The EMBO Journal vol.12 no.12 pp.4727–4737, 1993

Functional domains of the human splicing factor ASF/SF2

Ping Zuo[1] and James L.Manley

Department of Biological Sciences, Columbia University, New York, NY 10027, USA
[1]Present address: Department of Molecular Biology and Biochemistry, Harvard University, Cambridge, MA, USA

Communicated by I.Mattaj

The human splicing factor ASF/SF2 displays two predominant activities in *in vitro* splicing assays: (i) it is an essential factor apparently required for all splices and (ii) it is able to switch utilization of alternative 5′ splice sites in a concentration-dependent manner. ASF/SF2 is the prototype of a family of proteins typified by the presence of one or two RNP-type RNA binding domains (RBDs) and a region highly enriched in repeating arginine–serine dipeptides (RS regions). Here we describe a functional analysis of ASF/SF2, which defines several regions essential for one, or both, of its two principal activities, and provides insights into how this type of protein functions in splicing. Two isoforms of the protein, which arise from alternative splicing, are by themselves inactive, but each can block the activity of ASF/SF2, thereby functioning as splicing repressors. Some, but not all, mutations in the RS region prevent ASF/SF2 from functioning as an essential splicing factor. However, the entire RS region can be deleted without reducing splice site switching activity, indicating that it is not absolutely required for interaction with other splicing factors. Experiments with deletion and substitution mutants reveal that the protein contains two related, but highly diverged, RBDs, and that both are essential for activity. Each RBD by itself retains the ability to bind RNA, although optimal binding requires both domains.

Key words: human splicing factor/RNA binding domains/RNA splicing/splicing repressors

Introduction

A very large number of proteins, in addition to those that constitute the four essential snRNPs, are required for the process of pre-mRNA splicing (reviewed by Green, 1991; Moore *et al.*, 1992). In yeast, a combination of genetic and biochemical approaches has led to the identification of numerous non-snRNP splicing factors. Biochemical analyses have suggested that these yeast proteins participate in essentially all steps of the splicing reaction, from the earliest phases of spliceosome assembly to release of the spliced product (Guthrie, 1991; Ruby and Abelson, 1991). The yeast non-snRNP proteins characterized to date contain several types of sequence motifs that provide insights into their function. A few contain zinc finger-like structures that may function in RNA binding and as many as half a dozen contain 'DEAD' or 'DEAH' boxes that are the hallmark of ATP-dependent RNA helicases.

In higher eukaryotes, essential protein splicing factors have been identified exclusively by *in vitro* assays. Many of the proteins characterized to date are RNA binding proteins that appear to function very early in the splicing reaction. U2AF binds the polypyrimidine stretch associated with the pre-mRNA 3′ splice site–branch point region, and appears to function by aiding binding of U2 snRNP to the branch site (Ruskin *et al.*, 1988; Zamore and Green, 1989; Zamore *et al.*, 1992). Another polypyrimidine binding protein, PSF, has been described recently and shown to be essential for splicing and to function early in the reaction (Patton *et al.*, 1993). ASF/SF2 is required early in spliceosome assembly (Krainer and Maniatis, 1985; Krainer *et al.*, 1990a) and can also influence the selection of alternative 5′ splice sites (Ge and Manley, 1990; Krainer *et al.*, 1990b). Recent experiments have shown that purified ASF/SF2 can specifically recognize pre-mRNA 5′ splice sites (P.Zuo and J.L.Manley, manuscript submitted). SC35 is another protein required early in spliceosome assembly, perhaps facilitating an interaction between U1 and U2 snRNPs (Fu and Maniatis, 1992a). ASF/SF2 and SC35 appear to be interchangeable in several *in vitro* assays (Fu *et al.*, 1992) and are in fact members of an evolutionarily conserved family of proteins (Zahler *et al.*, 1992). Finally, a number of hnRNP proteins have been implicated in the splicing reaction, although it remains to be seen whether these proteins are essential (e.g. Choi *et al.*, 1986; Gil *et al.*, 1991; Patton *et al.*, 1991; Mayeda *et al.*, 1992).

The above proteins are all characterized by one or two common motifs. Firstly, all of the essential splicing factors (Ge *et al.*, 1991; Krainer *et al.*, 1991; Fu and Maniatis, 1992b; Zamore *et al.*, 1992; Patton *et al.*, 1993) contain one or more copies of an ~80 residue domain known as the RNP-type RNA binding domain (RBD; Bandziulis *et al.*, 1989) or RNA recognition motif (Kenan *et al.*, 1991). The RBD is a conserved structure found in a large number of RNA binding proteins, including several snRNP and hnRNP proteins. It is characterized by two well conserved motifs, the RNP-1 octamer and the RNP-2 hexamer. Based on the known structure of two RBDs, those from the U1A (Nagai *et al.*, 1990; Hoffman *et al.*, 1991) and hnRNP C (Wittekind *et al.*, 1992) proteins, both motifs appear to be juxtaposed at the center of a four-stranded β-sheet, with highly conserved aromatic residues proposed to contact the RNA. A second motif, found in all of the essential factors mentioned above (with the exception of PSF), is an unusual, highly-charged region rich in repeating arginine–serine dipeptides (RS repeats). A related region is found in the U1 snRNP 70 kDa protein (Query *et al.*, 1989; Mancebo *et al.*, 1990). Intriguingly, a number of genetically defined regulators of splicing in *Drosophila* also contain RBDs and/or RS repeats (Boggs *et al.*, 1987; Choi *et al.*, 1987; Amrein *et al.*, 1988; Bell *et al.*, 1988; Goralski *et al.*,

1989). Although it may be a function of the assays employed, it is notable that proteins similar to these have not, as yet, been found in yeast where alternative splicing is rare or non-existent.

The functional significance of these protein domains in splicing has not yet been investigated in depth. The RS region from one of the *Drosophila* regulators has been shown to target a heterologous protein to specific regions in the nucleus enriched in splicing factors, suggesting a role for this region in subcellular protein transport (Li and Bingham, 1991). However, a deletion that removed the comparable region from U2AF inactivated the protein in an *in vitro* splicing assay, but did not prevent it from binding the polypyrimidine stretch of a target RNA (Zamore *et al.*, 1992). U2AF also contains three stretches that match the RBD consensus, and it appears that this region must be intact, both for sequence-specific RNA binding and for splicing (Zamore *et al.*, 1992).

Here, we present an analysis of the regions of ASF/SF2 required for the protein to function in *in vitro* splicing assays and to interact with RNA is presented. A number of ASF/SF2 derivatives were tested to determine those regions of the protein necessary to shift selection of alternative splice sites, and/or to function as an essential splicing factor. We also show that ASF/SF2 contains two divergent RBDs that can bind RNA independently but which are both required for optimal RNA binding and for activity in splicing.

Results

We previously described two distinct assays to measure ASF/SF2 activity (Ge and Manley, 1990; Ge *et al.*, 1991; Harper and Manley, 1992). In the first, purified protein is added to splicing competent nuclear extracts (which contained endogenous ASF/SF2) and the effects on alternative splicing of an SV40 early pre-mRNA is determined. In the absence of exogenous ASF/SF2, splicing of this transcript utilizes predominantly an upstream 5′ splice site to generate large T RNA, while addition of ASF/SF2 results in a concentration-dependent shift in splice site selection such that the downstream, small t 5′ splice site is eventually used almost exclusively. In the second assay, ASF/SF2 is added to an inactive S-100 extract (which contains all of the factors required for pre-mRNA splicing except ASF/SF2) and the ability of the protein to activate splicing of the SV40 RNA is measured. Analogous assays (with different pre-mRNAs) have been used by others, and have led to similar conclusions (e.g. Krainer *et al.*, 1990a,b, 1991).

ASF/SF2 contains four possible functional domains (see Figure 1). At its N-terminus is a region with an excellent match to the RBD consensus. Following this is a glycine stretch that separates the RBD from a second region that has weak, but detectable, RBD homology (Kim *et al.*, 1992; Zahler *et al.*, 1992). Finally, the C-terminal 50 residues (the RS region) are highly enriched in arginine−serine dipeptides. Two additional forms of ASF/SF2, called ASF-2 and ASF-3, can be generated by alternative splicing (Ge and Manley, 1991; see Figure 1). Both lack the RS region and ~12 residues immediately upstream of it.

ASF-2 and ASF-3 can inhibit ASF/SF2

Figure 1A shows that, as expected (Ge *et al.*, 1991), purified recombinant ASF/SF2 was able to both (i) activate splicing of SV40 early pre-mRNA when added to HeLa cell S-100 extract, and (ii) shift splice site selection in favor of the small t 5′ splice site when added to HeLa nuclear extract. In this experiment, significant switching was observed only at the highest ASF/SF2 concentration. This may reflect a relatively high level of small t splicing in the absence of exogenous ASF/SF2; more gradual switching was seen in the experiments described below. Additionally, one or more unidentified bands were occasionally detected amongst the products of splicing in nuclear extract. It is likely that these represent utilization of a previously identified 5′ cryptic splice site (Noble *et al.*, 1988).

In contrast, ASF-2 and ASF-3 were inactive in both assays (Figure 1A; results with ASF-2 not shown), suggesting that either the RS region and/or residues immediately upstream of it are essential for both activities, or the substituted residues in ASF-2 and ASF-3 interfere with ASF/SF2 activity. In fact, in the nuclear extract ASF-3 appeared to have a weak but detectable effect opposite to that of ASF/SF2; large T splicing was slightly increased and small t splicing was slightly decreased. As seen more clearly below, this appears to reflect the ability of ASF-3 to block ASF/SF2 function.

To examine further the possibility that ASF-2 and/or ASF-3 can influence splicing *in vitro*, mixing experiments were performed in which increasing amounts of ASF-2 or -3 were added to S-100 extracts containing moderate amounts of ASF/SF2 (Figure 1B). Both proteins (results with ASF-2 not shown) inhibited ASF/SF2-activated splicing. Substantial inhibition of small t splicing was observed when the concentration of ASF-3 equalled that of ASF/SF2, although notably the low level of large T splicing was essentially unaffected. When the concentration of ASF-3 was increased to twice that of ASF/SF2, small t splicing was abolished and only an extremely low level of large T splicing remained. These results suggest that ASF-3 can function as a dominant inhibitor of ASF/SF2. Previous experiments have shown that very high concentrations of ASF/SF2 itself can also block splicing (Ge *et al.*, 1991; Krainer *et al.*, 1991). However, the ASF-3 induced inhibition is distinct from this, as it occurred when the ASF/SF2 + ASF-3 concentration was well below that required for ASF/SF2 self-inhibition (Figure 1B).

The ASF/SF2 RS region contains sequences essential for the activation of splicing, but not for splice site switching

Whether the RS region is essential for either of the two demonstrable activities of ASF/SF2 in splicing was then tested. Several deletion mutants that lack portions of the RS region were constructed and the purified proteins were tested for their activity in splicing assays. SDS gel analysis of several of the protein preparations used in this experiment, or elsewhere in this study, are shown in Figure 2. The results of splicing experiments, together with the structures of the RS region mutants analyzed, are shown in Figure 3. These results allow several conclusions to be drawn. First, the C-terminal half of the RS region can be deleted without affecting either of the splicing activities (mutant RS-b). In contrast, mutations that extend farther into the RS region (e.g. RS-c), or which leave the C-terminal half of the region intact but delete the N-terminal half (RS-a), are completely defective in the activation of splicing in S-100 extract. In sharp contrast, all these mutants are essentially unaffected in their ability to influence alternative splicing when added

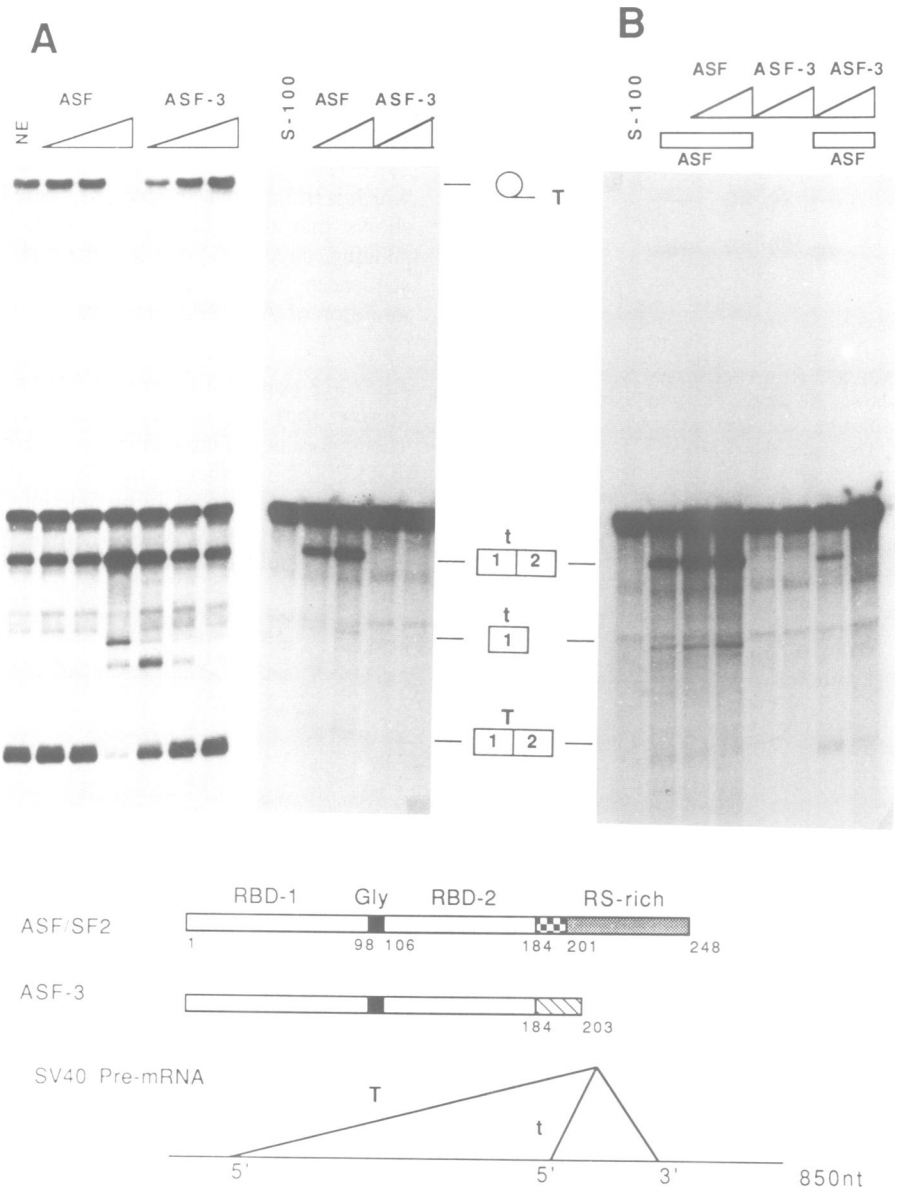

Fig. 1. Effects of ASF/SF2 and ASF-3 on splicing activation and splicing site switching. **(A)** In vitro splicing of [^{32}P]SV40 pre-mRNA in both nuclear and S-100 extracts was performed with increasing amounts of ASF/SF2 (ASF) or ASF-3 as indicated at the top of each lane. The resulting products were analyzed on a 5% denaturing polyacrylamide gel. (In this and other figures, the bottom ~1/3 of the gel, which contained the large T 5′ exon and/or small t intron, is not shown.) With the nuclear extract (NE), 30, 60 or 90 pmol of ASF/SF2 or ASF-3 was added to the reaction mixtures, as indicated. With S-100 extract, 40 or 90 pmol of ASF/SF2 or ASF-3 was added to the indicated reaction mixtures. The structures of the splicing products and intermediates are indicated schematically on the right. **(B)** Mixing experiments were performed in which increasing amounts (30 or 60 pmol) of ASF/SF2 (ASF) or ASF-3 were added to S-100 extracts containing 30 pmol of ASF/SF2, as indicated at the top of each lane. The same amounts (30 or 60 pmol) of ASF-3 were also added to S-100 extract in the absence of ASF/SF2 as negative controls. At the bottom of the figure, schematic diagrams of the primary structures of ASF/SF2 and ASF-3 are shown. Functional domains of ASF/SF2 are indicated above the diagram and the numbers represent the amino acid residues of the proteins. ASF-3 lacks the RS region found in ASF/SF2 and ~12 residues immediately upstream of it (indicated in ASF/SF2 by a checkered pattern). The approximate positions of two RNA binding domains (RBD-1 and RBD-2) and a glycine-rich region (Gly) are also indicated. The alternative splicing pattern and size of the SV40 pre-mRNA substrate is also shown.

to nuclear extract. Indeed, deletion of the entire RS region (ASFΔRS) did not reduce splice site switching; in fact, the activity of ASFΔRS in this assay was somewhat higher than that of full-length ASF/SF2 (compare the switching activities of ASF/SF2 and ASFΔRS observed at lower protein concentrations).

The above results show a clear difference in the requirement for the RS region in the two assays employed to monitor ASF/SF2 activity, perhaps reflecting an intrinsic difference in the ability of ASFΔRS to function in the assays, differential interactions with other splicing factors that might be present in nuclear extract but not S-100, or a combination of the two. To address this disparity, the effect of ASFΔRS when added to S-100 containing a small amount of ASF/SF2 (Figure 4) was tested. The results suggest that the difference in requirement may be a combination of both the above mentioned possibilities. ASFΔRS increased small t splicing in S-100, but not to the same extent as observed in nuclear extract, nor did it function as well as equivalent amounts of ASF/SF2 (Figure 4). However, another significant point emerges from this experiment. Although the positive effect of ASFΔRS was relatively small, it nonetheless contrasts

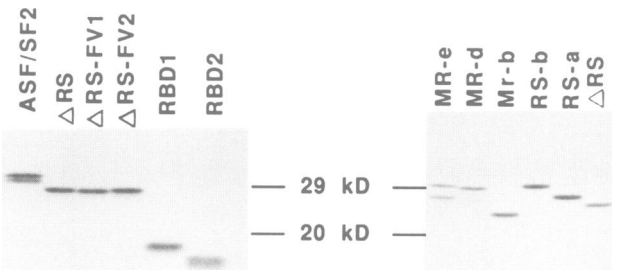

Fig. 2. Representative ASF/SF2 protein derivatives. Selected ASF/SF2 protein derivatives are shown. All the proteins were expressed as histidine fusions and purified by Ni^{2+} agarose chromatography (see Materials and methods). The figure shows two 15% SDS–polyacrylamide gels, stained with Coomassie blue, of purified ASF/SF2 protein derivatives in which each lane contained the same mass amount (10 μg) of protein. The identities of the proteins are indicated above each lane and the molecular weights of marker proteins are listed between the two gels.

sharply with that of the related protein ASF-3, which had a significant negative effect on ASF/SF2 function (see Figure 1B). To determine whether the inhibitory effect is caused by the unique sequences present in ASF-3 or by the lack of the relevant ASF/SF2 sequences (i.e. residues 184–198), the effects of an additional mutant, ASF MR-b, which is truncated at residue 181, were determined. Figure 4 shows that this mutant, like ASF-3, strongly inhibited splicing, suggesting that it is the lack of specific ASF/SF2 sequences that causes ASF-3 to behave as a dominant inhibitor of ASF/SF2 function.

ASF/SF2 contains a second RNP-type RNA binding domain that is essential for activity

The importance of the region just N-terminal to the RS region in ASF/SF2 is further exemplified by the behavior of several different mutants containing deletions encompassing these residues. Not unexpectedly, C-terminal truncations that extend beyond the RS region (MR-a and MR-b) were unable

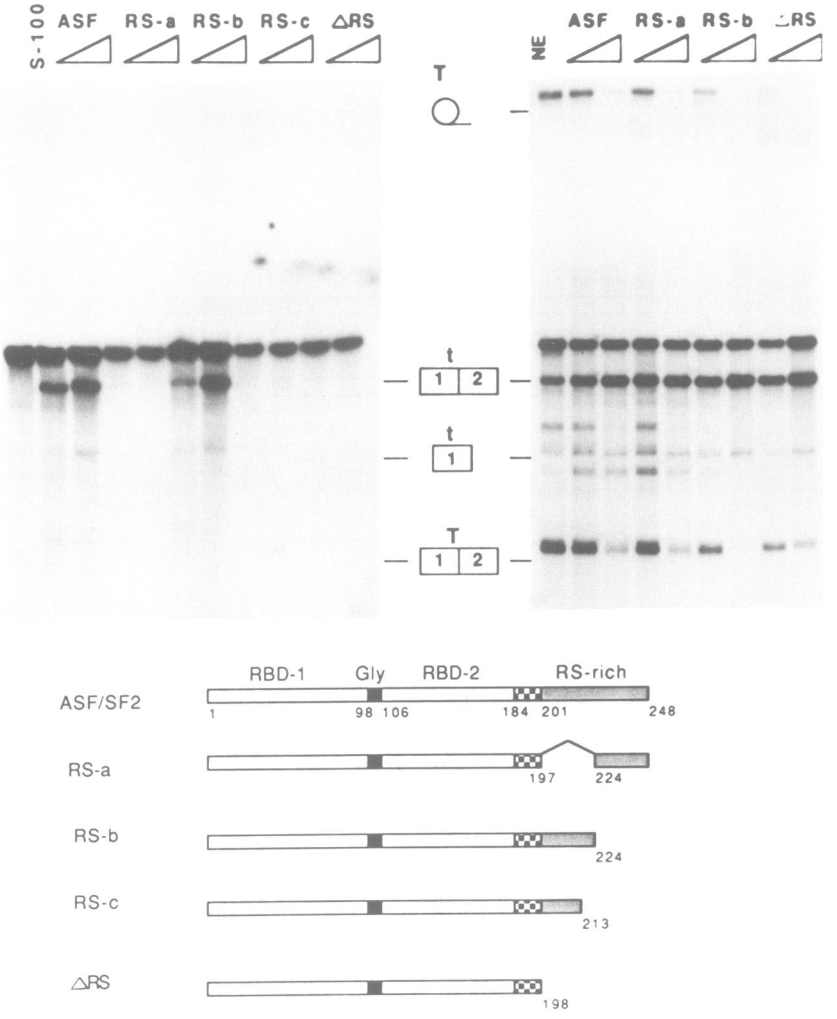

Fig. 3. Sequences within the RS region are essential for splicing activation but not switching. The left panel shows the results of *in vitro* splicing in S-100 extract with two amounts (30 or 80 pmol) of either full-length ASF/SF2 (ASF) or the indicated RS region deletion mutants, as shown above the lanes. The same pre-mRNA and splicing conditions as in Figure 1 were used. Positions of the major splicing products and intermediates are shown between the two panels. The right panel shows the products *in vitro* splicing in nuclear extract. Two amounts (30 or 60 pmol) of the indicated proteins were added to nuclear extract. Schematic diagrams of full-length ASF/SF2 derivatives and the mutant derivatives are displayed at the bottom of the figure.

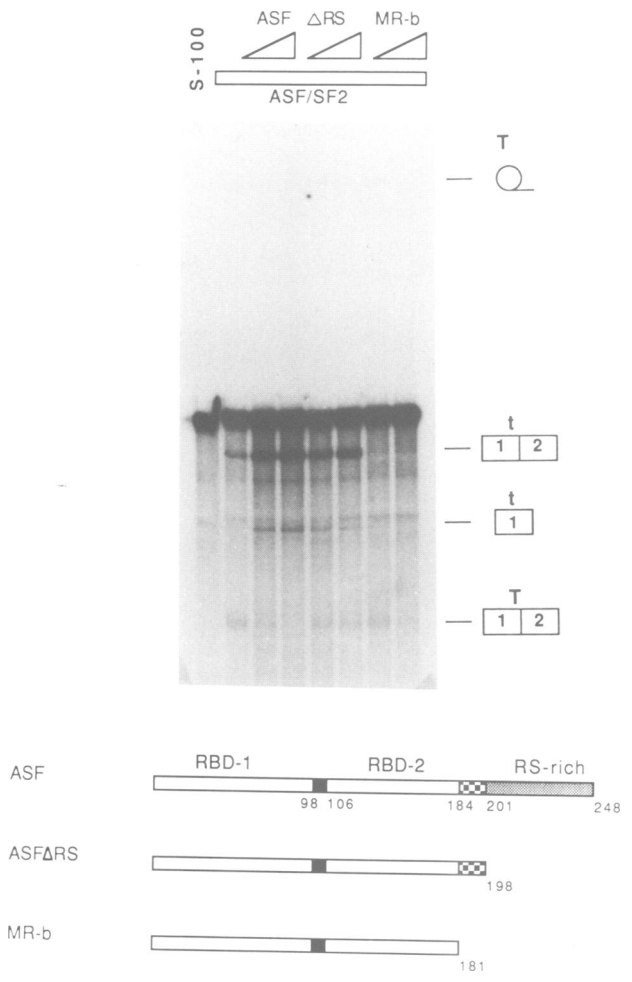

Fig. 4. Converting a splicing activator into a splicing repressor. *In vitro* splicing reactions in S-100 extract were carried out as in Figure 1 with the indicated ASF/SF2 derivatives. All reaction mixtures, except that shown in lane 1, contained 30 pmol of full-length ASF/SF2 and either 30 or 60 pmol of the ASF/SF2 derivative indicated above each lane. The structures of the splicing products and intermediates are listed to the right. Schematic diagrams of the ASF/SF2 derivatives are shown at the bottom of the figure.

to activate splicing in S-100 (results not shown). In addition, two small internal deletions, MR-d and MR-e, were also completely inactive in this assay (Figure 5, right panel), establishing a critical role for these sequences in the activation of splicing. When these mutants were assayed for splice site switching activity (Figure 5, left panels), MR-a, -b and -e were all essentially inactive. These results indicate that amino acid sequences just upstream of the RS region, but downstream of the position where ASF/SF2 and ASF-2 and -3 diverge (i.e. aa 184), are essential for ASF/SF2 activity. As described below, these residues appear to constitute part of a second RNP-type RNA binding domain.

The above results prompted the examination of the sequence of ASF/SF2 just N-terminal to the RS region. Intriguingly, a significant similarity was detected between this region and sequences at the C-terminus of the ASF/SF2 RBD (which, based on structural predictions [Kenan *et al.*, 1991], constitute the fourth strand of the RBD β-sheet). Indeed, an alignment can be made extending N-terminally that includes the entire RBD (Figure 6). However, the overall identity is relatively low (28%), and the characteristic RNP-1 octamer is a very poor match to the consensus. In an effort to assess the functional significance of this region (denoted RBD-2), a number of mutations were constructed in it, as well as in the classic N-terminal RBD (designated RBD-1), and their effects tested on ASF/SF2 function in splicing and, with some, RNA binding (see below). The results obtained with a collection of 10 deletion mutants encompassing portions of either RBD-1 or RBD-2 were similar (results with one [NR1] are shown in Figure 7A; other results not shown). All mutations abolished activity in the S-100 activation assay, suggesting that both regions are important for function as an essential splicing factor. In all cases, splice site switching was reduced or eliminated, although in some instances significant activity in this assay was retained (e.g. NR1). In contrast to these mutations, a deletion that precisely removed the glycine stretch separating the two RBDs had no detectable effect on either activity (Figure 7B; results with nuclear extract not shown). These findings indicate that the flexibility that would probably be imparted by the glycine residues is not essential for either of the protein's two functions in *in vitro* splicing.

To investigate further the significance of the putative RBDs, several small substitution mutants were constructed and analyzed. More specifically, we targeted the sequence FV, which constitutes the only match to the RNP-1 consensus in RBD-2 as well as the only position of identity between the RNP-1 motifs of RBD-1 and -2 (see Figure 6). In one mutant (FV1) the FV in RBD-1 was changed to SR, while mutant FV2 contained the same change in RBD-2. A more conservative single residue change (F→A) was also created in RBD-2 (AV). Strikingly, all three mutant proteins were unable to activate splicing in S-100 (Figure 8A). In addition, their activities in switching 5′ splice site selection in nuclear extract were greatly (FV2 and AV) or slightly (FV1) reduced (Figure 8B). These results provide strong support for the notion that both RBDs are functionally significant and are, in fact, essential for ASF/SF2 to work as a general splicing factor.

ASF/SF2 contains two functional RNA binding domains

The RNA binding properties of several ASF/SF2 derivatives were then examined. We recently described an RNA gel shift assay that we employed to show that ASF/SF2 can bind RNA with specificity for 5′ splice sites (P.Zuo and J.L.Manley, manuscript submitted). These studies showed that the ASFΔRS protein bound this RNA with higher specificity, but somewhat lower affinity, than did intact ASF/SF2, presumably because of non-specific interactions involving the highly charged RS region (P.Zuo and J.L.Manley, manuscript submitted; see also Zamore *et al.*, 1992). In this study this assay has been used with a 95 nt SV40 pre-mRNA fragment that spans the small t intron to examine the ability of several ASF/SF2 derivatives to bind RNA (Figure 9). Specifically, the FV1, FV2 and AV mutations described above were each transferred into the ΔRS background, the mutant proteins purified and their RNA binding abilities compared with that of ASFΔRS. The ability of each putative RBD to bind RNA by itself was also examined by constructing and purifying two additional mutant proteins.

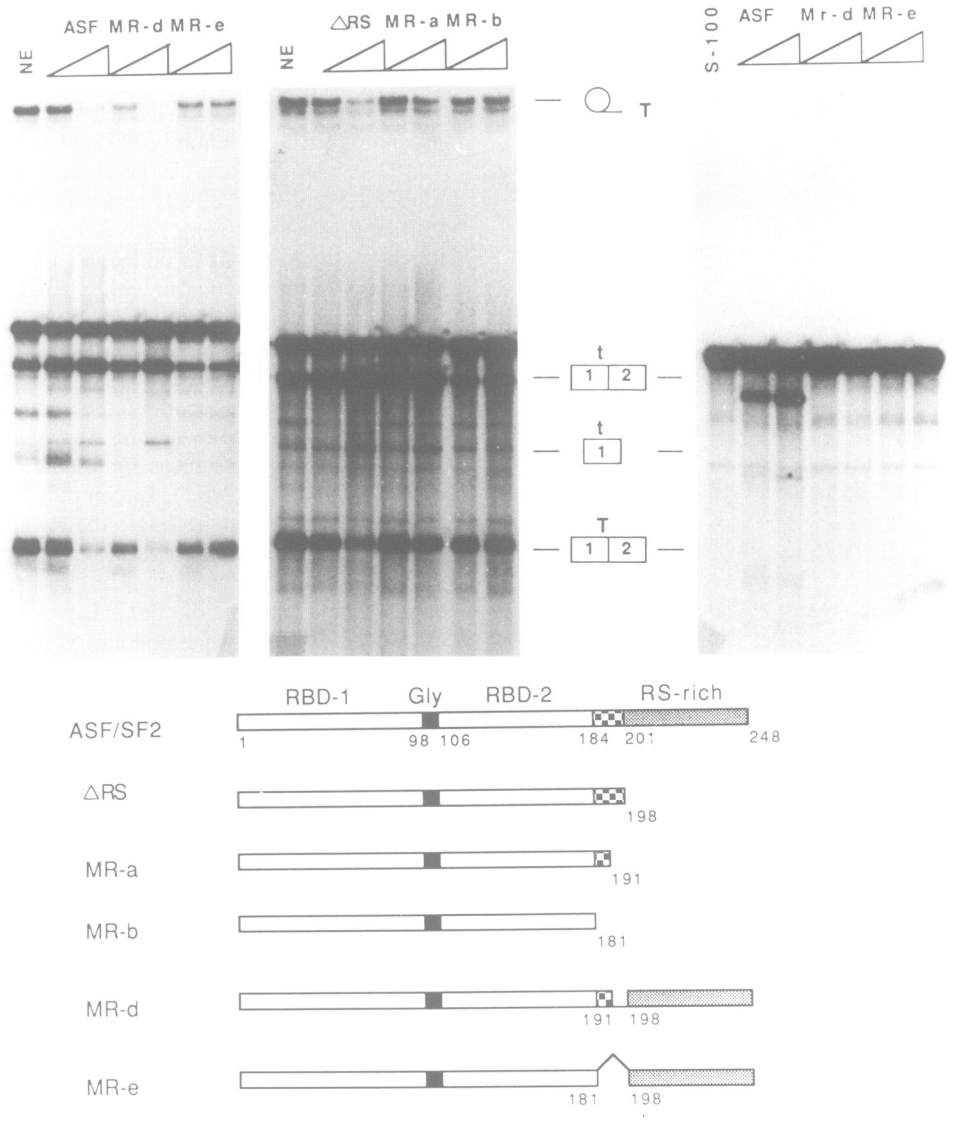

Fig. 5. An essential region of ASF/SF2 suggests a second RNA binding domain. The left panel shows the products of *in vitro* splicing reactions in nuclear extracts containing either 30 or 60 pmol of the indicated ASF/SF2 derivatives. The right panel shows the products of *in vitro* splicing reactions in S-100 extract. Full-length ASF/SF2, MR-d or MR-e mutant derivatives (40 or 80 pmol) were added to the S-100 extracts as indicated. The structures of the major splicing products and intermediates are shown between the two panels. Schematic diagrams of the ASF/SF2 derivatives are shown at the bottom of the figure.

The RBD-1 and RBD-2 proteins contain residues 1−108 and 106−198 respectively. As seen in Figure 9, the FV2 mutation significantly reduced (by a factor of ~5) RNA binding, whereas the FV1 and AV mutations (results with AV not shown) had no, or very slight, effects on binding. Importantly, each RBD by itself retained RNA binding activity, but in both cases the affinity was greatly reduced. Together, these results indicate that (i) ASF/SF2 contains two separable RNP-type RNA binding domains, and (ii) both domains are essential for optimal RNA binding and splicing activity.

Discussion

The experiments described above provide the first detailed structure−function analysis of a mammalian non-snRNP splicing factor. Our results suggest that each of three putative domains of ASF/SF2 plays an important role in one or both of the protein's functions during *in vitro* splicing. Below we

```
                    RNP2                                          RNP1
RBD1 ( 17- 93)  RIYVGNLPPDIRTKDIEDVFYKYGAIRDIDLKNRRGGPPFAFVEF
                :. :  :::     :. :     : .   :.  : :     ::
RBD2 (122-201)  RVVVSGLPPSGSWQDLKDHMREAGDVCYADV-YRDGTGVVEFVRK
                  | ||       |       | |                 ||
CONSENSUS         IYIKGL     D  L   F  YG V        R  KGFAFVXF
                  LFVGN      E  I      F  I           YG

RBD1 ( 17- 93)  EDPRDAEDAVYGRDGYDYDG---Y-RLRVEFPRSGR
                ::   :   .        .:   : :..:. :::
RBD2 (122-201)  EDMTYAVRKLDNTKFRSHEGETAYIRVKVDGPRSPS
                          |       |
CONSENSUS               A       L G       L
                                  I       I
                                          V
```

Fig. 6. Sequence comparison of ASF/SF2 RNA binding domains. ASF/SF2 RBD-1 (residues 17−93), RBD-2 (residues 122−201) and the RBD consensus sequence are compared. Positions at which identities or similarities between RBD-1 and RBD-2 occur are indicated by ':' and '.' respectively. RNP-1 (octamer) and RNP-2 sequences are underlined in each RNA binding domain. Highly conserved residues, which were deduced from compilations presented by Bandziulis *et al.* (1989) and Query *et al.* (1989), are listed below the two ASF/SF2 RBDs. Matches between the conserved residues and RBD-2 are indicated by lines.

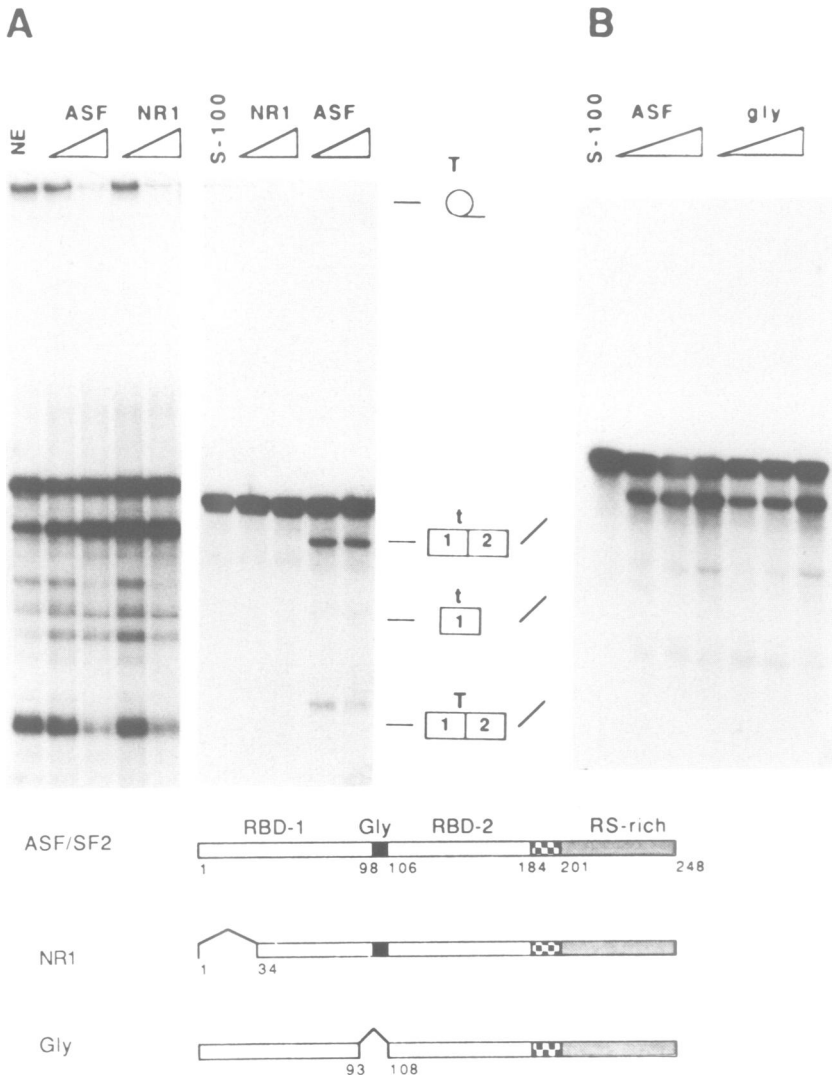

Fig. 7. A deletion in RBD-1 inactivates ASF/SF2, but the glycine hinge is dispensable. **(A)** The products of *in vitro* splicing, both in nuclear and S-100 extracts with increasing amounts of either ASF/SF2 or the mutant NR1, are shown, performed exactly as described in Figure 5. **(B)** The products of *in vitro* splicing in S-100 extract, with increasing amounts (30, 60 or 90 pmol) of either ASF/SF2 or the Gly mutant, are shown. The positions of major products and intermediates are indicated between **(A)** and **(B)**. A schematic diagram of the ASF/SF2 derivatives is shown at the bottom of the figure.

discuss how each of these regions may contribute to ASF/SF2 activity, and then present a model suggesting how the protein functions in splicing.

RS repeat regions are present in a number of different metazoan splicing factors. To date, all proteins of known function that contain such regions are involved in splicing, suggesting that RS repeats may be diagnostic of splicing proteins. But the question of what RS regions actually do remains unclear. The results presented here provide an insight into this question. As mentioned in the Introduction, one previous study showed that an RS region can target a heterologous protein, which is normally cytoplasmic, to specific subnuclear regions known to be enriched in several splicing factors (Li and Bingham, 1991). Based on these results and others, Li and Bingham (1991) argued that the principal function of RS repeats may be to target splicing factors to sites of splicing. Zamore *et al.* (1992) showed that a region encompassing the RS domain of the essential splicing factor U2AF is required for the protein to function in splicing *in vitro*. At first glance these results might appear contradictory, because nuclear targeting should certainly not be required for function in a soluble *in vitro* processing reaction. However, a variation of this model is that the RS region is essential to, in some way, nucleate or bring together many of the large number of factors essential for formation of the spliceosome. *In vivo*, it could be that such a function would result in discrete subcellular localization.

The fact that deletions of the RS region of ASF/SF2 can totally prevent the protein from functioning as an essential splicing factor can be interpreted in a similar way, i.e. that this region is important for bringing ASF/SF2 into association with other components of the splicing apparatus. However, the fact that complete deletion of the RS region is without detectable effect on the protein's ability to function in splice site switching suggests that the situation may be more complicated. Because ASFΔRS can have a significant effect on splicing, it must be assumed that the protein can associate with other splicing factors despite its lack of an RS region. If ASFΔRS can interact with such factors to function in alternative splicing, why would it not be able

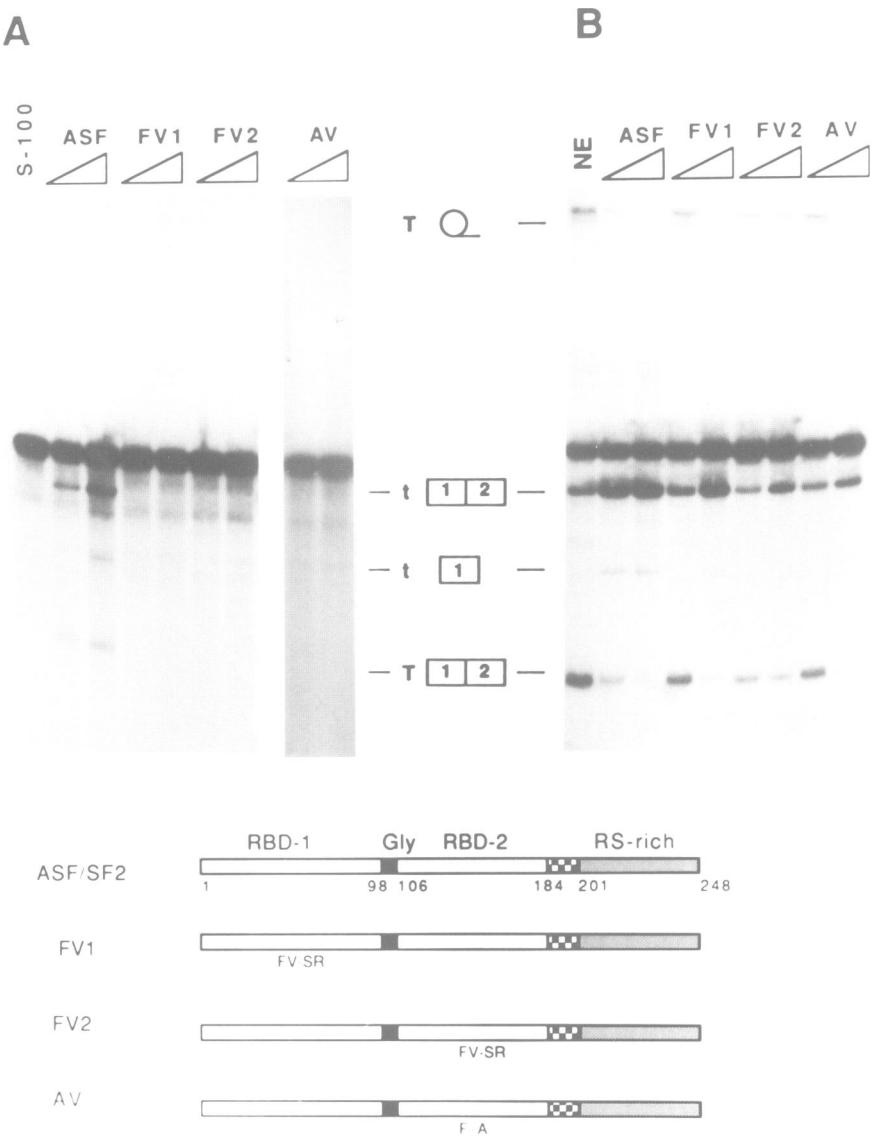

Fig. 8. Point mutations identify critical residues in RBD-1 and RBD-2. **(A)** shows the products of *in vitro* splicing in S-100 extract supplemented with either 30 or 70 pmol of ASF/SF2 or the indicated mutants. **(B)** shows the products of *in vitro* splicing reactions in nuclear extracts containing either 30 or 60 pmol of the ASF/SF2 derivatives indicated above the lanes. The structures of the major splicing products and intermediates are shown between the panels. The schematic diagram below the figure indicates the amino acid changes in the mutant derivatives employed.

to do so to function in general splicing? Our findings thus suggest that the function of the ASF/SF2 RS region involves something more specific than a general 'association' signal.

Our results also show that the entire RS region is not essential for activation of splicing, as the C-terminal 50% (24 residues) can be deleted without detectable effect. Residues 198–224 are necessary and sufficient for function of the RS domain in the *in vitro* assay employed here. This region contains eight consecutive RS dipeptides and we speculate that these are critical for the activation of splicing. Interestingly, this same part of the RS region can be specifically phosphorylated *in vitro* by a U1 snRNP-associated kinase activity (Woppmann *et al.*, 1993). Although the RS region can increase the affinity of ASF/SF2 for RNA, as well as facilitate interaction with other proteins (e.g. the U1 snRNP 70 kDa protein; J.Kohtz and J.L.Manley, unpublished data), precisely how this domain functions in splicing will be an important goal of future experiments.

The experiments described here have established that ASF/SF2 contains two distinct RNA binding domains each capable of functioning partially by itself, but with maximal affinity requiring both domains. The second RBD is significantly diverged from the first, and was not recognized as an RBD by standard homology searches (Ge *et al.*, 1991; Krainer *et al.*, 1991). For example, of 24 residues that are highly conserved among RBD proteins, RBD-1 of ASF/SF2 contains matches at 20 positions, but RBD-2 at only 10. The most characteristic feature of RBD proteins is the RNP-1 octamer, and within RNP-1 the most conserved position is the phenylalanine at position 5, with valine at position 6 being the second most conserved residue (Kenan *et al.*, 1991). The dramatic effect on RNA binding of the RBD-2 FV mutation on the one hand supports the suggestion that this domain is a bona fide RNP-type RNA binding domain. However, if one assumed this from the limited homology, then the fact that the non-conservative FV→SR mutation at these key residues affects binding is perhaps not unexpected. From

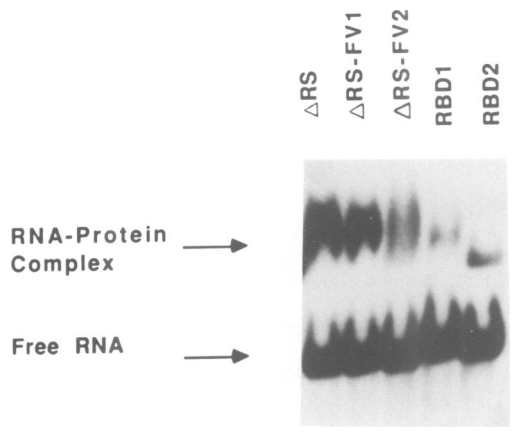

Fig. 9. RNA binding properties of ASF/SF2 derivatives. Gel shift assays were performed as described in Materials and methods. Identical molar amounts (20 pmol) of the indicated ASF/SF2 derivatives were incubated with 3 fmol of SV40 RNA substrate, and RNA−protein complexes were analyzed on a 6% native polyacrylamide gel. The ASF/SF2 derivative utilized is indicated above each lane. The positions of the free RNA probe and the RNA protein complexes are indicated.

this perspective, what might be more surprising is the lack of a significant effect when the identical mutation was introduced into RBD-1. It is conceivable that this may reflect the significantly higher match of RBD-1 to the consensus, perhaps with other residues able to compensate for the disruption of RNP-1. Mutation of the corresponding phenylalanine in a poly(A) binding protein (PABP) RBD was also found to be without detectable effect, although this was a more conservative F→L substitution (Sachs *et al.*, 1986).

A number of RBD-type RNA binding proteins appear to contain more than one RNA binding domain. The reasons for this are unclear and may vary from protein to protein. In only a limited number of instances have the properties of isolated single RBDs from such proteins been examined. The best studied of these multiple RBD proteins is PABP, which contains four RBDs. While it is clear that not all four RBDs are essential for efficient poly(A) binding, whether or not a single RBD can bind efficiently is unclear (Sachs *et al.*, 1986; Nietfeld *et al.*, 1990; Burd *et al.*, 1991). It also appears that one or two of the RBDs may bind preferentially to non-poly(A) RNA (Burd *et al.*, 1991) and that there may be cooperative interactions between individual RBDs (Nietfeld *et al.*, 1990). U2AF^{65} contains three RBDs, which all appear to be essential for high affinity binding to the polypyrimidine stretch (Zamore *et al.*, 1992).

The data presented here indicate that both RBDs of ASF/SF2 are required for optimal activity. While it is not surprising that RNA binding is necessary for the protein to be active in splicing, why does ASF/SF2 require two intact RBDs? The simple answer is that both domains are required for maximal RNA binding affinity, but the situation appears more complex. For example, the SR protein, SC35, has recently been shown to behave in a manner indistinguishable from ASF/SF2 in several *in vitro* assays (Fu *et al.*, 1992), although it contains only a single RBD (Fu and Maniatis, 1992b). While it may be that the SC35 RBD has an RNA binding affinity comparable with that of the two ASF/SF2 RBDs together, a second possibility is that the two proteins might behave differently with other substrates, or under different conditions. Consistent with this, differences in the activities of several SR proteins have recently been observed (Kim *et al.*, 1992; Zahler *et al.*, 1993).

The results described here not only provide evidence that both RBDs of ASF/SF2 are essential for function, but also suggest that this requirement may reflect a role in addition to RNA binding. Our results show that each of the ASF/SF2 RBDs can bind RNA by itself, but that both are required for high affinity binding. In keeping with this, the FV→SR mutation in the second diverged RBD greatly reduced both RNA binding and activity in splicing assays. However, the same mutation in RBD-1 had no significant effect on RNA binding, nor did the conservative F→A change in RBD-2, but each nonetheless markedly reduced activity in splicing. While it is conceivable that these mutations may influence RNA binding under different conditions than were tested here, another possibility is that these residues play a role during splicing distinct from binding the pre-mRNA. One hypothesis is that this involves a specific protein−protein interaction. Several RBD-containing proteins are known to interact with other proteins. These include hnRNP proteins (reviewed by Dreyfuss *et al.*, 1993), the U2 snRNP proteins B″ and A′ (Scherly *et al.*, 1990; Bentley and Keene, 1991), and the splicing regulators Tra and Tra-2 (Inoue *et al.*, 1992). In the case of the B″−A′ interaction, residues within the RBD of the B′ protein are known to be required. In addition, there are a number of examples of DNA binding transcription factors that contact other proteins through regions within their DNA binding domains (e.g. Miner and Yamamoto, 1992 and references therein).

The results presented here, coupled with our demonstration that ASF/SF2 can recognize 5′ splice sites (P.Zuo and J.L.Manley, manuscript submitted), allow us to suggest a model to explain how the protein functions. As an essential splicing factor, we propose that ASF/SF2 functions to help specify the 5′ splice site. It is suggested that the RS region functions by interacting with another splicing component, a simple hypothesis being that it facilitates stable binding of U1 snRNP to the pre-mRNA. This is consistent with the fact that ASF/SF2 is required for a very early step during spliceosome assembly (Krainer *et al.*, 1990a). To influence the selection of alternative 5′ splice sites, we propose that ASF/SF2 is capable of binding multiple sites on the same pre-mRNA, and when, at high ASF/SF2 concentrations, more than one site is bound, the 5′ splice site most proximal to the 3′ splice site is preferentially used simply because of steric considerations. While this model proposes that the recognition of a 5′ splice site by ASF/SF2 is an important determinant in its activation, it is clearly not the only one. Besides the obvious requirement of U1 snRNP binding, such factors as the influence of nearby exon sequences (e.g. Reed and Maniatis, 1986), intron size (e.g. Fu and Manley, 1987) and interactions involving factors at the branch site region (Noble *et al.*, 1988) may also play important roles. Furthermore, additional proteins can influence 5′ splice site selection (Harper and Manley, 1991; Mayeda and Krainer, 1992). One (or more) of these factors may help explain why distal 5′ splice sites are frequently favored at low concentrations of ASF/SF2.

Finally, we offer an explanation for our observation that the RS region is completely dispensable for splice site switching, which is that ASF/SF2 functions as a multimer. This could involve interaction with other ASF/SF2 monomers or with other SR proteins (Zahler *et al.*, 1992), and preliminary results indeed suggest that ASF/SF2 can

interact both with itself and with other RS domain-containing proteins (J.Kohtz and J.L.Manley, unpublished data). It is conceivable that not all subunits of such multimers would require an RS region for activity. This model also provides a feasible explanation for the dominant negative effects of ASF-3, i.e. that the protein interferes with the assembly or activity of functional multimers.

Materials and methods

DNA plasmids

All bacterial expression constructs used to produce ASF/SF2-related proteins were derived from pDS-ASF1 (Ge *et al.*, 1991). Most of the expression vectors were constructed by standard subcloning methods using restriction enzyme sites present in the ASF/SF2 cDNA. Portions of DNA around the junctions of all mutations were sequenced to confirm their identities. ASF-2 and -3 expression vectors were constructed using the appropriate cDNAs (Ge *et al.*, 1991). Detailed strategies for constructing all ASF/SF2 expression vectors are available on request. pDS-ASF-FV1, pDS-ASF-FV2 and pDS-ASF-AV2 were derived from pDS-ASF1 by oligonucleotide-directed mutagenesis using oligos 1, 2 and 3 respectively. Oligos 1 (CGC CCT TCG CCT CTA GAG AGT TCG AGG ACC) and 2 (GGT GTC GTG GAG TCG CGA CGG AAA GAA GAG) result in an FV→SR change in RNP-1 of RBD-1 and RBD-2 respectively. Oligo 3 (GTG TCG TGG AGG CTG TAC GGA AAG AAG AT) produces an F→A change in RNP-1 of RBD-2.

SV40 early pre-mRNA for *in vitro* splicing reactions was produced from pSVi66 (Ge and Manley, 1990). The small t intron RNA fragment used for RNA binding was derived from a 650 bp *DraI–PstI* fragment of pHd (Noble *et al.*, 1987). This fragment was inserted into the *HindIII–PstI* sites of pGem in the orientation of the SP6 promoter, producing pGt.

Expression and purification of ASF/SF2 proteins

Expression and purification of ASF/SF2 derivatives in *Escherichia coli* were carried out essentially as described previously (Ge *et al.*, 1991). Briefly, 100 ml cultures of JM101 were induced with 1 mM IPTG for 3 h and harvested cells were lysed by sonication. The his-tagged recombinant proteins were purified by Ni^{2+} agarose chromatography, dialyzed against buffer D + 0.5 M guanidine HCl, and stored at −70°C. Protein concentrations were determined by the Bradford method and all protein preparations were monitored by Coomassie blue staining of SDS–polyacrylamide gels. Prior to *in vitro* splicing and RNA binding assays, proteins were diluted with buffer D at 4°C to equal molar concentrations. ASF/SF2 with and without the his-tag, produced in *E.coli* and purified by elution and renaturation from SDS gels, were found to behave identically in the assays described here. All experiments were performed with at least two independent protein preparations and at least in triplicate.

In vitro splicing assays

Nuclear and cytoplasmic S-100 extracts were prepared from HeLa cells as described previously (Noble *et al.*, 1987; Ge *et al.*, 1991). SV40 early pre-mRNA was prepared from pSVi66 by *in vitro* transcription. *In vitro* splicing reactions, both in nuclear and S-100 extracts, were carried out essentially as described (Ge *et al.*, 1991). Purified recombinant ASF/SF2 derivatives (10–100 pmol) were added to each 25 μl splicing reaction. Concentrations of buffer D and guanidine HCl (<5 mM) were adjusted to be identical in all samples. This level of guanidine did not influence splicing detectably, although concentrations > ~20 mM could alter alternative splicing pathways.

RNA binding assays

The small t RNA fragment, labeled with [α-^{32}P]GTP and -UTP, was made by combining SP6 polymerase with the plasmid pGt linearized at a PflM I site, producing a 95 nt RNA encompassing the entire small t intron. Full-length transcripts were purified on denaturing polyacrylamide gels prior to RNA binding assays. Gel shift assays were performed with 2.5 fmol of radiolabeled RNA. RNA binding buffer contained 10 mM HEPES, pH 7.9, 200 mM KCl, 20 mM NaCl, 0.025% NP40, 1 mM DTT, 15 μg/ml BSA and 10% glycerol. The indicated ASF/SF2 derivatives (20 pmol) and 5 μg tRNA were added to 25 μl reaction mixtures prior to the labeled probes. Samples were incubated for 30 min at room temperature and analyzed on a 6% (acrylamide:bisacrylamide ratio of 40:1) polyacrylamide gel in 0.5 × TBE. The gel was electrophoresed in 0.5 × TBE buffer at 3.5 W for 4 h, and then dried and autoradiographed.

Acknowledgements

We thank H.Ge, J.Harper and J.Kohtz for helpful discussions, M.Wang for technical assistance, T.Raabe for help with photography, and W.Weast and M.Hosford for help with preparing the manuscript. This work was supported by NIH grant GM 48259.

References

Amrein,H. Gorman,M. and Nöthiger,R. (1988) *Cell*, **55**, 1025–1035.
Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) *Genes Dev.*, **3**, 431–437.
Bell,L.R., Maine,E.M., Schedl,P. and Cline,T.W. (1988) *Cell*, **55**, 1037–1046.
Bentley,R.C. and Keene,J.D. (1991) *Mol. Cell. Biol.*, **11**, 1829–1839.
Boggs,R.T., Gregor,P., Idriss,S., Belote,J.M. and McKeown,M. (1987) *Cell*, **50**, 739–747.
Burd,C.G., Matunis,E.L. and Dreyfuss,G. (1991) *Mol. Cell. Biol.*, **11**, 3419–3424.
Choi,Y.D., Grabowski,P.J., Sharp,P.A. and Dreyfuss,G. (1986) *Science*, **231**, 1534–1539.
Choi,T.-B., Zachar,Z. and Bingham,P.M. (1987) *EMBO J.*, **6**, 4095–4104.
Dreyfuss,G., Matunis,M.J., Piñol-Roma,S., Burd,C.G. (1993) *Annu. Rev. Biochem.*, in press.
Fu,X.-D. and Maniatis,T. (1992a) *Proc. Natl Acad. Sci. USA*, **89**, 1725–1729.
Fu,X.-D. and Maniatis,T. (1992b) *Science*, **256**, 535–538.
Fu,X.-Y. and Manley,J.L. (1987) *Mol. Cell. Biol.*, **7**, 738–748.
Fu,X.-D., Mayeda,A., Maniatis,T. and Krainer,A. (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.
Gil,A., Sharp,P.A., Jamison,S.F. and Garcia-Blanco,M.A. (1991) *Genes Dev.*, **5**, 1224–1236.
Goralski,T.J., Edström,J.E. and Baker,B.S. (1989) *Cell*, **56**, 1011–1118.
Green,M.R. (1991) *Annu. Rev. Cell Biol.*, **7**, 559–599.
Guthrie,C. (1991) *Science*, **253**, 157–163.
Harper,J.E. and Manley,J.L. (1991) *Mol. Cell. Biol.*, **11**, 5945–5953.
Harper,J.E. and Manley,J.L. (1992) *Gene Expression*, **2**, 19–29.
Hoffman,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.
Inoue,K., Hoshijima,K., Higuchi,I., Sakamoto,H. and Shimura,Y. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8092–8096.
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., Conway,G.C. and Kozak,D. (1990a) *Genes Dev.*, **4**, 1158–1171.
Krainer,A.R., Conway,G.C. and Kozak,D. (1990b) *Cell*, **62**, 35–42.
Krainer,A.R., Mayeda,A., Kozak,D. and Binns,G. (1991) *Cell*, **66**, 383–394.
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.
Mayeda,A. and Krainer,A.R. (1992) *Cell*, **68**, 365–375.
Mayeda,A., Zahler,A.M., Krainer,A.R. and Roth,M.B. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1301–1304.
Miner,J.N. and Yamamoto,K.R. (1992) *Genes Dev.*, **6**, 2491–2501.
Moore,M.J., Query,C.C. and Sharp,P.A. (1992) In Gesteland,R. and Atkins,J. (eds), *RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1–30.
Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P. (1990) *Nature*, **348**, 515–520.
Nietfeld,W., Mentzel,H. and Pieler,T. (1990) *EMBO J.*, **9**, 3699–3705.
Noble,J.C.S., Pan,Z.-Q., Prives,C. and Manley,J.L. (1987) *Cell*, **50**, 227–236.
Noble,J.C.S., Prives,C. and Manley,J.L. (1988) *Genes Dev.*, **2**, 1460–1475.
Patton,J.G., Mayer,S.A., Tempst,P. and Nadal-Ginard,B. (1991) *Genes Dev.*, **5**, 1224–1236.
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.

Reed,R. and Maniatis,T. (1986) *Cell*, **46**, 681–690.
Ruby,S.W. and Abelson,J. (1991) *Trends Genet.*, **7**, 79–85.
Ruskin,B., Zamore,P.D. and Green,M.R. (1988) *Cell*, **52**, 207–219.
Sachs,A.B., Bond,M.W. and Kornberg,R.D. (1986) *Cell*, **45**, 827–835.
Scherly,D., Dathan,N.A., Boelens,W., Van Venrooij,W.J. and Mattaj,I.W. (1990) *EMBO J.*, **9**, 3675–3681.
Wittekind,M., Görlach,M., Friedrichs,M., Dreyfuss,G. and Mueller,L. (1992) *Biochemistry*, **31**, 6254–6265.
Woppmann,A., Will,C.L., Kornstädt,U., Zuo,P., Manley,J.L. and Lührman,R. (1993) *Nucleic Acids Res.*, **21**, 2815–2822.
Zahler,A.M., Lane,W.S., Stalk,J.A. and Roth,M.B. (1992) *Genes Dev.*, **6**, 837–847.
Zahler,A.M., Neugenbauer,K.M., Lane,W.S. and Roth,M.B. (1993) *Science*, **260**, 219–222.
Zamore,P.D. and Green,M.R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9243–9247.
Zamore,P.D, Patton,J.G. and Green,M.R. (1992) *Nature*, **355**, 609–614.

Received on July 20, 1993; revised on August 12, 1993