RNA

Splicing factor SF1 from *Drosophila* and *Caenorhabditis*: presence of an N-terminal RS domain and requirement for viability

R. Mazroui, A. Puoti and A. Kramer

RNA 1999 5: 1615-1631

References

Article cited in:

<http://www.rnajournal.org/cgi/content/abstract/5/12/1615#otherarticles>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Notes

To subscribe to *RNA* go to:
<http://www.rnajournal.org/subscriptions/>

Splicing factor SF1 from *Drosophila* and *Caenorhabditis*: Presence of an N-terminal RS domain and requirement for viability

RACHID MAZROUI,¹ ALESSANDRO PUOTI,² and ANGELA KRÄMER¹

¹Département de Biologie Cellulaire, Université de Genève, 1211 Genève 4, Switzerland

²Institut de Zoologie, Université de Fribourg Péroles, CH-1700 Fribourg, Switzerland

ABSTRACT

Splicing factor SF1 contributes to the recognition of the 3' splice site by interacting with U2AF⁶⁵ and binding to the intron branch site during the formation of the early splicing complex E. These interactions and the essential functional domains of SF1 are highly conserved in *Saccharomyces cerevisiae*. We have isolated cDNAs encoding SF1 from *Drosophila* (Dm) and *Caenorhabditis* (Ce). The encoded proteins share the U2AF⁶⁵ interaction domain, a hnRNP K homology domain, and one or two zinc knuckles required for RNA binding as well as Pro-rich C-terminal sequences with their yeast and mammalian counterparts. In contrast to SF1 in other species, DmSF1 and CeSF1 are characterized by an N-terminal region enriched in Ser, Arg, Lys, and Asp residues with homology to the RS domains of several splicing proteins. These domains mediate protein–protein or protein–RNA interactions, suggesting an additional role for DmSF1 and CeSF1 in pre-mRNA splicing. Human (Hs), fly, and worm SF1 interact equally well with HsU2AF⁶⁵ or the *Drosophila* homolog DmU2AF⁵⁰. Moreover, DmSF1 lacking its N terminus is functional in prespliceosome formation in a HeLa splicing system, emphasizing the conserved nature of interactions at an early step in spliceosome assembly. The *Ce-SF1* gene is located in a polycistronic transcription unit downstream of the genes encoding U2AF³⁵ (*uaf-2*) and a cyclophilin (*cyp-13*), implying the coordinate transcriptional regulation of these genes. Injection of double-stranded RNA into *C. elegans* results in embryonic lethality; thus, the SF1 gene is essential not only in yeast but also in at least one metazoan.

Keywords: alternative splicing; KH domain; pre-mRNA splicing; protein–protein interaction; RNA interference; spliceosome; U2AF

INTRODUCTION

The spliceosome, which catalyzes the removal of introns from nuclear pre-mRNAs, is a multicomponent complex consisting of small nuclear ribonucleoprotein particles (snRNPs) and many non-snRNP proteins (for reviews see Krämer, 1996; Nilsen, 1998; Staley & Guthrie, 1998; Burge et al., 1999). Its assembly involves a series of specific RNA–RNA, RNA–protein and protein–protein interactions that align the reactive sites of the pre-mRNA for catalysis. Elements essential for the definition of 5' and 3' splice sites are first recognized during the earliest step in spliceosome assembly (the formation of complex E in mammals or a commitment complex in yeast). U1 snRNP interacts with the 5' splice site by base pairing and elements at the 3' splice site

are targets for the binding of splicing proteins (reviewed in Krämer, 1996; Burge et al., 1999). Although many of the components of and interactions within the spliceosome are conserved from yeast to man, early events in spliceosome assembly differ somewhat between the two organisms.

In humans, SF1 (HsSF1 or mBBP [for mammalian branch site binding protein]) and the large subunit of U2AF (U2AF⁶⁵) interact with one another and cooperatively bind to the pre-mRNA (Abovich & Rosbash, 1997; Berglund et al., 1998a; Rain et al., 1998). SF1 specifically recognizes the branch site (Berglund et al., 1997, 1998b) and U2AF⁶⁵ associates with the polypyrimidine tract (Zamore & Green, 1989). In addition, binding of the N-terminal Arg–Ser-rich domain of U2AF⁶⁵ to the branch site sequence contributes to 3' splice site recognition and facilitates base pairing between U2 snRNA and the branch site during the formation of presplicing complex A (Valcárcel et al., 1996).

Reprint requests to: Angela Krämer, Département de Biologie Cellulaire, Université de Genève, 30, quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland; e-mail: angela.kraemer@cellbio.unige.ch.

In addition to SF1 and U2AF, members of the SR family of splicing factors are essential for the assembly of complex E (for review see Fu, 1995; Manley & Tacke, 1996). These proteins are characterized by one or two N-terminal RNA recognition motifs (RRMs) and a C-terminal domain enriched in Arg-Ser or Ser-Arg dipeptides (RS domain). SR proteins function in constitutive and alternative splicing and mediate a number of protein-protein interactions during the course of spliceosome assembly (Wu & Maniatis, 1993; Kohtz et al., 1994; Roscigno & García-Blanco, 1995; Tronchère et al., 1997). In complex E, SR proteins promote the binding of U1 snRNP and U2AF to the pre-mRNA by simultaneous interactions with the U1 snRNP-specific 70K protein and the small subunit of U2AF, U2AF³⁵ (Wu & Maniatis, 1993). These interactions are mediated through the RS domain of the SR proteins and similar domains in U1-70K and U2AF³⁵ (Wu & Maniatis, 1993; Kohtz et al., 1994).

In contrast to humans, *S. cerevisiae* introns contain nearly invariant branch site sequences but lack extended pyrimidine-rich regions between the branch site and the 3' splice site, and sequences downstream of the branch site are not required until catalysis (reviewed in Rymond & Rosbash, 1992). Furthermore, the early complex or commitment complex (CC) in yeast is assembled in two distinguishable steps (Séraphin & Rosbash, 1989). The formation of CC1 involves binding of U1 snRNP to the 5' splice site. The *S. cerevisiae* homologs of human SF1 and U2AF⁶⁵, ScSF1 (or BBP) and Mud2p, respectively, are incorporated into the complex during the transition to CC2 (Abovich et al., 1994; Abovich & Rosbash, 1997; Rutz & Séraphin, 1999). As in the human system, ScSF1 and Mud2p interact with one another, and ScSF1 specifically recognizes the branch site (Abovich & Rosbash, 1997; Berglund et al., 1997, 1998b; Fromont-Racine et al., 1997; Rain et al., 1998). Mud2p also binds to the pre-mRNA, but, in contrast to U2AF⁶⁵, requires a functional branch site sequence for RNA binding (Abovich et al., 1994).

Functional domains in SF1 are well conserved between humans and *S. cerevisiae*. The N-terminal ~140 amino acids are required for the interaction with U2AF⁶⁵ or Mud2p (Rain et al., 1998). Binding to RNA and specific recognition of the branch site are mediated by a hnRNP K homology (KH) domain immediately C-terminal to the U2AF⁶⁵/Mud2p interaction domain in HsSF1/ScSF1 (Berglund et al., 1997, 1998b; Rain et al., 1998). In addition, one or two zinc knuckles in HsSF1 or ScSF1, respectively, increase the RNA-binding affinity of the proteins (Berglund et al., 1998b). SF1 exists in at least six isoforms in human and mouse cells and the corresponding mRNAs are expressed in a cell type-specific fashion (Toda et al., 1994; Arning et al., 1996; Caslini et al., 1997; Wrehlike et al., 1997; Krämer et al., 1998; Zhang & Childs, 1998). The isoforms are identical in their N-terminal halves (contain-

ing the U2AF⁶⁵ interaction domain, the KH domain, and the zinc knuckle), but differ in the length of a C-terminal Pro-rich region and contain distinct C termini. Recombinant SF1 that lacks the C-terminal half is functional in prespliceosome formation (Rain et al., 1998); thus, the role of these sequences in splicing is unclear.

Here we report the isolation of cDNAs encoding SF1 from *Drosophila melanogaster* (DmSF1) and *Caenorhabditis elegans* (CeSF1) and the partial characterization of the encoded proteins. This study was initiated to test the function of SF1 in vivo. Mutations in the chromosomal loci of the SF1 genes in *Drosophila* and *C. elegans* have not been reported; however we have used RNA interference to demonstrate that SF1 is essential for viability in worms. Moreover, DmSF1 and CeSF1 contain an N-terminal RS domain not found in the human or yeast SF1 proteins, suggesting a specialized or additional function for SF1 in *Drosophila* and *C. elegans*.

RESULTS

Drosophila SF1

cDNAs encoding DmSF1 were isolated by a combination of PCR amplification and library screening (see Materials and Methods). The longest insert (Dm-30) of 2,726 bp contained an open reading frame (ORF) encoding a protein of 773 amino acids with a calculated molecular mass of 86.5 kDa and an isoelectric point of 10.67 (accession number AJ243904). Conceptual translation indicated a domain structure very similar to that of SF1 in other species (see below). The ORF is flanked by 127 bp of 5' untranslated region (UTR). Although no in-frame stop codons are found in this region, several lines of evidence suggest that Dm-30 encodes the full-length DmSF1 cDNA. First, the cDNA corresponds in length to the major mRNA detected by Northern blotting (see below). Moreover, the in vitro translation product of Dm-30 and the major polypeptide detected in *Drosophila* extracts with anti-DmSF1-antibodies comigrate in SDS gels (see below). Second, a number of *Drosophila* Expressed Sequence Tag (EST) sequences start within the first 20 bp of Dm-30 and no ESTs extending further upstream have been found. Third, stop codons are present 55 bp upstream of the 5' end of Dm-30 in a partial genomic sequence of DmSF1 recently deposited in the database (clone BACR29H07, accession number AC007804) and no further upstream ATG initiation codons are present. The 3' UTR of Dm-30 is 272 bp in length and the cDNA ends in a short stretch of polyA 39 bp downstream of a canonical AATAAA polyadenylation signal. Other clones isolated were truncated at their 5' or 3' ends when compared to Dm-30.

Northern blot analysis of polyA⁺ RNA isolated from different developmental stages of *Drosophila* revealed

a major RNA species of ~2.8 kb (Fig. 1), in good agreement with the size of the Dm-30 cDNA. The 2.8-kb mRNA was present in high amounts in early embryos; its abundance started to decline between 6–12 h of embryogenesis and remained at a constant low level during later developmental stages. Thus, this SF1 mRNA is maternally deposited as shown for mRNAs of other splicing proteins, for example DmU2AF⁵⁰ (Kanaar et al., 1993) or the subunits of SF3a (R. Mazroui & A. Krämer, unpubl. results). Three RNAs of very low abundance (of approximately 5.0, 2.2, and 1.8 kb) were detected in 0–12 h embryos and in first- and second-instar larvae, which may be indicative of low amounts of alternatively spliced mRNAs. However, corresponding cDNAs have not been identified by restriction-enzyme digestion or sequencing of cDNAs or PCR-amplified phage inserts.

Combined in vitro transcription and translation of Dm-30 DNA in a reticulocyte lysate yielded a protein with an apparent molecular mass of 97 kDa, which was absent from reactions primed with a HsSF1 cDNA or a control luciferase construct (Fig. 2A). A polypeptide of the same size was detected in the reticulocyte lysate with anti-DmSF1 antibodies (Fig. 2B). This size is considerably larger than the calculated molecular mass of 86.5 kDa. The difference between the apparent and calculated sizes could be explained by posttranslational modifications or the highly charged nature of the N terminus of DmSF1 (see below). A major polypeptide that comigrates with in vitro translated DmSF1 was also decorated by the antibodies in a *Drosophila* Schneider cell nuclear extract and an embryo whole-cell extract (Fig. 2B). Only small amounts of the protein were detected in the corresponding cytoplasmic S100, whereas other S100 preparations did not contain any

FIGURE 2. Analysis of DmSF1 proteins. **A:** In vitro translation. A reticulocyte lysate was primed for in vitro transcription and translation with cDNAs encoding HsSF1 (lane 1), DmSF1 (lane 2), and CeSF1 (lane 3). A luciferase vector was included as a control (lane 4). Proteins were separated by 8% SDS PAGE and visualized by autoradiography. The migration of ¹⁴C-methylated protein markers (Amersham) in kilodaltons is indicated on the left. **B:** Western blot analysis. Proteins (2 μ g) from a *Drosophila* Schneider cell cytoplasmic S100 (S-C, lane 1) or nuclear extract (S-N, lane 2) and an embryo whole-cell extract (E, lane 3) were separated by SDS PAGE and transferred to nitrocellulose. The membrane was incubated with anti-DmSF1 antibodies and antigens were detected by ECL. Five microliters of a 100- μ L in vitro translation reaction primed with Dm-30 cDNA (IV-TL, lane 4) served as a control. The migration of protein markers in kilodaltons (Sigma) is shown on the left. Please note that the proteins markers in **A** and **B** migrate at slightly different positions.

FIGURE 1. Northern blot analysis. PolyA⁺ RNA from different developmental stages of *Drosophila* was probed with the full-length DmSF1 cDNA (Dm-30). The same amount of RNA was loaded in each lane except for the lane containing RNA from third instar larvae, which contained less RNA, as judged by ethidium bromide staining of ribosomal RNAs (not shown). Stages are indicated above the lanes, size markers (in kilobases) are shown on the right and arrows on the left indicate SF1 mRNAs.

detectable SF1 (not shown). Thus, as observed in HeLa cells (Krämer & Utans, 1991), DmSF1 may leak from the nuclei during extract preparation. In the Western blot of the whole-cell extract from embryos (and some preparations of Schneider cell nuclear extracts; data not shown) a larger protein was observed in addition to a series of smaller polypeptides. It is most likely that the smaller polypeptides represent proteolytic degradation products rather than cross-reacting proteins (with, for example, proteins containing related KH domains) or protein isoforms, because their abundance varied between preparations. Whether or not the larger polypeptide represents a cross-reacting protein or a DmSF1

protein isoform is unclear at present. Consistent with the detection of DmSF1 in nuclear extracts, the anti-SF1 antibody exclusively decorated nuclei of *Drosophila* embryos by immunofluorescence (not shown). The antibody also stained pole cells, demonstrating that DmSF1 is ubiquitously expressed during embryogenesis.

A Southern blot of *Drosophila* genomic DNA digested with *Bam*HI, *Eco*RI, *Hind*III, or *Xho*I was probed with a 457-bp fragment encoding sequences encompassing the KH domain and the first zinc knuckle of DmSF1 (data not shown). The probe hybridized to single DNA fragments of 8–10 kb of each digest. The gene for DmSF1 (*dSF1*) was localized to region 90B on chromosome 3R of *Drosophila* by in situ hybridization. This result is consistent with the presence of DmSF1 sequences in clone BACR29H07, which maps to region 90A–90B. Hybridization to other chromosomal loci was not observed. Together these results indicate that DmSF1 is encoded by a single gene and no pseudogenes are present.

A novel alternatively spliced SF1 mRNA in *Drosophila* and mammals

As mentioned above, we failed to isolate cDNAs representing alternatively spliced DmSF1 mRNAs in library screens. Furthermore, several attempts to detect variant cDNAs comparable to those identified for HsSF1 (which differ in the 3' half, see Introduction) by PCR approaches were unsuccessful. However, in database searches we found one EST (accession number AA697004) from a *Drosophila* ovary library that lacked 76 bp within the coding region. Amplification of *Drosophila* polyA⁺ RNA with primers flanking this region and sequencing of the PCR products demonstrated that a corresponding mRNA is expressed at all developmental stages, but at two- to tenfold lower levels than the form including the 76 bp (data not shown). A similar PCR product has also been detected in HeLa cells and a variety of mouse tissues (F. Mulhauser and A. Krämer, unpubl. data).

Comparison of the Dm-30 cDNA to the partial genomic sequence of DmSF1 (clone BACR29H07) indicated that the 76-bp fragment represents exon 4. These sequences correspond to exon 3 (76 bp) in the HsSF1 gene and exon 4 (73 bp) in the CeSF1 gene (Krämer et al., 1998; see below). These data strongly suggest that alternative splicing generates a novel SF1 mRNA, most likely by exon skipping. Conceptual translation of the alternatively spliced SF1 mRNAs yields proteins that are collinear with the full-length proteins up to sequences in the N-terminal portion of the U2AF⁶⁵ interaction domain (Rain et al., 1998; indicated in Fig. 4) and translation continues in an alternative reading frame in the following exon(s). The truncated proteins have calculated molecular masses of 37.5 kDa in *Drosophila*, 26.6 kDa in *C. elegans* and 12 kDa in mammals,

and encode unique C-terminal portions of 14, 16, and 73 amino acids, respectively, that bear no resemblance to sequences in other proteins (data not shown). A protein corresponding in size to the putative product of the alternatively spliced mRNA has not been detected in Schneider cell or *Drosophila* embryo extracts, although the anti-DmSF1 antibody cross-reacts with a recombinant version of the truncated protein in Western blots (R. Mazroui & A. Krämer, unpubl. data).

C. elegans SF1

Two SF1 cDNAs were isolated from *Caenorhabditis*. First, a PCR fragment containing a *C. briggsae* (Cb) EST sequence (accession number R02974) with high homology to human and DmSF1 (Arning et al., 1996) was sequenced. This cDNA (CbSF1; accession number AJ245633) contained a 5' truncated ORF of 401 amino acids. Second, a λ ZAP library prepared from mixed stages of *C. elegans* was screened with a restriction fragment containing the 5' part of the CbSF1 cDNA. The inserts of two double-positive phages contained long ORFs with high similarity to HsSF1 but they lacked the 5' ends. The 3' ends were found 326 and 337 bp downstream of the stop codon in A-rich sequences that contain a variation of the consensus polyadenylation signal AAAAAA that is found in a number of *C. elegans* cDNAs (Blumenthal & Steward, 1997). By the time these cDNAs had been analyzed, a working draft of the genomic sequence of CeSF1 became available in databases (CEY116A8; accession number Z98858). Comparison of the cDNA sequences with the genomic sequence revealed a 100% match to several CeSF1 exons. Despite the fact that *C. elegans* introns can easily be detected by their high content of A- and T-stretches (Blumenthal & Steward, 1997), upstream exons containing ORFs with homology to SF1 from other organisms could not be clearly identified. This was due to a complete divergence between species of the N-terminal sequences of SF1 (Arning et al., 1996; see below) and to mistakes in the initially deposited sequence.

About 25% of the *C. elegans* mRNAs are generated from polycistronic precursors by *trans*-splicing to short spliced leader (SL) RNAs (Blumenthal & Steward, 1997). As a general rule, the SL1 sequence is spliced to the 5' most mRNA in an operon, whereas the SL2 sequence (or a variation thereof) is spliced to downstream mRNAs. We therefore used primers complementary to the SL1 or SL2 sequences and a primer complementary to sequences close to the 3' end of the CeSF1 ORF (LP2) to PCR amplify *C. elegans* cDNA obtained by reverse transcription with 3' primer LP1 (Fig. 3A). In addition, two upstream primers complementary to potential 5' exon sequences (UP1 and UP2) were used in combination with primer LP2. PCR fragments of the sizes expected from putative exon sequences were detected with primers SL2 (1,734 bp), UP1 (1,712 bp), and UP2

FIGURE 3. Cloning of *C. elegans* SF1. **A:** Schematic representation of the *CeSF1* gene structure. The eight exons of the *CeSF1* gene are shown as open boxes drawn to size, introns as lines. Numbers indicate exon or intron sizes in base pairs. The last exon of the *cyp-13* gene located upstream of the SF1 gene is shown as a striped box. The initiating ATG and stop codon of *CeSF1* and a putative polyadenylation signal are indicated above the diagram. Primers used for RT-PCR are shown below. **B:** RT-PCR of *C. elegans* cDNA. PolyA⁺ RNA from mixed stages of *C. elegans* was reverse transcribed with primer LP1. cDNA was amplified with upstream primers UP2, UP1, SL1, and SL2 (as indicated above the figure) and downstream primer LP2. DNA markers (in kilobases) are shown on the left. **C:** Genomic and SL2 sequences upstream of the *CeSF1* coding sequence. The genomic sequence bordering exon 1 of the *CeSF1* gene is shown on top. The SL2 sequence is shown below; sequences derived from the primer are shown in lower case. The coding sequence is shown in bold.

(1,238 bp), but not with primer SL1 (Fig. 3B). After cloning and sequencing of the PCR products, the corresponding exons were clearly identified in the genomic sequence. The SL2 sequence was found immediately upstream of the initiating ATG, which is located at the very 5' end of exon 1 (Fig. 3C).

Conceptual translation of the combined *CeSF1* cDNA sequences yielded a protein of 699 amino acids with a calculated molecular mass of 75.3 kDa and an isoelectric point of 10.53 (accession number AJ243905). In vitro translation of the *CeSF1* cDNA in a reticulocyte lysate resulted in a protein with an apparent size of 86 kDa (Fig. 2A). Thus, similar to DmSF1, the apparent size in SDS gels is higher than the calculated molecular mass. Comparison of the database entry of the genomic sequence with that of the *CeSF1* cDNA revealed the presence of eight exons ranging in size from 73 to ~670 bp and spanning a total of ~6,770 bp (Fig. 3A). Intron sizes vary from 213 to 1,058 bp.

In agreement with the observation that SL2 is usually *trans*-spliced to downstream mRNAs in polycistronic transcripts of *C. elegans* (Blumenthal & Steward, 1997), the termination codon of the *cyp-13* gene, encoding an unusual cyclophilin with an N-terminal RRM, is located 155 bp upstream of the initiation codon of *CeSF1*. Interestingly, the gene encoding the *C. elegans* homolog of the small subunit of U2AF, *uaf-2*, is found 107 bp upstream of *cyp-13* (Zorio & Blumenthal, 1999). The *uaf-2* gene appears to be the first gene in this operon, because its mRNA is spliced to SL1, whereas the mRNA of *cyp-13* is spliced to SL2. Thus, *CeSF1* is encoded in a single operon together with the small subunit of U2AF and a cyclophilin-like protein of unknown function.

Conserved and novel domains in SF1 homologs

A multiple sequence alignment of SF1 from various species revealed a high conservation of structural domains but also important differences (Fig. 4; the alignment also includes the amino acid sequence of an EST encoding part of SF1 from *Arabidopsis thaliana* [accession number B97867]). As expected from the essential function of SF1 in branch site recognition, the central portion encompassing the domains required for RNA binding, the KH domain and the zinc knuckle(s) (Berglund et al., 1997, 1998b; Rain et al., 1998), is most highly conserved. Similar to KH domains in many other proteins, the N-terminal part of this domain, which is thought to directly contact the RNA, shows the highest degree of identity, whereas the C-terminal part, which is essential for the stability of the KH fold, is more diverged (Musco et al., 1996). All SF1 homologs contain one zinc knuckle C-terminal of the KH domain. A second zinc knuckle is found in the yeast and *Caenorhabditis* SF1 proteins, whereas DmSF1 contains a cryptic second zinc knuckle in which the otherwise highly conserved spacing between the first and second Cys residues is reduced from two amino acids to one (Darlix et al., 1995). Otherwise, the sequence of this zinc knuckle is almost identical to the corresponding motif in *CeSF1* and *CbSF1*. Another region of high identity between the SF1 proteins is a stretch of ~30 amino acids located between the KH domain and the zinc knuckle(s). This sequence is shared between SF1 and members of the STAR (signal transduction and activation of RNA) family of proteins (termed QUA2; Vernet & Artzt, 1997).

Significant homology is also found in the region N-terminal of the KH domain that mediates the interaction of HsSF1 or ScSF1 with U2AF⁶⁵ or the *S. cerevisiae* homolog Mud2p, respectively (Rain et al., 1998). The minimal Mud2p interaction domain of ScSF1 (amino acids 41–141) was derived from overlapping cDNA clones obtained in a yeast two-hybrid screen. The U2AF⁶⁵ interaction domain of HsSF1 extends further N-terminal and requires sequences between amino acids 8 and 28 of HsSF1 (Rain et al., 1998; Z. Rhani & A. Krämer, unpubl. data). Consistent with these results we have recently shown that Ser-20 of HsSF1 is located close to a site of specific contact with U2AF⁶⁵ (Wang et al., 1999). A few amino acids on both sides of Ser-20 of HsSF1 are conserved among the metazoan proteins, suggesting that this short sequence may also contribute to U2AF⁶⁵ binding in flies and worms. The N-terminal part of the ScSF1 Mud2p-interaction domain and the corresponding sequences in *Schizosaccharomyces pombe* (Sp) SF1 are less well conserved; thus, the interaction between yeast and metazoan SF1 and U2AF⁶⁵ homologs may in part rely on different principles.

HsSF1 is specifically phosphorylated at Ser-20 by cGMP-dependent protein kinase I (PKG) and phosphorylation interferes with the U2AF⁶⁵ interaction (Wang et al., 1999). Sequences N-terminal to the corresponding Ser-276 in DmSF1 conform to the optimal sequence context for phosphorylation by PKG or the highly related cAMP-dependent protein kinase (PKA), Arg/Lys-Arg/Lys-X-Ser/Thr (Tegge et al., 1995). In CeSF1 this sequence is Glu-Arg-Arg-Ser, thus deviating from the consensus in the first position. Three additional PKA/PKG consensus phosphorylation sites are present in the N-terminal domains of DmSF1 and CeSF1, and DmSF1 contains a fourth site in its Pro-rich C-terminal region.

The most striking difference between the amino acid sequences of DmSF1 and CeSF1 and those of mammals and yeast is found N-terminal to the U2AF⁶⁵ interaction domain. These sequences diverge completely between yeast and mammals and contain no obvious sequence motifs (Arning et al., 1996). In contrast, the N-terminal regions of DmSF1 and CeSF1 are much longer and are characterized by a domain (of 147 and 92 amino acids, respectively) with a high content (86%)

of charged amino acids that are interspersed with Ser residues. Many of these are arranged by alternating charge or as Arg-Ser or Ser-Arg dipeptides, which are hallmarks of SR and SR-related splicing factors (Fu, 1995; Manley & Tacke, 1996). Database searches retrieved several SR proteins, such as plant ASF/SF2 (Lazar et al., 1995), human, chicken, and plant SC35 (Fu & Maniatis, 1992; Vellard et al., 1992; Frugier et al., 1998), the putative *C. elegans* SRp20 (accession number Q18409), and human SRp75 (Zahler et al., 1993). In addition, significant homologies (ranging from 35–60% identity) were found with the RS domains of a number of other splicing factors. These are the *Nicotiana glauca* U2AF⁶⁵ homologs, which have longer N-terminal RS domains than the *Drosophila* or human U2AF⁶⁵ proteins (Zamore et al., 1992; Kanaar et al., 1993; Zorio et al., 1997; Domon et al., 1998) and the C-terminal RS domains of U1-70K from mammals, *Xenopus laevis*, *Drosophila*, and *A. thaliana* (Etzerodt et al., 1988; Query et al., 1989; Mancebo et al., 1990; Reddy et al., 1992). We also found good homology with the TRA protein, which is involved in alternative splicing events that regulate sex determination in *Drosophila* (Boggs et al., 1987). In addition, several other proteins contain RS domains similar to the N-termini of DmSF1 or CeSF1. These include a protein that interacts with the C-terminal domain of RNA polymerase II (Yuryev et al., 1996), two mammalian cyclophilin-like proteins (Page et al., 1995; Nestel et al., 1996; Mortillaro & Berezney, 1998), and several RNA helicases, for example the *C. elegans* ortholog (accession number Q09530) of yeast Prp22p or human HRH1, which function late during the splicing reaction (Company et al., 1991; Ono et al., 1994; Ohno & Shimura, 1996; Schwer & Gross, 1998; Wagner et al., 1998) and a rat RNA helicase that colocalizes with splicing proteins (Sukegawa & Blobel, 1995). The RS domains of many of these proteins have been shown to mediate protein–protein interactions (Fu, 1995; Manley & Tacke, 1996). In contrast the RS domain of human U2AF⁶⁵ binds to the branch site (Valcárcel et al., 1996). Thus, the presence of an RS domain in SF1 from flies and worms suggests a novel and possibly additional function for SF1 in these organisms.

FIGURE 4. Amino acid sequence alignment of SF1 orthologs. SF1 sequences were aligned with the BLAST program. Sequences identical or conserved between more than 50% of the proteins are shown with dark or light shading, respectively. Functional domains and structural motifs are boxed and indicated on top. The borders of the U2AF⁶⁵ interaction domain for human, *Drosophila*, and *C. elegans* SF1 are drawn according to the results obtained with HsSF1 (Rain et al., 1998; Z. Rhani & A. Krämer, unpubl. data), those for *S. cerevisiae* and *S. pombe* SF1 are according to results obtained in a two-hybrid screen with ScSF1 (Rain et al., 1998). The PKG-I phosphorylation site of HsSF1 and putative PKG/PKA phosphorylation sites in other species are boxed and in italics; the conserved PKG-phosphorylated Ser-20 of HsSF1 is indicated by an arrowhead. RGG motifs are boxed. Amino acids encoded by the common exon in HsSF1, DmSF1, and CeSF1 are indicated by x and a broken line above the alignment. Accession numbers for DmSF1, CeSF1, and CbSF1 are mentioned in the text. Other sequences have the following accession numbers: AtSF1, B97967; HsSF1, Y08765; ScSF1, U53877, gene YLR116W; SpSF1, Z78887, gene SPAC2C6.12.

Except for a short sequence of unknown function C-terminal to the zinc knuckles that is conserved in all SF1 proteins except ScSF1 (SF1 in Fig. 4), the C-termini of SF1 vary in length and sequence composition but are Pro-rich in all species (ranging from 31.4% Pro in HsSF1 to 13.4% in CbSF1). The sequence alignment shows three relatively well conserved blocks of multiple Pro residues. In addition, the C-terminal regions of the *Caenorhabditis* SF1 proteins are enriched in Gly residues (25.1% in CeSF1 and 18.4% in CbSF1). We would like to note that the C-terminal sequences differ to a certain extent between CeSF1 and CbSF1. These differences cannot be explained by alternative splicing because the divergent regions do not correspond to splice junctions in the CeSF1 pre-mRNA. CeSF1 and CbSF1 also contain several RGG motifs in this region and one RGG motif is found in the RS domain of DmSF1. These motifs have been implicated in RNA-binding and can be arginine dimethylated (Burd & Dreyfuss, 1994; Liu & Dreyfuss, 1995).

Functional studies

The above results demonstrate that regions of highest similarity among the SF1 homologs represent previously defined functional domains. We therefore expected not only a conservation of the role of individual domains, but also their exchangeability across species. To test this assumption, we first studied interactions of HsSF1, DmSF1, and CeSF1 with U2AF⁶⁵ homologs from humans and *Drosophila*. GST-tagged versions of

HsU2AF⁶⁵ and DmU2AF⁵⁰ were bound to glutathione agarose and incubated with in vitro-translated [³⁵S] methionine-labeled SF1 proteins. Bound proteins were eluted after washing, separated by SDS PAGE, and visualized by Coomassie-Blue staining (Fig. 5, top) and autoradiography (Fig. 5, bottom). Full-length (FL) DmSF1 and CeSF1 interact equally well with DmU2AF⁵⁰ and HsU2AF⁶⁵. Similarly, the binding of DmSF1 lacking the N-terminal RS domain (DmSF1ΔN) to DmU2AF⁵⁰ and HsU2AF⁶⁵ was comparable to that of HsSF1, which naturally does not contain an RS domain. These results indicate that structural features essential for protein-protein contacts are conserved in the SF1 as well as the U2AF⁶⁵ homologs. In addition, the RS domain of DmSF1 does not appear to be essential for binding to the U2AF⁶⁵ homologs, nor does this domain interfere with the interaction. Whether or not the RS domain of DmSF1 (or CeSF1) contributes to these interactions remains to be determined.

We have previously shown that the N-terminal half of HsSF1 encompassing sequences required for the interaction with U2AF⁶⁵ and RNA binding are sufficient to support prespliceosome assembly in vitro (Rain et al., 1998). Given the above results, we tested whether DmSF1 could substitute for HsSF1 in prespliceosome assembly. To date we have not succeeded in expressing full-length DmSF1 and CeSF1 proteins in bacteria. Therefore, DmSF1ΔN was used to complement a HeLa cell reconstitution assay (Krämer & Utans, 1991; Rain et al., 1998). In a control reaction with HeLa nuclear extract, a pre-mRNA derived from the adenovirus ma-

FIGURE 5. Interactions between SF1 and U2AF⁶⁵ homologs. GST (G) and GST-tagged DmU2AF⁵⁰ (Dm) or HsU2AF⁶⁵ (Hs) were bound to glutathione agarose and incubated with [³⁵S] methionine-labeled DmSF1-FL, DmSF1ΔN, HsSF1, and CeSF1-FL, as indicated above the panels. Bound proteins were eluted with SDS sample solution and separated by 10% SDS PAGE. Gels shown on top were stained with Coomassie-Blue and show the GST and GST-tagged proteins that were bound to glutathione agarose. The gel on the top left is representative for the experiment performed with DmSF1 and HsSF1 proteins; the gel on the top right corresponds to the experiment performed with CeSF1. Autoradiographs of the proteins bound to GST-U2AF⁶⁵ are presented on the bottom. Lanes marked IP represent 25% of the input of in vitro-translated proteins except for CeSF1-FL, where 50% of the input was loaded onto the gel.

for late transcription unit was assembled into presplicing complex A and splicing complexes B and C (Fig. 6). In the presence of partially purified fractions containing HeLa cell SF3a, SF3b, U2AF, SR proteins, and U1 and U2 snRNPs, only background levels of complex A were apparent. Recombinant DmSF1 Δ N (20 or 50 ng) efficiently reconstituted the assembly of complex A with an activity similar to that observed with partially purified HeLa SF1. We conclude that DmSF1 lacking the RS domain can functionally replace HsSF1 in prespliceosome assembly and all domains required for this activity are conserved between flies and mammals.

CeSF1 is essential for *C. elegans* development

The *S. cerevisiae* gene encoding SF1 is essential for viability (Abovich & Rosbash, 1997; Fromont-Racine

et al., 1997). Although the high conservation of SF1 and its role in branch site recognition suggest that the gene is also essential in metazoans, this has not been formally demonstrated. Mutations in the chromosomal loci of the *dSF1* or *Ce-SF1* genes have not been reported. We therefore used double-stranded RNA-mediated interference (RNAi), a procedure that generates a phenotype similar or identical to that of a null mutation in the corresponding gene in *C. elegans* (Fire et al., 1998; Hunter, 1999; Sharp, 1999). To interfere with a gene's function, in vitro-synthesized RNA is injected into the germline of wild-type hermaphrodites and the progeny of injected worms is analyzed for a mutant phenotype.

A 900-bp portion of the CeSF1 cDNA that encodes sequences C-terminal of the KH domain was used as a template for the in vitro synthesis of sense and antisense RNA. These sequences were chosen to avoid cross interference with other KH domain-containing *C. elegans* proteins, for example GLD-1 (Francis et al., 1995), a member of the STAR family of proteins that shows considerable homology on the DNA level to CeSF1 in the region of the KH domain. After annealing, the double-stranded CeSF1 cRNA was injected into the gonads of wild-type (N2) hermaphrodites. Twenty-eight worms were injected and 24 laid eggs. Of 2,048 embryos of the F1 progeny, 2,046 were arrested during gastrulation at a stage of approximately 200 blastomeres. Two eggs hatched, but larval development was very slow and the larvae eventually died. It is unlikely that injection of Ce-SF1 dsRNA into worms affected the function of *uaf-2*, an essential gene that encodes *C. elegans* U2AF³⁵ (Zorio & Blumenthal, 1999), because RNAi acts after RNA processing by an as yet unknown mechanism (see Hunter, 1999; Sharp, 1999). In addition, injection of antisense or dsRNA of *cyp-13*, which is located between the *uaf-2* and *Ce-SF1* genes, did not interfere with viability, indicating that the function of the neighboring genes was not affected by RNAi (Zorio & Blumenthal, 1999).

The early embryonic arrest of the progeny from worms injected with CeSF1 cRNA indicates that the protein is essential for key developmental events in *C. elegans*. This result is consistent with a role of CeSF1 in general pre-mRNA splicing and demonstrates for the first time that SF1 is an essential splicing factor in metazoans.

DISCUSSION

Conserved functional domains in SF1

We have isolated cDNAs encoding SF1 homologs of *Drosophila* and *C. elegans*. As expected from the structural conservation between mammalian and yeast SF1 proteins (Arning et al., 1996), domains shown to be essential for the function of SF1 are also conserved in flies and worms. The KH domain is essential for RNA binding and sequence-specific recognition of the branch

FIGURE 6. Prespliceosome assembly. Spliceosome formation was tested in the presence of HeLa cell nuclear extract (NE, lane 1) or partially purified splicing factors SF3a, SF3b, U2AF, and U1 and U2 snRNPs (lanes 2–5). These reactions were supplemented with buffer (lane 2), SF1 purified from HeLa cells (lane 3), or 20 and 50 ng of recombinant DmSF1 Δ N (lanes 4 and 5). Splicing complexes were separated in a nondenaturing 4% polyacrylamide gel. Presplicing complex A and splicing complexes B/C are indicated on the left.

site sequence (Berglund et al., 1997, 1998b; Rain et al., 1998). Consistent with this result, the KH domain represents the most highly conserved region in SF1 from different organisms. Zinc knuckles located C-terminal to the KH domain are also essential for RNA binding and increase SF1's affinity for RNA (Berglund et al., 1998b). Our results show that CeSF1, as SF1 in yeast, contains two zinc knuckles, whereas the first zinc knuckle is degenerate in DmSF1 and absent from HsSF1. In a model proposed by Berglund et al. (1998b), the KH domain of SF1 makes sequence-specific contacts with the bases of the branch site and two modules contact the RNA phosphate backbone. In humans, these two modules are the single zinc knuckle of HsSF1 and the N-terminal RS domain of U2AF⁶⁵ (Valcárcel et al., 1996), which was suggested to serve as a *trans*-acting affinity module for SF1 binding (Berglund et al., 1998b). In *S. cerevisiae*, two zinc knuckles may be required to increase the affinity of ScSF1 for the pre-mRNA, because Mud2p lacks the N-terminal RS domain present in other U2AF⁶⁵ homologs (Abovich et al., 1994). In addition, extended regions of pyrimidines are in general absent from *S. cerevisiae* introns, but the branch site sequence is nearly invariant (Rymond & Rosbash, 1992).

In extension of this model, we suggest that the presence of two zinc knuckles in SF1 of yeast and *C. elegans* and the loss of the first zinc knuckle during further evolution are related to the difference between species in the organization of the 3' splice site sequence. With the exception of Mud2p, all U2AF⁶⁵ homologs contain an N-terminal RS domain (Zamore et al., 1992; Kanaar et al., 1993; Potashkin et al., 1993; Zorio et al., 1997), suggesting that this region may serve to increase RNA-binding affinity in species other than humans as well. This might be particularly important in organisms that do not contain well-defined 3' splice site sequences. Branch site sequences in *S. pombe* are more variable than in *S. cerevisiae* and runs of pyrimidines between the branch site and the 3' splice site are present in some introns (Zhang & Marr, 1994; Romfo & Wise, 1997). In other *S. pombe* introns these sequences differ widely and can even be purine-rich. Despite this variability, SpU2AF⁵⁹ is required for the splicing of all *S. pombe* introns tested to date, independent of their pyrimidine-richness (Romfo et al., 1999). *C. elegans* introns are characterized by a highly conserved sequence at the 3' splice site (UUUUCAG), but do not contain pyrimidine-rich regions or discernable branch sites (Zhang & Blumenthal, 1996; Blumenthal & Steward, 1997). Efficient recognition of the 3' splice site in these organisms may necessitate the presence of three RNA affinity modules, that is, the two zinc knuckles of SF1 and the RS domain of the U2AF⁶⁵ homologs. Such a scenario may also contribute to tethering U2AF⁶⁵ proteins to the 3' splice site of introns with ill-defined or missing polypyrimidine tracts. *Drosophila* introns contain the sequence YUAAU (Y = C or T) as a consensus

branch site and regions between the branch site and the 3' splice site are enriched in pyrimidines (Mount et al., 1992). Thus, the situation in flies may be more relaxed and obviate the need for a consensus second zinc knuckle motif.

Sequences obtained in a two-hybrid screen with Mud2p defined amino acids 41–141 of ScSF1, located immediately N-terminal to the KH domain, as the minimal sequences required for the interaction between the two yeast proteins (Rain et al., 1998). Although sequences sufficient for the interaction between HsSF1 and U2AF⁶⁵ were located to a similar region, this interaction appeared to require sequences further N-terminal in HsSF1 (amino acids 8–137; Rain et al., 1998; Z. Rhani & A. Krämer, unpubl. results). The amino acid sequence alignment shown in Figure 4 and recent results indeed suggest that sequences in addition to those identified for the Mud2p interaction domain of ScSF1 are required for the interaction between metazoan SF1 and U2AF⁶⁵ homologs. Wang et al. (1999) demonstrated that mutation of Ser-20 to Ala in HsSF1 highly reduced its interaction with U2AF⁶⁵ and activity in prespliceosome assembly, indicating that Ser-20 of HsSF1 is located close to a site of contact with U2AF⁶⁵. Ser-20 and neighboring sequences are conserved in DmSF1 and CeSF1, but not in SF1 of *S. cerevisiae* or *S. pombe*, suggesting that these sequences may similarly be essential for the interaction with the U2AF⁶⁵ homologs in flies and worms. Binding experiments that address the exact sequences in SF1 homologs required for interaction with U2AF⁶⁵ proteins have not been performed. Such studies will be necessary to determine similarities and differences in interactions between the yeast and metazoan SF1 and U2AF⁶⁵ proteins.

Domain swap experiments have previously addressed the functional correlation between domains in *S. cerevisiae* and human SF1. For example, the KH domain of HsSF1 is at least partially functional in the context of ScSF1 in vivo, although this replacement resulted in a slow-growth phenotype of a strain disrupted in the *ySF1* gene (Rain et al., 1998). Replacement of the ScSF1 KH domain and the zinc knuckles with the corresponding sequences of HsSF1 did not result in complementation. This may be related to the observation that an exchange of the single HsSF1 zinc knuckle with the two zinc knuckles of ScSF1 resulted in increased RNA binding affinity in vitro (Berglund et al., 1997, 1998b). Because the *MUD2* gene is not essential for viability in yeast (Abovich et al., 1994), an in vivo function of the U2AF⁶⁵ interaction domain in the context of ScSF1 could not be tested. However, heterologous interactions in the two-hybrid system between ScSF1 and U2AF⁶⁵ or HsSF1 and Mud2p have not been observed (Rain et al., 1998). Our results indicate that metazoan SF1 proteins are not restricted to interactions with U2AF⁶⁵ proteins of the same species, but can also bind to

U2AF⁶⁵ homologs in other organisms. We therefore expect that sequences essential for contact are similar or identical in the metazoan proteins. The observation that DmSF1 can substitute for HsSF1 in prespliceosome assembly in the human system is consistent with the high structural conservation of those domains previously shown to be sufficient for function in splicing complex formation, that is, the U2AF⁶⁵ interaction domain, the KH domain, and the STAR-QUA2 domain (Rain et al., 1998).

The N-terminal RS domain of DmSF1 and CeSF1

DmSF1 and CeSF1 contain an extended N-terminal RS domain similar to the RS domain of SR or SR-related proteins (see Results). Re-inspection of the amino acid sequences of human and yeast SF1 revealed that the regions N-terminal of the U2AF⁶⁵ interaction domains in *S. pombe* and *S. cerevisiae*, but not in humans, are also particularly rich in charged amino acids and contain Ser residues. These sequences could represent remnants of an RS domain.

How could the RS domain of DmSF1 and CeSF1 contribute to their function? First, this domain could contact the branch site (or nearby sites in the intron) and stabilize the binding of SF1, similar to the RS domain of U2AF⁶⁵ (Valcárcel et al., 1996). This may be important in *C. elegans* introns where sequences that resemble branch sites could not be identified (Blumenthal & Spieth, 1996). Second, there are several possible roles for the RS domain in interactions with other splicing proteins. In *S. cerevisiae*, SF1 interacts with Mud2p and Prp40p, a component of the U1 snRNP, an interaction that is thought to bridge the 5' splice site and the branch site region (Abovich & Rosbash, 1997). A homolog of Prp40p has not been identified in mammals; however, HsSF1 may contact the U1 snRNP indirectly through interaction with a formin-binding protein (FBP21; Bedford et al., 1998). Whether homologs of Prp40p or FBP21 exist in flies or nematodes is not known. If SF1 was required for cross-intron bridging in these organisms, its RS domain could either directly interact with a U1 snRNP-associated protein, for example U1-70K, or indirectly through contacts with SR proteins. In another scenario, the RS domain of SF1 could bind to U2AF⁶⁵ in addition to the interaction reported previously (Abovich & Rosbash, 1997; Berglund et al., 1998a; Rain et al., 1998). This may help stabilize the binding of U2AF⁶⁵ to introns that do not contain extended polypyrimidine tracts, as is the case in *C. elegans* (Blumenthal & Spieth, 1996).

Studies to analyze the function of the RS domains of DmSF1 and CeSF1 are in progress. Furthermore, it needs to be determined whether this domain is essential for viability. This is not the case for the RS domain of U1-70K in *Drosophila* (S. Mount, pers. comm.). More-

over, Rudner et al. (1998) recently demonstrated that the presence of an RS domain in one of the U2AF subunits of *Drosophila* is sufficient for viability. The authors suggested that the RS domains of the U2AF heterodimer serve redundant functions in vivo and together perform the tasks of the individual domains previously assigned by in vitro experiments to the human U2AF subunits. Given the proximity of SF1 and U2AF once bound to the pre-mRNA, it will be interesting to test whether the RS domain of DmSF1 could compensate for the absence of the RS domain in one of the U2AF subunits.

Implications for the regulation of SF1 activity

The presence of an RS domain in DmSF1 and CeSF1 may also be relevant to the regulation of their activity. Ser residues in RS domains are often phosphorylated and kinases that target these domains have been identified (for references see Manley & Tacke, 1996; Xiao & Manley, 1998). Reversible phosphorylation regulates the activity of SR proteins during spliceosome assembly by modulating protein-protein and protein-RNA interactions. In addition, phosphorylation regulates the localization of SR proteins within the cell. In correlation with the effects of phosphorylation on the activity of SR proteins, we expect that a net change in the charge of the RS domain of DmSF1 and CeSF1 contributes to the regulation of their activity in splicing. Although we can only speculate about a function of the RS domain of SF1 in *Drosophila* and *C. elegans* at the moment, its presence in these species, but not in mammals or yeast, hints at a novel, additional role for SF1 in flies and worms. This role may not be required in other organisms or be carried out by different proteins.

Whereas phosphorylation of the RS domain of DmSF1 and CeSF1 would be expected to regulate their function during the course of the splicing reaction, DmSF1 (and possibly also CeSF1) is a likely target of PKG-1, which is activated by cGMP in response to extracellular stimuli (Wang & Robinson, 1997). We have recently shown that the interaction of HsSF1 with U2AF⁶⁵ and prespliceosome assembly are negatively regulated through PKG phosphorylation (Wang et al., 1999). The single phosphorylation site of HsSF1 is conserved in DmSF1 and additional consensus PKG/PKA phosphorylation sites are present in the N- and C-terminal regions. In vitro, DmSF1 and CeSF1 are phosphorylated by PKG (data not shown), but the exact phosphorylation site(s) have not been determined. Although an effect on the function of SF1 in *Drosophila* and *C. elegans* remains to be tested, we expect that the activity of SF1 is also regulated by PKG in flies and worms.

The C-terminal regions of SF1 in different species are Pro- or Pro-Gly-rich. The Pro-rich region of HsSF1 interacts with the WW domain of two formin-binding

proteins (FBP11 and FBP21) and the SH3 domain of Abl (Bedford et al., 1997, 1998). These proteins bind their partners through different Pro-rich motifs. FBP21 is present in prespliceosomal complex A and associated with the U2 snRNP (Bedford et al., 1998). In addition to SF1, FBP21 binds to the core snRNP proteins B and B' and the U1 snRNP-specific protein U1C, which suggested a role for FBP21 in splice-site bridging. The C-terminal Pro-rich region of HsSF1 is dispensable for prespliceosome assembly in vitro (Rain et al., 1998). Therefore, the relevance of these interactions for spliceosome assembly or their conservation in other organisms awaits the development of suitable assay systems. The variability of the C-terminal domain during evolution and the presence of different HsSF1 isoforms with variable C termini combined with the result that this region is dispensable for prespliceosome assembly in vitro (Rain et al., 1998) may imply regulatory functions for this domain. This regulation could at least in part be species specific.

A novel SF1 isoform

One aim of this study was to analyze whether alternatively spliced SF1 mRNAs similar to those found in mammals are also present in other species. Northern blot analysis of polyA⁺ RNA from different developmental stages of *Drosophila* suggested the presence of several low-abundance transcripts in addition to a major SF1 mRNA. However, sequencing and restriction analysis of a number of DmSF1 cDNAs as well as PCR approaches did not reveal the presence of variant SF1 mRNAs. Prompted by the finding of an EST that lacked 76 bp in the 5' part of the DmSF1 ORF, PCR analysis revealed the presence of an additional mRNA that is expressed at all developmental stages of *Drosophila*, but at a lower level than the full-size mRNA (data not shown). A corresponding product was observed in HeLa cells and several mouse tissues (F. Mulhauser & A. Krämer, unpubl. results). Comparison of the 76-bp sequence to the recently available partial DmSF1 genomic sequence indicated that the novel mRNA is generated by skipping of exon 4, which corresponds to exon 3 in the HsSF1 gene and exon 4 in *Ce-SF1* (Krämer et al., 1998; Fig. 3a). Exon 4 of DmSF1 is flanked by introns of about 600 and 1,000 nt. Similarly, the introns flanking exon 3 of HsSF1 are the longest introns in the pre-mRNA (~3,000 nt each; Krämer et al., 1998). In addition, the 5' splice site of the alternatively spliced exon deviates from the consensus sequence in both *Drosophila* and humans (GUUUGA and GUUGGG, respectively, as compared to the consensus GURAGU; R = A or G). Thus, both the length of the surrounding introns and the suboptimal 5' splice site may contribute to exon skipping. The introns bordering exon 4 in *C. elegans* are 213 and 445 nt long and the 5' splice site (GUAAGU) conforms to the consensus.

Whether this exon can also be skipped in worms remains to be tested; however, if the protein products fulfilled an essential function, we would expect that the alternative splicing event also occurred in *C. elegans*.

Proteins translated from the novel mRNAs would be contiguous with full-length SF1 up to the N-terminal part of the U2AF⁶⁵ interaction domain, but would contain only short C-termini due to termination of translation in a different reading frame. By Western blotting of HeLa or *Drosophila* extracts, we have thus far not detected proteins of the expected size. This may be related to low abundance, expression only in certain cell types or tissues, or failure of the antibodies to detect the truncated forms. The truncated proteins contain Ser-20, which is essential for the SF1-U2AF⁶⁵ interaction in humans (Wang et al., 1999) and would therefore compete with full-length SF1 for binding to U2AF⁶⁵. Similarly, the presence of the RS domain in the truncated form of DmSF1 could interfere with putative interactions with other spliceosomal components. Interference with the interaction between SF1 and U2AF⁶⁵ (or other proteins) may provide a means to regulate spliceosome assembly during development or in a tissue-specific manner.

Requirement for viability and genomic organization of *Ce-SF1*

The isolation of SF1 cDNAs from *Drosophila* and *C. elegans* was initiated with the goal in mind to establish a system that allows for testing of SF1 function in vivo. Previous studies demonstrated that SF1 is essential for viability in *S. cerevisiae* (Abovich & Rosbash, 1997; Fromont-Racine et al., 1997; Rain et al., 1998). In addition, the important and conserved role of SF1 in branch site recognition was a strong argument that the protein also provided an essential function in metazoans. We have used RNAi in *C. elegans* (Fire et al., 1998; Hunter, 1999; Sharp, 1999) to demonstrate that *Ce-SF1* is essential. Nearly 100% of the F1 progeny of injected hermaphrodite worms was arrested during gastrulation, demonstrating that CeSF1 provides an important function for early developmental decisions in *C. elegans*, consistent with the role of SF1 as an essential splicing factor in metazoans.

The *Ce-SF1* gene is located in a single polycistronic unit downstream of the *uaf-2* and *cyp-13* genes (Zorio & Blumenthal, 1999; this study). The three genes are transcribed from the same strand and are separated by short intercistronic regions. The U2AF³⁵ mRNA is *trans*-spliced to SL1 and represents the 5'-most transcription unit in the operon, whereas CYP-13 and CeSF1 mRNAs are *trans*-spliced to SL2, which is in general added to downstream mRNAs (Zorio & Blumenthal, 1999; this study). Whether other genes are located downstream of *Ce-SF1* is presently unknown. About 25% of the genes in *C. elegans* are arranged in operons and sev-

eral cases have been reported where unrelated proteins that are clustered in a single operon function in the same pathway (see Blumenthal, 1998). The operon containing *Ce-SF1* and *uaf-2* adds a novel example to this list. Both proteins are required for early events of spliceosome assembly (see Introduction). The presence of the *uaf-2* and *Ce-SF1* genes in a single operon implies that their expression is coordinated at the transcriptional level. Thus, pre-mRNA splicing in *C. elegans* may be tightly regulated by subjecting two essential splicing factors to the same transcriptional controls. This does not appear to be the case in *Drosophila* or humans, where the genes encoding SF1 and U2AF³⁵ are located on different chromosomes (Toda et al., 1994; Lalioti et al., 1996; Rudner et al., 1996; Krämer et al., 1998; this study).

The presence of the *cyp-13* gene in one polycistronic transcription unit together with *uaf-2* and *Ce-SF1* raises the question whether CYP-13 also plays a role in splicing. *cyp-13* encodes a cyclophilin with an N-terminal consensus RRM (Zorio & Blumenthal, 1999). Cyclophilins are characterized by a highly conserved domain of ~110 amino acids bordered by divergent N- and C-termini and function as peptidylpropyl-isomerases (for references, see Horowitz et al., 1997; Mortillaro & Berezney, 1998; Teigelkamp et al., 1998). They participate in a wide variety of cellular processes and are involved in protein folding and/or function as chaperones by facilitating protein targeting and macromolecular assembly. A canonical cyclophilin has recently been identified in the U4/U6 and U4/U6.U5 snRNPs and colocalizes with other splicing components in nuclear speckles (Horowitz et al., 1997; Teigelkamp et al., 1998). In addition, an RS domain-containing cyclophilin isolated independently in yeast two-hybrid screens with the Clk/Sty kinase and the C-terminal heptapeptide repeat of RNA polymerase II, and as a nuclear matrix-associated protein may function in some aspect of RNA processing (Nestel et al., 1996; Bourquin et al., 1997; Mortillaro & Berezney, 1998). Given that CYP-13 contains a consensus RRM, it may interact directly with the pre-mRNA and facilitate the binding of SF1 or U2AF⁶⁵ to the 3' splice site. Alternatively it may aid in protein-protein interactions that are essential for the formation of early splicing complexes. The fact that the *cyp-13* gene is not essential for viability in *C. elegans* does not exclude a function in splicing (Zorio & Blumenthal, 1999). As pointed out by these authors, the function of CYP-13 may be redundant in *C. elegans* or it may be required for efficient or accurate splicing without noticeable effects on viability.

In conclusion, the cloning of SF1 homologs from *Drosophila* and *C. elegans* has provided novel information regarding the conservation and divergence among species of structural and functional domains of an essential splicing factor. Future analysis of the role of the RS domain of SF1 in flies and worms may reveal perhaps

subtle differences in splicing between humans, flies, and worms. Moreover, the demonstration that SF1 is not only essential for viability in *S. cerevisiae*, but also in *C. elegans* (and most likely in other metazoans) provides a means to test the function of SF1 *in vivo*.

MATERIALS AND METHODS

Cloning procedures

For the cloning of a cDNA-encoding DmSF1, degenerate oligonucleotides were designed to cDNA sequences that are conserved between human and mouse SF1 and a *C. briggsae* EST (accession number R02974) with significant homology to SF1 in the region of the KH domain. Random-primed cDNA prepared from polyA⁺ RNA of 0–3 h *Drosophila* embryos was amplified with the upstream and downstream primers GMITTYAARCCICCGMGA and GCTCATIYCKC CIGGCA (Y = C or T; I = inosine; M = A or C; K = G or T), respectively. A fragment of an expected length of ~250 bp was gel purified and cloned into pMOSBlue (Amersham). The insert was sequenced with the T7 sequencing kit (Pharmacia LKB Biotech) and revealed high similarity to the corresponding region in HsSF1. A primer specific to sequences of this fragment (S1: AACCGCCAGTCACTCGTGTC) and a λ gt10 reverse primer (TTATGAGTATTTCTTCCAGGG) were used to isolate larger fragments from DNA of a *Drosophila* third instar cDNA library in λ gt10 (kindly provided by V. Pirotta) by touchdown PCR. After an initial incubation for 5 min at 95 °C, two cycles each were performed at annealing temperatures decreasing by 1 °C (from 59 to 46 °C) as follows: denaturation at 92 °C for 30 s, annealing for 1 min, extension at 72 °C for 6 min. These cycles were followed by 15 cycles at an annealing temperature of 45 °C, followed by 19 min extension at 72 °C. Several fragments were obtained and the largest one (457 bp, Dm-GT2) was cloned and sequenced.

Dm-GT2 DNA was random prime labeled (Boehringer Mannheim) and used to screen 6×10^5 phages of a *Drosophila* 0–18 h embryo cDNA library in λ gt10 (Clontech) under high stringency conditions (Sambrook et al., 1989). After three rounds of screening, 34 double-positive phage were obtained. Ten inserts were isolated, cloned into pBluescript KS (pBS; Stratagene) and sequenced. The longest inserts were sequenced on both strands from exonuclease III fragments (Erase-A-Base system, Promega).

A PCR fragment encoding a partial (1354 bp) *C. briggsae* SF1 cDNA (see above; kindly provided by M. Marra) was cloned into pMOSBlue and sequenced on both strands. The encoded amino acid sequence showed high homology to Hs and ScSF1 in a region encompassing the KH domain and neighboring sequences. A cDNA library (6×10^5 pfu) from mixed stages of *C. elegans* in λ ZAP (kindly provided by P. Nef) was screened with a *Bam*HI/*Cl*al restriction fragment representing the 5' part of the *C. briggsae* cDNA. The inserts of two double-positive phage were excised (Barstead & Waterston, 1989), cloned into pBS and sequenced.

Reverse transcription-PCR (RT-PCR)

Total RNA was prepared from *C. elegans* mixed stages grown on plates (Conrad et al., 1991). Contaminating DNA was re-

moved by treatment with 1 U of RQ DNase (Boehringer Mannheim) in 40 mM Tris-HCl (pH 7.5), 2 mM spermidine, 7 mM MgCl₂, 10 mM NaCl, 40 U RNasin, and 10 mM DTT for 45 min at 37 °C. After phenol-chloroform extraction and ethanol precipitation, the RNA was heated to 65 °C, cooled, and used for RT-PCR. The RNA (2 µg) was preincubated for 30 min at 37 °C with 200 ng of downstream primer LP1 (GCC TGCTGTTGCTGCTCTTCTTC; see Fig. 3A) in 50 mM Tris-HCl (pH 8.3), 75 mM KCl and 3 mM MgCl₂. After addition of M-MLV reverse transcriptase (Gibco-BRL) incubation was continued for 60 min at 37 °C and 15 min at 42 °C. Aliquots of the RT reaction (1 µL) were used for PCR amplification with the Expand High Fidelity kit (Boehringer). The downstream primer LP2 had the sequence CCTTTCGGTGGTCCAAATAAATCCC. Upstream primers were either complementary to the first 19 bp of the spliced leader sequences SL1 (GGTTTAATTAC CCAAGTTT) or SL2 (GGTTTAAACCCAGTTACTC) (Spieth et al., 1993) or to regions encoding putative upstream exons in the *C. elegans* cosmid CEY116A8 (UP1: ATGTCGAAAAC TGGTGGAAACACG; UP2: GAAGATTTGGCTGCGGCTGT TG). PCR conditions were as follows: 5 min at 95 °C, 40 cycles of 30 s at 92 °C, 30 s at 54 °C (LP2-SL1 and LP2-SL2), 64 °C (LP2-UP1) or 59 °C (LP2-UP2), and 3 min at 68 °C, followed by an additional incubation at 68 °C for 10 min. Reaction products of the expected sizes were gel purified, cloned into pMOS-Blue and sequenced.

Northern blotting

PolyA⁺ RNA was isolated from *Drosophila* embryos, larvae, pupae and adults as described (van Hille et al., 1993). RNA (5 µg) was separated in a 0.8% agarose/formaldehyde gel and transferred to a Zetabind membrane (Cuno). The membrane was prehybridized for 4 h at 42 °C as described above. ³²P random prime-labeled Dm-30 DNA was added and hybridization was performed for 18 h. The membrane was washed three times for 20 min in 2 × SSC/1% SDS at room temperature and twice for 30 min in 0.2 × SSC/1% SDS at 65 °C.

Western blotting

Proteins were separated by SDS PAGE and transferred to nitrocellulose membranes (Kyhse-Anderson, 1984). The membranes were incubated with rabbit antibodies raised against His₆-tagged DmSF1 (amino acids 237–773) at a dilution of 1:5,000, horseradish peroxidase-coupled swine anti-rabbit antibodies (1:2,000; Dako) and enhanced chemiluminescence reagents (ECL, Amersham Life Sciences) according to the supplier's instructions.

Tissue culture and extract preparation

Schneider cells were propagated at 27 °C in Schneider *Drosophila* medium supplemented with 10% heat-inactivated calf serum (Schneider et al., 1993). Small scale preparations of cytoplasmic and nuclear extracts were performed according to Krämer and Keller (1990). Extracts were dialyzed against D buffer (Dignam et al., 1983). *Drosophila* embryo extracts (1–24 h) were kindly provided by V. Pirota.

In vitro translation and bacterial expression of recombinant proteins

Constructs for in vitro translation and bacterial expression were generated by ligation of PCR products and restriction fragments derived from the isolated cDNAs into pGem3Z (Promega) and pTRC-His (Invitrogen), respectively. Correct construction was verified by sequencing and restriction analysis. Coupled in vitro transcription and translation was performed in 100-µL reactions in the TNT system (Promega) for 90 min at 30 °C in the presence of 0.01 µg/µL leupeptin and 0.25 mM phenylmethylsulfonyl fluoride. Recombinant His₆-tagged proteins were produced in the *Escherichia coli* strain TOP10 (Invitrogen) as described (Rain et al., 1998). Plasmids encoding HsU2AF⁶⁵ (Zamore et al., 1992) and DmU2AF⁵⁰ (Kanaar et al., 1993) were kind gifts of J. Valcárcel and D. Rio. The DmU2AF⁵⁰ insert was recloned into pGex2T by digestion of the original plasmid with *Hind*III. The *Hind*III site was filled in with Klenow polymerase, followed by digestion with *Bam*HI and ligation into the *Sma*I and *Bam*HI sites of pGex2T. The GST-tagged proteins were expressed in *E. coli* XL1-Blue cells for 3 h after induction with 0.1 mM IPTG. GST fusion proteins were purified as described (Smith & Corcoran, 1990).

Analysis of protein–protein interactions and spliceosome assembly

Interactions of in vitro-translated [³⁵S] methionine-labeled proteins with GST-tagged HsU2AF⁶⁵ and DmU2AF⁵⁰ were tested as described previously (Rain et al., 1998). Splicing complexes were analyzed in a HeLa cell nuclear extract or in a reconstituted system containing partially purified SF3a, SF3b, U2AF, SR proteins, U1 and U2 snRNPs, and SF1 purified from HeLa cells or recombinant DmSF1ΔN (Krämer & Utans, 1991; Rain et al., 1998).

Double-stranded RNA interference

A 900-bp *Nde*I-*Avr*II fragment encoding sequences C-terminal of the KH domain of CeSF1 was cloned into the *Sma*I site of pBS-KS. The plasmid was linearized with *Bam*HI or *Cla*I and sense or antisense RNAs were transcribed in separate reactions with T3 or T7 RNA polymerase (Promega), respectively (Fire et al., 1998). The RNA concentration was determined at OD₂₆₀. Equal amounts of sense and antisense RNA in H₂O were mixed with three volumes of injection buffer (20 mM KPO₄, pH 7.3, 3 mM K-citrate, pH 7.5, and 2% PEG 6000). The sample was incubated at 68 °C for 10 min followed by a 30-min incubation at 37 °C and stored at –70 °C. Successful annealing was verified by agarose gel electrophoresis.

Double-stranded cRNA was injected at a concentration of 1.9 µg/µL into both gonadal arms of young N2 adult hermaphrodites. The injected animals were passed onto fresh plates every day and the F1 progeny was scored during a time window of 22–24 h to 2–3 days after injection, that is, when the injected worms stopped laying eggs. During this time window on average about 150 eggs were laid by every single worm.

Computer analysis

Database searches and sequence alignments were performed with the FASTA, TFASTA, BESTFIT, and CLUSTALW pro-

grams of the GCG package of the University of Wisconsin (Devereux et al., 1984) or the BLAST program of the server of the National Center for Biotechnology Information.

ACKNOWLEDGMENTS

We are grateful to Rainer Maier (GeneData AG, Basel) for the initial amplification and isolation of a DmSF1 PCR fragment and Graeme Bilbe (Novartis, Basel) for providing RNAs from different developmental stages of *Drosophila* and helpful advice. We are thankful to M. Marra (Washington University School of Medicine, St. Louis) for providing the *C. briggsae* PCR fragment, Vincent Pirota (University of Geneva) for *Drosophila* embryo extracts, Patric Nef (Hoffmann LaRoche, Basel) for the *C. elegans* cDNA library, Don Rio (University of California, Berkeley) and Juan Valcárcel (European Molecular Biology Laboratory, Heidelberg) for providing plasmids. We would also like to thank François Karch and collaborators (University of Geneva) for help with in situ hybridization. This work was supported by grants from the Swiss National Science Foundation and the Canton of Geneva to A.K.

Received August 19, 1999; returned for revision
September 9, 1999; revised manuscript received
September 27, 1999

REFERENCES

- Abovich N, Liao XC, Rosbash M. 1994. The yeast MUD2 protein: An interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes & Dev* 8:843–854.
- Abovich N, Rosbash M. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89:403–412.
- Arning S, Grüter P, Bilbe G, Krämer A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* 2:794–810.
- Barstead RJ, Waterston RH. 1989. The basal component of the nematode dense-body is vinculin. *J Biol Chem* 264:10177–10185.
- Bedford M, Chan D, Leder P. 1997. FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. *EMBO J* 16:2376–2383.
- Bedford MT, Reed R, Leder P. 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.
- Berglund JA, Abovich N, Rosbash M. 1998a. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes & Dev* 12:858–867.
- Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89:781–787.
- Berglund JA, Fleming JL, Rosbash M. 1998b. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA* 4:998–1006.
- Blumenthal T. 1998. Gene clusters and polycistronic transcription in eukaryotes. *Bioessays* 20:480–487.
- Blumenthal T, Spieth J. 1996. Gene structure and organization in *Caenorhabditis elegans*. *Curr Opin Genet Dev* 6:692–698.
- Blumenthal T, Steward K. 1997. RNA processing and gene structure. In: Riddle D, Blumenthal T, Meyer B, Priess J, eds. *C. elegans II*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 117–145.
- Boggs RT, Gregor P, Idriss S, Belote JM, McKeown M. 1987. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the transformer gene. *Cell* 50:739–747.
- Bourquin JP, Stagliar I, Meier P, Moosmann P, Silke J, Baechi T, Georgiev O, Schaffner W. 1997. A serine/arginine-rich nuclear matrix cyclophilin interacts with the C-terminal domain of RNA polymerase II. *Nucleic Acids Res* 25:2055–2061.
- Burd CG, Dreyfuss G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265:615–621.
- Burge C, Tuschl T, Sharp P. 1999. Splicing of precursors to mRNA by the spliceosome. In: Gesteland R, Cech T, Atkins J, eds. *The RNA world*, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 525–560.
- Caslini C, Spinelli O, Cazzaniga G, Golay J, De Gioia L, Pedretti A, Brevario F, Amaru R, Barbui R, Biondi A, Introna M, Rambaldi A. 1997. Identification of two novel isoforms of the ZNF162 gene: A growing family of signal transduction and activator of RNA proteins. *Genomics* 42:268–277.
- Company M, Arenas J, Abelson J. 1991. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* 349:487–493.
- Conrad R, Thomas J, Spieth J, Blumenthal T. 1991. Insertion of part of an intron into the 5' untranslated region of a *Caenorhabditis elegans* gene converts it into a trans-spliced gene. *Mol Cell Biol* 11:1921–1926.
- Darlix J-L, Lapadat-Tapolsky M, de Rocquigny H, Roques B. 1995. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol* 254:523–537.
- Devereux J, Haeberli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387–395.
- Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11:1475–1489.
- Domon C, Lorkovic ZJ, Valcárcel J, Filipowicz W. 1998. Multiple forms of the U2 small nuclear ribonucleoprotein auxiliary factor U2AF subunits expressed in higher plants. *J Biol Chem* 273:34603–34610.
- Etzerodt M, Vignali R, Ciliberto G, Scherly D, Mattaj JW, Philipson L. 1988. Structure and expression of a *Xenopus* gene encoding an snRNP protein (U1 70K). *EMBO J* 7:4311–4321.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–811.
- Francis R, Barton MK, Kimble J, Schedl T. 1995. *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. *Genetics* 139:579–606.
- Fromont-Racine M, Rain J-C, Legrain P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet* 16:277–282.
- Frugier F, Kondorosi A, Crespi M. 1998. Identification of novel putative regulatory genes induced during alfalfa nodule development with a cold-plaque screening procedure. *Mol Plant Microbe Interact* 11:358–366.
- Fu X-D. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA* 1:663–680.
- Fu X-D, Maniatis T. 1992. Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science* 256:535–538.
- Horowitz DS, Kobayashi R, Krainer AR. 1997. A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA* 3:1374–1387.
- Hunter C. 1999. Genetics: A touch of elegans with RNAi. *Curr Biol* 9:R440–R442.
- Kanaar R, Roche SE, Beall EL, Green MR, Rio DC. 1993. The conserved pre-mRNA splicing factor U2AF from *Drosophila*: Requirement for viability. *Science* 262:569–573.
- Kohtz JD, Jamison SF, Will CL, Zuo P, Lüthmann R, Garcia-Blanco MA, Manley JL. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368:119–124.
- Krämer A. 1996. The structure and function of proteins involved in nuclear pre-mRNA splicing. *Annu Rev Biochem* 65:367–409.
- Krämer A, Keller W. 1990. Preparation and fractionation of mammalian extracts active in pre-mRNA splicing. *Methods Enzymol* 181:3–19.
- Krämer A, Quentin M, Mulhauser F. 1998. Diverse modes of alternative splicing of human splicing factor SF1 deduced from the exon-intron structure of the gene. *Gene* 211:29–37.
- Krämer A, Utans U. 1991. Three protein factors (SF1, SF3, and U2AF)

- function in pre-splicing complex formation in addition to snRNPs. *EMBO J* 10:1503–1509.
- Kyhse-Anderson J. 1984. Electrophoretic blotting of multiple gels. *J Biochem Biophys Methods* 10:203–209.
- Lalioti MD, Gos A, Green MR, Rossier C, Morris MA, Antonarakis SE. 1996. The gene for human U2 snRNP auxiliary factor small 35-kDa subunit (U2AF1) maps to the progressive myoclonus epilepsy (EPM1) critical region on chromosome 21q22.3. *Genomics* 33:298–300.
- Lazar G, Schaal T, Maniatis T, Goodman H. 1995. Identification of a plant serine-arginine-rich protein similar to the mammalian splicing factor SF2/ASF. *Proc Natl Acad Sci USA* 92:7672–7676.
- Liu Q, Dreyfuss G. 1995. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol Cell Biol* 15:2800–2808.
- Mancebo R, Lo PC, Mount SM. 1990. Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. *Mol Cell Biol* 10:2492–2502.
- Manley JL, Tacke R. 1996. SR proteins and splicing control. *Genes & Dev* 10:1569–1579.
- Mortillaro MJ, Berezney R. 1998. Matrin CYP, an SR-rich cyclophilin that associates with the nuclear matrix and splicing factors. *J Biol Chem* 273:8183–8192.
- Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C. 1992. Splicing signals in *Drosophila*: Intron size, information content, and consensus sequences. *Nucleic Acids Res* 20:4255–4262.
- Musco G, Stier G, Joseph C, Castiglione Morelli M, Nilges M, Gibson T, Pastore A. 1996. Three-dimensional structure and stability of the KH domain: Molecular insights into the fragile X syndrome. *Cell* 85:237–245.
- Nestel FP, Colwill K, Harper S, Pawson T, Anderson SK. 1996. RS cyclophilins: Identification of an NK-TR1-related cyclophilin. *Gene* 180:151–155.
- Nilsen T. 1998. RNA–RNA interactions in nuclear pre-mRNA splicing. In: Simons R, Grunberg-Manago M, eds. *RNA structure and function*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 279–307.
- Ohno M, Shimura Y. 1996. A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. *Genes & Dev* 10:997–1007.
- Ono Y, Ohno M, Shimura Y. 1994. Identification of a putative RNA helicase (HRH1), a human homolog of yeast Prp22. *Mol Cell Biol* 14:7611–7620.
- Page AP, Landry D, Wilson GG, Carlow CK. 1995. Molecular characterization of a cyclosporin A-insensitive cyclophilin from the parasitic nematode *Brugia malayi*. *Biochemistry* 34:11545–11550.
- Potashkin J, Naik K, Wentz-Hunter K. 1993. U2AF homolog required for splicing in vivo. *Science* 262:573–575.
- Query CC, Bentley RC, Keene JD. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* 57:89–101.
- Rain J-C, Rafi Z, Rhani Z, Legrain P, Krämer A. 1998. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* 4:551–565.
- Reddy AS, Czernik AJ, An G, Poovaiah BW. 1992. Cloning of the cDNA for U1 small nuclear ribonucleoprotein particle 70K protein from *Arabidopsis thaliana*. *Biochim Biophys Acta* 1171:88–92.
- Romfo CM, Lakhe-Reddy S, Wise JA. 1999. Molecular genetic analysis of U2AF⁵⁹ in *Schizosaccharomyces pombe*: Differential sensitivity of introns to mutational inactivation. *RNA* 5:49–65.
- Romfo CM, Wise JA. 1997. Both the polypyrimidine tract and the 3' splice site function prior to the first step of splicing in fission yeast. *Nucleic Acids Res* 25:4658–4665.
- Roscigno RF, García-Blanco MA. 1995. SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* 1:692–706.
- Rudner DZ, Breger KS, Rio DC. 1998. Molecular genetic analysis of the heterodimeric splicing factor U2AF: The RS domain on either the large or small *Drosophila* subunit is dispensable in vivo. *Genes & Dev* 12:1010–1021.
- Rudner DZ, Kanaar R, Breger KS, Rio DC. 1996. Mutations in the small subunit of the *Drosophila* U2AF splicing factor cause lethality and developmental defects. *Proc Natl Acad Sci USA* 93:10333–10337.
- Rutz B, Séraphin B. 1999. Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* 5:819–831.
- Rymond BC, Rosbash M. 1992. Yeast pre-mRNA splicing. In: Jones EW, Pringle JR, Broach JR, eds. *The molecular and cellular biology of the yeast Saccharomyces*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 143–192.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Schneider A, McNally KP, Agabian N. 1993. Splicing and 3'-processing of the tyrosine transfer RNA of *Trypanosoma brucei*. *J Biol Chem* 268:21868–21874.
- Schwer B, Gross CH. 1998. Prp22, a DEXH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J* 17:2086–2094.
- Séraphin B, Rosbash M. 1989. Identification of functional U1 snRNA–pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* 59:349–358.
- Sharp P. 1999. RNAi and double-stranded RNA. *Genes & Dev* 13:139–141.
- Smith DB, Corcoran LM. 1990. Expression and purification of glutathione-S-transferase fusion proteins. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JD, Smith JD, Struhl K, eds. *Current Protocols in molecular biology*. New York: John Wiley & Sons, Inc. pp 16.17.11–16.17.18.
- Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T. 1993. Operons in *C. elegans*: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell* 73:521–532.
- Staley JP, Guthrie C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* 92:315–326.
- Sukegawa J, Blobel G. 1995. A putative mammalian RNA helicase with an arginine–serine-rich domain colocalizes with a splicing factor. *J Biol Chem* 270:15702–15706.
- Tegge W, Frank R, Hofmann F, Dostmann WR. 1995. Determination of cyclic nucleotide-dependent protein kinase substrate specificity by the use of peptide libraries on cellulose paper. *Biochemistry* 34:10569–10577.
- Teigelkamp S, Achsel T, Mundt C, Gotthel SF, Cronshagen U, Lane WS, Marahiel M, Lührmann R. 1998. The 20kD protein of human (U4/U6.U5) tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA* 4:127–141.
- Toda T, Iida A, Miwa T, Nakamura Y, Imai T. 1994. Isolation and characterization of a novel gene encoding nuclear protein at a locus (*D11S636*) tightly linked to multiple endocrine neoplasia type 1 (MEN1). *Hum Mol Genet* 3:465–470.
- Tronchère H, Wang J, Fu X-D. 1997. A protein related to splicing factor U2AF³⁵ that interacts with U2AF⁶⁵ and SR proteins in splicing of pre-mRNA. *Nature* 388:397–400.
- Valcárcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF⁶⁵ RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* 273:1706–1709.
- van Hille B, Richener H, Evans DB, Green JR, Bilbe G. 1993. Identification of two subunit A isoforms of the vacuolar H⁺-ATPase in human osteoclastoma. *J Biol Chem* 268:7075–7080.
- Vellard M, Sureau A, Soret J, Martinier C, Perbal B. 1992. A potential splicing factor is encoded by the opposite strand of the trans-spliced *c-myc* exon. *Proc Natl Acad Sci USA* 89:2511–2525.
- Vernet C, Artzt K. 1997. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet* 13:479–484.
- Wagner JD, Jankowsky E, Company M, Pyle AM, Abelson JN. 1998. The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J* 17:2926–2937.
- Wang X, Bruderer S, Rafi Z, Xue J, Milburn P, Krämer A, Robinson P. 1999. Phosphorylation of splicing factor SF1 on Ser-20 by cGMP-dependent protein kinase regulates spliceosome assembly. *EMBO J* 18:4549–4559.
- Wang X, Robinson PJ. 1997. Cyclic GMP-dependent protein kinase and cellular signaling in the nervous system. *J Neurochem* 68:443–456.
- Wrehlke C, Schmitt-Wrede H, Qiao Z, Wunderlich F. 1997. Enhanced expression in spleen macrophages of the mouse homolog to the

- human putative tumor suppressor gene ZFM1. *DNA Cell Biol* 16:761–767.
- Wu JY, Maniatis T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75:1061–1070.
- Xiao SH, Manley JL. 1998. Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J* 17:6359–6367.
- Yuryev A, Patturajan M, Litingtung Y, Joshi RV, Gentile C, Gebara M, Corden JL. 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci USA* 93:6975–6980.
- Zahler AM, Neugebauer KM, Stolk JA, Roth MB. 1993. Human SR proteins and isolation of a cDNA encoding SRp75. *Mol Cell Biol* 13:4023–4028.
- Zamore PD, Green MR. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci USA* 86:9243–9247.
- Zamore PD, Patton JG, Green MR. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355:609–614.
- Zhang D, Childs G. 1998. Human ZFM1 protein is a transcriptional repressor that interacts with the transcription activation domain of stage-specific activator protein. *J Biol Chem* 273:6868–6877.
- Zhang H, Blumenthal T. 1996. Functional analysis of an intron 3' splice site in *Caenorhabditis elegans*. *RNA* 2:380–388.
- Zhang M, Marr T. 1994. Fission yeast gene structure and recognition. *Nucleic Acids Res* 22:1750–1759.
- Zorio DA, Blumenthal T. 1999. U2AF35 is encoded by an essential gene clustered in an operon with RRM/cyclophilin in *Caenorhabditis elegans*. *RNA* 5:487–494.
- Zorio DA, Lea K, Blumenthal T. 1997. Cloning of *Caenorhabditis* U2AF65: An alternatively spliced RNA containing a novel exon. *Mol Cell Biol* 17:946–953.