Pre-messenger RNA Processing Factors in the *Drosophila* Genome

Stephen M. Mount* and Helen K. Salz‡

*Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, College Park, Maryland 20742-5815; and ‡Department of Genetics, Case Western Reserve University, Cleveland, OH 44106-4955

In eukaryotes, messenger RNA genes are generated by a process that includes coordinated splicing and 3' end formation. Factors essential for the splicing of mRNA precursors (pre-mRNA)¹ in eukaryotes have been identified primarily through the study of nuclear extracts derived from mammalian cells and Saccharomyces cerevisiae genetics. Here, we identify homologues of most known pre-mRNA processing factors in the recently completed *Drosophila* genome. The set of proteins required for RNA processing shows remarkably little variation among eukaryotic species, and individual proteins are highly conserved. In general, proteins involved in the mechanisms of RNA processing are even more conserved than proteins involved in the interpretation of RNA processing signals. The genome does not appear to contain a gene for the U11 RNA, or for a protein unique to the U11 snRNP, which raises the possibility that the U12-dependent spliceosome functions without U11 in *Drosophila*.

Introduction

Most RNA processing factors have been identified in either nuclear splicing extracts derived from mammalian cells or in Saccharomyces cerevisiae (Burge et al., 1999; Kambach et al., 1999; M invielle-Sebastia and Keller, 1999). However, *Drosophila* is extensively used for genetic investigations of complex and regulated splicing. In this review, we survey the recently complete *Drosophila* sequence (Adams et al., 2000) for sequences related to factors identified in these other systems. In many cases, functional data for the *Drosophila* protein are not available, and our assignments are based on the best match among genomes. We have not included genes that have been identified in *Drosophila* for which there is evidence of a role in splicing (see, for example, the list presented in Burnette et al., 1999). This analysis yields a list of 27 genes that encode small nuclear RNAs (snRNAs; see Table I) and a list of 99 genes that encode proteins involved in RNA processing (see Table II). Our survey confirms that the components of the RNA processing machinery are highly conserved.

Analysis

Very few factors identified in other species are absent from the *Drosophila* genome. In general, the *Drosophila* proteins are more closely related to their vertebrate counterparts than to the Saccharomyces cerevisiae proteins.

Methods

Protein sequences of known yeast and human splicing factors were used to query the annotated set of predicted *Drosophila* proteins using BLASTP, and the nucleotide sequence of the genome using blastn, on the NCBI server (A ltschul et al., 1997; http://www.ncbi.nlm.nih.gov/). All identified *Drosophila* genes were used to query the nonredundant database to establish the optimal yeast and human matches. Alignments were generated using the blast two sequences option (Tatusova et al., 1999) or LALIGN (Huang and Miller, 1991). Cytological positions were taken from GadFly (http://flybase.bio.indiana.edu/) or flybase (http://flybase.bio.indiana.edu/), or deduced from the positions of flanking genes.

To identify snRNA genes, the *Drosophila* genome was queried using modified blastn parameters (parameter set A: -r 10 -q -11 -W 8 -G 100 -E 50; B: -r 10 -q -11 -W 7 -G 5 -E 20; C: -r 10 -q -11 -W 7 -G 15 -E 4; D: -r 7 -q -14 -W 7 -G 7 -E 3; and E: -r 4 -q -5 -W 8 -G 10 -E 2).

A curated database containing these results will be available at http://www.wam.umd.edu/~smount/DmRNA factors/table.html.

Results and Discussion

Major and minor snRNP components

*U* snRNAs. Two types of spliceosomes have been previously described (Burge et al., 1999). The more common U2-type spliceosome is responsible for splicing the majority of introns, and the U12-type spliceosome is responsible for splicing a minor class of rare introns (perhaps 0.1% in both humans and flies). The *Drosophila* genome contains multiple copies of the 5 U snRNA found in the major class spliceosomes. We found five genes for U1, six genes for U2, three genes for U4, seven genes for U5, and three genes for U6 (Table I). With the exception of U4-25F, and the U5 genes (which were previously known only by in situ hybridization), these genes had been described previously (A lonso et al., 1984; Das et al., 1987; Saba et al., 1986; Saluz et al., 1988; L o and M ount, 1990). The variant U4-25F has only 69% identity with the major form of fly U4 (Saba...
Table I. Genes for Drosophila snRNAs Involved in Splicing

| Drosophila gene | Accession | Cytological position |
|----------------|-----------|----------------------|
| U1a-21D        | AE003588  | 21D                  |
| U1a-82E        | X53542    | 82E                  |
| U1a-95C        | AE003745  | 95C                  |
| U1a-95C        | AE003745  | 95C                  |
| U1c-95C        | AE003745  | 95C                  |
| U1c-95C        | AE003745  | 95C                  |
| U2-14B         | AE003501  | 14B                  |
| U2-34A         | AE003639  | 34A-34B              |
| U2-34A         | AE003639  | 34A-34B              |
| U2-34A         | AE003639  | 34A-34B              |
| U2-38A         | AE003664  | 38A5-38B4            |
| U2-38A         | AE003664  | 38A5-38B4            |
| U4-38A         | AE003664  | 38A5-38B4            |
| U4-38A         | AE003664  | 38A5-38B4            |
| U4-39B         | AE003669  | 39B1-39B3            |
| U4-25F‡        | AE003610  | 25F                  |
| U5-14B         | AE003501  | 14B                  |
| U5-23D         | AE003581  | 23D                  |
| U5-34A         | AE003639  | 34A-34B              |
| U5-35D         | AE003648  | 35D                  |
| U5-38A         | AE003664  | 38A5-38B4            |
| U5-38A         | AE003664  | 38A5-38B4            |
| U5-63BC        | AE003477  | 63BC                 |
| U6-96Aa        | AE003748  | 96A                  |
| U6-96Aa        | AE003748  | 96A                  |
| U6-96Ab        | AE003748  | 96A                  |
| U6-96Ac        | AE003748  | 96A                  |
| U4atac-82E‡    | AE003603  | 82E                  |
| U6atac-29B‡    | AE003621  | 29B                  |
| U12*           | AE003526  | 73B                  |

*Although the human U12 is only 150 nucleotides, the fly U12 RNA is 238 nucleotides and the alignment, although fragmentary, includes both ends of the RNA.

†Hypothetical. No published data on Drosophila RNA supports these assignments.

et al., 1986), and 68% with human U 4. Although the possibility that these new snRNA genes are pseudogenes cannot be ruled out, they appear likely to be functional because of their highly conserved promoters. In the case of U 4-25F, some of the variation includes compensatory changes that allow formation of conserved stem loop structures. There are four clusters of snRNA genes, including one at 38A B with two U 2, one U 4, and two U 5 genes within 6 kb.

The Drosophila genome also contains introns that resemble the minor class (or U 12) introns first identified in mammals (A dams et al., 2000). These are recognized by the U12-type spliceosome including U 11, U 12, U 4atac, and U 6atac snRNA s in place of U 1, U 2, U 4, and U 6 (H all and Padgett, 1994; Tarn and Steitz, 1996). Identification of snRNA s for the U12-type spliceosome in the genome involved modification of the standard parameters for BLASTN (see Methods). It was possible to find one gene for U 12 snRNA , one gene for U 6atac snRNA , and one gene for U 4atac snRNA . These are almost certainly authentic genes, as critical sequences are conserved. In addition, the highly conserved snRNA promoter is present in each case, including a 9/10 or perfect match to the PSE consensus TAATTCCAA, which is ~52 nucleotides upstream of the start (J ensen et al., 1998; L o and M ount, 1990). In contrast, no gene for the U 11 snRNA was found. Consistent with the absence of a U 11 snRNA , we also failed to find the U 11 35-kD-specific protein (accession N o. NP_008951; W ill et al., 1999). In fact, the U 11 snRNA , which functions in 5’ splice site recognition, may not be required for splicing. The highly conserved minor class 5’ splice sites could be recognized by an unknown protein that acts during the early steps of splicing, by the U 6atac snRNA alone, or by both. This mechanism would be analogous to a situation seen in vitro, where certain vertebrate introns can be processed in the absence of U 1 snRNP if the 5’ splice sites can be recognized by U 6 snRNA (C rispino et al., 1994; T arn and S teitz, 1994).

snRNP Proteins. Each snRNP contains a set of Sm core proteins shared with the other snRNPs and a set of proteins that are specific to that snRNP. All 15 known proteins of the Sm family were identified and are highly conserved. These include seven Sm proteins that bind to the U 1, U 2, U 4, and U 5 snRNPs (B, D 1, D 2, D 3, E, F, and G); the seven related L Sm proteins found in the U 6 snRNP (L SM 2-L SM 8); and the CaSm/L SM 1p protein (B ouveret et al., 2000; T arun et al., 2000). These and subsequent matches are shown in Table II. The table reports, for each Drosophila gene, the GenBank accession number and expected value (the expected number of matches this good or better; A ltschul et al., 1997) for the best human and yeast (Saccharomyces cerevisiae) match.

In addition to the Sm proteins, each snRNP also contains a set of snRNP-specific proteins. As expected, orthologues of the proteins that are contained in both the vertebrate and Saccharomyces cerevisiae snRNPs are easily identified in Drosophila, based on their extensive sequence homology, except that a single Drosophila protein, encoded by the sans-fille (snf) gene, corresponds to both the U 1 snRNP-U 1A and U 2 snRNP-U 2B orthologues. (Polyarpou-Schwarz et al., 1996; Sitzinger et al., 1999), and no additional homologues were found.

Interestingly, the Saccharomyces cerevisiae U 1 snRNP is more complex than the vertebrate U 1 snRNP, with seven additional protein components that are not found in the purified vertebrate U 1 snRNP (G ottschalk et al., 1998; R igaut et al., 1999). Only two of these proteins, Luc7 and Prp40, have easily identifiable Drosophila orthologues. The Drosophila ortholog of Luc 7, CG 7564, is 33% identical to the entire Luc 7 protein (F ortes et al., 1999), and a second Drosophila Luc 7-related protein is 21% identical to the yeast protein (F ortes et al., 1999). We identified a single Prp40-like gene in the Drosophila database. CG 3542 shares 23% identity with the entire yeast Prp40 protein and 41% sequence identity over its entire length of 757 amino acids with the human protein, FB P11. FB P11 was initially identified because it also contains a tyrosine-rich WW domain and like Prp40, it interacts with the splicing factor SF1 (B edford et al., 1997). These observations suggest that the function of these proteins in forming bridges between 5’ splice sites and the branchpoint may be conserved (A bovich and R osbash, 1997). It is likely that these Drosophila proteins, like their human homologues, would not be found in purified U 1 snRNPs, but, nevertheless, do share a function with their yeast counterparts. A second human Prp40-like protein, FB P21, has been described in the literature (B edford et al., 1998). FB P21 is more closely related to the Drosophila CG 4291, with 28% identity over the entire length of the 338-amino acid protein. Because similarity between FB P11/CG 3542 and FB P21/CG 4291 is limited to the WW repeats, FB P21/
| Drosophila gene | Accession No. | Protein | Cytology | Human homologues, E value | Yeast (S.c.), E value |
|----------------|---------------|---------|----------|---------------------------|---------------------|
| snRNP proteins |               |         |          |                           |                     |
|               |               |         |          |                           |                     |
| snRNP proteins |               |         |          |                           |                     |
|               |               |         |          |                           |                     |
| snRNP70k (CG8749) | AAF52471 | U1-70k  | 27C9     | NP_003080, 2E-62          | Snp1, NP_012203, 7E-31 |
| snf (CG4528)* |               |         |          |                           |                     |
| CG5545 | AAF55616 | U1C      | 91E2     | NP_003084, 8E-28          | Yhl1, NP_013401, 4E-06 |
| snf (CG4528)* |               |         |          |                           |                     |
| CG7564† | AAF49330 | Luc7-like | 74C1    | BAA91737, 4E-84           | Luc7, NP_001916, 5E-24 |
| snf (CG4528)* |               |         |          |                           |                     |
| CG3198‡ | AAF46183 | Related to Luc7. | 6C8 | BAA90542, 3E-06          | Luc7, NP_001916, 3E-06 |
| snf (CG4528)* |               |         |          |                           |                     |
| CG1646‡ | AAF56816 | Prp39/Prp42-like | 98F4 | BAA92024, 2E-30          | Prp39, NP_013667, 2E-13 |
| snf (CG4528)* |               |         |          |                           |                     |
| CG5342 | AAF51161 | Prp40-like | 23C | FBPII1, AAD39463, 2E-98  |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG5422 | AAF56225 | Similar to Nam8 |     | NP_03243, 8E-82          | Pub1p, NP_014382, E-28 |
| snf (CG4528)* |               |         |          |                           |                     |
| CG2901 | AAF54466 | FBP21-like | 21D2 | FBPII1, AAC34810, 7E-28  |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG6841 | AAF51017 | 40 kD |       | NP_004809, E = “0”       |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG1953 | AAF58573 | 220 kD |       | NP_006436, E = “0”       |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG5025 | AAF55204 | 200 kD |       | NP_004805, E = “120”     |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG3436 | AAF51356 | 40 kD |       | NP_004805, E = “120”     |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG4849 | AAF6769 | 116 kD |       | NP_004238, E = “0”       |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG6841 | AAF49211 | Pp6-like | 75E1 | BAA37140, E = “0”       |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG10333 | AAF53680 | 100 kD |       | NP_004809, E = “0”       |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG3058 | AAF51017 | 15 kD |       | NP_006692, 5E-78         |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG7757 | AAF49097 | SAP90 |     | NP_004689, E = “144”    |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG6322 | AAF49331 | SAP60 |     | NP_004688, E = “138”    |                     |
| snf (CG4528)* |               |         |          |                           |                     |
| CG17266 | AAF57375 | 20 kD cyclophilin | 42C5 | NP_006338, 4E-80        |                     |

(Table II continues)
### Table II. (continued)

| Drosophila gene | Accession No. | Protein | Cytology | Human homologues, E value | Yeast (S. c.), E value |
|-----------------|---------------|---------|----------|---------------------------|-----------------------|
| **Prespliceosome proteins** |
| U2af38 (CG3582) | Q94535 | U2AF small subunit | 21B7 | NP_006749, 5E-88 | unknown |
| CG3294 | AAF50982 | Related to U2AF small subunit | 25B4-B5 | BAA08533, 1E-20 | unknown |
| U2af50 (CG9998) | Q24563 | U2AF small subunit | 14C1 | NP_009210, E-146 | Mud2, NP_012849, 0.22 |
| CG5836 | CAB64937 | SF1/BBP | 90B1 | CAA03883, E-106 | Bbp1, Ms5, NP_013217, 4E-62 |
| CG12357 | AAF55496 | CBP20 | 90E3-4 | CBP20, NP_031388, 4E-63 | Mud13, NP_015147, 6E-37 |
| Cbp80 (CG7035) | AAF45970 | CBP80 | 4C9-10 | CBP80, NP_002477, E = “0” | GCR3, NP_013844, 3E-23 |
| CG7907 | AAF55893 | CBP80-like | 36C | CBP80, NP_002477, 1E-96 | GCR3, NP_013844, 1E-10 |
| Hcl25E (CG7269) | AAF52261 | UAP56 | 25F1 | NP_004631, E = “0” | Sub2, NP_010199, E-141 |
| CG6627 | AAF48446 | Prp5-like | 15C5 | NP_055644, E = “0” | Psp5, NP_009796, 4E-94 |
| **Catalytically active spliceosome proteins** |
| CG6876 | AAF49655 | Prp31-like | 71B2 | CAB43677, E-141 | Psp31, NP_011605, 3E-22 |
| CG6905 | AAF37383 | Pombe, cdcs5-related | 61C | AAB61210, E = “0” | Cef1, NP_013940, 4E-49 |
| CG10689 | AAF53766 | Prp2-like | 37C | NP_003578, E = “0” | Psp2, NP_014408, E-172 |
| CG8241 | AAF58294 | Prp22-like | 50E | HRH1, NP_004932, E = “0” | Psp22, NP_010929, E=0 |
| CG1405 | AAF48355 | Prp16-like | 12D1 | NP_054722, E = “0” | Psp16, NP_013012, E-160 |
| CG6015 | AAF55949 | Prp17-like | 93F14 | AAC9730, E = “0” | Psp17, NP_016052, 1E-76 |
| CG1420 | AAF56845 | Slu7-like | 98F13 | NP_006416, E-128 | Slu7, NP_010373, 1E-12 |
| Prp18 (CG6011) | AAF55627 | Prp18-like | 91E4 | NP_006666, 4E-93 | Psp18, NP_011520, 4E-05 |
| **SR proteins** |
| SC35 (CG5442) | AAF53192 | SC35 | 33DE | SC35, NP_035488, 5E-53 | Npl3, NP_010720, 3E-12 |
| ASP/SF2 (CG6987) | AAF55300 | ASFSF2 | 89C | ASFSF2, NP_008855, 2E-80 | Npl3, NP_010720, 7E-15 |
| B52 (CG10851) | AAF54968 | B52 | 87F | SRp55, CAB43660, 2E-96; SRp75, NP_005617, 3E-96; Srp40/hrs, Q12343, 1E-78 | Npl3, NP_010720, 3E-14 |
| 9G8 (CG10203) | AAF52454 | 9G8 | 27C | 9G8, NP_006267, 4E-42 | Npl3, NP_010720, 7E-05 |
| RBP1 (CG17136) | AAF54555 | RBP1 | 86C12 | SRP20, NP_003008, 2E-38 | Nsr1, NP_011675, E-07 |
| RBP1-like (CG1987) | AAF48264 | RBP1-like | 11F3 | SRP20, NP_003008, 4E-31 | Nsr1, NP_011675, 4E-14 |
| SRp54 (CG4602) | AAF52825 | SRp54 | 30DE | SRp54, NP_004759, 2E-41 | unknown |
| **SR protein kinases** |
| CG8174 | AAF58140 | SRPKD1 | 52A1 | SRPK1, NP_003128, 9E-76; SRPK2, NP_003129, 3E-77; Sky1, NP_0013943, 3E-44 | Sky1, NP_0013943, 3E-44 |
| CG9085 | AAF51819 | SRPKD2 | 79E4 | SRPK1, NP_003128, 2E-63; SRPK2, NP_003129, 3E-66; Sky1, NP_0013943, 3E-49 | Sky1, NP_0013943, 3E-49 |
| CG8565 | AAF48523 | SRPKD3 | 13F3-4 | SRPK1, NP_003128, 8E-50; SRPK2, NP_003129, 3E-50; Sky1, NP_0013943, 2E-70 | Sky1, NP_0013943, 2E-70 |
| Doa (CG1658) | AAF56832 | LAMMER/CLK kinase | 98F6 | CLK2, NP_003984, 4E-151 | Kns1, NP_013081, 5E-61 |
| **Miscellaneous proteins** |
| Cm (CG3193) | AAF45760 | csr/CLF1 | 2F1 | NP_057736, E = “0” | Clip1, NP_013218, E-123 |
| Puf60 (CG12085) | AAF47501 | PUF60 (poly-U binding) | 62A | AAF56505, 1E-81 | unknown |
| non-l (CG10328) | AAF55347 | PSF | 89C7 | NP_005057, 5E-62 | unknown |
| Ptb (CG2094) | AAF57208 | Ptb/AnRNP I | 100F | NP_002810, E-150 | unknown |
| CG11107 | AAF59269 | Prp43-like | 43B2 | NP_001349, E = “0” | Pps43, NP_011395, E = “0” |
| CG11274 | AAF49848 | Srp160-like | 69F2 | NP_005830, 2E-43 | unknown |

(Table II continues)
CG4291 is unlikely to be related to Prp40. Consistent with this idea, human FBP21 has been found to stably associate with the U2 snRNPs and, therefore, may function at a later stage of spliceosome assembly than does Prp40 (Bedford et al., 1998).

Searches with the yeast U1 snRNP proteins Prp39 and Prp42 have identified only a single homologous sequence. Prp39 and Prp42 belong to a family of TPR repeat proteins (McLean and Rymond, 1998) and share 25% sequence identity with each other over a 270-amino acid region that includes several copies of the TPR repeat motif. We identified a single Drosophila protein, encoded by the CG1646 gene, that shares 25% sequence identity with Prp39 and Prp42 over the same 270 amino acids, and is the best match between the Saccharomyces cerevisiae and Drosophila genomes. The Drosophila crooked neck protein (Crn) is another TPR repeat protein, it’s yeast homologue has been shown to act later in spliceosome assembly (Chung et al. 1999).

Surprisingly, there are three Saccharomyces cerevisiae U1 snRNP proteins that have no clear counterparts in the Drosophila database. No Drosophila proteins, whose best match in the S. cerevisiae genome is Snu71, Snu56 or Nam8, were found. Recent work on Snu56 and Nam8 suggests that these proteins contact the pre-mRNA directly and may anchor the U1 snRNP onto the substrate (Puig et al., 1999; Zhang and Rosbash, 1999), a function that could be dispensable in metazoans because it could be provided by the SR proteins. Alternatively, a similar function may be provided by proteins, such as Drosophila rox8 in the case Nam8, that do not appear to be orthologues (Drosophila rox8 matches three other yeast proteins better than Nam8). Proteins in the U2 snRNP, U5 snRNP and U4/U6.U5 tri-snRNP are generally very conserved, and no
Proteins Required for Splice Site Selection

SR proteins are splicing factors that contain either one or two characteristic RNA-binding domains and an RS domain. These proteins are among the earliest acting proteins in spliceosome assembly (Zahler et al., 1999; Graveley et al., 1999; Tacke and Manley, 1999). There are 11 well characterized mammalian SR proteins: 9G8, SRp20, A SF/SF2, SC35, SRp30c, p54, SRp40, pSRp55, SRp75 (for review see M ount, 1997), N SSR1, and N SSR2 (K omatsu et al., 1999). Individual SR proteins differ with respect to the sequence specificity of their RNA-binding domains, and with respect to their ability to recognize and activate different exonic splicing enhancer sequences. We have identified seven SR protein genes in the Drosophila genome. These include the previously described B52, RBP1, SRp54 (K ennedy et al., 1998) and X 16/9G8 (V orbruggen et al., 2000) genes, as well as Drosophila orthologues of A SF/SF2 and SC35. In addition, we have identified a novel gene, C G1987, that is 95% identical to RBP1.

Phosphorylation of SR proteins is thought to play an important role in controlling spliceosome assembly (Stojdl and Bell, 1999; Y eakley et al., 1999). Both SRPK and L AMMER (or CLK) kinases phosphorylate SR proteins. We have identified three kinases of the SRPK type (C G 8174, C G 9085, and C G 8565) and only one L AMMER kinase, the previously described D oa kinase (D u et al., 1998).

A variety of proteins bind to pre-mRNA (also known as hnRNA), and many of these proteins, defined as hnRN P proteins, have been shown to influence splicing, typically by inhibition of splicing events near their binding sites (C hen et al., 1999). A number of Drosophila hnRN P proteins have been described (e.g., M atunis et al., 1992), and some nuclear RNA-binding proteins without clear homologues in mammalian species have unambiguous roles in the regulation of splicing (e.g., S XL). However, because it is impossible to determine from sequence alone whether a given RNA-binding protein is likely to function in splicing, or even to reside in the nucleus, we have not undertaken an analysis of these proteins. These proteins are discussed in the accompanying article by Lasko (2000).

Genome Contents: Parallels and Differences

The results of our search for RNA processing factors known from studies in mammalian extracts and Saccharomyces cerevisiae genetics indicate that very few RNA processing factors are absent from the Drosophila genome. Indeed, our survey reveals remarkably little variation in this list among yeast, flies, and mammals. As expected, Drosophila proteins are more closely related to their vertebrate counterparts than to the Saccharomyces cerevisiae proteins.

The extensive conservation of the components of the spliceosome between vertebrates and Drosophila supports the suggestion that the primary mode of regulating splicing takes place at the level of spliceosome assembly (L opez, 1998; S taley and G uthrie, 1998). Some of these factors, such as SR proteins, which regulate assembly of the spliceosome on many different RNA s, are well conserved and are easily identifiable. Missing from these tables are the factors that regulate the splicing of specific RNA s. These factors are less likely to be well conserved and, indeed, some may prove to be organism specific (e.g., S XL and T RA). Even the variation we observe among proteins and RNA s with clearly established roles in splicing, per se, is weighted towards the early events in splice site selection. For example, the Drosophila genome is missing a set of U1 snRN P proteins that are found in the yeast U1 snRN P but not in the vertebrate U1 snRN P, and the genome does not appear to contain a gene for the U11 RNA, or for the single known protein unique to the U11 snRN P, suggesting that U12 functions without U11 in Drosophila. Here again, variation is observed in components that function in splice site selection.

We thank Jonathan Roberts (University of Maryland, College Park, M D) for help automating the web searches and for help with the website. We thank J oanne Wise (Case Western Reserve University, Cleveland, O H), Javier Lopez (Carnegie Mellon University, Pittsburgh, PA), and Jim Man ley (Columbia University, New York, N Y) for discussions and comments on the manuscript. We apologize to those whose relevant publications could not be cited due to space limitations.

This work was supported by GM 37991-11 to SMM and NSF- M CB 9904565 to H. K. Salz.

Submitted: 30 M ay 2000
Revised: 22 J une 2000
A ccepted: 22 J une 2000

References

A boxivich, N., and M. Rosbash. 1997. Cross-intronic bridging interactions in the yeast commitment complex are conserved in mammals. Cell. 89:403-412.
A dams, M. D., S. E. Celinker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Maniatis, and J. C. Venter. 2000. The genomic sequence of Drosophila melanogaster. Science. 287:2185-2195.
A Isono, A. E. Beck, J. L. Jorcano, and B. Hovemann. 1984. Divergence of U2 snRNA sequences in the genus of D. melanogaster. Nucleic Acids Res. 12:9543-9550.
A Ischtsch, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.
Bedford, M. T., D. C. Chan, and P. Leder. 1997. FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. EMBO J. (Eur. Mol. Biol. Organ.) J. 16:2376-2383.
Bedford, M. T., R. Reed, and P. Leder. 1998. WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: The proline glycine and methionine-rich motif. Proc. Natl. Acad. Sci. USA. 95:10602-10607.
Bouveret, E., G. Rigaut, A. Shevchenko, M. Wilm, and B. Seraphin. 2000. A Sm-like protein complex that participates in mRNA degradation. EMBO J. (Eur. Mol. Biol. Organ.) J. 19:1661-1671.
Burge, C. B., T. Tuschi, and P. A. Sharp. 1999. Splicing of precursors to mRNAs by the spliceosome. In The RNA World. R. F. Gesteland and J. F. Atkins, editors. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 525-560.
Burnette, J. R., A. R. Hatton, and A. J. Lopez. 1999. Trans-acting factors required for inclusion of regulated exons in the ultrabithorax mRNAs of Drosophila melanogaster. Genetics 151:1517-1529.
C hen, C. D., R. Kobayashi, and D. M. Helfman. 1999. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. Genes Dev. 13:593-606.
Chung, S., M. R. McLean, and B. C. Rymond. 1999. Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6/U5 snRN P addition. RNA 5:1042-1054.
Crispino, J., B. J. Blencowe, and P. A. Sharp. 1994. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRN P. Science. 254:664-673.
D as, B., D. H enning, and R. Reddy. 1987. Structure, organization, and transcription of Drosophila U6 small nuclear RNA genes. J. Biol. Chem. 262:1187-1193.
Du, C. M. E., M. McGiffin, B. Dauwalder, L. Abinov, and W. M. Atwood. 1998. Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in Drosophila. Mol. Cell. 2:741-750.
Fortes, P., D. Bilbao-Cortes, M. Fornerod, G. Rigaut, W. Raymond, B.
