RNA (1996), 2:794–810. Cambridge University Press. Printed in the USA.

Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA

SILKE ARNING,[1,3] PATRIC GRÜTER,[1] GRAEME BILBE,[2] and ANGELA KRÄMER[1]

[1] Département de Biologie Cellulaire, Université de Genève, CH-1211 Genève 4, Switzerland
[2] Pharma Research, Ciba-Geigy Ltd., CH-4002 Basel, Switzerland

ABSTRACT

Mammalian splicing factor SF1 consists of a single polypeptide of 75 kDa and is required for the formation of the first ATP-dependent spliceosomal complex. Three cDNAs encoding variant forms of SF1 have been isolated and four highly related cDNAs have been found in current databases. Comparison of the cDNA sequences suggests that different SF1 mRNAs are generated by alternative splicing of a common pre-mRNA. In agreement with this idea, at least three mRNAs that are differentially expressed in different cell types have been detected by northern blot analysis. All SF1 cDNAs identified encode proteins with a common N-terminal half that contains two structural motifs implicated in RNA binding (an hnRNP K homology [KH] domain and a zinc knuckle), but the proteins differ in the length of a proline-rich region and have distinct C-termini. Three SF1 isoforms expressed in insect cells via baculovirus transfer vectors show comparable activities in the assembly of a pre-splicing complex. Consistent with the presence of a KH domain and a zinc knuckle, we show that SF1 binds directly to RNA. This interaction appears to be largely sequence-independent with a preference for guanosine- and uridine-rich sequences. The KH domain of SF1 is embedded in a 160-amino acid sequence that is shared with human Sam68, a target of Src during mitosis, as well as *Caenorhabditis elegans* GLD-1 and mouse Qkl, both of which play roles during cellular differentiation. The presence of this shared region in SF1 suggests functions in addition to its role in pre-spliceosome assembly.

Keywords: KH domain; pre-mRNA splicing; RNA binding; spliceosome; zinc knuckle

INTRODUCTION

The two transesterification reactions that result in the splicing of introns from nuclear pre-mRNA require the formation of a complex network of RNA–RNA interactions that involve the pre-mRNA and the RNA portions of five small nuclear ribonucleoprotein particles (snRNPs; for review, see Moore et al., 1993; Madhani & Guthrie, 1994; Ares & Weiser, 1995; Umen & Guthrie, 1995b). In addition, snRNP-associated as well as non-snRNP proteins engage in protein–RNA and protein–protein contacts and thus aid in the assembly of the active spliceosome, the site of intron removal. Although nuclear pre-mRNA splicing is almost certainly RNA-mediated, proteins play pivotal roles in this process and many essential splicing factors have been identified in mammalian cells and in yeast over the last decade (reviewed in Beggs, 1995; Krämer, 1995). Similar to the snRNPs that are incorporated into the spliceosome in a well-defined and step-wise fashion, protein factors interact with the spliceosome at defined stages of the reaction (see Bennett et al., 1992; Gozani et al., 1994 and references therein).

The first detectable splicing complex (complex E) formed in mammalian cells requires U2AF, which binds to the polypyrimidine tract at the 3′ splice site (Zamore & Green, 1989), and members of a family of proteins characterized by a C-terminal domain rich in arginine–serine dipeptides (SR proteins; for review see Fu, 1995) that act in concert with U1 snRNP to commit the pre-mRNA to the splicing pathway. SF3a and SF3b function in the formation of the pre-spliceosomal complex A by converting the inactive 12S U2 snRNP into the active 17S U2 particle (Brosi et al., 1993a, 1993b) that binds to the pre-mRNA branch site. Six of seven polypeptides associated with SF3a and SF3b interact

Reprint requests to: Angela Krämer, Département de Biologie Cellulaire, Sciences III, Université de Genève, 30, quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland; e-mail: kraemer@sc2a.unige.ch.

[3] Present address: Univ. Augenklinik Abt. II, Molekulargenetik, VG, Auf der Morgenstelle 15, D-72076 Tübingen, Germany.

directly with the pre-mRNA in the vicinity of the branch site and may thus tether U2 snRNP to the pre-mRNA (Champion-Arnaud & Reed, 1994; Gozani et al., 1996). The following step, the formation of splicing complex B, entails the binding of the U4/U6 and U5 snRNPs to the spliceosome in the form of a tri-snRNP. This step requires the tri-snRNP-specific polypeptides (Utans et al., 1992) and SR proteins (Roscigno & García-Blanco, 1995). The active spliceosome (complex C) is formed after a conformational rearrangement and splicing catalysis is initiated. SF4, a less well characterized protein, is essential for the formation of complex C and/or the first transesterification reaction (Utans & Krämer, 1990), whereas PSF is required for the second catalytic step (Gozani et al., 1994).

A number of mammalian splicing factors are associated with RNA-binding activities and structural domains that are essential for the interaction with RNA have been identified. Specific interactions with the pre-mRNA have been reported for U2AF, PSF, and PTB, a protein that functions in alternative splicing (Singh et al., 1995). These proteins bind to the polypyrimidine tract located at the 3′ splice site of most vertebrate introns. High-affinity binding of U2AF to the polypyrimidine tract requires the three RNA recognition motifs (RRM; Zamore et al., 1992) located in the C-terminal half of U2AF. RRMs are also present in the SR proteins and are similarly essential for the RNA-binding activity of these proteins (Fu, 1995). Initially, the SR proteins appeared to bind to RNA in a nonspecific fashion (Krainer et al., 1990); however, by selection of RNA sequences from pools of random oligoribonucleotides, it has been demonstrated that individual members of this protein family recognize different and specific RNA sequences (Heinrichs & Baker, 1995; Tacke & Manley, 1995). Other structural domains implicated in RNA-binding that have been found in splicing proteins are zinc fingers in SF3a60 (Chiara et al., 1994; Krämer et al., 1994) and SF3a66 (Bennett & Reed, 1993), zinc knuckles in the human SR protein 9G8 (Cavaloc et al., 1994) and in the yeast splicing factor Slu7p (Frank & Guthrie, 1992), the hnRNP K homology (KH) domain in the alternative splicing regulators PSI of *Drosophila* (Siebel et al., 1995) and yeast MER1 (Engebrecht & Roeder, 1990), and the arginine-glycine-glycine (RGG) motif in human PSF (Patton et al., 1993).

We have reported previously the purification of SF1, a heat-stable protein of 75 kDa (Krämer, 1992). SF1 is required for the assembly of pre-splicing complex A, but its exact function is unknown. Here we have isolated three cDNA clones encoding putative isoforms of SF1. Recombinant forms of SF1 expressed in insect cells support pre-splicing complex formation in vitro. We furthermore show that SF1 binds to RNA, which is in agreement with the presence of two structural motifs implicated in RNA-binding, a KH domain and a zinc knuckle.

RESULTS

cDNA cloning suggests the presence of multiple SF1 isoforms

cDNA libraries from human bone and HeLa cells were screened with degenerate oligonucleotides that were derived from amino acid sequences of tryptic peptides of purified SF1. A 2,298-bp cDNA (SF1-Bo; Fig. 1A) isolated from a bone cDNA library contains an open reading frame of 1,914 bp encoding a protein of 638 amino acids with a calculated molecular mass of 68.6 kDa and an isoelectric point of 9.92. Nine of eleven tryptic peptides derived from purified SF1 are present within the deduced amino acid sequence (see the Materials and methods; Fig. 2). The presumptive ATG initiation codon is located 295 bp from the 5′ end of the cDNA (not shown). The sequences surrounding this ATG codon conform to the consensus sequence for eukaryotic translation initiation (Kozak, 1989) and in-frame stop codons are present 120-bp and 264-bp upstream. The 3′ untranslated region of SF1-Bo is 87 bp in length and no polyA tail nor a consensus polyadenylation signal are present at the 3′ end.

Two cDNAs have been isolated from HeLa cDNA libraries (Fig. 1A). SF1-HL1 is 1,154 bp in length and ends at an internal *Eco*R I site within the protein-coding sequence. SF1-HL2 is 2,512 bp in length and is truncated at the 5′ end. In comparison to the SF1-Bo cDNA, it contains additional 3′ untranslated sequences and ends in an oligoA tail, which is preceded by a consensus AATAAA polyadenylation signal. The open reading frame of SF1-HL2 differs in two aspects from the SF1-HL1 and/or SF1-Bo sequences. First, SF1-HL2 lacks 21 bp within the protein-coding region that are present in the other cDNAs (Fig. 1A). Second, it contains a 432-bp insertion further downstream that is absent from SF1-Bo. Hence, the predicted amino acid sequence of SF1-HL2 differs from SF1-Bo and SF1-HL1 by the absence of seven internal amino acids (Fig. 2B). Furthermore, conceptual translation of SF1-HL2 continues into the 432-bp sequence up to a termination codon located 131 bp from the 5′ end of the insertion. This yields a C-terminal sequence of 43 amino acids for the predicted SF1-HL2 protein, which differs in composition from the 42-amino acid C-terminal sequence encoded in the SF1-Bo cDNA (Fig. 2B). (For the remainder of this report, we will refer to SF1-HL1 as a putative SF1 cDNA that contains both the 21-bp and the 432-bp insertion.)

cDNA sequences that are very similar to the SF1 cDNAs have been retrieved from current databases (Fig. 1). In a search for the gene responsible for multiple endocrine neoplasia type 1 (MEN1), Toda et al. (1994) cloned transcribable sequences from the human chromosomal locus 11q13. The composite cDNA sequence derived from overlapping clones was desig-

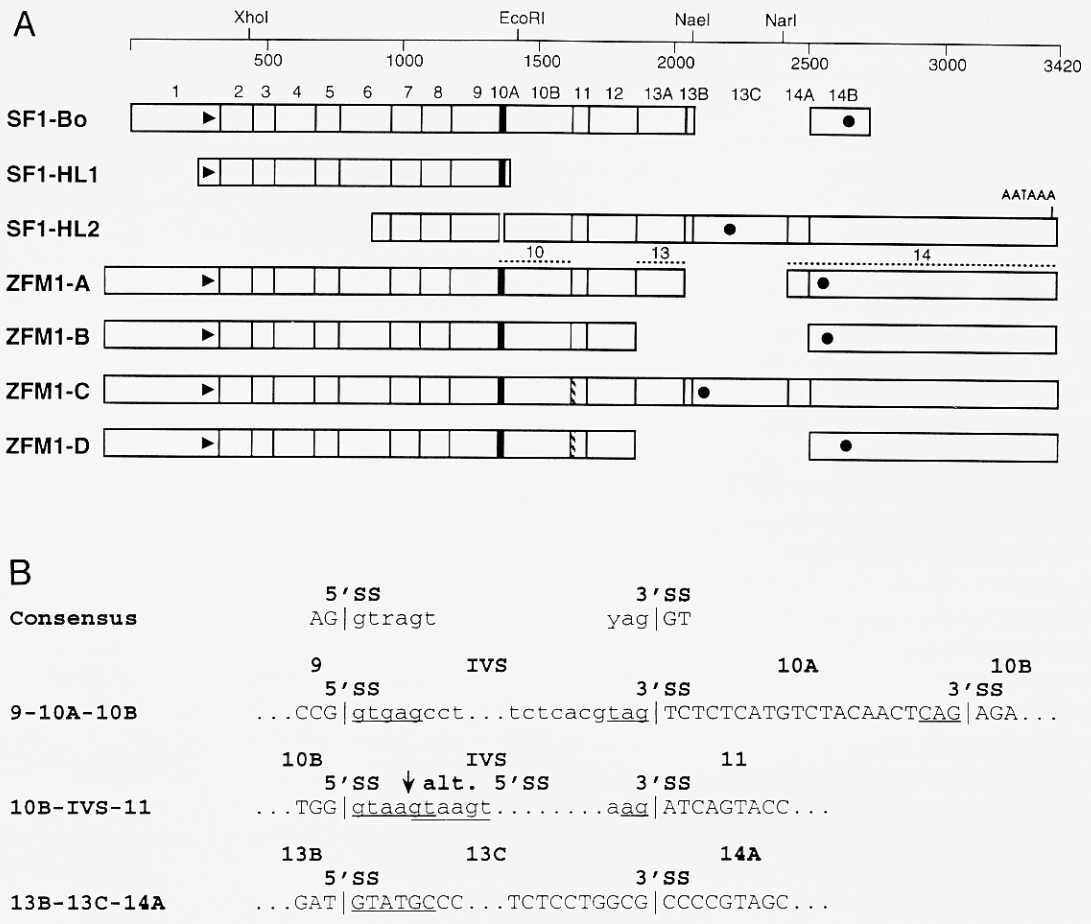

FIGURE 1. Schematic representation of SF1 cDNAs and exon–intron junctions. **A:** SF1 cDNAs isolated from human bone (SF1-Bo) and HeLa libraries (SF1-HL1 and SF1-HL2) as well as several related cDNAs (ZFM1-A to D) are shown. Exons are numbered above SF1-Bo according to Toda et al. (1994) and suffixes have been used to accommodate putative additional exons. The ATG initiation codons are indicated by triangles, the different termination codons are shown by circles. Restriction sites used for the preparation of hybridization probes are indicated on top. **B:** Sequences bordering exons 9-10B, 10B-11, and 13B-14A are compared to the consensus splice site sequences. Exon numbers are indicated above the sequences. Intron sequences are taken from Toda et al. (1994). Matches to the consensus splice sites in the relevant sequences are underlined.

nated ZFM1 (zinc finger gene in the *MEN1* locus) and will be referred to as ZFM1-A. Moreover, putative exon–intron junctions have been determined and intron sequences immediately adjacent to the splice sites have been published (Toda et al., 1994). The ZFM1-A sequence contains ~80 bp at the 5′ end that are not present in the SF1 cDNAs and is contiguous with SF1-Bo up to exon 13A in SF1 (or exon 13 in ZFM1-A; Fig. 1A). Exon 14 of the ZFM1-A cDNA contains 81 bp that are not present in SF1-Bo (indicated as exon 14A in Fig. 1A), but represent part of the 432-bp insertion of SF1-HL2. Conceptual translation of ZFM1-A yields a protein that is identical to SF1-Bo up to amino acid 586, but differs in its 37 C-terminal amino acids (Fig. 2). A second form of ZFM1 (Toda et al., 1994), which will be referred to as ZFM1-B, lacks exons 13A, 13B, and the SF1-HL2 insertion, and encodes 527 amino acids that are identical to SF1-Bo and a unique 21-amino acid C-terminus. A cDNA similar to ZFM1-B has also been isolated from mouse spleen and is derived from a testosterone-inducible transcript (EMBL accession no. X85802). The mouse ZFM1-B coding sequence shows an overall homology of 89% to the corresponding human cDNA. High sequence identity is also observed in the 5′ untranslated region and in the 3′ noncoding region up to 250 bp after the termination codon of SF1-Bo; the sequences then diverge, but significant homology is found for 330 bp preceding the polyA tail (not shown). The corresponding human and mouse proteins are 99.3% identical and differ from one another in four amino acid exchanges. Two further variants (referred to as ZFM1-C and ZFM1-D; EMBL accession no. L49380 and L49345) are almost identical to the SF1-HL1 and ZFM1-B cDNAs, but they contain a GTAA insertion at the border of exons 10B and 11 (Fig. 1A). The ZFM1-C cDNA codes for a protein that

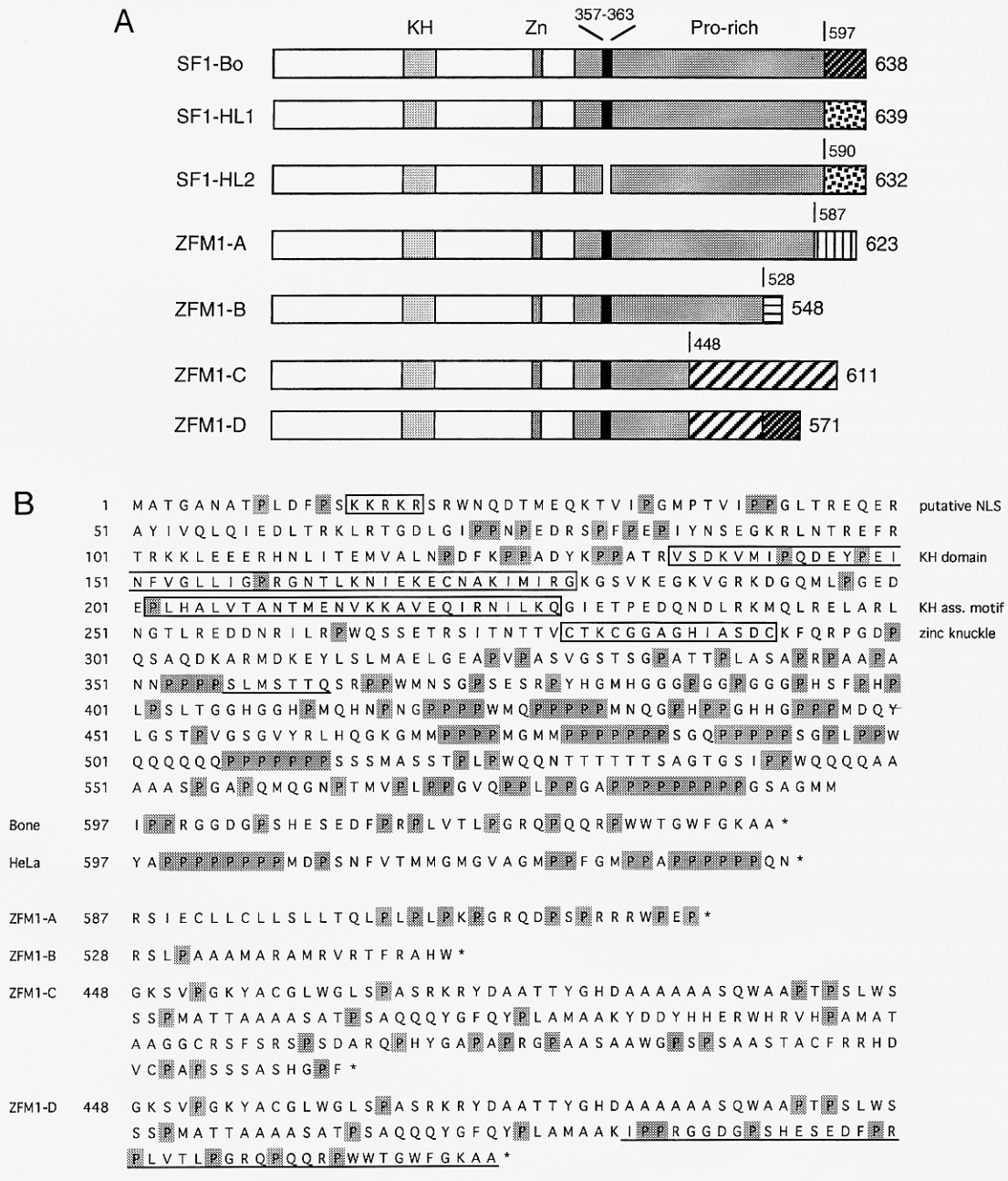

FIGURE 2. Structure of SF1 isoforms. **A:** Schematic representation of SF1 isoforms. Numbers above the boxes refer to the site of divergence between individual proteins. Numbers on the right show the lengths in amino acids. Structural domains in SF1 are indicated by shading. **B:** Amino acid sequence of the part of SF1 that is common to the bone and HeLa isoforms is shown on top; seven amino acids that are absent from SF1-HL2 are underlined. Structural motifs are boxed and indicated on the right, proline residues are shaded. The sequences of the variant C-terminal extensions are shown below; numbers on the left of the sequences indicate the site of divergence from the sequence common to the bone and HeLa isoforms. Underlined sequences in ZFM1-D are identical to the C-terminus of SF1-Bo.

has 447 amino acids in common with SF1-Bo and contains a unique C-terminal sequence of 164 amino acids (Fig. 2). The ZFM1-D sequence also diverges from the SF1-Bo sequence at amino acid 447 and encodes 82 amino acids that are identical to ZFM1-C, followed by a C-terminus that is identical to the last 42 amino acids of SF1-Bo (Fig. 2).

The various SF1 cDNAs most likely represent alternatively spliced transcripts that are derived from a common pre-mRNA. The 3′ end of the 21-bp sequence, which is included in exon 10 in the ZFM1 cDNA (Fig. 1; Toda et al., 1994), shows a good match to the consensus 3′ splice site sequence YAG| (Fig. 1B; Y = T or C, | indicates the intron/exon border;

Stephens & Schneider, 1992) and no similarity to the consensus 5′ splice site (|GTRAGT; R = A or G) is detected. Thus, inclusion or exclusion of the 21-bp sequence in different mRNAs may be caused by alternative splicing to a duplicated 3′ splice site. The 5′ end of the sequence present in SF1-HL2 cDNA (exon 13C in Fig. 1), but absent from other cDNA species matches the consensus 5′ splice site (Fig. 1B). This sequence ends in TGGCG, which was interpreted as an unusual 3′ splice site by Toda et al. (1994). However, recent descriptions of minor 3′ splice sites do not include the sequence GCG (Jackson, 1991; Hall & Padgett, 1994). Thus, the insertion present in SF1-HL2 could have been generated by splicing to an alternative 5′ splice site. The most likely explanation for the presence of the GTAA sequence in ZFM1-C and -D is the use of an alternative 5′ splice site that overlaps with the 5′ splice site used in the other mRNAs (Fig. 1B).

Taken together, these data strongly suggest that different isoforms of SF1 are generated by extensive alternative splicing of its pre-mRNA. The sites of divergence between the identified cDNA sequences are restricted to the 3′ half of the protein-coding sequence. This is consistent with results obtained by PCR amplification of HeLa cell cDNA with primers specific to different parts of the SF1 cDNA. Primers specific to exons 1 and 12 generate a single detectable PCR product, whereas amplification with primers specific to exons 12 and 14B results in at least seven PCR products (F. Mulhauser & A. Krämer, unpubl.).

SF1 mRNAs are expressed in a cell type-specific fashion

Hybridization of a northern blot containing polyA$^+$ RNA from HeLa cells with a *Xho* I/*Eco*R I fragment derived from the 5′ part of the SF1 cDNA that is common to all isolated forms, detected mRNAs of approximately 3.5, 2.9, and 2.5 kb (Fig. 3A). Thus, at least three SF1 mRNAs that differ in size are expressed in HeLa cells. The size of the 3.5-kb mRNA is in good agreement with the sizes predicted for the SF1-HL2 and ZFM1-C mRNAs. This assignment is consistent with the result that an *Nae* I/*Nar* I fragment derived from the sequence unique to these cDNAs hybridizes only to the 3.5-kb mRNA (Fig. 3A). The 2.9-kb mRNA most likely corresponds to mRNAs for the SF1-Bo and ZFM1-A forms, whereas the size of the third mRNA of ~2.5 kb is most closely related to the size predicted for ZFM1-B and ZFM1-D mRNAs.

Northern blot hybridization of total RNA from several human cell lines or polyA$^+$ RNA from human and murine tissues with the common *Xho* I/*Eco*R I fragment revealed that, in all cell lines and tissues examined, the two larger mRNAs of 2.9 and 3.5 kb represent the major species (Fig. 3B). The smaller 2.5-kb mRNA is present in three of the human cell lines analyzed (Wish, RPMI, and U937), in all mouse tissues, and in human liver. The northern blots were not standardized for the amount of RNA loaded in different lanes; it is, however, apparent that the mRNAs are expressed at different ratios depending on cell line or tissue. For example, in monocytes as well as in human heart and murine liver, the 2.9-kb mRNA is clearly more abundant than the 3.5-kb species. In human and mouse brain, the 3.5-kb mRNA is more abundant than the 2.9-kb mRNA species and these two mRNAs are expressed at approximately equimolar ratios in other tissues or cell lines. Taken together, these results indicate that three mRNAs of predicted sizes are expressed in various mammalian tissues and cell lines, suggesting that the cDNAs isolated originate from genuine mRNAs.

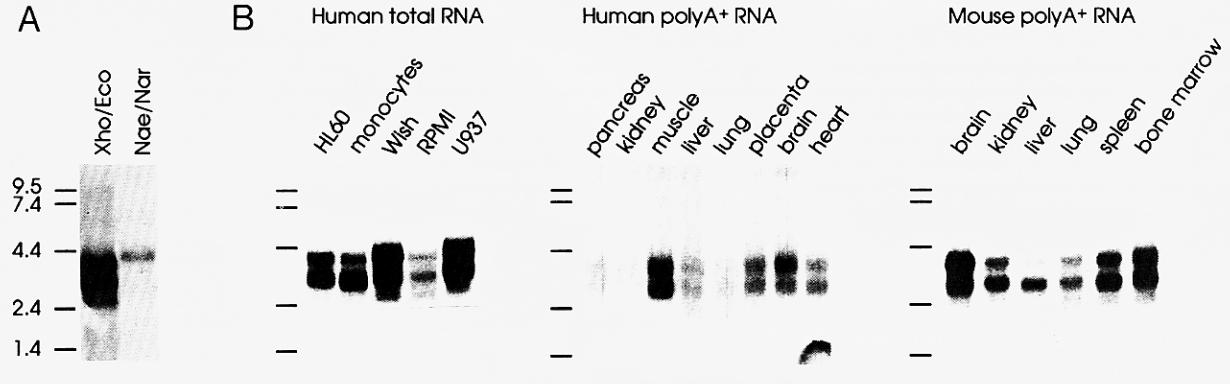

FIGURE 3. Northern blot analysis of SF1 mRNAs. **A:** PolyA$^+$ RNA from HeLa cells was probed with a *Xho* I/*Eco*R I restriction fragment derived from the 5′ half of SF1 that is common to all SF1 cDNAs isolated or with an *Nae* I/*Nar* I fragment derived from the insertion that is present in the SF1-HL2 cDNA. **B:** Total RNA from human cell lines and polyA$^+$ RNA from various human and mouse tissues was probed with a *Xho* I/*Eco*R I fragment. Cell lines and tissues are indicated on top of the figure. Relative migration of RNA size markers is shown on the left.

Structural organization of SF1

The SF1 protein can be divided into distinct structural domains. The N-terminal half (amino acids 1–323), common to all isoforms, is highly hydrophilic and contains 32% charged residues (Fig. 2B). It contains a putative nuclear localization signal (amino acids 15–19; Dingwall & Laskey, 1991), an hnRNP K homology (KH) domain (amino acids 136–179; Siomi et al., 1993a), and a zinc knuckle (amino acids 279–292; Katz & Jentoft, 1989). The central portion of the SF1-Bo and SF1-HL isoforms is rich in proline (32%) and glycine residues (13%). This region is shorter in the other isoforms and is followed by the individual C-termini of different lengths and amino acid composition (Fig. 2B).

The KH domain was first described in hnRNP K, a major polyC-binding protein in HeLa cells (Matunis et al., 1992; Siomi et al., 1993a) and similar domains have since been identified in a number of proteins that are in physical and/or functional contact with RNA or DNA (for examples see Gibson et al., 1993; Siomi et al., 1993b). The KH domain can be present in a protein in one or several copies, with vigilin containing an exceptional 14 KH domains (Schmidt et al., 1992).

In Figure 4A, the KH domain of SF1 is compared to the KH domains of hnRNP K and two splicing proteins, *Drosophila* PSI (Siebel et al., 1995) and *Saccharomyces cerevisiae* MER1 (Engebrecht & Roeder, 1990). All highly conserved positions of a typical KH domain (for examples see Gibson et al., 1993; Siomi et al., 1993b) are conserved in SF1. In addition, several positions that show a conservation in more than 33% of 52 eukaryotic KH domains examined (A. Krämer, unpubl.) are present in SF1. A deviation from most KH domains is represented by the insertion of six amino acids in the N-terminal part of the KH domain of SF1, allowing for an optimal sequence alignment. Similar insertions are found in the KH domains of the mouse *quaking* gene product QkI (Ebersole et al., 1996), *Caenorhabditis elegans* GLD-1 (Jones & Schedl, 1995), human Sam68 (Wong et al., 1992), and *Artemia salina* Grp33 (Cruz-Alvarez & Pellicer, 1987). Moreover, a protein encoded on chromosome XII of *S. cerevisiae* and the predicted amino acid sequence of an expressed sequence tag (EST) of *Caenorhabditis briggsae* (see below) contain KH domains with high homology to that of SF1. In addition to these examples, the N-terminal KH domains of NusA of *Escherichia coli* and *Bacillus subtilis* contain a five-amino acid insertion at the same position (Gibson et al., 1993).

Many, but not all KH domains are followed at a variable distance by a conserved sequence of five amino acids (Gibson et al., 1993) and a stretch of amino acids that can be folded into an amphipathic helix (Duncan et al., 1994; Castiglione Morelli et al., 1995). Similar sequences are found preceding the N-terminal KH domain of some proteins, for example, in vigilin (Schmidt et al., 1992) and PSI (Siebel et al., 1995), but not in hnRNP K (Matunis et al., 1992) or MER1 (Engebrecht & Roeder, 1990). A multiple sequence alignment (generated with the PILEUP program) of 57 sequences preceding or following eukaryotic KH domains (not shown) resulted in the consensus sequence presented in Figure 4B. We will refer to this sequence as the KH-associated motif. Seven positions in the KH-associated motif are conserved in ≥75% of 57 sequences examined and 13 positions show a conservation of ≥33% (Fig. 4B; A. Krämer, unpubl.). The predicted KH-associated motifs of human SF1 and the *S. cerevisiae* open reading frame (and, in part, the *C. briggsae* EST sequence) fit the consensus in five highly conserved positions with a one-amino acid insertion in the N-terminal part. A similar conservation is found for the KH-associated motifs of QkI, GLD-1, Sam68, Grp33; however, optimal sequence alignment requires an insertion of six amino acids when compared to a typical KH-associated motif.

Another conserved motif found in SF1 is a zinc knuckle (Fig. 4C), which is characteristic for retroviral nucleocapsid proteins (Berg, 1986; Rein, 1994) and is also present in the SR protein 9G8 (Cavaloc et al., 1994) and in the yeast splicing factor Slu7p (Frank & Guthrie, 1992; Fig. 4C). In contrast to zinc finger motifs, the spacing of the cysteine and histidine residues in zinc knuckles is strictly defined and additional amino acids are conserved (for alignments see Berg, 1986; Frank & Guthrie, 1992). In addition to the invariant cysteine and histidine residues, the sequence of the zinc knuckle in SF1 conforms to the consensus in four conserved positions. Furthermore, two positively charged amino acids are present immediately C-terminal to the zinc knuckle, as shown for other proteins containing this motif (Katz & Jentoft, 1989).

The central portion of the SF1-Bo and SF1-HL2 isoforms is particularly rich in proline residues, many of which are present as stretches of consecutive prolines (Fig. 2B). In addition, stretches of glutamine, threonine, and alanine residues are present and this region contains a high number of methionines. Proline-rich regions are found in several other splicing factors, for example, PSF (Patton et al., 1993), SF3a66 (Bennett & Reed, 1993), SF3a120 (Krämer et al., 1995), SAP49 (Champion-Arnaud & Reed, 1994), and SAP145 (Gozani et al., 1996); however, apart from the proline richness, no homology among these sequences is apparent.

Comparison of the unique C-terminal sequences of the potential SF1 isoforms to database entries did not yield significant matches to known proteins. In a search for amino acid sequence motifs, differences in potential phosphorylation sites among the individual isoforms were detected. In addition, the ZFM1-C and -D forms contain the sequence A/GXXXXGKS/T between amino acids 443 and 450 (Fig. 2B), which matches the consensus sequence for the ATP/GTP binding site motif A (or P-loop; reviewed in Saraste

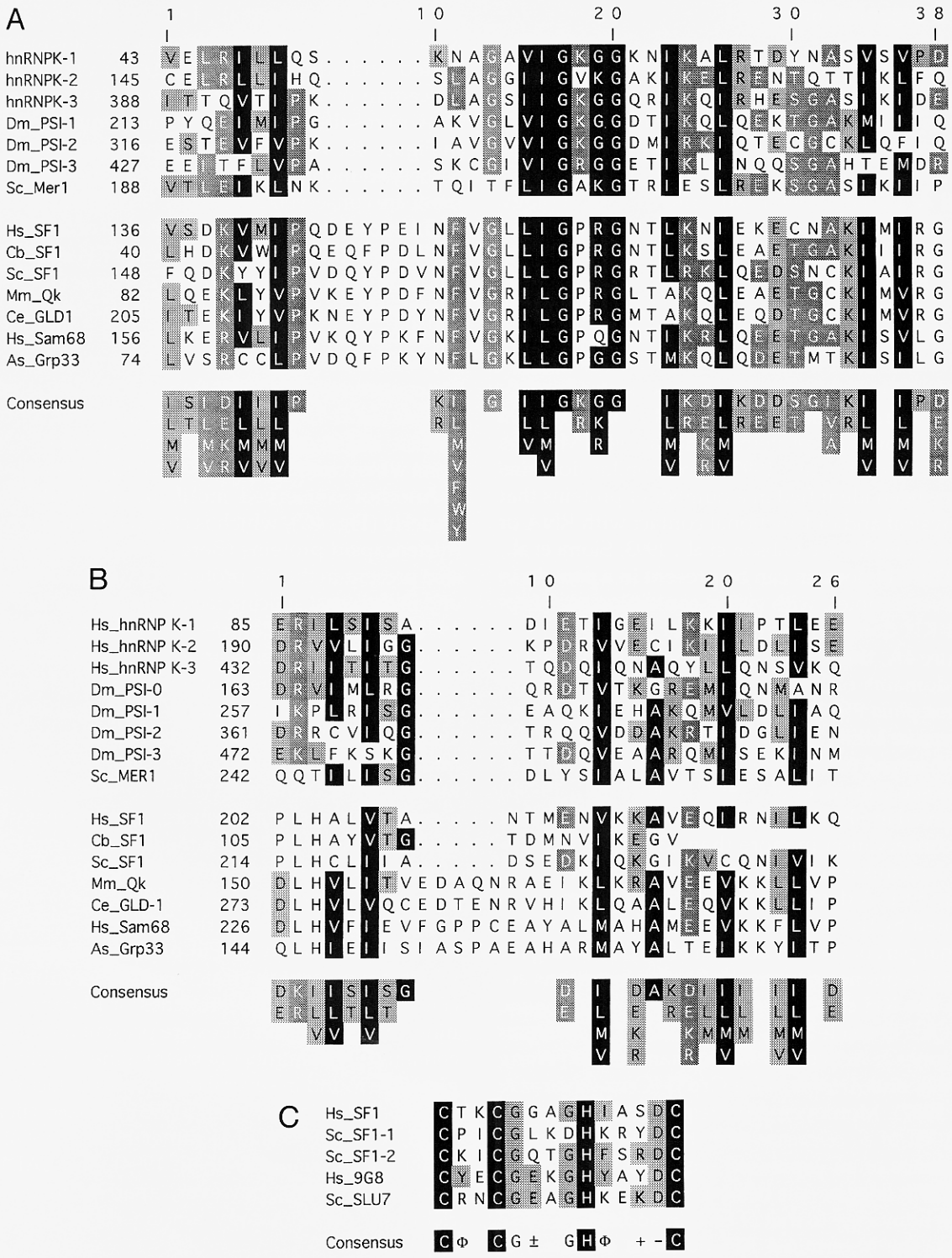

FIGURE 4. Amino acid sequence alignment of KH domains, KH-associated motifs, and zinc knuckles. **A:** An alignment of 52 eukaryotic KH domains was generated with the PILEUP program of the GCG package (A. Krämer, unpubl.) and the KH domains of human hnRNP K (Hs_hnRNP K), *Drosophila* PSI (Dm_PSI), yeast MER1 (Sc_MER1), human SF1 (Hs_SF1), putative SF1 homologues of *S. cerevisiae* (Sc_SF1) and *C. briggsae* (Cb_SF1), mouse QkI (Mm_Qk), *C. elegans* GLD-1 (Ce_GLD1), human Sam68 (Hs_Sam68), and *A. salina* Grp33 (As_Grp33) were selected. Positions with ≥75% sequence conservation in the original alignment are shown in white on black, sequence conservation of ≥50% is indicated in white on gray, and of ≥33% in black on gray. Conservative groupings were I = L = M = V, F = W = Y, D = E, K = R, and S = T. Numbers on the left indicate the first amino acid of the domains. The consensus sequence derived from the original alignment is shown on the bottom. **B:** An alignment of 57 KH-associated domains was generated with the PILEUP program and optimized by eye. The KH-associated domains of the proteins in A were selected and compared to the corresponding domain of SF1. **C:** Amino acid sequences of zinc knuckle motifs in SF1, yeast Slu7p, and human 9G8 are shown. The consensus sequence was derived from a further alignment of other zinc knuckle-containing proteins (Katz & Jentoft, 1989; Frank & Guthrie, 1992). Φ indicates a preference for hydrophobic amino acids, ±, +, and − designate preferences for charged amino acids.

et al., 1990). Whether this sequence has any relevance for the function of ZFM1-C and -D and whether these proteins indeed bind ATP or GTP remains to be analyzed. The SF1-HL2 protein differs from other isoforms by a deletion of seven amino acids (SLMSTTQ) in the central, common portion of SF1 (Fig. 2B). Again, no significant matches to this sequence were found in current databases.

Proteins with homology to SF1

Two protein sequences that share high homology with the mammalian SF1 sequence have been retrieved from current databases. The amino acid sequence of an uncharacterized open reading frame (477 amino acids) encoded on chromosome XII of *S. cerevisiae* (EMBL accession no. U53877; gene L9233.6) is 37% identical and 55% similar to human SF1. The yeast protein has a predicted molecular mass of 53 kDa and an isoelectric point of 10.49. The homology extends over the entire common portion of the proteins and is particularly high in the N-terminal half, including the KH domain and the KH-associated motif (Fig. 5). The C-terminal half of the *S. cerevisiae* protein contains two zinc knuckles that are separated by 12 amino acids, in contrast to a single zinc knuckle in SF1, and the remaining amino acids are enriched in proline and serine residues. Thus, except for the presence of an additional zinc knuckle, the structural organization of the yeast protein and mammalian SF1 is strikingly similar, suggesting that the open reading frame encodes the yeast homologue of human SF1.

Part of a putative *C. briggsae* homologue of SF1 is encoded by an expressed sequence tag (EMBL accession no. R02974). The predicted amino acid sequence is 58% identical to a region of 122 amino acids of SF1 that includes the KH domain (Figs. 4, 5).

In addition, database searches identified four protein sequences that share significant homology with SF1 over ~160 amino acids (Fig. 5): the QkI protein, which is encoded by the gene responsible for the quaking phenotype in mouse (Ebersole et al., 1996); *C. elegans* GLD-1, the product of a germ-line-specific tumor suppressor gene that is essential for oogenesis (Jones & Schedl, 1995); human Sam68, a major tyrosine-phosphorylated protein in transformed cells (originally termed p62; see Wong et al., 1992; Lock et al., 1996); and *A. salina* hnRNP protein Grp33 (Cruz-Alvarez & Pellicer, 1987). The similarity between SF1 and these proteins in the 160-amino acid region ranges from 52.8 to 65.5% (identities are 26.4–39.9%), whereas QkI, GLD-1, Sam68, and Grp33 show a higher conservation among one another (similarities of 57.5–82.9%, identities of 28.1–67.1%). These proteins contain a single KH domain, as well as a KH-associated motif, but no zinc knuckles (Figs. 4, 5). The C-terminal amino acid sequences of QkI, GLD-1, Sam68, and Grp33 are relatively glycine- and/or proline-rich; however, the sequence conservation in this part is limited.

Different SF1 isoforms are active in pre-splicing complex formation

The apparent size of SF1 purified from HeLa cells is 75 kDa (Krämer, 1992), whereas the sizes predicted for the SF1-Bo, SF1-HL1, and SF1-HL2 proteins are 68.6, 68.3, and 67.6 kDa, respectively. In vitro-translated SF1 proteins, however, migrate close to SF1 from HeLa cells (Fig. 6) with apparent sizes of 78–80 kDa, suggesting that the cDNAs isolated encode full-size proteins. The smaller polypeptides observed in this particular experiment are most likely proteolytic degradation products of SF1 or are caused by translation of prematurely terminated RNA transcripts synthesized in the coupled in vitro transcription/translation reaction. Small size differences are observed between the in vitro-translated proteins. Compared to the SF1-HeLa isoforms, the SF1-Bo protein migrates with a smaller size. It is unclear whether this is due to proteolytic degradation, the different amino acid composition of the C-terminus, or to posttranslational modification.

We note that it is not clear which isoform(s) of SF1 were purified from HeLa cells, because all peptides sequenced are derived from the common portion of SF1 (not shown). In the most purified SF1 preparations, a slightly smaller protein was sometimes observed that cofractionated with the major polypeptide (Krämer, 1992). The smaller protein was thought to represent a proteolytic product of SF1, but, given the fact that different SF1 mRNAs are expressed in HeLa cells (Fig. 3A), it is also possible that this minor protein represented another SF1 isoform.

To test a function of recombinant SF1 in vitro, we tried to express SF1-Bo, SF1-HL1, and SF1-HL2 in bacteria. Attempts to produce the full-size proteins from a variety of bacterial expression vectors have been unsuccessful; therefore, the protein-coding sequences of the three isoforms were integrated into baculovirus DNA via baculovirus transfer vectors, followed by expression of the proteins in insect cells. SDS-PAGE of nuclear extracts prepared from cells infected with recombinant virus encoding the three SF1 isoforms indicates that the proteins are expressed efficiently when compared to a control extract (Fig. 6). All three proteins migrate similar to HeLa cell SF1 and the size differences between the isoforms appear to be equivalent to those observed in the in vitro-translated SF1 proteins.

We have shown previously that SF1 is required for the formation of pre-splicing complex A (Krämer, 1992). Similarly, the recombinant SF1 proteins are active in complex formation, when tested in a reconstituted in vitro system in the presence of SF3a, SF3b, U2AF, U1, and U2 snRNPs (Fig. 7). (The SR proteins that are essential for this reaction [Fu, 1995] are pro-

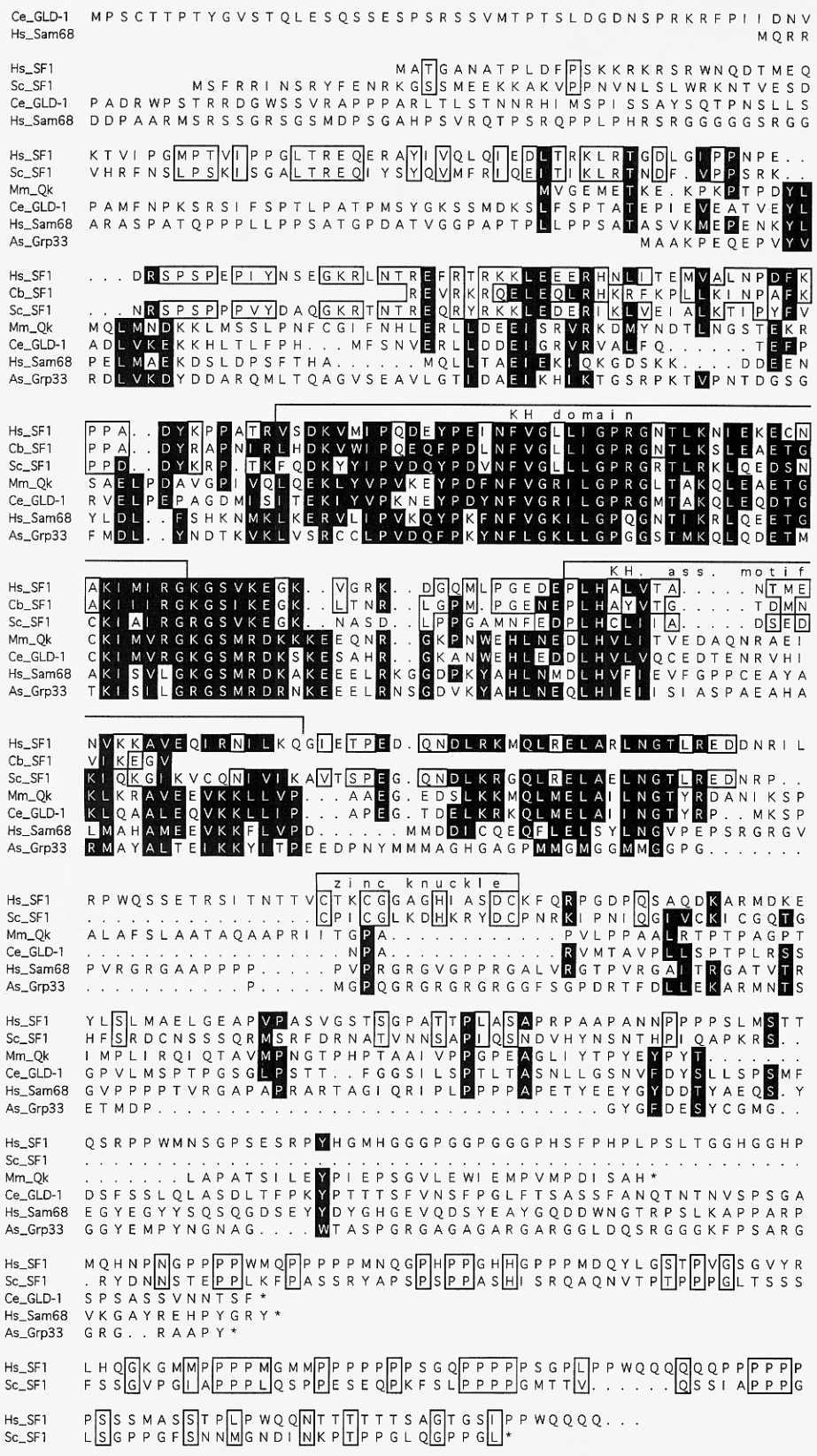

FIGURE 5. Amino acid sequence alignment of human SF1, putative yeast, and *C. briggsae* homologues of SF1, QkI, Sam68, GLD-1, and Grp33. The alignment was generated with the PILEUP program (gap creation penalty 3.00, gap extension penalty 0.05). Amino acids that are conserved in at least four proteins (see Fig. 4A) are shown in reverse print. Additional amino acids that are conserved between human SF1 and the related proteins in yeast and *C. briggsae* are boxed. The positions of the KH domain, the KH-associated domain, and the zinc knuckle of SF1 are indicated above the alignment.

vided with the partially purified snRNP preparations.) In a reaction lacking SF1, pre-splicing complex assembly is reduced to background levels (Fig. 7, lane 2). Addition of a nuclear extract prepared from insect cells infected with a virus encoding SF1-Bo results in the efficient formation of complex A in a concentration-dependent manner (Fig. 7, lanes 4–8), whereas a nuclear extract prepared from a control cell line infected

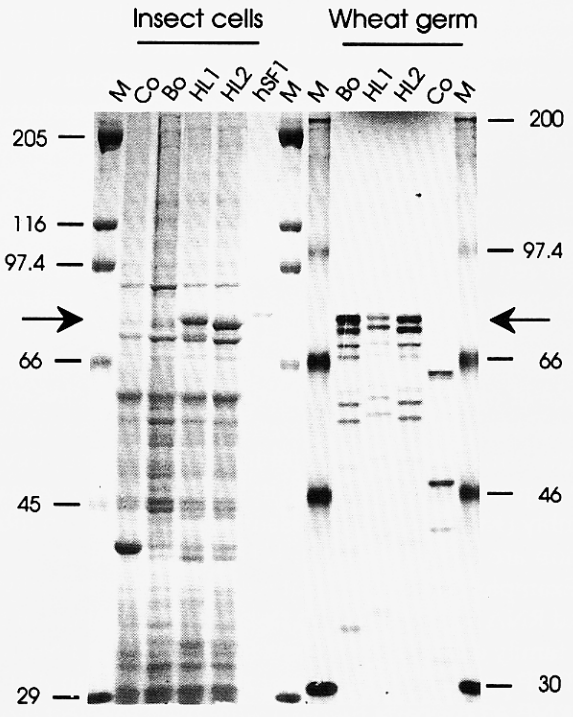

FIGURE 6. Expression of SF1 proteins in vitro and in vivo. Proteins were expressed in insect cells infected with baculovirus encoding the SF1-Bo, -HL1, and -HL2 isoforms or an unrelated protein. Total nuclear proteins were analyzed in a 10% SDS polyacrylamide gel and visualized by Coomassie-blue staining. Expression of proteins in wheat germ lysates was performed in a coupled transcription/translation system. The control reaction represents the translation of luciferase. In vitro-translated proteins were separated in the same gel as insect cell-expressed proteins and visualized by autoradiography. For comparison, HeLa cell-purified SF1 (hSF1) was separated in the same gel. The relative sizes of unlabeled (Sigma) and 14C-labeled (Amersham) size markers are given according to the suppliers' information.

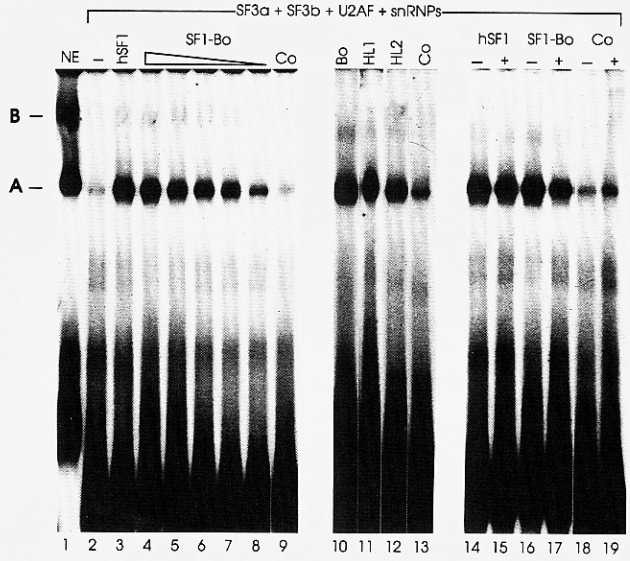

FIGURE 7. Pre-splicing complex formation in the presence of recombinant SF1. The activity of recombinant SF1 proteins was compared to the activity of SF1 isolated from HeLa cells (Blue-Sepharose fraction; Krämer, 1992) in the presence of partially purified splicing factors SF3a, SF3b, U2AF, U1, and U2 snRNPs (Krämer & Utans, 1991). The positions of pre-splicing complex A and splicing complex B are indicated on the left. Partially purified SF1 from HeLa cells (hSF1; lane 3), nuclear extract from insect cells infected with virus encoding SF1-Bo (2, 1, 0.3, 0.1, and 0.03 μL; lanes 4–8), or an unrelated virus (2 μL; lane 9) were added to reconstituted splicing reactions. Complex formation in the absence of SF1 is shown in lane 2; lane 1 represents splicing complex formation in a HeLa cell nuclear extract (NE). Lanes 10–13 show reactions in the presence of nuclear extracts (2 μL) from insect cells infected with virus encoding SF1-Bo, -HL1, and -HL2 or with control virus. In the reactions shown in lanes 14–19, HeLa cell SF1 (hSF1) or nuclear extracts from insect cells infected with virus encoding SF1-Bo or control virus were incubated for 15 min on ice (–) or at 100 °C (+). After centrifugation to remove denatured proteins, the supernatants were added to splicing reactions as indicated above the figure.

with an unrelated virus is inactive in complex assembly (Fig. 7, lane 9). SF1-HL1 and SF1-HL2 promote complex formation in a similar fashion (Fig. 7, lanes 11, 12). The activity of SF1 isolated from HeLa cells is resistant to incubation at 100 °C (Krämer, 1992) and recombinant SF1 shows similar properties, although a slight reduction in activity is observed (Fig. 7, lanes 14–19). Taken together, these results confirm that the cDNAs isolated encode SF1 and demonstrate that the three forms analyzed are functional in pre-spliceosome assembly.

SF1 binds directly to RNA

The presence of a KH domain and a zinc knuckle, two protein motifs implicated in RNA binding, suggested that SF1 can interact directly with RNA. To investigate this possibility, purified SF1 was incubated with radiolabeled RNA, the reactions were subjected to UV crosslinking and, after RNAse A digestion of unprotected RNA, crosslinked proteins were separated by SDS-PAGE and visualized by autoradiography. Figure 8A demonstrates that SF1 binds directly to U1 snRNA and this interaction is independent of time of incubation and ATP. Similarly, SF1 expressed in insect cells binds to U1 snRNA, whereas no crosslinked protein similar in size to SF1 is detected in an insect control extract (Fig. 8B). SF1 appears to bind to RNA in a nonspecific fashion, because it can be crosslinked to U1 or U2 snRNA, AdML pre-mRNA, or a transcript derived from pBluescript (Fig. 8C and data not shown). Also, the binding to either U1 snRNA or AdML pre-mRNA can be competed efficiently by an excess of different unlabeled transcripts (Fig. 8C). Some specificity becomes apparent when the RNA-binding activity of SF1 is tested in the presence of an excess of polyribonucleotides. Whereas polyG and polyU compete efficiently for the binding to U1 snRNA, no effect is seen with an excess of polyA or polyC (Fig. 8D). Thus, as predicted from the presence of two protein motifs implicated in RNA binding, purified or recombinant SF1 interact directly with RNA.

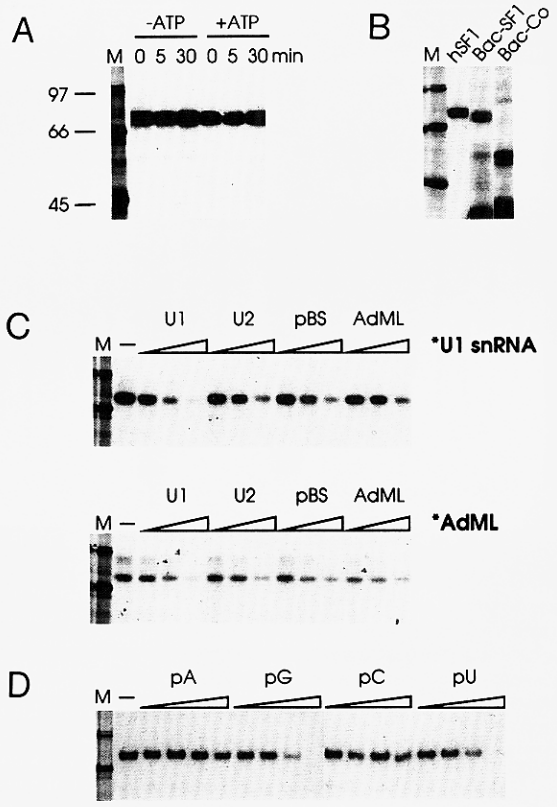

FIGURE 8. RNA binding activity of SF1. **A:** SF1 purified from HeLa cells (~150 ng) was incubated with radiolabeled U1 snRNA at 30 °C in the absence or presence of ATP for the times indicated above the figure. Proteins were UV crosslinked to the RNA and processed for SDS-PAGE as described in the Materials and methods. **B:** Purified SF1 or nuclear extracts from insect cells infected with virus encoding SF1-Bo (1 μL) or control virus (5 μL) were incubated with radiolabeled U1 snRNA for 30 min at 30 °C and subjected to UV crosslinking. **C:** Purified SF1 was incubated with radiolabeled U1 snRNA (80 fmol) or AdML pre-mRNA (55 fmol) in the presence of increasing amounts of unlabeled U1 RNA (1.5, 6, and 21 pmol), U2 snRNA (1.1, 4.4, and 15.4 pmol), pBS RNA (0.65, 2.6, and 9 pmol), or AdML pre-mRNA (0.85, 3.4, and 12 pmol) as indicated above the figure. **D:** Purified SF1 was incubated with radiolabeled U1 snRNA in the presence of increasing amounts (10, 30, 100, and 1,000 ng) of polyA, polyG, polyC, or polyU.

DISCUSSION

Three cDNAs encoding different forms of human splicing factor SF1 that vary in their C-terminal amino acid sequences have been isolated. The sequences of nine peptides of purified SF1 were found in the common portion of the predicted amino acid sequences. Moreover, recombinant SF1 expressed in insect cells can substitute for HeLa cell-purified SF1 in a splicing assay, demonstrating that bona fide cDNAs for SF1 have been cloned.

SF1 is required for the assembly of pre-splicing complex A, but its exact function is unknown (Krämer, 1992). Also, it is unclear at present whether SF1 plays a role in the formation of complex E, which forms prior to complex A in an ATP-independent fashion (Michaud & Reed, 1991; Jamison et al., 1992). A first hint at a possible function is provided by the result that SF1 interacts directly with RNA. The binding appears to be sequence-independent, because SF1 can be crosslinked equally well to the AdML pre-mRNA, U1, and U2 snRNAs, and to a transcript derived from pBluescript vector sequences. In competition experiments with polyribonucleotides, an inhibition of RNA binding was observed in the presence of polyG or polyU, but not with polyA or polyC, suggesting that SF1 interacts preferentially with G- and/or U-rich sequences. The UV crosslinking experiments were performed in the absence of other splicing activities, and it remains to be tested whether SF1 recognizes specific RNA sequences in the context of the assembling spliceosome. Given the requirement of SF1 activity for pre-spliceosome assembly (Krämer, 1992), possible partners for an interaction are the pre-mRNA, U1, or U2 snRNAs.

Examples of proteins that were thought originally to bind RNA in a nonspecific fashion are members of the SR family of splicing proteins (Krainer et al., 1990) and hnRNP proteins (Dreyfuss et al., 1993). When ASF or hnRNP A1 (or parts thereof) were used to select target sequences from pools of random oligoribonucleotides, specific high-affinity RNA-binding sites were defined (Burd & Dreyfuss, 1994; Tacke & Manley, 1995). Moreover, although U2AF65 and PTB both bind to the polypyrimidine tracts of many pre-mRNA introns and to polyU (García-Blanco et al., 1989; Zamore et al., 1992), these proteins selected different, but nevertheless uridine-rich target sequences from random oligoribonucleotide pools (Singh et al., 1995). Similar approaches could be used to determine whether SF1 is associated with any sequence-dependent RNA-binding activity.

Gozani et al. (1996) have recently shown that six of the seven subunits of splicing factors SF3a and SF3b, which interact with the spliceosome as part of the U2 snRNP (Brosi et al., 1993a; Staknis & Reed, 1994), bind directly to the pre-mRNA near the branch site in a defined spacial order. The specific arrangement of the polypeptides on the pre-mRNA remained unaltered when the nucleotide sequence bordering the branch site was changed (Gozani et al., 1996). These results suggest that the contacts with the pre-mRNA are dictated by the higher-order structure of the U2 snRNP and are not a consequence of specific sequence recognition. Although we consider this possibility less likely, SF1 could similarly be guided to a spliceosomal RNA by interaction with other components and then contact the RNA in a sequence-independent fashion at a specific position.

Several proteins that bind to the branch site region have been identified by site-specific crosslinking to a pre-mRNA containing a modified branch point adenosine (MacMillan et al., 1994). Crosslinking of an 80-kDa protein was observed early during the splicing

reaction and required intact U1 snRNP, but no ATP. The protein was not detected in pre-splicing complex A nor spliceosomal complexes B and C. The similar sizes of the 80-kDa protein and SF1 and the function of SF1 in the formation of pre-splicing complex A (or possibly earlier) raise the possibility that these proteins are identical. This issue could be solved by a direct comparison of SF1 with the 80-kDa protein in the specific crosslinking assay.

Three structural motifs in SF1 could be responsible for its RNA-binding activity. KH domains are conserved in a number of RNA-binding proteins in all taxonomic kingdoms (for review, see Gibson et al., 1993). An intact KH domain is essential for the RNA-binding activity of hnRNP K and FMR1, a protein involved in the fragile X syndrome (Siomi et al., 1994), and additional evidence exists that the KH domain is in direct contact with RNA (for references see Musco et al., 1996). Among the splicing proteins, the *Drosophila* protein PSI and the yeast MER1 protein contain three and one KH domains, respectively. Both proteins function in alternative splicing events and appear to bind RNA in a sequence-specific fashion (Nandabalan et al., 1993; Siebel et al., 1994, 1995; Nandabalan & Roeder, 1995); in these cases, it has, however, not been reported whether the KH domains are required for RNA binding. The KH domain of SF1 corresponds in all highly conserved residues to the consensus sequence derived from an alignment of 52 eukaryotic KH domains. For an optimal sequence alignment, an insertion of six amino acids in the N-terminal portion was required, a feature that is shared with QkI, GLD-1, Sam68, and Grp33.

Many eukaryotic KH domains are followed at a variable distance by the sequence I/L/V-S/T-I/L/V-S/T-G (amino acids 4–8 in Fig. 4B), which has been included as part of the KH domain in the multiple sequence alignment by Gibson et al. (1993). An alignment of 57 sequences N- or C-terminal to KH domains revealed additional conserved residues (A. Krämer unpubl.). Amino acids 11–26 in the alignment shown in Figure 4B correspond to the amphipathic helices described for the DNA-binding protein FBP (Duncan et al., 1994) and to an α-helical structure in human vigilin following the KH domain, as determined by NMR spectroscopy (Castiglione Morelli et al., 1995). Recent analysis of the three-dimensional solution structure of the KH domain has revealed that it consists of a stable $\beta\alpha\alpha\beta\beta\alpha$ fold (Musco et al., 1996), where the KH-associated domain described here corresponds to the C-terminal β-sheet and α-helix, and the predicted protein structure of the KH domain of SF1 (not shown) is in agreement with this assignment. According to the nomenclature of Musco et al. (1996), the complete $\beta\alpha\alpha\beta\beta\alpha$ fold is now termed maxi-KH domain. Information on a possible function of the C-terminal $\beta\alpha$ structure of the maxi-KH domain is limiting, but it appears to be required for the stability of the domain (Musco et al., 1996). Moreover, amino acid insertions into one of the C-terminal α-helices of a KH domain in FBP abolish or highly reduce its specific DNA-binding activity, suggesting that this structure is essential for nucleic acid interactions (Duncan et al., 1994).

Another motif present in SF1 implicated in RNA binding is a zinc knuckle. Zinc knuckles are sequence motifs characteristic of retroviral nucleocapsid proteins that are derived from the viral gag polyprotein precursor (for review see Katz & Jentoft, 1989; Rein, 1994). Mutational analyses suggest that the zinc knuckles are required for normal packaging of viral RNA during retrovirus assembly; however, specific recognition of the viral RNA genome does not appear to be mediated by the nucleocapsid part of the gag polyprotein. Among the splicing factors, the SR protein 9G8 contains a zinc knuckle that separates its RRM from the serine/arginine-rich domain (Cavaloc et al., 1994). The function of this protein motif in 9G8 is unknown. In addition, the yeast Slu7p, which was isolated in a screen for factors that genetically interact with U5 snRNA, contains a zinc knuckle (Frank & Guthrie, 1992; Frank et al., 1992). Slu7p interacts specifically with the 3′ splice site in yeast (Umen & Guthrie, 1995a). Although mutations in the first two cysteine residues of the SLU7 zinc knuckle did not result in a lethal phenotype, 3′ splice site selection was affected (Frank & Guthrie, 1992). Whether this effect is caused by a change in the binding of Slu7p to the 3′ splice site is unknown.

Taken together, the maxi-KH domain and the zinc knuckle are domains important for RNA binding. Experiments are in progress to determine whether these domains in SF1 can bind RNA independently of one another and whether they cooperate in specifying sequence-dependent interactions with RNA.

A protein of unknown function that is encoded on chromosome XII of *S. cerevisiae* is 37% identical to SF1 and probably represents the yeast homologue of human SF1. The yeast protein (ScSF1) contains a maxi-KH domain, but, in contrast to human SF1, two closely spaced zinc knuckles instead of one. These motifs are arranged in ScSF1 in a similar order and spacing as in human SF1. The C-terminal half of the yeast protein is proline-rich, a feature also found in human SF1. This region is shorter in ScSF1 and accounts for the observed difference in predicted molecular mass of ~15 kDa between the yeast and human proteins. The conservation of the overall structure suggests that not only the domains implicated in RNA binding, but also the proline-rich C-terminal half of SF1, are important for function of the protein.

In addition to the SF1 cDNAs described here, four mammalian cDNAs that appear to encode additional SF1 isoforms have been isolated. ZFM1 was first cloned by Toda et al. (1994) from human brain as a transcribable sequence that maps to human chromosome 11q13

close to the locus for the gene responsible for MEN1. Based on newly isolated markers, the MEN1 gene has been assigned to a region of ~900 kb, and the cosmid clones used for the isolation of ZFM1 lie outside of this region (Teh et al., 1995). Thus, an involvement of ZFM1 in MEN1 can most likely be excluded. Owing to the presence of a zinc knuckle, Toda et al. (1994) suggested that ZFM1 represents a transcription factor. Our results are consistent with a function of SF1 in pre-mRNA splicing (this paper; Krämer, 1992), and no evidence that SF1 participates in transcription is available at present. The ZFM1-B form was cloned from human brain (Toda et al., 1994) and from mouse (EMBL accession no. X85802), where it was detected as a testosterone-induced transcript in spleen cells, and cDNAs for the human ZFM1-C and -D forms (EMBL accession no. L49345 and L49380) were isolated from an acute myeloid leukemia cell line.

Inspection of the sites of divergence between the various SF1 cDNAs and comparison of these sequences to putative exon–intron junctions in ZFM1-A (Toda et al., 1994) suggest that the mRNAs are derived from a common precursor by alternative splicing. Hybridization of northern blots with a probe from the common portion of the different SF1 isoforms revealed three mRNAs of predicted sizes. Size differences of 100 nt or less could not be resolved with this method, but at least seven products are generated by PCR with SF1-specific primers from HeLa cDNA (F. Mulhauser & A. Krämer, unpubl.), suggesting that more than three variants of SF1 exist at least in HeLa cells. The mRNAs detected by northern blot analysis are expressed differentially in a cell-type specific pattern, suggesting that the alternatively spliced forms of SF1 are of functional significance.

Alternatively spliced mRNAs have been described for several essential splicing factors (Ge et al., 1991; Kim et al., 1992; Patton et al., 1993; Popielarz et al., 1995; Screaton et al., 1995); extensive alternative splicing within the coding region similar to that seen for SF1, however, is unprecedented among splicing proteins. With the exception of ASF-3, a truncated form of ASF/SF2 that lacks the serine/arginine-rich domain essential for splicing activity (Cáceres & Krainer, 1993; Zuo & Manley, 1993) and acts as a dominant inhibitor of ASF/SF2 in vitro (Zuo & Manley, 1993), roles for putative isoforms of splicing proteins have not been reported. We have shown that three SF1 isoforms tested (SF1-Bo, SF1-HL1, SF1-HL2) support pre-spliceosome assembly in vitro. Could these proteins (and possibly other isoforms) play additional roles, for example, in alternative splicing? All predicted SF1 proteins contain the KH domain and the zinc knuckle. It can be expected that these regions function in RNA binding; hence, interactions with a target RNA should be similar for the different forms. The predicted proteins vary in the extent of their proline-rich regions and are characterized by C-termini of different lengths and unique sequence contents. The proline-rich regions may be required for interactions with other splicing factors and such contacts could be influenced by the length of this region. Moreover, the unique C-termini differ in overall charge from −1 (SF1-HL2) to +7 (ZFM1-C), a feature that could modulate protein–protein interactions or contacts with RNA. Differential RNA–protein and protein–protein interactions are important determinants in the choice of alternative splice sites (for review see Valcárcel et al., 1995), and future experiments will address the question whether different SF1 isoforms participate in this process.

The mouse QkI (Ebersole et al., 1996), *C. elegans* GLD-1 (Jones & Schedl, 1995), mammalian Sam68 (Wong et al., 1992; Lock et al., 1996), and *A. salina* Grp33 (Cruz-Alvarez & Pellicer, 1987) share significant sequence homology with SF1 in a ~160-amino acid region that includes the maxi-KH domain. Owing to the diverse functions reported for these proteins, a role for the homologous domain is elusive. Mouse QkI is the product of the *quaking* gene that is essential in early embryogenesis and for myelination in the nervous system (see Ebersole et al., 1996). GLD-1 is encoded by a tumor suppressor gene and carries out diverse functions during germ line differentiation (Jones & Schedl, 1995). Sam68 associates with Src in normal and transformed cells and becomes tyrosine-phosphorylated during mitosis (Fumagalli et al., 1994; Taylor & Shalloway, 1994). Finally, Grp33 represents a hnRNP protein (Cruz-Alvarez & Pellicer, 1987).

Sam68 interacts, directly or indirectly, with nucleic acids (Wong et al., 1992; Taylor & Shalloway, 1994) and it does so only when it is dephosphorylated (Wang et al., 1995). Whether QkI, GLD-1, or Grp33 bind to RNA is unknown; however, given the presence of a maxi-KH domain and the result that SF1 binds to RNA, this may be expected. All of these proteins contain proline-rich regions that could be required for protein–protein interactions. In particular, Sam68 has been shown to bind to the SH3 domain of Src and other oncogene products (Fumagalli et al., 1994; Taylor & Shalloway, 1994; Weng et al., 1994; Richard et al., 1995), and potential SH3 binding sites are also present in QkI (Ebersole et al., 1996). Hence, phosphorylation/dephosphorylation events could regulate the RNA-binding activity of these proteins. Another interesting observation concerns Qk1. In normal mice, the myelin-associated glycoprotein is expressed in two isoforms of 67 and 72 kDa, resulting from the alternative inclusion or exclusion of an exon that contains an in-frame stop codon. In the *quaking* mouse, splicing appears to switch to exon inclusion and the larger isoform is expressed at highly reduced levels (Fujita et al., 1988, 1990). In correlation with the function of SF1 as a splicing factor that binds to RNA, the alternative splice site switch in the *quaking* mouse could be caused by a mutation in the Qk1 protein that modulates its interaction with a

relevant splice site. Taken together, the interaction of Sam68 with Src and its phosphorylation, the requirement of QkI and GLD-1 during cellular differentiation, the RNA-binding activities of Sam68 and SF1, and the function of SF1 in pre-mRNA splicing may be an indication that these proteins play equivalent roles in RNA binding, thus affecting alternative splicing events.

MATERIALS AND METHODS

Protein purification and microsequencing

SF1 was purified from HeLa cell nuclear extracts as described (Krämer, 1992). About 100 pmol of SF1 obtained after Superose 12 gel filtration were precipitated with acetone. The pellet was dissolved in 50 μL 8 M urea/0.4 M ammonium bicarbonate and incubated for 60 min at 37 °C in the presence of 4.5 mM DTT. Iodoacetamide was added to a final concentration of 10 mM, the sample was incubated in the dark for 30 min at room temperature, and diluted to 200 μL with water, followed by digestion with 0.75 μg trypsin at 37 °C overnight. The resulting peptides were separated on a C4 reverse phase column (VYDAC 218TP52, 2.1 × 250 mm) and eluted with a 90-min gradient of 0–60% (v/v) acetonitrile in 0.05% (w/v) trifluoroacetic acid. Several peptides were further purified on Aquapore RP-300 (1 × 50 mm) or Spheri 5 RP-18 (1 × 50 mm) columns that were eluted with acetonitrile gradients in 0.09% trifluoroacetic acid. Amino acid sequencing was performed by P. Jenö and T. Mini (Protein Chemistry Laboratory, Biozentrum, University of Basel) as described (Krämer et al., 1994).

Eleven peptide sequences were obtained. Nine of these were found in the SF1 amino acid sequence. The sequences of peptides T10/12 (amino acids 20–30; see Fig. 2B) and T44 (amino acids 136–150) were used to design oligonucleotides for library screening. The sequence of one additional peptide (KLENATPNLDTEMV) is somewhat similar to amino acids 104–114 of SF1 and one peptide sequence has not been found in any reading frame of SF1 nor in proteins in current databases.

cDNA cloning

Degenerate oligonucleotides were designed according to Lathe (1985) and Sambrook et al. (1989) as follows (R = A or G, Y = C or T, H = A or C or T, N = any nucleotide):

T44-DI: 5′-ATCTCAGGGTACTCATCCTGTGGGATCATC ACCTTGT-3′ (representing amino acids DKVMIPQDEY PEI in the antisense orientation),
T44-MP1: 5′-ATGATHCCRCARGAYGAHTAYCC-3′ and
T44-MP2: 5′-ATGATHCCYCARGAYGARTAYCC-3′ (amino acids MIPQDEYP),
T10/12-DK: 5′-GAYACNATGGARCARAAR-3′ (amino acids DTMEQK).

Recombinant phage (7.5×10^5) of a HeLa cDNA library in λgt11 (kindly provided by G. McMaster) were screened in duplicate with ^{32}P-end-labeled oligonucleotide T44-DI. Hybridization was performed as described (Krämer et al., 1994). The final wash of the filters was performed in 2× SSC/1% SDS at room temperature. Eighteen double-positive phage were purified by two further rounds of screening with oligonucleotides T10/12-DK and a 1:1 mixture of T44-MP1 and T44-MP2 under the same conditions as above. The DNA of one phage that was positive with all probes was isolated and restriction digests were subjected to Southern blot analysis. The 1.61-kb *Eco*R I insert was subcloned into pBluescript II KS$^+$ (pBS; Stratagene) and designated SF1-HL1 (Fig. 1A).

A cDNA containing the complete coding sequence of SF1 was isolated with standard methods (Sambrook et al., 1989) from a human bone cDNA library in λgt11 (made from a femoral head obtained following hip replacement surgery) by screening with a 973-bp *Xho* I/*Eco*R I fragment of SF1-HL1 (Fig. 1A). DNA from six positive phage were prepared and analyzed by Southern blot hybridization. Two *Eco*R I fragments of 1.4 and 0.9 kb of the phage containing the largest insert were subcloned separately into pBS. A plasmid harboring the complete coding sequence of SF1 was constructed by a partial *Eco*R I restriction digest of the isolated λ DNA. The 2.3-kb *Eco*R I fragment was subcloned into pBS and designated SF1-Bo (Fig. 1A).

A third cDNA was isolated from a size-selected HeLa cDNA library in λgt11 by screening with the *Xho* I/*Eco*R I fragment as above. Positive phage were rescreened with both *Eco*R I fragments of SF1-Bo. DNA of one positive phage was isolated and two *Eco*R I fragments of 0.5 and 2.0 kb were subcloned separately into pBS. The composite cDNA sequence of both *Eco*R I fragments was designated SF1-HL2 (Fig. 1A).

The *Eco*R I fragments of the SF1-Bo, SF1-HL1, and SF1-HL2 clones in pBS were sequenced entirely on both strands from subcloned restriction fragments and exonuclease III fragments generated with the Erase-a-Base system (Promega). Sequencing was performed with the Sequenase II (USB) or T7 sequencing kits (Pharmacia LKB Biotechnology).

Northern blot analysis

Northern blot analysis of total RNA and polyA$^+$ RNA was performed according to van Hille et al. (1993). A northern blot containing polyA$^+$ RNA from human tissues was purchased from Clontech. The blots were hybridized with ^{32}P-labeled restriction fragments derived from the 5′ part of the SF1 cDNAs (*Xho* I/*Eco*R I; Fig. 1A) or from the insertion in SF1-HL2 (*Nae* I/*Nar* I; Fig. 1A). After hybridization, the blots were washed in 0.2× SSC/1% SDS at 60–65 °C.

In vitro translation

A plasmid for the in vitro translation of SF1-Bo was constructed by PCR amplification of SF1-Bo and insertion of the PCR product into pGEM4 (Promega). Plasmids for expression of SF1-HL1 and SF1-HL2 were constructed by replacement of pGEM/Bo sequences with corresponding restriction fragments of the SF1-HL1 and SF1-HL2 clones in pBS. Plasmid DNAs were linearized with appropriate restriction enzymes in the polylinker of the vector followed by phenol-chloroform extraction and ethanol precipitation. Coupled in vitro transcription/translation reactions (50 μL) were performed in the TNT T7 wheat germ extract system (Promega) according to the supplier's instructions. In addition, the reactions contained 0.2 μg/μL each of chymostatin, antipain,

leupeptin, and pepstatin, and 0.5 mM PMSF. Aliquots (5 μL) were analyzed by SDS-PAGE followed by autoradiography.

Expression of SF1 in insect cells

For expression in insect cells, SF1 sequences were cloned into the baculovirus transfer vector pVL1393 (Pharmingen). Translation initiates at the SF1 initiation codon. For the construction of pVL/Bo, the insert of the pGEM/Bo construct was cloned into the *Bam*H I sites of pVL1393. Vectors for expression of the SF1-HL1 and SF1-HL2 cDNAs were generated by PCR amplification of the corresponding pGEM constructs and cloning the PCR products into the *Not* I and *Bgl* II sites of pVL1393 to result in pVL/HL1 and pVL/HL2. Correct construction of the plasmids was verified by restriction digests and sequencing of relevant regions of all plasmids. The pVL constructs were cotransfected into *Spodoptera frugiperda* (Sf9) cells with linearized baculovirus DNA (BaculoGold DNA, Pharmingen) according to the supplier's instructions. Recombinant virus isolates were obtained by plaque assays and positive plaques were identified in dot blot assays. *Trichoplusia ni* (Tn5) cells in serum protein-free Excell401 medium (Readysysteme) were infected at a MOI of >1 and grown for 2 days at 27 °C.

Extract preparation and pre-splicing complex formation

Nuclear and cytoplasmic extracts were prepared from infected Tn5 cells by the small-scale procedure described by Krämer and Keller (1990). Control extracts were prepared from cells infected with DNA encoding a protein unrelated to splicing (a generous gift from L. Bernasconi). Pre-splicing complex assembly was analyzed as described (Krämer & Utans, 1991) in the presence of partially purified SF3a and SF3b (=SF3; 1 μL), U2AF (1 μL), U1, and U2 snRNPs (0.5 μL each), and the amount of SF1 or insect extracts indicated in the figure legends.

RNA-binding assays

The AdML pre-mRNA is derived from the adenovirus major late transcription unit. The region coding for the first two leader exons of AdML were recloned from RNA 1 (Frendewey & Keller, 1985) into pBluescript. Transcription by T3 RNA polymerase results in a RNA that contains a first exon of 111 nt, comprising 70 nt of vector sequences and the entire first leader exon of the AdML transcription unit, an intron of 113 nt, and the second leader exon that is truncated at a *Sca* I site 38 nt downstream of the 3′ splice site. U1 snRNA was transcribed with SP6 RNA polymerase from plasmid pHU1A (a gift from R. Lührmann). A vector containing the gene for human U2 snRNA was a gift from M. Ares. The U2 snRNA coding region was amplified by PCR. The 5′ oligonucleotide contained a *Hind* III site and a T7 RNA polymerase promoter, the 3′ oligonucleotide contained a *Hpa* II site. The PCR fragment was ligated into the *Hind* III site of the pSP64 polyA vector (Promega). For transcription by T7 RNA polymerase, the plasmid was linearized with *Hpa* II. The resulting U2 snRNA contains two additional guanosine residues at the 5′ end. The control transcript was synthesized with T3 RNA polymerase from pBluescript linearized with *Pvu* II.

Labeled RNAs were incubated with purified SF1 (~150 ng; Superose fraction; Krämer, 1992) or nuclear extracts from insect cells infected with virus encoding SF1-Bo for 30 min at 30 °C under conditions appropriate for pre-splicing complex formation (Krämer & Utans, 1991). Deviations from the standard reaction conditions are indicated in the legend to Figure 8. After incubation, reactions were irradiated with 254-nm UV light (900 μWatts/cm^2) for 10 min on ice, followed by digestion with RNAse A (0.2 mg/mL) for 15 min at 37 °C. Proteins were precipitated with 13% trichloroacetic acid, washed twice with acetone, and separated in 8 or 12% SDS polyacrylamide gels.

Computer analysis

Structural analyses and sequence alignments were performed with programs of the GCG software package of the University of Wisconsin (Devereux et al., 1984). Nucleotide and protein sequences were compared with entries in the GenEmbl (release 46 plus updates) and SWISS-PROT (release 32) databases.

ACKNOWLEDGMENTS

We are indebted to Lilia Bernasconi and Bernard Allet (Glaxo Institute for Molecular Biology, Geneva) for introducing us to the baculovirus expression system and for help at the initial stages of protein expression in insect cells. We thank Gary McMaster (Ciba-Geigy Ltd., Basel) for a HeLa cDNA library and HeLa cDNA, Jim Gallagher (University of Liverpool) for a bone sample, Manuel Ares (University of California, Santa Cruz) for the human U2 snRNA gene, Karsten Gröning (Universität Marburg) for cloning the U2 sequence, and Reinhard Lührmann (Universität Marburg) for a vector containing the U1 snRNA. The technical assistance of Mireille Quentin at early stages of the project is gratefully acknowledged. This work was supported by grants from the Schweizerischer Nationalfonds and the Canton of Geneva.

Manuscript accepted without revision June 6, 1996

REFERENCES

Ares M, Weiser B. 1995. Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog Nucleic Acid Res Mol Biol 50*:131–159.

Beggs JD. 1995. Yeast splicing factors and genetic strategies for their analysis. In: Lamond AI, ed. *Pre-mRNA processing*. Austin, Texas: R.G. Landes Company. pp 79–95.

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

Bennett M, Reed R. 1993. Correspondence between a mammalian spliceosome component and an essential yeast splicing factor. *Science 262*:105–108.

Berg JM. 1986. Potential metal-binding domains in nucleic acid binding proteins. *Science 232*:485–487.

Brosi R, Gröning K, Behrens SE, Lührmann R, Krämer A. 1993a. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science 262*:102–105.

Brosi R, Hauri HP, Krämer A. 1993b. Separation of splicing factor SF3

into two components and purification of SF3a activity. *J Biol Chem 268*:17640–17646.
Burd CG, Dreyfuss G. 1994. RNA binding specificity of hnRNP A1: Significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J 13*:1197–1204.
Cáceres JF, Krainer AR. 1993. Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J 12*:4715–4726.
Castiglione Morelli MA, Stier G, Gibson T, Joseph C, Musco G, Pastore A, Travè G. 1995. The KH module has an $\alpha\beta$ fold. *FEBS Lett 358*:193–198.
Cavaloc Y, Popielarz M, Fuchs JP, Gattoni R, Stévenin J. 1994. Characterization and cloning of the human splicing factor 9G8: A novel 35 kDa factor of the serine/arginine protein family. *EMBO J 13*: 2639–2649.
Champion-Arnaud P, Reed R. 1994. The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes & Dev 8*:1974–1983.
Chiara MD, Champion-Arnaud P, Buvoli M, Nadal-Ginard B, Reed R. 1994. Specific protein–protein interactions between the essential mammalian spliceosome-associated proteins SAP 61 and SAP 114. *Proc Natl Acad Sci USA 91*:6403–6407.
Cruz-Alvarez M, Pellicer A. 1987. Cloning of a full-length complementary DNA for an *Artemia salina* glycine-rich protein. *J Biol Chem 262*:13377–13380.
Devereux J, Haeberli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res 12*: 387–395.
Dingwall C, Laskey R. 1991. Nuclear targeting sequences—A consensus? *Trends Biochem Sci 16*:478–481.
Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG. 1993. HnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem 62*:289–321.
Duncan R, Bazar L, Michelotti G, Tomonaga T, Krutzsch H, Avigan M, Levens D. 1994. A sequence-specific, single-strand binding protein activates the far upstream element of *c-myc* and defines a new DNA-binding motif. *Genes & Dev 8*:465–480.
Ebersole T, Chen Q, Justice M, Artzt K. 1996. The *quaking* gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nature Genetics 12*:260–265.
Engebrecht J, Roeder GS. 1990. MER1, a yeast gene required for chromosome pairing and genetic recombination, is induced in meiosis. *Mol Cell Biol 10*:2379–2389.
Frank D, Guthrie C. 1992. An essential splicing factor, SLU7, mediates 3′ splice site choice in yeast. *Genes & Dev 6*:2112–2124.
Frank D, Patterson B, Guthrie C. 1992. Synthetic lethal mutations suggest interactions between U5 small nuclear RNA and four proteins required for the second step of splicing. *Mol Cell Biol 12*:5197–5205.
Frendewey D, Keller W. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell 42*:355–367.
Fu XD. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA 1*:663–680.
Fujita N, Sato S, Ishiguro H, Inuzuka T, Baba H, Kurihara T, Takahashi Y, Miyatake T. 1990. The large isoform of myelin-associated glycoprotein is scarcely expressed in the quaking mouse brain. *J Neurochem 55*:1056–1059.
Fujita N, Sato S, Kurihara T, Inuzuka T, Takahashi Y, Miyatake T. 1988. Developmentally regulated alternative splicing of brain myelin-associated glycoprotein is lacking in the quaking mouse brain. *FEBS Lett 232*:323–327.
Fumagalli S, Totty N, Hsuan J, Courtneidge S. 1994. A target for Src in mitosis. *Nature 368*:871–874.
García-Blanco MA, Jamison S, Sharp PA. 1989. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes & Dev 3*:1874–1886.
Ge H, Zuo P, Manley JL. 1991. Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell 66*:373–382.
Gibson TJ, Thompson JD, Heringa J. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acids. *FEBS Lett 324*:361–366.
Gozani O, Feld R, Reed R. 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes & Dev 10*:233–243.
Gozani O, Patton JG, Reed R. 1994. A novel set of spliceosome-associated proteins (SAPs) and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. *EMBO J 13*:3356–3367.
Hall SL, Padgett RA. 1994. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J Mol Biol 239*:357–365.
Heinrichs V, Baker B. 1995. The *Drosophila* SR protein RBP1 contributes to the regulation of *doublesex* alternative splicing by recognizing RBP1 RNA target sequences. *EMBO J 14*:3987–4000.
Jackson IJ. 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res 19*:3795–3798.
Jamison SF, Crow A, García-Blanco MA. 1992. The spliceosome assembly pathway in mammalian extracts. *Mol Cell Biol 12*:4279–4287.
Jones AR, Schedl T. 1995. Mutations in *gld-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes & Dev 9*:1491–1504.
Katz RA, Jentoft JE. 1989. What is the role of the Cys-His motif in retroviral nucleocapsid (NC) proteins? *BioEssays 11*:176–181.
Kim YJ, Zuo P, Manley JL, Baker BS. 1992. The *Drosophila* RNA-binding protein RBP1 is localized to transcriptionally active sites of chromosomes and shows a functional similarity to human splicing factor ASF/SF2. *Genes & Dev 6*:2569–2579.
Kozak M. 1989. The scanning model for translation: An update. *J Cell Biol 108*:229–241.
Krainer AR, Conway GC, Kozak D. 1990. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes & Dev 4*:1158–1171.
Krämer A. 1992. Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a pre-splicing complex. *Mol Cell Biol 12*:4545–4552.
Krämer A. 1995. The biochemistry of pre-mRNA splicing. In: Lamond AI, ed. *Pre-mRNA processing*. Austin, Texas: R.G. Landes Company. pp 35–64.
Krämer A, Keller W. 1990. Preparation and fractionation of mammalian extracts active in pre-mRNA splicing. *Methods Enzymol 181*: 3–19.
Krämer A, Legrain P, Mulhauser F, Gröning K, Brosi R, Bilbe G. 1994. Splicing factor SF3a60 is the mammalian homologue of PRP9 of *S. cerevisiae*: The conserved zinc finger-like motif is functionally exchangeable in vivo. *Nucleic Acids Res 22*:5223–5228.
Krämer A, Mulhauser F, Wersig C, Gröning K, Bilbe G. 1995. Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is homologous to the essential splicing factor PRP21p of *S. cerevisiae*. *RNA 1*:260–272.
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.
Lathe R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data. Theoretical and practical considerations. *J Mol Biol 183*:1–12.
Lock P, Fumagalli S, Polakis P, McCormick F, Courtneidge S. 1996. The human p62 cDNA encodes Sam68 and not the RasGAP-associated p62 protein. *Cell 84*:23–24.
MacMillan AM, Query CC, Allerson CR, Chen S, Verdine GL, Sharp PA. 1994. Dynamic association of proteins with the pre-mRNA branch region. *Genes & Dev 8*:3008–3020.
Madhani HD, Guthrie C. 1994. Dynamic RNA–RNA interactions in the spliceosome. *Annu Rev Genet 28*:1–26.
Matunis MJ, Michael WM, Dreyfuss G. 1992. Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol Cell Biol 12*:164–171.
Michaud S, Reed R. 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes & Dev 5*:2534–2546.
Moore MJ, Query CC, Sharp PA. 1993. Splicing of precursors to mRNA by the spliceosome. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 303–357.
Musco G, Stier G, Joseph C, Castiglione Morelli M, Nilges M, Gibson T, Pastore A. 1996. Three-dimensional structure and stabil-

ity of the KH domain: Molecular insights into the Fragile X syndrome. *Cell 85*:237–245.
Nandabalan K, Price L, Roeder GS. 1993. Mutations in U1 snRNA bypass the requirement for a cell type-specific RNA splicing factor. *Cell 73*:407–415.
Nandabalan K, Roeder GS. 1995. Binding of a cell-type-specific RNA splicing factor to its target regulatory sequence. *Mol Cell Biol 15*:1953–1960.
Patton JG, Porro EB, Galceran J, Tempst P, Nadal-Ginard B. 1993. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes & Dev 7*:393–406.
Popielarz M, Cavaloc Y, Mattei MG, Gattoni R, Stévenin J. 1995. The gene encoding human splicing factor 9G8—Structure, chromosomal localization, and expression of alternatively processed transcripts. *J Biol Chem 270*:17830–17835.
Rein A. 1994. Retroviral RNA packaging: A review. *Arch Virol Suppl 9*:513–522.
Richard S, Yu D, Blumer KJ, Hausladen D, Olszowy MW, Connelly PA, Shaw AS. 1995. Association of p62, a multifunctional SH2- and SH3-domain-binding protein, with src family tyrosine kinases, Grb2, and phospholipase Cy-1. *Mol Cell Biol 15*:186–197.
Roscigno RF, García-Blanco MA. 1995. SR proteins escort the U4/U6·U5 tri-snRNP to the spliceosome. *RNA 1*:692–706.
Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
Saraste M, Sibbald P, Wittinghofer A. 1990. The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci 15*:430–434.
Schmidt D, Henkel B, Pöschl E, Zorbas H, Purschke WG, Gloe TR, Müller PK. 1992. Complete cDNA sequence of chicken vigilin, a novel protein with amplified and evolutionary conserved domains. *Eur J Biochem 206*:625–634.
Screaton G, Cáceres J, Mayeda A, Bell M, Plebanski M, Jackson D, Bell J, Krainer A. 1995. Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. *EMBO J 14*:4336–4349.
Siebel CW, Admon A, Rio DC. 1995. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes & Dev 9*:269–283.
Siebel CW, Kanaar R, Rio DC. 1994. Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes & Dev 8*:1713–1725.
Singh R, Valcárcel J, Green MR. 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science 268*:1173–1176.
Siomi H, Choi M, Siomi MC, Nussbaum RL, Dreyfuss G. 1994. Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell 77*:33–39.
Siomi H, Matunis MJ, Michael WM, Dreyfuss G. 1993a. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res 21*:1193–1198.
Siomi H, Siomi MC, Nussbaum RL, Dreyfuss G. 1993b. The protein product of the Fragile-X gene, FMR1, has characteristics of an RNA-binding protein. *Cell 74*:291–298.
Staknis D, Reed R. 1994. Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol Cell Biol 14*:2994–3005.
Stephens RM, Schneider TD. 1992. Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J Mol Biol 228*:1124–1136.
Tacke R, Manley J. 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J 14*:3540–3551.
Taylor S, Shalloway D. 1994. An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature 368*:367–371.
Teh BT, Cardinal J, Shepherd J, Hayward NK, Weber G, Cameron D, Larsson C. 1995. Genetic mapping of the multiple endocrine neoplasia type 1 locus at 11q13. *J Intern Med 238*:249–253.
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.
Umen JG, Guthrie C. 1995a. A novel role for a U5 snRNP protein in 3′ splice site selection. *Genes & Dev 9*:855–868.
Umen JG, Guthrie C. 1995b. The second catalytic step of pre-mRNA splicing. *RNA 1*:869–885.
Utans U, Behrens SE, Lührmann R, Kole R, Krämer A. 1992. A splicing factor that is inactivated during in vivo heat shock is functionally equivalent to the [U4/U6·U5] triple snRNP-specific proteins. *Genes & Dev 6*:631–641.
Utans U, Krämer A. 1990. Splicing factor SF4 is dispensable for the assembly of a functional splicing complex and participates in the subsequent steps of the splicing reaction. *EMBO J 9*:4119–4126.
Valcárcel J, Singh R, Green MR. 1995. Mechanisms of regulated pre-mRNA splicing. In: Lamond AI, ed. *Pre-mRNA processing*. Austin, Texas: R.G. Landes Company. pp 97–112.
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.
Wang LL, Richard S, Shaw AS. 1995. p62 association with RNA is regulated by tyrosine phosphorylation. *J Biol Chem 270*:2010–2013.
Weng Z, Thomas SM, Rickles RJ, Taylor JA, Brauer AW, Seidel-Dugan C, Michael WM, Dreyfuss G, Brugge JS. 1994. Identification of Src, Fyn, and Lyn SH3-binding proteins: Implications for a function of SH3 domains. *Mol Cell Biol 14*:4509–4521.
Wong G, Müller O, Clark R, Conroy L, Moran MF, Polakis P, McCormick F. 1992. Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell 69*:551–558.
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.
Zuo P, Manley JL. 1993. Functional domains of the human splicing factor ASF/SF2. *EMBO J 12*:4727–4737.