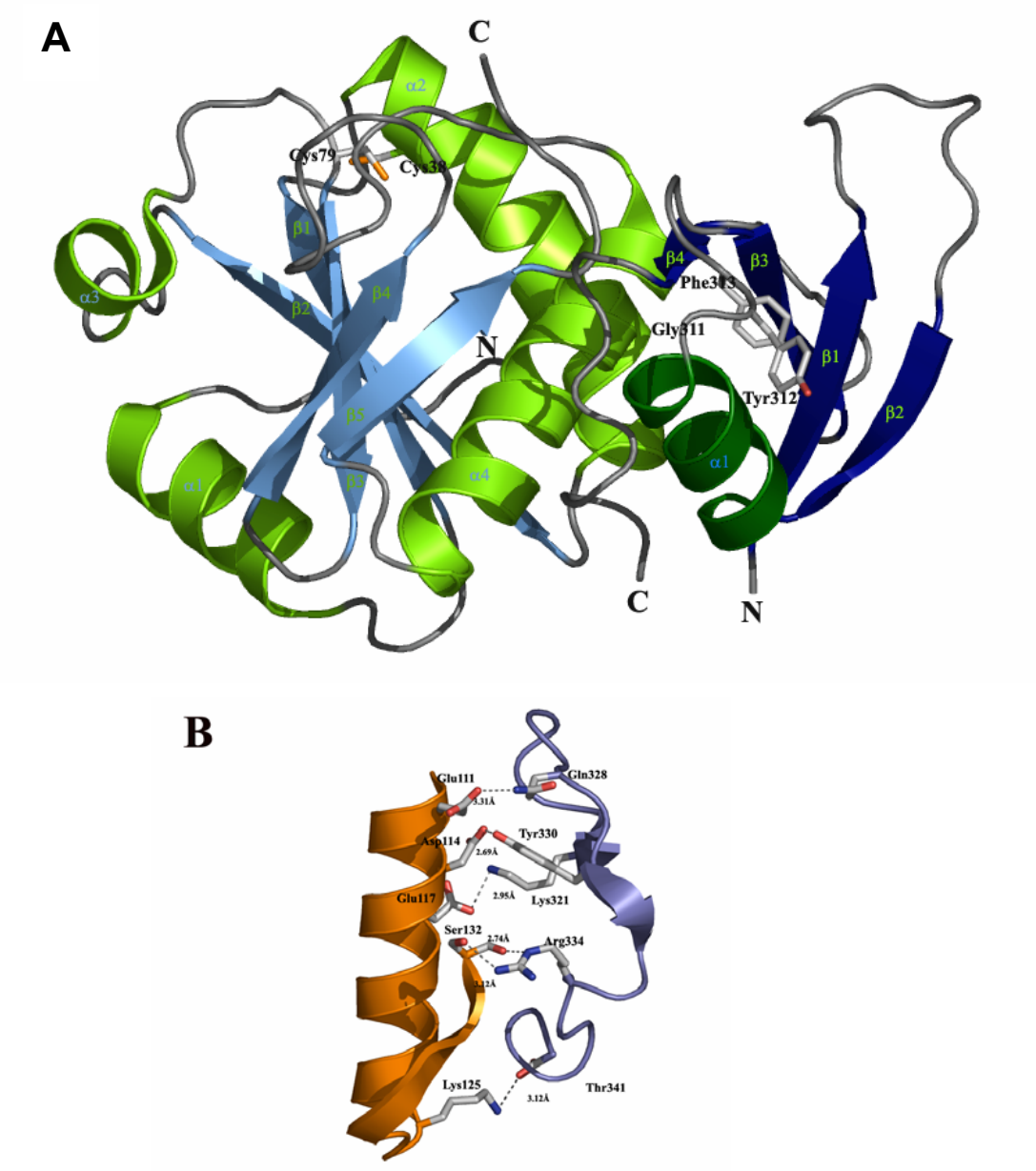


Figure 4.16. (A) Cartoon representation of the U5-15K:U5-52K complex. U5-15K is shown on the left in the lighter colours. Cysteine 38 and cysteine 79 are shown in ball-and-stick representation. The GYF domain of U5-52K is shown on the right in darker blue and green. Gly311, Tyr312 are Phe312 displayed in ball-and-stick representation. (B) A stereo picture of the polar interactions and the distances between the interacting atoms. U5-15K is shown in orange and U5-52K in blue. The contacts between the two molecules were determined with the program CONTACT from the CCP4 program package. The interactions are found in all three molecules. The distances are from molecules A+B, but the distances in C+D and E+F are almost the same. This figure is provided by Tine Kragh Nielsen in the laboratory of Prof. R. Ficner (unpublished data).

4.8 The interaction of the U4/U6-90K with U6-p110

The recycling factor p110 is associated with mammalian U6 and U4/U6 snRNPs, but not with U4/U6.U5 tri-snRNP. Recombinant p110 binds *in vitro* specifically to human U6 snRNA (Bell *et al*., 2002). To investigate the potential role of U4/U6-specific proteins in this respect, yeast two-hybrid assay was performed to detect the association between p110 and U4/U6 snRNP-specific 15.5K, 61K, 20K, 60K, and 90K proteins. Of these proteins, only the 90K protein showed an interaction with the full-length p110 protein (figure 4.17, left panel). The specific association of p110 and the U4/U6-90K protein was confirmed by a co-immunoprecipitation assay (see section 3.2.3.5 for methods). After incubation with recombinant His-tagged p110 protein, an *in vitro*-translated 90K protein could be efficiently co-immunoprecipitated with anti-p110 antibodies (figure 4.17, right panel). The recombinant N-terminal His-tagged p110 (amino acids 2 to 963) which has been expressed and purified from baculovirus-transfected SF21 cells was obtained from laboratory of Prof. A. Bindereif (Bell *et al*., 2002).

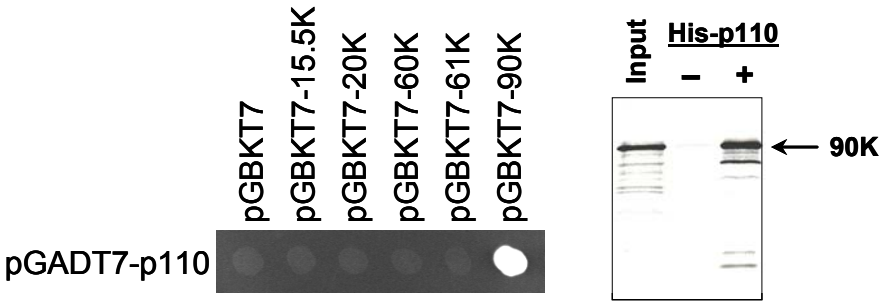


Figure 4.17. Y2H assay of interactions between p110 and U4/U6-specific proteins (left panel). p110 was constructed as a prey fusion, while U4/U6-specific proteins as bait fusions. Protein-protein interactions were assayed by selection on SD/-Leu-Trp-His-Ade medium at 30°C for 3 days. Analysis of the p110–90K protein interaction by *in vitro* binding assay (right panel). [^{35}S]-labelled c-Myc-tagged 90K was incubated with His-p110 protein, followed by precipitating with anti-p110 antibodies. Co-precipitated proteins were fractionated by SDS-PAGE and visualized by autoradiography. The input lane presents 10% of the 90K protein used in the reaction. The mobilities of the 90K protein is marked on the right.

The domain structure of the human U6-p110 protein consists of a large N-terminal region with at least seven tetratricopeptide repeat (TPR) motifs, two RNA recognition motifs (RRMs) in the C-terminal half of U6-p110, and a stretch of 10 highly conserved amino acids at the C terminus (Bell *et al*., 2002; Rader and Guthrie, 2002). A splice variant of the human protein has been reported as

a full-length cDNA (AF 387506) (Liu *et al*., 2002) which arises through the usage of a different 5' splice site in exon 7; this usage is expected to result in a truncated protein carrying only three of the seven TPR motifs. The p110 domain organization is conserved in many other eukaryotes, including *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, and *Drosophila melanogaster* (Bell *et al*., 2002; Rader and Guthrie, 2002); of the orthologues, only the well-known yeast Prp24 protein deviates from this domain organization, in that it lacks the entire N-terminal half with the TPR domain. Therefore, I was particularly interested in the possible role of the conserved TPR domain in U4/U6 snRNP formation and recycling.

To delineate the domain structure of p110, a series of p110 mutant derivatives were constructed for yeast two-hybrid assay (figure 4.18A, upper panel). p110 ΔN (amino acids 537 to 963) lacks the entire N-terminal TPR domain but contains the putative nuclear localization signal region, the two RRMs, and the C-terminal 10 amino acids. p110 ΔC (amino acids 2 to 688) carries the entire TPR domain, including the putative nuclear localization signal region, but not the two RRMs and the C-terminal 10 amino acids. A natural splice variant, AF 387506 (see above), is truncated within the fourth TPR motif and contains amino acids 2 to 350, followed by 14 additional amino acids, RSTTESKGFGFICT. The p110–90K protein interaction depended on the 90K protein component and was specific, since p110 ΔN gave a negative result; in contrast, both p110ΔC and the natural splice variant AF 387506 gave a positive result for the 90K protein interaction (figure 4.18A, lower panel).

Mutational analysis of the 90K protein showed in two-hybrid assays that the N-terminal half of the 90K protein (N442, N309 and N194) did not interact with p110. In contrast, C-terminal fragments of the 90K protein (C267, C381 and C489) were positive for the p110 interaction. When the p110 mutant derivative ΔC was used instead of full-length p110, identical results were obtained, consistent with the TPR domain mediating the interaction between p110 and the 90K protein (figure 4.18B). In summary, the results showed that U6-p110 interacts through its TPR domain with a C-terminal region (amino acids 417 to 683) of the U4/U6-90K protein.

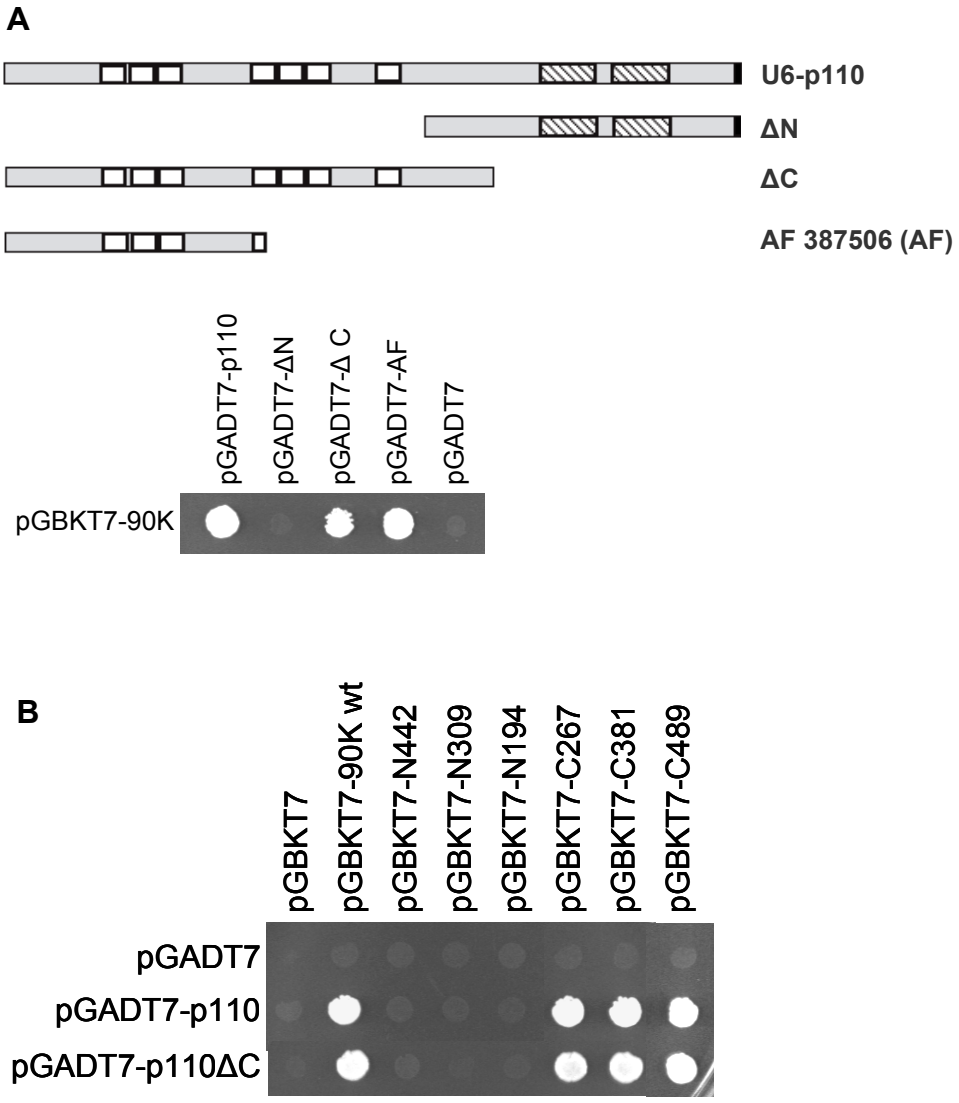


Figure 4.18. U6-p110 interacts through its TPR domain with a C-terminal region (amino acids 417 to 683) of the U4/U6 snRNP-specific 90K protein. (A) Upper panel: The p110 domain structure is schematically represented (RRMs, striped boxes; HAT or TPR domain, white boxes; conserved C-terminal region, black boxes). Lower panel: Y2H assay of interactions between p110 mutant derivatives and the U4/U6-specific 90K protein. Yeast strain AH109 was transformed with plasmid pGADT7-p110, pGADT7-ΔN, pGADT7-ΔC, pGADT7-AF 387506 (AF) or pGADT7 (as a control) in combination with plasmid pGBKT7-90K. Selection was done for 3 days at 30°C on SD/-Leu-Trp-His-Ade medium. (B) Y2H assay of interactions between p110 and U4/U6-specific 90K protein mutant derivatives. In the yeast strain AH109, p110 and its ΔC mutant were expressed as prey fusions, while 90K and its mutant derivatives were expressed as bait fusions. Interactions were observed on SD/-Leu-Trp-His-Ade medium at 30°C for 3 days.

4.9 The N-terminal region of the U4/U6-90K is required for binding of U2-associated protein SPF30/SMNrp

Recent studies suggested that SPF30/SMNrp is an essential protein-splicing factor whose function is required for the recruitment of the tri-snRNP to the A complex during spliceosome assembly (Meister *et al*., 2001; Rappsilber *et al*., 2001). The data from an *in vitro* binding assay indicated that SPF30/SMNrp interacts with the U4/U6.U5 tri-snRNP by a direct interaction with U4/U6-specific protein 90K (Meister *et al*., 2001). Here the *in vivo* Y2H system was used to investigate this interaction and the domain of 90K involved. For this purpose, the cDNA of SPF30/SMNrp was first cloned from a human Marathon-Ready cDNA library (Clontech) as described in Materials and Methods. The cloned full-length cDNA sequence which composes of 717 nucleotides is identical to the published data base from Neubauer *et al*. (1998, accession no. gi|3746841). NCBI protein-protein blast could find the potential orthologues in the organisms such as *Pongo pygmaeus, Pan troglodytes, Xenopus tropicalis, Drosophila melanogaster, Arabidopsis thaliana, Oryza sativa japonica, Gallus gallus, Macaca fascicularis and Schizosaccharomyces pombe*. However, no homologue could be detected in *Saccharomyces cerevisiae*. SPF30, also termed SMNrp, is the first and so far only reported homologue of SMN1, the gene responsible for spinal muscular atrophy. Both proteins share a central Tudor domain. Note that the Tudor domain is not conserved in the homologous protein of *Schizosaccharomyces pombe*. For the Y2H assay, SPF30/SMNrp was subcloned into yeast two-hybrid vectors as described in the Materials and Methods, and was subsequently tested for interacting with U4/U6-90K and its six truncated fragments which have been described above. SPF30/SMNrp indeed interacts with 90K protein as a prey or bait fusion, and mutational analysis showed that the N-terminus of 1-442 amino acids (N442) is responsible for the binding (figure 4.19). The interactions are specific since the other five truncated proteins did not display any signals. Interestingly, the Y2H signal of N-terminus is much stronger than the one of the full-length protein. In conclusion, the *in vivo* two-hybrid analysis successfully defines the SPF30/SMNrp binding domain in 90K, and further confirms the interaction demonstrated previously by the *in vitro* binding study (Meister *et al*., 2001).

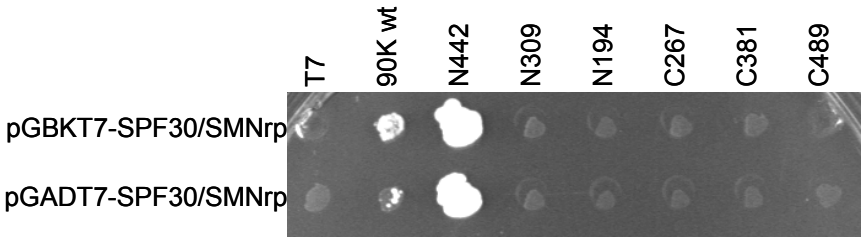


Figure 4.19. Y2H analysis of the interaction of U4/U6-90K with U2-SPF30/SMNrp. 90K wild-type (wt) and the truncated fragments as shown above the panel were incorporated into pGADT7 or pGBKT7 and co-transformed in combination with pGBKT7-SPF30/SMNrp or pGADT7-SPF30/SMNrp in turn. The positive colonies were selected on the SD/-Leu-Trp-His-Ade medium at 30°C for 7 days.

4.10 Analysis of point mutations related to the retinitis pigmentosa

Recent investigations first demonstrated that a number of mutations in the tri-snRNP proteins are related to the autosomal dominant retinitis pigmentosa (Vithana *et al.*, 2001; McKie *et al.*, 2001; Chakarova *et al.*, 2002). The identified genes and the related missense mutations were listed in the table 4.1. However, molecular mechanism for retinitis pigmentosa caused by mutations in these three tri-snRNP proteins is unclear at present. From the current interacting data, it is reasonable to ask whether the mutations have an effect on protein-protein interactions.

Table 4.1. Missense mutations related to retinitis pigmentosa

Protein	Gene	Missense mutation	Locus name	Publication
U4/U6-61K (hPrp31p)	*PRPF31*	A194E, A216P	RP11	Vithana *et al*., 2001
U4/U6-90K (hPrp9p)	*PRPF3 (HPRP3)*	P493S, T494M	RP18	Chakarova *et al*., 2002
U5-220K (hPrp8p)	*PRPF8 (PRPC8)*	P2301T, F2304L, H2309P, H2309R, R2310G, R2310K, F2314L	RP13	McKie *et al*., 2001

4.10.1 Analysis of U4/U6-61K mutations identified in retinitis pigmentosa

As described in section 4.4.1, protein 61K interacts specifically with protein 102K, and this interaction is critical for the formation of tri-snRNP (Makarova *et al.*, 2002; Schaffert *et al.*, 2004). Y2H analysis was used to investigate whether the mutations have an effect on such an interaction. Hence, the two mutant plasmids, pGBKT7-61K-A194E and pGBKT7-61K-A216P, were constructed as described in Materials and Methods. Subsequently they were co-transformed individually with pGADT7-102K in the yeast strain AH109 and were examined for the interaction. The results are shown in figure 4.20. While the clone bearing A216P shows a similar result to the wild type (compare spots 2 and 4 in middle and lower panels), the clone bearing A194E is hardly seen on the selective medium lacking histidine and adenine after 2 days (spot 3 in the middle panel) and has only grown slightly after 3 days at 30°C (spot 3, compare with spots 1 and 2 and 4 in the lower panel). As indicated on the growing medium of SD/-Leu-Trp, the upper panel shows that the cells are equally applied on the plates. In conclusion, the two-hybrid result suggested that one of the 61K mutants, namely A194E, significantly reduces the binding to 102K protein.

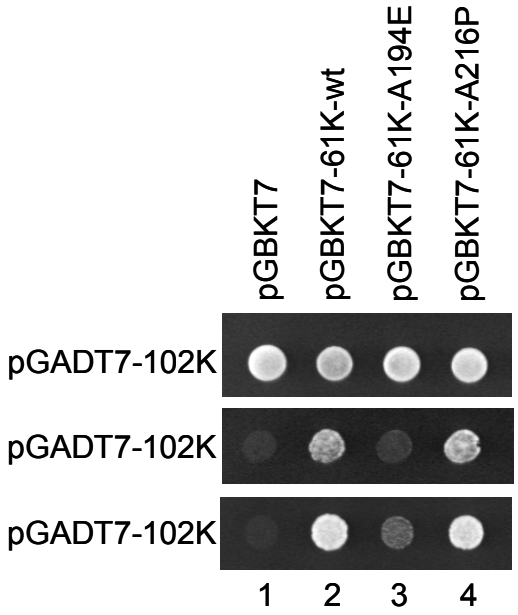


Figure 4.20. Mutational analysis of 61K interactions. Protein 102K was cloned into pGADT7 and cotransformed with pGBKT7 (as a control) and protein 61K wt and mutants which were cloned into pGBKT7, respectively. The co-transformants were replica plated on the growing medium SD/-Leu-Trp (upper panel), the selective medium SD/-Leu-Trp-His-Ade at 30°C for 2 days (middle panel) and for 3 days (lower panel).

4.10.2 Analysis of 90K mutations identified in retinitis pigmentosa

As described above, U4/U6-specific protein 90K interacts with U4/U6-60K (section 4.2.1), U5-102K (section 4.4.2), tri-snRNP-specific 110K (section 4.5), recycling factor U6-p110 (section 4.8) and U2-associated protein SPF30/SMNrp (section 4.9). Figure 4.21 summarises these interactions and the 90K domains involved. Obviously, the C-terminal domains containing the residues related to the two mutations are required for the binding of protein 102K, p110 and probably 110K, but are dispensable for the binding of SPF30/SMNrp. It is unlikely that the mutations are involved in the binding of 60K since the N-terminal region (N442) alone is sufficient for the binding.

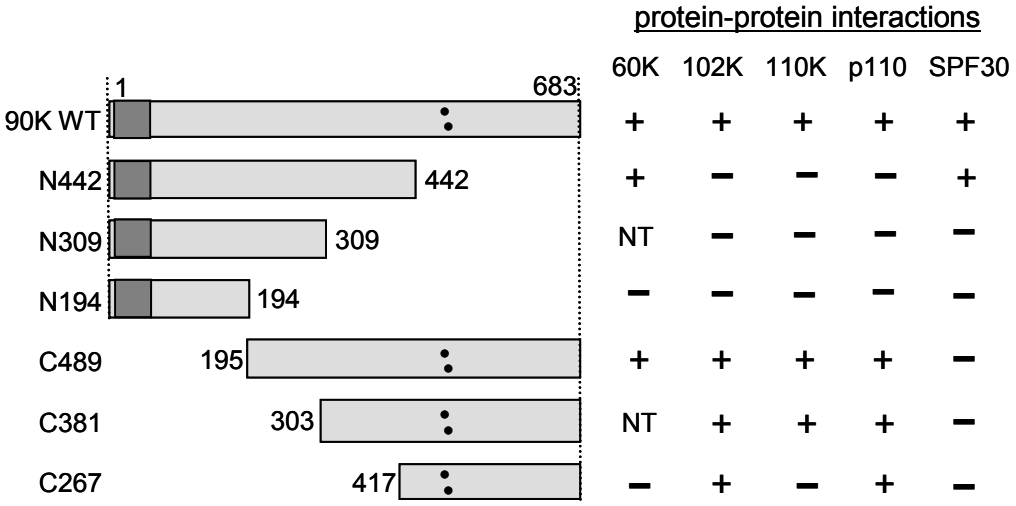
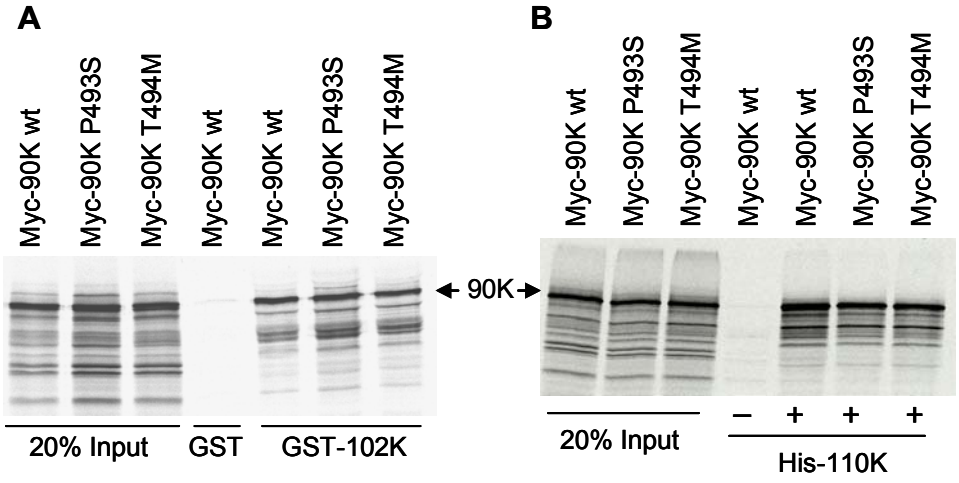


Figure 4.21. U4/U6-90K wild-type protein (683 amino acids [aa]) and mutant derivatives N442 (1-442 aa), N309 (1-309 aa), N194 (1-194 aa), C489 (195-683 aa), C381 (303-683 aa) and C267 (417-683 aa) are schematically represented. The dark grey boxes indicate the N-terminal PWI domain, and two black points show the positions of two point mutations (P493S and T494M) found in RP. The abilities of 90K wild type (WT) and mutant derivatives to interact with other proteins are shown on the right (+ or –). The 60K interaction data were published by Gonzalez-Santos *et al*. (2002), and the N309 and C381 were not tested (NT). The interactions of 90K with protein 110K, p110 and SPF30/SMNrp are observed in both the Y2H system and the *in vitro* binding assay, while the interactions with 60K and 102K are detected only in the *in vitro* binding assay.

Therefore, mutational analyses are focused on the interactions between the 90K and the 102K, p110 and 110K proteins. To do this, the two mutant clones, pGBKT7-90K-P493S and pGBKT7-90K-T494M, were generated from wild-type plasmid pGBKT7-90K using PCR as described in Materials and Methods. These two mutant constructs, along with the wild-type 90K protein, were used

for the Y2H assay and as templates for *in vitro* transcription/translation to produce [^{35}S]-labelled c-Myc-tagged proteins for the following *in vitro* binding assays. First, GST pull-down assay as described in section 3.2.3.2 was used for analysis of the interaction between 102K and 90K (figure 4.22A). GST-102K could precipitate both the recombinant [^{35}S]-labelled 90K wild-type protein and two mutant proteins (P493S or T494M). The binding affinities are not significantly different. Second, an immunoprecipitation experiment as described in section 3.2.3.5 was applied to analyse the interaction between 110K and 90K (figure 4.22B). In the presence of His-tagged 110K protein, anti-pep-110K antibodies could efficiently precipitate the [^{35}S]-labelled 90K wild-type protein and both mutant proteins (P493S or T494M) as well. Finally, the interaction between p110 and 90K was examined in both the Y2H and in the *in vitro* binding assay (figure 4.22C). In the two-hybrid assay, two mutant clones grow as well as the wild-type on the selective medium SD/-Leu-Trp-His-Ade (figure 4.22C, left panel). Similarly, immunoprecipitation (see section 3.2.3.5 for methods) shows that, in the presence of His-tagged p110 protein, anti-p110 antibodies could precipitate the [^{35}S]-labelled 90K wild-type protein and both mutant proteins (P493S or T494M) almost equally well (figure 4.22C, right panel). The specificity of the interaction was demonstrated by the fact that a 60K protein prepared by translation *in vitro*, another U4/U6-specific protein, was not co-immunoprecipitated. In conclusion, the *in vitro* binding assays, and also the Y2H method in the case of p110–90K, could not demonstrate any clear-cut effect of the two mutations P493S and T494M on the binding of protein 90K to either the 102K, the 110K or the p110 protein.



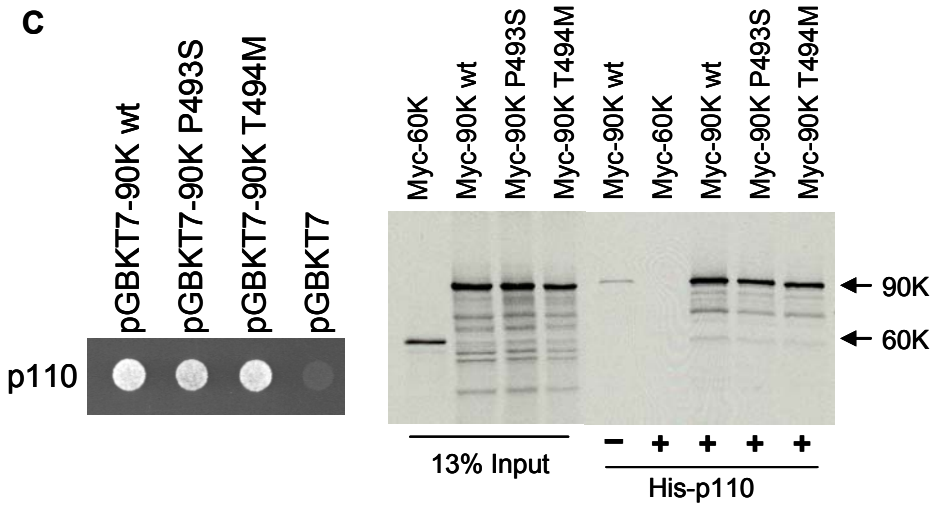


Figure 4.22. Mutational analysis of 90K interactions. (A) The interaction of GST-102K fusion protein and *in vitro* translated [^{35}S]-labelled c-Myc-tagged 90K mutant proteins; the wild-type 90K protein served as a control. 10 µl of [^{35}S]-labelled samples of interest were incubated with GST, or GST-102K immobilized on glutathione-Sepharose beads. Bound proteins were eluted in SDS sample buffer and analyzed with SDS-PAGE and visualized by autoradiography. The input lane contains 20% of the total amount of *in vitro* translated protein added to each reaction. (B) The interaction of His-110K fusion protein and *in vitro* translated [^{35}S]-labelled c-Myc-tagged 90K mutant proteins; the wild-type 90K protein served as a control. Purified His-tagged 110K was incubated with anti-pep-110 antibodies bound to protein A-Sepharose and [^{35}S]-labelled proteins. The precipitated proteins were fractionated by SDS-PAGE and visualized by autoradiography. The input lane contains 20% of the total amount of *in vitro* translated protein added to each reaction. (C) Left panel: Y2H assay of interactions between p110 and 90K mutant proteins. pGADT7-p110 was tested against mutant proteins pGBKT7-90K-P493S and pGBKT7-90K-T494M; the wild-type 90K protein and the empty vector served as controls. Interactions were observed on SD/-Leu-Trp-His-Ade medium at 30°C for 3 days. Right panel: The interaction of His-p110 fusion protein and *in vitro* translated [^{35}S]-labelled c-Myc-tagged 90K mutant proteins; the wild-type 90K protein and the U4/U6-60K protein served as controls. [^{35}S]-labelled 90K was incubated with His-p110 protein, followed by precipitating with anti-p110 antibodies. Co-precipitated proteins were fractionated by SDS-PAGE and visualized by autoradiography. The input lane presents 13% of the *in vitro* translated protein used in the reaction.

4.10.3 Analysis of 220K mutations found in retinitis pigmentosa

As shown in table 4.1, seven point mutations were found in the RP. Strikingly, all the mutations are related to the last 35 amino acids encoded by exon 42 of human 220K gene (*PRPF8*). Protein alignment at exon 42 between human PRP8 and its orthologues in other species was carried out previously by McKie *et al*. (2001) and again as shown below. Particularly striking is the level of homology around residues 2309 and 2310 and the absolute conservation of

these two residues throughout the species shown in the multiple alignment. Additionally, the proline residue at position 2301 has the same high degree of conservation. The remaining missense mutations alter the amino acid residues 2304 and 2314, which are also conserved, but to a lesser degree.

Figure 4.23. The amino acid sequence of human PRP8 exon 42 is aligned with that of *Drosophila melanogaster* (Dros), *Caenorhabditis elegans* (C.eleg), *Arabidopsis thaliana* (Arab), *Schizosaccharomyces pombe* (S.pombe), *Leishmania major* (Liesh) and *Trypanosoma bruceii* (Tryp) using the programme Clustal. The shaded boxes were created using the programme Mac-Boxshade with the default parameters set. The intensity of shading represents the relative degree of conservation and physico-chemical similarities. Hence the black boxes indicate amino acid identity or only highly conservative variation, while lesser shaded boxes show conservation but with more variation and with less conservative amino acid substitutions. The five amino acid residues mutated in the adRP cases are highlighted above with the codon number. Adapted from McKie *et al*., 2001.

As shown in the Y2H analysis in section 4.3.1, the 220K protein's C-terminal fragment 220-6 (amino acids 1986 to 2335) participates in the interaction with its own N-terminal fragment (220-1), fragment 4 of protein 200K (200-4) and the C-terminal fragment of protein 116K (116-3). This raises the question of whether the mutations affect these interactions. In an attempt to answer this question, the 220-6 fragment was first dissected into two pieces. ΔC (amino acids 1986 to 2300) lacks the last 35 amino acids related to seven point mutations; ΔN comprises the last 97 amino acids (2239 to 2335) containing the region related to the mutations. Figure 4.24A schematically represents the truncation strategy (also see figure 4.6A). ΔC and ΔN were subcloned into pGADT7 and were then tested for interaction with 116-3, 200-4 and 220-1. The plasmid pGADT7-220-6 was used as a positive control, and the blank pGADT7 vector as a negative control. Figure 4.24B shows that the ΔN, like 220-6, displays full binding activities in the three different combinations (compare spots 2 and 4 in the upper, middle and lower panels), whereas the ΔC completely lost

the binding abilities (compare spots 1 and 3). These results clearly suggest that the ΔN fragment is sufficient for the binding and imply that the mutations may have an effect on this. Therefore, the further experiments were done to check each of the seven point mutations. The PCR-based site directed mutagenesis method was used to generate the mutant constructs using pGADT7-ΔN plasmid as PCR template (see section 3.2.1.7 in the Materials and Methods). Subsequently, they were tested in the Y2H system. The strength of these interaction was estimated from comparing the growth of indicator yeast AH109 bearing wild-type and mutant alleles of 220-6 ΔN on medium lacking histidine and/or adenine, and containing different concentrations of 3-amino-1,2,4-triazole (3-AT). 3-AT inhibits the activity of the His3 reporter protein, and growth at higher concentrations of 3-AT indicates that more His3 is being synthesized and implies that the interaction is stronger. Typical results are shown in figure 4.24C. On the selective medium SD/-Leu-Trp-His-Ade, all mutant clones show similar growth patterns as the wild-type ΔN protein (upper panel). However, on the plates of SD/-Leu-Trp-His plus 3-AT, the clones bearing mutation R2310G (spot 7), R2310K (spot 8), or F2314L (spot 9) grow obviously slowly compared with the wild type (spot 2) in all three combinations; similar results were obtained for clones bearing P2301T (spot 3) and F2304L (spot 4) in the combination with 200-4, but only a slight – or no – difference was seen for the combination with 116-3 and 220-1. Instead, the clones bearing H2309P (spot 5) and H2309R (spot 6) appear to show faster growth than the wild-type clone (spot 2) does. To investigate the significance of this, a chemiluminescence ß-galactosidase assay was used. However, this method did not give consistent results when repeated (data not shown). In conclusion, the exon-42-encoded region of 220K protein related to all seven mutations identified in RP is indeed required for several protein-protein interactions. Most of the single point mutations affect the ability of this region to interact with its binding partners. However, the biological significances are still under investigation.

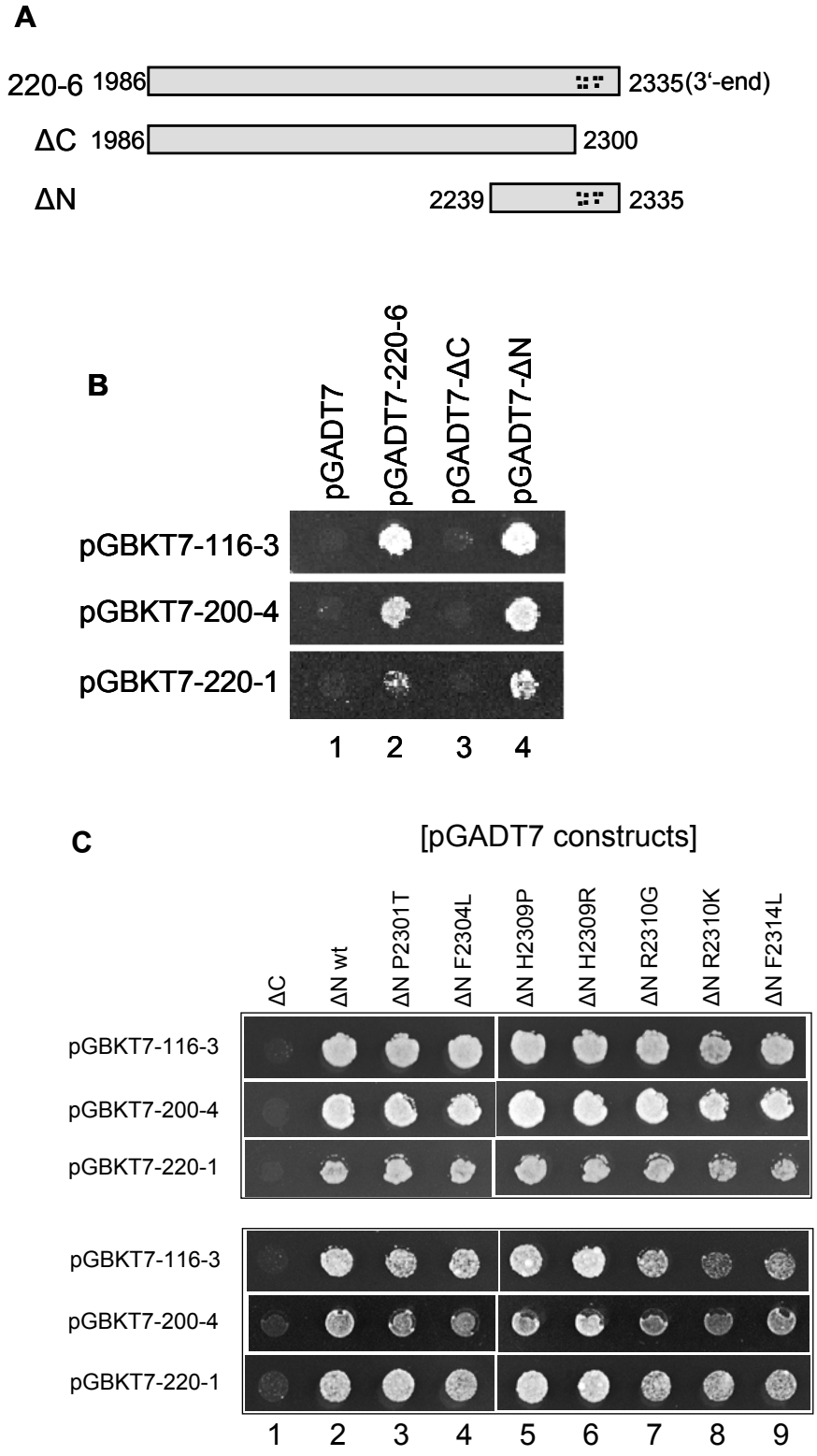


Figure 4.24. (A) Human U5-220K C-terminal fragment 220-6 (1986-2335 amino acids [aa]) and deletion mutations ΔC (1986-2300 aa), ΔN (2239-2335 aa) are schematically represented. The black points show the positions of the seven point mutations. (B) Y2H examination of the deletion mutation clones as indicated in the figure. Yeast cell growth was monitored on the selective medium of SD/-Leu-Trp-His-Ade for 3-4 days at 30°C. (C) Y2H examination of the point mutation clones as indicated above the panels. Yeast

cell growth was monitored on the selective medium of SD/-Leu-Trp-His-Ade for 3-4 days (upper panel), and on the medium of SD/-Leu-Trp-His+0.2 mM 3-AT for 3 days for the combinations with 116-3, SD/-leu-Trp-His+1.5 mM 3-AT for 3 days for the combinations with 200-4, and SD/-Leu-Trp-His+0.1 mM 3-AT for 6 days for the combinations with 220-1 (lower panel). All plates were incubated at 30°C and were observed from day to day. ΔC clone was included as a negative control.

4.11 Y2H screen with the N-terminal 434 amino acids of protein U5-200K

Protein 200K is a DExD/H box RNA helicase containing two helicase domains which play an important role in the splicing process. RNA helicases often have large amino- or carboxy-terminal extensions that are not typically found in DNA helicases. These extensions are presumed to have additional functions such as an additional enzymatic activity, nuclear signalling, or providing specificity through interaction with protein cofactors (Tanner and Linder, 2001). Indeed, the cloning of the full-length 200K protein in this work has identified an additional 435-residues sequence in the amino-terminus which is absent in the previously published 200K sequence (Lauber *et al*., 1998). This region is conserved among the orthologues of 200K protein, as shown in the multiple alignment in figure 4.1. For this region (also termed 200-1), a Y2H assay has been performed to identify the interaction within tri-snRNP particle as described above. However, no interaction has been found so far. To understand the potential function of the N-terminal 200-1 fragment, yeast two-hybrid screening was used to search for the putative interaction partners. The plasmid pGBKT7-200-1 used as bait was transformed into yeast strain AH109, and was subsequently mated with a pretransformed human HeLa cDNA library in yeast strain Y187 from Clontech. The mating efficiency was 1% and about 10^6 clones were screened. 235 His^+ and Ade^+ clones were selected after incubation at 30°C for 5-12 days on the 50 150-mm plates containing selective medium SD/-Leu-Trp-His-Ade. These clones were reselected by their ability to grow in the presence of 20 mM 3-amino-triazole (3-AT) and their ability to activate the MEL1 reporter gene, as detected in a α- galactosidase assay on the high-stringency medium SD/-Leu-Trp-His-Ade+20 mM 3-AT+X-α-Gal. From this step, 126 positive clones were selected for preparing plasmid DNA samples. The samples were then sequenced at Sequence Laboratories Göttingen GmbH.

NCBI Blast was used to analyse all the sequence data obtained. Based on the DNA sequence of their inserts, 26 putative clones were identified and shown in the table 4.2. The clones that express non-biological peptides, i.e. anti-sense, intergenic region or non-protein-encoding regions, are excluded from the table of results. Most candidates are fused to the Gal4 sequence in the reading frame of the human gene, while some of them are out of frame. However, it was found that at least in some cases, frame-shifted translation occurs (Fromont-Racine *et al*., 1997). Nine candidates related to ribosomal protein, mitochondrial protein, zinc finger protein, proteasome, etc., which are marked with an asterisk in table 4.2, are probably false positive readouts, as suggested by previous studies (see 'Table of false positives' at http://www.fccc.edu/research/labs/golemis/Table1.html). However, some putative interactors may have biological relevance. For example, the identification of the import receptor karyopherin α2 implies that protein 200K may be recognized by this protein via its N-terminal nuclear localization signal for transport through the NPC. Identification of the nuclear export protein p270/Tpr suggests that it may function for the mRNA export, and subsequently may be involved in protein expression by association with the eukaryotic translation elongation factor 1 and translation initiation factor 3. Clearly, further work is needed to elucidate these questions.

Table 4.2. Outcome of two-hybrid screening using the N-terminal fragment of protein 200K (200-1: 1-434 amino acids) fused to the Gal4 DNA binding domain. The first column shows the number of identified clones for each gene. The asterisks (*) show that the protein encoded by this gene often displays a false positive result in the two-hybrid screens (see next page).

Clone No.	Putative human gene	Accession No.	Frame of insert
1	kinesin family member 21A (KIF21A)	gi|38569483	in frame
3	kinesin family member 23 (KIF23)	gi|20143966	in frame
1	eukaryotic translation elongation factor 1, alpha 1 (eEF1A1)	gi|25453469	in frame
1	eukaryotic translation initiation factor 3, subunit 10 (EIF3S10)	gi|4503508	not clear
1	karyopherin alpha 2 (KPNA2) (RAG cohort 1; importin alpha 1)	gi|4504896	in frame
1	p270/Tpr (translocated promoter region)	gi|4507658	in frame
1	fanconi anemia, complementation group G (FANCG)	gi|4759335	in frame
1	androgen-induced proliferation inhibitor (APRIN)	gi|7657268	in frame
1	chromosome X open reading frame 53 (CXorf53)	gi|13236582	in frame
2	topoisomerase II binding protein-1 (TOP-BP1) *	gi|20143948	in frame
1	proteasome (prosome, macropain) subunit, alpha type 6 (PSMA6) *	gi|23110943	in frame
1	mitochondrial ATP synthase epsilon chain *	gi|21327678	in frame
2	Zinc finger protein 143 (ZNF 143) *	gi|18874091	in frame
5	unnamed protein product (OXR1-like?)	gi|21753814	in frame
1	NONO (p54nrb)	gi|34932413	not clear
2	unknown DNA sequence from chromosome 16	gi|15375156	not clear
1	DNA sequence from Chromosome 14	gi|12641568	not clear
1	splicing factor, arginine/serine-rich 5(SFRS5) (HRS,SRP40)	gi|33869323	frame-1
4	chromosome 10 open reading frame 61 (C10orf61)	gi|24308108	frame-1
5	nuclease sensitive element binding protein 1 (NSEP1)	gi|34098945	frame-1
1	S-adenosylhomocysteine hydrolase (AHCY)	gi|9951914	frame+1
1	threonyl-tRNA synthetase (TARS) *	gi|38202254	frame+1
1	voltage-dependent anion channel 2(VDAC2); (mitochondrion) *	gi|42476280	frame+1
1	NADH dehydrogenase (ubiquinone) Fe-S protein 3 (NADH-coenzyme Q reductase) (NDUFS3)*	gi|4758787	frame+1
1	ribosomal protein L28 (RPL28) *	gi|34486095	frame+1
1	ribosomal protein S15a (RPS15A) *	gi|34335150	frame+1

5 Discussion

The tri-snRNP is especially important for the splicing reaction as it contains many essential components of the spliceosome. However, the mechanism of the tri-snRNP assembly and its roles in the activation of the spliceosome and in catalysis are still poorly understood. There is evidence that the protein-protein interactions play a critical role in these processes.

In this work the yeast two-hybrid techniques and *in vitro* biochemical methods were established to investigate the protein-protein interactions within the individual U4/U6 and U5 snRNPs, as well as protein contacts between these two snRNPs in the tri-snRNP particle. All the positive results are summarised in table 5.1. The data comprise Y2H screening for 43 full-length proteins and protein fragments derived from 15 protein components of the human tri-snRNP and from the U5-52K protein, which associates with the U5 snRNP but not the tri-snRNP. In total, 46 pairs of interacting proteins were detected in Y2H analysis. Of these, 20 interacting pairs were examined by *in vitro* binding assays and 15 pairs were again verified, making up 75% of the total examined samples. Six additional pairs of interacting proteins were observed in *in vitro* binding assays only. Table 5.2 presents some additional data on protein-protein interactions between the 90K protein and the recycling factor U6-p110, and the U2-associated protein SPF30/SMNrp. Eleven interactions were identified in the Y2H assay (table 5.2).

Protein pair	Y2H interaction	*In vitro* biochemical interaction	Section in the Results
20K/60K	+	+	4.2
60K/90K	-	+	
116K/200-4	+	NT	4.3.1
116K/220-1	+	NT	
116K/220-6	+	NT	
116-2/200-4	+	NT	
116-2/220-1	+	NT	
116-2/220-6	+	NT	
116-3/200-4	+	NT	
116-3/220-6	+	NT	
200K/220-6	+	NT	
200-4/220-1	+	NT	
200-4/220-6	+	NT	
220-1/220-6	+	NT	
220-6/220-6	+	NT	
220K/40K	-	+	
102K/15K	+	+	4.3.2
102K/102K	+	+	
102K/116K	+	-	
102K/116-2	+	-	
102K/116-3	+	-	
102K/200-2	+	-	
102K/200-4	+	+	
102K/220-1	+	+	
102K/220-6	+	-	
102K/61K	+	+	4.4.1
102K/90K	-	+	4.4.2
102K/90K_C267	-	+	
102K/90K_C381	-	+	
102K/90K_489	-	+	
110K/90K	+	+	4.5
110K/90K_C381	+	NT	
110K/90K_C489	+	NT	
110K/102K	+	+	
110K/110K	+	? (high background)	
110K/200K	+	+	
110K/200-4	+	+	
110K/200-5	+	+	
110ΔRS/90K	+	NT	
110ΔRS/102K	+	NT	
110ΔRS/200-4	+	NT	
110ΔRS/200-5	+	NT	
102K_NTD/15K	+	NT	4.6
102K_TPR_M/61K	+	NT	
102K_TPR_M/110K	+	NT	
102K_TPR_M/200-4	+	NT	
102K_TPR_M/220-1	+	NT	
102K_TPR_C/61K	+	NT	
52K/102K	+	+	4.7
52K/15K	+	+	
52N/102K	+	+	
52C/15K	+	+	
Summary	46(+); 6(-)	21(+); 25(NT); 1(?); 5(-)	

Table 5.1. Summary of the protein-protein interactions within the U4/U6 and the U5 snRNP, and between them in the tri-snRNP particle. NT, not tested for *in vitro* binding in this work (grey background).

Protein pair	Y2H interaction	*In vitro* biochemical interaction	Section in the Results
90K/p110	+	+	4.8
90K/p110ΔC	+	NT	
90K/p110-AF	+	NT	
90K_C267/p110	+	+*	
90K_C381/p110	+	NT	
90K_C489/p110	+	NT	
90K_C267/p110ΔC	+	NT	
90K_C381/p110ΔC	+	NT	
90K_C489/p110ΔC	+	NT	
90K/SPF30	+	+ *	4.9
90K_N442/SPF30	+	NT	
Summary	11(+)	3(+); 8(NT)	

Table 5.2. Summary of the protein-protein interactions between 90K protein and the protein U6-p110 or SPF30/SMNrp. NT, not tested for the *in vitro* binding in this work (grey background). * The result was shown previously (Medenbach *et al*., 2004; Meister *et al*., 2001).

5.1 Protein associations within the U4/U6-snRNP particle

U4/U6 snRNP contains five particle-specific proteins, namely 15.5K, 61K and the 20K, 60K, 90K. Y2H and *in vitro* binding assays were used to investigate how they bind to one another in the absence of snRNAs.

The well known paucity of interactions among U4/U6 snRNP proteins was again clearly confirmed by Y2H and by *in vitro* binding experiments (section 4.2.1). The interactions of 60K with 90K (*in vitro* binding assay only) and the 20K protein (Y2H and *in vitro*) remain the only ones observed. These interactions establish the biochemically stable 20K•60K•90K heterotrimeric complex (Horowitz *et al.,* 1997; Teigelkamp *et al.,* 1998). Previous studies showed that the N-terminal SFM domain of 60K is essential for the binding of 20K protein (Horowitz *et al*., 2002; Reidt *et al*., 2003), while the analysis of a mutant clone in this study indicated that the C-terminal WD40 domain of 60K is involved in the binding of 90K protein (section 4.2.1). Consistently with the distinct domains involved, co-immunoprecipitation results showed that one interaction does not

affect the other in the 20K•60K•90K complex (figure 4.4A). The failure of Y2H to demonstrate an interaction between 60K and 90K could have a trivial reason, such as folding problems or lack of a necessary post-translational modification.

It has been shown that the U4/U6 snRNP-specific protein 15.5K binds directly to a specific sequence element of U4 snRNA through a novel RNA-binding domain (Vidovic *et al.,* 2000). The 20K•60K•90K complex and 61K follow after this initiating step only (Nottrott *et al.,* 2002). It is not known whether the initial step forms a protein–RNA "nucleus" for the attachment of the other proteins or whether the function of protein 15.5K is simply to assist the correct folding of the U4 snRNA. However, the general failure to detect strong pairwise interactions between the heterotrimer, 61K and 15.5K would rather seem to point to a possible co-operative activity in complex formation, presumably involving the U4 snRNA, that disfavours the formation of 1:1 complexes such as are detected in Y2H and pull-down experiments of the type conducted in this work. Indirect support for the latter interpretation is afforded by the fact that proteins 61K and 90K can both be cross-linked to U4 snRNA and U6 snRNA respectively by photolysis (Nottrott *et al*., 2002).

5.2 Protein associations within the U5-snRNP particle

In contrast to the U4/U6 snRNP, the U5-snRNP particle assembles mostly via protein-protein interactions, as demonstrated in this work. These proteins exhibit a wide range of binding affinities, and dissociate from the spliceosome at different stages.

5.2.1 Interactions of the RNA-free 220K•200K•116K•40K heterotetramer

By combining two-hybrid assay with mutational analysis, the interactions between 220K, 200K and 116K, and the domains contributions were for the first time described in detail (section 4.3.1). Currently, I have not yet found a conventional biochemical method to confirm these Y2H interactions, since there was no functional recombinant protein for 220K, 200K or 116K available (see section 4.3.1.1 for an example). However, several lines of evidence from previous studies indicate that these interactions are specific and biologically relevant. First, these three proteins, together with the 40K protein, can form a

stable RNA-free complex at high salt *in vitro* (Achsel *et al*., 1998). Second, one of the interactions between 220K and 116K has previously been observed by using far-western blotting (Achsel *et al*., 1998). Third, as discussed below, most of the binding domains were also identified in the interactions of the orthologous proteins of yeast: Prp8p, Brr2p and Snu114p.

In a previous Y2H study van Nues and Beggs used this technique to establish a protein network within a different set of yeast spliceosomal proteins, such as the tri-snRNP-specific protein Snu66p and the U5 snRNP-specific proteins Prp8p and Brr2p. From their data a picture emerges of Brr2p as the hub of numerous protein-protein interactions, and Brr2p is therefore assigned the role of a central organizer not only of the tri-snRNP, but also of spliceosomal protein dynamics in general. Apart from its task of unwinding the U4/U6 snRNA duplex, Brr2p, through its C-terminus, appears to be involved in second-step factor recruitment and other protein contacts to U1, U2 as well as tri-snRNP specific proteins. My two-hybrid experiments confirm the critical function of this C-terminus in protein-protein interactions for protein 200K, the human orthologue of Brr2p. This includes the interactions within the RNA-free heterotetramer (through interactions with 220K and 116K) and with 110K. The region in the C-terminal half of 200K required for these interactions seems to be conserved between yeast and man (van Nues and Beggs, 2001). Interactions of a full-length 200K protein were observed only with the 220K fragment 6 (220-6) and the 110K protein (see section 5.4.1).

The human 200K protein, like its yeast orthologue, follows the design of a C-terminal helicase domain with low sequence conservation in domains V and VI (following the nomenclature of Tanner and Linder, 2001). For a group of yeast spliceosomal DEAH RNA helicases, comprising Prp2p, Prp16p, Prp22p and Prp43p, it has been shown recently that the interaction domain with the spliceosome resides outside of the catalytic helicase domain (Edwalds-Gilbert *et al.,* 2004). For the yeast Brr2p it has been reported that only the N-terminal helicase domain is required for the unwinding of the U4/U6 snRNA duplex (Kim and Rossi, 1999; Raghunathan and Guthrie, 1998). The poorly conserved C-terminus of the second helicase domain therefore possibly serves exclusively

as a protein-protein interaction domain. A case of conversion of an RNA-binding domain into a protein-binding domain was recently reported for the Y14 –Mago interaction (Fribourg *et al.,* 2003). The 200-5 fragment of 200K interacts with 110K protein in Y2H and *in vitro* binding assays. The fragment comprises a SEC63 domain which through its high density of negative charge is suspected to mediate protein-protein interactions particularly in the context of RNA-protein complexes (Willer *et al.,* 2003; Ponting, 2000).

The 220K protein, like its yeast orthologue Prp8p, is a large protein with only few recognizable domain features. The observed Y2H interaction of yeast Prp8p with Lin1p (the orthologue of the human U5 snRNP 52K protein) is predicted to be based on the interaction of the GYF domain of Lin1p with the proline-rich N-terminus of yeast Prp8p (Bialkowska and Kurlandzka, 2002; Freund *et al.,* 2002). A small region in the centre of yeast Prp8p (aa 1166-1193) is reported to interact with Snp1p, the orthologue of the human U1-70K protein (Awasthi *et al.,* 2001), while a neighbouring region (aa 1066-1107) was identified as a modulator of yeast Brr2p helicase activity (Kuhn *et al.,* 2002). In addition, Prp8p interacts genetically with Prp28p, the helicase which removes the U1 snRNA from the 5'-SS, and with second-step splicing factors Prp16p, Slu7p, and Prp22p (Umen and Guthrie, 1995; Schneider *et al.,* 2004). In contrast to the paucity of reported protein contacts, protein 220K maintains a broad variety of RNA contacts with U5 and U6 snRNA (Dix *et al*., 1998; Vidal *et al*., 1999) as well as with the 5' and 3' splice site and the branchpoint of the pre-mRNA (Teigelkamp *et al*., 1995; Query and Konarska, 2004). The current data show that 220K protein, like the yeast Prp8p (van Nues and Beggs, 2001), establishes protein contacts exclusively through its immediate N- or C-termini (fragments 220-1 and 220-6). In this way the protein establishes contacts to 200K, 116K and 102K proteins within the U5 snRNP (figures 4.6 and 4.9). Although Y2H could not identify any interaction of protein 40K, protein 40K is able to bind the protein 220K efficiently in the *in vitro* co-immunoprecipitation experiment (figure 4.7), consistently with the result obtained by far western blotting (Achsel *et al*., 1998). Therefore, either protein 40K or its interaction partners may be folded incorrectly and thus fail to exhibit stable protein-protein interactions in the Y2H system.

These data confirm 220K as the centre of the heterotetrameric 220K•200K•116K•40K complex. These proteins remain stably associated throughout the spliceosomal cycle and the final release of the post-spliceosomal 35S complex (Makarov *et al*., 2002; Makarova *et al*., 2004). Query and Konarska recently identified yeast Prp8p mutations that act as suppressors of the branchpoint A -> G mutation by improving the second step of splicing, possibly through stabilization of the second-step core catalytic structure. The authors point out that the available genetic data most favourably supports a model in which Prp8p, acting co-ordinately with U6 snRNA and Prp16p, serves as the modulator for major conformational rearrangements occurring between steps one and two of the splicing reaction (Query and Konarska, 2004).

5.2.2 Poly-TPR protein 102K is stably bound to U5-snRNP particle via multiple protein-protein interactions

Previous work showed that protein 102K remains tightly associated with U5 snRNP particle at salt concentrations of 700 mM NaCl (Makarov *et al*., 2000). Further experiments demonstrated that both U5 snRNA and Sm core proteins are not required for the association. In contrast, protein 102K binds efficiently to the *in vitro* isolated RNA-free heteromeric complexes from U5 snRNP, namely the complex 220K•200K•116K•40K and 220K•116K•40K. These observations are strongly supported by the protein-protein interaction data obtained in this work. The Y2H assay revealed that protein 102K interacts with U5-specific protein 15K, and distinctive regions from each of the three proteins 116K, 200K, and 220K within the tri-snRNP particle (section 4.3.2). The specificity of the interactions was confirmed by *in vitro* binding assays for most of the proteins except protein 116K (section 4.3.2). Thus, the interaction between protein 102K and 116K (or its fragments) may be too weak to be detectable by the *in vitro* binding assay. The stable association of 102K with U5 snRNP is probably important for the subsequent functions of 102K in the spliceosome assembly; it will be discussed below that protein 102K also interacts with the U4/U6-specific proteins 61K and 90K as bridging factor, with the tri-snRNP specific protein 110K, and with protein 52K (which binds to the U5 particle only).

This broad spectrum of interactions of protein 102K is consistent with the 19 TPR repeats of the protein. This protein-binding motif was first recognized as a distinct TPR subfamily in the characterization of the *Drosophila melanogaster crooked neck* (CRN) gene product (Zhang *et al*., 1991), and appears restricted to RNA-processing proteins such as U1 snRNP proteins Prp39p and Prp42p, crn homologues (Clf1p in yeast and hCRN in human), the human recycling factor U6-p110 and U5-102K protein in this study (Chung *et al*., 1999; McLean *et al*., 1998). The crn-like TPR variant, also termed HAT (half a TPR) repeat (Preker and Keller, 1998), is distinguished from the more general TPR repeat most notably by the absence of the conserved glycine at position 8, by the substitution of a glutamic acid for a hydrophobic residue at position 11, and by the prevalence of basic residues at positions 19, 21, 26. There is also a bias towards leucine, isoleucine or valine at positions 3, 5, 23, 28 (figure 5.1). The unique sequence characteristics of this subfamily presumably configure a TPR structure with distinct protein folding or ligand selection properties. However, the semi-conserved hydrophobic amino acids of the HAT motif are slightly different in position from the generalized TPR consensus, but still lie within the predicted nonpolar core of the motif, suggesting a similar overall structure.

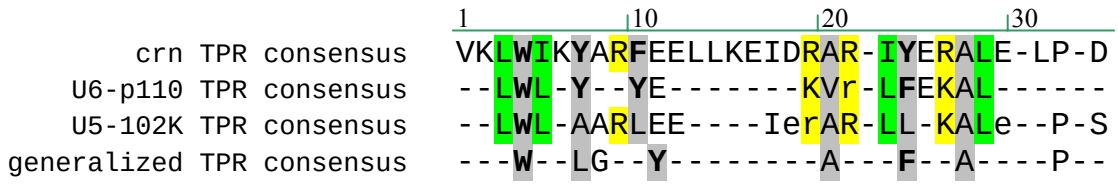


Figure 5.1. Comparison of the generalized TPR consensus with the consensus from a subtype of *Drosophila* crn, human U5-102K and U6-p110. The hydrophobic residues in the conserved or semi-conserved space through the TPR motifs are shaded in grey and the aromatic residues are highlighted in bold type. The leucine and isoleucine, existing mostly in the subfamily, are shaded in bright green, while the basic residues are shown in yellow.

The deletion approach in this work has shown that the TPR repeats of 102K protein are each dedicated to interactions with specific partners. All repeats participate in the interaction with the U4/U6-61K, whereas only the first nine repeats interact with 110K, the 200-4 fragment of 200K and the 220-1 fragment of 220K. Makarov *et al*. reported that antibodies raised against the C-terminal TPR repeats (TPR 10-19) efficiently precipitate free U5 snRNP, but not the U4/U6.U5 tri-snRNP, from HeLa nuclear extracts (Makarov *et al*., 2000).

Binding of the U4/U6-61K protein to the C-terminal domain could explain this observation. The N-terminal fragment of 102K, predicted to contain a coiled-coil domain (Makarov *et al.,* 2000), interacts with protein 15K.

Based on its highly repetitive TPR structure and the experimental data presented here, I propose that protein 102K probably acts as a scaffolding to drive the assembly of spliceosome by binding several other spliceosomal proteins.

5.2.3 The U5-52K protein interacts with the U5-specific proteins 102K and 15K, but dissociates upon tri-snRNP formation

The following evidence found in this study and other studies in the laboratory of Prof. Lührmann supports the conclusion that the 52K protein described here is indeed a U5 snRNP-specific protein (Laggerbauer *et al*., 2005; Behrens and Lührmann, 1991). First of all, antibodies against the recombinant 52K protein expressed in *E. coli* recognise only one protein of the appropriate size; this protein co-fractionates with U5 snRNP, and is also part of highly purified 20S U5 particles. Secondly, immunofluorescence studies with the same antibodies demonstrate that the 52K protein is found predominantly in the nucleus, and this finding was confirmed with an exogenous HA-tagged 52K protein. Thirdly, the 52K protein is stably associated with the purified U5 snRNP particle, and binds specifically to the U5 snRNP-specific proteins 102K and 15K. Therefore, unspecific aggregation of the 52K protein with snRNPs, e.g., during extract preparation, seems highly unlikely.

Database searches revealed that the sequence of the 52K protein is identical to that of a previously described protein, CD2BP2, which interacts with CD2, a surface marker of T cells, which facilitates activation of T lymphocytes on binding to its ligand CD58 (Laggerbauer *et al*., 2005; Nishizawa *et al*., 1998). Since CD2 resides in the plasma membrane of lymphocytes, the characterisation of the same protein as a component of nuclear U5 snRNPs raises interesting questions as to the mechanism(s) by which the 52K/CD2BP2 protein may exert different functions. In this respect it is interesting to note that the 52K protein is phosphorylated (Behrens and Lührmann, 1991) and post-

translational modification might well create isoforms with different association behaviour. Indeed, CD2 association may not be the only alternative function of the protein. The yeast orthologue of the 52K protein, Lin1p, interacts with factors involved in chromosome segregation, and Lin1p has therefore been proposed to play a role in this process (Bialkowska and Kurlandzka, 2002).

The distribution of the 52K protein in glycerol gradients and its behaviour in immunoaffinity purification experiments both demonstrate that it associates only with the U5 snRNP (Laggerbauer *et al.*, 2005). This raises the question of which contacts of the 52K protein with other components of the U5 snRNP might mediate its specific interaction with this particle. Recently, it was shown that the GYF domain of CD2BP2/52K interacts strongly with a particular polyproline motif that is also found in the common snRNP protein SmB. The GYF domain, overexpressed in mammalian cells, was shown to precipitate SmB from cell extracts, suggesting that the 52K protein might interact with all splicing snRNPs (Kofler *et al*., 2004). These findings are not easily reconcilable with our observation that the 52K protein does not associate with the other spliceosomal snRNPs, which each contain a set of the Sm proteins. In particular, the Sm proteins of the 12S U1 snRNP are largely exposed to the solvent (Stark *et al*., 2001). Nonetheless, we do not observe that the 52K protein co-fractionates with U1 in glycerol gradients. The yeast orthologue of the 52K protein, Lin1p, interacts with the yeast U5 snRNP protein Prp8p, presumably through the extensive polyproline stretches at the N terminus of Prp8p (Bialkowska and Kurlandzka, 2002). I do not detect an interaction of the 52K protein with the human Prp8/220K protein, nor with any of six overlapping fragments of the protein, at least by two-hybrid analysis, and, indeed, the proline-rich N terminus of Prp8 is absent in the mammalian Prp8 protein. Instead, employing yeast two-hybrid analysis and *in vitro* binding assays, I find specific contacts of the 52K protein with the 20S U5-specific 102K and 15K proteins. These two interactions may well occur simultaneously, as different domains of the 52K protein are involved; the N-terminal two-thirds bind to 102K while the C-terminus, including the GYF domain, binds to the 15K protein (figure 4.14). I therefore believe that these contacts are responsible for the specific integration of the 52K protein into the 20S U5 particle.

Considering that the GYF domain has been characterised as a polyproline-targeting molecule (Freund *et al*., 2003; Freund *et al*., 1999; Nishizawa *et al*., 1998), my observation that it forms a specific complex with the 15K protein came as a surprise, as 15K is devoid of potential proline-rich binding motifs (Reuter *et al*., 1999). Thus my data indicate for the first time that a GYF domain can also engage in specific protein-protein interactions in a proline-independent manner. A crystallography study of 52K GYF domain in the complex with U5-15K indeed showed that the distinct surface of GYF domain binds to the non-proline-rich 15K (figure 4.16).

The 52K protein is the only U5 snRNP protein that is not present in the U4/U6.U5 tri-snRNP particle. Since the same is true of its yeast orthologue (Stevens et al., 2001), this characteristic property appears to be universally conserved. As U5 snRNP enters the spliceosome as part of the U4/U6.U5 tri-snRNP, the 52K protein does not appear to have a function in the splicing reaction itself. Consistently, deletion of the 52K protein or the yeast homologue of the 52K protein (Lin1p) is not lethal (Laggerbauer et al., 2005; Bialkowska and Kurlandzka, 2002). This suggests that the 52K protein functions in a pathway that is auxiliary or redundant for the splicing process, and it is tempting to speculate that this function may have something to do with snRNP assembly/recycling; the 52K protein is the only U5 snRNP protein that leaves the 20S U5 snRNP when the latter is integrated into the U4/U6.U5 tri-snRNP. In this respect, it resembles the U6-p110 protein (Prp24p in yeast), which is a component of the U6 and U4/U6 snRNPs but not of the tri-snRNP (see section 5.5). Possibly, the 52K protein covers a protein–protein interaction site that is necessary for the association of the U5 and U4/U6 snRNPs. In support of this idea, the 52K protein interacts with the 102K protein in the U5 snRNP (figure 4.14). The 102K protein forms part of the bridge between the U5 and U4/U6 particles (Schaffert et al., 2004; Makarova et al., 2002) and may well need to be masked to prevent premature incorporation of U5 particles, which are not fully assembled, into the tri-snRNP (Makarov et al., 2000). Alternatively, the 52K protein could act as a chaperone for the 15K and 102K proteins. These two proteins form a dimer independently of the 52K protein (section 4.3.2), which 52K contacts via two independent interactions. Interestingly, both 15K and 102K

are not present in the 35S U5 snRNP which very likely represents a post-spliceosomal U5 intermediate in the spliceosomal disassembly process (Makarov et al., 2002). Therefore these proteins appear to leave the U5 part of the tri-snRNP before or during the catalytic steps of the splicing reactions. Since two-thirds of the 102K protein consist of tetratricopeptide repeats, which are protein motifs capable of accommodating several simultaneous protein-protein interactions, it is conceivable that this protein needs a chaperone to avoid unfavourable interactions or aggregation before it is reincorporated into the 20S U5 snRNP and then into the 25S U4/U6.U5 tri-snRNP. It will be interesting to see, in future studies, what role the 52K protein may play in the dynamic restructuring of U5 snRNP-particles in the context of the nuclear pre-mRNA splicing machinery.

5.3 The interaction between U5-102K and U4/U6-61K is critical for the formation of U4/U6.U5 tri-snRNP

The formation of the U4/U6.U5 tri-snRNP seems to be exclusively protein-mediated, since there is no interaction between the U5 and U4/U6 snRNAs. Previous experiments showed that the 102K protein, prepared by *in vitro* translation, is efficiently co-precipitated with 13S U4/U6 snRNP by antibodies directed against the U4/U6-specific 60K protein (Makarov *et al*., 2000), and that the [^{35}S]-labelled 61K protein prepared by *in vitro* translation binds to purified 20S U5 snRNP, as demonstrated by co-immunoprecipitation of 61K with U5 snRNP at low salt concentration using anti-Sm (Y12) antibodies (Makarova *et al*., 2002). This study is the first time that the direct one-to-one protein-protein interactions between the two particles have been identified; the U5 snRNP-specific protein 102K interacts specifically with U4/U6 snRNP-specific proteins 61K and 90K. Additional evidence from different experimental approaches supports the critical role of the 102K–61K interaction for tri-snRNP formation (Makarova *et al.*, 2002; Schaffert *et al*., 2004). First, removal of protein 61K from HeLa nuclear extracts by immunodepletion inhibits tri-snRNP formation, subsequent spliceosome assembly, and pre-mRNA splicing. Complementation with recombinant 61K protein restores each of these steps. However, the removal of protein 61K from nuclear extract does not affect the stability of the accumulating U4/U6 or U5 snRNP particles (Makarova *et al*., 2002). Second,

Schaffert *et al.* (2004) demonstrated that, after knockdown of the U4/U6-specific 61K or the U5-specific 102K protein in HeLa cells using RNA interference, tri-snRNP formation is inhibited and stable U5 snRNP and U4/U6 snRNP containing U4/U6 proteins and the U4/U6 recycling factor p110 accumulate. Thus, both U5-102K and U4/U6-61K are essential for the interaction between U5 snRNP and U4/U6 snRNP *in vivo*.

5.4 The recruitment of tri-snRNP to the pre-spliceosome

The final step in the formation of the mature spliceosome is the binding of the U4/U6.U5 tri-snRNP to the pre-spliceosome (complex A), leading to the complete spliceosome (complex B). The data from this study have confirmed that several tri-snRNP proteins are involved in this step.

5.4.1 The C-terminal region of tri-snRNP-110K is required for anchoring protein 110K to the tri-snRNP while the N-terminal RS domain potentially mediates the association with the pre-spliceosome

Unlike 102K and 61K proteins, the tri-snRNP specific protein 110K, like the tri-snRNP specific 65K protein, is not required for tri-snRNP stability (Makarova *et al.*, 2001). While Snu66p is not an essential protein in yeast (Stevens *et al*. 2001), proteins 110K and 65K are required for joining the human tri-snRNP to the pre-spliceosomal complex A. I was able to identify interactions of the 110K protein (hSnu66) with specific U4/U6- and U5-snRNP proteins, namely the 90K and the 102K, 200K proteins through the C-terminal region lacking RS domain. Since the absence of 110K protein does not compromise the stability of the tri-snRNP, it appears reasonable that anchoring 110K to these proteins is required to position protein 110K properly for its contribution in connecting the tri-snRNP to the A-complex. For all three tri-snRNP specific proteins, 110K, 65K and the 27K, this interaction is probably mediated through N-terminal RS domains (Makarova *et al.,* 2001; Fetzer *et al.,* 1997). It has been shown that RS domains function by directly interacting with the RS domains of other splicing factors. For example, the interaction between SR protein SF2/ASF and the SR-related protein U1-70K is mediated by their RS domains, which facilitating recruitment of U1 snRNP to the pre-mRNA substrate (Kohtz *et al*., 1994). This observation suggests that the interactions between SR proteins and one or more of the tri-

snRNP SR-related proteins may well contribute to the stable association of the U4/U6.U5 tri-snRNP complex with the pre-spliceosome. Consistently with this, indirect evidence has been provided previously that the SR proteins are necessary to recruit the tri-snRNP to the spliceosome (Roscigno and Garcia-Blanco, 1995). However, there was no interaction for 65K and 27K identified in two-hybrid assay. This might due to the folding problems or lack of a necessary post-translational modification.

5.4.2 The N-terminal region of U4/U6-90K is essential for the binding of U2-SPF30/SMNrp required for the formation of the mature spliceosome

The *in vivo* two-hybrid assay (this study) and *in vitro* binding assay (Meister *et al*., 2001) both point to a direct interaction between U4/U6-90K and the U2 snRNP-associated protein SPF30/SMNrp (figure 4.19). Mutational analysis further showed that the 90K N-terminal region, comprising amino acids 1 to 442, participates efficiently in the interaction. This region contains a conserved PWI motif at its N terminus (amino acids 3 to 76). The presence of PWI motifs in known splicing factors suggests that it may be important for numerous interactions within splicing complexes (Szymczyna *et al*., 2003; Blencowe and Ouzounis, 1999). However, it seems not to be required for, or at least not efficient in, the binding of protein 90K, since the shorter N-terminal fragment (amino acids 1–194) containing the PWI domain does not bind to the 90K. The 90K–SPF30/SMNrp interaction appears to be an important link between tri-snRNP and pre-spliceosomal A-complex, in addition to the RS-domain interactions of the tri-snRNP-specific proteins. The presence of SPF30/SMNrp in the BΔU1 complex, but its absence in the activated spliceosome of the B* complex, lends support to the idea of the 90K–SPF30/SMNrp interaction being an important temporary link to build a spliceosome of increasing molecular complexity (Makarov *et al*., 2002; Makarova *et al*., 2004). Removal of SPF30/SMNrp from nuclear extract by immunodepletion inhibits the first step of pre-mRNA splicing by preventing the formation of mature spliceosome (complex B). Re-addition of recombinant SPF30/SMNrp to the immunodepleted extract reconstitutes both spliceosome formation and splicing (Meister *et al*., 2001; Rappsilber *et al*., 2001). It is interesting to identify the U2 snRNP protein(s) to which SPF30/SMNrp binds, and to characterize the domains involved in the

event. Detailed knowledge about the conditions and the temporal order of these interactions will be required in order to clarify whether SPF30/SMNrp acts as a bridging factor between the two snRNPs, or whether it actively induces a structural rearrangement required for their stable association.

Since there is no known yeast orthologue of the human SPF30/SMNrp, and the yeast homologue proteins Snu66p (110K in human) and Sad1p (65K) lack RS domains, yeast probably uses a different mechanism to establish a connection between the tri-snRNP and the A complex. Experiments with ts mutants of yeast Prp31p, which is not required for tri-snRNP stability, were interpreted by assuming that in yeast Prp31p connects the tri-snRNP to complex A (Weidenhammer *et al.,* 1997).

5.5 The link between the C-terminal region of U4/U6-90K and the TPR domain of U6-p110 is necessary for the recycling of U4/U6 snRNPs

After each spliceosome cycle, the U4 and U6 snRNAs are released separately and are recycled to the functional U4/U6 snRNP, requiring in the mammalian system the U6-specific RNA-binding protein p110 (SART3). p110 contains in its C-terminal region two RNA recognition motifs (RRMs), and an extensive N-terminal TPR domain which is absent in the yeast orthologue protein Prp24p (figure 5.2).

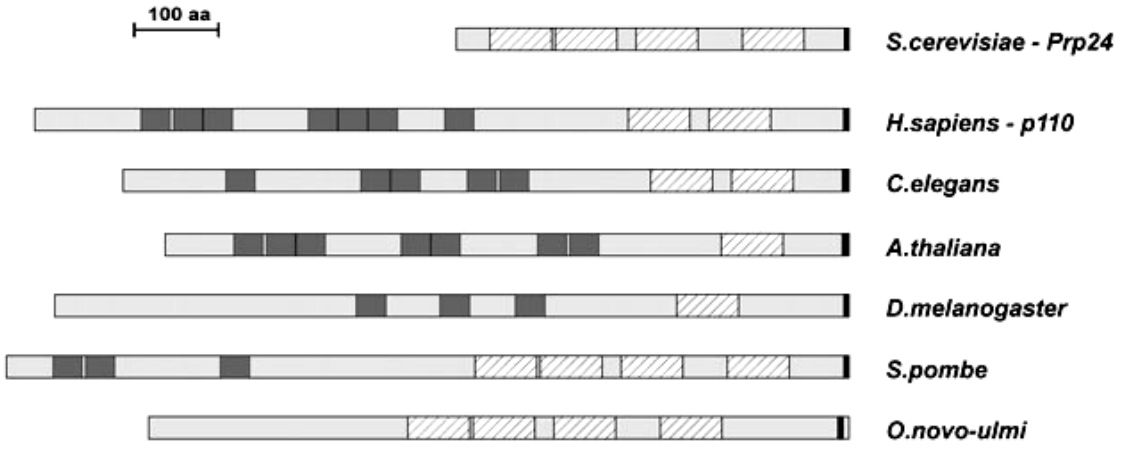


Figure 5.2. Domain structure of the *S. cerevisiae* Prp24 protein, human p110 and related proteins from other species. The proteins are aligned relative to their C-termini. The RRM (striped boxes) and TPR (HAT) motifs (in dark grey) as well as the short C-terminal region (in black) are indicated. Adapted from Bell *et al*., 2002.

TPR protein-protein interaction domain in p110 belong to the same subfamily of U5-102K, thus indicating that this protein may function by interacting with other splicing factors, probably the U4/U6 snRNP proteins. Indeed, two-hybrid screening with p110 allowed the identification of an interacting partner of protein 90K out of the five U4/U6-specific proteins. The specificity of the interaction was confirmed by an *in vitro* binding assay. Mutational analysis further showed that the N-terminal TPR domain, or the set of three N-terminal TPR motifs alone, appears to be sufficient for 90K protein binding. As no other protein motifs are predicted in this N-terminal part of p110, the results strongly suggest that the TPR motifs mediate the specific interaction with the 90K protein. There seems to be some redundancy in the interaction, since the number of TPR motifs varies at least between three and seven in the known orthologues (figure 5.2). These domain structure data would also be consistent with the finding that the three N-terminal TPR motifs present in the splice variant AF 387506 appear to suffice for interaction with the 90K protein (figure 4.18A). The latter finding raises the interesting question as to whether the splicing variant (i.e. AF 387506) may act as a negative regulator of p110 function, competing for the 90K protein interaction but lacking the RNA binding and recycling functions. Alternatively, p110 may associate some other unknown proteins through its additional TPR repeats for other functions.

In the 90K protein, the C-terminal region (amino acids 417–683) suffices for the p110 interaction (figure 4.18B). This region provides a novel bridge between the U6 snRNA-bound p110 and the U4/U6-specific protein components and most probably plays an important role in U4/U6 snRNP assembly. It is notable that the 90K protein – within the 20K/60K/90K complex – is in contact with the U6 snRNA in the stem II region (Nottrott *et al.*, 2002), and this event may stabilize the newly formed U4/U6 base-paired structure.

These data, along with the other data from the co-operating laboratory of Prof. Bindereif, were published recently in the journal of Molecular and Cellular Biology (Medenbach *et al.*, 2004). In this paper, it was further shown that the two RRM domains at the C terminus are required for U6 snRNA binding, while the N-terminal TPR domain is not. Moreover, the *in vitro*

depletion/complementation experiments showed that both the N-terminal TPR domain and the C-terminal RRMs are required for the U4/U6 snRNP recycling activity. This can explained by the observations that the TPR domain is required for the binding of U4/U6-90K, and that the RRMs domain is essential for the U6 snRNA binding.

How do these findings fit to the model of the U4/U6 snRNP assembly pathway? In contrast to the yeast protein, the mammalian p110 protein carries only two RRMs homologous to the highly conserved second and third RRMs of Prp24p. As proposed by others (Rader and Guthrie, 2002), the additional RRMs of Prp24p, 1 and 4, may recognize U4 snRNA and thereby promote the U4/U6 interaction. Perhaps this U4 interaction has shifted from an RNA-protein interaction (such as in yeast Prp24 RRMs 1 and 4 with the U4 snRNA) in the mammalian and other systems to a protein-protein interaction (p110 TPR domain with the 90K protein).

The recycling factor p110 is associated with human U6 and U4/U6 snRNPs, but it dissociates after the formation of tri-snRNP. What makes p110 leave the reassembled U4/U6 snRNP? Most probably, this signal is provided by the association with U5 snRNP, which results in U4/U6.U5 tri-snRNP formation. The data from this study suggest that U5-102K may play a role in the release of p110. Protein 102K is the only U5-specific protein that interacts with U4/U6-snRNPs, namely protein 61K and 90K (section 4.4). As discussed in section 5.3, the interaction between U5-102K and U4/U6-61K is critical for the formation of tri-snRNP. The interaction between U5-102K and U4/U6-90K may play a similar role. However, it is also possible that U5-102K binds 90K and its function is to replace the p110. This suggestion is supported by the fact that both proteins bind to the same region in 90K (amino acids 417-683) through the same domain structure of TPR motifs (see figure 5.1). Note that the yeast Prp6p (102K in human) behaves as a genuine U4/U6 snRNP protein (Stevens *et al.,* 2001) and not as a U5 snRNP protein such as protein 102K.

5.6 The role of tri-snRNP splicing factors in retinitis pigmentosa

Retinitis pigmentosa (RP) is a genetic disease characterized by progressive degeneration of the peripheral retina leading to night blindness and loss of visual fields. RP can be inherited in autosomal dominant, autosomal recessive, X-linked, mitochondrial and genetically more complex modes. To date, more than 30 genes have been found to be implicated in this disease. Recently, mutations in 61K, 90K and 220K, the respective gene expression products of PRPF31, PRPF3, and PRPF8, have been associated with autosomal dominant retinitis pigmentosa (adRP) (McKie *et al*., 2001; Withana *et al*., 2001; Chakarova *et al*., 2002). Given the overwhelming evidence, there is no doubt that mutations within these splicing factors result in RP; however, why ubiquitously expressed genes with such an apparently crucial function only cause a disease of the photoreceptors is still largely a mystery.

Results from initial functional studies on 61K protein have provided some clues to the effects of its mutations *in vivo* (Deery *et al*., 2002). An *in vivo* assay of splicing function in human cells using a bovine rod opsin splicing template did not detect any defect in splicing efficiency or accuracy attributable to either protein 61K mutation (A194E or A216P), suggesting that there is no dominant negative effect on the wild-type protein and that the pathology more probably arises from other causes as indicated by other experiments. Analysis in cell culture has offered some evidence of reduced protein solubility in those 61K proteins that incorporate missense mutations. If this observation holds true under the physiological conditions of the photoreceptors, it might result in insoluble protein aggregates accumulating over time, eventually causing apoptosis. Alternatively, some of the experimental data suggest that the lack of solubility might result in less protein being transported from the cytosol to its site of action in the nucleus. The authors therefore suggested that the reduction in concentration of functional 61K protein in the nucleus could lead to an insufficiency of splicing function. Photoreceptors shed their outer segments discs, where rhodopsin resides, on a daily basis and, therefore, require sufficient levels of rhodopsin transcript to replace the protein continually. Such transcripts obviously require splicing before translation. Rather than mutations in 61K causing a dominant-negative effect, the disease phenotype might be

related to thresholds in the levels of splicing. In other words, the photoreceptors of patients might survive until the level of splicing of rhodopsin, or other mRNAs, drops below a crucial level, at which point the cell would be irrevocably damaged, eventually leading to apoptosis. In support of this theory, families with 61K mutations exhibit an 'all or none' mode of incomplete penetration.

Y2H assay in this study showed that the yeast strain AH109 co-transformed with pGADT7-102K and pGBKT7-61K-A194E appears to grow when incubated on the selective medium lacking histidine and adenine at 30°C for more than two days. However, the strain grows much more slowly than the cells co-transformed with 102K and 61K wild-type or A216P. This indicates that the A194E mutant has a significant but relatively mild effect on the interaction between 102K and 61K. Consequently, this may reduce the formation of the functional U4/U6.U5 tri-snRNP that would lead to a lower the rate of splicing in the cell. This could result in the disease as proposed above by Deery *et al*. The weak interaction between protein 61K and 102K might be due to the lower solubility of the mutant protein as observed by Deery *et al*. However, the A216P mutant appears to have no such effect on the protein interaction, although one cannot exclude the possibility that it may reduce the binding to the U4 snRNA.

A similar explanation can be used for the observation of 220K mutants, since some 220K mutations such as R2310G, R2310K and F2314L display a weak interaction with protein 116K and 220K. Interestingly, the work from Beggs' group (Grainger *et al*., 2003) has revealed that, in yeast, Prp8-RP13 mutations affect the interaction between Prp8p and Brr2p.

5.7 The model of the human U4/U6.U5 tri-snRNP assembly

Based on currently available data, figure 5.3 presents the model of the assembly and the structural organisation of human tri-snRNP. In the U5 snRNP particle, three large proteins 220K, 200K and 116K associate with one another to form a basic protein-protein network for the subsequent assembly steps. The proteins 40K and 102K join this network by connecting with 220K (for 40K) or all the three proteins (for 102K). Protein 52K and 15K follow by binding to protein 102K. U5 snRNA binds to the protein component, probably to proteins 220K and 116K (Dix *et al.,* 1998). For the U4/U6-snRNP particle, a hierarchical assembly has been proposed previously (Nottrott *et al*., 2002) and is confirmed and discussed in this work (see section 5.1 for details). This study and others further showed that, at least under recycling conditions, protein p110 is required for the formation of U4/U6 snRNP (Bell *et al.,* 2002; Medenbach *et al.,* 2004). For this function, U6-p110 interacts with U4/U6-90K through the TPR domain and with U6 snRNA through the RRM motifs (see section 5.5 for details). During the assembly of the U4/U6.U5 tri-snRNP, the U5-102K protein binds to the U4/U6-specific proteins 61K and 90K and thus makes a connection between the U4/U6 and U5 snRNPs. In the meantime, the tri-snRNP-specific 110K protein (probably also 65K and 27K, which are not shown in this model) joins by binding to U5-102K, 200K and U4/U6-90K. These interactions promote the dissociation of U5-52K from U5-snRNP and U6-p110 from U4/U6-snRNP, resulting in the U4/U6.U5 tri-snRNP. The final step in the formation of the mature spliceosome (complex B) is the recruitment of the pre-assembled U4/U6.U5 tri-snRNPs to the pre-spliceosome (complex A). In this work, I confirmed that the U4/U6-90K mediates the process through interacting with U2-SPF30/SMNrp.

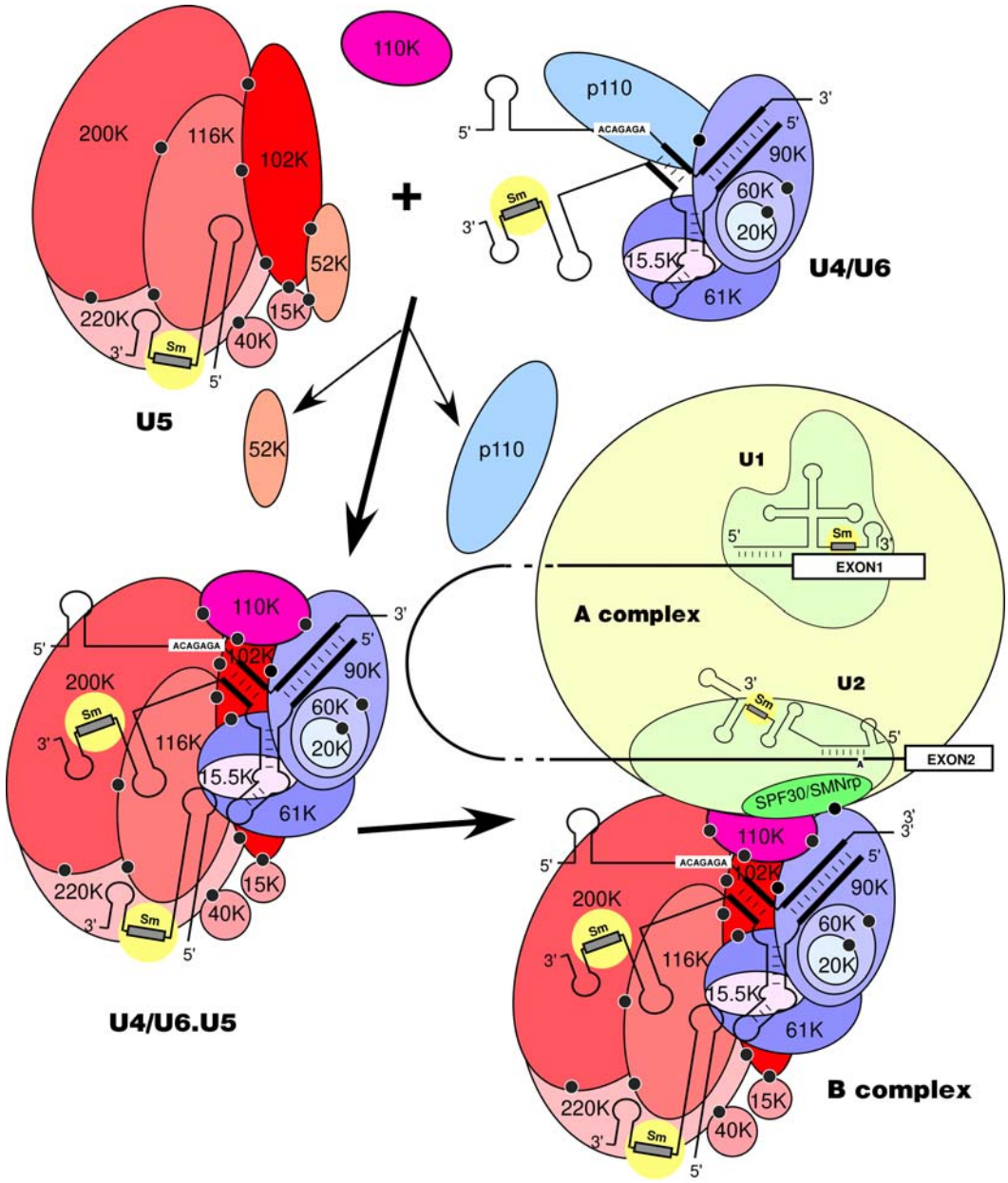


Figure 5.3. Model of the structural organisation and the assembly of human U4/U6.U5 tri-snRNP. The black spots denote the direct protein-protein interactions identified in this study.

5.8 Outlook

The protein-protein interaction data obtained in this work provide a basic knowledge on which further studies of the three-dimensional structure of the U4/U6, U5, and U4/U6.U5 snRNP particles by methods such as X-ray crystallography will be able to build. One example is the study of the U5-52K GYF domain in complex with U5-15K in the cooperation with the laboratory of Prof. R. Ficner at the University of Göttingen (figure 4.16). The crystallography study of the U4/U6 snRNP particle will answer the question of whether the 15.5K protein associates with 61K and 20K•60K•90K heterotrimer in the presence of U4/U6 snRNA and how this protein promotes the assembly of the U4/U6 snRNP.

As demonstrated in this work, the U5-102K and U6-p110 interact with the same region of U4/U6-90K, suggesting that the U5-102K interacts with 90K to replace the U6-p110. If this is true, the addition of the recombinant U5-102K protein to the nuclear extract under recycling conditions will probably inhibit the function of p110 in the recycling of the U4/U6 snRNPs and thus the splicing of pre-mRNA. The interaction between U5-102K and U4/U6-61K is critical for the formation of the U4/U6.U5 tri-snRNP (section 5.3). One of the 61K mutants (A194E) – one that contributes to the autosomal dominant retinitis pigmentosa – displays reduced interaction with 102K (figure 4.20). Recent work has demonstrated that, after knockdown of the U4/U6-61K protein in HeLa cells using RNAi, tri-snRNP formation is inhibited (Schaffert *et al.*, 2004). Therefore, I and my colleagues in the laboratory of Prof. Lührmann are working on the question of whether overexpression of the wild-type protein 61K or its mutants can restore the formation of tri-snRNP in HeLa cells lacking 61K.

The U4/U6-90K protein is involved in the assembly/reassembly of U4/U6 snRNP, the formation of tri-snRNP and the recruitment of tri-snRNP to the pre-spliceosome. Therefore, it will be interesting to see how the spliceosome assembles after the depletion/complementation of the full-length or truncated 90K protein.

The 52K protein is the only U5-specific protein that is not integrated into the tri-snRNP. Therefore, an attractive idea would be to purify the U5 snRNP particle by immunoaffinity selection using anti-52K antibodies and to analyse the protein component by mass spectrometry. This may yield purified U5 snRNP for the subsequent structural and functional studies.

6 Acknowledgments

First of all, I would like to thank Professor Reinhard Lührmann for giving me the opportunity to work in his laboratory, for providing excellent experimental facilities, for his support and consideration.

I would like to thank Hans-Peter Vornlocher for all his help in the first two years of my work, for giving me an excellent introduction into the exciting field of the yeast two-hybrid system.

I am grateful to Paul Woolley for checking the English of this dissertation and for his valuable suggestions and timely help.

I should like to express my appreciation to the University of Göttingen for accepting me as a PhD student. I wish to express my gratitude to the following professors for agreeing to be members of my examination committee: Prof. Ralf Ficner, Prof. Hans-Joachim Fritz, Prof. Dietrich Gradmann, PD Dr. Stefan Irniger, Prof. Oliver Einsle and Prof. Ernst A. Wimmer. I would like to give my special thanks to my referee Prof. Ficner for his advice and help during my study.

I wish to express my thanks to Hans-Peter Vornlocher, Patrizia Fabrizio, Evgeny Makarov, Olga Makarova, Stephanie Nottrott, Tilmann Achsel, Dierk Ingelfinger for providing me with original cDNA cloning; to Juliane Moses for her helpful work; to Christian Merz and Irene Öchsner for their help in lab 114; to Marion Killian, Gabi Heyne and Dagmar Meyer for the sequencing runs; to Klaus Hartmut, Berthold Kastner and the IT team for solving computer-related problems; to Reinhard Rauhut for the useful bio-information.

I would like to thank all the past and present members in the laboratory of Prof. Lührmann for their help and kindness.

I would like to thank Markus Wahl for his special interest, and also to his previous and present group members Thomas Conrad, Simon Trowitzsch, Gert Weber, Mihaela Diaconu, Nina Müllers, Catharina Netter, Elke Penka, Ulrich Reidt and Li-chi Chang for the warm and friendly working atmosphere.

I would also like to thank all my friends in the church for their support and prayer.

Lastly, I address my very special thanks to my wife Yan for her support and her love; thanks also to my mother-in-law and my grandmother-in-law for helping my wife to take care of my newborn son Joseph while I was writing this dissertation.

7 References

A

Abovich, N., Legrain, P. and Rosbash, M. (1990) The yeast PRP6 gene encodes a U4/U6 small nuclear ribonucleoprotein particle (snRNP) protein, and the PRP9 gene encodes a protein required for U2 snRNP binding. *Mol Cell Biol*, **10**, 6417-6425.

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S. and Lührmann, R. (1998) The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol Cell Biol*, **18**, 6756-6766.

Anthony, J.G., Weidenhammer, E.M. and Woolford, J.L., Jr. (1997) The yeast Prp3 protein is a U4/U6 snRNP protein necessary for integrity of the U4/U6 snRNP and the U4/U6.U5 tri-snRNP. *RNA*, **3**, 1143-1152.

Awasthi, S., Palmer, R., Castro, M., Mobarak, C.D. and Ruby, S.W. (2001) New roles for the Snp1 and Exo84 proteins in yeast pre-mRNA splicing. *J Biol Chem*, **276**, 31004-31015.

Ayadi, L., Callebaut, I., Saguez, C., Villa, T., Mornon, J.P. and Banroques, J. (1998) Functional and structural characterization of the prp3 binding domain of the yeast prp4 splicing factor. *J Mol Biol*, **284**, 673-687.

Ayadi, L., Miller, M. and Banroques, J. (1997) Mutations within the yeast U4/U6 snRNP protein Prp4 affect a late stage of spliceosome assembly. *RNA*, **3**, 197-209.

B

Bach, M. and Lührmann, R. (1991) Protein-RNA interactions in 20S U5 snRNPs. *Biochim Biophys Acta*, **1088**, 139-143.

Bach, M., Winkelmann, G. and Lührmann, R. (1989) 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc Natl Acad Sci U S A*, **86**, 6038-6042.

Banroques, J. and Abelson, J.N. (1989) PRP4: a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. *Mol Cell Biol*, **9**, 3710-3719.

Bartels, C., Klatt, C., Lührmann, R. and Fabrizio, P. (2002) The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome. *EMBO Rep*, **3**, 875-880.

Bartels, C., Urlaub, H., Lührmann, R. and Fabrizio, P. (2003) Mutagenesis suggests several roles of Snu114p in pre-mRNA splicing. *J Biol Chem*, **278**, 28324-28334.

Behrens, S.E. and Lührmann, R. (1991) Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev*, **5**, 1439-1452.

Bell, M., Schreiner, S., Damianov, A., Reddy, R. and Bindereif, A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J*, **21**, 2724-2735.

Bialkowska, A. and Kurlandzka, A. (2002) Proteins interacting with Lin 1p, a putative link between chromosome segregation, mRNA splicing and DNA replication in Saccharomyces cerevisiae. *Yeast*, **19**, 1323-1333.

Bjorn, S.P., Soltyk, A., Beggs, J.D. and Friesen, J.D. (1989) PRP4 (RNA4) from Saccharomyces cerevisiae: its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle. *Mol Cell Biol*, **9**, 3698-3709.

Black, D.L. and Pinto, A.L. (1989) U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Mol Cell Biol*, **9**, 3350-3359.

Blatch, G.L. and Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*, **21**, 932-939.

Blencowe, B.J. and Ouzounis, C.A. (1999) The PWI motif: a new protein domain in splicing factors. *Trends Biochem Sci*, **24**, 179-180.

Boelens, W.C., Palacios, I. and Mattaj, I.W. (1995) Nuclear retention of RNA as a mechanism for localization. *RNA*, **1**, 273-283.

Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Haendler, B. and Jacob, M. (1982) U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *EMBO J*, **1**, 1259-1265.

Brosi, R., Hauri, H.P. and Kramer, A. (1993) Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem*, **268**, 17640-17646.

Brow, D.A. (2002) Allosteric cascade of spliceosome activation. *Annu Rev Genet*, **36**, 333-360.

Burge, C.B., Tuschl, T. and Sharp, P.A. (1999) Splicing of Precursors to mRNAs by the Spliceosomes. In Gesteland, Cech and Atkins (eds.) The RNA world. Cold Spring Harbor Laboratory Press, New York, pp. 525-560.

C

Caspary, F. and Seraphin, B. (1998) The yeast U2A'/U2B complex is required for pre-spliceosome formation. *EMBO J*, **17**, 6348-6358.

Caspary, F., Shevchenko, A., Wilm, M. and Seraphin, 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.

Cech, T.R. (1990a) Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Biosci Rep*, **10**, 239-261.

Cech, T.R. (1990b) Self-splicing of group I introns. *Annu Rev Biochem*, **59**, 543-568.

Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L., Patel, R.J., Papaioannou, M.G., Inglehearn, C.F., Keen, T.J., Willis, C., Moore, A.T., Rosenberg, T., Webster, A.R., Bird, A.C., Gal, A., Hunt, D., Vithana, E.N. and Bhattacharya, S.S. (2002) Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet*, **11**, 87-92.

Chen, J.Y., Stands, L., Staley, J.P., Jackups, R.R., Jr., Latus, L.J. and Chang, T.H. (2001) Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol Cell*, **7**, 227-232.

Cheng, S.C. and Abelson, J. (1987) Spliceosome assembly in yeast. *Genes Dev*, **1**, 1014-1027.

Chung, S., McLean, M.R. and Rymond, B.C. (1999) Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA*, **5**, 1042-1054.

Collins, C.A. and Guthrie, C. (1999) Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev*, **13**, 1970-1982.

Collins, C.A. and Guthrie, C. (2000) The question remains: is the spliceosome a ribozyme? *Nat Struct Biol*, **7**, 850-854.

Cooper, M., Johnston, L.H. and Beggs, J.D. (1995) Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J*, **14**, 2066-2075.

D

Dahlberg, J.E. and Lund, E. (1991) How does III x II make U6? *Science*, **254**, 1462-1463.

Das, B.K., Xia, L., Palandjian, L., Gozani, O., Chyung, Y. and Reed, R. (1999) Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. *Mol Cell Biol*, **19**, 6796-6802.

Deery, E.C., Vithana, E.N., Newbold, R.J., Gallon, V.A., Bhattacharya, S.S., Warren, M.J., Hunt, D.M. and Wilkie, S.E. (2002) Disease mechanism for retinitis pigmentosa (RP11) caused by mutations in the splicing factor gene PRPF31. *Hum Mol Genet*, **11**, 3209-3219.

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.

Dziembowski, A., Ventura, A.P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprevote, O. and Seraphin, B. (2004) Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J*, **23**, 4847-4856.

E

Edwalds-Gilbert, G., Kim, D.H., Silverman, E. and Lin, R.J. (2004) Definition of a spliceosome interaction domain in yeast Prp2 ATPase. *RNA*, **10**, 210-220.

F

Fabrizio, P., Esser, S., Kastner, B. and Lührmann, R. (1994) Isolation of S. cerevisiae snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts. *Science*, **264**, 261-265.

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.

Fetzer, S., Lauber, J., Will, C.L. and Lührmann, R. (1997) The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase. *RNA*, **3**, 344-355.

Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340**, 245-246.

Fischer, U. and Lührmann, R. (1990) An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science*, **249**, 786-790.

Freund, C., Dotsch, V., Nishizawa, K., Reinherz, E.L. and Wagner, G. (1999) The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. *Nat Struct Biol*, **6**, 656-660.

Freund, C., Kuhne, R., Park, S., Thiemke, K., Reinherz, E.L. and Wagner, G. (2003) Structural investigations of a GYF domain covalently linked to a proline-rich peptide. *J Biomol NMR*, **27**, 143-149.

Freund, C., Kuhne, R., Yang, H., Park, S., Reinherz, E.L. and Wagner, G. (2002) Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *EMBO J*, **21**, 5985-5995.

Fribourg, S., Gatfield, D., Izaurralde, E. and Conti, E. (2003) A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat Struct Biol*, **10**, 433-439.

Fromont-Racine, M., Rain, J.C. and Legrain, P. (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet*, **16**, 277-282.

G

Galisson, F. and Legrain, P. (1993) The biochemical defects of prp4-1 and prp6-1 yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. *Nucleic Acids Res*, **21**, 1555-1562.

Gautier, T., Berges, T., Tollervey, D. and Hurt, E. (1997) Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol*, **17**, 7088-7098.

Gonzalez-Santos, J.M., Wang, A., Jones, J., Ushida, C., Liu, J. and Hu, J. (2002) Central region of the human splicing factor Hprp3p interacts with Hprp4p. *J Biol Chem*, **277**, 23764-23772.

Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Lührmann, R. and Fabrizio, P. (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J*, **18**, 4535-4548.

Gozani, O., Feld, R. and Reed, R. (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev*, **10**, 233-243.

Grainger, R.J. and Beggs J.D, (2003) Prp8 and the retinitis pigmentosa mutations in the yeast *Saccharromyces cerevisiae*. RNA 2003, 8th Annual Meeting of the RNA Society Vienna, Austria, pp357

H

Hall, S.L. and Padgett, R.A. (1996) Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science*, **271**, 1716-1718.

Hamm, J., Dathan, N.A., Scherly, D. and Mattaj, I.W. (1990) Multiple domains of U1 snRNA, including U1 specific protein binding sites, are required for splicing. *EMBO J*, **9**, 1237-1244.

Hermann, H., Fabrizio, P., Raker, V.A., Foulaki, K., Hornig, H., Brahms, H. and Lührmann, R. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J*, **14**, 2076-2088.

Horowitz, D.S., Kobayashi, R. and Krainer, A.R. (1997) A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA*, **3**, 1374-1387.

Horowitz, D.S., Lee, E.J., Mabon, S.A. and Misteli, T. (2002) A cyclophilin functions in pre-mRNA splicing. *EMBO J*, **21**, 470-480.

I
Igel, H., Wells, S., Perriman, R. and Ares, M., Jr. (1998) Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *RNA*, **4**, 1-10.

J
Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, **12**, 5-14.

K
Kim, D.H. and Rossi, J.J. (1999) The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. *RNA*, **5**, 959-971.

Kofler, M., Heuer, K., Zech, T. and Freund, C. (2004) Recognition sequences for the GYF domain reveal a possible spliceosomal function of CD2BP2. *J Biol Chem*, **279**, 28292-28297.

Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., Garcia-Blanco, M.A. and Manley, J.L. (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**, 119-124.

Konarska, M.M. and Sharp, P.A. (1986) Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell*, **46**, 845-855.

Konarska, M.M. and Sharp, P.A. (1987) Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell*, **49**, 763-774.

Konarska, M.M. and Sharp, P.A. (1988) Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. *Proc Natl Acad Sci U S A*, **85**, 5459-5462.

Krämer, A., Mulhauser, F., Wersig, C., Groning, K. and Bilbe, G. (1995) Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is homologous to the essential splicing factor PRP21p of Saccharomyces cerevisiae. *RNA*, **1**, 260-272.

Kuhn, A.N. and Brow, D.A. (2000) Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. *Genetics*, **155**, 1667-1682.

Kuhn, A.N., Li, Z. and Brow, D.A. (1999) Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol Cell*, **3**, 65-75.

Kuhn, A.N., Reichl, E.M. and Brow, D.A. (2002) Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. *Proc Natl Acad Sci U S A*, **99**, 9145-9149.

L

Laggerbauer, B., Achsel, T. and Lührmann, R. (1998) The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. *Proc Natl Acad Sci U S A*, **95**, 4188-4192.

Laggerbauer, B., Liu, S., Makarov, E., Vornlocher, H.P., Makarova, O., Ingelfinger, D., Achsel, T. and Lührmann R. (2005) The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation. *RNA*, **11**, 598-608.

Lauber, J., Fabrizio, P., Teigelkamp, S., Lane, W.S., Hartmann, E. and Lührmann, R. (1996) The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisiae are members of the DEXH-box protein family of putative RNA helicases. *EMBO J*, **15**, 4001-4015.

Lauber, J., Plessel, G., Prehn, S., Will, C.L., Fabrizio, P., Groning, K., Lane, W.S. and Lührmann, R. (1997) The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA*, **3**, 926-941.

Lerner, M.R. and Steitz, J.A. (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A*, **76**, 5495-5499.

Liu, Y., Li, J., Kim, B.O., Pace, B.S. and He, J.J. (2002) HIV-1 Tat protein-mediated transactivation of the HIV-1 long terminal repeat promoter is potentiated by a novel nuclear Tat-interacting protein of 110 kDa, Tip110. *J Biol Chem*, **277**, 23854-23863.

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.

Luo, H.R., Moreau, G.A., Levin, N. and Moore, M.J. (1999) The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA*, **5**, 893-908.

Lustig, A.J., Lin, R.J. and Abelson, J. (1986) The yeast RNA gene products are essential for mRNA splicing in vitro. *Cell*, **47**, 953-963.

M

Makarov, E.M., Makarova, O.V., Achsel, T. and Lührmann, R. (2000) The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions. *J Mol Biol*, **298**, 567-575.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M. and Lührmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, **298**, 2205-2208.

Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P. and Lührmann, R. (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J*, **21**, 1148-1157.

Makarova, O.V., Makarov, E.M. and Lührmann, R. (2001) The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *EMBO J*, **20**, 2553-2563.

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.

Mattaj, I.W. (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell*, **46**, 905-911.

Mattaj, I.W. and De Robertis, E.M. (1985) Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell*, **40**, 111-118.

Mattaj, I.W., Habets, W.J. and van Venrooij, W.J. (1986) Monospecific antibodies reveal details of U2 snRNP structure and interaction between U1 and U2 snRNPs. *EMBO J*, **5**, 997-1002.

McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E., Goliath, R., van Lith-Verhoeven, J.J., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P., Mackey, D.A., Bhattacharya, S.S., Bird, A.C., Markham, A.F. and Inglehearn, C.F. (2001) Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet*, **10**, 1555-1562.

McLean, M.R. and Rymond, B.C. (1998) Yeast pre-mRNA splicing requires a pair of U1 snRNP-associated tetratricopeptide repeat proteins. *Mol Cell Biol*, **18**, 353-360.

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.

Medenbach, J., Schreiner, S., Liu, S., Lührmann, R. and Bindereif, A. (2004) Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling. *Mol Cell Biol*, **24**, 7392-7401.

Meister, G., Hannus, S., Plottner, O., Baars, T., Hartmann, E., Fakan, S., Laggerbauer, B. and Fischer, U. (2001) SMNrp is an essential pre-mRNA splicing factor required for the formation of the mature spliceosome. *EMBO J*, **20**, 2304-2314.

Michel, F. and Jacquier, A. (1987) Long-range intron-exon and intron-intron pairings involved in self-splicing of class II catalytic introns. *Cold Spring Harb Symp Quant Biol*, **52**, 201-212.

N

Nagai, K. and Mattaj, I.W. (1994) RNA-Protein interactions (Nagai, K. and Mattaj, I.W., eds), pp. 150-177, Oxford University Press, Oxford

Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O. (1998) Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res*, **5**, 277-286.

Nelissen, R.L., Will, C.L., van Venrooij, W.J. and Lührmann, R. (1994) The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J*, **13**, 4113-4125.

Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A. and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet*, **20**, 46-50.

Nishikimi, A., Mukai, J., Kioka, N. and Yamada, M. (1999) A novel mammalian nuclear protein similar to Schizosaccharomyces pombe Prp1p/Zer1p and Saccharomyces cerevisiae Prp6p pre-mRNA splicing factors. *Biochim Biophys Acta*, **1435**, 147-152.

Nishizawa, K., Freund, C., Li, J., Wagner, G. and Reinherz, E.L. (1998) Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. *Proc Natl Acad Sci U S A*, **95**, 14897-14902.

Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayasi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K. and Tabata, S. (1994) Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res*, **1**, 27-35.

Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R. and Lührmann, R. (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J*, **18**, 6119-6133.

Nottrott, S., Urlaub, H. and Lührmann, R. (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J*, **21**, 5527-5538.

O

O'Keefe, R.T., Norman, C. and Newman, A.J. (1996) The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell*, **86**, 679-689.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432-438.

P

Padgett, R.A. and Shukla, G.C. (2002) A revised model for U4atac/U6atac snRNA base pairing. *RNA*, **8**, 125-128.

Patel, A.A. and Steitz, J.A. (2003) Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*, **4**, 960-970.

Ponting, C.P. (2000) Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. *Biochem J*, **351 Pt 2**, 527-535.

Preker, P.J. and Keller, W. (1998) The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem Sci*, **23**, 15-16.

Q

Query, C.C. and Konarska, M.M. (2004) Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol Cell*, **14**, 343-354.

Query, C.C., Strobel, S.A. and Sharp, P.A. (1996) Three recognition events at the branch-site adenine. *EMBO J*, **15**, 1392-1402.

R

Rader, S.D. and Guthrie, C. (2002) A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA*, **8**, 1378-1392.

Raghunathan, P.L. and Guthrie, C. (1998) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol*, **8**, 847-855.

Raker, V.A., Hartmuth, K., Kastner, B. and Lührmann, R. (1999) Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol Cell Biol*, **19**, 6554-6565.

Rappsilber, J., Ajuh, P., Lamond, A.I. and Mann, M. (2001) SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome. *J Biol Chem*, **276**, 31142-31150.

Reidt, U., Reuter, K., Achsel, T., Ingelfinger, D., Lührmann, R. and Ficner, R. (2000) Crystal structure of the human U4/U6 small nuclear ribonucleoprotein particle-specific SnuCyp-20, a nuclear cyclophilin. *J Biol Chem*, **275**, 7439-7442.

Reidt, U., Wahl, M.C., Fasshauer, D., Horowitz, D.S., Lührmann, R. and Ficner, R. (2003) Crystal structure of a complex between human spliceosomal cyclophilin H and a U4/U6 snRNP-60K peptide. *J Mol Biol*, **331**, 45-56.

Reuter, K. and Ficner, R. (1999) Overproduction, purification, crystallization and preliminary x-ray diffraction studies of the human spliceosomal protein U5-15kD. *Acta Crystallogr D Biol Crystallogr*, **55 (Pt 4)**, 888-890.

Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R. and Ficner, R. (1999) Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J Mol Biol*, **294**, 515-525.

Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J. and Konarska, M.M. (1999) The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA*, **5**, 167-179.

Rhode, B.M., Hartmuth, K., Urlaub, H. and Lührmann, R. (2003) Analysis of site-specific protein-RNA cross-links in isolated RNP complexes, combining affinity selection and mass spectrometry. *RNA*, **9**, 1542-1551.

Roscigno, R.F. and Garcia-Blanco, M.A. (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA*, **1**, 692-706.

Ruby, S.W. and Abelson, J. (1991) Pre-mRNA splicing in yeast. *Trends Genet*, **7**, 79-85.

S

Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T. and Lührmann, R. (2004) RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J*, **23**, 3000-3009.

Schneider, C., Will, C.L., Makarova, O.V., Makarov, E.M. and Lührmann, R. (2002) Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol Cell Biol*, **22**, 3219-3229.

Schneider, S., Campodonico, E. and Schwer, B. (2004) Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. *J Biol Chem*, **279**, 8617-8626.

Schwer, B. (2001) A new twist on RNA helicases: DExH/D box proteins as RNPases. *Nat Struct Biol*, **8**, 113-116.

Seraphin, B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J*, **14**, 2089-2098.

Sharp, P.A. (1987) Splicing of messenger RNA precursors. *Science*, **235**, 766-771.

Siatecka, M., Reyes, J.L. and Konarska, M.M. (1999) Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev*, **13**, 1983-1993.

Singh, R. and Reddy, R. (1989) Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci U S A*, **86**, 8280-8283.

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.

Staley, J.P. and Guthrie, C. (1999) An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell*, **3**, 55-64.

Stark, H., Dube, P., Lührmann, R. and Kastner, B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, **409**, 539-542.

Stevens, S.W. and Abelson, J. (1999) Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc Natl Acad Sci U S A*, **96**, 7226-7231.

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.

Szymczyna, B.R., Bowman, J., McCracken, S., Pineda-Lucena, A., Lu, Y., Cox, B., Lambermon, M., Graveley, B.R., Arrowsmith, C.H. and Blencowe, B.J. (2003) Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. *Genes Dev*, **17**, 461-475.

T

Tang, J., Abovich, N. and Rosbash, M. (1996) Identification and characterization of a yeast gene encoding the U2 small nuclear ribonucleoprotein particle B" protein. *Mol Cell Biol*, **16**, 2787-2795.

Tanner, N.K. and Linder, P. (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell*, **8**, 251-262.

Tarn, W.Y. and Steitz, J.A. (1996a) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell*, **84**, 801-811.

Tarn, W.Y. and Steitz, J.A. (1996b) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science*, **273**, 1824-1832.

Teigelkamp, S., Achsel, T., Mundt, C., Gothel, S.F., Cronshagen, U., Lane, W.S., Marahiel, M. and Lührmann, R. (1998) The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA*, **4**, 127-141.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C.L. and Lührmann, R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *RNA*, **3**, 1313-1326.

Teigelkamp, S., Newman, A.J. and Beggs, J.D. (1995) Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J*, **14**, 2602-2612.

Teigelkamp, S., Whittaker, E. and Beggs, J.D. (1995) Interaction of the yeast splicing factor PRP8 with substrate RNA during both steps of splicing. *Nucleic Acids Res*, **23**, 320-326.

Terns, M.P., Dahlberg, J.E. and Lund, E. (1993) Multiple cis-acting signals for export of pre-U1 snRNA from the nucleus. *Genes Dev*, **7**, 1898-1908.

Turner, I.A., Norman, C.M., Churcher, M.J. and Newman, A.J. (2004) Roles of the U5 snRNP in spliceosome dynamics and catalysis. *Biochem Soc Trans*, **32**, 928-931.

U

Umen, J.G. and Guthrie, C. (1995) Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA*, **1**, 584-597.

V

Valadkhan, S. and Manley, J.L. (2001) Splicing-related catalysis by protein-free snRNAs. *Nature*, **413**, 701-707.

Vankan, P., McGuigan, C. and Mattaj, I.W. (1990) Domains of U4 and U6 snRNAs required for snRNP assembly and splicing complementation in Xenopus oocytes. *EMBO J*, **9**, 3397-3404.

van Nues, R.W. and Beggs, J.D. (2001) Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of Saccharomyces cerevisiae. *Genetics*, **157**, 1451-1467.

van Venrooij, W.J. (1987) Autoantibodies against small nuclear ribonucleoprotein components. *J Rheumatol Suppl*, **14 Suppl 13**, 78-82.

Vidal, V.P., Verdone, L., Mayes, A.E. and Beggs, J.D. (1999) Characterization of U6 snRNA-protein interactions. *RNA*, **5**, 1470-1481.

Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R. and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell*, **6**, 1331-1342.

Villa, T., Pleiss, J.A. and Guthrie, C. (2002) Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell*, **109**, 149-152.

Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N.D., Willis, C., Moore, A.T., Bird, A.C., Hunt, D.M. and Bhattacharya, S.S. (2001) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell*, **8**, 375-381.

W

Wang, A., Forman-Kay, J., Luo, Y., Luo, M., Chow, Y.H., Plumb, J., Friesen, J.D., Tsui, L.C., Heng, H.H., Woolford, J.L., Jr. and Hu, J. (1997) Identification and characterization of human genes encoding Hprp3p and Hprp4p, interacting components of the spliceosome. *Hum Mol Genet*, **6**, 2117-2126.

Wang, Q. and Rymond, B.C. (2003) Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. *Mol Cell Biol*, **23**, 7339-7349.

Watkins, N.J., Dickmanns, A. and Lührmann, R. (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol*, **22**, 8342-8352.

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.

Weidenhammer, E.M., Ruiz-Noriega, M. and Woolford, J.L., Jr. (1997) Prp31p promotes the association of the U4/U6 x U5 tri-snRNP with prespliceosomes to form spliceosomes in Saccharomyces cerevisiae. *Mol Cell Biol*, **17**, 3580-3588.

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.

Will, C.L. and Lührmann, R. (1997) Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol*, **9**, 320-328.

Will, C.L. and Lührmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol*, **13**, 290-301.

Will, C.L., Schneider, C., Reed, R. and Lührmann, R. (1999) Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science*, **284**, 2003-2005.

Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M. and Lührmann, R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J*, **21**, 4978-4988.

Willer, M., Jermy, A.J., Young, B.P. and Stirling, C.J. (2003) Identification of novel protein-protein interactions at the cytosolic surface of the Sec63 complex in the yeast ER membrane. *Yeast*, **20**, 133-148.

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.

Y

Yean, S.L., Wuenschell, G., Termini, J. and Lin, R.J. (2000) Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature*, **408**, 881-884.

Z

Zhang, K., Smouse, D. and Perrimon, N. (1991) The crooked neck gene of Drosophila contains a motif found in a family of yeast cell cycle genes. *Genes Dev*, **5**, 1080-1091.

Zieve, G.W. and Sauterer, R.A. (1990) Cell biology of the snRNP particles. *Crit Rev Biochem Mol Biol*, **25**, 1-46.

8 Appendix

8.1 Abbreviations

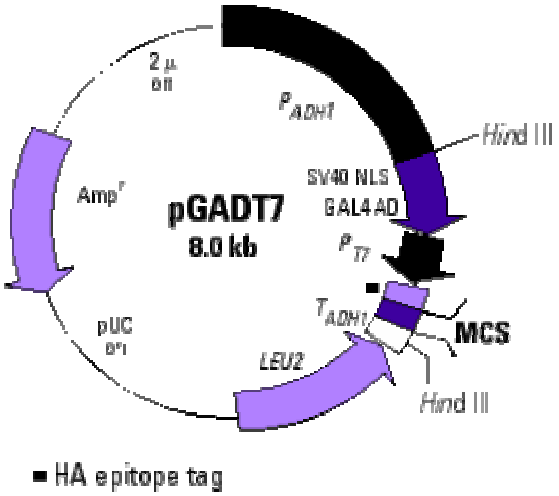
aa	amino acid
Ab	antibody
BLAST	Basic Local Alignment Search Tool
bp	base pair
β-gal	β-galactosidase
cDNA	complementary DNA
CSM	Complete Supplement Mixture
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DOB	Dropout Base medium
DOBA	Dropout Base with Agar
DTT	dithiothreitol
EDTA	ethylene-diamine-tetraacidic acid
EtBr	ethidium bromide
GST	glutathione *S*-transferase
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
kb	kilobases
kDa	kiloDalton
LB	Luria Bertani medium
MCS	multiple cloning site
mRNA	messenger RNA
MW	molecular weight
NP-40	Nonidet P-40
OD	optical density
Oligo	oligonucleotide
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
RBD	RNA-binding domain
rpm	revolutions per minute
RNA	ribonucleic acid
SD	Synthetic Defined medium
snRNP	small nuclear ribonucleoprotein particle
TBS	Tris-buffered saline
UV	ultraviolet
Y2H	yeast two-hybrid
YPD	yeast peptone dextrose medium

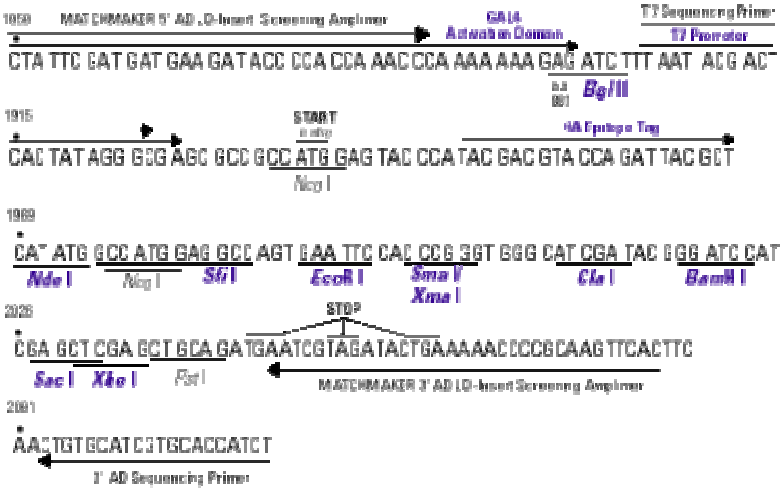
Amino acids

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

8.2 pGADT7 map and MCS

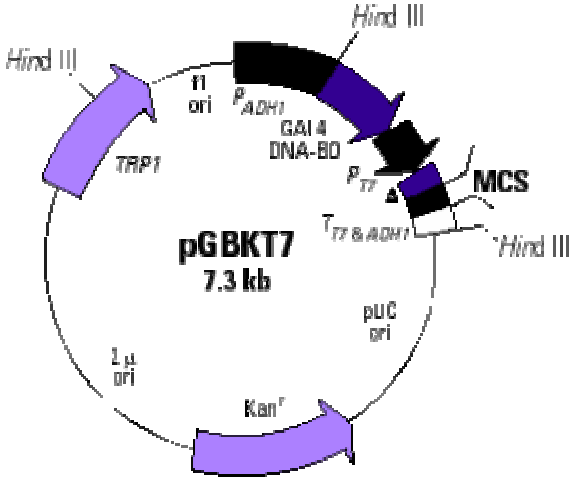
(from Clontech)





8.3 pGBKT7 map and MCS

(from Clontech)



8.4 cDNA sequence of U5-220K (ORF)

```
   1  ATGGCCGGAGTGTTTCCTTATCGAGGGCCGGGTAACCCGGTGCCTGGCCCTCTAGCCCCG
  61  CTACCGGACTACATGTCGGAGGAGAAGCTGCAGGAGAAAGCTCGAAAATGGCAGCAATTG
 121  CAGGCCAAGCGCTATGCAGAAAAGCGGAAGTTTGGGTTTGTGGATGCCCAGAAGGAAGAC
 181  ATGCCCCCAGAACATGTCAGGAAGATCATTCGAGACCATGGAGACATGACCAACAGGAAG
 241  TTCCGCCATGACAAAAGGGTTTACTTGGGTGCCCTAAAGTACATGCCCCACGCAGTCCTC
 301  AAACTCCTGGAGAACATGCCTATGCCTTGGGAGCAGATTCGGGATGTGCCTGTGCTGTAC
 361  CACATCACTGGAGCCATTTCCTTCGTCAATGAGATTCCCTGGGTCATTGAACCTGTCTAC
 421  ATCTCCCAGTGGGGGTCAATGTGGATTATGATGCGCCGAGAAAAAAGAGATAGGAGGCAT
 481  TTCAAGAGGATGCGTTTTCCCCCTTTTGATGATGAGGAGCCGCCCTTGGACTATGCTGAC
 541  AACATCCTAGATGTTGAGCCACTGGAGGCCATTCAGCTAGAGCTGGACCCTGAGGAGGAC
 601  GCCCCTGTGTTGGACTGGTTCTATGACCACCAGCCGTTGAGGGACAGCAGGAAGTATGTA
 661  AATGACTCCACTTACCAGCGCTGGCAGTTCACACTACCTATGATGTCGACTCTCTACCGC
 721  CTGGCTAATCAGCTCCTGACAGACTTGGTGGATGACAACTACTTCTACCTGTTTGATTTG
 781  AAGGCCTTCTTTACGTCCAAGGCACTCAATATGGCCATTCCTGGAGGCCCCAAATTTGAA
 841  CCTCTTGTTCGAGACATCAACCTACAGGATGAAGACTGGAATGAATTCAATGATATTAAC
 901  AAGATTATCATCCGGCAGCCTATCCGCACTGAGTACAAGATTGCTTTTCCTTACTTGTAC
 961  AACAATCTTCCACACCATGTCCACCTCACCTGGTACCATACTCCCAATGTTGTATTCATC
1021  AAAACTGAGGATCCTGACTTGCCAGCTTTCTACTTTGACCCTTTGATCAACCCAATCTCC
1081  CATAGGCACTCAGTCAAGAGCCAGGAACCATTGCCGGATGATGATGAGGAATTTGAGCTC
1141  CCGGAGTTTGTGGAGCCCTTCCTGAAGGACACACCCCTCTATACAGACAATACAGCCAAT
1201  GGCATTGCCCTGCTCTGGGCCCCGCGGCCCTTCAACCTACGCTCTGGTCGCACCCGTCGG
1261  GCCCTGGACATACCCCTTGTCAAGAACTGGTATCGGGAGCATTGTCCTGCCGGGCAGCCT
1321  GTGAAAGTGAGGGTCTCCTACCAGAAGCTGCTTAAGTACTATGTGCTGAATGCCCTGAAG
1381  CATCGGCCCCCTAAGGCTCAAAAGAAGAGGTATTTGTTCCGCTCCTTCAAAGCCACCAAA
1441  TTCTTTCAGTCCACAAAGCTGGACTGGGTGGAGGTTGGGCTCCAGGTTTGCCGCCAGGGC
1501  TACAACATGCTCAACCTTCTCATTCACCGCAAAAACCTCAACTACCTGCACCTGGACTAC
1561  AACTTCAACCTCAAGCCTGTGAAAACGCTCACCACCAAGGAAAGAAAGAAATCTCGTTTT
1621  GGGAATGCTTTCCACCTGTGTCGGGAAGTTCTGCGTTTGACTAAGCTGGTGGTGGATAGT
1681  CACGTGCAGTATCGGCTGGGCAATGTGGATGCCTTCCAGCTGGCAGATGGATTGCAGTAT
1741  ATATTTGCCCATGTTGGGCAGTTGACGGGCATGTATCGATACAAATACAAGCTGATGCGA
1801  CAGATTCGCATGTGCAAGGACCTGAAGCATCTCATCTATTATCGTTTCAACACGGGCCCT
1861  GTAGGGAAGGGTCCTGGCTGTGGCTTCTGGGCTGCCGGTTGGCGAGTCTGGCTCTTTTTC
1921  ATGCGTGGCATTACCCCTTTATTAGAGCGATGGCTTGGCAACCTCCTGGCCCGGCAGTTT
1981  GAAGGTCGACACTCAAAGGGGGTGGCAAAGACAGTAACAAAGCAGCGAGTGGAGTCACAT
2041  TTTGACCTTGAGCTGCGGGCAGCTGTGATGCATGATATTCTGGACATGATGCCTGAGGGG
2101  ATCAAGCAGAACAAGGCCCGGACAATCCTGCAGCACCTCAGTGAAGCCTGGCGCTGCTGG
2161  AAAGCCAACATTCCCTGGAAGGTCCCTGGGCTGCCGACGCCCATAGAGAATATGATCCTT
2221  CGATACGTGAAGGCCAAGGCTGACTGGTGGACCAACACTGCCCACTACAACCGAGAACGG
2281  ATCCGCCGAGGGGCCACTGTGGACAAGACTGTTTGTAAAAAGAATCTGGGCCGCCTCACC
2341  CGGCTCTATCTGAAGGCAGAACAGGAGCGGCAGCACAACTACCTGAAGGACGGGCCTTAC
2401  ATCACAGCGGAGGAAGCAGTGGCAGTATATACCACCACAGTGCATTGGTTGGAAAGCCGC
2461  AGGTTTTCACCCATCCCATTCCCCCCACTCTCCTATAAGCATGACACCAAGTTGCTCATC
2521  TTGGCATTGGAGCGGCTCAAGGAAGCTTATAGTGTGAAGTCTCGGTTGAACCAGTCTCAG
2581  AGGGAGGAGCTAGGTCTGATCGAGCAGGCCTACGATAACCCCCACGAGGCGCTGTCCCGC
2641  ATCAAGCGTCACCTCCTCACACAGAGAGCCTTCAAAGAGGTGGGCATTGAGTTCATGGAT
2701  CTGTATAGCCACCTCGTTCCAGTATATGATGTTGAGCCCCTGGAGAAGATAACTGATGCT
2761  TACCTGGACCAGTACCTGTGGTATGAAGCCGACAAGCGCCGCCTGTTCCCACCCTGGATT
2821  AAGCCTGCAGACACAGAACCACCTCCGCTGCTTGTTTACAAGTGGTGTCAAGGCATCAAT
2881  AACCTGCAGGACGTGTGGGAGACGAGTGAAGGCGAGTGCAATGTCATGCTGGAATCCCGC
2941  TTTGAGAAGATGTATGAGAAGATCGACTTGACTCTGCTCAACAGGCTGCTGCGCCTCATC
3001  GTGGACCACAACATAGCCGACTACATGACAGCCAAGAACAACGTCGTCATCAACTATAAG
3061  GACATGAACCATACGAATTCATATGGGATCATCAGAGGCCTGCAGTTTGCCTCATTCATC
3121  GTGCAGTATTATGGCCTGGTGATGGATTTGCTTGTATTGGGATTGCACCGGGCCAGTGAG
3181  ATGGCTGGGCCCCCTCAGATGCCAAATGACTTTCTCAGTTTCCAGGACATAGCCACTGAG
3241  GCTGCCCACCCCATCCGTCTCTTCTGCAGATACATTGATCGCATCCATATTTTTTTCAGG
3301  TTCACAGCAGATGAGGCTCGGGACCTGATTCAACGTTACCTGACAGAGCACCCTGACCCC
3361  AATAATGAAAACATCGTTGGCTATAATAACAAGAAGTGCTGGCCCCGAGATGCCCGCATG
3421  CGCCTCATGAAACATGATGTTAACTTAGGCCGGGCGGTATTCTGGGACATCAAGAACCGC
3481  TTGCCACGGTCAGTGACTACAGTTCAGTGGGAGAACAGCTTCGTGTCTGTGTACAGTAAG
3541  GACAACCCCAACCTGCTGTTCAACATGTGTGGCTTCGAGTGCCGCATCCTGCCTAAGTGC
```

```
3601  CGCACCAGCTATGAGGAGTTCACCCACAAGGACGGGGTCTGGAACCTGCAGAATGAGGTT
3661  ACTAAGGAGCGCACAGCTCAGTGTTTCCTGCGTGTGGACGATGAGTCAATGCAGCGCTTC
3721  CACAACCGCGTGCGTCAGATTCTCATGGCCTCTGGGTCCACCACCTTCACCAAGATTGTG
3781  AATAAGTGGAATACAGCTCTCATTGGCCTTATGACATACTTTCGGGAGGCTGTGGTGAAC
3841  ACCCAAGAGCTCTTGGACTTACTGGTGAAGTGTGAGAACAAAATCCAGACACGTATCAAG
3901  ATTGGACTCAACTCCAAGATGCCAAGTCGGTTCCCCCCGGTTGTGTTCTACACCCCTAAG
3961  GAGTTGGGTGGACTCGGCATGCTCTCAATGGGCCATGTGCTCATCCCCCAATCCGACCTC
4021  AGGTGGTCCAAACAGACAGATGTAGGTATCACACACTTTCGTTCAGGAATGAGCCATGAA
4081  GAAGACCAGCTCATTCCCAACTTGTACCGCTACATACAGCCATGGGAGAGCGAGTTCATT
4141  GATTCTCAGCGGGTCTGGGCTGAGTACGCACTCAAGAGGCAAGAGGCCATTGCTCAGAAC
4201  AGACGCCTGACTTTAGAAGACCTAGAAGATTCATGGGATCGTGGCATTCCTCGAATCAAT
4261  ACCCTCTTCCAGAAGGACCGGCACACACTGGCTTATGATAAGGGCTGGCGTGTCAGAACT
4321  GACTTTAAGCAGTATCAGGTTTTGAAGCAGAATCCGTTCTGGTGGACACACCAGCGGCAT
4381  GATGGGAAGCTCTGGAACCTGAACAACTACCGTACAGACATGATCCAGGCCCTGGGCGGT
4441  GTGGAAGGCATTCTGGAACACACACTCTTTAAGGGCACTTACTTCCCTACCTGGGAGGGG
4501  CTTTTCTGGGAGAAGGCCAGTGGCTTTGAGGAATCTATGAAGTGGAAGAAGCTAACTAAT
4561  GCTCAGCGATCAGGACTGAACCAGATTCCCAATCGTAGATTCACCCTCTGGTGGTCCCCG
4621  ACCATTAATCGAGCCAATGTATATGTAGGCTTTCAGGTGCAGCTAGACCTGACGGGTATC
4681  TTCATGCACGGCAAGATCCCCACGCTGAAGATCTCTCTCATCCAGATCTTCCGAGCTCAC
4741  TTGTGGCAGAAGATCCATGAGAGCATTGTTATGGACTTATGTCAGGTGTTTGACCAGGAA
4801  CTTGATGCACTGGAAATTGAGACAGTACAAAAGGAGACAATCCATCCCCGAAAGTCATAT
4861  AAGATGAACTCTTCCTGTGCAGATATCCTGCTCTTTGCCTCCTATAAGTGGAATGTCTCC
4921  CGGCCCTCATTGCTGGCTGACTCCAAGGATGTGATGGACAGCACCACCACCCAGAAATAC
4981  TGGATTGACATCCAGTTGCGCTGGGGGGACTATGATTCCCACGACATTGAGCGCTACGCC
5041  CGGGCCAAGTTCCTGGACTACACCACCGACAACATGAGTATCTACCCTTCGCCCACAGGT
5101  GTACTCATCGCCATTGACCTGGCCTATAACTTGCACAGTGCCTATGGAAACTGGTTCCCA
5161  GGCAGCAAGCCTCTCATACAACAGGCCATGGCCAAGATCATGAAGGCAAACCCTGCCCTG
5221  TATGTGTTACGTGAACGGATCCGCAAGGGGCTACAGCTCTATTCATCTGAACCCACTGAG
5281  CCTTATTTGTCTTCTCAGAACTATGGTGAGCTCTTCTCCAACCAGATTATCTGGTTTGTG
5341  GATGACACCAACGTCTACAGAGTGACTATTCACAAGACCTTTGAAGGGAACTTGACAACC
5401  AAGCCCATCAACGGAGCCATCTTCATCTTCAACCCACGCACAGGGCAGCTGTTCCTCAAG
5461  ATAATCCACACGTCCGTGTGGGCGGGACAGAAGCGTTTGGGGCAGTTGGCTAAGTGGAAG
5521  ACAGCTGAGGAGGTGGCCGCCCTGATCCGATCTCTGCCTGTGGAGGAGCAGCCCAAGCAG
5581  ATCATTGTCACCAGGAAGGGCATGCTGGACCCACTGGAGGTGCACTTACTGGACTTCCCC
5641  AATATTGTCATCAAAGGATCGGAGCTCCAACTCCCTTTCCAGGCGTGTCTCAAGGTGGAA
5701  AAATTCGGGGATCTCATCCTTAAAGCCACTGAGCCCCAGATGGTTCTCTTCAACCTCTAT
5761  GACGACTGGCTCAAGACTATTTCATCTTACACGGCCTTCTCCCGTCTCATCCTGATTCTG
5821  CGTGCCCTACATGTGAACAACGATCGGGCAAAAGTGATCCTGAAGCCAGACAAGACTACT
5881  ATTACAGAACCACACCACATCTGGCCCACTCTGACTGACGAAGAATGGATCAAGGTCGAG
5941  GTGCAGCTCAAGGATCTGATCTTGGCTGACTACGGCAAGAAAAACAATGTGAACGTGGCA
6001  TCACTGACACAATCAGAAATTCGAGACATCATCCTGGGTATGGAGATCTCGGCACCGTCA
6061  CAGCAGCGGCAGCAGATCGCTGAGATCGAGAAGCAGACCAAGGAACAATCGCAGCTGACG
6121  GCAACACAGACTCGCACTGTCAACAAGCATGGCGATGAGATCATCACCTCCACCACCAGC
6181  AACTATGAGACCCAGACTTTCTCATCCAAGACTGAGTGGAGGGTCAGGGCCATCTCTGCT
6241  GCCAACCTGCACCTAAGGACCAATCACATCTATGTTTCATCTGACGACATCAAGGAGACT
6301  GGCTACACCTACATCCTTCCCAAGAATGTGCTTAAGAAGTTCATCTGCATATCTGACCTT
6361  CGGGCCCAAATTGCAGGATACCTATATGGGGTGAGCCCACCAGATAACCCCCAGGTGAAG
6421  GAGATCCGCTGCATTGTGATGGTGCCGCAGTGGGGCACTCACCAGACCGTGCACCTGCCT
6481  GGCCAGCTGCCCCAGCATGAGTACCTCAAGGAGATGGAACCCTTAGGTTGGATCCACACT
6541  CAGCCCAATGAGTCCCCGCAGTTATCACCCCAGGATGTCACCACCCATGCCAAGATCATG
6601  GCTGACAACCCATCTTGGGATGGCGAGAAGACCATTATCATCACATGCAGCTTCACGCCA
6661  GGCTCCTGTACACTGACGGCCTACAAGCTGACCCCCAGTGGCTACGAATGGGGCCGCCAG
6721  AACACAGACAAGGGCAACAACCCCAAGGGCTACCTGCCTTCACACTATGAGAGGGTGCAG
6781  ATGCTGCTGTCGGACCGTTTCCTTGGCTTCTTCATGGTCCCTGCCCAGTCCTCGTGGAAC
6841  TACAACTTCATGGGTGTTCGGCATGACCCCAACATGAAATATGAGCTACAGCTGGCGAAC
6901  CCCAAAGAGTTCTACCACGAGGTGCACAGGCCCTCTCACTTCCTCAACTTTGCTCTCCTG
6961  CAGGAGGGGGAGGTTTACTCTGCGGATCGGGAGGACCTGTATGCCTGA
```

8.5 Curriculum vitae

PERSONAL DATA

First name	Sunbin
Family name	Liu
Date of birth	20. 01. 1963
Nationality	Chinese
Marital status	Married, with one child
Address	Immanuel-Kant-Str. 34, 37083 Göttingen

EDUCATION

1969 – 1975	Long-Bei Primary School, Pingtan, Fujian
1975 – 1977	Long-Nan School, Pingtan, Fujian
1977 – 1980	Pingtan No. 1 High School, Fujian, graduated in July 1980
1980 – 1984	Faculty of Biology at the University of Xiamen, Fujian Received B.Sc. in July 1984
1999 – 2005	PhD work under the direction of Prof. Reinhard Lührmann in the Institute of Molecular Biology and Tumor Research at the University of Marburg (1999-2000) and in the Department of Cellular Biochemistry in the Max-Planck-Institute for Biophysical Chemistry in Göttingen

WORKING EXPERIENCE

04. 1997 – 03. 1999	Laboratory of Endocrinology, Clinic of Großhadern, University of München *Gene expression of the IGF-I receptor, IGF-II and IGF-binding protein in human colon carcinomas*
04. 1996 – 03. 1997	Department of Molecular Biology, Max-Planck-Institute for Biochemistry in Martinsried bei München. *Functional study of phosphatase*
01.1996 – 03. 1996	Department of Endocrinology & Reproduction, Erasmus University Rotterdam in Holland

(I received a fellowship from the World Health Organization in Geneva)
Effect of FSH on the expression of leucine-rich primary response gene in primate testis

03. 1995 – 12. 1995 Institute for Reproductive Medicine at the University of Münster
(I received a fellowship from the World Health Organization in Geneva)
Gene expression of rat LHR and androgen receptor

08. 1984 – 03. 1995 Department of Human Genetics in the National Research Institute for Family Planning in Beijing
(World Health Organization Collaborating Centre for Research in Human Reproduction)
Environment and Genetics; Chromosome analysis; Male contraception; Spermatogenesis

PUBLICATIONS DURING MY PHD WORK

Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P. and Lührmann, R. (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6.U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J*, 21, 1148-1157.

Medenbach, J., Schreiner, S., Liu, S., Lührmann, R. and Bindereif, A. (2004) Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling. *Mol Cell Biol*, 24, 7392-7401.

Laggerbauer, B., Liu, S., Makarov, E., Vornlocher, H.P., Makarova, O., Ingelfinger, D., Achsel, T. and Lührmann R. (2005) The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation. *RNA*, 11, 598-608.

Nielsen, T.K., Liu, S., Lührmann, R. and Ficner, R. (2005) Structural basis of GYF-domain bifunctionality: the interaction between the proteins 15K and 52K of the spliceosomal U5 snRNP. In preparation.

Liu, S., Vornlocher, H.P., Rauhut, R. and Lührmann, R. (2005) A network of protein interactions in the human U4/U6.U5 tri-snRNP. In preparation.