RNA (1998), *4*:551–565. Cambridge University Press. Printed in the USA.

Conservation of functional domains involved in RNA binding and protein–protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1

JEAN-CHRISTOPHE RAIN,[1]* ZAHRA RAFI,[2]* ZAKARIA RHANI,[2]* PIERRE LEGRAIN,[1] and ANGELA KRÄMER[2]

[1]Laboratoire du Métabolisme des ARN, URA CNRS 1300, Département des Biotechnologies, Institut Pasteur, 28 rue du Dr Roux, 75724, Paris Cedex 15, France
[2]Département de Biologie Cellulaire, Université de Genève, 30 quai Ernest-Ansermet, 1211 Genève 4, Switzerland

ABSTRACT

The modular structure of splicing factor SF1 is conserved from yeast to man and SF1 acts at early stages of spliceosome assembly in both organisms. The hnRNP K homology (KH) domain of human (h) SF1 is the major determinant for RNA binding and is essential for the activity of hSF1 in spliceosome assembly, supporting the view that binding of SF1 to RNA is essential for its function. Sequences N-terminal to the KH domain mediate the interaction between hSF1 and U2AF^{65}, which binds to the polypyrimidine tract upstream of the 3′ splice site. Moreover, yeast (y) SF1 interacts with Mud2p, the presumptive U2AF^{65} homologue in yeast, and the interaction domain is conserved in ySF1. The C-terminal degenerate RRMs in U2AF^{65} and Mud2p mediate the association with hSF1 and ySF1, respectively. Analysis of chimeric constructs of hSF1 and ySF indicates that the KH domain may serve a similar function in both systems, whereas sequences C-terminal to the KH domain are not exchangeable. Thus, these results argue for hSF1 and ySF1, as well as U2AF^{65} and Mud2p, being functional homologues.

Keywords: commitment complex; KH domain; Mud2p; spliceosome assembly; two-hybrid system; U2AF^{65}

INTRODUCTION

The removal of introns from nuclear pre-mRNA is catalyzed by a multicomponent complex, the spliceosome, which consists of small nuclear ribonucleoprotein particles (snRNPs) and numerous proteins (reviewed in Moore et al., 1993; Madhani & Guthrie, 1994; Krämer, 1996). These components interact with the pre-mRNA and with one another in a highly dynamic fashion. The assembly of the active spliceosome progresses through several well-defined intermediate complexes and finally results in intron excision by two transesterification reactions.

Spliceosome assembly and catalysis are well conserved from yeast to mammals and in both systems the earliest stages of the reaction are important for the selection of the splice sites and the commitment of the pre-mRNA to the splicing pathway. The initial stage in spliceosome assembly has first been defined in the yeast system, where two commitment complexes (CC) can be distinguished (Legrain et al., 1988; Ruby & Abelson, 1988; Séraphin & Rosbash, 1989, 1991). The formation of CC1 requires the 5′ splice site, whereas the conversion to CC2 requires, in addition, the presence of the branch site sequence UACUAAC. This sequence is highly conserved in yeast and usually is located 15–50 nt upstream of the 3′ splice site (Rymond & Rosbash, 1992; J.-C. Rain, unpubl. obs.). Sequences downstream of the branch site appear to be dispensable for this step (Rymond & Rosbash, 1985). Commitment of the pre-mRNA to the splicing pathway has also been analyzed in vivo in an indirect assay for pre-mRNA export from the nucleus (Legrain & Rosbash, 1989), where export reflects the failure of intron recognition by the splicing machinery. With this assay, it has been demonstrated that a U residue at the position preceding the UACUAAC sequence is preferred over other nucleotides for intron recognition (Rain & Legrain, 1997). This U residue is conserved in natural yeast introns.

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: angela.kraemer@cellbio.unige.ch.

*The first three authors contributed equally to the work.

In the mammalian system, a comparable complex, pre-splicing complex E, has been identified (Michaud & Reed, 1991, 1993; Jamison et al., 1992). Although the 5′ splice site is of less importance than in the yeast system, formation of complex E is more efficient in the presence of both 5′ and 3′ splice sites. In addition, mammalian introns usually contain a polypyrimidine tract upstream of the 3′ splice site that is necessary for the formation of complex E. The mammalian branch site sequence YNCURAY is less well conserved than that in yeast introns (Keller & Noon, 1984; Green, 1986) and is apparently required only after the assembly of complex E (Champion-Arnaud et al., 1995).

During the formation of the mammalian complex E, U1 snRNP associates with the 5′ splice site and the large subunit of splicing factor U2AF (U2AF^{65}) binds to the polypyrimidine tract (Michaud & Reed, 1991, 1993; Bennett et al., 1992; Zamore et al., 1992). U2AF^{65} also contacts the branch site directly (Gaur et al., 1995; Valcárcel et al., 1996). Members of the SR family of splicing proteins (Fu, 1995) facilitate these interactions by binding to the pre-mRNA and engaging in protein–protein interactions between the U1 snRNP-specific 70K protein and U2AF^{35} (Wu & Maniatis, 1993; Kohtz et al., 1994; Staknis & Reed, 1994). In yeast, U1 snRNP binds to the 5′ splice site during the formation of CC1 (Séraphin & Rosbash, 1989). CC2 contains, in addition to U1 snRNP, Mud2p, a protein identified in a screen for synthetic lethality with a mutation in U1 snRNA (Abovich et al., 1994). Mud2p is not tightly associated with U1 snRNP and contacts the pre-mRNA directly in the presence of a functional branch site sequence. In addition, Mud2p is directly or indirectly involved in the recognition of the U residue immediately upstream of the branch site (Rain & Legrain, 1997).

Amino acid sequence comparisons suggested that Mud2p is distantly related to U2AF^{65}, which contains an N-terminal arginine–serine-rich (RS) domain and three RNA recognition motifs (RRMs, see Fig. 1; Zamore et al., 1992; Abovich et al., 1994). The first two RRMs are consensus RRMs, whereas the third one is degenerate (Birney et al., 1993). These features are conserved in U2AF^{65} homologues in *Drosophila*, *Caenorhabditis elegans*, and *Schizosaccharomyces pombe* (Kanaar et al., 1993; Potashkin et al., 1993; Zorio et al., 1997). Mud2p lacks an RS domain and shows highest homology to U2AF^{65} in the region of the degenerate RRM3 (31% identity and 49% similarity, see Fig. 1; Abovich et al., 1994).

We have previously purified human splicing factor SF1 (hSF1) based on its activity in the formation of pre-splicing complex A (Krämer, 1992). This complex forms by binding of U2 snRNP and associated proteins to complex E and involves base pairing of U2 snRNA with the branch site (reviewed in Hodges & Beggs, 1994; Reed, 1996). Inspection of the amino acid sequence of hSF1 (Arning et al., 1996) revealed

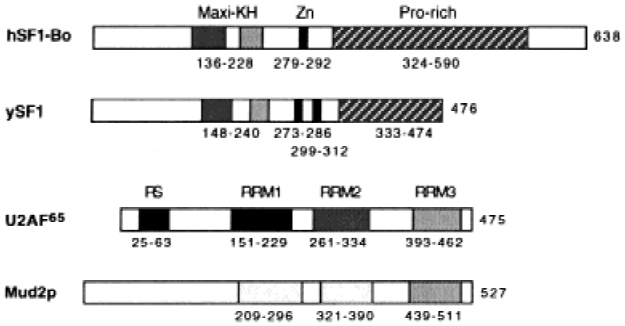

FIGURE 1. Schematic representation of hSF1, ySF1, U2AF^{65}, and Mud2p proteins. Key structural conserved features are indicated by shading and the amino acid positions are given in numbers below the diagrams. Structural domains are Maxi-KH, maxi-KH domain (the two shaded boxes represent the N- and C-terminal conserved portions); RRM, RNA recognition motif; RS, arginine–serine-rich domain; Pro-rich, proline-rich region; Zn, zinc knuckle. RRM1 and RRM2 of U2AF^{65} are not conserved in Mud2p and thus are shaded differently.

the presence of a maxi-KH domain, a motif found in a variety of proteins associated with RNA (Fig. 1; Siomi et al., 1993a; Musco et al., 1996), and a zinc knuckle (of the consensus sequence Cys-X_2-Cys-X_4-His-X_4-Cys), which is a characteristic motif in retroviral nucleocapsid proteins (Darlix et al., 1995). Consistent with the presence of these motifs, hSF1 was shown to bind RNA in a sequence-independent fashion, although a preference for binding to guanosine or uridine-rich RNA was observed (Arning et al., 1996). The C-terminal half of SF1 is rich in proline residues and recently has been shown to be the target for binding of formin-binding proteins and the oncogene product *abl* (Bedford et al., 1997). Different isoforms of hSF1 that are derived from alternatively spliced mRNAs are expressed in various mammalian cell types (Toda et al., 1994; Arning et al., 1996; Caslini et al., 1997; Wrehlke et al., 1997). These isoforms differ in the length of the proline-rich region and distinct C termini, the function of which is unknown. In the putative yeast homologue of hSF1 (ySF1) that was found in database searches (Fig. 1; 37% overall identity and 55% similarity; Arning et al., 1996), the maxi-KH domain is conserved and the yeast protein contains two zinc knuckles instead of one in the human protein. ySF1 also contains a shorter proline-rich C terminus, and the amino acid sequences of hSF1 and ySF1 immediately N-terminal to the maxi-KH domain are 47% identical over a stretch of 113 amino acids.

Recently, this putative homologue was independently found to interact with Mud2p in a synthetic lethal screen, and termed BBP for “branch point bridging protein” (Abovich & Rosbash, 1997), and in an exhaustive yeast two-hybrid screen with Mud2p (Fromont-Racine et al., 1997). In agreement with these findings, a direct physical interaction of ySF1 with Mud2p has been demon-

strated by biochemical means (Abovich & Rosbash, 1997). Moreover, ySF1 functions in the formation, and is part of the yeast CC2 complex and binds specifically to the yeast branch site (Abovich & Rosbash, 1997; Berglund et al., 1997). Similarly, hSF1 contacts U2AF^{65}, it also binds to the branch site, albeit with a lower affinity than its yeast counterpart, and has been detected in human complex E (Abovich & Rosbash, 1997; Berglund et al., 1997).

Given the modular structure of SF1, the role of ySF1/hSF1 in RNA binding, activity in spliceosome assembly, and interaction with Mud2p/U2AF^{65}, we have performed a mutational analysis of human and *Saccharomyces cerevisiae* SF1 to test whether functional domains are conserved. We have also analyzed the regions in U2AF^{65} and Mud2p that are responsible for direct contacts with hSF1 and ySF1, respectively. Based on amino acid sequence similarities and functional characteristics, these results confirm that ySF1 and hSF1, as well as Mud2p and U2AF^{65}, are functional homologues.

RESULTS

The maxi-KH domain is required for binding of hSF1 to RNA

To dissect functional domains in hSF1, N- or C-terminally truncated proteins and proteins carrying internal deletions or point mutations were expressed in *Escherichia coli* (Fig. 2A). Recombinant proteins containing N-terminal His_6-tags were purified by metal-chelate chromatography and incubated with a synthetic, uniformly labeled AdML pre-mRNA. The RNA-binding activity of mutant proteins was tested in a UV cross-linking assay. hSF1 lacking part or all of the proline-rich C-terminal sequences (C5 and C4) efficiently bound to RNA (Fig. 2B). In addition, C-terminal or internal deletions of the zinc knuckle (C3 and ΔZn) or a point mutation in the conserved histidine residue (data not shown) were fully active in this assay. In contrast, a C-terminal deletion of the maxi-KH domain (C2) abolished RNA-binding activity. N-terminal deletions of up to 93 amino acids (N1a and N1) were without effect on RNA binding, but removal of the N-terminal 134 amino acids (N2) resulted in a severe reduction in RNA binding. Further deletion of the maxi-KH domain (N4) abrogated the binding of hSF1 to RNA.

Analysis of the solution structure of KH domains of vigilin and FMR1 demonstrated that the domain consists of a stable $\beta\alpha\alpha\beta\beta\alpha$ fold in which the three α-helices are packed along the three-stranded antiparallel β-sheet (Musco et al., 1996, 1997). The two N-terminal α-helices are connected by a flexible loop of five amino acids that contains the highly conserved sequence GXXG. It has been suggested that this loop represents the RNA-binding surface of the domain with possible contributions of sequences in the neighboring α-helices. A second loop of variable length and sequence in different KH domains separates β-sheets 2 and 3. The C-terminal α-helix folds back onto the three-stranded antiparallel β-sheet and is believed to contribute to the stability of the KH fold by shielding the hydrophobic core of the domain. To test whether the regions of the KH domain that are separated by the variable loop are required for RNA binding, they were individually deleted from hSF1. The RNA-binding activity of the mutant proteins (ΔKH-N and ΔKH-C) was highly reduced, indicating that both conserved regions of the maxi-KH domain contribute to the binding to RNA. A function of the KH domain in RNA binding was further analyzed with point mutations in two highly conserved residues (Musco et al., 1996). Changing a residue located in the first flexible loop (G161D) reduced RNA binding to almost background levels (Fig. 2B), whereas the second point mutation (L164N) located in α-helix 2 only had a marginal effect. Taken together, these results indicate that the maxi-KH domain is essential for the RNA-binding function of hSF1.

Sequences of about 40 amino acids N-terminal to the KH domain appear to contribute to RNA binding. However, we cannot rule out the possibility that the KH domain in hSF1-N2 is not folded properly, because vector-derived sequences immediately border the KH domain. A role for the zinc knuckle in the interaction with RNA could not be established. In summary, these results locate the RNA-binding domain of hSF1 to amino acids 94–278.

The maxi-KH domain and the N terminus of hSF1 are essential for pre-splicing complex assembly

We next asked which portions of hSF1 are essential for the formation of pre-splicing complex A. In these experiments, an SF1-depleted DS500 fraction obtained from HeLa cell nuclear extracts (see Materials and Methods) was incubated in the presence of mutant SF1 proteins and spliceosome assembly was analyzed by nondenaturing PAGE. The depleted DS500 fraction shows residual activity in complex formation. For technical reasons, it was not possible to completely deplete hSF1 without compromising complex assembly in the reconstitution experiments. hSF1 with partial or complete deletions of the proline-rich region (C5 and C4) supported complex formation in a way similar to SF1 purified from HeLa cells, whereas an additional deletion of the zinc knuckle (C3) showed slightly reduced activity (Fig. 3A). Scanning of the autoradiographs and comparison between the reaction performed in the presence of HeLa cell SF1 and those with mutant proteins indicated a relative activity of 90–95% for C4 and C5, whereas the activity of C3 was about 60% of the control. Further deletion of the maxi-KH domain (C2) reduced complex assembly to background levels. In

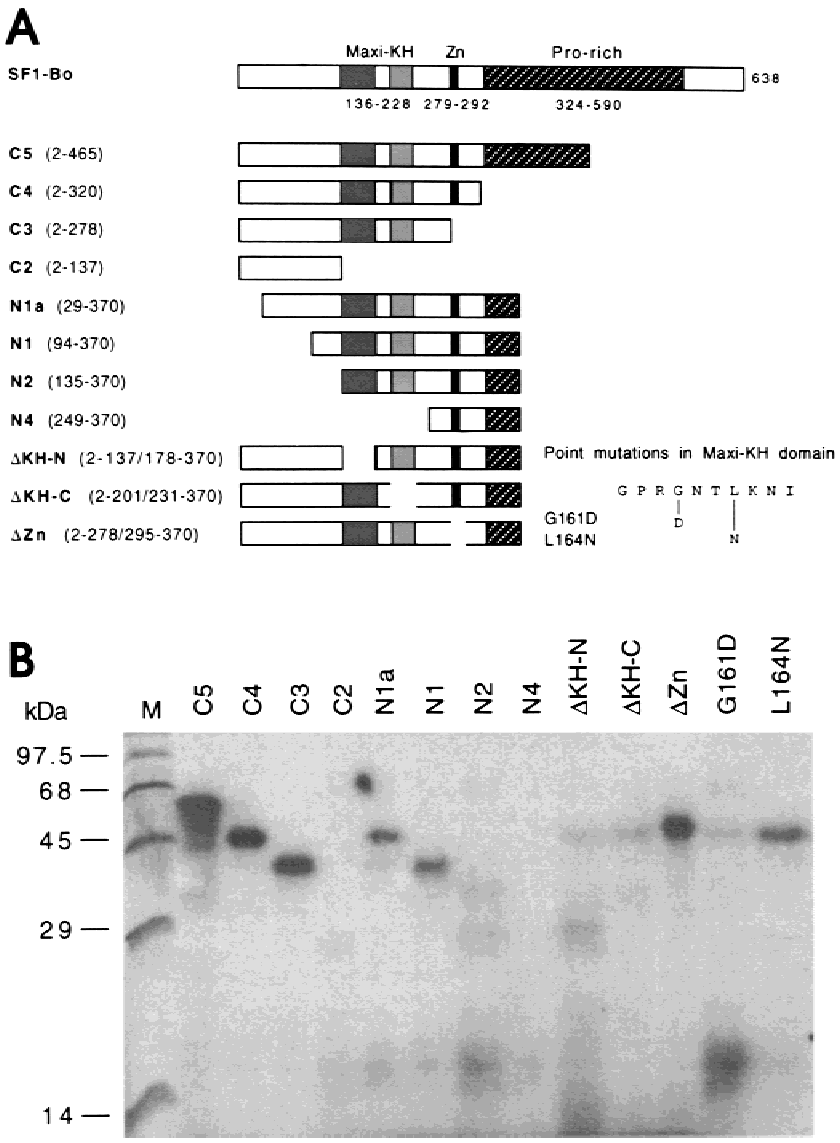

FIGURE 2. Maxi-KH domain of hSF1 is essential for RNA binding. **A:** Full-length hSF1 and mutant derivatives are represented schematically as in Figure 1. Mutant hSF1 proteins carry an N-terminal His_6-tag and contain the amino acids indicated. **B:** Recombinant His_6-tagged hSF1 proteins (0.2 μg each) were crosslinked to an in vitro-synthesized [^{32}P] UTP-labeled RNA and separated in a 15% SDS polyacrylamide gel. The gel was dried and exposed to X-ray film.

addition, pre-splicing complex formation with hSF1 proteins carrying internal deletions of the maxi-KH domain was 10–20% of the control (ΔKH-N and ΔKH-C; Fig. 3B). The point mutation G161D in the KH domain is highly reduced in complex formation (10%) and the mutation L164N shows decreased activity (50%). Consistent with the activity of hSF1-C3, which lacks the zinc knuckle in addition to the proline-rich region, an internal deletion of this region (ΔZn) or a point mutation of the conserved histidine (data not shown) are without effect on the activity of hSF1 in spliceosome assembly. These results are in agreement with the observations made when the same mutant proteins were tested for RNA-binding activity (see above). We conclude that the maxi-KH domain is essential for both RNA binding and activity in pre-splicing complex assembly. Therefore, the RNA-binding activity of hSF1 is an important feature for its function in spliceosome formation. Neither the zinc knuckle nor the proline-rich region of hSF1 appear to play a role during pre-spliceosome assembly. Furthermore, the divergent C termini of different hSF1 isoforms are dispensable for spliceosome assembly, which is consistent with the observation that different SF1 isoforms expressed in insect cells showed similar activity in complex A formation (Arning et al., 1996).

The activity of N-terminal deletion mutants hSF1-N1a and N1 was low and only apparent at the highest protein concentration tested (15–35% of the control; Fig. 3C). hSF1-N2 and N4 were inactive in complex

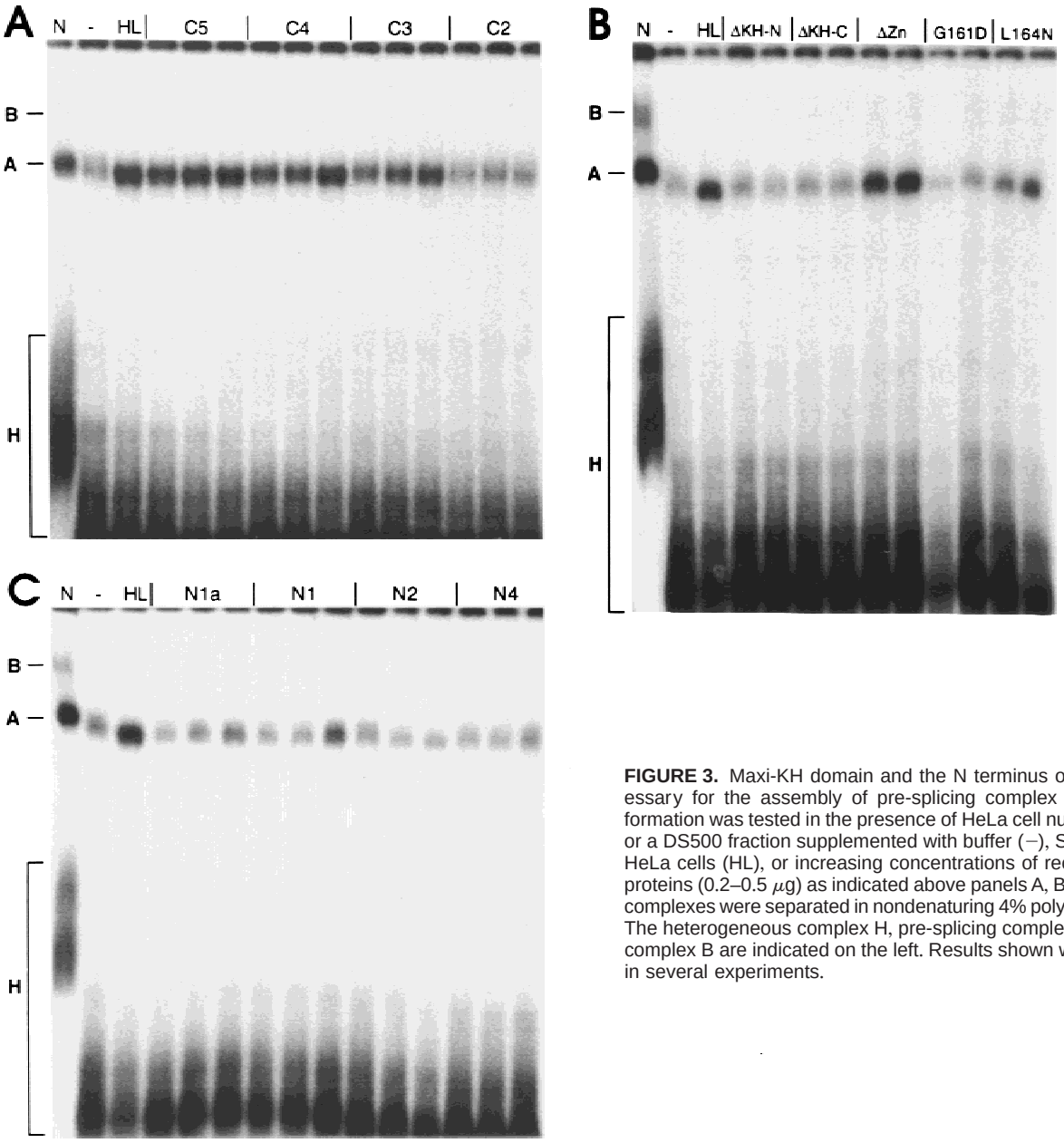

FIGURE 3. Maxi-KH domain and the N terminus of hSF1 are necessary for the assembly of pre-splicing complex A. Spliceosome formation was tested in the presence of HeLa cell nuclear extract (N) or a DS500 fraction supplemented with buffer (−), SF1 purified from HeLa cells (HL), or increasing concentrations of recombinant hSF1 proteins (0.2–0.5 μg) as indicated above panels A, B, and C. Splicing complexes were separated in nondenaturing 4% polyacrylamide gels. The heterogeneous complex H, pre-splicing complex A, and splicing complex B are indicated on the left. Results shown were reproduced in several experiments.

formation. Thus, although N-terminal sequences up to amino acid 93 are dispensable for binding to RNA, this region of hSF1 is essential for spliceosome assembly, suggesting an additional function of the N terminus of hSF1.

The N termini of hSF1 and ySF1 are involved in protein–protein interactions with U2AF^{65} and Mud2p, respectively

The yeast orthologue of hSF1 was identified previously in an exhaustive genomic two-hybrid screen performed with yeast Mud2p as bait (Fromont-Racine et al., 1997). Several ySF1 fragments were obtained as in-frame and out-of-frame fusions to the GAL4 activation domain, suggesting that the synthesis of low amounts of a portion of ySF1 is sufficient for an interaction with Mud2p and activation of reporter genes (Fig. 4). It has been shown previously that out-of-frame fusions can indeed be selected in two-hybrid screens via frameshifting events (Fromont-Racine et al., 1997; Sourdive et al., 1997). Analysis of the genomic ySF1 fragments obtained in the two-hybrid screen showed that the shortest region covered by all interacting fusion proteins extended from amino acids 41 to 141 of ySF1 (Fig. 4; corresponding to amino acids 28–128 of hSF1) and is located N-terminal to the maxi-KH domain. As shown

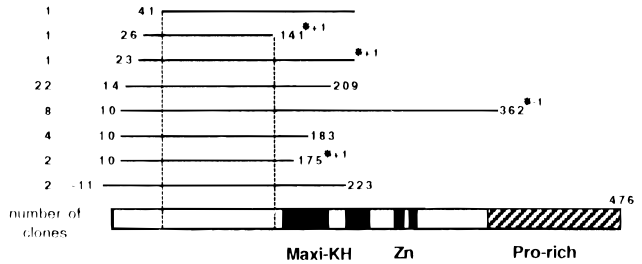

FIGURE 4. Characterization of the Mud2p-interaction domain of ySF1. ySF1 genomic fragments selected in a two-hybrid screen performed with Mud2p are indicated above a schematic view of the ySF1 sequence. Numbers of the first and last amino acids are indicated relative to the initiation codon. The 3′ ends of the inserts starting at amino acids 41 and 23 have not been determined to the nucleotide. Out-of-frame fusions with the GAL4 domain are marked with −1 and +1 for the two alternative reading frames. The thoroughness of the screen can be evaluated from the number of independent clones (indicated on the left) found in the screen.

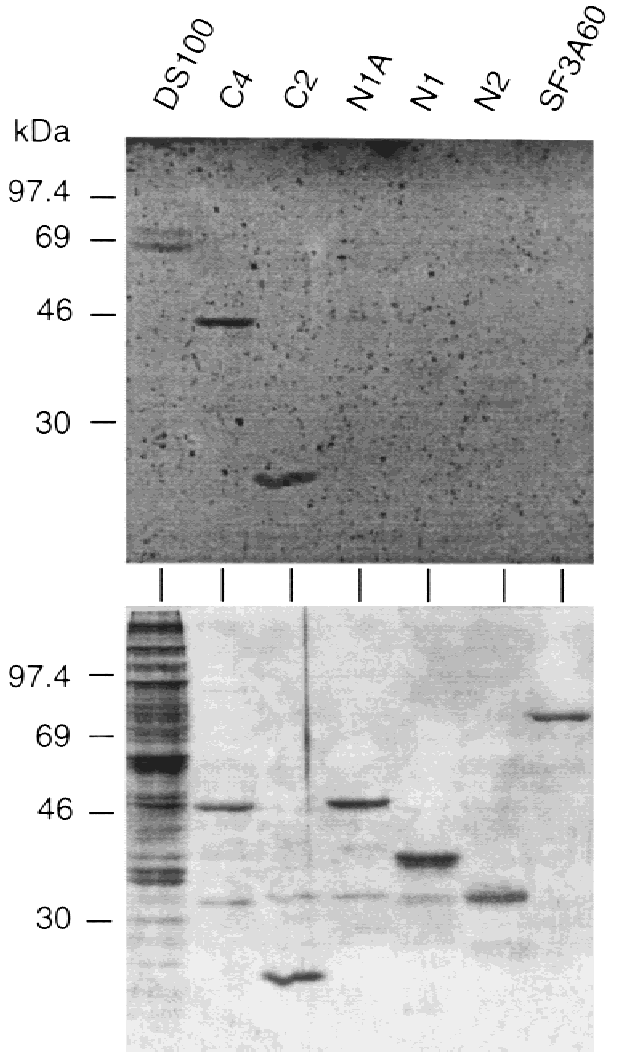

FIGURE 5. N terminus of hSF1 is required for an interaction with U2AF^{65}. SF1 partially purified from HeLa cells (DS100) or recombinant hSF1 and hSF3a60 (as indicated above the figure) were separated in a 10% SDS polyacrylamide gel and blotted to nitrocellulose. The membrane was incubated with [^{35}S] methionine-labeled GST-U2AF^{65}. Top: Far western analysis. Bottom: Silver-stained SDS polyacrylamide gel of SF1 and SF3a60 proteins.

above, this region is essential for spliceosome assembly in the human system.

In a yeast genomic two-hybrid screen performed with ySF1 as a bait, no clones encoding Mud2p were selected (Fromont-Racine et al., 1997). Yeast Mud2p and human U2AF^{65} are structurally related and have been considered to be potential homologues (Abovich et al., 1994). Thus, we performed a series of experiments to test for direct contacts between U2AF^{65} and hSF1. When the U2AF^{65} cDNA was cloned into the two-hybrid vector pACTII and analyzed for an interaction with hSF1, no yeast transformants were obtained, suggesting that U2AF^{65} exhibits strong toxicity in yeast. Therefore, hSF1, cloned into the two-hybrid vector pAS2ΔΔ, was used in a two-hybrid screen of a human liver cDNA library. Full-length hSF1 could not be used as a bait due to a strong transcriptional activation in the two-hybrid assay (data not shown). An hSF1 fragment encoding amino acids 1–441 was used instead, and about 15,000,000 interactions were tested using a mating strategy (Fromont-Racine et al., 1997). Of 117 colonies that grew on selective medium for the *HIS3* reporter gene, 47 scored positive for the LacZ reporter. Thirty-three of 45 sequenced clones corresponded to fragments encoding U2AF^{65}. Interestingly, all fragments were fused out of frame (both frames) to the GAL4 activation domain, confirming the finding that an in-frame construct encoding U2AF^{65} is toxic for yeast cells.

The domain of hSF1 required for the interaction with U2AF^{65} was analyzed in a far western assay. A crude fraction derived from HeLa cell nuclear extracts containing hSF1 (DS100; Krämer, 1992) and truncated recombinant hSF1 proteins were separated by SDS-PAGE, transferred to nitrocellulose, and incubated with in vitro-translated [^{35}S] methionine-labeled U2AF^{65} (Fig. 5). Whereas no interaction is apparent with an unrelated protein (SF3a60), in the DS100 fraction, U2AF^{65} specifically binds to two proteins of ~67 and 75 kDa, which correspond to SF1 isoforms as confirmed by western blotting of the same filter with an antibody directed against the N-terminal half of hSF1 (data not shown). Thus, the interaction demonstrated in the two-hybrid screen can be reproduced in vitro. The region in hSF1 required for the interaction with U2AF^{65} was determined with truncated hSF1 proteins. Both C-terminal hSF1 deletion mutants tested (C2 and C4) bind U2AF^{65} efficiently, indicating that the N-terminal 137 amino acids of hSF1 are sufficient for an interaction between the two proteins. Deletion of 28 amino acids or more from the N terminus of hSF1 (N1a, N1, and N2) abolishes the hSF1/U2AF^{65} interaction, suggesting that this region of hSF1 contributes to the bind-

ing between the proteins. In a sequence alignment, amino acid 28 of hSF1 corresponds to amino acid 41 in ySF1, which has been defined as the N-terminal border of the ySF1 Mud2p-interaction domain (see Fig. 4; Fromont-Racine et al., 1997). The results presented in Figure 5 suggest that either all or part of the N-terminal 28 amino acids of hSF1 are required for binding to U2AF^{65}. It is, however, possible that the N terminus of the mutant hSF1-N1a protein in the context of the recombinant protein is folded in a way that differs from the wild-type protein, thus interfering with the proper interaction between hSF1 and U2AF^{65}.

In summary, the data presented for the Mud2p and U2AF^{65}-interaction domains of ySF1 and hSF1, respectively, are in good agreement. We conclude that sequences of the SF1 orthologues N-terminal of the maxi-KH domain are necessary and sufficient for the interaction with Mud2p and U2AF^{65}.

The C-terminal RRMs of U2AF^{65} and Mud2p are essential for the interaction with the SF1 orthologues

In the yeast two-hybrid screen with hSF1 as bait, the smallest fragment of U2AF^{65} identified encoded the C-terminal region (residues 240–475), which includes part of RRM2 and the complete RRM3. Thus, this region of U2AF^{65} is sufficient for an interaction with hSF1 in vivo. To determine the domain in U2AF^{65} that is involved in the binding to hSF1 more precisely, full-length and truncated versions of U2AF^{65} were expressed as GST-fusion proteins in *E. coli* (Fig. 6A). The proteins were bound to glutathione agarose and incubated with in vitro-translated [^{35}S] methionine-labeled hSF1. After washing, bound proteins were eluted with SDS sample solution and separated by SDS-PAGE. Full-length U2AF^{65} and N-terminal deletion mutants lacking 55 or 94 amino acids bound hSF1 to similar extents, whereas proteins from which 45 or more C-terminal amino acids had been removed or GST alone did not interact with hSF1 (Fig. 6B). RNase-treated in vitro-translated hSF1 bound to the GST-U2AF^{65} proteins, suggesting that the interaction is not mediated by RNA (data not shown). To test whether RRM3 of U2AF^{65} was sufficient for binding of hSF1, U2AF^{65} mutant proteins containing either part of RRM2 and RRM3 (Δ1–287) or only RRM3 (Δ1–366) were used. hSF1 bound to both proteins (Fig. 6B) and the binding was affected in an identical way by increasing the salt concentration in the wash buffer (data not shown). The interaction was reduced in the presence of 200 mM NaCl and virtually abolished at 300 mM NaCl. These results demonstrate that RRM3 is sufficient for the in vitro interaction and RRM2 does not appear to contribute to the association between U2AF^{65} and hSF1.

The ySF1-interaction domain of Mud2p was defined in two-hybrid assays with truncated versions of Mud2p (cloned into pAS2ΔΔ) and the ySF1 fragment encoding

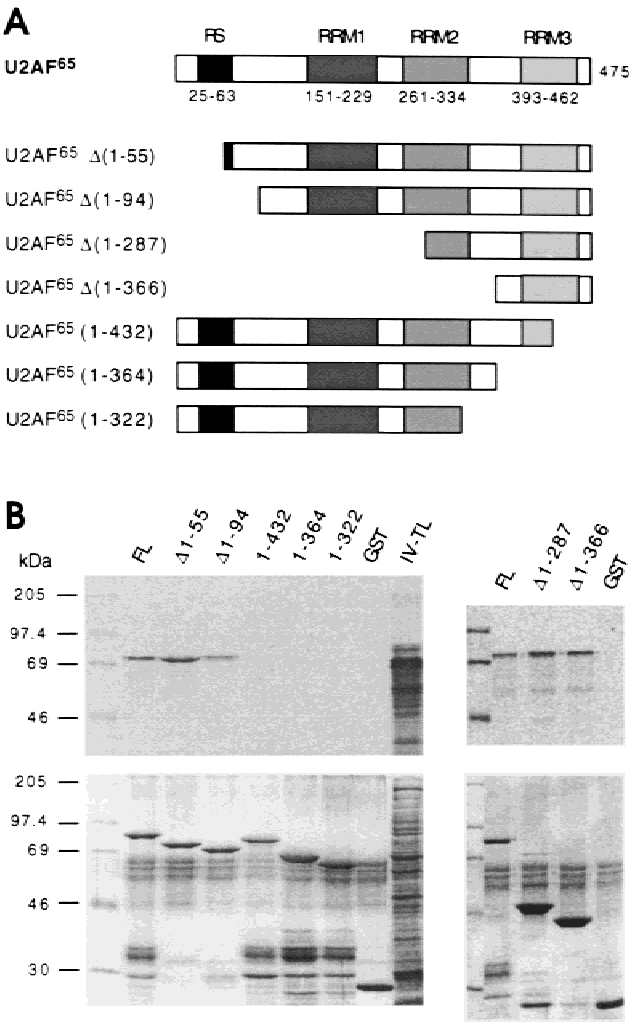

FIGURE 6. C terminus of U2AF^{65} is required for the interaction with hSF1. **A:** Full-length U2AF^{65} and mutant derivatives are represented schematically as in Figure 1. The proteins carry an N-terminal GST-tag and contain the amino acids indicated. **B:** GST and GST-tagged full-length (FL) or mutant U2AF^{65} proteins were bound to glutathione agarose and incubated with [^{35}S] methionine-labeled hSF1. Bound proteins were eluted with SDS sample solution and separated in a 10% SDS polyacrylamide gel. The gel was stained with Coomassie blue (bottom), dried, and autoradiographed (top). The GST-U2AF^{65} proteins used are indicated on top of the figure. The lane marked IV-TL represents 100% of the input of in vitro-translated SF1 used in the experiment.

amino acids 14–209 that was selected in the initial two-hybrid screen (Fig. 7). None of the C-terminal deletion mutants of Mud2p exhibited binding to this portion of ySF1. On the contrary, the N-terminal deletion mutant of Mud2p that contained amino acids 343–527 (i.e., part of RRM2 and the complete RRM3) interacted with ySF1. The binding to ySF1 was weaker than that of full-length Mud2p, but the quantitative value of two-hybrid assays cannot be directly related to the strength of the interaction. The stability of the fusion protein might also interfere with the final outcome of the quantitative assay. For example, similar variations were reported previously and discussed for interactions of Prp9p

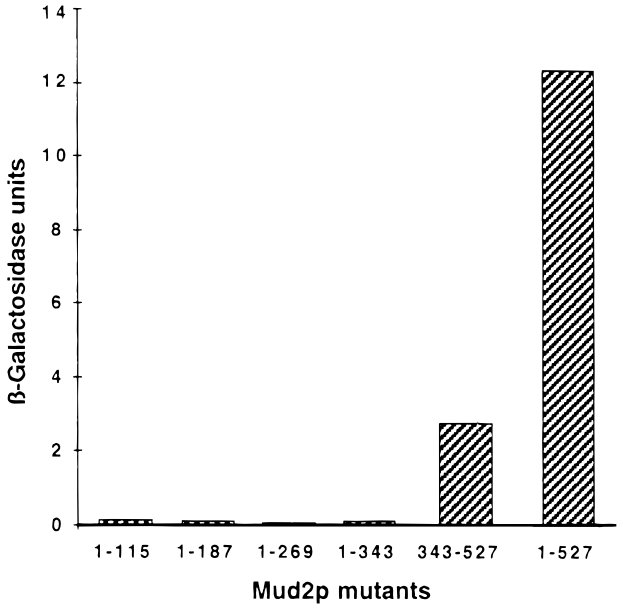

FIGURE 7. Mud2p interacts with ySF1 through its C terminus. Two-hybrid assays were performed in the Y546 yeast strain and the β-galactosidase activity of the LacZ reporter was assayed. Numbers of the first and last amino acids of full-length and mutant Mud2 proteins are indicated. Mean values for two independent transformants are given. Experimental variations were less than 5%. Values for C-terminal deletion mutants are at background level (less than 0.15 unit).

and Prp11p with Prp21p (Rain et al., 1996). In summary, our results demonstrate that U2AF^{65} and Mud2p share a functionally conserved domain encompassing RRM3, which is responsible for the interaction with the SF1 orthologues.

Functional complementation between human and yeast SF1 orthologous domains

We have shown above that full activity of hSF1 in spliceosome assembly requires the N-terminal U2AF^{65}-interaction domain and the maxi-KH domain involved in RNA binding. For comparison, we have also tested the minimal domain(s) of ySF1 required for viability in yeast (Fig. 8). First, we have shown that a clone selected in the initial two-hybrid screen with Mud2p, which encodes amino acids 10–362 of ySF1 fused to the GAL4 activation domain, complements the ySF1 disruption (Fromont-Racine et al., 1997). More extensive C-terminal truncations (that delete the zinc knuckles and part of the maxi-KH domain up to amino acid 183) of ySF1 do not complement the disruption. Second, to further analyze the functional complementation of the ySF1 disruption, sequences encoding chimeric constructs between the SF1 orthologues or

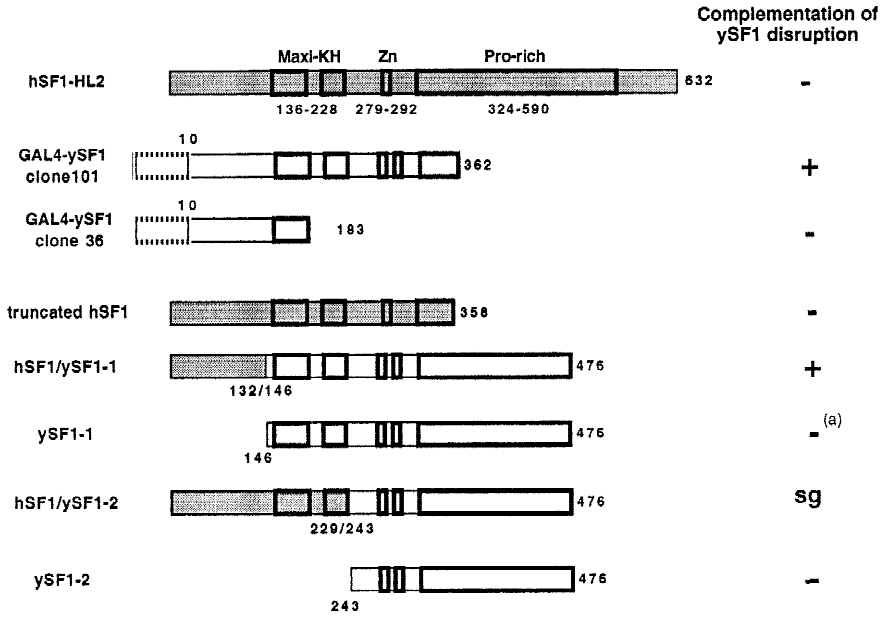

FIGURE 8. ySF1/hSF1 chimeric proteins partially complement the ySF1 disruption. A schematic view of the proteins assayed by functional complementation is shown with the phenotype of complementation indicated. The GAL4 domain of chimeric proteins selected in the screen appears as a hatched box. The complementation was assayed by tetrad analysis of a diploid heterozygous strain disrupted for one allele of the ySF1 gene (Fromont-Racine et al., 1997). The growth assay was performed at 25°C for five days. +, normal growth of [Trp$^+$] spores; sg, slow growth of [Trp$^+$] spores; −, no growth of [Trp$^+$] spores; (a), very small [Trp$^+$] colonies were observed after four weeks. No temperature-sensitive phenotype was observed at 15 or 37°C.

truncated versions of hSF1 were cloned into the vector pVT103U that provides an ADH gene promoter and terminator. Results of the complementation are shown in Figure 8. A complete restoration of viability is obtained with a chimeric protein containing the U2AF^{65}-interaction domain of hSF1 and the maxi-KH domain of ySF1 (hSF1/ySF1-1). Further replacement of the ySF1 maxi-KH domain by the human counterpart results in a slow-growth phenotype (hSF1/ySF1-2). The restoration of viability was due to the chimeric nature of these constructs because the fragments containing only the yeast parts did not restore growth when assayed under the same conditions (Fig. 8; ySF1-1 and ySF1-2). The replacement of ySF1 by a clone expressing amino acids 1–358 of hSF1 (which includes the single zinc knuckle of hSF1) or by the complete human SF1-HL2 cDNA (Arning et al., 1996) does not allow any complementation.

In conclusion, these results indicate that the maxi-KH domain of ySF1 can be partially replaced by the corresponding domain of hSF1, suggesting a similar function. The significance of the functional replacement of the Mud2p-interaction domain of ySF1 with the U2AF^{65}-interaction domain of hSF1 remains to be addressed with additional methods, because the MUD2 gene is not essential (Abovich et al., 1994). In this respect, we have not been able to detect significant heterologous interactions between ySF1 and U2AF^{65} or hSF1 and Mud2p using the two-hybrid assay (data not shown).

DISCUSSION

We have investigated functional domains in yeast and human SF1 with in vivo and in vitro approaches. In both proteins, the N-terminal halves, which show the highest degree of sequence conservation, are either essential for viability of yeast cells (ySF1; Fig. 8) or for activity in pre-spliceosome assembly (hSF1; Fig. 3), whereas the less-conserved C-terminal portions are not required in either assay. The N-terminal halves comprise an RNA-binding domain (Fig. 2) and a protein–protein interaction domain that mediates contacts between ySF1 and Mud2p or hSF1 and U2AF^{65} (Figs. 4, 5). Analysis of the regions in Mud2p and U2AF^{65} that are required for these interactions indicates an involvement of the C-terminal portions, which consist of a degenerate RNA recognition motif (Figs. 6, 7). This region shows the highest sequence conservation between Mud2p and U2AF^{65}. Based on the conservation of structural features and functional domains, we conclude that ySF1 and hSF1, as well as Mud2p and U2AF65, are true functional homologues.

The RNA-binding domain of hSF1 is located between amino acids 94 and 278. This region includes the maxi-KH domain, which was first described in the hnRNP K protein (Siomi et al., 1993a) and has been found since in a large number of proteins that function in association with RNA (for references, see Musco et al., 1996). A well-studied example is the product of the FMR1 gene, which is responsible for the fragile X syndrome, a common form of heritable mental retardation in human males (Oostra & Verkerk, 1992; Ashley et al., 1993; Siomi et al., 1993b). Among the splicing factors, mammalian KSRP (Min et al., 1997), *Drosophila* PSI (Siebel et al., 1995; Adams et al., 1997) and *S. cerevisiae* Mer1p (Nandabalan et al., 1993; Nandabalan & Roeder, 1995) contain one or more KH domains. These proteins act to promote or repress specific alternative splicing events by interaction with their target pre-mRNA sequences. It has not been reported whether the KH domain(s) in these proteins are required for activity.

The solution structure of two KH domains in human FMR1 and vigilin consists of a stable $\beta\alpha\alpha\beta\beta\alpha$ fold in which the three α-helices are packed along the three-stranded antiparallel β-sheet (Musco et al., 1996, 1997). Assuming that the hSF1 KH domain structure does not differ significantly from that reported for FMR1 and vigilin, the failure to support RNA binding of hSF1 proteins that carry mutations in the KH domain can be explained by the removal of sequences implicated in RNA binding or the stability of the KH domain. For example, in hSF1-ΔKH-N, which is inactive in RNA binding, the N-terminal $\beta\alpha\alpha\beta$ portion of the maxi-KH domain is deleted. This deletion removes a flexible loop (containing the highly conserved GXXG motif), which is located between the two N-terminal α-helices and has been suggested to provide the RNA binding surface of the domain (Musco et al., 1996, 1997). The mutation G161D in the GXXG sequence results in drastically reduced RNA-binding activity, which underscores the importance of the GXXG motif for the interaction with RNA. Similar deletion or point mutations in human hnRNP K, Sam68, and *Xenopus* Xqua reduce the binding to RNA homopolymers in vitro (Siomi et al., 1994; Chen et al., 1997; Zorn & Krieg, 1997), and missense mutations in this motif in the *C. elegans* Gld-1 and Mex-3 proteins lead to severe developmental defects (Jones & Schedl, 1995; Draper et al., 1996). α-Helix 2 may contribute to the RNA-binding function of the KH domain (Musco et al., 1996). The point mutation of L164N, which is located in this region, only marginally affected the RNA-binding activity of hSF1. The corresponding mutation in FMR1 (I304N) results in a severe case of the fragile X syndrome (DeBoulle et al., 1993). The binding of the mutated FMR1 protein to homopolymeric RNA was compromised in vitro (Siomi et al., 1994) and a similar mutation in Sam68 abolished RNA-binding activity (Chen et al., 1997). Less severe effects have been observed with analogous mutations in individual KH domains of hnRNP K (Siomi et al., 1994). Thus, the RNA-binding efficiency of different proteins carrying this mutation appears to vary. Deletion of the C-terminal conserved portion of the maxi-KH domain in hSF1-

ΔKH-C removes α-helix 3, which is essential for the stability of the KH fold (Musco et al., 1996, 1997). The inability of this mutant protein to bind RNA most likely results from a perturbation in the structure of the truncated KH domain. Notably, the original mutation in ySF1 (the *msl-5* allele; Gly 230 to Ser) that was isolated in a genetic synthetic-lethal screen with MUD2 is a point mutation within α-helix 3, and the mutant protein is inactive in commitment complex formation (Abovich & Rosbash, 1997).

Although these results suggest that the KH domain is a major determinant of the RNA-binding activity of hSF1, it may not be sufficient for this function; 40 amino acids N-terminal to the KH domain appear to contribute to the interaction with RNA. In agreement with this finding, Berglund et al. (1998) have reported that an hSF1 protein containing amino acids 135–308 binds to a 34-nt RNA containing the AdML 3′ splice site about 10-fold less efficiently than hSF1 containing amino acids 1–361. Moreover, we have not yet tested whether sequences C-terminal to the KH domain (up to amino acid 278) contribute to the RNA-binding activity of hSF1. Thus, all mutations in the KH domain that interfered with binding of hSF1 to RNA negatively affected pre-spliceosome assembly, demonstrating that the RNA-binding function of hSF1 is essential for splicing.

Our results did not reveal a function for the zinc knuckle of hSF1 in RNA binding or splicing complex formation; its role, if any, remains to be established. In the UV cross-linking assay, we monitored the sequence-independent binding of hSF1 to RNA (see Arning et al., 1996). Thus, we did not address the possibility that the zinc knuckle is necessary for the sequence-specific interaction of hSF1 with the branch site sequence (Berglund et al., 1997). Assuming that this specific interaction is essential for spliceosome formation, we consider an involvement of the zinc knuckle less likely, because deletions of this motif were without effect on pre-spliceosome assembly. Two other splicing proteins containing zinc knuckles have been isolated. A role for this motif in the human SR protein 9G8 has not been reported (Cavaloc et al., 1994), and genetic experiments suggested that a zinc knuckle in the yeast splicing factor Slu7p is not essential for function (Frank & Guthrie, 1992; Zhang & Schwer, 1997).

The binding of ySF1 to RNA has not been investigated in this study. However, given the evolutionary conservation of the SF1 maxi-KH domain (including residues not conserved in other KH domains; see Arning et al., 1996; Musco et al., 1996) and the result that ySF1 specifically binds to the branch site (Berglund et al., 1997), we expect that the maxi-KH domain also mediates the binding of ySF1 to RNA. The complementation of the ySF1 null allele with chimeric hSF1/ySF1 proteins suggests that the maxi-KH domain of hSF1 functions at least partially in yeast, reinforcing the structural and functional homology between the two proteins. The inability of full-length or truncated hSF1 to complement the yeast null allele probably reflects the divergence between the yeast and human systems, which may be related to differences in the splicing substrates and/or differences in the regulation of the activities of these splicing factors (see below).

Sequences N-terminal to the KH domains of hSF1 and ySF1 mediate the interaction with $U2AF^{65}$ and Mud2p, respectively. Amino acids 2–137 of hSF1 are sufficient for its association with $U2AF^{65}$, whereas the minimal Mud2p interaction domain of ySF1 is located between amino acids 41 and 141 (this report; Fromont-Racine et al., 1997). An alignment of SF1 sequences from *S. cerevisiae*, *S. pombe* (I. Witt & J. Potashkin, pers. comm.), *Caenorhabditis*, *Drosophila* (R. Mazroui & A. Krämer, unpubl. data), and human demonstrates a high evolutionary conservation in this region; for example, ySF1 and hSF1 are 50% identical and 60% similar in the interaction domain compared to 45% identity and 62% similarity in the maxi-KH domain. It remains to be shown whether the entire region N-terminal to the KH domain constitutes the interaction domain. Computer analysis of the N-terminal sequences of the SF1 orthologues with the programs Pepcoil (Lupas et al., 1991) and Proteinpredict (Rost & Sander, 1994) suggests the presence of two coiled coil or α-helical regions (Cohen & Parry, 1990; O'Shea et al., 1991). Inspection of the amino acid sequences in these regions of SF1 revealed a potential direct tandem repeat (Fig. 9). Future experiments are aimed at testing whether these sequences are necessary for contacts between hSF1 and $U2AF^{65}$ or ySF1 and Mud2p.

The C-terminal portion of $U2AF^{65}$ and Mud2p mediates the interaction with hSF1 and ySF1. In Mud2p, amino acids 342–527, i.e., half of RRM2 and the complete RRM3, are sufficient for the association with ySF1 in vivo. A two-hybrid screen with hSF1 identified amino acids 240–475 (i.e., RRM2 and RRM3) of $U2AF^{65}$ as the smallest interacting fragment. In vitro binding studies showed that the minimal interaction domain of $U2AF^{65}$ comprises amino acids 367–475, i.e., the entire RRM3. In addition, Berglund et al. (1998) have shown that amino acids 334–475 of $U2AF^{65}$ supported the association with hSF1 in vivo and in vitro, whereas an N-terminal truncation to amino acid 391 abolished the interaction. Moreover, removal of the C-terminal 43 amino acids of $U2AF^{65}$ is detrimental to the interaction with hSF1 (this report). Hence, it appears that the entire RRM3 mediates the binding to hSF1, and RRM2 is most likely not important. Interestingly, the highest and most convincing homology between $U2AF^{65}$ and Mud2p is present in RRM3, and this particular RRM deviates from a consensus RRM (Birney et al., 1993; Abovich et al., 1994).

All three RRMs of $U2AF^{65}$ contribute to the specific, high-affinity binding of the protein to the polypyrimidine tract, although the contribution of RRM3 to the effi-

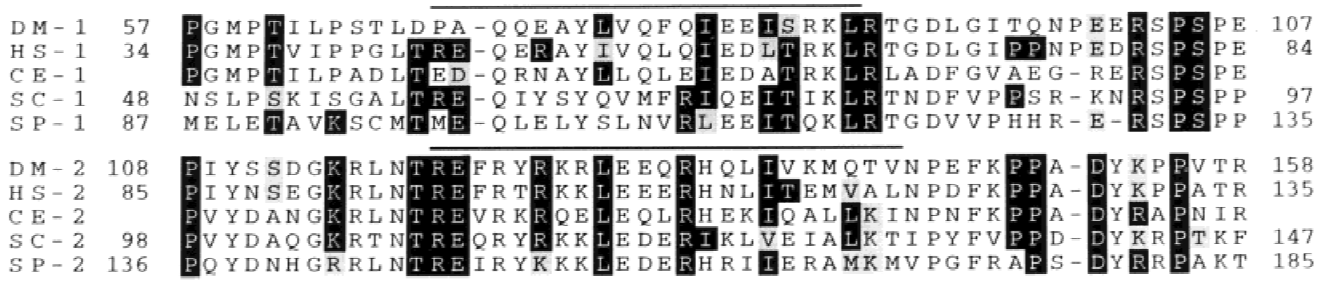

FIGURE 9. Comparison of SF1 homologues reveals a novel conserved feature. Amino acid sequence alignment of the N-terminal portions of SF1 from *C. elegans* (CE), *Drosophila* (DM), human (HS), *S. cerevisiae* (SC), and *S. pombe* (SP) was performed with the CLUSTALW program of the GCG package. Shading was done with the WWW Boxshade server of the ISREC Bioinformatics Group. Identical amino acids are shaded in black, conserved residues in gray. The two putative helical regions are indicated by a bar on top of the alignments.

ciency of RNA binding is lower than that of RRM1 (Zamore et al., 1992). Thus, RRM3 may play a dual role in contacting RNA and engaging in protein–protein interactions. If this is the case, it will be interesting to see whether the sequences that mediate the RNA–protein and protein–protein interactions within RRM3 are separate from one another or overlapping. Other examples of RRMs that mediate protein–protein contacts are the U2 snRNP B″ protein (Scherly et al., 1990) and *S. cerevisiae* Hsh49p, the homologue of the smallest subunit of mammalian SF3b (Igel et al., 1998). The binding of Cus1p is at least partially mediated through RRM1 of Hsh49p. Although it has not been shown that RRM1 of Hsh49p is required for the RNA-binding activity of the protein, mutations in putative RNA-binding residues of this domain are lethal, but do not prevent the interaction between the two proteins, which may indicate that protein interaction does not involve the RNA-binding surface of the RRM.

In addition to binding hSF1, U2AF^{65} has been isolated in a tight complex with U2AF^{35} (Zamore & Green, 1989) and also interacts with UAP56, which is required prior to binding of U2 snRNP to the branch site (Fleckner et al., 1997). It is unknown at present whether all three proteins interact with U2AF^{65} at the same time; however, because the interaction domains of U2AF^{35} and UAP56 are located in the N-terminal half of U2AF^{65}, the simultaneous binding of hSF1 to the C-terminal RRM3 is possible.

Our experiments did not yield any information regarding a role for the proline-rich region of SF1. First, the activity in spliceosome assembly of hSF1-C4 that lacks this region is comparable to that of the full-length protein purified from HeLa cells. Bedford et al. (1997) have recently shown that formin-binding proteins and *abl* bind specific sequences in the proline-rich region of hSF1. These interactions may be relevant to the regulation of hSF1 activity at a level that is not addressed with the rather simple assays used here. Second, a ySF1 mutant that is devoid of most of the proline-rich region can fully complement the null allele of the ySF1 gene. Except for the proline richness, the amino acid sequence in this region is not highly conserved between yeast and human (or other species; unpubl. results). Given this divergence, it is possible that a function of the proline-rich regions was adapted to the needs of the particular organisms or may have been lost in certain species.

Despite the progress in unravelling interactions within the spliceosome during the commitment of the pre-mRNA to the splicing pathway, one of the remaining key issues is to understand the dynamics of early spliceosome assembly, as well as apparent differences between the yeast and mammalian splicing systems. In yeast, ySF1 and Mud2p are involved in the conversion of CC1 to CC2 (Abovich et al., 1994; Abovich & Rosbash, 1997), which requires the presence of a functional branch site sequence (Séraphin & Rosbash, 1991). Mud2p binding to the pre-mRNA depends on the branch site (Abovich et al., 1994) and ySF1 interacts directly with this sequence (Berglund et al., 1997). Furthermore, a conserved residue at the position preceding the UACUAAC box is important during the commitment step in yeast, and Mud2p is involved directly or indirectly in the recognition of this nucleotide (Rain & Legrain, 1997). In the human system, U2AF^{65} binds the polypyrimidine tract via its RRMs (Zamore et al., 1992), and it contacts the branch site through its RS domain in the absence of other components (Gaur et al., 1995; Valcárcel et al., 1996). hSF1 is present in complex E and can specifically recognize the branch site sequence (Berglund et al., 1997; Z. Rafi & A. Krämer, unpubl. data).

Given the protein–protein interaction between ySF1/hSF1 and Mud2p/U2AF^{65}, these proteins may be considered as part of a common functional entity, a view that is supported by synergistic interactions of hSF1 and U2AF^{65} with the splicing substrate (Berglund et al., 1998). However, many questions remain to be solved. For example, both hSF1 and U2AF^{65} have been shown to bind to the branch site (Gaur et al., 1995; Valcárcel et al., 1996; Berglund et al., 1997, 1998). Do these in-

teractions occur at the same time, or sequentially? If both proteins bind at the same time, do the interactions involve the same nucleotides? Furthermore, if hSF1 and U2AF^{65} bind sequentially, in which order do the interactions occur? Because U2 snRNA base pairs with the branch site sequence during the assembly of pre-splicing complex A, this process most likely requires the dissociation of some splicing components. The association of U2AF^{65} with the assembling spliceosome is apparently weakened prior to complex A formation (Champion-Arnaud et al., 1995). Evidence for a possible dissociation of hSF1 from the spliceosome is indirect. First, a protein of 80 kDa that crosslinks to the branch site in an ATP-independent and U1 snRNP-dependent fashion does not crosslink at later stages of the splicing reaction (MacMillan et al., 1994). Second, a protein of 72 kDa has been found crosslinked to the branch site in complex E, but not in complex A (Chiara et al., 1996). Both of these proteins have the approximate size and functional characteristics of hSF1. Moreover, it remains to be shown whether hSF1 and/or U2AF^{65} (or their yeast homologues) interact with any of the proteins associated with the active U2 snRNP (Brosi et al., 1993) before base pairing of U2 snRNA with the pre-mRNA branch site. Clearly, further experiments are needed to clarify the dynamics of protein–protein and protein–RNA interactions during the early stages of spliceosome assembly.

MATERIALS AND METHODS

Plasmids and yeast strains

The following yeast strains were used: CG1945 from Clontech, Y526 (GAL4Δ GAL80Δ URA3::UASGAL1-LacZ his3 leu2 trp1), Y187 (GAL4Δ GAL80Δ ade2-101 his3 leu2-3,112 trp1-901 ura3-52 URA3::UASGAL1-LacZ), LMA1, a diploid strain containing a completely disrupted allele of the ySF1 gene (Fromont-Racine et al., 1997). Yeast manipulation, transformation, and selection on drop-out media were according to standard protocols. The lacZ assay was performed according to Transy and Legrain (1995). Yeast MUD2 and SF1 two-hybrid baits were cloned into the pAS2ΔΔ vector (Fromont-Racine et al., 1997). Human SF1-Bo DNA was digested with *Nco* I and cloned into pAS2ΔΔ (Arning et al., 1996). Mud2p C-terminal deletion constructs were derived from the pAS2ΔΔ-MUD2 plasmid by digestion with *Apa* I and *Pst* I (positions 1–115), *Sma* I and *Pst* I (1–187), *Bgl* II and *Pst* I (1–269), or *Kpn* I and *Pst* I (1–343), blunt-end formation, and religation. The N-terminal deletion was obtained after digestion with *Nco* I and *Kpn* I and religation with appropriate adapters (5′-CATGGAGGCCCCG GGTAC-3′ and 5′-CCGGGGCCTC-3′).

Two-hybrid yeast genomic and cDNA library screening

The yeast genomic library FRYL was screened according to Fromont-Racine et al. (1997). A human liver cDNA library from Clontech (HL4024AH) was introduced into the yeast strain Y187 by transformation. Eight million colonies were recovered, pooled, and frozen in aliquots. Two-hybrid screens were then performed as for the FRYL yeast genomic library (Fromont-Racine et al., 1997).

hSF1 and chimeric constructs for functional complementation

Full-length SF1-HL2 cDNA (Arning et al., 1996) was digested with *Bam*H I and cloned into pVT103U (Vernet et al., 1987). Chimeric constructs were made from PCR products generated with cloned PFU polymerase (Stratagene) on hSF1-Bo and pAS2ΔΔ-MUD2 plasmids. PCR products were digested and cloned into the pVT103U vector by three-way ligation. *Bam*H I, *Kpn* I, and *Hin*d III cloning sites were used for upstream, middle, and downstream junctions, respectively. The amino acids contributed by human and yeast SF1 sequences are shown schematically in Figure 8. The amino acids at the junctions are human P_{132}–G–yeast T_{146} and human G_{229}–G–yeast T_{243}, for hSF1/ySF1-1 and hSF1/ySF1-2, respectively. The yeast constructs ySF1-1 and ySF1-2 were derived after digestion with *Bam*H I and *Kpn* I and religation with a double-strand adapter (5′-GATCATGGCGACCGGTAC-3′) of the hSF1/ySF1-1 and hSF1/ySF1-2 constructs, respectively.

The yeast diploid strain LMA1 containing one disrupted allele of the ySF1 gene (TRP1 cassette) was transformed by the different constructs on URA3 plasmids, sporulated, and tetrads were dissected. Twelve tetrads were analyzed in each case and the 2:2 segregation of the sexual markers was controlled. All [Trp$^+$] spores were also [Ura$^+$]. For all tetrad analyses, most [Trp$^-$] spores were [Ura$^+$], demonstrating that all constructs could pass meiosis. For each construct, two independent transformants from *E. coli* were used for complementation and gave the same results.

Construction of hSF1 and U2AF^{65} mutant plasmids

cDNAs encoding truncated versions of hSF1 were amplified from pGEM-SF1-Bo (Arning et al., 1996) by PCR with Pwo polymerase. Plasmids were constructed by standard cloning procedures (Sambrook et al., 1989) and amplified DNAs were sequenced with the T7 DNA sequencing kit (Pharmacia). PCR fragments containing 5′ *Bam*H I and 3′ *Eco*R I restriction sites were cloned into the corresponding sites of the vector pTRCHisA (Invitrogen) for the expression in *E. coli* of recombinant proteins carrying N-terminal His_6-tags.

Internal deletions of the KH domain and the zinc knuckle were generated by PCR of DNA sequences upstream and downstream of the desired deletion and the amplified fragments were cloned into the *Bam*H I and *Eco*R I sites of pTRCHisA by three-way ligation. The recombinant proteins contain a *Kpn* I restriction site in place of the deleted DNA sequences; thus, in the corresponding proteins, the deleted amino acids are replaced by glycine and threonine.

For the generation of point mutations, 31- or 33-nt primers spanning the site of the desired mutation on both DNA strands were used. In these primers, the following codons were changed: G161D, GGG to GAC; L164N, CTG to AAC; H287L, CAC to CTC. Plasmid DNA was amplified by PCR, followed by digestion of the methylated, nonmutated, parental DNA in

the presence of 10 units of *Dpn* I for 1 h at 37 °C. Amplified DNA was transformed into XL1-Blue cells. Mutations were confirmed by sequencing. The mutant SF1 proteins are shown schematically in Figure 2A.

Plasmids encoding GST-tagged full-length U2AF^{65} (GST-U2AF^{65}; Zamore et al., 1992) and deletion mutants GST-U2AF$^{65}\Delta$(1–55) and GST-U2AF$^{65}\Delta$(1–94) (Valcárcel et al., 1996) were kindly provided by M. Green and J. Valcárcel. Plasmids encoding the N-terminal 323, 365, and 430 amino acids of U2AF^{65} [GST-U2AF^{65}(1–323), GST-U2AF^{65}(1–365), and GST-U2AF^{65}(1–430)] were generated by double digestion of GST-U2AF^{65} (in pGEX-3X) with *Sph* I, *Sac* I, and *Bgl* II, respectively, and *Eco*R I. Restriction sites were filled in, the fragments were gel-purified, religated, and transformed into *E. coli* XL1-Blue cells. N-terminal deletion mutants GST-U2AF$^{65}\Delta$(1–287) and GST-U2AF$^{65}\Delta$(1–366) were generated by digestion with *Stu* I and *Eco*R I or *Sac* I and *Eco*R I, respectively. Restrictions sites were filled in, the *Stu* I–*Eco*R I and *Sac* I–*Eco*R I fragments were gel-purified, ligated into the pGEX-1T vector (Pharmacia), and transformed into XL1-Blue cells. A plasmid for the in vitro translation of U2AF^{65} was constructed by cloning the *Bam*H I–*Eco*R I insert of GST-U2AF^{65} into the corresponding restriction sites of pGEM4 (Promega). The U2AF^{65} mutants are shown schematically in Figure 6A.

Expression and purification of recombinant proteins

hSF1 plasmids were transformed into *E. coli* strain TOP10 (Invitrogen) by electroporation. Typically, proteins were expressed in 100-mL cultures for 4 h at 37 °C after addition of IPTG to a final concentration of 1 mM. Cells were harvested by centrifugation at 6,000 × *g* for 15 min and lysed in lysis buffer (50 mM NaH_2PO_4, pH 8.0, 10 mM Tris-HCl, pH 8.0, 8 M urea, and 100 mM NaCl). Proteins were purified on a 1-mL column of Talon Metal Affinity Resin (Clontech) equilibrated in lysis buffer. Unbound proteins were removed with lysis buffer. His_6-tagged proteins were eluted with 0.1 M EDTA, pH 8.0, dialyzed against buffer D (Dignam et al., 1983) containing 3 mM $MgCl_2$, and stored at −80 °C. All purified proteins were soluble and of comparable purity (for an example, see Fig. 5).

GST-U2AF^{65} derivatives were expressed in *E. coli* XL1-Blue cells for 3 h after induction with 0.1 mM IPTG. GST-tagged SF3a60 expressed in insect cells was a kind gift from Dobrila Nesic. Recombinant proteins were purified on glutathione-agarose as described (Smith & Johnson, 1988).

UV crosslinking and splicing complex formation

Proteins were covalently crosslinked to an in vitro-transcribed uniformly labeled pre-mRNA derived from the Adenovirus major late (AdML) transcription unit (RNA1) as described by Arning et al. (1996), but omitting ATP and creatine-phosphate. Proteins were separated in 15% SDS polyacrylamide gels and visualized by autoradiography.

Nuclear extracts were fractionated on DEAE-Sepharose Fast Flow (Pharmacia) according to Krämer and Utans (1991). The DS500 fraction, which contains residual SF1 activity, was depleted by incubation with an affinity-purified rabbit antibody raised against hSF1-C4 and sheep anti-rabbit IgG-coated Dynabeads (Dynal). For splicing complex formation, the depleted DS500 fraction was incubated with partially purified SF1 (MonoS fraction; Krämer, 1992) or recombinant hSF1 proteins for 30 min at 30 °C. Analysis of the reaction products was as described previously (Krämer & Utans, 1991).

In vitro translation

Coupled in vitro transcription and translation reactions were performed in a total volume of 100 μL with the reticulocyte lysate TNT-T7 kit (Promega). Reactions were performed for 90 min at 30 °C according to the supplier's instructions in the presence of 0.01 μg/μL leupeptin and 0.25 mM phenylmethylsulfonyl fluoride. hSF1 was produced from plasmid pGEM-SF1-Bo linearized with *Kpn* I (Arning et al., 1996), U2AF^{65} was produced from plasmid pGEM-U2AF^{65} linearized with *Eco*R I. Reactions were diluted with 100 μL of buffer D (Dignam et al., 1983) containing 3 mM $MgCl_2$ and stored at −20 °C.

Far western analysis and GST binding assays

Proteins were separated in 10% SDS polyacrylamide gels and transferred to nitrocellulose (Kyhse-Anderson, 1984). The membranes were blocked, incubated with [^{35}S]-methionine-labeled in vitro-translated U2AF^{65}, and washed as described by Khotz et al. (1994). The denaturation/renaturation steps after the transfer were omitted.

The binding of hSF1 to GST or GST-U2AF^{65} fusion proteins was performed essentially as described (Xiao & Manley, 1997). In brief, 1 μg of protein was bound to 40 μL of glutathione-agarose beads in 300 μL of NETN (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.5% NP-40, 0.5 mM EDTA) for 30 min at 4 °C. The beads were washed twice with 500 μL NETN. In vitro-translated [^{35}S]-methionine-labeled hSF1 (2 μL) was incubated with the beads in a total volume of 200 μL NETN for 30 min at 4 °C followed by six washes with 200 μL NETN. Bound proteins were eluted from the beads by addition of 20 μL SDS gel loading buffer and incubation for 15 min at 65 °C. Proteins were separated in 10% SDS polyacrylamide gels. Gels were stained with Coomassie blue, dried, and exposed to X-ray film.

ACKNOWLEDGMENTS

We thank Catherine Transy for preparation of the human cDNA liver library cloned into yeast strain Y187, Juan Valcárcel and Michael Green for generous gifts of plasmids encoding U2AF^{65}, and Dobrila Nesic for SF3a60. Communication of results prior to publication by Nadja Abovich, Andy Berglund, Michael Rosbash, and Manny Ares is gratefully acknowledged. We also thank Graeme Bilbe and Guillaume Chanfreau for critical comments on the manuscript. This work was supported by European Union grants Biotech 950009 and CHRX-CT94-0677 and the Ministère de la Recherche (ACC-SV1) to P.L. and by grants of the Swiss National Fonds and the Canton of Geneva to A.K.

Manuscript accepted without revision March 9, 1998

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.

Adams M, Tarng R, Rio D. 1997. The alternative splicing factor PSI regulates P-element third intron splicing in vivo. *Genes & Dev 11*:129–138.

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.

Ashley CT, Wilkinson KD, Reines D, Warren ST. 1993. FMR1 protein: Conserved RNP family domains and selective RNA binding. *Science 262*:563–566.

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.

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

Berglund J, Abovich N, Rosbash M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes & Dev*. Forthcoming.

Berglund J, 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.

Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res 21*:5803–5816.

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

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.

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, Gozani O, Palandjian L, Reed R. 1995. Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. *Mol Cell Biol 15*:5750–5756.

Chen T, Damaj B, Herrera C, Lasko P, Richard S. 1997. Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and QK1: Role of the KH domain. *Mol Biol Cell 17*:5707–5718.

Chiara M, Gozani O, Bennett M, Champion-Arnaud P, Palandjian L, Reed R. 1996. Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Mol Cell Biol 16*:3317–3326.

Cohen C, Parry D. 1990. Alpha-helical coils and bundles: How to design alpha-helical proteins. *Proteins Struct Funct Genet 7*:1–15.

Darlix JL, 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.

DeBoulle K, Verkerk AJMH, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van den Bos F, De Graaff E, Oostra BA, Willems PJ. 1993. A point mutation in the *FMR-1* gene associated with fragile X mental retardation. *Nature Genet 3*:31–35.

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.

Draper BW, Mello CC, Bowerman B, Hardin J, Priess JR. 1996. MEX-3 is a KH domain protein that regulates blastomere identity in early *C. elegans* embryos. *Cell 87*:205–216.

Fleckner J, Zhang M, Valcárcel J, Green M. 1997. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes & Dev 11*:1864–1872.

Frank D, Guthrie C. 1992. An essential splicing factor, SLU7, mediates 3′ splice site choice in yeast. *Genes & Dev 6*:2112–2124.

Fromont-Racine M, Rain JC, Legrain P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet 16*:277–282.

Fu XD. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA 1*:663–680.

Gaur R, Valcárcel J, Green M. 1995. Sequential recognition of the pre-mRNA branch point by $U2AF^{65}$ and a novel spliceosome-associated 28-kDa protein. *RNA 1*:407–417.

Green MR. 1986. Pre-mRNA splicing. *Annu Rev Genet 20*:671–708.

Hodges PE, Beggs JD. 1994. RNA splicing. U2 fulfils a commitment. *Curr Biol 4*:264–267.

Igel H, Wells S, Perriman R, Ares M Jr. 1998. Conservation of structure and subunit interactions in yeast homologs of splicing factor 3b (SF3b) subunits. *RNA 4*:1–10.

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.

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.

Keller E, Noon W. 1984. Intron splicing: A conserved internal signal in introns of animal pre-mRNAs. *Proc Natl Acad Sci USA 81*:7417–7420.

Kohtz JD, Jamison SF, Will CL, Zuo P, Lührmann 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. 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. 1996. The structure and function of proteins involved in nuclear pre-mRNA splicing. *Annu Rev Biochem 65*:367–409.

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. Electroblotting of multiple gels. *J Biochem Biophys Methods 10*:203–209.

Legrain P, Rosbash M. 1989. Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell 57*:573–583.

Legrain P, Séraphin B, Rosbash M. 1988. Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol Cell Biol 8*:3755–3760.

Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coil from protein sequences. *Science 252*:1162–1164.

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.

Michaud S, Reed R. 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes & Dev 5*:2534–2546.

Michaud S, Reed R. 1993. A functional association between the 5′ and 3′ splice sites is established in the earliest prespliceosome complex (E) in mammals. *Genes & Dev 7*:1008–1020.

Min H, Turck C, Nikolic J, Black D. 1997. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes & Dev 11*:1023–1036.

Moore M, Query C, Sharp P. 1993. Splicing of precursors to mRNA by the spliceosome. In: Gesteland R, Atkins J, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 303–357.

Musco G, Kharrat A, Stier G, Fraternali F, Gibson T, Nilges M, Pastore A. 1997. The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nature Struct Biol 4*:712–716.

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.

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.

O'Shea E, Rutkowski R, Kim P. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded parallel coiled coil. *Science 254*: 539–554.

Oostra B, Verkerk A. 1992. The fragile X syndrome: Isolation of the FMR-1 gene and characterization of the fragile X mutation. *Chromosoma 101*:381–387.

Potashkin J, Naik K, Wentz-Hunter K. 1993. U2AF homolog required for splicing in vivo. *Science 262*:573–575.

Rain JC, Legrain P. 1997. In vivo commitment to splicing in yeast involves the nucleotide upstream from the branch site conserved sequence and the Mud2 protein. *EMBO J 16*:1759–1771.

Rain JC, Tartakoff AM, Krämer A, Legrain P. 1996. Essential domains of the PRP21 splicing factor are implicated in the binding to PRP9 and PRP11 proteins and are conserved through evolution. *RNA 2*:535–550.

Reed R. 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev 6*:215–220.

Rost B, Sander C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins Struct Funct Genet 19*:55–72.

Ruby SW, Abelson J. 1988. An early and hierarchic role of the U1 snRNP in spliceosome assembly. *Science 242*:1028–1035.

Rymond BC, Rosbash M. 1985. Cleavage of 5′ splice site and lariat formation are independent of 3′ splice site in yeast mRNA splicing. *Nature 317*:735–737.

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.

Scherly D, Dathan NA, Boelens W, van Venrooij WJ, Mattaj IW. 1990. The U2B″ RNP motif as a site of protein–protein interaction. *EMBO J 9*:3675–3681.

Séraphin B, Rosbash M. 1989. Identification of functional U1 snRNA–pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell 59*:349–358.

Séraphin B, Rosbash M. 1991. The yeast branchpoint sequence is not required for the formation of a stable U1 snRNA–pre-mRNA complex and is recognized in the absence of U2 snRNA. *EMBO J 10*:1209–1216.

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.

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.

Smith DB, Johnson KS. 1988. Single-step purification of polypeptides expressed in *E. coli* as fusions with glutathione-S-transferase. *Gene 67*:31–40.

Sourdive D, Transy C, Garbay S, Yaniv M. 1997. The bifunctional DCOH protein binds to HNF1 independently of its 4-alpha-carbinolamine dehydratase activity. *Nucleic Acids Res 25*:1476–1484.

Staknis D, Reed R. 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol Cell Biol 14*:7670–7682.

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.

Transy C, Legrain P. 1995. The two-hybrid: An in vivo protein–protein interaction assay. *Mol Biol Rep 21*:119–127.

Valcárcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF^{65} RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science 273*:1706–1709.

Vernet T, Dignard D, Thomas D. 1987. A family of yeast expression vectors containing the phage F1 intergenic region. *Gene 52*:225–233.

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 J. 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein–protein and RNA–protein interactions and is necessary for splicing. *Genes & Dev 11*:334–344.

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 X, Schwer B. 1997. Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Res 25*:2146–2152.

Zorio D, Lea K, Blumenthal T. 1997. Cloning of *Caenorhabditis* U2AF^{65}: An alternatively spliced RNA containing a novel exon. *Mol Cell Biol 17*:946–953.

Zorn A, Krieg P. 1997. The KH domain protein encoded by *quaking* functions as a dimer and is essential for notochord development in *Xenopus* embryos. *Genes & Dev 11*:2176–2190.