STUDIES OF PROTEINS INVOLVED IN PRE-mRNA SPLICING

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Olga Makarova MSc (Leningrad)

Department of Biochemistry

University of Leicester

September 1999

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I dedicate this work to my family.

…the only thing we require to be good philosophers is the faculty of wonder…

Jostein Gaarder

Abstract

Studies of proteins involved in pre-mRNA splicing.

Olga V. Makarova
1999

Human SF2/ASF and hnRNP A1 are splicing factors that modulate alternative splicing in antagonistic manner: hnRNP A1 activates distal 5' splice sites and promotes alternative exon skipping, whereas SF2/ASF activates proximal 5' splice sites and promotes exon inclusion. The possibility that antagonism is a result of the specific RNA binding properties of these proteins was investigated *in vitro* using recombinant proteins and pre-mRNA. The binding of SF2/ASF and hnRNP A1 to RNA was compared by assays based on RNase H, nitrocellulose filter binding, and UV-crosslinking. HnRNP A1 dissociated from RNA within 5 seconds whereas the dissociation rate of SF2/ASF was much slower. The main determinant of hnRNP A1 binding was co-operativity but, in contrast, SF2/ASF may discriminate between different binding sites on the pre-mRNA. Close thermodynamic characteristics and very different kinetic parameters of binding were rationalised in the model that explains how the competitive binding of SF2/ASF and hnRNP A1 determines splice site recognition.

Human Y chromosome-encoded RBM genes were identified on the basis of genetic analysis of infertile males. The structural features of RBM (RNA binding domain and SRGY-box repeats) evoked the possibility that it might function in splicing that was addressed here. Using antibodies, the expression of RBM was detected in human testis. Transient transfection of HeLa cells and immunofluorescence demonstrated a nuclear localisation of RBM with a pattern characteristic of splicing factors. Overexpression of RBM was harmful for HeLa and COS-1 cells, and it may induce appoptosis.

For the first time, human U4/U6 snRNP particles were isolated that are able to associate with U5 snRNPs into [U4/U6.U5] tri-snRNP particles. A novel 61kD protein was identified as an integral component of the U4/U6 snRNP, and it appeared to be an orthologue of yeast splicing factor Prp31p. The U5-102kD protein, an orthologue of yeast Prp6p, was characterised, and it is likely to be involved in the interaction between U5 and U4/U6 particles through its multiple tetratrico peptide repeats.

ACKNOWLEDGEMENTS

I am very grateful to my supervisor Dr. Ian Eperon who introduced to me the fascinating field of pre-mRNA splicing. I would like to thank him for his help and advice during my work at the bench and writing this manuscript.

I would also like to thank Dr. Reinhard Lührmann for giving me an opportunity to continue the work in splicing field, for his guidance and encourage.

I also thank the members of both laboratories for creating the friendly and supportive atmosphere. I shall always remember Katherine Saunders, Neil Thornton, George Butler, and Claire Taylor for our outdoor adventures and amateur musical entertainment. Thanks to Shern Chew, Julian Venables, and Lysa Baginsky for the comprehensive discussions and spiritual atmosphere in the lab.

I am thankful to all the members of AG Lührmann. Particularly, I would like to thank Cindy Will for tolerating me as a half-person in Lab 224 and my room-mates: Hans-Peter Vornlocher, Irene Öchsner, Axel Badouin, and Wini Lorenz who have been foolhardy enough to share with me the equipment and space.

In the life before "splicing" I was influenced by a number of people to whom I am glad to express my acknowledgement. First, there are my tutors of the Leningrad high school No 239 who gave me a taste of intellectual adventure and exploration. I am grateful to my first supervisor, Sergey Bulat, at the department of Biophysics, Leningrad Nuclear Physics Institute, for teaching me the basic skills in biochemistry and molecular biology during a study of the species-specific DNA repeats in fungi *Yarrowia lipolytica* and *Cochliobolus sativus* (1987-1990).

In America, I would like to thank the members of laboratory of Dr. Morton Bradbury, UC Davis, where I was working on chromatin structure in 1991-1993, and particularly, Peter Yau for introduction to mega-scale biochemistry.

I appreciate the help and guidance of Marc Dreyfus, Ecole Normale Superiore, Paris, who gave me the first insight into the RNA World during my work in his lab in 1993-1994.

I cannot imagine how my life would be without the unfailing love of my parents, the

constant criticism and challenge of my husband to whom I am obliged for travelling around the World, and my dear daughter with her strong character. I am really grateful to my family for their love, appreciation, and support in the course of my research work and during the writing of this manuscript.

Abbreviations

AMP	adenosine 5'-monophosphate
APS	ammonium peroxodisulfate
ATP	adenosine 5'-triphosphate
AZF	azoospermia factor
bGH	bovine growth hormone
BSA	bovine serum albumin
CDC	*cell cycle defect* genes
cDNA	complementary DNA
CMV	cytomegalovirus
cpm	counts per minute
CREM	cAMP-responsible element modulator
CTLs	cytotoxic T lymphocytes
Cy3	cyanine 3 (fluorochrome)
DAPI	4', 6'-diamidino-2-phenylindole
dATP	2'-deoxyadenosine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
DEAE cellulose	diethylaminoethyl cellulose
DNA	deoxyribonucleic acid
DMSO	dimethyl sulfoxide
dsx	*doublesex* gene
DTT	dithiotreitol
DTE	1,4 - dithioerythritol
ECL	enhanced chemiluminescence
EDTA	ethylenediamine-tetraacetic acid
EMSA	electrophoretic mobility shift assay

ESE	exon splicing enhancer
ESS	exon splicing silencer
EST	expressed sequence tag
FSH	follicular stimulating hormone
FITC	fluorescein isothiocyanate
GpppG	diguanosine triphosphate [G(5')ppp(5')G]
GS-MS	gas chromatography-mass spectrometry
GST	glutathione S-transferase
GTP	guanosine triphosphate
HEPES	hydroxyethyl-piperazine-ethanesulfonic acid
hnRNA	heterogeneous nuclear RNA
hnRNP	heterogeneous nuclear ribonucleoprotein
HPLC	high-performance liquid chromatography
Ig	immunoglobulin
IPTG	β-D-isopropyl-thiogalactopyranoside
IVS	intervening sequence
KH	hnRNP K homology
LSm	like Sm
LSCM	laser-scanning confocal microscope
m_3G	2,2,7-trimethylguanosine
m^7G	7-methylguanosine
MALDI-MS	matrix-assisted laser desorption/ionization mass spectrometry
mRNA	messenger RNA
NaSCN	sodium thiocyanate
NCFBA	nitrocellulose filter binding assay
NMR	nuclear magnetic resonance
NP40	nonylphenylpolyethylene gycol P40
nt	nucleotide(s)
ORF	open reading frame

PAS	protein A sepharose
PAGE	polyacrylamide gel electrophoresis
PCD	Programmed Cell Death
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
PRP	*pre-mRNA processing* genes
PSI	P-element somatic inhibitor (protein)
RBM	RNA-binding motif (protein)
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein
RRM	RNA-recognition motif
RS domain	arginine / serine-rich domain
SC35	spliceosome complex protein 35
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SELEX	systematic evolution of ligands by exponential enrichment
SF	splicing factor
SF2/ASF	splicing factor 2 / alternative splicing factor
Sm	Smith autoantigen
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SR protein	serin / arginine-rich protein
SRGY box	serin / arginine / glycine / tytosine box
Sxl	*Sex lethal* gene
TEMED	N, N, N', N'-tetramethylethylenediamine
TPR	tetratrico peptide repeat
tra	*transformer* gene
Trx	thioredoxin

tRNA	transfer ribonucleic acid
u	unit
UP1	unwinding protein 1
U2AF	U2 auxiliary factor
UTP	uridine 5'-triphophate
WT1	Wilms' Tumor suppressor gene
YRRM	Y chromosome RNA recognition motif (protein)

CONTENTS

CHAPTER 1

A general overview of pre-mRNA splicing

1.1. Split genes of eukaryotes

Most mammalian genes have a split gene organisation. Regions not represented in the mRNA were named introns or intervening sequences. Segments that do specify the mRNA were named exons (because they should be expressed). Exons include not only the protein coding sequences of the gene, but also the untranslated 5' leader and 3' trailer sequences present in mRNA. The poly(A) tail of mRNA is not encoded in the genome but added post-transcriptionally. Introns can be of many different lengths and sequences. They need not reside neatly between codons, often they split a triplet, so that the information needed to specify a particular amino acid lies thousands of base pairs apart on the chromosomal DNA. Unlike introns, exons are usually small, about 100 nt. However, if the nucleotide sequences of the mouse, rabbit, and human β-globin genes are compared in detail, little divergence is seen in their protein-coding regions, reflecting the evolutionary conservation of the β-globin polypeptide. The introns, by contrast, differ between species in both length and exact sequence. Yet, they occur at precisely the same positions within the β-globin-coding region of each species. In a few cases, the distribution of the residues between exons and introns is quite extraordinary. For instance, the chromosomal gene for mouse dihydrofolate reductase extends over about 31,000 base pairs, but it gives rise to an mRNA only about 1600 nucleotides, which in turn possesses only 558 coding residues. The number of introns per pre-mRNA can be large: the chicken collagen gene contains at least 50 exons, which divide its coding region into regular repeating units. However, some protein-coding genes are not interrupted. The examples are the multiple genes encoding the histone and interferon proteins, which only rarely have introns. In general, as one proceeds down the evolutionary ladder, fewer and fewer protein-coding genes are interrupted, and introns that do exist tend to be shorter. For instance, in the yeast *Saccharomyces cerevisiae*, a gene containing introns is the exception rather than the rule.

1.2. Splicing system

RNA splicing is a mechanism responsible for removing introns by RNA processing. This mechanism is exquisitely precise: if intron removal was off by even one nucleotide, then missense or frameshifted proteins would result. Usually, exons are not shuffled and their order in the mRNA is exactly the same as that in the gene from which it is derived.

The availability of RNA precursor molecules containing introns challenged biochemists to ask whether accurate splicing could be carried out in the test tube. From such studies, much has been learned about the biochemical mechanisms that achieve precise cutting at the splice junctions and the subsequent ligation of exon sequences. An understanding of the structural features required for splicing RNA precursor molecules has greatly evolved for the last 20 years. The comparison of many DNA sequences from split genes of many eukaryotic organisms has allowed the identification of common sequence elements. Both naturally occurring and artificially-induced mutations that affect splicing have been systematically scrutinised. These mutant analyses have confirmed original guesses about which sequences are necessary, have uncovered other sequences, and have raised additional questions about splicing mechanisms.

Signals crucial for the splicing of mammalian pre-mRNAs reside directly at the splice junctions themselves. Most of the introns begin with GU and end with AG (Breathnach *et al.*, 1978). The GU-AG rule is nearly always obeyed and even longer consensus sequences can be written for both 5' and 3' splice junctions. The eight-nucleotide consensus for 5' junctions extends two residues upstream and six residues downstream from the splice point (Fig.1.1). The consensus sequence for 3' splice junctions is composed of a pyrimidine-rich region of variable length (but usually greater than ten nucleotides) followed by a short consensus sequence extending only three residues upstream and one residue downstream from the splice point (Fig.1.1). Remarkably, splice site sequences from genes representing the whole range of eucaryotic species (yeast to human) conform well to the same 5' and 3' consensus sequences. This suggests that these splicing signals have been highly conserved over evolutionary time.

Recently, a minor class of introns with different splice junctions has been identified. These introns have AU-AC termini, breaking the almost universal GU-AG rule (Jackson,

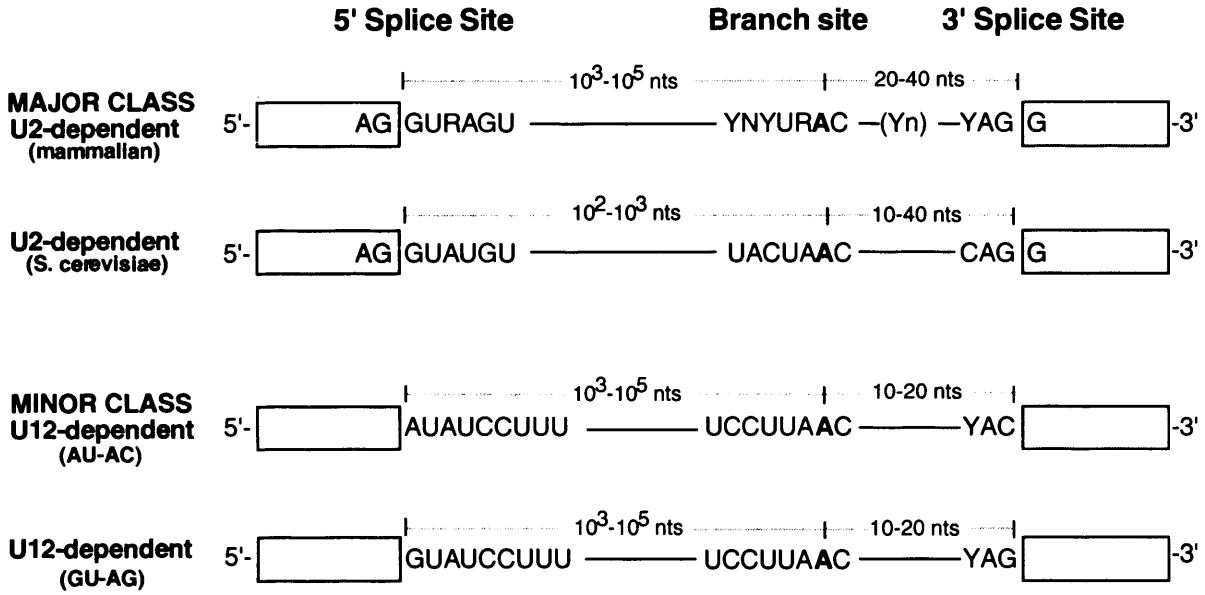


Figure 1.1. The conserved sequence elements recognised in a process of splicing. The exon sequences are boxed and introns are presented as a line. The distances between conserved elements are indicated at the top. Nucleotides (R - purine, Y - pyrimidine) that are conserved are shown in letters. (Adapted from Will and Lührmann, 1999)

1991). Members of this novel group of introns have strongly conserved 5' and 3' splice site sequences. To date, the evolutionary origins of the two types of intron remain unclear. It would not be unexpected if, when more genome sequence data becomes available, a third subclass of introns were discovered (Burge *et al.*, 1999).

Investigations of both cell free and *in vivo* splicing systems established that intron excision occurs via two transesterification reactions (Fig.1.2). A first step is a cut at the 5' splice junction accompanied by formation of an unusual 2'-5' phosphodiester bond between the 5'-terminal G residue of the intron and a site about 30 nucleotides upstream from the 3' end of the intron. This creates an RNA lariat structure in which an A residue invariably forms the branch point. The branched A nucleotide engages in three phosphodiester bonds: the two ordinary 3'-5' bonds plus a novel 2'-5' bond. In mammalian pre-mRNA, the sequence surrounding this A residue is only very loosely defined (YNYURAC, where Y is pyrimidine, N is any nucleotide and R is purine), and intron sequences can be deleted nearly at random. In yeast, on the other hand, the branch always occurs at the last A within a particular sequence (UACUAAC), which appears in every intron 20-40 nucleotides upstream of the 3' splice site. Deletion of this yeast sequence or certain mutations within it abolish excision of that intron. As a splicing reaction proceeds to the second step, the liberated 3'-OH group of the 5' exon forms a phosphodiester bond with the 5'-terminal phosphate of the 3' exon yielding the spliced product. The intron is thereby eliminated in its lariat form and, *in vivo*, is rapidly degraded. It should be noted that the transesterification reactions preserve the free energy of each cleaved phosphodiester bond through the concomitant formation of a new one.

1.3. Spliceosome and snRNPs

Splicing occurs in a large and dynamic ribonucleoprotein complex, the spliceosome. This 50-60S particle has a calculated molecular weight on the order of $3\text{-}5\text{x}10^6$ kDa, and is estimated to be 40-60 nm in diameter (Moore *et al.*, 1993). Five small nuclear RNAs (U1, U2, U4, U5 and U6 snRNAs) constitute key components of this machine. Packaged by proteins into ribonucleoproteins (snRNPs), these snRNAs assemble onto the intron of the pre-mRNA in an ordered, step-wise fashion. A conspicuous feature of this conserved assembly pathway is that many steps require the hydrolysis of ATP. Figure 1.3 summarises some of the

Figure 1.2

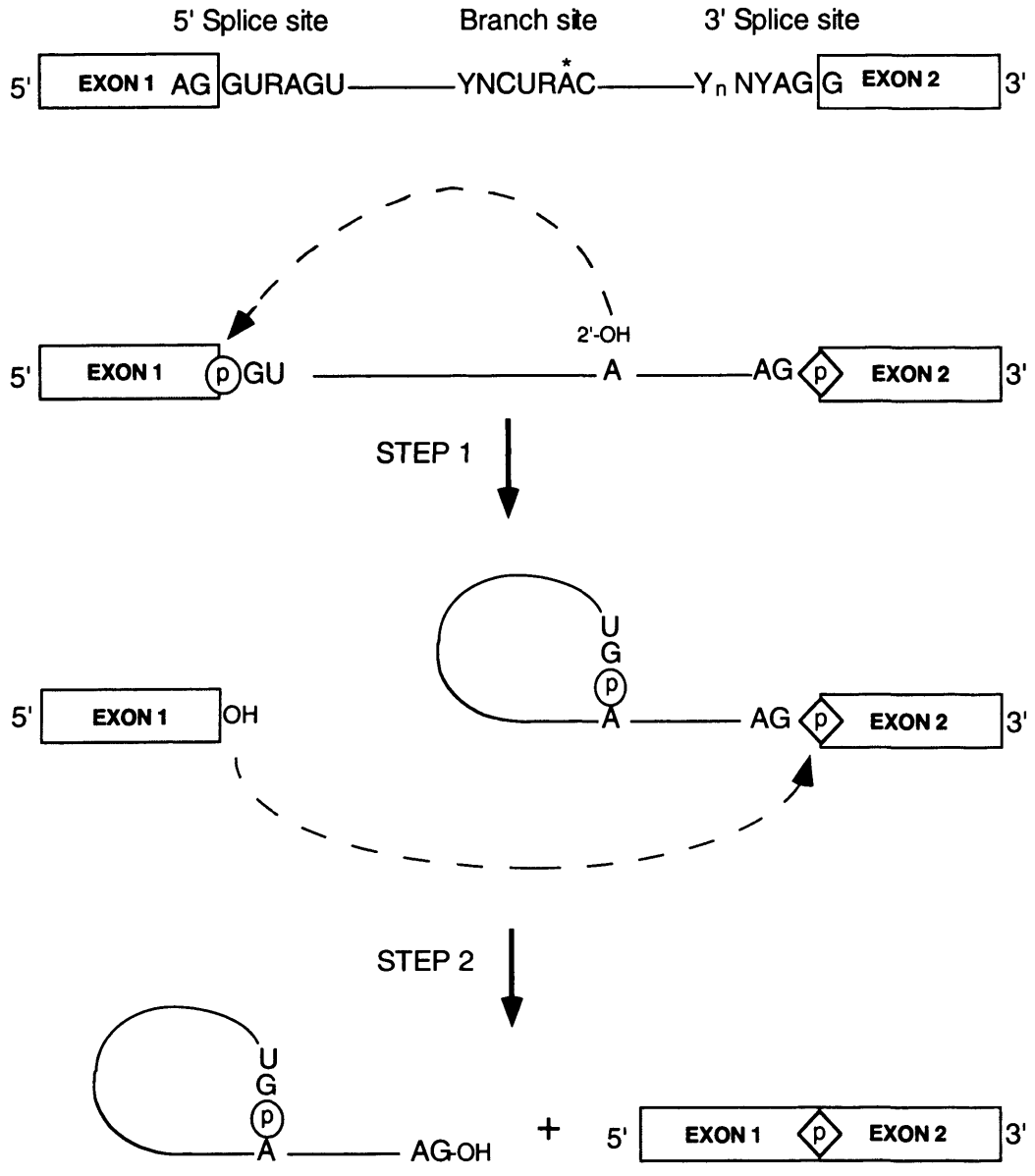


Figure 1.2. Two-step chemical mechanism for pre-mRNA splicing. Exons are represented by rectangles and an intron shown as a line. The phospodiester linkages that are broken or formed during the reaction are represented by the letter **p**. The branch adenosine (A), the intron termini (GU-AG), and 2' and 3' hydroxyl groups (OH) are also indicated.

more prominent events of this process, which include the initial binding of U1 and U2 snRNPs to the pre-mRNA, the binding of the tripartite [U4/U6.U5] snRNP, and the displacement of U4 snRNP just prior to two chemical steps of splicing. The general spliceosome assembly pathway is highly conserved in both metazoans and yeast, although some fundamental differences exist between individual proteins involved.

Recently, it has been shown that a minor class of introns is spliced by a novel type of spliceosome involving U11, U12, U4atac, U6atac and U5 snRNAs. The U11 and U12 snRNAs are the functional analogues of U1 and U2 snRNAs, whereas U4atac and U6atac are the analogues of U4 and U6 snRNAs. The discovery of the novel spliceosome has changed the classification of the introns. Introns are distinguished now on the basis of which spliceosome is responsible for their removal. Thus, U2-dependent introns are removed by the canonical spliceosome, and U12-dependent introns are excised by the novel spliceosome. The extent of the phylogenetic distribution of U12-type introns and spliceosomes is not known. They were detected in vertebrate, insect, and plant species, including human, mouse, *Xenopus*, *Drosophila*, and *Arabidopsis*. On the other hand, the yeast *S. cerevisiae* lacks U12-type introns and associated machinery. The accumulating data suggest that the two types of spliceosomes may assemble and disassemble in similar cycles and probably share some common protein components (Hall and Padgett, 1994; 1996; Tarn and Steitz, 1996a, b; Will *et al.*, 1999).

The discovery of RNA catalysis led to early speculations that nuclear pre-mRNA splicing might be a fundamentally RNA-catalysed process mediated by the spliceosomal snRNAs (Sharp, 1985; Cech, 1986). This hypothesis is supported by the observation that Group II self-splicing introns are removed by a two-step chemical pathway that is highly similar to nuclear pre-mRNA splicing (Konarska *et al.*, 1985; Peebles *et al.*, 1986; Ruskin *et al.*, 1984; van der Veen *et al.*, 1986; Sontheimer *et al.*, 1999). The notion that the spliceosome is an RNA enzyme suggests the existance of spliceosomal active site structure(s) composed of snRNAs and pre-mRNA. According to this supposition, the ATP-dependent spliceosome assembly pathway is needed to build these active sites and juxtapose catalytic groups with their respective substrates for the two chemical steps of splicing reaction.

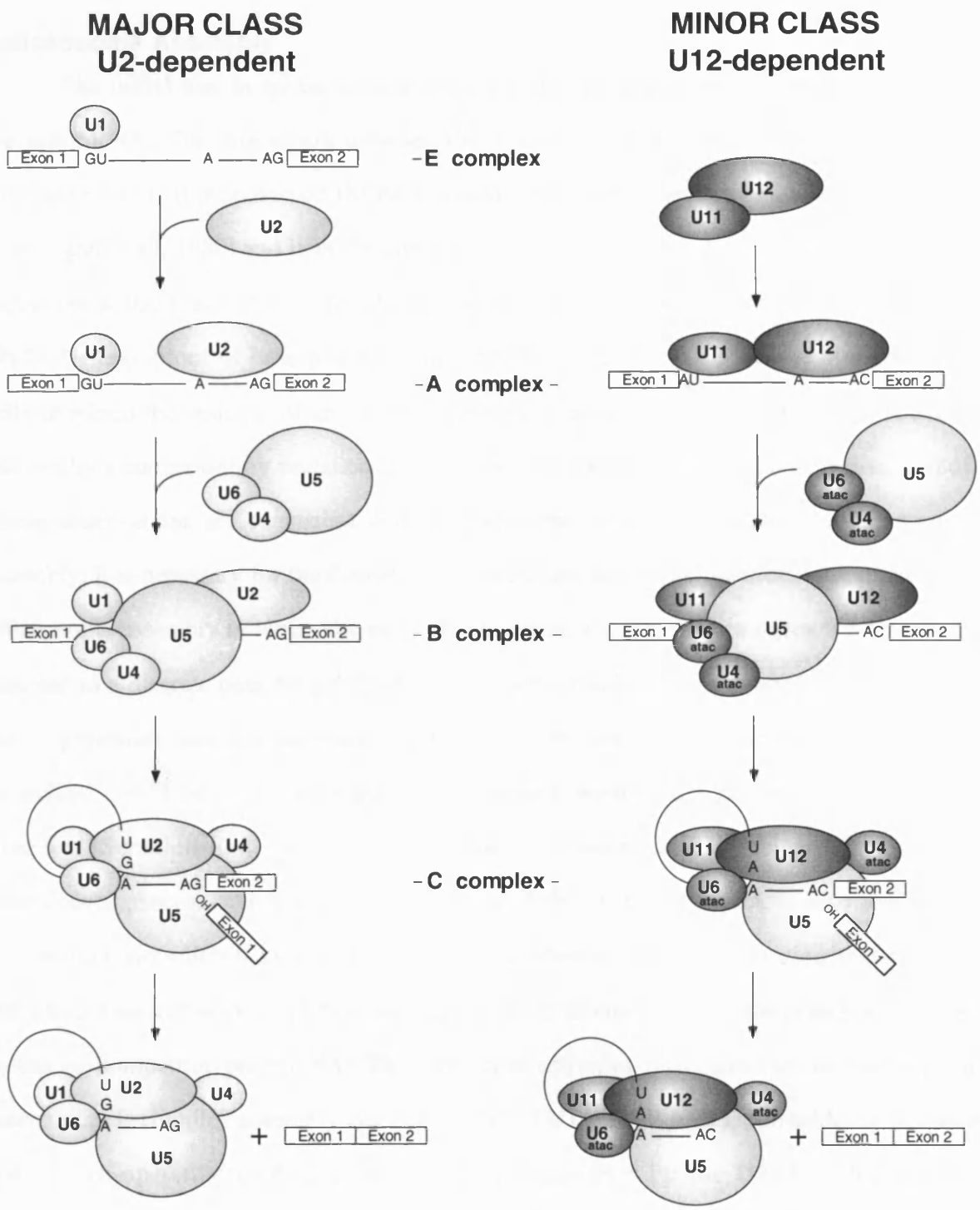


Figure 1.3. Stepwise assembly of the major and minor spliceosomes. The steps, which can be resolved by biochemical methods (e. g., native gel electrophoresis or gel filtration) under native conditions with mammalian splicing extracts, are shown. The snRNPs are indicated by ellipses. Exons and introns are indicated by boxes and lines, respectively. The various spliceosomal complexes are named according to the metazoan nomenclature. The assembly pathway of the U2- and U12-dependent spliceosome differ in that U11 and U12 snRNPs form a stable 18S di-snRNP particle, which is thought to interact with pre-mRNA as a pre-formed complex (Wassarman and Stetz, 1992). The subsequent steps are likely to be identical. (Adapted from Will and Lührmann, 1999).

Spliceosome Assembly

The initial step in spliceosome assembly is the recognition of splicing signals within the pre-mRNA. The interaction between the 5' end of U1 snRNA and the 5' splice site consensus was first proposed on the basis of sequence complementarity (Lerner *et al.*, 1980; Rogers and Wall, 1980) and later demonstrated *in vitro* (Mount *et al.*, 1983). The integrity of sequences at the 5' end of U1 snRNA was shown to be essential for splicing (Kramer *et al.*, 1984). Genetic proof of base-pairing came from the experiments performed in mammalian cells in which the splicing of an intron containing a mutation in the 5' splice site could be rescued by a compensatory mutation in the 5' end of U1 snRNA (Zhuang and Weiner, 1986). These observations are consistent with the hierarchic role of U1 snRNPs in spliceosome assembly; it is necessary for the formation of the earliest detectable functional intermediate in spliceosome assembly *in vitro*. The early (**E** - in mammals, Fig. 1.3) **complex** was originally detected as a discrete peak by gel filtration, and is not resolved from non-splicing-specific H (heterogeneous) complex on native gels or density gradients (reviewed by Reed and Palandjian, 1997). In yeast, native gel conditions were established to resolve the commitment complex (Seraphin and Rosbash, 1989). The formation of E complexes is an ATP-independent process and requires incubation at 30°C. The E complex is also defined as 'committing' pre-mRNA to spliceosome assembly based on the observation that the pre-mRNA in these complexes can be chased into spliced products despite the presence of a vast excess of competitor pre-mRNA. The functional activities recognised to be essential for mammalian E complex assembly are as follows: (i) U1 snRNP interaction with the 5' splice site, (ii) co-operative binding of the polypirimidine tract by the U2AF factor and the branchpoint by mBBP/SF1 (Berglund *et al.*, 1998). The SR proteins stimulate E complex formation by promoting binding of U1 snRNPs to the 5' splice site (Eperon *et al.*, 1993; Kohtz *et al.*, 1994; Staknis and Reed, 1994) and $U2AF^{65}$ to the 3' splice site (Wu and Maniatis, 1993). Although, the protein composition of the E complex is described (Bennett *et al.*, 1992a; Michaud and Reed, 1993; Staknis and Reed, 1994) and requirement of splicing signal sequences for its formation is demonstrated, the precise nature of this assembly is not well understood. It is thought that in E complexes the 5' and 3' splice sites should be brought in close proximity by functional association with each other. Figure 1.4 shows a model proposed

for the mammalian E complex (Wu and Maniatis, 1993). It is possible that in the yeast commitment complex, the nature of cross-intron interactions is different as distinct protein-protein contacts were found within the complex (Abovich and Rosbash, 1997).

The progression of spliceosome assembly is further denoted as a formation of **A complex** in mammals. The A complex is the first ATP-dependent complex in spliceosome assembly and can be detected by density gradient centrifugation, gel filtration and native gel electrophoresis (reviewed in Reed and Palandjian, 1997). The main functional significance of this complex for the spliceosome is that the interaction between U2 snRNP and pre-mRNA defines the intron branchpoint. The base-pairing of intron sequence around the branchpoint with U2 snRNA results in bulging out of the adenosine nucleophile that is required for its use in splicing (Query *et al.*, 1994). A number of factors were shown to promote the association of U2 snRNP with the branchpoint. They are named SF1, SF3a and SF3b in mammalian systems (Krämer and Utans, 1991; Brosi *et al.*, 1993).

The next complex to be detected is the **B complex**. It is formed upon ATP-dependent association of U4, U6, and U5 snRNAs as one [U4/U6.U5] tri-snRNP particle within the spliceosome (Cheng and Abelson, 1987; Konarska and Sharp, 1987). It is likely that the U1-5' splice site interaction is disrupted at this stage, through the action of the putative RNA helicase (Prp28p/U5-100kD; Staley and Guthrie, 1999). Concomitant several important snRNA rearrangements take place. First, the highly stable interaction between U4 and U6 snRNAs is disrupted such that U4 is released just prior to the chemical steps of reaction and U6 snRNA is activated for participation in catalysis. This again requires ATP and possibly the putative RNA helicase (Brr2/U5-200kD; Raghunathan and Guthrie, 1998a; Laggerbauer *et al.*, 1998). The dissociation of U4/U6 snRNAs requires the functions of Prp38p and Prp8p in yeast (Xie *et al.*, 1998; Kuhn *et al.*, 1999). Subsequently, U6 snRNA forms three helixes with U2 snRNA that requires a rearrangement of the structure of U2 snRNA (Sun and Manley, 1995). Also the U6-5' splice site interaction is established, which is mutually exclusive of the U1-5' splice site interaction (Konforti *et al.*, 1993; Hwang and Cohen, 1996; O'Mullane and Eperon, 1998). The formation of U6-U2 helixes and U6-5' splice site base pairing bring

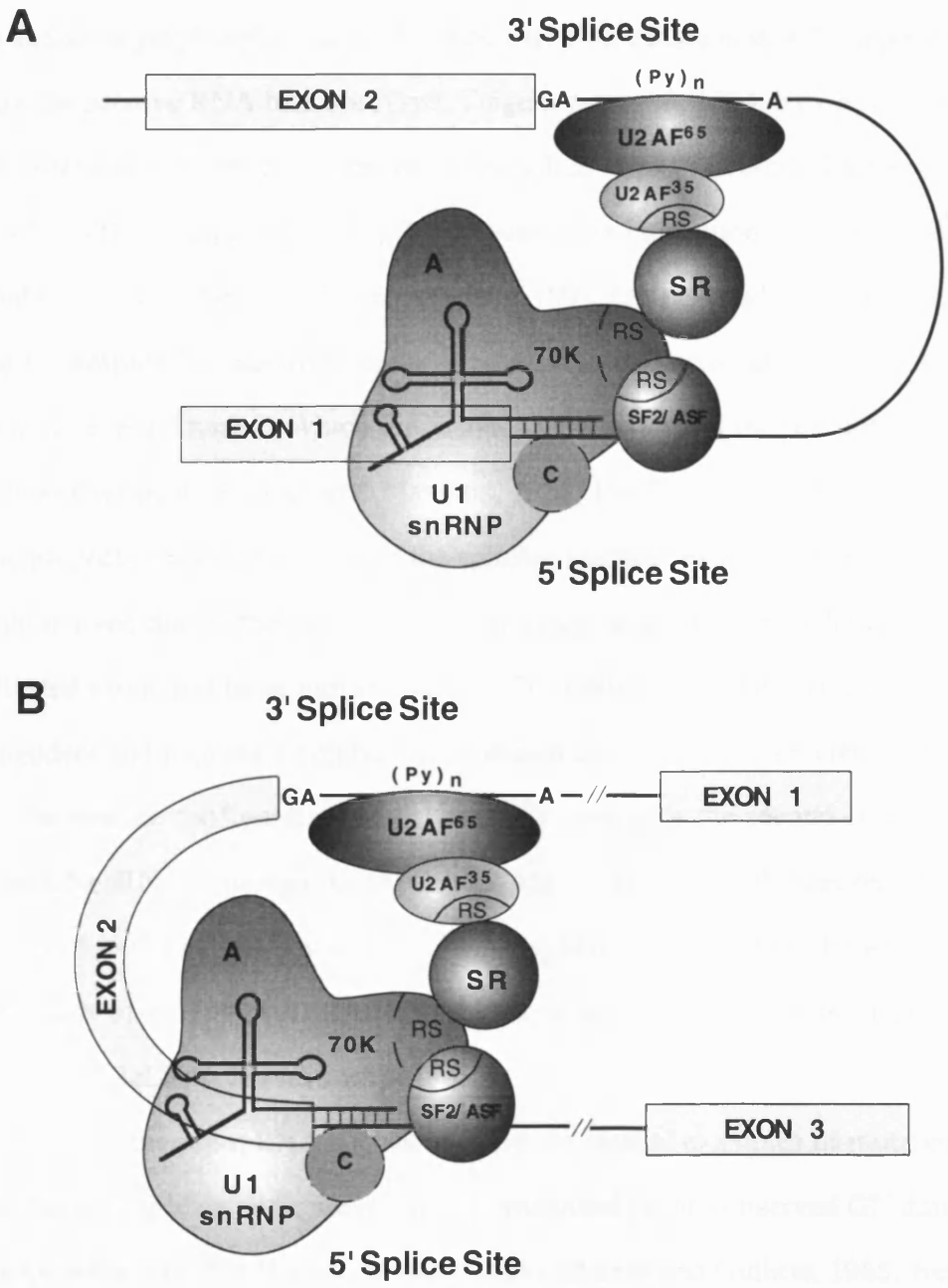


Figure 1.4. Network of protein-protein and protein-RNA interactions in the metazoan spliceosomal E complex. The U1 snRNA (thick solid line) base pairs with the pre-mRNA 5' splice site. This interaction is facilitated by the co-operation between RS domains of U1-70K, SF2/ASF and other SR proteins (discussed in Chapter 2). SR proteins are proposed to bridge the 5' and 3' splice sites by simultaneous interaction with U1 snRNP and U2AF. RS domains are shown by a dotted line. (Modified, with permission, from Will and Lührmann, 1999)

together the nucleophile in the first chemical step of the splicing reaction (the branchpoint adenosine) and its target phosphate at the 5' splice site. This transition is ATP-dependent and promoted by the putative RNA helicase (Prp2, Teigelkamp *et al.*, 1994; Roy *et al.*, 1995). The U5 snRNA interacts with exon sequences at the splice sites of pre-mRNA (Newman and Norman, 1991,1992; Wyatt *et al.*, 1992), but in yeast, this interaction is not essential for the first step and required for the second step (Newman, 1997; O'Keefe and Newman, 1998).

The **C complex** is detected with some substrates on native gels and was isolated by gel filtration under conditions in which the second catalytic step of the reaction was blocked by a mild heat treatment (Krainer and Maniatis, 1985; Reed *et al.*, 1988). The C complex contains the products of catalytic step I of the splicing reaction, exon 1 and lariat-exon 2, and it is very short-lived due to the rapid conversion of the splicing intermediates into spliced products, ligated exons and lariat intron (Fig.1.5). The transition between B and C complexes is ATP-dependent and requires a number of associated splicing factors (reviewed in Burge *et al.*, 1999). The precise configuration of the catalytic centre for the second chemical step is obscure. The U5 snRNA is thought to align the 5' and 3' exons through base pairing with the highly conserved loop I of U5 (reviewed in Madhani and Guthrie, 1994). However, a recent study by Segault *et al.* (1999) demonstrated that this conserved loop is dispensable for splicing *in vitro* in HeLa nuclear extract.

In respect to the pre-mRNA sequence, there are several examples of mutations in the pre-mRNA that uncouple catalytic steps I and II. Mutations in the conserved GU dinucleotide at the 5' splice site affect step II more severely than I (Parker and Guthrie, 1985; Newman *et al.*, 1985; Aebi *et al.*, 1986; Lamond *et al.*, 1987; Aebi *et al.*, 1987). Similarly, mutation of the branchpoint A to G or U residues results in lariats that are unable to undergo step II (Hornig *et al.*, 1986). The AG dinucleotide at the 3' splice site is essential for step II (Reed and Maniatis, 1985; Aebi *et al.*, 1986), and in most pre-mRNAs, the first AG downstream from the branchpoint is used (Reed and Maniatis, 1985; Lamond *et al.*, 1988; Smith *et al.*, 1993).

The most intriguing component of the spliceosome is an evolutionarily highly conserved U5-specific protein, Prp8p in yeast or U5-220kD in humans (Hodges *et al.*, 1995; Lucke *et al.*, 1997; Luo *et al.*, 1999). It was shown that Prp8p is required for

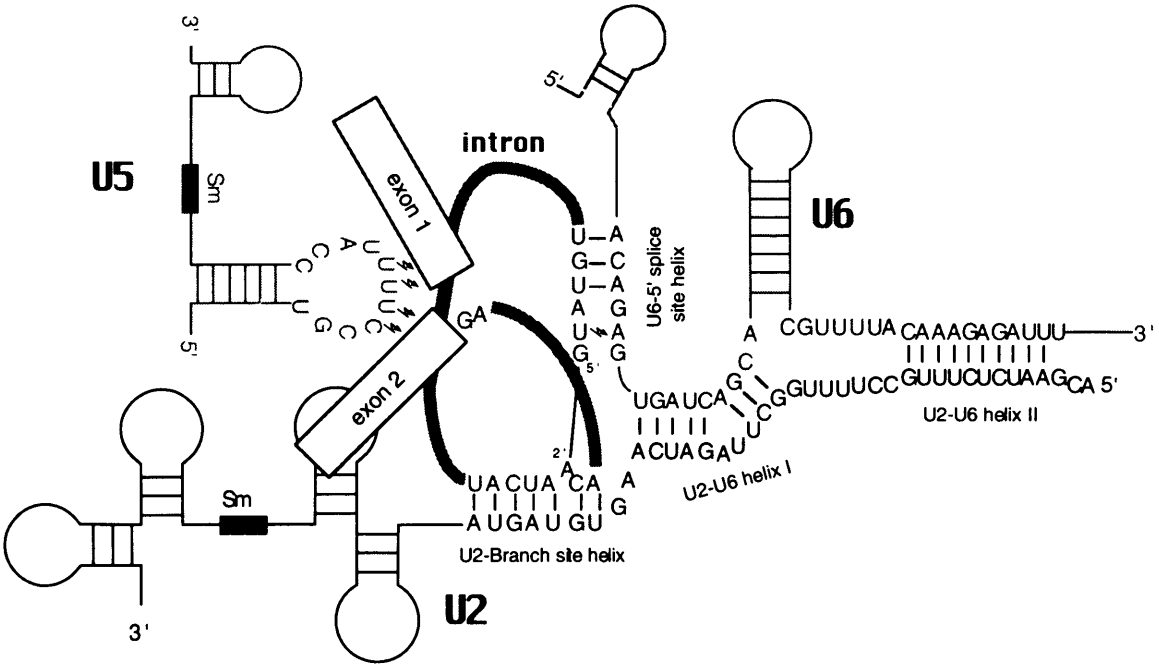


Figure 1.5. The network of RNA interactions occurring in the spliceosomal C complex. In this model, the pre-mRNA has undergone the first trans-esterification reaction and the cleaved exon 1 is held by the interactions with nucleotides of the U5 loop 1 sequence. Exons are represented by rectangles. A thick shaded line depicts intron sequences other than the conserved nucleotides at the splice and branch sites. The 2'-5' phosphodiester bond between the first intron nucleotide (G) and the branched adenosine is shown by a thin line. Watson-Crick base-pairing interactions are indicated by dashes and interactions confirmed by cross-linking are depicted by a wavy line. (Adapted from Wise, 1993).

spliceosome assembly before the first step (Brown and Beggs, 1992). Also, it is closely associated with sequences at the 5' splice site, branch site, and 3' splice site, implying a possibly direct role in catalysis or a function as a major scaffolding protein for aligning the splice sites (Collins and Guthrie, 1999; Siatecka *et al.*, 1999). Interestingly, there is no recognised subdomain structure in Prp8p and U5-220kD.

It was shown that spliced exon RNA is released from the spliceosome through the action of a helicase (Prp22/HRH1) and (perhaps) other factors (Company *et al.*, 1991; Ono *et al.*, 1994; Ohno and Shimura, 1996; Schwer and Gross, 1998). Another putative RNA helicase (Prp43) is important for the release of the lariat intron and U2, U5, and U6 snRNPs (Arenas and Abelson, 1997; Gee *et al.*, 1997). The released snRNPs are believed to reassemble and participate in new rounds of spliceosome assembly (Raghunathan and Guthrie, 1998b).

1.4. Small nuclear RNA-protein complexes

A distinctive feature of mammalian cells, as compared to bacteria, is their great abundance of small RNA species. These reside in the nucleus rather than in cytoplasm, and are relatively stable, having lifetimes at least as long as a cell generation. Those small nuclear RNAs (snRNAs) that are involved in splicing possess unusual trimethylated ($m_3^{2,2,7}G$) cap structures at their 5' ends (except for U6) and contain some other modified nucleotides. They all exist not as naked RNA molecules but stably complexed with certain abundant nuclear proteins. The resulting particles are referred to as small nuclear ribonuclearproteins (snRNPs). The snRNP particles that constitute a major type of spliceosome are very plentiful especially in mammalian cells, where the number of each type is equal to $2x10^5$-10^6 particles per cell.

Another important feature of these snRNPs is that their protein components and the sequence of RNAs are highly conserved across all higher eukaryotic species. Even in lower eukaryotes like yeast the organisation of cognate particles is very similar to that in mammals (Will and Lührmann, 1997b).

In higher eukaryotes two groups of splicing snRNPs have been described so far. The first group is involved in the removal of major class of introns (GT-AG) whereas another group functions within the minor (AT-AC) class. The abundance of the former is at least two

orders of magnitude higher than the latter, in agreement with the representation of major class of intron in comparison with the minor one.

The snRNP constituents of the major spliceosome have been best characterised in HeLa cells. It is possible to isolate U1, U2, U5 and [U4/U6.U5] snRNPs from nuclear extracts. The protein composition of these snRNPs is summarised in Table 1.1. Some of the polypeptides are shared by all splicing snRNPs, whereas others are unique to particular snRNPs. The existence of common and unique protein components was first recognised with the discovery that some patients with systemic lupus erythematosus (an autoimmune disease) produce autoantibodies directed against snRNP proteins. For instance, some autoantibodies precipitate U1, U2, U5 and U4/U6 snRNPs and therefore recognise shared polypeptides. The biochemical isolation and sequence analysis of these polypeptides revealed the existence of a common set (referred to as Sm proteins) of seven proteins, named B/B' (alternatively spliced isoforms), D3, D2, D1, E, F and G, with low molecular weights ranging from 9 to 29 kDa. These proteins form a very stable complex that associate with a specific Sm site on each snRNA. The Sm proteins seem to play an important role in the biogenesis of the snRNPs.

The second class of proteins associated with snRNPs is comprised of particle-specific proteins. The U1 snRNP particle contains three proteins, designated 70K, A and C. U2 snRNA is complexed with twelve specific proteins of molecular weights ranging from 33 to 160 kDa. The 25S [U4/U6.U5] tri-snRNPs possess the most complex protein composition of snRNPs; that is 15 specific proteins in addition to two sets of Sm proteins and one set of recently discovered Sm-like proteins, which belongs to U6 snRNA (Mayes *et al.*, 1999; Achsel *et al.*, 1999, in press).

The vast majority of proteins identified in isolated snRNPs are also present in the spliceosome (Reed and Palandjian, 1997), and thus, most spliceosomal proteins have a snRNP origin. The majority of the human snRNP proteins have now been cloned and several have been shown to perform essential functions during spliceosome assembly and the catalytic steps of splicing. For example, U1-70kD is involved in association of U1 snRNP with the 5' splice site and its reversible phosphorylation is necessary for the first cleavage reaction (Tazi *et al.*, 1993). Moreover, several snRNP proteins possess structural motifs that could enable

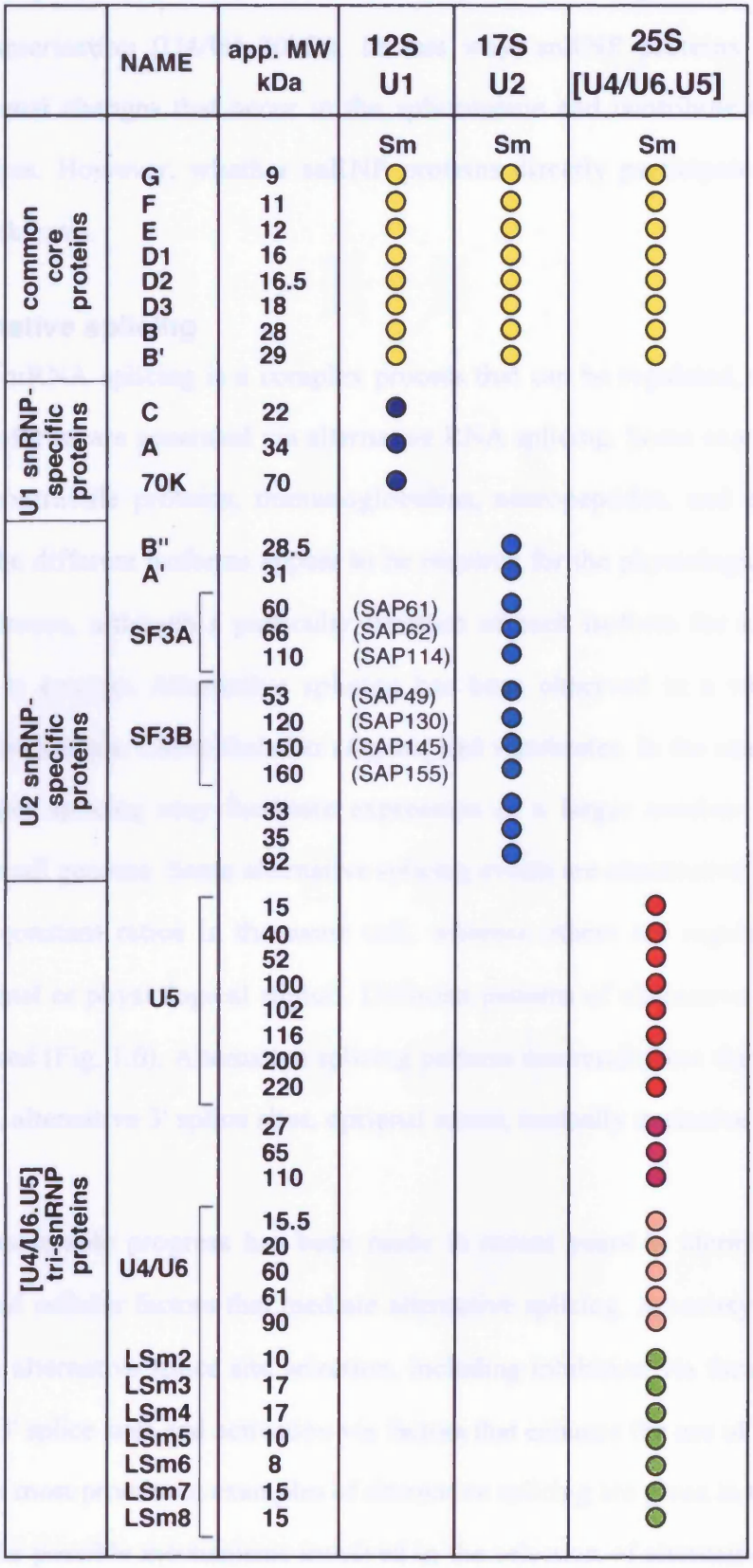


Table 1.1. Schematic presentation of the protein composition of HeLa 12S U1, 17S U2 and 25S [U4/U6.U5] snRNPs. The colour code indicates the proteins associated in one complex.

important enzymatic activities, such as RNA duplex unwinding (U5-200kD, U5-100kD) or protein isomerisation (U4/U6-20kD). In this way, snRNP proteins might drive many conformational changes that occur in the spliceosome and contribute to the formation of catalytic sites. However, whether snRNP proteins directly participate in the catalysis is currently unknown.

1.5. Alternative splicing

Pre-mRNA splicing is a complex process that can be regulated, and for some genes different isoforms are generated via alternative RNA splicing. Some examples include genes encoding contractile proteins, immunoglobulins, neuropeptides, and extracellular matrix proteins. The different isoforms appear to be required for the physiological needs of various cells and tissues, although a particular function of each isoform for a given gene is just beginning to emerge. Alternative splicing has been observed in a variety of organisms including *Drosophila*, *Caenorhabditis elegans*, and vertebrates. In the case of viruses, the use of alternative splicing may facilitate expression of a larger number of proteins from a relatively small genome. Some alternative splicing events are constitutive and mRNA variants coexist at constant ratios in the same cell, whereas others are regulated in response to developmental or physiological stimuli. Different patterns of alternative RNA splicing have been observed (Fig. 1.6). Alternative splicing patterns can result from the use of alternative 5' splice sites, alternative 3' splice sites, optional exons, mutually exclusive exons, and retained introns.

Considerable progress has been made in recent years in identifying the cis-acting elements and cellular factors that mediate alternative splicing. A variety of mechanisms are involved in alternative splice site selection, including inhibition via factors that specifically block 5' or 3' splice sites and activation via factors that enhance the use of specific splice sites. Some of the most prominent examples of alternative splicing are given in Chapter 3 and details of the possible mechanisms involved in the selection of alternative 5' splice sites are discussed in Chapter 2.

The cellular factors that regulate alternative splicing can be divided broadly into general splicing factors and cell type-specific factors. Tissue-specific regulators of splicing,

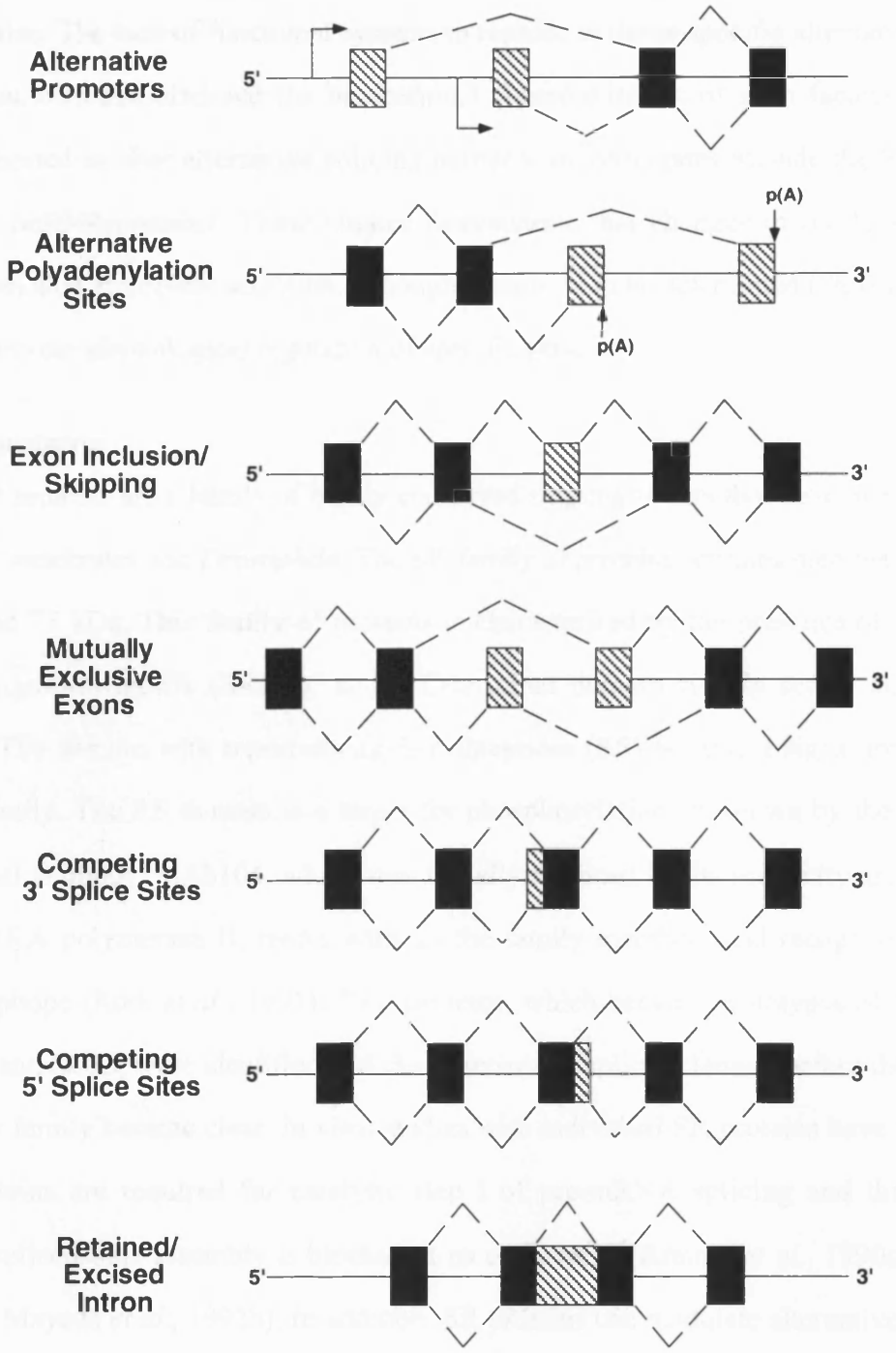


Figure 1.6. Patterns of alternative splicing. Boxes represent exons and lines represent introns. Constitutively spliced exons are shown as black boxes and alternatively spliced exons as shaded boxes. Splicing events are depicted by dashed line. Transcription start sites are indicated by an arrow and polyadenylation sites are denoted by p(A). (Adapted from Breitbart *et al.*, 1987).

similar to those identified in *Drosophila* systems (see Chapter 3), have not yet been identified in vertebrates. The lack of functional systems to reproduce tissue-specific alternative splicing partterns *in vitro* has hindered the biochemical characterisation of such factors. The only factors reported to alter alternative splicing partterns in vertebrates include the SR proteins and some hnRNP proteins. These studies demonstrate that changes in the level of these proteins can alter splice-site selection, although it remains to be determined if this observation is relevant to the physiological regulation of specific genes.

1.6. SR proteins

SR proteins are a family of highly conserved splicing factors that have been identified in various vertebrates and *Drosophila*. The SR family of proteins contains members of 20, 30, 40, 55, and 75 kDa. This family of proteins is characterised by the presence of one or two RNA recognition motifs (RRMs), and a C-terminal domain rich in serine and arginine residues. The domain with repeated Arg-Ser dipeptides (RS) became a signature of the SR protein family. The RS domain is a target for phosphorylation, as shown by the fact that a monoclonal antibody, mAb104, which was initially screened for its reactivity against active sites of RNA polymerase II, reacts with all the family members and recognises a shared phosphoepitope (Roth *et al.*, 1991). Two proteins, which became prototypes of the family, SF2/ASF and SC35, were identified and characterised as splicing factors before the existence of a larger family became clear. *In vitro* studies with individual SR proteins have shown that these proteins are required for catalytic step I of pre-mRNA splicing and that, in their absence, spliceosome assembly is blocked at an early stage (Krainer *et al.*, 1990a; Zahler *et al.*, 1992; Mayeda *et al.*, 1992b). In addition, SR proteins can modulate alternative splice site selection within a number of different pre-mRNAs in a concentration dependent manner (Fu *et al.*, 1992; Zahler *et al.*, 1993; Cavaloc *et al.*, 1994; Screaton *et al.*, 1995).

The RS domain is also found in other mammalian splicing factors, such as U2AF, U1-70K, U5-100K, U4/U6.U5-27K, and in *Drosophila* specific splicing factors such as SWAP, Tra, and Tra-2 proteins. The functional role of the RS domain was demonstrated (Wu and Maniatis, 1993; Jamison *et al.*, 1995; Xiao and Manley, 1997; Graveley and Maniatis, 1998b; Yeakley *et al.*, 1999) and it is to impel the protein-protein interactions with other RS domain containing proteins. The ability of SR proteins to establish the network of protein-protein

interactions plays an essential role in constitutive splicing in early events of spliceosome assembly and is most likely involved in bridging of splice sites (Fig. 1.4). The importance of bridging interactions is emphasised by the Exon/Intron definition model, which proposes that the splice sites are defined primarily by the interactions between factors bound at 5' or 3' splice sites (Robberson *et al.*, 1990). It was shown that, in some cases, SR proteins bypass the requirement for U1 snRNP in splicing *in vitro* (Crispino *et al.*, 1994, 1996; Tarn and Steitz, 1994). The phosphorylation state of the RS domain affects protein-protein interactions it mediates. Several studies showed that phosphorylation of SR proteins is critical for spliceosome assembly (Mermoud *et al.*, 1994; Roscigno and Garcia-Blanco, 1995; Xiao and Manley, 1997) whereas dephosphorylation is required for both catalytic steps of pre-mRNA splicing (Cao *et al.*, 1997; Kanopka *et al.*, 1998; Murray *et al.*, 1999). These observations suggest that phosphorylation modulates protein-protein interactions within the spliceosome, thereby contributing to dynamic structural reorganization of the spliceosome during splicing (reviewed by Misteli, 1999). It was also shown that phosphorylation affects the intranuclear distribution of SR proteins and other splicing factors in interphase cells and their reorganisation during mitosis (Misteli and Spector, 1996; Misteli *et al.*, 1997; Misteli *et al.*, 1998; Kuroyanagi *et al.*, 1998; Yeakley *et al.*, 1999; Koizumi *et al.*, 1999). Several serine kinases were identified that specifically phosphorylate SR proteins on RS domain serine residues (Gui *et al.*, 1994; Rossi *et al.*, 1996; Fetzer *et al.*, 1997; Wang *et al.*, 1998; Prasad *et al.*, 1999).

The RNA binding domains play a role in the association of SR proteins with the target RNA. In this respect, it was shown that SF2/ASF recognises the 5' splice site and enhances the U1 snRNP binding to it (Eperon *et al.*, 1993; Kohtz *et al.*, 1994; Staknis and Reed, 1994). Also, in certain metazoan pre-mRNA, splicing enhancers have been defined as exonic elements that bind *trans*-acting factors in a sequence specific manner and activate splicing of the introns located some distance away from the protein binding site. SR proteins mediate functional recognition of the enhancers, since they were shown to form specific complexes with RNA containing the enhancer element *in vitro* (Sun *et al.*, 1993; Tacke and Manley, 1995; Liu *et al.*, 1998; Schaal and Maniatis, 1999). Among the splicing enhancers that have been best characterised in terms of their bound trans-acting factors are those present in

specific exons of *Drosophila* dsx, human IgM, growth hormone, human fibronectin, and chicken cardiac troponin T pre-mRNAs (some examples are discussed in Chapter 3). The main features of these ESE are that they contain purine-rich sequences, and are often functionally interchangeable among different pre-mRNAs. Importantly, tissue-specific or developmentally regulated vertebrate splicing factors that modulate splicing efficiency have not been identified yet. Therefore, binding of splicing enhancers by individual SR proteins was proposed to be a general mechanism for regulating alternative splicing of pre-mRNA that contains these elements, whether of the purine-rich class or of a different class. It was shown that SR proteins might mediate interactions between the splicing enhancer and the 3' splice site through protein-protein interactions with small subunit U2AF35 (Wu and Maniatis, 1993; Zuo and Maniatis, 1996). Thus, it is thought that the recognition of ESEs by SR proteins results in the assembly of specific complexes that are able to interact with the components of splicing machinery. This should add to the complexity of the network of interactions established early during spliceosome formation and determine the pattern of alternative splicing.

1.7. HnRNP proteins

Following transcription, the nascent transcript becomes associated with a distinct set of proteins known as heterogeneous nuclear ribonucleoproteins (hnRNPs). The human hnRNP proteins are the best characterised and this family contains at least 20 major proteins that are resolved by two-dimensional gel electrophoresis; these are designated A1 (34 kDa) to U (120 kDa) (reviewed by Dreyfuss *et al.*, 1993). The hnRNP proteins, like histones, are among the most abundant proteins in the nucleus. Notably, most of the major hnRNP proteins (A through U) appear to be highly conserved among vertebrates, both immunologically and structurally. Diversity among hnRNP proteins is often generated through alternative splicing and by posttranslational modifications, including methylation of arginines and phosphorylation of serines and threonines. The functions of these modifications have not been determined, but they are likely to modulate specific interactions of the proteins with other proteins and with RNA and could affect the localisation of these proteins within the cell (Liu and Dreyfuss, 1995; Kim *et al.*, 1998; Shen *et al.*, 1998).

A general theme that has emerged is that most, if not all, hnRNP proteins that have been identified by both UV-crosslinking and immunopurification methods are RNA-binding proteins (Pinol-Roma *et al.*, 1988; Swanson and Dreyfuss, 1988). Therefore, these proteins might influence the structure of RNA, to which they are bound (hnRNA, for heterogeneous nuclear RNA), and facilitate or hinder the interactions of pre-mRNA sequences with other components that are needed for processing of hnRNA, thus affecting the fate of hnRNA. Accordingly, it was shown that hnRNP proteins can bind *in vitro* to many different single stranded ribo- and deoxyribo-polynucleotides (Conway *et al.*, 1988; Schenkel *et al.*, 1988).

The hnRNP proteins have a modular structure. They possess one or more RNA-binding modules and at least one auxiliary domain. The most common RNA-binding motif in hnRNP proteins is the RNA recognition motif (RRM, also referred to as RNP motif). Other types of RNA-binding domains include RGG box, KH motif and non-canonical RBDs found in hnRNP I and L proteins. The functional significance of auxiliary domains is a relatively unexplored area, but it is likely that these regions mediate protein-protein interactions, affect RNA-binding properties, and may also act to localise the proteins within the cell (Cartegni *et al.*, 1996; Abdul-Manan *et al.*, 1996; Del Gatto-Konczak *et al.*, 1999).

The RRM is an ancient and extremely common RNA-binding module, within which the RNP-2 hexamer and RNP-1 octamer submotifs are highly conserved (reviewed by Birney *et al.*, 1993). The crystal structure of RRMs from several proteins (U1-A, UP1/hnRNP A1, Sxl) were recently determined and the overall structure of different RRMs is similar (Nagai *et al.*, 1990; Xu *et al.*, 1997; Shamoo *et al.*, 1997; Handa *et al.*, 1999). The RRM has β1-α1-β2-β3-α2-β4 structure; the four β strands form antiparallel β sheet that packs against the two α helices. Notably, many proteins contain multiple RRMs, and it is conceivable that they can bind to more than one RNA segment simultaneously. The latter situation was realised in the co-crystal of the two-RRM domain of hnRNP A1 and single-stranded telomeric DNA (Ding *et al.*, 1999). Alternatively, two RRMs of the Sxl protein were bound to the 12-nucleotide RNA derived from the *tra* polypyrimidine tract sequentially (Handa *et al.*, 1999). Importantly, the determination of the overall structures of the RRM-RNA complexes showed that the RNA is exposed as it is bound by the RRM (as opposed to buried in a binding pocket) and thus

accessible to other pre-mRNA processing factors (Görlach *et al.*, 1992; Ding *et al.*, 1999; Handa *et al.*, 1999).

In addition to the general role of hnRNP proteins in substrate presentation, they may also have specific functions in the regulation of pre-mRNA splicing. The functional relatedness of hnRNP proteins and splicing factors is underscored by the observation that alternative splicing can be regulated by the interplay between hnRNP protein A1 and another RRM motif RNA-binding protein SF2/ASF (discussed in details in Chapter 2). The 57 kDa hnRNP I/PTB protein has also been implicated in pre-mRNA splicing (Chan and Black, 1997; Southby *et al.*, 1999; Lou *et al.*, 1999). The specific role of hnRNP I/PTB in the case of the splicing of a β-tropomyosin pre-mRNA is that binding of the protein to the downstream polypyrimidine tract causes the selection of this 3' splice site over the upstream splice site (Guo *et al.*, 1991; Mulligan *et al.*, 1992; Grossman *et al.*, 1998).

In addition, some hnRNP proteins, such as those in the A and B groups, shuttle between the nucleus and the cytoplasm (Pinol-Roma and Dreyfuss, 1992; Siomi *et al.*, 1997). The shuttling of some hnRNP proteins has important implications: it suggests that these proteins may also have functions in the cytoplasm and in nucleoplasmic transport of mRNA.

CHAPTER 2:

Comparative study of RNA-binding properties of hnRNP A1 and SF2/ASF proteins

2.1. INTRODUCTION: Much Ado about SF2/ASF and hnRNP A1

The selection of authentic splice sites in a wilderness of similar sequences is a task with which her majesty Nature deals quite successfully, but not human beings that challenge the same problem. It was discussed above that a little sequence conservation is a main feature of splice signals recognised in high eukaryotes. Nevertheless, there are numerous examples of the alternative use of splicing sites, which suggest the existence of certain mechanisms or rules that regulate this process. Most progress has been achieved in the study of 5' splice site selection, because historically the base pairing between the consensus 5' splice site and the 5' end of U1 snRNA was first proposed, whereas nothing was known about 3' splice site recognition. The first step forward in understanding the selection mechanism was postulating the idea that all potential 5' splice sites exist in competition with each other (Reed and Maniatis, 1986; Eperon *et al.*, 1986). The question of competition was first addressed in terms of strong or weak 5' splice site in respect to U1 snRNPs binding. The early results clearly demonstrated that increasing base pairing between the splice site and the U1 snRNP improves splicing to this particular site (Eperon *et al.*, 1986). The second parameter that influences the choice of splice site is its physical accessibility. For example, the sterical effects of sequestration of a potential splice site into secondary structures was shown to abolish the usage of the splice site (Eperon *et al.*, 1986; Blanchette and Chabot, 1997). It was also noted *in vitro* that the distance between two 5' splice sites can affect the splicing pattern. The effect of proximity was proposed based on the observation that, for two strong splice sites separated by more than 40 nt, the site which is closer to the 3' splice site is used (Reed and Maniatis, 1986; Cunningham *et al.*, 1991). This observation has brought into consideration the next step of spliceosome formation, emphasising that the ultimate goal of a 5' splice site is not only to be recognised by the machinery but rather to be processed.

The diversity of splice sites and their differential utilisation induced the search for additional trans-acting factors that influence a choice of the splice site.

SF2/ASF and hnRNP A1 antagonism

Human pre-mRNA splicing factor 2 (SF2) was first isolated from HeLa cell nuclear extract as an activity required to complement a HeLa S100 cytosolic extract, which is otherwise inactive in splicing. It was shown that SF2 is necessary for 5' splice site cleavage and lariat formation during pre-mRNA splicing *in vitro* and appear to be required for the assembly or stabilisation of the earliest specific pre-spliceosome complex (Krainer *et al.*, 1990a). In addition to its essential role in constitutive splicing, SF2 strongly influences 5' splice site selection. When pre-mRNAs containing multiple cis-competing 5' splice sites are spliced *in vitro*, high concentrations of purified SF2 promote the use of the 5' splice site closest to the 3' splice site (Krainer *et al.*, 1990b). The hypothesis was proposed that these effects of SF2 on splice site selection might reflect the cellular mechanisms that prevent exon skipping and ensure the accuracy of splicing. In addition, alterations in the concentration or activity of SF2 may serve to regulate alternative splicing *in vivo*.

Independently, a protein factor ASF (Alternative Splicing Factor) was identified and purified from 293 cells, on the basis of an alternative splicing assay (Ge and Manley, 1990). In this assay, ASF was added to the HeLa cell nuclear extract and caused a change in 5' splice site selection of the SV40 early pre-mRNA, favouring the downstream small t-antigen 5' splice site. When human SF2 and ASF cDNAs were isolated, they were found to be identical factors, and are now designated SF2/ASF or ASF/SF2.

Both laboratories were able to purify the recombinant SF2/ASF protein overexpressed in bacteria and show that it is active in *in vitro* splicing and stimulates proximal 5' splice sites (Krainer *et al.*, 1991; Ge *et al.*, 1991). The detailed characterisation of properties of SF2/ASF was carried out in both laboratories and discussed later.

A novel protein factor, DSF (Distal Splicing Factor), was fractionated and described that is required for the use of distal alternative 5' splice sites *in vitro* (Harper and Manley, 1991), but individual polypeptide was not identified. On the basis that in S100 extract complemented with SF2/ASF a proximal splice site is used whereas in nuclear extract - a

distal splice site, an activity that promotes distal site utilisation was caught. Krainer and co-workers isolated an activity, termed SF5, that specifically counteracts the effect of SF2 on splice-site selection for several model pre-mRNA substrates (Mayeda and Krainer, 1992). It is unlikely that SF5 acts as inhibitor of SF2/ASF, since an excess of SF5 over SF2/ASF had an effect of distal splice site selection but did not inhibit splicing in S100 extract, triggered by SF2/ASF. Thus, it was shown that the ratio of SF5 to SF2/ASF is a major determinant of splice-site selection. In general, an excess of SF5 favours distal 5' splice sites, whereas an excess of SF2/ASF results in utilisation of proximal 5' splice sites.

When SF5 was purified to apparent homogeneity, it was identified as hnRNP A1 (Mayeda and Krainer, 1992), a member of hnRNP protein family that had already been cloned and well-studied. The opposing effects of SF2/ASF and hnRNP A1 on 5' splice site selection were generalised to a concept of antagonism. The idea was put forward that a regulation of these antagonistic activities or variations in their intracellular levels may influence different modes of alternative splicing *in vivo* and may be a natural mechanism for tissue-specific or developmental regulation of gene expression. Comprehensive experimental data were obtained *in vitro* to support this hypothesis (Krainer *et al.*, 1991; Zahler *et al.*, 1992; Fu *et al.*, 1992; Mayeda and Krainer, 1992; Mayeda *et al.*, 1993).

In addition to the 5' splice site selection, the ratio of SF2/ASF to hnRNP A1 may also regulate another mode of alternative splicing such as exon skipping-inclusion (Mayeda *et al.*, 1993). For example, the last intron of bovine growth hormone (bGH) gene pre-mRNA is spliced inefficiently due to a weak 5' splice site. SF2/ASF stimulates *in vitro* splicing of bGH intron D through specific interaction with the exonic splicing enhancer (ESE) sequences. The stimulation of bGH intron D splicing by SF2/ASF is counteracted by the addition of hnRNP A1 (Sun *et al.*, 1993). Since the relative levels of SF2/ASF and hnRNP A1 influence the efficiency of bGH intron D splicing *in vitro*, it may be the underlying mechanism of this alternative pre-mRNA processing event *in vivo*.

The breakthrough in appreciation of this model was the demonstration of the antagonistic effects of hnRNP A1 and SF2/ASF *in vivo* (Caceres *et al.*, 1994; Wang and Manley, 1995). The transient overexpression of SF2/ASF in HeLa cells activated proximal 5' splice sites, promoted inclusion of a specific exon and prevented abnormal exon skipping.

Krainer and co-workers (Caceres *et al.*, 1994) insisted that there was an opposing effect of hnRNP A1 during transient overexpression of the protein in HeLa cells although, the documented effects were very minor. Nevertheless, the strong effect of SF2/ASF on splicing *in vivo* was unambiguous. Consistent with the antagonistic model it was observed that the protein levels of hnRNP A1 and SF2/ASF splicing factors vary naturally over a very wide range in different tissues and in immortal and transformed cell lines (Hanamura *et al.*, 1998; Kamma *et al.*, 1995; Chabot *et al.*, 1992). Despite the fact that the antagonistic model is commonly accepted the effect of hnRNP A1 was demonstrated only under certain circumstances, "with appropriate alternative exons" (Mayeda *et al.*, 1993). It was noted that the effect of hnRNP A1 depends on the size of the internal alternative exon and on the strength of the polypyrimidine tract in the preceding intron. Significantly, an excess of hnRNP A1 fails to promote inappropriate exon skipping of natural constitutively spliced pre-mRNA (Mayeda *et al.*, 1993).

Recent work of Bai *et al.* (1999) extended the list of antagonistic effects for hnRNP A1 and SF2/ASF by showing that the relative ratio of these factors influences the 3' splice site choice *in vivo*. On the transcripts that naturally undergo alternative 3' splicing, they showed that SF2/ASF promote proximal 3' splice site, resulting in the inclusion of downstream exon. HnRNP A1 antagonises this effect and promotes distal 3' splice site and exon skipping. The authors give an alternative name to the antagonism; they call it proximity/distance sensing (Bai *et al.*, 1999).

SF2/ASF

The breadth and importance of different physiological functions of SF2/ASF was demonstrated in experiments of targeted gene disruption in the chicken B-cell line DT40 (Wang *et al.*, 1996; Wang *et al.*, 1998). It was shown that SF2/ASF is required for cell viability but the RS domain is interchangeable. (The important implications of the RS domain in the variety of protein-protein interactions were already discussed in Chapter 1.)

SF2/ASF consists of 248 amino acid residues, including two copies of the RNA-recognition motif (RRM) at its N-terminus, separated by a glycine-rich hinge, and a 50 residue C-terminal region that is 80% arginine/serine-rich (RS domain). The second RRM of

SF2/ASF is atypical in that it shows poor conservation of the usually conserved RNP-2 and RNP-1 submotifs. However, the conservation of other typical RRM residues strongly suggest that its tertiary fold is similar to that of canonical RRMs. The C-terminal RS domain of SF2/ASF is rich in consecutive RS dipeptides, in contrast to other proteins, in which Arg and Ser residues are dispersed and do not show the same type of periodicity (reviewed in Birney *et al.*, 1993). Functional analysis of SF2/ASF structural domains demonstrated distinct role of the individual domains in the constitutive and alternative splicing (Caceres and Krainer, 1993; Zuo and Manley, 1993). Whereas both intact N-terminal RRMs and an RS domain are essential for the constitutive splicing activity, the RS domain is surprisingly dispensable for the *in vitro* alternative splicing activity. The requirement for an intact RS domain for complementation of inactive S100 extract seems to be very stringent, since SF2/ASF mutant derivatives with an RT or KS domain are inactive, demonstrating that the presence of positive charges interspersed with polar residues that may be phosphorylated is not sufficient for constitutive splicing (Caceres and Krainer, 1993). In addition, it has been shown that RRM1 and RRM2 synergize for binding to the RNA (Caceres and Krainer, 1993; Zuo and Manley, 1993).

It is believed that SF2/ASF as a typical RNA binding protein modulates alternative splicing through specific recognition of RNA sequences. Therefore, the question of RNA binding specificity for SF2/ASF was intensively studied. At the period of time when I was carrying out the experiments presented in this chapter the following results had been obtained in different laboratories. The specific recognition of 5' splice sites, generally purine-rich sequences, was documented in electrophoretic mobility shift assay and UV-crosslinking experiments (Zuo and Manley, 1994). The first data from SELEX experiments performed with the portion of the protein comprising the two RRMs were reported (Tacke and Manley, 1995). A general problem of aggregation of SR proteins due to the RS domain is a well-documented fact that, probably, did not allow the authors to use the full-length protein in this study. It was shown that SF2/ASF preferentially binds purine-rich sequences with a characteristic motif. Binding assays performed with the full-length protein confirmed this specificity, and at the same time provided evidence that the highly charged arginine-serine region of the protein is not a major determinant of the specificity. Further Tacke *et al.* (1995)

demonstrated that an element containing three copies of a high-affinity SF2/ASF binding site constitutes a powerful splicing enhancer when placed 80 nucleotides downstream of the 3' splice site of an intron with sub-optimal splicing signals. Thus, these experiments supported the hypothesis that recognition of such specific sequences is one of the functions of SF2/ASF in alternative splicing. My work had also broached the question of RNA binding specificity, and hence, this matter is addressed in Discussion (2.3).

Analysis of additional cDNA clones has shown that SF2/ASF pre-mRNA can be alternatively spliced giving rise to two more isoforms (Ge *et al.*, 1991). These two isoforms of SF2/ASF, ASF-2 and ASF-3 lack the RS domain. Whether they encode stable proteins *in vivo*, which may have a regulatory role, is unknown. Alternatively, these mRNA may serve to down-regulate the levels of authentic SF2/ASF.

The role of SF2/ASF in constitutive and alternative splicing is under intensive investigation up to now. Based on experimental data Eperon *et al.* (1993) proposed the first model for 5' splice site selection. According to this model, SF2/ASF increases the affinity of U1 snRNPs for 5' splice sites leading to the selection of the proximal 5' splice site (the model is discussed in detail below). When my research project had been already started several laboratories united their efforts to reveal the proposed effect of SF2/ASF on U1 snRNPs binding to the 5' splice site (Kohtz *et al.*, 1994; Jamison *et al.*, 1995). It was shown that purified U1 snRNP could interact with pre-mRNA containing a 5' splice site, but this binding is not stable. Only in the presence of SF2/ASF could stable complex formation be detected. It was noted that the order of addition is important because the effect of stabilisation could be achieved only when SF2/ASF was pre-incubated with pre-mRNA and the U1 snRNP is subsequently added. The RS domain of SF2/ASF is required for co-operative binding of the U1 snRNP and SF2/ASF to the 5' splice site as it is thought to interact with the RS domain of U1-70kD protein. In summary, SF2/ASF is believed to recognise the 5' splice site, recruit U1 snRNP to it and initiate the spliceosome assembly by imposing protein-protein interactions.

Subsequently, new data on functional effects of phosphorylation-dephosphorylation of RS domain have been reported. Xiao and Manley (1997) provided evidence that the RS domain of unphosphorylated recombinant SF2/ASF is necessary, but not sufficient, for binding to the U1-70kD protein *in vitro*. An apparent interaction of the isolated RS domain

with U1-70kD protein was observed if contaminating RNA was not removed, suggesting a non-specific bridging between the basic RS domain, RNA, and U1-70kD protein. *In vitro* phosphorylation of recombinant SF2/ASF both significantly enhanced binding to U1-70kD protein and eliminated the RS domain-RNA interaction. Garcia-Blanco and co-workers (Cao *et al.*, 1997) confirmed the necessity of SF2/ASF phosphorylation for activation of *in vitro* splicing and also showed that thiophosphorylated SF2/ASF, which cannot be readily dephosphorylated, promotes spliceosome assembly, but does not allow the first transesterification reaction to occur.

HnRNP A1

HnRNP A1 is a major component of heterogeneous ribonucleoprotein (hnRNP) particles, within which the messenger RNA precursor, heterogeneous nuclear RNA (hnRNA), is encapsulated by a set of proteins (reviewed in McAfee *et al.*, 1997). HnRNP A1 has been shown to accelerate the annealing of complementary DNA and RNA molecules and may well be a factor that facilitates a variety of intermolecular RNA-RNA interactions in the spliceosome (Guthrie, 1991). In addition to splicing, hnRNP A1 may function in the nucleocytoplasmic translocation of mRNA, since it has been found to shuttle between nucleus and cytoplasm, and is bound to mRNA in the cytoplasm (Pinol-Roma and Dreyfuss, 1992; Izaurralde *et al.*, 1997).

In agreement with its potentially wide repertory of physiological roles, hnRNP A1 shows an unusual structural diversity and a multiplicity of potential nucleic acid-binding surfaces. Human hnRNP A1 is 320 amino acid long (34 kDa); residues 2 through 195 are identical with calf thymus UP1 (Unwinding Protein), a nucleic acid helix-destabilising protein, which is the proteolytic product obtained when mammalian tissue is the source of protein. UP1 contains two globular subdomains, RRM 1 and 2 (discussed in Chapter 1), joined by a flexible linker. The crystal structure analysis (Shamoo *et al.*, 1997; Xu *et al.*, 1997) demonstrated that two independently folded RRMs in UP1 are held together in a fixed geometry, enabling the two RRMs to function as a single entity in binding RNA. The UP1 displays a clear preference for single-stranded nucleic acids and binds RNA through a non-cooperative mode (Casas-Finet *et al.*, 1993; Shamoo *et al.*, 1994).

The 135-residue C-terminal domain of hnRNP A1 is glycine-rich, and apparently more sensitive to proteolysis than the UP1 domains. The C-terminus of hnRNP A1 contains 45% glycine and the repeating sequence GN(F/Y)GG(S/G)RG, within which is found a regular spacing of both positively charged and aromatic residues (Cobianchi *et al.*, 1988). There are also six Arg-Gly-Gly (RGG) repeats termed the RGG box (Kiledjian and Dreyfuss, 1992). The RGG box is thought to be a general RNA binding motif and is often found in proteins that possess other RNA binding motifs (Burd and Dreyfuss, 1994a). The RNA-binding elements of the C-terminus appear to drive hnRNP A1 co-operative binding mode (Casas-Finet *et al.*, 1993) and its strand annealing activity (Kumar *et al.*, 1990; Munroe and Dong, 1992). Computer prediction based on amino acid sequence suggested that the C-terminal domain is very flexible, with high β-turn and random coil probability, and no predicted α-helices (Wilson *et al.*, 1987).

The dissimilar structure of the UP1 and C-terminal domains of hnRNP A1 suggests that they exhibit distinct contributions to the binding properties of the whole protein. The two RRMs of hnRNP A1 contribute about 50% of A1's free energy for RNA binding (Shamoo *et al.*, 1994) while the remaining binding energy is derived from co-operative A1-A1 and A1-nucleic acid interactions located in the carboxy-terminal domain (Nadler *et al.*, 1991; Kumar *et al.*, 1990; Shamoo *et al.*, 1994).

The co-operativity, mentioned above, is an important characteristic of RNA-binding properties of hnRNP A1. Based on the experimental data of Wilson and co-workers (Cobianchi *et al.*, 1988; Nadler *et al.*, 1991; Casas-Finet *et al.*, 1993), it is commonly accepted that hnRNP A1 binds to RNA co-operatively. The effect of co-operativity logically matches a concept of packaging of pre-mRNA, and therefore these data were not challenged. In the method employed, the relative changes in fluorescence of poly(ethenoadenylic acid) upon binding of recombinant hnRNP A1 or its derivatives were monitored. Binding isotherms were analysed by using the extended Scatchard equation according to the linear co-operative model of McGhee and von Hippel (1974). The typical bell-shaped Scatchard plot was considered as an evidence for co-operative binding. However, the authors themselves admit that the model may not be fully descriptive for the RNA-protein complexes formed in this case (Casas-Finet *et al.*, 1993).

The function of conserved domains of hnRNP A1 and other hnRNP A/B proteins in splicing was questioned using an alternative splicing assay, RNA-binding and annealing assays (Mayeda *et al.*, 1994; 1998). The structural requirements for hnRNP A1 function were determined by mutagenesis of the corresponding cDNA. It was proposed that two conserved Phe residues in the RNP-1 submotif of each of two RNA recognition motifs are involved in specific RNA-protein interactions and are essential for modulating alternative splicing, although, these residues were not required for general pre-mRNA binding or RNA annealing activity. The C-terminal Gly-rich domain was necessary for alternative splicing activity, for stable RNA binding and for optimal RNA annealing activity. Analysis of variant proteins with duplications, deletions, or swaps of the RRMs showed that although both RRMs are required for alternative splicing function, each RRM plays a distinct role, and their relative position is important. Moreover, RRM2 but not RRM1 could support alternative splicing switch when duplicated, despite their very similar structure.

HnRNP A1 corresponding pre-mRNA can be alternatively spliced by the inclusion of the additional exon producing the $A1^B$ isoform with a longer Gly-rich domain, which is essentially another constituent of the 40S hnRNP monoparticle, designated as B2 (Buvoli *et al.*, 1990a). This alternative isoform was shown to have more stronger RNA binding but limited alternative splicing activities compared to A1 (Mayeda *et al.*, 1994). And, conversely, hnRNP A2 and B1, which have 68% amino acid identity with hnRNP A1, bind more weakly to pre-mRNA and have stronger splice site switching activities than hnRNP A1 (Mayeda *et al.*, 1994).

The physiological role of A1 in the cell is a matter of long going dispute and mainly the question of physiologically relevant sequence-specific RNA-binding activity. From one side, the contiguous distribution of hnRNP complexes along the length of nascent transcripts and remarkably high intranuclear concentration of hnRNP A1 in actively growing cells (about 2 mM) suggested that hnRNP A1 had a limited role as an RNA packaging protein possessing no sequence specific RNA binding activity. On the other hand, starting from the work of Swanson and Dreyfuss (1988b), several laboratories began to search for hnRNP A1 high affinity sites with a distinct physiological role. In their work Swanson and Dreyfuss showed that monoclonal antibodies against hnRNP A1 immunoprecipitate from nuclear extract the

RNase T1-generated fragments of pre-mRNA containing the conserved polypyrimidine stretch between the branch site and the 3' splice site. Thus, some functional significance in hnRNP A1 specific binding to the 3' end of introns was proposed. Next, intensive quantitative biochemical studies were performed with recombinant hnRNP A1 to reveal its binding properties (Cobianchi *et al.*, 1988; Casas-Finet *et al.*, 1993; Burd and Dreyfuss, 1994; Abdul-Manan and Williams, 1996a, b). The effect of co-operativity was postulated (Cobianchi *et al.*, 1988; Casas-Finet *et al.*, 1993), although the actual values describing the parameters of binding (an affinity of only about $1.5x10^5$ M^{-1} for poly r(εA)) were too low for that the interaction can be detected by most non-equilibrium binding assays and account for co-operativity. Overall, the issue of whether there are functional high affinity sites is still obscure. This matter is discussed later in Discussion (2.3).

Objectives and Strategy

Essential splicing factor SF2/ASF regulates alternative splicing by promoting the use of the proximal 5' splice site, whereas hnRNP A1 exhibits the opposing effect by activation of the distal 5' splice site. The hypothesis was proposed that variations in the intracellular levels of these antagonistic factors influence different modes of alternative splicing and might be a natural mechanism for tissue specific or developmental regulation of gene expression (Caseres and Krainer, 1997). This hypothesis is commonly accepted but the underlying mechanism that explains it is still unknown.

I had started my project in 1994 when the experimental data were rationalised by a model proposed by Eperon *et al.* (1993). The model postulates that the occupancy of the potential 5' splice site by the U1 snRNP determines the use of the site. The occupancy, in turn, is determined by the affinity of U1 snRNP to a particular site, and thus, sites could be divided on strong or weak ones. When two competing 5' splice sites are strong both sites are occupied by U1 snRNPs simultaneously. The double occupancy leads to the use of the downstream site and this is the cause of the proximity effect seen with strong alternative splice sites. The proposed scheme suggests that for weak competing 5' splice sites the choice depend on the probability of the occupancy or, in other words, U1 snRNP dissociating rates. The experimental data suggested that SF2/ASF increase the general affinity of U1 snRNP for 5'

splice sites indiscriminately (Eperon *et al.*, 1993; Kohtz *et al.*, 1994). Thus, both alternative sites are occupied and the proximal site is used. This model considered the selection of 5' splice site mediated by U1 snRNPs and SF2/ASF but not the antagonism between SF2/ASF and hnRNP A1. The role of hnRNP A1 was not addressed specifically but it was proposed that hnRNP A1 weakens U1 snRNPs binding to the 5' splice site.

The objective of my work was to study the interaction of hnRNP A1 and SF2/ASF with the long RNA substrate that mimics pre-mRNA. This was a novel investigation because most of previous data were obtained for the short RNA substrate or RNA oligonucleotides. Not all the methods used for the short RNAs could be directly applied for the long substrates where multiple binding opportunities are present, therefore I tried to employ as many as possible different techniques and the emphasis was on the comparison of hnRNP A1 binding with SF2/ASF.

The project addresses the questions of the distribution of these proteins along the pre-mRNA; the thermodynamic and kinetic parameters of RNA-protein interaction and possible co-operativity in binding. The ambition was to investigate the possibility that the antagonistic effect of hnRNP A1 and SF2/ASF on splicing could be solely explained at the level of RNA-binding properties of these proteins.

2.2. RESULTS

The emphasis of this project was on *in vitro* characterisation of hnRNP A1 and SF2/ASF binding to pre-mRNA substrates. This RNA provides multiple binding opportunities for the splicing factors, in contrast to most of the previous investigations carried out with short RNAs or RNA oligonucleotides.

There are many methodological approaches available for studying the RNA-binding properties of proteins. It is possible to divide these approaches into qualitative and quantitative ones. The former includes a comparison of several proteins or mutants of one protein in their ability to bind to certain RNA substrate, or, conversely, a comparison of different RNA sequences in interaction with one protein. This comparison is judged qualitatively by the presence or absence of an effect. Quantitative approaches are devoted to measurements of the thermodynamic and kinetic parameters of interactions between two interacting partners.

There is a wide range of different methods, starting from the nitro-cellulose filter binding assay (NCFBA) and the electrophoretic mobility shift assay (EMSA), the conventional and relatively simple methods which require only basic laboratory equipment, to sophisticated biophysical methods such as the analytical ultracentrifugation (Ralston, 1993; McRorie and Voelker, 1993) and fluorescent spectroscopy (Jameson and Eccleston, 1997; Robertson *et al.*, 1988). The usage of that or another method is determined not only by the availability of the technique but also by the objective. The main drawback of the simple methods is a problem of equilibrium, which could be disturbed during detection of the complex formation. For example, some RNA-protein complexes could be destroyed during electrophoresis (EMSA), by washing procedure in the NCFBA, or applied force in ultracentrifugation. The choice of the particular method is influenced by the nature of interaction, namely, the number of binding sites, their affinities and interdependence.

To study the hnRNP A1 and SF2/ASF interaction with pre-mRNA a number of methods were employed, which are presented and discussed below in separate paragraphs.

2.2.1. RNase H assay

The RNase H assay was developed for the quantitative study of the RNA-protein interactions in the laboratory of Dr. I. Eperon. The main advantage of this assay is that it allows one to determine the specific site on pre-mRNA to which the protein is bound and to assess the dissociation rate from it, without being affected by interactions of proteins with other sites on RNA. RNase H is an enzyme, which recognises and cleaves RNA in the RNA-DNA heteroduplexes. In the splicing field, RNase H is commonly used in DNA-oligo directed extermination of specific RNA in nuclear extract. In a simplified *in vitro* system, consisting of RNA and the protein of interest, the sequence-specific DNA oligonucleotide (approximately 14 nt) would anneal to the cognate site on the RNA and compete with the protein for the binding to this site. If a protein is pre-bound to the site tested the RNase H cleavage is dependent on accessibility of the site for the DNA-oligo, and, therefore, the time course of RNase H cleavage reflects the dissociation rate of the protein from this site. In other words, if the protein interacts with RNA, the overall process of cleavage is expected to slow down if compared to the control reaction without a protein added.

HnRNP A1 and SF2/ASF affect 5' splice site selection and, therefore, the particular interest were to investigate the comparative binding of these proteins to the 5' splice site on the pre-mRNA. The interactions of these proteins with the branch-point and the 3' splice site were also investigated with the complementary oligonucleotides.

The pre-mRNA substrate chosen for this project, named C175G, was previously constructed in the laboratory and well characterised in a number of studies of the 5' splice site selection (Eperon *et al.*, 1986; Cunningham *et al.*, 1991; Eperon *et al.*, 1993). This substrate is 533 nt long and derived from rabbit β-globin IVS-2 by inserting an extra 5' consensus splice site (C) and a spacer sequence of 151 nt long upstream of the natural β-globin 5' splice site (G). Moreover, constructs originating from the rabbit β-globin IVS-2 were used by other laboratories investigating 5' splice site selection (Nelson and Green, 1988) and therefore this sequence is one of the most characterised. In addition, this substrate was used in experiments that led to the formulation of the model for 5' splice site selection by U1 snRNPs (Eperon *et al.*, 1993). Thus, the characterisation of binding of hnRNP A1 and SF2/ASF to C175G

substrate should shed light on the model. The RNA was co-transcriptionally labelled with [^{32}P]-UTP as described in the Materials and Methods.

Both hnRNP A1 and SF2/ASF were purified recombinant proteins expressed in *E. coli*. SF2/ASF was kindly provided by Dr. A. Krainer, Cold Spring Harbour Laboratory (Krainer *et al.*, 1991). HnRNP A1 was expressed from pET 9C plasmid, obtained from Dr. A. Krainer (Mayeda and Krainer, 1992), and purified according to Cobianchi *et al.* (1988) by the procedure described in the Materials and Methods.

All the experiments presented in this section were carried out in a similar manner as described below. The protein binding to the site tested was monitored by comparison of two reactions: (1) RNase H digestion of RNA in the presence of protein, (2) RNase H digestion in the absence of protein. In (1) the protein was preincubated with RNA in high excess over RNA at 30 °C before an addition of oligonucleotide. The RNase H digestion was triggered by the addition of an oligonucleotide. Aliquots were taken out at different time intervals and the reaction was stopped. When all time-points had been collected they were subjected to proteinase K treatment; RNA was ethanol precipitated and analysed by 6% urea-PAGE. The dried gel was exposed to a PhosphoImager and bands corresponding to the pre-mRNA and the products of digestion were quantified.

Figure 2.1 shows the results of hnRNP A1 binding to the 5' splice site and the branch-point monitored as a time-course of RNase H digestion. The scheme of RNA cleavage with the expected size of fragments is presented at the top of the figure. The panel A gives an example of image of the whole gel and the panel B shows results of quantification which are plotted as a percent of uncleaved RNA versus time. The result of this experiment is that hnRNP A1 protein does not protect RNA from the oligonucleotide-directed cleavage by RNase H at both the 5' splice site and the branch-point. There are several possible explanations of obtained results: (1) hnRNP A1 does not bind RNA at all (due to the poor preparation of recombinant protein); (2) hnRNP A1 does not interact with the 5' splice site and the branch-point but binds to other sequences on the RNA; (3) the RNase H assay is not able to detect hnRNP A1 interaction with RNA because the protein can be easily displaced by an oligonucleotide. It is important to note that the RNase H digestion is a non-equilibrium

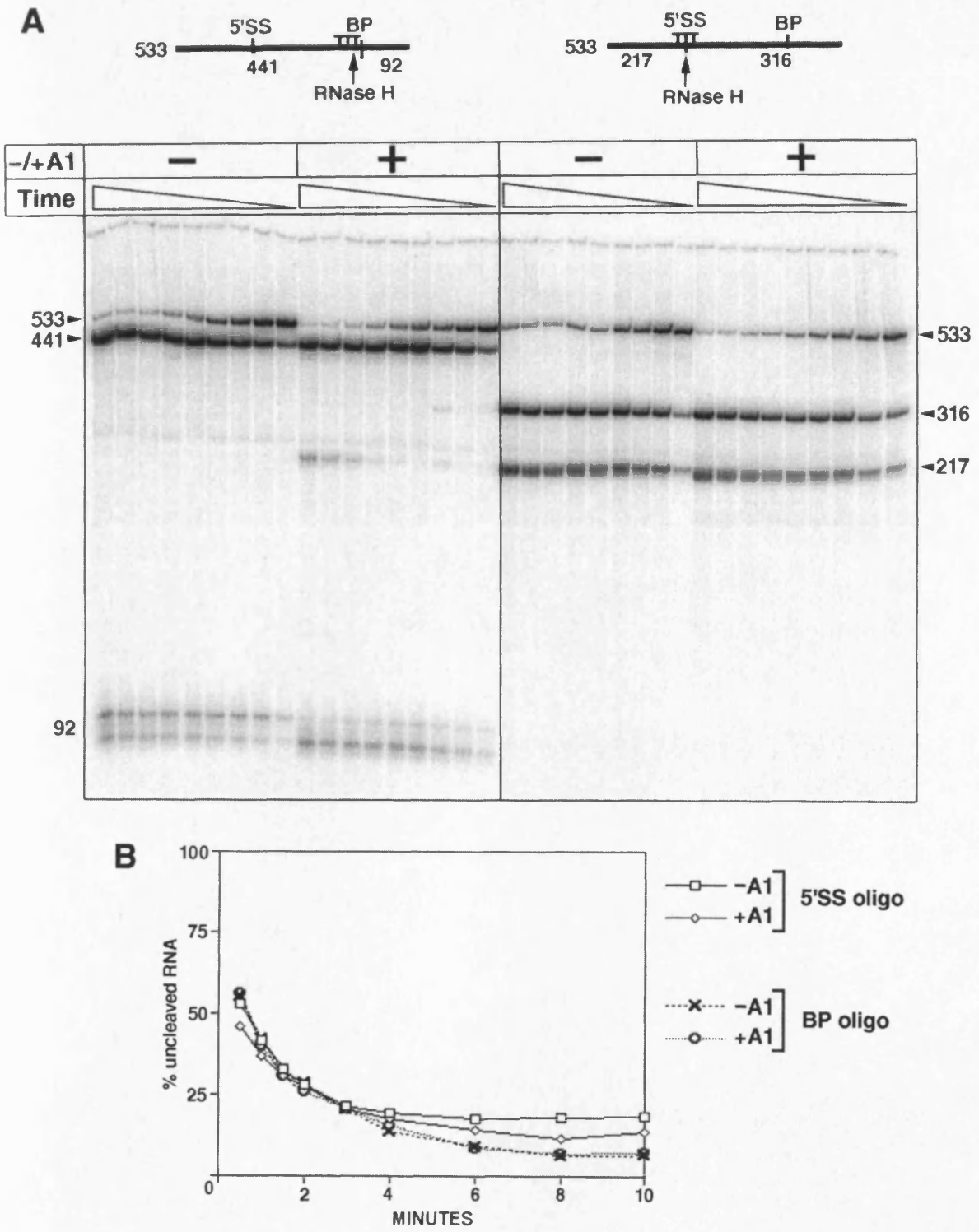


Figure 2.1. HnRNP A1 does not affect the RNase H cleavage of the pre-mRNA **A:** (*at the Top*) A scheme of RNase H cleavage is presented. The positions of the oligonucleotides (BP - branch point, 5'SS - 5' splice site) and the sizes of the generated fragments are indicated. (*Lower part*) The progression of RNase H cleavage at 0.5, 1, 1.5, 2, 3, 4, 6, 8, and 10 minutes was monitored, and the images of the full-sized gel obtained by exposure to a PhosphoImager are shown. The bands corresponding to the uncleaved RNA and the major product were quantified and the percent of uncleaved RNA was calculated as described in the Materials and Methods. **B:** The graphical representation of the results.

reaction, and as soon as the site on RNA is vacant, the oligonucleotide anneals and cleavage occurs. In contrast, binding of protein to RNA is an equilibrium reaction. Therefore, this RNase H assay reflects the dissociation rate of the protein. Thus, the absence of protection of RNA by hnRNP A1 could be interpreted that the protein has a high dissociation rate.

In the next experiment (Fig. 2.2) SF2/ASF protein was assayed by the same technique and products of RNase H digestion were analysed on the similar gel as depicted in Figure 2.1A. The panel A (Fig. 2.2) shows the enlarged part of the gel where the interaction of SF2/ASF with the branch-point is presented. In contrast to hnRNP A1, the protection of the site by SF2/ASF is evident. The same protection was obtained for the 5' splice site. The quantified results for both sites are presented in Panel B (Fig. 2.2). Interestingly, the protection of both sites by SF2/ASF appears to be the same indicating that binding of the protein to the 5' splice site and branch-point is similar. SF2/ASF binding to this RNA was expected but the similar protection of two unrelated sites raised certain questions. (1) It is possible that this preparation of recombinant protein binds to RNA promiscuously but it would contradict to the data of Zuo and Manley (1994). They demonstrated, using UV-crosslinking and gel shift assays, that point mutations in the 5' splice site consensus can reduce binding by as much as a factor of 100. (2) There is another possibility, that SF2/ASF can corrupt RNase H activity. The latter was investigated in the independent experiment (Fig. 2.3). RNase H was pre-incubated with SF2/ASF for 30 minutes and then a "ready to use" substrate, RNA with the annealed oligonucleotide, was added to the reaction. Comparison of the control RNase H digestion and in the presence of SF2/ASF did not reveal any differences, whereas pre-incubation of SF2/ASF with RNA confirmed the effect of protection. Thus, SF2/ASF does not affect ability of RNase H to cleave.

The problems of the RNase H assay, especially the inability to detect A1 binding, suggested that other techniques should be tested.

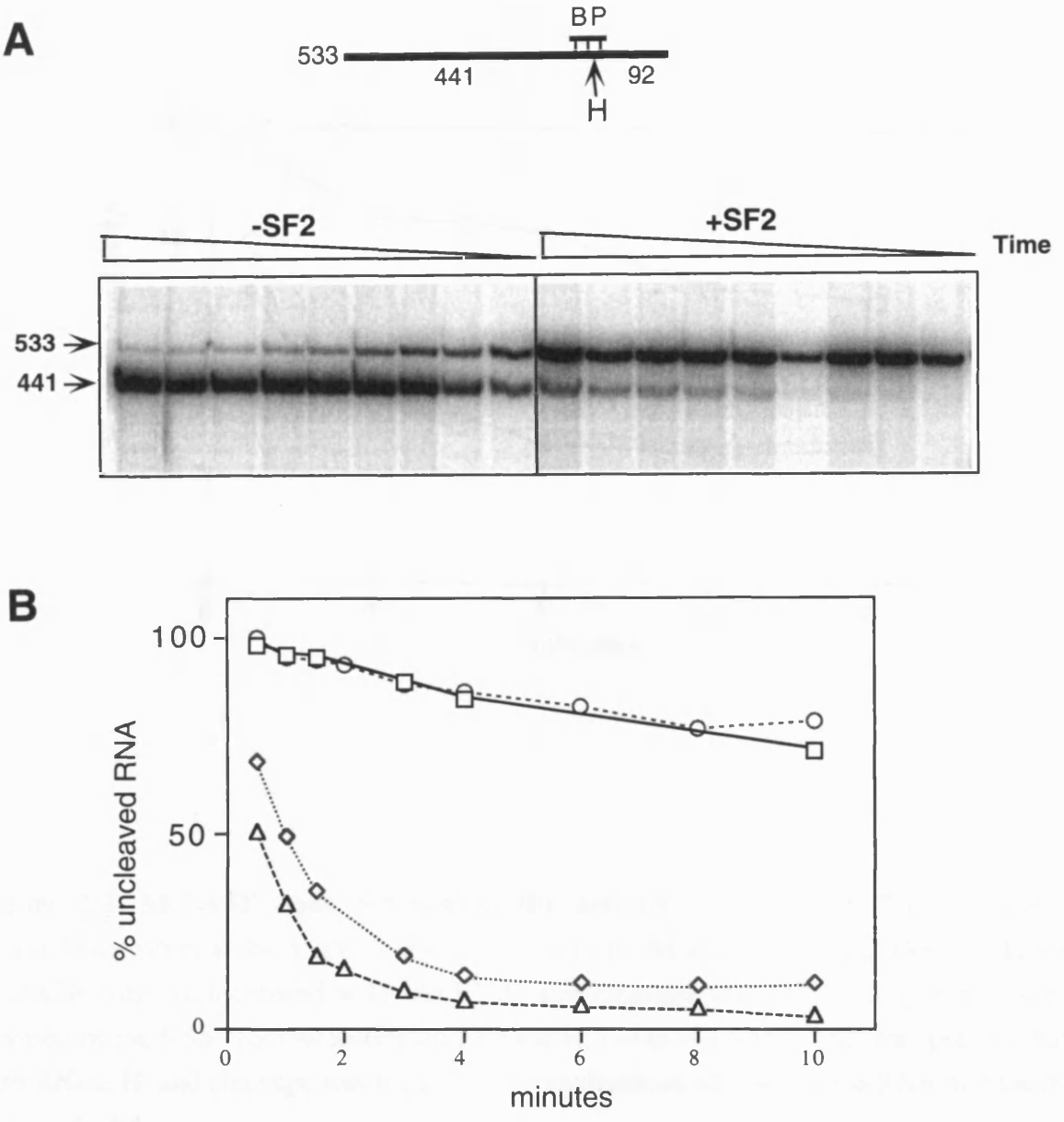


Figure. 2.2. Effect of SF2/ASF on RNase H digestion. A: Time course of RNase H digestion at the branch point analysed by 6% urea-PAGE. The image contains an enlarged part of the gel similar to that depicted in Fig. 2.1, and shows only the uncleaved RNA and the major product of digestion which were used for quantification. **B:** Quantification of RNase H digestion at the branch point (the gel is presented in panel A) and at the 5' splice site (the gel is not shown). (····○····) and (----Δ----) - cleavage of the RNA is triggered by addition of the branch point oligonucleotide in the presence or absence of SF2/ASF, respectively. (—□—) and (·····◇·····) - cleavage is triggered by the 5'SS oligonucleotide in the presence or absence of SF2/ASF, respectively.

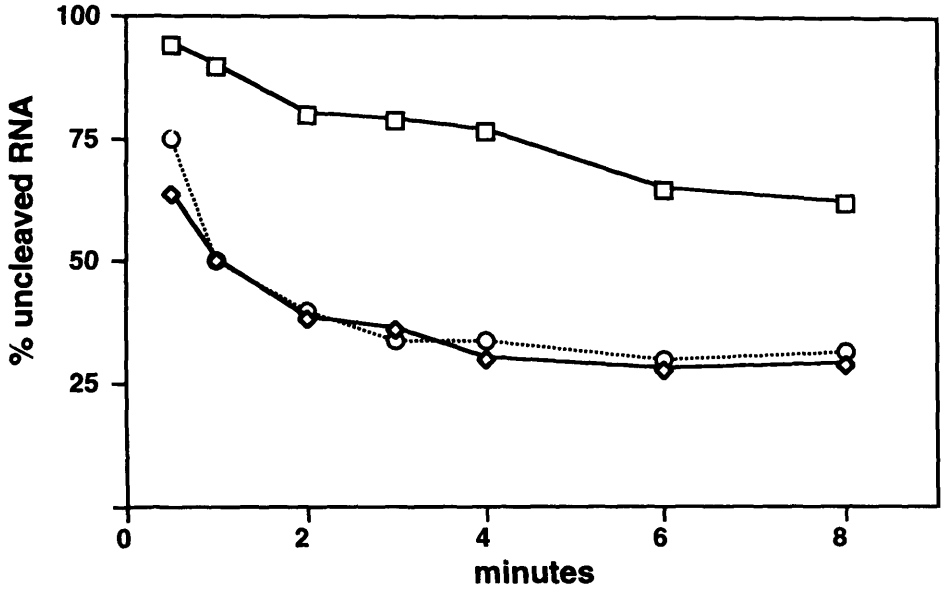


Figure 2.3. SF2/ASF does not inhibit the activity of RNase H. Time courses of RNase H digestion at the 5' splice site: (·····○·····) - in the absence of SF2/ASF; (—□—) - SF2/ASF was pre-incubated with the RNA, and cleavage was triggered by simultaneous addition of the 5'SS oligonucleotide and RNase H; (—◇—) - SF2/ASF was pre-incubated with RNase H, and cleavage was triggered by simultaneous addition of the RNA and the 5'SS oligonucleotide.

2.2.2. Nitrocellulose filter binding assay for interaction of hnRNP A1 with pre-mRNA

Since the recombinant hnRNP A1 did not affect the RNase H cleavage we had to rule out a possibility that this preparation of the protein is not functionally active in RNA binding. There are several conventional methods to measure an RNA binding activity of proteins, of which the electrophoretic mobility shift assay (EMSA) and the nitrocellulose filter binding assay (NCFBA) are most widely used. The gel retardation approach is usually applied for interaction between a protein and oligonucleotides or short RNA molecules (up to 200 nt) because the migration behaviour of long RNA molecules on native gels is complicated, the bands are usually diffuse and the separation of complexes is difficult to achieve (discussed in Black *et al.*, 1998). The NCFBA allows many experimental points to be obtained, and does not require complexes to remain stable during electrophoresis, and, in general, is a somewhat simpler assay than the EMSA. Moreover, all previous attempts in the laboratory to establish the EMSA for similar pre-mRNA had failed. Therefore, the NCFBA was used.

The nitrocellulose filter binding assay is based on a chemical property of nitrocellulose to absorb proteins and RNA-protein complexes but not free RNA molecules. In this experiment RNA was radioactively labelled as before and incubated with hnRNP A1 protein to achieve equilibrium in binding. During filtration unbound RNA pass through the nitrocellulose filter and the amount of radioactivity retained on filter was used to quantify the RNA-protein complex. Figure 2.4A shows the saturating binding of hnRNP A1 to the RNA carried out in a constant volume with increasing amounts of the protein. This experiment clearly demonstrated that this preparation of recombinant hnRNP A1 is active in binding to this RNA, and at least 70% of the RNA molecules could be bound. Bovine serum albumin (BSA) does not bind RNA and serves here as a negative control.

The filter binding assay is also quantitative, which allows an equilibrium association constant to be measured if the interaction can be described as a simple bi-molecular reaction (alternatively referred to as monovalent system, Klotz, 1986).

$$P+RNA=P\text{-}RNA$$

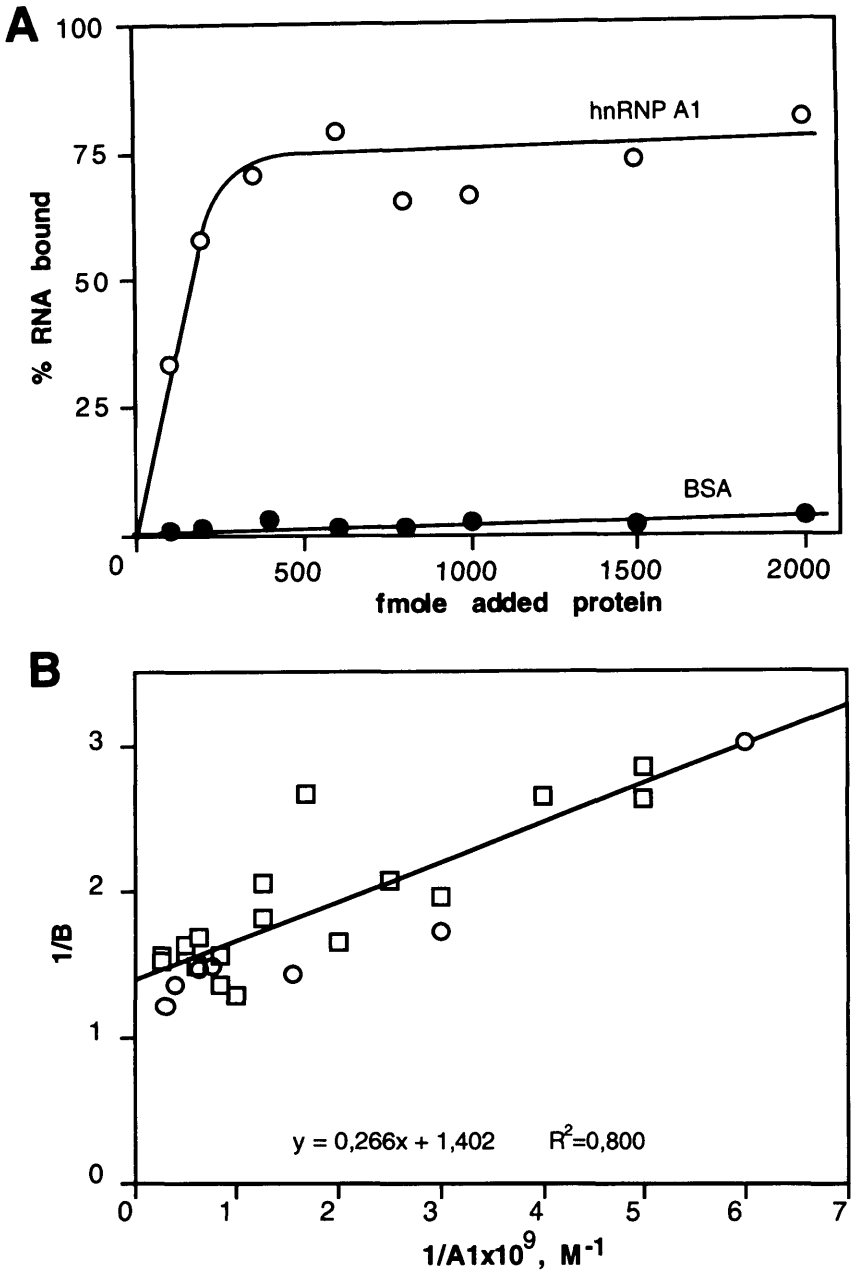


Figure 2.4. Nitrocellulose filter binding assay for the interaction of recombinant hnRNP A1 with pre-mRNA. A: Saturating binding of the hnRNP A1 to RNA. The incubation mixtures contained 2 fmol of labelled RNA with a specific activity of 2000 cpm/fmol and varied amounts of protein (100-2000 fmol) in 600 µl of buffer D-Glu, and were processed as described in the Materials and Methods. (-O-) - hnRNP A1, (-●-) - BSA. **B:** Binding of hnRNP A1 to RNA is presented in double reciprocal co-ordinates. (-O-) - points from the experiment depicted in (A). (-□-) - data obtained by the method of variable volume. The incubation mixtures of different volumes of buffer D-Glu (0.1-2.5 ml) contained 0.5 pmol of hRNP A1 and 2 fmol of labelled RNA, and were processed as described in the Materials and Methods.

This is the case when both protein and RNA contain only a single binding site for interaction to each other. However, if there are multiple binding sites on the RNA molecules, which is likely to represent the situation for interaction of hnRNP A1 with long pre-mRNAs, the NCFBA has the disadvantage of measuring only total binding (Black *et al.*, 1998). Nevertheless, I decided to estimate an apparent equilibrium association constant. The association constant is determined by the equation:

$$K=[P\text{-}RNA]/[P][RNA],$$

where [P-RNA] is a concentration of the RNA-protein complexes, [P] and [RNA] are concentrations of free protein and RNA, respectively. The ideal case of a single binding site is described by the equation

$$B=\gamma K[P]/(1+K[P])$$

where B is a fraction of bound RNA and γ is a fraction of RNA active in binding (Klotz, 1986; McGhee and von Hippel, 1974). This equation may be transformed into leaner form:

$$1/B=1/\gamma+1/(\gamma K[P])$$

A graph of 1/B versus 1/[P] is linear with an intercept on the ordinate axis of $1/\gamma$ and a slope of $1/(\gamma[K])$.

The association constant was measured in another type of experiment (Fig. 2.4B). To achieve different concentrations of components, the volume was varied, which is more accurate than the method of constant volume. The incubation mixtures contained constant amount of components (protein is in axcess over RNA), and changes in concentration of components are achieved by the variation of volume. It is important to stress that, despite the simplification, this approach gives the correct value of the active RNA, which does not depend on how many binding-sites are present on the RNA molecules, and the K value also includes the correction for the fraction of active RNA. The data presented in Figure 2.4B were analysed by the Cricket Graph computer program using linear extrapolation. It yields a K value of 5.3×10^9 M^{-1}, and that 71% of RNA molecules are active for binding to hnRNP A1. It is remarkable that this K value nicely correlates with the affinity of hnRNP A1 to the SELEX winner RNA oligo (Burd and Dreyfuss, 1994b). At the same time, this method does not allow to the following situations to be distinguished: (1) If both components contain only single site for interaction and, in this case, one molecule of protein binds to one RNA molecule with the

association constant determined above; (2) there are several independent A1-binding sites on RNA with fixed affinities (k), then the apparent K will be a sum of the individual k. (3) The more complicated situation, that is, if the binding of the protein to the RNA is co-operative. Then the measured constant does not characterise the nature of interaction.

2.2.3. UV-crosslinking

The method of the direct UV-crosslinking of proteins to continuously labelled pre-mRNA was applied for the following reasons. Since the RNase H assay could not be used, the site-specific binding of hnRNP A1 and SF2/ASF to the pre-mRNA was not addressed, and only total binding was assayed at this stage. UV-crosslinking has certain advantages over the NCFBA. It allows the visualisation of proteins being involved in the interaction with RNA by SDS-PAGE followed by autoradiography. The relative amount of crosslinked protein could be measured by quantitation on the PhosphoImager. Notably, if hnRNP A1 and SF2/ASF are present in the incubation mixture and interact with labelled RNA then both proteins could be visualised on the same gel simultaneously.

Despite the fact that the method of direct UV-crosslinking is widely used for the study of DNA- and RNA-protein interactions, the molecular mechanism of the reaction is not well understood. Data obtained during the study of DNA-protein crosslinking in isolated chromatin suggested the preferential involvement of thymine (Thy) and cytosine (Cyt) bases in crosslinking with various amino acids (Dizdaroglu *et al.*, 1989; Dizdaroglu and Gjewski, 1989). It is believed that the reaction is initiated by hydroxyl radical (OH radical) produced from water by ionising radiation (Mee and Adelstein, 1981). Evidence indicates that the chemical bonds involved in OH radical-induced DNA-protein crosslinks are covalent (Mee and Adelstein, 1981; Oleinick *et al.*, 1987). The OH radical-induced crosslinking of Thy and Cyt to amino acid tyrosine (Tyr) was intensively studied by variety of methods: gas chromatography-mass spectrometry (GC-MS), high-performance liquid chromatography (HPLC) and nuclear magnetic resonance (NMR). In this work the products of reaction induced by OH radicals were identified and the possible mechanisms of the Thy-Tyr crosslink were proposed (Margolis *et al.*, 1988; Gajewski and Dizdaroglu, 1990).

Regarding the RNA-protein crosslinks, a detailed analysis of the chemical reaction has not been carried out but it was postulated that mainly uracil and cytosine can be crosslinked to proteins. A lot of data on RNA-protein crosslinking were collected during structural studies of ribosomes (reviewed by Brimacombe, 1991). Recently, crosslinking sites at the amino acid and nucleotide level were determined by N-terminal amino acid sequence analysis in combination with matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) (Urlaub *et al.*, 1995; Urlaub *et al.*, 1997). Surprisingly, most of identified crosslinked amino acids were lysines but not tyrosines or phenylalanines. It is interesting, that, for proteins where the three-dimensional structure is available, the amino acids crosslinked to the RNA are located in the flexible loops but not in the structurally conserved elements like β-sheets or α-helixes (Urlaub *et al.*, 1995; Urlaub *et al.*, 1997).

2.2.3.1. The concentration dependence of the SF2 and A1 binding to pre-mRNA detected by UV-crosslinking

To gain information in what range of concentrations RNA-binding of hnRNP A1 and SF2/ASF can be readily detected by means of direct UV-crosslinking an equilibrium binding assay was performed. A set of reactions was set up, containing a constant amount of [^{32}P]-UTP labelled RNA (10^{-10} M) and different amounts of protein, ranging from 3 x 10^{-6} M to 10^{-9} M. After an equilibrium was reached, the reactions were UV-irradiated at 254 nm on the SpotCure device, which directs light into the well of a microtitre plate without dispersion of energy. During UV irradiation some of the protein molecules became covalently attached to the RNA, and therefore, after treatment with RNases, these proteins are radioactively labelled with small pieces of RNA, which do not change substantially the migration mobility of proteins on SDS-PAGE. Hence, the crosslinked protein can be visualised by exposure of the gel to a PhosphorImager and its amount can be quantified in relative units of detection (V). Figure 2.5A shows an example of autoradiographic images of the gel corresponding to crosslinked hnRNP A1 and SF2/ASF, and Figure 2.5B represents the quantification of these images. The data are plotted as V (arbitrary units of detection) versus logP, where P is the

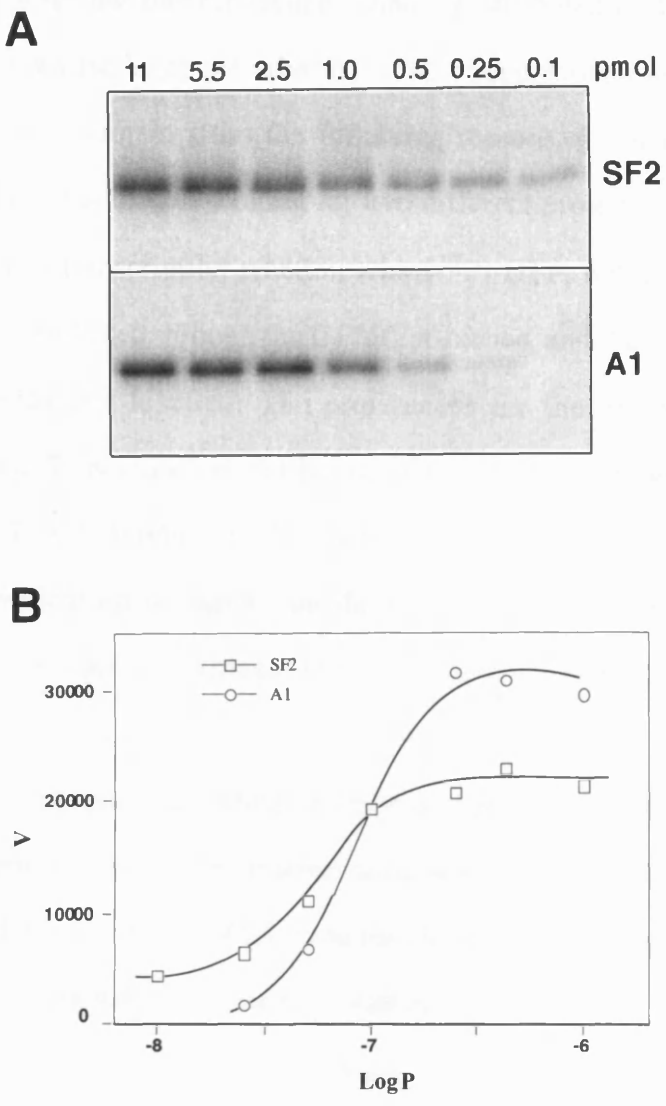


Figure 2.5. Comparison of the equilibrium binding of hnRNP A1 and SF2/ASF to pre-mRNA by UV-crosslinking. The incubation mixtures contained 5000 cpm of labelled RNA (approximately 1-2 fmol) and varied amounts of protein (0.1-11 pmol) in 10 µl of buffer D-Glu. The reactions were incubated for 15 min at 30°C, UV-irradiated for 30 seconds, treated with RNases and subjected to 12% SDS-PAGE followed by autoradiography as described in the Materials and Methods. **A:** The autoradiographic images of the gels for SF2/ASF and hnRNP A1. **B:** The amount of radioactivity in each band was quantified using a PhosphoImager, and the data are plotted as relative units of detection (V) versus log P, where P is the concentration of protein.

amount of protein added into the reaction. It should be noted, that V_{max} for hnRNP A1 is higher then for SF2/ASF, but this difference cannot be attributed to the stronger binding of hnRNP A1 to RNA because V, as the arbitrary units of detection, could not be transformed into the moles of bound protein from the following reasons. (1) It is difficult to provide evidence that the efficiency of crosslinking for two different proteins is equal. (2) The RNA was labelled in *in vitro* transcription reaction with [^{32}P]-UTP, but uridine residues are not absolutely randomly distributed through the C175G sequence, and therefore, the RNA can not be considered as uniformly labelled. The preferences for the hnRNP A1 and SF2/ASF binding, except high affinity sites, are not known and could be U-dependent, so the level of crosslinking will reflect U-binding preferences (Nikogosyan, 1990; Hockensmith *et al.*, 1991). (3) The quantification of signals on the gel by a PhosphorImager does not allow comparison of the amount of radioactivity present in different gels from different experiments.

Nevertheless, both proteins exhibit the typical sigmoidal (S-shaped) binding curves, and it opens an opportunity for further mathematical analysis of the UV-crosslinking data to reveal the RNA binding properties of the proteins (Klotz, 1986). The saturation in binding allows to introduce the parameter, fractional saturation (Y),

$$Y=V/V_{max}$$

which eliminates the arbitrary units, and therefore, the binding of hnRNP A1 could be compared to the SF2/ASF binding from the results of different experiments. Figure 2.6A shows a comparison of hnRNP A1 and SF2/ASF binding as Y versus log P.

By the method of UV-crosslinking, it is not possible to evaluate such parameter as the number of binding sites for the proteins on the RNA as discussed above. From the other side, the parameter Y actually describes the fraction of binding sites on RNA, occupied by protein. Thus, we can determine the concentrations of hnRNP A1 and SF2/ASF when half of their binding sites are occupied. Interestingly, these values are very similar for both proteins (approximately $6x10^{-8}$ M). It means that both proteins could be used in the same range of concentration in experiments when both proteins are present in the incubation mixture. The result, that both proteins occupied 50% of the sites at the same concentration, tempted us to speculate on its physiological significance. If the counteracting effect of hnRNP A1 and

SF2/ASF on alternative 5' splice site selection can be explained by their binding to the RNA, then small variations in the relative concentrations of these proteins would modulate the effect.

To analyse the binding curves on Figure 2.6A, it is convenient to visualise binding in terms of deviations from the ideal curve. If a monovalent equilibrium reaction is considered as an ideal binding situation, then fractional saturation Y could be expressed as

$$Y=[P]/(K+[P])$$

where K is a dissociation constant and equals the concentration of the free protein P at which half of the protein-binding sites are occupied. Substituting the experimentally defined value of $6x10^{-8}$ M into the equation yields Y for the single site binding situation. Thus, a theoretical binding curve was built and compared with those for hnRNP A1 and SF2/ASF. The curve obtained for SF2/ASF quite closely follows the theoretical one, whereas the dissociation curve for hnRNP A1 is very distinct. The difference in concentrations when the binding of hnRNP A1 is detected and reaches saturation is one order of magnitude. Thus, the shape of the curve obtained for hnRNP A1 binding, in comparison with theoretical (monovalent) one, suggests the existence of several binding sites on RNA and theirs positive co-operative relation (McGhee and von Hippel, 1974). The positive co-operativity in binding implies that binding of one molecule of protein to the RNA increases the binding of others. In contrast to hnRNP A1, the binding of SF2/ASF to the RNA proceeds in much wider interval of concentrations and detected at concentration as low as 10^{-9} M. Hence, this indicates that binding of SF2/ASF to RNA is more selective and there are sites for SF2/ASF binding with greater affinities than for hnRNP A1. The results support an idea that C175G RNA contains at least one high affinity binding site for SF2/ASF whereas the binding of hnRNP A1 to it is less specific but co-operative.

To emphasise the observed differences the data were converted into the Hill equations. The earliest attempt to analyse co-operative binding was formulated by Archibald Hill in 1910 in order to explain the oxygen dissociation curve of haemoglobin in whole blood. The Hill equation is useful to characterise the binding of **n** molecules of ligand to the substrate and

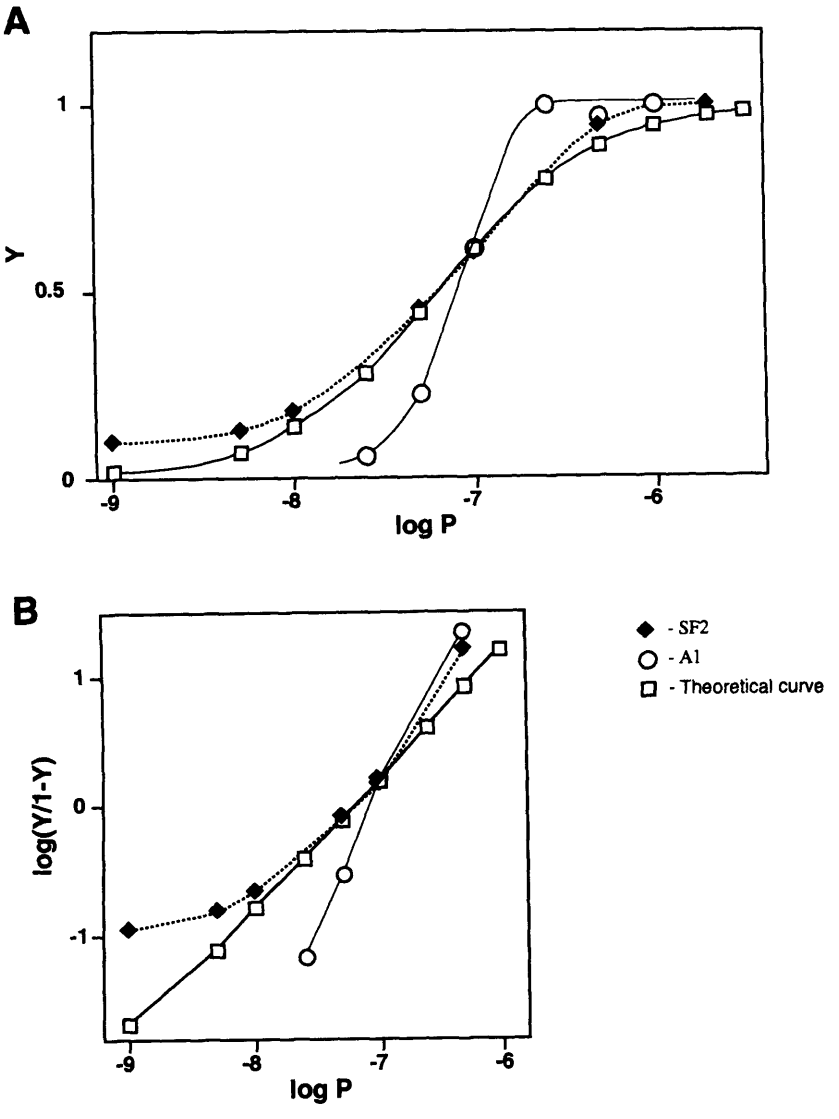


Figure 2.6. Graphical analysis of the equilibrium binding of hnRNP A1 and SF2/ASF to pre-mRNA. A: The data from several experiments similar to the one depicted in Fig. 2.5 are summarised and plotted as a function of fractional saturation (Y) versus logarithm of protein concentration (P). B: The data are presented as the Hill plot. The theoretical curve is generated for a monovalent reaction.

the co-operative behaviour of this binding. The Hill equation is derived with assumption of infinite ligand binding co-operativity,

$$nP+RNA=P_n\text{-}RNA$$

that is, either all or none of the ligand binding sites are occupied. The dissociation constant for this reaction is

$$K=[P]^n[RNA]/[P_n\text{-}RNA]$$

and its fractional saturation is expressed as

$$Y=n[P_n\text{-}RNA]/n([RNA]+[P_n\text{-}RNA])$$

or upon algebraic rearrangement becomes the Hill equation:

$$Y=[P]^n/(K+[P]^n).$$

The Hill equation can be rearranged as follows

$$Y/(1-Y)=[P]^n/K$$

and graphically presented as the Hill plot, the plot of log[Y/(1-Y)] versus log[P]. Figure 2.6B shows the Hill plots for hnRNP A1, SF2/ASF and the theoretical curve generated for the monovalent reaction. For the latter it is linear with a slope of n=1, as expected. The plot for hnRNP A1 is essentially linear with n>1. For SF2/ASF the individual points can not be fitted by a single straight line but rather there is a tendency of constantly changing **n** dependent on the concentration of the protein.

The Hill plot is a useful empirical curve-fitting relationship rather than an indicator of a particular model of ligand binding. In deriving the Hill equation, it was assumed that **n** molecules of ligand bind substrate in a single step. Although infinite binding co-operativity is a physical impossibility, nevertheless, **n** may be taken as a parameter (the Hill constant) related to the degree of co-operativity among interacting ligands rather than the number of binding sites. If n=1, the ligand-binding reaction is non-co-operative. A reaction with n>1 is described as positively co-operative: binding of one molecule of ligand increases the affinity of others for further ligand binding. Conversely, if n<1, the reaction is termed negatively co-operative: ligand binding reduces the affinity for subsequent ligand binding.

The linear Hill plot for hnRNP A1 with n>1 could be considered as an evidence for positive co-operativity in binding to the pre-mRNA. The character of the Hill plot for SF2/ASF does not allow a simplistic explanation of the nature of interaction. The binding

curve (Fig. 2.6A) for SF2/ASF probably reflects a very complex situation, that is, a collection of heterogeneous sites of fixed affinities exist on RNA, which also can produce either positive or negative effects on the other binding-sites with increasing occupancy by the protein (some examples are discussed by Klotz, 1986).

Although this method did not allow to determine the number of binding sites and their affinities it revealed drastic differences in binding properties of hnRNP A1 and SF2/ASF to the pre-mRNA. In practical terms, these data allowed experiments to be designed where both proteins are present in the incubation mixture with pre-mRNA and their binding can be monitored.

2.2.3.2. The rate of hnRNP A1 and SF2/ASF dissociation from the pre-mRNA

The nitrocellulose filter binding assay (2.2.2) has shown that the failure of hnRNP A1 to protect RNA against RNase H can not be attributed to a failure of the recombinant protein to bind. The only possibility left is that hnRNP A1 protein dissociates from RNA very rapidly and does not interfere with binding of an oligonucleotide. To investigate the time course of dissociation the method of direct UV-crosslinking was employed. The protein was pre-incubated with [^{32}P]-labelled RNA under splicing conditions to reach an equilibrium, and dissociation of protein from labelled RNA was induced by addition of excess of the same but unlabelled RNA. The reactions were incubated for the periods of time indicated in the Figures, then UV-irradiated for 30 seconds and analysed by 12% SDS-PAGE as described in the Materials and Methods. The quality of the competitor was checked by mixing it with the labelled RNA prior to the incubation with the protein and irradiation (data not shown).

Figure 2.7 shows the time course of dissociation of hnRNP A1 and SF2/ASF measured in time intervals between 30 seconds and 60 minutes. The panel A shows the autoradiographical images of the gels and the panel B presents the plotted data obtained by quantification of images on a PhosphoImager. The result of this experiment demonstrates that hnRNP A1 dissociates from RNA much faster than SF2/ASF. Unfortunately, the difference in dissociation rates for these proteins could not be calculated in this experiment. Surprisingly,

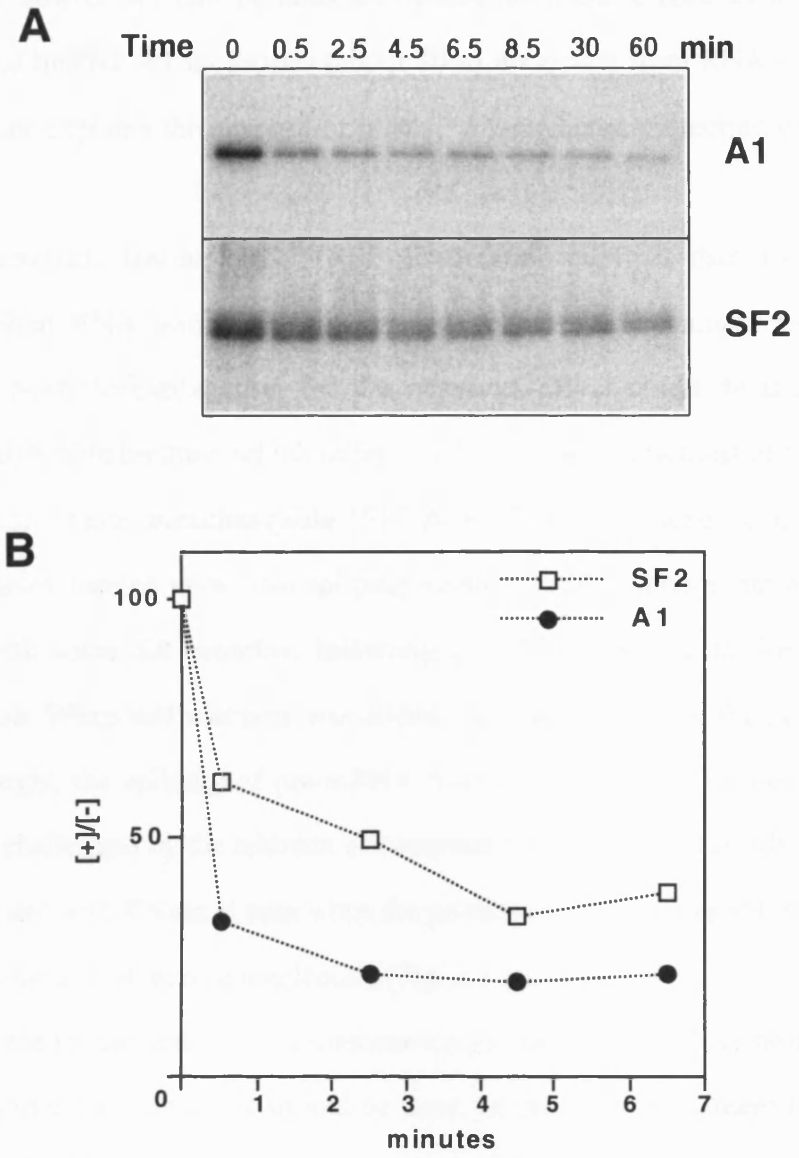


Figure 2.7. Time course of hnRNP A1 and SF2/ASF dissociation from pre-mRNA. The reaction mixtures, containing 5 pmol of hnRNP A1 or SF2/ASF and 5000 cpm of labelled RNA (approximately 1-2 fmol) in 10 µl of buffer D-Glu, were incubated for 15 min at 30°C. The dissociation was triggered by the addition of 1 µl of unlabelled RNA (2.4 mg/ml). The samples were UV-irradiated for 30 seconds after the indicated periods of time, digested with RNases and subjected to 12% SDS-PAGE followed by autoradiography. **A:** Autoradiographic images of the gels for hnRNP A1 and SF2/ASF. **B:** Quantification of the above gels using a PhosphoImager. The ratio between the signals in the presence [+] and absence [-] (time "0") of competitor RNA was plotted versus the time of dissociation.

in the first time point taken in 30 seconds after addition of competitor only the residual amount of bound hnRNP A1 can be detected (dissociation curve reaches the plateau). It means that most of hnRNP A1 molecules (about 80%) dissociate from RNA in less than 30 seconds. This result explains the absence of hnRNP A1-mediated protection of the RNA in RNase H assay.

The characteristic feature of SF2/ASF dissociation curve is that, for 30% of the protein, the labelled RNA bound to the protein can not be exchanged for unlabelled competitor. The possible explanation for the observed effect could be that these 30% represent extremely stable binding, which reflect the functional interactions of SF2/ASF with high affinity sites. These unexchangable [SF2/ASF-RNA] complexes could mimic the commitment complex formed in *in vitro* splicing reactions. It was shown that pre-incubation of pre-mRNA with some SR proteins, including SF2/ASF, resulted in the commitment complex formation. When this complex was added into nuclear extract, the pre-mRNA was spliced. Interestingly, the splicing of pre-mRNA from this artificially formed commitment complex was not challenged by the addition of competitor RNA (Fu, 1993). Also the obtained result is in agreement with RNase H data when the protection of RNA was still observed in 10 minutes after the addition of an oligonucleotide (Fig. 2.2 and 2.3).

However, the virtual absence of a dissociation profile for hnRNP A1 protein (Fig. 2.7) suggested that shorter time intervals should be used. In the next experiment (Fig. 2.8), the samples were incubated for 0, 5, 10, 15, 20, 40 and 120 seconds after an addition of the competitor and then UV-irradiated for 30 seconds. In order to measure the dissociation rate more precisely, an additional set of control reactions (without competitor) was introduced (Fig. 2.8A, "-"). The latter was used to control efficiency of crosslinking and normalise signals obtained in the presence of competitor (Fig. 2.8A, "+"). The result, plotted in Figure 2.8B, demonstrates that the protein dissociates completely within 5 seconds plus the irradiation time of 30 seconds.

The only feasible way to obtain a dissociation curve was to reduce the irradiation time. The irradiation time of 30 seconds in previous experiments was the minimum time required to reach saturation in crosslinking but avoid UV-induced damage of RNA and proteins. The

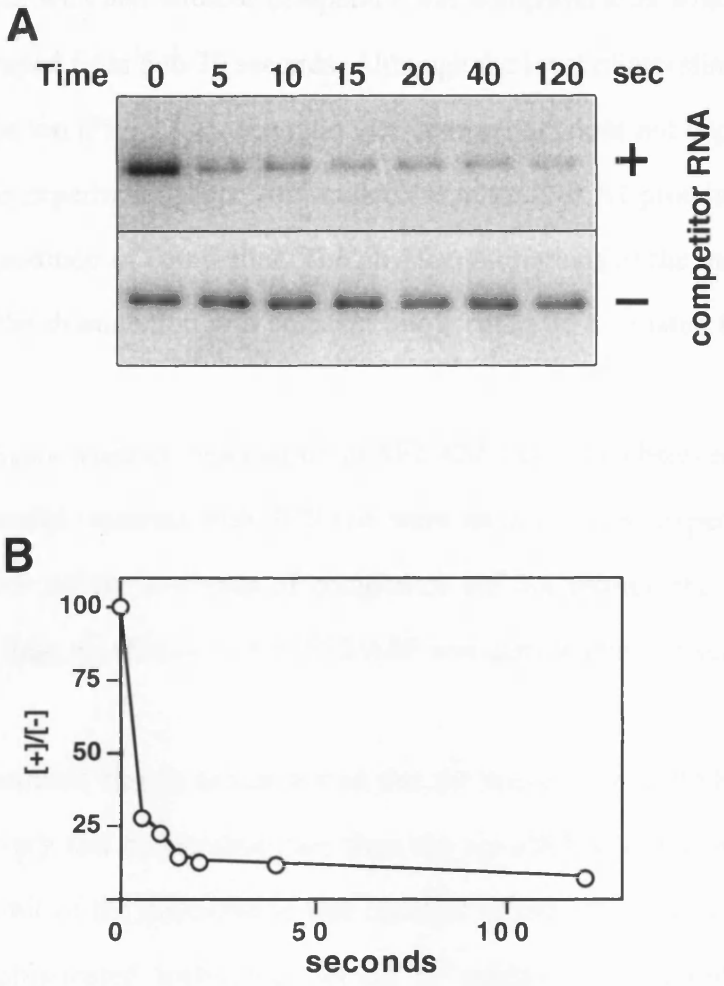


Figure 2.8. High rate of hnRNP A1 dissociation from pre-mRNA. The experiment was carried out essentially as in Fig. 2.7, but the time of dissociation (time between addition of competitor RNA and UV-irradiation) was shortened as indicated in the figure. **A:** Images of the gels, where "+" is the experiment carried out in the presence of competitor RNA, and "-" is the parallel experiment without competitor. The latter serves a control for efficiency of crosslinking asince corresponding samples (+ and -) were irradiated at the same time. **B:** Quantification of the above gels using a PhosphoImager. The data are plotted as in Fig. 2.7B.

minimum time required to mix components and start the irradiation was estimated as 5 seconds, and it is the incubation time in the following experiment (Fig. 2.9). The amount of crosslinked protein with and without competitor was compared after irradiation for different periods of time varied from 5 to 30 seconds. Although the level of crosslinking increases with duration of irradiation (Fig. 2.9A), the ratio (-/+ competitor) does not depend on the time of irradiation. In this experiment about 70% molecules of hnRNP A1 protein dissociated within 10 seconds after addition of competitor. The physical limitations of the method did not allow us to determine the dissociation rate constant but it could be estimated to be higher than 5 seconds.

To investigate whether dissociation of SF2/ASF could be observed within such short time intervals, parallel reactions with SF2/ASF were set in the same experiment (Fig. 2.9). In contrast to hnRNP A1 the presence of competitor did not change the relative amount of crosslinking and thus, no dissociation of SF2/ASF was detected during such a short period of time.

This experiment clearly demonstrated that the majority of hnRNP A1, in contrast to SF2/ASF, has a very fast dissociation rate from the pre-mRNA. It was possible to estimate only the lower limit of the dissociation rate constant of hnRNP A1 and to determine exact values more sophisticated techniques would be needed (for example, a fluorescence spectrometer equipped with stop-flow device).

In this work, results have been obtained for the dissociation rate constant for hnRNP A1 and, earlier (2.2.2 and 2.2.3), the equilibrium constant. The apparent dissociation equilibrium constants determined by both the NCFBA and UV-crosslinking are high (around 10^{-9} M), and as discussed above, these constants reflect only the total binding, including effects of co-operativity. The equilibrium constants of hnRNP A1 to the RNA-oligonucleotides selected by the SELEX procedure are the same order of magnitude, but the binding was considered to be non-cooperative (Burd and Dreyfuss, 1994). Taking into account the high values of the dissociation equilibrium constant (K_{eq}) and the dissociation rate constant (k_-), and their interdependence

$$K_{eq}=k_+/k_-$$

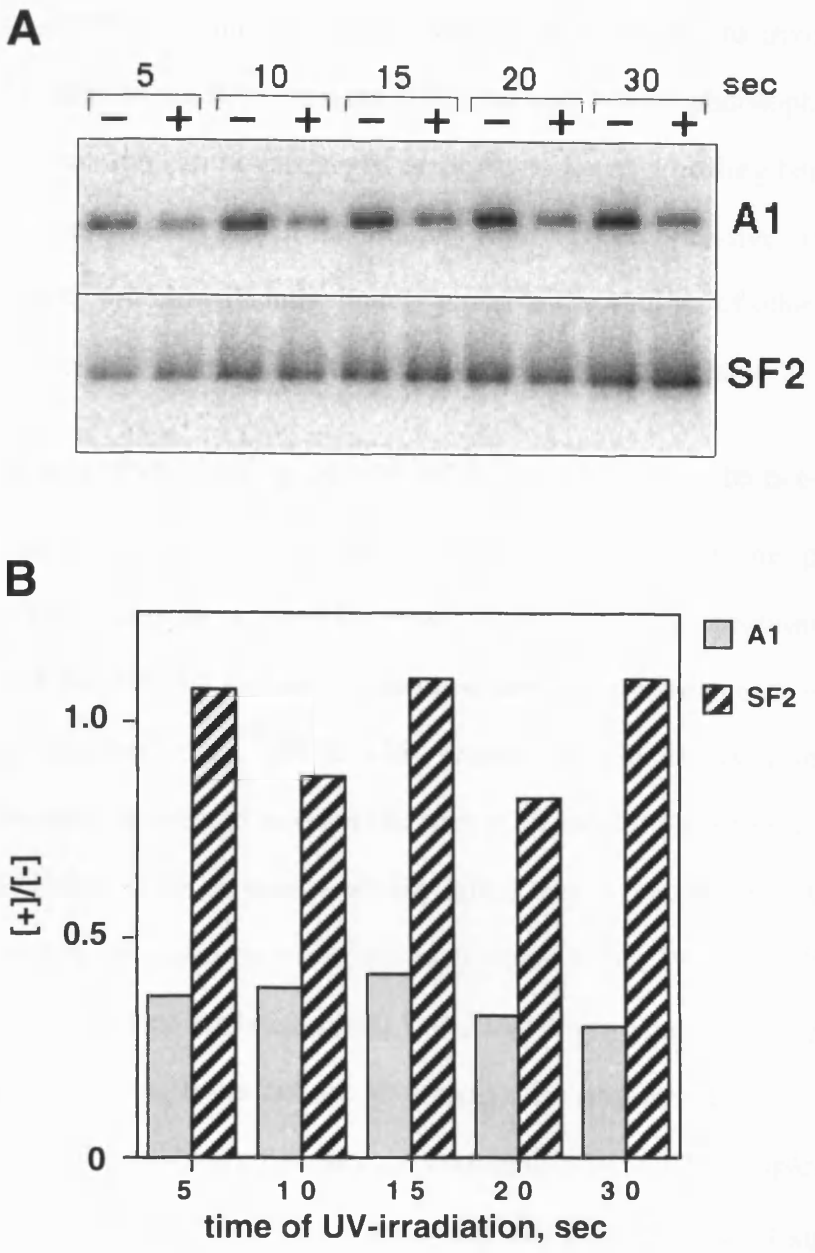


Figure 2.9. Comparison of hnRNP A1 and SF2/ASF dissociation at 5 seconds. The incubation mixtures were done as in Fig. 2.7, where "+" and "-" are the reactions with and without competitor RNA, respectively. At 5 seconds after addition of competitor RNA, the samples were UV-irradiated for different periods of time as indicated in the figure. **A:** Images of the gels obtained for hnRNP A1 and SF2/ASF. **B:** Quantification of the above gels using a PhosphoImager. The diagram shows the ratio between the amounts of radioactivity crosslinked in the presence and absence of competitor. The grey bar corresponds to hnRNP A1, and the striped bar corresponds to SF2/ASF.

the rate of binding (k+) should be extremely high. Such kinetic parameters are likely to reflect the mode of hnRNP A1 binding to pre-mRNA, which appears to involve binding by all molecules of protein to the RNA or none (a "Black and White" philosophy). In this case, the hnRNP A1 dissociation can be envisaged metaphorically as a peeling banana. This mode of interaction may be achieved only if the binding is strongly co-operative. Thereby, the binding of one molecule of protein will immediately promote the binding of others, and, conversely, dissociation of the first molecule will initiate the departure of the rest.

2.2.3.3. Counteractive binding of hnRNP A1 and SF2/ASF to pre-mRNA

The early studies of RNA processing indicated that the pre-mRNA is co-transcriptionally coated by a specific class of RNA-binding proteins to form hnRNP complexes, and hnRNP A1 is one of the most abundant constituents of such complexes (reviewed by McAfee *et al.*, 1997). This protein is also detected in the spliceosomal complexes although, at reduced amount (Bennet *et al*, 1992b; Neubauer *et al.*, 1998). On the other hand, SF2/ASF is an essential splicing factor that is required for the assembly of the earliest (E) specific pre-spliceosome complex (Krainer *et al.*, 1990; Fu, 1993). The detailed characterisation of these complexes using UV-crosslinking showed that SR proteins do not cross-link in hnRNP complexes but are abundant in E complexes (Staknis and Reed, 1994). Conversely, hnRNP crosslinking is largely excluded from the E complex. Based on these observations the following hypothesis is proposed. The pre-mRNA is completely covered by hnRNP A1 and other hnRNP proteins to prevent the formation of secondary structure. SF2/ASF recognises the specific sites on the RNA and binds there. Because the dissociation rate of SF2/ASF is much slower than hnRNP A1, the hnRNP A1 molecule can not bind to the same site again. The dissociation of the hnRNP A1 molecule from this site disturbs the co-operativity and forces surrounding molecules of hnRNP A1 to leave. By this action SF2/ASF frees long stretches of RNA from hnRNP proteins and makes them vacant for the interactions with snRNPs and splicing factors. This might be the first step in the commitment of pre-mRNA to the splicing pathway. The attractive feature of this hypothesis is that it proposes a physiological role for hnRNP A1 binding as a passive or temporary RNA cover and at the same time describes the mechanism of the counteracting effect of SF2/ASF on hnRNP A1

binding as an active process of RNA clearance from bulky non-specific binding which serves as a preparation step to the forthcoming events of splicing.

The binding properties of hnRNP A1 and SF2/ASF documented above do not contradict this hypothesis. The question was whether such an event can be simulated *in vitro* using recombinant proteins and *in vitro* transcribed pre-mRNA.

In the next experiment the pre-mRNA was initially pre-incubated with hnRNP A1 at a concentration 5 x 10^{-7} M, which already corresponds to saturation binding. Then a two-fold excess of SF2/ASF was added into reaction and the time course was monitored by UV-crosslinking to find out when the state of equilibrium in binding of both proteins to the pre-mRNA would be reached (Fig. 2.10A). The result is that SF2/ASF completely displaces hnRNP A1 within less than 5 seconds plus irradiation time (15 seconds). That is also in agreement with the fast dissociation rate observed in previous experiments and supports the "all or nothing" mode of hnRNP A1 binding discussed above. The complete absence of hnRNP A1 binding in this experiment can be explained by the existence of a concentration dependent barrier for an exclusive binding of one protein to the pre-mRNA.

Next, the simultaneous equilibrium binding of hnRNP A1 and SF2/ASF was monitored with anticipation to find the expected sharp, concentration dependent, switch from [A1-RNA] to [SF2-RNA] complex. Beyond expectation both proteins were detected bound to the RNA in the close range of concentrations. The relative amounts of crosslinked proteins in each lane were quantified and plotted as a function of added SF2/ASF (Fig. 2.10B, the lower panel). This curve clearly demonstrates antagonism in binding of hnRNP A1 and SF2/ASF to pre-mRNA. It should be noted that when just 6 pmol of SF2/ASF was added into the reaction the signal corresponding to hnRNP A1 was not detectable.

In fact, when both proteins are crosslinked to RNA, the situation can be interpreted in two different ways. First, if hnRNP A1 and SF2/ASF have distinct (not overlapping) binding preferences, molecules of both proteins can be simultaneously accommodated on the one pre-mRNA molecule. The second is the mutually exclusive binding, when hnRNP A1 molecules are bound to one pool of RNA molecules and SF2/ASF to another. The latter case would

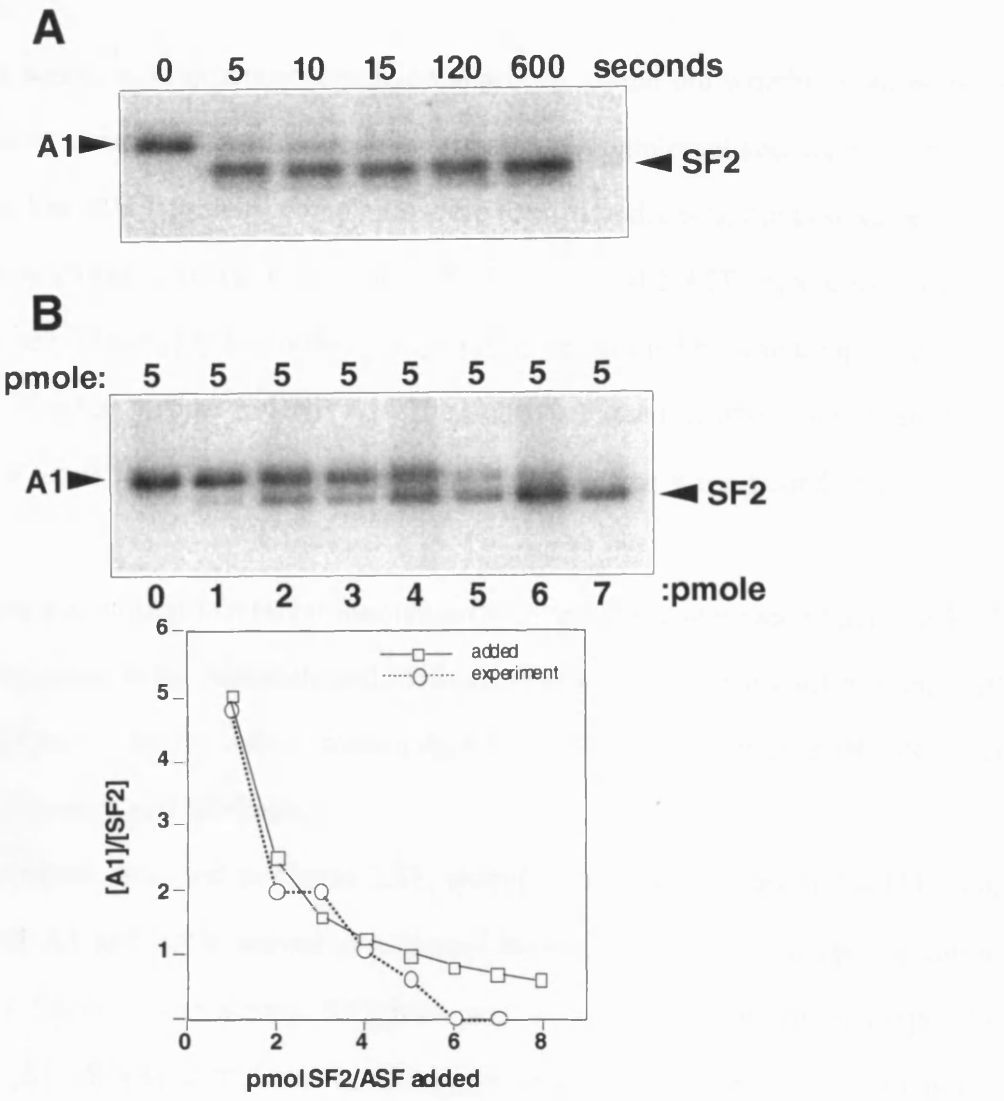


Figure 2.10. SF2/ASF chases off hnRNP A1 from pre-mRNA. A: Time course of the hnRNP A1 dissociation triggered by SF2/ASF. 5 pmol of hnRNP A1 were pre-incubated with 5000 cpm of labelled RNA in 10 µl of buffer D-Glu, and then 10 pmol of SF2/ASF were added. After incubation for the indicated time intervals, the samples were UV-irradiated for 15 seconds, digested with RNases, and subjected to 12% SDS-PAGE followed by autoradiography. **B:** (*Top*) Equilibrium binding of hnRNP A1 and SF2/ASF to pre-mRNA. 5 pmole of hnRNP A1 were mixed with 5000 cpm of labelled RNA in 10 µl of buffer D-Glu, and then 1-7 pmole of SF2/ASF was added to the reaction mixture as indicated in the figure. After incubation for 15 min at 30°C, the samples were processed as described above. (*Bottom*) Quantification of the gel depicted in (B). The ratio between signals corresponding to hnRNP A1 and SF2/ASF is plotted as a dashed line (·····o·····). The solid line (—□—) represents the calculated ratio between the actual concentrations of hnRNP A1 and SF2/ASF in the reaction mixture.

strongly support the hypothesis discussed above. Therefore, this question was addressed in the next experiment.

The idea was to use immunoprecipitation approach to find out whether both proteins can be accommodated on one RNA molecule, or are mutually intolerant and occupy different RNA molecules. The [RNA-protein] complexes were formed under conditions selected in the previous experiment (Fig. 2.10B), when both hnRNP A1 and SF2/ASF were crosslinked to RNA simultaneously. Then [RNA-protein] complexes were selected by immunoprecipitation with antibodies directed against hnRNP A1. The ability of α-A1 antibodies to precipitate SF2/ASF protein would be a strong argument that both proteins were bound to one RNA molecule.

Antibodies were raised in a rabbit immunised with the full length recombinant hnRNP A1 protein as described in the Materials and Methods. The antibodies were affinity purified on protein-A Sepharose and covalently cross-linked to the Affi-Gel Hz matrix (Bio-Rad) as described in the Materials and Methods.

The experiment, depicted in Figure 2.11, analysed three sets of reactions. (1) First, containing hnRNP A1 and RNA, served as a control for efficiency of immunoprecipitation (lanes 1-3). (2) Second, containing SF2/ASF and RNA, evaluated the non-specific precipitation of [SF2-RNA] complexes on the α-A1 beads (lanes 4-6). (3) Third is the experiment itself, when both proteins are present in the incubation mixture and interact with RNA (lanes 7-9). After the equilibrium binding was achieved, reaction mixtures were UV-irradiated and aliquots were subjected to immunoprecipitation. The samples were analysed by 12% SDS-PAGE. The gel was silver stained (the lower panel, Fig. 2.11) and then exposed to PhosphorImager (the upper panel, Fig. 2.11). The first lane in each set (lanes 1, 4, 7) shows the input material into immunoprecipitation. The second lane in the set (lanes 2, 5, 8) shows the material immunoprecipitated from the mixture treated with RNases. This should serve as a control for RNA-dependent immunoprecipitation of the proteins. The third lane in each set (lanes 3, 6, 9) represents immunoprecipitation of RNA-protein complexes without treatment with RNases. If molecules of both hnRNP A1 and SF2/ASF are bound to one RNA molecule, two bands corresponding to the crosslinked hnRNP A1 and SF2/ASF should be seen in lane 9.

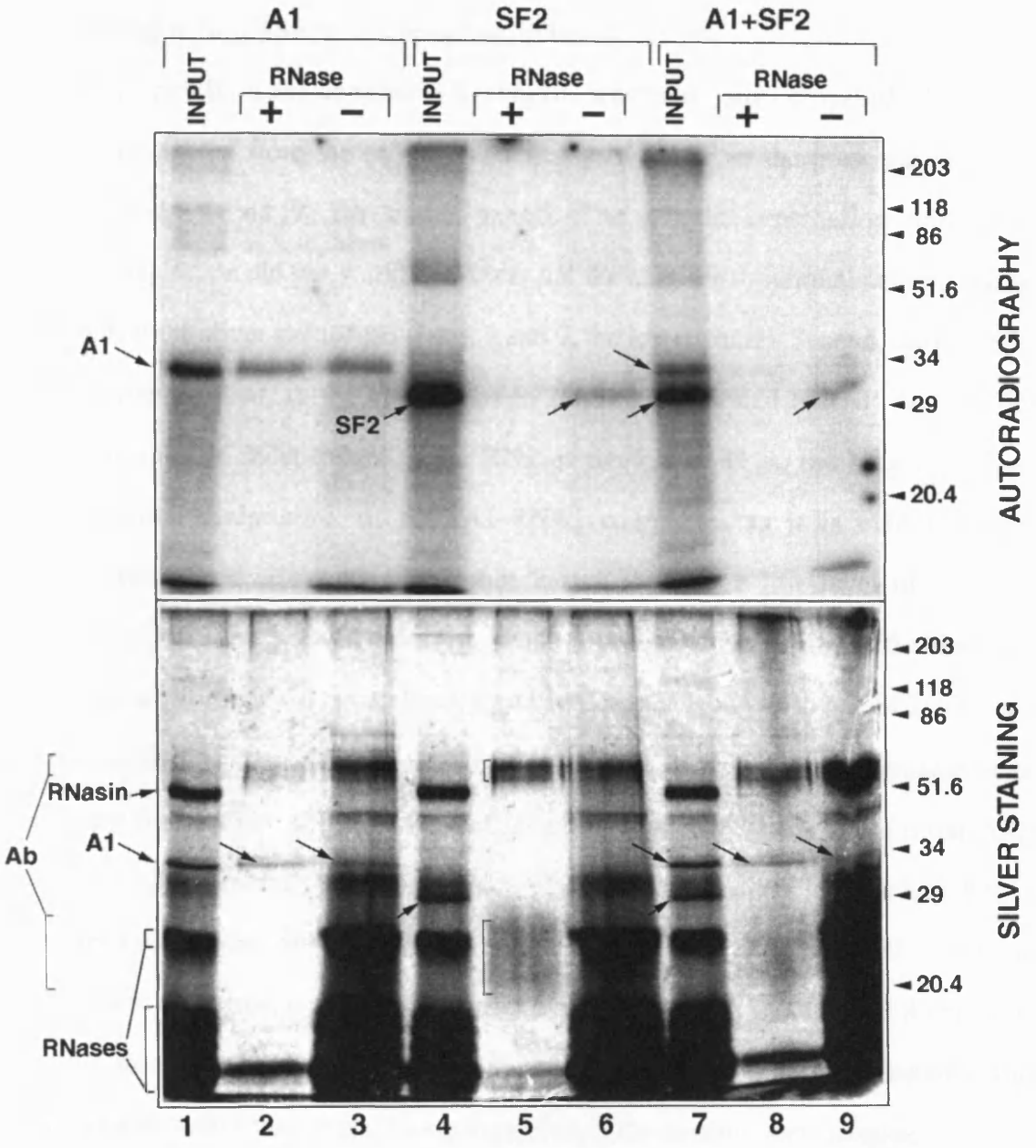


Figure 2.11. Immunoprecipitation of RNA-protein complexes with α-A1 antibodies. The reaction mixtures contained 5000 cpm of labelled RNA (approximately 1-2 fmol) and either 5 pmol of hnRNP A1 (lanes 1-3), 5 pmol of SF2/ASF (lanes 4-6), or 5 pmol of each of the proteins (lanes 7-9) in 10 µl of buffer D-Glu. The reactions were incubated for 15 min at 30°C and UV-irradiated for 15 seconds. Then, the samples were subjected to immunoprecipitation before (lanes 3, 6, 9) or after (lanes 2, 5, 8) RNase treatment (indicated as "-" and "+", respectively). The input material used for immunoprecipitation is shown in lanes 1, 4, and 7. (*Top*) Autoradiographic image of the gel. (*Bottom*) The silver stained gel.

Moreover, if co-immunoprecipitation of these proteins is RNA-dependent, the band corresponding to SF2/ASF should disappear in lane 8.

The result was unexpected. There was not any crosslinked hnRNP A1 immunoprecipitated from the mixture containing SF2/ASF in the presence or absence of RNases (lanes 8 and 9, the upper panel). The simple explanation would be that immunoprecipitation did not work, but it was not the case. First, immunoprecipitated hnRNP A1 is seen on the silver stained gel (lanes 8 and 9, the lower panel). Second, the control (lanes 1-3) demonstrates that, in the absence of SF2/ASF, crosslinked hnRNP A1 is efficiently immunoprecipitated independent of the RNases treatment. Thus, the long RNA does not hinder immunoprecipitation of the [A1-RNA] complexes as it is evident from both autoradiography and silver staining (lanes 2 and 3). Only a low level of the SF2/ASF precipitation (background) was observed, which is probably due to unspecific binding of the RNA to the α-A1 beads (compare lanes 6 and 9). The result shows that if SF2/ASF is added into the incubation mixture the crosslinked hnRNP A1 can not be immunoprecipitated whereas the free hnRNP A1 is precipitated. The only reasonable explanation is that SF2/ASF and hnRNP A1 are bound to one RNA molecule. It was unexpected that the [A1-RNA-SF2] complexes escape from immunoprecipitation. It might be that binding of SF2/ASF induces such dramatic conformational changes in [A1-RNA] complexes that hnRNP A1 is no longer accessible to α-A1 antibodies. These conformational changes or rearrangements could be similar to those, which take place during formation of the commitment complex.

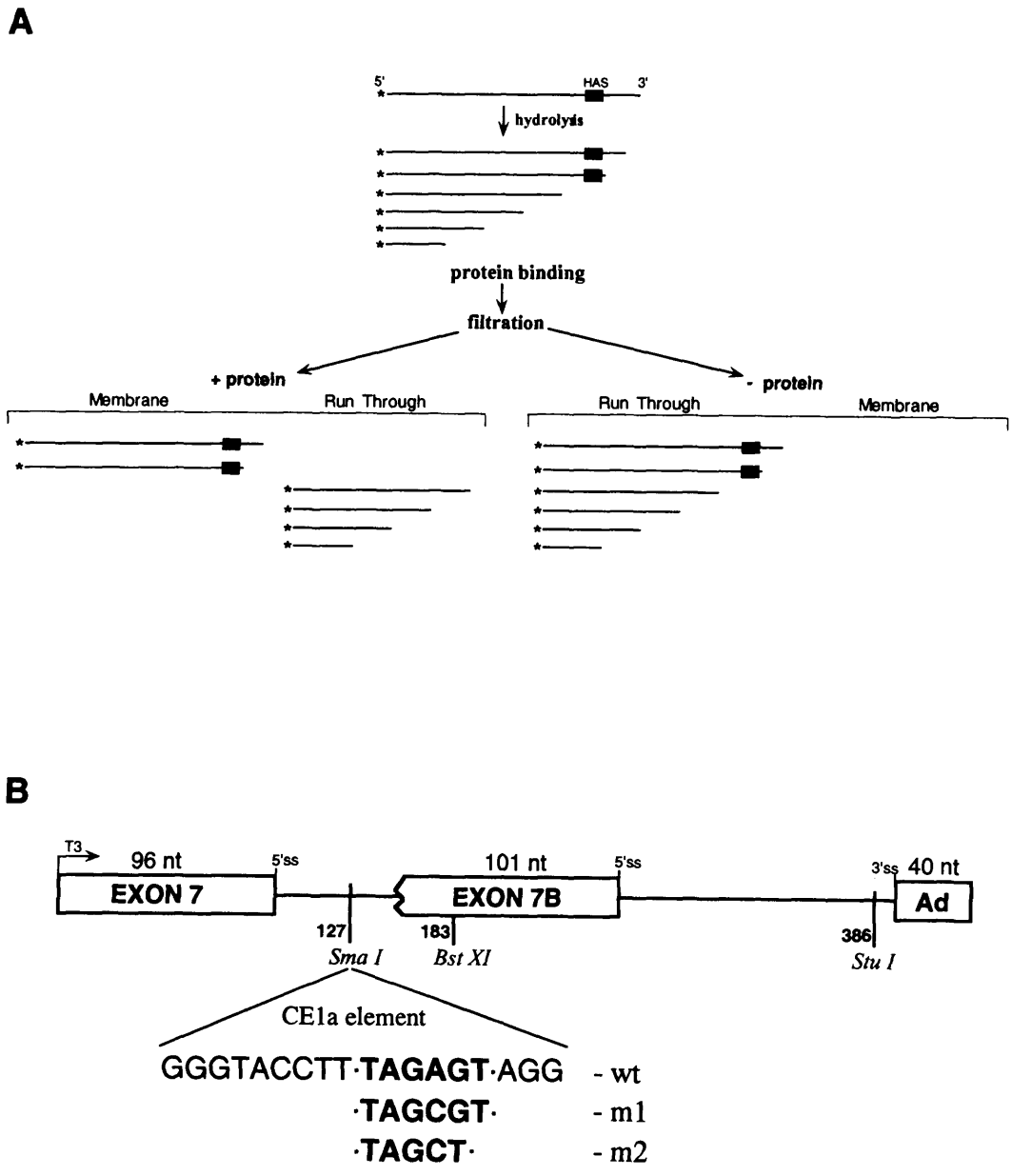
In conclusion, we could not mimic in this experiment the event proposed in the hypothesis. The result of immunoprecipitation does not support the "all or nothing" mode of counteraction of hnRNP A1 and SF2/ASF. We think, the reason for our failure to simulate the "all or nothing" effect is the length of RNA. It is possible that under the conditions of this experiment molecules of SF2/ASF bound to RNA does causes the displacement of hnRNP A1 from the RNA, but the cleared gap is smaller than 533 nt in length (C175G RNA). If the idea of mimicking the clearance of RNA from hnRNPs by SF2/ASF is further considered the shorter molecules of RNA should be tried. But this matter was not addressed at that time.

2.2.4. Study of the putative high affinity site for hnRNP A1

Since the results demonstrated that both hnRNP A1 and SF2/ASF could be accommodated on the same molecule of RNA and suggested the different binding preferences, the presence of high affinity sites for hnRNP A1 on the pre-mRNA was investigated. For this purpose an approach had been developed that was based on the combination of partial alkali hydrolysis of RNA and the nitrocellulose filter binding (Fig. 2.12A). *In vitro* transcribed RNA was labelled at 5' end with [γ-^{32}P] ATP and subjected to partial alkali hydrolysis in order to generate a statistical population of different RNA fragments. The recombinant hnRNP A1 protein was incubated with the pool of RNA molecules and the incubation mixture was filtered through the nitrocellulose. The run through material (RT fraction), unbound RNA, was analysed by urea-PAGE and labelled RNA fragments were visualised by autoradiography. The idea was as follows. If there is a high affinity site N nucleotides from the 5' end, at a certain concentration of the protein all molecules of RNA containing this site would be bound and therefore retained on the filter. Inasmuch as the 5' end of RNA is labelled, all molecules longer than N nt should not be detected in RT fraction. From these considerations, the cut off of RT indicates the position (N nt) of a high affinity site closest to the 5' end of RNA.

At that period of time, the laboratory of Dr. Benoit Chabot (University of Sherbrooke, Canada) had preliminary data indicating the existence of the functionally significant hnRNP A1 binding sites. The question was whether this site can be considered as a high affinity site in comparison to other sites on the pre-mRNA. If this is the case, it would be worth investigating the presence of high affinity sites on the pre-mRNA, on which the counteracting effect of hnRNP A1 and SF2/ASF can be modulated.

Chabot and co-workers identified a 17 nt sequence (CE1a) within a conserved intron element that is responsible for exclusion of the alternative exon 7B from the hnRNP A1 pre-mRNA (Chabot *et al.*, 1997). The preliminary data suggested that hnRNP A1 binds to the CE1a element and promotes exon skipping. CE1a contains the sequence UAGAGU, which resembles a SELEX winner for hnRNP A1, UAGGGU (Burd and Dreyfuss, 1994). We accepted that the CE1a element might serve an example of the hnRNP A1 high affinity



Fugure 2.12. An overview of the methodological approach used to study the putative high affinity site for hnRNP A1 binding. A: Strategy of the RNA fragmentation/selection method. Only RNA species labelled at the 5'-end are shown. The high affinity site (HAS) is boxed in grey. **B:** Schematic drawing of templates used to generate RNA substrates. Restriction enzymes and the length of the fragments are indicated. The putative high affinity site (CE1a element) inserted into the Sma I site is shown in letters, and the sequence resembling the SELEX winner is in bold.

binding site and therefore the RNA substrates generated from the mini-gene constructs, kindly provided by Dr. B. Chabot, were used in following experiments.

The RNA was generated by *in vitro* transcription from the model DNA templates, schematically depicted on Figure 2.12B. The constructs contain the 5' splice sites of exon 7 and 7B in competition for the 3' splice site of the adenovirus L2 exon. The basic model pre-mRNA lacks the CE1a element and is spliced almost exclusively to the proximal 5' splice site in *in vitro* splicing assays. Insertion of CE1a into the Sma I site of the intronic element between the two competing 5' splice sites stimulated splicing to the distal 5' splice site. The following nomenclature was used: (1) wt - RNA containing a 17 nt CE1a element with sequence UAGAGU, resembling the SELEX winner. (2) m1 - contains A to C substitution in putative high affinity site (UAGCGU). (3) m2 - AG to C substitution (UAGCU). (4) Δ - a basic RNA which does not contain CE1a element (no insertion into Sma I site). Different restriction enzymes were used to generate templates of appropriate lengths as depicted in the Figure 2.12.

In the first experiment, I investigated whether the method would produce the expected pattern on RNA containing a putative high affinity site. For these purposes, the wt-RNA of 400 nt long, which contains the putative high affinity site at the position of 143 nt, was hydrolysed, then allowed to interact with hnRNP A1 at two different concentrations (60 and 600 nM), and filtered. The run-through fractions were analysed by 8% urea-PAGE (Fig. 2.13). When 15 pmol of hnRNP A1 were added all RNA molecules longer than 120 nt were completely removed from the run-through fraction. The lower concentration of hnRNP A1 was not sufficient to achieve this effect. This result demonstrated that, in principle, the method allows detection of this high affinity site under selected conditions.

In the next experiment the wt-RNA was compared to the mutant m2-RNA, Δ-RNA, and the D5'ss-RNA under conditions selected in the previous experiment (Fig. 2.14). The D5'ss-RNA is a model splicing substrate with a duplicated 5' splice-site (Reed and Maniatis, 1986; Krainer *et al.*, 1990b) that has been widely used to demonstrate the counteracting effects of hnRNP A1 and SF2/ASF on 5' splice site selection (Mayeda and Krainer, 1992; Caceres and Krainer, 1993; Caceres *et al.*, 1994; Mayeda *et al.*, 1998). 8 pmol of Δ-RNA, 3 pmol of

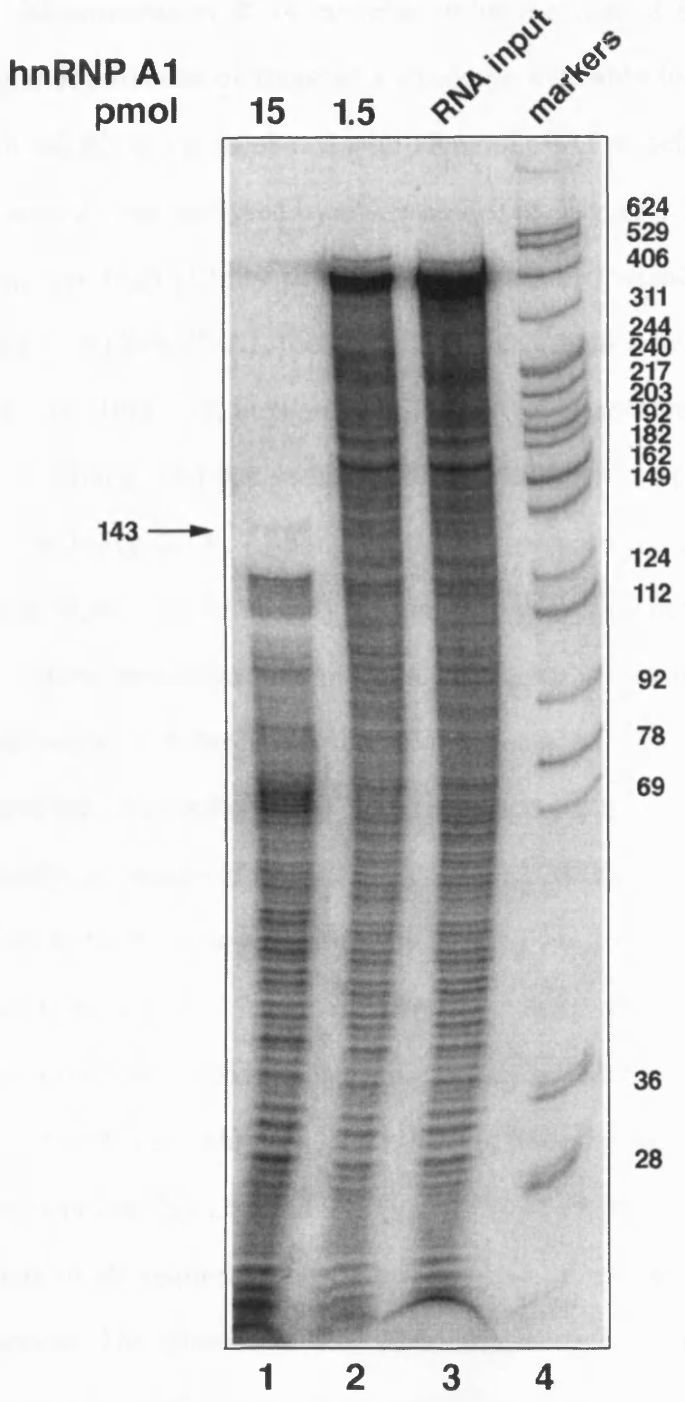


Figure 2.13. Visualisation of the high affinity site for hnRNP A1 binding.
3 pmol of the 5'-labelled wt-RNA containing the putative high affinity site at nucleotide position 143 (indicated by an arrow), was hydrolysed (lane 3) and incubated with 15 pmol (lane 1) or 1.5 pmol (lane 2) of hnRNP A1 in 25 µl of buffer D-Glu. The incubation mixture was filtered and the run through fraction was analysed by 8% urea-PAGE followed by autoradiography. The positions of DNA markers (lane 4) are indicated on the right.

wt-RNA, 3 pmol of m2-RNA, and 2 pmol of D5'SS RNA were subjected to hydrolysis. The differences in the amounts of RNA correlate with the size of substrates, and therefore, statistically equal populations of fragments would be available for interaction with hnRNP A1. The hydrolysed RNA was incubated with 15 pmol of A1 as before and then, filtered. The run through fractions were analysed by 6% urea-PAGE. Figure 2.14 shows that if the RNA contains the putative high affinity site UAGAGU (wt) or mutant sequence UAGCU (m2) then, in the presence of hnRNP A1, the high molecular weight RNA fragments were removed from the RT fraction. However, in contrast to the previous experiment, the clear cut off of all the fragments containing the putative high affinity site was not reproduced (compare lane 4, Fig. 2.14, and lane 1, Fig. 2.13). Although both experiments were carried out under the same conditions, the difference could be only attributed to the efficiency of RNA hydrolysis, and therefore, the relative concentrations of RNA fragments and protein were not identical in these two experiments. It is likely that the ratio between the RNA population and protein is crucial to observe the clear-cut effect as it was obtained in the experiment depicted in Figure 2.13. Unexpectedly, a change of AG for C in the putative high affinity site did not produce dramatic changes in the binding pattern (lanes 3 and 5, Fig.2.14). This indicates that SELEX winner sequence located in the 17nt CE1a element is not a determinant of high affinity.

The main effect observed in this experiment is that the RNA without high affinity sites (Δ-RNA) and D5'ss-RNA exhibit a completely different pattern from wt-RNA, as all fragments were removed from the RT fraction in the presence of hnRNP A1. It means that hnRNP A1 binds to all sequences on these RNAs and does not discriminate in favour of particular sequences. The question is why all the fragments are removed from RT fraction if the RNA does not contain the high affinity site, but if it is present then the small fragments are not retained on the filter. The effect is likely to be modulated by the relative concentrations of protein and RNA. In the case of non-specific binding the protein interacts with every fragment at this concentration. If there is a high affinity site, hnRNP A1 would bind to it and might induce strong co-operative binding of other protein molecules to the same RNA molecule. This situation will result in the decrease of the free concentration of protein relative to the situation when this site is absent, and, therefore, there would not be enough

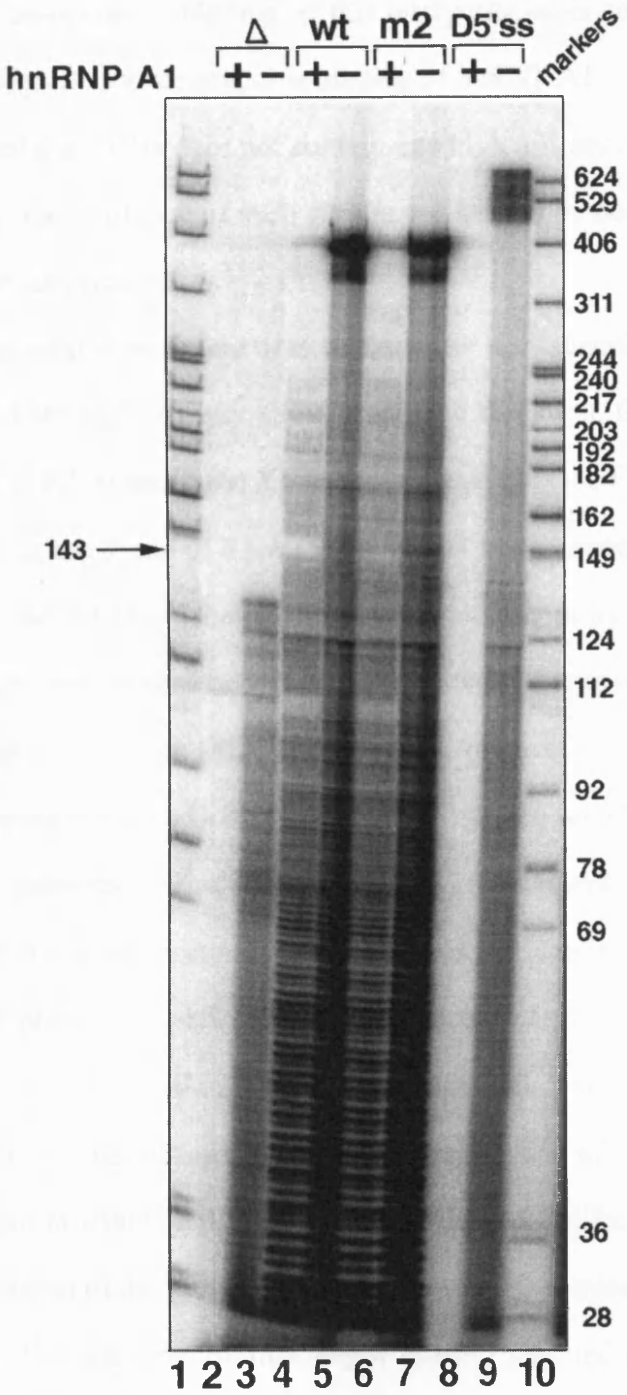


Figure 2.14. Comparison of hnRNP A1 binding to different RNA substrates. 8 pmol of Δ-RNA (lanes 2 and 3), 3 pmol of wt-RNA (lanes 4 and 5), 3 pmol of m2-RNA (lanes 6 and 7), and 2 pmol of D5'ss (lanes 8and 9) were incubated in the presence (+) or absence (-) of 15 pmol of hnRNP A1 in 25 µl of buffer D-Glu. The run through fractions were analysed by 6% urea-PAGE, and labelled RNAs were visualised by autoradiography. The positions of DNA markers (lanes 1 and 10) are indicated on the right.

protein to bind all the fragments. Thus, it is conceivable that high affinity binding sites modulate the effect of co-operative binding. If this interpretation is correct, the high affinity site serves as a nucleation site for co-operative binding of hnRNP A1.

It can be inferred that D5'ss does not contain any high affinity sites for the hnRNP A1 binding, and, therefore, the existence of such sites is unlikely to be necessary for modulating alternative 5' splice site selection.

The idea of the next experiment was to decrease non-specific binding in order to accentuate the effect of the high affinity site. To achieve this, first, the substrate RNA was shortened from 400 nt to 200 nt using Bst X1 restriction enzyme (see Figure 2.12) to bring the high affinity site closer to the 3' end of RNA (50 nt from 3' end); second, a 100-fold excess of unlabelled tRNA was added to the labelled RNA substrate before hydrolysis. The wt-RNA, m1-RNA, and Δ-RNA were examined in a parallel set of experiments to reveal the contribution of the high affinity site (Fig. 2.15, A, B, C, respectively). In these experiments both RT and membrane-retained fractions were analysed by 10% urea-PAGE. All three RNAs produced very similar patterns, in that most of the RNA was found in the RT fraction in a wide range of hnRNP A1 concentrations irrespective of the presence of high affinity site. Thus, in contrast to the previous experiments, in the presence of non-specific competitor there were no obvious differences detected in RT fractions of wt-, m1-, and Δ-RNA. The patterns of RNA recovered from the membrane were also similar for all three substrates. The comparative quantification analysis of corresponding lanes for different substrates did not reveal any specific retention of the particular RNA species with increasing amounts of hnRNP A1 (data not shown). The per cent of full-length species retained on the membrane was quantified for all three RNAs and plotted as a function of hnRNP A1 added in Figure 2.16. Although more wt-RNA molecules were retained on the membrane in comparison to the Δ-RNA or m1-RNA, the difference is not big enough to claim a significant contribution of a high affinity site in overall binding. Thus, the strong effect observed in Figure 2.13 was almost abolished in the presence of non-specific competitor, and therefore, the results presented here could not be considered as a support for existence of the high affinity site for hnRNP A1 on this RNA substrate.

Figure 2.15. Comparison of hnRNP A1 binding to different RNA substrates in the presence of non-specific competitor RNA. The incubation mixtures contained 2 pmol of the 5'-labelled RNA hydrolysed in the presence of 6 μg of total yeast tRNA and indicated amounts of hnRNP A1 in 100 μl of buffer D-Glu. Both the run through fractions (lane 9-14) and the material eluted from the filter (lane 2-7) were ethanol precipitated and analysed by 10% urea-PAGE followed by autoradiography.

A: Binding of hnRNP A1 to the wt-RNA.

B: Binding of hnRNP A1 to the m1-RNA.

C: Binding of hnRNP A1 to the Δ-RNA.

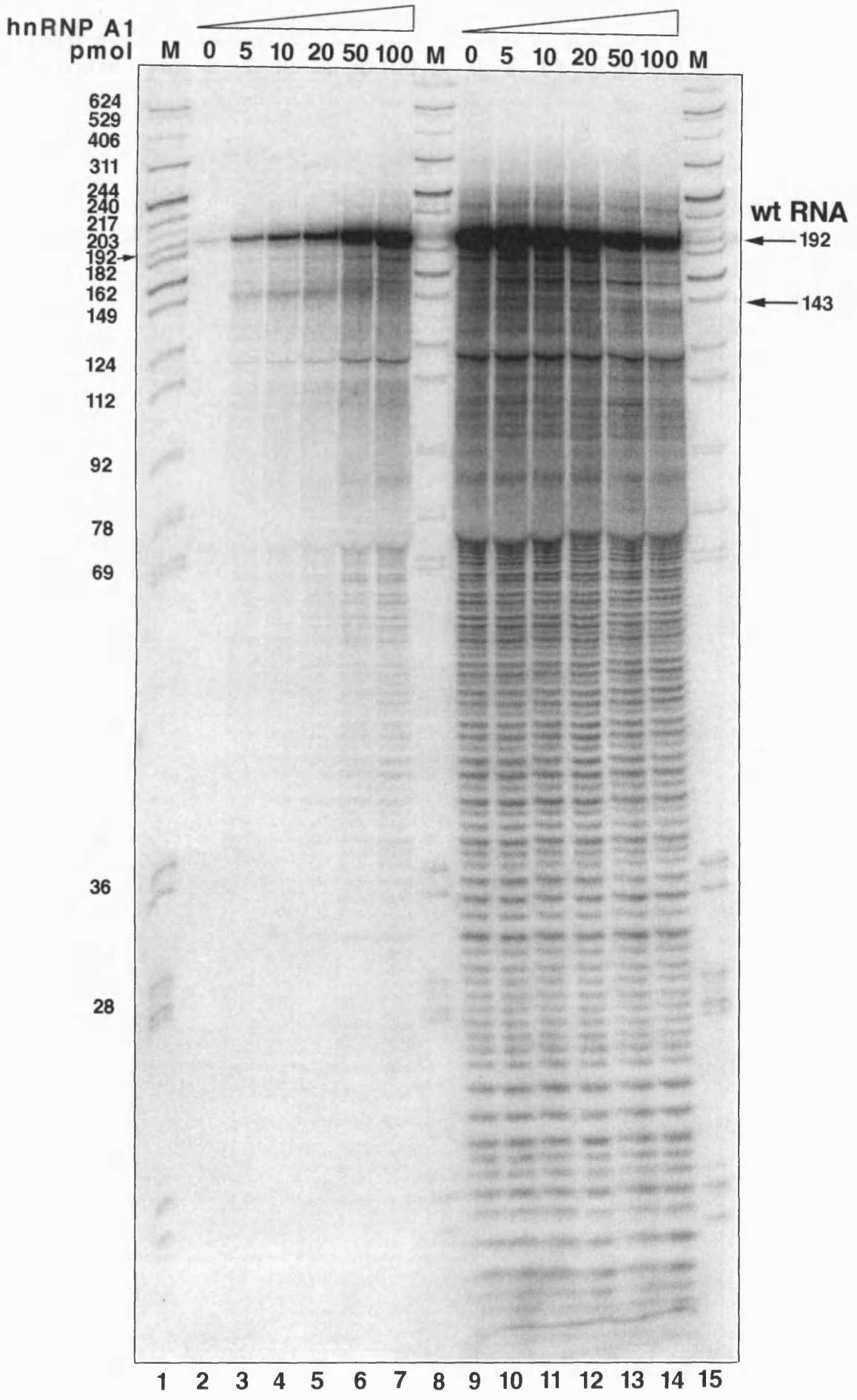


Figure 2.15A. Binding of hnRNP A1 to the wt-RNA.

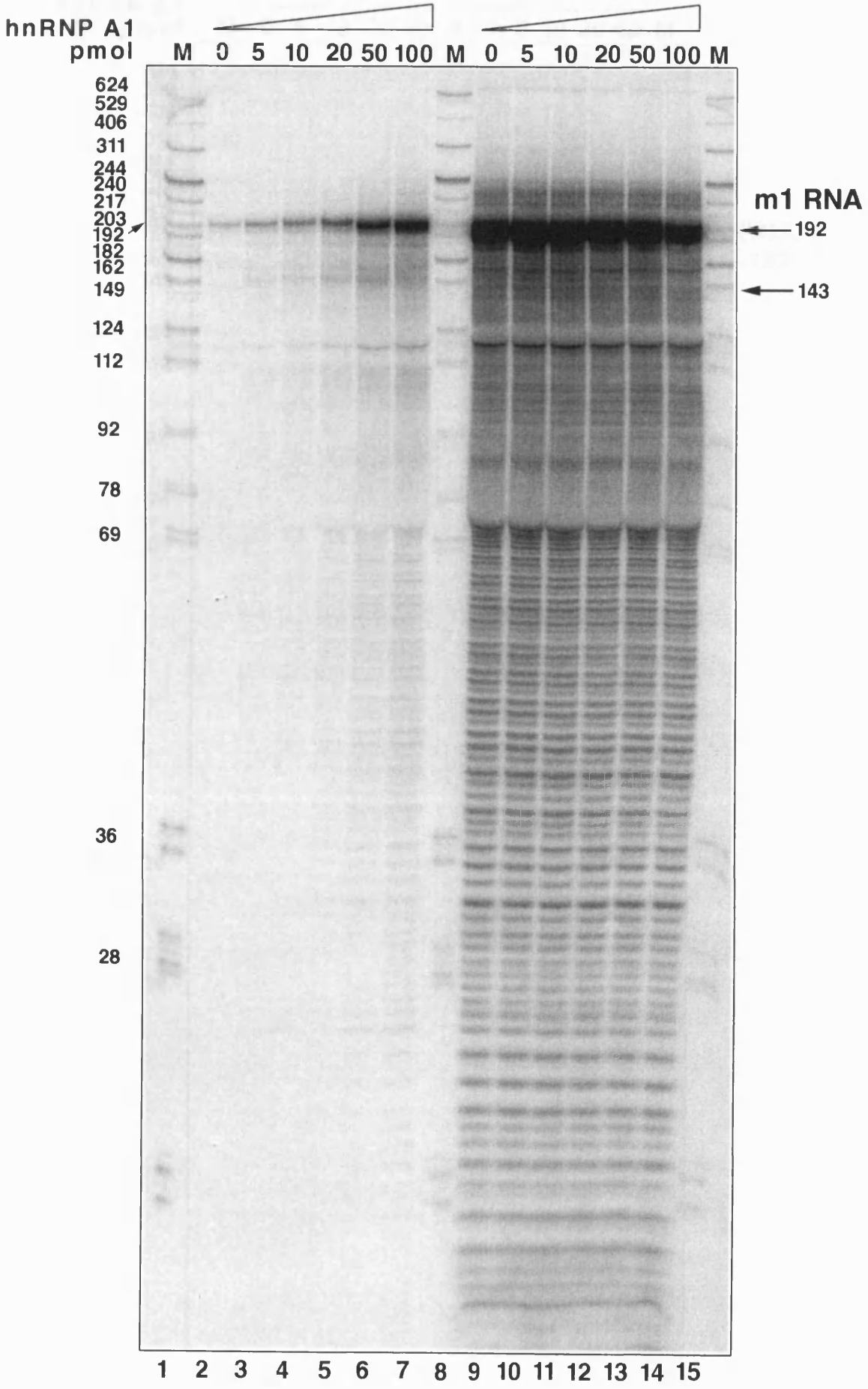


Figure 2.15B. Binding of hnRNP A1 to the m1-RNA.

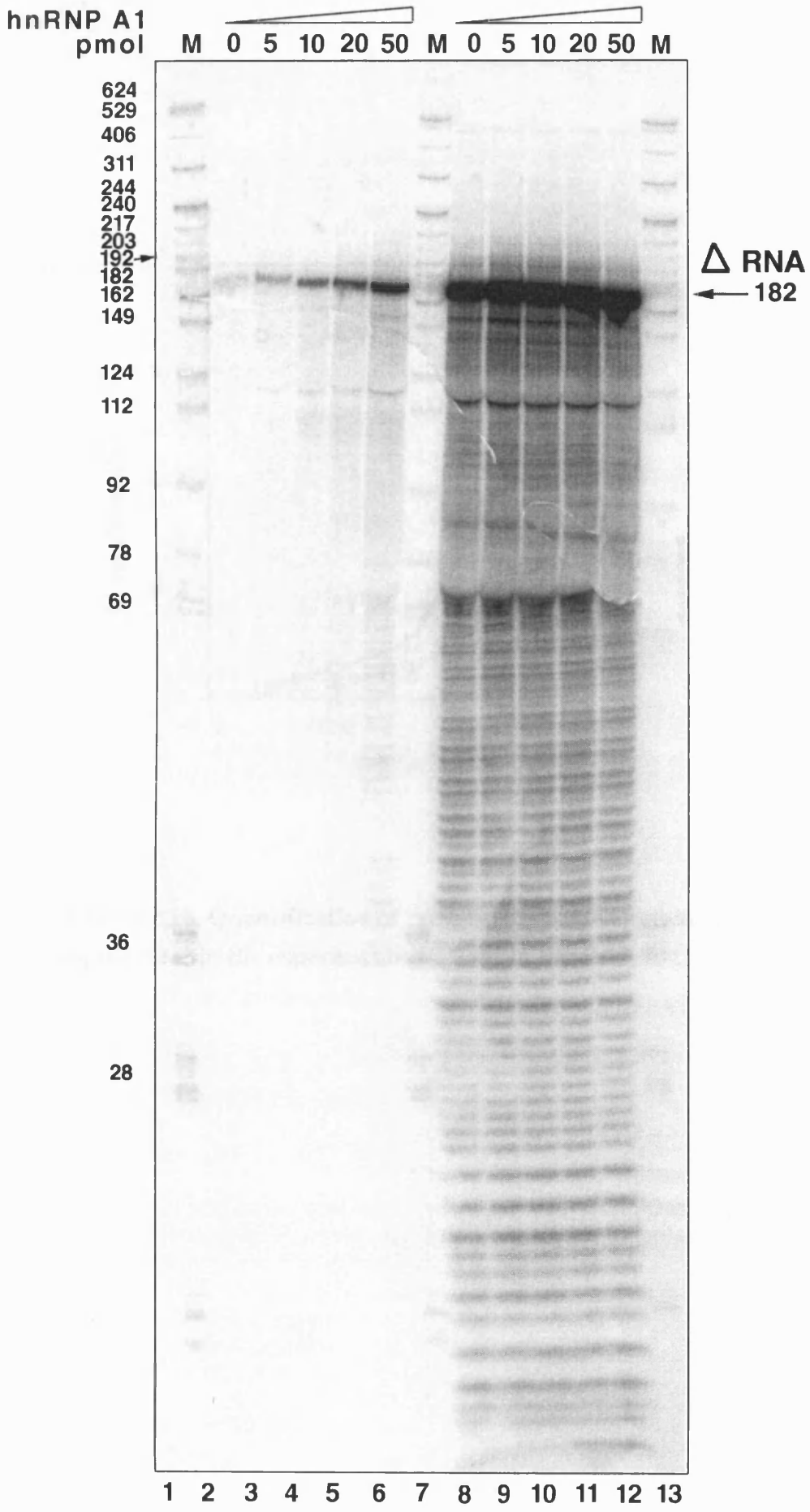
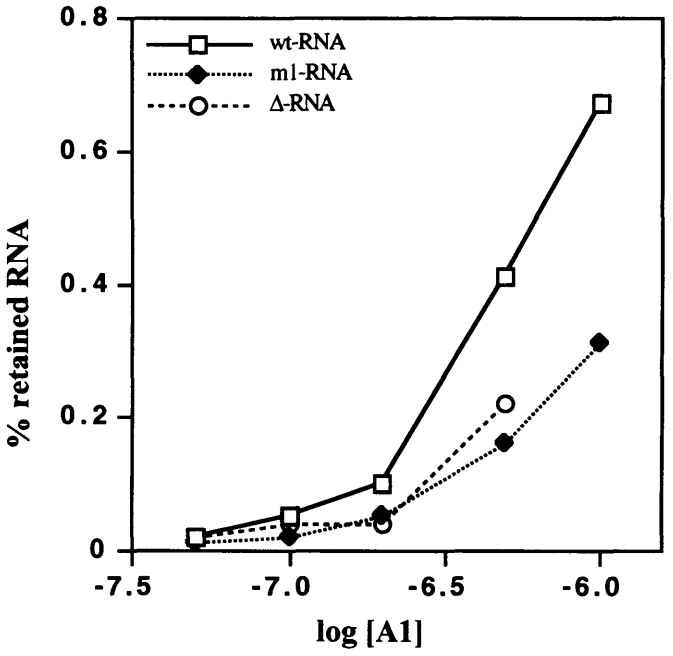


Figure 2.12C. Binding of hnRNP A1 to the Δ-RNA.



Fugure 2.16. Quantification of the full-length RNA retained on the filter in the experiments depicted in Fig. 2.15, A-C.

Despite some experiments that were carried out, the work has not been accomplished because of my departure from Leicester. At the present state of the art, the data are not conclusive, and therefore, can not be used neither in favour nor against the high affinity site.

The question of distribution of proteins along pre-mRNA and stoichiometry of binding were investigated by analytical ultracentrifugation in collaboration with the laboratory of Arthur Rowe (University of Leicester). The study did not yield any results due to methodological problems (*e.g.*, degradation of RNA) and, mainly, due to the absence of a mathematical apparatus for the interpretation of experimental data.

2.3. DISCUSSION: Much Ado about Little and Lots

The polar effects of hnRNP A1 and SF2/ASF in alternative splicing and the possibility that it might be modulated by variations in the relative levels of these factors brought about the concept of antagonism. This describes the observed phenomenon, although it tells little about the underlying mechanism. Based on the towering examples of multiple functions that SF2/ASF exhibits during both constitutive and alternative splicing, the proposed models try to explain the antagonism by the opposite effects of these factors on interactions of snRNPs with pre-mRNA during spliceosome formation. For example, it was shown that SF2/ASF increases U1 snRNP binding to the 5' splice site (Eperon *et al.*, 1993; Klotz *et al.*, 1994; Jamison *et al.*, 1995), and it was proposed that hnRNP A1 could be a factor weakening U1 snRNP interaction with RNA (Eperon *et al.*, 1993). Alternatively, SR proteins form stable complexes with ESE (Sun *et al.*, 1993; Ramchatesing *et al.*, 1995; Lynch and Maniatis, 1996) and hnRNPs oppose to this (Sun *et al.*, 1993) in a way not completely understood but could be similar as it is implied in 5' splice site selection. In this way, hnRNP A1 and SF2/ASF might be considered as effectors or modulators rather than the *prima facie* actors. The prospect can be invoked that the counteracting effect is due to the direct interaction between hnRNP A1 and SF2/ASF and their mutual inhibition. Although such a view is vulnerable because the intracellular protein concentrations vary a lot in different tissues (Kamma *et al.*, 1995; Hanamura *et al.*, 1998), we had tried to detect possible protein-protein interactions between hnRNP A1 and SF2/ASF, but the attempt failed (data not shown). Interestingly, a two-hybrid

screen did not detect an interaction between SF2/ASF and hnRNP A1 (Cartegni *et al.*, 1996). Nevertheless, Cartegni *et al.* (1996) proposed that SR protein recruitment to specific sites could be guided by their interaction with hnRNP proteins. They tested the idea by an *in vitro* pull-down assay using the GST-fusion of the C-terminus of hnRNP A1 (GST-Gly) and purified SR proteins. In this assay all SR proteins were bound to GST-Gly with the notable exception of SF2/ASF and SC35, although there was not enough evidence for the specificity of the interactions.

The main structural features of both proteins that imply their function are the N-terminal RNA binding domain. The presence of the different auxiliary domains is a secondary feature, and moreover, in both cases shared by the whole family of proteins. We decided to question whether the RNA-binding properties of hnRNP A1 and SF2/ASF alone, without additional factors, such as, for example, U1 snRNP, would be distinctive enough to explain their antagonistic effects. By this, we do not reject the importance of protein interaction with snRNP components of the spliceosome during splice site recognition and selection, but rather invert the hierarchy of the ascribed roles. In support of this idea, there were two observations: first, pre-incubation of these factors with pre-mRNA was a decisive prerequisite in some cases (Kohtz *et al.*, 1994; Fu, 1993), and second, sometimes the U1 snRNP could be substituted by the excess of SR proteins (Crispino *et al.*, 1994; Tarn and Steitz, 1994).

One of the major attributes of any binding is the question of specificity. In this respect, the existence of the high affinity binding sites for hnRNP A1 as well as their functional significance is a matter of intensive debate. On one hand, hnRNP A1 was identified as a member of a diverse hnRNP family involved in packaging of pre-mRNA. As pre-mRNA looks like randomly generated sequence (Senapathy, 1995) the hnRNP proteins are commonly considered to be non-specific RNA binding proteins (Svitkin *et al.*, 1996). This opinion is supported by the data obtained in several laboratories showing that *in vitro* hnRNP A1 binds to any if not all RNAs with certain affinities and even to single stranded DNA (Nadler *et al.*, 1991; Casas-Finet *et al.*, 1993). On another hand, there is accumulating amount of data demonstrating selective binding of hnRNP A1 to RNA targets. Swanson and Dreyfuss (1988b) showed that hnRNP A1 bind specifically to the 3' end of introns, and mutations which alter the conserved 3' splice site dinucleotide AG strongly impair the hnRNP A1

binding. Riva and co-workers (Buvoli *et al.*, 1990b) confirmed the preferential binding of hnRNP A1 to the 3' splice site of β-globin intron by UV-crosslinking. Also, the specific interactions of hnRNP A1 with the reiterated AUUUA sequences found in the 3' untranslated region of many labile mRNAs (Hamilton *et al.*, 1993) and human telomeric DNA and analogous RNA sequences (Ishikawa *et al.*, 1993) were reported. Burd and Dreyfuss (1994b), using a SELEX procedure, identified a consensus high affinity hnRNP A1 binding site, UAGGGA/U, that has some resemblance to consensus sequences for vertebrate 5' and 3' splice sites.

Recently Williams and co-workers published two articles, which seriously challenged the fact of specific binding of hnRNP A1 (Abdul-Manan *et al.*, 1996; Abdul-Manan and Williams, 1996). They designed 20 RNA and DNA oligos, which were reported to contain a high affinity sites, and compared the binding of hnRNP A1 to them by fluorescence technique. The affinity constants determined for naturally occurring RNA sequences, homo- and random- oligonucleotides ranged from a low of 1.1×10^6 M^{-1} (β-globin 3' splice site) to high of 4.2×10^7 M^{-1} (β-globin 5' splice site). The only exception was the original SELEX winner, which contained the two copies of UAGGGU/A separated by CU dinucleotide and had an affinity at least 10-times higher than others. Importantly, it was shown that the major determinants of hnRNP A1 specificity for the winner sequence reside in the Gly-rich C-terminal domain and not in RRMs. Moreover, it is likely that the winner oligonucleotides formed a G-tetrad, an unusual tetrameric structure. In this respect the SELEX winner may not have biological relevance. The conclusion was that hnRNP A1 binds promiscuously to oligoribonucleotides and contradicts to the results discussed above. It is very difficult to criticise these results from the methodological point of view. These data were obtained by the "forward" competition fluorescence assay, a quantitative approach, where the measurements of the affinity constant are carried out under real equilibrium conditions in contrast to the many previous studies.

Chabot and co-workers, studying the alternative splicing of the hnRNP A1 pre-mRNA, which yields two isoforms of hnRNP A1 differing in their ability to modulate 5' splice site selection, identified 17 nt (CE1a) and 24 nt (CE4) sequences within the conserved intron elements, which are responsible for exclusion of the alternative exon 7B from the

hnRNP A1 pre-mRNA (Chabot *et al.*, 1997; Blanchette and Chabot, 1999). The data suggest that hnRNP A1 binds to the CE1a and CE4 element and promotes exon skipping, and hence, autoregulates splicing of its own pre-mRNA. Both CE1a and CE4 contain the sequence UAGAGU, which closely matches a SELEX winner UAGGGU. Thus, these elements may serve as an example of natural occurrence of hnRNP A1 high affinity binding sites. The recent publication from the laboratory of Breathnach describes similar elements in HIV-1 tat exon 2 and the human fibroblast growth factor receptor 2 K-SAM exon that function as splicing silencers upon binding of hnRNP A1 to them (Del Gatto-Konczak *et al.*, 1999).

In arguing the specificity in binding of hnRNP A1, someone puts himself in front of a series of controversial data without clear conclusion. It appears that the issue of specificity is not more clear regarding SF2/ASF. Indeed, SF2/ASF was intensively studied in respect to its RNA-binding specificity and the existence of preferential binding for SF2/ASF has never been never challenged, in contrast to hnRNP A1. First, Zuo and Manley (1994) showed, using UV-crosslinking and electrophoretic mobility shift assays, that the RNA binding domains of recombinant SF2/ASF can interact with RNA in a sequence-specific manner, recognising the 5' splice site in different pre-mRNAs, and point-mutations in the 5' splice site consensus reduce binding by a factor of 100. Then the issue of RNA binding specificity was addressed by comparison of SF2/ASF and SC35, another essential splicing factor and the member of SR protein family. The SELEX procedure was applied, and both SF2/ASF and SC35 selected mainly purine-rich sequences (Tacke and Manley, 1995). Comparison of individual sequences indicated that the motifs recognised are different. The functional importance of the SELEX winners has been proven in *in vitro* splicing assay as an element, containing three copies of a high-affinity SF2/ASF binding site, constituted a powerful splicing enhancer. Later on, Tacke *et al.* (1997) showed that phosphorylation of the RS domain has a striking effect on specific RNA binding of the SRp40 protein, as in the SELEX procedure unphosphorylated SRp40 failed to select specific sequences. The immediate question that comes to mind is whether the same applies to other SR proteins and SF2/ASF particularly. Xiao and Manley (1997) addressed the importance of phosphorylation of SF2/ASF but only in respect of interaction of its RS domain with U1-70kD protein. They showed that phosphorylation of the RS domain enhanced the binding of SF2/ASF to the U1 snRNP and eliminated the RS domain-RNA

interaction, which is presumably unspecific. The question whether phosphorylation affects specific binding of SF2/ASF to the RNA remains unanswered.

The question of specific recognition of ESE sequences by the SR proteins and the functional significance of these elements were again addressed in a recent study by Krainer and co-workers (Liu *et al.*, 1998). In contrast to conventional SELEX procedure based on selection for high affinity binding (Tuerk and Gold, 1990; Tacke and Manley, 1995), they used functional selection (Tian and Kole, 1995; Coulter *et al.*, 1997) from randomised ESE sequences to identify those which activate splicing upon addition of single SR protein. Two approaches gave very different results. First, the consensus sequences obtained by these two approaches are very different. The SF2/ASF motifs determined by binding (Tacke and Manley, 1995) are a purine-rich sub-group of those defined by function, with relatively low scores. Second, many of the winner sequences obtained by binding protocols were not functional as ESEs. Third, the consensus sequences obtained from the binding selection are less degenerate than those from functional selection. The conclusion was drawn that degenerate sequence specificity is essential for a limited number of SR proteins to recognise a very large number of ESE-containing exons in different genes.

Thus, for the inexperienced reader it is quite confusing to read that "these ESEs are highly specific" and later on in the same article that "SR proteins that bind RNA with limited sequence specificity" (Liu *et al.*, 1998). Another example regarding to hnRNP A1 is "while the combined interactions of a constellation of hnRNP proteins may package nascent pre-mRNAs in a form that is primed to undergo many types of maturation events, interactions between individual hnRNP proteins bound at specific sites may modulate pre-mRNA conformation and, in the case of hnRNP A1, promote splice site pairing" (Blanchett and Chabot, 1999). In these and many other possible examples, it is unclear what the authors mean by "specific binding", and, in the case of hnRNP A1, whether it is thought that the specific or non-specific binding of proteins to the RNA determines a functional role.

The data summarised above are inconclusive as far as the specificity of hnRNP A1 and SF2/ASF binding are concerned. Thus, it is not clear whether the RNA-binding preferences can determine the function of these proteins. Here, we did not address the RNA-binding specificity but the aim was to compare the binding of hnRNP A1 and SF2/ASF to the

long RNA containing the splicing signals in contrast to many previous data obtained for short RNA or oligonucleotides. The binding properties of hnRNP A1 and SF2/ASF appeared to be very different. We observed that hnRNP A1 does not protect RNA against RNase H digestion and it was explained by the very high dissociation rate of hnRNP A1 from RNA. We could not measure the dissociation rate precisely by the technique employed but estimated that hnRNP A1 dissociates within 5 seconds. Contrarily, SF2/ASF strongly protects RNA from RNase H cleavage at several sites tested, and the dissociation rate of SF2/ASF is much slower. The titration curves for these proteins were very different, indicating that the main determinant of hnRNP A1 binding is co-operativity, whereas SF2/ASF is able to discriminate different binding sites on the RNA. Moreover, SF2/ASF can rapidly displace hnRNP A1 from RNA in the concentration dependent manner. Close thermodynamical characteristics (apparent equilibrium constant) and such different kinetic parameters of binding allowed us to propose a dynamic model for the interaction of hnRNP A1 and SF2/ASF with RNA. According to this model, many molecules of hnRNP A1 are bound to one RNA molecule and this binding is co-operative, which means that binding of all molecules of hnRNP A1 is interdependent. SF2/ASF recognises one or few specific sites on this RNA, and as soon as the site is vacant (due to the high rate of hnRNP A1 dissociation/association) it binds there and obstructs the re-binding of hnRNP A1 molecules. The departure of one hnRNP A1 molecule disrupts the co-operative binding to other molecules and initiates their dissociation (according to the metaphor of a peeling banana). By this action SF2/ASF liberates a stretch of RNA from hnRNP proteins. Despite the fact that SF2/ASF and hnRNP A1 have similar equilibrium constants and the total concentration of hnRNPs in the cell is much higher than SF2/ASF, it would take a relatively long period of time to reach an equilibrium because SF2/ASF has a low rate of dissociation. Thus, the gap of cleared RNA exists for a while, and it could be considered as a landing place for the snRNP particles (U1 or U2 snRNPs). The rationale for clearance is as follows. The snRNPs could compete with hnRNP proteins for binding to RNA, but particles are bulky and not manoeuvrable in comparison with a polypeptide, and the nature of interactions between snRNPs and pre-mRNA (RNA-RNA) is different from those of SF2/ASF or hnRNP A1 (RNA-protein). It might be energetically/kinetically advantageous for the entire process of splice site recognition that SF2/ASF should compete with hnRNPs, clear

the RNA, and allow the snRNPs to establish interactions with the RNA, rather than snRNPs would do it by themselves. The hypothesised process would not produce the protein-free gap at the 5' or 3' splice sites only. As the formation of the stretch is triggered by the SF2/ASF binding therefore, SF2/ASF determines the place of clearance and marks pre-mRNA for splicing. For instance, if SF2/ASF specifically binds to the sequences around 5' and 3' splice sites, it would make the landing places for U1 or U2 snRNPs, correspondingly. At the 3' splice site, SF2/ASF might be helped by U2AF, which is able to interact with it (Wu and Maniatis, 1993).

In terms of this hypothesis the counteracting effect of hnRNP A1 and SF2/ASF is not an antagonism. SF2/ASF plays an active role as a cleaner of specific places, whereas hnRNP A1 is a passive player (a social worker) that covers RNA unspecifically and protects RNA from secondary structure formation and degradation. The latter aspect of the hypothesis contradicts to those models which credit the active role for hnRNP A1 as an inhibitor of the splice site (Blanchette and Chabot, 1999; Del Gatto-Konczak *et al.*, 1999; Bai *et al.*, 1999).

The clearance proposed in this hypothesis was mimicked so in a model experiment. SF2/ASF displaced hnRNP A1 from RNA and it was anticipated that the binding of these proteins to RNA would be mutually exclusive. In other words, the binding of SF2/ASF was expected to clear from hnRNP A1 the whole RNA of 533 nt. But we have failed to prove such "all or nothing" effect. Our results suggest that both SF2/ASF and hnRNP A1 can be accommodated on one RNA molecule of this size. If the hypothesis stands then SF2/ASF makes shorter gaps of free RNA (for instance, 100 nt) but they should be big enough to serve as a landing place for the snRNPs (compare with a landing strip for aircraft).

This hypothesis is not supported with enough experimental data at present. In the future, it will be important to determine whether cleared stretches of RNA form by an independent technique and to determine the length of gaps formed upon binding of one molecule of SF2/ASF (or other SR protein). This length could correlate with the distance between exon splicing enhancer and the splice site. The experimental data showed that the efficiency of splicing inversely depends on the distance between ESE and 3' splice site (Tian and Maniatis, 1994; Graveley *et al.*, 1998a). If the exon is short the binding of SF2/ASF to it would clear both the 3' splice site of an upstream intron and 5' splice site of downstream

intron, and by these means, the exon is defined without bridging interactions across exon proposed by the exon definition model (Berget, 1995). Such a situation would eliminate the necessity to disrupt the bridging interaction. If exons and introns are long, then SF2/ASF binding would introduce non-overlapping gaps, and the 5' and 3' splice sites would be determined independently. Consistent with this, Hertel and Maniatis (1999) provided the first evidence that there is no obligatory dependence of SR proteins function on an exon, as they showed that RNA substrates containing only 1 nt of exon sequence can undergo the first step of splicing *in vitro* and that this activity requires SR proteins, particularly SF2/ASF.

Here, we presented a simple but attractive way to envisage the hnRNP A1 - SF2/ASF antagonism.

CHAPTER 3

Initial characterisation of testis specific RBM protein, a possible candidate for tissue specific regulator of alternative splicing

3.1. INTRODUCTION

It is commonly accepted that the distinct structural motifs found in certain protein sequences provide an initial clue for understanding the protein function and determine the methodological approaches that aimed to reveal a molecular mechanism of the protein action. Under certain circumstances, particularly in the case of medical background when the protein of interest is encoded by a gene linked to genetic disorder or diseases, the only way to launch a project is based on the hypothesis that the proteins with similar structural domains would be involved in similar cellular processes. This chapter represents such an example as a study of the protein containing a classical RNA Binding Motif (RBM), whose function is unknown, but this protein is believed to play an important role in spermatogenesis because the gene was identified in the course of genetic analysis of male infertility.

Fertility is not purely of academic interest: despite the desire to have children, 2-7% of couples have not had a child at the end of their reproductive life (reviewed by Hargreave, 1997). Male infertility is often associated with either gross reduction of the number of sperm (oligozoospermia) or their complete absence (azoospermia) in the ejaculate. The individuals with oligo- and azoospermia are usually otherwise healthy, suggesting that any genes involved in these defects should either be only expressed in spermatogenesis, or be functionally required for spermatogenesis. The process, generating male germ cells, involves the unique form of cell division - meiosis - that produces haploid daughter cells rather than the diploids produced by cell division in the rest of the organism. All the major stages of spermatogenesis in the adult occur in the testis, which is anatomically subdivided into seminiferous tubules. Each of these developmental stages is found concurrently in the seminiferous tubules of the adult testis. The earliest germ cells (spermatogonia) are located at the periphery of the tubules and progressively more mature germ cells are found more interior to the tubule. Spermatogonia act both as stem cells to renew themselves and to provide precursors for the later stages of spermatogenesis. The most mature spermatogonial cells

replicate their DNA and differentiate into meiotic cells - spermatocytes. The initial phase of meiosis, meiotic prophase, is the critical period during which genetic recombination takes place. Then, spermatocytes undergo two sequential meiotic divisions to generate first, secondary spermatocytes and then haploid cells - round spermatids. Round spermatids next differentiate into elongating spermatids and then spermatozoa, in a process called spermiogenesis. During spermiogenesis the genome is re-packaged with protamines rather than histones, which is necessary to reduce the volume of the genetic payload from the relatively bulky round spermatid to the streamlined spermatozoa (reviewed by Elliott and Cook, 1997).

Spermatogenesis is such a complicated process that there are many points at which the normal production of sperm can be disrupted. The initial work for understanding of genetic influences of male infertility was guided by cytogenetic observations, principally on the Y chromosome. The Y chromosome is the smallest human chromosome, contributing only 2-3% of the haploid genome, and is restricted to males. At the cytogenetic level, the Y chromosome has a long and a short arm, demarcated by a centromeric region essential for chromosome segregation (Fig. 3.1). Only small regions homologous to the X chromosome at the tips of the Y chromosome undergo genetic recombination - the pseudoautosomal regions. The remainder of the chromosome is the only haploid compartment of the human genome and this should have evolutionary consequences. In practical terms, it was useful for genetic mapping as there are individuals without Y chromosomes and therefore, any deletions are viable and mutations cannot be masked by a normal allele.

In 1976 Tiepolo and Zuffardi using light microscope, observed the loss of the most distal segment of Yq, including all fluorescent heterochromatin (Yqh) in a number of infertile men. At the end of the eighties, more precise deletion mapping was done that was still based on cytogenetic observations of fairly large deletions of the Y chromosome. This study indicated that a putative gene controlling spermatogenesis, the azoospermia factor (AZF), named after the disease for which it is thought to be responsible, is located within interval 6 of the Y chromosome long arm (Fig. 3.1, Vergnaud *et al.*, 1986; Andersson *et al.*, 1988; Chandley *et al.*, 1989). Several individuals with a cytologically normal Y chromosome, but

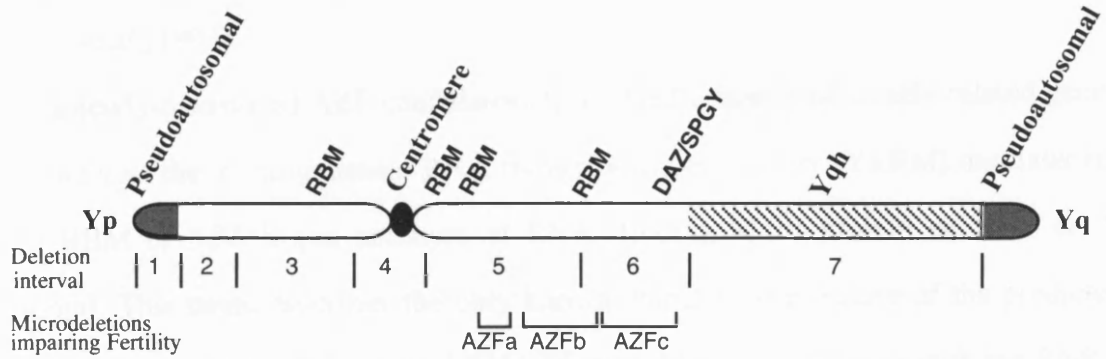


Figure 3.1. Map of the human Y chromosome showing candidate genes controlling spermatogenesis. The long and short arms of Y chromosome are called Yq and Yp, respectively. The map shows deletion intervals according to Vergnaud *et al.* (1986). The end of the long arm is composed of heterochromatin, called Yqh (hatched region). A microdeletion analysis identified at least three regions on the human Y chromosome, AZFa, AZFb, and AZFc, required for proper spermatogenesis as each might be independently deleted in infertile men (Vogt *et al.*, 1996). Despite the fact that RBM family members are found in different locations on the Y chromosome, only deletions in the AZFb region prevent the expression of the RBM protein (Ma *et al.*, 1993). Deletion of AZFc removes a small family of genes, individual members of which have been cloned and referred to as DAZ/SPGY (Reijo *et al.*, 1995). Both RBM and DAZ/SPGY contain a RNP-type RNA binding motif and internal sequence repeats, called SRGY and DAZ.

who are phenotypically oligo- or azoospermic men, were found and molecular screening by Southern blotting with a specific probe revealed microdeletions in interval 6, a region of approximately 200 kb (Ma *et al.*, 1992). Finally, at the end of 1993, the first candidate for the azoospermia factor controlling human spermatogenesis was identified by positional molecular cloning (Ma *et al.*, 1993).

The newly-discovered AZF candidates represented a family of closely related genes named initially as the Y chromosome RNA recognition motif family (YRRM) and later re-named to RBM or RBMY (an acronym of RNA Binding Motif protein encoded on Y chromosome). This name describes the only known characteristic feature of the predicted amino acid sequence that is the classical RNA Binding Motif of RNP type with the RNP-1 and RNP-2 consensuses. It is remarkable that a single N-terminal RRM domain of RBM exhibits 83% identity (90% similarity) to that of heterogeneous nuclear ribonucleoprotein G (hnRNP G) (Fig. 3.2), whose amino acid sequence was determined at the same time as RBM (Soulard *et al.*, 1993). The hnRNP G protein was first described as a component of hnRNP complexes (Pinol-Roma *et al.*, 1988), and independently as a novel autoantigen, p43 (Soulard *et al.*, 1991; 1993). The p43 is a glycoprotein, found in a variety of mammalian cells, was detected by circulating autoantibodies from dogs with lupus-like syndrome. Later hnRNP G and p43 appeared to be the same protein (Soulard *et al.*, 1993). A high overall sequence homology between RBM and hnRNP G (60% identity, 76% similarity) found throughout the entire sequence was recently explained as that mammalian spermatogenesis gene RBM evolved on the Y chromosome from hnRNP G encoded on the X chromosome (Delbridge et al., 1999; Mazeyrat *et al.*, 1999). The major difference between human RBM and hnRNP G is the additional central part present in RBM, which contains tandem repeats of 37 amino acids, called the SRGY boxes for their high content of serine (15%), arginine (20%), glycine (9%) and tyrosine (14%) (Fig. 3.2). Interestingly, hnRNP G contains only one box, the last repeat of 37 amino acid. Since each SRGY box has several SR/RS dipeptides, this protein might be assigned to the SR-like protein family (discussed in Chapter 1 and 2). Thus, a pure sequence analysis implies a function of RBM in RNA processing.

The high homology between RBM and hnRNP G could imply that these proteins perform similar roles although, the function of hnRNP G is not known. The important

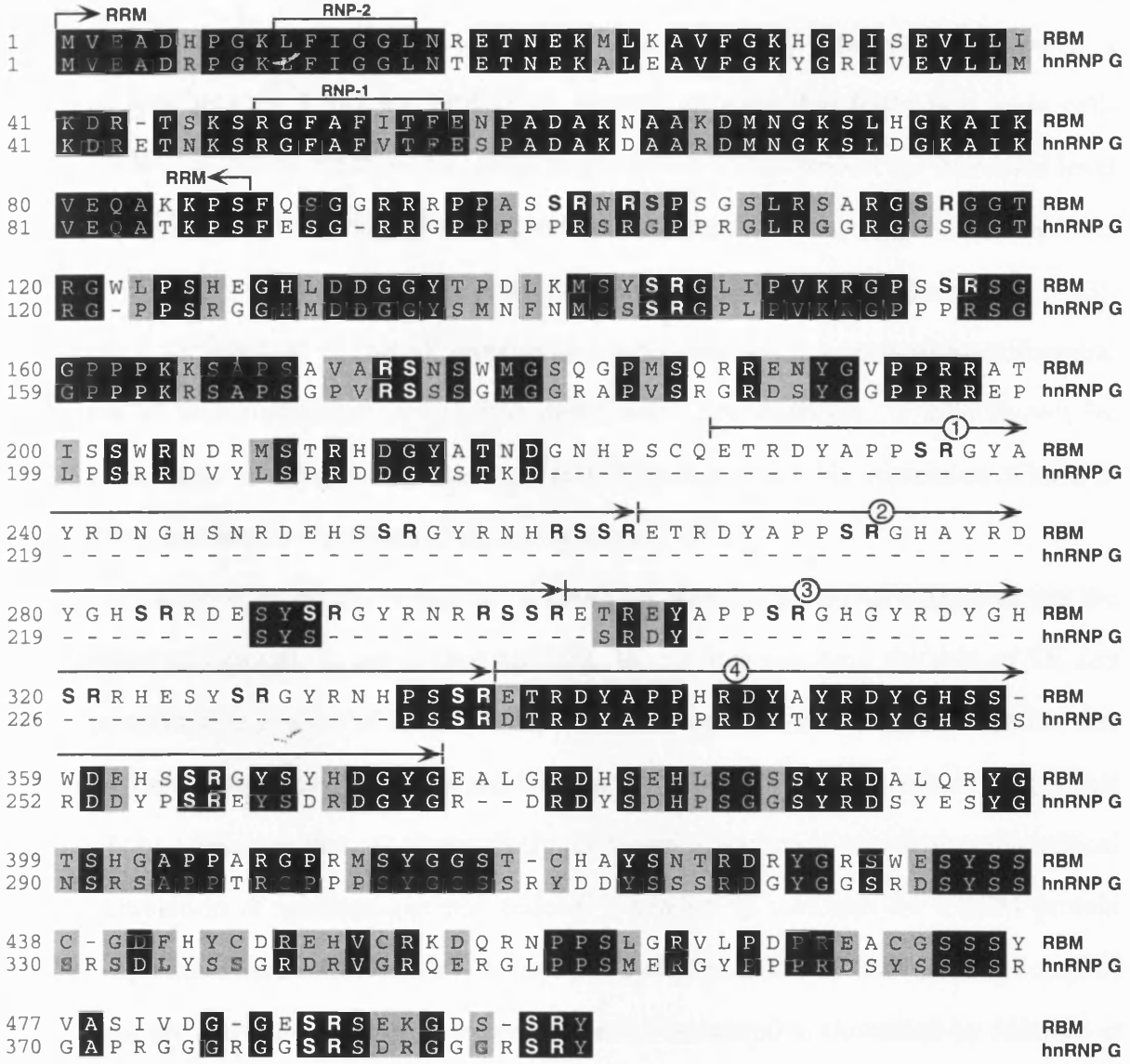


Figure 3.2. Comparison of the primary structure of the human RBM and hnRNP G proteins. The accession gene bank numbers of RBM and hnRNP G are NM_005058 and NM_005058, respectively. The alignment was carried out by the cluster method using the MegAlign program. The identical residues are boxed in black and the conserved residues are shaded grey. SR/RS dipeptides are shown in bold. Arrows indicate the SRGY repeats. Conserved amino acids are combined in the standard functional groups: acidic - DE; basic - RKH; hydrophobic - AFILMPVW; polar - CGNQSTY.

difference between RBM and hnRNP G is that expression of the former is restricted to testis only and, more specifically, to germ cells (Elliot *et al.*, 1998), whereas the expression of hnRNP G is thought to be ubiquitous (Dreyfuss *et al.*, 1993). Therefore, if there is a functional link between RBM and hnRNP G, one can propose that RBM is a germ cell-specific hnRNP G and its expression in testis might reflect a requirement for increased level of hnRNP G-like proteins during spermatogenesis. Consistent with this hypothesis, the level of hnRNP G is known to be elevated in actively dividing somatic cells (Soulard *et al.*, 1993). In addition, a participation of hnRNP proteins in spermatogenesis is supported by differential expression of some proteins in specific germ cells. For example, it was shown by immunohistological staining with monoclonal antibodies that detectable expression of hnRNP A1 in testis is restricted to spermatogonia (Kamma *et al.*, 1995).

It is plausible, especially taking into account the presence of SR/RS dipeptides that the RBM protein participates in pre-mRNA splicing. In previous chapters, the role of SR and hnRNP proteins in regulation of alternative splicing has been discussed. The observation that RBM is expressed specifically in testis and the possibility of its function in splicing argue that there might be transcripts that are alternatively processed in testis that encode proteins critical for the completion of spermatogenesis. Indeed, a number of isoforms for CREM protein (cAMP-responsible element modulator) are synthesised as products of alternative splicing, of which some are activators and some are repressors of transcription (reviewed by Monaco *et al.*, 1996). Prior to puberty, repressor isoforms of CREM are present at low levels in germ cells. At puberty, Follicular Stimulating Hormone (FSH) secreted by the pituitary gland induces a change in polyadenylation site, stabilising the mRNA encoding the transcriptional activator form, CREMτ. This stimulates then the transcription of a number of structural genes in round spermatids, such as protamine genes, that have cyclic AMP responsive elements in their promotors.

Another example of a germline-specific alternative splicing pattern serves the well-documented phenomenon of *Drosophila* P-element, what is known as intron retention. In *Drosophila*, certain diseases, which affect exclusively the germline cells including sterility due to abnormal gonad development, are thought to be results of the action of P transposable elements. P-element encodes a transposase protein which binds to specific DNA sequences

and catalyses its transposition (reviewed by Rio, 1991; Rio, 1993). In germ cells, all introns are properly removed and a full-length functional transposase is synthesised, whereas in somatic cells, the third intron (IVS3) is retained and a truncated protein is produced due to the presence of a stop codon in this intron. The truncated protein inhibits transposition; this explains why transpositions occur only in germ tissues. The molecular events in this case are already understood in some detail. It was found that exon 3, upstream of IVS3, contains two sequences which closely resemble a 5' splice site consensus, and it was shown that the U1 snRNP binds to the pseudo 5' splice site sequence in *Drosophila* cell extracts. The proposed, although not proven, mechanism is that the pseudo 5' splice site sequence competes with the authentic downstream 5' splice site despite not being used as a cryptic site. Two proteins, the *Drosophila* heterogeneous nuclear ribonucleoprotein particle protein hrp48, similar to the mammalian splicing factor hnRNP A1, and the P-element somatic inhibitor protein (PSI), containing three KH-type RNA-binding domains, were identified that specifically interact with exon 3. It was demonstrated using immunodepletion-reconstitution assays that these proteins are essential for inhibition of IVS3 splicing (Siebel *et al.*, 1994; Siebel *et al.*, 1995). A current hypothesis, supported already by *in vivo* data, is that hrp48 and PSI cooperatively drive the binding of U1 snRNP away from the downstream authentic 5' splice site, and thus, inhibit excision of IVS3 in somatic cells (Adams *et al.*, 1997; Hammond *et al.*, 1997).

The tissue-specific expression of RBM implies that it might be a cell-specific splicing factor. Until now, only general splicing factors that are expressed ubiquitously were shown to affect alternative splicing in vertebrates. Thus, the best-characterised example of alternative splicing that is modulated by regulatory factors are the genes involved in sex determination in *Drosophila* (reviewed by Lopez, 1998). In this case, there are four genes: Sex-lethal (*Sxl*), transformer (*tra*), transformer 2 (*tra-2*) and doublesex (*dsx*), involved in a cascade. Splicing of the Sxl, Tra and Dsx pre-mRNAs are non-regulated events in males, which require only the general splicing machinery and produce default splicing patterns. The protein product of the sex-lethal gene (*sxl*) is expressed in female but not male embryos. The Sxl protein regulates its own splicing as well as that of the *tra* gene. The default, male-specific splice of Tra, results in the non-functional RNA. Thus, the *sxl* product is resposible for the Tra protein production exclusively in females. The Tra protein, along with the unregulated product of the *tra-2* gene,

dictates female-specific splicing of the *dsx* gene. The female Dsx protein represses male differentiation, whereas the male protein represses female differentiation. The *sxl* and *tra-2* genes encode proteins with RNA-binding domains, and it was shown that the regulation of splicing is achieved through the binding of these proteins to specific sequences in the corresponding pre-mRNAs and through the interaction with other proteins. The Sxl protein binds to the stronger polypyrimidine tract of the upstream 3' splice site of exon 2 of the Tra pre-mRNA, thereby inhibiting the binding of U2AF there and allowing splicing at a weaker downstream 3' splice site. Both Tra and Tra-2 are SR proteins and interact cooperatively with a sequence located within exon 4 of the Dsx pre-mRNA, the dsxRE, which includes six repeats of a 13 nucleotide sequence and an 18 nucleotide purine-rich enhancer (Lynch and Maniatis, 1995). They recruit other SR proteins, as well as U2AF, to the upstream 3' splice site, and thus, directly activate usage of the weak 3' splice site of exon 4 of the Dsx pre-mRNA (Lynch and Maniatis, 1996).

In summary, there are logical reasons to anticipate that an action of RBM in spermatogenesis could be realised at the level of RNA processing, particularly as a tissue-specific factor, that is responsible for alternative splicing events. To understand the function of RBM, first, the biochemical characterisation of the protein is required which could be then combined with cytological and immunohistological observations on the way to reveal the origin of the genetic disorder.

Objectives and Strategy

In general, the characterisation of RBM as a protein, particularly its interaction with specific cellular components and RNA, requires a strong biochemical background and experience in the RNA processing field. The RBM gene was identified by joint efforts of several UK laboratories, which are mainly involved in medical and cytological research. Thus, it was natural to approach RBM function via collaboration with a laboratory specialised in the study of RNA processing. Such a collaboration was established between the laboratories of Dr. Ian Eperon, University of Leicester, and Dr. Howard Cooke, Medical Research Council Human Genetics Unit, Edinburgh.

In 1994, the sequence of RBM was known, but the only information about expression of RBM was at the RNA level. Northern blot analysis revealed a transcript in total mRNA from adult testis with no detectable signal in male brain, heart, kidney, liver, lung, prostate gland or skeletal muscle (Ma *et al.*, 1993). Therefore, my initial task was to reveal the expression of RBM on a protein level. For these purposes good antibodies are necessary that would allow the detection of protein in different tissues and would help to follow the cellular location of the protein during different stages of spermatogenesis. At the beginning of such a study, polyclonal antibodies raised in animals immunised with a whole protein have certain advantages for cytological experiments over monoclonal antibodies or antibodies directed against a peptide, because the latter is often not able to recognise a native protein if an epitope is buried inside a protein molecule. The best way is to use a recombinant protein as an antigene for immunisation. The recombinant RBM protein would be also necessary for investigation of its possible effect on pre-mRNA splicing as a lot of information in the field was obtained from *in vitro* splicing assays.

Some work had been done in the laboratory before I started the RBM project. A cDNA clone, encoding the putative RBM protein, had been received from Dr. H. Cooke. The recombinant protein was expressed in *E. coli* from the plasmid pYRT3, a pRSET A (Invitrogen) derivative, carrying RBM under the control of a T7 promotor with an N-terminal fusion of six histidine residues to the 8th amino acid of RBM. The recombinant protein was found in inclusion bodies and, therefore, was purified by affinity chromotography on a Ni^{+}-column in the presence of 6 M urea or guanidine hydrochloride. All attempts to remove denaturants completely led to precipitation of the protein. Thus, this preparation of recombinant RBM could not be used for a functional study. Nevertheless, the recombinant protein solubilised in urea solution was assayed for possible effects on *in vitro* splicing, presuming that after addition of the protein solution into the nuclear extract some of the molecules will fold properly in order to find the functional partners and produce an effect. The result of the experiment showed that this recombinant RBM protein inhibited splicing of two different pre-mRNAs in a concentration dependent manner whereas urea present in the reaction at the same concentrations did not. This result was considered as promising for further investigations of a possible role of RBM protein in RNA splicing, bearing in mind the

problem of insolubility. If the usefulness of this protein preparation for *in vitro* assay was rather questionable, it was perfectly suitable for raising polyclonal antibodies - the stage where I inherited the project.

The project was methodologically divided in two main streams: *in vivo* and *in vitro*. The *in vivo* part utilised antibodies against RBM and was based on transfection of HeLa cells with plasmids encoding RBM. I intended to investigate the cellular localisation of RBM after transient transfection of HeLa cells using both biochemical approaches for separation of cellular compartments as well as immunostaining of permeablised HeLa cells. The histological immunostaining of tissue samples was carried out by our collaborators (MRC, Edinburgh) using our antibodies. I also planned the construction of a stable HeLa cell line expressing RBM in order to study the involvement of the protein in different cellular processes.

The *in vitro* approach was designed to identify the potential interacting partners for RBM hoping that they would be known factors with determined function. This approach mainly relies on the availability of the soluble recombinant protein and utilises co-immunoprecipitaion technique. Independently, immunoprecipitation of RBM from extracts of the transiently transfected HeLa cells could be used to precipitate a complex of RBM with its interacting partners. The availability of purified recombinant protein was also essential for the investigation of its possible effects on splicing *in vitro*. Therefore, my initial task was to try different expression systems to isolate the soluble recombinant RBM. Alternatively, the protein can be obtained by *in vitro* translation in the rabbit reticulocyte lysate. This system has advantages as it is more likely that a human protein synthesised in eukaryotic environment would be functionally active because some post-translational modifications could be required (*e.g.*, phosphorylation). On the other hand, the total amount of the protein produced could be a limiting factor to obtain the results.

3.2. RESULTS AND DISCUSSION

3.2.1. Construction of the plasmid encoding RBM

To carry out the proposed experiments that do not utilise the recombinant protein, RBM has to be introduced into the plasmids, which allowed: 1) transient transfection of HeLa

cells; 2) selection of stable cell line; 3) *in vitro* translation. The vector plasmid pcDNA3 (Invitrogen) had been chosen to re-clone the RBM cDNA because the same plasmid can be used for all purposes. This plasmid utilises the cytomegalovirus (CMV) efficient enhancer-promoter, which gives a high yield of expression in mammalian cells. The plasmid carries two selective markers, ampicillin resistance for easy propagation in *E. coli*, and neomycin resistance for selection of stable eukaryotic cell line. A T7 promoter could be used for *in vitro* transcription to make mRNA for *in vitro* translation, or alternatively, the plasmid is suitable for the TNT coupled reticulocyte lysate system (Promega). Moreover, the plasmid pcDNA3 has a multiple cloning site (MCS) containing the EcoRI site. The original cDNA clone, pMK5 (Ma *et al.*, 1993) encoding the putative RBM protein, is a derivative of pBluescript SK II(+) and contains the cDNA cloned into the EcoRI site. Therefore, the cDNA of RBM was transferred from pMK5 into pcDNA3 vector using the EcoRI site and variants were selected for the proper orientation of the gene. The selected clones were sequenced at junctions for about 300 nucleotides from each side. The constructed plasmid encoded a full-length RBM protein and was named pOM6.

3.3.2. Biochemical analysis of the transient expression of RMB in HeLa cells

The polyclonal antibodies were raised in rabbit immunised with the recombinant RBM protein that had been expressed in *E. coli* and purified in the presence of 6 M urea as described in the Materials and Methods. The initial question was whether this antiserum could be used to detect expression of the RBM gene at the protein level in human tissues or not. Based on northern blot analysis, the expression of RBM should be restricted to the human testis only, but human tissues are not widely available, and our laboratory did not have access to samples of the human testis during the initial stages of this project. Therefore, for characterisation of the antibody, transient transfection of eukaryotic cells that do not express RBM was employed. HeLa cells had been chosen for transfection due to its female origin and simplicity of cultivation and manipulation.

First, it was confirmed that HeLa cells do not produce any proteins of molecular weight around 60 kDa, which would be recognised by α-RBM antibody (Fig. 3.3A). For transient expression of RBM, HeLa cells were transfected with pOM6 plasmid by calcium-phosphate procedure as described in the Materials and Methods, and cells were harvested 24

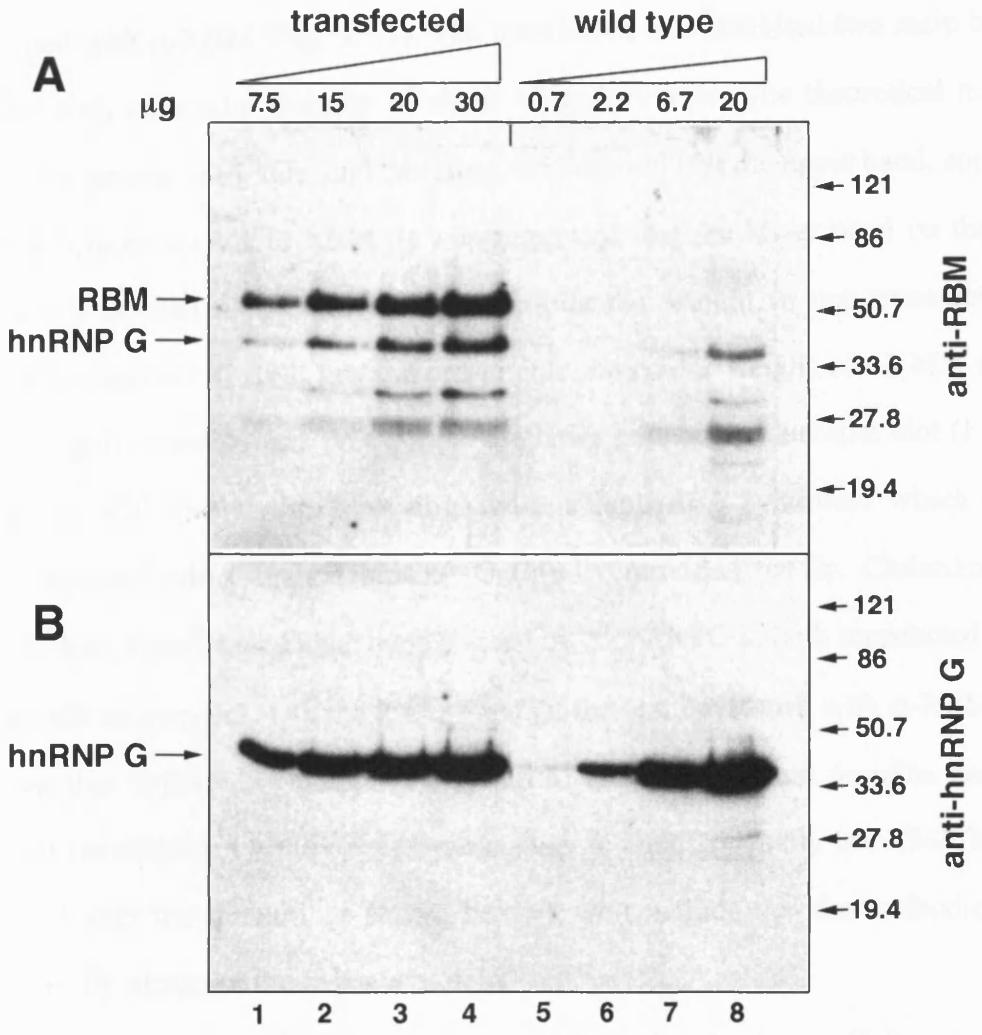


Figure 3.3. Analysis of transient expression of RBM in HeLa cells by western blotting. The indicated amounts of the whole cell lysate prepared from HeLa cells transfected with pOM6 (lanes 1-4) or untransfected HeLa cells (lanes 5-8) were applied on 10% SDS-PAGE, transferred on a membrane, and immunostained with α-RBM (**A**) or with the serum of a dog with lupus-like syndrome (**B**). Both membranes were developed with protein A conjugated to horseradish peroxidase (Sigma) followed by ECL-detection (Amersham).

hours after transfection. The total cell lysate was prepared in buffer containing 6 M urea and subjected to 10% SDS-PAGE followed by electro-blotting. The nitrocellulose membrane was immunostained with α-RBM (Fig. 3.3A). The transfected cell exhibited two main bands on western blot with molecular weights of about 40 and 60 kDa. The theoretical molecular weight of RBM protein is 61 kDa, and therefore, we believed that the upper band, appeared in trasfected cells, corresponds to RBM. It was suggested that the lower band on the blot of transfected cell as well as a band of similar molecular weight in non-transfected cells corresponded to hnRNP G as it has the reasonable molecular weight of 43 kDa and both proteins are highly homologous. To examine the latter possibility a parallel blot (Fig. 3.3B) was developed with the serum from dogs with a lupus-like syndrome which contains circulating autoantibodies against hnRNP G (kindly provided by Dr. Christian-Jacques Larsen, INSERM, Paris). Indeed, the band stained by α-hnRNPG in both transfected and wild type HeLa cells co-migrates with the lower band on the blot developed with α-RBM. Taking into account that hnRNP G belongs to a group of abundant nuclear proteins and that an efficiency of transfection by the calcium-phosphate method is usually low (5-20% of cells express RBM after transfection, as shown below), we conclude that the antibodies against RBM specifically recognise the cognate protein.

The important issue for whether the protein is involved in splicing or not is its subcellular localisation. Since splicing takes place in the nucleus all factors involved in this process have a predominantly nuclear localisation. Therefore, if the RBM would not be localised in the nucleus of transiently transfected cells, it would be a fateful moment for the project as one must abandon the idea of RBM involvement in splicing. At this point, a subcellular localisation of RBM was first approached biochemically.

A small-scale procedure for nuclear extract preparation described by Lee *et al.* (1988) was used for a crude separation of the cell lysates on a nuclear and a cytosolic fractions. This procedure is based on the same principles that were first established by Dignam *et al.* (1983) for the nuclear extract preparation active in transcription, but it was applied to the small amount of cells that could be scraped off the bottom of petri dish (described in the Materials and Methods). Figure 3.4A shows the western blot analysis of subcellular fractions from the cells transfected with pOM6 (+) and pcDNA3 (-), a mock transfection with a parental vector.

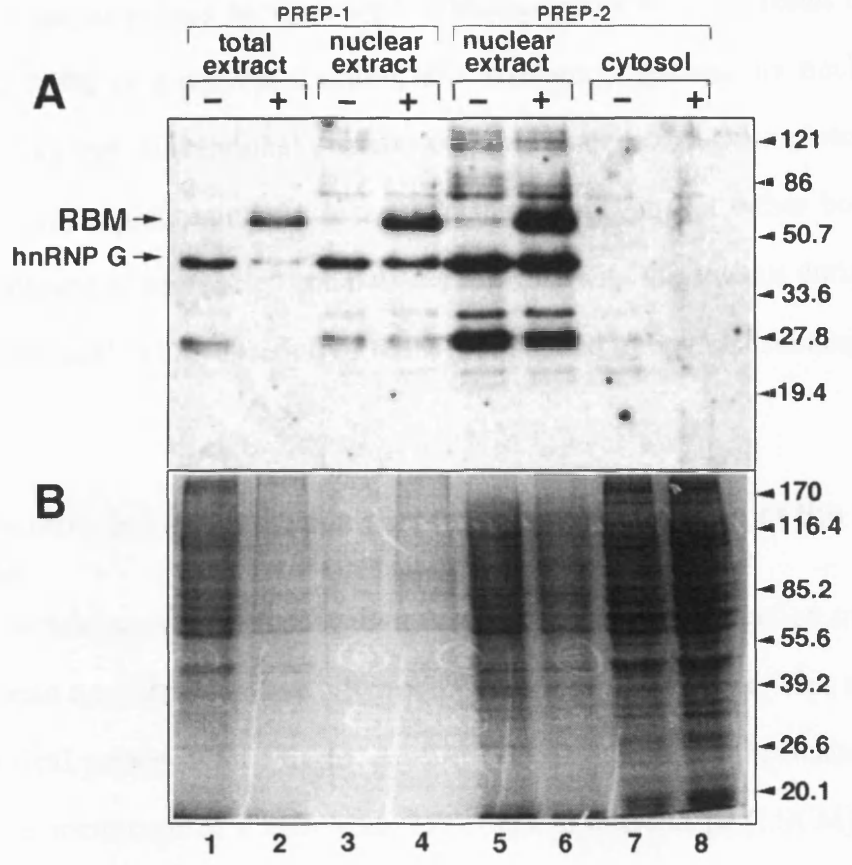


Figure 3.4. Western blot analysis of the nuclear and cytosolic fractions of HeLa cells after transient transfection with RMB. Two preparations of HeLa nuclear extracts were analysed by 10% SDS-PAGE followed by Coomassie staining (**B**) or western blotting (**A**). Lanes 1-4 represent extracts from one experiment and lanes 5-8 correspond to the separate experiment. The (+) and (-) signs are attributed to the transfections carried out with pOM6 (containing RBM gene) and pcDNA3 (parental vector), respectively. In the first experiment (lanes 1-4), the total cell extract was compared to the nuclear extract (3 μg of proteins was applied to lanes 1-4). In the second experiment (lanes 5-8), a nuclear extract is compared with the complementary cytosolic fraction (10 μg of proteins was applied to lanes 5-8).

In this experiment, the supernatant left after pelleting the nuclei was used as a cytosolic fraction. The signals corresponding to the RBM were detected only in the nuclear fractions of the cells transfected with RBM even though much higher amounts of cytosolic proteins were applied on a gel, as judged by Coomassie staining (panel B). This result demonstrated the presence of RBM in a nuclear fraction, and strongly suggested its nuclear localisation. However, using this conventional procedure one can not rule out completely the possibility that the protein is predominantly located in the cytoplasm but either bound to the outer nuclear membrane or aggregated and pelleted together with the nucleus during centrifugation. Therefore, a subcellular localisation of RBM was studied by immunostaining of permeablised HeLa cells.

3.2.3. Immunocytological analysis of transient expression of the RBM protein in HeLa cells

To determine subcellular location of the RBM after transfection of HeLa cells with pOM6 plasmid the indirect immunofluorescence technique was used. This technique is based on the physical principles of fluorescence optics and utilises sophisticated equipment as a fluorescence microscope or a laser-scanning confocal microscope (LSCM) (described in the Materials and Methods).

For specific staining of cellular proteins with fluorescent dyes, the indirect technique is usually used. In this method the first antibodies raised against specific protein are unlabelled, and after their interaction with the antigen, the secondary antibodies, produced against immunoglobulins of the species in which the first antibodies are made, would bind the first antibodies. The secondary antibodies are coupled to a fluorophore, which would emit the light of distinctive colour after illumination with the primary light. Thus, the distribution of the antigen in the cell, immobilised on a glass support, would be visualised in a fluorescence microscope equipped with the appropriate filters. The second antibodies are commercially available and usually conjugated with FITC (fluorescein isothiocyanate) producing a green light or with the red fluorophores, Texas Red and Cy3 (cyanine 3). If the first antibodies were raised against specific proteins in the different species, for instance, in mouse and in rabbit, and the secondary antibodies, recognising immunoglobulins of these species, are conjugated

to the fluorophores emitting light of different wavelength, several antigens could be visualised simultaneously using the appropriate filters.

The usage of antibodies to localise the proteins within cells requires that the cells are chemically fixed and rendered permeable to the antibody molecules. There are two basic fixation procedures, "weak" fixation with inorganic solvents (acetone, methanol) and "strong" fixation with aldehydes. It is generally accepted that stronger fixation usually gives better structural preservation but leads to weaker staining with antibodies. Both the Formaldehyde/Triton method and the Acetone/Methanol methods had been tried (as described in the Materials and Methods), and it was found that the latter produced more diffuse staining with higher background. The images presented below were obtained from cells fixed by the Formaldehyde/Triton method.

To immunostain RMB in HeLa cells after transfection, a secondary antibody conjugated to Texas Red was chosen taking into account the filter sets available. Both in Leicester and in Marburg, the LSCM facilities were accessible for analysis of our specimens, which were initially screened on the laboratory fluorescence microscope. The laser scanning of the specimen is combined with the photodetection of emitted signal from every point and then analysed by the electronic image processing system of the LSCM. The images of the same specimen could be recorded using different filters, which allows the detection of the specific flourophores. The images could be converted into computer files in formats (for instance, TIFF) compatible with the personal computer programs for analysis of photographic images. In this work, the Adobe PhotoShop program was used to simulate the green and red fluorescence colours and carry out the overlay of one image over another.

All the experiments documented in this section were carried out in the following manner. HeLa cells were transfected with the pOM6 plasmid, encoding the RBM, fixed on the coverslips, and then were stained with the two antibodies, one of which was the protein-A-purified rabbit antibody against RBM and another was the monoclonal mouse antibody against the protein to which the immunostaining pattern was compared. The fluorescence was brought by a pair of the secondary antibodies. The deep red colour is characteristic to the Texas Red conjugated anti-rabbit IgG antibodies, and therefore, is attributed to the location of

the RBM protein. The green colour is produced by the FITC conjugated anti-mouse antibodies and it determines the location of the second protein.

Figure 3.5 shows the LSCM images obtained in experiments of the double staining with α-RBM and KSm2 (the monoclonal antibodies against Sm proteins of snRNPs were kindly provided by Dr. D. Williams). These experiments unambiguously demonstrated that the RBM localises in the nucleus of transfected cells. The left panel (Fig. 3.5, panel 1) allows the efficiency of transfection to be estimated. In the field shown, two out of the eleven cells were transfected and this is a representative picture of many experiments. The efficiency of transfection was usually between 5% and 20% which correlates with literature data for the calcium phosphate method of transfection.

The images were enlarged (Fig. 3.5, panel 2) to compare the RBM and Sm immunostaining, as the latter represents a classical nuclear staining of the main splicing components, snRNPs. There are similarities and differences in staining with these two antibodies. Neither antibody stains nucleoli, producing 2-4 big dark holes in the nuclear staining. In nucleoli, the processing of ribosomal RNAs and ribosome assembly take place, but components of the splicing machinery are excluded from these structures. Another feature observed is that both α-RBM and α-Sm antibodies produce a speckled pattern, although these patterns are not the same. It seems that RBM is more homogeneously distributed throughout the nucleus, although the speckles are seen better when the brightness of the images is reduced. α-Sm produced more defined speckled pattern, with 3-5 small very bright spots in the nucleoplasm, known as coiled bodies. In spite of being first identified in 1903 at the light microscope level, coiled bodies were characterised on molecular level much later and their functions are still not defined well (reviewed by Lamond and Carmo-Fonseca, 1993; Huang and Spector, 1997). Coiled bodies are round structures of 0.5-1.0 mm in diameter and consist of coiled fibrillar strands. It is well established now that snRNPs are concentrated in coiled bodies (Carmo-Fonseca *et al.*, 1992; Carmo-Fonseca *et al.*, 1993), though their presence in the nucleus is not restricted to bright foci. However, the absence of newly transcribed RNA, hnRNP proteins, and several non-snRNP splicing factors in coiled bodies supports the idea that these inclusions are not involved directly in pre-mRNA splicing but rather may be essential for biogenesis of snRNAs, storage or recycling (discussed by Huang and Spector,

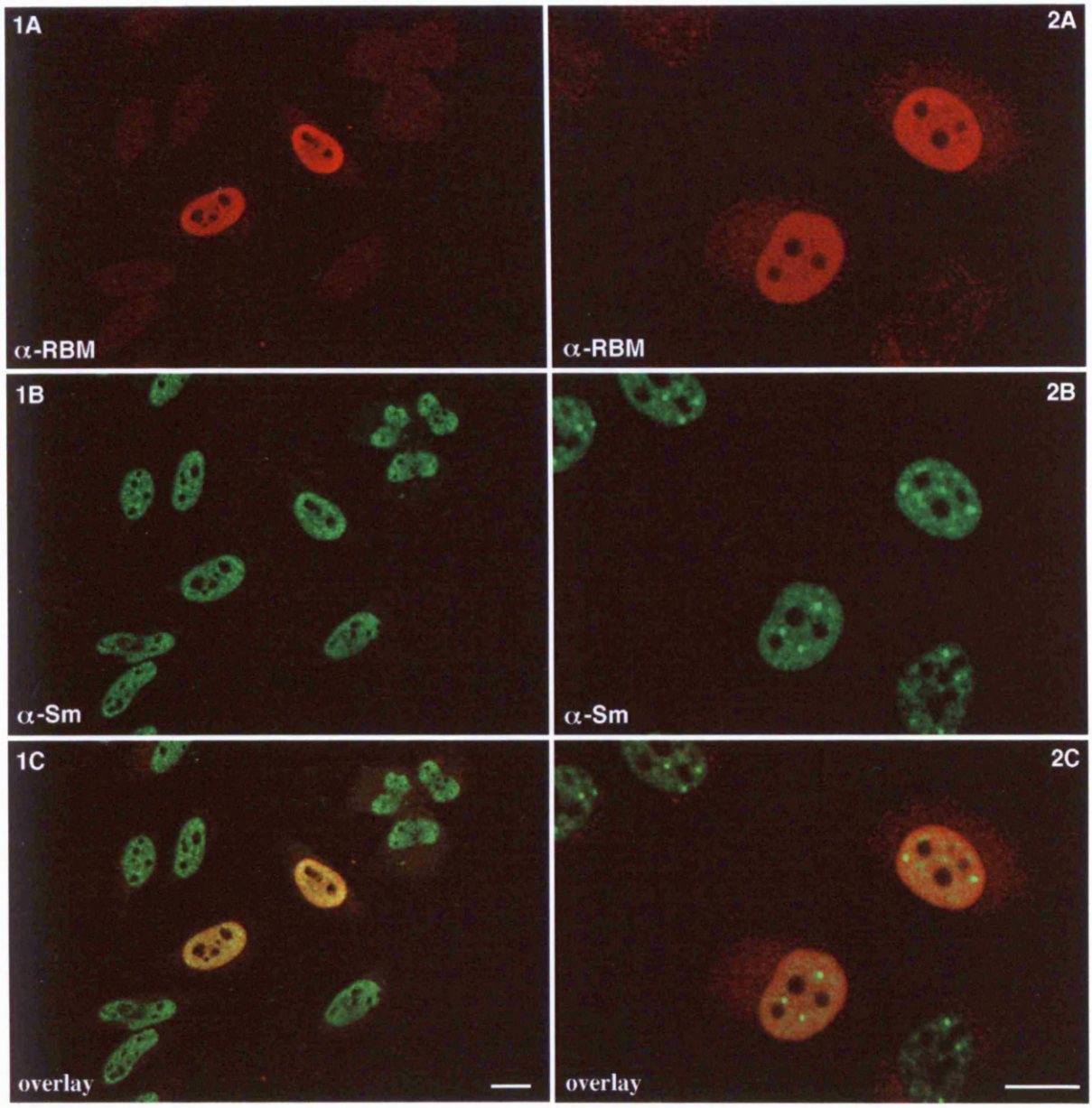


Figure 3.5. Immunofluorescence staining of the RBM transfected HeLa cells with α-RBM and α-Sm antibodies. Two sets of images (1 and 2) show HeLa cells transfected with pOM6 and double-stained with α-RBM (A) and α-Sm (B). The red fluorescence (A) corresponds to the RBM protein and produced by secondary antibody conjugated to Texas Red. The green fluorescence (B) corresponds to the Sm proteins and produced by the secondary antibody conjugated to FITC. The images A and B were recorded independently and combined in an overlay image C resulting in yellow staining of structures decorated by both antibodies. The bar indicates 10 μm.

1997). It is evident from the overlay that RBM is not present in coiled bodies. The speckled pattern (25-50 more or less defined speckles) produced by both α-RBM and α-Sm is characteristic of splicing factors and corresponds to perichromatin fibrils and interchromatin granule clusters at the electron microscopic level. Several studies have shown that highly active genes and their transcripts are associated with the periphery of speckled domains (reviewed by Huang and Spector, 1997). Therefore, speckles are considered as sites of splicing or storage of splicing factors.

Since the hypothesis of RBM function in RNA processing was not contradicted by the immunofluorescence approach, it was interesting to compare the nuclear distribution of RBM with other splicing components to which monoclonal antibodies were available in the laboratory. Figure 3.6 shows a double staining of the RBM transfected Hela cells with α-RBM and α-SF2/ASF (kindly provided by Dr. A. Krainer, Cold Spring Harbor Laboratory) and Figure 3.7 – with α-RBM and α-SR (3CSH4 monoclonal antibody, kindly provided by Dr. B. Turner, Birmingham). In all these cases staining is restricted to the nucleus, does not highlight nucleoli and coiled bodies, and the patterns produced are more similar to each other than to the one produced by α-Sm. Particularly, the speckled pattern obtained with α-RBM and α-SR are the most similar as the superimposition of the red and green staining causes the regions to be yellow (Fig. 3.7). This observation is in agreement with the presence of RS dipeptides in RBM, and supports the possibility that the RBM protein can have a function similar to SR proteins.

The nuclear localisation of the RBM protein with a pattern similar to SR proteins firmly convinced us to continue investigating the possible role of this protein in splicing. Analysis of gene function frequently requires the formation of mammalian cell lines that contain the studied gene in a stably integrated form. This general methodological approach was chosen to study RBM protein on cytological level, and since RBM is not expressed in HeLa cell, an attempt to create a HeLa cell line stably expressing RBM protein has been undertaken.

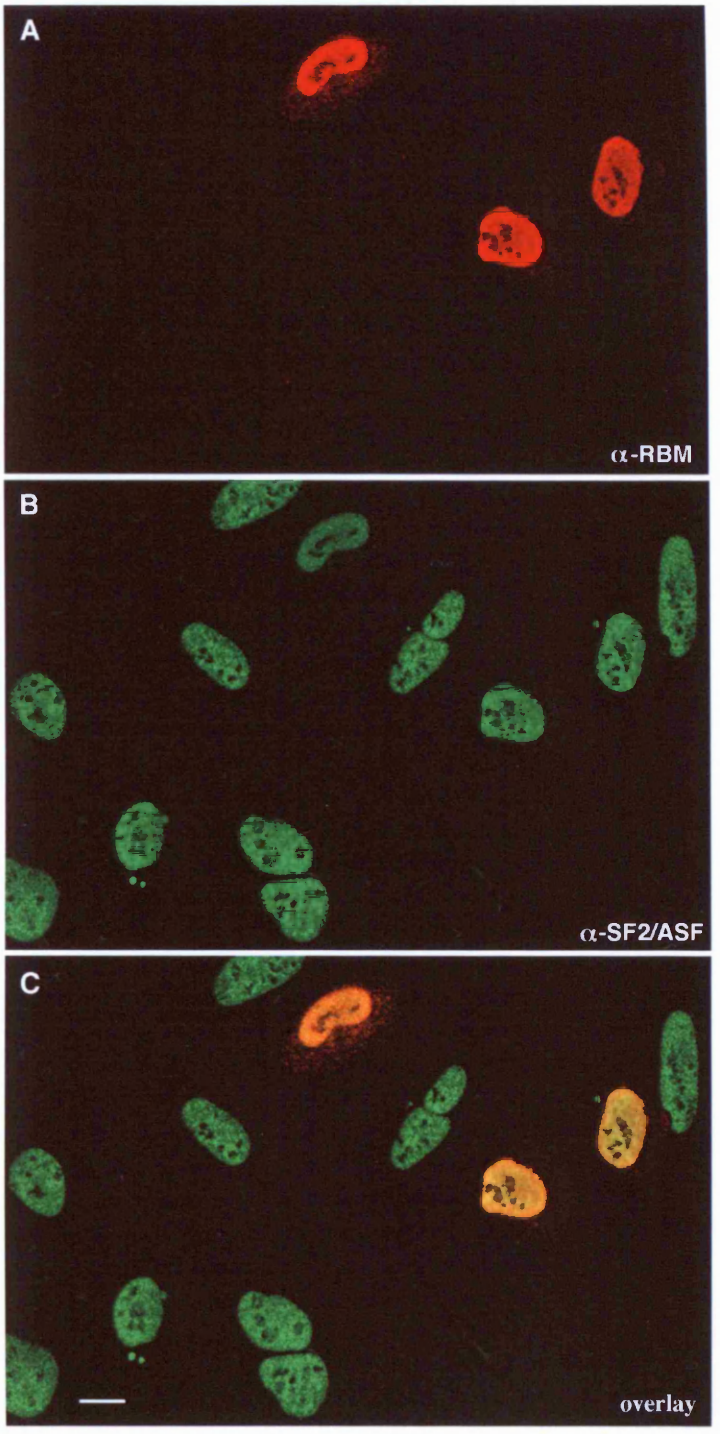


Figure 3.6. Immunofluorescence staining of the RBM transfected HeLa cells with α-RBM and α-SF2/ASF antibodies. HeLa cells were transfected with pOM6 and double-stained with α-RBM (A) and α-SF2/ASF (B). The red fluorescence (A) corresponds to the RBM protein and produced by secondary antibody conjugated to Texas Red. The green fluorescence (B) corresponds to the SF2/ASF protein and produced by secondary antibody conjugated to FITC. The images A and B were recorded independently and combined in an overlay image (C). The bar indicates 10 μm.

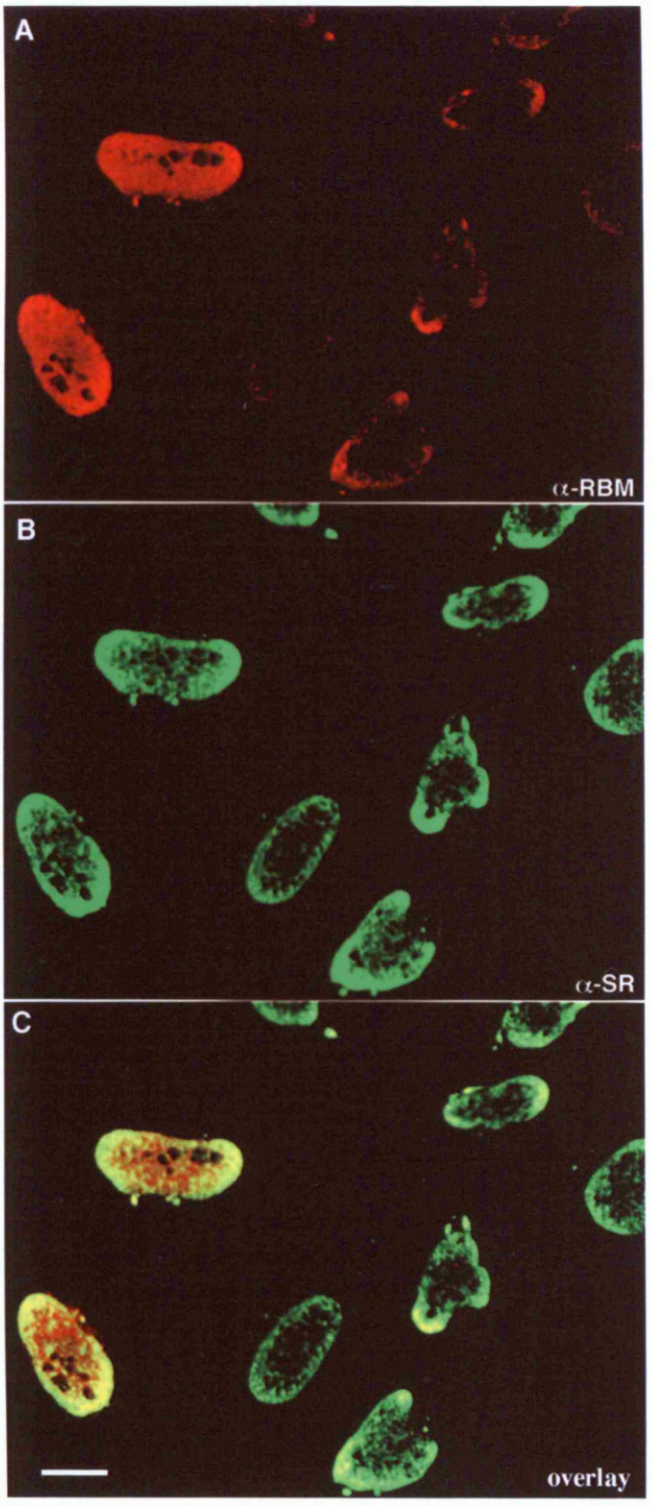


Figure 3.7. Immunofluorescence staining of the RBM transfected HeLa cells with α-RBM and α-SR antibodies. HeLa cells were transfected with pOM6 and double-stained with α-RBM (A) and α-SR (B). The red fluorescence (A) corresponds to the RBM protein and produced by secondary antibody conjugated to Texas Red. The green fluorescence (B) corresponds to SR proteins and produced by secondary antibody conjugated to FITC. The images A and B were recorded independently and combined in an overlay image (C). The bar indicates 10 μm.

3.2.4. Construction of a cell line expressing RBM protein

In order to analyse the effect of RBM protein on RNA metabolism of living cells we hoped to select HeLa cells that, after transient transfection, would incorporate the RBM gene into a chromosome via recombination processes and would continuously express RBM protein. Approximately one in 10,000 cells in a transfection would stably integrate DNA and a dominant selectable marker is used to permit isolation of stable transfectants. The parental vector pcDNA3 contains the bacterial gene aminoglycoside phosphotransferase (APH or *neo*) as a dominant selectable marker that could be used in the presence of G418 (GIBCO-BRL), an aminoglycoside, similar in structure to neomycin, gentamycin and kanamycin. G418 inhibits protein synthesis in mammalian cells by interfering with ribosomal function. The expression of APH gene in mammalian cells therefore results in detoxification of G418.

Mammalian cells differ in their sensitivity to G418, and moreover, different batches of G418 can have different potencies. Therefore, it was first necessary to select the concentration of G418 in medium that results in complete killing of the wild type HeLa cells. Different concentrations of G418, ranged between 0.1 and 0.8 mg/ml, were tested. At 0.5 mg/ml, all HeLa cells died after the period of time that is necessary for approximately two cell divisions.

To create a cell line, HeLa cells were transfected with pOM6 plasmid by the calcium-phosphate procedure and selection on resistance to G418 was performed as described in the Materials and Methods. In two weeks after transfection 30 clones were selected (approximately 6 clones from each of five 90 mm petri dishes, initially subjected to transfection). After additional cultivation for about two weeks 6 clones, which looked more healthy and faster growing, were analysed by immunostaining with α-RBM. Surprisingly, the results of immunostaining were similar for all the chosen colonies, and only about 10% of the cells on the stained coverslip had bright red staining as in Figures 3.5-3.7. It is difficult to compare the intensity of staining on different coverslips even if they are processed in parallel but the pictures looked very similar to those of the transiently transfected cells. It was assumed that selected clones were not homogeneous and contained mixtures of cells expressing RBM and cells that do not express RBM but are G418 resistant due to the integration of the APH gene only. This was not a plausible explanation but the presence of the RBM-positive cells suggested that it might be possible to select homogeneous colonies.

Therefore, single colonies were transferred to new plates. After one month of maintenance in the presence of 0.5 mg/ml G418, new colonies were assessed for the presence of the RBM protein: the picture stayed the same as before. One more round of selection was used, this time in the presence of 0.7 mg/ml G418. But the result of immunostaining looked again similar to that of the transient transfection despite the fact that these cells had been growing for at least three months in the presence of the antibiotic. At this stage the selection procedure was stopped.

During the selection procedure the chosen clones were regularly analysed for the presence of RBM. Figure 3.8 shows the western blot analysis of three different clones collected after each round of selection. In the whole cell lysate of each clone, the antibody against RBM (Fig.3.8A) recognises a protein with apparent molecular mass of 60 kD which is not present in the wild type HeLa cells. It is tempting to assume that this band corresponds to the RBM protein, but no further evidence supporting this was obtained. Likewise, the explanation why this protein migrates on SDS-PAGE faster than the protein detected in the cells after 48 hours from transient transfection (lane 1 vs lane 3-5) was not found. The relative amount of putative RBM expressed in these cells was normalised using western blotting with anti-bodies against hnRNP A1 (Fig.3.8B). All three clones presented here express similar amounts of the putative RBM, but in other experiments the relative intensity of the RBM band varied for different clones with a different degree of reproducibility. Significantly, in numerous experiments with different clones the expression of RBM in the transiently transfected cells was always higher than in any selected "cell lines". This observation correlates with the immunofluorescence staining as fewer cells were found RBM-positive in the selected clones than in recently transfected cells. Overall, the results were ambiguous and inconclusive and the attempt to select the cell line expressing RBM failed.

3.2.5. Expression of the RBM protein has a profound effect on HeLa and COS cells

During the multiple transfections and prolonged selection of the cell lines it became evident that cells transfected with RBM looked different in comparison to untransfected cells or mock-transfected cells. Several observations suggested that expression of RBM is harmful for HeLa and COS cells. First, the growth of HeLa cells during selection of cell lines was

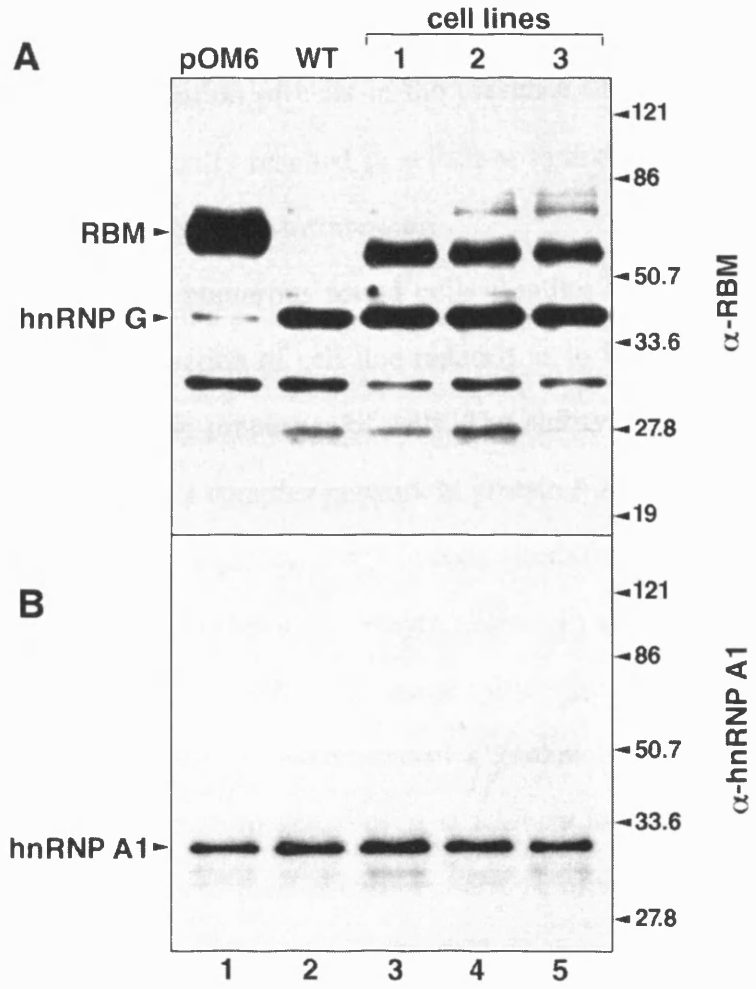


Figure 3.8. Western blot analysis of putative HeLa cell lines expressing RBM. **A:** Expression of RBM in putative HeLa cells lines was analysed by 10% SDS-PAGE followed by western blot analysis with α-RBM. 10 μg of proteins of each the whole cell extract from the cells transiently transfected with pOM6 (lane 1), untransfected cells (lane 2), and three independent clones of putative cell lines (lanes 3-5) were analysed. **B:** Identical western blot was developed with α-hnRNP A1.

affected as the doubling time was increased in comparison with the mock-transfected cells. Second, the death rate of cells transfected with RBM was increased as a higher proportion of cells was floating in the medium than in the cases of mock-transfected cells. A similar effect was observed during propagation of cells in the presence of G418. Thus, it is likely that selection of the cell line actually resulted in selection against RBM and led to creation of neomycin resistant cells without RBM expression.

The presence of the numerous round cells floating in the medium after transient transfections and during selection of cell line induced us to investigate a possible effect of RBM expression on adhesion properties of cells. The ability of eukaryotic cells to adopt a variety of shapes depends on a complex network of protein filaments, named the cytoskeleton that extends throughout the cytoplasm. It was hypothesised that RBM might affect processing of the pre-mRNA encoding a protein of the cytoskeleton in a way that the cells were losing their ability to be attached to the surface of a tissue culture dish.

In the cortex of animal cells, actin molecules continually polymerise and depolymerise to generate cell-surface protrusions and sites of attachment to other cells or solid phase. The cortical actin filaments are organised in different arrays, of which stress fibres are the most prominent components of the cytoskeleton of cells in culture. At one end stress fibers insert into the cytoplasma membrane at special sites named focal contacts, where the external face of the cell is closely attached to the extracellular matrix. At the other end they insert into a second focal contact or into a meshwork of intermediate filaments that surround the nucleus. When cells grow on the bottom of tissue culture dish, most of the cell surface is separated from the solid phase by a gap of more than 50 nm, whereas at focal contacts this gap is reduced to 10-15 nm. Here the cytoplasmic membrane is attached to components of the extracellular matrix that have become absorbed to the tissue culture dish. In a chain of protein-protein interactions between the extracellular matrix and cytoskeleton, vinculin associates with the actin-binding protein, α-actinin, that forms the anchorage for the ends of stress fibres and it is also the main cross-linker of actin filaments in stress fibres.

To assess a possible effect of RBM on the cytoskeleton and particularly, the focal contacts, stress fibres and vinculin were fluorescently stained in the HeLa cells transfected with RBM. The monoclonal antibodies against vinculin (kindly provided by Dr. D. Critchley,

Leicester) in combination with the secondary FITC-conjugated antibodies, produced the typical staining of focal contacts that resemble the shape of a painting-brush (data not shown). Stress fibres were stained with FITC-conjugated phalloidin, a toxin isolated from *Amanita* mushrooms that binds all along the side of actin filaments but does not recognise the monomers (data not shown). Several coverslips containing at least 100 transfected cells were scrutinised and nothing was observed that would indicate any effect of RBM expression on staining of stress fibres and focal contacts.

Instead, a detailed analysis of these coverslips revealed that some of the cells contained the multinucleate structures. The multinucleate formations were present in a small percentage of cells (<10%) expressing RBM but they were not detected in untransfected cells.

The transient transfection of COS-1 cells with pOM6 plasmid resulted even in a higher percentage of transfected cells that contained the multinucleate structures. Moreover, the number of pseudo-nuclei was increased such that 20-30 of them could be seen in one cell. Figure 3.9 shows six different examples of COS-1 cells expressing RBM. The higher number of multinucleate COS cells in comparison with HeLa cells is likely to be due to the presence of SV40 origin of replication in the pcDNA plasmid that could be used for episomal replication in cell lines expressing the large T antigen like COS-1 and COS-7 cells.

To determine whether the multinucleate structures contained DNA, HeLa cells, expressing RBM, were stained with DAPI, a DNA intercalating reagent that is widely used to visualise chromosomal DNA in a fluorescence microscope. The colour images were obtained with a fluorescence microscope equipped with the filter that allows detection of the blue light from DAPI and the red light from Texas Red. It was clear that these membrane-wrapped structures do contain DNA (data not shown).

Apoptosis is a multistep process that leads to activation of a proteolytic cascade, which leads to digestion of structural proteins in the cytoplasm, degradation of chromosomal DNA and, finally, death of the cell (reviewed by Ashkenazi and Dixit, 1999; Li and Yuan, 1999). Apoptotic cells decrease in size and their DNA is degraded at the internucleosomal linker sites yielding small DNA fragments of aproximately 180 base pairs which are membrane-wrapped that makes an elusion of the multinucleate formation. Therefore, it is more likely that multiple nuclei (20-30) in COS cells and several in HeLa (3-4) are formed

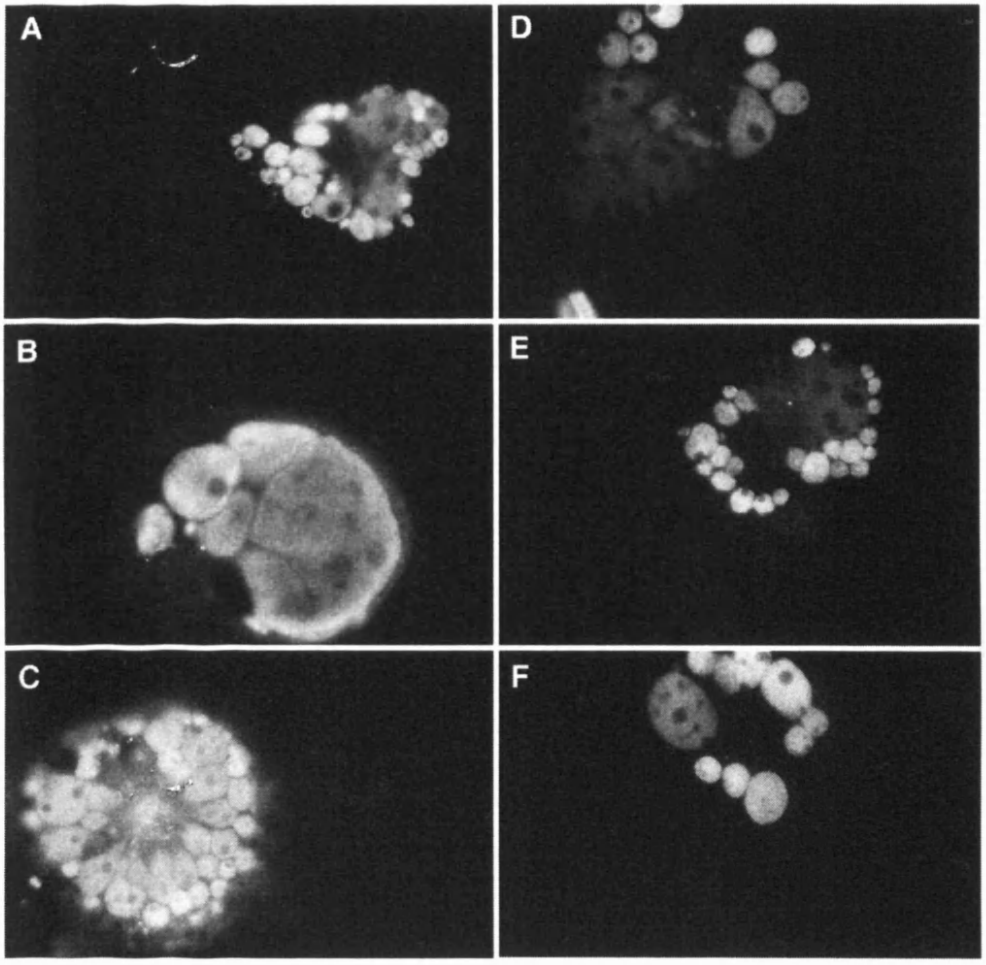


Figure 3.9. Multiple nucleate structures in COS-1 cells transfected with RBM. Cells were transfected with pOM6 and immunostained with a-RBM. Black and white images of selected cells were taken on Zeiss fluorescence microscope with a build-in camera.

due to apoptosis. Probably, cells became round and lost their attachment to solid phase, because stress fibres were disintegrated in apoptotic development.

A variety of physiological death signals, as well as pathological cellular insults, trigger apoptosis. The hypothesis that RBM can induce apoptosis in transcfected cells does not seem to be illogical in the view of possible effects of RBM on alternative splicing. The expression and function of a large number of genes involved in apoptosis are regulated by alternative splicing and the ratio between these isoforms play a pivotal role in deciding whether a cell will live or die (Jiang and Wu, 1999). One of the best documented examples is Bcl-x gene, a member of the Bcl-2 family, which has two alternatively spliced forms, Bcl-xL and Bcl-xS (Minn *et al.*, 1996). Bcl-xL, like Bcl-2, is able to protect cells from a wide variety of apoptotic stimuli. As a result of alternative splicing, Bcl-xS lacks 63 amino acids that comprise the region of greatest amino acid identity between Bcl-xL and Bcl-2. These amino acids contain the highly conserved BH1 and BH2 regions, which have been used to define the Bcl-2 family. In contrast to Bcl-xL, Bcl-xS isoform encodes a protein capable of accelerating cell death. Thus, alternative splicing of Bcl-x pre-mRNA can modulate cell death. The detailed mechanism of this alternative splicing is not known and could be regulated by specific RNA-binding proteins. It is unlikely that RBM is naturally involved in regulation of the Bcl-x pre-mRNA processing because of its tissue specificity, but the overexpression of RBM in transfected cells could perturb normal processing if RBM has an affinity for the particular sites on this pre-mRNA.

Recently the alternative splicing of another gene involved in apoptosis was investigated in more detail (Jiang *et al.*, 1998). Ich-1 encodes a member of the caspase family of proteases, caspase-2. As in the case of Bcl-x gene, two isoforms of Ich-1 generated by alternative splicing produce the opposite effect on apoptosis. Ich-1L, resulting from alternative exclusion of a 61-bp exon, causes apoptosis, and Ich-1S, including this exon, inhibits apoptosis. The authors demonstrated that alternative splicing of Ich-1 pre-mRNA can be regulated by general splicing factors: SF2/ASF, SC35, and hnRNP A1. It was shown that SR-proteins SC35 and ASF/SF2 promote exon skipping, decreasing the ratio of Ich-1S to Ich-1L transcripts, whereas hnRNP A1 facilitates exon inclusion, increasing this ratio. Interestingly, in cultured cells, overexpression of SC35 increases apoptosis whereas

overexpression of hnRNP A1 decreases apoptosis. These results suggest that alternative splicing may be an important regulatory mechanism for apoptosis.

The apoptotic scenario is also could explain the reduced size of the RBM protein in cell lines as following. The expression of RBM in transfected cells triggers apoptosis, and therefore result in increase of proteolytic activity in these cells that digest RBM itself.

In conclusion, if expression of RBM causes apoptosis in transfected cells, a plausible explanation of this phenomenon would be an effect of RBM on alternative splicing of one of the genes involved in apoptotyc pathway.

3.2.6. Existence of RBM protein in nature and its tissue and species specificity

At the beginning of this study the only information about expression of RBM gene was reported at the RNA level (Ma *et al.*, 1993). Thus, one of the initial tasks was to detect the RBM protein. For this purpose, polyclonal antibodies were raised and characterised using transfected cultured cells and these antibodies appeared to be suitable for the detection of RBM expression in naturally occurring tissues. An analysis of RBM expression in different species was carried out using a collection of testis tissue samples from chicken, mouse, dog, man and also aberrantly developed human testicles arising from androgen insensitivity syndrome (AIS) or testicular feminization. The presence of the RBM protein in the tissue was analysed by western blot (Fig. 3.10A). Whole cell lysates from these samples were prepared in the urea-containing buffer as described in the Materials and Methods. Equal amounts of proteins extracted from each tissue were subjected to 10% SDS-PAGE, blotted and immunostained with α-RBM. The lysate of HeLa cells served as a negative control. Figure 3.10A shows an image of the overexposed blot in order to see specific and unspecific staining. Two abundant bands are highlighted in the lane corresponding to the sample of human testis with molecular weights around 60 kD. These two bands seem to be the most specific bands on the entire blot. Taking into account that the theoretical molecular weight of RBM is 61 kD, it is likely that these two bands correspond to the RBM protein or its isoforms. The band of approximately 40 kD which is seen in the samples of dog testis, HeLa cells and AID is likely to correspond to hnRNP G (as discussed above, Fig. 3.3). Since the short exposure shows bands only in the sample of human testis but not in other lanes, it indicates that human RBM is unique as it is restricted to the human tissue and also developmentally

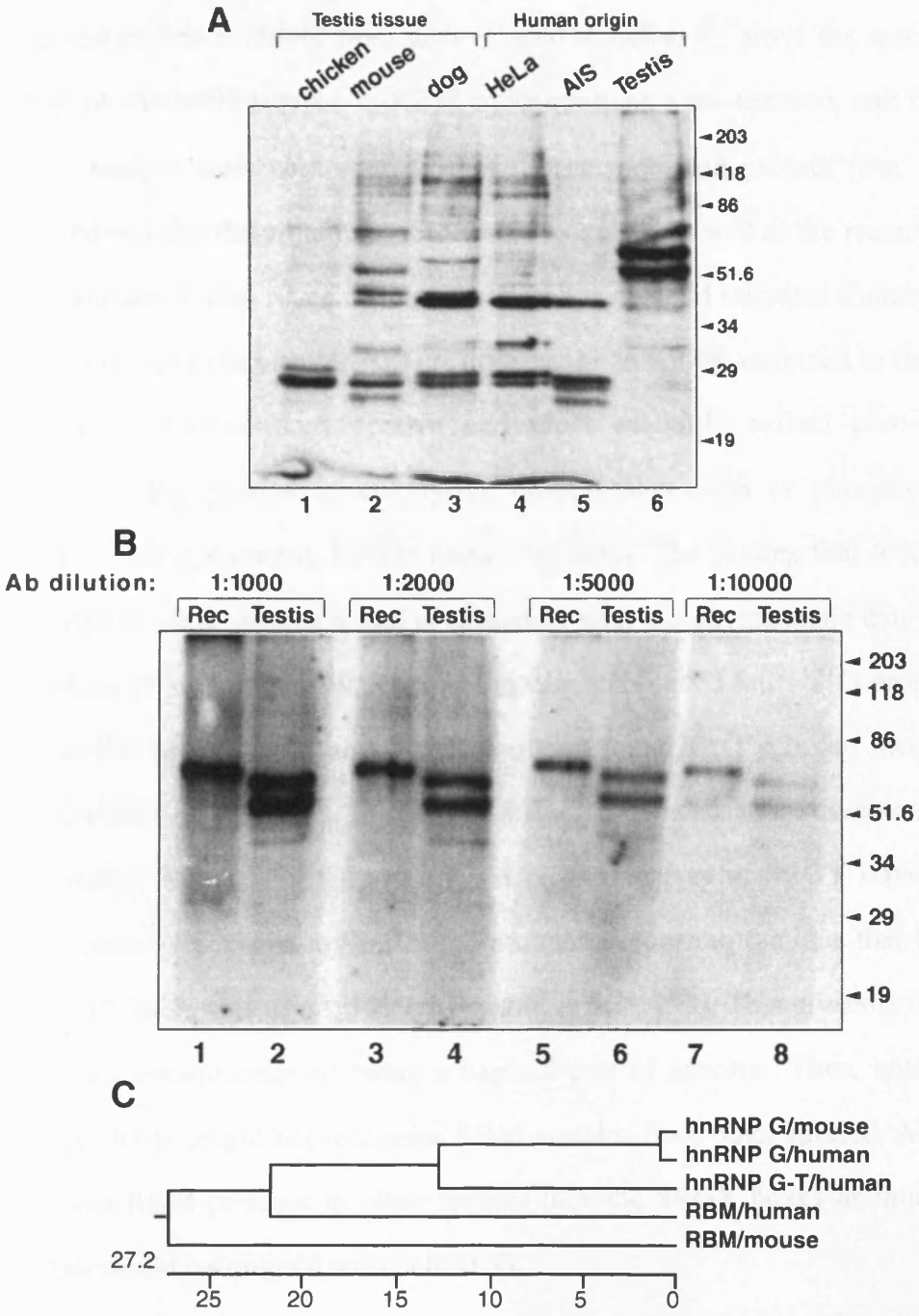


Figure 3.10. Existence of the human RBM protein in nature. A: Western blot analysis of tissue samples indicated in the figure with α-RBM antibodies. 10 μg of proteins extracted from different tissues were subjected to 10% SDS-PAGE, transferred to a membrane and immunostained with α-RBM. **B:** Recognition of RBM by antibodies in human testis was compared with recognition of recombinant RBM used as immunogen. 0.1 μg of recombinant protein (lane 1, 3, 5 and 7) and 5 μg of proteins from human testis (lane 2, 4, 6 and 8) were subjected to 10% SDS-PAGE, blotted and immunostained with α-RBM at different dilutions as indicated in the figure. **C:** Sequence comparison of the human RBM protein and its homologues. The alignment and phylogenetic tree were done with MegAlign program. The accession numbers of the proteins are as follows: hnRNP G/mouse – AJ237847, hnRNP G/human – NM_002139, hnRNP G-T/human – AF069682, RBM/human – NM_005058, RBM/mouse – U36929.

regulated as the protein is absent from undeveloped testicles. To prove the specificity of the signal in the human testis sample, α-RBM concentrations were reduced, and the signals in human testis sample were compared to those of recombinant protein (Fig. 3.10B). The experiment showed that the protein from testis is recognised as well as the recombinant which served as an antigen. It also revealed that the recombinant RBM migrates slightly slower then both protein isoforms in human testis. This difference can not be attributed to the presence of 6xHis-tag in the recombinant protein and most probably reflect post-translational modifications of the protein in eukaryotic environment such as phosphorylation and glycosylation which is common for this class of proteins. The finding that α-RBM does not recognise RBM in other species tested is consistent with the phylogenetic data (Fig. 3.10C). The comparison of amino acid sequences of mouse and human hnRNP G proteins revealed only 4 conserved substitutions and that is common for general splicing factors (*e.g.*, see Chapter 4, U4/U6-61kD protein). In contrast, RBM proteins are more diverse and share less than 40% identity. Interestingly, mouse RBM lacks the array of SRGY boxes and, in this respect, it is more reminiscent of hnRNP G proteins supporting the idea that RBM evolved from hnRNP G (Delbridge *et al.*, 1999; Mazeyrat *et al.*, 1999). This diversity of RBM could be evolutionary consequence of being a haploid part of genome. Thus, antibodies raised against human RBM might not recognise RBM proteins from other species. Alternatively, it is possible that RBM proteins in other species lack the SRGY boxes as mouse RBM and therefore they would co-migrate with hnRNP G.

The two bands observed in human testis suggest the existence of two isoforms of the human RBM protein. These isoforms might be generated either from different copies of the RBM gene (Prosser *et al.*, 1995), or by alternative splicing of the RBM pre-mRNA. Original cloning of the RBM gene resulted in isolation of two cDNAs, MK5 and MK29 (Ma *et al.*, 1993). The MK5 cDNA encoded a longer protein of 496 amino acids, which was designated as the RBM protein. The MK29 cDNA showed a 7 nt substitution and 5 nt deletion relative to the MK5. The deletion causes a translational frameshift that truncates the MK29 open reading frame, giving a predicted protein of 419 amino acids. Thus, the two bands detected on western blot would match two polypeptides of predicted molecular masses of 61kD and 54kD. At the same time, we are not able to rule out the possibility that the smaller protein is a degradation

product of the large and the only form of RBM. Interestingly, the smaller band has mobility similar to the RBM protein expressed in "cell lines".

The existence of two RBM isoforms could have a functional significance as it was shown for the WT1protein (Wilms' Tumor suppressor). WT1 is associated with nephropathy, gonadal and genital abnormalities that cause Denys-Drash (DDS) and Frasier syndromes. In DDS there is a predisposition to Wilm's tumor. The WT1 pre-mRNA is alternatively spliced, and insertion of three amino acids, KTS, plays an important role. Significantly, Frasier syndrome is associated with WT1 mutations at splice-donor sites that prevent generation of +KTS isoforms. The potential roles for WT1 isoforms came from examination of its cellular localisation. Immunostaining of WT1-expressing cells revealed co-localisation of the +KTS isoforms and splicing factors whereas the protein without KTS co-localised with transcription factors (Larsson *et al.*, 1995). Recently, it has been shown that WT1 specifically interacts with U2AF65, in a manner enhanced by KTS-insertion, and moreover, WT1 can be directly incorporated into the spliceosomes (Davies *et al.*, 1998). Despite the absence of experimental data, it is easy to envisage how different RBM isoforms might correspond to populations of protein with different functions.

3.2.7. Expression and isolation of recombinant RBM using the ThioFusion expression system

The availability of soluble recombinant protein is important for most *in vitro* studies, as discussed in the Introduction. For this reason a new attempt to obtain the recombinant protein was undertaken as the previous one, based on the pRSET plasmid, which drives expression from the T7 promoter (Invitrogen), had resulted in purification of a His-tagged RBM protein that was soluble only in the presence of urea or guanidine hydrochloride. The ThioFusion Expression System (Invitrogen) was chosen for the following reasons. This system was just launched on the market, and was the only one commercially available at that time that specifically addressed the problem of solubility of recombinant proteins expressed in *E. coli*. The ThioFusion system offers high-level expression of the heterologous protein as a C-terminal fusion to thioredoxin, a chaperone that facilitates proper folding of other proteins. The thioredoxin moiety appears to confer solubility to formerly insoluble heterologous proteins, and it was the main attractive feature of the system. Another aspect was that I

wanted to avoid using the T7 promoter, which had already been tried without success. The plasmids driving expression from the T7 promoter are commonly used for expression of heterologous proteins in *E. coli* but at the same time, the drawbacks of this system are also known. The T7 RNA polymerase is approximately 8-fold faster than the ribosomes that translate its message, creating a continuously expanding ribosome-free mRNA stretch behind itself (Iost *et al.*, 1992). Naked mRNAs are subjected to nuclease attacts that results in instability of the T7 polymerase transcripts. This leads to general correlation that the yield of full-length protein per transcript from any RNA polymerase, which is faster then the authentic *E. coli* polymerase, inversely depends upon the polymerase speed (Makarova *et al.*, 1995). The accumulation of degraded T7 RNA polymerase transcripts may cause the appearance of a large amount of truncated proteins that can unpredictably affect the expression and folding of the protein of interest. The pTrxFus vector drives expression of thioredoxin fusions from the bacteriophage λ P_L promoter, one of the most efficient promoters for initiation of the *E. coli* RNA polymerase transcripts. Thus, a natural coupling between transcription and translation, the mechanism regulating gene expression in *E. coli*, would be applied to a foreign gene in the ThioFusion system. This P_L promoter is in addition very tightly regulated. The bacteriophage λ *cI* repressor binds to the operator region in front of the P_L promoter and controls the level of transcription from this promoter. Expression of the *cI* repressor is regulated in the strains GI698 or GI724 (designed for pTrxFus expression), where the *cI* repressor gene is located on a chromosome under control of *trp* promoter. When cells are grown in tryptophan-free medium, the *cI* repressor gene is transcribed, and the *cI* repressor protein binds to the P_L promoter preventing transcription. Adding to the medium tryptophan, which shuts down *cI* repressor synthesis, induces expression.

The ThioFusion expression system also offers two easy purification procedures. The fusion of thioredoxin to foreign proteins may confer a unique cellular location to the fused protein allowing easy purification by osmotic shock. Native thioredoxin is localised at particular sites on the cytoplasmic side of the inner membrane known as adhesion zones or Bayer's patches, and is quantitatively released during osmotic shock into the shock fluid. An alternative one-step purification procedure of the thioredoxin fusion proteins is affinity chromatography on ThioBond resin. This resin is specially designed for the affinity

purification of proteins containing vicinal dithiols. If the cysteines are close to each other, either in the primary amino acid sequence or in the tertiary structure, the free thiol groups are said to be vicinal. These thiol groups may bind metal ions such as arsenic. The reaction mechanism of thioredoxin involves a reduction-oxidation event between two vicinal dithiols in the active site, -Cys-Gly-Pro-Cys-. Thus, thioredoxin and thioredoxin fusions contain vicinal dithiols which could reversibly bind the phenylarsine oxide in the ThioBond resin. During procedure other proteins are washed away, and purified thioredoxin fusion protein is eluted from the resin using β-mercaptoethanol (β-ME).

To construct the Trx-RBM plasmid, the coding sequence of RBM was amplified by PCR from cDNA and cloned into KpnI/BamHI sites of the pTrxFus expression vector, resulting in Trx-RBM plasmid. The Trx-RBM plasmid encodes the protein consisting of the 12 kD thioredoxin at the N-terminus fused to the first methionine of the RBM. The absence of mistakes introduced during cloning was confirmed by sequencing. Both strains GI698 and GI724 were examined at two temperatures, 30°C and 37°C, for expression of the fusion protein. The empirically selected procedure was as follows. Cells from a single colony of freshly transformed GI724 cells were grown at 37°C until the mid-log phase, and then induced with tryptophan and transferred to 30°C. To analyse the expression of the Trx-RBM fusion protein, aliquots of the cell culture were taken at different time intervals after induction. Cells were pelleted, lysed by sonication, and the lysates were centrifuged at 13,000 rpm for 5 minutes to remove cell debris and insoluble material. Figure 3.11A shows the SDS-PAGE analysis of the soluble fraction. The band of approximately 75 kD, which is in agreement with the expected molecular weight for fusion protein, appeared and accumulated after induction. An equal distribution of the Trx-RBM fusion protein between the supernatants and pelletted fractions was observed (data not shown). The presence of RBM in supernatant indicated that either the thioredoxin moiety did confer solubility to the formerly insoluble protein or the coupling between transcription and translation was responsible for proper folding of the protein.

Since a reasonable amount of the protein was found in a soluble fraction, the osmotic shock procedure was tried first for the protein purification. The cells were subjected to osmotic shock by transferring them from a high ionic strength buffer to a low ionic strength

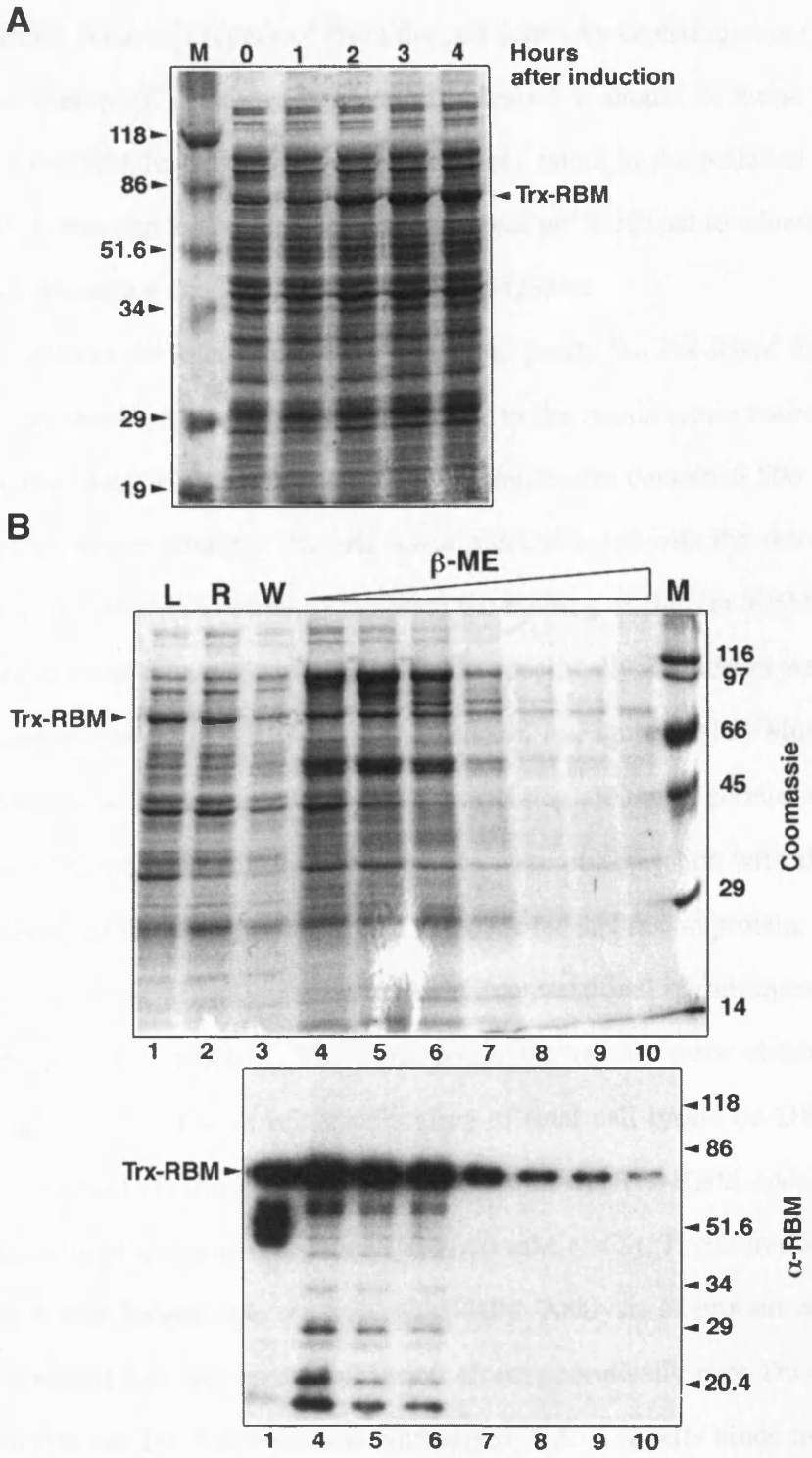


Figure 3.11. Expression and isolation of the recombinant Trx-RBM protein. A: Expression of Trx-RBM in *E. coli*. 10 μg of total cell lysates prepared at different time intervals after induction of the cells with Trp were separated by 10% SDS-PAGE and stained with Coomassie blue. **B:** Purification of Trx-RBM by affinity chromatography on ThioBond resin. Different fractions were analysed by 10% SDS-PAGE, and proteins were visualised by Coomassie staining (upper panel) or western blotting with α-RBM (lower panel). Lane 1 shows the material loaded onto the column, lane 2 is the run through fraction, lane 3 is the material washed from the column with running buffer containing 1 mM of β-ME, lanes 4-10 are fractions eluted with 5, 10, 50, 100, 200, 500 and 1000 mM β-ME, respectively, as described in the Materials and Methods.

buffer, and the shock fluid was separated from the cell debris by centrifugation (described in the Materials and Methods). If the fusion protein is released, it should be found in the shock fluid. Since the Trx-RBM fusion protein was exclusively found in the pelletted cell fraction (data not shown), it was concluded that the Trx-RBM was not localised to adhesion zone and the osmotic shock procedure cannot be used for the purification.

Next, the affinity chromatography was applied to purify the Trx-RBM fusion protein. The purification procedure was carried out according to the manufacture recommendations and described in the Materials and Methods. The running buffer contained 500 mM NaCl to reduce non-specific protein binding. The cell lysate was incubated with the activated resin in a batch with constant rotation in order to facilitate the binding of the Trx-RBM to the resin. After washing the column from proteins that bind non-specifically, the fusion was eluted with increasing concentrations of β-ME. The result is shown in Figure 3.11B. Most of the Trx-RBM fusion protein was found in the flow through fraction although, certain amount of the protein elutes at 5-50 mM β-ME but these fractions were not enriched with the Trx-RBM. This result suggests that the ThioBond resin did not work for this fusion protein.

After the failure of affinity chromatography, conventional chromtographic methods were tried to purify the Trx-RBM. The most promising results were obtained by anion exchange chromatography. The direct fractionation of total cell lysate on DEAE cellulose column (DE52, Whatmann) showed that some amount of the Trx-RBM specifically elutes from the column at high salt concentrations (500-800 mM NaCl). These fractions contained mainly RNA as it was judged spectrophotometerically. Analysis of protein composition of these fractions revealed that they contained almost electrophoretically pure Trx-RBM protein. The observation that the Trx-RBM protein synthesised in *E. coli* cells binds to certain RNA species might be an explanation why the fusion protein was not released by osmotic shock and did not bind to the ThioBond resin. Since the concentration of RNA in the fractions eluted from DE52 column at 500-800 mM NaCl was high, 1-2 mg/ml, whereas the protein concentration estimated by Bradford assay was low, only 0.1 mg/ml, the idea was first to treat the sample with RNases, and then concentrate the protein on a Mono Q or Phenyl Superose columns. However, after the treatment with RNases no protein was eluted from the column (data not shown). The results suggest that as soon as the protein was released from RNA it

became prone to aggregation and was retained on the column filter. In comparison, when an untreated sample was loaded on the same column, some Trx-RBM was eluted from the column but without any enrichment versus RNA.

Thus, the attempt to isolate the recombinant RBM protein from *E. coli* using the ThioFusion expression system failed. It is likely that the main problem was due to the presence of the SRGY domain in the RBM protein as the problem of solubility is common for all SR proteins. In most recent publications recombinant SR proteins are expressed and purified using baculovirus systems (Mayeda *et al.*, 1999). It seems that post-translational modifications of SR proteins, and particularly phosphorylation of the RS domain, are important for keeping the protein in a functionally active conformation. In agreement with our observation that the non-phosphorylated RBM protein avidly binds RNA, it was recently shown that the phosphorylation of the RS domain of SRp40 and SF2/ASF decreases non-specific RNA-binding (Tacke and Manley, 1997; Xiao and Manley, 1997).

The absence of recombinant RBM made the *in vitro* study of this protein more difficult. Apparently, the protein has to be synthesised in eukaryotic cells or in a eukaryotic *in vitro* translation system. Since RBM was expressed in HeLa cells transiently transfected with a pOM6 plasmid, extracts from these cells were tried as a source of a functionally active RBM protein.

3.2.8. Immunoprecipitation of RBM protein from transiently transfected HeLa cells

A function of a protein might be deduced if the roles of the components with which it interacts are known. To identify interacting partners for RBM, the method of immunoprecipitation from transiently transfected HeLa cells was exploited. The idea was to immunoprecipitate the complex of RBM with other proteins using α-RBM antibodies immobilised on beads, and then identify interacting partners either by peptide microsequencing if the unique bands are visible on the stained gel or using antibodies against splicing components.

At the beginning it was necessary to investigate different procedures of extract preparation from HeLa cells transfected with RBM that would allow precipitation of RBM. The method tried first was a small-scale procedure for nuclear extract preparation, described

by Green and co-workers (Lee *et al.*, 1988). This procedure is based on the same principles and buffers which were established by Dignam *et al.* (1983) for preparation of extracts from a large quantity of cells, but it was applied to a small number of cells that could be scraped off the bottom of standard 90 mm tissue culture petri dish. Hence, HeLa cells were transiently transfected with pOM6 and the nuclear extract was prepared as described in the Materials and Methods. The western blot analysis of different fractions revealed that the RBM protein was completely removed from the extract during the last centrifugation for 5 min at 13,000 rpm after an addition of buffer C (data not shown). It indicated that either RBM aggregated at approximately 250 mM NaCl or stayed bound to chromatin and was not extracted with buffer C. Therefore, the splicing nuclear extract is not suitable for immunoprecipitation of the RBM. It was reasonable to try alternative methods, which utilise the extraction in the presence of high salt.

Next, the preparation of whole cell extract was tried in accordance to the method of Jiang and Eberhardt (1995). This rapid, one-step procedure was developed for detection and quantitative analysis of DNA binding activities from small number of cells, and it appeared to be ideal for the evaluation of up- and down-modulation of transcription factors in transfected cells. Jiang and Eberhardt compared the extracts prepared by their procedure and the Dignam (Dignam *et al.*, 1983) or Manley (Manley *et al.*, 1980) procedures, and found using gel shift analysis and footprinting that their method improved the yield of DNA-binding activity and provided a more consistent extraction of activities when quantitative comparison is desired.

Whole cell extracts from HeLa cells transfected with pOM6 plasmid or with parental vector pcDNA 3 (mock) were prepared as described in the Materials and Methods. The cells were resuspended in the high salt buffer containing 0.5 M KCl, and subjected to one freeze-thaw cycle. The extract was transparent and there were no visible pellet obtained after centrifugation. When the whole cell extract was diluted to reduce the salt concentration to that of Dignam buffer D (100 mM KCl), the solution became cloudy, and the precipitate, which is considered to be the chromatin fraction, was removed by centrifugation. The supernatant and the whole cell extract were used for immunoprecipitation with α-RBM antibodies as described in the Materials and Methods. It is important to stress that for these experiments α-RBM antibodies were covalently coupled to the beads because if the protein A Sepharose is

used the antibodies would be also eluted from the beads and produce a smear on the stained gel or western blot. Therefore, protein A-purified α-RBM antibodies were covalently coupled to Affi-Gel 10 (Bio-Rad) or Affi-Prep Hz Hydrazide support (Bio-Rad) according to the manufacturer's recommendations (described in the Materials and Methods).

Figure 3.12A shows that RBM protein could be immunoprecipitated from the whole cell extract, but that centrifugation of diluted extract removes RBM from the solution. Thus, it indicates again that either the RBM is not soluble at low salt or it is bound to the chromatin. In the experiment depicted in Figure 3.12A, immunoprecipitates were eluted from Affi-Gel 10 beads, charged with α-RBM, using SDS-PAGE loading buffer containing DTT, and, therefore, non-crosslinked heavy and light chains of immunoglobulins were also eluted. The eluted material was also analysed by SDS-PAGE and visualised by coomassie staining (Fig. 3.12B). There were no abundant bands detected. The ladder of very faint bands was seen in immunoprecipitates from both the RBM transfected and mock transfected cells, but none of these bands was significantly more abundant in the RBM-containing samples. Even RBM itself was immunoprecipitated in an amount sufficient only for detection by western blot and not by Coomassie or silver staining. Thus, the material eluted from the beads was analysed by western blot using a collection of antibodies available in the laboratory. In total, antibodies against twelve antigens were used: hnRNP A1, hnRNP C, hnRNP I, hnRNP (K+J), hnRNP (F+H), hnRNP G, Sm proteins (KSm2 and KSm5), U1-70kD, U1-A, SR-proteins, and SF2/ASF (the source of antibodies is described in the Materials and Methods). None of these antibodies produced signals above background (mock transfected cells) (data not shown). Therefore, if RBM was precipitated specifically, then none of the proteins listed above are the interacting partners, although, the possibility that RBM aggregates and precipitates independent of antibodies was not rule out.

Several attempts were made to make a nuclear extract from surgically obtained human testis. It was observed that as soon as non-denaturing buffers without urea or guanidine hydrochloride, were used to extract RBM from the tissue, the protein degraded, and therefore, immunoprecipitation experiments were not possible (data not shown).

Based on the results of immunoprecipitations described above, it is not apparent whether RBM is free from nucleic acids or stays bound to RNA or DNA molecules after

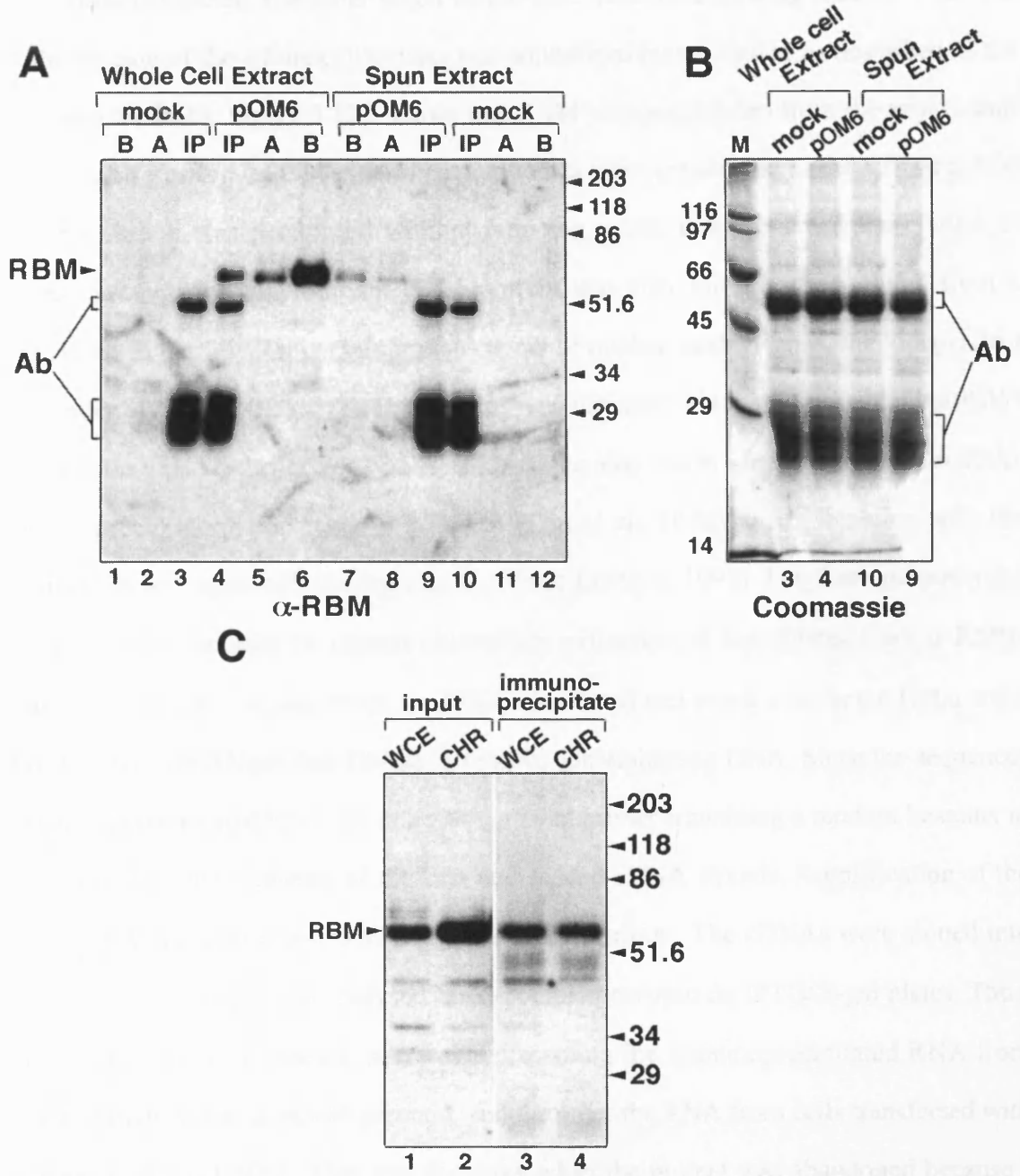


Figure 3.12. Immunoprecipitation of RBM from extracts of transiently transfected HeLa cells. HeLa cells were transfected with pOM6 or with pcDNA 3 (mock). Extracts were prepared and immunoprecipitation was carried out as described in the Materials and Methods. **A:** Aliquots from the extracts before (lane 1, 6, 7 and 12) or after (lane 2, 5, 8 and 11) immunoprecipitation and the material eluted from the beads (lane 3, 4, 9 and 10) were subjected to 10% SDS-PAGE, blotted and immunostained with α-RBM. **B:** Proteins eluted from the beads were analysed by 10% SDS-PAGE and visualised by Coomassie staining. **C:** Immunoprecipitation from the chromatin fraction.

extraction from the nuclei. The latter might be the case from the following reasons. First, the chromatin fraction of the whole cell extract was solubilised by sonication as described in the Materials and Methods. Figure 3.12C shows that RBM was precipitated from the supernatant of the chromatin fraction. In this experiment antibodies were crosslinked to the Affi-Prep Hz-support and elution was performed without reducing agent, this prevented the elution of antibodies. Second, the recombinant RBM protein was also soluble when eluted from a DEAE column in the fraction containing the excess of nucleic acids. Hence, the independent project to identify the RNA target for the RBM was initiated. Since it is possible that RBM stays soluble only in a complex with RNA or DNA, the idea was to identify any specific RNA via immunoprecipitation and random RT-PCR (Chu *et al.*, 1996), in combination with the "subtractive library" approach (Lisitsyn et al., 1993; Lisitsyn, 1995). For these purposes the nucleic acids were isolated by phenol-chloroform extraction of the eluates from α-RMB-beads after immunoprecipitaion from the RBM transfected and mock transfected HeLa cells, and then treated with RNase-free DNase to remove contaminating DNA. Since the sequences of the immunoprecipitated RNA are unknown, a 26 nt primer containing a random hexamer at its 3' end was used for synthesis of the first and second cDNA strands. Amplification of the resulting cDNA was performed with a 20 nt universal primer. The cDNAs were cloned into pBluescript KS(+) plasmid and analysed for α-complementation on IPTG/X-gal plates. Thus, two cDNA libraries were created, one was representing the immunoprecipitated RNA from HeLa cells transfected with pOM6 plasmid, and the other the RNA from cells transfected with the parental vector pcDNA3. That was the stage when the project was abandoned because I left Leicester. The further strategy would have been as follows. To identify the RNA sequences specifically immunoprecipitated with the RBM, the mock transfected library should be subtracted from the RBM transfected library. In simple terms, colony hybridisation could be applied. When the RBM-transfected library is plated on IPTG/X-gal, those white colonies, which did not hybridise with the radioactive probe raised by PCR amplification from the mock-transfected library could be considered as clones representing the RNA targets for the RBM. If the RNA sequences identified in such screen did correspond to pre-mRNA of known genes, it would be possible to design a model pre-mRNA substrate to investigate a potential role of RBM in regulation of splicing of the potential target pre-mRNA.

3.2.9. Study of RBM using the *in vitro* translated protein

Since the attempts to isolate recombinant RBM protein in a soluble form failed, the only source of protein suitable for functional assays was *in vitro* translation. The RBM protein was *in vitro* translated in the coupled transcription and translation reactions from pOM6 plasmid using TNT (T7) rabbit reticulocyte lysate system (Promega). The [^{35}S]methionine-labelled *in vitro* translated RBM protein produces one band on the SDS-PAGE (Fig. 3.13A) which is more diffuse than two other *in vitro* translated proteins used in experiments described in Chapter 4. The diffuseness is a characteristic feature of proteins that are post-translationally modified, and it indicates that the protein is phosphorylated. Figure 3.13B shows that *in vitro* translated RBM protein can be efficiently precipitated in a wide range of salt concentration with α-RBM antibodies immobilised on Protein-A Sepharose. This demonstrates that the RBM protein synthesised *in vitro* in eukaryotic system is soluble and could be used for functional assays. It is also suggests that the post-translational modifications (probably phosphorylation due to the presence of RS dipeptides) are necessary for protein solubility.

If RBM is involved in regulation of splicing the effect might be achieved not only through direct binding to the specific pre-mRNA but also due to the interactions with components of the spliceosome. In this work, it was possible to investigate whether the RBM protein interacts with the snRNPs or not. For this purpose, a "pull-down" assay was used. This method utilises the ability of the monoclonal Y12 antibody directed against Sm proteins to immunoprecipitate any snRNP complexes containing the Sm-core (Teigelkamp *et al.*, 1998). Thus, an interaction of the *in vitro*-translated [^{35}S]methionine-labelled RBM protein with snRNPs was monitored by co-immunoprecipitation with Y12 antibody (the snRNP particles used in this experiment are described in Chapter 4). Figure 3.14A shows that in the absence of snRNP particles labelled protein did not precipitate much with Y12 antibody (lane 8), and only weak signals were detected in the presence of the core U1 and U5 snRNPs, containing only the Sm-proteins and snRNAs (lane 3 and 5, respectively). When labelled RBM protein was pre-incubated with the intact snRNPs, abundant signals corresponding to [^{35}S]-RBM were detected. It indicates that the *in vitro* translated RBM binds to the U1, U2, U5 snRNPs

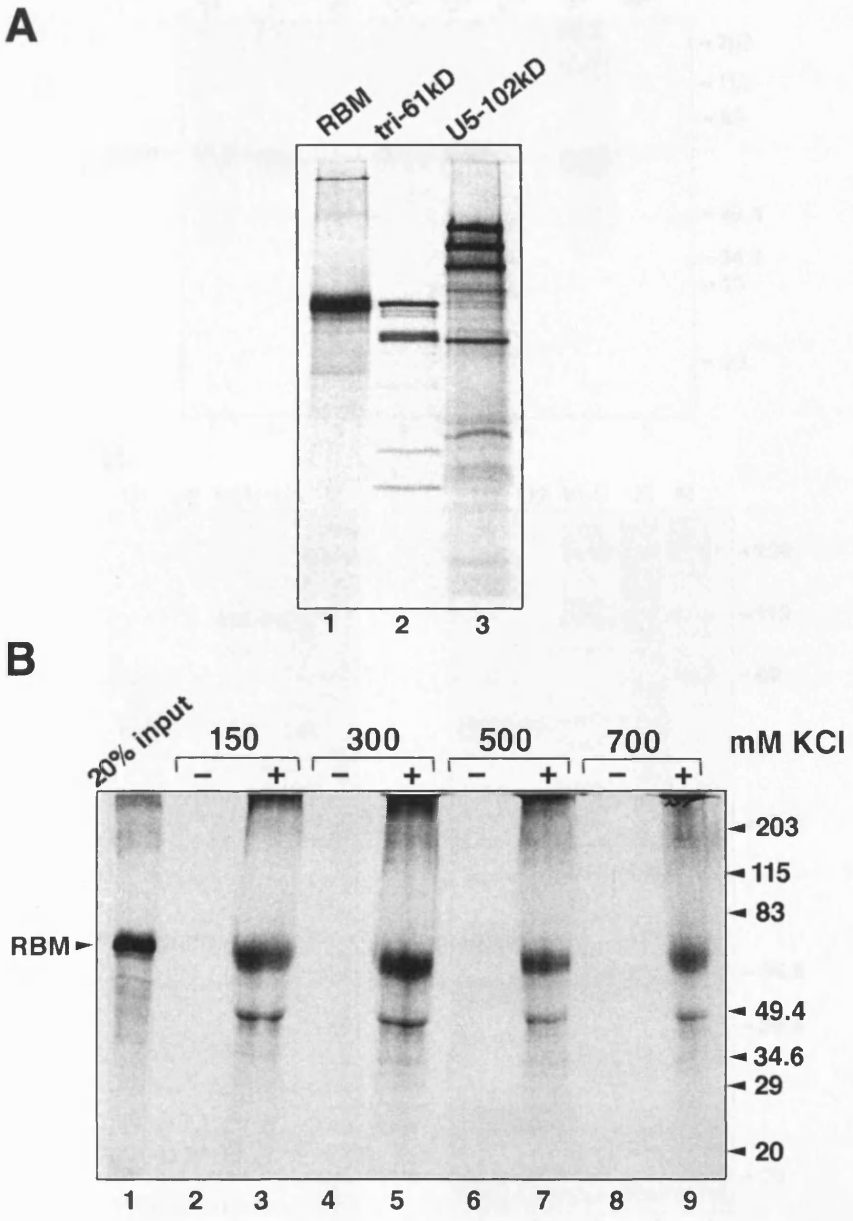


Figure 3.13. Characterisation of the *in vitro*-translated RBM protein. A: Comparison of the *in vitro*-translated RBM protein and proteins of [U4/U6.U5] tri-snRNP characterised in Chapter 4. 1 µl of each *in vitro*-translation reaction (TNT, Promega) was subjected to 10% SDS-PAGE and the [^{35}S]-labelled proteins were visualised by fluorography. **B:** Immunoprecipitation of the *in vitro*-translated RBM with α-RBM. The *in vitro*-translated protein was incubated with Protein A Sepharose (-) or with Protein A Sepharose charged with α-RBM (+) at different salt concentrations as indicated in the figure. The material eluted from the beads was analysed by 10% SDS-PAGE followed by fluorography. The position of pre-stained molecular weight markers (Bio-Rad) is indicated on the right.

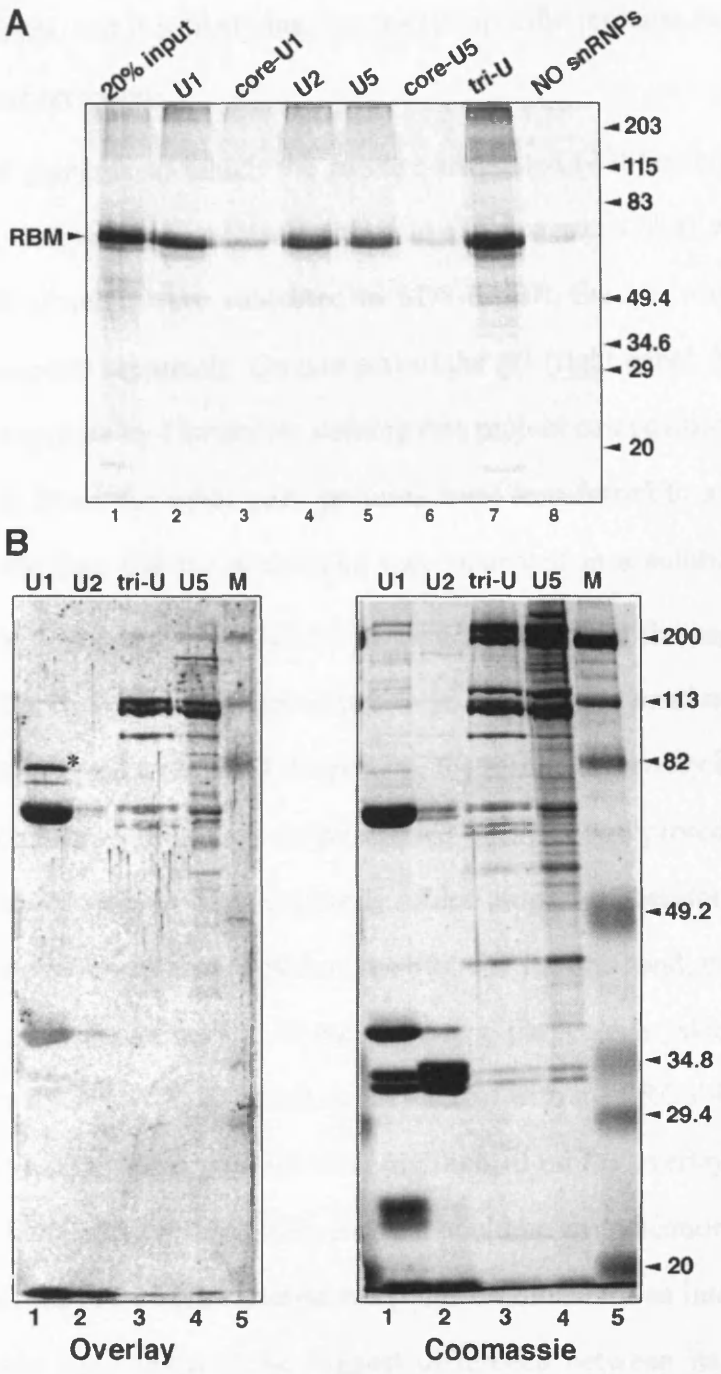


Figure 3.14. Interaction of the *in vitro* -translated RBM protein with snRNPs. **A:** Co-precipitation of the [^{35}S]-labelled RBM protein with snRNPs. RBM was incubated with individual snRNP particles (U1 snRNPs in lane 2, U1 core snRNPs in lane 3, U2 snRNPs in lane 4, U5 snRNPs in lane 5, U5 core snRNPs in lane 6, [U4/U6.U5] tri-snRNPs in lane 7) or without (lane 8). Complex formation was analysed by co-immunoprecipitation with Y12 antibodies. The precipitated material was analysed by 10% SDS-PAGE and visualised by fluorography. In lane 1, 20% of input material for each immunoprecipitaton reaction is shown. **B:** snRNPs (designated as in **A**) were subjected to 10% SDS-PAGE in duplicates. One part of the gel was stained with Coomassie, and the other was transferred to a Hybond P membrane. The membrane was incubated with [^{35}S]-labelled RBM protein, and the proteins were visualised by fluorography.

and [U4/U6.U5] tri-snRNPs, and it is likely that the snRNP-specific proteins, rather than Sm-proteins, are involved in interactions.

To visualise the proteins to which the *in vitro*-translated [^{35}S]methionine-labelled RBM protein can bind, an overlay blot (also referred to as far-western blot) was employed. Samples of the snRNP proteins were subjected to SDS-PAGE, the gel was cut and two identical parts were processed separately. On one part of the gel (right panel, Fig.3.14B), the snRNP proteins were visualised by Coomassie staining (the protein composition of snRNPs is reviewed in Table 1.1). From the other part, proteins were transferred to a nitrocellulose membrane by electro-blotting and the membrane was incubated in a solution, containing [^{35}S]labelled-RBM. The proteins that absorb RBM were visualised by fluorography. Figure 3.14B shows that U1-70kD, and two U5-proteins around 100kD, and band marked with an asterisk were selectively stained with RBM. In general, the results of overlay blots cannot be considered as the conclusive evidence for an interaction between two proteins because the proteins transferred to membrane are not necessarily folded properly. Therefore, it is difficult to judge how specific the interactions detected on the blot. On the one hand, binding of RBM to U1-70kD, which is an abundant protein, is not surprising, particularly taking into account the presence of RS domain in U1-70kD which could interact with the SRGY-boxes of RBM. On the other hand, not all abundant proteins were highlighted on the overlay blot, amongst them B/B', U1/C, U2-A', U5-220 and U5-200, and that could be an indication of specificity. In this respect, the band marked with an asterisk is a prime candidate for an interacting partner with RBM because this band showed the biggest difference between its intensities on autoradiographic film and the Coomassie stained gel. Further experiments to identify this protein were not carried out, but it is likely to be the 69 kD protein, weakly associated with the snRNPs (Hackl *et al.*, 1994). It was shown that the 69 kD protein interacts with Sm-core of all snRNPs, either under *in vitro* reconstitution conditions or upon injection of the snRNAs into the cytoplasm of *Xenopus* oocytes. However, when snRNPs are isolated from HeLa nuclear extract, the only significant proportion of the 69 kD protein is found in association with U1 snRNPs, but little or none with U2 snRNPs or [U4/U6.U5] tri-snRNPs. Sequence analysis of the 69 kD protein revealed: one canonical RNP-type RNA-binding domain, three RGG-rich RNA-binding motifs, and one zinc finger domain, a characteristic of many DNA-

binding proteins (Hackle and Lührmann, 1996). Recently, Wu and Green (1997), using site-specific photo-crosslinking, identified a direct contact between the 69 kD protein and the conserved AG dinucleotide at the 3' splice site. In conclusion, the experiment depicted in Figure 3.14 shows that RBM might interact with components of splicing machinery and therefore, supports a hypothesis of direct involvement of RBM in splicing.

The availability of [^{35}S]methionine-labelled *in vitro*-translated RBM protein allowed to investigate its effect on splicing *in vitro*. Figure 3.15 shows the time course of *in vitro* splicing in the absence or presence of RMB. As much as 25% of an incubation mixture (35% of which is HeLa nuclear extract) could be given to rabbit reticulocyte lysate without significant inhibition of splicing. The experiment did not reveal any effect of RBM on splicing of this particular pre-mRNA. The utilisation of a cryptic site on MINX pre-mRNA (the product is marked with asterisk) is not due to the presence of RBM as the same site is activated with mock rabbit reticulocyte lysate. If RBM has, in principle, an ability to affect splicing *in vitro*, it is likely that a specific substrate or much higher concentrations of the protein would be needed to reveal an effect.

3.3. CONCLUSIONS

The RBM genes were identified on the basis of genetic analysis of infertile males. It is evident that the RBM protein plays an important role in spermatogenesis, but its function is still unknown. The protein acquires its name from its primary sequence, the presence of canonical RNP-type RNA Binding Domain. RBM is also characterised by the presence of SRGY-box sequence elements, which contain SR/RS dipeptides. Both distinctive features are the main attribute of the SR protein family of splicing factors, and therefore, an important indication that such a protein is involved in RNA processing.

The experiments described in this Chapter represent the very beginning of an intensive study of the RBM genes in which several laboratories are engaged now. The idea that RBM was involved in splicing could have been eliminated at the early stages of investigation if the transfection of HeLa cells with the plasmid encoding RBM had resulted in accumulation of the protein in the cytosol. However, both biochemical and immunocytological approaches demonstrated the nuclear localisation of RBM, and this encouraged the further studies with an

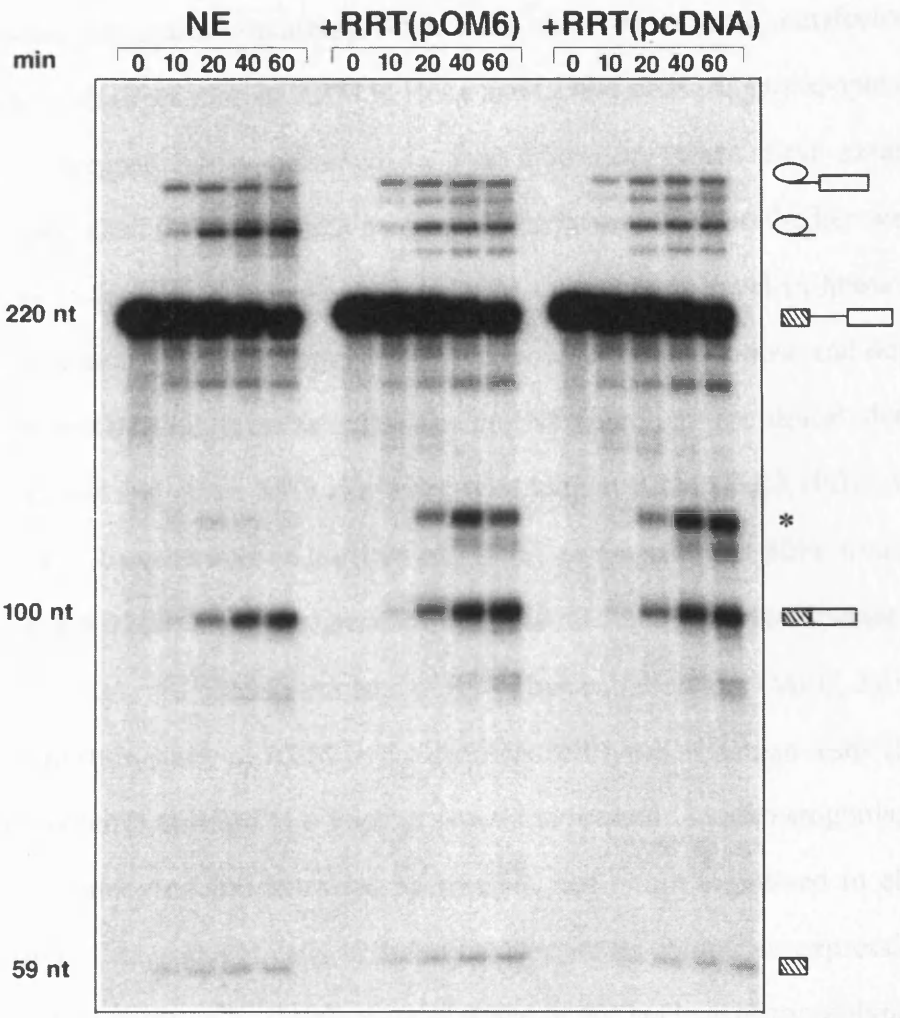


Figure 3.15. The *in vitro*-translated RBM protein does not affect splicing *in vitro*. Time course of MINX pre-mRNA splicing in HeLa nuclear extract (NE), and in the presence of rabbit reticulocyte translation mixture (RRT) containing RBM (from pOM6) or without RBM (pcDNA 3). The splicing reaction contained 35% of HeLa nuclear extract and 25% of *in vitro* translation mixture. Products of reaction were analysed by 14% urea-PAGE and visualised by autoradiography.

emphasis on RNA processing. The expression of RBM appeared to be harmful for HeLa cells, and that explains the failure to select a stable cell line expressing RBM. Moreover, the unusual multinucleate structures containing DNA were observed in some transfected cells, and it could indicate that expression of RBM in HeLa and COS-1 cells triggers apoptosis.

Antibodies against RBM obtained in our laboratory were first extensively characterised and then used for histological analysis. In this work, these antibodies were used in western blotting to identify the expression of RBM on a protein level in human testis. Interestingly, the antibodies do not recognise RBM homologues in the mouse and dog testis, although react with hnRNP G in these samples. The hnRNP G proteins are almost identical in both mouse and human and share 80% similarity with human RBM (Fig.3.10B). A mouse RBM homologue has been cloned (Elliott *et al.*, 1996) and it exhibits 60% similarity to human protein, but α-RBM seems to be specific enough to distinguish the two species.

These antibodies were used in the laboratory of our collaborators (MRC, Edinburgh) for an immunolocalisation study of RBM in the different cell types of human testis (Elliott *et al.*, 1997). It was shown that RBM is a nuclear protein expressed in spermatogonia, in both early and late spermatocytes and in round spermatids, but is not expressed in elongated spermatids or in any of the somatic cells in the testis. Despite its ubiquitous expression in all transcriptionally active germ cells, distribution of RBM in the nucleus changes dynamically over the course of germ cell development. In spermatogonia and spermatocytes, RBM associates with regions that do not look similar to nuclear speckles where the splicing factors are localised. Indeed, it was shown that splicing factors are also targeted to the same large structures as RBM during early stages of germ cell development, but in round spermatids RBM became more diffuse. Neither the transcription factor Sp1 nor hnRNP proteins were targeted to the same nuclear domains enriched with splicing factors and RBM. Co-localisation of RBM with splicing factors supports the idea that RBM is actively involved in splicing at early stages of germ cells development.

The attempts to identify interacting partners for RBM using immunoprecipitation from extracts prepared from transiently transfected HeLa cells have failed. There are several possible explanations for the negative outcome. In general, the low efficiency of transfection could produce insufficient RBM protein for immunoprecipitaion. It is not clear how the

procedure should be scaled-up, if, for one immunoprecipitation described in this chapter, 10 standard 90 mm tissue culture petri dishes were taken for transfection. Another aspect that could affect immunoprecipitations is that RBM is prone to aggregation at a low salt concentration and might be soluble only as a complex with RNA or DNA. If the salt concentration is increased, as in the whole cell extract prepared by the method of Jiang and Eberhardt (1995), it might be that protein-protein interactions involving RBM are already destroyed. In this respect it could be more promising to concentrate efforts on identification of an RNA target for RBM using the "immunoprecipitation-RNA-random PCR" method (Chu *et al.*, 1996).

The experiments devoted to a possible interaction between RBM and snRNPs produced interesting results, although they are only preliminary. The 69kD protein, which transiently associates with all snRNPs, could be considered as a "prime suspect" for an interaction. Unfortunately, a precise role of this protein in splicing is not known, and therefore, this interaction gives no clue for the RBM function. The interactions with U1-70kD and U5-proteins (around 100 kD) envisage the possibility that RBM might influence spliceosome assembly at specific sites on a pre-mRNA.

An independent project was started in Dr. Eperon's laboratory to identify the proteins that interact with RBM. It was based on an alternative technique, the yeast two-hybrid screen, and this approach turned out to be more fruitful then immunoprecipitations. J. Venables used RBM as a bait to fish in a cDNA library of human testis, and found several proteins which are potential candidates for RBM-interacting partners. The most frequent isolate encoded a novel protein with KH-type RNA-binding domain, named T-STAR, that is closely related to Sam68, a Src-associated protein of unknown function (Venables *et al.*, 1999). The interaction between T-STAR and RBM was confirmed by *in vitro* pull down experiments. Interestingly, both proteins are expressed in spermatocytes and therefore, this interaction could have a place *in vivo* in human testis. Since very little is know about functions of the STAR family of proteins, there is no evidence at present that the RBM-STAR interaction can modulate any specific event in RNA processing.

Another putative interacting partner for RBM isolated in the yeast two-hybrid screen was Tra2β, a human homologue of the Drosophila splicing factor Tra2 (Venables *et al.*,

submitted). In *Drosophila*, Tra2 is an important player in a sex determination cascade. Tra2 and Tra, both SR proteins, interact co-operatively with a specific ESE, recruit other SR proteins and finally, activate the usage of a female-specific exon (discussed in Introduction, 3.1). Recently, it was shown that the human Tra2α and Tra2β are able to increase the efficiency of splicing *in vitro* by direct binding to the model exon enhancer (Tacke *et al.*, 1998), but in vivo targets for human Tra2 proteins are unknown and therefore, the function of human Tra2 is obscure. Nevertheless, the interaction of RBM with Tra2β is a first undoubted support for a possible role of RBM in regulation of splicing. Interestingly, it was shown that SRGY domain inhibit Tra2β-associated splicing (Venables *et al.*, submitted). Thus, one can envisage a hypothetical cascade in which at a certain stage of spermatogenesis (presumably, in spermatocytes) RBM and Tra2β interact co-operatively with the specific ESE, alter the splicing pattern and produce a specific protein isoform that is necessary for next stages of germ cell development.

Interactions between RBM and Tra2β were comprehensively studied by *in vitro* experiments (Venables *et al.*, submitted). Interestingly, in the GST-pull down assay, where the C-terminal portion of one protein is fused to glutathione-S-transferase and the partner is an *in vitro*-translated [^{35}S]-labelled protein, the level of binding was significantly increased in the presence of RNases. It indicates that RNA-binding of one of these proteins interferes with protein-protein interaction, and it could be a reason for failure to identify interacting partners in immunoprecipitation experiments.

Currently, the tissue-specific or developmental regulation of gene expression at the level of alternative splicing is a matter of intensive investigations. The involvement of general splicing factors in the modulation of alternative splicing in vertebrates have been documented, but surprisingly, no tissue specific proteins have been identified in mammalian systems whose effect on alternative splicing is comprehensively proven (as in *Drosophila*). Thus, RBM remains a good contender for a protein in mammals whose function is the developmental regulation of alternative splicing.

CHAPTER 4

Investigation of the organization of [U4/U6.U5] tri-snRNP

4.1. INTRODUCTION

Pre-mRNA splicing occurs via two sequential transesterification reactions in a large ribonucleoprotein complex, the spliceosome, which assembles on pre-mRNA by an ordered recruitment of U1, U2, U5, and U4/U6 small nuclear ribonucleoprotein particles (snRNPs) accompanied by other non-snRNP proteins (discussed in Chapter 1). The snRNPs consist of specific snRNAs, which are responsible for the alignment of pre-mRNA splice sites at the catalytic centre of the spliceosome, and a certain number of proteins. During the last decade substantial progress has been achieved in the characterisation of individual proteins involved in splicing. Over 40 different gene products named PRP, for pre-mRNA processing, have been genetically identified in yeast *Saccharomyces cerevisae.* Originally, the first PRP genes (PRP2-PRP11) were identified from temperature-sensitive mutant strains in which, at non-permissive temperatures, pre-mRNA from intron-containing genes accumulated (reviewed by Beggs, 1995). Additional PRP genes were identified through genetic interactions with the first PRP genes. There are now several laboratories involved in screening of large collections of temperature-sensitive yeast strains obtained by mutagenesis and they continue to report new genes required for pre-mRNA splicing in yeast (discussed by Maddock *et al.*, 1996).

In contrast to the fruitful genetic approach, a biochemical characterisation of stable components of snRNPs from yeast is rather difficult due to their low abundance. Higher eukaryotes contain about 3 to 4 orders of magnitude more copies of snRNAs per cell than yeast, and, therefore, HeLa cells became a primary source for isolation of the stable snRNP complexes. A breakthrough in purification of snRNPs was achieved with the introduction of immunoaffinity chromatography using the monoclonal antibody H20, directed against the unique snRNA cap structure (m_3G) (Bochnig *et al.*, 1987; Bach *et al.*, 1990). The total snRNPs pulled out of a HeLa nuclear extract by anti-m3G chromatography could be further fractionated into 12S U1/U2, 20S U5 and 25S [U4/U6.U5] snRNPs by glycerol gradient centrifugation under mild conditions (Laggerbauer *et al.*, 1996). Alternatively, the

individual 12S U1, 12S U2, and 20S U5 snRNPs could be isolated under more stringent conditions by MonoQ anion exchange chromatography (Will *et al.*, 1994). The purification procedures of snRNPs from HeLa nuclear extracts have been scaled up to such an extent that specific proteins are represented as individual bands on a Coomassie-stained gel in amounts sufficient for peptide sequencing. The partial amino acid sequences obtained by microsequencing are valuable information for identification of cDNA encoding the protein.

Based on this approach, individual proteins have been assigned to particular snRNP particles and cDNAs encoding many of the proteins have been cloned. It is always attractive to discover homology between the proteins isolated as components of human snRNPs and the yeast splicing factors because this knowledge complements the absence of *in vivo* approaches in HeLa cells to prove a functional significance of the particular protein in splicing. Remarkably, a list of human homologues to the yeast splicing factors is growing, and the tri-snRNP proteins are the recent contributors.

[U4/U6.U5] tri-snRNP

The [U4/U6.U5] tri-snRNP particle has a most complex protein composition of all snRNPs. Figure 4.1 shows a Coomassie-stained gel of the 25S particle isolated from a HeLa nuclear extract; the protein composition is schematically summarised in the Table 4.1. In addition to one set of U5 canonical Sm proteins and one set of U4 Sm proteins, this particle contains seven recently identified LSm, Sm-like, proteins (Mayes *et al.*, 1999; Achsel *et al.*, 1999) as well as fifteen specific proteins. The latter are named according to their putative molecular weight determined by their migration on a gel (*e.g.*, U5-220kD). Among these specific proteins, those with molecular masses of 220, 200, 116, 102, 100, 40, and 15 kDa were assigned as U5 proteins because they co-fractionate with the 20S U5 particles. In contrast, the association of 90, 60, 20, and 15.5 kDa proteins with U4/U6 snRNAs was demonstrated primarily by immunoprecipitation. The 90kD and 60kD proteins, and cyclophilin, the 20kD protein, form a very stable heteromer (Horowitz *et al.*, 1997; Teigelkamp *et al.*, 1998). The 15.5kD protein directly binds to the 5' stem loop of U4 snRNA (Nottrott *et al.*, 1999). Some of the proteins, namely 110, 65, 61, and 27 kDa, have not yet been characterised in detail and therefore their association behaviour is less clear. For most of the identified proteins, putative

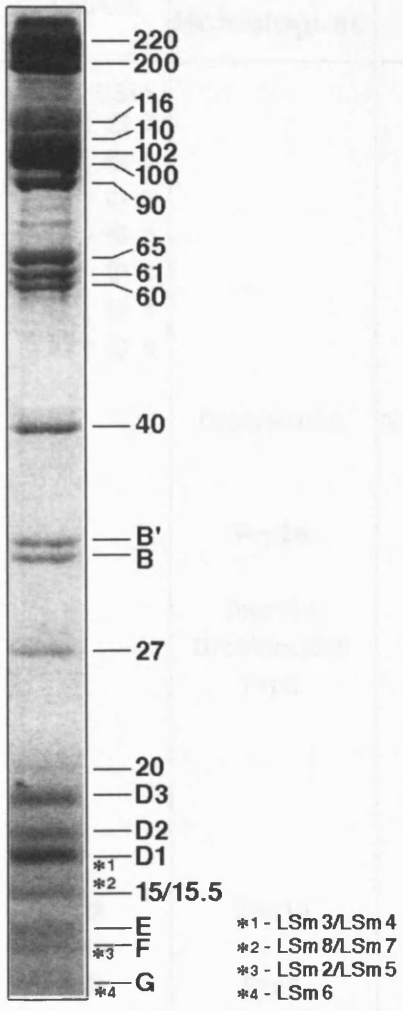


Figure 4.1. Protein composition of 25S [U4/U6.U5] tri-snRNPs. Glycerol gradient purified 25S [U4/U6.U5] tri-snRNPs were analysed by 13% SDS-PAGE and visualised by Coomassie staining. The specific proteins are assigned according to their apparent molecular mass. Sm proteins are lettered and LSm proteins are marked with asterisks.

app. MW kDa	25S [U4/U6.U5]		20S U5	U4/U6		Yeast Homologues	Sequence Motif
	Sm	LSm	Sm	Sm	LSm		Sm-motif
G	○	◎ 7	○	○	◎ 7		
F	○	◎ 6	○	○	◎ 6		
E	○	◎ 5	○	○	◎ 5		
D1	○	◎ 2	○	○	◎ 2		
D2	○	◎ 3	○	○	◎ 3		
D3	○	◎ 4	○	○	◎ 4		
B/B'	○	◎ 8	○	○	◎ 8		
15	●		●			Dib1/Snu16	Thioredoxin-like fold
40	●		●				WD-repeats
52			○				
100	●		●			Prp28	SR, DEAD-box
102	○		○				
116	●		●			Snu114	EF-2 homolog
200	●		●			Brr2/Snu246	DEXH-box
220	●		●			Prp8	
27	●						SR
65	○						
110	○						
15.5	◎			◎		Snu13	RNA-binding
20	◎			◎			Cyclophilin-like
60	◎			◎		Prp4	WD-repeats
61	○			○			
90	◎			◎		Prp3	

Table 4.1. Complexity of mammalian [U4/U6.U5] tri-snRNPs. Schematic presentation of protein composition of the gradient purified 25S [U4/U6.U5] tri-snRNP and 20S U5 snRNPs. The U4/U6-specific proteins were assigned by immunoprecipitation. Open circles represent proteins whose cDNAs had not been identified.

orthologues were found in yeast. U5-220kD is a human homologue of the essential yeast factor Prp8p, which is also a component of yeast U5 snRNP (Lossky *et al.*, 1987; Pinto and Steitz, 1989). It was shown that both proteins play important roles in both the first and second chemical steps of splicing (Teigelkamp *et al.*, 1995; Umen and Guthrie, 1995; Wyatt *et al.*, 1992; MacMillan *et al.*, 1994; Collins and Guthrie, 1999; Siatecka *et al.*, 1999). U5-100kD is a human homologue of the yeast splicing factor Prp28p (Teigelkamp *et al.*, 1997); both proteins are members of the DEAD-box family of putative ATP-dependent RNA helicases, but human protein contains the RS domain in addition. The two human U4/U6-specific proteins with molecular masses of 60 kDa and 90 kDa appear to be homologues of yeast Prp4p and Prp3p, respectively (Lauber *et al.*, 1997).

Currently very little is known in mammalian systems about the roles of the tri-snRNP proteins in tri-snRNP biogenesis. The tri-snRNP is a major subunit of the spliceosome. During splicing cycle, it integrates into pre-spliceosome and converts it into the spliceosome. Consequently, tri-snRNP dissociates and then reforms for the next cycle. The first questions to address might be the roles of individual proteins in tri-snRNP formation and the roles of tri-snRNP proteins in integration of the whole particle into the spliceosome.

Remarkably, several proteins of tri-snRNP have been found to contain sequence motifs such as DEAH/DEXH (100kD, 200kD), GTP-binding (116kD), or peptidyl-prolyl isomerase (20kD). The presence of these domains suggests the roles for proteins in RNA and protein conformational changes. However, these possibilities and their significance for splicing still have to be examined.

Objectives and Strategy

My project was part of a program to identify and characterize the [U4/U6.U5] tri-snRNP proteins. First, the individual snRNP particles are isolated in a preparative amount. The proteins are separated from RNA by phenol extraction and precipitated with acetone. Preparative SDS-PAGE is carried out in order to achieve single band separation of individual polypeptides. If there are several proteins that migrate similarly (like the four proteins of the tri-snRNP around 100 kDa), two-dimensional gel electrophoresis is applied. The bands or spots are cut off and send for peptide sequencing. If the individual protein is obtained in such

amount that it could be seen on a preparative gel (1 mm) stained with Coomassie then, conventional N-terminal peptide sequencing by the Edman method (Edman and Beggs, 1967; Niall, 1973) is used after proteolytic digestion. If a protein is seen only on a silver stained gel, the MALDI-MS method is applied in combination with an EST (expressed sequence tag) database search (Shevchenko *et al.*, 1996; Gottschalk *et al.*, 1998). Information obtained from peptide sequencing is used to begin the identification of corresponding cDNA. The initial task is to clone the cDNA. Next, it is necessary to prove that the cloned cDNA encodes the protein of interest. The availability of a cDNA clone affords an opportunity to characterize properties of the protein using antibodies, *in vitro*-translated or recombinant protein with the ultimate goal to discover the function of this protein in splicing or spliceosome formation.

This chapter describes my work on the 61 kDa protein (61kD) of tri-snRNP and my part in characterisation of the newly identified 102kD protein.

4.2. RESULTS AND DISCUSSION

4.2.1. Identification and characterization of the [U4/U6.U5] tri-snRNP specific protein of 61 kDa

Spliceosomal snRNPs (U1, U2, U4/U6 and U5) were purified from a HeLa nuclear extract by immunoaffinity chromatography under low-salt conditions (250 mM NaCl) using the monoclonal antibody H20 directed against m3G-cap of snRNAs (Bochnig *et al.*, 1987; Bach *et al.*, 1990). Total snRNPs were subsequently separated by density gradient centrifugation. Figure 4.2 shows the protein and RNA compositions of different fractions from a 10-30% glycerol gradient, panel A and B, respectively. A protein with apparent molecular weight of 61 kDa co-migrates in fractions corresponding to the 25S [U4/U6.U5] tri-snRNP. These fractions were combined and 500 µg of proteins was subjected to preparative SDS-PAGE.

4.2.1.1. Identification of cDNA encoding the 61kD protein

A band corresponding to the 61kD protein was cut out of a Coomassie stained gel and sent for microsequencing, which was carried out as described by Aebersold *et al.* (1987). The peptide sequences obtained were the following:

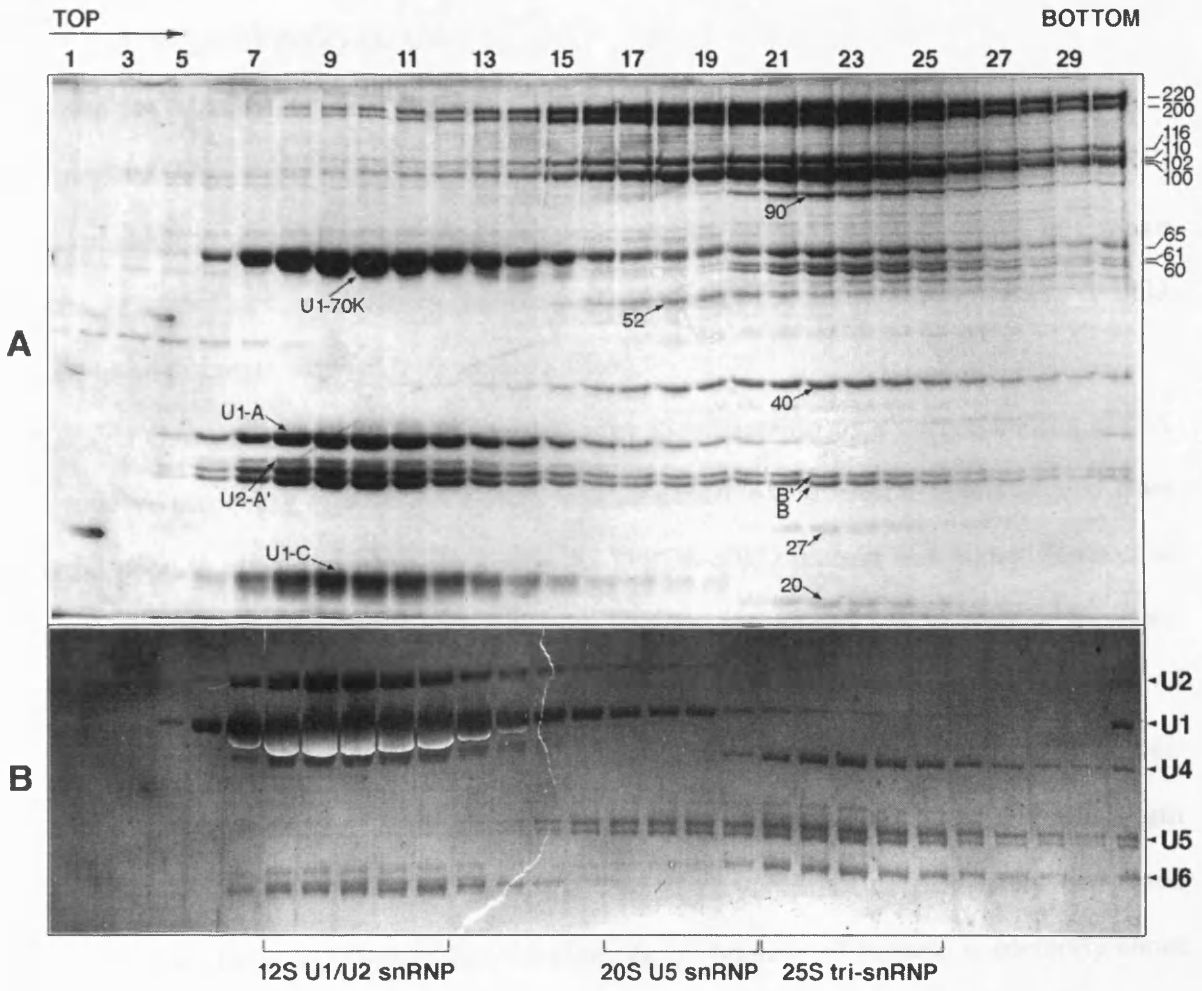


Figure 4.2. Fractionation of snRNP particles by density gradient centrifugation. 1.0 mg (0.4 ml) of total snRNPs, eluted from an anti-m_3G affinity column, were layered onto a linear 4 ml, 10-30% glycerol gradient prepared with buffer Na-150. The gradients were centrifuged in a Beckman SW60 rotor at 4°C for 14 hours at 29,000 rpm. Twenty-nine fractions of 150 µl were harvested manually from top to bottom. **A:** Proteins from 50 µl aliquots were analysed by 10% SDS-PAGE and visualised by Coomassie staining. **B:** RNA was recovered from 100 µl aliquots by phenol extraction and ethanol precipitation, analysed by 10% urea-PAGE, and visualised by silver staining. The identity of the RNAs is indicated on the right.

pep.1: RFPELESLVPNALDYIRTVK

pep.2: YFS**E**MAEFL**E**R**K**

pep.3: AEFLKVKGEKSGLM

pep.4: MKI**F**EYIS**G**QAKASE

pep.5: GVAGGLTNLSK

pep.6: FAEI

It is important to note that microsequencing by the Edman method could not make assignments with absolute certainty. In this particular case the amino acids marked in bold appeared to be different when cDNA was identified.

Even five years ago a general strategy for identification of a corresponding cDNA would involve screening of a cDNA library with degenerate oligonucleotides designed from the partial peptide sequences. In such a way the U4/U6-60kD protein was cloned (Lauber *et al.*, 1997). Nowadays a database search with peptide sequences has proven to be more efficient. The aim is to deduce a putative full-length cDNA sequence from the database. If overlapping ESTs are available they could be obtained commercially and pasted together. Alternatively, primers could be designed for the extreme 5' and 3' ends and the full-length cDNA would be amplified by PCR from a cDNA library. With a constitutively expressed protein of molecular weight lower than 60 kDa, the probability of success is currently about 90%.

A database search for open reading frames containing the peptide sequences listed above yielded several ESTs derived from cDNA libraries of different human and mouse tissues. This was an indication that these ESTs correspond to a general factor expressed in different cells. The phylogenetic conservation of splicing components is very high. Therefore, a comparison of human and mouse sequences serves as an intrinsic control for the correct open reading frame (ORF). The following ESTs were selected, obtained commercially and sequenced:

1. EST Id - mn96d01, gene bank (GB#) accession number AA087405 was isolated from a cDNA library of mouse lung.

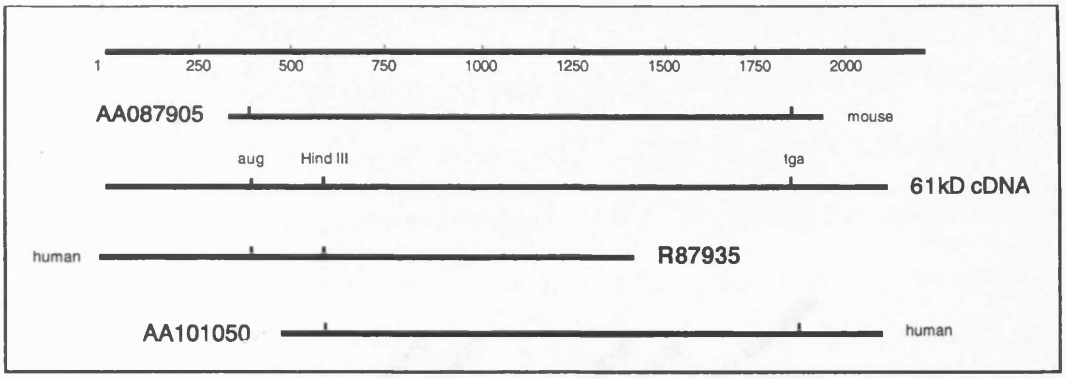
2. EST Id - yo47b11, GB# R87935(6), from human adult brain cDNA library.

3. EST Id - zm27h11, GB# AA101050(1), from human pancreas cDNA library.

The mouse EST contained an ORF of 499 amino acids, which is reasonable for a protein of apparent molecular weight of 61 kDa. At that time, there were no human ESTs, containing the full-length protein. Thus, two human ESTs presented above were the longest overlapping ones. These ESTs contain a unique Hind III site in the overlapping region that was used to paste them together. Figure 4.3B shows the cDNA sequence corresponding to the 61kD protein. This sequence contains an ORF of 499 amino acids with calculated molecular weight of 55,447 Da. A comparison of human and mouse amino acid sequences revealed only 5 amino acid substitutions $(T\text{-}S)_{20}$, $(K\text{-}N)_{76}$, $(E\text{-}D)_{181}$, $(D\text{-}E)_{188}$, $(L\text{-}T)_{496}$ that result in 99% identity and 99.6% similarity. All six peptides obtained by microsequencing were found in this cDNA sequence (underlined in Fig. 4.3B).

The following considerations suggest that this sequence is a full-length clone. First, the ORF encodes a protein with a theoretical molecular weight of 55 kDa, and that is within error margins of the protein size determined by gel electrophoresis. Second, we concluded that the second AUG triplet found in the cDNA sequence is an initiation codon because (i) there is an in frame stop codon 11 triplets upstream of this initiation codon (ii) nucleotides preceding this AUG match the Kozak consensus sequence (Kozak, 1987) much better than the first AUG (iii) the first AUG would produce a peptide of only 18 amino acids. Third, the sequence AUUAAA is located 259 nt downstream of the stop codon and 11 nt upstream of the poly(A) tail. This sequence is the only common natural variant of poly(A) signal that is nearly as active as the canonical sequence AAUAAA (Wilusz *et al.*, 1989; Wickens, 1990).

The authenticity of the cDNA was demonstrated using antiserum raised against the C-terminal peptide 3 (amino acids 484 to 497) coupled to chicken ovalbumin as described in the Materials and Methods. The antibodies were affinity purified on a Sulfolink column (Pierce) with immobilised peptide (described in the Materials and Methods). Specificity of the affinity purified peptide antibodies is comparable to that of monoclonal antibodies. Figure 4.4 shows that these antibodies strongly and specifically recognise a single protein in nuclear extracts (lane 3) and 25S [U4/U6.U5] tri-snRNPs but not in 20S U5 snRNPs (lanes 5 and 4, respectively). This result suggests that the 61kD protein designated as tri-snRNP-specific is either a component of the U4/U6 particle or associates only with the entire tri-snRNP particle as the protein does not present in isolated U5 snRNPs.



B

```
   1 gaattcggcacgaggcgaatccctgcttaaaggccttgctttcttgtctaacgccgcaaccagtcctctgagttgccaacgtctttcttcttgtctcgac 100

 101 gccccgtcgtccggccacagcgattctctgcttagcaggatcggtccacagcgggacgtgagtccctttcctcctcgcggcttaccgcctctctccgcct 200

 201 agtgccaggtgctaataaagttgttgtttcaaatgcggccaggaacatcgcgagcggggaccaatcagagagtagctttgcctctataacggcgcgagag 300

 301 tgagacgtcatcggtgagcgactaacgdtagaaacagtggtgcgcggagaggagaggcctcggg ATG TCT CTG GCA GAT GAG CTC TTA GCT 391
   1                                                                M   S   L   A   D   E   L   L   A   9

 392 GAT CTC GAA GAG GCA GCA GAA GAG GAG GAA GGA GGA AGC TAT GGG GAG GAA GAA GAG GAG CCA GCG ATC GAG GAT 466
  10 D   L   E   E   A   A   E   E   E   E   G   G   S   Y   G   E   E   E   E   E   P   A   I   E   D  34

 467 GTG CAG GAG GAG ACA CAG CTG GAT CTT TCC GGG GAT TCA GTC AAG ACC ATC GCC AAG CTA TGG GAT AGT AAG ATG 541
  35 V   Q   E   E   T   Q   L   D   L   S   G   D   S   V   K   T   I   A   K   L   W   D   S   K   M  59

 542 TTT GCT GAG ATT ATG ATG AAG ATT GAG GAG TAT ATC AGC AAG CAA GCC AAA GCT TCA GAA GTG ATG GGA CCA GTG 616
  60 F   A   E   I   M   M   K   I   E   E   Y   I   S   K   Q   A   K   A   S   E   V   M   G   P   V  84

 617 GAG GCC GCG CCT GAA TAC CGC GTC ATC GTG GAT GCC AAC AAC CTG ACC GTG GAG ATC GAA AAC GAG CTG AAC ATC 691
  85 E   A   A   P   E   Y   R   V   I   V   D   A   N   N   L   T   V   E   I   E   N   E   L   N   I 109

 692 ATC CAT AAG TTC ATC CGG GAT AAG TAC TCA AAG AGA TTC CCT GAA CTG GAG TCC TTG GTC CCC AAT GCA CTG GAT 766
 110 I   H   K   F   I   R   D   K   Y   S   K   R   F   P   E   L   E   S   L   V   P   N   A   L   D 134

 767 TAC ATC CGC ACG GTC AAG GAG CTG GGC AAC AGC CTG GAC AAG TGC AAG AAC AAT GAG AAC CTG CAG CAG ATC CTC 841
 135 Y   I   R   T   V   K   E   L   G   N   S   L   D   K   C   K   N   N   E   N   L   Q   Q   I   L 159

 842 ACC AAT GCC ACC ATC ATG GTC GTC AGC GTC ACC GCC TCC ACC ACC CAG GGG CAG CAG CTG TCG GAG GAG GAG CTG 916
 160 T   N   A   T   I   M   V   V   S   V   T   A   S   T   T   Q   G   Q   Q   L   S   E   E   E   L 184

 917 GAG CGG CTG GAT GAG GCC TGC GAC ATG GCG CTG GAG CTG AAC GCC TCC AAG CAC CGC ATC TAC GAG TAT GTG GAG 991
 185 E   R   L   D   E   A   C   D   M   A   L   E   L   N   A   S   K   H   R   I   Y   E   Y   V   E 209

 992 TCC CGG ATG TCC TTC ATC GCA CCC AAC CTG TCC ATC ATT ATC GGG GCA TCC ACG GCC GCC AAG ATC ATG GGT GTG 1066
 210 S   R   M   S   F   I   A   P   N   L   S   I   I   I   G   A   S   T   A   A   K   I   M   G   V 234

1067 GCC GGC GGC CTG ACC AAC CTC TCC AAG ATG CCC GCC TGC AAC ATC ATG CTG CTC GGG GCC CAG CGC AAG ACG CTG 1141
 235 A   G   G   L   T   N   L   S   K   M   P   A   C   N   I   M   L   L   G   A   Q   R   K   T   L 259

1142 TCG GGC TTC TCG TCT ACC TCA GTG CTG CCC CAC ACC GGC TAC ATC TAC CAC AGT GAC ATC GTG CAG TCC CTG CCA 1216
 260 S   G   F   S   S   T   S   V   L   P   H   T   G   Y   I   Y   H   S   D   I   V   Q   S   L   P 284

1217 CCG GAT CTG CGG CGG AAA GCG GCC CGG CTG GTG GCC GCC AAG TGC ACA CTG GCA GCC CGT GTG GAC AGT TTC CAC 1291
 285 P   D   L   R   R   K   A   A   R   L   V   A   A   K   C   T   L   A   A   R   V   D   S   F   H 309

1292 GAG AGC ACA GAA GGG AAG GTG GGC TAC GAA CTG AAG GAT GAG ATC GAG CGC AAA TTC GAC AAG TGG CAG GAG CCG 1366
 310 E   S   T   E   G   K   V   G   Y   E   L   K   D   E   I   E   R   K   F   D   K   W   Q   E   P 334

1367 CCG CCT GTG AAG CAG GTG AAG CCG CTG CCT GCG CCC CTG GAT GGA CAG CGG AAG AAG CGA GGC GGC CGC AGG TAC 1441
 335 P   P   V   K   Q   V   K   P   L   P   A   P   L   D   G   Q   R   K   K   R   G   G   R   R   Y 359

1442 CGC AAG ATG AAG GAG CGG CTG GGG CTG ACG GAG ATC CGG AAG CAG GCC AAC CGT ATG AGC TTC GGA GAG ATC GAG 1516
 360 R   K   M   K   E   R   L   G   L   T   E   I   R   K   Q   A   N   R   M   S   F   G   E   I   E 384

1517 GAG GAC GCC TAC CAG GAG GAC CTG GGA TTC AGC CTG GGC CAC CTG GGC AAG TCG GGC AGT GGG CGT GTG CGG CAG 1591
 385 E   D   A   Y   Q   E   D   L   G   F   S   L   G   H   L   G   K   S   G   S   G   R   V   R   Q 409

1592 ACA CAG GTA AAC GAG GCC ACC AAG GCC AGG ATC TCC AAG ACG CTG CAG CGG ACC CTG CAG AAG CAG AGC GTC GTA 1666
 410 T   Q   V   N   E   A   T   K   A   R   I   S   K   T   L   Q   R   T   L   Q   K   Q   S   V   V 434

1667 TAT GGC GGG AAG TCC ACC ATC CGC GAC CGC TCC TCG GGC ACG GCC TCC AGC GTG GCC TTC ACC CCA CTC CAG GGC 1741
 435 Y   G   G   K   S   T   I   R   D   R   S   S   G   T   A   S   S   V   A   F   T   P   L   Q   G 459

1742 CTG GAG ATT GTG AAC CCA CAG GCG GCA GAG AAG AAG GTG GCT GAG GCC AAC CAG AAG TAT TTC TCC AGC ATG GCT 1816
 460 L   E   I   V   N   P   Q   A   A   E   K   K   V   A   E   A   N   Q   K   Y   F   S   S   M   A 484

1817 GAG TTC CTC AAG GTC AAG GGC GAG AAG AGT GGC CTT ATG TCC ACC TGA atgactgcgtgtgtccaaggtggcttcccactgaag 1900
 485 E   F   L   K   V   K   G   E   K   S   G   L   M   S   T   *                                      500

1901 ggacacagaggtccagtccttctgaaggctaggatcgggttctggcagggagaacctgccctgccactggccccattgctgggactgcccagggaggag 2000

2001 gccttggaagagtccggcctggcctcccccaggaccgagatcaccgcccagtatgggctagagcaggttttcatcatgccttgtctttttttaactgagaa 2100

2101 aggagatttttgaaaagagtacaattaaaaggacattgtcaaaaaaaaaaaaaaaaaaactcgag                                    2166
```

Figure 4.3. The cDNA encoding the human 61kD tri-snRNP protein. A: Schematic presentation of the cloning strategy. Two human ESTs, acc. No R87935 and AA101050, were pasted together using Hind III site. **B:** Nucleotide and predicted amino acid sequences of the 61kD protein. Peptide sequences obtained by microsequencing are underlined. The termination codon is marked by an asterisk, the in-frame stop codon in the 5'-UTR is boxed, and the polyadenilation signal in the 3'-UTR is shown in bold.

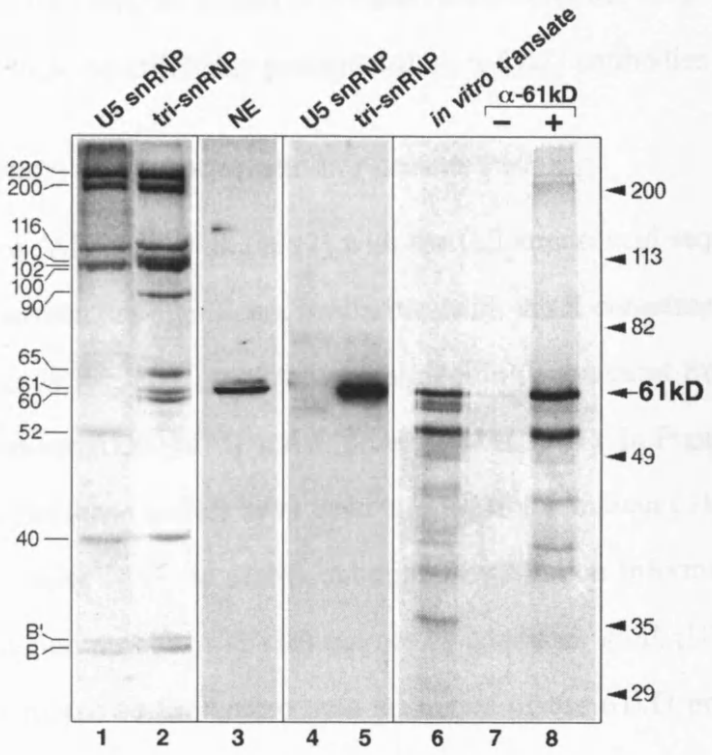


Figure 4.4. Characterisation of the cDNA encoding the 61kD protein. Various protein samples were subjected to 10% SDS-PAGE. The gel was divided and different parts were processed as follows. Left panel: the U5 snRNP proteins (lane 1) and the tri-snRNP proteins (lane 2) were visualised by Coomassie staining. Middle panels: total HeLa nuclear extract proteins (lane 3), U5 proteins (lane 4), and tri-snRNP proteins (lane 5) were immunostained with affinity-purified α-61kD antibodies. Right panel: the 61kD cDNA was translated *in vitro*, immunoprecipitated and visualised by fluorography. Lane 6 shows the *in vitro* translate (20% of the amount used for immunoprecipitation), and lane 7 and 8 are the proteins precipitated by Protein-A-Sepharose in the absence (-) or presence of α-61kD (+), respectively. The positions of the pre-stained molecular weight markers (Bio-Rad) are indicated on the right.

Another conventional criteria for a cDNA identity was met in the experiments using the *in vitro*-translated [^{35}S]methionine-labelled protein. The cDNA of the 61kD protein was cloned in pBluescript SK(-) under the control of T3 promoter. The *in vitro*-translated 61kD protein was produced in rabbit reticulocyte lysate of TNT coupled system (Promega) from this plasmid. Figure 4.4 shows that the largest *in vitro*-translated product co-migrates with the native 61kD protein and could be efficiently precipitated by α-61kD antibodies (lanes 6-8).

4.2.1.2. The 61kD protein is phylogenetically conserved

Database searches (Altschul *et al.*, 1997) with the full amino acid sequence of 61kD revealed that this novel protein has significant similarities with the *S. cerevisae* splicing factor Prp31p (Maddock *et al.*, 1996), and predicted protein coding sequences from *A. thailana* (gene bank accession number AC004473) and *S. pombe* (AL022117). In Figure 4.5A, amino acid sequences from the database entries have been aligned to the human 61kD protein. The sequence of Prp31p corresponds to the cDNA submitted by Munich Information centre for Protein Sequences (accession number Z72876) but not by Maddock *et al.* (U31970) because the former has a better match to the amino acid sequence of the 61kD protein. The two sequences for Prp31p differ in a stretch of 34 amino acids that arises from insertion followed by deletion of a single nucleotide which does not introduce a stop codon. The alignment (Fig. 4.5A) demonstrates that regions of homology between these four proteins are not restricted to a particular domain, but rather distributed throughout the entire sequence. All ORFs from given species appear to be derived from the same gene origin and probably represent orthologues.

The *S. cerevisiae* Prp31 protein was identified and characterised as an essential splicing factor in the laboratory of John Woolford (Maddock *et al.*, 1996; Weidenhammer *et al.*, 1996). The initial analysis of Prp31p indicated that it is required before or concomitant with the first cleavage reaction of splicing both *in vivo* and *in vitro*. Further, it was shown that cell extracts derived from a temperature sensitive *prp31-1* strain fail to form mature spliceosomes that contain [U4/U6.U5] tri-snRNPs upon heat inactivation, although commitment and pre-spliceosomal complexes were detected (Weidenhammer *et al.*, 1997). Co-immunoprecipitation experiments with HA-tagged Prp31 protein revealed its association

Figure 4.5. Sequence analysis of the human 61kD protein. A: Comparison of the primary structure of the human 61kD protein and its putative orthologs from *Arabidopsis thaliana* (acc. no AC004473), *Schizosaccharomyces pombe* (acc. no AL022117), and *Saccharomyces cerevisiae* Prp31p (acc. no Z72876). The alignment was carried out by the cluster method using the MegAlign program. Residues that are identical to the 61kD protein are boxed in black and those that are conserved are shaded grey. Conserved amino acids are combined in the standard functional groups: acidic - DE; basic - RKH; hydrophobic - AFILMPVW; polar - CGNQSTY. **B:** Alignments of the human 61kD protein with other highly conserved proteins: *Pisum sativum* SAR DNA-binding protein (acc. no AF061962), *Mus musculus* SIK (Src-related intestinal kinase) similar protein (acc. no AF053232), and *Human* nucleolar proteins NOP58 and NOP56 (acc. no AF123534 and Y12065, respectively). The alignment was carried out by the cluster method using the MegAlign program. Residues that are identical in at least two proteins are boxed in black and those that are conserved between at least two proteins are shaded grey. Conserved amino acids are combined as in (A). **C:** Table of sequence distances. **D:** Phylogenetic tree, generated by the MegAlign program.

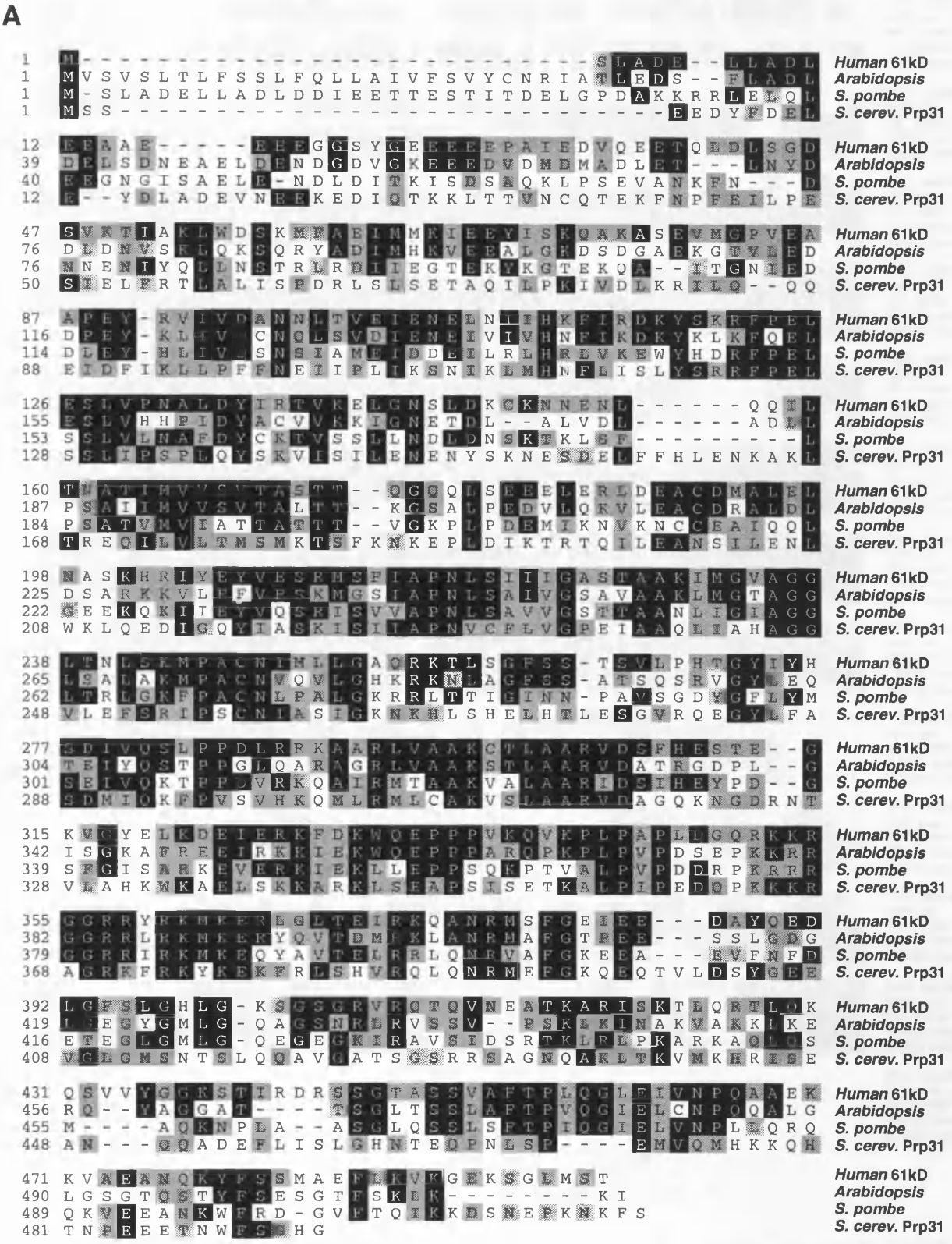


Figure 4.5A. Sequence analysis of the human 61kD protein.

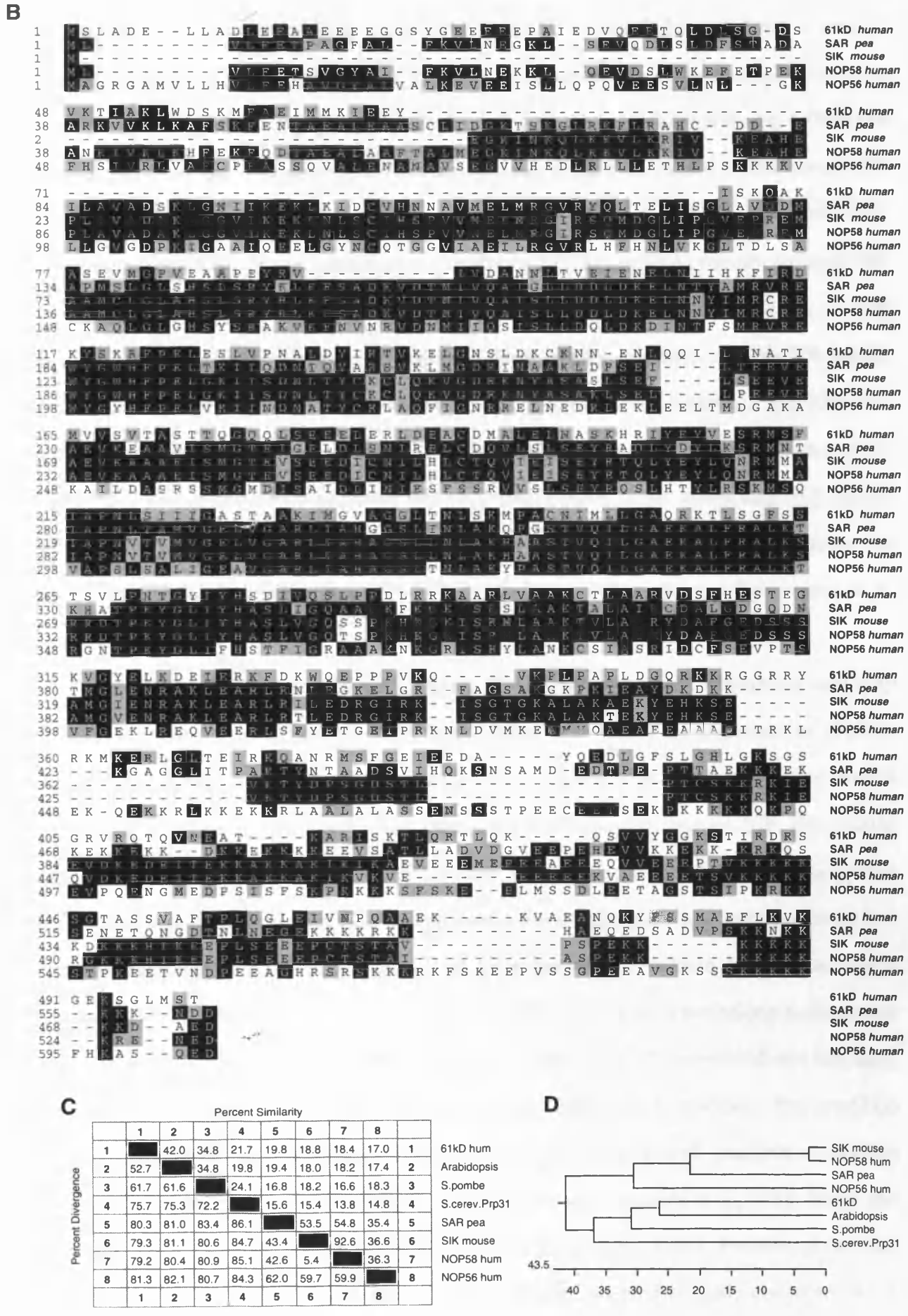


Figure 4.5. (B, C, D) Sequence analysis of the human 61kD protein.

with [U4/U6.U5] tri-snRNPs but not with U5 snRNPs. The failure to form a mature spliceosome in heat-inactivated extracts could not be explained by induced destabilisation of tri-snRNPs, as shown for Prp3p, Prp4p, Prp6p or Prp8p, because there were no differences detected between active and heat-inactivated splicing extracts in regard to the tri-snRNPs content. The only distinction observed in *prp31-1* extracts upon shift to the non-permissive temperature was that free U4 snRNA (not complexed with U6 snRNA) rapidly accumulated, but this phenomenon did not get a clear explanation. Finally, Weidenhammer *et al.* showed that pre-mRNA could be immunoprecipitated from a splicing reaction with HA-tagged Prp31p and this precipitation is dependent upon the presence of intact U2 snRNPs. The authors put together results obtained with different extracts (extract prepared from a strain harbouring the temperature sensitive *prp31-1* allele and extract containing HA-tagged Prp31p) and rationalised their findings with the conclusion that Prp31p promotes the association of [U4/U6.U5] tri-snRNP with pre-spliceosome to form spliceosomes in *S. cerevisiae*.

In addition to four putative orthologues, BLAST Search revealed reasonable similarity to another group of highly conserved proteins. Figure 4.5B shows the alignment of 61kD to the most prominent members of this group: human nucleolar proteins Nop56 and Nop58 (accession numbers Y12065 and AF123534, correspondingly), mouse SIK (Src-related intestinal kinase) similar protein (AF053232) and *Pisum sativum* SAR DNA-binding protein (AF061962). It is evident that the latter proteins are homologous and contain distinctive structural motifs (which are recognised as individual domains in ProDom database; accession numbers: PD008439, PD005332, PD004104). The 61kD protein does not belong to this group and shares homology only within short stretches of amino acids. This could indicate that there are particular properties in common, such as nucleic acid binding. A hypothesis that the 61kD protein might be an RNA-binding protein arose from the fact that SAR (scaffold attachment region) binding proteins recognise specific DNA structures (Izaurralde *et al.*, 1988; Strick and Laemmli, 1995; Gohring and Fackelmayer, 1997; Gohring *et al.*, 1997). Proteins of the Nop family are components of small nucleolar RNP particles and bind a highly conserved RNA structural motif in all CD box snoRNPs (Gautier *et al.*, 1997; Lafontaine and Tollervey, 1999). Moreover, Nop56/Nop58 could be crosslinked to the CD box snoRNAs (Watkins *et*

al., 1998; Watkins, personal communication). If the functions of these homologues proteins are really to recognise and bind to specific RNA structures, it is, therefore, likely that these proteins have evolved from common ancestor.

4.2.1.3. Immunoprecipitation of U4/U6 snRNAs with antibodies against the 61kD protein

Yeast splicing factors were identified through genetic screens and their association into snRNP complexes was assessed by co-immunoprecipitation of snRNAs from yeast extracts with the protein tested. For comparison of the human 61kD protein with its apparent yeast orthologue, Prp31p, immunoprecipitation from HeLa nuclear extracts was performed using affinity-purified antibodies against the C-terminal peptide (pep. 3). To examine the strength of association with snRNAs, immunoprecipitations were performed in buffers of increasing salt concentrations, and compared to that of well-characterised tri-snRNP proteins, U4/U6-60kD and U5-116kD. Co-immunoprecipitated RNA species were extracted and analysed by both pCp-labelling and northern blot with radiolabelled fragments of snRNA genes as probes (Figure 4.6, A and B, respectively). The labelling of RNA 3' ends with [^{32}P]-pCp allows to visualise all RNA species, except U6 snRNA (England and Uhlenbeck, 1978; Lund and Dahlberg, 1992). Therefore, a northern blot was used to detect U6 snRNAs. Y12, monoclonal anti-Sm antibodies, serve as a control for specific precipitation of all snRNPs. At low salt concentration the U4, U5 and U6 snRNAs were co-immunoprecipitated with the 61kD protein. At higher salt concentrations, when tri-snRNPs are dissociated, only the U4 and U6 snRNAs are precipitated. The immunoprecipitation profile of the 61kD protein is almost identical to that of the U4/U6-60kD protein and differs from U5-116kD, which precipitates only U5 snRNAs at high salt concentrations. This result demonstrates that the 61kD protein associates with U4/U6 snRNAs. It is also tempting to suggest that the 61kD protein might interact with the [90/60/20] complex. Co-immunoprecipitation of snRNAs with Prp31p described by Weidenhammer *et al.* (1997) is very similar to that observed for 61kD protein.

4.2.1.4. Investigation of possible RNA-binding activity by the 61kD protein

Despite the absence of any known RNA-binding motifs in the 61kD protein, a similarity to the Nop and SAR-binding protein families as well as co-immunoprecipitation of

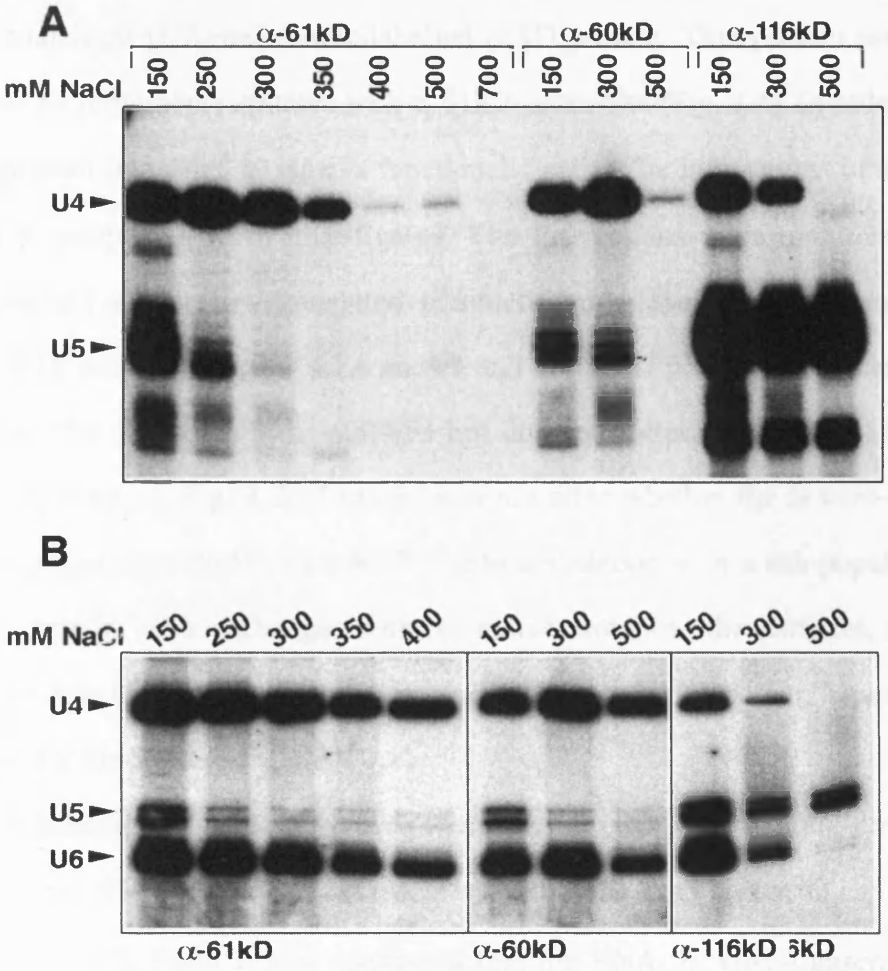


Figure 4.6. Co-immunoprecipitation of snRNAs from HeLa nuclear extract with antibody against the 61kD protein. Immunoprecipitations from HeLa nuclear extract were performed at different salt concentrations as indicated above each lane. α-116kD and α-60kD are used as controls for the precipitation of U5 snRNPs and U4/U6 snRNPs, respectively. **A:** Precipitated RNA was 3' end-labelled with [^{32}P]-pCp and analysed by 10% urea-PAGE. **B:** RNA was separated by 10% urea-PAGE and snRNAs were visualised by Northern hybridisation using probes specific for the U4, U5 and U6 snRNAs.

U4/U6 snRNAs prompted a study of RNA-binding activity by the 61kD protein. Various attempts to isolate recombinant protein expressed in E. *coli* failed, and experiments were done using *in vitro*-translated [^{35}S]methionine-labelled 61kD protein. This protein seems to be soluble, as it can be immunoprecipitated with α-61kD antibodies (Fig. 4.4). In order to prove that the 61kD protein translated *in vitro* is functionally active the interactions of 61kD with different snRNP complexes were investigated. The interactions were monitored by co-immunoprecipitation of *in vitro*-translated [^{35}S]methionine-labelled 61kD protein with snRNPs using Y12 antibody. Figure 4.7A shows that the 61kD protein associates with the gradient-purified 25S [U4/U6.U5] tri-snRNPs but does not interact with the 12S U1/U2 snRNPs (fractions 9 to 11, Fig. 4.2). Though, it is not clear whether the *in vitro*-translated 61kD protein incorporates into the tri-snRNP due to association with a sub-population that lacks the 61kD protein, or it exchanges with the native protein in the particles, this result demonstrates the specific binding by the *in vitro*-translated 61kD protein. Therefore, this protein was used for further investigations.

An RNA-binding ability of 61kD was studied by another type of "pull-down assay" with the monoclonal H20 antibody which reacts with both the m_3G and m^7G cap structures (Bochning *et al.*, 1987). First, it was confirmed that the RNA, *in vitro*-transcribed under conditions when $m^7G(5')ppp(5')G$ is incorporated at the 5' end, could be immunoprecipitated with H20 antibody (Fig. 4.7B). Then, binding of the *in vitro*-translated [^{35}S]methionine-labelled 61kD protein to different *in vitro*-transcribed and capped RNA substrates was monitored by co-immunoprecipitation with H20 antibody. All five snRNAs (U1, U2, U4, U5 and U6 snRNAs) as well as the model pre-mRNA MINX (Zillmann *et al.*, 1988) were examined for binding. When incubation mixtures of 50 µl contained 1 pmol of RNA and 7 µl of [^{35}S]-labelled 61kD, the binding of protein to any of the RNAs was not detected (data not shown). In the next experiment, increasing concentrations (up to 1 µM) of U1, U4, U5 and U6 *in vitro*-transcribed snRNAs were used in order to detect interactions as weak as ones with affinity of 10^6 M^{-1}. Figure 4.7C shows that none of the RNA substrates could precipitate the 61kD protein as the signals obtained in the absence or presence of RNA are the same intensity and could be considered as a background.

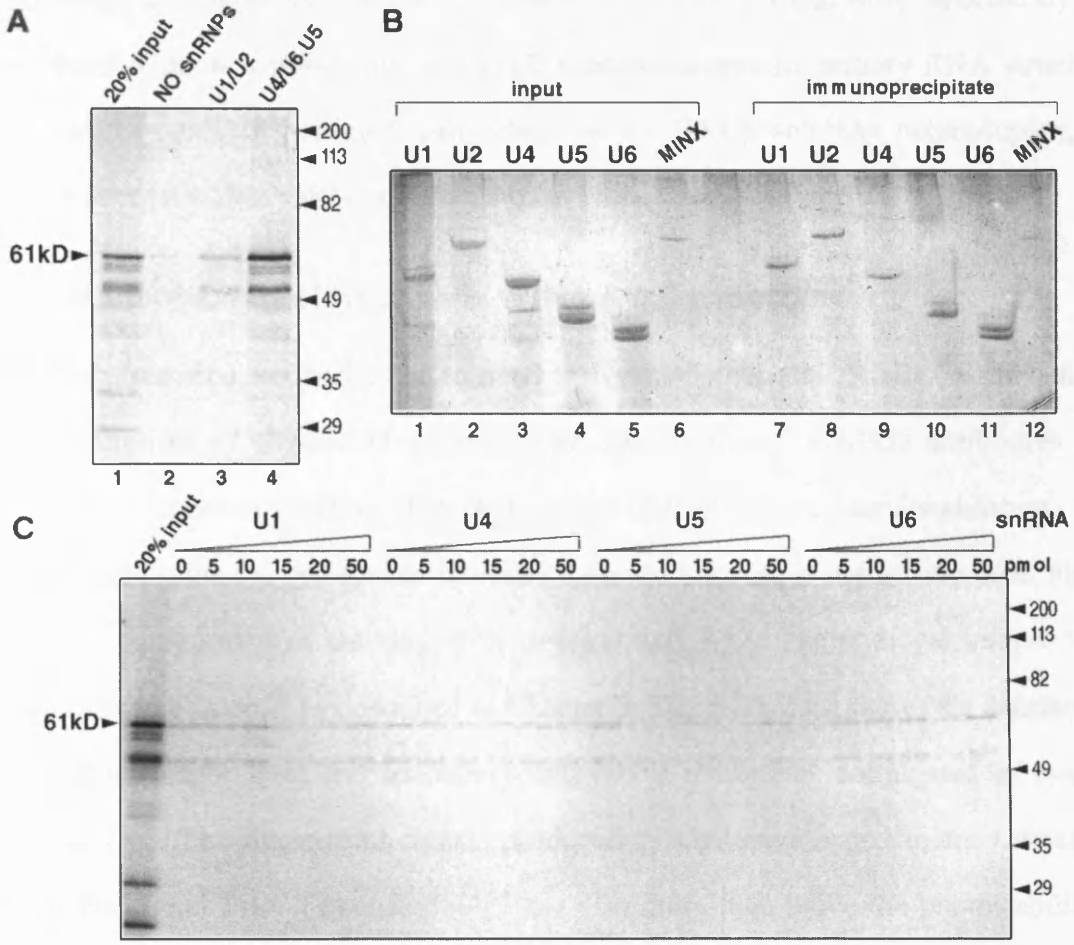


Figure 4.7. Binding of the 61kD protein to snRNPs and RNAs. A: Specific association of the 61kD protein with [U4/U6.U5] snRNPs. The *in vitro* translated, [^{35}S]-labelled 61kD protein was incubated with a mixture of U1 and U2 snRNPs (lane 3), or tri-snRNPs (lane 4). The snRNPs were immunoprecipitated with Y12 antibodies, and co-precipitated proteins were analysed by 10% SDS-PAGE followed by fluorography. Lane 1 represents 20% of the *in vitro* translate added to each immunoprecipitation. Lane 2 is a control for the co-precipitation of the 61kD protein with Protein-A-Sepharose beads charged with Y12 antibody, in the absence of snRNPs. The migration of pre-stained molecular weight markers is indicated on the right. **B:** Precipitation of the *in vitro* transcribed, m^{7}G-capped RNAs with H20 antibody. The U1, U2, U4, U5, U6 snRNAs and MINX pre-mRNA were transcribed *in vitro* in the presence of m^{7}G(5')ppp(5')G, and used for immunoprecipitation with H20 antibody. Antibody-bound RNA was recovered from Protein-A-Sepharose beads by phenol extraction and ethanol precipitation, analysed by 10% urea-PAGE, and visualised by silver staining. Lanes 1-6 represent the material added to each immunoprecipitation reaction and lanes 7-12 show the precipitated RNAs. **C:** The *in vitro* translated 61kD protein does not bind the *in vitro* transcribed RNAs. The [^{35}S]-labelled 61kD protein was incubated with m^{7}G-capped U1, U4, U5, U6 snRNAs. Incubation mixtures contained increasing amounts of RNA as indicated in the figure. The snRNAs were immunoprecipitated with H20 antibody, and co-precipitated proteins were analysed by 10% SDS-PAGE followed by fluorography. The positions of molecular weight protein markers are indicated on the right.

Although no interactions between RNAs and the 61kD protein were detected by this approach, there remains a possibility that 61kD recognises specific tertiary RNA structures that are present in snRNP particles, particularly in the U4/U6 snRNAs heteroduplex, but might not be formed within snRNAs transcribed *in vitro*.

4.2.1.5. Immunostaining of HeLa cells with α-61kD antibodies

Affinity purified antibodies were used to immunostain HeLa cells and reveal the cellular distribution of the 61kD protein. The specificity of α-61kD antibodies was demonstrated by western blotting (Fig. 4.4). As a control for nuclear localisation, Y12 antibodies (kindly provided by Iain Mattaj) that recognise Sm core proteins were used. Figure 4.8 shows the comparison of staining with α-61kD and Y12. The confocal images were obtained in the same manner as described in Chapter 3. The red colour shows the distribution of 61kD and corresponds to the secondary anti-rabbit antibodies conjugated to cyanine fluorochrome Cy3. The fluorescent signals produced by Cy3 have approximately twice the intensity of the signal from Texas Red. Cy3 has also more than twice the photostability of other fluorophores. The green colour corresponds to Y12 antibodies, detected by secondary anti-mouse antibodies conjugated to FITC. Significantly, the pattern produced by α-61kD antibodies is identical to the Y12 staining, and is different from the staining obtained with antibodies against most of other snRNP proteins. Both α-61kD and Y12 produce clear nuclear staining and highlight 2-3 coiled bodies as well as nuclear speckles in addition to diffuse staining through out the nucleoplasm but not nucleoli. The bright staining of coiled bodies with α-61kD antibodies is a distinctive feature of the protein in comparison with other snRNP proteins. Whether it implies the involvement of 61kD in the snRNA biogenesis has yet to be investigated.

4.2.2. Characterisation of U5 snRNP-specific protein of 102 kDa

The cDNA encoding the U5-102kD protein was identified in the laboratory (Makarov and Lührmann, unpublished results). I was involved in characterisation of the protein, and results are described below.

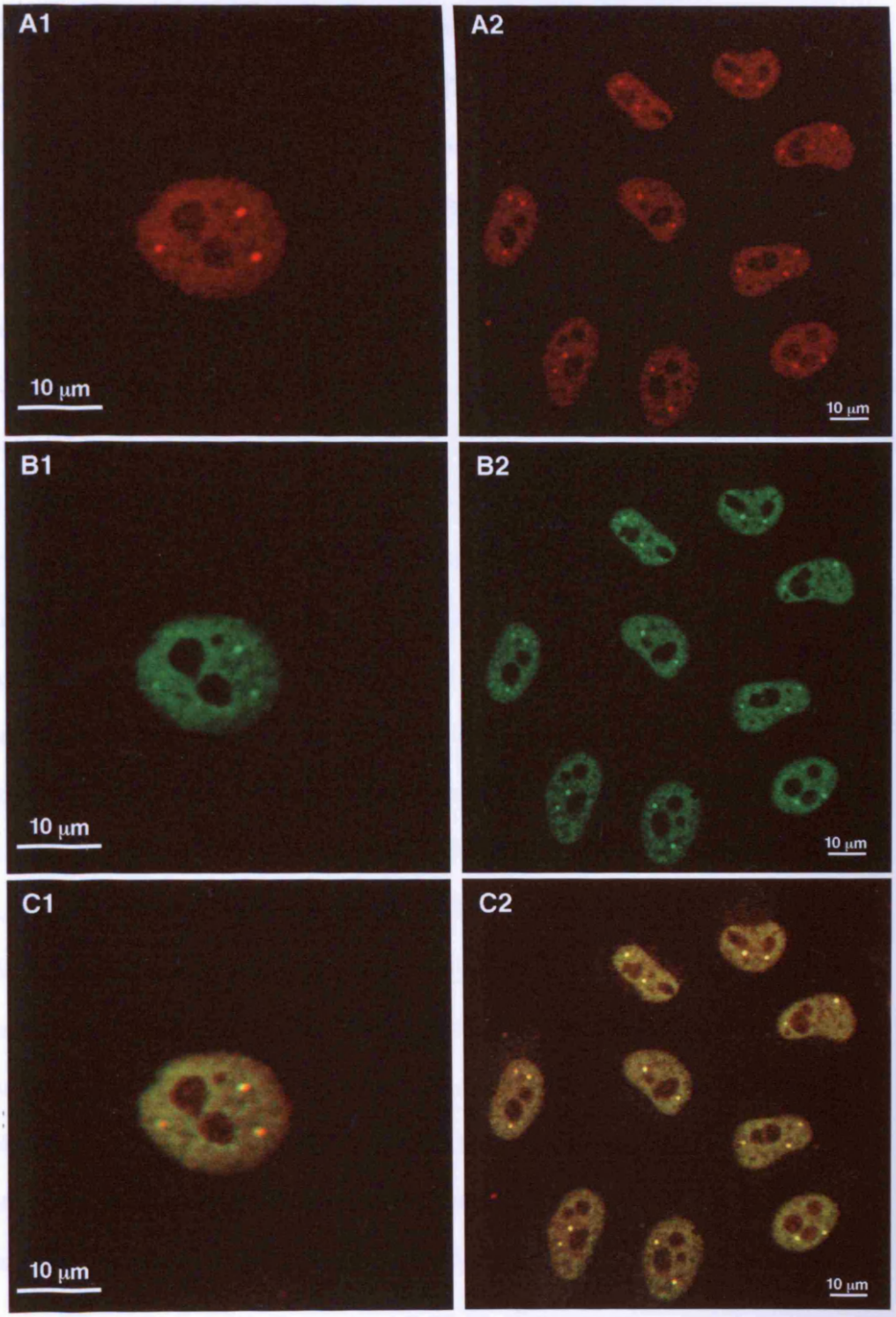


Figure 4.8. Immunofluorescence staining of the 61kD protein in HeLa cells. Two sets of images (1 and 2) show HeLa cells double-stained with α-RBM (A) and Y12 antibodies (B). The red fluorescence (A) corresponds to the 61kD protein and produced by secondary antibody conjugated to Cy3. The green fluorescence (B) corresponds to the Sm proteins and produced by the secondary antibody conjugated to FITC. The images A and B were recorded independently and combined in an overlay image C resulting in yellow staining of structures decorated by both antibodies. The bar indicates 10 µm.

4.2.2.1. Characterisation of the cDNA encoding U5-102kD protein

A comparison of the complete U5-102kD amino acid sequence to the sequences in database revealed that U5-102kD is a novel protein homologous to the splicing factors Prp6p from bakers yeast *Saccharomyces cerevisae* (27% identity, 45% similarity) and Prp1$^+$p from fission yeast *Schizosaccharomyces pombe* (44% identity, 62% similarity). The closest homologue of U5-102kD (57% identity, 72% similarity) found in database was a plant one: an ORF for the protein of 106 kDa was deduced from the complete nucleotide sequence of chromosome IV from *Arabidopsis thaliana* (nt 72099-74930, AF0712527). It is important to stress that regions of homology between these four proteins are not restricted to a particular domain, but rather distributed throughout the entire sequence (Fig. 4.9A).

The most distinctive structural feature of the U5-102kD protein as well as its homologues is a repeated tetratrico peptide (TPR) in the C-terminal part of the sequence (Fig. 4.9B). The TPR is a highly degenerate 34 amino acid motif found in around forty proteins that are evolutionary spread from bacteria to humans, and which are involved in a wide variety of processes such as cell cycle control, transcription repression, stress response, protein kinase inhibition, protein transport and neurogenesis (Sikorski *et al.*, 1990; Hirano *et al.*, 1990; Lamb *et al.*, 1995). Multiple sequence comparison of the TPRs revealed eight loosely conserved consensus residues (Fig. 4.9F). Although none of these residues is an invariant one, a tryptophan and proline are usually located at position 8 and 32, respectively, and small hydrophobic residues are commonly observed at positions 8, 20 and 27 within the motif. Typical of TPR proteins, these motif elements form a closely spaced array. Since definition of the TPR is rather subjective due to its degeneracy, the entire two-thirds part of the sequences is aligned in a TPR-like fashion in Figure 4.9 (C-E). The sequences of U5-102kD and Prp1$^+$p align more closely to the TPR consensus than does Prp6p, and we propose that the entire carboxy terminus of U5-102kD, starting from amino acid 309, consists of TPR-like structures. Interestingly, the temperature sensitive mutations in PRP6, which led to a dramatic decrease in the protein's abundance at the restrictive temperature, were mapped to the TPR13-18 region (Legrain and Choulika, 1990), and three amino acid substitutions detected in a *prp6 ts* mutant were in the residues which are identical in U5-102kD and Prp1p. In particular, crystal structures of TPRs (Das *et al.*, 1998) suggest that the change of a small hydrophobic alanine

Figure 4.9. Sequence analysis of the human U5-102kD protein. A: Comparison of the primary structure of the human U5-102kD protein with its putative orthologs in *Arabidopsis thaliana* (acc. no AF0712527; nt 72099-74930), *S.pombe* (Prp1$^+$p), and *S. cerevisiae* (Prp6p). Alignment was carried out by the cluster method using the MegAlign program. Residues that are identical in at least two sequences are boxed in black, whereas those that are conserved between at least two sequences are shaded grey. Arrows indicate the TPR-like elements of U5-102kD. **B:** Block diagram of the domain structure of U5-102kD and its yeast homologues. The TPR-like elements within each protein are boxed and those that are close to the TPR motif consensus sequence (see bottom line in F) are shaded grey. **C, D, E:** Alignments of the TPR-like elements of U5-102kD, *S.pombe* Prp1$^+$p and *S. cerevisiae* Prp6p, respectively. Residues that are identical or conserved in at least 9 out of 19 sequences are boxed in black or grey, respectively. Conserved amino acids are combined in the standard functional groups as in Fig.4.5. **F:** Alignment of the first TPR from U5-102kD with its homologues and other TPR containing yeast splicing factors. Ylr117 - Gene Bank accession number for the *S. cerevisiae* homologue of the Crn protein (Brian Rymond, personal communication). Helix A and Helix B were demonstrated by crystallography and their schematic presentation is adapted from Das *et al.* (1998).

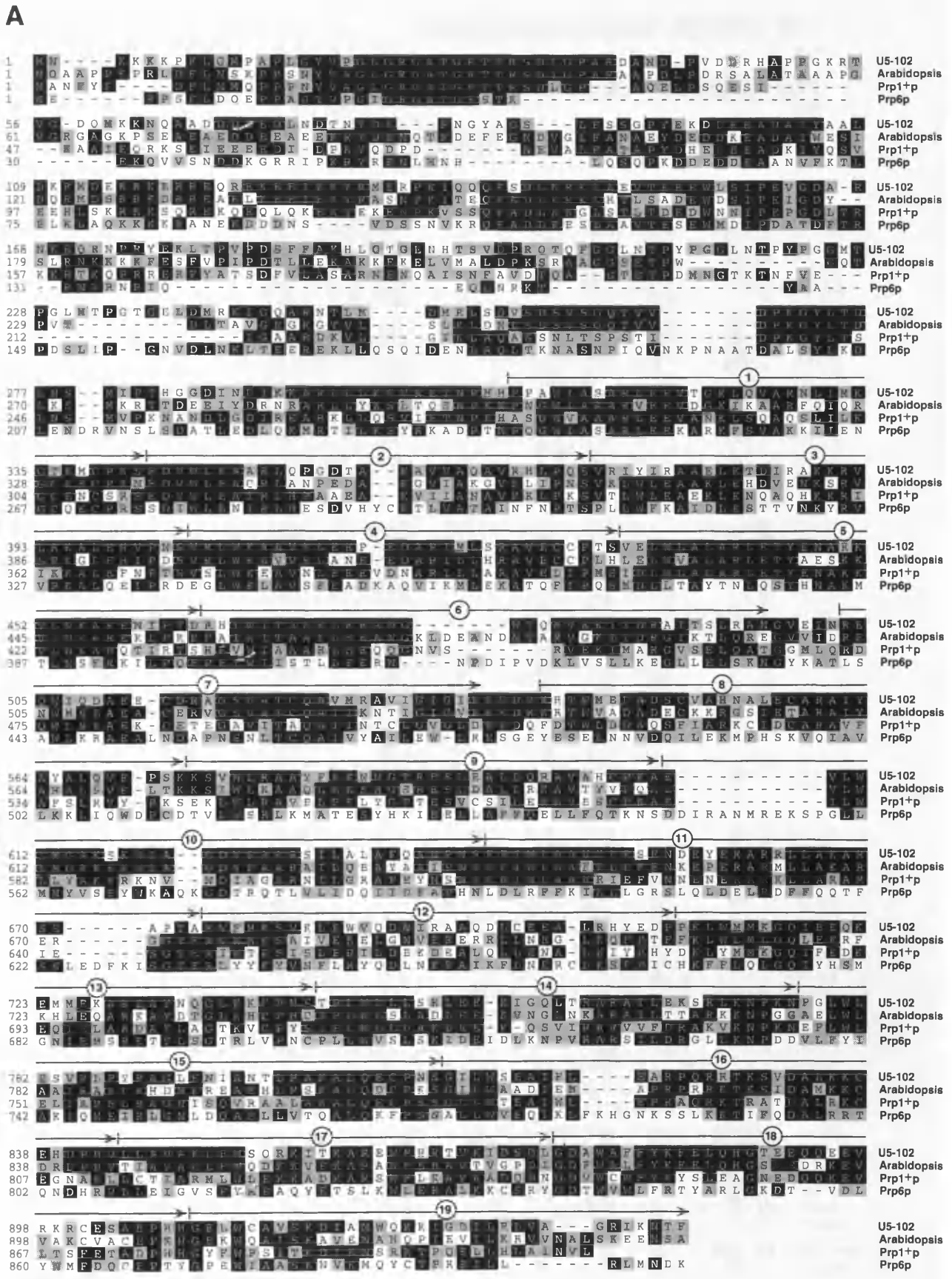


Figure 4.9. (A) Sequence analysis of the human U5-102kD protein.

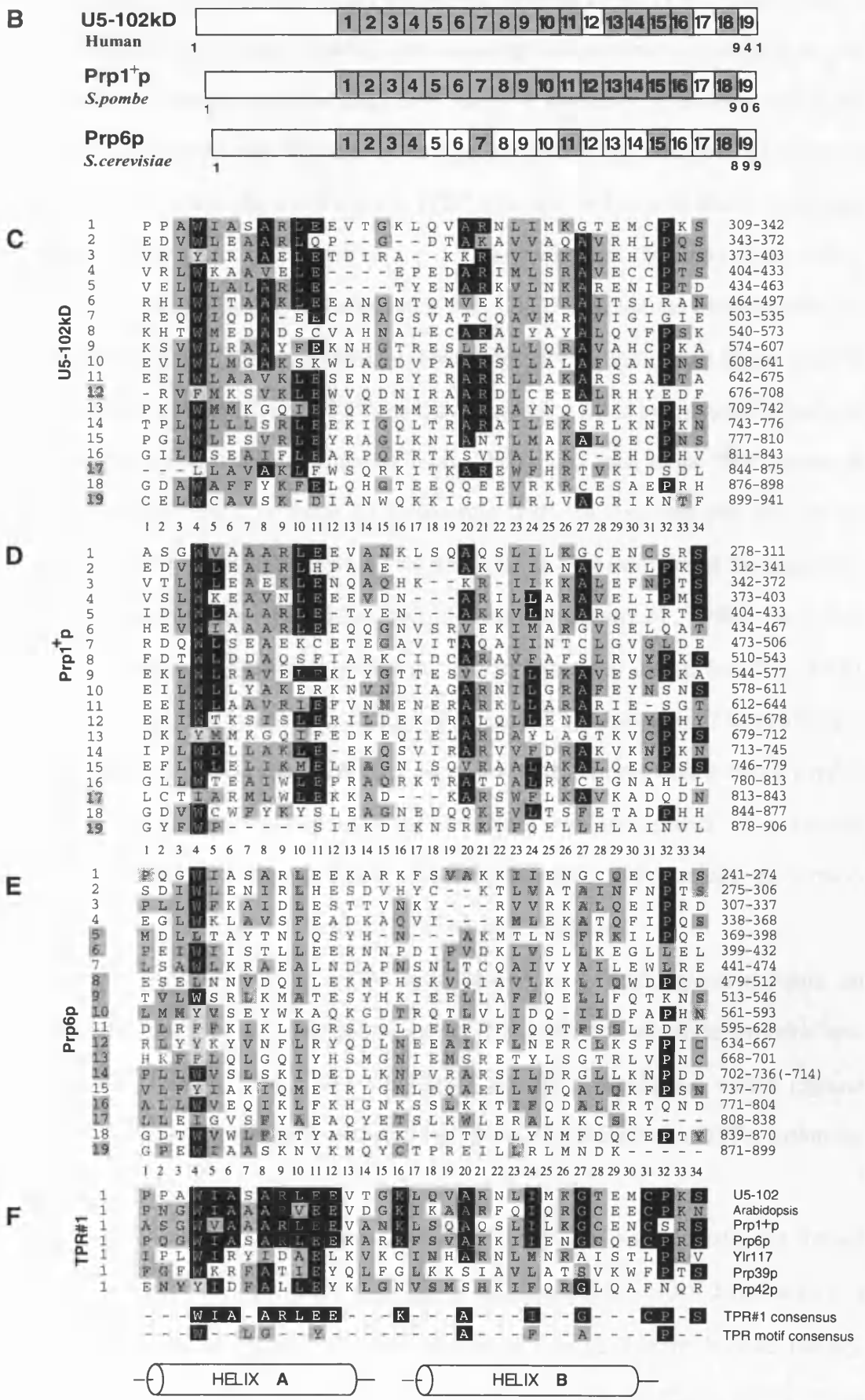


Figure 4.9. (B, C, D, E, F) Sequence analysis of the human U5-102kD protein.

(Ala 763) for a bulky polar threonine at the conserved position 27 of TPR15 might disrupt proper protein folding. There is accumulating data showing that proteins with multiple copies of TPRs function as scaffolding proteins and co-ordinate the assembly of proteins into multi-subunit complexes. Protein-protein interactions in these complexes are thought to be mediated by individual TPRs. For example, a mutation in TPR7 of Cdc27 reduced its ability to interact with Cdc23, but did not affect interaction with Cdc16 or with itself (Sikorski *et al.*, 1990), suggesting that TPR7 is involved in the interaction with Cdc23 whereas the other interactions (Cdc27-Cdc16, Cdc27-Cdc27) are mediated by other domains of Cdc27, most likely by other TPRs. A detailed analysis of individual TPRs of the proteins known to be involved in splicing revealed that the first TPRs share much higher mutual homology than to the TPR consensus only (Fig. 4.9F), suggesting that these TPRs interact with the domains that are similar amongst interacting partners of these proteins. Recently, the crystal structure of the triple TPR domain of protein phosphatase 5 (PP5) has been determined (Das *et al.*, 1998). Each TPR motif comprised a pair of antiparallel α-helices, termed helices A and B (see Fig. 4.9F), associated together with a packing angle of approximately 24° between axes. The structure of each TPR motif is virtually identical. Three consecutive TPRs are organised into a parallel arrangement, such that sequentially adjacent α-helices are antiparallel in a manner reminiscent of a concertina. Knowing the structure of the triple TPR domain, the authors constructed a model of 12 tandem TPR motifs, which would be folded
into a right handed super-helical structure with continuous helical groove suitable for accommodation of α-helix of an interacting partner. The predicted structure explains how proteins containing multiple TPRs, like U5-102kD and its homologues, would interact simultaneously with several target proteins, utilising specific combination of TPRs within the super-helix (Das *et al.*, 1998).

Comparing other structural features of U5-102kD and its homologues, the former contains a canonical leucine zipper pattern (L_{745}-$(X)_6$-L_{752}-$(X)_6$-L_{759}-$(X)_6$-L_{766}) which is present in many transcription factors (Jun/AP1 family and C-myc, L-myc, N-myc and Fos oncogenes; reviewed by Landschulz *et al.*, 1988; Busch and Sassone-Corsi, 1990). Prp6p contains a domain (-C_{268}-...-C_{271}-...-H_{285}-...-H_{290}-), reminiscent of zinc finger structures found in the transcription factor TFIIIA, Prp11p and the human U1C protein of U1 snRNP

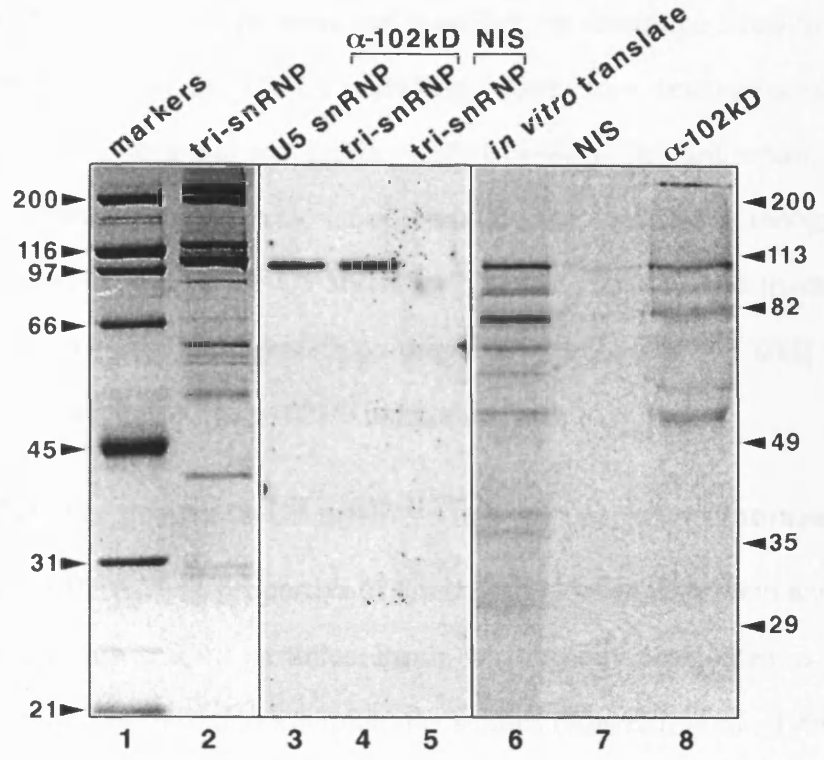


Figure 4.10. Authenticity of the U5-102kD cDNA. Various protein samples were subjected to 10% SDS-PAGE. The gel was cut and different parts were processed as follows. Left panel: marker proteins (lane 1) and tri-snRNP proteins (lane 2) were visualised by Coomassie staining. The molecular weights of the marker proteins are shown on the left. Middle panel: U5 proteins (lane 3) and tri-snRNP proteins (lanes 4 and 5) were transferred to the membrane and immunostained with α-102kD or pre-immune serum (NIS) as indicated. Right panel: the U5-102kD cDNA was translated *in vitro* and proteins were immunoprecipitated and visualised by fluorography. Lane 6 shows the *in vitro* translate added to the reaction, and lane 7 and 8 - those proteins precipitated by pre-immune serum and α-102kD, respectively. The positions of pre-stained molecular weight markers (Bio-Rad) are indicated on the right.

(discussed by Legrain and Choulika, 1990). However, the homologues do not conserve both motifs. Moreover, the leucine zipper of U5-102kD is immersed completely in TPR14, and none of mutations in a zinc knuckle of Prp6p led to its functional impairment (Legrain and Choulika, 1990). Therefore, it is possible that the spacing of amino acids characteristic of both motifs is fortuitous in these proteins and is unlikely to determine a functional property.

The authenticity of the cDNA encoding 102kD was demonstrated by raising an antiserum against the C-terminal part (amino acids 614-941). This antiserum, but not the pre-immune serum derived from the same rabbit, strongly and specifically recognised the native 102kD protein present in purified U5 snRNPs (Fig. 4.10, lane 3) and tri-snRNPs (lane 4). Moreover, the *in vitro*-translated protein co-migrates with the native 102kD protein (lane 6) and is efficiently precipitated by α-102kD antibodies (lane 8).

4.2.2.2. U5-102kD is bound to U5 snRNPs in a salt resistant manner

The main difference in properties of the human U5-102kD protein and yeast Prp6p is their localisation within snRNP particles. Prp6p was initially designated as U4/U6-snRNP-specific protein based on immunoprecipitations studies (Abovich *et al.*, 1990; Galisson and Legrain, 1993), whereas the 102kD protein was defined as a component of isolated 20S U5 snRNPs (Bach *et al.*, 1989). Therefore, we were interested to elucidate the interactions that were responsible for the binding of the 102kD protein to either snRNP. First, we looked at the stability of 102kD association with the U5 snRNP by increasing the ionic strength of the solution. In the experiment depicted in Figure 4.11, the fractions containing 25S [U4/U6.U5] tri-snRNPs were pelleted by centrifugation, resuspended in buffer containing 0.7 M NaCl and separated by glycerol gradient centrifugation at the same salt concentration. Under such conditions the [U4/U6.U5] tri-snRNPs dissociate almost completely into U4/U6 and U5 particles (Fig. 4.11E). Protein analysis demonstrates that at least two proteins in the 100 kDa range sediment as particle-free proteins on the top of the gradient, while the other proteins co-migrate with U5 snRNA (Fig. 4.11A). Western blot demonstrates that the 102kD protein stays entirely associated with the U5 particle (Fig. 4.11B) as well as U5-116kD (Fig.4.11D), whereas the 100kD protein is dissociated from the U5 particle (Fig. 4.11C). This experiment demonstrates that the 102kD protein is very strongly associated with the U5 snRNP particle,

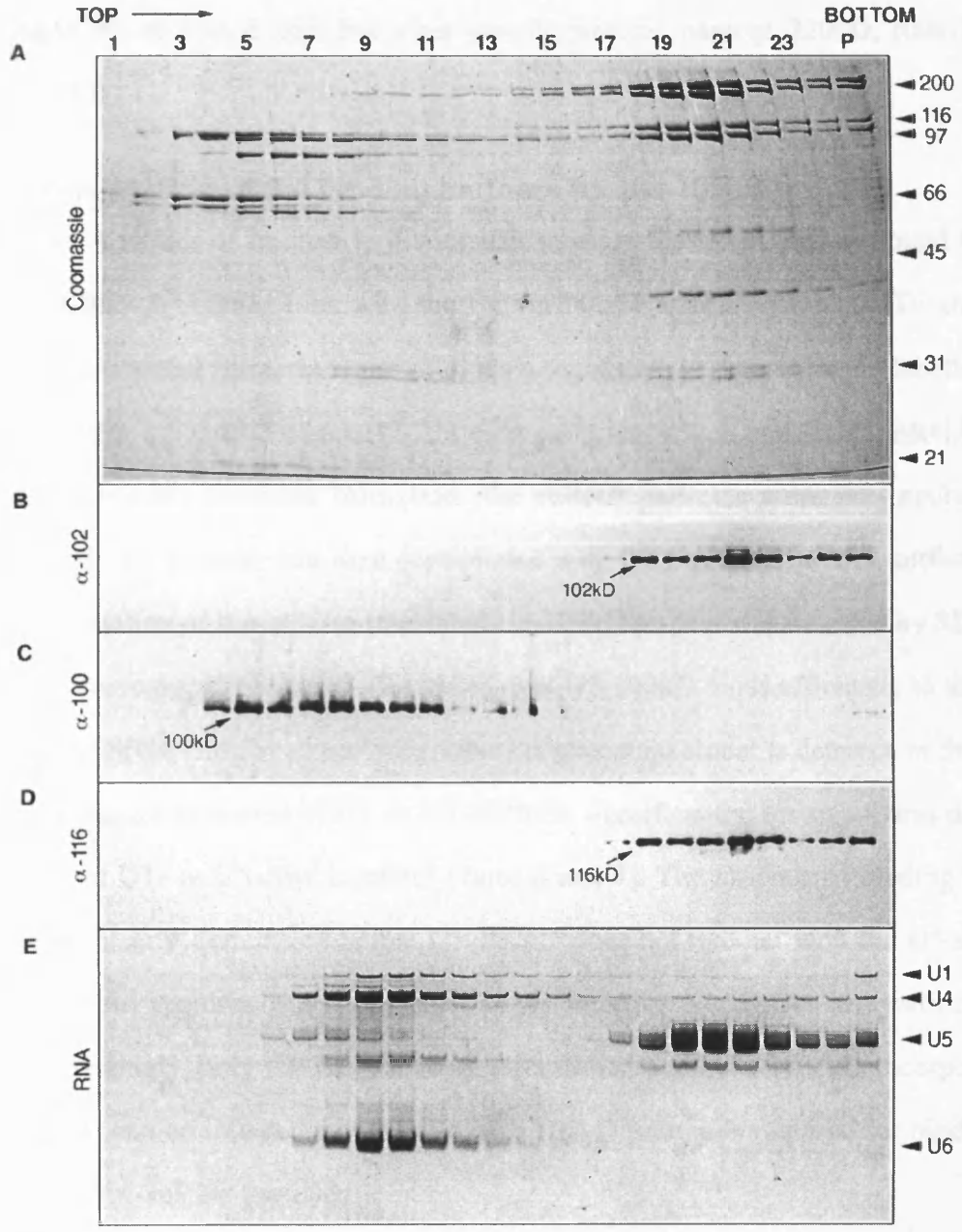


Figure 4.11. Dissociation of [U4/U6.U5] tri-snRNPs in the presence of 700 mM NaCl. Glycerol gradient purified [U4/U6.U5] tri-snRNPs (see Fig. 4.2) were pelleted by centrifugation in a Beckman TLA 100.3 rotor for 6 hours at 70,000 rpm. Pellets were resuspended in buffer Na-700, containing 700 mM NaCl. 200 µg (400 µl) of tri-snRNPs were applied to a linear 4 ml, 5-20% glycerol gradient in buffer Na-700 and centrifuged in a Beckman SW60 rotor for 15 h 20 min at 34,000 rpm. Twenty-four fractions of 175 µl each were harvested manually from top to bottom, and the pellet (P) was resuspended in 150 µl of buffer Na-700 containing 1% SDS. Aliquots from each fraction were analysed as follows. **A:** Proteins were fractionated by 10% SDS-PAGE and visualised by Coomassie staining. The molecular weight markers are indicated on the right. **B, C, D:** Proteins were fractionated as in (**A**), transferred to a Hybond-P membrane and visualised by immunostaining with α-102kD, α-100kD and α-116kD antibodies, respectively. **E:** RNA was phenol extracted, ethanol precipitated, and fractionated by 10% urea-PAGE and visualised by silver staining. The positions of the snRNAs are shown on the right.

resembling in this respect at least five other specific proteins, namely, 220kD, 200kD, 116kD, 52kD and 40kD.

4.2.2.3. Investigation of the binding partners for the 102kD protein

It was a matter of interest to distinguish whether the U5-102kD is bound to the U5 particle primarily by interactions with the U5 snRNA or with U5 proteins. To answer this question, we compared the association of *in vitro*-translated, [^{35}S]methionine-labelled 102kD protein with 20S U5 snRNPs and 10S U5 core particles, which consist of snRNA and Sm proteins. To monitor complex formation, the snRNP particles were pre-incubated with labelled U5-102kD protein, and then precipitated with the Sm-specific Y12 antibodies. Co-immunoprecipitation of the *in vitro*-translated U5-102kD protein was detected by SDS-PAGE followed by fluorography. Figure 4.12A shows that U5-102kD binds efficiently to the 20S U5 snRNPs and 25S[U4/U6.U5] tri-snRNPs, whereas almost no signal is detected in the absence of snRNPs or in the presence of U1 or U2 snRNPs. Significantly, no signal was detected in the presence of U1- or U5-core particles (lanes 6 and 7). The absence of binding to the U5 core particles clearly demonstrates that U5-102kD does not interact with the U5 snRNA or Sm proteins, but requires other U5-specific proteins for successful integration into the particle. Interestingly, only the longest *in vitro*-translated product of 102kD incorporates into the particles. It can be inferred that a full-length 102kD protein is required for binding to the U5 snRNP and tri-snRNP particles.

The RNA binding activity was further investigated in the series of experiments employing the *in vitro*-translated 102kD protein and *in vitro*-transcribed snRNAs or MINX pre-mRNA. The binding of the 102kD protein to either RNA species was detected by co-immunoprecipitation with RNA precipitated via an m^{7}G cap, introduced during *in vitro* transcription. As with the 61kD protein (4.2.1.4), no significant binding was detected (data not shown). Therefore, the conclusion was drawn that the U5-102kD protein does not exhibit any RNA-binding activity and must be integrated into the U5 particle through contacts to other U5-specific proteins.

The absence of interaction between the 102kD protein and U5 core particles persuaded us to search for a binding partner among the specific proteins which are also stably associated with the U5 particle. It was recently shown that the U5-220kD protein interacts with the U5-

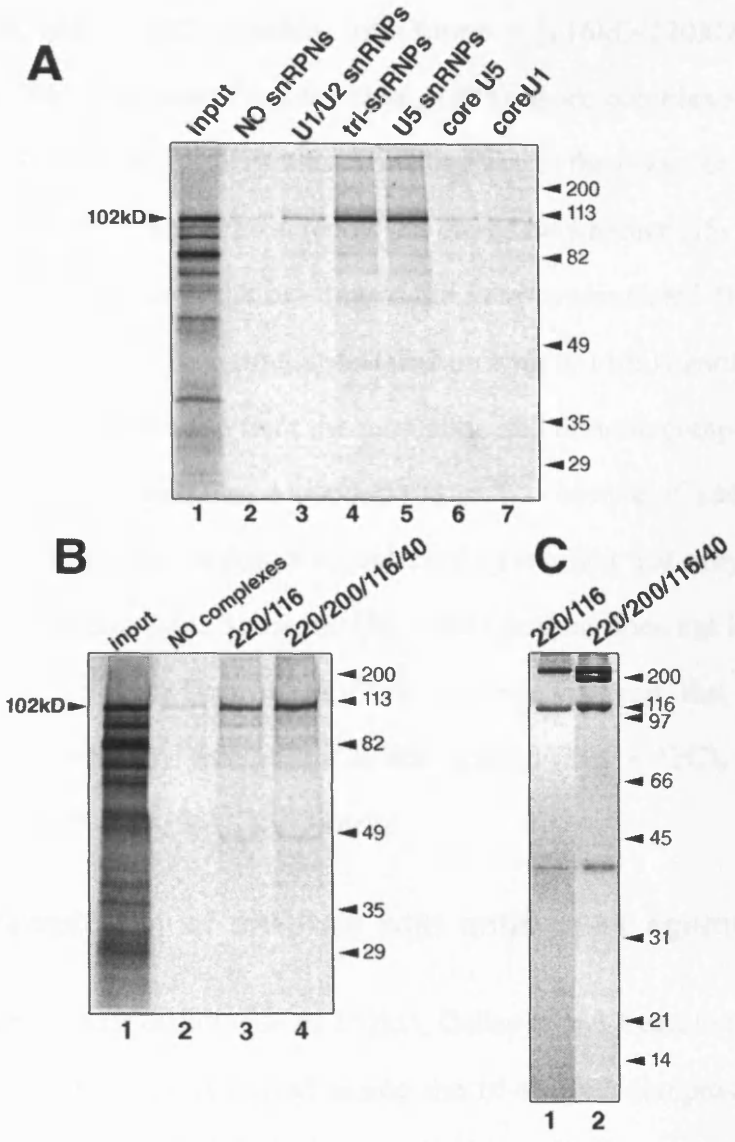


Figure 4.12. Interaction of U5-102kD with snRNPs and protein complexes. A: Specific association of the U5-102kD protein with the U5 and [U4/U6.U5] tri-snRNPs. *In vitro* translated, [^{35}S]-labelled 102kD protein was incubated with a mixture of U1 and U2 snRNPs (lane 3), tri-snRNPs (lane 4), U5 snRNPs (lane 5), U5 core particles (lane 6) or U1 core particles (lane 7). The snRNPs were immunoprecipitated with Y12 antibodies, and co-precipitated proteins were analysed by 10% SDS-PAGE followed by fluorography. Lane 1 shows 20% of the input material added to each immunoprecipitation. The position of pre-stained molecular weight markers is indicated on the right. **B:** The U5-102kD protein associates with the RNA-free [220/116] and [220/200/116/40] complexes. *In vitro* translated, [^{35}S]-labelled 102kD protein was incubated with fractions containing purified [116kD/220kD] dimer (lane 3) or [40kD/116kD/200kD/220kD] tetramer (lane 4). Complexes were immunoprecipitated with α-116kD antibody and co-precipitated proteins were analysed by 10% SDS-PAGE followed by fluorography. Lane 1 represents 20% of the *in vitro* translate added to each immunoprecipitation reaction. Lane 2 is a control for co-precipitation of the 102kD protein in the absence of additional proteins. **C:** The protein composition of the complexes used in (**B**) was analysed by 10% SDS-PAGE and visualised by Coomassie staining. The position of molecular weight markers is indicated on the right.

specific 40kD, 200kD, and 116kD proteins, and forms a [116kD/220kD] dimer and a [40kD/116kD/200kD/220kD] tetramer (Achsel *et al.*, 1998). Both complexes are RNA-free and could be isolated by glycerol gradient ultracentrifugation in the presence of a chaotropic agent, sodium thiocyanate, of different stringency. To elucidate whether U5-102kD interacts with these complexes, we monitored the binding of the *in vitro*-translated 102kD protein to the isolated tetramer and dimer by co-immunoprecipitation with α-116kD antibody. As shown in Figure 4.12B, U5-102kD is bound to both the tetrameric and dimeric complex (lanes 3 and 4), whereas no background precipitation was observed in the absence of additional proteins (lane 2). The specificity of the interaction is accentuated by the fact that only the full-length U5-102kD protein is co-precipitated. Since the U5-102kD protein does not bind to *in vitro*-translated U5-40kD or U5-116kD proteins (data not shown) and the preparation of [116kD/220kD] dimer is virtually free of the 200kD protein (Fig. 4.12C), it is likely that 102kD interacts directly with the U5-220kD protein.

4.2.2.4. Immunoprecipitaion of snRNAs with antibodies against the 102kD protein

With regard to the yeast orthologue of 102kD, Galisson and Legrain (1993) proposed that the C-terminal part of Prp6p is buried inside the tri-snRNP complexes because co-immunoprecipitation of snRNAs with polyclonal α-Prp6p antibodies, raised against the C-terminal domain of the protein (amino acids 639 to 899), required the presence of at least 0.02% SDS (Legrain and Choulika, 1990). Also taking into account that the TPR proteins often interact simultaneously with several proteins, it might be anticipated that U5-102kD should be able to bind to other proteins in addition to the [116kD/220kD] dimer, and possibly, to components of the U4/U6 particle. To test this idea, we first tried to co-immunoprecipitate the [U4/U6.U5] tri-snRNPs from HeLa nuclear extracts with the antibodies directed against the C-terminal domain of 102kD (amino acids 614-941). Figure 4.13 demonstrates that this serum does efficiently precipitate U5 snRNAs from nuclear extracts, but not tri-snRNPs (lanes 3, 6, 9). It should be noted that the pCp-labelling procedure does not visualise U6 snRNA and, therefore, only U4 snRNA is evidence for U4/U6 or tri-snRNP precipitation; the presence of U6 snRNA was confirmed by northern blot (data not shown). In contrast, the antiserum against the U5-116kD protein precipitated both U5 and tri-snRNPs at low and

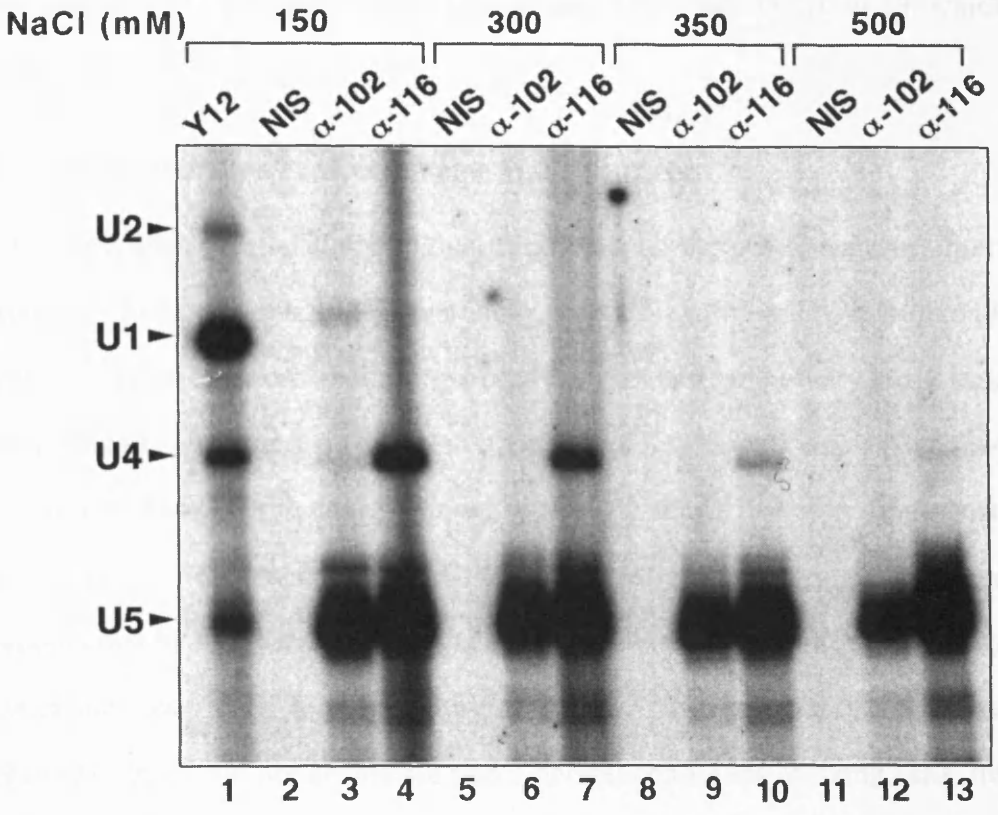


Figure 4.13. Co-immunoprecipitation of snRNAs from HeLa nuclear extract with α-102kD. Immunoprecipitations from HeLa nuclear extract with pre-immune serum (NIS: lanes 2, 5, 8, 11) or with α-102kD (lanes 3, 6, 9, 12) were performed at different salt concentrations as indicated in the figure. Y12 antibody (lane 1) and α-116kD (lanes 4, 7, 10, 13) were used as positive controls. Precipitated RNAs were 3' end-labelled with [^{32}P]-pCp and analysed by 10% urea-PAGE. The positions of the snRNAs are indicated on the left.

moderate salt concentrations (lanes 4, 7, 10), while no tri-snRNPs were observed at 500 mM NaCl, because the tri-snRNP dissociates under these conditions (lanes 13). The failure of α-102kD antibodies to precipitate tri-snRNPs under native conditions is consistent with inaccessibility of the C-terminal domain (comprising TPR elements 10 to 19, which served as the antigen).

4.2.2.5. Subcellular localisation of the 102kD protein

The data from cellular immunolocalisation studies support the above observation that the epitope for α-102kD antibodies is not easily accessible in tri-snRNPs. In fixed HeLa cells, the 102kD antiserum showed no staining of the nucleus and only a very weak staining of the cytoplasm which is characteristic of a non-specific binding of antibodies (Fig. 4.14B). When the cells, immobilised on the cover slip, were treated with SDS before immunostaining, the nucleus was stained very strongly (Fig. 4.14C). The uneven staining pattern coincided largely with that obtained by anti-Sm antibodies (Fig. 4.14D) with the main difference that the coiled bodies were not specifically highlighted by α-102kD. This result demonstrates that the TPRs 10 to 19 of U5-102kD are not accessible under normal conditions in living cells. Interestingly, this also might indicate that almost all U5 snRNPs are in the [U4/U6.U5] tri-snRNP complexes, or alternatively, the remaining free U5 particles are engaged in interactions with other components of nucleus through U5-102kD.

In agreement with the immunoprecipitation and immunostaining, attempts to inhibit splicing *in vitro* (or immunodeplete nuclear extracts) with α-102kD have failed (data not shown). This suggests that the amount of pre-formed [U4/U6.U5] tri-snRNP complexes in nuclear extract is sufficient for *in vitro* splicing.

4.2.3. Investigation of the subunit organisation of [U4/U6.U5] tri-snRNP particles

The mammalian tri-snRNP is a complex ribonucleoprotein particle, encompassing three snRNAs and at least thirty different polypeptides. The structure, stability and function of oligomeric complexes are determined by their constituent subunits and their interactions. The investigations of inter-subunit interactions using a reversible dissociation-association

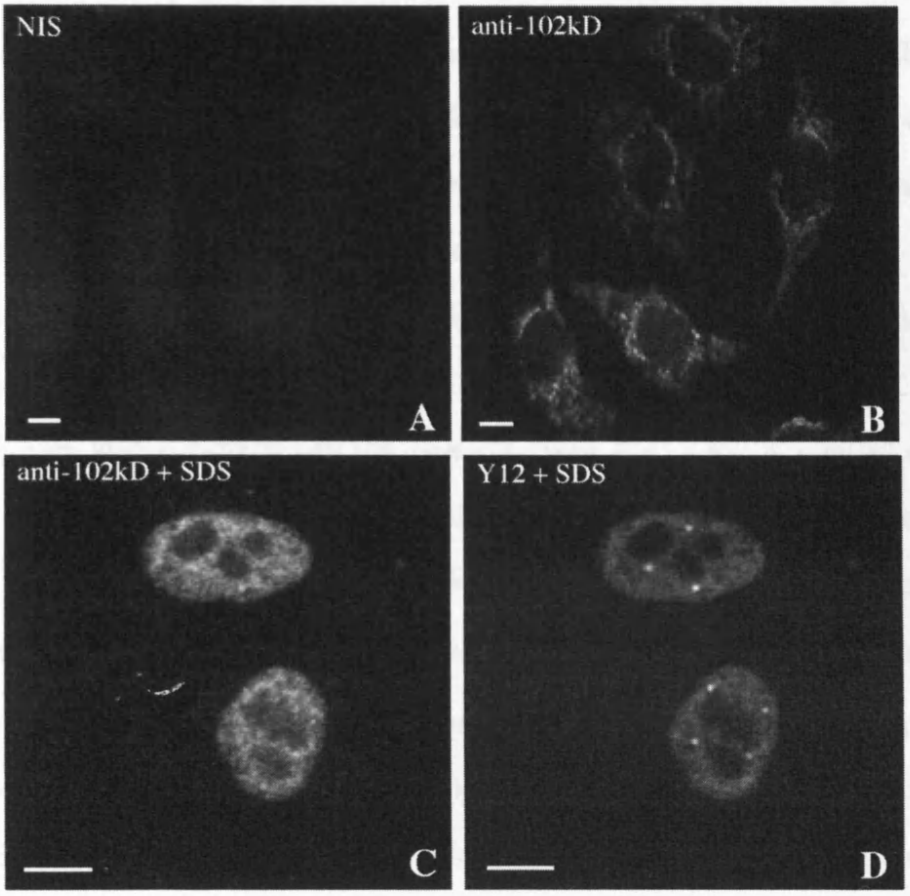


Figure 4.14. Subcellular localisation of U5-102kD in HeLa cells. HeLa cells were immobilised on glass coverslips and immunostained with non-immune serum (**A**) or with antiserum against U5-102kD (**B**). In panels C and D, the fixed cells were treated with 1% SDS for 10 min prior to immunostaining and then double stained with α-102kD (**C**) and Y12 (**D**) antibodies. The images in panels C and D were recorded independently using appropriate filters to separate the red fluorescence from Cy3 (**C**) and the green fluorescence from FITC (**D**). The bar indicates 10 μm.

approach could add to our understanding of how these complexes interact with each other. We attempted to dissociate [U4/U6.U5] tri-snRNPs into subunits that re-associate.

Chemical and physical treatments have been used to dissociate various complex enzymes into subunits. Many procedures make use of the differential solubility of distinct subunits under specific conditions. Changes in temperature, pH, ionic strength, or the addition of denaturants or specific "co-factors" have been used for the successful dissociation of many multisubunit complexes (discussed by Einstein and Schachman, 1989).

It is commonly accepted, based on immunoprecipitations from nuclear extracts that with the increase of ionic strength of the solution, tri-snRNPs dissociate into a U5 snRNA-containing part and a U4/U6 snRNA complex (see Fig. 4.6). A similar approach, the variations of ionic strength, was applied to the affinity-purified snRNPs ten years ago (Bach *et al.*, 1989; Behrens and Lührmann, 1991). It was observed using density gradient centrifugation that total snRNPs eluted from the anti-m3G column at 420 mM KCl are not able to form the [U4/U6.U5] tri-snRNP complexes at reduced salt concentration (150 mM KCl), at which in nuclear extracts almost all U4/U6 snRNAs are found associated with U5 snRNPs (Behrens and Lührmann, 1991). Alternatively, MonoQ chromatography of snRNPs allowed isolation of the 10S U4/U6 snRNPs, containing snRNAs and Sm proteins, as well as the 20S U5 snRNPs, containing a set of specific proteins in addition to the U5 snRNA and Sm proteins (Bach *et al.*, 1989). Importantly, in an *in vitro* reconstitution assay the purified 20S U5 and 10S U4/U6 snRNPs could not form a complex, unless free tri-snRNP-specific proteins (micrococcal nuclease treated [U4/U6.U5] tri-snRNPs) were added (Behrens and Lührmann, 1991). Thus, it was demonstrated that the association between U5 and U4/U6 snRNPs requires specific proteins.

The existence of some sub-complexes inside the [U4/U6.U5] tri-snRNP has been already demonstrated. This is the Sm core, comprising seven polypeptide chains, which assembled in a compact doughnut-like structure and occupied an Sm site on the U5 and U4 snRNAs (Kambach *et al.*, 1999). Recently, the LSm (Like Sm) proteins were isolated from tri-snRNPs as a complex similar to the canonical Sm proteins (Achsel *et al.*, 1999; Mayes *et al.*, 1999). This LSm complex binds to the U-rich site at the 3' end of U6 snRNA and forms doughnut-like structures as it was visualised by electron microscopy (Achsel *et al.*, 1999).

Stable RNA-free protein complexes, a dimer [220kD/116kD] and a tetramer [220kD/200kD/116kD/40kD] (see Fig. 4.12) were isolated from U5 snRNPs using differential treatment with sodium thiocyanate (Achsel *et al.*, 1998). A stable complex of three tri-snRNP specific proteins, [90kD/60kD/20kD], was isolated from a nuclear extract (Horowitz *et al.*, 1997), and is thought to be an integral part of U4/U6 snRNP (Lauber *et al.*, 1997; Teigelkamp *et al.*, 1998). However, the isolation of a U4/U6 snRNP particle that is active in tri-snRNP formation has not been successful.

4.2.3.1. Isolation and characterisation of the 14S U4/U6 snRNP particles

In contrast to the studies described above, I was not limited by the amount of material available and could investigate the effect of the increase of ionic strength on tri-snRNPs in more detail. The [U4/U6.U5] tri-snRNP particles were purified by a glycerol gradient from the total snRNP preparation obtained by affinity chromatography at low salt (250 mM).

To study the dissociation of tri-snRNPs, particles from the fractions around 25S (Fig. 4.2, fractions 21 to 25) were pelleted by ultracentrifugation, resuspended in a buffer of desired salt concentration, and then separated on a glycerol gradient with the same salt concentration. An experiment done in the presence of 700 mM NaCl is depicted in Figure 4.11. At this salt concentration the tri-snRNPs dissociated completely, as judged by the RNA gel (lower panel). As discussed above (4.2.2.2.), the U5 snRNA is still associated with a number of specific proteins, namely, 220kD, 200kD, 116kD, 102kD, 52kD, 40kD. In contrast, the U4/U6 snRNAs do not co-fractionate with any of the specific proteins and associate only with Sm core proteins. The association of canonical Sm proteins with the Sm site of snRNAs is inherently strong and resists, *e.g.*, CsCl centrifugation. These U4/U6 particles are similar to the 10S U4/U6 complexes described in Bach *et al.* (1989) and are characterised below (Fig. 4.17). Taking into account that at 150 mM NaCl all U4/U6 snRNAs are associated with U5 in the tri-snRNP complex but at 700 mM this complex is completely dissociated, different salt concentrations were examined in order to find the conditions when U4/U6 snRNAs could be separated in a complex with U4/U6-specific proteins. It was observed that, with decreasing salt concentration, the percent of intact tri-snRNPs increased, but unexpectedly, the U4/U6

snRNAs released from tri-snRNPs were always found in the 10S form, essentially without any specific proteins. Therefore, another approach was tried.

As an alternative, a known chaotropic agent, NaSCN, was used to destabilise tri-snRNPs. This solvent is very effective and causes dissociation of complex enzymes with little unfolding of the polypeptide chains (Einstein and Schachman, 1989). The effect of differential NaSCN treatment on U5 snRNPs has been extensively studied and resulted in isolation of stable RNA-free protein complexes (Achsel *et al.*, 1998). A similar strategy was applied, and conditions were selected for NaSCN treatment and density gradient centrifugation that, finally, resulted in isolation of the U4/U6 snRNAs in the complex with specific proteins. The procedure is described below.

The [U4/U6.U5] tri-snRNPs (fractions 21 to 25, Fig. 4.2) were pelleted by ultracentrifugation, resuspended in buffer with 0.2 M NaSCN and loaded on 5-20% glycerol gradient without NaSCN but containing 150 mM NaCl. The conditions of centrifugation (RPM and time) were selected empirically to separate complexes with sedimentation coefficients between 10S and 20S. Figure 4.15 shows analysis of the protein and RNA fractions from a gradient. Fractions 14-16 contain both the proteins (90kD, 61kD, 60kD, 20kD, B/B'), visualised by the 10% Coomassie-stained gel, and the U4/U6 snRNAs, revealed by northern blot. The presence of low molecular weight proteins that are not separated by 10% SDS-PAGE was revealed by 13% SDS-PAGE followed by western blot. The identity of 90kD, 61kD, and 60kD was also confirmed by western blot. Thus, the following proteins were found to co-migrate on a gradient with U4/U6 snRNAs: the Sm proteins (data not shown, but B/B' are seen on the Coomassie-stained gel), the Sm like proteins (LSm4), 15.5kD, 20kD, 60kD, 61kD, and 90kD.

It is important to note that most of the material applied on a gradient was found in the translucent pellet at the bottom of a centrifuge tube. This pellet was analysed by standard glycerol gradient centrifugation (as in Fig. 4.2) and it apparently contained [U4/U6.U5] tri-snRNPs. Therefore, only small fraction of U4/U6 snRNPs was retarded on the gradient during centrifugation. Proteins with apparent molecular masses of 110kD, 100kD, 65kD and 27kD are found on the top of the gradient. The loading material contained 0.2 M NaSCN, which

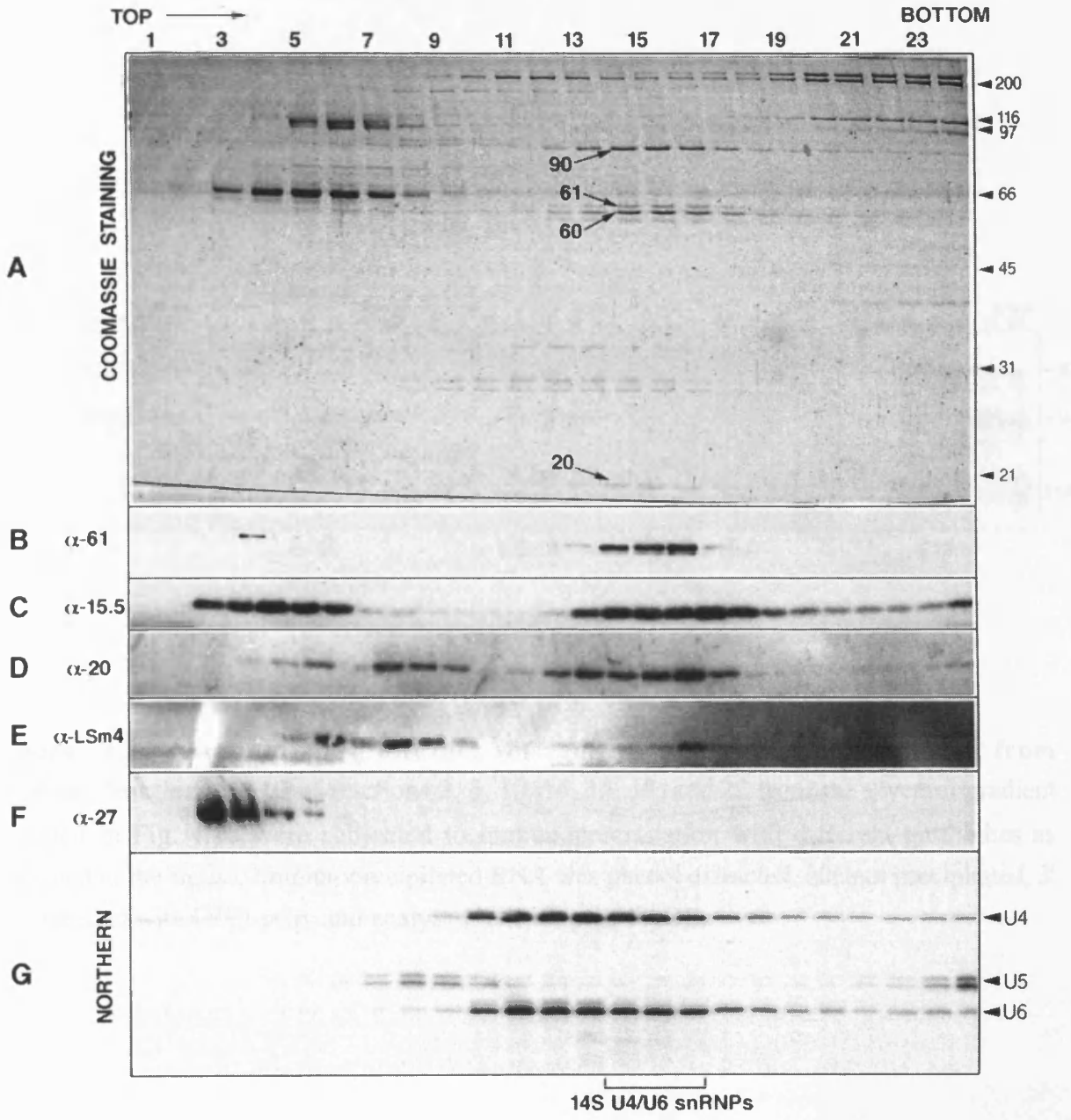


Figure 4.15. Isolation of the 14S U4/U6 snRNP. Glycerol gradient purified [U4/U6.U5] tri-snRNPs (see Fig. 4.2) were pelleted by centrifugation at 4°C in a Beckman TLA 100.3 rotor for 6 hours at 70,000 rpm. Pellets were resusupended in buffer SCN-200. 200 µg (400 µl) of tri-snRNPs were layered onto a linear 4 ml, 5-20% glycerol gradient containing buffer Na-150, and centrifuged at 4°C in a Beckman SW60 rotor for 14 h 25 min at 35,000 rpm. Twenty-four fractions of 175 µl were harvested manually from top to bottom. Aliquots from each fraction were analysed as follows. **A:** Proteins were separated by 10% SDS-PAGE and visualised by Coomassie staining. The molecular weight markers are indicated on the right. **B, C, D, E, F:** Proteins were processed as in (A), blotted to a Hybond P membrane and visualised by immunostaining with α-61kD, α-15.5kD, α-20kD, α-LSm4 and α-27kD antibodies, respectively. **G:** RNA was phenol extracted, ethanol precipitated, and fractionated by 10% urea-PAGE. The U4, U5 and U6 snRNAs were visualised by northern hybridisation with specific probes. The positions of the snRNAs are shown on the right.

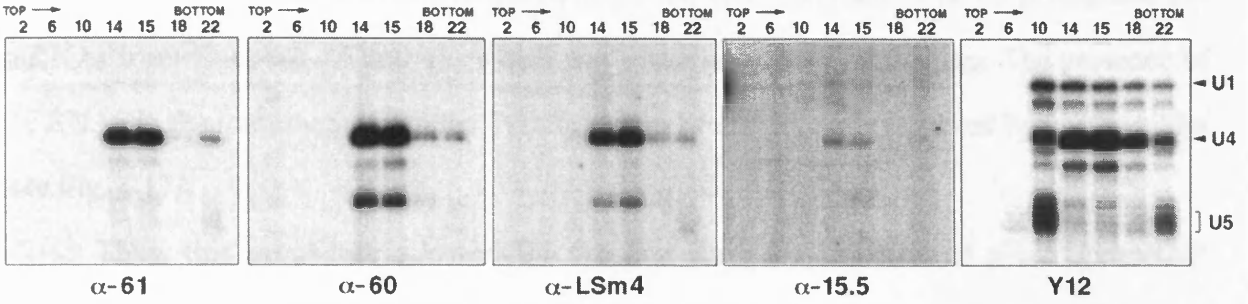


Figure 4.16. Immunoprecipitation of 14S U4/U6 snRNP particles from gradient fractions. 40 µl of fractions 2, 6, 10, 14, 15, 18, and 22 from the glycerol gradient depicted in Fig. 4.15 were subjected to immunoprecipitation with different antibodies as indicated in the figure. Immunoprecipitated RNA was phenol extracted, ethanol precipitated, 3' end-labelled with [^{32}P]-pCp, and analysed by 10% urea-PAGE.

penetrated up to the fraction 10 during centrifugation as it was determined by absorption at 260 nm.

The co-migration of U4/U6 snRNAs with a number of proteins (90kD, 61kD, 60kD, 20kD, 15.5kD, Sm and LSm) on the gradient indicates that they form a complex. The existence of the snRNP particle was proven by immunoprecipitations from gradient fractions. Figure 4.16 shows that α-61kD, α-60kD, α-LSm4, α-15.5kD and Y12 co-precipitate U4 snRNAs from fractions 14 and 15, which was visualised by pCp-labelling. The presence of U6 RNAs in the immunoprecipitates from fractions 14 and 15 was confirmed by northern blot (see Fig. 4.17).

Thus, this procedure allowed for the first time the isolation of a U4/U6 snRNP particle, containing U4/U6-specific proteins. To estimate the sedimentation coefficient (S value) for U4/U6 snRNPs, a mixture of 5S, 16S and 23S rRNAs extracted from *E. coli* 70S ribosomes (kindly provided by Dr. R. Brimacombe, MPI, Berlin) was subjected to the identical glycerol gradient centrifugation and served as markers. The S-value is a linear function of a migration distance on a linear gradient (Young, 1984). Thus, fractions 14 and 15, containing U4/U6 snRNPs, were found to correspond to the 14S. Thereafter this particle will be referred to as the 14S U4/U6 snRNP.

4.2.3.2. Reconstitution of the tri-snRNP from the isolated U4/U6 and U5 particles

The MonoQ-purified 10S U4/U6 snRNPs were not able to reconstitute [U4/U6.U5] tri-snRNPs when mixed with U5 snRNPs in the absence of specific proteins (Behrens and Lührmann, 1991). It was interesting to investigate whether the 14S U4/U6 snRNP contains all the proteins required for the association with U5 snRNP. The reconstitution was assayed by determining whether U4/U6 snRNPs co-immunoprecipitate with U5 snRNPs using either U5-specific antibodies directed against the 116kD protein or U4/U6-specific antibodies directed against the 61kD protein. The immunoprecipitation assay was reliable as α-61kD does not precipitate U5 snRNPs whereas α-116kD does not precipitate U4/U6 snRNPs. The U5 snRNPs used for a reconstitution were purified by 5-20% glycerol gradient in the presence of 700 mM NaCl (Fig. 4.11, fractions 20 and 21). This procedure removes cross-contamination of U5 snRNPs with U4/U6 snRNPs but also causes the loss of weakly associated proteins,

110kD, 100kD, and 65kD. Moreover, the preparation of U5 snRNPs obtained by high salt treatment was more stable and less prone to aggregation in comparison to the U5 particles isolated at low salt.

The simple combination of fractions containing the 14S U4/U6 snRNPs and the high salt treated U5 snRNPs was enough to reconstitute tri-snRNPs as both α-61kD and α-116kD could co-immunoprecipitate tri-snRNP complexes (Fig. 4.17A). Figure 4.17B shows that, in contrast to the 14S U4/U6 snRNPs, 10S U4/U6 snRNPs obtained by a glycerol gradient in the presence of 700 mM NaCl (Figure 4.11, fractions 9 and 10) were not able to reconstitute tri-snRNP particles. Panels A and B of Figure 4.17 present co-immunoprecipitated snRNAs visualised by pCp-labelling, which does not detect U6 snRNA. Figure 4.17C demonstrates the northern blot analysis of reconstitution and confirms the association of U6 snRNA. In addition, the protein compositions of the 14S and 10S U4/U6 snRNPs were compared, and apparently, the 14S U4/U6 snRNPs do contain the set of specific proteins that is absent in the 10S U4/U6 snRNPs.

The different conditions that could affect the reconstitution were examined. The association of the 14S U4/U6 snRNPs with U5 snRNPs takes place even at 0°C, and addition of ATP does not change the level of reconstitution (data not shown).

Here, it was demonstrated that the set of the 14S U4/U6 snRNP specific proteins, 90kD, 61kD, 60kD, 20kD, 15.5kD and LSm, is sufficient to promote the association of U4/U6 and U5 particles, but which of them are necessary for reconstitution is a matter of further investigations. On the other hand, the 110kD, 65kD and 27kD tri-snRNP specific proteins as well as the U5-100kD protein are dispensable for tri-snRNP formation, since these proteins are not present in the U5 snRNP obtained by high salt treatment and the 14S U4/U6 snRNP.

4.3. CONCLUSIONS

In this study two novel [U4/U6.U5] tri-snRNP proteins were characterised. The cDNA for the 61kD protein was identified and this protein appeared to be an orthologue of the yeast-splicing factor Prp31p. There are no well characterised domains in the amino acid sequence of the 61kD protein but the homology to the Nop5 (58 and 56 proteins) and SAR-binding protein families may argue that 61kD recognises specific RNA structures and be an RNA binding

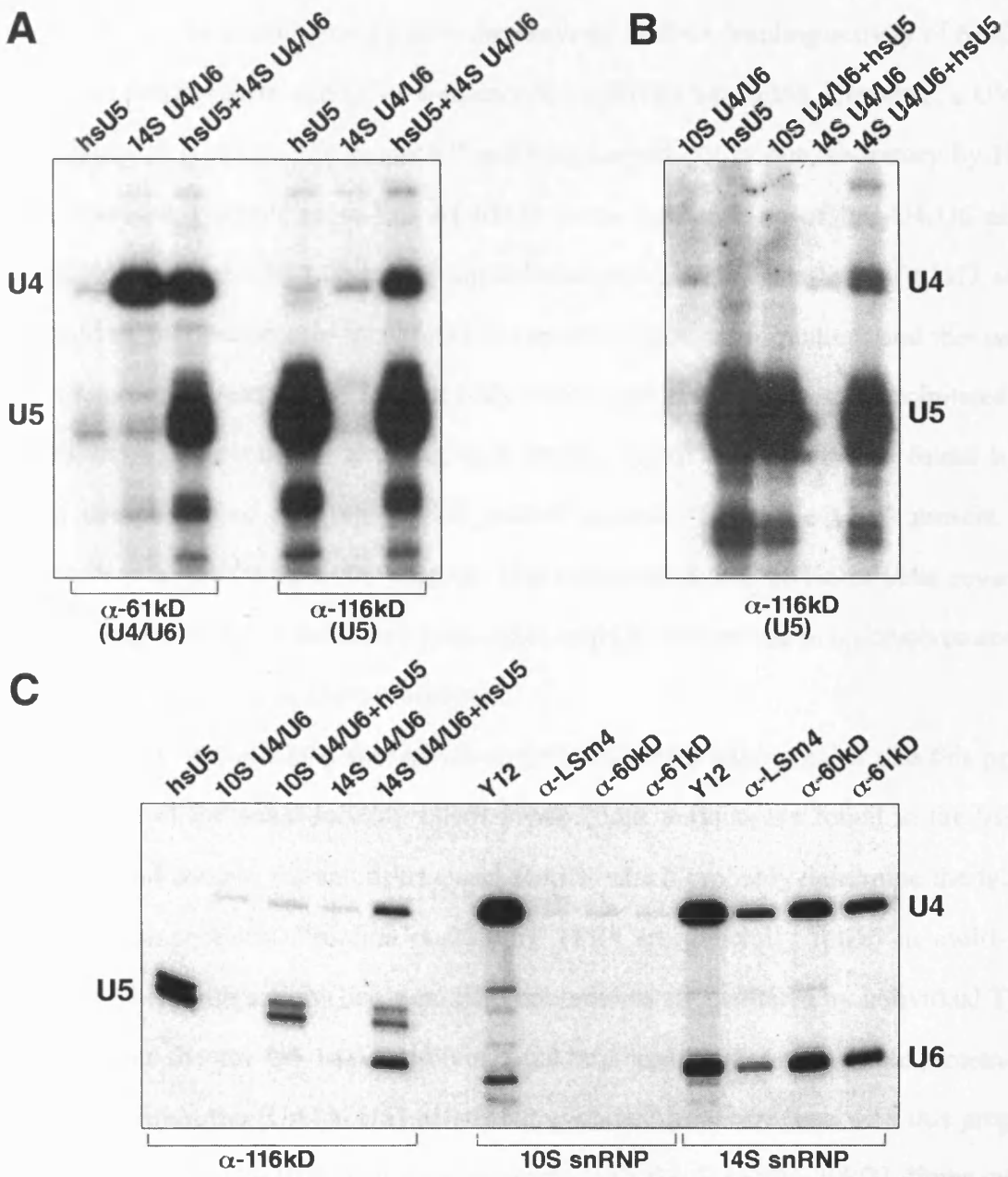


Figure 4.17. Reconstitution of [U4/U6.U5] tri-snRNPs from isolated U4/U6 and U5 snRNP particles. The high salt resistant U5 snRNPs (hsU5) are originated from the gradient fractions 20-21, Fig. 4.11. The 14S U4/U6 snRNPs are from the gradient fractions 14-15, Fig. 4.15. The 10S U4/U6 snRNPs are from the gradient fractions 9-10, Fig. 4.11. The isolated U5 and U4/U6 particles were used in immunoprecipitation either separately or after they were combined and incubated for 15 min at 30°C. **A:** Analysis of association of the high salt resistant U5 snRNP and the 14S U4/U6 snRNP by immunoprecipitation with both U4/U6-specific α-61kD and U5-specific α-116kD antibodies. Precipitated RNA was 3' end-labelled with [^{32}P]-pCp and analysed by 10% urea-PAGE. **B:** Comparison of 10S U4/U6 and 14S U4/U6 snRNP particles on their ability to associate with U5 snRNPs. RNA was precipitated with α-116kD, 3' end-labelled with [^{32}P]-pCp and analysed by 10% urea-PAGE. **C:** 10S U4/U6 and 14S U4/U6 snRNP particles, and their association with U5 snRNPs were analysed by immunoprecipitation followed by northern hybridisation.

protein with a novel motif. An attempt to demonstrate an RNA-binding activity of 61kD using *in vitro*-translated protein and *in vitro*-transcribed snRNAs has failed. However, a UV cross-linking study of [U4/U6.U5] tri-snRNP particles carried out in our laboratory by Henning Urlaub revealed a direct cross-link of 61kD to the central loop of the U4/U6 snRNA's heteroduplex (Urlaub and Lührmann, unpublished data). This cross-link of 61kD could be interpreted as that the protein recognises the specific RNA conformation, and this issue is a matter of current investigation. The antibody raised against 61kD immunoprecipitated U4/U6 snRNAs from HeLa nuclear extracts, and finally, the 61kD protein was found to be an integral component of the 14S U4/U6 snRNP particle. Thus, the 61kD protein can be designated as a U4/U6-specific protein. The immunostaining of HeLa cells revealed an unusual strong staining of the coiled bodies that could be interpreted as an involvement of this protein in the biogenesis of U4/U6 snRNAs.

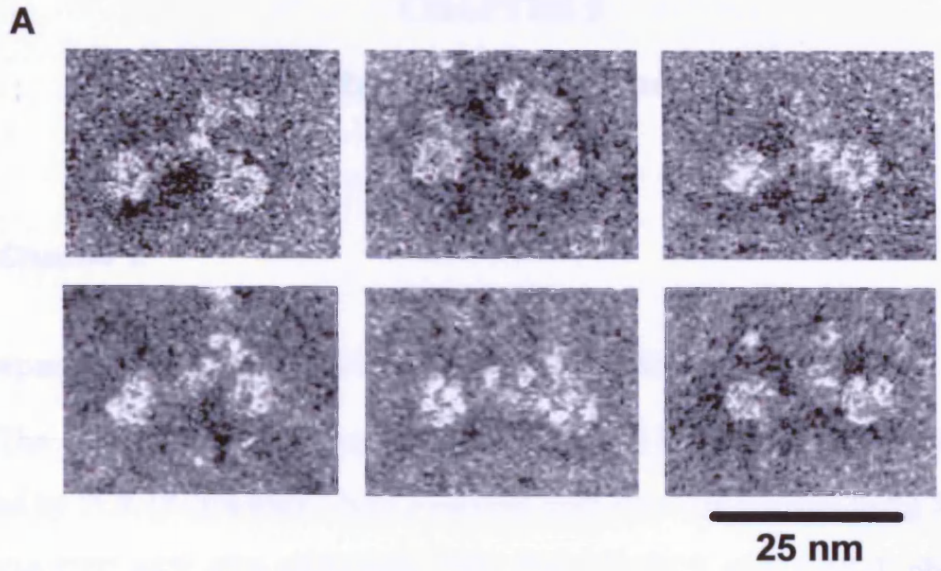
Analysis of the cDNA for the U5-specific102kD protein revealed that this protein is an orthologue of the yeast-splicing factor Prp6p. Both proteins are found in the tri-snRNP complexes, and contain the same structural motifs, which probably determine the functional roles for these proteins. Proteins containing TPRs are generally found in multi-subunit complexes where interactions between different proteins are mediated by individual TPRs, so it was proposed that the U5-102kD protein could be a co-ordinator of dynamic protein-protein interactions within the [U4/U6.U5] tri-snRNP complex. In accordance with this proposal, it was found that the U5-102kD protein interacts with the [116kD/220kD] dimer, probably through direct contact with U5-220kD. In addition, the C-terminal TPRs (10 to 19) are not accessible in [U4/U6.U5] tri-snRNPs under native conditions, which suggests that U5-102kD might interact with the U4/U6-specific proteins. In the future, it is important to determine the interacting partners of U5-102kD on the U4/U6 side. This should argue that the protein forms a bridge between the U5 and U4/U6 particles, and thus plays an important role in tri-snRNP assembly.

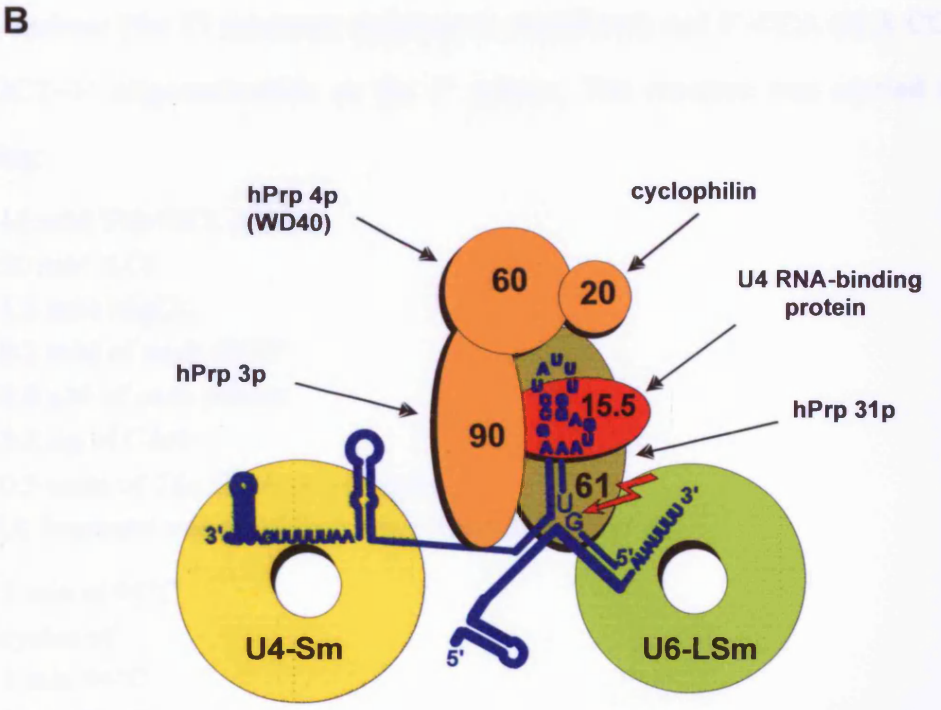
A method has been developed that allowed for the first time the isolation of 14S U4/U6 snRNP particle, functionally active in association with U5 snRNP. The 14S U4/U6 snRNPs contain a set of specific proteins, namely, 90kD, 61kD, 60kD, 20kD, 15.5kD and the complex of LSm proteins, in addition to the 10S U4/U6 snRNP comprising the U4/U6

snRNAs heteroduplex and canonical Sm proteins. The specific proteins associated with the 14S U4/U6 particle are sufficient to interact with U5 snRNP to form the [U4/U6.U5] tri-snRNP complex. Employing the reconstitution assay, it was shown that the 110kD, 100kD, 65kD and 27kD proteins are dispensable for tri-snRNP formation. However, a recent study has demonstrated that both 110kD and 65kD proteins are essential for splicing *in vitro* and are likely to promote integration of tri-snRNP into the spliceosome (Makarova and Lührmann, unpublished data).

Dr. B. Kastner (Marburg) has obtained electron micrographs of negatively stained 14S U4/U6 snRNPs. Figure 4.18A shows six bicycle-like structures characteristic for 14S U4/U6 snRNPs. Previous studies demonstrated that Sm proteins as well as LSm are organized in specific higher order assemblies that appear as a round projection with a diameter of about 8 nm and resemble doughnuts (Kambach *et al.*, 1999; Achsel *et al.*, 1999). Based on this, the 14S U4/U6 snRNP appearance could be interpreted as following: one wheel-like globular domain is the complex of Sm proteins and another globular domain is the LSm complex. There is also an additional mass in between two globular domains, which is likely to represent a set of U4/U6 specific proteins.

The current knowledge about organisation of the 14S U4/U6 particle has been rationalised in a model (Fig. 4.18B). According to this model, Sm proteins are bound to the Sm site of U4 RNA, and the complex of LSm proteins interacts with the uridine-rich site at the 3' end of U6 snRNA. It was shown that 5'-stem-loop is important for Prp4p (60kD) binding and function (Bordonne *et al.*, 1990; Hu *et al.*, 1995). Therefore, the [90kD/60kD/20kD] heteromer is shown associated with the 5' stem-loop of U4 snRNA. The 15.5kD protein directly binds to the 5' stem loop of U4 snRNA, and it is thought to be a nucleation factor for the assembly of other U4/U6-specific proteins (Nottrott *et al.*, 1999). The data obtained in our laboratory demonstrated that the 61kD protein could be directly crosslinked to U4 snRNA at the central loop of the U4/U6 helix.





Fgure 4.18. Structure of the 14S U4/U6 snRNP. A: Electron micrographs of negatively stained 14S U4/U6 snRNPs obtained by Dr. B. Kastner. **B:** Model of the 14S U4/U6 snRNP particle. The U4 and U6 snRNAs are represented by thick blu lines, with sequences to which proteins bind shown in letters. The position of the 61kD protein crosslink is indicated by the UG nucleotides in the central loop of the U4/U6 helix. The complexes of Sm and LSm proteins are shown as large holed circles, and the proteins are depicted by disks of different size and shape. Notable features of these proteins are indicated.

CHAPTER 5

Materials and Methods

Chapter 2

5.1. Preparation of the C175G DNA template by PCR

The DNA template for transcription of the C175G RNA-substrate *in vitro* was generated by PCR (Polymerase Chain Reaction) from the CΔr6 plasmid using 5'-AAA TTA ATA CGA CTC ACT ATA GGG GGC TGC TGG TTGTCT ACC CAT-3' oligonucleotide as the 5'-primer (the T7 promoter sequence is underlined) and 5'-CCA GCA CGTTGC CCA GGA GCT–3' oligonucleotide as the 3' primer. The reaction was carried out in 50 µl containing:

10 mM Tris-HCl, pH 8.0
50 mM KCl
1.5 mM $MgCl_2$
0.2 mM of each dNTP
0.6 µM of each primer
5.2 µg of CΔr6
0.5 units of Taq DNA polymerase

The DNA fragment was amplified using a standard program:

2 min at 94°C

and 30 cycles of

1 min 94°C
1 min 55°C
1 min 72°C

The product was analysed and subsequently purified on a 1.5% agarose gel. The DNA fragment was recovered using a QIAEX gel extraction kit (Qiagen) and finally eluted in 50 µl of deionised water.

The C175G sequence was as follows:

GGC TGC TGG TTG TCT ACC CAT GGA CCC AGA GGT TCT TCC AGT CCT TTG GGG ACC TGT CCT CTG CAA ATG CTG TTA TGA ACA ATC CTA AGG TGA AGG CTC ATG GCA AGA AGG TGC TGG CTG CCT TCA GTG AGG GTC TGA GTC ACC TGG ACA ACC TCA AAG GCA CCT TTG CTA AGC TGA GTG AAC TGC ACT GTG ACA AGC TGC ACG TGG ATC CCC **AGG TAA GTC** GAT CTT CGG TAC ACC ACG TTT TCA GCA TGA GAA ATT GCC TCC GTG GGC CAG GTG GTA AGC ACT TCC

TGC TCA ATC TGT TTA ACG GCT TCA TTC ATG GAC GGC ATC TGC GTC TTC TGC GGA TGG TTA TGC CAC AGG GAC ATC GCC ACC AGA AAC GCG CCG AAC TTC **AGG GTG AGT** TTG GGG ACC CTT GAT TGT TCT TTC TTT TTC GCT ATT GTA AAA TTC ATG TTA TAT GGA GGG GCA AAG TTT GCT A**A**C CAT GTT CAT GCC TTC TTC TTT TTC CTA C**AG** CTC CTG GGC AAC GTG CTG G

The splicing signals (two alternative 5' splice sites, a branched adenosine A, and a 3' splice site) are depicted in bold, and the spacer sequence between the two 5' splice sites is underlined.

5.2 Preparation of [^{32}P]-labelled C175G RNA substrate by transcription *in vitro*

This preparation of RNA was used for the RNase H assay, nitrocellulose filter binding assay and UV-crosslinking study. The DNA template, prepared by PCR amplification as described above, contained a bacteriophage T7 promoter sequence for T7 RNA polymerase. The usual incubation mixture of 10-100 µl contained:

80 mM HEPES-KOH, pH 7.6
6 mM $MgCl_2$
10 mM DTT
0.8 units/µl RNasin
1 mM rCTP
1 mM rATP
0.5 mM rGTP
0.1 mM rUTP
0.8 mM GpppG (Pharmacia)
3.08 µM [α-^{32}P]-UTP (650 Ci/mmol)
0.01 µg/µl DNA-template
1 unit/µl T7 RNA polymerase

The reaction mixture was incubated for 1-2 hours at 37°C and analysed by 4% urea-PAGE. The position of the band corresponding to C175G was determined by a short exposure of the gel to an X-Ray film. The band was cut out of the gel and RNA was eluted overnight at 4°C in buffer containing 0.5 M sodium acetate, pH 5.6, 1 mM EDTA, and 0.2% SDS. After ethanol precipitation, RNA was dissolved in H_2O in a volume corresponding to the reaction mixture.

The specific activity of the RNA was calculated in the following way. The C175G is 533 nt long and contains 145 uridine residues. As the ratio between labelled and cold rUTP in the transcription reaction is 1:32.5, on average 4.5 nt of the RNA will be labelled. Taking into account the specific activity of the commercial [α-^{32}P]-UTP (650 Ci/mmol=1.43x10^6

dpm/pmol), the specific activity of the transcribed RNA will be $6.4x10^6$ dpm/pmol (this value was corrected according to the half-life of [^{32}P] on the date of the experiment). To calculate the concentration of RNA, 1-2 µl of the RNA solution was measured by scintillation counting. The yield of gel-purified RNA was 1-10 fmol/µl.

5.3. Expression and Isolation of recombinant hnRNP A1

The purification procedure was based on the method described by Cobianchi *et al.* (1988). The plasmid pET9d-A1 (Mayeda and Krainer, 1992), encoding a full-length hnRNP A1 under the control of a T7 promoter, was kindly provided by Dr. A. Krainer, Cold Spring Harbor Laboratory.

Special buffers:

Buffer A:

0.35 M potassium phosphate, pH7.5
2 mM DTT
1 mM EDTA
1 mM PMSF
10 mM $Na_2S_2O_5$
1 µg/ml pepstatin A

Buffer B:

1.0 M potassium phosphate, pH7.5
2 mM DTT
1 mM EDTA
1 mM PMSF
10 mM $Na_2S_2O_5$
1 µg/ml pepstatin A

Buffer C:

50 mM Tris-HCl, pH 7.5
0.5 mM DTT

The *E. coli* strain BL21(DE3) was transformed with pET9d-A1. A single colony from the fresh plate was used to initiate a culture. Cells were grown at 37°C in LB medium containing 30 µg/ml of kanamycin until the OD_{600} reached 0.4-0.5. The expression of hnRNP A1 was induced by addition of IPTG to a final concentration of 0.5 mM. The cells from 2 litre of culture were harvested 3 hours later, and washed in 100 ml of ice-cold PBS. The cells were resuspended in 40 ml of buffer A, and opened by sonication. Debris and unopened cells were removed by low speed centrifugation for 15 min at 3,000 rpm in a bench-top centrifuge. The

cell lysate was then subjected to high-speed centrifugation for 20 min at 15,000 rpm in a Beckman 50.2Ti rotor. The supernatant was loaded at a flow rate of 25 ml/hour onto a 40 ml column (1.6x19 cm AC16, Pharmacia column) containing DEAE-cellulose (DE52, Whathman) equilibrated with buffer A. The flow through fraction was directly applied to a 15 ml (5 g dry weight) column of DNA-cellulose (Sigma). The column was washed initally with 50 ml of buffer A and then with 50 ml of buffer B. Protein was eluted from the column with a linear gradient of NaCl (0 - 1.0 M) in buffer C (V=250 ml + 250 ml) with a flow rate of 20 ml/hour. The A1 protein was eluted in a single broad peak at 0.4-0.6 M NaCl. The protein analysed by 12% SDS-PAGE was essentially (electrophoretically) pure.

5.4. RNase H assay as a tool for the detection of protein-RNA interactions

General procedures and all manipulations are based on principles described by Eperon and Krainer (1994). The RNase H assay was carried out under *in vitro* splicing conditions in D-Glu buffer supplemented with ATP, creatine phosphate, PVA and RNasin (5.16, splicing *in vitro*).

In order to decrease the margin of error, the starting mixture, containing RNase H and RNA was divided into two reactions. In the first reaction, protein (hnRNP A1 or SF2/ASF) was added, whereas the second one served as a control in the absence of protein. Both reactions were designed to accommodate the desired number of time points. The incubation mixture of 15 µl contained approximately 1-10 fmol of RNA (50,000 cpm), 5 pmol of hnRNP A1 or SF2/ASF, and 0.8 units of RNase H (Pharmacia), and was incubated for 20 min at 30°C to allow the formation of RNA-protein complexes. The RNase H cleavage was triggered by the addition of 300 pmol of a specific DNA oligonucleotide (5'SS oligo or BP oligo), and 1 µl aliquots were taken out at the designated time intervals. The reaction was stopped by the addition of ice-cold buffer (0.1 M Tris-HCl, pH 7.5, 150 mM NaCl, 1% SDS, 12.5 mM EDTA) containing proteinase K (0.4 mg/ml). When all time-points had been collected, they were subjected to proteinase K digestion for 15 min at 37°C. The RNA was ethanol precipitated and analysed by 6% urea-PAGE.

For analysis of the RNase H assay, the bands corresponding to pre-mRNA and the largest cleavage product were quantified on a PhosphorImager (Molecular Dynamics). The percent of uncleaved RNA was calculated according to the equation:

$$A\% = A/(fxB+A)$$

Where A is uncleaved RNA, B is the largest product of cleavage, and f is a normalisation factor which takes into account the ratio between the number of uridine residues in B and A. For the cleavage directed by the 5'SS oligo, the product contains 93 out of 145 uridines and, therefore, the factor is 145/93, and for the BP oligo, the factor is 145/122. The 5'SS oligo was 5'-CGA CTT ACC TGG GG-3' and directed the cleavage at a position 216 nt from the 5' end of the C175G transcript. The BP oligo was 5'-TCC ATA TAA CAT GAA T-3' and directed the cleavage at a position 452 nt from the 5' end of the transcript.

5.5. Nitrocellulose filter binding assay

Two methods were employed that differ in the manner in which variations in the concentrations of the components were achieved. In the method of constant volumes, reactions were carried in 600 µl of D-Glu buffer and contained 1-2 fmol of [^{32}P]-labelled C175G (5,000 cpm); the amount of hnRNP A1 varied from 100 to 2,000 fmol. The mixtures were incubated for 30 min at 30°C to achieve equilibrium. A Costar nitrocellulose filter unit (8 wells) connected to a water pump was used for filtration. Filters were pre-soaked before filtration in the incubation buffer. Then, mixtures were applied directly to the filters and passed through. Filters were washed with 1 ml of D-Glu buffer and air-dried. The retained radioactivity was measured in a Beckman scintilation cocktail in a scintilation counter (LKB). Alternatively, the method of variable volumes was employed. In this case, incubation mixtures with different volumes of D-Glu buffer (0.1-2.5 ml) contained constant amounts of hnRNP A1 (0.5 pmol) and RNA (1-2 fmol). Mixtures were processed as described above.

5.6. Direct UV-crosslinking of proteins to the RNA

All incubations were carried out under *in vitro* splicing conditions in D-Glu buffer supplemented with ATP, creatine phosphate and RNasin (5.16, splicing *in vitro*) in a 96-well microtitre plate. The 10 µl reaction contained a constant amount of [^{32}P]-labelled C175G

RNA (5000 cpm, 1-2 fmol), and the desired amount of protein(s), which varied in different experiments from 10 fmol to 30 pmol. The mixtures were incubated for 15 min at 30°C, and then UV-irradiated, using a SpotCure device (UPV), at 254 nm for 30 seconds or as stated in the Figure legends. After crosslinking, 4 µl of an RNase mixture (4 ng RNase A, 11.2 units RNase T1, 48 ng yeast RNA) were added to each 10 µl reaction, and digestion was carried out for 15 min at 37°C. Reactions were analysed by 12% SDS-PAGE. The gel was electro-blotted onto a Hybond C membrane (Amersham), and the signals were quantified with a PhosphoImager using Molecular Dynamics Software.

5.7. RNA fragmentation – selection assay for the investigation of a putative high affinity site for hnRNP A1 binding

The RNA substrates used for this assay were prepared by large scale RNA synthesis according the Promega Protocols and Applications Guide using as templates plasmids (kindly provided by Dr. B. Chabot) digested with suitable restriction enzymes (Fig.2.12).

The transcription reaction was performed in a total volume of 50 µl and contained:

80 mM HEPES-KOH, pH7.5
24 mM $MgCl_2$
2 mM spermidine
40 mM DTT
7.5 mM of each rNTP
3 µg of linear DNA template
5 µl of T3 RNA polymerase (20 u/µl, Promega)
50 units of RNasin (Promega)

The reaction was incubated for 3 hours at 37°C and then ethanol precipitated and dissolved in 325 µl of deionised water. To remove the 5' terminal phosphate, RNA was incubated with 35 units of CIAP (Calf Intestinal Alkaline Phosphatase, Promega) in a total volume 400 µl containing 50 mM Tris-HCl, pH 9.3, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, and 1 mM spermidine, for 30 minutes at 37°C. The reaction was stopped by adding EDTA to a final concentration of 5 mM and incubating for 10 minutes at 80°C. The RNA was phenol-chloroform extracted and ethanol precipitated.

After de-phosphorylation, RNA was 5' end-labelled with [γ-^{32}P]ATP using T4 PNK (Polynucleotide Kinase, Promega). The 30 µl incubation mixture contained:

50 mM Tris-HCl, pH 7.6
10 mM $MgCl_2$

5 mM DTT
0.1 mM spermidine
0.5-1 μM [γ-^{32}P]ATP (3,000Ci/mmol, 10mCi/ml)
0.4 units/μl T4 PNK (10 u/μl)
30 pmol RNA

The reaction was incubated for 30 minutes at 37°C, and RNA was purified by 6% urea-PAGE. After elution, the concentration of RNA was determined by a spectrophotometer. The typical yield was approximately 1 μg of RNA whose specific activity varied from 50,000 to 200,000 cpm/pmol for the different RNA-substrates.

For partial hydrolysis, RNA substrates were incubated for 15 minutes in 25 μl of 50 mM $NaHCO_3$ at 90°C and the reaction was stopped by ethanol precipitation. The binding of hnRNP A1 to the partially hydrolysed RNA was carried out as described for the UV-crosslinking studies (5.6). After 15 minutes of incubation at room temperature, the reactions were filtered through pre-wet Costar nitrocellulose filter units (8 wells) assembled on a microtitre plate, by centrifugation at 1,000 rpm for 5 minutes in a bench-top centrifuge. The run through (RT) fractions were collected and ethanol precipitated in the presence of tRNA (1 μl of 10 μg/μl) as a carrier. The membranes were washed with 1 ml of D-Glu buffer, and the bound RNA was eluted with 50-100 μl of elution buffer (7 M Urea, 20 mM sodium citrate pH 5.0, 1 mM EDTA) and ethanol precipitated in the presence of tRNA as a carrier. The RT and membrane retained fractions were analysed by urea-PAGE; gels were dried and subjected to autoradiography.

Chapter 3

5.8. Isolation of the His-tagged recombinant RBM protein and immunisation of a rabbit

Special buffers:

1) Buffer 1: 20 mM Tris-HCl, pH 7.9; 0.5 M NaCl, 8% glycerol, 5 mM DTT.

2) Loading buffer: 20 mM Tris-HCl, pH 7.9, 0.5 M NaCl, 6 M urea, 1 mM imidazole, pH 7.9.

3) Washing buffer: loading buffer with 20 mM imidazole.

4) Elution buffer: loading buffer with 80 mM imidazole.

Procedure (all volumes correspond to the amount of cells from a 1 litre culture):

The *E. coli* BL21(DE3) cells were transformed with the plasmid pYTR3 and plated on LB agar containing 100 μg/ml ampicillin. A single colony from the fresh plate of transformed cells was picked and the cells were grown in LB medium, containing 100 μg/ml ampicillin, until the OD_{600} had reached 0.3-0.4. Expression of RBM was then induced by addition of IPTG to a final concentration of 0.5 mM. The cells were harvested 1 hour after induction by centrifugation at 6,000 rpm for 10 minutes (Sorvall GSA rotor), washed twice with ice-cold buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA) and resuspended in 25 ml of Buffer 1. The cells were lysed by sonication, and debris and unopened cells were removed subsequently by low speed centrifugation for 15 min at 2,500 rpm in a bench-top centrifuge ("Jouan"). The cell lysate was centrifuged for 30 min at 30,000 g (Beckman 50.2Ti rotor at 18,000 rpm) to prepare a so-called S-30 extract. Although RBM was present in the S-30 extract, the supernatant was discarded because RBM could not be recovered by affinity chromatography on a Ni^{+}-column.

The pellet from the S-30 preparation was dissolved in 5 ml of loading buffer and applied on a 2 ml Ni-NTA agarose (Qiagen) column equilibrated with loading buffer. The column was washed with 10 volumes of washing buffer, and RMB was released from the column with elution buffer. The peak fractions contained RBM at a concentration of 1.5-3.0 mg/ml, as measured by Bradford assay using BSA as a standard. The purity of RBM was controlled by 10% SDS-PAGE and estimated to be at least 90%.

For immunological purposes, fractions containing RBM in 6 M urea solution, were dialysed exhaustively against 6% acidic acid containing 5 mM β-mercaptoethanol and then lyophilised. The lyophilised material was not soluble in aqueous buffers and could be solubilised only in the presence of 6 M urea or guanidine chloride.

Immunisation procedure:

The protocol was designed according to Harlow and Lane (1988). A pre-immune serum was collected one day before the primary injection. The lyophilised RBM was dissolved in a sterile solution (20 mM Tris-HCl, pH 7.5 and 6 M urea) at a concentration 6 mg/ml. Before injection, the protein solution was diluted with sterile water to 1 mg/ml, mixed with complete Freund's adjuvant for the primary injection or with incomplete Freund's adjuvant for subsequent injections. 1 mg of protein was used per injection and applied to several areas of

the rabbit. In total, five injections were carried out, separated by a two-week interval. Bleeding tests were carried out every two weeks, and serum was tested for its ability to recognise the recombinant protein on western blots at different dilutions. When an increase in the titre was no longer observed, the rabbit was exsanguinated.

5.9. Preparation of rabbit antiserum and purification of polyclonal antibodies

After the collection in the animal house, the blood was transferred to the laboratory and left on the bench for 1-2 hours or until a clot had formed. The clot was detached from sides of the tube with a long Pasteur pipette, and was left at 4°C overnight to allow complete contraction of the clot. The tube was centrifuged for 10 min at 4000 rpm, and the serum was carefully withdrawn using a pipette. A portion of the serum was stored at 4°C in the presence of sodium azide (0.1% final concentration), and the rest was frozen at -20°C or -70°C in aliquots (1-7 ml). This serum was used directly for western blot analysis. For immunofluorescence experiments and for immunoprecipitations, the polyclonal antibodies were first affinity purified on Protein A sepharose (Pharmacia).

The antiserum raised against the recombinant RBM and hnRNP A1 proteins in Leicester, was treated with an acetone-dried extract prepared from *E. coli* BL21(DE3) cells. The antiserum against a GST-fusion of the C-terminus of U5-102kD protein (Marburg) was directly purified on Protein A Sepharose. The antibody against the C-terminal peptide of the U4/U6-61kD protein (Marburg) was affinity purified on a SulfoLink column (Pierce) containing the same peptide as described below. The antibodies against other snRNP proteins raised in the laboratory of R. Lührmann were used directly without purification. All other antibodies used in this work were generous gifts from other laboratories and were used without any further purification. All antibodies are itemised in the table at the end of this chapter.

5.10. Preparation of plasmid DNA for the transfection of mammalian cells

For routine cloning manipulations, the plasmid DNA was usually isolated according to a miniprep protocol (Sambrook *et al.*, 1989). Commercial DNA purification kits (Promega, Qiagen) were used to prepare plasmid DNA for automatic sequencing. However, for transfection of mammalian cells, plasmid DNA was prepared according to a protocol with a

LiCl step. This inexpensive method allows the isolation of large quantities of a high quality plasmid DNA that could be directly used in the transfection experiments.

Stock solutions:

1) GTE: 50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA.

2) SDS/NaOH solution: 0.2 M NaOH, 1% SDS.

3) 5 M potassium acetate, pH 4.8.

4) 4.4 M LiCl.

5) RNase A, 10 mg/ml.

All steps in this procedure were carried out on ice with ice-cold solutions. *E. coli* cells from a 50 ml overnight culture were pelleted by centrifugation and resuspended in 4 ml of GTE. 8 ml of the SDS/NaOH solution was added, mixed well and left on ice for 10 minutes. Then, 6 ml of 5 M potassium acetate (pH4.8) was added, mixed and kept on ice for 10 minutes. To remove the cell debris, the lysate was centrifuged for 15 minutes at 4000 rpm in a bench-top centrifuge, and the supernatant was filtered through a 0.45-μm acrodisc filter (Gelman Sciences). 17 ml of propanol-2 were added to the supernatant, and the mixture was chilled for 15 minutes at -70°C. The precipitate obtained by centrifugation for 10 minutes at 4000 rpm, was re-suspended in 2 ml of TE buffer. 2.5 ml of 4.4 M lithium chloride was added and the solution was incubated for 10 minutes on ice. After centrifugation, the supernatant was collected and DNA was precipitated by the addition of 10 ml of cold ethanol, washed with 80% ethanol and re-suspended in 400 μl of TE buffer. 10 μl of RNAse A, treated before use for 10 minutes at 70°C, was added and incubated for 15 minutes at 37°C. 20 μl of 10% SDS was added and the solution was heated at 70°C for 10 minutes. The DNA was extracted with phenol, followed by phenol-chloroform, and precipitated with ethanol without chilling. The pellet was resuspended in water.

5.11. Transient transfection of RBM gene into HeLa cells using the calcium phosphate method

1). Tissue culture:

Tissue culture manipulations were carried out according to Celis (1994). For transient transfections, HeLa cells were grown as a monolayer in tissue culture flasks or petri dishes (Nunc) in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% newborn

calf serum, and 1% P/S solution (10,000 penicillin/streptomycin, Gibco-BRL) at 37°C in the presence of 4-5% CO_2. The medium was supplemented with fetal calf serum if cells started growing from the stock kept in liquid nitrogen or directly after transfection. The medium was changed daily. Stock cells were grown in 200 ml Nunc tissue culture flasks until 80% confluent, and then split using a trypsinisation procedure. According to this procedure, the medium was removed by aspiration and cells were rinsed twice with PBS. Then, 1.5 ml of a pre-warmed (37°C) 1xTypsin-EDTA solution (Gibco-BRL) was added and the flask was incubated for 5 minutes at 37°C or until the cells were completely detached. To continue growing a cell line for routine purposes, the detached cells were diluted with fresh medium and distributed between 6 tissue culture flasks of the same size. For transfection experiments, detached cells were quickly counted using a haemocytometer to distribute the cells at the desirable density.

2). Transfection by the calcium phosphate method:

Transfections were carried out using the protocol of Gorman (1985).

Special Solutions:

a) 10xHBS (HEPES-buffered saline). The 10x stock solution contains 8.18% NaCl, 5.94% HEPES, 0.2% Na_2HPO_4 (all w/v). This stock was used to prepare a 2xHBS solution whose pH was adjusted to 7.12 directly before transfection (a correct pH value is critical for the formation of the plasmid containing precipitate).

b) 2 M $CaCl_2$

c) 15% glycerol/HBS was made by mixing 30 ml of 50% glycerol (w/v), 50 ml of 2xHBS, pH 7.12, and 20 ml of water.

d) 0.5 M sodium butyrate.

All stock solutions were sterilised by filtration through a 0.2-μm membrane.

Procedure:

One day before transfection, cells were replated on a standard 90 mm tissue culture petri dish at a density of approximately $5x10^5$ cells per dish as determined by haemocytometer counting (100 HeLa cells per haemocytometer field corresponds to 10^6 cells per ml).

On the next day, the medium was replaced with fresh medium containing 10% fetal calf serum and the plasmid DNA was added to the cells 3 hours later. The calcium phosphate-DNA precipitate was prepared just before use from the stock solutions warmed-up to room temperature. In tube A, 5 μg of DNA solution in 219 μl of water was mixed with 31 μl of 2 M $CaCl_2$. Tube B contained 250 μl of 2xHBS. To make the precipitate, the contents of tube A were added to the contents of tube B in a drop-wise manner. As the solution became cloudy and a precipitate was formed, it was immediately dispersed over the medium in the petri dish in a drop-wise manner. Cells were incubated subsequently for 3-4 hours. The aggregation of precipitated material on the cell surfaces was examined using a light microscope. If the cells were covered with small black grains, a glycerol shock was performed. The medium was removed by aspiration and 0.5 ml of 15% glycerol/HBS was added per 90-mm petri dish. The dish was placed under the light microscope and the cell condition was followed. When the cells started to shrink (within 2 min after addition of glycerol), the glycerol was rapidly washed away with PBS, and the cells were supplied with medium containing 10% fetal calf serum.

The treatment of the transfected cells with butyrate had been shown to increase at least 3-fold the number of cells that can express the incoming DNA. Usually, the butyrate was added to the cells, but its effect on the expression of RBM has not been confirmed. After the glycerol shock, the sodium butyrate solution was added to the fresh medium to a final concentration of 10 mM (200 μl of the stock solution into 10 ml of medium in a 90 mm petri dish). The sodium butyrate containing medium remained with the cells overnight. On the next day, the butyrate was removed, cells were washed with PBS and re-fed with medium containing 10% newborn calf serum.

The transient expression of RBM in HeLa cells was analysed 40-48 hours after transfection by immunostaining and western blotting. The percent of the cells expressing RBM was estimated by immunostaining and the value varied between 5 and 20%.

5.12. Stable transfer of RBM gene into HeLa cells

Transfection by the calcium phosphate method was carried out according to the protocol of Gorman (1985). Clones resistant to antibiotic were selected as described in Solowska *et al* (1989).

Procedure:

HeLa cells growing on five 90 mm petri dishes were transfected by the calcium phosphate method as described above. On the next morning after transfection, the medium containing butyrate was replaced with the standard DMEM medium containing 10% newborn calf serum (NBCS). The cells were incubated for 48 hours in the absence of the antibiotic. On the third day after transfection, the cells were trypsinized and diluted 1:5. The incubation was continued in DMEM+10% NBCS containing 0.5 mg/ml of G418 antibiotic (50 µl of a 100 mg/ml stock solution were added to 10 ml of medium covering a 90 mm petri dish). At this stage, a small aliquot of trypsinised cells was transferred to a 6 well (35 mm diameter) tissue culture plate (Nunc) containing sterile glass coverslips and grown overnight. The cells bound to the coverslip were fixed and immunostained with α-RBM to check whether the transient transfection was successful.

Fresh medium containing 0.5 mg/ml of G418 antibiotic was added every 3-4 days. Cells began to look sick within 3-4 days. They rounded up and became detached from the dish. By 14 days after transfection, the plates were nearly devoid of cells except for the small islands of resistant cells. Each of the 5 starting plates contained about 10-20 such islands. At this stage, the clones were selected. The cloning cylinders were made from yellow tips by cutting off a wide end (5-7 mm long) and were sterilised. The medium was removed from the plate and the cells were washed twice with PBS. Using sterile forceps, the wide end of the cloning cylinder (yellow tip) was slightly immersed into sterile vaseline and then was gently pressed onto the plate to form a seal over the colony. Two drops of the pre-warmed (37°C) 1xTypsin-EDTA solution were added to the cylinder. The plate was left for 5 min at 37°C, and then examined under the microscope. After the cells had detached, they were pipetted up and down to break up clumps, and then transferred to one well of a 6 well (35 mm diameter) tissue culture plate (Nunc) containing DMEM+10%NBCS. After 6 hours when the living cells were attached to the plates, the medium was removed and the cells were washed

thoroughly with PBS in order to remove bits of vaseline, and then fed with medium containing 0.1 mg/ml of G418 antibiotic. Using the procedure described above, only 20% of the transferred clones survived. The cells grew very slowly and it took about two weeks to reach semi-confluency. At this moment, the cells were subcultured, and aliquots were taken for western blot analysis and immunostaining.

5.13. Immunofluorescence microscopy

The procedures and all manipulations carried out in the immunocytochemical part of my work are based on recommendations of Celis (1994). The indirect immunostaining technique was used for locating the specific proteins to a subcellular compartment. The procedure includes the following steps:

1) Fixation:

Usually two days before the experiment, HeLa cells were trypsinised and transferred to a 6 well (35-mm diameter) tissue culture plate (Nunc) containing sterile glass coverslips. Cells were grown until 80% confluence. On the day of immunostaining, the cells were washed twice with PBS and fixed using the Acetone/Methanol method or the Formaldehyde/Triton method.

Acetone/Methanol Method: The solution, containing 50% acetone and 50% methanol, was pre-chilled to -10°C to -20°C, and then pipetted into the wells containing coverslips. After incubation for 90 seconds, the coverslips were withdrawn with forceps, drained and left to air dry.

Formaldehyde/Triton Method: The coverslips were incubated for 15 min in 3.7% formaldehyde in PBS buffer (Leicester) or 2% paraformaldehyde in PBS (Marburg), washed twice with PBS and permeablised by incubation with 0.2% Triton-X-100 in PBS for 5 min.

The Acetone/Methanol method is much faster than the Formaldehyde/Triton method, and therefore was routinely used for multiple sample analysis. However, the Acetone/Methanol method usually gives much higher background staining, and the pictures presented were obtained from cells fixed by the Formaldehyde/Triton method.

2) Interaction with antibodies

The cells were pre-blocked with 1% BSA/PBS for 30 min at 37°C, and then incubated for 60 min at room temperature with the primary antibody diluted 1:500, washed three times with PBS and incubated for 30 min with the secondary antibody diluted 1:500 in BSA/PBS. Alternatively, permeablised cells were additionally denatured by incubation in 1% SDS/PBS for 10 min at room temperature, washed three times with PBS, and then processed as described above. Fluorescence images were recorded using a Zeiss fluorescent microscope, or Leica DM/IRB microscope equipped with a Kappa videocamera 100x objective, or a Zeiss confocal laser scanning microscope. For double staining, two primary antibodies, and later, two secondary antibodies were added together. In this case, the images were recorded and processed using a Zeiss confocal laser scanning microscope.

5.14. Preparation of a total HeLa cell lysate in a urea buffer

This procedure was used for the quick preparation of whole cell lysates for the western blot analysis of the transient expression of RMB in HeLa cells. HeLa cells were grown on the bottom of standard 90 mm diameter tissue culture plates. Medium was removed by aspiration, and cells were washed twice with PBS, and harvested into 1 ml of PBS using a rubber policeman, and transfered to an Eppendorf tube. The cells were pelleted by centrifugation at 13,000 rpm for 20 sec, the supernatant was discarded and the cells were resuspended in 100 µl of urea lysis buffer (50 mM Tris-HCl, pH 7.5, 6M urea, 5 mM DTT). The suspension was sonicated on ice using a small probe with 3 bursts of 10 seconds each or until it became clear. The protein concentration was measured by Bradford assay using BSA as a standard. The lysates were directly subjected to SDS-PAGE or were aliquoted and stored at -70 °C.

5.15. Preparation of HeLa cell nuclear extracts

The preparation of HeLa cell nuclear extracts for *in vitro* splicing assays and immunoprecipitations was carried out essentially by the Dignam method (Dignam *et al.*, 1983). Two protocols were used, depending on the source of culture. When HeLa cells were grown in DMEM medium in a standard 3-litre spinner flask, and harvested at a density of $5x10^5$ cells per ml, extracts were prepared as described by Eperon and Krainer (1994). Alternatively, HeLa cells, obtained from the Computer Cell Culture Centre (CCCC, Mons,

Belgium) were grown in CCCC medium in a 10-litre bio-reactor (Applikon BIOTEK, Germany), and were harvested at a density of $5x10^6$ cells per ml. The important parameters which determine whether or not an extract is active for splicing, is how low the proportion of dead cells is and, in general, how healthy the cells are. Therefore, it is essential to control the growth of cells by haemocytometer counting (100 HeLa cells per haemocytometer field corresponds to 10^6 cells per ml) to be sure that the cell population doubles every 24 hours.

For the extract preparation the following **Stock solutions** were used:

1) Buffer A:

10 mM HEPES-KOH, pH 7.9
1.5 mM $MgCl_2$
10 mM KCl
0.5 mM DTE

2) Buffer C:

20 mM HEPES-KOH, pH 7.9
0.42 M NaCl
1.5 mM $MgCl_2$
25% glycerol
0.2 mM EDTA (potassium salt)
0.5 mM DTE
0.5 mM PMSF

3) Buffer D:

20 mM HEPES-KOH, pH 7.9
0.1 M KCl
1.5 mM $MgCl_2$
10% glycerol
0.2 mM EDTA (potassium salt)
0.5 mM DTE
0.5 mM PMSF

The following protocol is designed for a 4-litre culture taken out of the bio-reactor at a density of $5x10^6$ cells per ml. All procedures are carried out on ice or in a cold room.

A 4-litre culture was divided among 6x1000 ml plastic bottles and centrifuged at 2,000 rpm for 10 minutes. Cell pellets were rapidly resuspended in 6x60 ml of ice-cold PBS by pipetting. The suspensions were combined and distributed among ten 50 ml disposable tubes (Falkon) and centrifuged in a bench-top centrifuge for 10 min at 2,100 rpm. The supernatant was discarded. At this point, it is important to estimate (by eye) the packed cell volume in order to determine the number of cells and accurately calculate the volume of Buffer C that

should be added in the following steps. It was determined empirically that 10 ml of pelleted cells corresponds to $3x10^9$ cells. The packed cell volume from a 4-litre culture was usually 70 ml (10x7 ml). Next, cells were resuspended in 5 packed cell volumes of ice-cold Buffer A (total 350 ml) by pipetting and centrifuged for 10 min at 2,100 rpm. A supernatant was removed by gentle vacuum aspiration. Subsequently, cells were resuspended in 2 volumes of ice-cold Buffer A (total 140 ml), transferred to a Dounce homogeniser and lysed with 10 gentle strokes. The suspension was transferred to 35 ml centrifuge tubes and spun in an SS34 rotor (Sorval) for 10 min at 3,000 rpm. In order to remove ribosomes, 2/3 of the supernatant were removed by gentle aspiration, and then, the same tubes were further centrifuged for 20 min at 16,000 rpm. The supernatant was completely removed and discarded.

The volume of the nuclear pellets was estimated by eye, and an equal volume of Buffer C was added. The salt concentration is considered crucial for opening the nuclei and extraction of splicing components from the chromatin. Therefore, the volume of added Buffer C was double-checked taking into account that not more than 3 ml of Buffer C should be added per 10^9 cells. Thus, 60 ml of Buffer C was typically added to the total nuclear pellet obtained from a 4-litre culture.

The suspension of nuclei was transferred to a Dounce homogeniser and the nuclei were opened with 10 gentle strokes. The homogenate was transferred to a 50-ml disposable tube, and rotated "head-over-tail" for 30 min in a cold room. Then, the suspension was transferred to 35-ml centrifuge tubes and spun in a SS34 rotor (Sorval) for 30 min at 16,000 rpm. The supernatant (*i. e.*, the nuclear extract) was either directly aliquoted and quick-frozen in liquid nitrogen (high salt nuclear extract) or was first dialysed against 2 litres of buffer D for 4 hours with one change of buffer (2x2 L) and then frozen. The extracts were stored at -70°C. The high-salt nuclear extracts were usually used for immunoprecipitation or immunodepletion experiments, where a higher salt concentration is necessary, and therefore, the dialysis is not required. Alternatively, the high salt extract was dialysed rapidly against buffer D just before an experiment and used for splicing *in vitro*. The precipitate that forms during dialysis was removed by centrifugation for 30 seconds at 13,000 rpm just before addition to the splicing reaction mixture.

5.16. In vitro splicing reactions

In vitro splicing assays were carried out essentially as described by Eperon and Krainer (1994) with minor modifications.

1). Leicester:

The HeLa cell nuclear extract was dialysed against buffer D-Glu.

Buffer D-Glu:

20 mM Triethanolamine-HCl, pH 7.9
80 mM monopotassium glutamate
5% glycerol
0.2 mM EDTA (potassium salt)
0.5 mM DTT
0.5 mM PMSF

Reaction mixtures contained 42% (v/v) of nuclear extract in D-Glu buffer supplemented with $MgCl_2$, ATP, creatine phosphate (CP) and polyvinyl alcohol (PVA). Final concentrations were as follows:

13 mM Triethanolamine-HCl, pH 7.9
50 mM monopotassium glutamate
3.2 mM $MgCl_2$
3% glycerol
0.13 mM EDTA (potassium salt)
0.3 mM DTT
0.3 mM PMSF
20 mM CP
1.6 mM ATP
2.6% PVA (w/v)

A standard one-point, 10 µl reaction contained approximately 2.5 fmol of [^{32}P]-labelled pre-mRNA (10.000-15.000 cpm) that had been prepared by *in vitro* transcription and gel-purified.

2). Marburg:

The HeLa cell nuclear extract was dialysed against buffer D as described in the paragraph above. Reaction mixtures contained 35% (v/v) of nuclear extract in buffer D supplemented with $MgCl_2$, ATP, and CP to the following final concentrations:

10 mM HEPES-KOH, pH 7.9
50 mM KCl
3.25 mM $MgCl_2$
3.5% glycerol

0.07 mM EDTA (potassium salt)
0.175 mM DTE
0.175 mM PMSF
20 mM CP
2 mM ATP

A standard one-point, 20 µl reaction contained approximately 10 fmol of [^{32}P]-labelled pre-mRNA (25.000-30.000 cpm) that had been prepared by *in vitro* transcription and gel-purified.

The splicing was carried out at 30°C and the reaction mixtures were processed using microtitre plates as described by Eperon and Krainer (1994). At defined time intervals aliquots of 10-20 µl were transferred to the wells of a microtitre plate, which was kept on ice. When all time points were collected, 2 µg of proteinase K in 50 µl of proteinase K buffer (0.1 M Tris-HCl, pH7.5, 0.15 M NaCl, 12.5 mM EDTA, 1% SDS) were added to the wells and the mixtures were incubated at 37°C for 5 min. 150 µl of ethanol was added to each well, and the plate was immediately centrifuged in a lab-top centrifuge at 2100 rpm for 20 min. The supernatant was removed by aspiration. Pellets were washed with 200 µl of ethanol, air dried and dissolved in 10 µl of formamide dye mixture. The samples were subjected to urea-PAGE and the RNA was visualised by autoradiography.

5.17. Mini-scale preparation of HeLa cell nuclear extracts

A mini-scale method was developed by Lee *et al.* (1988) for the preparation of nuclear extract from monolayer cells. Monolayer HeLa cells from a standard 90 mm tissue culture petri dish at 80% confluence were washed with PBS and harvested into 1-2 ml of PBS using a rubber policeman. The cells were pelleted by centrifugation in a microcentrifuge for 20 seconds at 13,000 rpm, resuspended gently in 1 packed cell volume of ice cold buffer A and allowed to swell on ice for 15 minutes. The cells were then lysed by passing them rapidly 5 times through a 23 gauge hypodermic needle using a 1 ml syringe. The homogenate was spun in a microcentrifuge for 20 seconds. The crude nuclear pellet was resuspended in a 2/3 packed cell volume of ice cold buffer C and incubated on ice for 30 minutes. Insoluble matter was pelleted by spinning for 5 min at 13,000 rpm in a microcentrifuge. The supernatant, (*i. e.*, the nuclear extract) was either directly aliquoted and quick-frozen in liquid nitrogen or first

dialysed against buffer D for 2 hours and then frozen. Usually, 5 standard 90 mm tissue culture petri dishes yield approximately 0.2 ml of nuclear extract.

5.18. Mini-scale preparation of whole HeLa cell extract using high salt buffer

This mini-scale method is based on a rapid, one-step procedure developed by Jiang and Eberhard (1995) for the detection and quantitative analysis of DNA binding activities.

Extraction buffer:

20 mM HEPES-KOH, pH 7.9
0.5 M KCl
1.5 mM $MgCl_2$
0.2 mM EDTA
0.5 mM DTT
20% glycerol

Dilution buffer:

20 mM HEPES-KOH, pH 7.9
5 mM KCl
0.5 mM $MgCl_2$
0.5 mM DTT

Monolayer HeLa cells, growing on the bottom of a standard 90 mm tissue culture petri dish, were rinsed with PBS and harvested into 1-2 ml of PBS using a rubber policeman. The cells were pelleted by centrifugation in a microcentrifuge for 20 seconds at 13,000 rpm, and PBS was removed by aspiration. The cells (50 µl) were resuspended by vortexing in 3 packed cell volumes of extraction buffer. To achieve complete lysis, the suspension was subjected to one cycle of freezing on dry ice and immediately thawing at 37°C in a water bath. The samples were centrifuged for 1 min at 13,000 (no visible pellet was formed) and the supernatant was taken and used as the whole cell extract.

In some experiments, the whole cell extract was diluted with 4 volumes of dilution buffer to reduce the salt concentration. During addition of the dilution buffer, the solution became unevenly condensed. The samples were centrifuged for 1 min at 13,000 and the condensed fraction settled down, but a pellet was not formed. The condensed fraction was designated the chromatin fraction. The clear supernatant was removed. Two packed cell volumes of dilution buffer were added to the chromatin fraction, and the sample was sonicated for 25-30 seconds (3x10 sec) using a small probe. During sonication the solution

became transparent and there was no visible pellet formed after centrifugation for 1 min at 13,000 rpm. The supernatant was transferred to another tube and was used for immunoprecipitation as a chromatin fraction.

5.19. Immunoprecipitation of RBM from extracts of transiently transformed HeLa cells

IP (immunoprecipitation) buffer:

20 mM HEPES-KOH, pH 7.9
0.17 M KCl
1.5 mM $MgCl_2$
0.2 mM EDTA
0.5 mM DTT
5% glycerol

Elution buffer:

20 mM sodium citrate, pH 5.0
7 M urea
1 mM EDTA

α-RBM antibodies were covalently immobilised on Affi-Gel 10 or Affi-Prep Hz support as described below. The beads were washed with IP buffer and pre-blocked with IP buffer containing 0.1% BSA at 4°C overnight or at room temperature for 1-2 hours. Then, the beads were equilibrated with the same buffer as one of the extract (*e. g.*, diluted whole cell extract). 0.1 ml bed volume of pre-blocked beads was mixed with 1 ml of the extract (*e. g.*, the whole cell extract or the chromatin fraction isolated from the same cells). The mixtures were rotated gently "head-over-tail" for 2 hours at room temperature. Subsequently, the beads were allowed to settle down by gravity and the extract (supernatant) was removed. The beads were washed 3 times with 0.5 ml of IP buffer. The suspension of beads was not centrifuged during the washing procedures and the beads were allowed to settle down by gravity. After the last wash the sample was centrifuged for 30 seconds at 13,000 rpm and the liquid was removed completely. The material bound to the beads was eluted with 50 μl of either SDS-PAGE loading buffer (20 mM Tris-HCl, pH 6.8, 1% SDS, 0.1 mM DTT, 10% glycerol) or elution buffer, and analysed by SDS-PAGE, followed by staining with Coomassie or silver and western blotting.

5.20. Coupling of antibodies to Affi-Gel 10

Affi-Gel 10 (Bio-Rad) support was used according to the recommendations of the manufacture to couple antibody directed against RBM to a solid phase. The beads charged with α-RBM were used for immunoprecipitations. Affi-Gel 10 is an agarose-based matrix containing a neutral 10-atom spacer arm with the chemically reactive N-hydroxysuccinimide at the end. Ligands with primary amino groups couple spontaneously to Affi-Gel 10 in aqueous solutions at pH 6.5-8.5. Upon addition of ligand, N-hydroxysuccinimide is displaced, and a covalent amide bond is formed.

Special solutions:

Coupling buffer: 0.1 M MOPS, pH 7.5

Blocking buffer: 1 M ethanolamine-HCl, pH 8.0

Procedure:

Antiserum was affinity purified on Protein A Sepharose as described below. 2 ml of the eluate from the Protein A column (approximately 2 mg/ml) was dialysed against 2 litres of coupling buffer overnight at 4°C. 0.5 ml bed volume of Affi-Gel 10 slurry, supplied in isopropanol, was transferred to a disposable Econo-column (Bio-Rad), washed with 10 ml of ice cold water, and immediately mixed with 2 ml of the antibody solution. The mixture was rotated gently "head-over-tail" for 1 hour at room temperature. To block any remaining active esters, 50 µl of 1 M ethanolamine-HCl, pH 8.0 was added and incubation continued for 30 min at room temperature. The mixture was transferred to a disposable Econo-column and washed with 10 ml of PBS. The column was stored at 4°C in PBS containing 0.02% sodium azide.

5.21. Coupling of antibodies to Affi-Prep Hz hydrazide support

Affi-Prep Hz Hydrazide support (Bio-Rad) was used according to the recommendations of the manufacture to couple antibody directed against RBM to a solid phase. The beads charged with α-RBM were used for immunoprecipitations. The procedure consists of two steps. The first step is the periodate oxidation of vicinal hydroxyls in the sugar residues of the carbohydrates localised on the Fc region of IgG to form aldehydes. The second is a covalent coupling between aldehyde groups of IgG and hydrazide groups of the Affi-Prep Hz support.

Special solutions:

1) Oxidation buffer:

20 mM sodium acetate, pH 5.0
0.15 M NaCl

2) Coupling buffer:

0.1 M sodium acetate, pH 4.5
1.0 M NaCl

3) Washing buffer:

20 mM sodium phosphate, pH 7.0
0.5 M NaCl

4) Sodium periodate stock solution:

0.5 M $NaIO_4$

5) 50% glycerol

Procedure:

Antiserum was affinity purified on Protein A Sepharose as described below. 2 ml of the eluate from the Protein A column (approx. 2 mg/ml) was dialysed against 2 litres of oxidation buffer overnight at 4°C. To oxidise the IgG, 40 µl of the sodium periodate stock solution was added to 2 ml of antibodies and the mixture was incubated for 45 min at room temperature in a tube wrapped in foil. The reaction was stopped by the addition of 0.2 ml of 50% glycerol. The oxidised IgG was dialysed against 2 litres of coupling buffer at 4°C overnight and then mixed with 1 ml of Affi-Prep Hz beads equilibrated with coupling buffer just prior to this step. The mixture was rotated gently "head-over-tail" overnight at 4°C. The Affi-Prep Hz/IgG slurry was transferred to a disposable Econo-column (Bio-Rad) and washed with 10 ml of washing buffer. The column was stored at 4°C in washing buffer containing 0.02% sodium azide.

5.22. Preparation of extracts from the testis tissues for western blotting

The testis tissues from chicken, mouse, dog and human were obtained surgically and stored at -70 °C. Frozen tissues were directly ground to a powder in an electrical tissue homogeniser. The ground tissues were washed to remove traces of blood in ice-cold PBS until the solution was colourless. They were then sonicated in urea buffer (50 mM Tris-HCl, pH 7.5, 7.5 M urea, 0.5 M NaCl, 1 mM EDTA, 2 mM DTT) with 3 bursts of 15 seconds. Debris

was removed by centrifugation. The concentration of proteins in the supernatant was measured by Bradford assay.

5.23. Expression and purification of the Trx-RBM protein

All procedures were performed according to the recommendations of the manufacture (Invitrogen).

1). Cloning:

The PCR product containing the full-length RBM coding sequence was amplified from the MK5 plasmid using the primers:

5'-Kpn I/RBM: GCAGGTTGGTACCCATGGTAGAAGCAGATCATC

3'-RBM/Bam HI: GGATCGGGATCCTTAATATCTGCTCGAGTC,

and cloned into the Kpn I/Bam HI sites of the pTrxFus plasmid. The resulting construct was sequenced to confirm the absence of mistakes, and designated pTrx-RBM.

2). Expression:

Stock solutions:

a) 10xM9 Salts: 60 g Na_2HPO_4, 30 g KH_2PO_4, 5 g NaCl, 10 g NH_4Cl per 1 litre. The chemicals were dissolved in water and the pH 7.4 was adjusted with NaOH. The solution was autoclaved and stored at room temperature.

b) 1 M $MgCl_2$.

c) 50% glucose sterilised by filtration.

d) Buffer 1: 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 0.1 mM DTT.

RGM medium: 1xM9 Salts, 0.2% Casamino Acids, 0.5% glucose, 1 mM $MgCl_2$.

GI724 cells were transformed with the pTrx-RBM plasmid by electroporation. A single colony was used to start the overnight culture. Cells were grown at 30°C in RGM medium containing 0.1 mg/ml ampicillin. 10 ml of fresh RGM-Amp medium was inoculated with 0.5 ml of the overnight culture and cells were grown at 30°C until an OD_{550} of 0.5. To induce the expression of Trx-RBM protein, tryptophan was added to a final concentration of 0.1 mg/ml and the cultures were further incubated at 37°C. To analyse the expression, 1.5 ml of culture was withdrawn every hour after induction. Cells were pelleted by centrifugation at 13.000 rpm for 1 minute, resuspended in 0.2 ml Buffer 1 and lysed by sonication. The total

cell extract was centrifuged for 5 min at 13,000 rpm, and the supernatant was analysed by 10% SDS-PAGE. The protein concentration was estimated by Bradford assay and 10 μg were applied per each lane.

3). Purification by osmotic shock:

Osmotic Shock Solutions were supplied by the manufacture:

a) Solution 1: 20 mM Tris-HCl, pH 8.0, 2.5 mM EDTA, 20% sucrose.

b) Solution 2: 20 mM Tris-HCl, pH 8.0, 2.5 mM EDTA.

After induction with tryptophan, 1.5 ml aliquots were taken every hour as described above. Cells were pelleted and stored on ice (osmotic shock works best on fresh cells). Cell pellets were resuspended in Solution 1 to an OD_{550} of 5.0 (0.15 ml of Solution 1 per 1.5 ml of culture with an OD_{550}=0.5). Cells were incubated in Solution 1 for 10 min on ice, then centrifuged for 1 min. Cell pellets were resuspended in Solution 2 using the same volume as for Solution 1, incubated on ice for 10 min, and centrifuged for 10 min. The supernatant was removed and designated the shock fluid. The extracts from the cell pellets were prepared by sonication as described above. The presence of the Trx-RBM protein in the shock fluid or in the cell pellet was analysed by 10% SDS-PAGE. The protein was found exclusively in the cell pellet.

4) Purification by affinity chromatography using ThioBond resin:

Running buffer: 50 mM Tris-HCl, pH 7.9, 500 mM NaCl, 1 mM EDTA.

The ThioBond resin, approximately 2 ml (bed volume), was activated in 10 ml of running buffer containing 20 mM β-ME for 60 min at room temperature by "head-over-tail" rotation, and then equilibrated with 5 volumes of running buffer without β-ME. Cells were grown in 200 ml of culture and induced as described above. At 3 hours after induction, cells were collected by centrifugation, washed in 200 ml of running buffer, and resuspended in 10 ml of running buffer containing 1 mM β-ME. Cells were opened by sonication and centrifuged for 10 min at 4000 rpm in a bench-top centrifuge. The total amount of protein in the supernatant was approximately 50 mg as estimated by Bradford assay.

The cell lysate from a 200 ml culture (10 ml) was added to 2ml bed volume of the equilibrated ThioBond resin, and incubated in batch at room temperature for 60 minutes with "head-over-tail" rotation. Then, the resin was transferred to the column and washed with 10

ml of running buffer containing 1 mM β-ME. Proteins were eluted from the column by stepwise with 6 ml of running buffer containing increasing concentrations of β-ME: 5, 10, 50, 100, 200, 500, and 1000 mM. All of the fractions were analysed by 10% SDS-PAGE followed by Coomassie staining or western blot with α-RBM.

Chapter 4

5.24. Isolation of snRNPs from HeLa nuclear extracts

The basic protocol is described in detail by Bach *et al.* (1990) and Will *et al.* (1994). Nuclear extracts were prepared from HeLa cells by the method of Dignam *et al.* (1983) as described above. The nuclear extract in Dignam buffer C was centrifuged at 55,000 rpm in a Beckman Ti70 rotor for 30 min at 4°C to remove large aggregates. The supernatant was filtered through a 1.2 μm membrane and diluted with C buffer (20 mM HEPES, pH 7.9, 250 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTE, 0.5 mM PMSF) without glycerol to reduce the concentration of glycerol to 10%. Alternatively, the extract was overlayed with 2-butanol, centrifuged and the supernatant was carefully removed with a syringe. The supernatant was loaded onto a CNBr-activated Sepharose 4B column with covalently linked H20 anti-m_3G monoclonal antibody. Basic calculations are that a 1 ml column, containing approximately 6 mg of antibody per 1 ml of Sepharose, is loaded with 15 ml of extract. The column was washed with 5 volumes of C buffer, containing 5% glycerol, and the snRNPs were eluted with 1 volume of 20 mM m^7G nucleoside.

Total snRNPs, containing U1, U2, U4/U6 and U5 snRNPs, were separated into 12S U1/U2, 20S U5 and 25S [U4/U6.U5] snRNPs by 10-30% glycerol gradient centrifugation under conditions described in Laggerbauer *et al.* (1996). For preparative fractionation, 4 mg of total snRNPs (measured by Bio-Rad protein assay) were layered onto a 10-30% (w/w) gradient in buffer K150 (20 mM HEPES, pH 7.9, 150 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTE), and centrifuged at 28,000 rpm in a Beckman SW28 rotor for 17 hours at 4°C. 0.5 ml fractions were collected and 0.1 ml aliquots were withdrawn from every second fraction for the PAGE analysis of the protein and RNA content.

For analytical purposes, 1 mg of total snRNPs were applied to a linear 4 ml, 10-30% glycerol gradient in Na150 buffer (20 mM HEPES-KOH, pH 7.9, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM DTE) and centrifuged in a Beckman SW60 rotor at 29,000 rpm for 14 hour at 4°C. 29 fractions of 150 µl each were harvested manually from top to bottom. 50 µl aliquots were directly subjected to 10% SDS-PAGE followed by Coomassie staining. RNA was recovered from 100 µl aliquots RNA by phenol-chloroform extraction followed by ethanol precipitation. RNA was analysed by 10% urea-PAGE and visualised by silver staining.

5.25. Preparation of antigens for immunisation by coupling peptides to a carrier protein

α-61kD antibodies were raised in rabbits immunised with a peptide conjugated with ovalbumin using *m*-Maleimidobenzoyl-*N*-hydroxysuccinimide ester (MBS) according to the recommendations of the manufacture (Pierce). MBS contains an *N*-hydroxysuccinimide (NHS) ester that reacts with primary amines (only N-terminus and lysine residues react significantly with the NHS-ester). The maleimide group reacts selectively with sulfhydryl groups. The procedure consists of two steps: 1) modification of ovalbumin with MBS, and 2) coupling of cysteine residues of the peptide to maleimide groups of the MBS-modified ovalbumin. In order to introduce a sulfhydryl group, a designed peptide contains an additional cysteine at its N- or C-terminus.

Stock solution:

1) 6.6 mg/ml albumin from chicken eggs (Sigma, A-7641) in PBS, pH7.0.

2) 3.3 mg/ml MBS (Pierce, cat#22311, in water-free DMSO).

3) 20 mg/ml peptide in water (usually, a 16 amino acid peptide containing an N- or C-terminal cysteine).

Procedure:

3 ml of ovalbumin stock solution were mixed with 0.3 ml of a freshly prepared MBS stock solution and incubated "head-over-tail" at room temperature for 30 min. The protein was separated from MBS by gel filtration on a PD-10 column (Pharmacia) equilibrated with PBS, pH 7.0. Peak fractions were combined. 3 ml of modified ovalbumin with a concentration of 1.7-2.0 mg/ml was recovered. 1 ml of modified ovalbumin was mixed with 0.3 ml of

peptide stock solution and incubated "head-over-tail" at room temperature for 30 min. The protein-peptide conjugates were separated from free peptide by gel filtration on a PD-10 column (Pharmacia) equilibrated with PBS, pH7.0. 3 ml of protein with a concentration 0.7-0.8 mg/ml was recovered. The crosslinking of peptide to ovalbumin was confirmed by 10% SDS-PAGE. For immunisation of a rabbit, 0.5 mg of protein was injected 4 times on day 0, 28, 42, and 56, and the serum was collected by exsanguination (day 70).

5.26. Affinity purification of peptide antibodies

Peptide antibodies against the tri-snRNP-specific 61kD protein were affinity purified using SulfoLink coupling gel according to the recommendations of the manufacture (Pierce). SulfoLink coupling gel consists of immobilised iodoacetyl on a crosslinked agarose support, and thus, peptides or proteins containing sulfhydryl groups bind specifically to the SulfoLink support. The 12-atom spacer arm reduces steric hindrance, and this long arm is thought to be ideal for conjugating small peptides containing C- or N-terminal cysteine residues.

Special solutions:

1) Sample buffer: 0.1 M sodium phosphate, pH 6.0, 5 mM EDTA.

2) Coupling buffer: 50 mM Tris-HCl, pH 8.5, 5 mM EDTA.

3) Washing solution: 1.0 M NaCl.

4) Cysteine solution (0.05 M): 15.8 mg L-cysteine-HCl in 2 ml of Coupling buffer.

Procedure:

A) *Coupling of peptide to SulfoLink support.*

Dried peptide was dissolved in coupling buffer and used directly for immobilisation on the support. If a cysteine containing peptide or protein was stored in an aqueous solution, the SH groups were reduced by incubation in the presence of 1% β-mercapthoethanol (β-ME) for 30 min at room temperature. To remove the excess of β-ME, 5 mg of reduced peptide in 0.5 ml of PBS/β-ME was applied to a 5-ml desalting column (Pierce) equilibrated with Coupling buffer. Peak fractions were combined (3 ml), mixed with 2 ml (bed volume) SulfoLink support, and incubated "head-over-tail" for 30 min at room temperature. The column was washed with 6 ml of coupling buffer. Non-specific binding sites on the support were then blocked by cysteine. 2 ml of the cysteine solution were applied onto the column and incubated

for 30 min at room temperature. The column was washed with 4x4 ml of washing solution and 3x4 ml 0.05% NaN_3. The column was stored at 4°C in 0.05% NaN_3.

B). *Affinity purification of antibodies.*

This procedure was carried out at room temperature. The column was washed with 6 ml of PBS, pH 8.0. 0.5 ml of 10xPBS was added to 5 ml of rabbit antiserum. The antiserum was filtered through a 0.2 μ filter, divided into 4 portions of 1.3 ml. Each portion was applied sequentially onto the 2 ml SulfoLink column with immobilised peptide, incubated for 30 min, and finally washed with 16 ml of PBS. The antibodies were eluted with 100 mM glycine, pH 2.7. 0.5 ml fractions were collected and immediately neutralised with 25 μl of 1 M Tris-HCl, pH 9.5. The peak fractions were combined, dialysed against Dignam buffer D (20 mM HEPES, pH 7.9, 100 mM KCl, 1.5 mM $MgCl_2$), and stored in 100-μl aliquots at -20°C. The column was washed with PBS, equilibrated with 0.05% NaN_3 and stored at 4°C. The purification procedure was repeated 3 times with the same column.

5.27. Production of anti-U5-102kD antibody and immunoblotting

Polyclonal antibodies were raised in a rabbit, immunised with recombinant protein that corresponds to amino acids 614-941 of U5-102kD fused to GST through a glycine hinge (GSGGG) introduced by PCR. The fragment, amplified from plasmid DNA of the EST# H15523 using the 5'Gly/BamHI primer (5'-CGT GGA TCC GGA GGA GGC GCC AAG TCC AAG TGG CTG-3') and the 3'/EcoRI primer (5'-CGG GAA TTC CGA AGG TGT TCT TGA TGC GGC C-3'), was cloned into BamHI/EcoRI sites of the pGEX-His6 plasmid (the derivative from pGEX-4T-2, Pharmacia), containing a 6xHis-tag inserted between the Sal I and Not I sites. The constructed plasmid encoded a protein containing GST at the N-terminus and a His6-tag at the C-terminus.

The recombinant protein was expressed in *E. coli* BL21(DE3)pLysS cells, purified from a soluble cell extract on a Glutathione Sepharose (Pharmacia) according to the recommendations of the manufacture. A single colony from the freshly transformed cells was used to inoculate an overnight culture. 5 ml of overnight culture was transferred to 500 ml of LB medium containing 100 μg/ml ampicillin. Cells were grown at 30°C until an OD_{600} of 0.6 and then induced by the addition of IPTG to a final concentration of 0.5 mM. At 4 hours

after induction, cells were harvested by centrifugation at 6,000 rpm for 10 minutes (Sorvall GSA rotor), washed twice with ice-cold PBS buffer, and resuspend in 50 ml of PBS. The cells were lysed by sonication and Triton X-100 was added to the cell lysate to a final concentration of 1%. The cell lysate was clarified by centrifugation at 2,500 rpm for 15 min in a bench-top centrifuge. The supernatant was added to 1 ml of 50% slurry of Glutathione Sepharose 4B equilibrated with PBS. The suspension was rotated gently at 4°C for one hour and then transferred to a Poly-Prep chromatography column (Bio-Rad), and washed with 20 ml of PBS. The fusion protein was eluted with Glutathione elution buffer (10 mM reduced glutathione in 50 mM Tris-HCl, pH 8.0). Fractions of 0.5 ml were collected and the protein concentration was determined by Bio-Rad protein assay. The protein composition of fractions was analysed by 10% SDS-PAGE. This fusion protein was used for immunisation of a rabbit according to the protocol described for RBM.

5.28. Immunoprecipitation of snRNPs from HeLa nuclear extract

This procedure is based on the protocol described by Will *et al.* (1994). To immunoprecipitate snRNPs from HeLa nuclear extract with antibodies specific for snRNP proteins (α-61kD, α-U4/U6-60kD, α-U5-102kD and α-U5-116kD were used in this work), 100-200 µl of antiserum were coupled to 50 µl of protein A Sepharose beads (PAS, Pharmacia) in 900 µl of PBS containing 0.1 mg/ml BSA overnight at 4°C. The beads were washed three times with PBS, aliquoted to 10 µl (bed volume), and diluted with 400 µl of IPP buffer (20 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 0.05% NP-40), containing different concentrations of NaCl (150-500 mM as indicated in the figures) to vary the stringency of immunoprecipitation. 60 µl of nuclear extract were then added to each aliquot and the PAS-antibody conjugates were incubated at 4°C for 2 hours with constant rotation. Subsequently, the beads were washed five times (the tubes were change once) with IPP, containing the corresponding NaCl concentration, and the RNAs bound to the beads were extracted with phenol-chloroform and precipitated with ethanol. Precipitated RNAs were 3' end-labelled with [^{32}P]-pCp according to England and Uhlenbeck (1978), separated by 10% urea-PAGE, and visualised by autoradiography. Alternatively, precipitated RNAs were subjected to 10% urea-PAGE followed by northern blotting.

5.29. *In vitro* translation and radioimmunoprecipitation assays

To produce the [^{35}S]-labelled proteins, coupled transcription and translation reactions were performed *in vitro* using the Promega TNT system according to the recommendations of the manufacture. A typical incubation mixture of 50 µl contained:

25 µl of TNT rabbit reticulocyte lysate
2 µl of TNT reaction buffer
1 µl of amino acid mixture minus methionine, 1mM of each amino acid except Met
4 µl [^{35}S]-methionine, 10 mCi/ml (Amersham)
1 µl RNasin (Promega)
1 mg of DNA template
1µl of T7 RNA polymerase

The reaction was incubated for 2 hours at 30°C, aliquoted and stored at –80°C.

Radioimmunoprecipitations of the labelled, *in vitro*-translated proteins, incorporated into snRNP particles or as specific protein complexes, were performed with a monoclonal Y12 antibody, directed against Sm-proteins, or with α-U5-116. 5 to 10 µl of the [^{35}S]-labelled *in vitro* translated protein was pre-incubated at room temperature for 30 min with 1 pmol of snRNP particles or 400 ng of the protein complexes (a [220/116] dimer or a [220/200/116/40] tetramer) in 60 µl of IPP_{150} buffer (20 mM HEPES, pH 8.0, 150 mM NaCl, 1.5 $MgCl_2$, 0.5 mM DTT, 0.05% NP40). The complexes were precipitated with 10 µl of PAS beads (Pharmacia) coupled to antibodies for 30 min at room temperature in 400 µl IPP buffer (alternatively, for 2 hours at 4°C). The beads were washed as described above (for immunoprecipitation of snRNPs from nuclear extract) and then dried. Precipitated proteins were eluted from the beads in SDS loading buffer, separated on 10% SDS-PAGE, and visualised by fluorography.

5.30. Fractionation of the [U4/U6.U5] components on a high salt glycerol gradient

Fractions containing 25S [U4/U6.U5] tri-snRNPs from a preparative gradient of total snRNPs, were combined, pelleted in a Beckman TLA100.3 rotor at 70,000 rpm for 4-6 hours, and resuspended in Na700 buffer (20 mM HEPES-KOH, pH 7.9, 700 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM DTE). 200 µg of tri-snRNPs (measured by Bio-Rad protein assay) were applied to a linear 4 ml, 5-20% glycerol gradient in Na700 buffer and centrifuged in a

Beckman SW60 rotor at 34,000 rpm for 14 hours 20 min. 24 fractions of 175 µl each were harvested manually from top to bottom, and the pellet was resuspended in 150 µl of Na700 buffer, containing 1% SDS. 20 µl aliquots were directly analysed by 10% SDS-PAGE followed by either Coomassie staining or electroblotting to a membrane which was then immunostained. RNA was extracted from 50 µl aliquots with phenol-chloroform, ethanol precipitated, separated by 10% urea-PAGE, and visualised by silver staining or northern blotting.

5.31. Isolation of the 14S U4/U6 snRNP particle

Fractions containing 25S [U4/U6.U5] tri-snRNPs from a preparative gradient of total snRNPs were combined, pelleted in a Beckman TLA100.3 rotor at 70,000 rpm for 4-6 hours, and resuspended in SCN200 buffer (20 mM HEPES-KOH, pH 7.9, 200 mM NaSCN, 1.5 mM $MgCl_2$, 0.5 mM DTE). Approximately 200 µg of tri-snRNPs in 400 µl of SCN200 buffer were layered onto a linear 4 ml, 5-20% glycerol gradient in Na150 buffer (20 mM HEPES-KOH, pH 7.9, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM DTE) and centrifuged in a Beckman SW60 rotor at 35,000 rpm for 14 hours 25 minutes. 24 fractions of 175 µl each were harvested manually from top to bottom. Identical fractions from 6 gradients (6 buckets is the maximum capacity of the Beckman SW60 rotor) were combined. 60 µl aliquots were directly analysed by 10% SDS-PAGE and proteins were visualised by Coomassie or silver staining. To analyse the low molecular weight protein composition, 60 µl aliquots were subjected to 13% SDS-PAGE followed by electroblotting and immunostaining. RNA was extracted from 100 µl aliquots with phenol-chloroform, ethanol precipitated, separated by 10% urea-PAGE, and visualised by northern blotting with [^{32}P]-labelled probes specific to the snRNAs as described below.

5.32. Isolation of RNA-free protein complexes from U5 snRNPs

Fractionation of U5 snRNPs components in thiocyanate-containing glycerol gradients was carried out as described by Achsel *et al.* (1998). Fractions containing 20S U5 snRNPs were pelleted by centrifugation as described above and resuspended in SCN400 buffer (20 mM HEPES-KOH, pH 7.9, 400 mM NaSCN, 1.5 mM $MgCl_2$). To isolate the [220/116]

dimer, 200 μg of U5 snRNPs in 400 μl of SCN400 buffer were layered onto a linear 4 ml, 5-20% glycerol gradient prepared with SCN400 buffer. For purification of the [220/200/116/40] tetramer, 200 μg of U5 snRNPs in 400 μl of SCN400 buffer were layered onto a linear 4 ml, 10-30% glycerol gradients prepared with SCN200 buffer (20 mM HEPES-KOH, pH 7.9, 200 mM NaSCN, 1.5 mM $MgCl_2$). The gradients were centrifuged in a Beckman SW60 rotor at 33,000 rpm for 15 hours. 29 fractions of 150 μl each were harvested manually from top to bottom, and 50 μl aliquots were analysed by 10% SDS-PAGE. The typical gradient profiles for isolation of the [220/116] dimer and the [220/200/116/40] tetramer are shown in Achsel *et al.* (1998) in Fig.1 and Fig.2, respectively.

5.33. The tri-snRNP reconstitution assay

The 14S U4/U6, 10S U4/U6, and U5 snRNP particles were obtained by treatment of tri-snRNPs with NaSCN or NaCl, followed by glycerol gradient centrifugation (Fig.4.15, and Fig.4.11, respectively). Fractions containing 14S (14-15, Fig.4.15) or 10S snRNPs (9-12, Fig.4.11) were pooled together. For the reconstitution, 60 μl of 14S or 10S U4/U6 (0.5 pmol) were combined with 3 μl of U5 (0.5 pmol; fractions 19-21, Fig.4.11). The reconstitution was carried out for 30 minutes at 30°C, and analysed by co-immunoprecipitation with α-116kD or α-61kD antibodies as described previously.

Common Procedures

5.34. Different gel systems

1) Urea-PAGE

For the analysis of RNA, the sequencing type of gel was used. The stock polyacrylamide (PAA) solution contained 38% polyacrylamide with 2% of bisacrylamide (19:1). The gel mixture contained 7 M urea, 1xTBE, and polyacrylamide at the desired concentration. For the purification of RNA from a transcription reaction, 4-5% of PAA was used. For the analysis of splicing, 10-14% of PAA was used. The gel mixture was polymerised by the addition of 1/100 volume of 10% amoniumpersulfate (APS) and 1/1000 of TEMED. 0.5xTBE was used as the running buffer.

2) SDS-PAGE

For the analysis of proteins, SDS-PAGE was carried out. The stock PAA solution contained 30% acrylamide with 0.8% of bisacrylamide (38.5:1). The separation gel contained 8-13% of PAA, 1xSeparationBuffer, pH 8.8, and was cast by the addition of 1/300 volume of both 10% APS and TEMED. The stacking gel contained 5% of PAA, 1xStackingBuffer, pH 6.8, and was cast by the addition of 1/200 volume of 10% APS and 1/400 volume of TEMED. 1xTGS was used as a running buffer. The gel was run at 25 mA until samples were in the stacking gel, and then at 50 mA. (All buffers are found in Basic Solutions).

3) Native gel electrophoresis

The composite gel was used for the analysis of splicing complexes. The gel-form was prepared with 1 mm spacers. The gel mixture contained 0.5% agarose and 3.5% PAA. It was prepared in the following manner. 12.6 ml of 20% PAA (80:1 Bis), 8.3 ml of 87% glycerol, 2.2 ml of 10xTBE, 0.96 ml of 10% APS, and 11.9 ml of water were combined in an Erlenmeyer flask. 36 ml of 1% agarose dissolved in water and still hot were quickly mixed with the above solution. 35 µl of TEMED was added, and the gel was poured. The gel was left to set for 45 minutes at room temperature. Samples were prepared by pipetting 10 µl of splicing reaction into the tube containing 2 µl of heparin (5 mg/ml) and 2 µl of glycerol (87%). Samples were kept on ice and, before loading, were incubated 10 minutes at room temperature. The gel was run in 0.3xTBE at 25 mA for 5 hours. Autoradiography was performed overnight with a wet gel at –80°C.

5.35. Western and Far-Western blot analysis of proteins

Special solutions:

1) Transfer buffer: 48 mM Tris-base, 39 mM Glycine, 0.0375% SDS, 20% Methanol. To prepare 100 ml transfer buffer, 10 ml of a 10xStock solution (56 g Tris-base, 29.3 g Glycine, 37.5 ml 10% SDS per 1 litre) were mixed with 70 ml of water and 20 ml of methanol. The solution was chilled before use.

2) Incubation buffers. Two incubation buffers were used: D-Glu buffer (Leicester) and TBS buffer (Marburg), and no differences were observed between these two buffers. D-Glu buffer: 20 mM Triethanolamine, pH 7.9, 80 mM Potassium glutamate, 0.2 mM EDTA, 5% Glycerol. TBS buffer: 20 mM Tris-HCl, pH 7.5, 150 mM NaCl.

Procedure:

Proteins were separated by SDS-PAGE (pre-stained molecular weight markers from Bio-Rad were applied onto one lane in order to control for blotting) and transferred to Hybond C or Hybond P membrane using a semi-dry blotting apparatus (Biometra). The gel, membrane and sheets of 3MM paper (Whatman) were soaked in transfer buffer, and the sandwich containing the gel and membrane was placed between sheets of 3MM paper, 9 on each side. The transfer was performed for 40 min at 10 W. The filter was stained with Ponceau S solution (1 % Ponceau S, 2.5 % Trichloroacetic acid, 5 % acetic acid or ready-to-use Serva Ponceau S solution) to visualise the transferred protein. To decrease non-specific binding of antibodies, the membrane was blocked in incubation buffer (D-Glu or TBS) containing 0.1% Tween-20 and 5% non-fat dried milk for 2 hours at room temperature or at 4°C overnight on a rocking platform. All further manipulations were carried out in incubation buffer containing 0.1% Tween-20 and 1% non-fat dried milk. After blocking, the filter was washed 3 times, and then incubated with primary antibody against the protein of interest for 1 hour at room temperature on a rocking platform. 0.3-0.4 ml of solution was used per 1 cm^2 of the membrane. 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 3 times for 15 min each and then incubated with either a secondary antibody or protein A, conjugated to horseradish peroxidase (HRP), in a manner as described for a 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 3 times and developed with enhanced chemiluminescence (ECL) detection reagents (Amersham). The Amersham ECL system is based on the emission of blue light during the HRP/H_2O_2 catalysed oxidation of luminol. Equal volumes of Amersham ECL detection reagents 1 and 2 (50-100 µl per 1 cm^2 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 the previous one. For this purpose, the membrane was incubated in 62.5 mM Tris-HCl, pH 6.8, 2% SDS, and 100 mM β-mercaptoethanol for 30 min at 50°C.

For far western (or overlay), a pre-blocked membrane was incubated with *in vitro* translated protein (for 10 ml of incubation buffer 50 µl of *in vitro* translation-reaction was used) for 2-3 hours at 4°C. The membrane was washed three times with incubation buffer, dried and exposed to a PhosphoImager or X-ray film.

5.36. Purification of antibodies on Protein A Sepharose

Protein A Sepharose CL-4B (Pharmacia) contains protein A immobilised by the CNBr method. Protein A binds to the Fc region of immunoglobulins through interactions with the heavy chain. The binding of Protein A has been well documented for IgG from a variety of mammalian species and for some IgMs and IgAs as well. The protein A Sepharose used was pre-swollen and supplied in 20% ethanol. Before use, it was washed with the binding buffer, PBS or 50 mM Tris, pH 7.5.

The serum was buffered by adding 1/10 volume of 10xPBS, pH8.0, and added to the beads in a 2/1 ratio. The binding was performed in a cold room for 2 hours in batch with constant rotation. Then, the beads were packed into the column. The column was washed with 10 volumes of PBS and antibody was eluted with 0.1 M glycine, pH3.0 in 0.5 ml steps. The pH was neutralised by the addition of 50 µl of 1 M Tris-HCl, pH9.0. The concentration of antibodies was determined by Bradford assay.

5.37. Preparation of protein A bead- antibody affinity columns

Alternatively, after the binding of antibodies to the protein A sepharose, they were covalently crosslinked to the protein A using dimethylpimelimidate. For this purpose, the beads were washed after binding with 10 volumes of 0.2 M sodium borate, pH 9.0. The beads were resuspended in 10 volumes of 0.2 M sodium borate, pH 9.0 and the equivalent of 10 µl of beads was removed for future analysis. Solid dimethylpimelimidate was added to a final concentration of 20 mM, and coupling was carried out for 30 min at room temperature with constant mixing. After coupling, the equivalent of 10 µl beads was removed for subsequent analysis. The reaction was stopped by washing the beads once in 0.2 M of ethanolamine, pH 8.0 and then incubated for 2 hours at room temperature in 0.2 M ethanolamine with gentle mixing. The beads were washed 3 times with 10 volumes of PBS and resuspended in PBS

with 0.01% azide. The efficiency of coupling was checked by boiling the samples of beads taken before and after coupling in SDS-PAGE loading buffer and analysing them by 10% SDS-PAGE. Good coupling is indicated by heavy-chain bands (55 kDa) in the 'before' but not in the 'after' lanes. Additionally, the beads could be washed with 0.1 M glycine, pH3.0, to remove any antibodies that were not bound covalently to protein A.

5.38. Northern blot analysis of snRNAs

This procedure was used to visualise snRNAs immunoprecipitated from nuclear extract or in the reconstitution assay.

1) Transfer to nylon membrane

RNA was precipitated with ethanol and separated by 10% urea-PAGE on a 1-mm thick gel. RNA was transferred to a nylon membrane (PALL, GmbH) by the wet procedure using a PROTAN apparatus (Bio-Rad). The transfer was carried out in transfer buffer (25 mM sodium phosphate, pH 6.5), at 1-2 W, overnight at 4°C. After transfer, the RNA was crosslinked to the membrane by UV- irradiation. The membrane was stored dry until the hybridisation procedure.

2) Preparation of radioactive probes

Plasmids encoding the snRNAs (described below) were used as DNA templates to generate the [^{32}P]-labelled probes using the random primer labelling kit, Prime-It II (Stratagene), according to the recommendations of the manufacture. The procedure relies on the ability of random nonamer primers (9-mer) to anneal to multiple sites along the length of a DNA template, and on the 3' exonuclease-deficient mutant of the Klenow fragment of DNA polymerase I (Exo(-) Klenow), the enzyme that synthesises DNA by incorporating nucleotide monophosphates at the free 3'-OH group provided by the primer.

For annealing, a mixture of 25 ng of DNA template and 10 µl of random 9-mer primers (27 OD units/ml) in 34 µl of water was heated in a boiling water-bath for 5 min, and left for 10 min at room temperature. Then, 10 µl of 5x*dATP buffer, containing 0.1 mM of dCTP, dGTP, and dTTP, 5 µl of [α-^{32}P]dATP (10 mCi/ml; 3,000 Ci/mmol), and 1 µl of Exo(-) Klenow enzyme (5 U/µl) were added to the annealing mixture, and the reaction was

incubated at 37°C for 5 min. Non-incorporated nucleotides were removed using PROBE spin columns (Pharmacia) according to the recommendations of the manufacture.

3). Pre-hybridisation and hybridisation

Hybridisation solution:

1.25 ml 0.5 M sodium phosphate, pH 6.5
7.5 ml 20x SSC
1.25 ml 100x Denhardt's
1.25 ml 10% SDS
12.5 ml formamide
1.25 ml water

0.25 ml of salmon sperm DNA (10 mg/ml) boiled for 5 min was added to the solution pre-warmed at 42°C.

After the transfer, the nylon membrane (10x10 cm) was pre-hybridised in 12.5 ml of the hybridisation solution for 2 hours at 42°C in a hybridisation tube with constant rotation. The radioactive probe ($5x10^6$ cpm) was denatured in 1 ml of hybridisation solution for 5 min at 95°C, and added to 12.5 ml of fresh hybridisation solution pre-warmed to 42°C. Hybridisation was carried out for 24 hours at 42°C.

The membrane was washed as follows:

a) 2 times (2x) with 250 ml of 2x SSC, 0.5% SDS for 5 min at room temperature with constant shaking;

b) 2x with 250 ml of 2x SSC, 0.1% SDS for 5 min at room temperature with constant shaking;

c) 1x with 300 ml of 2x SSC, 0.1% SDS for 30 min at 50°C with constant shaking.

The wet membrane was wrapped in Saran wrap and exposed to film for 4-6 hours.

4). De-hybridisation

Prior to hybridisation with a different probe, the membrane was stripped by boiling for 10 minutes in 0.1xSSC, 0.1% SDS. The wet membrane was stored in a plastic bag at –20°C.

5.39. Labelling of RNA with [^{32}P]-pCp

Immunoprecipitated RNA was visualised by [^{32}P]-pCp labelling according to the procedure described by England and Ulenbeck (1978). In particular, RNA was recovered from the beads by phenol-chlorophorm extraction, and ethanol precipitated in the presence of 10 µg of glycogen. The pellet was resuspended on ice in 10 µl of a cold mixture containing:

1 µl 10xBuffer for T4 RNA ligase (New England BioLabs)
0.5 µl [^{32}P]-pCp (10mCi/ml; 3000 Ci/mmol, Amersham)
2.5 µl DMSO
0.5 µl T4 RNA ligase (20 units/µl, New England BioLabs)

The reaction was incubated overnight on ice in a cold room and stopped by addition of 10 µl of Formamide RNA loading buffer. RNA was analysed by 10% urea-PAGE followed by autoradiography.

5.40. Staining of polyacrylamide gels

1) Coomassie Blue R-250 staining of protein gels

Staining solution: 0.4 g Coomassie brilliant blue R-250, 360 ml methanol, 360 ml water, 73 ml glacial acetic acid.

De-staining solution: 5% methanol, 7% glacial acetic acid.

The gel was soaked for 2 hours in the staining solution on a gentle rocking platform. The gel was de-stained in the de-staining solution until the background was clear. Gels could be stored in this solution for one year at room temperature (without fading of the protein bands).

2) Silver staining of protein gels

The standard silver staining procedure was performed as described by Blum *et al.* (1987). In this procedure, at least a 10-fold volume of the gel is required for all solutions. The gel was fixed in 50% methanol, 12% acetic acid for at least two hours (preferably overnight). To remove the acetic acid, the gel was washed twice with 50% ethanol and once in 30% ethanol for 20 minutes each to avoid decomposition of thiosulfate and subsequent Ag_2S formation. It was then incubated for exactly 60 seconds with 0.8 mM $Na_2S_2O_3$ (0.02%) and washed (3x20 sec) with water to remove thiosulfate from the surface. The gel was impregnated for 20 minutes with 2g/l $AgNO_3$, 0.026% formaldehyde. The formaldehyde in this step increases the sensitivity. During this step, the gel appeares yellowish due to thiosulfate impregnation. The gel was washed subsequently with water (3x20 sec) and developed with 60 g/l Na_2CO_3, 0.0185% formaldehyde, 16 µM $Na_2S_2O_3$. If the developing solution turned brown, it was replaced immediately with fresh solution. Typically, gels were

developed for 2 to 10 minutes. Finally, the gel was washed with water for 2 minutes and development was stopped with 50% methanol, 12% acetic acid.

Another procedure described by Nesterenko *et al.* (1994) is a modification of the Blum silver stain method (Blum *et al.*, 1987) that allows the detection of proteins in polyacrylamide gels in 30 minutes. In this case, the protein gel was fixed in a solution containing 30 ml of water, 30 ml of acetone, 0.75 ml of 100% TCA, and 25 µl of 37% formaldehyde by incubating on a gentle rocking platform for 10 min with one change of the solution (5 min + 5 min). The gel was washed 3 times (3x5sec) with water, incubated for 5 min in water, and then, for 5 min in a solution containing 50 ml of acetone and 50 ml of water. Next, the gel was soaked for 1 min in a solution containing 100 ml of water, and 170 µl of 6.4 % $Na_2S_2O_3$, and washed 3 times (3x5sec) with water. The gel was stained for 8 min with a solution containing 100 ml of water, 1 ml of 37% formaldehyde, and 0.28 g of $AgNO_3$ (or 1.33 ml of 20% $AgNO_3$). After washing with water (2x5sec), the gel was developed for 10-20 sec with a solution containing 100 ml of water, 2 g of Na_2CO_3, and 42 µl of 37% formaldehyde, 42 µl of 6.4 % $Na_2S_2O_3$. The gel was quickly washed with water and the staining was stopped with 1% glacial acetic acid.

3). Silver staining of RNA gels

The silver staining of RNA gels was carried out according to the procedure described by Merril and Goldman (1982). The gel was fixed with 40% methanol, 10% acetic acid for 30 minutes and then twice with 10% ethanol, 5% acetic acid for 15 minutes. After a quick wash with water, the gel was soaked in 12 mM of $AgNO_3$ for 20 minutes. The gel was washed briefly, first, with water and then with 0.28 M Na_2CO_3, 0.0185% formaldehyde. The gel was developed with 0.28 M Na_2CO_3, 0.0185% formaldehyde for as long as necessary and staining was stopped by incubating in 5% acidic acid for 15 minutes.

DNA constructs

1) MINX is a pre-mRNA substrate originating from the Adenovirus major late transcript and was cloned into the EcoR I/BamH I sites of the pSP65 vector (Promega). *In vitro* transcription from the SP6 promoter (depicted in italics) results in the sequence shown in capital letters. The exon sequences are shown in bold, and the branch site is underlined.

a tta ggt gac act ata GGG CGA ATT CGA GCT CGC CCA **CTC TTG GAT CGG AAA CCC GTC GGC CTC CGA ACG** GTA AGA GCC TAG CAT GTA GAA CTG GTT ACC TGC AGC CCA AGC TTG CTG CAC GTC TAG GGC GCA GTA GTC CAG GGT TTC CTT GAT GAT GTC ATA CTT ATC CTG TCC CTT TTT TTT CCA CAG **CTC GCG GTT GAG GAC AAA CTC TTC GCG GTC TTT CCA GT**g gg gat cc

2) phU5

gg gag ATA CTC TGG TTT CTC TTC AGA TCG CAT AAA TCT TTC GCC TTT TAC TAA AGA TTT CCG TGG AGA GGA ACA ACT CTG AGT CTT AAC CCA ATT TTT TGA GCC TTG CCT TGG CAA GGC Ta

The genomic sequence of human U5 snRNA (depicted in capital letters) fused to the T7 promoter was cloned into the EcoR I/Hind III sites of pEMBL8 (+). The run-off product of the *in vitro* transcription reaction is shown. This plasmid was kindly provided by Dr. A. Bindereif.

3) phU6

gaa GUG CUC GCU UCG GCA GCA CAU AUA CUA AAA UUG GAA CGA UAC AGA GAA GAU UAG CAU GGC CCC UGC GCA AGG AUG ACA CGC AAA UUC GUG AAG CGU UCC AUA UUU UU g

The genomic sequence of human U6 snRNA (depicted in capital letters) was cloned into EcoR I/BamH I sites of pUC13 under SP6 promoter. The run off product of *in vitro* transcription reaction is shown. The plasmid was kindly provided by Dr. M. Green.

4) phU4

gaAGC TTT GCG CAG TGG CAG TAT CGT AGC CAA TGA GGT TTA TCC GAG GCG CGA TTA TTG CTA ATT GAA AAC TTT TCC CAA TAC CCC GCC ATG ACG ACT TGA AAT ATA GTC GGC ATT GGC AAT TTT TGA CAG TCT CTA CGG AGA CTG aattttt

The genomic sequence of human U4 snRNA (depicted in capital letters) was cloned into the EcoR I/Dra I sites of pUC13 downstream of an SP6 promoter. The run off product of the in vitro transcription reaction is shown. This plasmid was kindly provided by Dr. M. Green.

5) pHU1a

g AUA CTT ACC TGG CAG GGG AGA TAC CAT GAT CAC GAA GGT GGT TTT CCC AGG GCG AGG CTT ATC CAT TGC ACT CCG GAT GTG CTG ACC CCT GCG ATT TCC CCA AAT GTG GGA AAC TCG ACT GCA TAA TTT GTG GTA GTG GGG GAC TGC GTT CGC GCT TTC CCC TG ggc

The genomic sequence of human U1 snRNA (depicted in capital letters) was cloned as described in Patton *et al.* (1987). The run-off product of the *in vitro* transcription reaction of

Pst I-digested pHU1a using SP6 RNA polymerase is shown. This plasmid was kindly provided by Dr. T. Pederson.

6) pMRG3U2-27

ggg ATC GCT TCT CGG CCT TTT GGC TAA GATCAA GTG TAG TAT CTG TTC TTA TCA GTT TAA TAT CTG ATA CGT CCT CTA TCC GAG GAC AAT ATA TTA AAT GGA TTT TTG GAG CAG GGA GAT GGA ATA GGA GCT TGC TCC GTC CAC TCC ACG CAT CGA CCT GGT ATT GCA GTA CCT CCA GGA ACG GTG CAC Ccc

The genomic sequence of mature human U2 snRNA (depicted in capital letters) was cloned as described in Jacobson *et al.* (1993). The run off product of *in vitro* transcription reaction of Sma I-digested pMRG3U2-27 using T7 RNA polymerase is shown. This plasmid was kindly provided by Dr. T. Pederson.

Table of antibodies

Target	Name	Comments	Source
RBM		rabbit antiserum	I. Eperon
hnRNP A1		rabbit antiserum	I. Eperon
Sm	KSm2	mouse monoclonal	D. Williams
Sm	KSm5	mouse monoclonal	D. Williams
Sm	Y12	mouse monoclonal	I. Mattaj
HnRNP I		mouse monoclonal	G. Dreyfuss
HnRNP (K+J)		mouse monoclonal	G. Dreyfuss
HnRNP (F+H)		mouse monoclonal	G. Dreyfuss
HnRNP C	4F4	mouse monoclonal	G. Dreyfuss
HnRNP G		dog serum	C.-J. Larsen
SF2	α-SF2	mouse monoclonal	A. Krainer
SR proteins	3CSH4	mouse monoclonal	B. Turner
U1-A	α-U1A	mouse monoclonal	I. Mattaj
U1-70K		mouse monoclonal	S. Hoch
U5-116kD		rabbit antiserum	R. Lührmann

U5-102kD		rabbit antiserum	R. Lührmann
U5-100kD		rabbit antiserum	R. Lührmann
U4/U6-60kD		rabbit antiserum	R. Lührmann
U4/U6-90kD		rabbit antiserum	R. Lührmann
U4/U6-20kD		rabbit antiserum	R. Lührmann
U4/U6-61kD		rabbit antiserum	R. Lührmann
U4-15.5kD		rabbit antiserum	R. Lührmann
U4/U6.U5-27kD		rabbit antiserum	R. Lührmann
Lsm4		rabbit antiserum	R. Lührmann
peroxidase-conjugated goat anti-rabbit IgG		Cat# 111-035-144	Dianova
peroxidase-conjugated goat anti-rabbit IgG		Cat# A0545	Sigma
sheep anti-mouse IgG, FITC linked whole Ab		Cat# N1031	Amersham
donkey anti-rabbit IgG,Texas Red linked whole Ab			Amersham
goat anti-rabbit IgG, Cy3 linked		Cat# PA 43004	Amersham

Basic Solutions:

5xBradford reagent

100 mg of Coomassie brilliant blue G250, 50 ml ethanol, 100 ml phosphoric acid; add water to 200 ml and filter through 0.45 μm membrane.

10x Blot Transfer Buffer

58 g of Tris-HCl, 29.25 g of glycine, 37.5 ml of 10% SDS in 1 litre of water.

1x Blot Transfer Buffer

48mM Tris-HCl, 39 mM glycine, 0.0375% SDS, 20% methanol.

50x Denhardt's Reagent

1% Ficoll type 400, 1% polyvinylpyrrolidone, 1% bovine serum albumin (BSA) (fraction V).

D-Glu buffer

20 mM Triethanolamine-HCl, pH 7.9, 80 mM monopotassium glutamate, 5% glycerol, 0.2 mM EDTA (potassium salt), 0.5 mM DTT, 0.5 mM PMSF.

Formamide RNA loading buffer

90% formamide, 50 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol.

IPP_{150}

20 mM HEPES, pH 8.0, 150 mM NaCl, 1.5 $MgCl_2$, 0.5 mM DTT, 0.05% NP40.

LB medium

170 mM sodium chloride, 1% bactotryptone, 0.5% yeast extract; adjust pH 7.0 with NaOH.

10x PCR buffer

100 mM Tris-HCl, pH 8.3, 500 mM KCl, 20 mM $MgCl_2$.

Phosphate buffered saline (PBS):

150 mM NaCl, 2.5 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.0-8.0.

1x PK buffer

100 mM Tris-HCl, pH 7.5, 25 mM EDTA, 150 mM NaCl, 1 % SDS.

0.1 M PMSF

1.742 g of PMSF is dissolved in 100 ml of ethanol; should be used fresh.

RNA elution buffer

0.5 M sodium acetate, pH 5.0, 1 mM EDTA, 0.2% SDS.

1x SDS-PAGE loading buffer

50 mM Tris-HCl, pH 6.8, 100 mM DTT, 2% SDS, 10% glycerol, 0.01% bromophenol blue.

4x Separation Gel Buffer (SDS-PAGE)

181.7 g of Tris-HCl, 4 g of SDS in 1 litre of water; pH is adjusted to 8.8 with HCl.

4x Stacking Gel Buffer (SDS-PAGE)

30.3 g of Tris-HCl, 2 g of SDS in 0.5 litre of water; pH is adjusted to 6.8 with HCl.

20x SSC

4.9 M NaCl, 0.3 M sodium citrate.

10x TBE buffer

90 mM Tris-borate, pH 8, 2 mM EDTA.

10x TBS buffer

200 mM Tris-HCl, pH 7.5, 1500 mM NaCl.

1x TE buffer

10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA.

1x TGS

25 mM Tris-HCl, 192 mM glycine, 0.1% SDS.

TNES RNA extraction buffer

20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5mM EDTA, 0.1% SDS.

REFERENCES

Abdul-Manan, N., O'Malley, S. M., and Williams, K. R. (1996a). Origins of binding specificity of the A1 heterogeneous nuclear ribonucleoprotein. Biochemistry *35*, 3545-54.

Abdul-Manan, N., and Williams, K. R. (1996b). hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base- specific binding. Nucleic Acids Res *24*, 4063-70.

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*, 6417-25.

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403-12.

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Luhrmann, 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*, 6756-66.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Luhrmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'- end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. Embo J *18*, 5789-802.

Adams, M. D., Tarng, R. S., and Rio, D. C. (1997). The alternative splicing factor PSI regulates P-element third intron splicing in vivo. Genes Dev *11*, 129-38.

Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., and Kent, S. B. (1987). Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. Proc Natl Acad Sci U S A *84*, 6970-4.

Aebi, M., Hornig, H., Padgett, R. A., Reiser, J., and Weissmann, C. (1986). Sequence requirements for splicing of higher eukaryotic nuclear pre- mRNA. Cell *47*, 555-65.

Aebi, M., Hornig, H., and Weissmann, C. (1987). 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. Cell *50*, 237-46.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-402.

Andersson, M., Page, D. C., Pettay, D., Subrt, I., Turleau, C., de Grouchy, J., and de la Chapelle, A. (1988). Y autosome translocations and mosaicism in the aetiology of 45,X maleness: assignment of fertility factor to distal Yq11. Hum Genet *79*, 2-7.

Andrade, L. E., Chan, E. K., Raska, I., Peebles, C. L., Roos, G., and Tan, E. M. (1991). Human autoantibody to a novel protein of the nuclear coiled body: immunological characterization and cDNA cloning of p80-coilin. J Exp Med *173*, 1407-19.

Arenas, J. E., and Abelson, J. N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proc Natl Acad Sci U S A *94*, 11798-802.

Ashkenazi, A., and Dixit, V. M. (1999). Apoptosis control by death and decoy receptors. Curr Opin Cell Biol *11*, 255-60.

Bach, M., Bringmann, P., and Luhrmann, R. (1990). Purification of small nuclear ribonucleoprotein particles with antibodies against modified nucleosides of small nuclear RNAs. Methods Enzymol *181*, 232-57.

Bach, M., Winkelmann, G., and Luhrmann, R. (1989). 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. Proc Natl Acad Sci U S A *86*, 6038-42.

Bai, Y., Lee, D., Yu, T., and Chasin, L. A. (1999). Control of 3' splice site choice in vivo by ASF/SF2 and hnRNP A1. Nucleic Acids Res *27*, 1126-34.

Beggs, J. D. (1995). Yeast splicing factors and genetic strategies for their analysis. In Pre-mRNA Processing, A. Lamond, ed.: R.G.Landes Co.,TX), pp. 79-95.

Behrens, S. E., and Luhrmann, R. (1991). Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. Genes Dev *5*, 1439-52.

Bennett, M., Michaud, S., Kingston, J., and Reed, R. (1992a). Protein components specifically associated with prespliceosome and spliceosome complexes. Genes Dev *6*, 1986-2000.

Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G., and Reed, R. (1992b). Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. Mol Cell Biol *12*, 3165-75.

Berget, S. M. (1995). Exon recognition in vertebrate splicing. J Biol Chem *270*, 2411-4.

Berglund, J. A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev *12*, 858-67.

Birney, E., Kumar, S., and Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res *21*, 5803-16.

Black, D. L., Chan, R., Min, H., Wang, J., and Bell, L. (1998). The electrophoretic mobility shift assay for RNA binding proteins. In RNA:Protein Interactions, S. W. J. Smith, ed.: Oxford University Press), pp. 109-136.

Blanchette, M., and Chabot, B. (1997). A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. Rna *3*, 405-19.

Blanchette, M., and Chabot, B. (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. Embo J *18*, 1939-1952.

Blum, H., Beier, H., and Gross, H. J. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis *8*, 93-99.

Bochnig, P., Reuter, R., Bringmann, P., and Luhrmann, R. (1987). A monoclonal antibody against 2,2,7-trimethylguanosine that reacts with intact, class U, small nuclear ribonucleoproteins as well as with 7- methylguanosine-capped RNAs. Eur J Biochem *168*, 461-7.

Bordonne, R., Banroques, J., Abelson, J., and Guthrie, C. (1990). Domains of yeast U4 spliceosomal RNA required for PRP4 protein binding, snRNP-snRNP interactions, and pre-mRNA splicing in vivo. Genes Dev *4*, 1185-96.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc Natl Acad Sci U S A *75*, 4853-7.

Breitbart, R. E., Andreadis, A., and Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu Rev Biochem *56*, 467-95.

Brimacombe, R. (1991). RNA-protein interactions in the Escherichia coli ribosome. Biochimie *73*, 927-36.

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*, 17640-6.

Brown, J. D., and Beggs, J. D. (1992). Roles of PRP8 protein in the assembly of splicing complexes. Embo J *11*, 3721-9.

Burd, C. G., and Dreyfuss, G. (1994a). Conserved structures and diversity of functions of RNA-binding proteins. Science *265*, 615-21.

Burd, C. G., and Dreyfuss, G. (1994b). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high- affinity binding sites in pre-mRNA splicing. Embo J *13*, 1197-204.

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

Busch, S. J., and Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. Trends Genet *6*, 36-40.

Buvoli, M., Cobianchi, F., Bestagno, M. G., Mangiarotti, A., Bassi, M. T., Biamonti, G., and Riva, S. (1990a). Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein. Embo J *9*, 1229-35.

Buvoli, M., Cobianchi, F., Biamonti, G., and Riva, S. (1990b). Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. Nucleic Acids Res *18*, 6595-600.

Caceres, J. F., and Krainer, A. R. (1993). Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. Embo J *12*, 4715-26.

Caceres, J. F., and Krainer, A. R. (1997). Mammalian pre-mRNA splicing factors. In Eukaryotic mRNA processing, A. R. Krainer, ed.: IRL Press), pp. 174-203.

Caceres, J. F., Stamm, S., Helfman, D. M., and Krainer, A. R. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science *265*, 1706-9.

Cao, W., Jamison, S. F., and Garcia-Blanco, M. A. (1997). Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. Rna *3*, 1456-67.

Carmo-Fonseca, M., Ferreira, J., and Lamond, A. I. (1993). Assembly of snRNP-containing coiled bodies is regulated in interphase and mitosis--evidence that the coiled body is a kinetic nuclear structure. J Cell Biol *120*, 841-52.

Carmo-Fonseca, M., Pepperkok, R., Carvalho, M. T., and Lamond, A. I. (1992). Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. J Cell Biol *117*, 1-14.

Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S., and Biamonti, G. (1996). hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. J Mol Biol *259*, 337-48.

Casas-Finet, J. R., Smith, J. D., Jr., Kumar, A., Kim, J. G., Wilson, S. H., and Karpel, R. L. (1993). Mammalian heterogeneous ribonucleoprotein A1 and its constituent domains. Nucleic acid interaction, structural stability and self- association. J Mol Biol *229*, 873-89.

Cavaloc, Y., Popielarz, M., Fuchs, J. P., Gattoni, R., and Stevenin, J. (1994). Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. Embo J *13*, 2639-49.

Cech, T. R. (1986). The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell *44*, 207-10.

Celis, J. E. (1994). Cell Biology-A laboratory handbook (San Diego : Academic Press, Inc.)

Chabot, B., Blanchette, M., Lapierre, I., and La Branche, H. (1997). An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. Mol Cell Biol *17*, 1776-86.

Chabot, B., Frappier, D., and La Branche, H. (1992). Differential ASF/SF2 activity in extracts from normal WI38 and transformed WI38VA13 cells. Nucleic Acids Res *20*, 5197-204.

Chan, R. C., and Black, D. L. (1997). The polypyrimidine tract binding protein binds upstream of neural cell- specific c-src exon N1 to repress the splicing of the intron downstream. Mol Cell Biol *17*, 4667-76.

Chandley, A. C., Gosden, J. R., Hargreave, T. B., Spowart, G., Speed, R. M., and McBeath, S. (1989). Deleted Yq in the sterile son of a man with a satellited Y chromosome (Yqs). J Med Genet *26*, 145-53.

Cheng, S. C., and Abelson, J. (1987). Spliceosome assembly in yeast. Genes Dev *1*, 1014-27.

Chu, E., Cogliati, T., Copur, S. M., Borre, A., Voeller, D. M., Allegra, C. J., and Segal, S. (1996). Identification of in vivo target RNA sequences bound by thymidylate synthase. Nucleic Acids Res *24*, 3222-8.

Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., and Wilson, S. H. (1988). Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. Large-scale overproduction in Escherichia coli and cooperative binding to single-stranded nucleic acids. J Biol Chem *263*, 1063-71.

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*, 1970-82.

Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature *349*, 487-93.

Conway, G., Wooley, J., Bibring, T., and LeStourgeon, W. M. (1988). Ribonucleoproteins package 700 nucleotides of pre-mRNA into a repeating array of regular particles. Mol Cell Biol *8*, 2884-95.

Cooper, T. A. (1998). Muscle-specific splicing of a heterologous exon mediated by a single muscle-specific splicing enhancer from the cardiac troponin T gene. Mol Cell Biol *18*, 4519-25.

Crispino, J. D., Blencowe, B. J., and Sharp, P. A. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. Science *265*, 1866-9.

Crispino, J. D., Mermoud, J. E., Lamond, A. I., and Sharp, P. A. (1996). Cis-acting elements distinct from the 5' splice site promote U1- independent pre-mRNA splicing. Rna *2*, 664-73.

Cunningham, S. A., Else, A. J., Potter, B. V., and Eperon, I. C. (1991). Influences of separation and adjacent sequences on the use of alternative 5' splice sites. J Mol Biol *217*, 265-81.

Das, A. K., Cohen, P. W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. Embo J *17*, 1192-9.

Davies, R. C., Calvio, C., Bratt, E., Larsson, S. H., Lamond, A. I., and Hastie, N. D. (1998). WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. Genes Dev *12*, 3217-25.

Del Gatto-Konczak, F., Olive, M., Gesnel, M. C., and Breathnach, R. (1999). hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. Mol Cell Biol *19*, 251-60.

Delbridge, M. L., Lingenfelter, P. A., Disteche, C. M., and Graves, J. A. (1999). The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. Nat Genet *22*, 223-4.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-89.

Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev *13*, 1102-15.

Dizdaroglu, M., and Gajewski, E. (1989). Structure and mechanism of hydroxyl radical-induced formation of a DNA- protein cross-link involving thymine and lysine in nucleohistone. Cancer Res *49*, 3463-7.

Dizdaroglu, M., Gajewski, E., Reddy, P., and Margolis, S. A. (1989). Structure of a hydroxyl radical induced DNA-protein cross-link involving thymine and tyrosine in nucleohistone. Biochemistry *28*, 3625-8.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem *62*, 289-321.

Edman, P., and Begg, G. (1967). A protein sequenator. Eur J Biochem *1*, 80-91.

Eisenstein, E., and Schachman, H. K. (1989). Dertermining the roles of subunits in protein function. In Protein function, T. E. Creighton, ed.: IRL Press), pp. 135-176.

Elliott, D. J., and Cooke, H. J. (1997). The molecular genetics of male infertility. Bioessays *19*, 801-9.

Elliott, D. J., Ma, K., Kerr, S. M., Thakrar, R., Speed, R., Chandley, A. C., and Cooke, H. (1996). An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells. Hum Mol Genet *5*, 869-74.

Elliott, D. J., Oghene, K., Makarov, G., Makarova, O., Hargreave, T. B., Chandley, A. C., Eperon, I. C., and Cooke, H. J. (1998). Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development. J Cell Sci *111*, 1255-65.

England, T. E., and Uhlenbeck, O. C. (1978). 3'-terminal labelling of RNA with T4 RNA ligase. Nature *275*, 560-1.

Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A., and Krainer, A. R. (1993). Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. Embo J *12*, 3607-17.

Eperon, I. C., and Krainer, A. R. (1994). Splicing of mRNA precursors in mammalian cells. In RNA Processing, S. J. Higgins and B. D. Hames, eds. (Oxford: IRL Press), pp. 57-101.

Eperon, L. P., Estibeiro, J. P., and Eperon, I. C. (1986). The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. Nature *324*, 280-2.

Fetzer, S., Lauber, J., Will, C. L., and Luhrmann, 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*, 344-55.

Fu, X. D. (1993). Specific commitment of different pre-mRNAs to splicing by single SR proteins [see comments]. Nature *365*, 82-5.

Fu, X. D., Mayeda, A., Maniatis, T., and Krainer, A. R. (1992). General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5' and 3' splice site selection. Proc Natl Acad Sci U S A *89*, 11224-8.

Gajewski, E., and Dizdaroglu, M. (1990). Hydroxyl radical induced cross-linking of cytosine and tyrosine in nucleohistone. Biochemistry *29*, 977-80.

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*, 1555-62.

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*, 7088-98.

Ge, H., and Manley, J. L. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell *62*, 25-34.

Ge, H., Zuo, P., and Manley, J. L. (1991). Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. Cell *66*, 373-82.

Gee, S., Krauss, S. W., Miller, E., Aoyagi, K., Arenas, J., and Conboy, J. G. (1997). Cloning of mDEAH9, a putative RNA helicase and mammalian homologue of Saccharomyces cerevisiae splicing factor Prp43. Proc Natl Acad Sci U S A *94*, 11803-7.

Gohring, F., and Fackelmayer, F. O. (1997). The scaffold/matrix attachment region binding protein hnRNP-U (SAF-A) is directly bound to chromosomal DNA in vivo: a chemical cross-linking study. Biochemistry *36*, 8276-83.

Gohring, F., Schwab, B. L., Nicotera, P., Leist, M., and Fackelmayer, F. O. (1997). The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown. Embo J *16*, 7361-71.

Gorlach, M., Wittekind, M., Beckman, R. A., Mueller, L., and Dreyfuss, G. (1992). Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. Embo J *11*, 3289-95.

Gorman, C. (1985). High efficiency gene transfer into mammalian cells. In DNA cloning, D. M. Glover, ed. (Oxford: IRL Press), pp. 143-190.

Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer, G., Colot, H. V., Mann, M., Seraphin, B., Rosbash, M., Luhrmann, R., and Fabrizio, P. (1998). A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. Rna *4*, 374-93.

Graveley, B. R., Hertel, K. J., and Maniatis, T. (1998a). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. Embo J *17*, 6747-56.

Graveley, B. R., and Maniatis, T. (1998b). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol Cell *1*, 765-71.

Grossman, J. S., Meyer, M. I., Wang, Y. C., Mulligan, G. J., Kobayashi, R., and Helfman, D. M. (1998). The use of antibodies to the polypyrimidine tract binding protein (PTB) to analyze the protein components that assemble on alternatively spliced pre-mRNAs that use distant branch points. Rna *4*, 613-25.

Gui, J. F., Tronchere, H., Chandler, S. D., and Fu, X. D. (1994). Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. Proc Natl Acad Sci U S A *91*, 10824-8.

Guo, W., Mulligan, G. J., Wormsley, S., and Helfman, D. M. (1991). Alternative splicing of beta-tropomyosin pre-mRNA: cis-acting elements and cellular factors that block the use of a skeletal muscle exon in nonmuscle cells. Genes Dev *5*, 2096-107.

Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science *253*, 157-63.

Hackl, W., Fischer, U., and Luhrmann, R. (1994). A 69-kD protein that associates reversibly with the Sm core domain of several spliceosomal snRNP species. J Cell Biol *124*, 261-72.

Hackl, W., and Luhrmann, R. (1996). Molecular cloning and subcellular localisation of the snRNP-associated protein 69KD, a structural homologue of the proto-oncoproteins TLS and EWS with RNA and DNA-binding properties. J Mol Biol *264*, 843-51.

Hall, S. L., and Padgett, R. A. (1994). Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. J Mol Biol *239*, 357-65.

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*, 1716-8.

Hamilton, B. J., Nagy, E., Malter, J. S., Arrick, B. A., and Rigby, W. F. (1993). Association of heterogeneous nuclear ribonucleoprotein A1 and C proteins with reiterated AUUUA sequences. J Biol Chem *268*, 8881-7.

Hammond, L. E., Rudner, D. Z., Kanaar, R., and Rio, D. C. (1997). Mutations in the hrp48 gene, which encodes a Drosophila heterogeneous nuclear ribonucleoprotein particle protein, cause lethality and developmental defects and affect P-element third-intron splicing in vivo. Mol Cell Biol *17*, 7260-7.

Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., Jr., and Krainer, A. R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. Rna *4*, 430-44.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex- lethal protein. Nature *398*, 579-85.

Hargreave, T. B. (1997). Human infertility. In Male infertility, T. B. Hargreave, ed.: Springer Verlag), pp. 1-16.

Harlow, E., and Lane, D. (1988). Antibodies-A laboratory manual: Cold Spring Harbor Laboratory.

Harper, J. E., and Manley, J. L. (1991). A novel protein factor is required for use of distal alternative 5' splice sites in vitro. Mol Cell Biol *11*, 5945-53.

Hertel, K. J., and Maniatis, T. (1999). Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. Proc Natl Acad Sci U S A *96*, 2651-2655.

Hirano, T., Kinoshita, N., Morikawa, K., and Yanagida, M. (1990). Snap helix with knob and hole: essential repeats in S. pombe nuclear protein nuc2+. Cell *60*, 319-28.

Hockensmith, J. W., Kubasek, W. L., Vorachek, W. R., Evertsz, E. M., and von Hippel, P. H. (1991). Laser cross-linking of protein-nucleic acid complexes. Methods Enzymol *208*, 211-36.

Hodges, P. E., Jackson, S. P., Brown, J. D., and Beggs, J. D. (1995). Extraordinary sequence conservation of the PRP8 splicing factor. Yeast *11*, 337-42.

Hornig, H., Aebi, M., and Weissmann, C. (1986). Effect of mutations at the lariat branch acceptor site on beta-globin pre-mRNA splicing in vitro. Nature *324*, 589-91.

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*, 1374-87.

Hu, J., Xu, D., Schappert, K., Xu, Y., and Friesen, J. D. (1995). Mutational analysis of Saccharomyces cerevisiae U4 small nuclear RNA identifies functionally important domains. Mol Cell Biol *15*, 1274-85.

Huang, S., and Spector, D. L. (1997). Nuclear organization of pre-mRNA splicing factors and substrates. In Eukaryotic mRNA processing, A. R. Krainer, ed. (Oxford: IRL Press), pp. 37-67.

Hwang, D. Y., and Cohen, J. B. (1996). U1 snRNA promotes the selection of nearby 5' splice sites by U6 snRNA in mammalian cells. Genes Dev *10*, 338-50.

Iost, I., Guillerez, J., and Dreyfus, M. (1992). Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. J Bacteriol *174*, 619-22.

Ishikawa, F., Matunis, M. J., Dreyfuss, G., and Cech, T. R. (1993). Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. Mol Cell Biol *13*, 4301-10.

Izaurralde, E., Jarmolowski, A., Beisel, C., Mattaj, I. W., Dreyfuss, G., and Fischer, U. (1997). A role for the M9 transport signal of hnRNP A1 in mRNA nuclear export. J Cell Biol *137*, 27-35.

Izaurralde, E., Mirkovitch, J., and Laemmli, U. K. (1988). Interaction of DNA with nuclear scaffolds in vitro. J Mol Biol *200*, 111-25.

Jackson, I. J. (1991). A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Res *19*, 3795-8.

Jacobson, M. R., Rhoadhouse, M., and Pederson, T. (1993). U2 small nuclear RNA 3' end formation is directed by a critical internal structure distinct from the processing site. Mol Cell Biol *13*, 1119-29.

Jameson, D. M., and Eccleston, J. F. (1997). Fluorescent nucleotide analogs: synthesis and applications. Methods Enzymol *278*, 363-90.

Jamison, S. F., Pasman, Z., Wang, J., Will, C., Luhrmann, R., Manley, J. L., and Garcia-Blanco, M. A. (1995). U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. Nucleic Acids Res *23*, 3260-7.

Jiang, S. W., and Eberhardt, N. L. (1995). A micro-scale method to isolate DNA-binding proteins suitable for quantitative comparison of expression levels from transfected cells. Nucleic Acids Res *23*, 3607-8.

Jiang, Z. H., and Wu, J. Y. (1999). Alternative splicing and programmed cell death. Proc Soc Exp Biol Med *220*, 64-72.

Jiang, Z. H., Zhang, W. J., Rao, Y., and Wu, J. Y. (1998). Regulation of Ich-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors. Proc Natl Acad Sci U S A *95*, 9155-60.

Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Luhrmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell *96*, 375-87.

Kamma, H., Portman, D. S., and Dreyfuss, G. (1995). Cell type-specific expression of hnRNP proteins. Exp Cell Res *221*, 187-96.

Kanopka, A., Muhlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C., and Akusjarvi, G. (1998). Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. Nature *393*, 185-7.

Kiledjian, M., and Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. Embo J *11*, 2655-64.

Kim, S., Park, G. H., and Paik, W. K. (1998). Recent advances in protein methylation: enzymatic methylation of nucleic acid binding proteins. Amino Acids *15*, 291-306.

Klotz, I. M. (1986). Ligand-Receptor Interactions. In Introduction to Biomolecular Energetics: Academic Press, Inc.), pp. 103-133.

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., and Manley, J. L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature *368*, 119-24.

Koizumi, J., Okamoto, Y., Onogi, H., Mayeda, A., Krainer, A. R., and Hagiwara, M. (1999). The subcellular localization of SF2/ASF is regulated by direct interaction with SR protein kinases (SRPKs). J Biol Chem *274*, 11125-31.

Konarska, M. M., Grabowski, P. J., Padgett, R. A., and Sharp, P. A. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature *313*, 552-7.

Konarska, M. M., and Sharp, P. A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell *49*, 763-74.

Konforti, B. B., Koziolkiewicz, M. J., and Konarska, M. M. (1993). Disruption of base pairing between the 5' splice site and the 5' end of U1 snRNA is required for spliceosome assembly. Cell *75*, 863-73.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res *15*, 8125-48.

Krainer, A. R., Conway, G. C., and Kozak, D. (1990a). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. Genes Dev *4*, 1158-71.

Krainer, A. R., Conway, G. C., and Kozak, D. (1990b). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. Cell *62*, 35-42.

Krainer, A. R., and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. Cell *42*, 725-36.

Krainer, A. R., Mayeda, A., Kozak, D., and Binns, G. (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. Cell *66*, 383-94.

Kramer, A., Keller, W., Appel, B., and Luhrmann, R. (1984). The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. Cell *38*, 299-307.

Kramer, A., and Utans, U. (1991). Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. Embo J *10*, 1503-9.

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.

Kumar, A., Casas-Finet, J. R., Luneau, C. J., Karpel, R. L., Merrill, B. M., Williams, K. R., and Wilson, S. H. (1990). Mammalian heterogeneous nuclear ribonucleoprotein A1. Nucleic acid binding properties of the COOH-terminal domain. J Biol Chem *265*, 17094-100.

Kumar, A., and Wilson, S. H. (1990). Studies of the strand-annealing activity of mammalian hnRNP complex protein A1. Biochemistry *29*, 10717-22.

Kuroyanagi, N., Onogi, H., Wakabayashi, T., and Hagiwara, M. (1998). Novel SR-protein-specific kinase, SRPK2, disassembles nuclear speckles. Biochem Biophys Res Commun *242*, 357-64.

Lafontaine, D. L., and Tollervey, D. (1999). Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. Rna *5*, 455-67.

Laggerbauer, B., Lauber, J., and Luhrmann, R. (1996). Identification of an RNA-dependent ATPase activity in mammalian U5 snRNPs. Nucleic Acids Res *24*, 868-75.

Laggerbauer, B., Achsel, T., and Luhrmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. Proc Natl Acad Sci U S A *95*, 4188-92.

Lamb, J. R., Tugendreich, S., and Hieter, P. (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR? Trends Biochem Sci *20*, 257-9.

Lamond, A. I., and Carmo-Fonseca, M. (1993). Localisation of splicing snRNPs in mammalian cells. Mol Biol Rep *18*, 127-33.

Lamond, A. I., Konarska, M. M., Grabowski, P. J., and Sharp, P. A. (1988). Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. Proc Natl Acad Sci U S A *85*, 411-5.

Lamond, A. I., Konarska, M. M., and Sharp, P. A. (1987). A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. Genes Dev *1*, 532-43.

Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science *243*, 1681-8.

Larsson, S. H., Charlieu, J. P., Miyagawa, K., Engelkamp, D., Rassoulzadegan, M., Ross, A., Cuzin, F., van Heyningen, V., and Hastie, N. D. (1995). Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. Cell *81*, 391-401.

Lauber, J., Plessel, G., Prehn, S., Will, C. L., Fabrizio, P., Groning, K., Lane, W. S., and Luhrmann, R. (1997). The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p [published erratum appears in RNA 1997 Oct;3(10):1204-6]. Rna *3*, 926-41.

Lee, K. A., Bindereif, A., and Green, M. R. (1988). A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. Gene Anal Tech *5*, 22-31.

Legrain, P., and Choulika, A. (1990). The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cysteine/histidine motif common to several splicing factors. Embo J *9*, 2775-81.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L., and Steitz, J. A. (1980). Are snRNPs involved in splicing? Nature *283*, 220-4.

Li, H., and Yuan, J. (1999). Deciphering the pathways of life and death. Curr. Opin. Cell Biol. *11*, 261-266.

Lisitsyn, N. A. (1995). Representational difference analysis: finding the differences between genomes. Trends Genet *11*, 303-7.

Lisitsyn, N. A., Rosenberg, M. V., Launer, G. A., Wagner, L. L., Potapov, V. K., Kolesnik, T. B., and Sverdlov, E. D. (1993). A method for isolation of sequences missing in one of two related genomes. Mol Gen Mikrobiol Virusol, 26-9.

Liu, H. X., Zhang, M., and Krainer, A. R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev *12*, 1998-2012.

Liu, Q., and Dreyfuss, G. (1995). In vivo and in vitro arginine methylation of RNA-binding proteins. Mol Cell Biol *15*, 2800-8.

Lopez, A. J. (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. Annu Rev Genet *32*, 279-305.

Lossky, M., Anderson, G. J., Jackson, S. P., and Beggs, J. (1987). Identification of a yeast snRNP protein and detection of snRNP-snRNP interactions. Cell *51*, 1019-26.

Lou, H., Helfman, D. M., Gagel, R. F., and Berget, S. M. (1999). Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. Mol Cell Biol *19*, 78-85.

Lucke, S., Klockner, T., Palfi, Z., Boshart, M., and Bindereif, A. (1997). Trans mRNA splicing in trypanosomes: cloning and analysis of a PRP8- homologous gene from Trypanosoma brucei provides evidence for a U5- analogous RNP. Embo J *16*, 4433-40.

Lund, E., and Dahlberg, J. E. (1992). Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. Science *255*, 327-30.

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.

Lynch, K. W., and Maniatis, T. (1996). Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. Genes Dev *10*, 2089-101.

Lynch, K. W., and Maniatis, T. (1995). Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev *9*, 284-93.

Ma, K., Inglis, J. D., Sharkey, A., Bickmore, W. A., Hill, R. E., Prosser, E. J., Speed, R. M., Thomson, E. J., Jobling, M., Taylor, K., and et al. (1993). A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. Cell *75*, 1287-95.

Ma, K., Sharkey, A., Kirsch, S., Vogt, P., Keil, R., Hargreave, T. B., McBeath, S., and Chandley, A. C. (1992). Towards the molecular localisation of the AZF locus: mapping of microdeletions in azoospermic men within 14 subintervals of interval 6 of the human Y chromosome. Hum Mol Genet *1*, 29-33.

MacMillan, A. M., Query, C. C., Allerson, C. R., Chen, S., Verdine, G. L., and Sharp, P. A. (1994). Dynamic association of proteins with the pre-mRNA branch region. Genes Dev *8*, 3008-20.

Maddock, J. R., Roy, J., and Woolford, J. L., Jr. (1996). Six novel genes necessary for pre-mRNA splicing in Saccharomyces cerevisiae. Nucleic Acids Res *24*, 1037-44.

Madhani, H. D., and Guthrie, C. (1994). Dynamic RNA-RNA interactions in the spliceosome. Annu Rev Genet *28*, 1-26.

Makarova, O. V., Makarov, E. M., Sousa, R., and Dreyfus, M. (1995). Transcribing of Escherichia coli genes with mutant T7 RNA polymerases: stability of lacZ mRNA inversely correlates with polymerase speed. Proc Natl Acad Sci U S A *92*, 12250-4.

Manley, J. L., Fire, A., Cano, A., Sharp, P. A., and Gefter, M. L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc Natl Acad Sci U S A *77*, 3855-9.

Margolis, S. A., Coxon, B., Gajewski, E., and Dizdaroglu, M. (1988). Structure of a hydroxyl radical induced cross-link of thymine and tyrosine. Biochemistry *27*, 6353-9.

Mattaj, I. W., and Hamm, J. (1989). Regulated splicing in early development and stage-specific U snRNPs. Development *105*, 183-9.

Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M. Q., Gardiner, E. M., and Krainer, A. R. (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. Embo J *18*, 4560-70.

Mayeda, A., Munroe, S. H., Xu, R. M., and Krainer, A. R. (1998). Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. Rna *4*, 1111-23.

Mayeda, A., Munroe, S. H., Caceres, J. F., and Krainer, A. R. (1994). Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. Embo J *13*, 5483-95.

Mayeda, A., Helfman, D. M., and Krainer, A. R. (1993). Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol Cell Biol *13*, 2993-3001.

Mayeda, A., and Krainer, A. R. (1992a). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. Cell *68*, 365-75.

Mayeda, A., Zahler, A. M., Krainer, A. R., and Roth, M. B. (1992b). Two members of a conserved family of nuclear phosphoproteins are involved in pre-mRNA splicing. Proc Natl Acad Sci U S A *89*, 1301-4.

Mayes, A. E., Verdone, L., Legrain, P., and Beggs, J. D. (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. Embo J *18*, 4321-4331.

Mazeyrat, S., Saut, N., Mattei, M. G., and Mitchell, M. J. (1999). RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. Nat Genet *22*, 224-6.

McAfee, J. G., Huang, M., Soltaninassab, S., Rech, J. E., Iyengar, S., and Lestourgeon, W. M. (1997). The packaging of pre-mRNA. In Eukaryotic mRNA processing, A. R. Krainer, ed.: IRL Press), pp. 68-95.

McGhee, J. D., and Hippel, P. H. v. (1974). Theoretical aspects of DNA-protein interactions: co-operative and non- co-operative binding of large ligands to a one-dimensional homogeneous lattice. J Mol Biol *86*, 469-89.

McRorie, D. K., and Voelker, P. J. (1993). Self-associating systems in the analytical ultracentrifugation (Fullerton, USA: Beckman Instruments Press).

Mee, L. K., and Adelstein, S. J. (1981). Predominance of core histones in formation of DNA--protein crosslinks in gamma-irradiated chromatin. Proc Natl Acad Sci U S A *78*, 2194-8.

Mermoud, J. E., Cohen, P. T., and Lamond, A. I. (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. Embo J *13*, 5679-88.

Merril, C. R., and Goldman, D. (1982). Quantitative two-dimensional protein electrophoresis for studies of inborn errors of metabolism. Clin Chem *28*, 1015-20.

Michaud, S., and Reed, R. (1993). A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. Genes Dev *7*, 1008-20.

Minn, A. J., Boise, L. H., and Thompson, C. B. (1996). Bcl-x(S) anatagonizes the protective effects of Bcl-x(L). J Biol Chem *271*, 6306-12.

Misteli, T. (1999). RNA splicing: What has phosphorylation got to do with it? Curr Biol *9*, R198-200.

Misteli, T., Caceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. J Cell Biol *143*, 297-307.

Misteli, T., Caceres, J. F., and Spector, D. L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. Nature *387*, 523-7.

Misteli, T., and Spector, D. L. (1996). Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. Mol Biol Cell *7*, 1559-72.

Monaco, L., Nantel, F., Foulkes, N. S., and Sassone-Corsi, P. (1996). CREM: A transcriptional master switch govering the cAMP response in the testis. In Signal Transduction in Testicular Cells. Basic and Clinical Aspects, V. Hansson, F. O. Levy and K. Tasken, eds. (New York: Springer).

Moore, M. J., Query, C. C., and Sharp, P. A. (1993). Splicing of Precursors to mRNAs by the spliceosome. In The RNA World, R. F. Gesteland and J. F. Atkins, eds. (New York: Cold Spring Harbor Laboratory Press), pp. 303-357.

Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A., and Steitz, J. A. (1983). The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. Cell *33*, 509-18.

Mulligan, G. J., Guo, W., Wormsley, S., and Helfman, D. M. (1992). Polypyrimidine tract binding protein interacts with sequences involved in alternative splicing of beta-tropomyosin pre-mRNA. J Biol Chem *267*, 25480-7.

Munroe, S. H., and Dong, X. F. (1992). Heterogeneous nuclear ribonucleoprotein A1 catalyzes RNA.RNA annealing. Proc Natl Acad Sci U S A *89*, 895-9.

Murray, M. V., Kobayashi, R., and Krainer, A. R. (1999). The type 2C Ser/Thr phosphatase PP2Cgamma is a pre-mRNA splicing factor. Genes Dev *13*, 87-97.

Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F., Wilson, S. H., and Williams, K. R. (1991). Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. Biochemistry *30*, 2968-76.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A [see comments]. Nature *348*, 515-20.

Nelson, K. K., and Green, M. R. (1988). Splice site selection and ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. Genes Dev *2*, 319-29.

Nesterenko, M. V., Tilley, M., and Upton, S. J. (1994). A simple modification of Blum's silver stain method allows for 30 minute detection of proteins in polyacrylamide gels. J Biochem Biophys Methods *28*, 239-42.

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.

Newman, A. J. (1997). The role of U5 snRNP in pre-mRNA splicing. Embo J *16*, 5797-800.

Newman, A. J., and Norman, C. (1992). U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell *68*, 743-54.

Newman, A., and Norman, C. (1991). Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. Cell *65*, 115-23.

Newman, A. J., Lin, R. J., Cheng, S. C., and Abelson, J. (1985). Molecular consequences of specific intron mutations on yeast mRNA splicing in vivo and in vitro. Cell *42*, 335-44.

Niall, H. D. (1973). Automated Edman degradation: the protein sequenator. Methods Enzymol *27*, 942-1010.

Nikogosyan, D. N. (1990). Two-quantum UV photochemistry of nucleic acids: comparison with conventional low-intensity UV photochemistry and radiation chemistry. Int J Radiat Biol *57*, 233-99.

Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Luhrmann, 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*, 6119-33.

Ohno, M., and Shimura, Y. (1996). A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. Genes Dev *10*, 997-1007.

O'Keefe, R. T., and Newman, A. J. (1998). Functional analysis of the U5 snRNA loop 1 in the second catalytic step of yeast pre-mRNA splicing. Embo J *17*, 565-74.

O'Mullane, L., and Eperon, I. C. (1998). The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. Mol Cell Biol *18*, 7510-20.

Oleinick, N. L., Chiu, S. M., Ramakrishnan, N., and Xue, L. Y. (1987). The formation, identification, and significance of DNA-protein cross- links in mammalian cells. Br J Cancer Suppl *8*, 135-40.

Ono, Y., Ohno, M., and Shimura, Y. (1994). Identification of a putative RNA helicase (HRH1), a human homolog of yeast Prp22. Mol Cell Biol *14*, 7611-20.

Patton, J. R., Patterson, R. J., and Pederson, T. (1987). Reconstitution of the U1 small nuclear ribonucleoprotein particle. Mol Cell Biol *7*, 4030-7.

Parker, R., and Guthrie, C. (1985). A point mutation in the conserved hexanucleotide at a yeast 5' splice junction uncouples recognition, cleavage, and ligation. Cell *41*, 107-18.

Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., and Cheng, H. L. (1986). A self-splicing RNA excises an intron lariat. Cell *44*, 213-23.

Pinol-Roma, S., Choi, Y. D., Matunis, M. J., and Dreyfuss, G. (1988). Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins [published erratum appears in Genes Dev 1988 Apr;2(4):490]. Genes Dev *2*, 215-27.

Pinol-Roma, S., and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. Nature *355*, 730-2.

Pinto, A. L., and Steitz, J. A. (1989). The mammalian analogue of the yeast PRP8 splicing protein is present in the U4/5/6 small nuclear ribonucleoprotein particle and the spliceosome. Proc Natl Acad Sci U S A *86*, 8742-6.

Prasad, J., Colwill, K., Pawson, T., and Manley, J. L. (1999). The Protein Kinase Clk/Sty Directly Modulates SR Protein Activity: Both Hyper- and Hypophosphorylation Inhibit Splicing. Mol Cell Biol *19*, 6991-7000.

Prosser, J., Inglis, J. D., Condi, A., Ma, K., Kerr, S., Thakrar, R., Taylor, K., Cameron, J. M., and Cooke, H. J. (1995). Degeneracy in human multicopy RBM (YRRM), a candidate spermatogenesis gene. Mamm Genom *7*, 835-842.

Query, C. C., Moore, M. J., and Sharp, P. A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev *8*, 587-97.

Raghunathan, P. L., and Guthrie, C. (1998a). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr Biol *8*, 847-55.

Raghunathan, P. L., and Guthrie, C. (1998b). A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. Science *279*, 857-60.

Ralston, G. (1993). Introduction to analytical ultracentrifugation (Fullerton, USA: Beckman Instruments Press).

Ramchatesingh, J., Zahler, A. M., Neugebauer, K. M., Roth, M. B., and Cooper, T. A. (1995). A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. Mol Cell Biol *15*, 4898-907.

Reddy, R. (1988). Compilation of small RNA sequences. Nucleic Acids Res *16*, r71-85.

Reed, R., Griffith, J., and Maniatis, T. (1988). Purification and visualization of native spliceosomes. Cell *53*, 949-61.

Reed, R., and Maniatis, T. (1986). A role for exon sequences and splice-site proximity in splice-site selection. Cell *46*, 681-90.

Reed, R., and Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. Cell *41*, 95-105.

Reed, R., and Palandjian, L. (1997). Spliceosome assembly. In Eukaryotic mRNA Processing, A. R. Krainer, ed.: IRL Press).

Reijo, R., Lee, T. Y., Salo, P., Alagappan, R., Brown, L. G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., and et al. (1995). Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet *10*, 383-93.

Rio, D. C. (1991). Regulation of Drosophila P element transposition. Trends Genet *7*, 282-7.

Rio, D. C. (1993). Splicing of pre-mRNA: mechanism, regulation and role in development. Curr Opin Genet Dev *3*, 574-84.

Robberson, B. L., Cote, G. J., and Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol *10*, 84-94.

Robertson, J. M., Paulsen, H., and Wintermeyer, W. (1988). Pre-steady-state kinetic studies on ribosomal translocation. Methods Enzymol *164*, 581-97.

Rogers, J., and Wall, R. (1980). A mechanism for RNA splicing. Proc Natl Acad Sci U S A *77*, 1877-9.

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.

Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C., and Tazi, J. (1996). Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. Nature *381*, 80-2.

Roth, M. B., Zahler, A. M., and Stolk, J. A. (1991). A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. J Cell Biol *115*, 587-96.

Roy, J., Kim, K., Maddock, J. R., Anthony, J. G., and Woolford, J. L., Jr. (1995). The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. Rna *1*, 375-90.

Ruskin, B., Krainer, A. R., Maniatis, T., and Green, M. R. (1984). Excision of an intact intron as a novel lariat structure during pre- mRNA splicing in vitro. Cell *38*, 317-31.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning-A laboratory manual (Cold Spring Harbor: Cold Spring Harbor Laboratory).

Schaal, T. D., and Maniatis, T. (1999). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol Cell Biol *19*, 1705-19.

Schenkel, J., Sekeris, C. E., Alonso, A., and Bautz, E. K. (1988). RNA-binding properties of hnRNP proteins. Eur J Biochem *171*, 565-9.

Schwer, B., and Gross, C. H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre- mRNA splicing. Embo J *17*, 2086-94.

Screaton, G. R., Caceres, J. F., Mayeda, A., Bell, M. V., Plebanski, M., Jackson, D. G., Bell, J. I., and Krainer, A. R. (1995). Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. Embo J *14*, 4336-49.

Segault, V., Will, C. L., Polycarpou-Schwarz, M., Mattaj, I. W., Branlant, C., and Luhrmann, R. (1999). Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic steps of pre-mRNA splicing in HeLa nuclear extracts. Mol Cell Biol *19*, 2782-90.

Senapathy, P. (1995). Introns and the origin of protein-coding genes [letter; comment]. Science *268*, 1366-7; discussion 1367-9.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell *59*, 349-58.

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R., and Steitz, T. A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. Nat Struct Biol *4*, 215-22.

Shamoo, Y., Abdul-Manan, N., Patten, A. M., Crawford, J. K., Pellegrini, M. C., and Williams, K. R. (1994). Both RNA-binding domains in heterogenous nuclear

ribonucleoprotein A1 contribute toward single-stranded-RNA binding. Biochemistry *33*, 8272-81.

Sharp, P. A. (1985). On the origin of RNA splicing and introns. Cell *42*, 397-400.

Shen, E. C., Henry, M. F., Weiss, V. H., Valentini, S. R., Silver, P. A., and Lee, M. S. (1998). Arginine methylation facilitates the nuclear export of hnRNP proteins. Genes Dev *12*, 679-91.

Shevchenko, A., Wilm, M., Vorm, O., Jensen, O. N., Podtelejnikov, A. V., Neubauer, G., Mortensen, P., and Mann, M. (1996). A strategy for identifying gel-separated proteins in sequence databases by MS alone. Biochem Soc Trans *24*, 893-6.

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*, 1983-93.

Siebel, C. W., Admon, A., and Rio, D. C. (1995). Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. Genes Dev *9*, 269-83.

Siebel, C. W., Kanaar, R., and Rio, D. C. (1994). Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. Genes Dev *8*, 1713-25.

Sikorski, R. S., Boguski, M. S., Goebl, M., and Hieter, P. (1990). A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell *60*, 307-17.

Siomi, M. C., Eder, P. S., Kataoka, N., Wan, L., Liu, Q., and Dreyfuss, G. (1997). Transportin-mediated nuclear import of heterogeneous nuclear RNP proteins. J Cell Biol *138*, 1181-92.

Smith, C. W., Chu, T. T., and Nadal-Ginard, B. (1993). Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Mol Cell Biol *13*, 4939-52.

Solowska, J., Guan, J. L., Marcantonio, E. E., Trevithick, J. E., Buck, C. A., and Hynes, R. O. (1989). Expression of normal and mutant avian integrin subunits in rodent cells. J Cell Biol *109*, 853-861.

Sontheimer, E. J., Gordon, P. M., and Piccirilli, J. A. (1999). Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. Genes Dev *13*, 1729-41.

Soulard, M., Barque, J. P., Della Valle, V., Hernandez-Verdun, D., Masson, C., Danon, F., and Larsen, C. J. (1991). A novel 43-kDa glycoprotein is detected in the nucleus of mammalian cells by autoantibodies from dogs with autoimmune disorders. Exp Cell Res *193*, 59-71.

Soulard, M., Della Valle, V., Siomi, M. C., Pinol-Roma, S., Codogno, P., Bauvy, C., Bellini, M., Lacroix, J. C., Monod, G., Dreyfuss, G., and et al. (1993). hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. Nucleic Acids Res *21*, 4210-7.

Southby, J., Gooding, C., and Smith, C. W. (1999). Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of alpha-actinin mutally exclusive exons. Mol Cell Biol *19*, 2699-711.

Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol Cell Biol *14*, 7670-82.

Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-26.

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.

Strick, R., and Laemmli, U. K. (1995). SARs are cis DNA elements of chromosome dynamics: synthesis of a SAR repressor protein. Cell *83*, 1137-48.

Sun, J. S., and Manley, J. L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. Genes Dev *9*, 843-54.

Sun, Q., Mayeda, A., Hampson, R. K., Krainer, A. R., and Rottman, F. M. (1993). General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev *7*, 2598-608.

Svitkin, Y. V., Ovchinnikov, L. P., Dreyfuss, G., and Sonenberg, N. (1996). General RNA binding proteins render translation cap dependent [published erratum appears in EMBO J 1997 Feb 17;16(4):896]. Embo J *15*, 7147-55.

Swanson, M. S., and Dreyfuss, G. (1988). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Mol Cell Biol *8*, 2237-41.

Swanson, M. S., and Dreyfuss, G. (1988). RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. Embo J *7*, 3519-29.

Tacke, R., Chen, Y., and Manley, J. L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proc Natl Acad Sci U S A *94*, 1148-53.

Tacke, R., and Manley, J. L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. Embo J *14*, 3540-51.

Tacke, R., Tohyama, M., Ogawa, S., and Manley, J. L. (1998). Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. Cell *93*, 139-48.

Tarn, W. Y., and Steitz, J. A. (1996a). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. Science *273*, 1824-32.

Tarn, W. Y., and Steitz, J. A. (1996b). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell *84*, 801-11.

Tarn, W. Y., and Steitz, J. A. (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. Genes Dev *8*, 2704-17.

Tazi, J., Kornstadt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C., and Luhrmann, R. (1993). Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. Nature *363*, 283-6.

Teigelkamp, S., Achsel, T., Mundt, C., Gothel, S. F., Cronshagen, U., Lane, W. S., Marahiel, M., and Luhrmann, 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*, 127-41.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C. L., and Luhrmann, 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*, 1313-26.

Teigelkamp, S., McGarvey, M., Plumpton, M., and Beggs, J. D. (1994). The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. Embo J *13*, 888-97.

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*, 320-6.

Tian, H., and Kole, R. (1995). Selection of novel exon recognition elements from a pool of random sequences. Mol Cell Biol *15*, 6291-8.

Tian, M., and Maniatis, T. (1994). A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev *8*, 1703-12.

Tiepolo, L., and Zuffardi, O. (1976). Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. Hum Genet *34*, 119-24.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-10.

Umen, J. G., and Guthrie, C. (1995). A novel role for a U5 snRNP protein in 3' splice site selection. Genes Dev *9*, 855-68.

Urlaub, H., Kruft, V., Bischof, O., Muller, E. C., and Wittmann-Liebold, B. (1995). Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. Embo J *14*, 4578-88.

Urlaub, H., Thiede, B., Muller, E. C., Brimacombe, R., and Wittmann-Liebold, B. (1997). Identification and sequence analysis of contact sites between ribosomal proteins and rRNA in Escherichia coli 30 S subunits by a new approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing. J Biol Chem *272*, 14547-55.

Valcarcel, J., Gaur, R. K., Singh, R., and Green, M. R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. Science *273*, 1706-9.

van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., and Grivell, L. A. (1986). Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. Cell *44*, 225-34.

Venables, J. P., Elliott, D. J., Makarova, O. V., Makarov, E. M., Chew, S. L., Malygin, A. A., Bulygin, D. M., Graifer, G. G., Karpova, G. G., Cooke, H. J., and Eperon, I. C. (1999). RBM, a probable human spermatogenesis factor, is a tissue-specific splicing modulator that interacts with Tra2ß. *submitted.*

Venables, J. P., Vernet, C., Chew, S. L., Elliott, D. J., Cowmeadow, R. B., Wu, J., Cooke, H. J., Artzt, K., and Eperon, I. C. (1999). T-STAR/ETOILE: a novel relative of SAM68 that

interacts with an RNA- binding protein implicated in spermatogenesis. Hum Mol Genet *8*, 959-69.

Vergnaud, G., Page, D. C., Simmler, M. C., Brown, L., Rouyer, F., Noel, B., Botstein, D., de la Chapelle, A., and Weissenbach, J. (1986). A deletion map of the human Y chromosome based on DNA hybridization. Am J Hum Genet *38*, 109-24.

Vogt, P. H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Kohn, F. M., Schill, W. B., Farah, S., Ramos, C., Hartmann, M., Hartschuh, W., Meschede, D., Behre, H. M., Castel, A., Nieschlag, E., Weidner, W., Grone, H. J., Jung, A., Engel, W., and Haidl, G. (1996). Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum Mol Genet *5*, 933-43.

Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C., and Fu, X. D. (1998). SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. J Cell Biol *140*, 737-50.

Wang, J., and Manley, J. L. (1995). Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways. Rna *1*, 335-46.

Wang, J., Takagaki, Y., and Manley, J. L. (1996). Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. Genes Dev *10*, 2588-99.

Wang, J., Xiao, S. H., and Manley, J. L. (1998). Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. Genes Dev *12*, 2222-33.

Wassarman, K. M., and Steitz, J. A. (1992). The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. Mol Cell Biol *12*, 1276-85.

Watkins, N. J., Newman, D. R., Kuhn, J. F., and Maxwell, E. S. (1998). In vitro assembly of the mouse U14 snoRNP core complex and identification of a 65-kDa box C/D-binding protein. Rna *4*, 582-93.

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*, 3580-8.

Weidenhammer, E. M., Singh, M., Ruiz-Noriega, M., and Woolford, J. L., Jr. (1996). The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccharomyces cerevisiae. Nucleic Acids Res *24*, 1164-70.

Wickens, M. (1990). How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biochem Sci *15*, 277-81.

Will, C. L., Kastner, B., and Lührmann, R. (1994). Analysis of ribonucleoprotein interactions. In RNA Processing, S. J. Higgins and B. D. Hames, eds. (Oxford: IRL Press), pp. 141-177.

Will, C. L., and Luhrmann, R. (1997a). Protein functions in pre-mRNA splicing. Curr Opin Cell Biol *9*, 320-8.

Will, C. L., and Lührmann, R. (1997b). Mammalian pre-mRNA splicing factors. In Eukaryotic mRNA processing, (A. R. Krainer, ed.: IRL Press), pp. 130-173.

Will, C. L., Schneider, C., Reed, R., and Luhrmann, R. (1999). Identification of both shared and distinct proteins in the major and minor spliceosomes. Science *284*, 2003-5.

Wilson, S. H., Cobianchi, F., and Guy H. R., (1987). cDNA cloning and structure-function relationships of a mammalian helix destabilizing protein: hnRNP particle core protein A1. In DNA: Protein Interaction and Gene Regulation, (Thompson, E. D. and Papaconstantinou, J., ed.:U. Texas Press, Austin, TX), pp. 129-146.

Wilusz, J., Pettine, S. M., and Shenk, T. (1989). Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. Nucleic Acids Res *17*, 3899-908.

Wise, J. A. (1993). Guides to the heart of the spliceosome. Science *262*, 1978-9.

Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-70.

Wu, S., and Green, M. R. (1997). Identification of a human protein that recognizes the 3' splice site during the second step of pre-mRNA splicing. Embo J *16*, 4421-32.

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*, 2542-53.

Xiao, S. H., and Manley, J. L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. Genes Dev *11*, 334-44.

Xie, J., Beickman, K., Otte, E., and Rymond, B. C. (1998). Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. Embo J *17*, 2938-46.

Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A., and Krainer, A. R. (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. Structure *5*, 559-70.

Yeakley, J. M., Tronchere, H., Olesen, J., Dyck, J. A., Wang, H. Y., and Fu, X. D. (1999). Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. J Cell Biol *145*, 447-55.

Young, B. D. (1984). Measurement of sedimentation coefficients and computer simulation of rate-zonal separations. In Centrifugation-A practicle approach, D. Rickwood, ed. (Oxford.Washington DC: IRL Press), pp. 127-160.

Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev *6*, 837-47.

Zahler, A. M., Neugebauer, K. M., Lane, W. S., and Roth, M. B. (1993). Distinct functions of SR proteins in alternative pre-mRNA splicing. Science *260*, 219-22.

Zhuang, Y., and Weiner, A. M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell *46*, 827-35.

Zillmann, M., Zapp, M. L., and Berget, S. M. (1988). Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. Mol Cell Biol *8*, 814-21.

Zuo, P., and Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev *10*, 1356-68.

Zuo, P., and Manley, J. L. (1993). Functional domains of the human splicing factor ASF/SF2. Embo J *12*, 4727-37.

Zuo, P., and Manley, J. L. (1994). The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. Proc Natl Acad Sci U S A *91*, 3363-7.