MOLECULAR AND CELLULAR BIOLOGY, Jan. 2009, p. 281–301 Vol. 29, No. 1
0270-7306/09/$08.00+0 doi:10.1128/MCB.01415-08


# Conservation of the Protein Composition and Electron Microscopy Structure of *Drosophila melanogaster* and Human Spliceosomal Complexes▽†

Nadine Herold,[1] Cindy L. Will,[1] Elmar Wolf,[1] Berthold Kastner,[1] Henning Urlaub,[2] and Reinhard Lührmann[1]*

*Department of Cellular Biochemistry[1] and Bioanalytical Mass Spectrometry Group,[2] MPI for Biophysical Chemistry, D-37077 Göttingen, Germany*

Received 9 September 2008/Returned for modification 14 October 2008/Accepted 22 October 2008

**Comprehensive proteomics analyses of spliceosomal complexes are currently limited to those in humans, and thus, it is unclear to what extent the spliceosome's highly complex composition and compositional dynamics are conserved among metazoans. Here we affinity purified *Drosophila melanogaster* spliceosomal B and C complexes formed in Kc cell nuclear extract. Mass spectrometry revealed that their composition is highly similar to that of human B and C complexes. Nonetheless, a number of *Drosophila*-specific proteins were identified, suggesting that there may be novel factors contributing specifically to splicing in flies. Protein recruitment and release events during the B-to-C transition were also very similar in both organisms. Electron microscopy of *Drosophila* B complexes revealed a high degree of structural similarity with human B complexes, indicating that higher-order interactions are also largely conserved. A comparison of *Drosophila* spliceosomes formed on a short versus long intron revealed only small differences in protein composition but, nonetheless, clear structural differences under the electron microscope. Finally, the characterization of affinity-purified *Drosophila* mRNPs indicated that exon junction complex proteins are recruited in a splicing-dependent manner during C complex formation. These studies provide insights into the evolutionarily conserved composition and structure of the metazoan spliceosome, as well as its compositional dynamics during catalytic activation.**

Pre-mRNA splicing is catalyzed by the spliceosome, a large RNP molecular machine consisting of the U1, U2, U4/U6, and U5 snRNPs plus a multitude of non-snRNP proteins (reviewed in reference 59). Spliceosome assembly is an ordered, highly dynamic process that leads to the stepwise formation of the catalytic site(s) responsible for catalyzing the excision of an intron of a pre-mRNA and ligation of its flanking exons (reviewed in reference 59). Initially, the U1 snRNP interacts with the conserved 5′ splice site (5′ ss) of the pre-mRNA, forming the spliceosomal E complex, and in the next step the U2 snRNP stably interacts with the pre-mRNA's branch site, leading to the A complex (i.e., prespliceosome). The preformed U4/U6.U5 tri-snRNP particle then interacts with the A complex to generate the precatalytic B complex. The latter is converted to the catalytically activated B* complex, a process involving major RNP rearrangements, including destabilization or loss of the U1 and U4 snRNPs. The first step of splicing subsequently ensues and involves a nucleophilic attack of the branch point adenosine at the 5′ ss, generating a cleaved 5′ exon and intron-3′ exon lariat intermediate. The spliceosomal C complex is formed at this time and subsequently catalyzes the second step of splicing, which entails intron excision and concomitant ligation of the 5′ and 3′ exons to form mRNA. The mRNA, in the form of an RNP, is then exported to the cytoplasm for translation by the ribosome.

During spliceosome assembly, a complex RNA-RNA network involving the snRNAs and pre-mRNA is formed (reviewed in references 37 and 51). This RNA-RNA network plays a central role in juxapositing the reactive groups of the pre-mRNA (i.e., the 5′ ss, 3′ ss, and branch site). RNA structures involving U2 and U6 snRNA play crucial roles in the catalytic core of the spliceosome, with nucleotides of U6 directly involved in the catalysis of pre-mRNA splicing (reviewed in reference 55).

Proteins also play an essential role in splicing, both during spliceosome assembly and the catalytic steps of splicing. Recent analyses have demonstrated that the spliceosome is an extremely protein-rich molecular machine (reviewed in references 23 and 59). In humans, more than 50 proteins are recruited to the spliceosome as stable components of the spliceosomal snRNPs. In addition to these proteins, over 100 non-snRNP proteins are associated with human spliceosomes, as determined by mass spectrometry (MS) of affinity-purified spliceosomal complexes (11, 20, 22, 30, 31, 40, 61). MS studies of human spliceosomal complexes isolated at defined assembly/functional stages, including the A, B, and C complexes (3, 5, 11, 20, 22, 30, 31), further revealed that the spliceosome's protein composition is highly dynamic, with a major compositional change occurring during the transition from the precatalytic B complex to the catalytically active C complex. The ability to isolate human spliceosomes has also led to major advances in the elucidation of their structure. Higher-order structures at a resolution of 30 to 40Å have been obtained for affinity-purified human spliceosomal A, B, BΔU1 (i.e., B lacking U1), and C complexes by using single-particle electron microscopy (EM) (3, 7, 11, 24).

The spliceosome assembly pathway and catalytic mecha-

* Corresponding author. Mailing address: Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. Phone: 49-551-2011407. Fax: 49-551-2011197. E-mail: Reinhard.Luehrmann@mpi-bpc.mpg.de.

† Supplemental material for this article may be found at http://mcb.asm.org/.

▽ Published ahead of print on 3 November 2008.

nisms of pre-mRNA splicing are generally conserved among higher and lower eukaryotes. Splice site and branch site consensus sequences are also largely conserved among metazoans, including *Drosophila melanogaster* and humans (29, 35, 48). However, there are some differences between vertebrates and *Drosophila*, including (i) a lower incidence in *Drosophila* of a G nucleotide preceding the branched nucleotide in the branch point sequence (BPS) and (ii) the absence of a polypyrimidine tract between the BPS and 3′ ss in many *Drosophila* introns (35). Furthermore, in contrast to the situation in humans, where more than 90% of introns are longer than 130 nucleotides (nt), a large percentage (54%) of *Drosophila* introns are less than 80 nt in length, with most having lengths from 59 to 67 nt (29, 35). Most of these "short" introns cannot be spliced in mammals, where the minimum intron length required for splicing is ~80 nt (18, 58). In the case of long introns, splice site pairing is thought to be first established across the exon (so-called exon definition) and, at a later stage, cross-intron interactions are established (4, 41). In contrast, the initial pairing of the splice sites of short introns is thought to occur across the intron (52).

Whereas biochemical approaches have been extensively used to identify spliceosomal proteins in mammals, *Drosophila* has been primarily used to study regulated splicing events by using genetic approaches (reviewed in reference 6). Indeed, a large percentage of *Drosophila* pre-mRNAs are subject to alternative splicing (8), and several regulated splicing events in *Drosophila* have been well characterized (reviewed in references 14 and 39). With the completion of the sequence of the *Drosophila* genome, bioinformatics approaches confirmed the presence of *D. melanogaster* homologues of most human snRNP proteins, as well as many non-snRNP spliceosome-associated proteins (36). These proteins are generally highly conserved and more closely related to their human, as opposed to *Saccharomyces cerevisiae* yeast, counterparts (36). Biochemical/genetic approaches have also identified *Drosophila* orthologues of a subset of human splicing factors. However, to date, *Drosophila* spliceosomes have not been affinity purified, nor has their protein composition been determined. Thus, it is not clear to what extent the protein compositions of human and fly spliceosomes are conserved. Likewise, it is not known whether the exchange of proteins, particularly during catalytic activation of the spliceosome, is conserved between these two evolutionarily distant organisms. MS analyses of spliceosomes from *Drosophila melanogaster* might provide insights into the "core proteome" of the metazoan spliceosome.

During splicing, the mRNA is bound by several proteins that affect its subsequent metabolism, including its export from the nucleus, its localization in the cytoplasm, the efficiency of its translation in the cytoplasm, and its susceptibility to nonsense-mediated decay (NMD) (reviewed in reference 34). In vertebrates, a multiprotein complex, termed the exon junction complex (EJC), associates with the mRNA in a splicing-dependent but sequence-unspecific manner approximately 20 to 24 nt upstream of the exon-exon junction, protecting this region against RNase H cleavage (28). The EJC consists of four stably associated core proteins, eIF4AIII, MLN51, Magoh, and Y14, plus a number of factors that bind transiently or less stably (53, 54). *D. melanogaster* homologues of the main components of the human EJC have been identified, and several factors have been studied extensively (16, 19, 38, 49). However, it is not known in the fruit fly whether an EJC is deposited onto the mRNA as a consequence of pre-mRNA splicing.

To learn more about the evolutionarily conserved composition of the spliceosome, as well as the dynamics of spliceosomal protein recruitment/release, we have affinity purified *Drosophila* spliceosomal B and C complexes assembled in Kc cell nuclear extract and determined their composition by using MS. For comparative purposes, we also isolated B complexes formed on an identical substrate (the *fushi tarazu* pre-mRNA) in HeLa nuclear extract. These studies revealed a remarkable conservation of spliceosome-associated proteins in two evolutionarily distant organisms but also identified apparent *Drosophila*-specific splicing factors. Negative-stain EM revealed that the two-dimensional (2D) structure of the *Drosophila* B complex is also highly similar to that of the human B complex, consistent with a conservation of higher-order interactions within both spliceosomes. We also isolated spliceosomal B complexes formed on a short *Drosophila* intron (*Zeste62*) and show that although their protein composition differs only minimally from Ftz B complexes, structural differences are observed under the electron microscope. A comparison of *Drosophila* B and C complexes further revealed that the recruitment and release of factors during the conversion of the catalytically inactive B complex to the catalytically active C complex is also largely conserved between humans and flies. Lastly, we affinity purified spliced and unspliced mRNPs from *Drosophila* Kc cell nuclear extract and determined their protein composition by MS. These studies revealed that, as in humans, components of the EJC are deposited onto the mRNA in a splicing-dependent manner.

## MATERIALS AND METHODS

**Construction of plasmids and in vitro transcription.** PCR products containing a fragment of the *D. melanogaster fushi tarazu* gene consisting of either exon 1, intron 2, and exon 2 (Ftz) or solely exon 1 and exon 2 (FtzΔI) plus three MS2 RNA aptamers at their 3′ end were generated by PCR-based techniques and used as templates for the transcription of Ftz-M3 pre-mRNA or Ftz-M3 mRNA, respectively. Ftz pre-mRNA lacking a 3′ ss and 3′ exon but containing an elongated (29 nt) polypyrimidine tract was generated from the wild-type Ftz pre-mRNA template (pFtz-M3) by PCR-based techniques. The resulting PCR product was then cloned into a plasmid (pUC18-MS2) downstream of three MS2 RNA aptamers, generating pM3-Ftz-longPYT. A fragment of the *D. melanogaster Zeste* gene, comprising part of exon 2, the 62-nt-long intron 2, and a portion of exon 3, was amplified from genomic DNA by PCR using the primers 5′GGGGTACCA ATGATCTGCTGCACTTCAAGACAG-3′ (ZesteForKpn1) and 5′-CGGGATCC GGCGGTAATTGTGGCCATTGG-3′ (ZesteRevBamH1). The resulting PCR product was then cloned into pUC18-MS2 upstream of three MS2 RNA aptamers, generating pZeste62-M3. Uniformly $^{32}$P-labeled m$^7$G(5′)ppp(5′)G-capped pre-mRNAs were synthesized in vitro by T7 runoff transcription in the presence of [$\alpha^{32}$P]UTP (GE Healthcare).

**In vitro splicing.** *Drosophila* Kc cells were cultivated in suspension at room temperature in EX-Cell 420 serum-free medium for insect cells. Kc cell and HeLa nuclear extract were prepared essentially according to the method in reference 12. In vitro splicing reactions contained 40% Kc cell or HeLa nuclear extract in buffer D (20 mM HEPES-KOH, pH 7.9, 20% [vol/vol] glycerol, 100 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1 mM dithioerythritol, 0.5 mM phenylmethylsulfonyl fluoride, and in the case of Kc cells, 0.05% NP-40), 5 mM ATP, 20 mM creatine phosphate, and 4.5 to 7.5 nM of $^{32}$P-labeled pre-mRNA or mRNA (specific activity, ~100 dpm/fmol). The final concentration of $MgCl_2$ was adjusted to 4 mM (Kc) or 3 mM (HeLa), and HEPES-KOH (pH 7.9) was adjusted to 22 mM in both cases. Splicing reaction mixtures were incubated at 22°C (Kc) or 30°C (HeLa). For analysis of splicing, RNA was recovered and separated on an 8 M urea–12% polyacrylamide gel and detected by autoradiog-

raphy. Spliceosomal complexes were analyzed on native 3.5% polyacrylamide gels essentially as previously described (27).

**MS2 affinity purification.** Recombinant MBP-MS2 fusion protein was expressed in *Escherichia coli* and purified essentially as described previously (22). To isolate spliceosomal complexes or mRNPs, pre-mRNA or mRNA was incubated with a 20-fold molar excess of purified MBP-MS2 fusion protein for 30 min on ice prior to addition to the splicing reaction mixture. To isolate B complexes, Ftz-M3 or Zeste62-M3 pre-mRNA was incubated in Kc nuclear extract for 8 or 20 min, respectively, or Ftz-M3 was incubated in HeLa nuclear extract for 6 min under splicing conditions. To isolate C complexes or mRNPs, pre-mRNA or mRNA was incubated for 180 min in Kc nuclear extract under splicing conditions. To isolate C complexes, two DNA oligonucleotides complementary to the Ftz 5′ exon (at nt −6 to −17 [5′-TCGATGTGCGAC-3′] and nt −18 to −30 [5′-GCCAAAGTCTCC-3′] relative to the 5′ ss) or, to isolate mRNPs, DNA oligonucleotides complementary to intronic sequences of Ftz-M3 pre-mRNA (5′-GTTGTTAATCGTGTGT-3′ and 5′-GTAAGCATAAGCAAAG-3′) were added subsequently at a 100-fold molar excess, and the reaction mixture was incubated for an additional 20 min at 22°C to allow cleavage by RNase H present in the extract. In vitro splicing reaction mixtures were then loaded onto linear 10-to-30% (vol/vol) glycerol gradients containing GP150 buffer (20 mM HEPES-KOH, pH 7.9, 150 mM NaCl, 1.5 mM $MgCl_2$). The gradients were centrifuged at 23,400 rpm in a Surespin 630 rotor (Sorvall) for 15 h 45 min at 4°C and harvested manually in 1.5-ml fractions from the top. The peak fractions containing B/C complexes or mRNPs were determined by Cherenkov counting and then pooled prior to being applied to an amylose-Sepharose column. After being washed with 40 column volumes of GP150 buffer, bound complexes were eluted with GP150 buffer containing 15 mM maltose. RNA and protein of the eluates were recovered and analyzed by denaturing polyacrylamide gel electrophoresis (PAGE) or sodium dodecyl sulfate-PAGE, respectively. RNA was visualized by silver staining or autoradiography and protein by staining with Coomassie blue. The stoichiometry of the snRNAs and pre-mRNA/splicing intermediates was determined by visual inspection of the silver-stained bands, taking into consideration the length of the individual RNAs.

**MS and BLAST searches.** To identify proteins by MS, proteins were separated by one-dimensional (1D) PAGE (4 to 12% NuPAGE bis-Tris; Invitrogen) and entire lanes of the Coomassie-stained gels were cut into 25 slices of equal dimensions, irrespective of the visual presence or absence of a protein band. MS was performed after digestion of proteins in-gel, and peptides were analyzed on a liquid chromatography (LC)-coupled electrospray ionization quadropole time-of-flight (Q-Tof; Waters) mass spectrometer and/or a linear ion trap (4000 QTrap; Applied Biosystems) mass spectrometer as previously described (11). Human homologues of the identified *Drosophila* proteins (or vice versa) were identified by performing BLAST searches using the NCBI server.

**EM.** For EM, eluted complexes were fixed by being loaded onto a 4.4-ml, linear 10-to-30% glycerol gradient containing 0.1% gluteraldehyde in the 30% gradient solution (26) and centrifuged for 107 min at 60,000 rpm (374,000 × *g*). Gradients were subsequently fractionated from the bottom by using a fraction collector. Particles were negatively stained by the double-carbon-film method (25). Images were recorded with a calibrated CM120 EM (FEI, Eindhoven, The Netherlands) at a magnification of 88,000 or 115,000 in eucentric height at a defocus of approximately 1.5 μm. The microscope was equipped with a TemCam F224A digital camera (TVIPS, Gauting, Germany). Following a reference-free alignment (13), images were subjected to multivariate statistical analysis and classification (13, 56, 57). The resulting 2D class averages were in turn used as a reference during subsequent rounds of alignment until stable classes were achieved.

## RESULTS

**Kinetics of splicing of Ftz-M3 pre-mRNA in Kc and HeLa nuclear extract.** To isolate *Drosophila melanogaster* spliceosomal complexes, the *fushi tarazu* (*Ftz*) pre-mRNA substrate, which is derived from the naturally occurring *Drosophila fushi tarazu* gene and was shown previously to be spliced efficiently in both HeLa (43) and *Drosophila* nuclear extract (45), was used. For MS2 affinity purification, three MS2 binding sites were added to the 3′ end, generating Ftz-M3 pre-mRNA (Fig. 1A). In vitro splicing was performed for 0 to 180 min with nuclear extract prepared from *Drosophila melanogaster* Kc cells or HeLa cells and in vitro-transcribed $^{32}$P-labeled Ftz-M3 pre-mRNA. RNA was recovered and analyzed on a denaturing polyacrylamide gel. Splicing intermediates were first observed after 10 min and 8 min in Kc and HeLa nuclear extract, respectively, whereas splicing products were visible after 12 min (Kc) and 10 min (HeLa) (Fig. 1B; see Fig. S1A in the supplemental material). As determined by native gel electrophoresis, spliceosomal B and C complexes were formed in Kc nuclear extract after 4 min and 12 min of incubation, respectively (Fig. 1C, lanes 4 and 8). Similar kinetics of splicing complex formation (4 min and 10 min for B and C, respectively) were observed with HeLa nuclear extract (see Fig. S1B in the supplemental material). To isolate B complexes with little or no C complex contamination, splicing was thus performed for 8 min in Kc nuclear extract and for 6 min in HeLa nuclear extract.

**Isolation of spliceosomal B complexes assembled on MS2-tagged Ftz pre-mRNA in Kc or HeLa nuclear extract.** For the preparative isolation of human and *Drosophila* B complexes, the Ftz-M3 substrate was first incubated with MS2-MBP fusion protein prior to its addition to a splicing reaction mixture. Spliceosomes were then allowed to form for 8 (Kc) or 6 (HeLa) min, and the splicing reaction mixture was subsequently subjected to glycerol gradient centrifugation in order to separate B complexes from other spliceosomal complexes (such as the A complex) and also from excess MS2-MBP protein. The distribution of spliceosomal complexes in the gradient was determined by measuring the amount of radiolabeled RNA in each fraction. As shown in Fig. 2A, two main peaks are apparent, one in fractions 9 to 13 (*Drosophila*) or 11 to 15 (human) which corresponds mainly to a mixture of A and H complexes and a second peak in fractions 17 to 19 which corresponds to the B complex. Thus, *Drosophila* and human spliceosomal B complexes exhibit very similar sedimentation coefficients, with complexes from both species migrating in the 40S to 45S region of the gradient. In contrast, human H and/or A complexes exhibit a slightly higher sedimentation coefficient (~25S) than their *Drosophila melanogaster* counterparts (~20S).

Fractions 17 to 19, i.e., the peak fractions containing B complexes, were pooled and applied to amylose-coated Sepharose beads. After extensive washing, spliceosomes were eluted under native conditions with an excess of maltose, and their RNA and protein were subsequently analyzed by denaturing PAGE. The eluate of both *Drosophila* and human Ftz-M3 complexes contained nearly equimolar amounts of uncleaved pre-mRNA, as well as U1, U2, U4, U5, and U6 snRNA (Fig. 2B, upper panel, lanes 3 and 6), consistent with their designation as B complexes. The near-stoichiometric levels of U4 and U1 indicate that predominantly B and not activated B* complexes were isolated. Only small amounts of intron-lariat (5 to 6% as quantitated by phosphorimager analysis) but no splicing products were detected after autoradiography (Fig. 2B, lower panel, lanes 3 and 6). Thus, precatalytic spliceosomal B complexes of high purity (as judged by the absence of contaminating RNAs), with only minor amounts of contaminating B* and C complex, could be purified from Kc and HeLa cell nuclear extract.

**Protein composition of affinity-purified *Drosophila* and human B complexes formed on Ftz pre-mRNA.** Proteins isolated from affinity-purified *Drosophila* and human B complexes were separated by 1D PAGE and visualized by Coomassie staining. Proteins were subsequently identified by LC-tandem MS (MS-

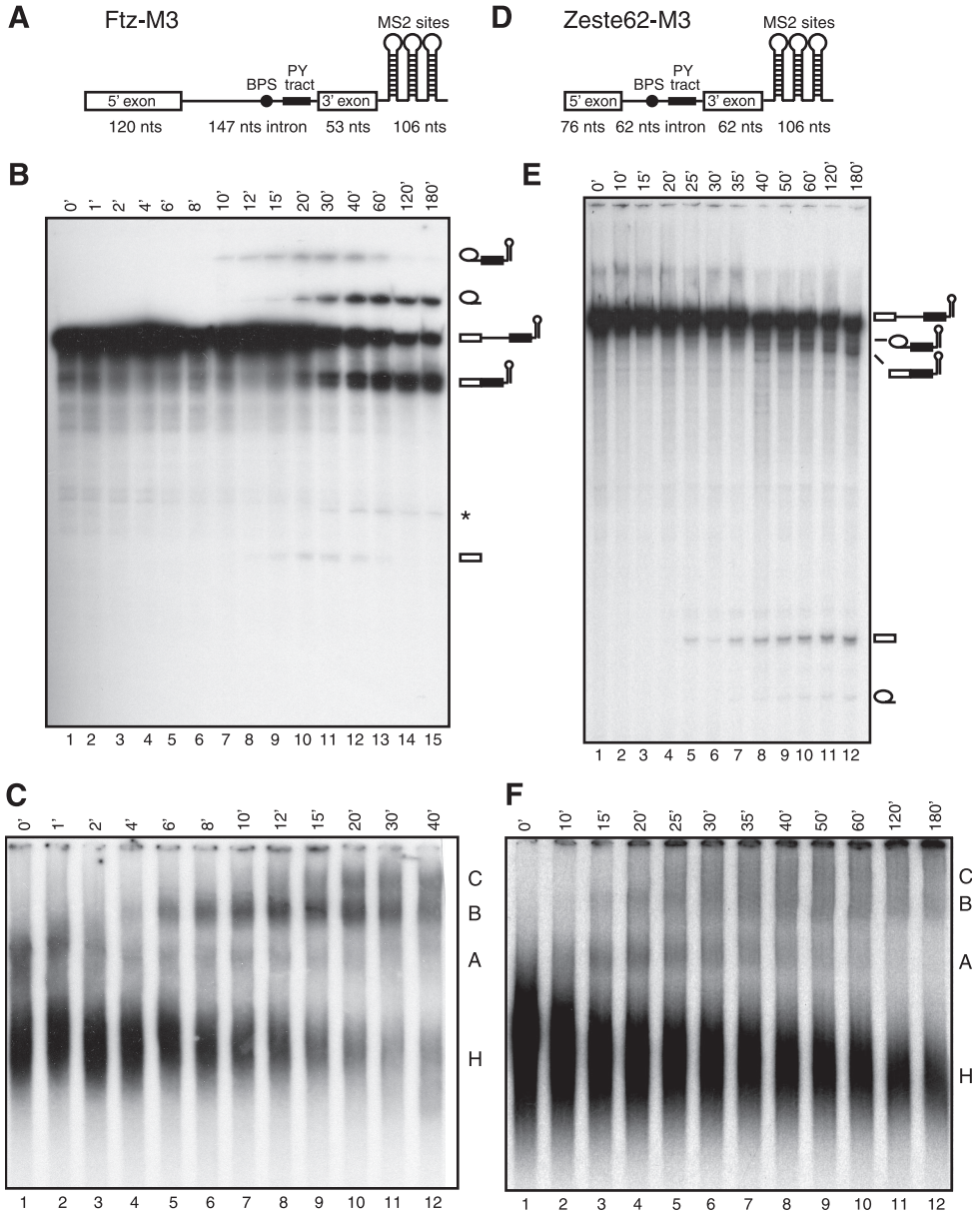


FIG. 1. Kinetics of splicing and splicing complex formation with Fushi tarazu (Ftz) and Zeste62 pre-mRNA. (A and D) Schematic representations of the Ftz-M3 and Zeste62-M3 pre-mRNA substrates. PY tract, polypyrimidine tract; nts, nucleotides. (B, C, E, and F) Splicing was performed with Ftz-M3 pre-mRNA (B and C) or Zeste62-M3 pre-mRNA (E and F) in Kc cell nuclear extract for the times indicated (in minutes) above each lane. In the experiments whose results are shown in panels B and E, RNA was recovered, analyzed on a 12% polyacrylamide–8 M urea gel, and visualized by autoradiography. Splicing substrates, intermediates, and products are indicated on the right. *, the band likely corresponds to debranched, spliced-out intron. Panels C and F show results for splicing complexes analyzed on a 3.5% native polyacrylamide gel and visualized by autoradiography. The positions of the H, A, B, and C complexes are indicated on the right.

MS) and scored by the absolute number of peptides found in each preparation. The protein composition of precatalytic spliceosomal B complexes formed in Kc and HeLa extract was determined in two independent experiments and is summarized in Table 1. Only proteins detected reproducibly by MS (i.e., in two out of two preparations) are shown in Table 1; proteins detected in only one preparation (see Table S1 in the supplemental material) are more likely to be contaminants and are therefore not included. The human homologue of the identified *Drosophila* protein (or vice versa) was determined by BLAST searches, and for simplicity's sake we use the name of the human homologue of a given *Drosophila* protein.

More than 120 proteins were identified reproducibly in *Drosophila* and human B complexes assembled on MS2-tagged Ftz pre-mRNA (Table 1). Most proteins identified in *Drosophila* B complexes were also found in human B complexes. The proteins common to *Drosophila* and human B complexes included nearly all U1, U2, U4/U6, and U5 snRNP-associated proteins. A large number of non-snRNP proteins were also shared by *Drosophila* and human Ftz B complexes. For example, several

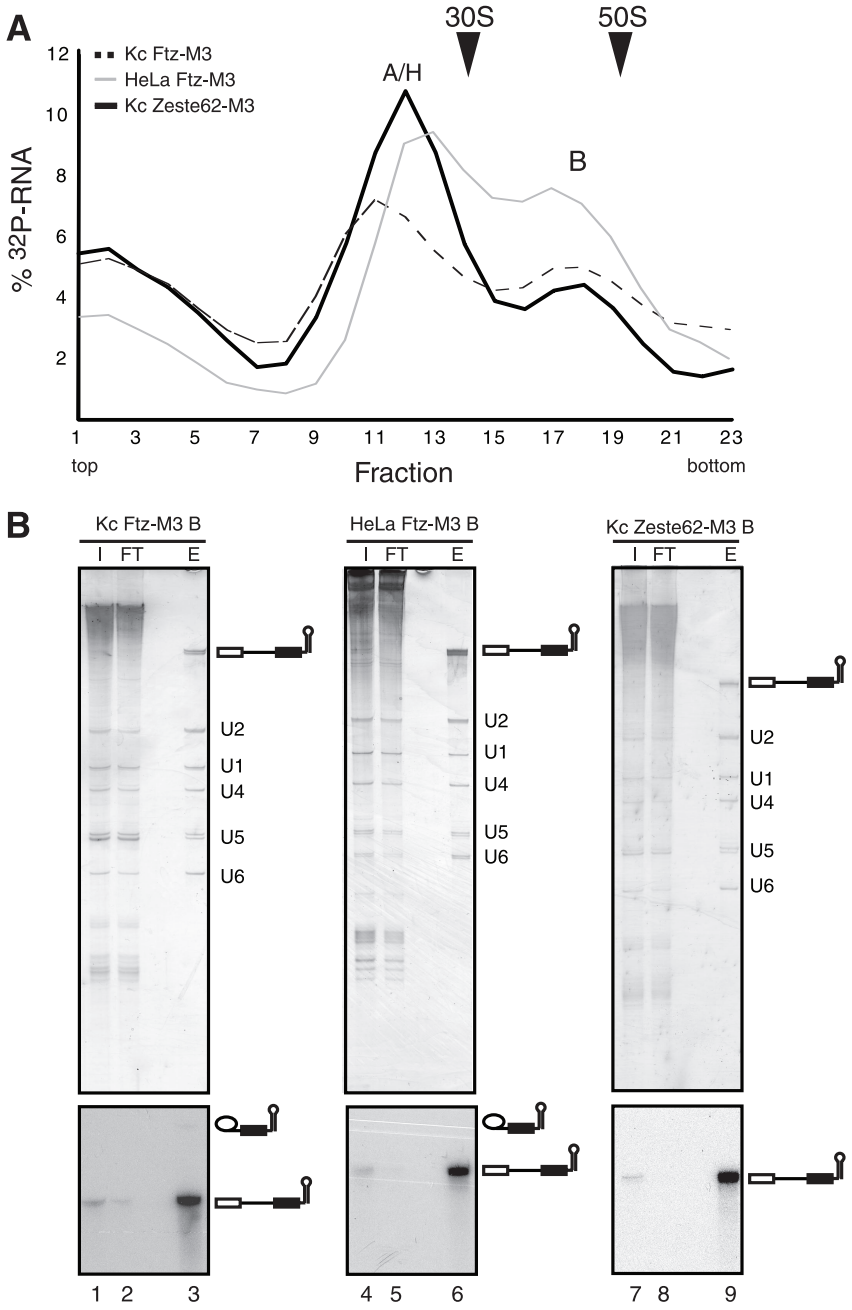


FIG. 2. MS2 affinity selection of spliceosomal B complexes. (A) Splicing reaction mixtures containing Ftz-M3 pre-mRNA and Kc or HeLa nuclear extract or Zeste62-M3 pre-mRNA and Kc nuclear extract were subjected to glycerol gradient centrifugation. The radioactivity contained in each gradient fraction was determined by Cherenkov counting and expressed as the percentage of the total $^{32}$P-RNA loaded onto the gradient. Sedimentation coefficients (indicated by arrowheads) were determined by analyzing the UV absorbance of fractions of a reference gradient containing prokaryotic ribosomal subunits. (B) Gradient fractions 17 to 19 containing Ftz B complexes formed in Kc (lanes 1 to 3) or HeLa (lanes 4 to 6) extract or Zeste62-M3 B complexes formed in Kc extract (lanes 7 to 9) were pooled and subjected to MS2 affinity selection. Bound complexes were eluted with maltose, and RNA was recovered, separated by denaturing PAGE, and visualized by silver staining (upper panel) or by autoradiography (lower panel). Lanes 1, 4, and 7, 2% of the input (I) pooled gradient fractions. Lanes 2, 5, and 8, 2% of the flowthrough (FT) of the amylose column. Lanes 3, 6, and 9, 25% of the eluted (E) spliceosomal complexes. The positions of the snRNAs, unspliced pre-mRNA, and splicing products/intermediates are indicated on the right.

members of the SR protein family were identified in both, including SF2/ASF and SRp55. Nearly all components of the Prp19/CDC5 complex could be identified in both *Drosophila* and human B complexes. With the exception of Cyp-E (the latter was detected solely in *Drosophila* Ftz complexes), essentially all proteins designated Prp19/CDC5 related were also present in both. A number of non-snRNP proteins previously detected in purified human A and/or B complexes were also shared by both human and *Drosophila* Ftz B complexes. In contrast, most proteins required for the first or second step of

TABLE 1. Protein composition of *Drosophila* spliceosomal B and C complexes[a]

| Human protein[b] | GenBank accession no. | *Drosophila melanogaster* protein | | | Presence of or absolute no. of peptides sequenced for protein[c] in indicated complex in: | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | HeLa nuclear extract | Kc nuclear extract | | | | | | | | HeLa extract, C, PM5[d] |
| | | CG no. | Gene name | Mol mass (kDa) | B, MINX-M3[d] | B, Ftz-M3 | | B, Ftz-M3 | | B, Zeste62-M3 | | | C, M3-Ftz-longPYT | |
| | | | | | | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | |
| Sm proteins | | | | | | | | | | | | | | | |
| B | gi\|17136806 | CG5352 | *SmB* | 21.0 | ● | 10 | 12 | 22 | 11 | 11 | 12 | 12 | 12 | 11 | ● |
| D1 | gi\|17864386 | CG10753 | *snRNP69D* | 13.8 | ● | 7 | 9 | 3 | 2 | 4 | 5 | 4 | 4 | 5 | ● |
| D2 | gi\|21357623 | CG1249 | *SmD2* | 13.5 | ● | 17 | 17 | 19 | 10 | 18 | 16 | 10 | 14 | 12 | ● |
| D3 | gi\|24652901 | CG8427 | *SmD3* | 15.6 | ● | 9 | 15 | 18 | 11 | 17 | 11 | 15 | 7 | 4 | ● |
| E | gi\|24582631 | CG18591 | | 11.1 | ● | 9 | 5 | 5 | 8 | 11 | 6 | 9 | 8 | 6 | ● |
| F | gi\|22024001 | CG16792 | *DebB* | 9.7 | ● | 11 | 13 | 5 | | | | | 2 | 2 | ● |
| G | gi\|29428065 | CG9742 | | 12.7 | ● | 2 | 3 | 6 | 3 | 4 | 3 | 4 | 3 | 3 | ● |
| U1 snRNP | | | | | | | | | | | | | | | |
| U1-70K | gi\|17137278 | CG8749 | *U1-70K* | 52.9 | ● | 1 | 4 | 17 | 5 | 5 | 7 | 6 | 1 | | |
| U1-A | gi\|17737284 | CG4528 | *snf* | 24.6 | ● | | | 12 | 4 | 4 | 2 | 4 | 5 | 5 | |
| U1-associated | | | | | | | | | | | | | | | |
| S164/RBM25 | gi\|18857955 | CG4119 | | 112.7 | | 1 | 1 | 18 | 7 | 11 | 13 | 12 | | | |
| FBP11 | gi\|24581324 | CG3542 | | 91.3 | | | | 9 | 6 | 8 | 13 | 11 | | | |
| U2 snRNP | | | | | | | | | | | | | | | |
| U2A′ | gi\|19921760 | CG1406 | *U2A* | 29.7 | ● | 18 | 20 | 9 | 2 | 9 | 8 | 9 | 8 | 5 | ● |
| U2B″ | gi\|17737284 | CG4528 | *snf* | 24.6 | ● | 5 | 7 | 12 | 4 | 4 | 2 | | 5 | 5 | ● |
| SF3a120 | gi\|24647566 | CG16941 | | 88.1 | ● | 26 | 32 | 84 | 21 | 28 | 26 | 20 | 20 | 9 | ● |
| SF3a66 | gi\|24663500 | CG10754 | | 30.3 | ● | 2 | 5 | 4 | 3 | 6 | 5 | 4 | | 2 | ● |
| SF3a60 | gi\|17137118 | CG2925 | *noi* | 58.4 | ● | 14 | 19 | 28 | 16 | 15 | 15 | 17 | 9 | 3 | ● |
| SF3b155 | gi\|45550087 | CG2807 | | 149.6 | ● | 37 | 54 | 102 | 29 | 48 | 49 | 54 | 15 | 16 | ● |
| SF3b145 | gi\|19920622 | CG3605 | | 84.8 | ● | 20 | 30 | 25 | 13 | 17 | 22 | 13 | 10 | 9 | ● |
| SF3b130 | gi\|24654874 | CG13900 | | 136.6 | ● | 50 | 71 | 66 | 36 | 68 | 50 | 60 | 28 | 30 | ● |
| SF3b49 | gi\|17530817 | CG3780 | *spx* | 39.7 | ● | 1 | 2 | | | | 1 | | | | ● |
| SF3b14a/p14 | gi\|21357707 | CG13298 | | 14.2 | ● | 14 | 6 | 7 | 8 | 8 | 6 | 6 | 5 | 6 | ● |
| SF3b14b | gi\|24582237 | CG9548 | | 12.5 | ● | | 6 | 9 | 3 | 5 | 6 | 5 | 4 | 3 | ● |
| SF3b10 | gi\|24645240 | CG11985 | | 10.0 | ● | 1 | 2 | 4 | | 1 | | | | | |
| U2-related | | | | | | | | | | | | | | | |
| hPRP43 | gi\|19921728 | CG11107 | | 82.7 | ● | 7 | 11 | 3 | | 2 | 8 | 3 | 3 | 4 | ● |
| SPF45 | gi\|116007460 | CG17540 | | 44.4 | ● | | | | | 3 | 1 | | | | |
| SPF30 | gi\|16769358 | CG17454 | *SPF30* | 28.2 | ● | 2 | 2 | 4 | 2 | 7 | 7 | 5 | | | |
| U2AF65 | gi\|17136764 | CG9998 | *U2af50* | 46.7 | ● | 6 | 2 | 8 | 7 | 9 | 11 | 8 | | | |
| U2AF35 | gi\|17137284 | CG3582 | *U2af38* | 29.9 | ● | 2 | | 4 | 1 | 3 | 2 | | | | |
| PUF60 | gi\|24655242 | CG12085 | *pUf68* | 67.9 | | | 3 | 20 | 9 | 25 | 17 | 16 | 29 | 12 | |
| A proteins | | | | | | | | | | | | | | | |
| RBM10 | gi\|24580815 | CG4887 | | 114.2 | ● | | | | | 1 | | 1 | | | |
| SF1 | gi\|24647704 | CG5836 | *SF1* | 87.4 | | | | 3 | 1 | 1 | 1 | | | | |
| A/B proteins | | | | | | | | | | | | | | | |
| DDX9 | gi\|62474635 | CG11680 | *mle* | 143.7 | ● | 24 | 33 | | | | | | | | |
| ZC3H18/LOC124245 | gi\|24640344 | CG1677 | | 109.1 | ● | 10 | 12 | 23 | 9 | 12 | 9 | 12 | 3 | | ● |
| ASR2B | gi\|24585960 | CG7843 | | 107.2 | ● | 15 | 23 | 50 | 18 | 24 | 22 | 29 | 8 | | ● |
| TCERG1 (CA150) | gi\|116007968 | CG33097 | | 84.8 | ● | 1 | 5 | 14 | 2 | 2 | | 1 | 3 | 2 | |
| p68/DDX5 or DDX17 | gi\|45551833 | CG10279 | *Rm62* | 78.5 | ● | | 4 | 19 | 10 | 18 | 14 | 19 | 15 | 14 | ● |
| ELAV (HuR) | gi\|62471589 | CG3151 | *Rbp9* | 70.5 | ● | 1 | 4 | | | | | | | | ● |
| RBM39/RNPC2 | gi\|45550943 | CG11266 | | 66.5 | ● | 7 | 6 | 20 | | 16 | 14 | 23 | 4 | 1 | ● |
| NF45/ILF2 | gi\|21357109 | CG5641 | | 43.7 | ● | 15 | 31 | 23 | 6 | 3 | 2 | 1 | 1 | | ● |
| YBX-1 | gi\|24663131 | CG5654 | *yps* | 38.9 | ● | 16 | 15 | 11 | 4 | 4 | 3 | 6 | 2 | 9 | ● |
| TLS/FUS | gi\|24642436 | CG3606 | *caz* | 38.8 | ● | | | | | | | | 3 | 6 | |
| HCNGP | gi\|19921898 | CG2063 | | 36.2 | | | | | | 1 | 1 | | 1 | | |
| SR proteins | | | | | | | | | | | | | | | |
| SFRS12 | gi\|17137528 | CG4602 | *Srp54* | 58.5 | | | | 10 | 2 | 7 | 5 | 3 | 13 | 7 | |
| SRp55 | gi\|28571701 | CG10851 | *B52* | 42.8 | ● | 2 | 1 | 8 | 3 | 8 | 10 | 6 | 9 | 3 | ● |

*Continued on following page*

TABLE 1—*Continued*

| Human protein[b] | GenBank accession no. | *Drosophila melanogaster* protein | | | Presence of or absolute no. of peptides sequenced for protein[c] in indicated complex in: | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | HeLa nuclear extract | Kc nuclear extract | | | | | | | | HeLa extract, C, PM5[d] |
| | | CG no. | Gene name | Mol mass (kDa) | B, MINX-M3[d] | B, Ftz-M3 | | B, Ftz-M3 | | B, Zeste62-M3 | | | C, M3-Ftz-longPYT | | |
| | | | | | | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | |
| hTra-2 b/SFRS10 | gi\|45552623 | CG10128 | *Tra2* | 31.0 | ● | | 3 | | | 3 | 2 | | | | ● |
| SF2/ASF | gi\|21358097 | CG6987 | *SF2* | 28.4 | ● | 4 | 15 | 4 | 9 | 12 | 8 | 6 | 8 | 7 | ● |
| 9G8 | gi\|24582360 | CG10203 | *xl6* | 27.9 | ● | 7 | 15 | | | 1 | 1 | | 2 | 3 | ● |
| SRp20 | gi\|24641772 | CG1987 | *Rbp1-like* | 16.8 | | | 1 | | | 2 | 1 | | | | ● |
| SRp30c | ▲[e] | | | | ● | 1 | 5 | | | | | | | | ● |
| SRp38 | ▲ | | | | ● | 3 | 8 | | | | | | | | ● |
| SR-related | | | | | | | | | | | | | | | |
| SRm300 | gi\|161080550 | CG7971 | | 127.3 | | 9 | 13 | 19 | | 3 | 3 | 1 | 41 | 19 | ● |
| SRm160 | gi\|24663664 | CG11274 | *SRm160* | 107.6 | ● | | | 3 | | 1 | 1 | 3 | 2 | | ● |
| Cap binding complex | | | | | | | | | | | | | | | |
| CBP80 | gi\|24639721 | CG7035 | *Cbp80* | 93.2 | ● | 25 | 35 | 37 | 9 | 21 | 24 | 21 | 13 | 8 | ● |
| CBP20 | gi\|17738031 | CG12357 | *Cbp20* | 17.7 | ● | 11 | 11 | 7 | 1 | 2 | 2 | 2 | 4 | 1 | ● |
| hnRNP | | | | | | | | | | | | | | | |
| hnRNP A0 | | | | | ● | 3 | 4 | | | | | | | | |
| hnRNP A1 | —[f] | | | | ● | 11 | 16 | | | | | | | | ● |
| hnRNP A3 | — | | | | ● | 4 | 8 | | | | | | | | ● |
| hnRNP AB | — | | | | ● | 4 | 7 | | | | | | | | |
| hnRNP A2/B1 | — | | | | ● | 10 | 23 | | | | | | | | |
| hnRNP C | — | | | | ● | 22 | 49 | | | | | | | | ● |
| hnRNP D | — | | | | | 4 | 4 | | | | | | | | ● |
| hnRNP G | — | | | | ● | 13 | 11 | | | | | | | | ● |
| hnRNP K | — | | | | | 1 | 3 | | | | | | | | |
| hnRNP L | — | | | | | 27 | 46 | | | | | | | | |
| hnRNP Q | — | | | | ● | 13 | 23 | | | | | | | | ● |
| hnRNP R | — | | | | ● | 44 | 41 | | | | | | | | ● |
| hnRNP U | — | | | | | 4 | 6 | | | | | | | | ● |
| hnRPUL2/SAF-A2 | — | | | | | 4 | 2 | | | | | | | | |
| PCBP1 | — | | | | ● | 3 | 7 | | | | | | | | ● |
| RALY | — | | | | ● | 5 | 8 | | | | | | | | ● |
| — | gi\|22024201 | CG30122 | | 81.0 | | | | | | 1 | | 2 | | | |
| — | gi\|161078453 | CG17838 | | 76.9 | | | | 8 | 1 | 2 | 7 | 6 | | | |
| — | gi\|24646609 | CG16901 | *sqd* | 63.2 | | | | | | | | | 2 | 5 | |
| — | gi\|11040 | | *hrp40.2* | 39.5 | | | | | | 2 | 2 | | | | |
| U5 | | | | | | | | | | | | | | | |
| 220K | gi\|20129897 | CG8877 | *prp8* | 279.4 | ● | 92 | 140 | 149 | 71 | 123 | 150 | 161 | 202 | 104 | ● |
| 200K | gi\|28574898 | CG5931 | | 244.5 | ● | 86 | 120 | 134 | 54 | 136 | 99 | 113 | 171 | 108 | ● |
| 116K | gi\|21357743 | CG4849 | | 110.7 | ● | 33 | 46 | 61 | 42 | 41 | 34 | 46 | 90 | 43 | ● |
| 102K | gi\|24666532 | CG6841 | | 105.2 | ● | 30 | 75 | 44 | 21 | 30 | 21 | 26 | 6 | 3 | ● |
| 100K | gi\|24584994 | CG10333 | | 94.6 | ● | 26 | 27 | 26 | 19 | 23 | 21 | 18 | 3 | 2 | ● |
| 52K | gi\|24583433 | CG5198 | *LIN1* | 36.8 | ● | 2 | 5 | | | | | | | | |
| 40K | gi\|24580577 | CG3436 | | 38.9 | ● | 16 | 10 | 23 | 7 | 6 | 9 | 10 | 24 | 17 | ● |
| 15K | gi\|28574029 | CG3058 | | 16.8 | ● | 3 | 8 | 4 | 3 | 3 | 5 | 4 | | | |
| LSm proteins | | | | | | | | | | | | | | | |
| LSm2 | gi\|21358381 | CG10418 | | 10.7 | ● | 9 | 6 | 5 | 2 | 2 | 7 | 5 | | 1 | ● |
| LSm3 | gi\|24649486 | CG5926 | | 11.9 | ● | 2 | 4 | 2 | 2 | | | | | | ● |
| LSm4 | gi\|78706862 | CG17764 | | 30.6 | ● | 2 | 6 | 7 | | 2 | 1 | 3 | | | |
| LSm5[j] | gi\|21358059 | CG6610 | | 10.1 | | | | | | | | | | | |
| LSm6[j] | gi\|24656550 | CG9344 | | 9.0 | ● | | | | | | | | | | ● |
| LSm7 | gi\|20129565 | CG13277 | | 12.2 | ● | 2 | 3 | 3 | 2 | 3 | 3 | 4 | 2 | 2 | ● |
| LSm8 | gi\|24655606 | CG2021 | | 10.4 | ● | 4 | 1 | | | | 1 | | | | ● |
| U4/U6 snRNP | | | | | | | | | | | | | | | |
| 90K | gi\|24667067 | CG7757 | | 67.4 | ● | 27 | 34 | 41 | 15 | 12 | 17 | 15 | | | ● |
| 60K | gi\|21355245 | CG6322 | | 61.6 | ● | 16 | 25 | 22 | 16 | 13 | 22 | 24 | 2 | | ● |
| 61K | gi\|21357435 | CG6876 | | 55.5 | ● | 15 | 17 | 42 | 27 | 15 | 22 | 22 | | | ● |

*Continued on following page*

TABLE 1—*Continued*

| Human protein[b] | GenBank accession no. | *Drosophila melanogaster* protein | | | Presence of or absolute no. of peptides sequenced for protein[c] in indicated complex in: | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | HeLa nuclear extract | Kc nuclear extract | | | | | | | | HeLa extract, C, PM5[d] |
| | | CG no. | Gene name | Mol mass (kDa) | B, MINX-M3[d] | B, Ftz-M3 | | B, Ftz-M3 | | B, Zeste62-M3 | | | C, M3-Ftz-longPYT | |
| | | | | | | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | |
| 20K | gi\|24586038 | CG17266 | | 20.2 | ● | 10 | 12 | 11 | 8 | 10 | 6 | 3 | | | ● |
| 15.5K | gi\|17864298 | CG3949 | *hoi-polloi* | 13.8 | ● | 3 | 1 | 3 | 3 | 2 | 3 | 3 | | | ● |
| U4/U6.U5 | | | | | | | | | | | | | | | |
| 110K | gi\|24583764 | CG6686 | | 112.6 | ● | 37 | 31 | 19 | 12 | 12 | 12 | 11 | | | ● |
| 65K | gi\|18859907 | CG7288 | | 57.5 | ● | 11 | 18 | 20 | 9 | 14 | 16 | 16 | | | ● |
| 27K | gi\|19922846 | CG18426 | *ytr* | 21.5 | | 2 | 5 | 10 | | 6 | 4 | 3 | | | |
| TFIP11 | gi\|24582006 | CG7238 | *sip1* | 94.8 | ● | 2 | 4 | 6 | | 3 | 5 | 4 | 4 | 1 | ● |
| hPRP38 | gi\|22024089 | CG30342 | | 40.1 | ● | 13 | 10 | 19 | 7 | 7 | 11 | 8 | 7 | 6 | ● |
| hPRP19/CDC5L | | | | | | | | | | | | | | | |
| CDC5L | gi\|19922992 | CG6905 | *cdc5-like* | 93.1 | ● | 18 | 20 | 21 | 7 | 24 | 19 | 16 | 46 | 28 | ● |
| Hsp73 (CCAP1) | gi\|28571721 | CG4264 | *Hsc70-4* | 71.1 | ● | 4 | 5 | | | | | 2 | 3 | 3 | ● |
| hPRP19 | gi\|17647459 | CG5519 | *Gbp* | 55.2 | ● | 15 | 24 | 37 | 18 | 22 | 21 | 16 | 49 | 44 | ● |
| PRL1 | gi\|18859793 | CG1796 | *Tango4* | 52.7 | ● | 9 | 13 | 8 | 5 | 4 | 7 | 14 | 29 | 19 | ● |
| SPF27 | gi\|24650745 | CG4980 | | 31.3 | ● | 6 | 4 | 11 | 2 | 5 | 7 | 5 | 11 | 17 | ● |
| AD002 (CCAP2) | gi\|21358119 | CG12135 | *c12.1* | 28.4 | ● | 2 | | 3 | | 3 | 2 | | 8 | 13 | ● |
| Npw38BP | gi\|20128957 | CG2685 | | 61.0 | ● | 9 | 10 | 6 | | | | | 2 | 1 | |
| Npw38 | gi\|21356239 | CG11820 | | 27.1 | ● | 2 | 5 | 2 | 2 | 2 | | 1 | 1 | | |
| hPRP19/CDC5L related | | | | | | | | | | | | | | | |
| KIAA0560 | gi\|45553341 | CG31368 | | 172.0 | ● | 11 | 7 | 63 | 6 | 24 | 30 | 35 | 102 | 65 | ● |
| hSYF1 (XAB2) | gi\|20129961 | CG6197 | | 103.3 | ● | 13 | 4 | 13 | 1 | 11 | 13 | 11 | 39 | 36 | ● |
| CRNKL1/Clf1 | gi\|17137126 | CG3193 | *crn* | 84.3 | ● | 14 | 12 | 17 | 1 | 16 | 17 | 29 | 62 | 40 | ● |
| SKIP | gi\|24640727 | CG8264 | *BX42* | 61.2 | ● | 13 | 12 | 19 | 6 | 3 | 7 | 8 | 40 | 28 | ● |
| hECM2 (fSAP47) | gi\|28573264 | CG14641 | | 47.2 | ● | 5 | 6 | 22 | 2 | 3 | 3 | 4 | 29 | 17 | ● |
| Cyp-E | gi\|17647301 | CG4886 | *cyp33* | 33.3 | ● | | | 5 | 3 | 7 | 6 | 6 | 20 | 21 | ● |
| hIsy1 | gi\|78709073 | CG9667 | | 31.4 | ● | 3 | 2 | 2 | | 6 | 4 | 8 | 13 | 12 | ● |
| PPIL1 | gi\|17986117 | CG13892 | *Cypl* | 19.5 | ● | 3 | 4 | 7 | 2 | 6 | 6 | 2 | 11 | 7 | ● |
| G10 | gi\|17737310 | CG1639 | *l(1)10Bb* | 17.0 | ● | 9 | 8 | 5 | 2 | 5 | 6 | 7 | 12 | 9 | ● |
| RES complex | | | | | | | | | | | | | | | |
| MGC13125 | gi\|21355513 | CG13625 | | 76.5 | ● | 7 | 3 | | | | | | 2 | 2 | ● |
| SNIP1 | gi\|62862214 | CG17168 | | 49.8 | ● | 4 | 5 | | | | | | | | ● |
| CGI-79 | gi\|20129641 | CG10466 | | 17.8 | ● | 4 | 1 | 1 | | | | | | 1 | ● |
| B/B* proteins | | | | | | | | | | | | | | | |
| hPRP4-kinase/PRPF4B | gi\|19922978 | CG7028 | | 104.0 | ● | 7 | 10 | 18 | 4 | 5 | 10 | 6 | 2 | | ● |
| hPRP2 | gi\|19921526 | CG10689 | | 102.8 | ● | 3 | | 8 | | 7 | 6 | 5 | 5 | 7 | ● |
| PPIL4 | gi\|21355677 | CG5808 | | 75.5 | ● | 2 | | | | 1 | 1 | 1 | 1 | | |
| RED | gi\|21355769 | CG18005 | | 61.4 | ● | 10 | 8 | 17 | 3 | 9 | 5 | 9 | 9 | 1 | ● |
| PPIL2/Cyp-60 | gi\|19922376 | CG7747 | | 59.0 | ● | 7 | | | | | | | 5 | 1 | ● |
| hSmu-1 | gi\|21356791 | CG5451 | | 57.3 | ● | 20 | 22 | 16 | 8 | 7 | 7 | 14 | 12 | 4 | ● |
| MFAP1 | gi\|21358323 | CG1017 | | 56.1 | ● | 7 | 6 | 5 | | 1 | 2 | 2 | 4 | | ● |
| FBP21 | gi\|24580725 | CG4291 | | 38.8 | ● | 2 | 2 | 6 | | 7 | 9 | 3 | | | |
| UBL5 | gi\|24586084 | CG3450 | *ubl* | 8.6 | ● | 3 | 4 | | | | | | | | |
| C proteins | | | | | | | | | | | | | | | |
| hPRP22 | gi\|20129977 | CG8241 | *pea* | 142.0 | | | | 2 | | 3 | 2 | 5 | 54 | 28 | ● |
| PPIG, SR-Cyp | gi\|28571910 | CG1866 | *Moca-cyp* | 116.4 | | | | 1 | 1 | 2 | 1 | 2 | 12 | 2 | ● |
| Q9BRR8 | gi\|24663949 | CG8833 | *GPTC1* | 107.9 | | | | | | 1 | | 1 | 6 | 4 | ● |
| cactin (c19orf29) | gi\|45549144 | CG1676 | *cactin* | 91.3 | | | | | | | | | 17 | 10 | ● |
| DDX35 | gi\|19920696 | CG3225 | | 76.1 | | | | 3 | | 3 | 2 | 3 | 12 | 8 | ● |
| Abstrakt | gi\|17977678 | CG14637 | *abs* | 69.5 | ● | | | | | 1 | 3 | 4 | 42 | 17 | ● |
| hPRP17 | gi\|21355805 | CG6015 | | 65.4 | ● | 5 | 2 | 17 | 2 | 6 | 11 | 15 | 35 | 18 | ● |
| GCIP p29 | gi\|24652559 | CG12343 | | 26.5 | ● | | | 1 | | 4 | 4 | 2 | 18 | 14 | ● |
| EJC/mRNP | | | | | | | | | | | | | | | |
| eIF4AIII | gi\|24645031 | CG7483 | *eIF4AIII* | 45.6 | | 3 | | 12 | 3 | 4 | 7 | 8 | 21 | 16 | ● |
| Y14 | gi\|24651979 | CG8781 | *tsu* | 19.0 | | | | | | | | | 3 | 1 | ● |

*Continued on following page*

TABLE 1—*Continued*

| Human protein[b] | GenBank accession no. | *Drosophila melanogaster* protein | | | Presence of or absolute no. of peptides sequenced for protein[c] in indicated complex in: | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | HeLa nuclear extract | | | Kc nuclear extract | | | | | | HeLa extract, C, PM5[d] |
| | | CG no. | Gene name | Mol mass (kDa) | B, MINX-M3[d] | B, Ftz-M3 | | B, Ftz-M3 | | B, Zeste62-M3 | | | C, M3-Ftz-longPYT | |
| | | | | | | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | |
| Magoh | gi\|17136332 | CG9401 | *mago* | 17.3 | ● | | | 1 | | 1 | | | 2 | 2 | ● |
| Acinus | gi\|24585110 | CG10473 | *hkl* | 83.7 | | | | | | | 3 | 1 | 2 | | ● |
| UAP56 | gi\|24581956 | CG7269 | *Hel25E* | 48.7 | ● | 2 | 5 | | | | | | | | ● |
| Pinin | gi\|161078140 | CG8383 | *Pnn* | 35.9 | | | | 1 | | | | | 2 | 1 | ● |
| Aly/REF (THOC4) | gi\|21356157 | CG1101 | *Aly* | 27.9 | | 1 | 2 | | | | 1 | | 1 | | ● |
| SAP18 | gi\|17738001 | CG6046 | *bin1* | 17.3 | | | | 3 | | | | | 2 | 1 | ● |
| ELG | | NH | | | | 1 | 1 | | | | | | | | |
| Miscellaneous *Drosophila* + HeLa Spliceosomal[g] | | | | | | | | | | | | | | | |
| SKIV2L2 | gi\|17864608 | CG4152 | *lethal(2)35Df* | 118.9 | | 2 | 5 | | | | | | 6 | 1 | ● |
| GCFC | gi\|24644714 | CG1965 | | 102.6 | ● | 2 | | 3 | 1 | 7 | 6 | 8 | 3 | 3 | ● |
| ZFR2 | gi\|28574893 | CG5215 | *Zn72D* | 96.0 | | 12 | 16 | 21 | 8 | 2 | 1 | | 1 | | |
| RBM15B | gi\|24586450 | CG2910 | *nito* | 89.1 | | 2 | | 3 | 1 | 5 | 7 | | | | |
| PABP | gi\|45552715 | CG5119 | *PABP* | 69.9 | ● | 2 | 6 | 14 | 12 | 13 | 13 | 22 | 15 | 10 | ● |
| HsKin17 | gi\|24667403 | CG5649 | *kin17* | 45.4 | ● | 2 | 3 | 1 | 1 | | | | | | ● |
| RACK1 (GNB2L1) | gi\|17137396 | CG7111 | *RACK1* | 35.6 | | 1 | 1 | 1 | | 1 | 2 | 1 | 2 | 3 | ● |
| RBM42 | gi\|21357051 | CG2931 | | 33.6 | | 1 | 2 | 17 | 8 | 7 | 8 | 8 | 1 | | ● |
| Novel[h] | | | | | | | | | | | | | | | |
| IGF2BP1 | gi\|116007146 | CG1691 | *IMP-1* | 63.4 | | 10 | 12 | 40 | 18 | 10 | 18 | 20 | 10 | 4 | |
| PABP II | gi\|24586513 | CG2163 | *Pabp2* | 25.0 | | | | | | | 1 | 1 | 1 | | |
| Miscellaneous *Drosophila* only Spliceosomal[g] | | | | | | | | | | | | | | | |
| PSF | gi\|62473776 | CG4211 | *Bj6* | 82.0 | | | | 8 | 2 | 9 | 8 | 7 | 8 | 2 | |
| NY-CO-10 | gi\|21357249 | CG10907 | | 56.6 | ● | | 1 | 4 | 1 | 2 | | 1 | 3 | 2 | ● |
| DGCR14 | gi\|18858195 | CG1474 | *ES2* | 56.0 | | | | | | | | | 12 | 6 | ● |
| NKAP | gi\|24650530 | CG6066 | | 52.9 | | | | | | | | | 8 | 2 | ● |
| RBM4 | gi\|24659981 | CG8597 | *lark* | 39.9 | ● | | | 25 | 14 | 9 | 11 | 6 | 5 | 2 | |
| CDK10, PISSLRE | gi\|24652305 | CG1362 | *cdc2rk* | 39.4 | | | | | | 1 | | | 2 | 2 | ● |
| NOSIP | gi\|24643007 | CG6179 | *NOSIPL* | 34.2 | | | | | | | | | 3 | 1 | ● |
| MORG1 | gi\|24584567 | CG4935 | | 34.1 | | | | | | 2 | 2 | 3 | 7 | 3 | ● |
| CXorf56 | gi\|20129505 | CG16865 | | 27.7 | | | | | | | | | 4 | 4 | ● |
| Novel[h] | | | | | | | | | | | | | | | |
| CDC27 | gi\|24659892 | CG8610 | *Cdc27* | 101.3 | | | | | 2 | 11 | 7 | 8 | | | |
| PIWIL1 | gi\|161076864 | CG6137 | *aub* | 98.6 | | | | 7 | | 2 | 1 | 2 | 2 | 3 | |
| APTX | gi\|24648185 | CG5316 | | 76.5 | | | | 8 | 3 | | 2 | 1 | | | |
| MTHFSD | gi\|161078016 | CG14648 | *growl* | 58.8 | | | | 20 | 3 | 10 | 7 | 9 | 10 | 2 | |
| LUC7-like | gi\|24665973 | CG7564 | | 48.4 | | | | 3 | 2 | 2 | 5 | 4 | | | |
| KHDRBS2 | gi\|24658086 | CG3613 | *QKR58E-1* | 42.3 | | | | 4 | 1 | 3 | 2 | 2 | 1 | 1 | |
| PSME3 | gi\|18860055 | CG1591 | *REG* | 28.1 | | | | | | 1 | 1 | 1 | 1 | 1 | |
| NH | gi\|45553181 | CG6143 | *pep* | 78.1 | | | | 4 | 1 | 6 | 9 | 6 | 2 | | |
| NH | gi\|45550712 | CG9684 | | 71.9 | | | | 78 | 42 | 8 | 12 | 6 | 1 | | |
| NH | gi\|24656457 | CG8994 | *exu* | 58.0 | | | | 1 | 3 | | 3 | 2 | | | |
| NH | gi\|24641752 | CG15747 | | 44.7 | | | | 2 | | | 1 | | 1 | 3 | |
| NH | gi\|19922308 | CG11808 | | 25.7 | | | | | | 3 | | | 1 | 2 | |
| Miscellaneous *Drosophila* Ftz-M3 only, novel[h] | | | | | | | | | | | | | | | |
| EIF2AK4 | gi\|17137328 | CG1609 | *Gcn2* | 178.5 | | | | 1 | | | | | 1 | 1 | |
| G3BP2 | gi\|24646611 | CG9412 | *rin* | 75.0 | | | | 4 | 2 | | | 1 | 1 | | |
| ▽[i] | gi\|24649433 | CG13597 | | 59.2 | | | | 2 | 1 | | | | | | |
| PRPF38B | gi\|24641727 | CG1622 | | 45.4 | | | | 1 | 1 | | | | | | |
| NH | gi\|28571193 | CG5877 | | 114.2 | | | | | | | | | 2 | 1 | |

*Continued on following page*

TABLE 1—*Continued*

| Human protein[b] | GenBank accession no. | *Drosophila melanogaster* protein | | | Presence of or absolute no. of peptides sequenced for protein[c] in indicated complex in: | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | HeLa nuclear extract | | Kc nuclear extract | | | | | | | HeLa extract, C, PM5[d] |
| | | CG no. | Gene name | Mol mass (kDa) | B, MINX-M3[d] | B, Ftz-M3 | B, Ftz-M3 | | B, Zeste62-M3 | | | C, M3-Ftz-longPYT | | |
| | | | | | | 1 2 | 1 | 2 | 1 | 2 | 3 | 1 | 2 | |
| NH | gi\|24653446 | CG6209 | *mst101(2)* | 69.0 | | | 1 | | | | | 4 | 9 | |
| NH | gi\|24650092 | CG12250 | | 16.5 | | | 4 | | | | | 3 | 5 | |
| Miscellaneous *Drosophila* Zeste62-M3 only | | | | | | | | | | | | | | |
| Spliceosomal[g]: AGGF1 | gi\|24653802 | CG8079 | | 66.9 | | | | | 1 | 1 | | | | |
| Novel[h] | | | | | | | | | | | | | | |
| SLC7A2 | gi\|116007820 | CG7255 | | 85.6 | | | | | | 1 | 1 | | | |
| NCOA5 | gi\|17647723 | CG8614 | *Neos* | 42.6 | | | 1 | | 2 | 2 | | | | |
| DNAJC17 | gi\|24645889 | CG17187 | | 34.7 | | | | | 2 | 3 | | | | |
| TFIIB | gi\|19921082 | CG5343 | | 23.3 | | | | | 4 | 3 | | 1 | | |
| Miscellaneous HeLa only | | | | | | | | | | | | | | |
| Spliceosomal[g] | | | | | | | | | | | | | | |
| GEMIN3/DDX20 | gi\|17647335 | CG6539 | *Dhh1* | 116.5 | | 1 1 | | | | | | | | |
| SON3 | gi\|24645429 | CG8273 | | 88.6 | | 2 2 | | | | | | | | |
| GNL3 | gi\|28572990 | CG3983 | *GNL3* | 66.0 | | 4 5 | | | | | | | | ● |
| STRBP | ▽ | | | | | 8 19 | | | | | | | | |
| RBM7 | ▽ | | | | ● | 1 2 | | | | | | | | ● |
| ZCC2 | ▽ | | | | | 4 5 | | | | | | | | |
| Novel[h] | | | | | | | | | | | | | | |
| Spen | gi\|24580579 | CG18497 | *spen* | 600.0 | | 7 15 | | | | | | | | |
| RIMS1 | gi\|62472639 | CG33547 | *Rim* | 314.8 | | 6 3 | | | | | | | | |
| EXOSC10 | gi\|161078302 | CG7292 | *Rrp6* | 102.9 | | 1 2 | | | | | | | | |
| ZNF346, JAZ | gi\|24583489 | CG17098 | | 73.2 | | 3 7 | | | | | | | | |
| SNED1 | ▽ | | | | | 6 2 | | | | | | | | |

[a] Proteins were identified by LC–MS-MS after separation by PAGE. Proteins identified in two out of two or two out of three preparations are shown (preparations are numbered 1, 2, and 3). Proteins not reproducibly detected are summarized in Table S1 in the supplemental material. Proteins generally accepted to be common contaminants, such as ribosomal proteins, are not shown. NH, no apparent/unambiguous homologue could be detected by BLAST searches at NCBI.

[b] The name of the human protein is given to aid comparison with previous studies of human spliceosomal complexes. Proteins are grouped in organizational and/or functional subgroups.

[c] The presence of a protein is indicated by a number which represents the absolute number of peptides sequenced for that protein (see Materials and Methods for details).

[d] Data from previous proteomics studies of MS2-affinity-selected human B and C complexes formed on an adenovirus-derived pre-mRNA substrate (MINX) (11) or PM5 pre-mRNA (5), respectively, are also included. ●, presence of a given protein in these previously analyzed complexes.

[e] ▲, homologues could not be assigned on the basis of BLAST data.

[f] —, extensive homology between protein family members prevents assignment of *Drosophila* homologues on the basis of BLAST data.

[g] Human homologue previously detected in one or more human spliceosomal complex.

[h] Never detected in human spliceosomal complexes.

[i] ▽, homology with several proteins, but only within conserved domains.

[j] LSm5 and LSm6 are difficult to detect by MS and have been included for completeness.

splicing (except Prp17) and most components of the EJC and TREX complexes were absent or not consistently identified. These data indicate that both human and *Drosophila* Ftz B complexes are largely devoid of any contaminating C complexes. Taken together, our results demonstrate that *Drosophila* homologues of the majority of proteins found in human spliceosomes are also present in *Drosophila* spliceosomes.

Despite similarities in their compositions, some notable differences also exist. Human B complexes contain a few proteins not found in their *Drosophila* counterparts, including, among others, the DEAD box protein DDX9, ELAV, components of the RES (retention and splicing) complex, and the mRNP/EJC-associated proteins Aly/REF, UAP56, and ELG (Table 1). Differences between *Drosophila* and human B complexes formed on Ftz-M3 pre-mRNA are also seen in the subset of SR proteins that are detected. Whereas SFRS12 was exclusively present in *Drosophila* B complexes, 9G8, SRp38, and SRp30c appear to be specifically associated with human Ftz B complexes. The extensive homology between members of the hnRNP protein family does not allow a conclusive assignment of *Drosophila* homologues on the basis of data from BLAST searches. Nonetheless, it is clear that the *Drosophila* B complex contains significantly fewer proteins of the hnRNP protein family, despite the fact that the hnRNP protein family in *Drosophila* appears to be nearly as complex as that in humans (32). Only the hnRNP protein CG17838 was consistently found in purified *Drosophila* Ftz B complexes, in comparison with over 15 different hnRNP proteins in human complexes. This indi-

cates that, in general, fewer hnRNP proteins associate with the pre-mRNA in *Drosophila* nuclear extract, which could explain why *Drosophila* H/A complexes exhibit a lower S value than their human counterparts (Fig. 2). Interestingly, several proteins, including IGF2BP1, RACK1, ZFR2, RBM15B, RBM42, and the hnRNP proteins D, K, L, U, and UL2, were associated with human B complexes formed on Ftz-M3 pre-mRNA but were not detected previously in affinity-purified human B complexes formed on the MINX-M3 substrate (11). As different pre-mRNA substrates were used in these studies but the conditions used to isolate B complexes were identical, these proteins might represent substrate-specific components or, alternatively, contaminants.

A number of proteins that were detected in *Drosophila* Ftz B complexes were absent from human Ftz B complexes. These include, among others, the U1-associated protein FBP11 and the splicing factor SF1. Interestingly, several *Drosophila*-specific proteins were identified whose human homologue either was not previously found in spliceosomal complexes or simply could not be identified by BLAST searches. Several of these newly identified, spliceosome-associated proteins possess motifs found in many human spliceosomal proteins, such as an RRM or zinc finger (see Discussion for details). Many of them were also found in B complexes formed on a second pre-mRNA substrate (Zeste62) and/or in *Drosophila* C complexes, suggesting that they are bona fide spliceosomal proteins and not contaminants. These data provide first insights into the *Drosophila* spliceosome at the level of the B complex and illustrate that its protein composition is very similar to that of humans. However, despite this remarkable conservation in composition, novel factors were also identified, some of which have no apparent counterparts in humans.

**Affinity selection of *D. melanogaster* B complexes assembled on the short intron of the Zeste gene.** To purify *D. melanogaster* spliceosomal B complexes assembled on a short intron, we constructed a pre-mRNA substrate containing the 62-nt Zeste intron plus flanking exon sequences and three MS2 binding sites at its 3′ end (Fig. 1D). The *Drosophila* 62-nt Zeste intron contains a polypyrimidine tract (in contrast to many other short *Drosophila* introns), and it has been shown previously to be spliced in vitro in *Drosophila* nuclear extract (52). When in vitro splicing was performed with Kc nuclear extract, splicing intermediates were first observed after 25 min (Fig. 1E, lane 5) and products after 50 min (lane 9). Native gel analysis revealed that B complexes first formed after 15 min of incubation (Fig. 1F, lane 3). A diffuse band migrating above the B complex appeared after 30 min and most likely corresponds to the C complex (Fig. 1F, from lane 6). Therefore, the 20-min time point was chosen to purify Zeste62 B complexes.

B complexes formed on the Zeste62-M3 pre-mRNA substrate were affinity-selected as described above for Ftz B complexes, except that splicing was performed for 20 min. *Drosophila* spliceosomes formed on Zeste62-M3 pre-mRNA exhibited a sedimentation behavior identical to that of those formed on Ftz-M3 pre-mRNA (Fig. 2A), i.e., H and A complexes peaked in fractions 9 to 13 (~20S to 25S) and B complex in fractions 17 to 19 (~45S). Affinity-purified complexes (Fig. 2B, left panels) contained nearly equimolar amounts of uncleaved pre-mRNA, as well as U1, U2, U4, U5, and U6 snRNAs. Neither splicing intermediates nor products could be identified in the eluate by autoradiography, confirming that highly pure precatalytic B complexes had been isolated.

**Protein composition of affinity-purified *Drosophila* B complexes formed on Zeste62-M3.** Proteins recovered from purified Zeste62-M3 B complexes were separated by 1D PAGE and subsequently identified by LC–MS-MS. The protein composition of *Drosophila* B complexes assembled on Zeste62-M3 and Ftz-M3 pre-mRNA is highly similar (Table 1). In addition to those proteins shared by *Drosophila* and human Ftz B complexes, most proteins present in *Drosophila* Ftz-M3 B complexes but absent from human Ftz-M3 B complexes were also found in Zeste-M3 B complexes. Proteins identified in both *Drosophila* B complexes are strong candidates for being bona fide spliceosomal proteins commonly found in all *Drosophila* B complexes, irrespective of the pre-mRNA substrate.

Zeste62-M3 B complexes contained several proteins not detected in *Drosophila* Ftz-M3 B complexes. These include proteins whose homologue was previously detected in human spliceosomes, such as RBM10; HCNGP; AGGF1 (an A complex-associated protein in humans; 3); SPF45, a U2-associated protein with a dual function in alternative splicing and DNA repair (10); PPIL4; Abstrakt; Q9BRR8; MORG1; and the EJC-associated protein Acinus. Zeste62-M3 B complexes also contained three additional SR proteins (9G8, Tra2, and Rbp1-like/Srp20), as well as the hnRNP proteins hrp40.2 and CG30122. Proteins not previously detected in human spliceosomal complexes that were solely found in Zeste B complexes include the *Drosophila* homologues of SLC7A2, NCOA5, and DNAJC17, as well as the *Drosophila* protein CG5343. Whether these proteins contribute specifically to the assembly/function of spliceosomes on short introns is currently not clear.

In contrast, only a few proteins were reproducibly detected in *Drosophila Ftz* B complexes but were absent from Zeste62-M3 B complexes. These include Kin17 (hsKin17), which was previously detected in human spliceosomal complexes, and the newly identified proteins PRPF38B and CG13597, a protein of unknown function with no clear human homologue. Both of these proteins are only represented by one or two peptides in Ftz-M3 B complexes, so it is possible that they have been partially lost during purification or that they might simply be contaminants. On the other hand, these differences in composition may in some cases be due to substrate-specific differences, including intron length. Taken together, there appear to be only small differences in the composition of B complexes assembled on the "short" Zeste62 intron and the "long" Ftz intron.

**EM of *D. melanogaster* spliceosomal B complexes formed on Ftz-M3 pre-mRNA.** To gain first insights into the structure of *Drosophila melanogaster* spliceosomes, MS2 affinity-purified Ftz B complexes were loaded onto a glycerol gradient containing glutaraldehyde in order to fix the particles (26). Particles from the B complex peak gradient fractions were then negatively stained with uranyl formate for EM analysis. In Fig. 3A, a typical EM overview is shown on the left. The particles are evenly dispersed and appear to be intact, as no smaller fragments are visible. Individual particles of the most-prominent view are shown on the right, together with a schematic drawing of their general morphology. Drosophila B complexes exhibited a maximum dimension of ~40 nm. In the most-prominent view of the particles, two domains can be distinguished: a triangular body, which consists of three branches connected to a central mass, and a globular head region posi-

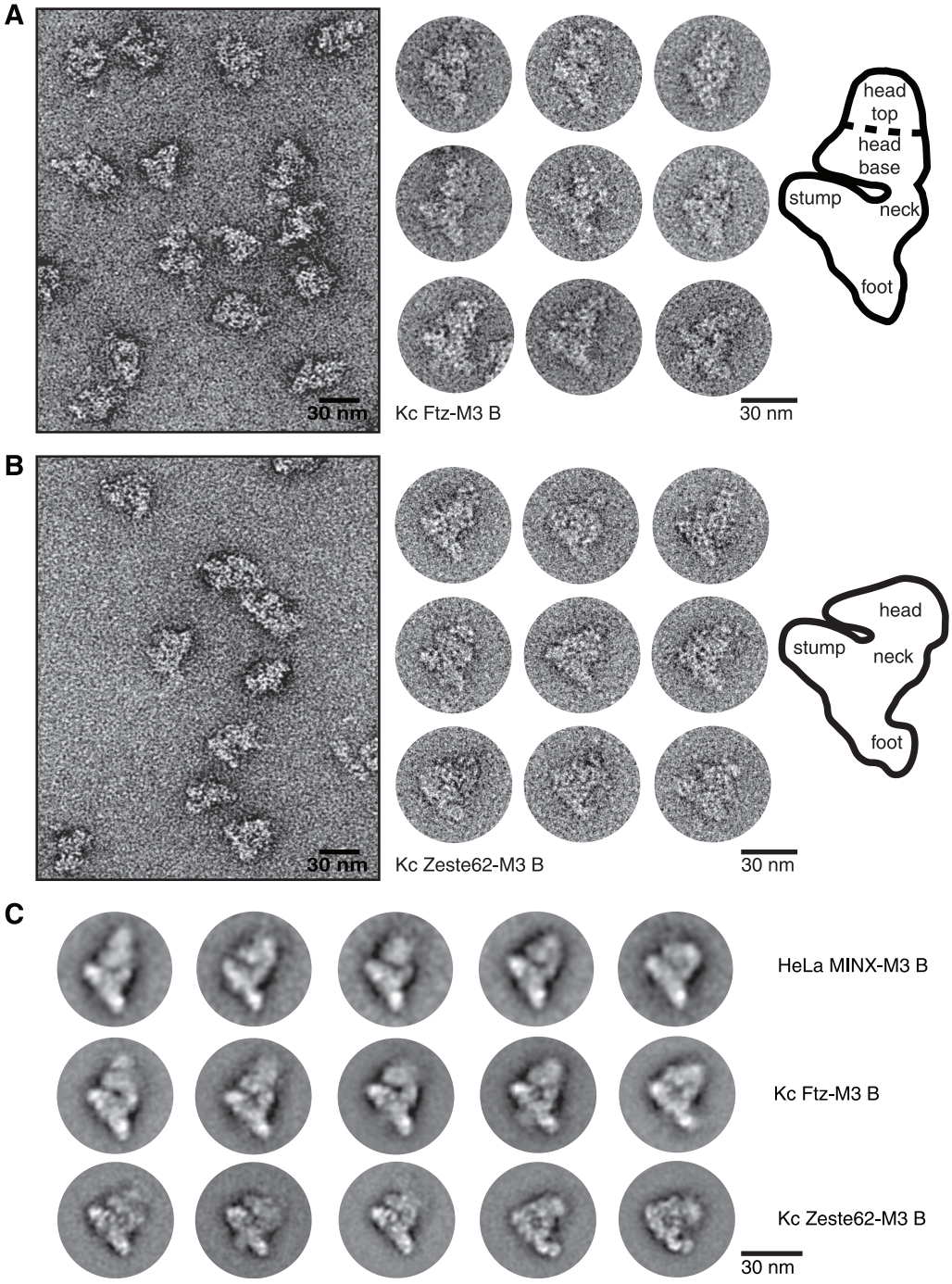


FIG. 3. EM of affinity-purified *Drosophila* B complexes reveals structural similarities with human B complexes. (A and B) EM overviews and single-particle images of negatively stained *Drosophila* B complexes formed on Ftz-M3 (A) or Zeste62-M3 (B) pre-mRNA in Kc nuclear extract. A schematic of the predominant view of *Drosophila* Ftz or Zeste62 B complex with the various regions/domains indicated is shown at the right. (C) Galleries of representative class averages of human B complexes formed on an adenovirus-derived pre-mRNA (top row), *Drosophila* B complexes formed on Ftz-M3 (middle row), or Zeste62-M3 (bottom row) pre-mRNA. For each class, approximately 40 to 50 individual images were averaged. The bar corresponds to 30 nm.

tioned toward one side of the body. The lower branch of the triangular body is referred to as the "foot," the massive upper branch as the "stump," and the slim branch as the "neck" (Fig. 3A, right). The head is attached on the "neck-stump" side and consists of two regions: the elongated "head base" located close to the body and a globular domain "head top" located further away from the body. 2D class averages were generated, and a gallery of several representative class averages is shown in Fig. 3C (middle row). Importantly, the general morphological features of the individual particles

are also observed in the 2D class averages. A visual comparison with 2D class averages of human B complexes formed on an adenovirus-derived pre-mRNA substrate (MINX-M3) (Fig. 3C, upper row) reveals that the *Drosophila* B complex closely resembles its human counterpart in size, shape, and morphology, including apparent conformational heterogeneity in the head region as previously reported for the human B complex (11). Thus, not only the protein composition but also the structure of the *Drosophila* spliceosomal B complex is strikingly similar to its human counterpart. These data show a remarkable conservation of spliceosome composition and appearance across two evolutionarily distant organisms.

**EM analyses of *D. melanogaster* spliceosomal B complexes formed on Zeste62-M3 pre-mRNA.** To investigate whether there are differences in the appearance of *Drosophila* spliceosomal B complexes assembled on long versus short introns, B complexes formed on the Zeste62-M3 pre-mRNA were also subjected to EM analyses. The EM overview (Fig. 3B, left panel) shows evenly dispersed particles of comparable size. A gallery of individual images of the most-predominant view of the particle, with a schematic drawing, is shown on the right. In contrast to *Drosophila* B complexes formed on the Ftz-M3 pre-mRNA, which appear rhombic in shape, Zeste62-M3 B complexes exhibit a more-compact, triangular appearance and also have slightly smaller dimensions (~35-nm maximum); this is particularly apparent in the 2D class averages (Fig. 3C, lower panel). Nevertheless, the massive axis formed by the foot, central mass, and stump is clearly distinguishable. In addition, the head region of *Drosophila* Zeste62-M3 B complexes is smaller than this region of their human or *Drosophila* Ftz-M3 counterparts. Furthermore, an additional, crooked domain is visible at the distal tip of the foot in several class averages. Compared to Ftz-M3 B complexes, Zeste62-M3 B complexes exhibit more well-defined internal structures, indicating that they are structurally more homogeneous. In summary, the 2D projections of *Drosophila* B complexes assembled on Zeste62-M3 pre-mRNA have significant similarities with those of Ftz-M3 *Drosophila* B complexes. The observed structural dissimilarities could result from slight differences in the protein composition of these complexes, the smaller intron size, and/or differences in the orientation of particles on the carbon film.

**Isolation of *D. melanogaster* spliceosomal C complexes assembled on an MS2-tagged Ftz pre-mRNA substrate lacking a 3′ exon.** To determine whether there is an exchange of *D. melanogaster* spliceosomal proteins during the B-to-C transition, we next set out to isolate *Drosophila* C complexes. Recently, highly pure and catalytically active human C complexes assembled on an MS2-tagged pre-mRNA substrate lacking a 3′ ss and 3′ exon were purified from HeLa nuclear extract and analyzed by MS (5). To isolate *Drosophila* C complexes, we modified the Ftz pre-mRNA substrate such that it contained a 5′ exon with three MS2 binding sites, an intron with a branch site, and a 29-nt pyrimidine tract but lacked a 3′ ss and 3′ exon (Fig. 4A). To test whether this substrate, designated M3-Ftz-longPYT, was able to undergo the first step of splicing, in vitro splicing was performed with Kc nuclear extract. Splicing intermediates (i.e., excised 5′ exon and lariat intron) were first observed after 30 min (Fig. 4b, lane 4) and accumulated at later time points. As expected, no splicing products were observed. Analysis of splicing complexes on a native polyacrylamide gel revealed that first A and B, and later, C complexes were formed (Fig. 4C). After 180 min of incubation, a significant portion of unspliced pre-mRNA was still present and most likely represents earlier stages of spliceosomal assembly, i.e., H, E, A, B, and B* complexes. In order to reduce the amount of these contaminating complexes, RNase H digestion with DNA oligonucleotides complementary to nt −6 to −17 and nt −18 to −30 of the 5′ exon (relative to the 5′ ss) was performed. The levels of pre-mRNA, but not cleaved 5′ exon, were reduced (Fig. 4B, compare lanes 7 and 9). Visualization of splicing complexes on a native gel prior to and after cleavage by RNase H confirmed that the levels of H, A, and B complex, but not C complex, were reduced (Fig. 4C, cf. lanes 2 and 7 and lanes 6 and 8).

C complexes formed on the M3-Ftz-longPYT pre-mRNA were affinity purified under conditions identical to those described above for the purification of B complexes (i.e., at 150 mM salt and in the absence of heparin), except that splicing was performed for 180 min and was followed by RNase H digestion. Spliceosomal C complexes peaked in fractions 19 to 22 after glycerol gradient centrifugation, indicating that *Drosophila* C complexes exhibit a slightly higher sedimentation coefficient (~50S) than B complexes (40S to 45S) (Fig. 4E). Affinity-purified complexes contained splicing intermediates, U2, U5, and U6 but no U1 and U4 (Fig. 4D, upper panel), consistent with their designation as C complexes. However, in addition to splicing intermediates, purified complexes also contained unspliced pre-mRNA (Fig. 4D, lower panel). As both U1 and U4 were essentially absent, the latter complexes most likely represent activated B* spliceosomal complexes. Quantification of the radioactive bands corresponding to first-step intermediates and residual unspliced pre-mRNA revealed that the eluate contained ~75% C complex, with the remaining 25% presumably B* complexes.

**Comparison of the protein compositions of *Drosophila* spliceosomal B and C complexes.** Proteins present in affinity-purified *Drosophila* C complexes were separated by protein gel electrophoresis and subsequently identified by LC–MS-MS. Approximately 100 proteins were reproducibly identified, homologues of which were in most cases also previously found in human C complexes (Table 1) (5). Sixty of the C complex-associated proteins were also consistently observed in purified *Drosophila* Ftz B complexes. These include nearly all U2 and U5 snRNP proteins, Prp19/CDC5 complex and Prp19-related proteins, and most SR proteins, plus numerous other non-snRNP proteins. If the protein composition of purified *Drosophila Ftz* B and C complexes is compared, several differences are apparent. As expected, U1 and U4/U6 snRNP proteins are lost during C complex formation, consistent with U1 and U4 snRNA dissociation. Note that the *Drosophila* protein sans-fille (snf) is the functional homologue of both U1-A and U2-B″ and is therefore still represented in the table as a U1 snRNP protein. The apparent absence of LSm proteins is consistent with the results of previous studies in humans and yeast demonstrating the loss of these proteins during catalytic activation (5, 9). Additional proteins lost include 17S U2-related proteins (e.g., SPF30 and U2AF35/65), tri-snRNP-specific proteins (110K, 65K, and 27K), and a number of non-snRNP proteins. Interestingly, there appears to be some general reduction in the peptide num-

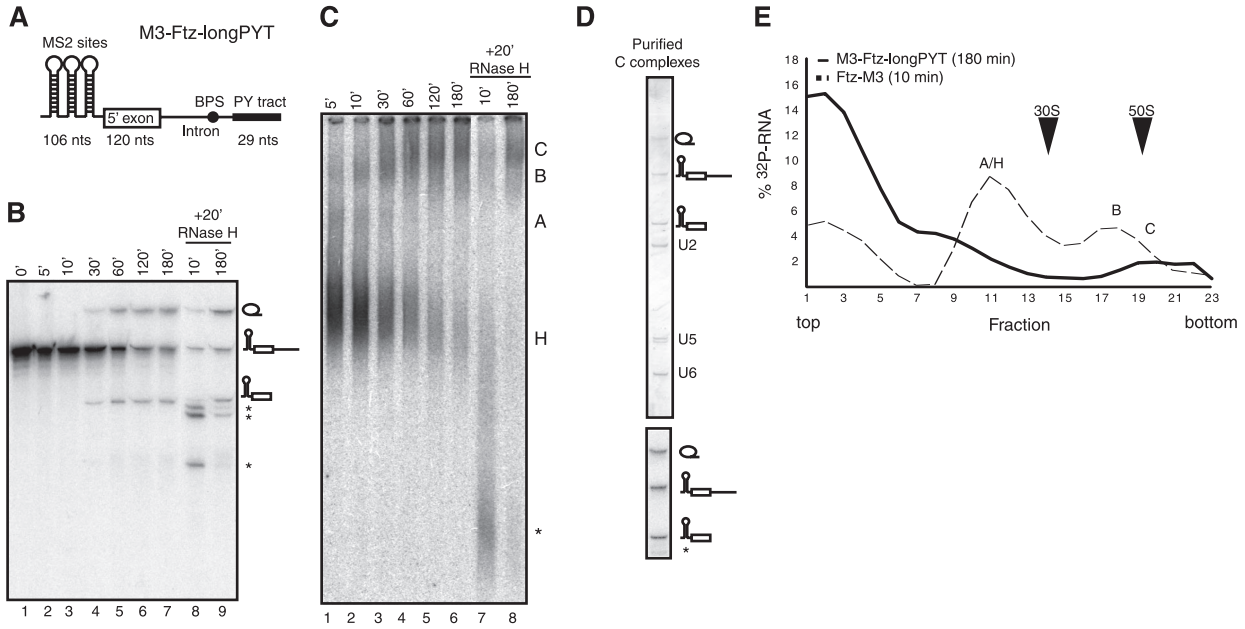


FIG. 4. Affinity purification of *Drosophila* C complexes. (A) Schematic representation of the M3-Ftz-longPYT pre-mRNA substrate. PY tract, polypyrimidine tract; nts, nucleotides. (B) Splicing was performed with M3-Ftz-longPYT pre-mRNA in Kc cell nuclear extract for the times indicated (in minutes) above each lane. In lanes 8 and 9, DNA oligonucleotides complementary to the Ftz 5′ exon (nt +6 to +17 and nt +18 to +30 relative to the 5′ ss) were added after 10 min or 180 min of splicing, and the reaction mixture was incubated for an additional 20 min to allow RNase H digestion of early spliceosomal complexes. RNA was recovered, analyzed on a 12% polyacrylamide–8 M urea gel, and visualized by autoradiography. Splicing substrates, intermediates, and products are indicated on the right. *, RNase H digestion product. (C) Splicing complexes were analyzed on a 3.5% native polyacrylamide gel and visualized by autoradiography. The positions of the H, A, B, and C complexes and RNase H digestion products (*) are indicated on the right. (D) RNA composition of MS2-affinity-purified C complexes. Gradient containing Ftz C complexes formed in Kc extract (fractions 19 to 22) was pooled and subjected to MS2 affinity selection. Bound complexes were eluted with maltose, and RNA was recovered, separated by denaturing PAGE, and visualized by silver staining (upper panel) or by autoradiography (lower panel). The positions of the snRNAs, unspliced pre-mRNA, and splicing products/intermediates are indicated on the right. (E) Splicing reaction mixtures containing Ftz-M3 pre-mRNA or M3-Ftz-longPYT pre-mRNA were subjected to glycerol gradient centrifugation. The radioactivity contained in each gradient fraction was determined by Cherenkov counting and expressed as the percentage of the total $^{32}$P-RNA loaded onto the gradient. Sedimentation coefficients, indicated by arrowheads, were determined by analyzing the UV absorbance of fractions of a reference gradient containing prokaryotic ribosomal subunits.

bers observed for SF3a/b proteins, suggesting that, similar to the situation in humans (5), these proteins are also destabilized in *Drosophila* during the B-to-C transition (and thus less abundant in C complexes). However, additional studies, i.e., immunoblotting, will be required to clarify whether these proteins are truly underrepresented.

A number of proteins, on the other hand, were solely/predominantly identified in C complexes, indicating that they are recruited during catalytic activation or the first step of splicing. These include the known second-step factor Prp22 and DEAD box helicases and peptidyl-prolyl isomerases (PPIases) that were also identified in purified human C complexes (e.g., Abstrakt, DDX35, PPIL2, and PPIG) (5). Most factors comprising the core of the EJC in vertebrates (i.e., eIF4A3, Magoh, and Y14), as well as additional EJC-associated proteins, were either first detected in C complexes or appear to be more abundant in C than in B, as evidenced by an increase in the number of peptides identified. In addition, several proteins that were previously identified in purified human C complexes join *Drosophila* spliceosomes during the transition from B to C complex, namely cactin, SKIV2L2, NOSIP, CXorf56, DGCR14, NKAP, and CDK10. Further, there are additional putative novel spliceosomal proteins which appear to associate with the *Drosophila* spliceosome during the B-to-C transition. These include, among others, three *Drosophila* proteins (CG11808, CG6209, and CG5877) without apparent homologues in humans. Previous studies reported that components of the Prp19/CDC5 complex and related proteins are more abundant in the human C complex than in the B complex, based on an increase in the number of peptides sequenced for these proteins and immunoblotting data (5). The apparent stabilization of these proteins during the B-to-C transition appears to occur in *Drosophila* as well, as there are consistently more peptides of these proteins detected in the *Drosophila* C complex than in the B complex. In summary, the compositional dynamics of the spliceosome during the transition from the precatalytic B complex to the catalytically active C complex are largely conserved between humans and flies.

**MS2 affinity purification of spliced and unspliced *Drosophila* mRNPs.** To determine whether EJC proteins are deposited on *Drosophila melanogaster* mRNAs in a splicing-dependent manner, we also affinity purified spliced and unspliced Ftz mRNPs. To isolate unspliced Ftz mRNPs, we constructed a substrate lacking the intronic region of Ftz-M3 pre-mRNA (i.e., corresponding to spliced Ftz-M3 mRNA; FtzΔI-M3) (Fig. 5A). $^{32}$P-labeled Ftz-M3 pre-mRNA or Ftz-M3 mRNA (from the construct lacking an intron) was incubated for 180 min under splicing conditions in *Drosophila* Kc cell nuclear extract.

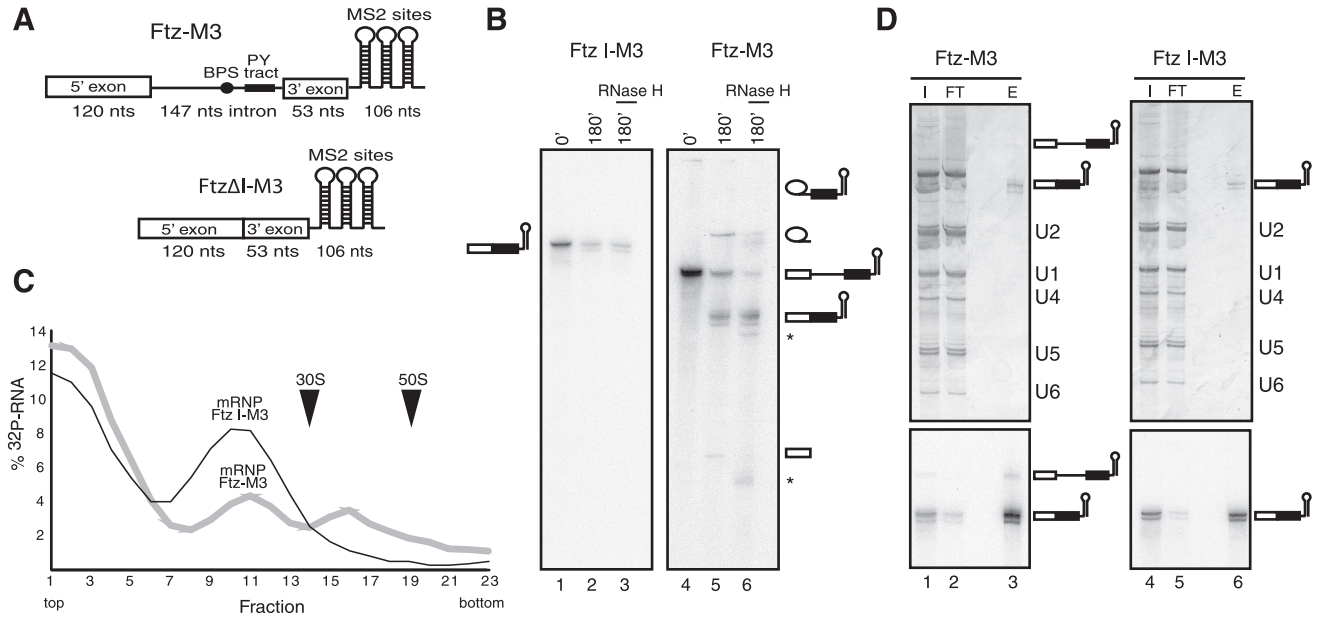


FIG. 5. MS2 affinity selection of in vitro spliced and unspliced mRNPs. (A) Schematic representation of the Ftz-M3 pre-mRNA and FtzΔI-M3 mRNA. PY tract, polypyrimidine tract; nts, nucleotides. (B) FtzΔI-M3 or Ftz-M3 RNA was incubated for 180 min under splicing conditions in Kc cell nuclear extract, followed by oligonucleotide-directed cleavage by RNase H. Splicing substrates, intermediates, and products are indicated on the right. (C) Splicing reaction mixtures containing Ftz-M3 or FtzΔI-M3 were loaded onto a linear 10-to-30% (vol/vol) glycerol gradient containing G150 buffer. The radioactivity contained in each fraction was determined by Cherenkov counting. The profiles show the percentage of radioactivity in each fraction. Sedimentation coefficients (indicated by arrowheads) were determined by analyzing the UV absorbance of fractions of a reference gradient containing prokaryotic ribosomal subunits. (D) Gradient fractions 10 to 12 were pooled and subjected to MS2 affinity selection. Bound complexes were eluted with maltose, and RNA was recovered, separated by 12% denaturing PAGE, and visualized by silver staining (upper panel) or by autoradiography to detect radioactive species (lower panel). Lanes 1 and 4, 2.5% of the input (I) pooled gradient fractions. Lanes 2 and 5, 2.5% of the flowthrough (FT) of the amylose column. Lanes 3 and 6, 10% of the eluted (E) mRNPs. The positions of snRNAs, unspliced pre-mRNA, and splicing products are indicated on the right.

After 180 min, a large portion of the pre-mRNA had been converted to mRNA (Fig. 5B, cf. lanes 4 and 5). To reduce the amount of unspliced pre-mRNA or intron-containing intermediates, RNase H digestion was carried out with two oligonucleotides complementary to Ftz intron sequences. As shown in Fig. 5B, the addition of the oligonucleotides led to a reduction in pre-mRNA and intron lariat levels (cf. lanes 5 and 6). The levels of spliced and unspliced Ftz-M3 mRNA, on the other hand, remained unaffected.

*Drosophila* mRNPs were affinity purified under conditions identical to those described above for the purification of B and C complexes. mRNPs formed on spliced and unspliced Ftz-M3 mRNA exhibit a similar sedimentation behavior, with a main peak in both cases in fractions 10 to 12 (~20S to 25S) (Fig. 5C). In reactions performed with Ftz-M3 pre-mRNA, a second peak was observed in fractions 15 to 17 (~35S) and appeared to contain spliceosomal complexes resistant to RNase H-mediated cleavage (data not shown). The mRNP peak gradient fractions (i.e., fractions 10 to 12) were pooled and subjected to MS2 affinity selection. The eluates contained essentially only mRNA and, in reactions performed with Ftz-M3 pre-mRNA, small amounts of pre-mRNA (<8%) (Fig. 5D), confirming that relatively pure mRNPs had been affinity purified.

**Protein composition of purified *Drosophila melanogaster* mRNPs.** Proteins were recovered from the eluates, separated by 1D PAGE, and subsequently identified by LC–MS-MS (Table 2). More than 75 proteins were reproducibly detected in spliced Ftz mRNPs. However, a subset of these, in particular copurifying snRNP proteins or components of the 3′-end processing machinery (e.g., CPSF1 to -6), likely do not represent bona fide mRNP proteins, but rather contaminants (see Discussion). The majority of proteins identified in spliced mRNPs were also found in unspliced mRNPs, indicating that they can be deposited onto the mRNA in a splicing-independent manner. Proteins shared by both include the cap binding proteins CBP20 and CBP80, SR proteins, several snRNP proteins, and a large number of proteins previously detected in human spliceosomes and/or in *Drosophila* spliceosomes (see Table 1). A total of 19 proteins were exclusively detected in spliced mRNPs, indicating that they are deposited on the mRNA as a consequence of splicing. These included core components of the mammalian EJC, such as MLN51 (Barentsz in *Drosophila*), Magoh, and Y14. eIF4AIII was also detected in two out of three preparations of unspliced mRNP, but based on peptide numbers, it appears to be more abundant in spliced mRNPs. Only two components of the THO/TREX complex (THOC2 and THOC5), which play a role in mRNA export, were identified in spliced mRNPs. Interestingly, homologues of the human EJC-associated factor Upf2, and also Upf1, which are both involved in NMD, were identified solely in spliced mRNPs. Homologues of three proteins previously identified in human spliceosomes (SKIV2L2, EIF2C2, and RANBP2) were also specifically associated with spliced *Drosophila* mRNPs, together with a few proteins not previously detected in human spliceosomal complexes (e.g.,

TABLE 2. Protein composition of affinity-selected *Drosophila* mRNPs[a]

| Human protein[b] | *Drosophila melanogaster* protein | | | | Presence of or absolute no. of peptides sequenced for protein[c] in indicated mRNP | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GenBank accession no. | CG no. | Gene name | Mol. mass (kDa) | Ftz-M3 | | | FtzΔI-M3 | | | AdML-M3[d] | AdMLΔI-M3[d] |
| | | | | | 1 | 2 | 3 | 1 | 2 | 3 | | |
| EJC and TREX | | | | | | | | | | | | |
| eIF4AIII | gi\|24645031 | CG7483 | *eIF4AIII* | 45.6 | 18 | 21 | 18 | 6 | | 8 | ● | |
| MLN51 | gi\|45551990 | CG12878 | *btz* | 86.1 | 11 | 6 | 7 | | | | | |
| Magoh | gi\|17136332 | CG9401 | *mago* | 17.3 | 4 | 3 | 4 | | | | ● | |
| Y14 | gi\|24651979 | CG8781 | *tsu* | 19.0 | 4 | 3 | 2 | | | | ● | |
| Acinus | gi\|24585110 | CG10473 | *hkl* | 83.7 | 8 | 3 | 5 | 15 | 2 | 11 | ● | |
| SAP18 | gi\|17738001 | CG6046 | *bin1* | 17.3 | 3 | 2 | 1 | 4 | 2 | 4 | ● | |
| THOC2 | gi\|24581012 | CG31671 | *Tho2* | 188.5 | 1 | 1 | 1 | | | | ● | |
| THOC5 | gi\|19922862 | CG2980 | *thoc5* | 70.9 | 2 | | 2 | 1 | | | ● | |
| Upf1 | gi\|18859757 | CG1559 | *Upf1* | 129.9 | 2 | 2 | 7 | | | | | |
| Upf2 | gi\|24640488 | CG2253 | *Upf2* | 140.1 | 7 | 4 | 8 | | | | | |
| Cap binding complex | | | | | | | | | | | | |
| CBP20 | gi\|17738031 | CG12357 | *Cbp20* | 17.7 | 4 | 2 | | 1 | | 2 | ● | ● |
| CBP80 | gi\|24639721 | CG7035 | *Cbp80* | 93.2 | 11 | 16 | 11 | 22 | 15 | 11 | ● | ● |
| hnRNP: —[e] | gi\|22024201 | CG30122 | | 81.0 | 3 | 1 | 1 | 6 | 3 | 5 | | |
| SR proteins, SR related | | | | | | | | | | | | |
| SF2/ASF | gi\|21358097 | CG6987 | *SF2* | 28.4 | 25 | 13 | 12 | 26 | 18 | 14 | ● | |
| 9G8 | gi\|24582360 | CG10203 | *xl6* | 27.9 | 13 | 6 | 7 | 13 | 12 | 9 | ● | |
| SRp20 | gi\|24641772 | CG1987 | *Rbp1-like* | 16.8 | | 7 | 5 | 4 | 10 | 7 | ● | ● |
| SRp55 | gi\|28571701 | CG10851 | *B52* | 42.8 | 14 | 11 | 8 | 22 | 14 | 20 | ● | |
| SC35 (SFRS2) | gi\|45552333 | CG5442 | *SC35* | 21.4 | 9 | 7 | 6 | 10 | 11 | 10 | | |
| RSRC2 | gi\|24642204 | CG6340 | | 56.7 | 4 | | 1 | 3 | | 3 | | |
| SRm300 | gi\|161080550 | CG7971 | | 127.3 | | | | 9 | 2 | 3 | | |
| SRm160 | gi\|24663664 | CG11274 | *SRm160* | 107.6 | | | 1 | 1 | 2 | | | |
| NH | gi\|28571148 | CG7065 | | 136.8 | 3 | | 2 | 3 | | 2 | | |
| Copurifiying snRNP proteins | | | | | | | | | | | | |
| B | gi\|17136806 | CG5352 | *SmB* | 21.0 | 3 | | 1 | 3 | 1 | 2 | | |
| D2 | gi\|21357623 | CG1249 | *SmD2* | 13.5 | | | | | 1 | 3 | ● | |
| U1-70K | gi\|17137278 | CG8749 | *U1-70K* | 52.9 | | | 3 | | 1 | 3 | ● | |
| U1-A/U2-B″ | gi\|17737284 | CG4528 | *snf* | 24.6 | 2 | | 1 | 1 | 1 | 1 | | |
| SF3b130 | gi\|24654874 | CG13900 | | 136.6 | | 1 | | 3 | | 2 | ● | |
| PUF60 | gi\|24655242 | CG12085 | *pUf68* | 67.9 | 4 | | 2 | 10 | 3 | 9 | | |
| U5-220K | gi\|20129897 | CG8877 | *prp8* | 279.4 | 6 | 4 | 2 | 17 | 1 | 6 | ● | ● |
| U5-200K | gi\|28574898 | CG5931 | | 244.5 | | | | 10 | 1 | 5 | ● | ● |
| U5-116K | gi\|21357743 | CG4849 | | 110.7 | 1 | | 3 | 6 | 1 | 2 | ● | |
| U5-102K | gi\|24666532 | CG6841 | | 105.2 | 5 | 1 | | 5 | 1 | 3 | ● | |
| U5-100K | gi\|24584994 | CG10333 | | 94.6 | 2 | 2 | 2 | 3 | 2 | 5 | ● | |
| U4/U6-60K | gi\|21355245 | CG6322 | | 61.6 | 1 | | | 1 | | | ● | |
| Copurifiying spliceosomal proteins | | | | | | | | | | | | |
| MGC2803 | gi\|24643383 | CG18809 | | 11.4 | 1 | 1 | 1 | 1 | | | | |
| RBM10 | gi\|24580815 | CG4887 | | 114.2 | 2 | 3 | | 5 | 2 | 5 | | |
| RBM39/RNPC2 | gi\|45550943 | CG11266 | | 66.5 | 1 | | | 2 | | 4 | | |
| TLS/FUS | gi\|24642436 | CG3606 | *caz* | 38.8 | 7 | 3 | 3 | 9 | 2 | 5 | | |
| ZC3H18/LOC124245 | gi\|24640344 | CG1677 | | 109.1 | 7 | 1 | 3 | 6 | | 2 | ● | |
| YBX-1 | gi\|24663131 | CG5654 | *yps* | 38.9 | 14 | 5 | 11 | 16 | 7 | 2 | ● | ● |
| ASR2B | gi\|24585960 | CG7843 | | 107.2 | 14 | 10 | 9 | 16 | 8 | 19 | ● | ● |
| p68/DDX5 or DDX17 | gi\|45551833 | CG10279 | *Rm62* | 78.5 | 3 | | 6 | 15 | 12 | 16 | ● | |
| Abstrakt | gi\|17977678 | CG14637 | *abs* | 69.5 | 1 | | | 4 | | 6 | | |
| KIAA1604 (fSAPb) | gi\|24584968 | CG12750 | *ncm* | 151.6 | | | | 3 | 2 | | | |
| hPRP22 | gi\|20129977 | CG8241 | *pea* | 142.0 | | 2 | 1 | 2 | | 2 | | |
| Hsp73 (CCAP1) | gi\|28571721 | CG4264 | *Hsc70-4* | 71.1 | 11 | 5 | | 5 | 3 | | | |
| hPRP4-kinase/PRPF4B | gi\|19922978 | CG7028 | | 104.0 | | | | 3 | 3 | 4 | | |
| PPIG, SR-Cyp | gi\|28571910 | CG1866 | *Moca-cyp* | 116.4 | 1 | 1 | 1 | 1 | | 2 | | |
| PABP | gi\|45552715 | CG5119 | *PABP* | 69.9 | 18 | 8 | 18 | 13 | 7 | 10 | | |
| RBM4 | gi\|24659981 | CG8597 | *lark* | 39.9 | 14 | 6 | 8 | 9 | 8 | 10 | | |
| ZFR2 | gi\|28574893 | CG5215 | *Zn72D* | 96.0 | 1 | 1 | 1 | 3 | 1 | 2 | | |
| WDR33 | gi\|24644363 | CG1109 | | 90.5 | 5 | 4 | 7 | 8 | 4 | 6 | | |

*Continued on following page*

TABLE 2—*Continued*

| Human protein[b] | *Drosophila melanogaster* protein | | | | Presence of or absolute no. of peptides sequenced for protein[c] in indicated mRNP | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GenBank accession no. | CG no. | Gene name | Mol. mass (kDa) | Ftz-M3 | | | FtzΔI-M3 | | | AdML-M3[d] | AdMLΔI-M3[d] |
| | | | | | 1 | 2 | 3 | 1 | 2 | 3 | | |
| RBM15B | gi\|24586450 | CG2910 | *nito* | 89.1 | 8 | 3 | 3 | 12 | 4 | 8 | | |
| Copurifiying spliceosomal proteins | | | | | | | | | | | | |
| CPSF1 | gi\|45552619 | CG10110 | *CPSF* | 164.7 | 19 | 15 | 17 | 32 | 11 | 19 | | |
| CPSF2 | gi\|24650920 | CG1957 | | 85.4 | 9 | 3 | 4 | 8 | 5 | 6 | | |
| CPSF3 | gi\|24648013 | CG7698 | | 76.8 | 5 | 4 | 10 | 8 | 7 | 8 | | |
| CPSF5 | gi\|116007798 | CG3689 | | 23.0 | 27 | 14 | 14 | 21 | 13 | 19 | | |
| CPSF6 | gi\|21355973 | CG7185 | | 71.1 | 30 | 16 | 22 | 40 | 29 | 27 | | |
| GEMIN5 | gi\|45550470 | CG30149 | *rig* | 137.8 | 7 | 5 | 11 | | 3 | 5 | | |
| FIP1L1 | gi\|24644016 | CG1078 | | 78.6 | 4 | | 3 | 4 | 1 | 4 | | |
| NY-CO-10 | gi\|21357249 | CG10907 | | 56.6 | | | | | 2 | 1 | | |
| CPSF4 | gi\|17137188 | CG3642 | *clp* | 33.5 | | | 2 | 2 | 2 | 3 | | |
| CDK10, PISSLRE | gi\|24652305 | CG1362 | *cdc2rk* | 39.4 | 2 | | | 3 | | 2 | | |
| SKIV2L2 | gi\|17864608 | CG4152 | *lethal(2)35Df* | 118.9 | | 1 | 1 | | | | | |
| EIF2C2 | gi\|24664668 | CG7439 | *Ago2* | 136.9 | 4 | | 3 | | | | | |
| RANBP2 | gi\|45550830 | CG11856 | *Nup358* | 296.4 | | 4 | 3 | | | | | |
| MTHFSD | gi\|161078016 | CG14648 | *growl* | 58.8 | 10 | 31 | 10 | 13 | 14 | 13 | | |
| SPN | gi\|45550712 | CG9684 | | 71.9 | 53 | | 48 | 40 | 37 | 26 | | |
| IGF2BP1 | gi\|116007146 | CG1691 | *imp* | 63.4 | 16 | 7 | 16 | 4 | 3 | 2 | | |
| PIWIL1 | gi\|161076864 | CG6137 | *aub* | 98.6 | 6 | 6 | 10 | 8 | | 6 | | |
| PRPF38B | gi\|24641727 | CG1622 | | 45.4 | | 2 | 2 | 3 | | 3 | | |
| NCOA5 | gi\|17647723 | CG8614 | *Neos* | 42.6 | 4 | | 2 | 4 | | 3 | | |
| C2orf30 | gi\|24583799 | CG6766 | | 59.8 | 5 | 11 | 15 | | 12 | 19 | | |
| APTX | gi\|24648185 | CG5316 | | 76.5 | 8 | 6 | 9 | 5 | 2 | 3 | | |
| LUC7-like | gi\|24665973 | CG7564 | | 48.4 | 2 | | | 6 | 6 | | | |
| PSME3 | gi\|18860055 | CG1591 | *REG* | 28.1 | 1 | | 1 | | | 1 | | |
| NH | gi\|24653446 | CG6209 | *mst101(2)* | 69.0 | 10 | 10 | 9 | 12 | 10 | 10 | | |
| NH | gi\|24656457 | CG8994 | *exu* | 58.0 | 7 | 6 | 13 | 6 | 12 | 8 | | |
| Miscellaneous in both | | | | | | | | | | | | |
| NXF2 | gi\|45551570 | CG4118 | *nxf2* | 96.6 | 4 | 1 | | 1 | | 1 | | |
| Symplekin | gi\|24644386 | CG2097 | | 132.1 | 10 | 11 | 14 | 17 | 9 | 20 | | |
| NH | gi\|24641127 | CG2186 | | 122.7 | 16 | 5 | 10 | 4 | | 2 | | |
| RBM26 | gi\|24585249 | CG10084 | *swm* | 115.7 | 2 | | 4 | 2 | | 3 | | |
| PYGM | gi\|78706832 | CG7254 | *GlyP* | 97.0 | 1 | | 3 | 1 | | 1 | | |
| NH | gi\|62484342 | CG9776 | | 140.3 | 5 | 5 | 5 | 12 | 4 | 6 | | |
| NH | gi\|24664166 | CG3836 | *stwl* | 112.9 | 9 | 8 | 7 | 6 | 6 | 13 | | |
| NH | gi\|45550425 | CG6701 | | 143.3 | 2 | 3 | 6 | 8 | 9 | 7 | | |
| Miscellaneous in unspliced only | | | | | | | | | | | | |
| RAPTOR | gi\|24640048 | CG4320 | *raptor* | 177.5 | | | 5 | 5 | | 3 | | |
| NH | gi\|24653724 | CG10205 | | 13.6 | 2 | | | | 2 | 2 | | |
| NH | gi\|24649669 | CG17780 | | 39.5 | 5 | | | 5 | | 3 | | |
| NH | gi\|28573576 | CG8929 | | 68.0 | | | | 1 | | 1 | | |
| Miscellaneous in spliced only | | | | | | | | | | | | |
| DNAJA4 | gi\|24646562 | CG8863 | | 45.2 | 5 | 1 | 5 | | | 1 | | |
| NH | gi\|24644051 | CG12001 | | 59.8 | 11 | 13 | 13 | | 12 | | | |
| SLC26A11 | gi\|19922482 | CG5002 | | 65.1 | | 2 | 4 | | | | | |
| DDX43 | gi\|21355075 | CG7878 | | 78.6 | 1 | | 2 | | | | | |
| NH | gi\|24668137 | CG7597 | | 128.4 | 2 | | 2 | 5 | | | | |
| ADAM10 | gi\|24651113 | CG1964 | *kul* | 168.8 | 2 | | 2 | | | | | |
| KIAA0265 | gi\|24655044 | CG5186 | *slim* | 69.7 | 1 | | 2 | | | | | |
| NH | gi\|24583347 | CG5381 | | 76.9 | 3 | 2 | 5 | | | | | |

[a] Proteins were identified by LC–MS-MS after separation by PAGE. Proteins identified in at least two out of three preparations are shown (preparations are numbered 1, 2, and 3). Proteins not reproducibly detected are summarized in Table S2 in the supplemental material. Proteins generally accepted to be common contaminants, such as ribosomal proteins, are not shown. NH, no apparent/unambiguous homologue could be detected by BLAST searches at NCBI.

[b] The name of the human protein is given to aid comparison with previous studies of human mRNP complexes. Proteins are grouped in organizational and/or functional subgroups.

[c] The presence of a protein is indicated by a number which represents the absolute number of peptides sequenced for that protein (see Materials and Methods for details).

[d] Data from a previous proteomics study of MS2 affinity-selected spliced (AdML-M3) or unspliced (AdMLΔI-M3) human mRNPs formed on adenovirus (Ad)-derived pre-mRNA substrate (33) are also included; a dot represents the presence of a given protein in these previously analyzed complexes.

[e] —, extensive homology between protein family members prevents assignment of *Drosophila* homologues on the basis of BLAST data.

DNAJA4, SLC26A11, DDX43, ADAM10, KIAA0265, and CG5381, which has no apparent human homologue). Taken together, these data indicate that, as in humans, many components of the EJC are deposited onto the mRNA in a splicing-dependent manner.

## DISCUSSION

Here we have shown that the composition of spliceosomal B and C complexes is largely conserved between humans and flies. We also identified several *Drosophila*-specific spliceosome-associated proteins which may be novel splicing factors. Our studies revealed that the compositional dynamics during the transition from the precatalytic B to catalytically active C are also very similar in both organisms. By performing EM, we further demonstrated that human and *Drosophila* B complexes share a nearly identical 2D structure. A comparison of *Drosophila* spliceosomes formed on short versus long introns indicated that although there are only small differences in their protein compositions, their 2D structures clearly differ. Finally, by comparing the composition of unspliced and spliced *Drosophila* mRNPs, we provide evidence that many components of the EJC are deposited onto *Drosophila* mRNAs as a consequence of splicing. Our studies reveal a remarkable conservation of the composition and structure of human and *Drosophila* spliceosomes, as well as their compositional dynamics during catalytic activation.

**Most spliceosomal proteins are evolutionarily conserved between humans and flies.** Homologues of the vast majority of human spliceosomal proteins were detected in affinity-purified *Drosophila* B and C complexes. Due to the large number of proteins identified in human spliceosomes (over 150), there has been much discussion as to whether all of the identified factors are bona fide spliceosomal proteins. Most of those proteins identified in both human and *Drosophila* spliceosomes very likely comprise the set of functionally important factors common to most metazoan spliceosomes. It is unlikely that all proteins detected in both organisms are directly involved in splicing. Rather, some of them may have important roles in coupling the splicing machinery to other macromolecular complexes in the nucleus, including those involved in transcription and polyadenylation. Proteins found in only human or *Drosophila* spliceosomes are more likely to be contaminants, although some of them may have species-specific roles in splicing (see below). The conserved composition of *Drosophila* and human spliceosomes is also reflected in their 2D structure as evidenced by EM. That is, a highly similar morphology was observed for both B complexes (Fig. 3). This suggests that higher-order interactions and the general spatial organization of spliceosomal subunits are also conserved between these evolutionarily distant organisms.

**Identification of putative, novel spliceosomal proteins in *Drosophila*.** A number of novel proteins, whose homologues have not been detected to date in human spliceosomal complexes, were identified in *Drosophila* spliceosomal complexes (summarized in Table S3 in the supplemental material), suggesting that they might be specifically involved in *Drosophila* splicing or in coupling the *Drosophila* spliceosome to other molecular machines in the nucleus. Several proteins were found in *Drosophila* complexes assembled on both the Zeste62 and Ftz pre-mRNA and thus could represent *Drosophila*-specific splicing factors required for the splicing of most introns. It should be noted, however, that some of the compositional differences observed been Drosophila and human spliceosomes might not be species-specific differences but rather arise due to differences in the cell lines used, the proteins expressed in them, and/or the extracts derived from them.

Some of the newly detected proteins have no known function but contain motifs or domains often found in splicing factors, adding credence to the idea that they are bona fide spliceosomal proteins. For example, CG7564 is closely related to the yeast U1-associated protein LUC7, which is involved in 5′ ss recognition in yeast (15). QKR58E-1, whose function is unknown, contains a KH domain, which is found in many proteins involved in RNA metabolism. CG1622 contains a Prp38 domain and is closely related to the tri-snRNP-associated *Saccharomyces cerevisiae* protein Prp38. No functions are known for the proteins rin and neos, but both contain an RRM motif, which is also characteristic of proteins involved in RNA metabolism. Aub and pep are also of particular interest as both proteins are thought to be involved in mRNA metabolism (http://beta.uniprot.org/). Aub has been shown to be required for proper localization of oskar mRNA in developing oocytes (reviewed in reference 21), and pep contains a C2H2-type zinc finger and is associated with hnRNP complexes (1). Many of the other proteins identified have been implicated in other cellular processes, such as DNA repair, cell cycle regulation, or protein transport (http://beta.uniprot.org/) (see Table S3 in the supplemental material). While this suggests that they might be contaminants, it is also possible that they have dual functions (i.e., an additional role in splicing) and/or are involved in coupling the spliceosome to other nuclear processes. Further studies are required to determine which of the newly identified proteins represent true components of *Drosophila* spliceosomes.

**The compositional dynamics of the spliceosome are evolutionarily conserved.** Comparative proteomics of human spliceosomes isolated at different stages has revealed that the spliceosome's composition is highly dynamic (reviewed in references 23 and 59). The most-dramatic change appears to occur during catalytic activation, which is accompanied by large RNP remodeling events (5). A comparison of the composition of precatalytic B and catalytic C complexes from *Drosophila* revealed that the recruitment and release of factors during catalytic activation is highly conserved between humans and flies. As in humans, ~40 B complex proteins are lost during the B-to-C transition in *Drosophila*, suggesting that these factors are required prior to the first step of splicing. On the other hand, ~35 factors are recruited at this stage. As is the case in humans, factors required for the catalytic steps of splicing (e.g., Prp22), as well as components of the EJC complex, are first detected in the C complex (see below). However, in contrast to the situation in humans, the second-step factors Prp16 and Prp18 were not detected in affinity-purified *Drosophila* C complexes and Slu7 was detected in only one *Drosophila* C complex preparation. Thus, these factors may be less stably associated and for the most part lost during the purification of *Drosophila* C complexes or, alternatively, they may be generally difficult to detect by MS. As in humans, the second-step factor hPrp17 and the first-step factor hPrp2, on the other

hand, were found already in *Drosophila* B complexes, suggesting that these factors are recruited prior to step I of splicing. Taken together, our results indicate that several functionally important protein recruitment events taking place after B complex formation are conserved among humans and flies.

The results of recent studies indicate that two major destabilization and stabilization events accompany the B-to-C transition in humans. Specifically, marked changes in the absolute number of peptides sequenced during MS coupled with the results of immunoblotting experiments indicated that there is a reduction in proteins comprising the U2-associated heteromeric complexes SF3a and SF3b (5). In contrast, proteins of the Prp19/CDC5 complex (and related proteins) appear to be more abundant (5). This suggests that the former proteins are destabilized, whereas the latter are stabilized, during catalytic activation—an indication that RNA-protein and protein-protein interactions are remodelled during this stage. A comparison of the peptides sequenced for SF3a/b and Prp19/CDC5-associated proteins in *Drosophila* B versus C complexes suggests that similar destabilization/stabilization events (i.e., RNP remodeling) are also occurring in *Drosophila*. The apparent conservation of these remodeling events underscores their importance during the catalytic steps of pre-mRNA splicing.

**Compositional and structural differences between B complexes formed on a long versus short *Drosophila* intron.** In contrast to mammals, a large percentage of introns in *Drosophila* pre-mRNAs are less than 80 nt in length. These short introns cannot be spliced in mammals and may accordingly exhibit compositional differences from "long" *Drosophila* introns, which can be spliced in HeLa nuclear extract. It is thought that there are mechanistic differences between the initial recognition of splice sites defining small versus long *Drosophila* introns (52). Here we compared the composition of spliceosomal complexes formed on the 62-nt Zeste intron with that of complexes formed on the 147-nt-long Ftz intron. Interestingly, the results of previous studies indicate that the sequence requirements for A complex formation differ between Ftz and Zeste62 in that the latter requires both the 3′ and 5′ ss for complex formation (52), whereas, similar to vertebrate introns, an A-like complex forms on the isolated Ftz 3′ exon plus flanking sequences, including the BPS and polypyrimidine tract (50). As both introns have similar 3′ ss sequences and uninterrupted polypyrimidine tracts of similar length, i.e., 12 (Zeste) versus 11 (Ftz), the mechanistic basis for this difference is not clear, but this region of Zeste62 apparently does not support stable U2 snRNP association on its own. Initial attempts to isolate pure Zeste62 and Ftz A complexes were not successful, and thus, we instead affinity selected B complexes.

Only a few proteins were differentially associated with Zeste62 versus Ftz B complexes. One particularly interesting difference is the presence of three additional SR proteins (9G8, SRp20, and SFRS10) in Zeste62 B complexes. SR proteins play important roles during early splicing complex formation and are involved in splice site recognition, bridging the 5′ and 3′ ends of the intron and stabilizing the interaction of snRNPs with the pre-mRNA (reviewed in reference 47). Thus, the expanded subset of SR proteins found in Zeste62 B complexes could potentially reflect unique cross-intron interactions in short introns that are required for stable A complex formation. Zeste62 complexes also specifically contained four novel proteins (CG7255, neos, CG17187, and CG5343). CG17187 is a member of the DNAJ family of proteins, several members of which are found in human snRNPs or spliceosomes (e.g., DNAJC6 or SPF31), and neos contains an RNA binding motif (RRM). To date, no functions have been described for either of these proteins. CG7255 contains an amino acid permease domain (http://www.expasy.org/prosite/) and is probably involved in the transport of amino acids, whereas CG5343 contains a domain of unknown function (DUF667) and has been suggested to be a transcription factor. As the sequences of Zeste62 and Ftz differ, the recruitment of unique proteins may be related purely to sequence differences rather than to intron length. Thus, it is presently not clear if one or more of these proteins are specifically involved in forming spliceosomal complexes on short introns. Clarification of this issue will require a detailed dissection of the function of those proteins specifically associated with Zeste62 spliceosomal complexes.

Despite their nearly identical protein composition, EM studies indicated that there are structural differences between Zeste62 and Ftz B complexes. The former appear more compact and have a smaller head region. Interestingly, the head domain of the human B complex has been proposed to contain the main components of the A complex (46). Consistent with apparent mechanistic differences between long and short introns at the early stages of splicing complex formation, a smaller head domain suggests differences in the spatial organization and/or composition of components comprising this domain. In the future, a more-detailed analysis, including EM immunolabeling of spliceosomal components and 3D reconstructions, may reveal in more detail structural differences between spliceosomes formed on short and long *Drosophila* introns.

**EJC complex proteins are recruited in a splicing-dependent manner after B complex formation.** In addition to B and C complexes, we also affinity selected *Drosophila* mRNPs formed on Ftz mRNA. MS analyses detected ~75 proteins copurifying with affinity-purified spliced or unspliced mRNPs. Given that both mRNPs have an S value of ~22S, a subset of the identified proteins likely represent contaminants and/or are present in substoichiometric amounts. These include, among others, CPSFs (cleavage/polyadenylation specificity factors) and many snRNP proteins, which are not considered to be mRNP associated. Despite the large number of proteins identified, a comparison of spliced and unspliced mRNPs can still provide information about proteins that are recruited to mRNPs as a consequence of splicing.

Spliced *Drosophila* mRNPs contained several proteins not present in their unspliced counterparts, including components of the EJC. Specifically, we detected *Drosophila* homologues of the core components of the mammalian EJC (i.e., eIF4AIII, MLN51, Magoh, and Y14). With the exception of eIF4AIII, these proteins were exclusively found in preparations of spliced Ftz-M3 mRNP. The presence of eIF4AIII in unspliced mRNP is consistent with the results of recent pull-down experiments performed with extracts from *Drosophila* S2 cells expressing a glutathione *S*-transferase-tagged version of eIF4AIII; these studies revealed the association of eIF4AIII with both spliced and unspliced pre-mRNA/RNAs, indicating that the binding of eIF4AIII is not splicing dependent in *Drosophila* (60). Human eIF4AIII, Magoh, and Y14 form a trimeric complex in the

absence of MLN51 (Barentsz in *Drosophila*), suggesting that they may bind independently of MLN51 (2). Indeed, purified, human spliced mRNPs contained eIF4AIII, Magoh, and Y14 but not MLN51 (33). Interestingly, eIF4AIII, Magoh, and Y14, but not Barentsz, were detected in *Drosophila* C complexes (Table 1), suggesting that Barentsz (MLN51) associates independently of the other core EJC factors at a stage after C complex formation.

Acinus and SAP18, which in humans are components of the EJC outer shell (53), were also detected but in all mRNP preparations, indicating that their association with mRNA is not necessarily dependent on splicing. Pinin could only be identified in one preparation of spliced mRNP, indicating that it might be lost during purification or is highly underrepresented. Homologues of the human EJC-associated factor Upf2 and the non-EJC factor Upf1, which are involved in NMD, were also identified in spliced *Drosophila* mRNPs; the NMD factor Upf3, however, was detected in only one preparation (see Table S2 in the supplemental material). Interestingly, different from mammals, in *Drosophila* the presence of an EJC is not required for NMD (17). Our data indicate that Upf1 and Upf2 are nonetheless recruited to *Drosophila* mRNPs in a splicing-dependent manner. Surprisingly, homologues of the human EJC proteins RNSP1, Aly/REF, and UAP56 could not be identified in any of our *Drosophila* mRNP preparations. The latter two are also components of the THO/TREX complex, which plays a role in mRNA export in humans (reviewed in reference 42). The other TREX-associated proteins (i.e., THOC1, THOC2, THOC3, THOC5, and THOC7) which were present in human mRNPs (33) were absent or highly underrepresented (based on absolute peptide numbers) in purified *Drosophila* mRNPs. The apparent absence of the THO/TREX complex in *Drosophila* mRNPs is consistent with previous reports in which the knockdown of THO/TREX factors (with the exception of UAP56) did not affect the export of most mRNAs in *Drosophila* (44).

Our results are consistent with the idea that a stable EJC also forms upstream of the exon-exon junction of spliced *Drosophila* mRNPs. However, formal proof for such a complex is still lacking. Although the results of recent pull-down experiments suggest that there are differences in the splicing-dependent loading of EJC factors in humans and *Drosophila* (60), our results indicate that the temporal association of several EJC proteins with the pre-mRNA is similar in both organisms. That is, most of the known EJC proteins identified in *Drosophila* spliced mRNPs were first detected in C complexes, suggesting that, as in humans, they associate during or subsequent to catalytic activation. Taken together, our analyses of *Drosophila* splicing complexes and spliced mRNPs suggest a remarkable degree of conservation among metazoans at the level of their protein composition and in terms of their compositional dynamics.

## ACKNOWLEDGMENTS

We are grateful to Gabi Heyne, Thomas Conrad, and Hossein Kohansal for excellent technical assistance. We also thank Monika Raabe and Uwe Plessmann for their excellent help in MS analysis, Reinhard Rauhut for help in homology searches, and Christian Merz for help in preparing MS2-MBP and also for providing PCR template for transcription of the Ftz-M3 pre-mRNA.

This work was supported by grants from the DFG, the European Commission (EURASNET-518238), Fonds der Chemischen Industrie, and the Ernst Jung Stiftung to R.L. and a YIP grant from EURASNET to H.U.

## REFERENCES

1. **Amero, S. A., M. J. Matunis, E. L. Matunis, J. W. Hockensmith, G. Raychaudhuri, and A. L. Beyer.** 1993. A unique ribonucleoprotein complex assembles preferentially on ecdysone-responsive sites in *Drosophila melanogaster*. Mol. Cell Biol. **13:**5323–5330.
2. **Ballut, L., B. Marchadier, A. Baguet, C. Tomasetto, B. Séraphin, and H. Le Hir.** 2005. The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. Nat. Struct. Mol. Biol. **12:**861–869.
3. **Behzadnia, N., M. M. Golas, K. Hartmuth, B. Sander, B. Kastner, J. Deckert, P. Dube, C. L. Will, H. Urlaub, H. Stark, and R. Lührmann.** 2007. Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J. **26:**1737–1748.
4. **Berget, S. M.** 1995. Exon recognition in vertebrate splicing. J. Biol. Chem. **270:**2411–2414.
5. **Bessonov, S., M. Anokhina, C. L. Will, H. Urlaub, and R. Lührmann.** 2008. Isolation of an active step I spliceosome and composition of its RNP core. Nature **452:**846–850.
6. **Black, D. L.** 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. **72:**291–336.
7. **Boehringer, D., E. M. Makarov, B. Sander, O. V. Makarova, B. Kastner, R. Lührmann, and H. Stark.** 2004. Three-dimensional structure of a pre-catalytic human spliceosomal complex B. Nat. Struct. Mol. Biol. **11:**463–468.
8. **Celniker, S. E., and G. M. Rubin.** 2003. The *Drosophila melanogaster* genome. Annu. Rev. Genomics Hum. Genet. **4:**89–117.
9. **Chan, S. P., D. I. Kao, W. Y. Tsai, and S. C. Cheng.** 2003. The Prp19p-associated complex in spliceosome activation. Science **302:**279–282.
10. **Chaouki, A. S., and H. K. Salz.** 2006. *Drosophila* SPF45: a bifunctional protein with roles in both splicing and DNA repair. PLoS Genet. **2:**1974–1983.
11. **Deckert, J., K. Hartmuth, D. Boehringer, N. Behzadnia, C. L. Will, B. Kastner, H. Stark, H. Urlaub, and R. Lührmann.** 2006. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol. Cell. Biol. **26:**5528–5543.
12. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11:**1475–1489.
13. **Dube, P., R. Tavares, R. Lurz, and M. van Heel.** 1993. The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. EMBO J. **12:**1303–1309.
14. **Förch, P., and J. Valcárcel.** 2003. Splicing regulation in *Drosophila* sex determination. Prog. Mol. Subcell. Biol. **31:**127–151.
15. **Fortes, P., D. Bilbao-Cortés, M. Fornerod, G. Rigaut, W. Raymond, B. Séraphin, and I. W. Mattaj.** 1999. Luc7p, a novel yeast U1 snRNP protein with a role in 5′ splice site recognition. Genes Dev. **13:**2425–2438.
16. **Fribourg, S., D. Gatfield, E. Izaurralde, and E. Conti.** 2003. A novel mode of RBD-protein recognition in the Y14-Mago complex. Nat. Struct. Biol. **10:**433–439.
17. **Gatfield, D., L. Unterholzner, F. D. Ciccarelli, P. Bork, and E. Izaurralde.** 2003. Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. EMBO J. **22:**3960–3970.
18. **Guo, M., P. C. Lo, and S. M. Mount.** 1993. Species-specific signals for the splicing of a short *Drosophila* intron in vitro. Mol. Cell. Biol. **13:**1104–1118.
19. **Hachet, O., and A. Ephrussi.** 2001. *Drosophila* Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. Curr. Biol. **11:**1666–1674.
20. **Hartmuth, K., H. Urlaub, H.-P. Vornlocher, C. L. Will, M. Gentzel, M. Wilm, and R. Lührmann.** 2002. Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc. Natl. Acad. Sci. USA **99:**16719–16724.
21. **Johnstone, O., and P. Lasko.** 2001. Translational regulation and RNA localization in *Drosophila* oocytes and embryos. Annu. Rev. Genet. **35:**365–406.
22. **Jurica, M. S., L. J. Licklider, S. R. Gygi, N. Grigorieff, and M. J. Moore.** 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA **8:**426–439.
23. **Jurica, M. S., and M. J. Moore.** 2003. Pre-mRNA splicing: awash in a sea of proteins. Mol. Cell **12:**5–14.
24. **Jurica, M. S., D. Sousa, M. J. Moore, and N. Grigorieff.** 2004. Three-dimensional structure of C complex spliceosomes by electron microscopy. Nat. Struct. Mol. Biol. **11:**265–269.
25. **Kastner, B., and R. Lührmann.** 1989. Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5′ RNA terminus. EMBO J. **8:**277–286.
26. **Kastner, B., N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, J., F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F.**

**Herzog, J. M. Peters, D. Poerschke, R. Lührmann, and H. Stark.** 2008. GraFix: sample preparation for single-particle electron cryomicroscopy. Nat. Methods **5:**53–55.
27. **Konarska, M. M.** 1989. Analysis of splicing complexes and small nuclear ribonucleoprotein particles by native gel electrophoresis. Methods Enzymol. **180:**442–453.
28. **Le Hir, H., E. Izaurralde, L. E. Maquat, and M. J. Moore.** 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. EMBO J. **19:**6860–6869.
29. **Lim, L. P., and C. Burge.** 2001. A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci. USA **98:**11193–11198.
30. **Makarov, E. M., O. V. Makarova, H. Urlaub, M. Gentzel, C. L. Will, M. Wilm, and R. Lührmann.** 2002. Small nuclear ribonucleoprotein remodelling during catalytic activation of the spliceosome. Science **298:**2205–2208.
31. **Makarova, O. V., E. M. Makarov, H. Urlaub, C. L. Will, M. Gentzel, M. Wilm, and R. Lührmann.** 2004. A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. EMBO J. **23:**2381–2391.
32. **Matunis, M. J., E. L. Matunis, and G. Dreyfuss.** 1994. Isolation and characterization of RNA-binding proteins from *Drosophila melanogaster*. Methods Cell Biol. **44:**191–205.
33. **Merz, C., H. Urlaub, C. L. Will, and R. Lührmann.** 2007. Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. RNA **13:**116–128.
34. **Moore, M. J.** 2005. From birth to death: the complex lives of eukaryotic mRNAs. Science **309:**1514–1518.
35. **Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White, and C. Fields.** 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Nucleic Acids Res. **20:**4255–4262.
36. **Mount, S. M., and H. K. Salz.** 2000. Pre-messenger RNA processing factors in the *Drosophila* genome. J. Cell Biol. **150:**F37–F44.
37. **Nilsen, T. W.** 1998. RNA-RNA interactions in nuclear pre-mRNA splicing, p. 279–307. *In* R. W. Simons and M. Grunberg-Manago (ed.), RNA structure and function. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
38. **Palacios, I. M., D. Gatfield, D. St. Johnston, and E. Izaurralde.** 2004. An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. Nature **427:**753–757.
39. **Park, J. W., and B. R. Graveley.** 2007. Complex alternative splicing. Adv. Exp. Med. Biol. **623:**50–63.
40. **Rappsilber, J., U. Ryder, A. I. Lamond, and M. Mann.** 2002. Large-scale proteomic analysis of the human spliceosome. Genome Res. **12:**1231–1245.
41. **Reed, R.** 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. Curr. Opin. Genet. Dev. **6:**215–220.
42. **Reed, R., and H. Cheng.** 2005. TREX, SR proteins and export of mRNA. Curr. Opin. Cell Biol. **17:**269–273.
43. **Reed, R., and T. Maniatis.** 1985. Intron sequences involved in lariat formation during pre-mRNA splicing. Cell **41:**95–105.
44. **Rehwinkel, J., A. Herold, K. Gari, T. Köcher, M. Rode, F. L. Ciccarelli, M. Wilm, and E. Izaurralde.** 2004. Genome-wide analysis of mRNAs regulated by the THO complex in *Drosophila melanogaster*. Nat. Struct. Mol. Biol. **11:**558–566.
45. **Rio, D. C.** 1988. Accurate and efficient pre-mRNA splicing in *Drosophila* cell-free extracts. Proc. Natl. Acad. Sci. USA **85:**2904–2908.
46. **Sander, B., M. M. Golas, E. M. Makarov, H. Brahms, B. Kastner, R. Lührmann, and H. Stark.** 2006. Organization of core spliceosomal components U5 snRNA loop I and U4/U6 di-snRNP within U4/U6.U5 tri-snRNP as revealed by electron cryomicroscopy. Mol. Cell **24:**267–278.
47. **Sanford, J. R., J. Ellis, and J. F. Cáceres.** 2005. Multiple roles of arginine/serine-rich splicing factors in RNA processing. Biochem. Soc. Trans. **33:**443–446.
48. **Sheth, N., X. Roca, M. L. Hastings, T. Roeder, A. R. Krainer, and R. Sachidanandam.** 2006. Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Res. **34:**3955–3967.
49. **Shi, H., and R. M. Xu.** 2003. Crystal structure of the *Drosophila* Mago nashi-Y14 complex. Genes Dev. **17:**971–976.
50. **Spikes, D., and P. M. Bingham.** 1992. Analysis of spliceosome assembly and the structure of a regulated intron in Drosophila in vitro splicing extracts. Nucleic Acids Res. **20:**5719–5727.
51. **Staley, J. P., and C. Guthrie.** 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell **92:**315–326.
52. **Talerico, M., and S. M. Berget.** 1994. Intron definition in splicing of small *Drosophila* introns. Mol. Cell. Biol. **14:**3434–3445.
53. **Tange, T. Ø., T. Shibuya, M. S. Jurica, and M. J. Moore.** 2005. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. RNA **11:**1869–1883.
54. **Tange, T. Ø., A. Nott, and M. J. Moore.** 2004. The ever-increasing complexities of the exon junction complex. Curr. Opin. Cell Biol. **16:**279–284.
55. **Valadkhan, S.** 2005. snRNAs as the catalysts of pre-mRNA splicing. Curr. Opin. Chem. Biol. **9:**603–608.
56. **van Heel, M., and J. Frank.** 1981. Use of multivariate statistics in analysing the images of biological macromolecules. Ultramicroscopy **6:**187–194.
57. **van Heel, M.** 1989. Classification of very large electron microscopical image data sets. Optik **82:**114–126.
58. **Wieringa, B., E. Hofer, and C. Weissmann.** 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit beta-globin intron. Cell **37:**915–925.
59. **Will, C. L., and R. Lührmann.** 2006. Spliceosome structure and function, p. 369–400. *In* R. F. Gesteland, T. R. Cech, and J. F. Atkins (ed.), The RNA world, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
60. **Zhang, Z., and A. R. Krainer.** 2007. Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. Proc. Natl. Acad. Sci. USA **104:**11574–11579.
61. **Zhou, Z., L. J. Licklider, S. P. Gygi, and R. Reed.** 2002. Comprehensive proteomic analysis of the human spliceosome. Nature **419:**182–185.