The EMBO Journal vol.12 no.12 pp.4715-4726, 1993

Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains

Javier F.Cáceres and Adrian R.Krainer[1]

Cold Spring Harbor Laboratory, PO Box 100, 1 Bungtown Road, Cold Spring Harbor, NY 11724-2208, USA

[1]Corresponding author

Communicated by I.W.Mattaj

Human pre-mRNA splicing factor SF2/ASF has an activity required for general splicing *in vitro* and promotes utilization of proximal alternative 5′ splice sites in a concentration-dependent manner by opposing hnRNP A1. We introduced selected mutations in the N-terminal RNA recognition motif (RRM) and the C-terminal Arg/Ser (RS) domain of SF2/ASF, and assayed the resulting recombinant proteins for constitutive and alternative splicing *in vitro* and for binding to pre-mRNA and mRNA. Mutants inactive in constitutive splicing can affect alternative splice site selection, demonstrating that these activities involve distinct molecular interactions. Specific protein–RNA contacts mediated by Phe56 and Phe58 in the RNP-1 submotif of the SF2/ASF RRM are essential for constitutive splicing, although they are not required for RRM-mediated binding to pre-mRNA. The RS domain is also required for constitutive splicing activity and both Arg and Ser residues are important. Analysis of domain deletion mutants demonstrated strong synergy between the RRM and a central degenerate RRM repeat in binding to RNA. These two domains are sufficient for alternative splicing activity in the absence of an RS domain.

Key words: alternative splicing/pre-mRNA splicing/RNA–protein interactions/RS domain/SF2/ASF

Introduction

Multiple protein and nucleoprotein components are necessary to catalyze pre-mRNA splicing and to select the correct splice sites. The components of spliceosomes, within which splicing takes place, include the small nuclear ribonucleoprotein particles (snRNPs) U1, U2 and U4/U5/U6, the polypeptides that associate with hnRNA to form hnRNP particles and a number of essential protein splicing factors (reviewed in Rio, 1992; Lamm and Lamond, 1993). In mammalian cells several essential splicing factors identified biochemically have been partially or extensively purified and cDNA clones have been obtained for SF2/ASF (Ge *et al.*, 1991; Krainer *et al.*, 1991), for the two subunits of U2AF (Zamore *et al.*, 1992; Zhang *et al.*, 1992), for SC35 (Fu and Maniatis, 1992a) and for PSF (Patton *et al.*, 1993). Sequence analysis showed that these splicing factors share conserved structural features, including one or more copies of the RNP-type RNA recognition motif or RNA binding domain (RRM or RBD), which is characteristic of many RNA binding proteins and, except in the case of PSF, an Arg/Ser-rich region (RS domain), which is also present in several genetically identified *Drosophila* splicing regulators.

SF2/ASF is an RNA binding protein that has an activity essential for constitutive pre-mRNA splicing *in vitro* and in addition strongly influences splice site selection in a concentration-dependent manner when multiple alternative 5′ splice sites are present (Ge and Manley, 1990; Ge *et al.*, 1991; Krainer *et al.*, 1990a,b). An activity that opposes SF2/ASF in the selection of alternative 5′ splice sites is also present in HeLa cell nuclear extract (Krainer *et al.*, 1990a; Harper and Manley, 1991) and has been shown to be identical to hnRNP A1 (Mayeda and Krainer, 1992). An excess of hnRNP A1 favors the utilization of distal 5′ splice sites when two competing 5′ splice sites are present, whereas high levels of SF2/ASF favor the use of proximal 5′ splice sites (relative to the 3′ splice site). Thus, the counteracting activities of SF2/ASF and hnRNP A1 are important determinants of 5′ splice site selection *in vitro* and the intracellular ratio of SF2/ASF to hnRNP A1 may influence splice site selection *in vivo* (Mayeda and Krainer, 1992).

A family of nuclear phosphoproteins associated with active sites of RNA polymerase II transcription and with amphibian B-snurposomes has been recently described (Roth *et al.*, 1991). It comprises immunologically and sequence-related proteins with approximate molecular masses of 20, 30, 40, 55 and 75 kDa. These proteins, which are present in vertebrate and invertebrate somatic tissues, contain RS domains and RRMs (Zahler *et al.*, 1992). SF2/ASF, also termed SRp30a, is a member of this family, which also includes SRp20 (known as X16 in mouse and RBP1 in *Drosophila*), SC35 (also known as SRp30b or PR264), SRp40, SRp55 (a *Drosophila* variant of which is known as B52) and SRp75. There is strong sequence conservation of the SR proteins among different species (Ayane *et al.*, 1991; Champlin *et al.*, 1991; Ge *et al.*, 1991; Krainer *et al.*, 1991; Roth *et al.*, 1991; Fu and Maniatis, 1992a; Kim *et al.*, 1992; Mayeda *et al.*, 1992; Vellard *et al.*, 1992; Zahler *et al.*, 1992, 1993b). Most, perhaps all, members of this family have similar activities *in vitro*, as shown by their ability to complement the same inactive splicing extract and to influence splice site selection, although not all proteins have been tested in both assays (Fu *et al.*, 1992; Kim *et al.*, 1992; Mayeda *et al.*, 1992; Zahler *et al.*, 1992, 1993a). Therefore, an important question is whether individual members of the SR family are functionally redundant, or whether each has unique specificities in constitutive and alternative splicing *in vivo*. Quantitative differences in the relative specific activities of some of these proteins in constitutive and alternative *in vitro* splicing assays have been recently reported (Kim *et al.*, 1992; Zahler *et al.*, 1993a).

The primary sequence of SF2/ASF includes several motifs and homologies that are good candidates for regions of the protein that may be responsible for its splicing and RNA binding activities. These regions include: (i) a 76 amino acid RRM near the N-terminus, which includes the well conserved

RNP-1 and RNP-2 submotifs; (ii) an eight amino acid peptide, EFEDPRDA, which is identical to a region of unknown function in the U1-70K snRNP polypeptide; (iii) a variable loop region preceding RNP-1, which corresponds to a region of the first RRM of the U1-A and U2-B″ snRNP polypeptides that is a major determinant of specific RNA binding (Scherly *et al.*, 1990; Bentley and Keene, 1991); and (iv) a C-terminal RS domain that is mainly composed of RS or SR dipeptides, including a continuous stretch of 16 alternating R and S residues (Ge *et al.*, 1991; Krainer *et al.*, 1991).

The RRM is a conserved region of ~80 amino acids, which is present in one or several copies in a large family of RNA binding proteins (Bandziulis *et al.*, 1989; Kenan *et al.*, 1991; Haynes, 1992; Nagai, 1992; Mattaj, 1993). Substitution of conserved RRM residues within the RNP-1 and RNP-2 submotifs of several RNA binding proteins, including basic and aromatic amino acids, can eliminate RNA binding. For example, substitution of Gly-Ser in place of residues Ile12-Tyr13 in RNP-2, or Arg52-Gly53 in RNP-1, abolishes binding of the U1-A polypeptide to stem–loop II of U1 snRNA (Scherly *et al.*, 1989). Similarly, substitution of Arg143 with Pro, or Gly144 with Ser in RNP-2, eliminates binding of U1-70K polypeptide to stem–loop I of U1 snRNA (Surowy *et al.*, 1989). Basic residues within the RRM, but outside RNP-1 and RNP-2, are also important for U1-A binding (Nagai *et al.*, 1990), and in some cases sequences flanking an RRM contribute to RNA binding, probably by providing additional RNA contacts or by stabilizing the RRM structure (Kenan *et al.*, 1991).

A structural model for the N-terminal RRM of the U1-A snRNP protein has been proposed based on X-ray diffraction (Nagai *et al.*, 1990) and NMR studies (Hoffman *et al.*, 1990). It consists of four antiparallel β-strands and two α-helices, with RNP-1 and RNP-2 submotifs lying adjacent on the central β1 and β3 strands. This RRM model structure includes a hydrophobic core of conserved residues, which is probably involved in assuming and maintaining the correct tertiary structure, and several solvent-exposed aromatic residues that are probably involved in RNA binding. The high conservation of correctly spaced hydrophobic residues among members of this family of proteins suggests that other RRM proteins have a similar tertiary structure (Kenan *et al.*, 1991), a prediction borne out by the NMR structure of hnRNP C (Wittekind *et al.*, 1992).

Whereas the highly conserved RNP-1 and RNP-2 submotifs are thought to provide a general RNA binding surface common to all RRMs, the specificity of binding is probably mediated by non-conserved solvent-exposed residues present in the β strands and the linking loop regions. Sequences upstream of RNP-1, in variable loop-3 of the first two RRMs in U1-A and U2-B″, are partly responsible for the specificity of binding of these proteins to U1 and U2 snRNAs, respectively. Thus, the exchange of short stretches of amino acids (within loop-3 and β2) between U1-A and U2-B″ reversed the binding specificity of each RRM (Scherly *et al.*, 1990; Bentley and Keene, 1991). However, these two polypeptides have an overall homology of 85%, which vastly exceeds the similarity found between unrelated RRMs. Thus, in other RRMs, additional residues are likely to contribute to sequence-specific binding.

The octapeptide EFEDPRDA, which overlaps the RNP-1 submotif of the SF2/ASF RRM, is also present in human and *Xenopus laevis* U1-70K polypeptides, but in this case it is found upstream of the RRM, in a region not required for RNA binding (Etzerodt *et al.*, 1988; Query *et al.*, 1989). It is also present, with occasional substitutions, at homologous positions in proteins of the SR family (Fu and Maniatis, 1992a; Mayeda *et al.*, 1992; Zahler *et al.*, 1992, 1993b) and in *Drosophila* U1-70K (Mancebo *et al.*, 1990), but not in the *Saccharomyces cerevisiae* U1-70K homolog (Smith and Barrell, 1991). The function of this octapeptide is unknown, but it is not found in any other proteins or open reading frames in the current databases.

RS domains similar to that of SF2/ASF are present in the *Drosophila* proteins encoded by the *tra*, *tra2* and *su(w^a)* genes, all of which are involved in the regulation of alternative splicing pathways (reviewed in Bingham *et al.*, 1988; Mattox *et al.*, 1992) and also in the U1-70K snRNP polypeptide (Mancebo *et al.*, 1990), in other members of the SR protein family (Zahler *et al.*, 1992) and in both subunits of the splicing factor U2AF (Zamore *et al.*, 1992; Zhang *et al.*, 1992). RS domains may mediate binding to RNA by electrostatic interactions, by analogy to the Arg-rich motif found in HIV regulatory proteins and in bacterial transcription antiterminators (Lazinski *et al.*, 1989). The RS domain of su(w^a) is responsible for its *in vivo* localization to the nucleoplasmic speckled region and can be replaced by the tra RS domain (Li and Bingham, 1991). In the case of U2AF^{65}, deletion of the RS domain abolishes complementation for splicing in U2AF-depleted extract but does not affect RNA binding (Zamore *et al.*, 1992).

To carry out a detailed structural and functional analysis of SF2/ASF and to identify domains of this protein that are essential for constitutive splicing activity, for 5′ splice site selection and for RNA binding, we introduced selected mutations in specific domains of SF2/ASF. We expressed wild type (wt) and mutant proteins in *Escherichia coli* and then analyzed the purified proteins for splicing complementation and 5′ splice site selection *in vitro*, as well as for RNA binding. We report that both the RRM and the RS domain are essential for the constitutive splicing activity of SF2/ASF. In addition, we show that SF2/ASF mutant proteins that are inactive in constitutive splicing can affect alternative 5′ splice site selection. Therefore, constitutive and alternative splicing activities map to different regions of SF2/ASF. Finally, we describe the domains of SF2/ASF that are responsible for mediating alternative 5′ splice site selection and binding to RNA.

Results

Design of SF2/ASF mutants

An initial set of nine SF2/ASF mutant proteins was designed in such a way as to minimize structural or folding perturbations. Thus, we targeted individual domains of SF2/ASF by substituting specific amino acid residues (Figure 1), rather than by deleting small regions or entire domains, which could have a profound effect on protein structure. The high conservation of hydrophobic residues among known RRMs suggests that the tertiary structure described for RRM1 of the U1-A snRNP protein (Hoffman *et al.*, 1990; Nagai *et al.*, 1990) and more recently for hnRNP C (Wittekind *et al.*, 1992), is largely conserved. Hence, based on the coordinates of the U1-A RRM1 crystal structure (Nagai *et al.*, 1990), we were able to predict whether a particular

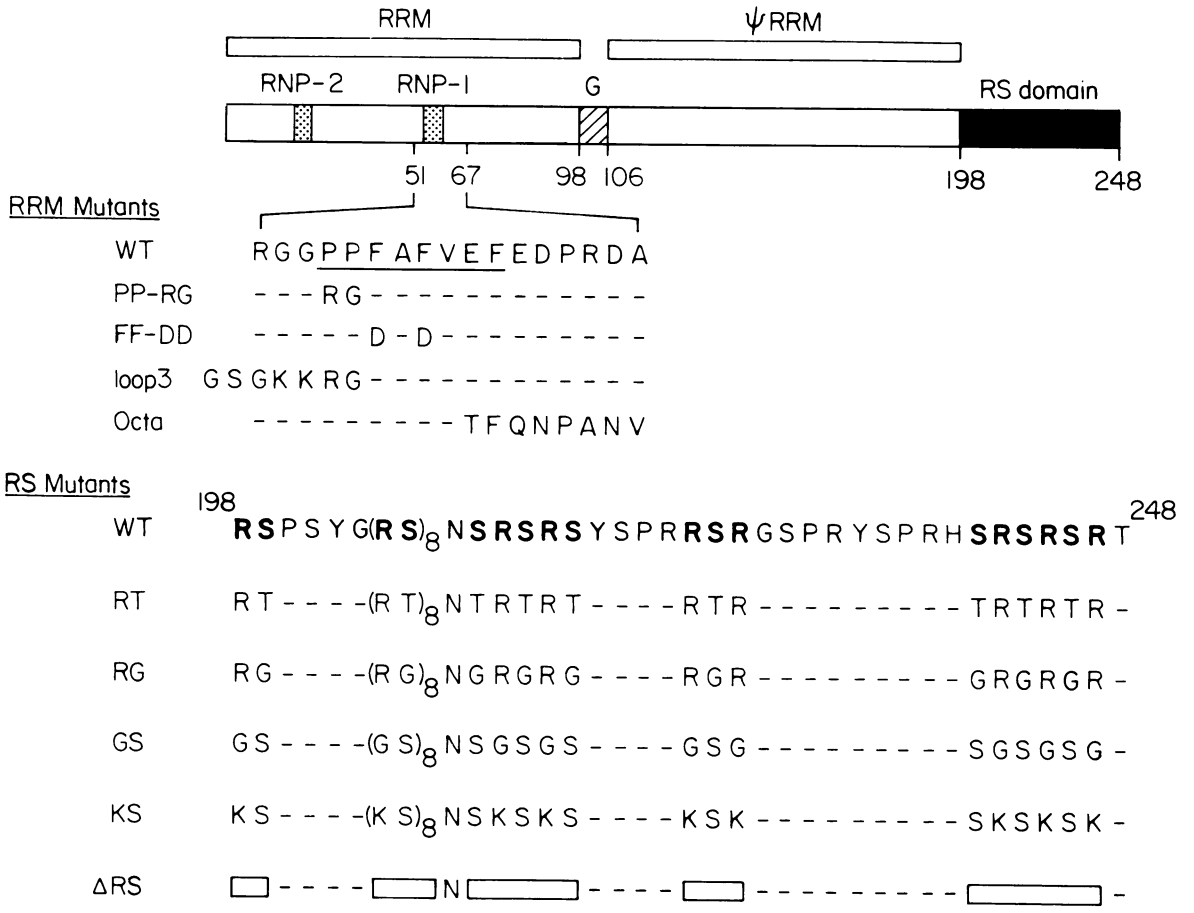

Fig. 1. Mutations in the RRM and RS domain of SF2/ASF. The RNP-2 and RNP-1 submotifs of the RRM, the Gly-rich hinge and a degenerate RRM repeat (ψRRM) are indicated in the diagram and described in the text. Changes of specific residues relative to the SF2/ASF wt amino acid sequence are displayed. The dashes indicate positions in which no changes were introduced. The open boxes in the ΔRS mutant indicate residues that were deleted. All proteins contain the same N-terminal His tag derived from the expression vector.

amino acid change within the RRM of SF2/ASF would affect the overall structure of the protein (data not shown).

N-terminal RRM.

(i) *PP-RG mutant.* The RNP-1 submotif of SF2/ASF deviates from the consensus at the first two positions: Pro54-Pro55 are present instead of the consensus Arg-Gly. As discussed above, each of these residues is critical for sequence-specific binding by U1-A and U1-70K polypeptides (Scherly *et al.*, 1989; Surowy *et al.*, 1989). As shown in Figure 1, we substituted Pro54-Pro55 in RNP-1 of SF2/ASF with the consensus Arg-Gly dipeptide present in most known RRMs (Kenan *et al.*, 1991; Krainer *et al.*, 1991).
(ii) *FF-DD mutant.* The structural model of the U1-A RRM (Hoffman *et al.*, 1990; Nagai *et al.*, 1990) predicts that solvent-exposed aromatic residues on β1 and β3 strands mediate RNA binding through ring-stacking interactions with the bases. In the hnRNP A1 protein, Phe residues in RNP-1 and RNP-2 of both RRMs crosslink to oligo(dT) (Merrill *et al.*, 1988). In an attempt to inactivate the RRM of SF2/ASF, we substituted Phe56 and Phe58 in RNP-1 with Asp residues. These acidic residues should remain solvent-exposed but are expected to interact very differently with the bases, or to repel the phosphate backbone of RNA.
(iii) *Loop3 mutant.* We tested whether the function and RNA binding properties of SF2/ASF could be altered by replacing the variable loop region upstream of RNP-1 by the variable loop of a different RRM. We arbitrarily chose the variable loop of the second RRM of hnRNP A1, in part because the RNP-1 submotifs of these two proteins are similar. The peptide RGGPP (residues 51–55), which includes the two RNP-1 Pro residues changed in the PP-RG mutant, was replaced by the longer peptide GSGKKRG from hnRNP A1.
(iv) *Octa mutant.* The octapeptide EFEDPRDA (residues 60–67), which is conserved in several SR proteins and in U1-70K, was changed to TFQNPANV. The substitutions were designed by taking into account the expected location of the octapeptide within the three-dimensional model RRM structure and attempting to preserve putative intramolecular protein–protein interactions, while eliminating the charged character of this sequence.

C-terminal RS domain.

ΔRS, RT, RG, GS and KS mutants. We deleted all the Arg and Ser residues in the RS domain of SF2/ASF (residues 198–248), while retaining other potentially important amino acids present within this domain, as illustrated in Figure 1. In addition, we substituted either Arg or Ser residues in the RS domain to generate RT, RG, GS and KS mutants. Thus, we were able to analyze the requirement for the RS domain in constitutive and alternative splicing and in RNA binding, as well as the specific requirements for Arg and Ser residues,

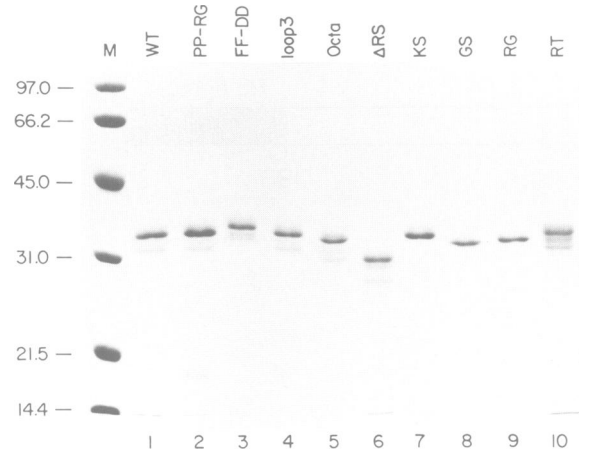

Fig. 2. Purification of recombinant wt and mutant SF2/ASF proteins. His-tagged proteins overproduced in *E.coli* were purified by affinity chromatography on a Ni^{2+}–NTA resin and elution with imidazole. The purified proteins were analyzed by 12% SDS–PAGE and Coomassie blue R250 staining. The relative mobilities of the molecular weight markers are indicated (lane M).

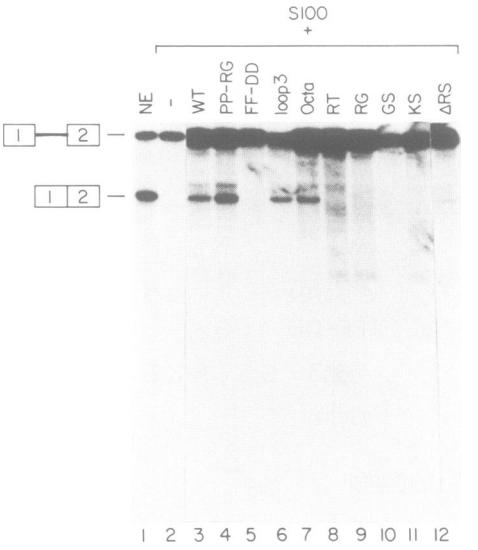

Fig. 3. Effect of SF2/ASF mutations on constitutive splicing *in vitro*. The constitutive splicing activity of wt and mutant SF2/ASF proteins was assayed by biochemical complementation with a HeLa S100 extract that contains all other general splicing factors. Pre-mRNA containing the first two exons and first intron of wt human β-globin was spliced *in vitro* with HeLa nuclear extract (NE) or with HeLa cytoplasmic S100 extract and recombinant SF2/ASF proteins. 0.2 μg of wt or mutant SF2/ASF in 5 μl were used to complement 7 μl of S100 extract. The products were analyzed by PAGE–urea and autoradiography. The structures and relative mobilities of the pre-mRNA and spliced mRNA are indicated at left. The band just above spliced mRNA is the lariat–exon 2 intermediate.

positive charge, hydroxyl groups and potential sites of phosphorylation by Ser/Thr protein kinases. Only four RS or SR dipeptides present outside the C-terminal RS domain are retained in these mutants, at positions 90–91, 116–119 and 181–182 (Krainer *et al.*, 1991).

Expression and purification of SF2/ASF mutant proteins

Mutations were introduced into cloned wt SF2/ASF cDNA by replacing selected restriction fragments with synthetic dsDNA or PCR fragments containing the desired mutations and appropriate restriction sites (see Materials and methods). wt and mutant SF2/ASF cDNAs were subcloned into pET-19b, a T7 vector designed for inducible expression of fusion proteins in *E.coli* (Studier *et al.*, 1990; see Materials and methods). The fusion proteins contain 10 His residues at the N-terminus, allowing a simple purification by metal chelate affinity chromatography (Hochuli *et al.*, 1987). This vector was chosen because expression of SF2/ASF with an authentic N-terminus, e.g. in pET-9c, is inefficient (Krainer *et al.*, 1991). The level of induction of the His-tagged SF2/ASF wt and mutant proteins was very high in all cases (data not shown). However, all of the above recombinant proteins produced in this way formed inclusion bodies and therefore they were denatured in guanidine hydrochloride and renatured after affinity purification (see Materials and methods; Ge *et al.*, 1991). We note that in comparing the activities of the different mutants, the results should not be interpreted in a strict quantitative manner, because slight differences may be attributed to variability in the process of renaturation. Figure 2 shows the SDS–PAGE analysis of recombinant SF2/ASF wt and mutant proteins after purification and renaturation. In most cases, amino acid substitutions caused slight mobility shifts that can be attributed to small variations in SDS binding.

Functional assays of SF2/ASF mutant proteins

Constitutive splicing. SF2/ASF mutant proteins were assayed for general splicing activity using human β-globin pre-mRNA and complementation with a cytosolic S100 HeLa extract (Krainer and Maniatis, 1985; Krainer *et al.*, 1990b). This extract provides all components required for splicing except SF2/ASF activity and is therefore a useful system to study domains of SF2/ASF that are essential for constitutive splicing. Recombinant wt SF2/ASF protein can restore splicing activity to the inactive S100 extract (Figure 3, lanes 2 and 3; Ge *et al.*, 1991; Krainer *et al.*, 1991). Surprisingly, the PP-RG, loop3 and Octa mutants also complemented S100 extract efficiently (lanes 4, 6 and 7), demonstrating that none of the residues affected by these mutations is essential for the constitutive splicing activity of SF2/ASF.

In contrast, substitution of Phe56 and Phe58 in the RNP-1 submotif of SF2/ASF completely abolished the splicing activity of the FF-DD mutant protein (lane 5). According to current RRM structural models, these two aromatic residues in RNP-1, together with a Tyr or Phe residue in RNP-2, form a cluster of solvent-exposed aromatic residues involved in RNA binding. This experiment demonstrates that the exposed aromatic residues in RNP-1, which are predicted to be involved in protein–RNA interactions, are essential for the constitutive splicing activity of SF2/ASF.

All of the RS mutants (RT, RG, GS, KS and ΔRS) were also essentially inactive in splicing complementation (lanes 8–12). In some experiments we have occasionally detected traces of activity with certain preparations of GS and ΔRS mutants (data not shown). Hence, we conclude that mutations in the RS domain greatly diminish, if not completely abolish, the splicing activity of SF2/ASF. Both kinds of amino acid side-chains within the RS domain appear to be specifically required for constitutive splicing, as substitution of the Arg residues for Lys or Gly, or of the Ser residues for Thr or Gly resulted in loss of function.

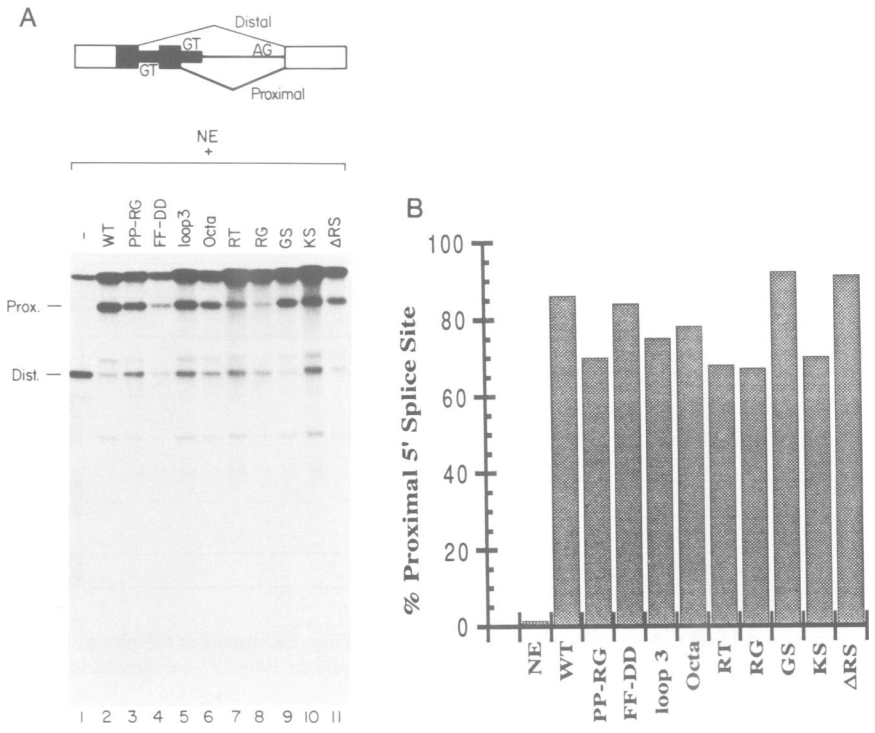

Fig. 4. Effect of SF2/ASF mutations on alternative splicing *in vitro*. The alternative splicing activity of wt and mutant SF2/ASF proteins was assayed by measuring their ability to switch splice site use in nuclear extract from the distal to the proximal 5′ splice site of a model β-globin pre-mRNA with duplicated 5′ splice sites. The products were analyzed by PAGE−urea and autoradiography. The relative mobilities of spliced mRNAs generated by proximal and distal 5′ splice site use are indicated at left. (**A**) 5 μl of NE were supplemented with 5 μl (0.2 μg) of the indicated recombinant SF2/ASF wt or mutant proteins. (**B**) The same gel was scanned with a Fujix/Bas 2000 phosphor imager. The radioactivity present in each spliced mRNA band corresponding to proximal (P) and distal (D) 5′ splice site use was determined and after background subtraction the percentage of proximal 5′ splice site use was calculated for each reaction as 100 × P/(P + D). The small differences between mutants are probably not significant.

In summary, we conclude that both the N-terminal RRM and the C-terminal RS domain are essential for the constitutive splicing activity of SF2/ASF. Intact versions of each region are required within a single polypeptide, since we failed to observe intra-cistronic complementation between separate FF-DD and RS mutants (data not shown).

Alternative splicing. We analyzed the same SF2/ASF mutants for alternative splicing activity, using a model β-globin pre-mRNA that has duplicated 5′ splice sites (Reed and Maniatis, 1986; Krainer *et al.*, 1990a). When this substrate is spliced in HeLa cell nuclear extract, the distal 5′ splice site is preferentially selected, reflecting the ratio of SF2/ASF to hnRNP A1 in this extract (Mayeda and Krainer, 1992) (Figure 4A, lane 1). However, when nuclear extract is supplemented with excess wt SF2/ASF, we observe the expected switch from distal to proximal 5′ splice site (Figure 4A, lane 2) (Krainer *et al.*, 1990a, 1991). Thus, this assay can be used to map the domains of SF2/ASF that are responsible for selecting proximal alternative 5′ splice sites by counteracting the activity of hnRNP A1. In this assay two populations of SF2/ASF are present: (i) the endogenous wt SF2/ASF already present in the extract, which provides the essential activity required for constitutive splicing, but which is not present in sufficient amounts to counteract distal 5′ splice site activation by hnRNP A1; and (ii) the exogenous SF2/ASF mutant proteins.

Surprisingly, all of the above SF2/ASF mutants were active in the alternative splicing assay (Figure 4A, lanes 3−11). In particular, SF2/ASF mutants that were unable to complement S100 extract (FF-DD, RT, RG, GS, KS and ΔRS; see Figure 3) were nevertheless able to stimulate proximal 5′ splice site selection (Figure 4A, lanes 4 and 7−11). Alternative splice site selection was also assayed in the absence of detectable endogenous wt SF2/ASF activity by splicing the same substrate in S100 extract. SF2/ASF wt, PP-RG, loop3 and Octa proteins complemented the extract and selected the proximal 5′ splice site exclusively, whereas in the presence of the remaining mutant proteins no splicing was observed, in agreement with the above results (data not shown).

The overall splicing efficiency and pre-mRNA and mRNA stabilities varied in reactions with different mutant proteins. We believe that these differences at least partly reflect the different degrees to which some of the mutants compete with endogenous wt SF2/ASF through the different steps of spliceosome assembly and catalysis. However, the ratio of proximal to distal 5′ splice site utilization is an indicator of the relative activities of the different mutants in alternative splicing. By this measure all the SF2/ASF mutants showed comparable activity with the wt (Figure 4A and B). This experiment shows that inactivation of domains essential for constitutive splicing (RRM and RS domain) does not affect alternative splicing activity, and therefore the two splicing activities of SF2/ASF map to different regions of the protein.

To search further for a region of SF2/ASF that may be necessary and perhaps sufficient for alternative splice site switching, we constructed additional mutants. First we constructed double mutants to address the possibility that the RRM and the RS domain have redundant functions in

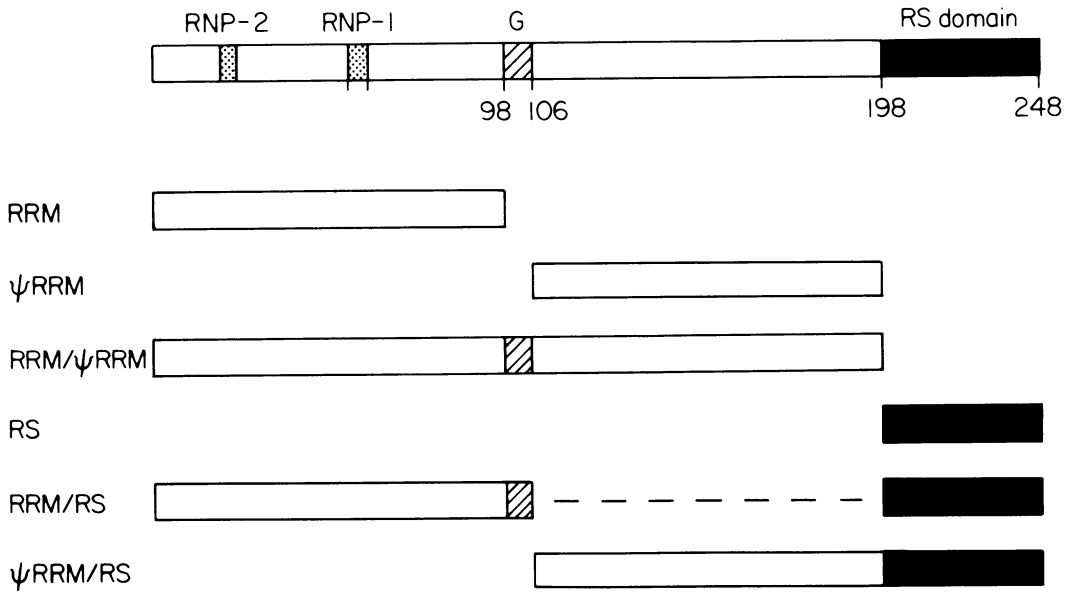

Fig. 5. Structure of SF2/ASF domain deletion mutants. The open and solid boxes indicate the regions of the protein present in each mutant, relative to the 248 amino acid wt structure shown at the top. The dashed line indicates that residues 106–197 are missing in the RRM/RS protein. All proteins contain the same N-terminal His tag derived from the expression vector.

alternative splicing. The following double mutants were constructed by combining some of the single mutants described above: (i) FF-DD/ΔRS combines FF-DD and ΔRS mutations; (ii) FF-DD/GS combines FF-DD and GS mutations.

In addition, we constructed SF2/ASF fragments encompassing individual domains of the protein or combinations thereof (Figure 5). The deletion boundaries were selected on the basis of the modular primary structure of SF2/ASF, which includes the N-terminal RRM, a Gly hinge consisting of nine consecutive Gly residues, a degenerate RRM (designated ψRRM) and the RS domain (Ge *et al.*, 1991; Krainer *et al.*, 1991). The Gly hinge probably serves to provide flexibility between the RRM and the rest of the protein. The ψRRM is an atypical RRM-like repeat present in some but not all SR proteins (Zahler *et al.*, 1992), which may represent a duplicated RRM whose sequence and function diverged during evolution. This ψRRM lacks the conserved aromatic residues present in RNP-1 and RNP-2, but contains all of the conserved amino acids that form the hydrophobic core of an RRM (E.Birney, S.Kumar and A.R.Krainer, submitted). The structures of the domain deletion proteins designated RRM, ψRRM, RRM/ψRRM, RS, RRM/RS and ψRRM/RS are shown in Figure 5. The RRM/ψRRM protein is similar to the ΔRS protein (Figure 1), but lacks the entire RS domain (residues 198–248), including 19 residues of unknown function interspersed with the Arg-Ser dipeptides.

The double mutant and domain deletion proteins were expressed in the pET-19b system and affinity-purified as described above. The domain deletion proteins were efficiently expressed in soluble form and were affinity-purified under native conditions (data not shown; Hoffmann and Roeder, 1991).

When the FF-DD/ΔRS and FF-DD/GS double mutant proteins were used to supplement nuclear extract, we observed that each protein retained about half of the switching activity of the wt protein (Figure 6A, lanes 1, 3 and 4, Figure 6B). Since individual FF-DD, ΔRS or GS mutations do not affect alternative splicing activity (Figure 4), whereas different preparations of the double mutant proteins are reproducibly less active than wt SF2/ASF, it appears that the RS domain and the Phe residues in the RRM may play a role, albeit limited, in alternative splicing but that these roles are redundant.

When the domain deletion proteins were used to supplement nuclear extract, we observed that the alternative splicing activity was affected in all cases (Figure 6A, lanes 5–10) except for the RRM/ψRRM protein (lane 11). Splice site switching activity, estimated as the ratio of proximal to distal 5′ splice site utilization relative to that in unsupplemented nuclear extract (lane 2), was severely decreased with RRM and ψRRM proteins (lanes 8 and 9) and almost absent with RRM/RS, ψRRM/RS and RS proteins (lanes 5–7; Figure 6B). Therefore, individual RRM, ψRRM or RS domain fragments of SF2/ASF are not sufficient for efficient alternative splicing activity and both the RRM and ψRRM regions appear to be important. Since the RRM/ψRRM protein, which consists of the RRM and the ψRRM separated by a flexible Gly hinge, was fully active in switching (Figure 6A, lane 11), we conclude that the RRM and the ψRRM, or elements embedded within these domains, act synergistically to promote proximal 5′ splice site selection. This synergy requires that the RRM and the ψRRM be present in the same polypeptide, since a mixture of the separate domains is no more active than the individual domains (Figure 6A, lane 10, Figure 6B).

RNA binding properties of SF2/ASF mutants

The intrinsic RNA binding properties of wt and mutant SF2/ASF in the absence of other splicing factors were investigated by two complementary techniques: (i) UV light-induced crosslinking between radiolabeled RNA and purified protein (Greenberg, 1980) and (ii) retention of protein-bound radiolabeled RNA on nitrocellulose filters (Riggs *et al.*, 1970). In both cases RNA and protein were pre-incubated under splicing buffer conditions for 30 min,

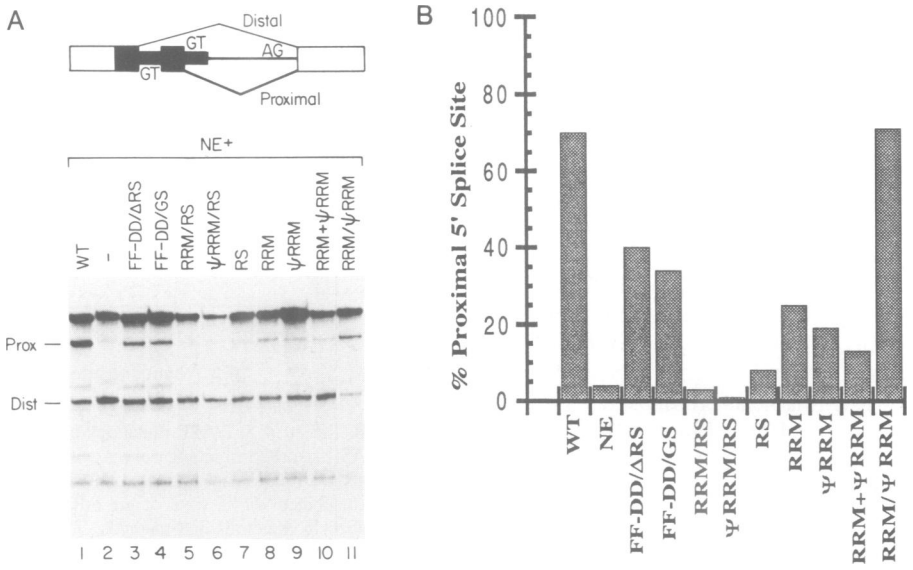

Fig. 6. Effect of SF2/ASF double mutants and domain deletions on alternative splicing *in vitro*. The FF-DD/△RS and FF-DD/GS double mutants combine the FF-DD mutation with the ΔRS and GS mutations, respectively. The alternative splicing activity of wt and mutant SF2/ASF proteins was assayed as described in Figure 4. (**A**) 5 μl of nuclear extract were supplemented with 5 μl (0.2 μg) of the indicated mutant protein and the spliced products were analyzed by PAGE−urea and autoradiography. In lane 10 equal amounts of RRM and ψRRM fragments were used. (**B**) Phosphor image analysis of the same gel. The percentage of proximal 5′ splice site use was calculated as described in Figure 4B.

which was determined to be sufficient for equilibrium binding (data not shown).

UV crosslinking. SF2/ASF has general RNA binding activity, as indicated by the observation that purified SF2/ASF can be UV crosslinked to β-globin pre-mRNA, mRNA and anti-sense β-globin RNA (Krainer *et al.*, 1990b; data not shown). Crosslinking efficiency is a function of binding and of the proximity of reactive bases and amino acid side chains. One limitation of the UV crosslinking approach is that quantitative comparisons are affected by differential photoreactivities of individual amino acid side chains and bases at a given wavelength (Smith, 1976). On the other hand, an advantage of this method is that the identity of the RNA binding polypeptide is clearly established. Moreover, weak or transient interactions of possible functional significance can be detected.

The crosslinking efficiency of SF2/ASF to β-globin pre-mRNA was decreased in the case of the FF-DD mutant and enhanced in the case of the RS domain mutants (Figure 7A, lanes 2, 4 and 7−11). The RS domain probably binds to RNA by electrostatic interactions between the Arg side chains and the phosphate backbone. RS-mediated RNA binding may compete or interfere with RRM-mediated RNA binding (by the same or by a different molecule of SF2/ASF) and thus prevent efficient crosslinking via conserved aromatic residues in the RNP-1 and RNP-2 submotifs of the RRM. However, all the RS domain substitution mutants showed increased crosslinking efficiency compared with the wt, including those that still have Arg or basic residues (RG, RT, KS) and those that are uncharged (GS, ΔRS). These results imply that binding to RNA is not simply through electrostatic interactions and/or that the residues interspersed with the Arg residues influence the conformation of the domain and its ability to bind to the phosphate backbone. The intact RS domain may also be involved in intra- or intermolecular protein−protein interactions that mask the

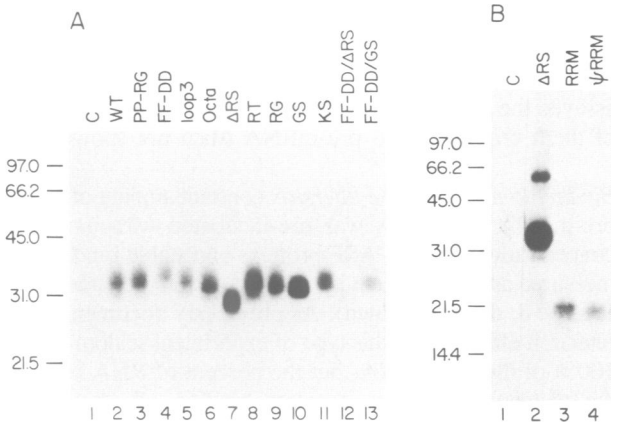

Fig. 7. Effect of SF2/ASF mutations on UV light-induced crosslinking to pre-mRNA. (**A**) ^{32}P-Labeled human β-globin pre-mRNA was incubated for 30 min with nuclease-free BSA (lane 1) or with the indicated purified recombinant SF2/ASF proteins (lanes 2−13) and irradiated at 254 nm. After digestion with RNase A, the products were analyzed by 12% SDS−PAGE and autoradiography. (**B**) As above but the labeled substrate was incubated with RRM and ψRRM protein fragments (or ΔRS protein as a control) and the products were resolved by 15% SDS−PAGE. The slow migrating band in lane 2 is probably a protein dimer arising either from protein−protein crosslinking or from incomplete digestion of an RNA fragment bound to two protein molecules.

rest of the protein. Each of the above RS domain mutations, which also severely inhibit constitutive splicing activity, may open up the conformation of the protein, allowing more efficient RRM−RNA contacts that lead to crosslinking.

Whereas mutation of Phe56 and Phe58 in RNP-1 decreases crosslinking activity only slightly (lane 4) and mutations in the RS domain enhanced crosslinking (lanes 7−11), the combination of both mutations in the double mutants FF-DD/ΔRS and FF-DD/GS drastically reduced the efficiency of crosslinking to RNA (lanes 12 and 13). Thus, in the

wt protein, RS domain-mediated RNA binding appears to mask RRM-mediated RNA binding and therefore, the importance of Phe56 and Phe58 for crosslinking to RNA becomes evident only after the RS domain is either removed (FF-DD/ΔRS) or mutated (FF-DD/GS). These experiments suggest that RNA binding is independently mediated by elements in the RRM and in the RS domain.

When β-globin mRNA was used as a substrate for crosslinking of the wt and mutant proteins, identical results were obtained (data not shown), indicating that individual RNA binding domains in SF2/ASF lack strong specificity for intronic sequences, at least in the absence of other components.

Both RRM and ψRRM fragments crosslinked very weakly to β-globin pre-mRNA or mRNA (Figure 7B, lanes 3 and 4; data not shown). The efficiency of RNA crosslinking was slightly higher for the RRM fragment, but this experiment shows that at least some binding to RNA can be mediated by the ψRRM, despite its atypical character as an RRM, compared with control proteins (including some His-tagged proteins) that do not bind RNA (lane 1; data not shown). The efficiency of RNA crosslinking by ΔRS protein or by RRM/ψRRM protein, is much greater than the sum of the crosslinking efficiencies of separate RRM and ψRRM fragments (lane 2; data not shown). Since the same amino acid residues are available for crosslinking, in particular the conserved aromatic residues in RNP-2 and RNP-1 submotifs of the RRM, this result demonstrates synergy between the RRM and the ψRRM in binding to RNA. In addition, we assayed the RS, RRM/RS and ψRRM/RS proteins, but none of them crosslinked to pre-mRNA (data not shown).

Binding to nitrocellulose filters. A constant amount of labeled pre-mRNA or mRNA was pre-incubated with increasing concentrations of SF2/ASF proteins and stable binding was measured as the radioactivity retained on nitrocellulose filters (Figure 8; data not shown). As previously documented, the retention efficiency in this type of experiment seldom reaches 100% of the input RNA, but the percent of RNA bound at the plateau represents complete binding of active RNA (Carey *et al.*, 1983).

Whereas wt and FF-DD proteins showed comparable binding efficiencies, binding of ΔRS and of FF-DD/ΔRS was decreased to the same extent. The decreased binding efficiency of ΔRS compared with wt, and of FF-DD/ΔRS compared with FF-DD, demonstrates that the RS domain mediates RNA binding. The binding efficiencies of FF-DD/ΔRS and ΔRS are similar, but the double mutant showed greatly reduced crosslinking (Figure 7A). Therefore, Phe56 and Phe58 are predominant sites of crosslinking, as expected (Merrill *et al.*, 1988), but surprisingly these residues are not required for general RNA binding involving the N-terminal RRM of SF2/ASF. However, the loss of constitutive splicing activity in the FF-DD mutants suggests that specific protein–RNA contacts mediated by the solvent-exposed Phe56 and Phe58 residues are essential for function.

The RRM and ψRRM protein fragments showed virtually no binding to RNA, whereas the ΔRS mutant exhibited much stronger binding. These data are consistent with the UV crosslinking results and demonstrate synergy between the RRM and the ψRRM in binding to RNA. The RS, RRM/RS and ψRRM/RS proteins showed comparable binding with pre-mRNA in the filter binding assay, at a level slightly below that of the ΔRS protein (data not shown), confirming that the RS domain can bind RNA. Since these proteins have very weak or no alternative splicing activity, we conclude that RNA binding *per se* is not sufficient for the alternative splicing activity of SF2/ASF.

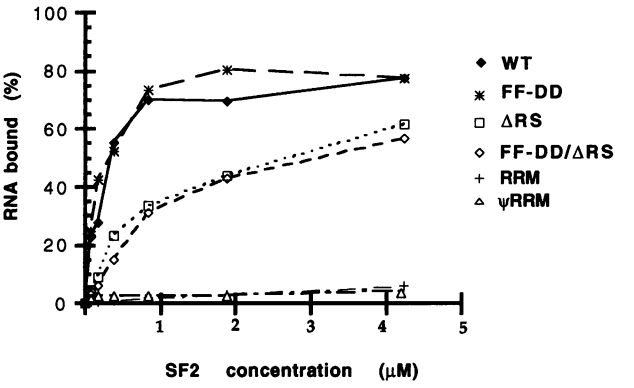

Fig. 8. Effect of SF2/ASF mutations on binding to pre-mRNA. SF2/ASF saturation binding curves with ^{32}P-labeled human β-globin pre-mRNA. Increasing amounts of SF2/ASF wt or mutant proteins were incubated with 5 fmol of pre-mRNA for 30 min at 30°C, as described in Materials and methods. The reactions were filtered through nitrocellulose membranes and the percentage of input label remaining on the filters was determined.

Discussion

The modular multidomain primary structure of SF2/ASF suggests that this factor is capable of establishing multiple interactions with one or more RNAs, which may account for its biochemical activities in constitutive and alternative splicing. By analyzing the biochemical properties of mutations in single and multiple domains we have defined distinct regions of the protein involved in each reaction and in RNA binding (Figure 9). We found that the N-terminal RRM is essential for constitutive splicing and that the C-terminal RS domain is also very important for this activity. Although we were unable to identify a single specific motif or region of SF2/ASF required for alternative 5′ splice site selection, we found that the RRM and the ψRRM act synergistically and are sufficient for this activity. We obtained several mutants that were inactive in constitutive splicing but active in alternative splicing, indicating that the two functions involve distinct molecular interactions. These results argue against the possibility that the alternative splicing activity of SF2/ASF is a reflection of its essential splicing activity, somehow related to the concentration dependence. It is possible that the same molecular interactions involved in alternative splicing are necessary but not sufficient for constitutive splicing activity.

Substitution of Phe56 and Phe58 in the RNP-1 submotif of the N-terminal RRM abolished the constitutive splicing activity of SF2/ASF, strongly suggesting that protein–RNA contacts mediated by the RRM are required for this activity. These two conserved aromatic residues present on the surface of the prototype RRM are thought to interact directly with the RNA ligand by ring stacking interactions with the bases (Merrill *et al.*, 1988; Nagai *et al.*, 1990; Kenan *et al.*, 1991). The same mutation had no effect on alternative splicing in the context of the intact protein, but did partially affect

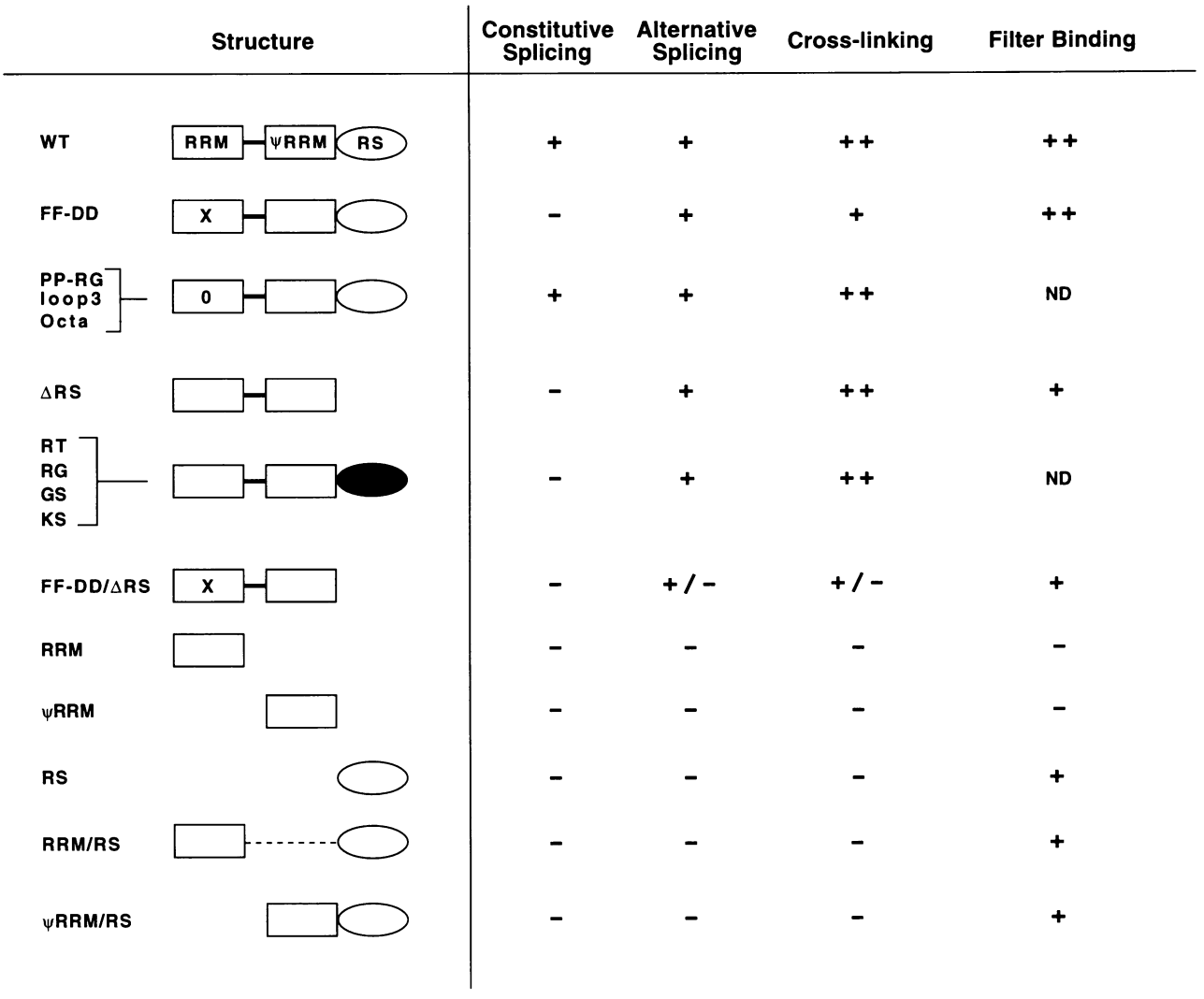

Fig. 9. Summary of the effects of SF2/ASF mutations. The structures of the wt and mutant proteins are shown schematically at left. The X symbolizes mutation of Phe56 and Phe58 in the RNP-1 submotif to Asp residues. The 0 represents each of three mutations of the RNP-1 submotif and/or adjacent residues, which have similar phenotypes: Pro54-Pro55 to Arg-Gly (PP-RG), replacement of the RRM variable loop (loop3) and substitutions within the octapeptide motif (Octa). The black ellipse denotes each of four mutations within the RS domain, which have similar phenotypes: replacement of all Ser residues by Thr (RT) or Gly (RG), or of all Arg residues by Gly (GS) or Lys (KS). The effects of single and double mutations and domain deletions of SF2/ASF on constitutive splicing, alternative 5′ splice site selection, UV crosslinking to RNA and retention of RNA on nitrocellulose filters are indicated. Mutants giving <25% proximal 5′ splice site use, as defined in Figures 4B and 6B, were considered negative in the alternative splicing assay. ND, not determined.

this reaction when the RS domain was also mutated. The double mutant bound pre-mRNA as well as the RS domain mutants, indicating that Phe56 and Phe58 are not the major determinants of general RRM-mediated RNA binding. However, the loss of crosslinking efficiency confirms the expectation that Phe56 and Phe58 contact RNA. It appears that for alternative but not for constitutive splicing, the absence of protein−RNA contacts mediated by Phe56 and Phe58 can be partially compensated by the presence of the RS domain, which may provide multiple contacts with RNA. Otherwise the N-terminal RRM appears to function in conjunction with the central ψRRM to bind RNA and stimulate proximal 5′ splice site use, and in this instance the RS domain is redundant. We also found that an RS domain by itself is unable to influence alternative 5′ splice site selection.

Mutations in the RS domain severely impaired the general splicing activity of SF2/ASF and in the context of a weakened RRM also partially affected its alternative splicing activity. We found that both Arg and Ser side-chains are specifically required for RS domain function, although we have not determined a minimum number of required residues or whether the alternating sequence, as opposed to the overall Arg/Ser content, is important. The fact that the RS domain is required for constitutive splicing activity *in vitro* demonstrates that in the case of SF2/ASF the function of this domain is not limited to determining subnuclear localization, a function that was demonstrated for the RS domains of *su(w^a)* and *tra* gene products (Li and Bingham, 1991). In the case of U2AF^{65}, deletion of the N-terminal RS domain eliminated complementation of U2AF-depleted nuclear extract but did not affect RNA binding (Zamore *et al.*, 1992). This splicing factor also contains three C-terminal RRMs and deletion of the first or third of these abolished both RNA binding and splicing activity (Zamore *et al.*, 1992).

We showed that specific protein–RNA contacts mediated in a synergistic manner by the RRM and the ψRRM, rather than overall RNA binding, are important for the alternative splicing activity of SF2/ASF. It remains possible that there is a specific region of SF2/ASF, overlapping with, but distinct from, the RRM and ψRRM, which may be required for alternative splicing activity. Such a region may mediate additional contacts with RNA, or protein–protein interactions, e.g. with hnRNP A1, snRNPs or with other SR proteins. We would expect that such a region would be recognizable as a conserved motif shared by all of the SR proteins that have been shown to affect 5′ splice site selection (see below). The octapeptide motif found in U1-70K and in SF2/ASF is shared by other SR proteins, but our mutational analysis did not uncover a function for this motif in alternative or constitutive splicing *in vitro*. We have not found other distinct motifs present in all available SR protein sequences.

When sequences upstream of the RNP-1 submotif of SF2/ASF were replaced by the corresponding variable loop of RRM2 in hnRNP A1, no major changes in constitutive or alternative splicing activities, or in crosslinking to pre-mRNA or mRNA were observed. These observations suggest that any determinants of specific binding or function reside in other regions of the RRM or of SF2/ASF. It is also possible that these particular RRMs have limited or no sequence specificity in RNA binding. Finally, it is possible that despite their different primary sequences, the SF2/ASF and hnRNP A1 variable loops contribute similar binding or functional specificities, since the two proteins have antagonistic functions and hence may share a common binding site.

Binding of SF2/ASF to RNA is mediated by three domains: the N-terminal RRM, the central ψRRM and the C-terminal RS domain. The RS domain may act as an auxiliary or effector domain to promote protein–protein or protein–RNA contacts with other factors or snRNPs within the spliceosome and also to promote sequence-independent RNA binding through salt bridges between Arg residues and the phosphate backbone. The RRM may be involved in binding to the pre-mRNA substrate or to an snRNP at one or multiple sites and in determining specific protein–RNA contacts. However, unlike other RRMs, the isolated SF2/ASF RRM binds RNA only inefficiently. Efficient RNA binding results from synergistic interactions with the ψRRM and significantly, a mutant SF2/ASF protein containing both domains and lacking the RS domain is active in alternative splicing, whereas separate RRM and ψRRM fragments are inactive. It appears that multi-RRM-containing proteins fall into two classes: those in which each RRM is capable of independent RNA binding (Sachs *et al.*, 1987; Scherly *et al.*, 1989; Hoffman *et al.*, 1990; Nagai *et al.*, 1990; Nietfeld *et al.*, 1990) and those that require synergy between RRMs or RRM-like regions (this study; Burd *et al.*, 1991; Zhang *et al.*, 1992). The same protein may behave differently in this respect depending on the target RNA or the binding conditions (Sachs *et al.*, 1987; Burd *et al.*, 1991).

We have shown that the RNA binding domains of SF2/ASF are competent for general RNA binding in the absence of other factors. The intrinsic RNA binding activities of the purified SF2/ASF mutant proteins were assayed by UV crosslinking and by a nitrocellulose filter binding assay. SF2/ASF bound to both β-globin pre-mRNA and mRNA, showing no preferential recognition of intronic sequences. Moreover, when a splicing reaction was incubated with anti-SF2/ASF antibodies, the mature mRNA was enriched in the immunoprecipitate, consistent with direct or indirect association of wt SF2/ASF with exonic sequences under splicing conditions (A.Mayeda and A.R.Krainer, unpublished data). The lack of preference for intronic sequences in the crosslinking and filter binding assays was not due to the preponderance of non-specific interactions mediated by the RS domain, since similar lack of discrimination between pre-mRNA and mRNA was obtained with all the RS mutant proteins. Non-specific interactions of SF2/ASF at multiple sites on pre-mRNA and mRNA could potentially mask high affinity sequence-specific interactions, e.g. with 5′ or 3′ splice sites. Preliminary gel retardation experiments with purified HeLa SF2/ASF and oligoribonucleotides showed binding to both 5′ and 3′ splice sites, although binding to 3′ splice sites was more resistant to heparin, suggesting differential contacts (E.Birney, A.Mayeda and A.R.Krainer, unpublished data).

In this study we measured protein–RNA interactions between purified SF2/ASF wt or mutant proteins and pre-mRNA or mRNA. So far we have not addressed the possibility that the splicing activities of SF2/ASF are mediated by direct protein–RNA contacts with snRNAs and that interactions with the substrate have limited sequence specificity, serving primarily to position SF2/ASF at or near the splice sites. For example, the splicing factor SC35, which has biochemical activities that are very similar to those of SF2/ASF (Fu *et al.*, 1992), interacts directly or indirectly with U1 and U2 snRNAs at the 3′ splice site (Fu and Maniatis, 1992b). Likewise, SF2/ASF stabilizes the interactions of U1 snRNP with 5′ splice sites, which may be the basis for its alternative splicing activity (Eperon *et al.*, 1993). Specific binding of SF2/ASF to pre-mRNA may result from kinetic competition with other RNA binding proteins, or from protein–protein interactions.

The present RNA binding studies were carried out with purified recombinant proteins produced in *E.coli* and therefore did not address the likely influence of other factors, or the effects of phosphorylation on the RNA binding properties of SF2/ASF. It is thought that the Ser residues in the RS domains of SR proteins become phosphorylated, since all are phosphoproteins (Krainer *et al.*, 1991; Zahler *et al.*, 1992) and the analogous domain in U1-70K is phosphorylated at multiple serines *in vivo* (Woppmann *et al.*, 1990). Moreover, synthetic random co-polymers of Arg and Ser are excellent substrates for many Ser/Thr protein kinases (Racker, 1991). Phosphorylation would tend to neutralize the strong positive charge of this domain and may have a regulatory role, although it is unclear at present whether modified and/or unmodified SF2/ASF is the active form of the protein.

Several members of the SR family of nuclear phosphoproteins have recently been shown to possess constitutive and alternative splicing activities that are very similar to those of SF2/ASF (Fu *et al.*, 1992; Kim *et al.*, 1992; Mayeda *et al.*, 1992; Zahler *et al.*, 1992, 1993a,b). The entire set of SR proteins is highly conserved phylogenetically and individual members from the same or from different species are highly related to each other in sequence and in the organization of RRM and RS domains, even though the apparent sizes of SR proteins range from 20 to 75 kDa (Zahler *et al.*, 1992). Interestingly, some of these proteins, e.g. SC35 and SRp20, lack the central ψRRM, yet SC35

and a *Drosophila* SRp20 homolog, RBP1, are active in both constitutive and alternative splicing assays (Fu *et al.*, 1992; Kim *et al.*, 1992). This structural difference may be explained if the RRM domain of these proteins functions autonomously, unlike the SF2/ASF RRM, which requires the ψRRM for binding and for alternative splicing activity. These structural differences among SR proteins may contribute to the expected functional differences among family members, such as *in vivo* substrate specificities. In addition to the above gene family diversity, two alternatively spliced isoforms of SF2/ASF have been reported (Ge *et al.*, 1991; Tacke *et al.*, 1992), which delete or replace the RS domain and also replace sequences corresponding to the β4 strand of the ψRRM, according to the prototype RRM structure. Based on our mutational analysis of SF2/ASF these isoforms would be expected to be inactive in constitutive splicing, since they lack an RS domain (Ge *et al.*, 1991). Indeed both isoforms were recently found to be inactive in constitutive as well as in alternative splicing (Zuo and Manley, 1993). Their inability to promote proximal 5′ splice site utilization suggests that the β4 sequence changes affect the structure, RNA binding properties and/or function of the ψRRM.

Materials and methods

Oligonucleotides

Oligonucleotides were synthesized by the Cold Spring Harbor Oligonucleotide Facility. The sequences were as follows: oligo # 1, tcgccgcgggggacgcggcttcgccttcg; oligo # 2, tcgccgcgggggaccgcccgatgccgatgtt; oligo # 3, tcgccgcggcagcggcaaaaaacgtggtttcgccttcgttgag; oligo # 4, cttctttccgtacaaactccacgacac; oligo # 5, tcgccgcgggggaccgcccttcgccttcgttaccttccagaacccggccaacgtggaagac; oligo # 6, cagacggtacccatcgtaatcatagccgtcgcgaccatacaccgcgtcttccacgttggccggt; oligo # 7, catgggccccgtactccgagctatggccgtactcgcactcgtacccgcacccgtactcgcacccgcactcgtactaacactcgtacccgtacttat; oligo # 8, cttggatccttaggtacggtacgagtgcgagtgtgacgcgggctataacgcggagagccacgagtgcgacgcgggctataagtacggtacgagtgcgagtgt; oligo # 9, catgggccccgtggtccgagctatggccgtggtcgcggtcgtggccgcggccgcggtcgtggtaacggtcgtggccgtggttat; oligo # 10, cttggatccttaggtacggccacgaccgcgaccgtgacgcgggctataacgcggagagccacgaccgcgacgcggcgctataaccacggccacgaccgt; oligo # 11, catgggcccggttctccgagctatggcggtagcggcagcggttctggcagcggttctggctctggcagcggttctaacagcggttctggtagctat; oligo # 12, cttggatccttaggtaccgctaccagagccgctgtgacgcgggctataacgcggagagccaccagagccacgcgggctatagctaccagaaccgctgt; oligo # 13, catgggcccaagagcccgagctatggcaagagcaaaagcaagtctaaatctaagagcaaatctaaaagcaagagcaacagcaagtctaagagctat; oligo # 14, cttggatccttaggtcttagacttgcttttgctgtgacgcgggctataacgcggagagccctgcttttacgcgggctatagctcttagacttgctgt; oligo # 15, catgggccccgagctatggcaactatagcccgcgtggctctccgcg; oligo # 16, cttggatccttaggtgtgacgcgggctataacg cggagagccacgcgggct; oligo # 17, cagcatatgtcgggaggtggtgt; oligo # 18, tcg ggatcctttatcggcctgttccacggcc; oligo # 19, tcgcatatggctccccgaggtcgctat; oligo # 20, tcgggatccttatccataacttggacttct; oligo # 21, cagcatatgcgtagcccgagctat; oligo # 22, cttggatccttaggtacgaga; oligo # 23, tcgggctacgacctcggggagctcc; oligo # 24, tccccgaggtcgtagcccgagctat.

Generation of SF2/ASF mutants

The different SF2/ASF mutants were generated by replacement of restriction fragments of a cloned wt SF2/ASF cDNA with either PCR amplified fragments or with synthetic DNA.

For the PP-RG, FF-DD and loop3 mutants, discrete PCR-amplified DNA fragments generated with oligonucleotides # 1 − # 4, # 2 − # 4 and # 3 − # 4, respectively, were purified on 10% non-denaturing polyacrylamide gels, digested with *Sac*I and *Sac*II and subcloned into the corresponding sites of pET9c-SF2 (Krainer *et al.*, 1991).

For the RS domain and the Octa mutants, the mutations were introduced by partial gene replacement using complementary mutant oligonucleotides. The following oligonucleotides were used in each case: # 5 and # 6 (Octa); # 7 and # 8 (RT); # 9 and # 10 (RG); # 11 and # 12 (GS); # 13 and # 14 (KS); and # 15 and # 16 (ΔRS). Half a microgram of each partially complementary oligonucleotide was incubated in 86 μl of 10 mM Tris−HCl pH 7.4 and 1 mM EDTA at 95°C for 2 min and annealed by slow cooling. The annealed oligonucleotides were filled-in with Sequenase 2.0 (USBC) and digested with *Apa*I and *Bam*HI (except for the Octa mutant, for which *Sac*II and *Kpn*I were used instead). The resulting restriction fragments were gel-purified and subcloned as above.

After optimizing protein expression using the pET-19b expression system (Novagen), all the SF2/ASF templates described above were subcloned as *Nde*I−*Bam*HI fragments into the corresponding sites of pET-19b. To express RRM and ψRRM fragments, discrete PCR-amplified DNA fragments generated with oligos # 17 and # 18 and # 19 and # 20, respectively, were gel-purified as above, digested with *Nde*I and *Bam*HI, and subcloned into the corresponding sites of pET-19b. Similarly, for the RRM/ψRRM, ψRRM/RS and RS constructs, discrete PCR-amplified DNA fragments generated with oligo pairs # 17 and # 20, # 19 and # 22, and # 21 and # 22, respectively, were processed as above. For the RRM/RS construct two steps of PCR amplification were employed. First, wt SF2/ASF cDNA was amplified in separate reactions with oligo pairs # 17 and # 23, and # 24 and # 22. The two PCR products, which have a 20 bp overlap, were gel-purified and used in equimolar amounts for a second PCR amplification with oligos # 17 and # 22. The resulting PCR product was purified and subcloned as above.

The double mutants FF-DD/ΔRS and FF-DD/GS were constructed by combining some of the mutants described above. For the FF-DD/ΔRS and FF-DD/GS mutants, *Sac*I−*Bam*HI fragments of the ΔRS and GS mutants were respectively subcloned into the corresponding sites of the FF-DD plasmid.

All subcloning steps were carried out using *E.coli* DH5α. Amino acid substitutions were confirmed by nucleotide sequencing of plasmid DNA using a Sequenase II kit (USBC). PCR conditions with Vent polymerase (New England Biolabs) were as previously described by Krainer *et al.* (1991).

Bacterial expression and purification of recombinant SF2/ASF proteins

pET-19b plasmids containing wt or mutant SF2/ASF coding sequences were re-transformed into the *E.coli* strain BL21(DE3)pLysS (Novagen) and transformants were selected with ampicillin and chloramphenicol. Bacterial cultures grown in M9ZB (Studier *et al.*, 1990) were induced for 3 h with 0.4 mM IPTG when their OD_{600} reached 0.6. The cells were harvested, washed in 10 mM Tris−HCl (pH 8.0), 0.1 M NaCl, 0.2 mM EDTA and 1 mM DTT, and frozen at −70°C. Bacterial extracts were prepared from 1 l cultures by resuspending the frozen cell pellet in buffer A (20 mM HEPES−Na (pH 8.0), 0.1 M NaCl, 5% glycerol, 0.5 mM DTT, 0.2 mM EDTA and 1 mM PMSF), followed by sonication and centrifugation at 12 000 r.p.m. for 20 min in an SS34 rotor (Sorvall). The pellets containing the inclusion bodies were resuspended in 12 ml of buffer B (buffer A containing 6 M guanidine hydrochloride and 0.5 M NaCl final concentrations, pH readjusted to 8.0), sonicated again and incubated at 4°C for 1 h with gentle shaking, followed by centrifugation at 12 000 r.p.m. for 10 min in an SS34 rotor. The supernatant fractions containing the SF2/ASF proteins were mixed with 1 ml of Ni^{2+}-nitrilotriacetic acid resin (Qiagen; Hochuli *et al.*, 1987) and rocked for 2 h at 4°C. Following binding of His-tagged SF2/ASF in buffer B, the resin was washed stepwise with 6 ml of buffer B containing decreasing amounts of guanidine hydrochloride (6, 2 and 0.5 M) and eluted in 0.4 ml of buffer B and 0.5 M guanidine hydrochloride containing 200 mM imidazole. Samples were finally dialyzed against 0.5 M guanidine hydrochloride in buffer A without PMSF and adjusted to a protein concentration of 0.4 mg/ml in the same buffer. Protein concentration was measured by the dye binding method (Bio-Rad) with BSA as a standard and verified by SDS−PAGE and Coomassie blue R250 staining. Each protein was diluted 10-fold in buffer A without PMSF immediately prior to use in splicing reactions or in RNA crosslinking assays.

The RRM, ψRRM, RRM/ψRRM, RRM/RS, ψRRM/RS and RS proteins were expressed as soluble proteins and were purified from the bacterial lysate supernatants omitting the denaturation/renaturation procedure (Hoffmann and Roeder, 1991). The cleared supernatants were mixed with the Ni^{2+} resin as before, except that the washes were done with buffer A containing 1 M NaCl and elution was in buffer A containing 200 M imidazole. Final dialysis was against buffer A lacking PMSF.

Several preparations of each mutant protein were analyzed and comparisons of the activities of different mutants were carried out multiple times. In addition, a range of concentrations of each mutant protein was tested in each assay.

Splicing reactions

Standard conditions were employed for the splicing reaction (Krainer *et al.*, 1984; Krainer and Maniatis, 1985) except that the amounts of nuclear extract, S100 extract and purified HeLa cell or recombinant SF2/ASF indicated in the figure legends were employed. In each 25 μl reaction the total volume of extract and SF2/ASF was brought up to 15 μl with buffer A. The synthesis of labeled pre-mRNA substrates, RNA extraction and

electrophoretic conditions have been described (Krainer *et al.*, 1984; Krainer and Maniatis, 1985).

The wt human β-globin pre-mRNA was synthesized from plasmid SP64-HβΔ6 (Krainer *et al.*, 1984) and the model β-globin pre-mRNA with duplicated 5′ splice sites was synthesized from plasmid 5′D16X (Reed and Maniatis, 1986). Both plasmids were linearized with *Bam*HI. Nuclear and cytoplasmic S100 extracts were prepared as described (Dignam *et al.*, 1983; Krainer *et al.*, 1984).

UV crosslinking

Recombinant SF2/ASF proteins (20 or 200 pmol) or 1 μg of nuclease-free BSA (Boehringer) in buffer A were incubated for 30 min at 30°C with 10^5 c.p.m. of globin pre-mRNA or mRNA, labeled with all four [^{32}P]rNTPs to a specific activity of 1.6×10^7 c.p.m./μg, in a 15 μl reaction in splicing buffer conditions. The samples were irradiated with 254 nm light at 4.32×10^6 μJ/cm^2 in a Spectrolinker XL1000 UV crosslinker (Spectronics Corporation). Ten microliters of 0.2 or 2 mg/ml RNase A were added and the samples were incubated for 30 min at 30°C. The samples were fractionated on 12 or 15% SDS–polyacrylamide gels, which were stained with Coomassie Blue R250, destained and autoradiographed.

Nitrocellulose filter binding assays

SF2/ASF proteins were diluted to appropriate concentrations in buffer A and 50 μg/ml BSA. Each dilution was incubated in a 170 μl reaction in splicing buffer conditions with 5 fmol of ^{32}P-labeled human β-globin pre-mRNA or mRNA, which was heated to 90°C for 3 min and cooled on ice prior to use and 5 μg of *E.coli* tRNA (Boehringer). After incubation at 30°C for 30 min, reactions were filtered through Millipore HAWP filters (0.45 μM) (Mayeda *et al.*, 1986) using a Millipore 1225 sampling manifold connected to house vacuum. The filters were washed with 600 μl of buffer A containing 3.2 mM $MgCl_2$ and dried. The radioactivity was measured in a Beckman liquid scintillation counter. Background obtained in the absence of protein (<5% of the input radioactivity) was subtracted.

Acknowledgements

We thank A.Mayeda for valuable advice and reagents; E.Birney and S.Stamm for helpful discussions; E.Birney, A.Mayeda, D.Horowitz, D.Frendewey and S.Munroe for critical reading of the manuscript. We are grateful to J.Pflugrath for advice on protein structure and design of mutants, and to P.Evans for providing the U1-A coordinates. This work was supported by the National Institutes of Health. A.R.K. is a Pew Scholar in the Biomedical Sciences.

References

Ayane,M., Preuss,U., Köhler,G. and Nielsen,P.J. (1991) *Nucleic Acids Res.*, **19**, 1273–1278.
Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) *Genes Dev.*, **3**, 431–437.
Bentley,R.C. and Keene,J.D. (1991) *Mol. Cell. Biol.*, **11**, 1829–1839.
Bingham,P.M., Chou,T.B., Mims,I. and Zachar,Z. (1988) *Trends Genet.*, **4**, 134–138.
Burd,C.G., Matunis,E.L. and Dreyfuss,G. (1991) *Mol. Cell. Biol.*, **11**, 3419–3424.
Carey,J., Cameron,V., de Haseth,P.L. and Uhlenbeck,O.C. (1983) *Biochemistry*, **22**, 2601–2610.
Champlin,D.T., Frasch,M., Saumweber,H. and Lis,J.T. (1991) *Genes Dev.*, **5**, 1611–1621.
Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
Eperon,I.C., Ireland,D.C., Smith,R.A., Mayeda,A. and Krainer,A.R. (1993) *EMBO J.*, **12**, 3607–3617.
Etzerodt,M., Vignali,R., Ciliberto,G., Scherly,D., Mattaj,I.W. and Philipson,L. (1988) *EMBO J.*, **7**, 4311–4321.
Fu,X.-D. and Maniatis,T. (1992a) *Science*, **256**, 535–538.
Fu,X.-D. and Maniatis,T. (1992b) *Proc. Natl Acad. Sci. USA*, **89**, 1725–1729.
Fu,X.-D., Mayeda,A., Maniatis,T. and Krainer,A.R. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 11224–11228.
Ge,H. and Manley,J.L. (1990) *Cell*, **62**, 25–34.
Ge,H., Zuo,P. and Manley,J.L. (1991) *Cell*, **66**, 373–382.
Greenberg,J.R. (1980) *Nucleic Acids Res.*, **8**, 5685–5701.
Harper,J.E. and Manley,J.L. (1991) *Mol. Cell. Biol.*, **11**, 5945–5953.
Haynes,S.R. (1992) *New Biol.*, **4**, 421–429.
Hochuli,E., Döbeli,H. and Schacher,A. (1987) *J. Chromatogr.*, **411**, 177–184.
Hoffmann,A. and Roeder,R.G. (1991) *Nucleic Acids Res.*, **19**, 6337–6338.
Hoffman,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1990) *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.
Kenan,D.J., Query,C.C. and Keene,J.D. (1991) *Trends Biochem. Sci.*, **16**, 214–220.
Kim,Y.-J., Zuo,P., Manley,J.L. and Baker,B.S. (1992) *Genes Dev.*, **6**, 2569–2579.
Krainer,A.R. and Maniatis,T. (1985) *Cell*, **42**, 725–736.
Krainer,A.R., Maniatis,T., Ruskin,B. and Green,M.R. (1984) *Cell*, **36**, 993–1005.
Krainer,A.R., Conway,G.C. and Kozak,D. (1990a) *Cell*, **62**, 35–42.
Krainer,A.R., Conway,G.C. and Kozak,D. (1990b) *Genes Dev.*, **4**, 1158–1171.
Krainer,A.R., Mayeda,A., Kozak,D. and Binns,G. (1991) *Cell*, **66**, 383–394.
Lamm,G.M. and Lamond,A.I. (1993) *Biochim. Biophys. Acta*, **1173**, 247–265.
Lazinski,D., Grzadzielska,E. and Das,A. (1989) *Cell*, **59**, 207–218.
Li,H. and Bingham,P.M. (1991) *Cell*, **67**, 335–342.
Mancebo,R., Lo,P.C.H. and Mount,S.M. (1990) *Mol. Cell. Biol.*, **10**, 2492–2502.
Mattaj,I.W. (1993) *Cell*, **73**, 837–840.
Mattox,R., Ryner,L. and Baker,B.S. (1992) *J. Biol. Chem.*, **267**, 19023–19026.
Mayeda,A. and Krainer,A.R. (1992) *Cell*, **68**, 365–375.
Mayeda,A., Tatei,K., Kitayama,H., Takemura,K. and Ohshima,Y. (1986) *Nucleic Acids Res.*, **14**, 3045–3057.
Mayeda,A., Zahler,A.M., Krainer,A.R. and Roth,M.B. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1301–1304.
Merrill,B.M., Stone,K.L., Cobianchi,F., Wilson,S.H. and Williams,K.R. (1988) *J. Biol. Chem.*, **263**, 3307–3313.
Nagai,K. (1992) *Curr. Biol.*, **2**, 131–137.
Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) *Nature*, **348**, 515–520.
Nietfeld,W., Mentzel,H. and Pieler,T. (1990) *EMBO J.*, **9**, 3699–3705.
Patton,J.G., Porro,E.B., Galceran,J., Tempst,P. and Nadal-Ginard,B. (1993) *Genes Dev.*, **7**, 393–406.
Query,C.C., Bentley,R.C. and Keene,J.D. (1989) *Cell*, **57**, 89–101.
Racker,E. (1991) *Methods Enzymol.*, **200**, 107–111.
Reed,R. and Maniatis,T. (1986) *Cell*, **46**, 681–690.
Riggs,A.D., Suzuki,H. and Bourgeois,S. (1970) *J. Mol. Biol.*, **48**, 67–83.
Rio,D. (1992) *Curr. Biol.*, **4**, 444–452.
Roth,M.B., Zahler,A.M. and Stolk,J.A. (1991) *J. Cell Biol.*, **115**, 587–596.
Sachs,A.B., Davis,R.W. and Kornberg,R.D. (1987) *Mol. Cell. Biol.*, **7**, 3268–3276.
Scherly,D., Boelens,W., van Venrooij,W.J., Dathan,N., Hamm,J. and Mattaj,I.W. (1989) *EMBO J.*, **8**, 4163–4170.
Scherly,D., Boelens,W., Dathan,N.A., van Venrooij,W.J. and Mattaj,I.W. (1990) *Nature*, **345**, 502–506.
Smith,K.C. (1976) In Wans,S.T. (ed.), *Photochemistry and Photobiology of Nucleic Acids*. Academic Press, New York, pp. 187–218.
Smith,V. and Barrell,B.G. (1991) *EMBO J.*, **10**, 2627–2634.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
Surowy,C.S., Van Santen,V.L., Scheib-Wixted,S.M. and Spritz,R.A. (1989) *Mol. Cell. Biol.*, **9**, 4179–4186.
Tacke,R., Boned,A. and Goridis,C. (1992) *Nucleic Acids Res.*, **20**, 5482.
Vellard,M., Surea,A., Soret,J., Martinerie,C. and Perbal,B. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2511–2515.
Wittekind,M., Gorlach,M., Friedrichs,M., Dreyfuss,G. and Mueller,L. (1992) *Biochemistry*, **31**, 6254–6265.
Woppmann,A., Patschinsky,T., Bringmann,P., Godt,F. and Lührmann,R. (1990) *Nucleic Acids Res.*, **18**, 4427–4438.
Zahler,A.M., Lane,W.S., Stolk,J.A. and Roth,M.B. (1992) *Genes Dev.*, **6**, 837–847.
Zahler,A.M., Neugebauer,K.M., Lane,W.S. and Roth,M.B. (1993a) *Science*, **260**, 219–222.
Zahler,A.M., Neugebauer,K.M., Stolk,J.A. and Roth,M.B. (1993b) *Mol. Cell. Biol.*, **13**, 4023–4028.
Zamore,P.D., Patton,J.G. and Green,M.R. (1992) *Nature*, **355**, 609–614.
Zhang,M., Zamore,P.D., Carmo-Fonseca,M., Lamond,A.I. and Green,M.R. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8769–8773.
Zuo,P. and Manley,J.L. (1993) *EMBO J.*, **12**, 000–000.

Received on July 20, 1993; revised on August 10, 1993