This is to certify that the
dissertation entitled

Characterization of galectin-3-snRNP complexes and mechanism of galectin entry into the splicing pathway

presented by

Kevin C. Haudek

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Biochemistry and Molecular Biology

Ronald J. Patterson
Major Professor's Signature

10/15/07
Date

Characterization of galectin-3-snRNP complexes and mechanism of galectin entry into the splicing pathway

By

Kevin C. Haudek

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biology

2007

Abstract

CHARACTERIZATION OF GALECTIN-3-SNRNP COMPLEXES AND MECHANISM OF GALECTIN ENTRY INTO THE SPLICING PATHWAY

By

Kevin C. Haudek

In previous studies, galectin-1 and galectin-3 have been found to be redundant pre-mRNA splicing factors using a cell free splicing system. In addition, immunoprecipitation experiments showed both galectin-1 and -3 assembled onto the spliceosome, from early to late stages of complex formation. Results presented here focus on the role of galectin-3 outside of the spliceosome and the mechanism of its entry into the pre-mRNA splicing pathway. Immunoprecipitation experiments of HeLa nuclear extract showed the association of galectin-3 with 5 snRNAs (U1, U2, U4, U5 and U6) and associated snRNP proteins (Sm core polypeptides and U1 specific protein U1-70K). In addition, galectin-3 was complexed with other RNA processing proteins including PSF (PTB-associated splicing factor), general transcription factor TFII-I and the SMN (survival of motor neuron) protein. Fractionation of HeLa nuclear extract on glycerol gradients showed a co-sedimentation of galectin-3 with multiple snRNP complexes. In particular, one complex (~10S) showed an RNase A sensitive association between U1 snRNP and galectin-3. Because U1 snRNP is known to bind the pre-mRNA first at the 5' splice site in a step-wise assembly of spliceosomes, we tested whether this isolated galectin-3-U1 complex was sufficient to recognize and load galectin-3 onto pre-mRNA. Our results suggest that galectin-3 enters the pre-mRNA splicing pathway with U1 under conditions that allow early splicing complexes to form.

We then tested whether galectin-3 is a *bona fide* component of the early splicing complex (E complex). To test this, we used a galectin-independent means of spliceosome selection. A novel pre-mRNA containing three recognition sites for the MS2 bacteriophage protein has been engineered and used to select specific splicing complexes. Using this system and differing incubation conditions, we investigated whether we could isolate unique splicing complexes formed on the pre-mRNA. Of particular interest was the isolation of the early spliceosome complex containing U1 snRNP. Once we had confirmed the specificity of complex formation and isolation, we analyzed proteins selected on the pre-mRNA formed under conditions to allow the formation of E complex. Galectin-3 was identified as a protein member of the E complex. Our results from RNA selection methods validate galectin-3 as a member of early splicing complexes containing U1 snRNP.

Acknowledgements

I would like to thank my mentor, Dr. Ron Patterson, for his invaluable support and guidance in my graduate education. I appreciate his help and understanding through all situations during my graduate career.

I would also like to thank Dr. John Wang for his assistance in developing me as a scientist. I also appreciate his help in organizing and finalizing this thesis.

Finally, I would like to thank Penny and my family and friends for their encouragement and support during this endeavor.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 1

Literature Review

I. Galectins

A. Family overview

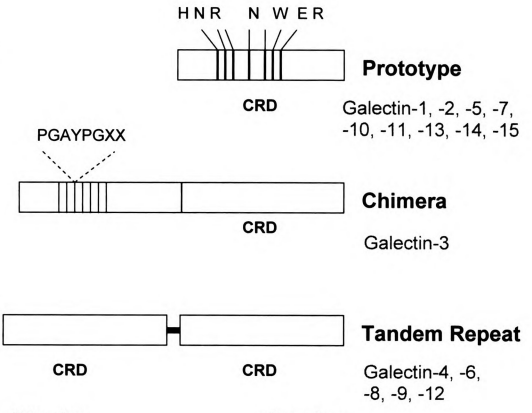
Members of the galectin family of proteins share two essential characteristics: one, they have binding affinity for β-galactosides and two, they contain conserved residues in the carbohydrate binding domain [1]. So far, fifteen mammalian proteins have been classified as galectins. Other organisms, ranging from amphibians to fungi, also have been found to contain galectins. However, other organisms vary in the number and identity of galectins they contain [2].

The galectin family is divided into three sub-groups based on their domain architecture (see Figure 1). The Prototype sub-group consists of members possessing a single carbohydrate recognition domain (CRD) of approximately 130 amino acids. The Tandem Repeat sub-group consists of members possessing two CRDs fused together by a short linker region. The final sub-group, the Chimera, consists of a sole member, galectin-3 (Gal3) which contains a CRD with a unique N-terminal domain (approximately 130 amino acids). This N-terminal domain consists of a series of proline- and glycine-rich repeats. Differential scanning calorimetry has determined the melting temperature of the N-domain (~40°C) and the CRD (~55°C) of Gal3, which suggest the domains fold independently of one another [3]. However, recent NMR evidence suggests that portions of the N-domain may contact the CRD [4].

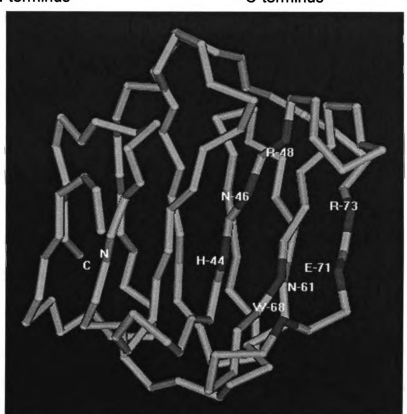
X-ray crystallography has been used to determine the three-dimensional structure of the CRD from a range of galectin family members, including members of all three

Figure 1. (A) Schematic diagram illustrating the domain structure of the galectin family subgroups. Conserved amino acid residues in the CRD are shown. The unique N-terminal domain of galectin-3 contains repeats rich in proline and glycine residues. The single-letter amino acid code is used, where X denotes any amino acid. (B) Illustration of the X-ray crystal structure of human galectin-7. The overall structure of the CRD when bound to ligand is shown by the polypeptide backbone (gray). Highlighted and numbered residues are involved in binding of saccharide ligands. The saccharide ligand is not shown in the figure.

A.


HNR N WER
Prototype
CRD
Galectin-1, -2, -5, -7,
-10, -11, -13, -14, -15
PGAYPGXX
Chimera
CRD
Galectin-3
Tandem Repeat
CRD
CRD
Galectin-4, -6,
-8, -9, -12

B.


N-terminus
C-terminus
R-48
N-46
R-73
N
C
H-44
E-71
N-61
W-68

subgroups [5-7]. The overall structure contains two anti-parallel β-sheets in a sandwich-like structure (see Figure 1B). The carbohydrate binding site consists of amino acids contained on the side of the sandwich-like structure. The highlighted residues in Fig. 1B show the amino acids in human galectin-7 that are conserved among the galectin family and make contact with the saccharide ligand [5]. It should be noted that the affinity of galectins for the monosaccharide galactose is fairly weak [8]. Larger saccharides, such as the disaccharide lactose or some oligosaccharides, bind to galectins with greater affinity. This may indicate that these larger ligands can make additional contacts with the galectin outside of the conserved residues in the CRD. However, these additional contacts are far less conserved in the galectin polypeptide and different galectins exhibit different affinities for these larger ligands [9].

Interestingly, many members of the galectin family exhibit dual-localization, that is, they can be found both intra- and extracellularly. Inside the cell, galectins have been found to localize both in the cytoplasm and the nucleus [9, 10]. The mechanism by which the galectins are moved to the extracellular space is poorly understood as none of the galectins contain an identifiable signal sequence for externalization via the endomembrane system [10, 11].

B. Intracellular galectin-1

Galectin-1 (Gal1) is a member of the Prototype sub-group and has been found in the cytoplasm of various cell types [12-15]. It should be noted that the localization of Gal1 can change on the basis of cell differentiation. For example, Cooper and Barondes stained myoblasts for Gal1 and showed an intracellular distribution of the protein [16]. As myoblasts fused into myotubes, Gal1 was exported to the extracellular space via a

novel secretory pathway involving blebs of the plasma membrane. In studies using H-Ras(12V)-transformed Rat-1 (EJ) cells, Gal1 has been shown to interact with H-Ras(12V) in the cytoplasm [17]. This interaction leads to membrane anchorage of H-Ras and promotes cell transformation. The interaction between Gal1 and H-Ras is thought to be direct due to several lines of evidence including UV cross-linking and reciprocal immunoprecipitations [17, 18]. To prevent cell transformation by H-Ras(12V), anti-sense RNA for Gal1 was transfected into baby hamster kidney cells. These cells showed a reduction in H-Ras(12V) clustering in non-raft microdomains at the plasma membrane [17, 18]. This is consistent with the fact that membrane anchorage of H-Ras is necessary for cell transformation. Finally, a mutant form of Gal1 (L11A) shows normal carbohydrate binding activity but mislocalizes H-Ras(12V) and prevents H-Ras(12V) GTP-loading in COS-7 cells [19]. Transfections of Rat-1 cells with this Gal1 mutant show reduced levels of cell transformation [19]. A recent report indicates that like Gal1, Gal3 may have a similar binding pocket to interact with Ras [20].

Other reports have found Gal1 localized in the nucleus of cells as well. Choi et al. reported the association of Gal1 with the nuclear matrix of rat calvarial osteoblasts [21]. This interaction is also dependent on cell development, as Gal1 was found in nuclear matrix preparations only in differentiated osteoblasts. In other studies, HeLa cervical carcinoma cells showed Gal1 in the nucleus and co-localized with known splicing factors (such as the Sm core proteins of snRNPs, see below) by immunofluorescence [22]. Using the Gal1 protein as bait in a yeast-two-hybrid system and screening a HeLa cell cDNA library identified the C-terminal 50 amino acids of gemin4 (Gemin4C50) as a binding partner [23]. This was confirmed by direct binding assays of Gal1 and a fusion

protein containing Gemin4C50. Gemin4 can be found both in the cytoplasm and nucleus of HeLa cells. However immunoprecipitation experiments of HeLa nuclear extracts (NE) using antibodies against Gal1 co-immunoprecipitate other known gemin4 interacting proteins such as the Survival of Motor Neuron protein (SMN, see below) and gemin2 [23]. This strongly suggests that Gal1 does interact with gemin4 in the nucleus, but does not rule out the possibility of additional interactions in the cytoplasm. Gal1 also has been shown to play a role in pre-mRNA splicing using HeLa NE [24]. Extracts depleted of both Gal1 and Gal3 cannot catalyze the splicing reaction of an exogenous pre-mRNA. Adding back recombinant Gal1 protein reconstitutes the ability of the extract to carry out pre-mRNA splicing. As well, addition of known carbohydrate ligands of the galectins to HeLa NE inhibits the splicing ability of that extract [25].

C. Intracellular galectin-3

Several reports place Gal3 in both the cytoplasm and nucleus of cells using a variety of microscopy techniques [26-28]. This cellular distribution of Gal3 can change according to cell proliferation state [29]. Phosphorylation of Gal3 also affects its cellular distribution. In 3T3 cells, Gal3 exists in two variants, a non-phosphorylated form which is entirely nuclear, and a phosphorylated form that can be found in both the nucleus and cytoplasm [30]. It has also been shown that the phosphorylated form of nuclear Gal3 is quickly exported to the cytoplasm as part of a large complex, indicating the phosphorylation of Gal3 may play a key role in nuclear export [31]. Further studies of the intracellular transport of Gal3 have determined it to be a shuttling protein, that is, it moves back and forth between the nucleus and cytoplasm of cells [32]. This was demonstrated by creating heterodikaryons of mouse and human cells and monitoring the

appearance of the human Gal3 polypeptide in the mouse nucleus [32]. Additional research has been focused on the nuclear import and export signal of Gal3. It has been found that the necessary nuclear import signal lies in a conserved sequence of amino acids, ITLT, beginning at residue 253 [33]. Site-directed mutants of this sequence show loss of nuclear accumulation of Gal3. An overlapping and neighboring leucine-rich region, at residues 240-255, mediates the nuclear export of Gal3 [34]. Using this export signal, a green fluorescent protein (GFP) construct, normally constrained to the nucleus, is efficiently exported to the cytoplasm. GFP-Gal3 mutants of this export signal show nuclear accumulation by fluorescent microscopy [34]. This leucine-rich export signal indicates that Gal3 may be exported in a CRM-1 dependant manner, consistent with previous reports [32]. In contrast, a report by Nakahara et al. found that Gal3 is imported into the nucleus via importin-alpha, using a nuclear localization signal at residues 223-228 [35].

Gal3 has been documented as having several intracellular binding partners. . Known binding partners for Gal3 include Bcl-2, Chrp, TTF-I, beta-catenin, CBP70, cytokeratin and gemin4. The apoptosis repressor Bcl-2 contains two regions of similarity to Gal3; the N-terminus is proline and glycine rich and the C-terminal portion shares the amino acid sequence NWGR [36]. Bcl-2 binds directly to Gal3 *in vitro* and mutants of Bcl-2 in the NWGR sequence eliminate its anti-apoptotic activity [37]. Chrp was originally identified as a Gal3 binding partner by a yeast-two-hybrid screen of a cDNA library from mouse 3T3 cells [38]. This interaction was then confirmed by *in vitro* binding assays and immunoprecipitations. It also was found that Chrp interacts at the CRD of Gal3 and this interaction is not inhibitable by carbohydrate ligands of Gal3 [39].

It should also be noted Chrp may only be a cytosolic binding partner of Gal3, in that immunofluorescence shows Chrp localized in the perinuclear space and cytoplasm of 3T3 cells [38]. The thyroid-specific transcription factor TTF-I has been identified as a nuclear binding partner of Gal3 in papillary thyroid cancer cells [40]. Direct interaction between the homeodomain of TTF-I and Gal3 was shown using GST-pull down assays. This association with Gal3 enhances the ability of TTF-I to bind DNA in a gel-retardation assay [40]. Gal3 was co-purified with another intracellular lectin, CBP70, from HL60 nuclei when isolated on saccharide affinity beads [41]. Binding of Gal3 to lactose disrupts the interaction between CBP70 and Gal3 [42]. Although some functions have been assigned to CBP70, the significance of the Gal3-CBP70 interaction remains to be uncovered. Gal-3 also has been reported to bind beta-catenin, determined by immunoprecipitations and *in vitro* binding assays [43]. Localization studies of these two proteins show a strong correlation and Gal-3 may activate beta-catenin stimulation of cyclin and c-myc expression [28]. There is evidence to suggest an *in vitro* association of cytokeratins with Gal3, due to novel glycosylated residues on the cytokeratins [44]. This would implicate the cytokeratins as true intracellular carbohydrate ligands of Gal3 [10]. Like Gal1, Gal3 has been described as a component of the nuclear matrix, however its binding partner has not been identified [45]

Gal3, similar to Gal1, also has been found to bind a fusion construct of Gemin4C50 directly, using an *in vitro* binding assay [23]. Whether this indicates an involvement of Gal3 (similar to Gal1) with the SMN protein and complex remains to be investigated. However, adding the purified N-domain of Gal3 to an *in vitro* splicing assay inhibits the splicing reaction, suggesting that Gal3 makes important contacts

necessary for pre-mRNA splicing [23]. Equally interesting, GST-Gemin4C50 also acts in a dominant negative manner when added to splicing extracts.

II. SnRNPs

A. Biogenesis and assembly

Five small nuclear RNAs (snRNAs) participate directly in pre-mRNA splicing; U1, U2, U4, U5 and U6. These five snRNAs are closely conserved among vertebrates and contain a large percentage of uracil bases and are commonly referred to as U snRNAs. These snRNAs are commonly found complexed with proteins in the nucleus, creating a small nuclear ribonucleoprotein particle (snRNP) [46]. Each snRNP contains an unique snRNA, a common set of Sm core proteins and snRNP-specific proteins unique only to that type of snRNP when isolated under stringent conditions [47]. The snRNAs undergo an extensive biogenesis pathway before they can participate in pre-mRNA splicing [48].

Transcription of most of the U snRNAs is carried out by RNA Polymerase II (Pol II). U6 snRNA, which is a RNA Polymerase III (Pol III) transcript, is the lone exception (see below). The Pol II transcribed U snRNAs are made as long primary transcripts (pre-snRNAs) with a monomethyl-guanosine cap structure [49]. Trimming of the pre-snRNAs requires both phosphorylation of the C-terminal domain of Pol II and a *cis*-acting signal in the transcript [50]. The snRNAs are then packaged for export with proteins including the cap-binding complex (CBC); PHAX, which is found phosphorylated in the nucleus; and CRM1/RanGTP. After export from the nucleus, hydrolysis of Ran-bound GTP and dephosphorylation of PHAX induces disassembly of the snRNA export complex [51, 52].

Assembly of the snRNAs into snRNPs occurs in the cytoplasm. The key factor in the cytoplasmic assembly and biogenesis of snRNPs is the SMN protein complex [53, 54]. This complex is responsible for loading the full complement of Sm proteins (B/B', D1-3, E, F, G) onto the consensus snRNA Sm site to form the core snRNP [54, 55]. The Sm proteins are known to form a heptamer ring when bound to the snRNA. However, the Sm proteins are loaded onto the snRNA in discrete sub-complexes, not in a fully assembled heptamer ring [56, 57]. Some evidence suggests that the SMN protein, using its Tudor domain, binds the Sm proteins, by recognizing the symmetrical dimethylated arginines contained in the Sm proteins [58-60]. Another member of the SMN complex, Gemin5, is responsible for binding the snRNAs during loading of the Sm core proteins [61, 62]. The binding of the Sm proteins to the snRNA allows further modification of the snRNP including hypermethylation of the cap structure. The monomethyl guanosine cap is modified to become a trimethyl guanosine cap by Tgs1, a methyltransferase, which has also been shown to interact with the SMN complex [63, 64].

After the snRNA has been loaded with Sm proteins and had its cap hypermethylated, it is ready to be imported back to the nucleus. The snRNPs are imported into the nucleus via a series of adapter proteins [65]. The snurportin-1 protein (SPN1) is responsible for recognizing and binding the hypermethylated cap of the snRNPs [66]. SPN1 binds the snRNP cap most likely near the nuclear pore complex (NPC) and has been found to interact with a NPC protein, Nup214 [67]. SPN1 also has a domain for interacting with importin-β, a nuclear import factor, and at some point SPN1 binds importin-β and releases Nup214 [65, 67]. Importin-β then mediates another interaction with the NPC to allow translocation of the snRNP into the nucleus [67]. In

other importin-β systems, importin-β binds RanGTP in the nucleus to dissociate its cargo; however, this does not seem to be the case for snRNP import [68]. The mechanism of dissociation of the snRNPs after import remains under investigation. Some evidence suggests that the cytoplasmic SMN complex (or sub-complexes thereof) is transported along with snRNPs into the nucleus through its affinity for the Sm proteins [69, 70]. Other evidence suggests a second import pathway for snRNPs which recognizes the assembled Sm core on the snRNP as opposed to the hypermethylated cap structure [71].

After import into the nucleus, the snRNPs are trafficked to the Cajal bodies. Coilin is a key protein component of the Cajal body and can, in most cell types, directly bind the SMN protein in Cajal bodies [72]. It is hypothesized that the SMN binding event to coilin is what carries the imported snRNPs to the Cajal bodies [73]. Once at the Cajal body the imported snRNPs bind any necessary snRNP-specific protein, which enter the nucleus independent of the snRNP. This completes assembly of the protein components with the snRNAs.

Another important step in snRNP biogenesis at the Cajal body is modification of the snRNA. SnRNAs can be either 2'-O-methylated or pseudouridylated by small Cajal body-specific RNAs (scaRNA) [74]. These scaRNAs resemble small nucleolar RNAs (responsible for modifying ribosomal RNA) in their sequence and structure [75]. The modifications performed on snRNPs are essential for their function in pre-mRNA splicing [76]. After the snRNPs have been assembled and modified they are moved to nuclear speckles, which are thought of as storage sites for various splicing factors [73, 77].

In contrast to the other U snRNAs, biogenesis of U6 snRNA is confined to the nucleus and transcribed by Pol III. The transcribed U6 snRNA is given a unique γ-monomethyl cap structure, which it keeps throughout its lifecycle [48]. Transcription of the U6 gene is terminated by a short stretch of uridines. There is evidence to suggest the La protein binds this 3' U-stretch to give stability to the U6 snRNA and allow the core snRNP to assemble [78, 79]. Another unique aspect of U6 is that it binds a set of proteins called the Lsm (Like-Sm) proteins, Lsm2-8, in place of the Sm proteins [80]. These Lsm proteins form a heptamer ring around the 3' U-stretch in the U6 snRNA and share homology with the Sm proteins [81, 82]. There is some evidence that links the Lsm proteins to recycling or regenerating snRNPs between rounds of pre-mRNA splicing [83]. Before being transported to nuclear speckles with the other U snRNAs, U6 also visits the Cajal bodies and nucleolus [84, 85]. Binding of the Lsm proteins to U6 is required to target the U6 snRNA to the nucleolus to undergo base modifications directed by snoRNPs [85, 86].

B. Nuclear multi-snRNP complexes

SnRNPs are found in other nuclear complexes in addition to those described during their biogenesis and role in the active spliceosome. In fact, snRNPs often associate with one another in the nucleus in discrete complexes outside of the spliceosome. These multi-snRNP complexes are loaded as a large particle when used in pre-mRNA splicing, in particular the U4/U6.U5 particle.

1. U4/U6

The U4/U6 di-snRNP was the first multi-snRNP complex to be identified [87]. The importance of this particle became obvious when it was found that disruption of the

basepairing between U4 and U6 or cleavage of the individual snRNAs halted pre-mRNA splicing [88]. The basepaired U4/U6 exists as a 12S particle and contains specific polypeptides, termed U4/U6-specific proteins, not found on either snRNP alone [89]. A U4-specific protein, 15.5K, must be bound to U4 before loading of the U4/U6-specific proteins, suggesting it may be responsible for additional protein loading [90]. p110 (SART3) has been identified as the protein responsible for annealing free U6 snRNP with U4 to form the di-snRNP complex [91]. This is an important process to create functional U4/U6 particles between rounds of pre-mRNA splicing. The formation of these U4/U6 particles occurs in Cajal bodies, a conclusion based on the results of microscopy and mutation studies [92, 93].

2. U4/U6.U5 tri-snRNP

A U4/U6.U5 particle was originally isolated using immuoaffinity and sedimentation centrifugation [94]. This tri-snRNP particle was found to be about 25S and could not be reconstituted by mixing free U5 particle with the 12S U4/U6 particle. This is because the tri-snRNP contains proteins only associated with the U4/U6.U5 particle and not with any individual snRNP alone [94, 95]. It is thought that assembly of this U4/U6.U5 tri-snRNP occurs in the Cajal bodies. Using RNAi to knock down specific U4/U6 or U5 proteins prevents the assembly of the tri-snRNP, while accumulating assembled U4/U6 particles in the Cajal bodies [96]. Recent reports have described a tri-snRNP assembly factor associated with U5 [97]. Interestingly, this protein is associated with the free U5 particle, but is not found associated with the 25S tri-snRNP. It is suggested that this protein participates in loading U5 and the tri-snRNP specific proteins onto U4/U6 then is released from the complex. Other studies have

elucidated a complete protein-protein interaction map of the tri-snRNP by doing a series of yeast-two-hybrid screens [98]. Evidence suggests that the U5 interaction with U4/U6 is based solely on protein-protein contacts, as no RNA-RNA interactions could be detected.

3. Pseudospliceosome

A large complex containing four of the snRNPs, U2, U4, U5 and U6, has been isolated from HeLa cell nuclear extract [99, 100]. This complex, termed the pseudospliceosome is formed in the presence of ATP and incubation at 30°C, conditions similar to those used for *in vitro* splicing assays. Other research suggests the pseudosplicesome forms due to the presence of a 5' splice site [99]. However, U1 is dispensable for the formation of the U2/U4/U5/U6 complex. The pseudospliceosome, although similar to the B splicing complex (see below), is unique from *bona fide* splicing complexes based on native gel mobility shifts and salt conditions used for isolation [100].

C. Other nuclear snRNP complexes

Using less stringent conditions to isolate snRNPs and their associated proteins allows the characterization of several other nuclear complexes. These particles show the association of multiple snRNAs or snRNP-proteins with non-snRNP proteins in large complexes.

1. SMN

The SMN protein is responsible for assembling snRNP particles in the cytoplasm and mutant forms of the SMN protein are the causative agent of spinal muscular atrophy (SMA) [54, 101, 102]. A complex involving the full length SMN polypeptide and several other proteins was originally identified in the nucleus, in a novel nuclear structure

termed a gem (for "gemini of Cajal bodies") [103]. As this nuclear SMN complex was characterized it was found that the complex consisted of SMN, gemins 2-8, Sm core proteins and several of the hnRNP proteins [104-106]. Other interacting proteins in an SMN complex can be detected using less stringent salt conditions for isolation [23, 107].

In the nucleus, the SMN protein can be found in at least three distinct complexes distinguished in sedimentation and chromatography experiments: an 18S complex, a 20S complex and a complex > 20S [108]. The protein composition of the 20S complex contains SMN, gemins2-4, and a subset of the Sm core proteins. It is now thought that the >20S SMN complex represents the SMN complex in its entirety, including the other gemin and Sm core proteins not found in the 20S complex. The full composition of the 18S SMN complex has yet to be investigated [108].

It is of some contention whether the full nuclear SMN complex contains the U snRNAs. Some reports using immunoprecipitation assays at high salt failed to find the U snRNAs with a nuclear SMN complex [103, 108]. Xu et al. however, showed that a GST-SMN fusion protein did pull-down all 5 U snRNAs used in pre-mRNA splicing as well as U7 snRNA from nuclear extract [72]. It was also demonstrated that SMN interacts with coilin, which associates with the snRNAs [72]. There is strong evidence that the SMN interaction with snRNAs may be sensitive to the high salt conditions of the earlier reports. It also should be noted that proteins present in the SMN complex, gemin3 and gemin4, have been found in complexes with unidentified small nuclear RNAs of several hundred nucleotides or U snRNAs [109, 110]. Whether these snRNA complexes are truly exclusive of the nuclear SMN complex or are only the result of using less stringent isolation conditions remains to be investigated.

Although the role of SMN and the SMN complex in the cytoplasm has been well documented in assembling nascent U snRNAs into functional snRNP particles, their role in the nucleus is open to speculation. Antibodies against the SMN protein inhibited pre-mRNA splicing only when pre-incubated with nuclear extract using an *in vitro* splicing assay [111]. Similar antibody inhibition experiments showed that antibodies against SMN prevented active splicing complexes from forming [108]. These results suggested that the SMN protein was responsible for either assembling or delivering snRNPs to the spliceosome to form active splicing complexes. This is further backed by the association of the SMN complex with several required splicing factors [106, 112]. There is also evidence to suggest that SMN interacts with coilin in the Cajal bodies, possibly indicating a role in modifying or assembling nascent snRNPs [113]. In the nucleus, SMN has taken on the moniker of "master assembler" due to its interaction with many different types of small, nuclear RNPs [72, 114-118].

2. Penta-snRNP

Stevens et al. introduced a new model for spliceosome assembly by identifying and characterizing a complex termed a penta-snRNP [119]. Using *S. cerevisiae*, it was found that a 45S complex could be isolated containing all 5 snRNPs, with the snRNAs in a stoichiometric ratio, that is equal amounts of each of the 5 snRNAs were found in the 45S complex. Furthermore, it was found that when the both the protein and snRNA components of snRNPs were labeled, the penta-snRNP complex did not exchange a snRNP in the penta-snRNP for free snRNPs in the extract. This penta-snRNP complex was able to excise an intron from an exogenous pre-mRNA in a cell-free system when complemented with nuclease treated yeast extract. The penta-snRNP complex

contained the Sm core proteins and the snRNP specific proteins. Interestingly, the penta-snRNP also contained other splicing factors, including all 8 members of the Prp19 complex, a complex important for formation of an active spliceosome from early splicing complexes [120, 121]. Using these findings, it was suggested that the penta-snRNP represented a pre-formed spliceosome that could assemble immediately onto a pre-mRNA when encountered.

Other studies in *Schizosaccharomyces pombe* found a similar penta-snRNP complex [122]. Yeast extracts were subjected to sedimentation on glycerol gradients and fractions containing 5 snRNAs were immunoprecipitated by antibodies against two splicing factors, Prp1 and Prp 31. These two proteins were shown to be in a pre-spliceosomal complex with 5 snRNAs. Interestingly, several studies indicate there may be a similar mammalian penta-snRNP complex [123, 124]. Using HeLa NE, it was found that U1 snRNA paired at the 5' splice site in a large complex containing all 5 snRNAs, and that this association was abolished when a mutated splice site was used or U1 was degraded [123]. Using a co-transcriptional/splicing assay in human A431 cells, Listerman et al. found that both U1 and U5 snRNP specific proteins could be present on a nascent pre-mRNA after only the 5' splice site in the pre-mRNA had been transcribed. [124]. This transcripition/splicing assay relied on chromatin immunoprecipitation techniques. Antibodies against known splicing factors and/or snRNP specific proteins were used to immunoprecipitate a nascent mRNA still attached to the DNA template by RNA Pol II. By using a series of reverse transcription primers, the researchers could determine where on the gene the RNA polymerase was located. The key result from these experiments was that antibodies against U5 (and U1) were able to precipitate the

nascent RNA, even though the RNA polymerase had not transcribed the branchpoint or 3' splice site fo the gene. The association of U1 and U5 with the 5' splice site before transcription of the 3' splice site suggests a multi-snRNP mammalian complex for spliceosome assembly.

3. PCC

Further evidence for a mammalian penta-snRNP was provided by the isolation of a complex containing all 5 snRNAs from HeLa nuclear extract using low salt conditions [110]. This complex is called the PCC (PSF-containing complex) and contains many RNA processing factors in addition to the snRNAs, snRNP proteins and PSF (PTB-associated splicing factor). Proteomics uncovered the presence of many hnRNP proteins, SMN, gemin4, RNA helicases, pre-mRNA splicing factors, several transcription factors and most important for our studies, galectin-3. PCC is formed in the absence of pre-mRNA and under incubation conditions that do not allow active spliceosomes to form. Using velocity sedimentation, Peng et al [110] found the size of the PCC to be near 60S, similar to the size of a spliceosome [125]. It is suggested that the PCC may represent a pre-formed pre-spliceosome containing all 5 snRNAs. Since the PCC contains components of both the above described SMN complex and the penta-snRNP, it is possible that the PCC represents an assembly of the penta-snRNP with or by the SMN complex. However, there is currently no functional evidence to suggest how PCC, penta-snRNP and the nuclear SMN complex relate to each other.

III. Pre-mRNA splicing

A. Spliceosome

Pre-mRNA splicing is an essential process for eukaryotic cells. Newly transcribed pre-mRNA undergoes splicing to remove intervening sequences (introns). Pre-mRNA splicing is carried out by a large complex consisting of both RNA and many proteins, termed the spliceosome [126]. This process uses two transesterification reactions to excise an intron and join two exons. The pre-mRNA contains important information regarding the splice sites in its sequence. The most critical sequence motifs are absolutely conserved: a 5' splice site, a 3' splice site and a branchpoint adenine [127-129]. Other pre-mRNA sequences, such as the polypyrimidine tract and splicing silencers and enhancers, also play a role in identifying the correct splice sites, but these can vary between gene transcripts.

The 5 U snRNAs are members of the spliceosome as snRNPs and serve an important role for basepairing the pre-mRNA at the correct sites [130]. Attempts at cataloging the proteins in or associated with the spliceosome have shown over 300 distinct polypeptides involved in this large complex [131]. The numerous protein components and 5 described snRNPs involved with the spliceosome suggest a large assembly. Active spliceosomes have been sedimented to 60S, confirming the magnitude of this complex [125]. There is some evidence that spliceosomes may oligomerize forming even larger supraspliceosomes of 200S [123, 132] .

The process of pre-mRNA splicing is ATP-dependent. This is likely due to the numerous RNA helicases involved in pre-mRNA splicing, as the actual chemical reactions necessary for intron removal are energetically favorable [133].

Another important finding is that the spliceosome may be considered another example of a ribozyme. Using protein-free RNA fragments of U2, U6 and branchpoint

RNA the products of the first transesterification reaction can be detected, suggesting only the properly assembled RNA components of the spliceosome are needed for the reaction [134, 135]. Protein components of the spliceosome may serve such roles as a scaffold for loading the snRNPs correctly, unwinding RNA to assure proper base-pairing or recruiting other factors necessary for mRNA processing following pre-mRNA splicing. In addition, protein components of the spliceosome play important roles in alternative splicing by helping choose the correct splice site [136, 137].

B. Assembly

The canonical model of spliceosome assembly is termed the step-wise assembly model. It entails the addition of U snRNPs in an ordered, sequential fashion [138]. Because of this ordered addition of factors, distinct complexes with varying snRNA and protein components can be isolated. The common feature in all these complexes is that they are formed on a pre-mRNA. These distinct complexes were most often found using a cell-free splicing system and were originally described by their mobility in native gel electrophoresis [139].

The first snRNA containing complex in the step-wise assembly model is the E, or early, complex. In yeast, it is commonly referred to as the commitment complex, CC [128]. This complex forms in the absence of ATP and at temperatures as low as 4°C. The key feature of this complex is the base-pairing of U1 snRNP to the 5' splice site [140]. In addition to the base-pairing of the U1 snRNA to the 5' splice site of the pre-mRNA, many important protein-pre-mRNA contacts are detected as well [141]. There is good evidence that the specificity of the U1 interaction at the 5' splice site may be due in part to protein-RNA interactions in addition to the U1-pre-mRNA base-pairing [142]. In

addition, the U2 snRNP may be present in the early E complex; however, it has not yet base-paired with the pre-mRNA [143, 144]. It is currently not known if specific protein contacts on the pre-mRNA are necessary before the loading of U1 in the E complex. Kent et al. have data to suggest an even earlier complex in the spliceosome assembly pathway, termed the E' complex, which may represent a precursor to the E complex [145]. This E' complex contains the U1 snRNA bound to the pre-mRNA at the 5' splice site but lacks U2AF, a protein component of the E complex.

The spliceosome then proceeds to the ATP-dependent A complex. In this complex, U2 snRNA binds to the branchpoint sequence [146, 147]. The ATP-dependence is due to UAP56, a member of the DExD/H-box helicase family of proteins. UAP56 is required for the association of U2 with the pre-mRNA branchpoint [148]. There is some evidence to suggest that UAP56 is actually removing other proteins from the branchpoint sequence, allowing U2 to bind there, as opposed to changing the RNA structure of U2 [149]. Other proteins also add at this complex, including some that may allow U1 and U2 to associate through a protein bridge [150, 151]. Recent studies have also shown that the ends of the U1 and U2 snRNAs are in proximity with one another at this stage of assembly, which may be another mechanism of communication between snRNPs in the spliceosome [152].

The next complex to form on the pre-mRNA is the B complex. This occurs with the addition of the tri-snRNP (U4/U6.U5) [153]. The spliceosome now contains all 5 snRNAs; however it is not catalytically active. To become catalytically active, U4/U6 must unwind its basepairing and U1 must be displaced from the 5' splice site. The 5' splice site is then free to pair with U6. There is also evidence to suggest that U5 makes

ATP-dependent contacts at the 5' splice site and 5' exon, even before U1 leaves the complex [154, 155]. However, it is the switch of basepairing from U1 to U6 at the 5' splice site that is thought to be catalytically important [133]. Afterwards, U1 and U4 are both released to form the active spliceosome, termed the C complex.

For the unwinding of U4 and U6 to happen in the B complex more RNA helicases are required [156]. It is also thought that this unwinding requires ATP. Without the unwinding of U4/U6, splicing cannot proceed presumably because U6 cannot switch for U1 at the 5' splice site [157]. It is also at the B complex where the Prp19 complex, also identified in the penta-snRNP, adds to the spliceosome [121, 158]. The switch to the active catalytic spliceosome is thought to occur when U6, now at the 5' splice site, pairs with U2. With U6 positioned at the 5' splice site and U2 at the branch point, the pairing of U2 and U6 brings the participants in the first transesterification reaction in proximity. The U2/U6 complex adopts a structure of a four-helix junction, involving U2-U6 intermolecular helices I, II, III and an U6 intrastemloop (ISL) [159]. Interestingly, this structure bears great similarity to the structure of a group II self-splicing intron [160]. The U6 intra-stem loop contains the base critical for binding the necessary Mg^{2+} involved in splicing catalysis [161]. The structure of the pairing of U2 and U6 at the 5' splice site also suggests that catalysis of pre-mRNA splicing is due to RNA, as experimental evidence has suggested [134, 160]. Conserved bases on U6 are essential for both transesterification reactions [162, 163]. U5 seems to serve a more structural role, by contacting both exons and keeping them in proximity for the second transesterification reaction to occur [155, 164]. This idea is supported by making deletion or insertion mutants in an invariable loop of U5 and testing these mutants by *in vitro* splicing assays.

These U5 mutants led to defects in splicing, due to misalignment of the exons in the splicing reaction [165]. The U5 particle also undergoes remodeling with respect to its protein components during spliceosome activation, although the significance of this is not yet understood [166].

An alternative model of spliceosome assembly using the previously described penta-snRNP has been proposed [119]. The penta-snRNP could load onto a pre-mRNA at once, containing all 5 snRNAs and many necessary splicing factors. This penta-snRNP addition would resemble the B complex of the step-wise addition, with respect to snRNA composition. For the penta-snRNP assembled on the pre-mRNA, all that would be necessary is snRNA unwinding and correct base-pairing (plus additional necessary protein contacts, if any) to proceed to the catalytic spliceosome [167, 168]. These two models of spliceosome assembly are not necessarily mutually exclusive [169]. It may be the case that either is capable of carrying out pre-mRNA splicing and both mechanisms of assembling spliceosomes may be used *in vivo* [110].

C. Proteomics

Recently, there have been several attempts to characterize the protein components of spliceosomes using tandem mass spectrometry. Rappsilber et al. used two different biotinylated pre-mRNA substrates and gel filtration to analyze a mixture of spliceosomal complexes [131]. Similarly, Zhou et al. used two different pre-mRNA substrates each containing an affinity sequence for the MS2 viral coat protein [170]. These selected complexes were than separated by gel filtration chromatography. These two large proteomic studies used conditions that allowed the pre-mRNA to be associated in many spliceosomal complexes. Subjecting the isolated complexes to tandem mass

spectrometry, these studies identified approximately 300 [131] and 150 [170] unique proteins associated with the spliceosome, respectively. The most striking result however, was the association of many proteins not before linked to pre-mRNA splicing [171]. Included in this group are transcription factors, translation factors, histone acetyl-transferases and many ribosomal proteins.

Other proteomic studies have attempted to isolate a single spliceosomal complex by stopping spliceosome assembly at a distinct step. Makarov et al. assembled B complexes and immunoselected them with an antibody against a specific B complex factor [166]. Further purification of this complex and subsequent tandem mass spectrometric analysis identified approximately 100 proteins associated with the spliceosomal B complex. Fewer novel proteins were identified than previous studies. From the identified proteins, the researchers concluded that they indeed had isolated a pure B complex due to the lack of identified C complex proteins. A second major discovery was a significant rearrangement of the 35S U5 particle upon addition to the spliceosome. This included dissociation of numerous proteins from the U5 particle after U5 had bound the spliceosome, indicating a subset of snRNP proteins interacts with U5 outside of the spliceosome, but not in the spliceosome [166]. A second complex isolation was performed by Jurica et al of the catalytically active C complex [172]. These complexes were isolated by affinity selection and size exclusion chromatography. Important in this study was the finding of unique C complex proteins. These proteins associated with the spliceosome only in the active C complex, as opposed to ATP-independent early complexes.

Finally, recent attempts have been made to characterize spliceosomes under "physiological conditions" [173, 174]. Isolation procedures that lack heparin and use lower salt concentrations in buffers were performed. These are unique in that previous studies employed both heparin and salt concentrations ranging from 100 to 250 mM NaCl [131, 166, 170]. The B complex was isolated and characterized under conditions of lower salt and lacking heparin [173]. Similarly, the pre-spliceosomal A complex was purified and submitted to tandem mass spectrometry [174]. In addition to finding many of the same proteins identified previously, these studies did find heretofore unidentified proteins associated with the spliceosome. This leads one to believe that by changing isolation conditions, unique proteins may be found that were missed during the early studies. However, the true significance is somewhat uncertain as the authors argue that some of the identified proteins are probably contaminants due to the low stringency conditions [173] . A new *in vivo* method of isolating spliceosomes from chicken and human cells has recently been used [175]. This technique involves adding a tandem-affinity purification (TAP) tag to the SmD3 protein at the native genomic location [176]. This tagged native protein and its associated proteins and RNA were then isolated using the developed affinity tag and the resulting complexes were subjected to glycerol gradient sedimentation followed by tandem mass spectrometry. This method allowed the identification of splicing complexes assembled *in vivo*. The most striking result from this study was the identification of numerous nuclear proteins not isolated by the cell-free splicing system applied earlier [131, 170]. However, many of these newly identified proteins have functions outside of pre-mRNA splicing (such as nuclear transport and 3' mRNA processing) and probably reflect the full life-cycle of the mRNA in the nucleus

[175]. Relatively few new splicing factors were identified in comparison to previous proteomic studies. In conclusion, the results serve as a reminder of the closely coupled nature of RNA biogenesis and processing *in vivo*. Other reports also indicate a close linking of pre-mRNA splicing, transcription and RNA processing [124, 177, 178].

IV. Galectins and pre-mRNA splicing

Numerous lines of evidence have supported the idea that Gal1 and Gal3 are redundant pre-mRNA splicing factors. For example, experimentation has shown co-localization of the galectins with known splicing factors, sedimentation of Gal3 with RNP complexes, depletion and reconstitution of cell free splicing assays and association of galectins with known splicing factors [22-25, 179]. Recently, antisera against Gal1 and Gal3 have been used to immunoprecipitate complexes in an *in vitro* splicing reaction [180]. Co-precipitated with the respective galectin proteins were all the mRNA species expected in a splicing reaction; pre-mRNA, mRNA, excised intron-lariat and intermediates. As well as the RNA components, splicing factors such as the Sm core proteins and Slu7 also were co-precipitated, suggesting that the galectins were interacting with the assembled spliceosome. Other important findings were that galectins do not bind RNA directly, the interaction between the galectins and the spliceosome was sensitive to salt conditions and Gal1 and Gal3 existed in mutually exclusive spliceosomes [180].

In addition to the association of galectins with the spliceosome, there are indications that Gal1 and Gal3 may interact with splicing factors or the splicing machinery outside the assembled spliceosome. Key points of evidence include the galectin association with the SMN protein in nuclear extracts and the sedimentation of

Gal3 in cesium sulfate gradients with RNP complexes [23, 179]. Both these studies lack the exogenous pre-mRNA scaffold necessary for cell-free splicing. Finally, galectin association with the spliceosome appears to be at a very early step in spliceosome formation based on two observations; one, when extracts are depleted of galectins, active splicing complexes cannot form and two, antibodies against the galectins can immunoprecipitate pre-mRNA from early complexes almost immediately after adding the pre-mRNA scaffold [25, 180]. Taken as a whole, these data suggest the possibility that galectins may interact with splicing factors outside the spliceosome and load onto the pre-mRNA with other necessary early factors.

References

1. Barondes, S.H., Castronovo, V., Cooper, D.N., Cummings, R.D., Drickamer, K., Feizi, T., Gitt, M.A., Hirabayashi, J., Hughes, C., Kasai, K., and et al. (1994). Galectins: a family of animal beta-galactoside-binding lectins. Cell *76*, 597-598.

2. Cooper, D.N., and Barondes, S.H. (1999). God must love galectins; he made so many of them. Glycobiology *9*, 979-984.

3. Agrwal, N., Sun, Q., Wang, S.Y., and Wang, J.L. (1993). Carbohydrate-binding protein 35. I. Properties of the recombinant polypeptide and the individuality of the domains. J Biol Chem *268*, 14932-14939.

4. Birdsall, B., Feeney, J., Burdett, I.D., Bawumia, S., Barboni, E.A., and Hughes, R.C. (2001). NMR solution studies of hamster galectin-3 and electron microscopic visualization of surface-adsorbed complexes: evidence for interactions between the N- and C-terminal domains. Biochemistry *40*, 4859-4866.

5. Leonidas, D.D., Vatzaki, E.H., Vorum, H., Celis, J.E., Madsen, P., and Acharya, K.R. (1998). Structural basis for the recognition of carbohydrates by human galectin-7. Biochemistry *37*, 13930-13940.

6. Nagae, M., Nishi, N., Murata, T., Usui, T., Nakamura, T., Wakatsuki, S., and Kato, R. (2006). Crystal structure of the galectin-9 N-terminal carbohydrate recognition domain from Mus musculus reveals the basic mechanism of carbohydrate recognition. J Biol Chem *281*, 35884-35893.

7. Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S.H., and Rini, J.M. (1998). X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-A resolution. J Biol Chem *273*, 13047-13052.

8. Leffler, H., and Barondes, S.H. (1986). Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian beta-galactosides. J Biol Chem *261*, 10119-10126.

9. Wang, J.L., Gray, R.M., Haudek, K.C., and Patterson, R.J. (2004). Nucleocytoplasmic lectins. Biochim Biophys Acta *1673*, 75-93.

10. Liu, F.T., Patterson, R.J., and Wang, J.L. (2002). Intracellular functions of galectins. Biochim Biophys Acta *1572*, 263-273.

11. Menon, R.P., and Hughes, R.C. (1999). Determinants in the N-terminal domains of galectin-3 for secretion by a novel pathway circumventing the endoplasmic reticulum-Golgi complex. Eur J Biochem *264*, 569-576.

12. Cho, M., and Cummings, R.D. (1995). Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. II. Localization and biosynthesis. J Biol Chem *270*, 5207-5212.

13. Cho, M., and Cummings, R.D. (1995). Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. I. Physical and chemical characterization. J Biol Chem *270*, 5198-5206.

14. Savin, S.B., Cvejic, D.S., and Jankovic, M.M. (2003). Expression of galectin-1 and galectin-3 in human fetal thyroid gland. J Histochem Cytochem *51*, 479-483.

15. Shimonishi, T., Miyazaki, K., Kono, N., Sabit, H., Tuneyama, K., Harada, K., Hirabayashi, J., Kasai, K., and Nakanuma, Y. (2001). Expression of endogenous galectin-1 and galectin-3 in intrahepatic cholangiocarcinoma. Hum Pathol *32*, 302-310.

16. Cooper, D.N., and Barondes, S.H. (1990). Evidence for export of a muscle lectin from cytosol to extracellular matrix and for a novel secretory mechanism. J Cell Biol *110*, 1681-1691.

17. Paz, A., Haklai, R., Elad-Sfadia, G., Ballan, E., and Kloog, Y. (2001). Galectin-1 binds oncogenic H-Ras to mediate Ras membrane anchorage and cell transformation. Oncogene *20*, 7486-7493.

18. Elad-Sfadia, G., Haklai, R., Ballan, E., Gabius, H.J., and Kloog, Y. (2002). Galectin-1 augments Ras activation and diverts Ras signals to Raf-1 at the expense of phosphoinositide 3-kinase. J Biol Chem *277*, 37169-37175.

19. Rotblat, B., Niv, H., Andre, S., Kaltner, H., Gabius, H.J., and Kloog, Y. (2004). Galectin-1(L11A) predicted from a computed galectin-1 farnesyl-binding pocket selectively inhibits Ras-GTP. Cancer Res *64*, 3112-3118.

20. Ashery, U., Yizhar, O., Rotblat, B., Elad-Sfadia, G., Barkan, B., Haklai, R., and Kloog, Y. (2006). Spatiotemporal organization of Ras signaling: rasosomes and the galectin switch. Cell Mol Neurobiol *26*, 471-495.

21. Choi, J.Y., van Wijnen, A.J., Aslam, F., Leszyk, J.D., Stein, J.L., Stein, G.S., Lian, J.B., and Penman, S. (1998). Developmental association of the beta-galactoside-binding protein galectin-1 with the nuclear matrix of rat calvarial osteoblasts. J Cell Sci *111 (Pt 20)*, 3035-3043.

22. Vyakarnam, A., Lenneman, A.J., Lakkides, K.M., Patterson, R.J., and Wang, J.L. (1998). A comparative nuclear localization study of galectin-1 with other splicing components. Exp Cell Res *242*, 419-428.

23. Park, J.W., Voss, P.G., Grabski, S., Wang, J.L., and Patterson, R.J. (2001). Association of galectin-1 and galectin-3 with Gemin4 in complexes containing the SMN protein. Nucleic Acids Res *29*, 3595-3602.

24. Vyakarnam, A., Dagher, S.F., Wang, J.L., and Patterson, R.J. (1997). Evidence for a role for galectin-1 in pre-mRNA splicing. Mol Cell Biol *17*, 4730-4737.

25. Dagher, S.F., Wang, J.L., and Patterson, R.J. (1995). Identification of galectin-3 as a factor in pre-mRNA splicing. Proc Natl Acad Sci U S A *92*, 1213-1217.

26. Hubert, M., Wang, S.Y., Wang, J.L., Seve, A.P., and Hubert, J. (1995). Intranuclear distribution of galectin-3 in mouse 3T3 fibroblasts: comparative analyses by immunofluorescence and immunoelectron microscopy. Exp Cell Res *220*, 397-406.

27. Moutsatsos, I.K., Davis, J.M., and Wang, J.L. (1986). Endogenous lectins from cultured cells: subcellular localization of carbohydrate-binding protein 35 in 3T3 fibroblasts. J Cell Biol *102*, 477-483.

28. Weinberger, P.M., Adam, B.L., Gourin, C.G., Moretz, W.H., 3rd, Bollag, R.J., Wang, B.Y., Liu, Z., Lee, J.R., and Terris, D.J. (2007). Association of nuclear, cytoplasmic expression of galectin-3 with beta-catenin/Wnt-pathway activation in thyroid carcinoma. Arch Otolaryngol Head Neck Surg *133*, 503-510.

29. Moutsatsos, I.K., Wade, M., Schindler, M., and Wang, J.L. (1987). Endogenous lectins from cultured cells: nuclear localization of carbohydrate-binding protein 35 in proliferating 3T3 fibroblasts. Proc Natl Acad Sci U S A *84*, 6452-6456.

30. Cowles, E.A., Agrwal, N., Anderson, R.L., and Wang, J.L. (1990). Carbohydrate-binding protein 35. Isoelectric points of the polypeptide and a phosphorylated derivative. J Biol Chem *265*, 17706-17712.

31. Tsay, Y.G., Lin, N.Y., Voss, P.G., Patterson, R.J., and Wang, J.L. (1999). Export of galectin-3 from nuclei of digitonin-permeabilized mouse 3T3 fibroblasts. Exp Cell Res *252*, 250-261.

32. Davidson, P.J., Davis, M.J., Patterson, R.J., Ripoche, M.A., Poirier, F., and Wang, J.L. (2002). Shuttling of galectin-3 between the nucleus and cytoplasm. Glycobiology *12*, 329-337.

33. Davidson, P.J., Li, S.Y., Lohse, A.G., Vandergaast, R., Verde, E., Pearson, A., Patterson, R.J., Wang, J.L., and Arnoys, E.J. (2006). Transport of galectin-3 between the nucleus and cytoplasm. I. Conditions and signals for nuclear import. Glycobiology *16*, 602-611.

34. Li, S.Y., Davidson, P.J., Lin, N.Y., Patterson, R.J., Wang, J.L., and Arnoys, E.J. (2006). Transport of galectin-3 between the nucleus and cytoplasm. II. Identification of the signal for nuclear export. Glycobiology *16*, 612-622.

35. Nakahara, S., Hogan, V., Inohara, H., and Raz, A. (2006). Importin-mediated nuclear translocation of galectin-3. J Biol Chem *281*, 39649-39659.

36. Akahani, S., Nangia-Makker, P., Inohara, H., Kim, H.R., and Raz, A. (1997). Galectin-3: a novel antiapoptotic molecule with a functional BH1 (NWGR) domain of Bcl-2 family. Cancer Res *57*, 5272-5276.

37. Yang, R.Y., Hsu, D.K., and Liu, F.T. (1996). Expression of galectin-3 modulates T-cell growth and apoptosis. Proc Natl Acad Sci U S A *93*, 6737-6742.

38. Menon, R.P., Strom, M., and Hughes, R.C. (2000). Interaction of a novel cysteine and histidine-rich cytoplasmic protein with galectin-3 in a carbohydrate-independent manner. FEBS Lett *470*, 227-231.

39. Bawumia, S., Barboni, E.A., Menon, R.P., and Hughes, R.C. (2003). Specificity of interactions of galectin-3 with Chrp, a cysteine- and histidine-rich cytoplasmic protein. Biochimie *85*, 189-194.

40. Paron, I., Scaloni, A., Pines, A., Bachi, A., Liu, F.T., Puppin, C., Pandolfi, M., Ledda, L., Di Loreto, C., Damante, G., and Tell, G. (2003). Nuclear localization of Galectin-3 in transformed thyroid cells: a role in transcriptional regulation. Biochem Biophys Res Commun *302*, 545-553.

41. Seve, A.P., Felin, M., Doyennette-Moyne, M.A., Sahraoui, T., Aubery, M., and Hubert, J. (1993). Evidence for a lactose-mediated association between two nuclear carbohydrate-binding proteins. Glycobiology *3*, 23-30.

42. Seve, A.P., Hadj-Sahraoui, Y., Felin, M., Doyennette-Moyne, M.A., Aubery, M., and Hubert, J. (1994). Evidence that lactose binding to CBP35 disrupts its interaction with CBP70 in isolated HL60 cell nuclei. Exp Cell Res *213*, 191-197.

43. Shimura, T., Takenaka, Y., Tsutsumi, S., Hogan, V., Kikuchi, A., and Raz, A. (2004). Galectin-3, a novel binding partner of beta-catenin. Cancer Res *64*, 6363-6367.

44. Goletz, S., Hanisch, F.G., and Karsten, U. (1997). Novel alphaGalNAc containing glycans on cytokeratins are recognized invitro by galectins with type II carbohydrate recognition domains. J Cell Sci *110 (Pt 14)*, 1585-1596.

45. Wang, L., Inohara, H., Pienta, K.J., and Raz, A. (1995). Galectin-3 is a nuclear matrix protein which binds RNA. Biochem Biophys Res Commun *217*, 292-303.

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

47. Hinterberger, M., Pettersson, I., and Steitz, J.A. (1983). Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs. J Biol Chem *258*, 2604-2613.

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

49. Kiss, T. (2004). Biogenesis of small nuclear RNPs. J Cell Sci *117*, 5949-5951.

50. Medlin, J.E., Uguen, P., Taylor, A., Bentley, D.L., and Murphy, S. (2003). The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. Embo J *22*, 925-934.

51. Segref, A., Mattaj, I.W., and Ohno, M. (2001). The evolutionarily conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export. Rna *7*, 351-360.

52. Ohno, M., Segref, A., Bachi, A., Wilm, M., and Mattaj, I.W. (2000). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell *101*, 187-198.

53. Yong, J., Golembe, T.J., Battle, D.J., Pellizzoni, L., and Dreyfuss, G. (2004). snRNAs contain specific SMN-binding domains that are essential for snRNP assembly. Mol Cell Biol *24*, 2747-2756.

54. Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002). Essential role for the SMN complex in the specificity of snRNP assembly. Science *298*, 1775-1779.

55. Urlaub, H., Raker, V.A., Kostka, S., and Luhrmann, R. (2001). Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. Embo J *20*, 187-196.

56. Raker, V.A., Plessel, G., and Luhrmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. Embo J *15*, 2256-2269.

57. 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*, 375-387.

58. Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L., and Luhrmann, R. (2000). The C-terminal RG dipeptide repeats of the spliceosomal

Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. J Biol Chem *275*, 17122-17129.

59. Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., and Luhrmann, R. (2001). Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. Rna *7*, 1531-1542.

60. Cote, J., and Richard, S. (2005). Tudor domains bind symmetrical dimethylated arginines. J Biol Chem *280*, 28476-28483.

61. Battle, D.J., Lau, C.K., Wan, L., Deng, H., Lotti, F., and Dreyfuss, G. (2006). The Gemin5 protein of the SMN complex identifies snRNAs. Mol Cell *23*, 273-279.

62. Golembe, T.J., Yong, J., and Dreyfuss, G. (2005). Specific sequence features, recognized by the SMN complex, identify snRNAs and determine their fate as snRNPs. Mol Cell Biol *25*, 10989-11004.

63. Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A.G., Bertrand, E., Tazi, J., and Bordonne, R. (2003). Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. EMBO Rep *4*, 616-622.

64. Hausmann, S., and Shuman, S. (2005). Specificity and mechanism of RNA cap guanine-N2 methyltransferase (Tgs1). J Biol Chem *280*, 4021-4024.

65. Rollenhagen, C., and Pante, N. (2006). Nuclear import of spliceosomal snRNPs. Can J Physiol Pharmacol *84*, 367-376.

66. Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M., and Luhrmann, R. (1998). Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. Embo J *17*, 4114-4126.

67. Rollenhagen, C., Muhlhausser, P., Kutay, U., and Pante, N. (2003). Importin beta-depending nuclear import pathways: role of the adapter proteins in the docking and releasing steps. Mol Biol Cell *14*, 2104-2115.

68. Huber, J., Dickmanns, A., and Luhrmann, R. (2002). The importin-beta binding domain of snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs in vitro. J Cell Biol *156*, 467-479.

69. Narayanan, U., Achsel, T., Luhrmann, R., and Matera, A.G. (2004). Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. Mol Cell *16*, 223-234.

70. Narayanan, U., Ospina, J.K., Frey, M.R., Hebert, M.D., and Matera, A.G. (2002). SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. Hum Mol Genet *11*, 1785-1795.

71. Marshallsay, C., and Luhrmann, R. (1994). In vitro nuclear import of snRNPs: cytosolic factors mediate m3G-cap dependence of U1 and U2 snRNP transport. Embo J *13*, 222-231.

72. Xu, H., Pillai, R.S., Azzouz, T.N., Shpargel, K.B., Kambach, C., Hebert, M.D., Schumperli, D., and Matera, A.G. (2005). The C-terminal domain of coilin interacts with Sm proteins and U snRNPs. Chromosoma *114*, 155-166.

73. Stanek, D., and Neugebauer, K.M. (2006). The Cajal body: a meeting place for spliceosomal snRNPs in the nuclear maze. Chromosoma *115*, 343-354.

74. Jady, B.E., Darzacq, X., Tucker, K.E., Matera, A.G., Bertrand, E., and Kiss, T. (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. Embo J *22*, 1878-1888.

75. Jady, B.E., and Kiss, T. (2001). A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. Embo J *20*, 541-551.

76. Yu, Y.T., Shu, M.D., and Steitz, J.A. (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. Embo J *17*, 5783-5795.

77. Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. Nat Rev Mol Cell Biol *4*, 605-612.

78. Xue, D., Rubinson, D.A., Pannone, B.K., Yoo, C.J., and Wolin, S.L. (2000). U snRNP assembly in yeast involves the La protein. Embo J *19*, 1650-1660.

79. Wolin, S.L., and Cedervall, T. (2002). The La protein. Annu Rev Biochem *71*, 375-403.

80. 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.

81. Karaduman, R., Fabrizio, P., Hartmuth, K., Urlaub, H., and Luhrmann, R. (2006). RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. J Mol Biol *356*, 1248-1262.

82. 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*, 5789-5802.

83. Verdone, L., Galardi, S., Page, D., and Beggs, J.D. (2004). Lsm proteins promote regeneration of pre-mRNA splicing activity. Curr Biol *14*, 1487-1491.

84. Lange, T.S., and Gerbi, S.A. (2000). Transient nucleolar localization Of U6 small nuclear RNA in Xenopus Laevis oocytes. Mol Biol Cell *11*, 2419-2428.

85. Ganot, P., Jady, B.E., Bortolin, M.L., Darzacq, X., and Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. Mol Cell Biol *19*, 6906-6917.

86. Spiller, M.P., Boon, K.L., Reijns, M.A., and Beggs, J.D. (2007). The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. Nucleic Acids Res *35*, 923-929.

87. Hashimoto, C., and Steitz, J.A. (1984). U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. Nucleic Acids Res *12*, 3283-3293.

88. Black, D.L., and Steitz, J.A. (1986). Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. Cell *46*, 697-704.

89. Gozani, O., Patton, J.G., and Reed, R. (1994). A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. Embo J *13*, 3356-3367.

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

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

92. Stanek, D., and Neugebauer, K.M. (2004). Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. J Cell Biol *166*, 1015-1025.

93. Stanek, D., Rader, S.D., Klingauf, M., and Neugebauer, K.M. (2003). Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. J Cell Biol *160*, 505-516.

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

95. 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*, 344-355.

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

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

98. Liu, S., Rauhut, R., Vornlocher, H.P., and Luhrmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. Rna *12*, 1418-1430.

99. Hall, K.B., and Konarska, M.M. (1992). The 5' splice site consensus RNA oligonucleotide induces assembly of U2/U4/U5/U6 small nuclear ribonucleoprotein complexes. Proc Natl Acad Sci U S A *89*, 10969-10973.

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

101. Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G., and Melki, J. (1997). Correlation between severity and SMN protein level in spinal muscular atrophy. Nat Genet *16*, 265-269.

102. Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., and et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80*, 155-165.

103. Liu, Q., and Dreyfuss, G. (1996). A novel nuclear structure containing the survival of motor neurons protein. Embo J *15*, 3555-3565.

104. Carissimi, C., Saieva, L., Baccon, J., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J., and Pellizzoni, L. (2006). Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly. J Biol Chem *281*, 8126-8134.

105. Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. Cell *90*, 1013-1021.

106. Mourelatos, Z., Abel, L., Yong, J., Kataoka, N., and Dreyfuss, G. (2001). SMN interacts with a novel family of hnRNP and spliceosomal proteins. Embo J *20*, 5443-5452.

107. Grimmler, M., Otter, S., Peter, C., Muller, F., Chari, A., and Fischer, U. (2005). Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization. Hum Mol Genet *14*, 3099-3111.

108. Meister, G., Buhler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F., and Fischer, U. (2000). Characterization of a nuclear 20S complex containing the survival of motor neurons (SMN) protein and a specific subset of spliceosomal Sm proteins. Hum Mol Genet *9*, 1977-1986.

109. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev *16*, 720-728.

110. Peng, R., Hawkins, I., Link, A.J., and Patton, J.G. (2006). The splicing factor PSF is part of a large complex that assembles in the absence of pre-mRNA and contains all five snRNPs. RNA Biol *3*, 69-76.

111. Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998). A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell *95*, 615-624.

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

113. Boisvert, F.M., Cote, J., Boulanger, M.C., Cleroux, P., Bachand, F., Autexier, C., and Richard, S. (2002). Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing. J Cell Biol *159*, 957-969.

114. Terns, M.P., and Terns, R.M. (2001). Macromolecular complexes: SMN--the master assembler. Curr Biol *11*, R862-864.

115. Friesen, W.J., and Dreyfuss, G. (2000). Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN). J Biol Chem *275*, 26370-26375.

116. Jones, K.W., Gorzynski, K., Hales, C.M., Fischer, U., Badbanchi, F., Terns, R.M., and Terns, M.P. (2001). Direct interaction of the spinal muscular atrophy disease

protein SMN with the small nucleolar RNA-associated protein fibrillarin. J Biol Chem *276*, 38645-38651.

117. Pellizzoni, L., Charroux, B., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001). A functional interaction between the survival motor neuron complex and RNA polymerase II. J Cell Biol *152*, 75-85.

118. Pillai, R.S., Grimmler, M., Meister, G., Will, C.L., Luhrmann, R., Fischer, U., and Schumperli, D. (2003). Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev *17*, 2321-2333.

119. Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell *9*, 31-44.

120. Tarn, W.Y., Lee, K.R., and Cheng, S.C. (1993). Yeast precursor mRNA processing protein PRP19 associates with the spliceosome concomitant with or just after dissociation of U4 small nuclear RNA. Proc Natl Acad Sci U S A *90*, 10821-10825.

121. Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p-associated complex in spliceosome activation. Science *302*, 279-282.

122. Bottner, C.A., Schmidt, H., Vogel, S., Michele, M., and Kaufer, N.F. (2005). Multiple genetic and biochemical interactions of Brr2, Prp8, Prp31, Prp1 and Prp4 kinase suggest a function in the control of the activation of spliceosomes in Schizosaccharomyces pombe. Curr Genet *48*, 151-161.

123. Malca, H., Shomron, N., and Ast, G. (2003). The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. Mol Cell Biol *23*, 3442-3455.

124. Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat Struct Mol Biol *13*, 815-822.

125. Grabowski, P.J., Seiler, S.R., and Sharp, P.A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell *42*, 345-353.

126. Abmayr, S.M., Reed, R., and Maniatis, T. (1988). Identification of a functional mammalian spliceosome containing unspliced pre-mRNA. Proc Natl Acad Sci U S A *85*, 7216-7220.

127. Reed, R., and Maniatis, T. (1988). The role of the mammalian branchpoint sequence in pre-mRNA splicing. Genes Dev *2*, 1268-1276.

128. Rosbash, M., and Seraphin, B. (1991). Who's on first? The U1 snRNP-5' splice site interaction and splicing. Trends Biochem Sci *16*, 187-190.

129. Ruskin, B., and Green, M.R. (1985). Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. Nature *317*, 732-734.

130. Maniatis, T., and Reed, R. (1987). The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature *325*, 673-678.

131. Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245.

132. Azubel, M., Habib, N., Sperling, R., and Sperling, J. (2006). Native spliceosomes assemble with pre-mRNA to form supraspliceosomes. J Mol Biol *356*, 955-966.

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

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

135. Valadkhan, S., and Manley, J.L. (2003). Characterization of the catalytic activity of U2 and U6 snRNAs. Rna *9*, 892-904.

136. Shen, H., and Green, M.R. (2006). RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. Genes Dev *20*, 1755-1765.

137. Shen, H., Kan, J.L., and Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell *13*, 367-376.

138. Bindereif, A., and Green, M.R. (1987). An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. Embo J *6*, 2415-2424.

139. Das, R., and Reed, R. (1999). Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. Rna *5*, 1504-1508.

140. Ruby, S.W., and Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science *242*, 1028-1035.

141. Puig, O., Gottschalk, A., Fabrizio, P., and Seraphin, B. (1999). Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. Genes Dev *13*, 569-580.

142. Stutz, F., Liao, X.C., and Rosbash, M. (1993). U1 small nuclear ribonucleoprotein particle-protein interactions are revealed in Saccharomyces cerevisiae by in vivo competition assays. Mol Cell Biol *13*, 2126-2133.

143. Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol Cell *5*, 779-787.

144. Hong, W., Bennett, M., Xiao, Y., Feld Kramer, R., Wang, C., and Reed, R. (1997). Association of U2 snRNP with the spliceosomal complex E. Nucleic Acids Res *25*, 354-361.

145. Kent, O.A., Ritchie, D.B., and Macmillan, A.M. (2005). Characterization of a U2AF-independent commitment complex (E') in the mammalian spliceosome assembly pathway. Mol Cell Biol *25*, 233-240.

146. Newnham, C.M., and Query, C.C. (2001). The ATP requirement for U2 snRNP addition is linked to the pre-mRNA region 5' to the branch site. Rna *7*, 1298-1309.

147. Sashital, D.G., Venditti, V., Angers, C.G., Cornilescu, G., and Butcher, S.E. (2007). Structure and thermodynamics of a conserved U2 snRNA domain from yeast and human. Rna *13*, 328-338.

148. Fleckner, J., Zhang, M., Valcarcel, J., and Green, M.R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes Dev *11*, 1864-1872.

149. Zhang, M., and Green, M.R. (2001). Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. Genes Dev *15*, 30-35.

150. Xu, Y.Z., Newnham, C.M., Kameoka, S., Huang, T., Konarska, M.M., and Query, C.C. (2004). Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. Embo J *23*, 376-385.

151. Dybkov, O., Will, C.L., Deckert, J., Behzadnia, N., Hartmuth, K., and Luhrmann, R. (2006). U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. Mol Cell Biol *26*, 2803-2816.

152. Donmez, G., Hartmuth, K., Kastner, B., Will, C.L., and Luhrmann, R. (2007). The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. Mol Cell *25*, 399-411.

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

154. Alvi, R.K., Lund, M., and Okeefe, R.T. (2001). ATP-dependent interaction of yeast U5 snRNA loop 1 with the 5' splice site. Rna *7*, 1013-1023.

155. Newman, A.J., Teigelkamp, S., and Beggs, J.D. (1995). snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. Rna *1*, 968-980.

156. 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.

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

158. Chen, C.H., Yu, W.C., Tsao, T.Y., Wang, L.Y., Chen, H.R., Lin, J.Y., Tsai, W.Y., and Cheng, S.C. (2002). Functional and physical interactions between components of the Prp19p-associated complex. Nucleic Acids Res *30*, 1029-1037.

159. Sashital, D.G., Cornilescu, G., McManus, C.J., Brow, D.A., and Butcher, S.E. (2004). U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. Nat Struct Mol Biol *11*, 1237-1242.

160. Butcher, S.E., and Brow, D.A. (2005). Towards understanding the catalytic core structure of the spliceosome. Biochem Soc Trans *33*, 447-449.

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

162. Umen, J.G., and Guthrie, C. (1995). The second catalytic step of pre-mRNA splicing. Rna *1*, 869-885.

163. Datta, B., and Weiner, A.M. (1993). The phylogenetically invariant ACAGAGA and AGC sequences of U6 small nuclear RNA are more tolerant of mutation in human cells than in Saccharomyces cerevisiae. Mol Cell Biol *13*, 5377-5382.

164. Sontheimer, E.J., and Steitz, J.A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science *262*, 1989-1996.

165. 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*, 565-574.

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

167. Nilsen, T.W. (2002). The spliceosome: no assembly required? Mol Cell *9*, 8-9.

168. Tardiff, D.F., and Rosbash, M. (2006). Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. Rna *12*, 968-979.

169. Behzadnia, N., Hartmuth, K., Will, C.L., and Luhrmann, R. (2006). Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. Rna *12*, 1738-1746.

170. Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182-185.

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

172. Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. Rna *8*, 426-439.

173. Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol Cell Biol *26*, 5528-5543.

174. Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. Embo J *26*, 1737-1748.

175. Chen, Y.I., Moore, R.E., Ge, H.Y., Young, M.K., Lee, T.D., and Stevens, S.W. (2007). Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. Nucleic Acids Res.

176. Chen, Y.I., Maika, S.D., and Stevens, S.W. (2006). Epitope tagging of proteins at the native chromosomal loci of genes in mice and in cultured vertebrate cells. J Mol Biol *361*, 412-419.

177. Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. Genes Dev *20*, 1100-1109.

178. Bird, G., Zorio, D.A., and Bentley, D.L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Mol Cell Biol *24*, 8963-8969.

179. Laing, J.G., and Wang, J.L. (1988). Identification of carbohydrate binding protein 35 in heterogeneous nuclear ribonucleoprotein complex. Biochemistry *27*, 5329-5334.

180. Wang, W., Park, J.W., Wang, J.L., and Patterson, R.J. (2006). Immunoprecipitation of spliceosomal RNAs by antisera to galectin-1 and galectin-3. Nucleic Acids Res *34*, 5166-5174.

Chapter 2

Physico-chemical characterization of galectin-3-snRNP complexes and a role in pre-mRNA binding

ABSTRACT

Previously, we showed that galectin-1 and galectin-3 are redundant pre-mRNA splicing factors associated with the spliceosome throughout the splicing pathway. Here we present evidence for the association of galectin-3 with snRNPs outside of the spliceosome. Immunoprecipitation of HeLa nuclear extract using buffers optimal for *in vitro* pre-mRNA splicing with anti-galectin-3 resulted in the coprecipitation of the 5 snRNAs needed for pre-mRNA splicing, snRNP proteins and other RNA processing factors including the SMN protein and PSF. Galectin-3 also co-sedimented with snRNP complexes when nuclear extract was fractionated on glycerol gradients. This co-sedimentation represents *bona fide* galectin-3-snRNP complexes as immunoprecipitation of gradient fractions with anti-galectin-3 yielded both snRNA and associated proteins. In particular, one fraction at approximately 10S showed an association of galectin-3 with U1 snRNP that was sensitive to treatment with ribonuclease A. We tested the ability of this U1 particle to recognize an exogenous pre-mRNA substrate. We found this isolated galectin-3-U1 snRNP particle was sufficient to load galectin-3 onto the pre-mRNA substrate under conditions that allow early splicing complexes to form. These data suggest a mechanism for the entry of galectin-3 into the pre-mRNA splicing pathway and describe a new class of polypeptides associated with spliceosomal snRNPs.

INTRODUCTION

Pre-mRNA splicing involves nearly 300 proteins and 5 snRNAs [1-4]. These components are assembled into the machinery used to perform the splicing chemistry of intron removal and exon ligation. The canonical model for the assembly process involves the stepwise addition of the snRNPs into early, commitment and active complexes. U1 snRNP assembles onto the pre-mRNA at the 5' splice site in the absence of ATP. Addition of ATP allows U2 snRNP to recognize U2AF at the branchpoint and form a stable commitment complex. Finally the U4/U6.U5 tri-snRNP particle binds at the 3' splice site resulting in the active spliceosome [5, 6]. In addition, various protein cofactors are differentially incorporated into the complexes and then disassemble once an intron is removed and exons are ligated. In this step-wise assembly model, individual U1 and U2 snRNPs recycle directly while U4, U5 and U6 must reassemble into the tri-snRNP for reutilization in subsequent rounds of splicing.

Recently another model of spliceosome assembly has been described. In this model, a large complex containing all 5 snRNPs and many splicing proteins assembles in the absence of a splicing substrate scaffold [7-9]. This penta-snRNP complex then assembles onto the splicing substrate. A series of remodeling events with addition and release of other proteins ensues to generate the catalytically active spliceosome. Following splicing chemistry and release of products, the unit either remodels to form an active penta-snRNP complex or disassembles before reforming another penta-snRNP particle.

We have shown previously that galectin-1 (Gal1) and galectin-3 (Gal3) were required and redundant splicing factors using a cell free splicing assay. The key findings were: (a) depletion of both galectins from HeLa nuclear extracts (NE) abolished splicing

activity and blocked spliceosome formation at an early complex; (b) both splicing activity and spliceosome formation were restored by addition of recombinant Gal1 or Gal3; (c) each galectin was a component of early and active splicing complexes as determined by co-immunoprecipitation of splicing substrate at early times and all mRNA species as they appeared in active complexes by galectin-specific antisera; (d) each galectin was incorporated into spliceosomes in a mutually exclusive manner and (e) neither galectin bound directly to the pre-mRNA substrate [10-12].

How are galectins assembled into early splicing complexes? In this manuscript we report that in the absence of splicing substrate, Gal3 is associated with several snRNP particles and, in particular, the U1 snRNP under conditions of the *in vitro* splicing assay. We present evidence that one mechanism of Gal3 incorporation into the splicing pathway is mediated by the U1 snRNP particle.

MATERIALS AND METHODS

Antibodies

Polyclonal rabbit antibodies directed against Gal3 [13] have been previously described. These antibodies were covalently cross-linked to protein A-Sepharose CL-4B beads (Amersham Biosciences) as previously described [11] using a 2:1 ratio of antiserum to protein-A beads. A mouse monoclonal antibody against trimethylguanosine (TMG; K121) was purchased as an agarose bead conjugate (Calbiochem). These antibodies were used for immunoprecipitation experiments.

For immunoblotting, polyclonal rabbit antibodies against Gal1 have been described previously [12]. The following antibodies were purchased for immunoblotting: Affinity purified rabbit anti-PSF (PTB-associated splicing factor), goat anti-Slu7 (Santa

Cruz Biotechnology); rabbit anti-TFII-I, rabbit anti-Gemin4 (Bethyl Laboratories); mouse monoclonals anti-SMN (survival of motor neurons protein) and anti-Ran (BD Transduction Laboratories); mouse monoclonal anti-U1-70K (Synaptic Systems); and human autoimmune sera ENA anti-Sm (The Binding Site). Rabbit anti-RAP30 was a kind gift from Dr. Zach Burton (Michigan State University). Rabbit anti-Gal3, rat monoclonal anti-Mac2 [14], or mouse monoclonal NCLGAL3 (Vector Laboratories) were used for blotting Gal3 protein. Primary mouse monoclonal antibodies were detected by goat anti-mouse IgG light chain specific-HRP (horseradish peroxidase) conjugates (Jackson ImmunoResearch Laboratories). When blotting the bound fractions of immunoprecipitation experiments, primary polyclonal rabbit blotting antibodies were probed with secondary monoclonal mouse anti-rabbit IgG light chain specific-HRP conjugates (Jackson ImmunoResearch Laboratories). All other secondary antibody-HRP conjugates (Pierce Biotechnology) were directed against both the heavy and light chains of the primary blotting antibodies.

Polyacrylamide gel electrophoresis and western blotting

For protein analysis, samples were subjected to 10% or 12% SDS-PAGE as described by Laemmli [15]. Proteins were electrophoretically transferred from the gel onto Hybond nitrocellulose membrane (Amersham Biosciences) in transfer buffer (25 mM Tris, 193 mM glycine and 20% methanol, pH 8.3). Following transfer, membranes were blocked overnight in 10% nonfat dry milk in Tris-buffered saline containing Tween-20 (10 mM Tris, pH 7.5, 0.5 M NaCl, 0.05% Tween 20, T-TBS). Primary antibodies for immunoblotting were diluted in 1% milk-T-TBS and incubated on the membrane for 1 hour at room temperature. After washing four times, 15 minutes each, in T-TBS, the

appropriate secondary antibody conjugated to HRP was added in 1% milk-T-TBS for 1 hour. Following four T-TBS washes as above, proteins were visualized using the Western Lightning Chemiluminescence System (Perkin Elmer Life Sciences).

For RNA samples, the RNA was extracted as described below and precipitated with 3 volumes of ethanol at -80°C. The precipitated RNA was dissolved in 10 µl of sample buffer (9:1/formamide:bromophenol blue) and subjected to electrophoresis through 13% polyacrylamide (bisacrylamide–acrylamide 1.9:50 [wt/wt])-8.3 M urea gels, run in 1X TBE (90 mM Tris base, 90 mM boric acid and 2.5 mM EDTA, pH 8.0). The radioactive RNA species were revealed by autoradiography. For northern analysis the RNA was transferred via wicking in 20X SSC (3 M NaCl, 0.3 M sodium citrate) overnight onto a nylon membrane (Hybond-N, Amersham Biosciences) and cross-linked by exposure to UV light (1200 µJoules).

Immunoprecipitation

Nuclear extract was prepared as described by Dignam et al. [16] from HeLa S3 cells obtained from the National Cell Culture Center (Minneapolis, MN) and was dialyzed into buffer D (20 mM HEPES, pH7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol). HeLa NE (30 µl) was incubated with 3 mM $MgCl_2$, 0.5 mM ATP and 20 mM creatine phosphate in a 50 µl reaction at 30° C for 30 minutes before use in immunoprecipitations [17]. After incubation, the reaction was diluted in 0.5 ml 60% buffer D (D60) with 0.05% Triton-X 100 (TX) and incubated with 15 - 20 µl of antibody coupled protein-A beads for 1 hour at 4°C with rocking. The beads were then washed with 1 ml D60+0.05% TX three times and eluted twice with 20 µl SDS-PAGE sample buffer. The elution was divided into two aliquots for analysis of proteins and RNA.

RNA was extracted by treating the sample with proteinase K (4 mg/ml final concentration) at 37°C for 20 minutes and diluting to 100 µl with 125 mM Tris (pH 8), 1 mM EDTA, 300 mM sodium acetate. RNA was extracted by mixing with 200 µl of phenol–chloroform (50:50 [vol/vol]), followed by 100 µl of chloroform.

For immunoprecipitation with anti-TMG, 20 µl antibody beads were mixed with 40 µl of pre-incubated NE for 1 hour at 4°C with rocking. Beads were washed 3 times with 0.5 ml D60+0.05% TX and eluted with 20 µl of 2X SDS sample buffer. Protein-G agarose beads without antibody were used as a negative control. Protein analysis was carried out with 15 µl of the eluted material as described above. For the sequential immunoprecipitation, 25 µl of the coupled anti-TMG beads were incubated with 50 µl of pre-incubated NE and treated as above. The beads were eluted twice with 15 µl 5 mM cap structure analog (m^7G(5')ppp(5')G; New England Biolabs) in D60. The elutions were combined and mixed with 20 µl of anti-Gal3 coupled beads and immunoprecipitation was carried out as above. The bound material was eluted with 40 µl 2X SDS sample buffer and proteins were analyzed as described above.

Northern analysis

The following oligonucleotides were synthesized at Research Technology Support Facility at Michigan State University: U1: TCCCCTGCCAGGTAAGTATC, U2: TTAGCCAAAAGGCCGAGAAGCGAT, U4: GGGGTATTGGGAAAAGTTTTC, U5: GATTTATGCGATCTGAAGAGAAACC, U6: TTCTCTGTATCGTTCCAAT, 5S rRNA: TTCAGGGTGGTATGGCCGTAGAC. The oligonucleotide probes were labeled with $^{32}PO_4$ using T4 Polynucleotide Kinase (Invitrogen). The cross-linked nylon membrane was pre-hybridized in 20 ml of hybridization solution containing 20%

deionized formamide, 3X SSC, 5X Denhardt's, 50 mM Na_2HPO_4/NaH_2PO_4, pH 6.8, 0.1% SDS and 0.1 mg/ml heat denatured herring sperm DNA for about 2 hours at room temperature. After pre-hybridization, the appropriate ^{32}P-labeled probes (~2 pmol) were added to the hybridization solution and the membrane was hybridized overnight (~16 hours) at room temperature. The membrane was then washed once with 30 ml 2X SSC + 0.1% SDS at room temperature for 20 minutes. Northern hybridization was quantitated by phosphorimage analysis (Molecular Dynamics).

Glycerol gradient

For glycerol gradient sedimentation, a 250 µl reaction containing 150 µl HeLa NE, 3 mM $MgCl_2$, 0.5 mM ATP and 20 mM creatine phosphate was incubated for 30 minutes at 30°C and loaded onto a 5 ml 12-32% glycerol gradient in D60 (minus glycerol). The gradient was centrifuged in a Beckman SW50.1 rotor at 44,000 rpm, at 4°C for 3.5 to 4 hours. Gradient fractions were collected manually as 250 µl aliquots from the top. For RNA and protein analysis, 10% of each fraction was subjected to gel electrophoresis and analyzed as described above. Size markers separated on parallel gradients include immunoglobulin G (7S), thyroglobulin (19S) and post-nuclear supernatant for ribosomal subunits (40S and 60S).

Gradient complexes

For immunoprecipitation experiments from gradient fractions, 200 µl of the indicated fractions were pooled and the total volume brought up to 600 µl with D60. The total reaction was incubated with 20 µl antibody coupled beads at 4°C for 1 hour with rocking, washed 5 times with 0.5 ml D60+0.05% TX and eluted with 40 µl 2X SDS

sample buffer. The elutions were split into 20 µl aliquots and analyzed for protein and RNA as described above.

To test ribonuclease (RNase) sensitivity of gradient complexes, 50 µl of the pooled indicated fraction were treated with 2 µg RNase A for 30 minutes at 30°C, after which it was mixed with 15 µl anti-Gal3 coupled beads for 1 hour at 4°C with rocking in 0.5 ml D60+0.05% TX. The beads were washed 5 times with 0.5 ml binding buffer and eluted with 30 µl 2X SDS sample buffer.

To test the association of Gal3 with pre-mRNA substrate the indicated pooled fractions were subjected to digestion by micrococcal nuclease. Each reaction assembled included 45 µl of pooled fractions 3 and 4, 3 mM $MgCl_2$, 4 mM $CaCl_2$ and 4000 gel units of micrococcal nuclease (New England Biolabs). Control reactions without micrococcal nuclease were supplemented by equal volumes of D60. Reactions were incubated at 37°C for 30 minutes. Micrococcal nuclease activity was stopped by adding EGTA to a final concentration of 8 mM. MINX pre-mRNA has been described previously and the plasmid containing this construct was a gift from Dr. Sue Berget (Baylor College of Medicine, Houston, TX) [18]. The MINX pre-mRNA was labeled with [^{32}P]GTP and a monomethyl cap added during *in vitro* transcription by SP6 RNA polymerase [19]. After stopping the digestion reaction, ^{32}P-labeled MINX pre-mRNA was added to each tube, incubated for 15 minutes at 30°C and incubated with 15 µl antibody-coupled beads in 0.5 ml D60+0.05% TX for 1 hour at 4°C with rocking. The beads were washed five times with 0.5 ml binding buffer and eluted with 40 µl 2X SDS sample buffer. Bound MINX pre-mRNA was detected by subjecting an aliquot of the bound material to electrophoresis, as described above, drying the gel and phosphorimage analysis.

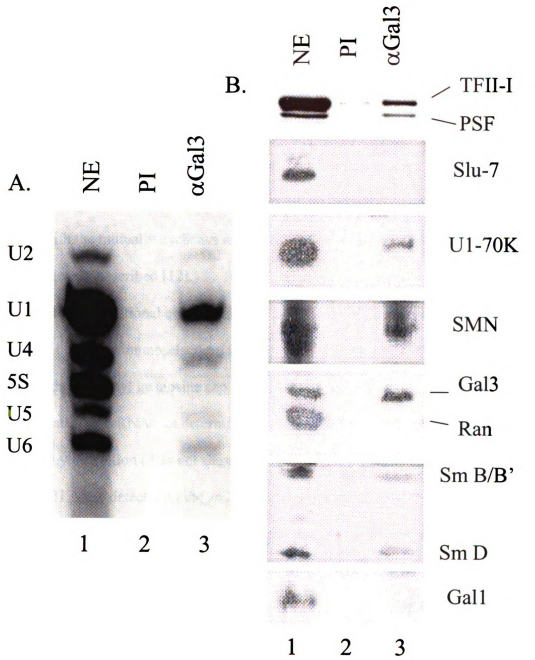
RESULTS

Galectin-3 is associated with snRNPs in the absence of splicing substrate

NE was incubated *without* splicing substrate at 30 °C with ATP for 30 minutes to disassemble endogenous splicing complexes [17]. This NE was then subjected to immunoprecipitation with anti-Gal3 serum. The bound material was analyzed for specific RNA species by northern hybridization (Figure 1A). First, anti-Gal3 co-precipitated U1, U2, U4, U5 and U6 snRNAs whereas pre-immune serum did not. Second, 5S rRNA, a prominent RNA species in NE, was not observed at all in the anti-Gal3 precipitate. Phosphorimage quantitation revealed that, at the most, 0.1% of the nuclear 5S rRNA was found in the anti-Gal3 precipitated fraction. In contrast, approximately 2-10% of each of the U-rich snRNAs was co-precipitated by anti-Gal3. Finally, since the RNA probes for the hybridization were of the same specific radioactivity, direct comparisons of the U-rich snRNA could be made between any two lanes of Figure 1A. Such a comparison indicated that the ratios of the snRNAs in the anti-Gal3 precipitate were different from the corresponding ratios in the NE subjected to precipitation. For example, the ratio of U4 to U5 snRNA in the input is 2.3, but in the ant-Gal3 bound material, this ratio changes to 4.8 (see lanes 1 and 3; Fig. 1A). All of these results suggest that the association of snRNAs with the anti-Gal3 precipitate was specific.

A similar conclusion regarding the specificity of the anti-Gal3 precipitation was obtained through analysis of protein components by western blotting (Fig. 1B). In addition to its cognate antigen, anti-Gal3 co-precipitated several key proteins that were

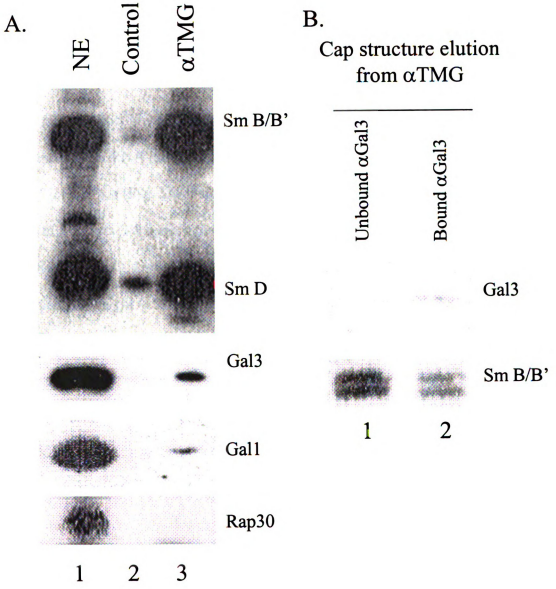
Figure 1. Analysis of nuclear RNA and proteins immunoprecipitated by anti-Gal3. NE was pre-incubated as described in Materials and Methods. The reaction was then incubated with antibody-coupled beads, either Gal3 (αGal3) or pre-immune serum (PI). Panel A shows a northern blot of 25% of the bound material probed with ^{32}P-labelled oligonucleotides complementary to the RNA species indicated on the left. Lane 1: NE representing 12% of the amount subjected to immunoprecipitation. Lane 2: RNA species bound by pre-immune serum; lane 3: RNA species bound by anti-Gal3. Panel B shows western blots of the bound material. Antibodies used to detect the proteins are indicated at right.


A.
NE
PI
αGal3
U2
U1
U4
5S
U5
U6
1
2
3
B.
NE
PI
αGal3
TFII-I
PSF
Slu-7
U1-70K
SMN
Gal3
Ran
Sm B/B'
Sm D
Gal1
1
2
3

not found in the corresponding fraction of pre-immune serum. These include: (a) the Sm core polypeptides B/B' and D of snRNPs and the 70K protein specific to U1 snRNP; (b) the splicing factor PSF [8] and the general transcription factor TFII-I, both of which have been recently identified to be in complexes containing Gal3 (P. Voss, J.L. Wang, unpublished observations); and (c) the Survival of Motor Neuron (SMN) protein. In contrast, Slu7, a second-step splicing factor, was not detected in the anti-Gal3 precipitate, ruling out the possibility that anti-Gal3 precipitated any endogenous active spliceosomes present in the NE. Similarly, the nuclear transport factor Ran was not found in the anti-Gal3 bound fraction. Finally, the anti-Gal3 precipitate did not contain Gal1, a finding consistent with the mutually exclusive association of either galectin with spliceosomes that we previously described [12].

To provide additional evidence for the association of Gal3 with nuclear snRNPs independent of Gal3 immunoselection, snRNPs were affinity purified with antiserum specific for the trimethyl guanosine cap of U1, U2, U4 and U5. This procedure precipitated all five snRNAs, as determined by ethidium bromide staining of the gel and by northern hybridization (data not shown). The core Sm B/B' and D proteins, as well as Gal3 and Gal1, were detected in the anti-TMG bound material (Fig. 2A, lane 3). In contrast, RAP30, a subunit of the general transcription factor TFII-F, was not detected in the anti-TMG precipitate (Fig. 2A, lane 3). Neither galectin could be detected in the bound material from control beads (Fig. 2A, lane 2). Lastly, the anti-TMG selected snRNPs were eluted with soluble cap analogue and then subjected to further precipitation by anti-Gal3. Panel B of Figure 2 shows that Sm B/B' proteins are co-precipitated by anti-Gal3 antibodies in the sequential immunoprecipitation protocol. Thus, these

Figure 2. Analysis of proteins immunoprecipitated by anti-TMG. NE was pre-incubated for 30 minutes at 30°C, then passed over anti-TMG-coupled beads (αTMG) or naked protein-G agarose beads (Control). Panel A: Western blots of the bound material for the proteins indicated at right; lane 1; NE represents 20% of the amount subjected to immunoprecipitation; lane 2, proteins bound by the control beads; lane 3, proteins bound by anti-TMG beads. Panel B shows the results of a double immunoselection. The pre-incubated NE was first incubated with anti-TMG beads. The αTMG bound material was eluted with cap structure analog and then incubated with anti-Gal3 beads. Material from both the unbound (lane 1) and bound (lane 2) fractions of the anti-Gal3 column was western blotted for the proteins indicated at right. Approximately 85% of both the unbound and bound material is shown in Panel B.


A.
NE
Control
αTMG
Sm B/B'
Sm D
Gal3
Gal1
Rap30
1
2
3
B.
Cap structure elution
from αTMG
Unbound αGal3
Bound αGal3
Gal3
Sm B/B'
1
2

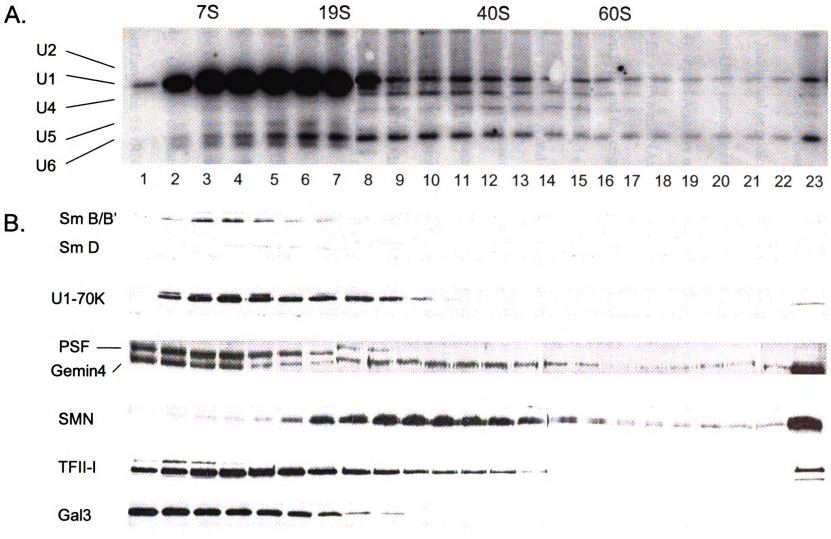
reciprocal co-immunoprecipitation experiments suggest an association of Gal3 with snRNP complexes, in the absence of any pre-mRNA scaffold.

Galectin-3 is associated with multiple snRNP complexes

Multiple snRNP complexes exist in the nucleus including the U4/U6 di-snRNP and the U4/U6.U5 tri-snRNP as well as the mono-snRNPs U1 and U2. Recently, complexes containing all five snRNPs have been described [7, 8, 20]. In light of the different ratios of snRNAs immunoprecipitated from NE with anti-Gal3, we hypothesized Gal3 is associated with complexes containing multiple snRNPs. To test this, pre-incubated NE was fractioned by glycerol gradient sedimentation and the distribution of snRNAs and several RNA processing proteins were determined. Both the northern blotting of U1 snRNA (Fig. 3A) and the western blotting of U1-70K protein (Fig. 3B) indicate that fractions 3-5 contain the mono U1 snRNP predominantly, consistent with previous findings that U1 snRNP sediments at about 10S (see Table I) [21]. Fractions at higher molecular weight (~19S and greater) contained various combinations of the snRNAs (Fig. 3A).

Several key points should be highlighted regarding the distribution profile of Gal3, our protein of interest. First, fraction 1 yielded the most intense Gal3 staining, at the top of the gradient (Fig. 3B). Second, the protein could be detected through fraction 10 (~30S). Finally, although Gal3 was not detected by direct western blotting of an aliquot of individual fractions beyond fraction 10, we did find the protein at least through fraction 18 (>60S) after enrichment by anti-Gal3 immunoprecipitation (see below). A similar observation was made for the Sm core proteins and U1-70K. The distribution of

Figure 3. Analysis of nuclear RNA and proteins separated on a 12-32% glycerol gradient. NE was treated as in Figure 1. Reactions were then loaded onto gradients and centrifuged for 3.5 hours at 44,000 rpm at 4°C. The gradients were fractionated by collecting 250 μl aliqouts from the top. Panel A shows a northern blot of 10% of each fraction using ^{32}P-labelled oligonucleotides specific for each snRNA. Size markers run in parallel are indicated at the top. Panel B shows western blots of 10% of each fraction for the proteins indicated at left.


A.
7S
19S
40S
60S
U2
U1
U4
U5
U6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
B.
Sm B/B'
Sm D
U1-70K
PSF
Gemin4
SMN
TFII-I
Gal3

the SMN protein across the gradient suggests there is little uncomplexed SMN in NE, an observation made previously [22].

Individual or pooled fractions throughout the gradient were immunoprecipitated with anti-Gal3 antibodies to determine the distribution of Gal3-containing complexes (Fig. 4). Panel A shows northern analysis of the immunoprecipitated material using probes for all five snRNAs; panel B shows the western blotting results for SMN, U1-70K, Sm B/B' and the cognate antigen Gal3. The anti-Gal3 precipitate of pooled fractions 3 and 4 yielded a single snRNA, U1, and associated proteins Sm B/B' and U1 70K (lane 3, panels A and B, Fig. 4). This suggests that at least some of the Gal3 in fractions 3 and 4 was associated with the mono U1 snRNP. On the other hand, the absence of U1 70K protein (Fig. 3B) and the miniscule amount of U1 snRNA in fraction 1, relative to the much more intense U1 bands in fractions 3 and 4 (Fig. 3A), both suggest that the majority of Gal3 is not associated with a snRNP complex. Consistent with this notion, fraction 1 yielded no snRNA or other proteins in the anti-Gal3 precipitate (lane 2, panels A and B, Fig. 4).

The anti-Gal 3 precipitate of fractions 6 and 7 (~19S; lane 4), fractions 9-11 (lane 5), fractions 13 and 14 (>40S; lane 6) and fractions 16-18 at ~60S (lane 7) yielded multiple snRNAs in various proportions along with Sm B/B' and U1 70K polypeptides (Fig. 4). Gal3, the cognate antigen, was found in all immunoprecipitated material. On the other hand, the 5S rRNA was not detected in the anti-Gal3 precipitated material (data not shown). All of these results suggest that Gal3 is associated with numerous and distinct endogenous snRNP particles in NE. In very high molecular weight fractions (fractions 10 and beyond), we also found the SMN protein co-precipitated by anti-Gal3

Table I. Listing of selected RNPs involved in pre-mRNA splicing and reported S values.

RNP	Reported S value	Reference
U1 snRNP	10S	[21]
U4/U6 di-snRNP	12S	[23]
U2 snRNP	17S	[24]
U5 snRNP	20S, 35S	[25, 26]
U4/U6.U5 tri-snRNP	25S	[25]
S. cerevisiae penta-snRNP	45S	[7]
PCC	50 – 60S	[8]
Spliceosome (various complexes)	40 – 60S	[3, 26, 27]

(Fig. 4B). Although these precipitates also contain various snRNAs, we do not know whether all three components are in the same complex or whether there are separate complexes.

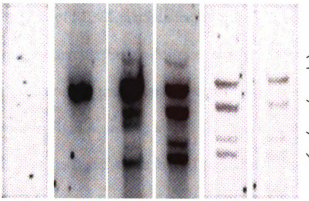
The U1 snRNP contains the core Sm proteins and the U1-specific proteins 70K, A1 and C1 assembled onto the U1 snRNA. Our hypothesis was that the association of Gal3 with U1 snRNP would be dependent on the integrity of the U1 snRNA. To test this, pooled material from fractions 3 and 4 was immunoprecipitated with anti-Gal3 before and after RNase A treatment. Fig. 5 shows that co-precipitation of the U1-70K protein by anti-Gal3 is abolished following degradation of U1 snRNA (compare lane 3 to lane 2). In contrast, the co-precipitation of the SMN protein from fractions containing large snRNA complexes was unaffected by RNase A treatment (Fig. 5, lanes 4 and 5). As expected, the immunoprecipitation of Gal3 itself is unaffected by RNase A treatment (data not shown).

The U1 snRNP mediates the loading of Gal3 onto pre-mRNA

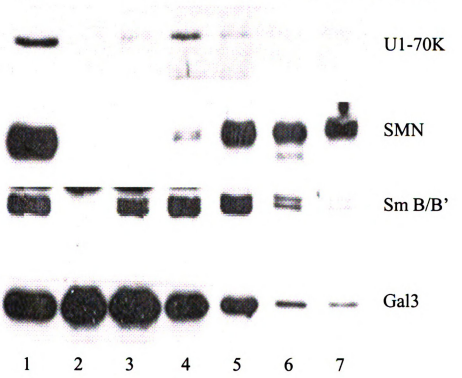
The binding of U1 snRNP to the pre-mRNA substrate at the 5' splice site results in the formation of the early (E) complex in spliceosome assembly [18, 28, 29]. We had previously documented that either Gal1 or Gal3 is associated with E-complexes [12]. Thus, our present findings that at least some Gal3 is bound to U1 snRNP prompted the question whether such Gal3-containing U1 snRNPs in fractions 3 and 4 could bind to ^{32}P-labeled pre-mRNA substrate under conditions that lead to E complex formation. Fig. 6 shows that, in the presence of fractions 3 and 4, anti-Gal3 co-precipitated exogenously added MINX pre-mRNA substrate (lane 4) whereas pre-immune serum failed to yield the same result (lane 3). That this interaction is RNA dependent is shown in lane 5 in which

Figure 4. Analysis of RNA and proteins immunoprecipitated by anti-Gal3 from glycerol gradient fractions. Indicated fractions (pooled, if necessary) from glycerol gradient fractionation of NE (see Fig. 3) were subjected to immunoprecipitation by anti-Gal3 coupled beads. Bound material was collected and analyzed for RNA (Panel A) or proteins (Panel B). Panel A shows a northern blot of the bound material (~50%) of the anti-Gal3 beads from each immunoprecipitation. ^{32}P-labeled oligonucleotides specific for each snRNA were used to detect the RNA species indicated at right. Panel B shows a western blot for proteins immunoprecipitated from the indicated fractions by anti-Gal3 beads. Material subjected to western blotting represents 50% of the bound fraction; lane 2, immunoprecipitate from fraction 1; lane 3, immunoprecipitate of fractions 3 & 4; lane 4, immunoprecipitate of fractions 6 & 7; lane 5, immunoprecipitate of fractions 9, 10, 11; lane 6, immunoprecipitate of fractions 13 & 14 and lane 7, immunoprecipitate of fractions 16, 17, 18. NE (lane 1) is included to verify the identity and migration of the proteins in western blotting. Respective antibodies against the proteins indicated at right were used in western blotting.

A.


U2
U1
U4
U5
U6

B.

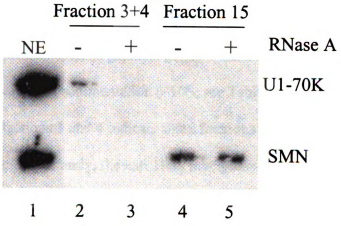

NE
1
3+4
6+7
9-11
13+14
16-18
Fraction
U1-70K
SMN
Sm B/B'
Gal3
1
2
3
4
5
6
7

fractions 3 and 4 were treated with micrococcal nuclease (whose activity was subsequently inhibited by the addition of EGTA) prior to incubation with MINX. After nuclease digestion, the level of MINX precipitated by anti-Gal3 bound was reduced to levels bound to pre-immune serum. However, residual micrococcal nuclease activity did not degrade the MINX substrate as approximately equal amounts of MINX were used for the anti-Gal3 precipitation (Fig. 6A, lanes 1 and 2). Northern blotting analysis showed that the U1 snRNA in fractions 3 and 4 was indeed degraded by the micrococcal nuclease treatment (Fig. 6B, lanes 1 and 2). Examination of the precipitated material shows the presence of U1 snRNA in the anti-Gal3 precipitate, but not in the pre-immune control or in fractions first treated with micrococcal nuclease and precipitated with anti-Gal3 (Fig. 6B, lanes 3-5). Finally, blotting for the Gal3 polypeptide in the bound material shows no significant change in precipitation of the protein after nuclease treatment (Fig. 6C, lanes 4 and 5). These results suggest that Gal3 is first assembled onto the pre-mRNA substrate via its interaction with the U1 snRNP particle.

DISCUSSION

Previous experiments had documented that Gal1 and Gal3 are factors involved in pre-mRNA splicing assayed in a cell-free system [10, 11]. Depletion of the galectins resulted in an arrest of spliceosome assembly at an early step, corresponding to the H-/E-complex. Given that neither Gal1 nor Gal3 interacts directly with pre-mRNA [12], it was of some interest to define how and at what step either protein is brought into the spliceosome. The most important conclusion derived from the present series of studies is, therefore, that the U1 snRNP can mediate the loading of Gal3 onto the pre-mRNA substrate during spliceosome assembly (see Figure 7). This association of U1 snRNP

Figure 5. Effect of RNase A treatment on Gal3 association with U1 snRNP. Fractions 3 and 4 or fraction 15 of NE separated on a glycerol gradient (see Fig. 3) were subjected to treatment with RNase A or a control mock treatment at 30°C for 30 minutes. After treatment, the reactions were incubated with anti-Gal3 coupled beads and the bound material analyzed for protein. Bound material from the mock treated (lanes 2 and 4) or RNAse treated (lane 3 and 5) fractions was western blotted for the proteins indicated on the right.

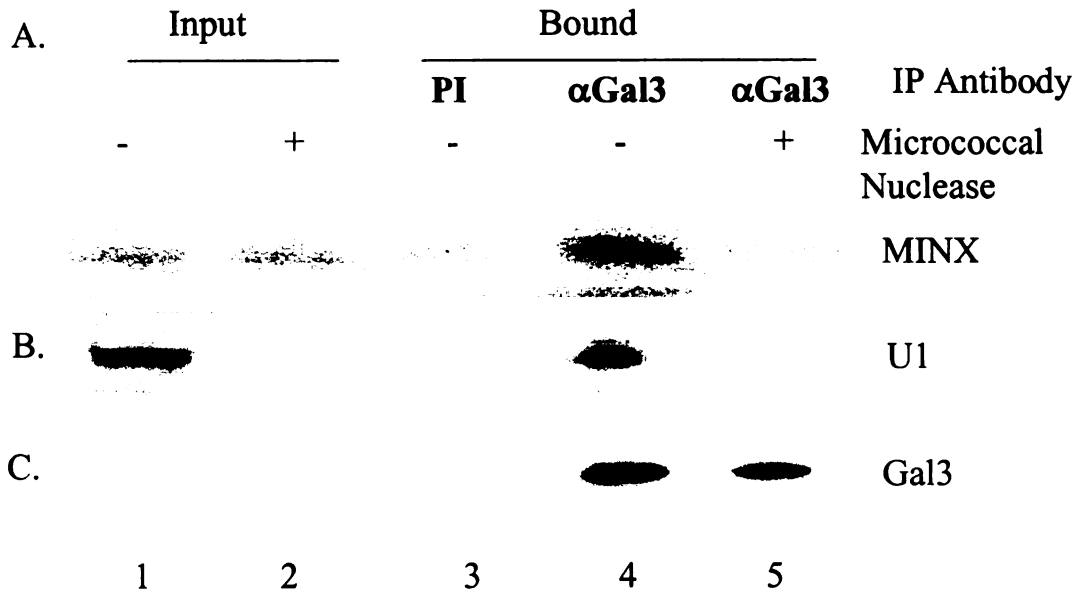

Fraction 3+4
Fraction 15
NE
-
+
-
+
RNase A
U1-70K
SMN
1
2
3
4
5

with Gal3 also exists outside of the pre-mRNA splicing pathway, i.e. in HeLa NE without any pre-mRNA scaffold. In addition to associating with U1 snRNP outside of the splicesome, we have found that Gal3 exists in larger complexes containing multiple snRNAs and other RNA processing factors (Figure 7). The involvement of Gal3 in the pre-mRNA splicing pathway, and in particular, with the spliceosome has been shown previously [12].

The conclusion that U1 snRNP facilitates the loading of Gal3 onto the pre-mRNA scaffold is based on several key considerations. First, Gal3 is associated with multiple snRNP complexes in the absence of splicing substrate as deduced from immunoprecipitation analysis of glycerol gradient fractions derived from NE. In particular, the position of sedimentation (~10S; see Table I) and the RNA and protein composition of fractions 3 and 4 indicate these fractions contain, predominantly, the mono U1 snRNP. Importantly, the anti-Gal3 precipitate of fractions 3 and 4 contained U1 snRNA and the U1-specific polypeptide U1-70K, suggesting that some of the Gal3 in these fractions was associated with the mono U1 snRNP.

Second, purified U1 snRNP binds selectively to the 5' splice site [29] and in the canonical step-wise scheme of spliceosome assembly, the U1 snRNP participates early during assembly and is required for the stable association of other snRNPs with pre-mRNA [18, 30]. A discrete ATP-independent complex containing U1 snRNP, designated the E-complex, is the first specific complex detected by gel filtration [31] and it constitutes the functional precursor to the ATP-dependent A complex, committing the pre-mRNA to the spliceosome assembly pathway [6].

Figure 6. Immunoprecipitation by anti-Gal3 of radio-labeled pre-mRNA after incubation with glycerol gradient fractions. Pooled fractions 3 and 4 were either treated with micrococcal nuclease or mock treated at 37°C for 30 minutes and then nuclease activity stopped by addition of EGTA. ^{32}P-labeled MINX pre-mRNA was added to the reactions and incubated for 15 minutes at 30°C. The reactions were incubated with antibody coupled beads, either anti-Gal3 (αGal3) or pre-immune serum (PI). Panel A shows the MINX pre-mRNA by autoradiography. Panel B shows a northern blot for the U1 snRNA. Panel C shows a western blot for the Gal3 protein. Input represents the fraction material and ^{32}P-MINX subjected to immunoprecipitation in reactions either mock treated (lane 1) or treated with micrococcal nuclease (lane 2). Bound material from immunoprecipitation of mock-treated fractions by pre-immune serum (lane 3) and anti-Gal3 (lane 4) is shown. Bound material from immunoprecipitation by anti-Gal3 beads using fractions first treated with micrococcal nuclease is shown in lane 5.

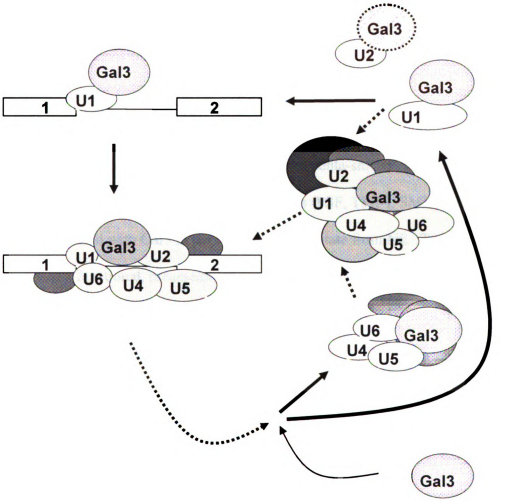

A.
Input
Bound
PI
αGal3
αGal3
IP Antibody
-
+
-
-
+
Micrococcal Nuclease
MINX
B.
U1
C.
Gal3
1
2
3
4
5

Finally, we have taken advantage of this initial, ATP-independent recognition of the 5' splice site by U1 snRNP as a functional test of the Gal3-containing U1 snRNP complex in fractions 3 and 4. Under conditions for E-complex formation (30° C in the absence of ATP), incubation of splicing substrates with fractions 3 and 4 should result in the association of Gal3 with the pre-mRNA. Indeed, we were able to demonstrate this by immunoprecipitating fractions 3 and 4 with antibodies against Gal3 and detecting radioactive MINX pre-mRNA as well as U1 snRNA.

A key caveat to our conclusion is that previous studies on the protein composition of purified U1 snRNP all describe the same core Sm proteins as well as U1-specific polypeptides: (a) U1 70K (~70 kD); (b) U1-specific A (~34 kD); (c) Sm B/B' (~28 kD); (d) U1-specific C (~22 kD); (e) Sm D (~16 kD); (f) Sm E (~12 kD); (g) Sm F (~11 kD); and (h) Sm G (~9 kD). The SDS gels of these studies did not reveal any band (~30 kD) that might correspond to Gal3 [32-34]. To reconcile this apparent discrepancy with our present observations, we note that the protocols used for the purification of U1 snRNP all involved the use of buffers of high ionic strength (150 mM NaCl or 175 mM NH_4Cl). We had shown, however, that antibodies directed against Gal1 or Gal3 can immunoprecipitate radioactive spliceosomal RNA species under conditions of the splicing assay (60 mM KCl) but that higher salt concentrations (130 mM or greater) release the galectins from spliceosomal complexes [12]. In the context of the present study, we have also observed that the co-precipitation of snRNAs by anti-Gal3 was sensitive to disruption by high ionic strength (K.C. Haudek and R.J. Patterson, unpublished observations). It seems reasonable to expect, therefore, that snRNPs and spliceosomes isolated under splicing conditions (60 mM salt) would contain Gal3.

Figure 7. Diagram showing the association of Gal3 with snRNPs and the pre-mRNA splicing substrate. The pre-mRNA is shown as two rectangular exons joined by a single line intron. Newly identified associations of Gal3 with snRNPs outside the splicesome are indicated on the right. Gal3 can enter the splicing pathway at an early assembly stage via its association with U1 snRNP. In addition, Gal3 does associate with multiple snRNPs in larger complexes outside of the spliceosome, shown here as the U4/U6.U5 tri-snRNP. Although U2 exists as a single snRNP particle in the nucleus, it is unknown whether Gal3 can interact with this single snRNP alone, illustrated by a semi-hatched Gal3. Whether Gal3 interacts with snRNPs being recycled between rounds of splicing or nascent snRNPs in the nucleus is unknown. Solid arrows indicate steps supported by data shown previously or contained in this report. Dashed arrows represent hypothetical galectin involvement in assemblies or mechanisms of snRNP loading onto pre-mRNA. Unlabeled shaded ovals indicated previously identified and novel proteins associated with snRNPs isolated under conditions optimal for *in vitro* splicing activity

Pre-mRNA splicing Galectin-snRNP complexes



Peng et al. recently reported the isolation, under conditions of the *in vitro* splicing assay (60 mM salt), of a macromolecular complex that assembles in the absence of pre-mRNA substrate and contains all five uracil-rich snRNAs [8]. Mass spectrometry analysis of this complex, designated as the PCC (PSF-containing complex; see Table I), revealed a protein composition similar to, but nevertheless distinct from, the corresponding composition of fully assembled active spliceosomes. Gal3 was found as a component of the PCC. Our present findings, that all five snRNAs are co-precipitated by anti-Gal3, either from complete NE or from selected gradient fractions, raises the possibility that we also have isolated a mammalian penta-snRNP, possibly similar if not identical to the PCC. In fact, multiple components (PSF, TFII-I, SMN) identified in the PCC were detected with Gal3 in the anti-Gal3 precipitate from NE. Together with the discovery of a functional, preassembled penta-snRNP in yeast [7], these results suggest that, under certain conditions, spliceosome assembly may be facilitated through the association of large ribonucleoprotein complexes that are already preformed in the absence of a pre-mRNA scaffold. In addition, the identification of novel proteins complexed with snRNAs under less stringent isolation conditions, characterized in this report and previously [7, 8], raises the possibility that the network of snRNA-interacting proteins outside of the spliceosome contains more members than originally described.

Although we have been able to demonstrate a function for the Gal3-U1 snRNP complex in terms of 5' splice recognition and E-complex formation, the role(s) of the higher molecular weight multi-snRNP complexes containing Gal3 remains as a challenge for future studies.

REFERENCES

1. Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245.

2. Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182-185.

3. Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. Rna *8*, 426-439.

4. Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci U S A *99*, 16719-16724.

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

6. Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev *5*, 2534-2546.

7. Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell *9*, 31-44.

8. Peng, R., Hawkins, I., Link, A.J., and Patton, J.G. (2006). The splicing factor PSF is part of a large complex that assembles in the absence of pre-mRNA and contains all five snRNPs. RNA Biol *3*, 69-76.

9. Malca, H., Shomron, N., and Ast, G. (2003). The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. Mol Cell Biol *23*, 3442-3455.

10. Dagher, S.F., Wang, J.L., and Patterson, R.J. (1995). Identification of galectin-3 as a factor in pre-mRNA splicing. Proc Natl Acad Sci U S A *92*, 1213-1217.

11. Vyakarnam, A., Dagher, S.F., Wang, J.L., and Patterson, R.J. (1997). Evidence for a role for galectin-1 in pre-mRNA splicing. Mol Cell Biol *17*, 4730-4737.

12. Wang, W., Park, J.W., Wang, J.L., and Patterson, R.J. (2006). Immunoprecipitation of spliceosomal RNAs by antisera to galectin-1 and galectin-3. Nucleic Acids Res *34*, 5166-5174.

13. Agrwal, N., Sun, Q., Wang, S.Y., and Wang, J.L. (1993). Carbohydrate-binding protein 35. I. Properties of the recombinant polypeptide and the individuality of the domains. J Biol Chem *268*, 14932-14939.

14. Ho, M.K., and Springer, T.A. (1982). Mac-2, a novel 32,000 Mr mouse macrophage subpopulation-specific antigen defined by monoclonal antibodies. J Immunol *128*, 1221-1228.

15. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

16. 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*, 1475-1489.

17. Conway, G.C., Krainer, A.R., Spector, D.L., and Roberts, R.J. (1989). Multiple splicing factors are released from endogenous complexes during in vitro pre-mRNA splicing. Mol Cell Biol *9*, 5273-5280.

18. 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*, 814-821.

19. Park, J.W., Voss, P.G., Grabski, S., Wang, J.L., and Patterson, R.J. (2001). Association of galectin-1 and galectin-3 with Gemin4 in complexes containing the SMN protein. Nucleic Acids Res *29*, 3595-3602.

20. Bottner, C.A., Schmidt, H., Vogel, S., Michele, M., and Kaufer, N.F. (2005). Multiple genetic and biochemical interactions of Brr2, Prp8, Prp31, Prp1 and Prp4 kinase suggest a function in the control of the activation of spliceosomes in Schizosaccharomyces pombe. Curr Genet *48*, 151-161.

21. Patton, J.R., Patterson, R.J., and Pederson, T. (1987). Reconstitution of the U1 small nuclear ribonucleoprotein particle. Mol Cell Biol *7*, 4030-4037.

22. Meister, G., Buhler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F., and Fischer, U. (2000). Characterization of a nuclear 20S complex containing the survival of motor neurons (SMN) protein and a specific subset of spliceosomal Sm proteins. Hum Mol Genet *9*, 1977-1986.

23. Gozani, O., Patton, J.G., and Reed, R. (1994). A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. Embo J *13*, 3356-3367.

24. Behrens, S.E., Tyc, K., Kastner, B., Reichelt, J., and Luhrmann, R. (1993). Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and

has a bipartite RNP structure under splicing conditions. Mol Cell Biol *13*, 307-319.

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

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

27. Grabowski, P.J., Seiler, S.R., and Sharp, P.A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell *42*, 345-353.

28. Bindereif, A., and Green, M.R. (1987). An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. Embo J *6*, 2415-2424.

29. 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*, 509-518.

30. Zillmann, M., Rose, S.D., and Berget, S.M. (1987). U1 small nuclear ribonucleoproteins are required early during spliceosome assembly. Mol Cell Biol *7*, 2877-2883.

31. Reed, R. (1990). Protein composition of mammalian spliceosomes assembled in vitro. Proc Natl Acad Sci U S A *87*, 8031-8035.

32. Hinterberger, M., Pettersson, I., and Steitz, J.A. (1983). Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs. J Biol Chem *258*, 2604-2613.

33. Pettersson, I., Hinterberger, M., Mimori, T., Gottlieb, E., and Steitz, J.A. (1984). The structure of mammalian small nuclear ribonucleoproteins. Identification of multiple protein components reactive with anti-(U1)ribonucleoprotein and anti-Sm autoantibodies. J Biol Chem *259*, 5907-5914.

34. Bringmann, P., and Luhrmann, R. (1986). Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. Embo J *5*, 3509-3516.

Chapter 3

Identification of galectin-3 in aptamer-selected early splicing complexes

ABSTRACT

We have shown that galectin-1 and galectin-3 are associated with the spliceosome from early complex formation (loading of U1 snRNP) through active complexes by immunoprecipitations using galectin-specific antisera. Here we use a galectin-independent method to document an association of galectin-3 with the early spliceosome (E complex). Using a pre-mRNA engineered to contain three recognition sites for the MS2 bacteriophage protein, assembled splicing complexes can be isolated and analyzed. We show that our isolation system selected only a single pre-mRNA containing our aptamer and that *bona fide* splicing factors could recognize and assemble on this pre-mRNA. Furthermore, the protein constituents assembled on the pre-mRNA varied, as expected, with incubation conditions to allow the assembly of unique splicing complexes. Characterization of selected complexes showed that isolated E complexes contained U1 snRNA, but not U6 snRNA. Subsequent analysis of proteins showed galectin-3 associated with the E complex. These results verify the association of galectin-3 with the spliceosome using an isolation method independent of galectin selection.

INTRODUCTION

Splicing cofactors containing the U snRNAs (snRNPs) exist either as single entities (the U1 and U2 snRNPs) or as larger complexes, such as with 3 snRNAs (the U4/U6.U5 tri-snRNP) [1, 2]. These snRNPs assemble onto a pre-mRNA scaffold to carry out pre-mRNA splicing. During spliceosome assembly, the first snRNP to bind the pre-mRNA is the U1 snRNP [3, 4]. This U1 binding event is the marker for early

complex formation (E-complex) and can be detected by gel mobility shifts [5, 6]. The binding of U1 to pre-mRNA is ATP independent and allows the step-wise assembly of higher-order splicing complexes (A/B/C) with the addition of ATP and incubation [7]. Each of these active complexes can be distinguished based on gel mobility shifts and on protein and RNA constituents.

We recently have identified galectin-3 (Gal3) as a snRNP-associated protein in HeLa nuclear extract (NE) (see Chapter 2). Immunoprecipitation of NE incubated in the absence of a splicing substrate with antiserum specific for Gal3 revealed the 5 splicing snRNAs along with several snRNA core proteins. Using glycerol gradient sedimentation to separate the endogenous nuclear snRNPs, we could detect a Gal3-containing mono-snRNP (U1) at ~10S. Using an *in vitro* binding assay, we showed that the Gal3-containing U1 snRNP formed a Gal3-immunoprecipitable complex on an exogenous pre-mRNA. Pretreatment of the Gal3-containing U1 snRNP with micrococcal nuclease abolished pre-mRNA binding activity. This complex was reminiscent of early splicing complexes that incorporated Gal3 (or galectin-1 (Gal1)) when splicing substrate is incubated with pre-mRNA in a complete NE [8]. Both types of early complexes relied on galectin-specific antiserum in co-immunoprecipitation protocols.

To confirm these data, a method independent of galectin selection was explored to show that Gal3 was indeed associated with U1 snRNP and pre-mRNA in an early splicing complex. Methods to select nucleic acids and associated proteins from complex mixtures have been evaluated. One powerful approach is to engineer an aptamer sequence into the nucleic acid of interest and subsequently select the aptamer-containing nucleic acid via interaction with its specific ligand. To purify specific RNP complexes

involved in splicing, Zhou et al. [9] developed the selection system schematically illustrated in Figure 1. First, a pre-mRNA substrate derived from the adenovirus major late gene (AdML-M3) was engineered such that the 3'-end contained three hairpin loops that bind to the bacteriophage MS2 protein with high affinity. In parallel, a fusion protein containing MS2 and maltose-binding protein (MBP) was expressed and purified on the basis of its binding to amylose beads and specific elution using soluble maltose. The purified MS2-MBP fusion protein, bound to fresh amylose beads, is then used to select for RNP complexes formed by incubation of the AdML-M3 pre-mRNA substrate with NEs under various conditions. The isolated RNP complexes can be similarly released from the beads using soluble maltose. Using this (or similar) aptamer selection protocol, other groups have isolated and purified various splicing complexes which were subjected to protein identification by mass spectrometry [10-13]. These latter studies have revealed that the spliceosome proteome consists of more than 300 proteins.

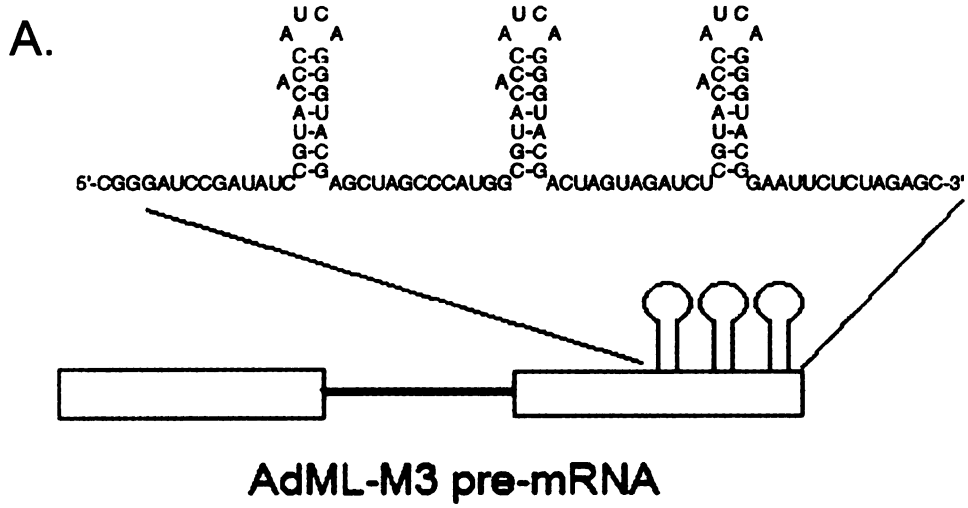
Using this MS2 aptamer selection protocol, we now show that early splicing complexes assembled in a complete NE contain Gal3 in addition to the Sm core polypeptide Sm B/B' and Sm D and the U1 snRNA.

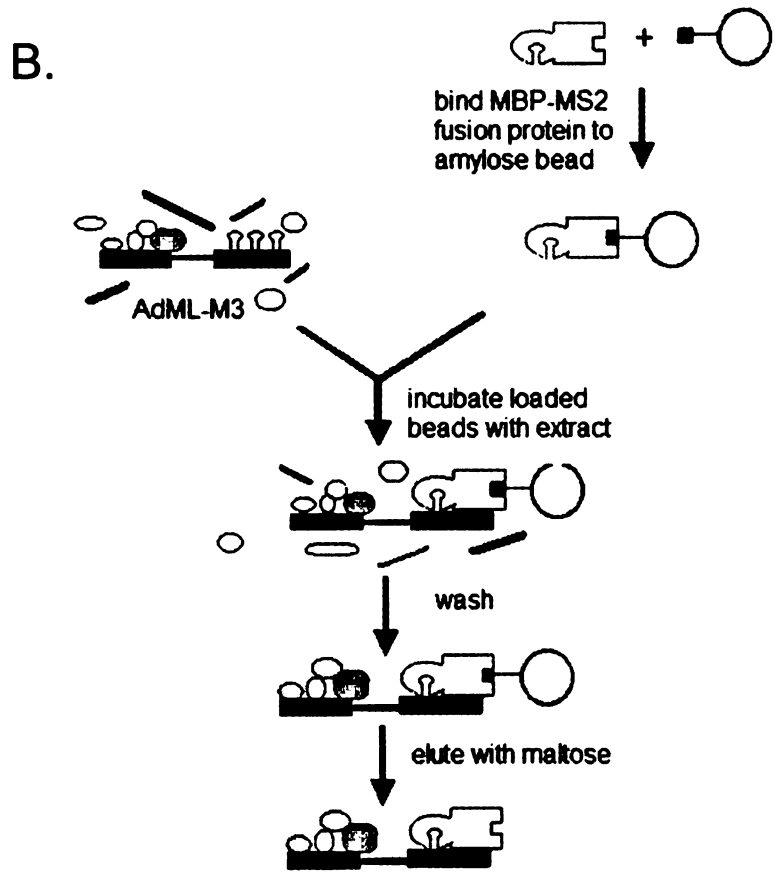
MATERIALS AND METHODS

Antibodies

Polyclonal rabbit antibodies directed against Gal3 [14] were described previously. The following antibodies were purchased for immunoblotting: rabbit anti-hnRNP C1/C2, goat anti-Slu7 (Santa Cruz Biotechnology); rabbit anti-MBP (Chemicon International); monoclonal mouse anti-SMN (BD Transduction Lab) and human autoimmune sera ENA anti-Sm (The Binding Site). Rabbit anti-Gal3, described previously [14], was used for

Figure 1. Schematic diagram illustrating the structure of AdML-M3 pre-mRNA and its use in purifying spliceosomal components assembled on the RNA. (A) The AdML-M3 pre-mRNA consists of two exons (rectangles) joined by a single intron (dark line) and three hairpin structures recognized by the MS2 protein. The MS2 recognition hairpins are situated at the 3' end of exon 2. (B) The MBP-MS2 fusion protein can bind to amylose beads via the maltose binding protein (MBP) and to AdML-M3 through MS2 recognition of the hairpins. Components assembled on the AdML-M3 pre-mRNA can thus be affinity selected on the amylose beads and subsequently eluted with soluble maltose.


A.
5'-CGGGAUCCGAUAUC AGCUAGCCCAUGG ACUAGUAGAUCU GAAUUCUCUAGAGC-3'
AdML-M3 pre-mRNA


B.
bind MBP-MS2 fusion protein to amylose bead
AdML-M3
incubate loaded beads with extract
wash
elute with maltose

blotting Gal3 protein. Primary mouse monoclonal antibodies were detected by goat anti-mouse IgG light chain specific-HRP conjugates (Jackson ImmunoResearch Laboratories). All other secondary antibody-HRP conjugates (Pierce Biotechnology) were directed against both the heavy and light chains of the primary blotting antibodies.

Plasmids and *in vitro* transcription

Plasmids encoding AdML-M3 pre-mRNA and MBP-MS2 protein were kindly provided by Dr. Robin Reed (Harvard University) [9, 12]. Plasmids containing anti-sense constructs for U1 and U6 snRNA for use as northern blot probes were gifts from Dr. Jeff Patton (University of South Carolina). The plasmid containing MINX pre-mRNA [6] and *in vitro* transcription reactions have been described previously [15]. For ^{32}P-labeled AdML-M3, the plasmid was digested with Xba I and *in vitro* transcribed using T7 RNA polymerase. For anti-sense probes against AdML-M3, the plasmid was digested with Cla I and *in vitro* transcribed using SP6 RNA polymerase. For U1 and U6 anti-sense probes, plasmids were digested with Nae I and *in vitro* transcribed with ^{32}P-GTP using SP6 RNA polymerase for U6 anti-sense and T7 RNA polymerase for U1 anti-sense. For large quantities of non-radioactive AdML-M3, transcription using AmpliScribe T7 High Yield Transcription kit was used (Epicentre). Amylose-conjugated agarose beads were purchased from New England Biolabs and the MBP-MS2 fusion protein was isolated according to the manufacturer's protocol and dialyzed into 60% buffer D (buffer D is 20 mM HEPES, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol).

Polyacrylamide gel electrophoresis, immunoblotting and northern blotting

For protein analysis, samples were subjected to 12% SDS-PAGE as described by Laemmli [16]. Proteins were electrophoretically transferred from the gel onto Hybond nitrocellulose membrane (Amersham Biosciences) or Biotrace PVDF membrane (Pall) in transfer buffer (25 mM Tris, 193 mM glycine and 20% methanol, pH 8.3). Following transfer, membranes were blocked overnight in 10% nonfat dry milk in Tris-buffered saline containing Tween-20 (10 mM Tris, pH 7.5, 0.5 M NaCl, 0.05% Tween 20, T-TBS). Primary antibodies for immunoblotting were diluted in 1% milk-T-TBS and incubated on the membrane for 1 hour at room temperature. After washing four times, 15 minutes each, in T-TBS, the appropriate secondary antibody conjugated to horseradish peroxidase was added in 1% milk-T-TBS for 1 hour. Following four T-TBS washes as above, proteins were visualized using the Western Lightning Chemiluminescence System (Perkin Elmer Life Sciences).

For RNA samples, the RNA was extracted as described below and precipitated with 3 volumes of ethanol at -80°C. The precipitated RNA was dissolved in 10 µl of sample buffer (9:1/formamide:bromophenol blue) and subjected to electrophoresis through 13% polyacrylamide (bisacrylamide –acrylamide 1.9:50 [wt/wt])-8.3 M urea gels, run in 1X TBE (90 mM Tris base, 90 mM boric acid and 2.5 mM EDTA, pH 8.0). The radioactive RNA species were revealed by autoradiography.

For Northern analysis the RNA was transferred via wicking in 2X SSC (0.3 M NaCl, 0.03 M sodium citrate) overnight onto a nylon membrane (Hybond-N, Amersham Biosciences) and cross-linked by exposure to UV. After cross-linking, the nylon membrane was pre-hybridized in 20 ml of hybridization solution containing 20% deionized formamide, 3X SSC, 5X Denhardt's, 50 mM Na_2HPO_4/NaH_2PO_4, pH 6.8, 0.1%

SDS and 0.1 mg/ml herring sperm DNA for about 4 hours at 42°C. After pre-hybridization, ^{32}P-labeled anti-sense probes were added to the hybridization solution and the membrane was hybridized overnight (~16 hours) at 42°C. The membrane was then washed once with 30 ml 2X SSC + 0.1% SDS at 50°C for 30 minutes. Northern hybridization was imaged by phosphorimage analysis (Molecular Dynamics).

Selection of AdML-M3

Nuclear extract (NE) from HeLa S3 cells (National Cell Culture Center, Minneapolis, MN) was prepared as described in Dignam et al [17]. To test the selectivity of the MS2 system, 10 μg of MBP-MS2 were loaded onto 20 μl of amylose beads in 0.5 ml binding buffer (20 mM HEPES, pH 7.9, 60 mM NaCl, 1 mM DTT) at 4°C for 1 hour. Then 0.8 pmol of *in vitro* transcribed MINX and AdML-M3 pre-mRNA were mixed and incubated in an 80 μl reaction in binding buffer with 10 μl HeLa NE for 20 minutes at 4°C and incubated with the loaded beads. The beads were washed three times with 0.5 ml binding buffer containing 0.05% Triton-X 100 (TX) and eluted with 20 mM maltose in binding buffer. Eluted material was split into aliquots and analyzed for protein and RNA, as described above. RNA was extracted by adding proteinase K (4 mg/ml final concentration) to the sample and incubating at 37°C for 30 minutes, then diluting to 100 μl with 125 mM Tris (pH 8), 1 mM EDTA, 300 mM sodium acetate. RNA was extracted by mixing with 200 μl of phenol–chloroform (50:50 [vol/vol]), followed by 100 μl of chloroform.

Pre-mRNA complex isolation

For early and active complex selection, 1 μg of *in vitro* transcribed AdML-M3 was incubated with roughly 5 μg of MBP-MS2 and 20 μl of amylose beads for one hour

at 4°C with rocking. The beads were washed once with binding buffer. A 150 µl reaction consisting of 90 µl HeLa NE, 3 mM $MgCl_2$, 40 U RNasin (Promega) was incubated with the AdML-M3 -loaded amylose beads at 30°C for 30 minutes to form E complex. Similarly, active complexes were formed with an identical reaction, also containing 0.5 mM ATP and 20 mM creatine phosphate. Active complexes were formed by incubating the reaction with the AdML-M3 -loaded amylose beads for 30 minutes at 30°C. Control beads had the same amount of MBP-MS2 loaded onto the amylose beads and were incubated with NE under active complex conditions. After incubation, the volume was brought up to 0.5 ml with binding buffer and mixed for one hour at 4°C. After binding the complexes, beads were washed 3 times in 1 ml of binding buffer containing 0.05% TX and eluted with 20 mM maltose in binding buffer.

E complex analysis

For analysis of the E complex, 1 µg of *in vitro* transcribed AdML-M3 was incubated with roughly 4 µg of MBP-MS2 in a total of 20 µl of binding buffer on ice for 30 minutes. After incubation, the protein and AdML-M3 were added to a 150 µl reaction containing 100 µl NE, 3 mM MgCl2 and 80 U RNasin and incubated for 20 minutes at 30°C. The whole reaction was then incubated with 30 µl of amylose beads for one hour at 4°C with rocking, with the total volume being brought up to 0.5 ml with binding buffer. A control reaction lacking only the AdML-M3 pre-mRNA was performed. After binding the beads were washed 3 times with 1 ml of binding buffer + 0.05% TX and eluted with 20 mM maltose in binding buffer.

RESULTS

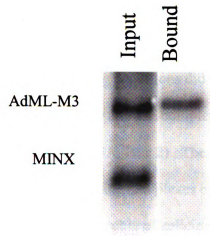
Selection of RNA species through MS2 recognition of its cognate RNA hairpin

We have adopted the AdML-M3 selection system as developed by Zhou et al. and outlined in Figure 1 [9]. We tested this selection system by comparing AdML-M3 against MINX, another derivative of the adenovirus major late gene but lacking the MS2 binding sites. The two RNAs, both *in vitro* transcribed and ^{32}P-labeled, were added in an equimolar ratio (Figure 2, lane 1) to HeLa nuclear extract, incubated and passed over amylose beads loaded with MBP-MS2 fusion protein. After washing, the bound fraction was eluted and subjected to gel electrophoresis. The bound fraction (Figure 2, lane 2) showed a strong selection for AdML-M3 over MINX pre-mRNA. There was nearly a 15-fold enrichment for the AdML-M3 substrate over the starting ratio of the mRNAs in the input. These results indicate that our experimental set-up, including the fusion protein and AdML-M3 transcript, is functional and that this MS2 selection allows a high degree of specificity in RNA purification.

Isolation of splicing complexes assembled on AdML-M3

To test if we could isolate distinct splicing complexes, AdML-M3 was incubated with HeLa NE, with and without added ATP, in order to form two unique splicing complexes: the ATP-independent early (E-) complex and catalytically active spliceosomes that require ATP [18]. Splicing complexes containing AdML-M3 were selected by passing the incubated extract over amylose beads loaded with MBP-MS2 fusion protein. A control reaction lacking the AdML-M3 substrate was carried along under conditions to form active splicing complexes. The bound and eluted material was western blotted for the Sm proteins to confirm the selection of splicing complexes (Figure 3A, lanes 2 and 3). Under both incubation conditions, early and active, we found the association of Sm proteins with AdML-M3, suggesting that snRNPs had been isolated

Figure 2. Specificity of selection by MBP-MS2 fusion protein. Two different ^{32}P-labeled pre-mRNAs, AdML-M3, containing three MS2 recognition sites and MINX, lacking any MS2 recognition sites, were incubated together with NE and passed over amylose beads loaded with MBP-MS2. The bound material was eluted with maltose. Input (lane 1) shows the starting amount of AdML-M3 and MINX in the reaction. Material eluted off the MBP-MS2 column is shown as bound (lane 2).



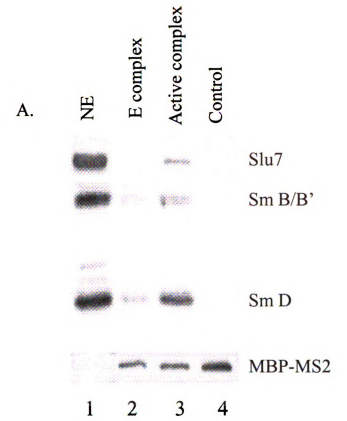
1 2

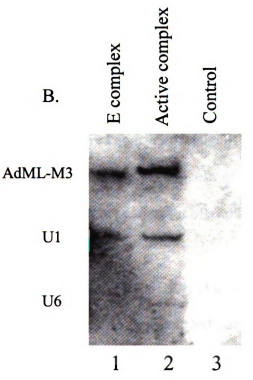
with the AdML-M3 substrate. To test whether we could distinguish between early and active complexes we probed for Slu7, a required second-step splicing factor that has only been found to associate with catalytically active spliceosomes [10, 19]. The AdML-M3 incubated under active splicing conditions did indeed pull out Slu7 (Figure 3A, lane 3). However, conditions that only allow E complex formation did not assemble Slu7 onto the AdML-M3 substrate (Figure 3A, lane 2). These results confirm that we can distinguish between active and early splicing complexes by AdML-M3 and that incubation conditions can control the assembly of splicing complexes.

Blotting for both Slu7 and Sm proteins failed to reveal any of these proteins in the eluted material from the control column (Figure 3A, lane 4). A final control was performed by blotting for the SMN protein. SMN is a nuclear protein that interacts with a variety of RNPs but has failed to be identified as directly interacting with the spliceosome. Immunoblotting for the SMN protein showed that it is not selected by AdML-M3 under any condition despite there being plenty of SMN present in the NE (data not shown). Blotting with antibodies against MBP confirm an equal amount of fusion protein in all samples.

Because U1 snRNA binding to the 5' splice site of a pre-mRNA is the hallmark of early complex formation, we examined the eluted material for the presence of snRNA. Northern blots revealed the presence of U1 snRNA in both the AdML-M3 assembled E complex and active complex, but U6 snRNA only in the active complex elution (Figure 3B). The U6 snRNP is necessary for a catalytically active spliceosome and its incorporation onto the spliceosome is ATP-dependent. Neither U1 nor U6 could be

Figure 3. Analysis of protein and RNA components of early and active splicing complexes selected via the AdML-M3 pre-mRNA. Amylose beads, loaded with AdML-M3 pre-mRNA and MBP-MS2 protein, were mixed with NE diluted to splicing conditions and incubated to form early or active splicing complexes. A control column lacking AdML was incubated under conditions that give rise to active complexes. The beads were washed and eluted with maltose. Panel A shows western blots of the bound material from the early complex (lane 2), active complex (lane 3) and control column (lane 4) for the proteins indicated at right. Panel B shows a northern blot of the eluted material from early complex (lane 1), active complexes (lane 2) and control column (lane 3) using anti-sense RNA probes against AdML-M3, U1 and U6 snRNAs.


A.
NE
E complex
Active complex
Control
Slu7
Sm B/B'
Sm D
MBP-MS2
1
2
3
4


B.
E complex
Active complex
Control
AdML-M3
U1
U6
1
2
3

detected eluted from control beads (Figure 3B, lane 3). All of these results indicate the selection of complexes by AdML-M3 to be specific and that the association of AdML-M3 complexes can be controlled by incubation conditions. In particular, selection of complexes under E conditions allow the selection of U1 snRNA, an important early complex marker, but not Slu7 protein or U6 snRNA, both components of active splicing complexes.

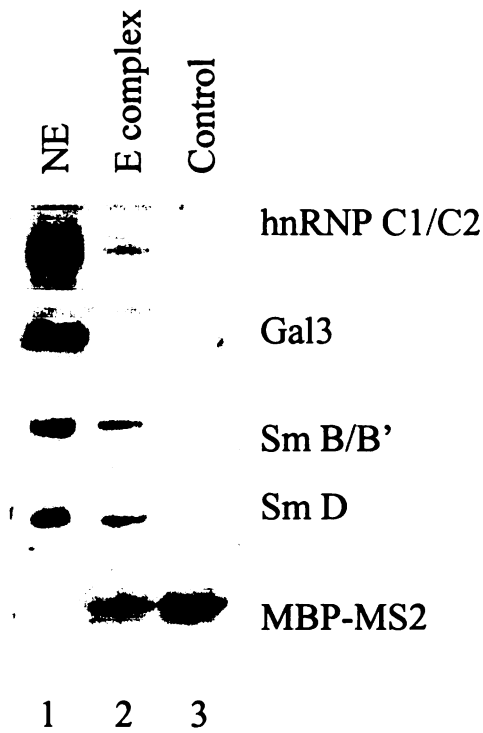
E-complexes assembled on AdML-M3 contain Gal3

Based on our previous data showing the association of Gal3 with U1 snRNP (see Chapter 2), our hypothesis was that Gal3 is a *bona fide* component of early splicing complexes, in particular the E complex. To test this hypothesis, we performed AdML-M3 selection of E complexes and the bound material was blotted for known E complex factors. The Sm proteins along with hnRNP C1/C2 were detected in our AdML-M3 selected material but not under our control conditions (Figure 4, lanes 2 and 3), allowing us to conclude we had successfully pulled out an E complex. We then blotted for the Gal3 polypeptide, which revealed the protein in our E complex specific pull-out but not in the control experiment (Figure. 4, lanes 2 and 3). As a final control, blotting with antibodies against MBP showed an equal amount of fusion protein loaded in both the experimental and control conditions. The results of this experiment confirmed our hypothesis that Gal3 is a protein component of a splicing E complex.

DISCUSSION

The high affinity binding (K_d ~3 nM) of the coat protein of bacteriophage R17 to a hairpin loop on its genomic RNA has been extensively studied by Uhlenbeck and co-workers [20, 21]. Bardwell and Wickens took advantage of this system in developing

Figure 4. Analysis of proteins associated with early splicing complexes. AdML-M3 pre-mRNA was bound to MBP-MS2 then added to a splicing reaction and incubated under conditions to give early complexes. The reaction was then passed over amylose beads and the bound material eluted with maltose. A reaction lacking AdML-M3 pre-mRNA was used as a control. Bound material from both the early complex (lane 2) and control reaction (lane 3) was western blotted for the indicated proteins at right.


NE
E complex
Control
hnRNP C1/C2
Gal3
Sm B/B'
Sm D
MBP-MS2
1
2
3

a purification scheme for RNA and RNA complexes but noted that the binding of an RNA containing non-R17 sequences required two recognition sites in tandem [22]. The system was further refined in the AdML-M3 pre-mRNA construct to contain three hairpin loops and in the use of MBP-MS2 fusion protein and amylose beads for the selection of assembled spliceosomal complexes that were subsequently visualized by electron microscopy [9] and analyzed by proteomics [10, 12]. For these latter studies on spliceosome assembly as well as for our present work, it is of particular importance to note that neither the presence of the hairpin loops nor the binding of the fusion protein affected either splicing complex formation or the splicing reaction [9, 23].

Our present application of this aptamer selection system has provided key confirmatory evidence to two previous experiments. First, we showed that either anti-Gal1 or anti-Gal3 can immunoprecipitate ^{32}P-labeled MINX pre-mRNA under conditions (incubation with NE at 30°C in the absence of ATP) that would lead to the formation of only early complexes [8]. Second, we have shown that an isolated Gal3-U1 snRNP complex is sufficient to load Gal3 onto a pre-mRNA under conditions that allow the formation of early splicing complexes (see Chapter 2) and basepairing of U1 at the 5' splice site [24, 25]. Using aptamer recognition to select spliceosomal complexes assembled on the AdML-M3 pre-mRNA under the same conditions, we have now documented the detection of Gal3 in these early spliceosomal complexes containing the U1 snRNA.

These results implicate that Gal3 enters the splicing reaction early in the spliceosome assembly process. This notion is consistent with three additional lines of evidence. First, we had reported that NEs depleted of the galectins by affinity adsorption

on lactose-agarose beads failed to form active spliceosomal complexes and gel mobility shift assays of ^{32}P-labeled pre-mRNA revealed only bands migrating in the H-/E-complex region [26]. The activities of the galectin-depleted extract, in forming active splicing complexes and in performing the *in vitro* splicing reaction, were reconstituted by the addition of recombinant Gal3 with similar dose-response curves.

Second, we have expressed and purified a fragment of the murine Gal3 polypeptide containing residues 1-137, corresponding to the amino-terminal domain (ND) bearing multiple repeats of the nine-residue motif, PGAYPGXXX [14]. When the splicing assay was carried out in the presence exogenously added Gal3 ND, we observed a dose-dependent inhibition of product formation [15]. This apparent dominant negative effect of the ND was associated with the arrest of spliceosome assembly at the H-/E-complex. In both the splicing reaction and in the assembly of active spliceosomes, parallel additions of either the full-length Gal3 polypeptide or the carboxyl terminal carbohydrate recognition domain failed to yield the same effect. Finally, we have recently found that addition of the NCL-GAL3 monoclonal antibody directed against Gal3 to a splicing competent NE inhibited the splicing reaction [27]. Again, native gel electrophoresis showed that NCL-GAL3 exerted its effect early in the spliceosome assembly process, blocking the progression of H-/E-complexes into active spliceosomes.

A key question posed by our studies is whether there is any difference between two complexes formed on the pre-mRNA substrate: (a) the classical E-complex assembled upon incubation of NE with pre-mRNA at 30°C in the absence of ATP and selected by the MS2 aptamer; and (b) the complex resulting from the binding of Gal3-U1 snRNP (in fractions 3 and 4 derived from glycerol gradient fractionation of NE; see

Chapter 2) to the pre-mRNA selected by antibodies against Gal3. The aptamer selection procedure documented in the present studies, coupled with additional gel filtration (size fractionation) steps, could provide the requisite extent of purification such that the two complexes can be compared in terms of protein composition by proteomic analysis.

REFERENCES

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

2. Bringmann, P., and Luhrmann, R. (1986). Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. EMBO J *5*, 3509-3516.

3. Zillmann, M., Rose, S.D., and Berget, S.M. (1987). U1 small nuclear ribonucleoproteins are required early during spliceosome assembly. Mol Cell Biol *7*, 2877-2883.

4. Abmayr, S.M., Reed, R., and Maniatis, T. (1988). Identification of a functional mammalian spliceosome containing unspliced pre-mRNA. Proc Natl Acad Sci U S A *85*, 7216-7220.

5. Ruby, S.W., and Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science *242*, 1028-1035.

6. 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*, 814-821.

7. Bindereif, A., and Green, M.R. (1987). An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. EMBO J *6*, 2415-2424.

8. Wang, W., Park, J.W., Wang, J.L., and Patterson, R.J. (2006). Immunoprecipitation of spliceosomal RNAs by antisera to galectin-1 and galectin-3. Nucleic Acids Res *34*, 5166-5174.

9. Zhou, Z., Sim, J., Griffith, J., and Reed, R. (2002). Purification and electron microscopic visualization of functional human spliceosomes. Proc Natl Acad Sci U S A *99*, 12203-12207.

10. Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA *8*, 426-439.

11. Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci U S A *99*, 16719-16724.

12. Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182-185.

13. Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245.

14. Agrwal, N., Sun, Q., Wang, S.Y., and Wang, J.L. (1993). Carbohydrate-binding protein 35. I. Properties of the recombinant polypeptide and the individuality of the domains. J Biol Chem *268*, 14932-14939.

15. Park, J.W., Voss, P.G., Grabski, S., Wang, J.L., and Patterson, R.J. (2001). Association of galectin-1 and galectin-3 with Gemin4 in complexes containing the SMN protein. Nucleic Acids Res *29*, 3595-3602.

16. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

17. 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*, 1475-1489.

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

19. Chua, K., and Reed, R. (1999). Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. Genes Dev *13*, 841-850.

20. Romaniuk, P.J., Lowary, P., Wu, H.N., Stormo, G., and Uhlenbeck, O.C. (1987). RNA binding site of R17 coat protein. Biochemistry *26*, 1563-1568.

21. Carey, J., Cameron, V., de Haseth, P.L., and Uhlenbeck, O.C. (1983). Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. Biochemistry *22*, 2601-2610.

22. Bardwell, V.J., and Wickens, M. (1990). Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. Nucleic Acids Res *18*, 6587-6594.

23. Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol Cell *5*, 779-787.

24. 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*, 509-518.

25. Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev *5*, 2534-2546.

26. Dagher, S.F., Wang, J.L., and Patterson, R.J. (1995). Identification of galectin-3 as a factor in pre-mRNA splicing. Proc Natl Acad Sci U S A *92*, 1213-1217.

27. Gray, R.M., Davis, M.D., Ruby, K.M., Voss, P.G., Patterson, R.J., and Wang, J.L. (2007). Galectin Domains and pre-mRNA Splicing. II. Epitope for the Mac-2 Monoclonal Antibody and Its Functional Significance in the Splicing Activity of Galectin-3. RNA Biol *submitted for publication*.

Chapter 4
Concluding statements

Previous work in our laboratory had focused on galectin-1 (Gal1) and galectin-3 (Gal3) as essential pre-mRNA splicing factors. In addition, recent work has shown both Gal1 and Gal3 to be assembled on the spliceosome.

My work presented here moves our understanding of nuclear galectins forward in several significant ways: first, Gal3 is a member of numerous nuclear complexes containing both proteins and RNA; second, nuclear Gal3 interacts with snRNPs outside of the spliceosome; third, Gal3 interaction with snRNPs leads to the entry of Gal3 into the pre-mRNA splicing pathway. These new data fit well with the existing evidence of galectins as pre-mRNA splicing factors. However, more experiments must be performed to fully understand the role of galectins in snRNP biogenesis and their function in pre-mRNA splicing.

Firstly, it is of great interest to determine the binding partner of Gal3 on the snRNPs. Does Gal3 interact directly with the Sm core polypeptides, a snRNP-specific protein or does another non-snRNP protein mediate this interaction? A series of *in vitro* binding assays may reveal the binding partner of galectins. One likely protein candidate to mediate the snRNP interaction may be the general transcription factor TFII-I, as recent work shows this protein is associated (i) with Gal3 in a GST-binding assay (TFII-I can also interact with Gal1) and (ii) with the snRNP-containing complex, PCC. This interaction must be tested to see if the Gal3-TFII-I interaction is direct and if so, whether TFII-I can bind snRNPs directly.

Another interesting finding is the association of Gal3 with snRNPs in large complexes. What exactly are these large multi-snRNP complexes? Are they a

mammalian penta-snRNP complex? These complexes should be tested for functional pre-mRNA binding and splicing activity. Gradient fractions containing this penta-snRNP may be sufficient to compliment an extract depleted of functional snRNPs in an *in vitro* splicing assay. The identity of several splicing factors with the yeast penta-snRNP and mammalian PCC points to the idea that these large complexes are involved in spliceosome formation at some level.

The presence of the SMN protein in these large complexes is intriguing in what it suggests about the assembly of these complexes. Is the nuclear function of the SMN protein to assemble pre-formed spliceosomes? Or is SMN in these complexes due to its association with another snRNP related protein, such as coilin? It would be interesting to test whether these large multi-snRNP complexes are involved in performing the snRNA modifications that are carried out at the Cajal body. Methods of assessing this include searching the snRNAs for modified bases, blotting for the presence of small Cajal body RNAs (scaRNAs) and immunoblotting for coilin, the protein marker of Cajal bodies and other known snRNA modifying enzymes.

A mechanism for galectin entry into the splicing pathway is a meaningful step forward to determine the role of galectins in splicing. Although numerous lines of evidence show galectin-inhibited splicing reactions halt at an early complex stage, the exact mechanism for this inhibition is not known. Armed with these new data and the binding assay demonstrated in this thesis, perhaps the association of galectins and/or U1 loading onto a pre-mRNA in an inhibited nuclear extract and/or gradient fraction can be assayed. Could this be the step at which pre-mRNA splicing is blocked?

Another important question to ask is how the Gal3-U1 snRNP complex relates to the early splicing complex (E) assembled on the pre-mRNA. Proteomic analysis using mass spectrometry may allow the identification of unique factors in both complexes. The Gal3-U1 snRNP complex also should be tested for functional assembly of active spliceosomes. This may be assayed by either gel mobility shift assays or complementation assays of U1 depleted extracts.

Recent work has shown that Gal1 and Gal3 are mutually exclusive in spliceosomes. That is, a spliceosome may contain either Gal1 or Gal3, but not both galectins simultaneously. A logical extension of this mutually exclusive finding is that the loading of the galectins onto the spliceosome should show this exclusivity. Therefore, I believe it is important to document which snRNPs Gal1 is associated with. If Gal1 and Gal3 show similar snRNP binding and a similar Gal1-U1 snRNP is found, it too should be tested for pre-mRNA binding and the presence of Gal3. My hypothesis is that the exclusivity seen in the spliceosome is actually conferred at the galectin binding of the snRNP before entry into the spliceosome. For this to be the case, a snRNP or snRNP complex would have to contain only a single galectin binding site, recognizable by both Gal1 and Gal3.

The foundation for numerous and interesting studies further investigating these questions has been laid by the data presented in this thesis. In the future, the role of nuclear galectins in snRNP and spliceosome assembly can be elucidated.