RNA (2002), *8*:426–439. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S1355838202021088

Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis

MELISSA S. JURICA,[1] LAWRENCE J. LICKLIDER,[2] STEVEN P. GYGI,[2] NIKOLAUS GRIGORIEFF,[1] and MELISSA J. MOORE[1]

[1]Howard Hughes Medical Institute, Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02454, USA
[2]Taplin Biological Mass Spectrometry Facility, Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

ABSTRACT

We describe characterization of spliceosomes affinity purified under native conditions. These spliceosomes consist largely of C complex containing splicing intermediates. After C complex assembly on an MS2 affinity-tagged pre-mRNA substrate containing a 3′ splice site mutation, followed by RNase H digestion of earlier complexes, spliceosomes were purified by size exclusion and affinity selection. This protocol yielded 40S C complexes in sufficient quantities to visualize in negative stain by electron microscopy. Complexes purified in this way contain U2, U5, and U6 snRNAs, but very little U1 or U4 snRNA. Analysis by tandem mass spectrometry confirmed the presence of core snRNP proteins (SM and LSM), U2 and U5 snRNP-specific proteins, and the second step factors Prp16, Prp17, Slu7, and Prp22. In contrast, proteins specific to earlier splicing complexes, such as U2AF and U1 snRNP components, were not detected in C complex, but were present in similarly purified H complex. Images of these spliceosomes revealed single particles with dimensions of approximately 270 × 240 Å that assort into well-defined classes. These images represent an important first step toward attaining a comprehensive three-dimensional understanding of pre-mRNA splicing.

Keywords: electron microscopy; mass spectroscopy; pre-mRNA splicing; spliceosomes

INTRODUCTION

In eukaryotes, the removal of introns from nascent transcripts is mediated by a highly dynamic, macromolecular machine called the spliceosome. In size and complexity, spliceosomes are comparable to ribosomes. Like ribosomes, spliceosomes are composed of distinct subunits containing stable RNA molecules in tight association with multiple proteins. Five of these U-rich small nuclear RNPs (U1, U2, U4, U5, and U6 snRNPs) function together with a plethora of non-snRNP protein factors to accurately identify the termini of each intron, followed by assembly of an active spliceosome to catalyze intron excision (Staley & Guthrie, 1998; Burge et al., 1999). Although the complete set of spliceosomal components is not yet known, it can be estimated that >70 polypeptides contribute to the functional complex (Bennett et al., 1992a; Gozani et al., 1994; Neubauer et al., 1998).

Recent determination of high-resolution structures of ribosomes has significantly increased our understanding of how these amazing machines operate (Moore, 1998; Wimberly et al., 2000; Yusupov et al., 2001). For spliceosomes, however, the only available three-dimensional information consists of a few single components and subassemblies. For example, X-ray crystallography has yielded structures of U1A protein bound to its cognate stem loop in U1 snRNA (Oubridge et al., 1994), U2A′ and U2B″ proteins bound to a portion of U2 snRNA (Price et al., 1998) and tri-snRNP 15.5 kDa protein bound to a U4 snRNA fragment (Vidovic et al., 2000). X-ray structures of Sm D1/D2 complexes and Sm D3/B have suggested that the core proteins shared by U1, U2, U4, and U5 snRNPs exist as a seven-membered ring (Kambach et al., 1999), an ultrastructure also visible by electron microscopy (Raker et al., 1999). Finally, a cryo-electron microscopy structure of the entire U1 snRNP, which contains U1 snRNA,

Reprint requests to: Melissa J. Moore, Howard Hughes Medical Institute, Brandeis University, Biochemistry Department, 415 South Street, Waltham, Massachusetts 02454, USA; e-mail: mmoore@brandeis.edu.

the Sm core, and three U1-specific proteins, was recently determined to 15 Å resolution (Stark et al., 2001). To date, however, there has been no report of any three-dimensional structure for an intact spliceosome.

One of the challenges inherent to structural studies of spliceosomes is isolating this remarkably dynamic machine in a single state. Recognition and removal of each new intron in vitro appears to require full reassembly of a spliceosome from its component parts. This assembly occurs in a series of distinct steps, primarily defined by the addition and release of the U snRNPs (Fig. 1A). Spliceosome assembly initiates with 5′ splice site recognition by U1 snRNP and subsequent recruitment of U2 snRNP to the branch site to form a species termed CC or E (commitment or early) complex. The U2 snRNP:branch site interaction becomes stabilized with the formation of A complex, and addition of U4, U5, and U6 as a preassembled tri-snRNP results in B complex. Subsequent formation of the catalytically competent C complex involves recruitment of additional protein factors along with significant structural rearrangements that destabilize the association of U1 and U4 snRNPs with the catalytic core (Lamond et al., 1988; Gozani et al., 1994; Staley & Guthrie, 1999). The two chemical steps of splicing, 5′ splice site cleavage and lariat formation followed by 3′ splice site cleavage and exon ligation, occur within C complex (Staley & Guthrie, 1998; Burge et al., 1999).

This article describes our progress in isolating a spliceosomal C complex suitable for future three-dimensional structure determination by single-particle reconstruction of electron microscopic (EM) images. Because of our long-standing interest in the second step of splicing (Anderson & Moore, 1997, 2000; Chen et al., 2000), we decided to focus on C complex, which is the catalytically competent form of the spliceosome and can be

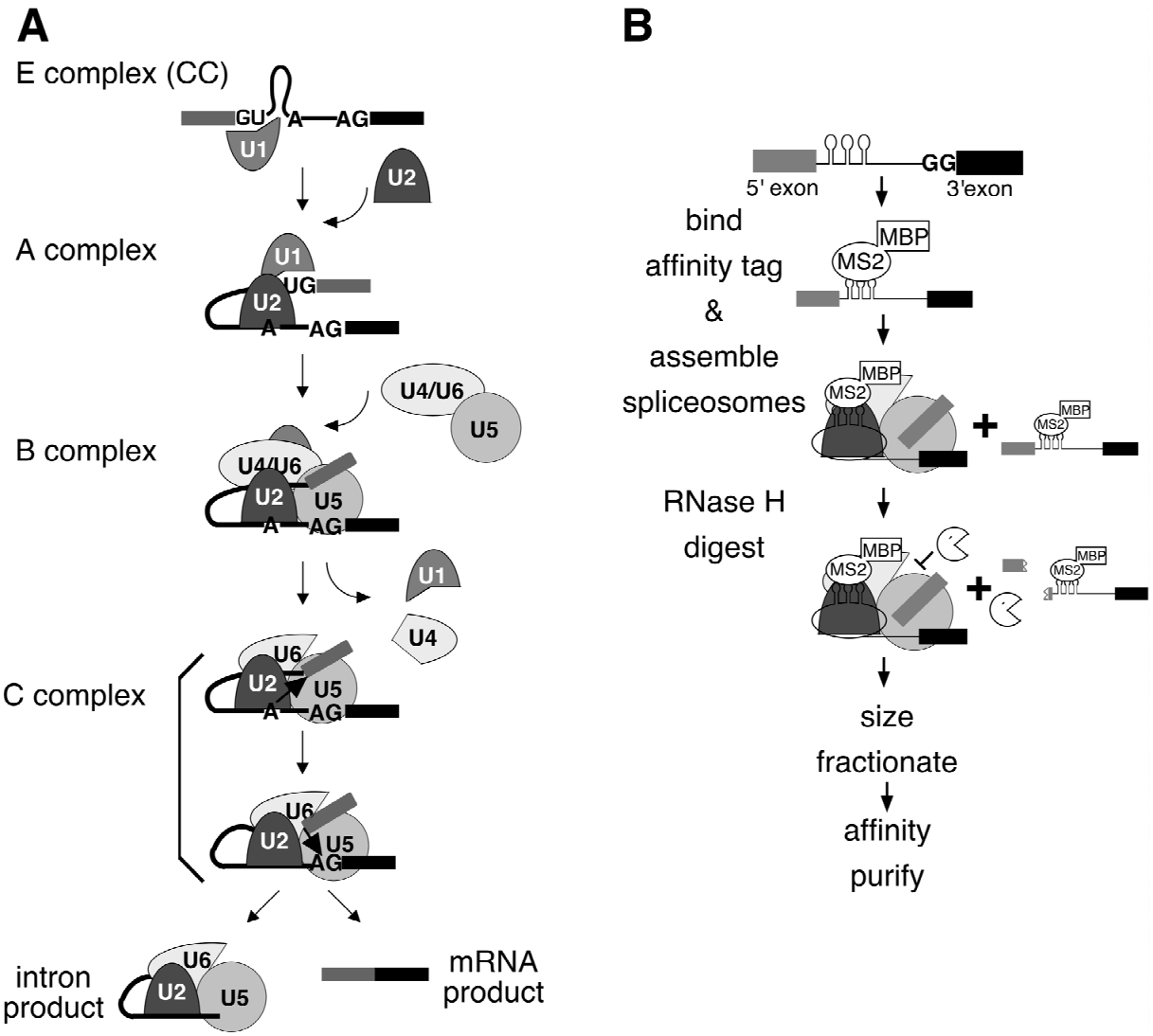

FIGURE 1. A: Complex formation along the splicing pathway (adapted from Moore et al., 1993). **B:** Outline of purification strategy.

readily accumulated on a substrate containing a 3′ splice site mutation. The protocol we developed and present here in detail to isolate this complex has the advantages of small scale (~1 mL splicing reaction) and speed (less than 1 day). Using a combination of size exclusion and affinity chromatographies, we successfully purified native spliceosomes that contain splicing intermediates. As expected for core C complex, our preparation contains U2, U5, and U6 snRNAs, and is essentially free of U1 and U4 snRNAs. Mass spectrometry confirmed the presence of U2 and U5 snRNP proteins, as well as several proteins required for the second step of splicing. However, no proteins specific to earlier splicing complexes were detected in C complex. Electron micrographs contain homogenous and monodisperse particles with dimensions ~270 × 240 Å. Class averages derived from ~4,000 single-particle images display views representative of individual particles. The ability to prepare and image these particles thus represents an important first step toward obtaining a high-resolution, three-dimensional structure of the fully assembled spliceosome.

RESULTS

Purification of intermediate-containing spliceosomes

Isolation of native, C complex spliceosomes required four steps: (1) complex assembly, (2) RNase H digestion, (3) size exclusion, and (4) affinity selection (Fig. 1). Splicing complexes were assembled in HeLa cell nuclear extract on a uniformly labeled T7 pre-mRNA transcript derived from the AdML transcription unit. This construct, which consists of a single intron flanked by two exons, contains an AG-to-GG mutation at the 3′ splice site that blocks splicing prior to the second catalytic step. Thus spliceosomes that have largely completed the first step of splicing accumulate on this substrate (Gozani et al., 1994). Three copies of a binding site for MS2 coat protein were placed in the intron between the 5′ splice site and branchpoint adenosine. When used in conjunction with an MS2-MBP (maltose-binding protein) fusion protein, these sites serve as tags for affinity purification (Das et al., 2000).

To maximize the yield of spliceosomes, we experimented with how best to prepare the MS2-MBP and at what point to bind it to MS2-tagged RNA (data not shown). Although purification on an amylose column of recombinant MS2-MBP expressed in *Escherichia coli* yielded a single band on a Coomassie-stained gel, spectroscopic analysis revealed significant absorbance at 260 nm that could indicate nucleic acid copurifying with the protein. This contaminant not only affected the protein's stability and RNA-binding efficiency, it also inhibited splicing when included with MS2-MBP in splicing reactions. To eliminate this material, we added heparin chromatography as a second purification step. MS2-MBP eluted from a heparin column has an A_{280}/A_{260} ratio of 2.0, is quite stable, and has no adverse effects on splicing efficiency. We obtained optimal yields of spliceosomes by prebinding an excess of this doubly purified MS2-MBP to the pre-mRNA substrate prior to splicing.

After spliceosome assembly, we used RNase H digestion to reduce the amount of residual species preceding C complex. In C complex, the last 30 nt of the 5′ exon are almost fully protected from RNase H digestion, but this region is susceptible to degradation in earlier complexes that have not yet completed lariat formation (Reichert, 2001). Cleavage of the 5′ exon in these earlier complexes facilitates further degradation of unspliced pre-mRNA by endogenous exonucleases, thereby greatly increasing the fraction of complexes containing splicing intermediates (as determined by the ratio of 5′ exon plus lariat intermediate to pre-mRNA and digestion products; Fig. 2A, compare the 60- and 80-min time points). Typically, after the pre-mRNA substrate was incubated under splicing conditions for 60 min and then subjected to RNase H digestion for an additional 20 min, approximately 70% of the labeled RNA remaining had undergone lariat formation.

Following RNase H digestion, splicing reactions were applied to a small (5 mL) S-400 sizing column. Whereas the bulk of nuclear proteins and unbound MS2-MBP were retained on this column, splicing intermediates eluted in the void volume (Fig. 2B). In comparison to previous protocols for fractionating spliceosomal complexes on large S-500 gel filtration columns (Reed et al., 1988; Das et al., 2000), we found that size exclusion chromatography is a faster and simpler method for separating spliceosomes from the bulk of proteins in nuclear extract and excess unbound MS2-MBP without significant dilution of the complexes. Pooled splicing intermediate-containing fractions from the S-400 void volume were subsequently applied to a small (250 μL) amylose column. After extensive washing, MBP-tagged splicing complexes were eluted with maltose.

By scaling down the affinity column, spliceosomes can be eluted at a high enough concentration for structural and biochemical analysis without further concentration. Starting with a 1-mL splicing reaction containing 10 nM pre-mRNA substrate, this purification procedure generally yielded a peak elution fraction (100 μL) containing 5- to 8-nM splicing intermediates. The 5′ exon and lariat intermediate were present in equal amounts, representing 75–85% of total labeled RNA (Fig. 2A). SDS-PAGE analysis of purified complexes (Figs. 2B and 4C) exhibited a complexity in line with the number of polypeptides (~70) previously shown to purify with C complex under conditions that only permitted elution by denaturation (Gozani et al., 1994; Neubauer et al., 1998).

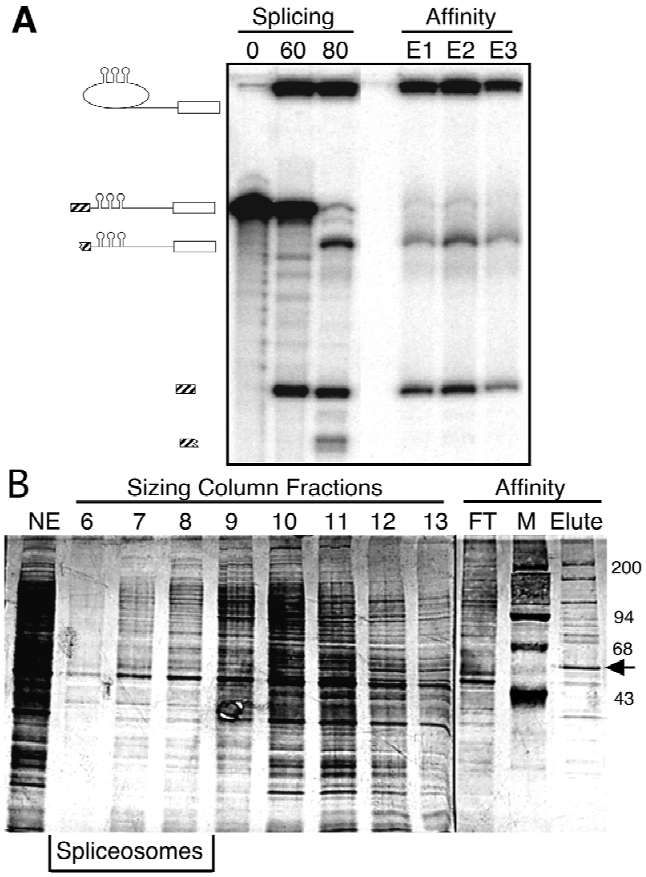

FIGURE 2. A: Denaturing gel (15% polyacrylamide) analysis of the substrate RNA during purification. Shown are a splicing reaction (0- and 60-min time points), RNase H digestion (80-min time point), and affinity selection elutions (E1–E3). The positions of MS2-tagged pre-mRNA, lariat intermediate, 5′ exon, and RNase H digestions products are indicated. The presence of 5′ exon in quantities equimolar to the tagged lariat intermediate throughout the purification indicates the presence of intact C complex. **B:** Silver stained SDS-PAGE (10% polyacrylamide) analysis of proteins during purification. Most proteins from pooled size exclusion spliceosome fractions (6–8) flow through the affinity column, whereas a relatively simple banding pattern is present in the affinity elution. The arrow indicates MS2-MBP.

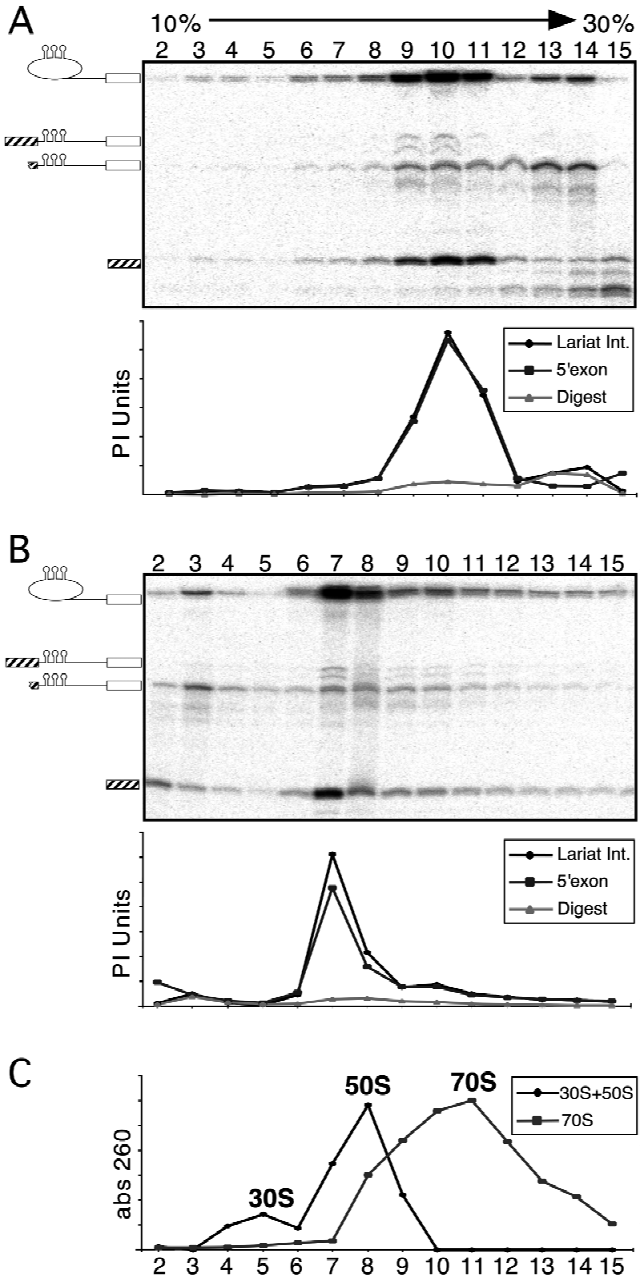

FIGURE 3. A: Glycerol gradient fractionation (10–30%) of a splicing reaction. RNA from gradient fractions was analyzed by denaturing gel (15% polyacrylamide) electrophoresis and quantified with a phosphorimager. The peak fractions containing equal amounts of splicing intermediates (lariat intermediate and 5′ exon) is at fraction 10 and lies between the sedimentation of 70S and 50S ribosomes fractionated in parallel gradients. **B:** Glycerol gradient fractionation of purified C complexes assembled on substrate RNA containing MS2 sites in the intron. The peak of intermediates is at fraction 7, between the 50S and 30S ribosomal subunits. **C:** Fractionation of 70S ribosomes, and 50S and 30S ribosomal subunits in 10–30% glycerol gradients. Fractions were monitored for absorbance at 260 nm.

Characterization of purified C complex

The approximate size and aggregation state of purified C complexes were assessed by glycerol gradient fractionation (Fig. 3). After RNase H digestion, but prior to purification, the majority of splicing intermediates in nuclear extract fractionate as a single peak between 50S and 70S ribosomes. Initially, we performed C complex purification with buffers containing 2 mM magnesium acetate. Under these conditions, affinity-purified intermediate-containing complexes fractionated on glycerol gradients as a wide peak with higher density than 70S ribosomes, and particles appeared aggregated in electron micrographs (data not shown). However, when we eliminated the magnesium and included instead 1 mM EDTA during the size exclusion and affinity steps, the purified complexes fractionated as a single peak between the 30S and 50S ribosomal subunits. This ~40S peak contained equimolar amounts of 5′ exon and lariat intermediates that accounted for 96% of the

total labeled RNA. EDTA has been shown previously not to affect the integrity of preassembled spliceosomes (Abmayr et al., 1988).

The complement of RNAs present in purified C complex were characterized by direct nucleic acid staining and northern blot analysis (Fig. 4A). Direct staining of the purified complexes with SYBR-gold revealed six strong and three weaker bands (Fig. 4A, lane *). Autoradiography showed that three of the strong bands corresponded to 5′ exon and lariat intermediates, and the RNase H digestion product from unspliced pre-mRNA was one of the weak bands (Fig. 4A, prehybridization lane C). Northern blotting with probes against the five spliceosomal snRNAs identified the remaining strong SYBR-gold bands as U2, U5, and U6 snRNAs. The other two weak SYBR-gold bands were U1 and U4 snRNAs (Fig. 4A, northern blot, lane C). When fractionated on a glycerol gradient, the C complex peak fraction retained U2, U5, and U6, but contained no detectable U1 or U4 (Fig. 4A, gradient fraction 7). The intensity of U2, U5, and U6 snRNA bands stained with SYBR-gold appeared equal to those of splicing intermediates.

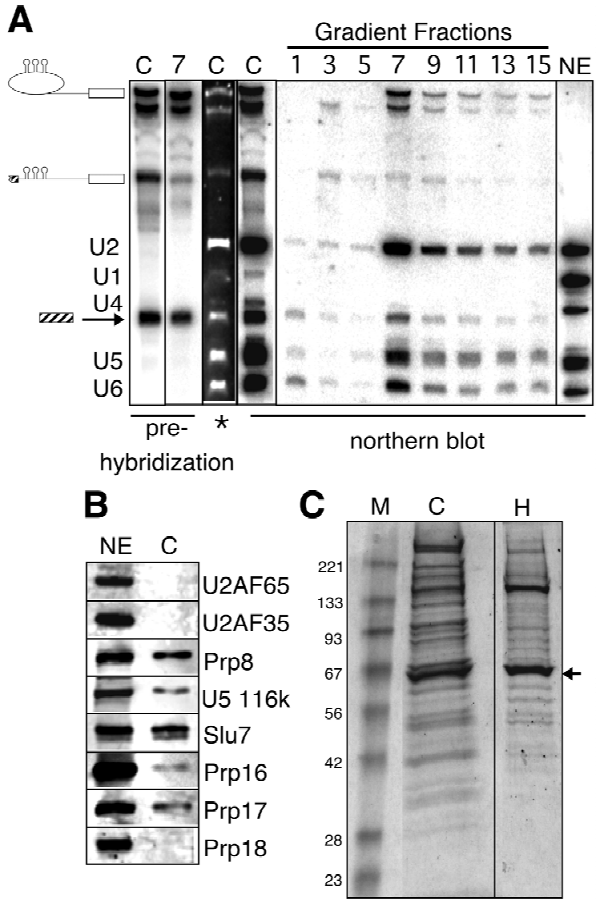

FIGURE 4. A: Northern analysis of purified splicing complexes probing for the five spliceosomal U snRNAs. The position of signal from the ^{32}P body-labeled substrate is indicated by the prehybridization exposure shown for purified C complex and glycerol gradient fraction 7 in the first and second lanes. (The second band just below the lariat intermediate is an exonucleolytic degradation product.) Nuclear extract (NE; last lane) contains all five U snRNAs. Affinity elutions of purified C complex contain primarily U2, U5, and U6 snRNAs. U4 and U1 show a very weak signal in purified C complex preparations, with the U4 band running slightly higher than the 5′ exon. Direct staining of purified C complex in a denaturing gel (10% polyacrylamide) with SYBR-gold (lane 3) reveals the presence of the same RNA species. Probing across a glycerol gradient fractionation of purified C complex reveals a peak at fraction 7 containing only U2, U5, and U6 snRNAs. **B:** Western analysis of purified C complex reveals the presence of Prp8, U5 116kDa protein, Slu7, Prp16, and Prp17. There is no detectable signal for Prp18, U2AF65, or U2AF35. **C:** Coomassie-G stained SDS-PAGE gel (10% polyacrylamide) of C complex and H complex used for mass spectrometry analysis. Results of the analysis are presented in Table 1. The large band present in both complexes is MS2-MBP (indicated by arrow).

As a first pass, western blotting was used to probe the protein complement of purified C complex. This analysis revealed that purified C complexes were essentially free of U2AF65 and U2AF35, proteins found in similarly purified E complex (Das et al., 2000) that are known to leave the spliceosome concomitant with A complex formation (Staknis & Reed, 1994; Fig. 4B). In contrast, purified C complexes did contain significant amounts of the 220 (Prp8) and 116 kDa U5 snRNP proteins, as well as the second step factors Slu7, Prp16, and Prp17. A fourth second-step factor, Prp18, was not detected. This is consistent with previous findings that Prp18 only associates weakly with intermediate-containing spliceosomes (Horowitz & Krainer, 1997).

Mass spectrometry of purified H and C complexes

To identify all the proteins associated with C complex and determine which are specific to splicing, we performed nanoscale microcapillary liquid chromatography tandem mass spectrometry (LC-MS/MS) on purified C and H complexes (Table 1). H complex is the complement of proteins that bind to pre-mRNAs in an ATP- and splicing-independent manner. For our purposes, we assembled H complex on the same MS2-tagged substrate used to accumulate C complex by incubating this substrate under splicing conditions at 4 °C in the absence of ATP. H complex was then purified under conditions identical to that of the C complex. Purified complexes were separated by SDS-PAGE and visualized by Coomassie staining (Fig. 4C). The entire gel lane was excised in 10 discrete sections, subjected to in-gel digestion with trypsin, and analyzed automatically by LC-MS/MS and database searching (Peng & Gygi, 2001).

For each complex, >100 proteins were identified by two or more unique peptides. Numerous small proteins (e.g., SmD3, LSM2, and LSM3) were additionally identified by a single statistically significant peptide. Comparison of the H and C protein lists allowed us to disregard all obvious contaminants (e.g., keratins, tubulins, abundant translation factors, heat shock pro-

TABLE 1. Proteins identified by mass spectroscopy.

Human name	*S. cerevisae* name	SwissProt accession	MW	H[a]	C[a]
Cap binding and hnRNP proteins					
CBP80	Cbc1p	Q01081	92.8	4	12
hnRNP A1		P09651	38.8	5	3
hnRNPA2/B2		P22626	41.4	3	1
hnRNP A3		P51991	39.7	4	2
hnRNP C		P07910	32.3	11	12
hnRNP F		P52597	45.7	4	1
hnRNP G		P38159	47.4	3	2
hnRNP H		P31943	49.2	4	2
hnRNP K		Q07244	51.0	3	2
hnRNP L		P14866	60.2	6	—
hnRNP M		P52272	77.5	9	14
hnRNP R		O43390	70.9	5	6
hnRNP U		Q00839	90.5	13	2
hnRNP RALY		Q9BQX6	32.2	2	4
PAB2	Pab2p	Q15097	32.7	2	1
snRNP core proteins					
SmB/B′	Smb1p	P14678	23.7	2	4
SmD1	Smd1p	P13641	13.3	—	3
SmD2	Smd2p	P43330	13.5	—	3
SmD3	Smd3p	P43331	13.9	1	1
SmE1	Sme1p	P08578	10.8	—	2
SmF1	Smf1p	Q15356	9.7	—	2
SmG1	Smg1p	Q15357	8.5	—	2
LSM2	Lsm2p	Q9Y333	10.8	—	1
LSM3	Lsm3p	Q9Y4Z1	11.8	—	1
snRNP specific proteins					
U1-A	Mud1p	P09012	31.3	1	—
U2-A′	Lea1p	P09661	28.4	1	9
U2-B″	Msl1p	P08579	25.5	—	2
SF3a60	Prp9p	Q12874	58.8	1	5
SF3a66	Prp11p	Q15428	49.2	—	2
SF3a120	Prp21p	Q15459	88.9	3	8
SF3b150	Cus1p	Q13435	97.6	3	8
SF3b130	Rse1p	Q15393	135.6	2	8
SF3b160		O75533	145.8	2	7
U5-220	Prp8p	Q9BW78	273.8	1	39
U5-200	Brr2p	O75643	206.0	6	50
U5-116	Snu114p	Q15029	109.4	10	26
U5-102	Prp6p	O94906	106.9	1	13
U5-100	Prp28p	Q9BUQ8	95.6	1	9
U5-40		O95320	39.3	1	11
First step factors					
U2AF65	Mud2p	P26368	53.5	5	—
ASF/SF2		Q07955	27.7	—	1
PRP19	Prp19p	Q9UMS4	55.2	4	21
Second step factors					
PRP16	Prp16p	Q92620	140.4	—	9
PRP17	Prp17p	O60508	65.5	—	24
SLU7	Slu7p	O95391	68.4	—	9
PRP22	Prp22p	Q14562	139.3	—	29
Other splicing related proteins					
SYF1	Syf1p	Q9HCS7	100.1	—	28
CRN	Clf1p	Q9NYD8	99.2	—	25
CDC5L	Cef1p	Q99459	92.3	6	18
SKIP	Prp45p	Q13573	61.5	2	18
CypE		Q9UNP9	33.4	—	18
SRm300		Q9UHA8	252.0	—	17
PRL1	Prl1p,	043660	57.5	1	16
Dbp1	Prp43p	P42285	92.8	8	8
ISY1	Isy1p	Q9BT05	33.0	1	7
ECM2	Slt11p	Q9NW64	46.9	1	6
SART1	Snu66p	O43290	90.3	—	4
p68	Dbp2p	P17844	69.1	5	4
Y14		Q9Y5S9	19.9	—	3
Srm160		O60585	93.5	—	1
Ref	Yrap1	O43672	26.9	1	1
Magoh		P50606	17.1	—	1
HPRP4	Prp4p	Q13523	58.5	—	1
Gry-Rbp		O60506	69.6	3	1
PTB		P26599	57.2	15	—

Name or homolog	Motifs or localization	SwissProt accession	MW	H[a]	C[a]
Proteins found only in C complex					
—	deadbox	O60306	164.0	—	36
Abstrakt	deadbox	Q9UJV9	69.7	—	23
—	ppiase, WD	Q15002	73.6	—	19
—	MIF4G, MA3	Q9HCG8	108f[b]	—	16
GCIP-IP	nuclear	O95926	28.7	—	13
eIF4a-like	deadbox, nuc	P38919	46.8	—	10
Cactin		Q9BTA6	99f[b]	—	8
—	G-patch	Q9UBB9	96.8	—	8
NMP-IP-19	deadbox	Q9H4H7	78.9	—	7
—		Q9BTZ4	52.4f[b]	—	6
Pinin	nuclear	Q99738	84.6	—	6
—	WD	Q9BRX9	34.3	—	6
HSPC148		Q9P013	26.6	—	6
—	G-patch	Q9BRR8	103.3	—	5
NY-CO-10	ppiase, NLS	O60529	42.8	—	5
FRG1	lys-rich	Q14331	29.2	—	5
—	PSP	Q9NSS2	51.1	—	5
PPIL1	ppiase	Q9Y3C6	18.2	—	5
PPIL3b	ppiase	Q9H2H8	18.2	—	5
Proteins found only in H complex					
Ku70	DNA helicase	P12956	69.7	7	—
TopIIa	DNA helicase	P11388	174.4	6	—
—	G-patch	O60369	101f[b]	5	—

[a]H, C**:** The number of unique peptides identify the protein in these complexes.
[b]f: fragment molecular weight.

teins, etc.), and revealed a few proteins specific to H complex as well as a large number specific to C complex. All detected proteins previously linked to splicing and/or RNA metabolism are listed in Table 1 along with the number of unique peptides identified for each. Also shown are proteins specific to each complex that were identified by five or more unique peptides. Although the number of peptides loosely correlates to both protein size and relative stoichiometry, these numbers should not be interpreted as strictly quantitative. References for each protein can be accessed via the SwissProt Database number (Bairoch & Apweiler, 2000).

Proteins shared between the H and C complexes included the large subunit (CBP80) of the nuclear cap-binding complex and numerous hnRNP proteins. Additional proteins detected in C complex included all of the core snRNP Sm proteins and two LSM proteins, as well as all of the U2- and U5-specific snRNP proteins (with the exception of U5-15, which, like the other LSMs, may have been missed because of its small size). Some snRNP proteins were also detected in H complex, although by fewer peptides. This is consistent with northern blots showing that AdML H complex does contain some U2 and U5 snRNA, but at ~10-fold lower levels than in C complex (data not shown). An ATP-independent association of U2 and U5 snRNPs with H complex was also reported previously (Chabot et al., 1985; Das et al., 2000). Other splicing-related proteins identified in both complexes include Prp19 (Cheng et al., 1993), CDC5 (Burns et al., 1999), SKIP (Neubauer et al., 1998), PRL1 (Neubauer et al., 1998), DBP1 (Arenas & Abelson, 1997), ISY1 (Dix et al., 1999), ECM2 (Xu et al., 1998), p68 (Neubauer et al., 1998), REF/Aly (Zhou et al., 2000), and Gry-Rbp (Neubauer et al., 1998).

Proteins previously linked to splicing that were found in C complex, but not in H, included the second-step splicing factors Prp16, Prp17, Slu7, and Prp22 (Umen & Guthrie, 1995), as well as CRN (Chung et al., 1999), SRm160, SRm300 (Blencowe et al., 1998), Cyclophilin E (Seghezzi et al., 1998), the human homolog of Syf1 (Ben-Yehuda et al., 2000), SART1 (Makarova et al., 2001), Y14, and MAGOH (Le Hir et al., 2000, 2001a; Kataoka et al., 2001; Kim et al., 2001). Proteins identified in C complex only that were not previously linked to splicing include a number of DEAD box proteins (O60306, Abstrakt, the nuclear eIF4a-like protein, and NMP-IP-19), *cis–trans* prolyl-isomerases (Q15002, PPIL3b, and PPIL1), and several hypothetical proteins. H complex-specific proteins linked to splicing include U1A, U2AF65, and PTB/hnRNP I (Burge et al., 1999). Other proteins identified only in H complex include hnRNP L, DNA helicases Ku70 and Top IIa, and hypothetical protein O60369.

Electron microscopy of purified C complexes

Electron microscope images of C complexes negatively stained with uranyl acetate revealed mostly monodisperse particles. The most distinct population, which represented the majority, were of homogeneous size (approximately 270 × 240 Å; Fig. 5A). A minor fraction appeared to be significantly smaller in size and may represent complexes associated with the small amount of RNase H-digested substrate remaining after purification. These smaller particles were clearly distinguishable from the majority population and not subjected to further image analysis. Thirty-four separate micrographs each yielded over 100 images of the large particles, and 3,955 of these were used for further image processing. Images were aligned and classified using the program IMAGIC-5 (van Heel et al., 1996) to gen-

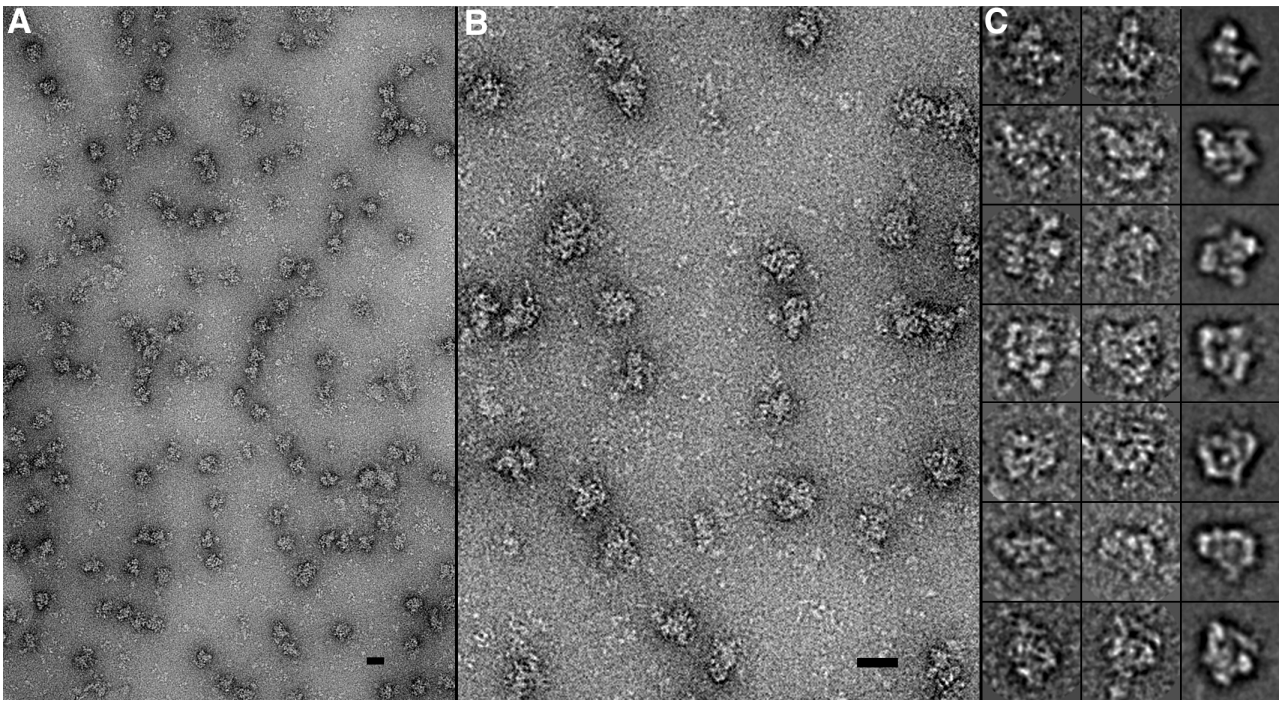

FIGURE 5. A,B: Fields of purified splicing complexes in negative stain at different magnifications. The bar represents 300 Å. **C:** Band-pass filtered images of particles (left images) present in representative class averages (right image).

erate class averages. These averages revealed a number of obviously unique views of a highly asymmetric particle that appears to be composed of several globular domains (Fig. 5C). Comparison of the class averages with individual images shows that the class averages reflect features clearly visible in the raw member images.

To confirm that the large particles do contain RNAs associated with C complex, we performed gold labeling via either a biotinylated oligo complementary to the 5′ exon or an antibody recognizing the tri-methyl guanosine cap structure found in the U2 and U5 snRNAs (Fig. 6). Both of these labels clearly became associated with the large particles. Because the streptavidin gold had a tendency to aggregate, we often observed 1–5 gold clusters labeling a single site on individual spliceosomes. In contrast, because the anti-IgG gold did not aggregate, the majority of spliceosomes labeled via snRNA cap were associated with only one gold cluster.

DISCUSSION

We have succeeded in purifying a homogeneous population of native, intermediate-containing C complex spliceosomes. Our purification procedure yields up to 1 pmol of spliceosomes per 1 mL splicing reaction. This quantity is sufficient to visualize fields of spliceosomes as single particles in electron micrographs that can be used for image processing. Thus we have completed a crucial first step toward three-dimensional structure determination of intact spliceosomes.

An important concern for purifying individual splicing complexes for structural analysis is isolating a single species along the splicing reaction pathway. Because of their transient nature, each form of the spliceosome constitutes only a minor fraction of the total complexes present in a typical splicing reaction (Konarska & Sharp, 1986). We employed two strategies to enrich splicing reactions for C complex. First, we used a substrate containing a 3′ splice site mutation (Gozani et al., 1994) to accumulate C complex on ~45% of the total pre-mRNA substrate in splicing reactions. Subsequent degradation of the 5′ exon with RNase H in complexes preceding C, followed by size exclusion and affinity purification, increased this percentage to 75–85%. When applied to a glycerol gradient, our preparation yielded a single major peak (Fig. 3B) and 96% of the labeled RNA in this peak constituted intact splicing intermediates. When examined by electron microscopy, the majority of particles in this preparation were of uniform size, and this population became labeled with gold attached to species that bound both the 5′ exon and the snRNA cap. We conclude that this major population of particles is C complex.

In nuclear extracts, C complex fractionates in glycerol gradients at a size previously characterized as 60S

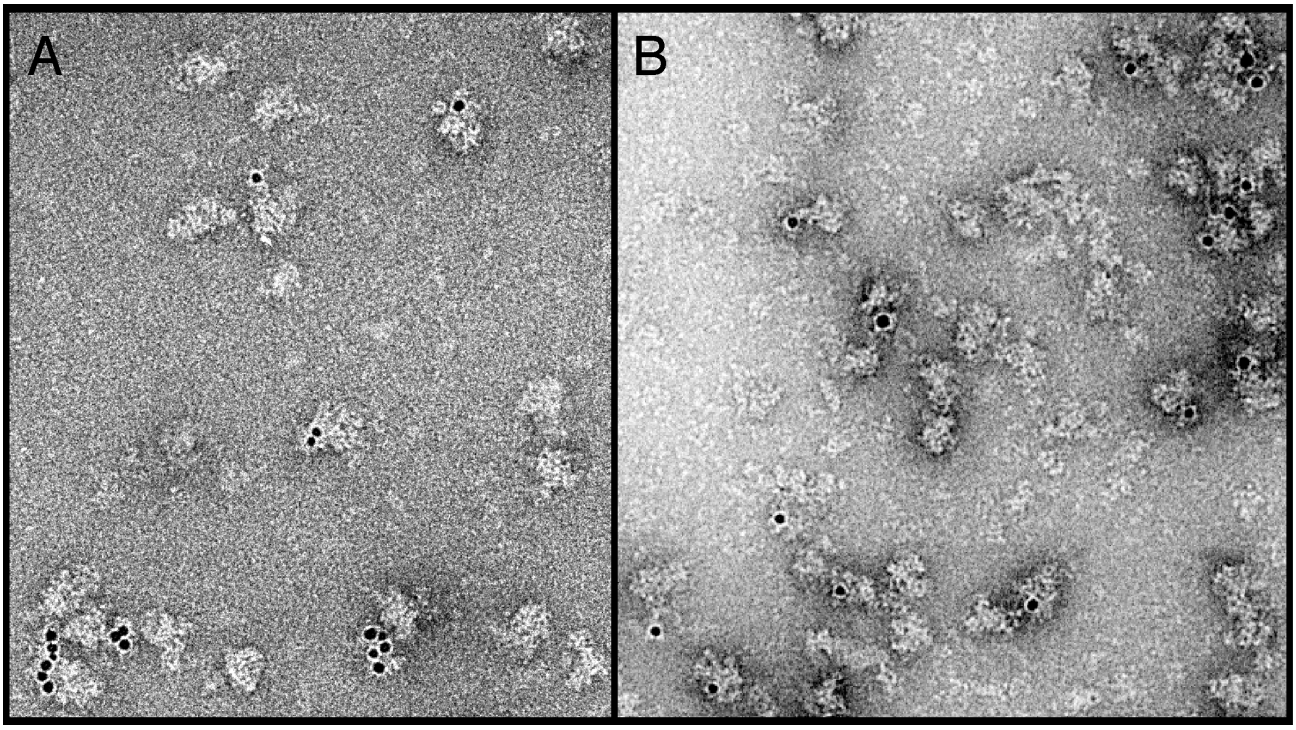

FIGURE 6. Gold labeling of spliceosomes. **A:** Labeling directed to the 5′ exon of the pre-mRNA substrate via a biotinylated oligo and streptavidin-conjugated gold (5 nm). **B:** Labeling directed to the tri-methyl guanosine cap of the U2 and U5 snRNAs via a monoclonal antibody and anti-IgG-conjugated gold.

(Grabowski et al., 1985). However, our purified C complexes appear to be smaller, fractionating between 30S and 50S ribosomes. Because splicing intermediates are present at equimolar ratios in our purified C complexes, and the U2, U5, and U6 snRNAs are also present at similar levels, it seems unlikely that this apparent size reduction is due to any large-scale dissociation of the complexes. Therefore, the size reduction in going from crude nuclear extract to a purified population most likely reflects loss of some peripheral species during purification. Such species might include SR proteins and RNA Pol II, which were not detected in our mass spectrometry analysis of purified C complex, but are known to be associated with splicing complexes in nuclear extract (Mortillaro et al., 1996; Hirose & Manley, 2000). It is also possible that the EDTA treatment we employed to obtain monodisperse particles contributed to the loss of some proteins present in 60 S spliceosomes. We are currently analyzing the protein composition of C complexes purified in the absence of EDTA to address this issue.

Analysis of C complex by mass spectrometry was made possible by direct analysis of complex peptide mixtures. Rather than focus on individual proteins, the entire complex (gel lane) was excised as 10 discrete pieces. Following in-gel digestion, the resulting peptide mixtures were subjected to sequence analysis by LC-MS/MS techniques and database searching (Peng & Gygi, 2001). Interpretation of the large data set was performed computationally. Using this approach, there is a high confidence level for proteins identified by two or more unique, statistically significant peptides. However, smaller proteins yield fewer unique peptides and are therefore more difficult to detect. For those small proteins reported here as being represented by a single peptide (e.g., SmD3, LSM2, and LSM3), assignments were confirmed manually by close inspection of the identifying tandem mass spectrum. To further assess which proteins were specific to C complex, we also determined the complement of proteins that our tagged pre-mRNA substrate in nuclear extract in the absence of splicing (H complex).

Our C complex preparation contained nearly all the proteins expected for core spliceosomes. These include all of the Sm proteins, two U6 snRNP-specific LSM proteins, and all of the known U2 and U5 snRNP-specific proteins (with the exception of U5-15, which likely escaped detection because of its small size). Importantly, C complex but not H complex contained several proteins known to function specifically in the second step of splicing [Prp16, Prp17, Slu7, and Prp22 (Umen & Guthrie, 1995)]. Absence of a fifth second-step factor, Prp18, can be attributed to its previously demonstrated low affinity for intermediate-containing spliceosomes (Horowitz & Krainer, 1997). However, another protein previously shown to associate stably with A/B and C complexes, PSF (Gozani et al., 1994), was also undetectable in our purified C complexes, as were some of the known tri-snRNP proteins. It is certainly possible that some proteins require an intact 3′ splice site AG or the presence of magnesium for efficient incorporation or retention.

Numerous other proteins previously linked to splicing were also detected in our C complex preparation. Five of these (Syf1, CRN, CypeE, SRm300, and SART1; for references, see Results) were found in C but not H complex. Several members of the exon-junction complex (Y14, Srm160, REF, and Magoh) recently shown to assemble on mRNAs as a consequence of splicing were also found (Le Hir et al., 2000, 2001a; Zhou et al., 2000; Kataoka et al., 2001; Kim et al., 2001). A surprise was the large number of proteins with no previously characterized role in splicing that were identified in C complex only. A number of these contain recognizable sequence motifs, such as the DEAD box associated with helicase function and signature sequences for cyclophilins that often have prolyl-isomerase (ppiase) activity. Some of the proteins in this category are known to exhibit nuclear localization, but many have neither a demonstrated expression pattern nor any identified function. It is likely that some of these proteins represent the C complex-specific proteins previously described but not identified by Gozani et al. (1994).

As expected, many hnRNP proteins (Bennett et al., 1992b; Calvio et al., 1995) as well as the nuclear cap binding protein (CBP80) were present in both H and C complexes. H complex also contained a number of U1, U2, and U5 snRNP proteins. The presence of U1 and U2 snRNP components could indicate that the splicing-specific E complex assembled on a fraction of the substrate in our H complex preparation. Like H, E complex forms in the absence of ATP, but E assembles more efficiently at 30 °C (Das et al., 2000). It has been further reported that U5 snRNP exhibits some ATP-independent association with the 3′ splice site of pre-mRNA at 4 °C (Chabot et al., 1985). Interestingly, the pattern of unique peptides identifying U5 snRNP components differed significantly between the H and C complexes. Whereas U5-116 was readily identifiable by multiple unique peptides in both complexes, only a single peptide was detected for the much larger Prp8 protein in H complex compared to 39 Prp8 peptides in C complex. This could potentially indicate that some portion of U5-116 in nuclear extract is not strongly associated with other U5 components.

Two proteins that recognize the polypyrimidine tract prior to spliceosome assembly, U2AF65 and PTB, were present exclusively in H complex. The absence of these proteins in C complex, as well as most other proteins involved in the early spliceosome assembly (e.g., U1 snRNP components and U2AF35) further underscores the validity of our C complex preparation (see below).

It is of interest to compare the proteins we identified here with those identified in previous affinity purifica-

tions of splicing-related complexes (Table 2). Bennett et al. (1992a) first identified proteins associated primarily with B complex based on two-dimensional gel electrophoresis and western analysis. Of 32 spliceosomal-associated proteins (SAPs) observed in those two-dimensional gels, reagents were available to identify 21. Mass spectrometry analysis of two-dimensional gel spots from a similar preparation (primarily B complex) later identified 19 additional proteins associated with spliceosomes (Neubauer et al., 1998).

TABLE 2. Comparison of purifications of spliceosomal complexes.

Human	*S. cerevisiae*	1	2	3	4	5	6	7
snRNP core proteins								
SmB/B′	Smb1p	○	●	●	●	●		●
SmD1	Smd1p				●	●	●	●
SmD2	Smd2p				●	●	●	●
SmD3	Smd3p	○			●	●	●	○
SmE1	Sme1p				●	●		●
SmF1	Smf1p				●	●		●
SmG1	Smg1p				●	●		●
LSM2	Lsm2P				●	●		○
LSM3	Lsm3p							○
LSM4	Lsm4p				●	●		
LSM5	Lsm5p				●	●		
LSM6	Lsm6p				●	●		
LSM7	Lsm7p				●			
LSM8	Lsm8p				●	●		
U1 snRNP proteins								
U1-70	Snp1p		●	●				
U1-A	Mud1p	○	●	●			●	
U2 snRNP proteins								
U2-A′	Lea1p	○	●	●			●	●
U2-B″	Msl1p		●	●				●
SF3a60	Prp9p	○	●	●				●
SF3a66	Prp11p		●	●				●
SF3a120	Prp21p	●	●				●	●
SF3b53	Hsh49p		●	●			●	●
SF3b150	Cus1p	●	●	●			●	●
SF3b130	Rse1p	●	●	●				●
SF3b160		●	●				●	●
U5 snRNP proteins								
U5-220	Prp8p	○			●	●		●
U5-200	Brr2p	●	●		●	●		●
U5-116	Snu114p	●	●			●		●
U5-102	Prp6p	○	●			●		●
U5-100	Prp28p	○	●	●				●
U5-40		○	●	●				●
U5-15	Dib1p				●	●		
tri-snRNP snRNP proteins								
HPRP3	Prp3p		●	●	●	●		
HPRP4	Prp4p		●	●	●	●		○
SART1	Snu66p				●	●		●
PRP31	Prp31p				●	●		
SNU13	Snu13p				●	●		
	Snu23p				●	●		
	Prp38p				●			
	Spp381p				●			

Human	*S. cerevisiae*	1	2	3	4	5	6	7
First-step factors								
U2AF65	Mud2p	●	●	●				
U2AF35				●				
SF1	Bbp			●				
ASF/SF2							●	○
PRP19	Prp19p	●		●			●	●
Dbp1	Prp2	●						●
Second-step factors								
PRP16	Prp16p							●
SLU7	Slu7p							●
PRP17	Prp17p							●
PRP22	Prp22p							●
Other splicing related proteins								
PTB		●	●					
PSF		○	●					
CRN	C1f1p							●
SKIP	Prp45p	●		●				●
p68	Dbp2p	●		●				●
CDC5L	Cef1p	●		●			●	●
PRL1	Prl1p	○					●	●
SPF27	Nuf1			●			●	
ISY1	Isy1p	○						●
SYF1	Syf1p							●
Gry-Rbp		●		●				○
BCAS2			●					
CBP80	Cbc1p	●						●
PAB2	Pab2p	●		●				○
SRm160							●	○
REF	Yra1p	○		●				○
Y14								●
SPF45				●				
SPF30				●				
p54nrb				●				
RPS4	Rps4p			●				
O00302				●				
CA150				●				
F23858				●				
WDC146				●				
HSPC148							●	●
SC35							●	
CCAP1	Ssa2p						●	
CCAP4							●	
CCAP5							●	
CCAP6	Gcn1p						●	
CCAP7							●	
CCAP8							●	
NF45		●					●	
TOP IIa		●					●	

1. H-complex (this study).
2. A/B-complex (Bennett et al., 1992a).
3. A/B-complex (Neubauer et al., 1998).
4. *S. cerevisiae* tri-snRNP (Gottschalk et al., 1999).
5. *S. cerevisiae* tri-snRNP (Stevens & Abelson, 1999).
6. Human CDC5 complex (Ajuh et al., 2000).
7. C-complex (this study).

○: proteins identified in this study by a single unique peptide only.

The complexes characterized by Bennett et al. and Neubauer et al. both contained a number of proteins involved in early spliceosome assembly, but not found in our C complex preparation. These include some U1 snRNP proteins and first-step factors (see above). B and C complexes do share the core U2 and U5 snRNP proteins, as well as Prp19, SKIP, p68, CDC5, and PRL1. However, a number of uncharacterized proteins found by Neubauer et al. were not identified in C complex.

Several proteins shared between B and C complexes were also found associated with the tri-snRNP (U4/U6.U5) from *Saccharomyces cerevisiae* by mass spectrometry (Gottschalk et al., 1999; Stevens & Abelson, 1999). These studies found the majority of Sm, LSM, and U5 snRNP proteins, as well as several new tri-snRNP proteins. One of these, SART1, is present in C complex but was not identified in B complex. The tri-snRNP associated proteins Prp3 and Prp4 are found in B complex, but Prp4 was identified by only a single peptide in C complex. A U4/U6 associated protein, Prp4 has been shown to join the spliceosome as part of the tri-snRNP, but it appears to depart with U4 snRNA during the transition to C complex (Ayadi et al., 1997).

Also listed in Table 2 are proteins associated with the CDC5 protein, first identified as a novel spliceosome component in the mass spectrometry analysis of B complexes (Neubauer et al., 1998). This protein has also been shown to associate with a 40S complex from *Schizosaccharomyces pombe* where it is essential for pre-mRNA splicing (Burns et al., 1999; McDonald et al., 1999). The 40S *S. pombe* complex also contains homologs to U5-220 (Prp8) and U5-116, as well as SmD2, Prp19, Prp5, Syf1, and Ecm2 (Slt11). Like C complex, the 40S *S. pombe* complex contains U2, U5, and U6 snRNAs, but no U1 or U4 (McDonald et al., 1999). Mass spectrometry of immunopurified CDC5 complex from HeLa cells identified Sm proteins and U1 and U2 snRNP proteins, as well as a number of other splicing-associated proteins (Ajuh et al., 2000). In the mammalian system, no U5 or tri-snRNP proteins were identified, although splicing intermediates were immunoprecipitated with CDC5. Of the proteins associated with mammalian CDC5 that had not been previously linked to splicing, only one, HSP148, was also found in our C complex preparation.

In electron micrographs, we observe that C complex particles are large asymmetric structures, with average dimensions of ~270 × 240 Å. This is comparable in size to the 70S ribosome, with dimensions of ~210 × 210 Å (Moore, 1998). Both of these RNA–protein assemblies have estimated molecular weights on the order of 2.5 MDa, so the spliceosome particles are of appropriate dimensions. It should be noted that our C complex particles differ in size and appearance from previous images of splicing complexes visualized by rotary shadowing (Reed et al., 1988). Those particles, isolated by size fractionation of a splicing reaction, had dimensions of 400–600 Å and a more lobed appearance.

On the support grid used for negative staining, C complex particles adopt a variety of orientations that are best categorized by 30 to 50 classes. Class averages derived from ~4,000 images represent different views of the particles and reveal distinct structural features. These features suggest that the particles are composed of globular domains, which may be indicative of snRNP subunits. However, labeling of individual spliceosomal components will be necessary to confirm this. We were able to label the spliceosomes with large gold clusters easily visible in negative stain using both streptavidin-conjugated gold with a biotinylated oligo complementary to the 5′ exon of the pre-mRNA substrate and anti-IgG-conjugated gold with an antibody to the tri-methyl cap structure of U2 and U5 snRNAs (Fig. 6). With both labels, the large gold clusters were almost exclusively associated with the large particles we identified as C complex. However, the large diameter of the gold clusters, although easy to visualize in negative staining, made it difficult to localize the binding sites of their conjugates.

Due to their large size and the relatively limited quantities that can be obtained, electron microscopy is well suited for initial structural analysis of the complexes we have purified. Decades of structural work on the ribosome demonstrate the power of electron microscopy to determine the ultrastructure of large complexes, as well as yield detailed mechanistic information about the translation elongation cycle through analysis of distinct subcomplexes (Agrawal & Frank, 1999). Our purification and initial images of spliceosomal C complex represent an important first step in pursuing a similar structural analysis of the pre-mRNA splicing cycle.

MATERIALS AND METHODS

Substrate and affinity tag

Substrate pre-mRNAs were derived from an AdML gene construct (HMS389) generously provided by Robin Reed. This construct contains an extended polypyrimidine tract and an AG→GG 3′ splice site mutation (Gozani et al., 1994; Anderson & Moore, 1997) and three MS2 sites in the 3′ exon. We changed the branch site sequence to the *S. cerevisiae* UACUA**A**C consensus, which significantly increased first-step splicing efficiency (Reichert, 2001). We also moved the three MS2 binding sites to the *Xho*I site between the 5′ splice and branch sites in the intron to create the construct (pMJM273) used in the experiments presented here. Uniformly [^{32}P]-labeled, GpppG-capped MJM273 splicing substrate was synthesized by T7 runoff transcription followed by gel purification. For use as an affinity tag, an MS2-MBP fusion protein (construct a gift from Josep Vilardell) was expressed in *E. coli*. This fusion places MBP N-terminal to MS2, and the MS2 portion carries a double mutation (V75Q and A81G) that prevents oligomerization (LeCuyer et al., 1995).

The fusion protein was purified first over an amylose (NEB) column in 20 mM HEPES, pH 7.9, 200 mM KCl, 1 mM EDTA, and then over a heparin column using a KCl gradient in the same buffer. Only protein that bound and eluted from the heparin column was used for subsequent affinity purification of spliceosomes.

Splicing and purification

For C complex, the MJM273 substrate was preincubated with a 50- to 100-fold molar excess of purified MS2-MBP prior to splicing. Splicing reactions (usually 1.0 mL) contained 10 nM pre-mRNA, 80 mM potassium glutamate, 2 mM magnesium acetate, 2 mM ATP, 5 mM creatine phosphate, 0.05 mg/mL tRNA, and 40% HeLa cell nuclear extract (Dignam et al., 1983; Abmayr et al., 1988; Reichert & Moore, 2000). C complexes were assembled by incubation at 30 °C for 60 min. Reactions were then supplemented with two 12-nt DNA oligonucleotides (500 nM each) complementary to 5′ exon positions −6 to −30 relative to the 5′ splice site to induce cleavage of the substrate in non-C complexes by endogenous RNase H. After an additional 20 min at 30 °C, reactions were treated with 0.5 mg/mL heparin for 5 min at 30 °C prior to loading onto a 5-mL S-400 sizing column (Amersham Pharmacia) equilibrated with sizing column buffer (SCB1: 250 mM KCl, 5 mM EDTA, 1 mM DTT, 20 mM Tris, pH 7.9) plus 0.05% NP-40 at room temperature. Pooled spliceosome-containing fractions from the S-400 void volume were subsequently loaded onto a 250-μL amylose column equilibrated with SCB1 alone (i.e., no NP-40) at 4 °C. The amylose column was washed with 10 mL SCB1, and then complexes eluted with SCB1 containing 10 mM maltose. Aliquots taken throughout the purification were either subjected to SDS-PAGE followed by silver staining, or extracted with phenol/chloroform, ethanol precipitated, and electrophoresed through a 15% denaturing polyacrylamide gel that was visualized with a phosphorimager (Molecular Dynamics). H complex was purified under identical conditions except that the splicing reactions lacked ATP and creatine phosphate, were incubated on ice for 30 min, and were not subjected to oligo-mediated RNase H digestion.

Glycerol gradient fractionation

Aliquots from splicing reactions or affinity elutions were loaded on 10–30% glycerol gradients in SCB1. Gradients were centrifuged for 2.5 h at 38,000 rpm in a SW55 (Beckman) rotor and then divided into 15 fractions. Labeled RNAs were extracted as above and separated on 15% denaturing polyacrylamide gels. *E. coli* ribosomes (a kind gift from Joanne Chapple, Cubist Pharmaceuticals, Inc.) were fractionated on similar gradients in either 2 mM (to separate the 50S and 30S subunits) or 20 mM magnesium acetate (where 70S ribosomes remain intact). Ribosomal subunit peak positions were determined by monitoring the A_{260} of gradient fractions.

Northern analysis and RNA staining

For northern blots, 50 fmol of purified complexes were electrophoresed through a denaturing 10% polyacrylamide gel, transferred to a nylon membrane (ICN), and hybridized with [^{32}P]-labeled probes complementary to the U snRNAs (Konarska & Sharp, 1987). For direct RNA staining, 150 fmol of purified complexes were electrophoresed through a denaturing 10% polyacrylamide gel, and the gel stained with SYBR-gold (Molecular Probes).

Antibodies and western blots

For western blots, 30 fmol of complexes were separated on a 5–15% gradient SDS-PAGE, transferred to a nitrocellulose membrane, and probed with the following: α-U2AF65 and α-U2AF35 (rabbit polyclonal from T. Maniatis; Zuo & Maniatis, 1996), α-U5 116k (rabbit polyclonal from R. Lührmann; Fabrizio et al., 1997), α-hSlu7, α-Prp16, α-Prp17 (rabbit polyclonal from R. Reed; Zhou & Reed, 1998; Chua & Reed, 1999), α-Prp18 (rabbit polycolonal from D. Horowitz; Horowitz & Krainer, 1997), and α-Prp8 (rabbit polyclonal; Luo et al., 1999).

Mass spectrometry

For peptide sequence analysis by mass spectrometry, 1 pmol of purified complexes were separated by 10% SDS-PAGE and stained with Coomassie-G. The gel lane was cut into 10 sections, and each section subjected to in-gel trypsin (Promega) digestion overnight (16 h). Sections were individually extracted using three aliquots of 50% acetonitrile:5% formic acid and transferred to autosampler vials and dried. Each sample was analyzed by automated nanoscale microcapillary reverse-phase chromatography/electrospray ionization-ion trap mass spectrometry (L. Licklider, in prep.). Overall more than 10,000 sequencing attempts for each complex were acquired automatically and matched to peptides from a nonredundant human database by the Sequest algorithm (Eng et al., 1994). Peptide candidates returned by Sequest were filtered by established criteria to determine successful matches (Gygi et al., 2002).

Electron microscopy

Concentrations of spliceosomes in purified fractions were determined from the specific activity of labeled splicing intermediates. For EM, 3 μL of 5 nM complexes was applied to carbon-coated grids and stained with 2% uranyl acetate solution. Samples were analyzed in a Philips CM12 microscope operating at 120 kV. Micrographs were exposed at 60,000× magnification and digitized with a SCAI scanner (Zeiss) at 7 μm/pixel with subsequent 4 × 4 pixel averaging. Individual particles were picked with the program Ximdisp (Smith, 1999) and separated into classes using the IMAGIC-5 software package (van Heel et al., 1996).

Gold labeling

During a standard purification, an excess of a 15-mer 2′*O*-methyl oligo complimentary to the 5′ exon 68 nt upstream of the 5′ splice site was passed over the amylose affinity column containing bound C complex. The oligo contained three biotin residues. After washing, streptavidin-conjugated 5-nm gold particles (Sigma) were applied to the column. The col-

umn was washed again, and complexes were eluted and prepared for EM analysis as above. In a similar manner, C complexes were also labeled with mouse monoclonal antibodies against the snRNA tri-methyl guanosine cap (Calbiochem) and 5-nm gold particles conjugated to anti-mouse IgG (Sigma).

ACKNOWLEDGMENTS

We thank Jen Hurwitz and Natasha Levin for technical assistance, Robin Reed and Josep Vilardell for supplying reagents as noted in the text, and Hervé Le Hir for discussion. M.J.M and N.G. are assistant investigators with the Howard Hughes Medical Institute; M.S.J. is a Paul Sigler/Agouron Institute Fellow of the Helen Hay Whitney Foundation. This work was supported, in part, by National Institutes of Health Grants GM53007 (M.J.M.), GM62580 (N.G.), and HG00041 (S.P.G.), funding from the Keck Foundation (M.J.M. and N.G.), and the Giovanni Armenise-Harvard Foundation (S.P.G.).

Received January 2, 2002; accepted without revision January 23, 2002

REFERENCES

Abmayr SM, Reed R, Maniatis T. 1988. Identification of a functional mammalian spliceosome containing unspliced pre-mRNA. *Proc Natl Acad Sci USA 85*:7216–7220.

Agrawal RK, Frank J. 1999. Structural studies of the translational apparatus. *Curr Opin Struct Biol 9*:215–221.

Ajuh P, Kuster B, Panov K, Zomerdijk JC, Mann M, Lamond AI. 2000. Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J 19*: 6569–6581.

Anderson K, Moore MJ. 1997. Bimolecular exon ligation by the human spliceosome. *Science 276*:1712–1716.

Anderson K, Moore MJ. 2000. Bimolecular exon ligation by the human spliceosome bypasses early 3′ splice site AG recognition and requires NTP hydrolysis. *RNA 6*:16–25.

Arenas JE, Abelson JN. 1997. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci USA 94*:11798–11802.

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

Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res 28*:45–48.

Ben-Yehuda S, Dix I, Russell CS, McGarvey M, Beggs JD, Kupiec M. 2000. Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in *Saccharomyces cerevisiae*. *Genetics 156*:1503–1517.

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

Bennett M, Pinol-Roma S, Staknis D, Dreyfuss G, Reed R. 1992b. Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol Cell Biol 12*:3165–3175.

Blencowe BJ, Issner R, Nickerson JA, Sharp PA. 1998. A coactivator of pre-mRNA splicing. *Genes & Dev 12*:996–1009.

Burge CB, Tuschl T, Sharp PA. 1999. Splicing of precursors to mRNAs by the spliceosomes. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA world*, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 525–560.

Burns CG, Ohi R, Krainer AR, Gould KL. 1999. Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. *Proc Natl Acad Sci USA 96*:13789–13794.

Calvio C, Neubauer G, Mann M, Lamond AI. 1995. Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. *RNA 1*:724–733.

Chabot B, Black DL, LeMaster DM, Steitz JA. 1985. The 3′ splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. *Science 230*:1344–1349.

Chen S, Anderson K, Moore MJ. 2000. Evidence for a linear search in bimolecular 3′ splice site AG selection. *Proc Natl Acad Sci USA 97*:593–598.

Cheng SC, Tarn WY, Tsao TY, Abelson J. 1993. PRP19: A novel spliceosomal component. *Mol Cell Biol 13*:1876–1882.

Chua K, 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.

Chung S, McLean MR, Rymond BC. 1999. Yeast ortholog of the *Drosophila* crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *RNA 5*:1042–1054.

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

Dignam JD, Lebovitz RM, Roeder RD. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res 11*:1475–1489.

Dix I, Russell C, Yehuda SB, Kupiec M, Beggs JD. 1999. The identification and characterization of a novel splicing protein, Isy1p, of *Saccharomyces cerevisiae*. *RNA 5*:360–368.

Eng JK, McCormack AL, Yates JR. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spec 5*:976–989.

Fabrizio P, Laggerbauer B, Lauber J, Lane WS, Lührmann R. 1997. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J 16*:4092–4106.

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

Gozani O, Patton JG, 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.

Grabowski PJ, Seiler SR, Sharp PA. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell 42*:345–353.

Gygi SP, Rist B, Griffin TJ, Eng J, Aebersold R. 2002. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope coded affinity tags. *J Proteome Res.* In press.

Hirose Y, Manley JL. 2000. RNA polymerase II and the integration of nuclear events. *Genes & Dev 14*:1415–1429.

Horowitz DS, Krainer AR. 1997. A human protein required for the second step of pre-mRNA splicing is functionally related to a yeast splicing factor. *Genes & Dev 11*:139–151.

Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Lührmann R, Li J, Nagai K. 1999. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell 96*:375–387.

Kataoka N, Diem MD, Kim VN, Yong J, Dreyfuss G. 2001. Magoh, a human homolog of *Drosophila* mago nashi protein, is a component of the splicing-dependent exon–exon junction complex. *EMBO J 20*:6424–6433.

Kim VN, Yong J, Kataoka N, Abel L, Diem MD, Dreyfuss G. 2001. The Y14 protein communicates to the cytoplasm the position of exon–exon junctions. *EMBO J 20*:2062–2068.

Konarska MM, Sharp PA. 1986. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell 46*:845–855.

Konarska MM, Sharp PA. 1987. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell 49*:763–774.

Lamond AI, Konarska MM, Grabowski PJ, Sharp PA. 1988. Spliceosome assembly involves binding and release of U4 small nuclear ribonucleoprotein. *Proc Natl Acad Sci USA 85*:411–415.

Le Hir H, Gatfield D, Braun IC, Forler D, Izaurralde E. 2001a. The

protein Mago provides a link between splicing and mRNA localization. *EMBO Rep 2*:1119–1124.
Le Hir H, Gatfield D, Izaurralde E, Moore MJ. 2001b. The exon–exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J 20*:4987–4997.
Le Hir H, Izaurralde E, Maquat LE, Moore MJ. 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. *EMBO J 19*:6860–6869.
LeCuyer KA, Gehlen LS, Uhlenbeck OC. 1995. Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. *Biochemistry 34*:10600–10606.
Luo HR, Moreau GA, Levin N, Moore MJ. 1999. The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA 5*:893–908.
Makarova OV, Makarov EM, Lührmann R. 2001. The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *EMBO J 20*:2553–2563.
McDonald WH, Ohi R, Smelkova N, Frendewey D, Gould KL. 1999. Myb-related fission yeast cdc5p is a component of a 40S snRNP-containing complex and is essential for pre-mRNA splicing. *Mol Cell Biol 19*:5352–5362.
Moore MJ, Query CC, Sharp PA. 1993. Splicing of precursors to mRNA by the spliceosome. In: Gesteland R, Atkins J, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 303–357.
Moore PB. 1998. The three-dimensional structure of the ribosome and its components. *Ann Rev Biophys Biomol Struct 27*:35–58.
Mortillaro MJ, Blencowe BJ, Wei X, Nakayasu H, Du L, Warren SL, Sharp PA, Berezney R. 1996. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci USA 93*:8253–8257.
Neubauer G, King A, Rappsilber J, Calvio C, Watson M, Ajuh P, Sleeman J, Lamond A, Mann M. 1998. Mass spectrometry and EST-database searching allows characterization of the multiprotein spliceosome complex. *Nat Genet 20*:46–50.
Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 372*:432–438.
Peng J, Gygi SP. 2001. Proteomics: The move to mixtures. *J Mass Spectrom 36*:1083–1091.
Price SR, Evans PR, Nagai K. 1998. Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature 394*:645–650.
Raker VA, Hartmuth K, Kastner B, Lührmann R. 1999. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol Cell Biol 19*:6554–6565.
Reed R, Griffith J, Maniatis T. 1988. Purification and visualization of native spliceosomes. *Cell 53*:949–961.
Reichert V. 2001. Structural and functional studies of mammalian splicing complexes. PhD thesis, Biophysics Graduate Program, Brandeis University, Waltham, Massachusetts.
Reichert V, Moore MJ. 2000. Better conditions for mammalian in vitro splicing provided by acetate and glutamate as potassium counterions. *Nucleic Acids Res 28*:416–423.
Seghezzi W, Chua K, Shanahan F, Gozani O, Reed R, Lees E. 1998. Cyclin E associates with components of the pre-mRNA splicing machinery in mammalian cells. *Mol Cell Biol 18*:4526–4536.
Smith JM. 1999. Ximdisp—A visualization tool to aid structure determination from electron microscope images. *J Struct Biol 125*: 223–228.
Staknis D, Reed R. 1994. Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol Cell Biol 14*:2994–3005.
Staley JP, Guthrie C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell 92*:315–326.
Staley JP, Guthrie C. 1999. An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell 3*: 55–64.
Stark H, Dube P, Lührmann R, Kastner B. 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature 409*:539–542.
Stevens SW, Abelson J. 1999. Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc Natl Acad Sci USA 96*:7226–7231.
Umen JG, Guthrie C. 1995. The second catalytic step of pre-mRNA splicing. *RNA 1*:869–885.
van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M. 1996. A new generation of IMAGIC image processing system. *J Struct Biol 116*:17–24.
Vidovic I, Nottrott S, Hartmuth K, Lührmann R, Ficner R. 2000. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell 6*:1331–1342.
Wimberly BT, Brodersen DE, Clemons WM, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature 407*:327–339.
Xu D, Field DJ, Tang SJ, Moris A, Bobechko BP, Friesen JD. 1998. Synthetic lethality of yeast slt mutations with U2 small nuclear RNA mutations suggests functional interactions between U2 and U5 snRNPs that are important for both steps of pre-mRNA splicing. *Mol Cell Biol 18*:2055–2066.
Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science 29*:29.
Zhou Z, Luo MJ, Straesser K, Katahira J, Hurt E, Reed R. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature 407*:401–405.
Zhou Z, Reed R. 1998. Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *EMBO J 17*:2095–2106.
Zuo P, Maniatis T. 1996. The splicing factor U2AF35 mediates critical protein–protein interactions in constitutive and enhancer-dependent splicing. *Genes & Dev 10*:1356–1368.