The EMBO Journal vol.13 no.22 pp.5483–5495, 1994

Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins

Akila Mayeda, Stephen H.Munroe[1], Javier F.Cáceres and Adrian R.Krainer[2]

Cold Spring Harbor Laboratory, PO Box 100, 1 Bungtown Road, Cold Spring Harbor, NY 11724-2208 and [1]Department of Biology, Marquette University, 530 North 15th Street, Milwaukee, WI 53233, USA
[2]Corresponding author

Communicated by I.W.Mattaj

hnRNP A1 is a pre-mRNA binding protein that antagonizes the alternative splicing activity of splicing factors SF2/ASF or SC35, causing activation of distal 5′ splice sites. The structural requirements for hnRNP A1 function were determined by mutagenesis of recombinant human hnRNP A1. Two conserved Phe residues in the RNP-1 submotif of each of two RNA recognition motifs appear to be involved in specific RNA–protein interactions and are essential for modulating alternative splicing. These residues are not required for general pre-mRNA binding or RNA annealing activity. The C-terminal Gly-rich domain is necessary for alternative splicing activity, for stable RNA binding and for optimal RNA annealing activity. hnRNP $A1^B$, which is an alternatively spliced isoform of hnRNP A1 with a longer Gly-rich domain, binds more strongly to pre-mRNA but has only limited alternative splicing activity. In contrast, hnRNP A2 and B1, which have 68% amino acid identity with hnRNP A1, bind more weakly to pre-mRNA and have stronger splice site switching activities than hnRNP A1. We propose that specific combinations of antagonistic hnRNP A/B and SR proteins are involved in regulating alternative splicing of distinct subsets of cellular pre-mRNAs.

Key words: alternative splicing/hnRNP A1/hnRNP A or B group proteins/RNA–protein interactions/RNA–RNA annealing

Introduction

Nascent mRNA precursors (pre-mRNAs) in metazoan cell nuclei are associated with protein in large complexes known as heterogeneous nuclear ribonucleoprotein (hnRNP). RNase treatment of isolated hnRNP complexes generates monoparticles that are 20 nm in diameter, sediment at 40S and package ~700 nt of RNA. Individual monoparticles are enriched in the core hnRNP proteins A1, A2, B1, B2, C1 and C2 (Beyer *et al.*, 1977). Immunopurification and two-dimensional gel electrophoresis led to the discovery of >20 hnRNP proteins, including the above six core proteins, that are components of hnRNP complexes in human cells (Piñol-Roma *et al.*, 1988).

Several functional properties of hnRNP proteins are potentially relevant to splicing. Splicing of pre-mRNA involves an ordered assembly of components, including hnRNP proteins, small nuclear ribonucleoprotein particles (snRNPs) and several additional protein factors, resulting in the formation of a large 50–60S complex, termed the spliceosome (reviewed in Dreyfuss *et al.*, 1993; Moore *et al.*, 1993). Reconstituted 40S hnRNP complexes generated with purified core proteins and β-globin pre-mRNA are functional substrates for splicing *in vitro* (Huang *et al.*, 1994). The presence of different sets of hnRNP proteins in transient pre-spliceosome complexes assembled on different pre-mRNAs *in vitro* suggested the involvement of these proteins in splice site selection (Bennett *et al.*, 1992a,b). Preferential sequence-specific binding of hnRNP proteins A1, C1, C2, type C, D and I (also known as PTB) to the polypyrimidine-rich tract upstream of most 3′ splice sites has been reported (Kumar *et al.*, 1987; Swanson and Dreyfuss, 1988; Buvoli *et al.*, 1990b; Gil *et al.*, 1991; Patton *et al.*, 1991; Ghetti *et al.*, 1992; Mulligan *et al.*, 1992). The hnRNP proteins I/PTB and A1 have also been shown to associate with pre-mRNA 5′ splice sites (Stolow and Berget, 1991; Stolow, 1992; Wyatt *et al.*, 1992). Specific binding of hnRNP I/PTB to a rat β-tropomyosin intron region may repress exon-inclusion (Mulligan *et al.*, 1992). Finally, inhibition of splicing *in vitro* by anti-hnRNP antibodies suggested a requirement for hnRNP C proteins in splicing (Choi *et al.*, 1986; Sierakowska *et al.*, 1986).

Direct evidence for the role of an hnRNP protein in splicing was obtained with the purification of an activity that influences alternative 5′ splice site selection *in vitro*, and its identification as hnRNP A1 (Mayeda and Krainer, 1992). The human general splicing factor SF2/ASF (Ge and Manley, 1990; Krainer *et al.*, 1990b, 1991; Ge *et al.*, 1991), which is a member of the SR protein family of nuclear phosphoproteins (Zahler *et al.*, 1992, 1993b) and hnRNP A1 regulate alternative pre-mRNA splicing *in vitro* through antagonistic effects on 5′ splice site selection (Mayeda and Krainer, 1992). Several SR proteins, including at least SF2/ASF (SRp30a), SC35 (SRp30b/PR264), SRp55 (B52) and RBP1 (a *Drosophila* homolog of SRp20), promote use of proximal alternative 5′ splice sites in a concentration-dependent manner (Ge and Manley, 1990; Krainer *et al.*, 1990b; Fu *et al.*, 1992; Kim *et al.*, 1992; Mayeda *et al.*, 1992; Zahler *et al.*, 1993a; reviewed in Horowitz and Krainer, 1994). hnRNP A1 antagonizes the activities of SF2/ASF and SC35 (and perhaps other SR proteins) in 5′ splice site selection, leading to the activation of distal 5′ splice sites. However, hnRNP A1 has no influence on the constitutive splicing and alternative 3′ splice site selection activities of SF2/ASF and SC35 (Mayeda and Krainer, 1992; Fu *et al.*, 1992). Altering the ratio of hnRNP A1 and SF2/ASF can also modulate

© Oxford University Press

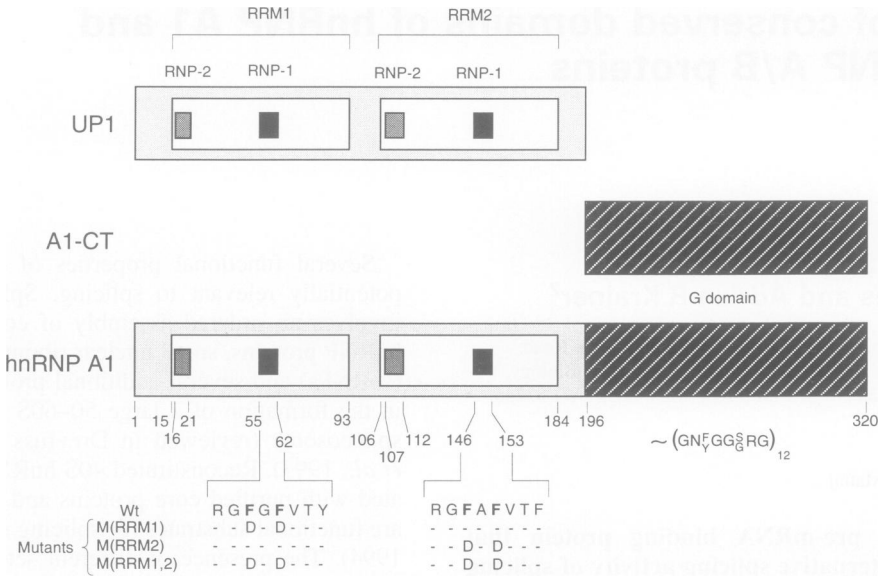

Fig. 1. Structure of human hnRNP A1 and mutant derivatives. The RNP-2 and RNP-1 conserved submotifs of RRM1 and RRM2, and the Gly-rich domain (G domain) are diagrammed. The RNP-1 sequences are shown at the bottom, and amino acid substitutions relative to the wild-type (Wt) hnRNP A1 sequence are indicated. The dashes indicate positions in which no changes were introduced. The numbers indicate amino acid positions from the initiation codon Met1. The G domain consists of 12 imperfect repeats of the sequence shown below the diagram (Cobianchi *et al.*, 1988). All recombinant proteins are in authentic form, except that natural hnRNP A1 contains *N*-acetyl Ser (Ser2) at the mature N-terminus and Arg194 is dimethylated (Williams *et al.*, 1985; Cobianchi *et al.*, 1988). The cloned hnRNP A1 cDNA used here contains the Tyr128 and Arg146 polymorphisms (Buvoli *et al.*, 1988).

alternative exon-skipping/inclusion *in vitro*, provided that the alternative exons are of appropriate length and the competing splice sites have balanced strengths (Mayeda *et al.*, 1993). Recently, overexpression of hnRNP A1 or SF2/ASF was shown to affect alternative splicing *in vivo* (Cáceres *et al.*, 1994; Yang *et al.*, 1994).

hnRNP A1 facilitates the efficient annealing of complementary RNA or DNA molecules *in vitro* (Kumar and Wilson, 1990; Pontius and Berg, 1990, 1992; Munroe and Dong, 1992). RNA–RNA annealing activities may directly facilitate base-pairing interactions between snRNA and pre-mRNA, or they may indirectly facilitate splicing by altering the conformation of the pre-mRNA (Krainer *et al.*, 1990a; Munroe and Dong, 1992; Lee *et al.*, 1993; Portman and Dreyfuss, 1994). Several studies have suggested the existence of specific interactions between hnRNP proteins and snRNP particles. Efficient ultraviolet light (UV)-induced crosslinking of hnRNP proteins A1 and C to pre-mRNA requires the presence of intact U1 and U2 snRNAs in nuclear extract (Mayrand and Pederson, 1990). Recently, hnRNP A1 was shown to bind stably to U2 and U4 snRNPs but not to U1 snRNP, suggesting a potential role for hnRNP A1 in the binding of U2 snRNA to the branch site in the early stages of pre-spliceosome assembly (Buvoli *et al.*, 1992).

Understanding the mechanisms of alternative splicing and its regulation will require detailed analysis of the structural features of these splicing factors that are important for their biochemical activities. The structural elements of general splicing factor SF2/ASF that are required for its constitutive and alternative splicing activities, as well as for general RNA binding, have been described (Cáceres and Krainer, 1993; Zuo and Manley, 1993). The experiments reported here address the structural requirements of an SF2/ASF antagonist, hnRNP A1, for general RNA binding, RNA annealing and alternative 5′ splice site switching. In addition, functional analysis of several of the hnRNP A/B group proteins, which are similar in structure to hnRNP A1, led to the discovery that each of these proteins modulates alternative splicing with different efficiencies.

Results

Design of hnRNP A1 mutants

The hnRNP A1 protein consists of two copies of the RNA recognition motif (RRM), also known as RNP consensus RNA-binding domain (CS-RBD), at the N-terminal portion, and a Gly-rich auxiliary domain (G domain) containing RGG repeats, at the C-terminal portion (Figure 1; reviewed in Dreyfuss *et al.*, 1993). The structure of the N-terminal RRM of polypeptide A of U1 snRNP and of the RRM of hnRNP C suggested that solvent-exposed aromatic residues in RNP-1 and RNP-2 on two antiparallel β-strands contribute to RNA binding through ring-stacking interactions with the bases (Nagai *et al.*, 1990; Hoffman *et al.*, 1991; Görlach *et al.*, 1992). Recently, the secondary structures of RRM1 (Garrett *et al.*, 1994) and RRM2 (S.Yokoyama, personal communication) of hnRNP A1 were determined by NMR spectroscopy, and showed essentially the same overall folding as that of U1-A RRM1 and the hnRNP C RRM, as expected from the conservation of hydrophobic core and other residues among many different RRMs (Kenan *et al.*, 1991; Birney *et al.*, 1993). An earlier study showed that conserved Phe residues in the RNP-1 and RNP-2 submotifs of both RRMs of hnRNP A1 crosslink to oligo(dT) (Merrill *et al.*, 1988). In addition, recent structural studies of splicing factor SF2/ASF demonstrated that mutation of two solvent-exposed Phe residues in the RNP-1 submotif of the N-terminal RRM completely

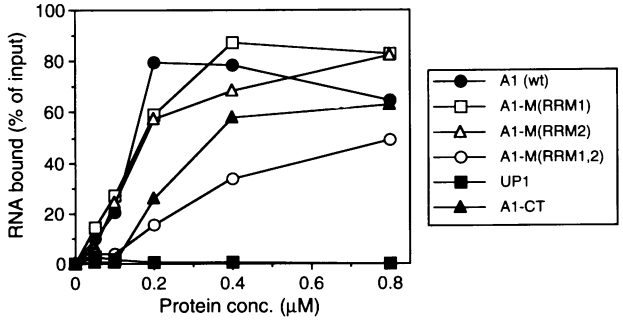

Fig. 2. RNA binding properties of wild-type and mutant hnRNP A1 measured by a filter binding assay. ^{32}P-labeled human β-globin pre-mRNA was incubated under splicing conditions with the indicated concentrations of each protein. The radioactivity of protein-bound labeled RNA retained on the filter was measured and expressed as a percentage of the input radioactivity.

abolishes its function in constitutive, but not in alternative, splicing (Cáceres and Krainer, 1993). Mutation of one of these solvent-exposed Phe residues together with an adjacent hydrophobic core Val residue also had a similar effect (Zuo and Manley, 1993).

In an attempt to inactivate the RRMs of hnRNP A1, we mutated each pair of conserved RNP-1 Phe residues to Asp residues in either RRM1 (Phe57 and 59) or RRM2 (Phe148 and 150), or in both RRM1 and RRM2. The corresponding mutant proteins are designated A1-M(RRM1), A1-M(RRM2) and A1-M(RRM1,2), respectively (Figure 1). The acidic Asp residues should remain solvent-exposed but would be expected to affect substantially the interactions between RNP-1 and the bases or the phosphate backbone of RNA (Cáceres and Krainer, 1993).

To study the function of the characteristic C-terminal G domain, we constructed a mutant, previously termed UP1 (Herrick and Alberts, 1976), that lacks this domain (Figure 1). This deletion mutant has exactly the same truncation site as reported for UP1 polypeptide isolated as a natural proteolytic product of hnRNP A1 (Williams *et al.*, 1985). The same pairs of Phe to Asp point mutations as in hnRNP A1 were also introduced individually or in combination in the RRMs of UP1, giving rise to UP1-M(RRM1), UP1-M(RRM2) and UP1-M(RRM1,2) mutant proteins (Figure 1). We also generated a polypeptide consisting of only the C-terminal G domain, termed A1-CT (Figure 1). All nine recombinant proteins were over-expressed in *Escherichia coli*, purified to apparent homogeneity and characterized in the functional assays.

RNA binding and annealing properties of hnRNP A1 mutant proteins

RNA–protein interactions are undoubtedly an essential feature of the mechanism by which hnRNP A1 regulates alternative splicing. Therefore, the intrinsic RNA-binding properties of hnRNP A1 wild-type and mutant proteins in the absence of other splicing factors were analyzed. The ability of these proteins to promote annealing of complementary RNA molecules was also examined. These studies were carried out with β-globin pre-mRNA as a substrate for binding and annealing under conditions of buffer, salt and temperature optimal for pre-mRNA splicing.

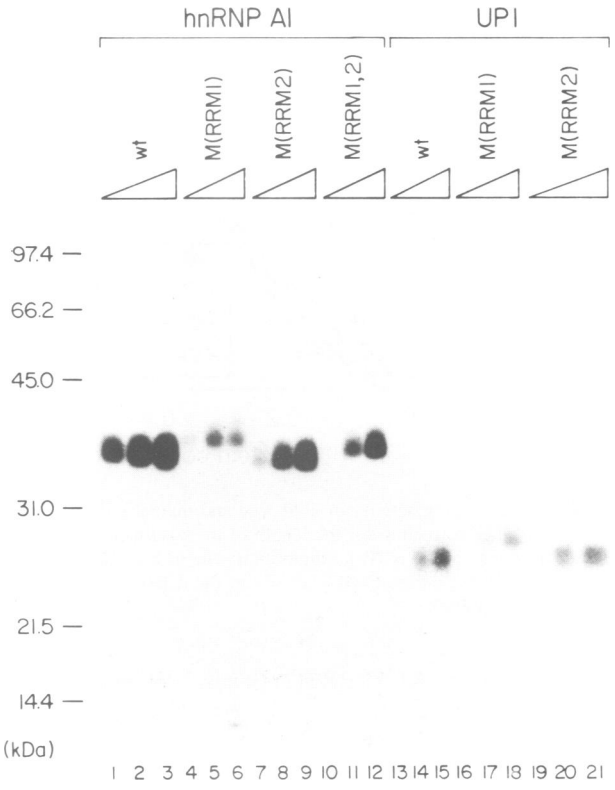

Fig. 3. RNA binding by wild-type and mutant hnRNP A1 measured by a UV crosslinking assay. ^{32}P-labeled human β-globin pre-mRNA was incubated under splicing conditions with increasing amounts of the indicated protein and irradiated with 254 nm UV light. After digestion with RNase A, the products were analyzed by SDS–PAGE and autoradiography. Each set of three lanes corresponds to 0.2, 0.6 and 1.8 μM final protein concentration from left to right. The mobilities of protein molecular weight markers are indicated.

RNA binding was assayed by two techniques that provide different measures of RNA–protein interactions. The first of these was a filter binding assay, which has been used extensively as a rapid, quantitative measure of relatively stable RNA–protein interactions (e.g. Carey *et al.*, 1983; Mayeda *et al.*, 1986). Strong binding to β-globin pre-mRNA was observed with wild-type hnRNP A1 and with the single mutants, A1-M(RRM1) and A1-M(RRM2) (Figure 2). The double mutant A1-M(RRM1,2) bound significantly more weakly. Therefore, either RRM1 or RRM2 is required for maximal stable binding in the context of the full protein. The C-terminal G domain fragment, A1-CT, had significant affinity for the globin RNA, albeit lower than that of intact hnRNP A1. In contrast, UP1 protein showed no significant binding to this RNA by the filter binding assay, even at very low salt concentrations (Figure 2; unpublished data).

To measure weaker or more transient binding to RNA, we used UV-induced crosslinking to compare binding of wild-type and mutant proteins. With this technique, relatively unstable or transient nucleic acid–protein interactions can be detected, depending upon the reactivity of the bases and amino acids that are in direct contact (reviewed in Smith, 1976). Wild-type hnRNP A1 protein crosslinked efficiently to β-globin pre-mRNA (Figure 3, lanes 1–3). A substantial decrease in crosslinking efficiency was observed for A1-M(RRM1) (Figure 3, lanes 4–

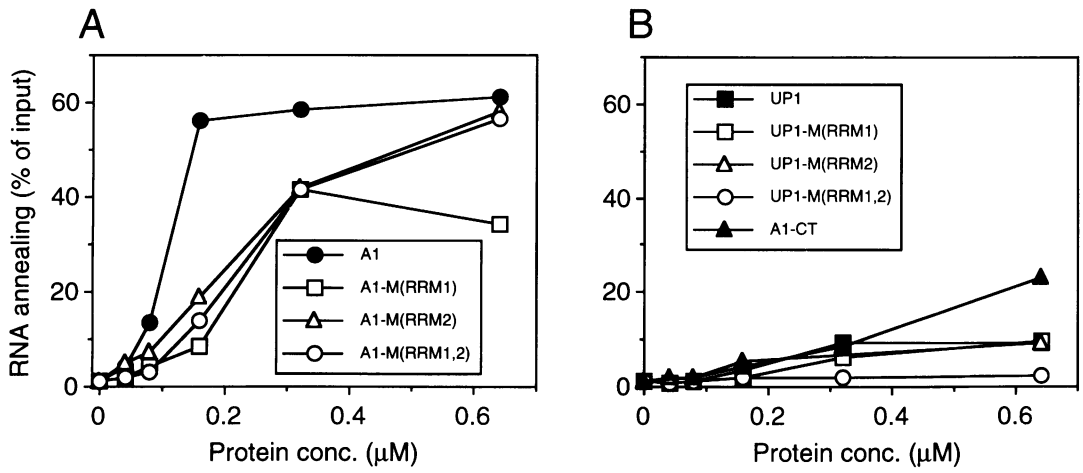

Fig. 4. RNA annealing activity of wild-type and mutant hnRNP A1 proteins. Activity is expressed as percentage of RNase T1-resistant ^{32}P-labeled sense RNA, after correcting for the length of the base-paired region. (**A**) Activity of full-length hnRNP A1 and mutants containing double Phe to Asp replacements in the RNP-1 submotifs of one or both RRMs. In other experiments the annealing activity of A1-M(RRM1) did not drop at the highest protein concentration. (**B**) Activity of the deletion mutants UP1 (N-terminal fragment, with or without Phe to Asp substitutions) and A1-CT (C-terminal fragment).

6), and less so for A1-M(RRM2) (lanes 7–9). Surprisingly, we detected more efficient UV crosslinking with the double mutant A1-M(RRM1,2) than with the single mutant A1-M(RRM1) (Figure 3, lanes 4–6 and 10–12). The UP1 proteins with wild-type or individually mutated RRMs also crosslinked to RNA, but with drastically reduced efficiencies (Figure 3, lanes 13–21, compare with lanes 1–9), whereas UP1-M(RRM1,2) gave no detectable crosslinking (unpublished data). The isolated G domain, A1-CT, showed detectable crosslinking (unpublished data), in agreement with previous studies (Kumar *et al.*, 1990).

Very similar results were obtained for all the proteins with β-globin mRNA instead of pre-mRNA, using either the filter binding or the crosslinking assay (unpublished data). Therefore, these assays do not detect any strong preference of hnRNP A1 for intron sequences over exon sequences, at least in the context of these long RNAs (497 and 367 nt), in contrast to earlier results obtained in crude extracts (Swanson and Dreyfuss, 1988). Although the above assays appear to be measuring general RNA binding with apparent K_ds in the 0.1 μM range, higher affinity, sequence-specific RNA binding by hnRNP A1 has been well documented, in either the presence or absence of additional components (see Discussion).

As a further measure of the interactions of hnRNP A1 with RNA, the ability of wild-type and mutant proteins to promote RNA−RNA annealing was assayed using a 208 nt antisense RNA complementary to the 5′ end of β-globin pre-mRNA (Munroe and Dong, 1992). RNA annealing was measured as the amount of RNase-resistant duplex formed in the presence of 0–0.64 μM protein under splicing conditions (Figure 4). Using these conditions, up to 70% of the RNA forms a duplex within 2–5 min, whereas <0.2% of the pre-mRNA forms duplexes in 16 min in the absence of protein (Munroe and Dong, 1992).

Wild-type hnRNP A1 promoted nearly maximal annealing at 0.16 μM protein (Figure 4A). A1-M(RRM1) and A1-M(RRM2) showed substantial annealing activity, with maximal annealing observed at 0.64 μM protein, at a level similar to that of wild-type hnRNP A1. In contrast to the filter binding results, A1-M(RRM1,2) had almost the same annealing activity as the single mutants. Hence, simultaneous replacement of four conserved Phe residues, which mediate contacts between RRM1 and RRM2 in wild-type hnRNP A1 (Merrill *et al.*, 1988), has little effect on the RNA annealing activity of hnRNP A1. Both A1-CT and UP1 had RNA annealing activity, but at significantly lower levels than either the wild-type or mutant forms of the intact protein (Figure 4B). The annealing activities of the RRM1 and RRM2 mutants were also examined in the context of UP1. UP1-M(RRM1) and UP1-M(RRM2) had similar levels of annealing activity as wild-type UP1. In contrast, the double mutant UP1-M(RRM1,2) showed a virtually complete loss of RNA annealing activity. When the G domain of hnRNP A1 was replaced by the RS domain of SF2/ASF, resulting in a chimeric protein (A1/RS) that bound RNA strongly, RNA annealing activity was fully restored to the level of wild-type hnRNP A1 (unpublished data; see Figure 8).

Alternative splicing activity of hnRNP A1 mutant proteins

hnRNP A1 mutant proteins were assayed for alternative splicing activity using a model β-globin pre-mRNA that has a duplicated 5′ splice site (see diagram in Figure 5; Reed and Maniatis, 1986; Krainer *et al.*, 1990b). When this substrate is spliced in HeLa cell nuclear extract, the distal 5′ splice site is preferentially used, whereas the proximal 5′ splice site is almost exclusively selected in HeLa cell cytosolic S100 extract supplemented with SF2/ASF (Krainer *et al.*, 1990b), reflecting the ratio of SF2/ASF to hnRNP A1 in these extracts (Mayeda *et al.*, 1993). Therefore, S100 extract plus SF2/ASF was used to detect the switch from proximal to distal 5′ splice site that results from the influence of hnRNP A1 on the selection of alternative 5′ splice sites (Figure 5, lanes 1–5; Fu *et al.*, 1992; Mayeda and Krainer, 1992). Using the variant proteins with mutated or deleted domains (Figure 1), we mapped the domains of hnRNP A1 that are responsible

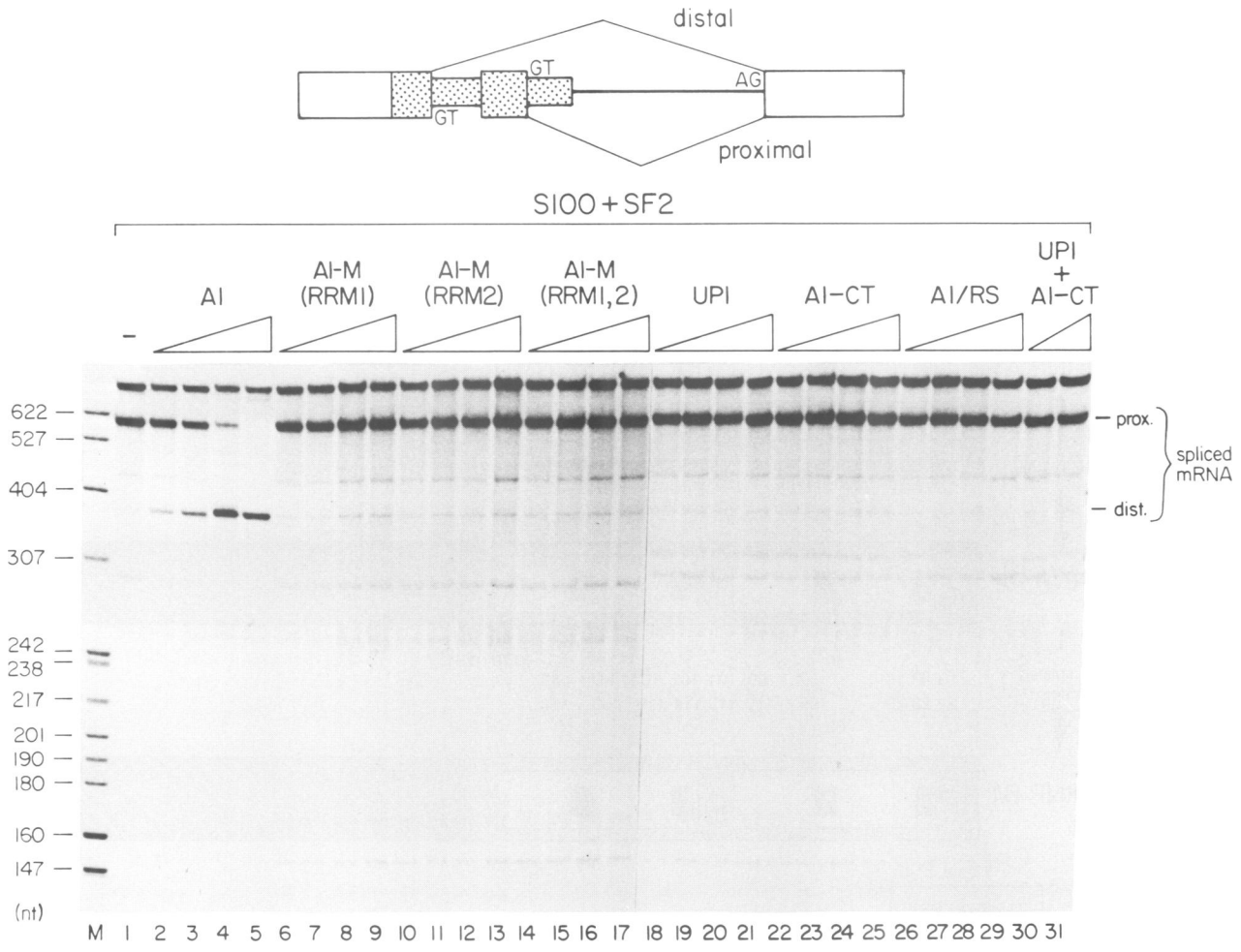

Fig. 5. Effect of hnRNP A1 mutations on alternative 5′ splice site switching. The structure of the model β-globin pre-mRNA with duplicated 5′ splice sites and the two possible splicing paths are shown schematically at the top. Splicing reactions contained 5 μl of S100 extract complemented with 0.4 μM SF2/ASF purified from HeLa cells, plus increasing amounts of the indicated hnRNP A1 proteins: [−, none (lanes 1); each set of four lanes (lanes 2–29) corresponds to 0.1, 0.2, 0.4 and 0.8 μM final protein concentration from left to right; in lanes 30 and 31, UP1 and A1-CT were each present at 0.2 μM and 0.8 μM, respectively]. The positions of the spliced mRNAs generated by selection of proximal or distal 5′ splice sites are indicated. pBR322/*Hpa*II DNA markers are shown (lanes M) and their sizes indicated.

for selecting distal alternative 5′ splice sites by counteracting the activity of SF2/ASF. In parallel experiments, we determined that both wild-type and mutant hnRNP A1 proteins were stable in the S100 extract during the splicing reaction (unpublished data).

Surprisingly, each of the A1-M(RRM1), A1-M(RRM2) and A1-M(RRM1,2) mutant proteins completely lacked alternative splicing activity (Figure 5, lanes 6–17). Likewise, the UP1 and A1-CT fragments had no activity (lanes 18–25). We conclude that both RRMs and the G domain are required for alternative splicing activity. We did not detect intra-cistronic complementation between separate UP1 and A1-CT polypeptides (lanes 30 and 31), indicating that the RRMs and the G domain of hnRNP A1 cooperate only when present within the same polypeptide.

The chimeric A1/RS protein, which consists of UP1 joined to the C-terminal RS domain from SF2/ASF, also lacked detectable 5′ splice site switching activity (Figure 5, lanes 26–29). Thus, the G domain of hnRNP A1 has a unique and specific function that apparently cannot be provided by an RS domain. The A1/RS protein also lacked functions specific to SF2/ASF, i.e. distal to proximal 5′ splice site switching activity and constitutive splicing activity (unpublished data). This result suggests that the N-terminal pairs of RRMs in hnRNP A1 and SF2/ASF have different specificities and functions.

In contrast to the results obtained in the general RNA binding and RNA annealing assays, the loss of switching activity in A1-M(RRM1), A1-M(RRM2), UP1, A1-CT and A1/RS suggests that specific RNA−protein contacts mediated cooperatively by both the G domain and the two pairs of solvent-exposed Phe residues in RRM1 and RRM2 are essential for the alternative splicing function of hnRNP A1 (summarized in Figure 8).

Functional assays of other hnRNP A/B group proteins

Since other hnRNP A/B proteins have the same domain structure as hnRNP A1 (Figure 6; reviewed in Dreyfuss *et al.*, 1993), we characterized the biochemical activities of all the human proteins of this class, except for hnRNP B2, whose amino acid sequence is unknown. In our original purification of the alternative splicing activity, we did not detect strong activity in fractions containing other

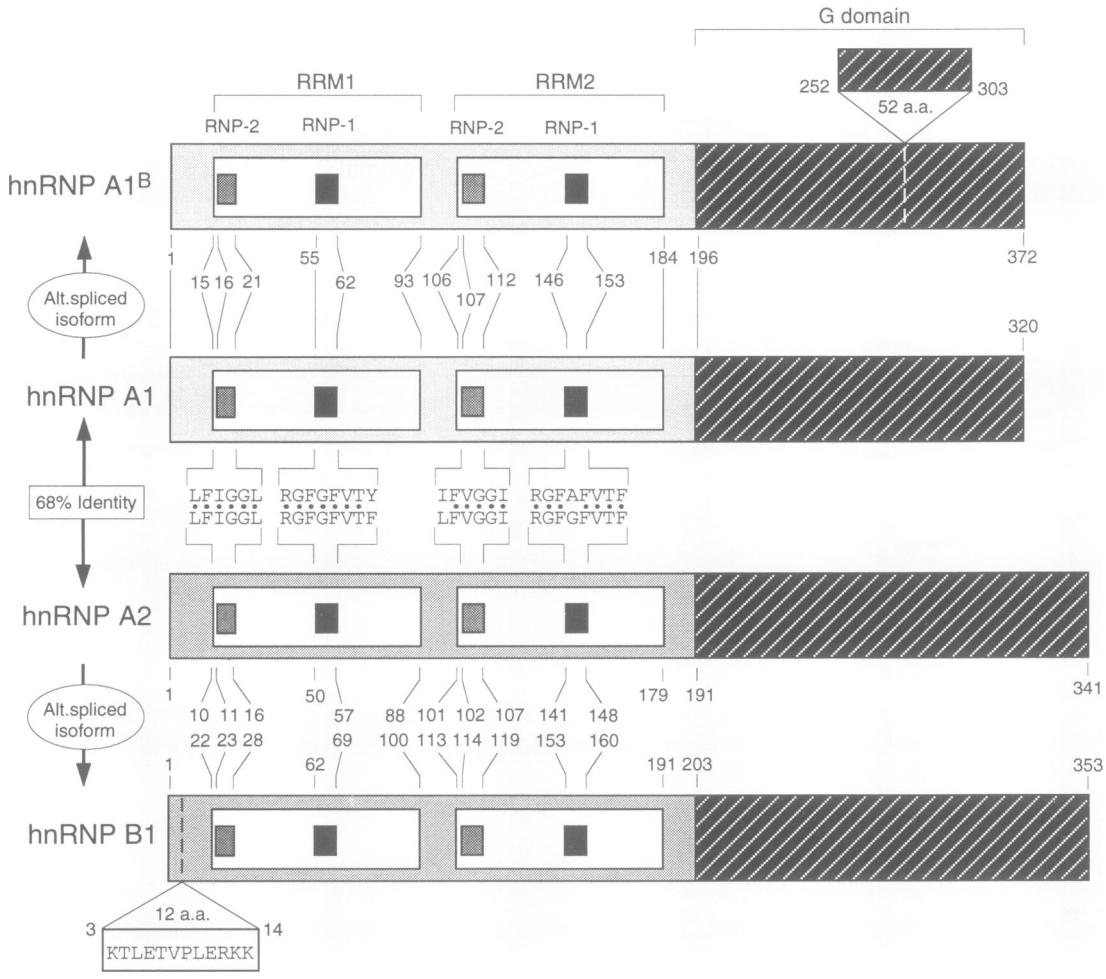

Fig. 6. Structure of human core hnRNP proteins A1, A1^B, A2 and B1. The RNP-2 and RNP-1 conserved submotifs of RRM1 and RRM2, and the G domains of each protein are shown. hnRNP A1 and A1^B or hnRNP A2 and B1 are identical except for extra amino acid regions indicated by boxes. The sequences of the RNP-1 and RNP-2 submotifs are aligned. The dots in the alignment indicate amino acid identities. All recombinant proteins are in authentic form except for post-translational modifications (see Figure 1 legend). The numbers indicate the position of amino acid residues from the initiation codon Met1. These diagrams are based on published cDNA sequences (Burd *et al.*, 1989; Buvoli *et al.*, 1990a).

hnRNP proteins (Mayeda and Krainer, 1992). However, this may have been due to their partial loss during purification, the lower abundance of these proteins in the active fractions, or their co-fractionation with interfering activities. Therefore, we assayed other hnRNP A/B proteins individually after expressing their cloned cDNAs in *E.coli*. The cDNAs and genes encoding human hnRNP A1, A1^B, A2 and B1 have been cloned and sequenced (Buvoli *et al.*, 1988, 1990a; Burd *et al.*, 1989; Biamonti *et al.*, 1989, 1994). hnRNP A1^B is identical to hnRNP A1 except for the presence of an additional 52 amino acid insert within the G domain, which is generated from the same pre-mRNA by inclusion of an alternative exon (Buvoli *et al.*, 1990a). hnRNP A2 and B1 cDNAs are identical except for 36 in-frame nucleotides present only in hnRNP B1, which result from alternative splicing and give rise to a 12 amino acid insert near the N-terminus of hnRNP B1 (Burd *et al.*, 1989; Biamonti *et al.*, 1994).

Human hnRNP A1^B, A2 and B1 were over-expressed in *E.coli*, extensively purified and assayed for general RNA binding, RNA annealing and 5′ splice site switching activities (Figure 7 and unpublished data; summarized in Figure 8). hnRNP A1^B bound RNA more strongly than hnRNP A1, whereas hnRNP A2 and B1 bound RNA more weakly. The first result is consistent with previous data showing that hnRNP A1^B has higher affinity than hnRNP A1 for single-stranded DNA (Buvoli *et al.*, 1990a). hnRNP A2 and B1 had substantial RNA annealing activity, although less than hnRNP A1 and A1^B (Figure 8). hnRNP A1^B has a longer G domain (Figure 6; Buvoli *et al.*, 1990a) whereas the G domains of hnRNP A2 and B1 are considerably divergent from that of hnRNP A1 (Burd *et al.*, 1989). Taken together with the UP1 results (Figures 2, 4 and 8), these experiments show that both general RNA binding and RNA annealing activities of the hnRNP A/B proteins are dependent on the length and sequence of the C-terminal G domain.

Alternative splicing activity was assayed using the same β-globin pre-mRNA derivative as in Figure 5. Surprisingly, hnRNP A1^B caused only a slight switch from proximal to distal alternative 5′ splice site use (Figure 7A, compare lanes 1–7 and 8–11), despite the fact that it bound RNA more strongly than hnRNP A1 (Figure 8). Hence, the extended Gly-rich region of hnRNP A1^B has a significant repressive effect on alternative splicing.

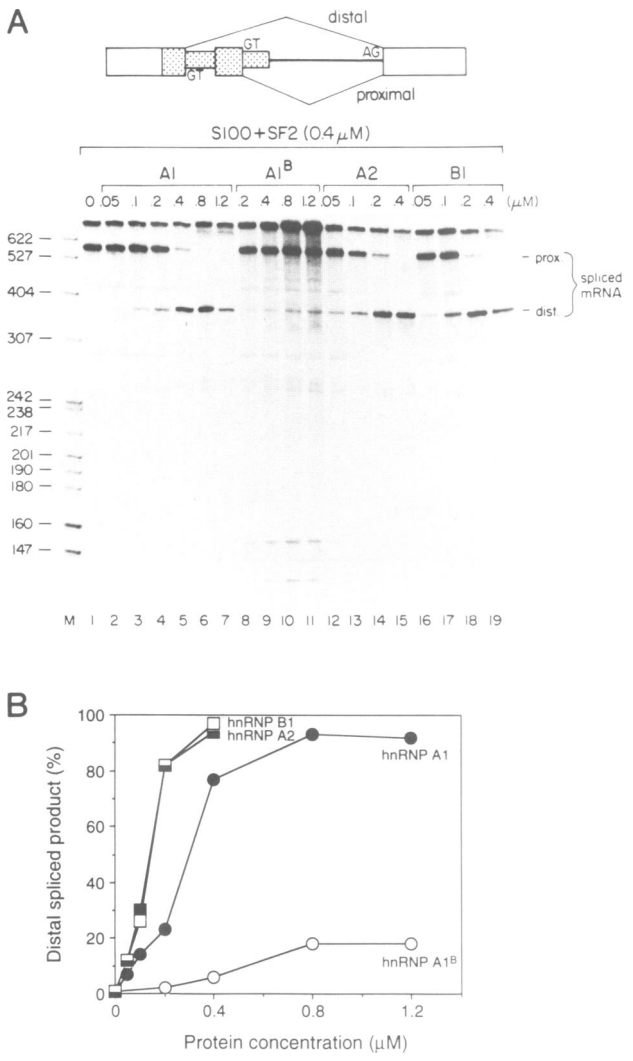

Fig. 7. Activities of hnRNP A/B core proteins on alternative 5′ splice site selection. (**A**) Alternative splicing assays were carried out as in Figure 5, in the presence of the indicated concentration of each hnRNP protein. (**B**) Quantification of 5′ splice site switching specific activities. The percentage of total spliced mRNAs resulting from use of the distal 5′ splice site is plotted as a function of the concentration of each protein, based on the data in (A).

In contrast to hnRNP $A1^B$, hnRNP A2 and B1, which have 68% amino acid identity with hnRNP A1, had stronger switching activity than hnRNP A1 (Figure 7A, compare lanes 1–5, 12–15 and 16–19). By measuring the amounts of spliced products generated using serial dilutions of the proteins, we estimate that hnRNP $A1^B$ has 5-fold lower specific activity than hnRNP A1, whereas hnRNP A2 and B1 have 2-fold higher specific activity than hnRNP A1 (Figure 7B; compare percent distal spliced product: 20 pmol hnRNP $A1^B$ is equivalent to ~4 pmol of hnRNP A1, and 10 pmol hnRNP A2 or B1 is comparable with 20 pmol hnRNP A1). A similar hierarchy of the specific activities of hnRNP A1, $A1^B$, A2 and B1 was observed with an adenovirus E1A pre-mRNA (unpublished data). Hence, one or more of these proteins may correspond to DSF, an activity that stimulates E1A distal 5′ splice sites (Harper and Manley, 1991).

Discussion

We have analyzed the role of individual domains of hnRNP A1 with respect to RNA binding, RNA annealing and alternative splice site selection. Each of the three domains of hnRNP A1 contributes to the RNA binding and annealing activities of the intact protein, and each plays a unique and essential role in directing the selection of alternative 5′ splice sites. However, the different functions of hnRNP A1 appear to have different requirements for RNA binding specificity. Non-conservative replacements of two conserved Phe residues in either RRM1 or RRM2 had minimal effects on non-specific RNA–protein interactions or RNA annealing, but led to a complete loss of the ability of hnRNP A1 to direct splicing to the distal 5′ splice site. Thus, the concentration-dependent effects of hnRNP A1 on alternative splicing, unlike general RNA binding and annealing, appear to involve specific interactions between RNA and aromatic residues in both RRMs. Taken together, these results suggest that hnRNP A1 displays a range of specificities in its interactions with different sites on pre-mRNAs *in vivo*. The specificity of binding to a particular site on RNA probably reflects multiple RNA–protein interactions within individual domains of hnRNP A1 and interactions between the domains within a single protein, in addition to interactions with other proteins bound to adjacent sites. We also found that three other closely related hnRNP A/B proteins differ quantitatively from hnRNP A1 in their RNA binding affinity and their ability to modulate alternative splice site selection.

Effect of hnRNP A1 mutations on general RNA binding

Our results confirm previous studies that showed that both the N-terminal fragment, UP1, and the C-terminal fragment bind to nucleic acids, although this binding is substantially weaker than that observed with intact hnRNP A1 protein (Herrick and Alberts, 1976; Kumar *et al.*, 1986, 1990; Cobianchi *et al.*, 1988; Merrill *et al.*, 1988; Nadler *et al.*, 1991; Casas-Finet *et al.*, 1993). We investigated the role of individual RRMs using a series of mutant proteins altered by non-conservative substitution of two Phe residues within the highly conserved RNP-1 submotif of each RRM. These solvent-exposed residues have been shown to play an important role in RNA binding (Merrill *et al.*, 1988; Nagai *et al.*, 1990; Hoffman *et al.*, 1991; Cáceres and Krainer, 1993). The results of filter binding studies showed only a small effect of replacing these two Phe residues with Asp in either of the two RRMs. Since an RRM forms extensive interactions with bound RNA (Görlach *et al.*, 1992), general RNA binding may be relatively insensitive to these substitutions in a single RRM. However, a substantial reduction in RNA binding was observed when the pairs of Phe residues were substituted in both RRMs. Thus, the filter binding results suggest that these tandem RRMs play an essential but redundant role in RNA binding.

Substitution of analogous Phe residues within RRM1 or RRM2 had markedly different effects on the UV-induced crosslinking of the mutant proteins to RNA. This observation highlights possible differences in RNA–protein contacts within each RRM. This result is consistent

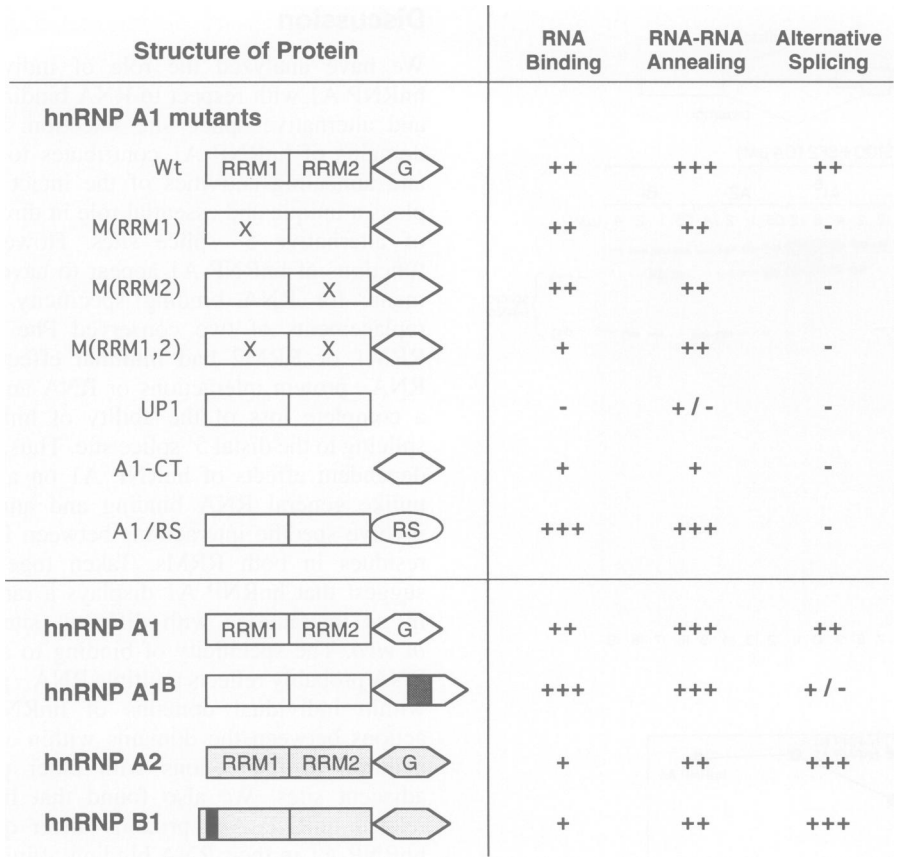

Fig. 8. Summary of the activities of hnRNP A1 mutants and other hnRNP A/B proteins. The structures of the wild-type (Wt) and mutant proteins are shown schematically at left. The RRMs are indicated by boxes, the G domains by hexagons, and the RS domain from SF2/ASF by an ellipse. The X represents a double Phe to Asp replacement in the RNP-1 submotif of the indicated RRM (see Figure 1 for details). The extra amino acid inserts in hnRNP $A1^B$ and B1 are indicated by gray or black shading, respectively. The relative specific activities for general RNA binding (measured by a filter binding assay), for RNA annealing, and for alternative 5′ splice site switching are indicated by the number of + signs; +/− denotes trace activity and − indicates no detectable activity. See Figures 2, 4, 5 and 7 for data used in scoring these values.

with the finding of Merrill *et al.* (1988) that residues Phe17 and Phe59 within RRM1 account for 75% of total UV crosslinking to oligo(dT), the remaining 25% being due to Phe108 and Phe150 in RRM2. Thus, substitutions of Phe residues at these crosslinking sites in RRM1, or near them in the tertiary structure (Garrett *et al.*, 1994), had a more drastic effect on RNA−protein crosslinking than the analogous substitutions in RRM2. We obtained similar results with the corresponding UP1 mutants, demonstrating that the relative crosslinking efficiencies of the RRMs are independent of the C-terminal G domain.

The observation that the double mutant, A1-M(RRM1,2), crosslinks more efficiently than A1-M(RRM1) suggests that wild-type RRM2 interferes directly or indirectly with RNA crosslinking to RRM1. One possible explanation, which is consistent with our filter binding studies, is that individual domains may compete for binding to RNA, with only one of the two RRMs of A1 forming tight contacts at any given time. Thus, when only RRM1 was mutated, crosslinking but not filter binding dropped, because more protein bound via RRM2 and/or the G domain, which crosslink relatively inefficiently to RNA. This model for hnRNP A1 binding is in agreement with a model proposed by Casas-Finet *et al.* (1991, 1993), based on biophysical studies that suggested that each molecule of UP1 binds to nucleic acid homopolymers through only one of its two RRMs.

RNA annealing activity of hnRNP A/B proteins

A number of RNA-binding proteins recently have been shown to promote rapid annealing of complementary nucleic acid strands. These include four hnRNP A/B proteins: hnRNP A1 (Kumar and Wilson, 1990; Pontius and Berg, 1990; Munroe and Dong, 1992), $A1^B$, A2 and B1 (this study), as well as hnRNP U, hnRNP C1 (Portman and Dreyfuss, 1994) and the splicing factors SF2/ASF (Krainer *et al.*, 1990a) and $U2AF^{65}$ (Lee *et al.*, 1993). All these proteins have a terminal auxiliary domain characterized by a quasi-repetitive motif, and all except hnRNP U include at least one RRM. Previous studies have suggested that each of these types of domains plays a direct role in facilitating annealing. The relatively unstructured, quasi-repetitive domains of these proteins, such as the G domain found in the hnRNP A/B proteins, may promote annealing by mediating transient interactions between complementary strands, which enhance the nucleation of annealing (Pontius and Berg, 1990; Munroe and Dong, 1992; Pontius, 1993). The highly structured RRMs, on the other hand, may facilitate annealing by binding specifically to single-stranded nucleic acids and unwinding intramolecular secondary structure, which provides a kinetic barrier to formation of stable, intermolecular pairing (Munroe and Dong, 1992; Portman and Dreyfuss, 1994). Our results show that these two types of domains contribute to the annealing activity of hnRNP A1. Both

the highly structured RRMs present in UP1 and the repetitive G domain of the A1-CT polypeptide displayed annealing activity. Also, substitution of the RS domain of SF2/ASF for the G domain of hnRNP A1 resulted in wild-type levels of annealing.

We observed a general correlation between the RNA annealing and RNA binding activities of the four hnRNP A/B proteins and the hnRNP A1 and UP1 mutants (Figure 8). This correlation suggests that both activities, which involve relatively non-specific RNA–protein interactions, have similar structural requirements. However, there were exceptions to this general correlation. The single RRM mutants bound RNA almost as well as the wild-type protein, but their annealing activities were reduced. This result suggests that the two RRMs of hnRNP A1 cooperate with each other, and/or with the G domain, to promote annealing activity. The RNA annealing activity of the double RRM mutant was similar to that of each single mutant, although its binding to RNA was substantially weaker. Also, the double RRM mutant had substantially more annealing activity than the A1-CT polypeptide consisting of only the G domain. Since A1-CT bound RNA somewhat more tightly than A1-M(RRM1,2) in the filter binding assay, it appears that annealing is not a simple function of RNA binding. These results also indicate that the mutated RRMs retain substantial RNA annealing activity, or that the activity of the C-terminal domain is enhanced through its attachment to the N-terminal portion of hnRNP A1. The RRMs and the G domain cooperated to give maximal RNA annealing activity only when present in the same polypeptide.

Structural requirements for alternative splicing activity

The complete loss of alternative splicing activity observed upon replacing two conserved Phe residues in either RRM contrasts sharply with the residual binding and annealing activity observed with the single and double RRM mutants of hnRNP A1. We conclude that general RNA annealing activity is not sufficient for the alternative splicing activity of hnRNP A1, although our results do not rule out a direct role for RNA annealing activity in this process. These results also indicate that RNA binding alone is insufficient for the splice selection activity of hnRNP A1. This conclusion is reinforced by the negative correlation observed between RNA binding and alternative splicing activity of the four hnRNP A/B proteins studied. hnRNP A2 and B1, which bind RNA more weakly, are more active than hnRNP A1 in promoting selection of distal 5′ splice sites. In contrast, hnRNP $A1^B$, which binds RNA more strongly than hnRNP A1, is much less active. The stringent requirement for the conserved Phe residues suggests that the alternative splicing activity of hnRNP A1 may require more specific interactions with the pre-mRNA substrate than general RNA binding or annealing. Taken together with the conservation of the substituted residues in most RRMs, their solvent-exposed position at the center of the RNA binding surface, and their close proximity to bound RNA, our results suggest that these aromatic residues participate in sequence-specific RNA–protein interactions that are essential for alternative splicing activity.

At present we do not know the nature of the proposed specific RNA–protein interactions disrupted by replacement of the RNP-1 Phe residues. Several recent reports have demonstrated that hnRNP A/B proteins can bind nucleic acids in a sequence-specific manner. Preferential binding of hnRNP A1 or UP1 to 5′ or 3′ splice sites has been obtained, both in crude extracts and with purified protein (Swanson and Dreyfuss, 1988; Buvoli *et al.*, 1990b; Stolow and Berget, 1991; Stolow, 1992). The hnRNP A/B proteins were recently identified as some of the major mammalian nuclear proteins that bind with high affinity to telomeric DNA repeats $(TTAGGG)_n$ and to the RNA equivalent, $(UUAGGG)_n$ (McKay and Cooke, 1992; Ishikawa *et al.*, 1993). SELEX experiments with purified hnRNP A1 yielded the high-affinity binding site consensus UAGGGA/U, which shows some resemblance to both consensus 5′ and 3′ splice sites (Burd and Dreyfuss, 1994). A different study, which implicated cytoplasmic hnRNP A1 in mRNA turnover, demonstrated specific binding of purified hnRNP A1 to RNAs containing AUUUA sequences (Hamilton *et al.*, 1993).

It will be interesting to see if the RRM1 and RRM2 mutations have greater effects on binding of hnRNP A1 to high affinity sites present in short RNAs than on overall binding to long pre-mRNAs, and whether binding to specific sites on the latter can be enhanced by the presence of other proteins in the nuclear extract. Relatively long pre-mRNAs, such as those used in this study, probably contain multiple binding sites with a range of affinities. Only a small subset of these binding sites may be important for splice site selection.

Recent detailed structural studies of SF2/ASF activities demonstrated that the conserved Phe residues in the RNP-1 submotif of the N-terminal RRM are essential for the constitutive splicing activity of this protein, but not, in contrast to the above results with hnRNP A1, for alternative 5′ splice site selection (Cáceres and Krainer, 1993). The RS domain of SF2/ASF, and specifically both Arg and Ser residues, are also required for efficient constitutive splicing activity, whereas the N-terminal RRM and a central atypical RRM are sufficient to promote selection of proximal 5′ splice sites in the absence of an RS domain (Cáceres and Krainer, 1993; Zuo and Manley, 1993). Therefore, analogous mutations in the RRMs of hnRNP A1 and SF2/ASF have different effects on the alternative splicing activities of these two proteins, and the requirement for the respective C-terminal auxiliary domains also differs. These observations suggest that the RNA–protein contacts that mediate 5′ splice site modulation by SF2/ASF and by hnRNP A1 are not equivalent.

Regulation of alternative splicing by hnRNP A/B proteins

The observation that four closely related human hnRNP A/B group proteins display quantitatively different activities in promoting selection of distal 5′ splice sites raises a number of interesting points. hnRNP $A1^B$, which is an alternatively spliced isoform of hnRNP A1, displays only very limited ability to alter 5′ splice site selection. This decrease in the alternative splicing activity of hnRNP $A1^B$, relative to hnRNP A1, was observed with two different substrates *in vitro*, and is consistent with a recent study of the alternative splicing activities of these proteins *in vivo* (Yang *et al.*, 1994). The reduced activity of hnRNP

$A1^B$ may be due to the fact that this protein binds these pre-mRNAs more tightly than hnRNP A1, and thus may interfere with subsequent recognition of the substrate by other splicing factors. It is possible that hnRNP $A1^B$ has strong activity with certain pre-mRNA substrates to which it binds less tightly. Differential tissue-specific expression of hnRNP A1 and $A1^B$ was detected at both mRNA and protein levels (Buvoli *et al.*, 1990a; A.Hanamura and A.R.Krainer, in preparation). The differential expression of these isoforms, together with their different specific activities, suggests that alternative splicing of hnRNP $A1/A1^B$ pre-mRNA may itself provide a mechanism to modulate alternative splicing *in vivo* in a tissue-specific manner.

In contrast to hnRNP $A1^B$, hnRNP A2 and B1 promote distal 5′ splice site selection with greater efficiency than hnRNP A1. These two protein isoforms, which are generated by alternative splicing (Burd *et al.*, 1989; Biamonti *et al.*, 1994), had equivalent alternative splicing activities with two different substrates. The significance of the existence of two isoforms that are nearly identical in structure is unknown. One possibility is that A2 and B1 differ in their substrate specificities, but these differences are not apparent with every pre-mRNA, or cannot be reproduced *in vitro*. Another possibility is that hnRNP A/B proteins have other functions besides modulating alternative 5′ splice site selection and hence, the individual proteins may be partially but not completely redundant.

A similar situation applies to the SR protein family (reviewed in Horowitz and Krainer, 1994). In a number of *in vitro* assays of constitutive and alternative splicing, several of the SR proteins appear to be functionally interchangeable (Fu *et al.*, 1992; Mayeda *et al.*, 1992; Zahler *et al.*, 1992, 1993b). However, more recent experiments have uncovered apparent substrate-specific differences between individual SR proteins (Kim *et al.*, 1992; Fu, 1993; Sun *et al.*, 1993; Tian and Maniatis, 1993; Zahler *et al.*, 1993a). Such functional differences would explain the remarkable phylogenetic sequence conservation of every member of both the SR protein family and the hnRNP A/B group proteins (reviewed in Birney *et al.*, 1993). Although more distant family members exhibit substantial functional overlap, at least *in vitro*, the existence of at least one essential gene whose expression can only be properly regulated by a single SR and hnRNP A/B protein pair would provide the necessary selective pressure to prevent sequence divergence or loss of those family members.

With respect to other potential functions of the hnRNP A/B proteins, it appears that these proteins are bound to many, if not all, pre-mRNAs, including those that do not undergo alternative splicing (reviewed in LeStourgeon *et al.*, 1990; Dreyfuss *et al.*, 1993). It is possible that at least one of these proteins is necessary for general splicing. A mouse cell line has been isolated, which has greatly reduced expression of hnRNP A1, as a result of deletion of one allele and retroviral integration in the 3′ untranslated region of the remaining allele (Ben-David *et al.*, 1992; Yang *et al.*, 1994). This finding shows that hnRNP A1 is dispensable, at least in cell culture. It is likely, however, that hnRNP A2 and B1 can functionally compensate for the absence of hnRNP A1, although this might not be the case at the organismal level. For example, genetic analysis of two different hnRNP A/B proteins in *Drosophila* has uncovered essential roles for these proteins in spermatogenesis or oogenesis (Karsch-Mizrachi and Haynes, 1993; Matunis *et al.*, 1994).

The apparent functional redundancy among individual members of the hnRNP A/B group proteins and the SR protein family may reflect the potential diversity of alternatively spliced pre-mRNAs that are subject to regulation by these proteins *in vivo*. The finding that multiple hnRNP A/B proteins can each antagonize one or more SR proteins increases the number of pairwise combinations of these factors that can potentially modulate alternative splicing *in vivo*. We propose that specific pairs of antagonistic hnRNP A/B and SR proteins may have evolved optimally to regulate the expression patterns of specific subsets of alternatively spliced pre-mRNAs.

Materials and methods

Construction of wild-type and mutant hnRNP A/B cDNA expression plasmids

Conditions for PCR or RT-PCR were essentially as described (Krainer *et al.*, 1991), with annealing temperatures optimized for maximal specific product yield in each reaction. The proofreading Vent polymerase (New England Biolabs) was used to minimize unwanted mutations. Plasmid constructs with the correct sequences were identified by nucleotide sequencing and/or the electrophoretic mobility of the expressed proteins (see below). The *E.coli* strain DH5α was used for all plasmid constructs.

The hnRNP A1 cDNA insert in the plasmid pET9d-hnRNP A1 (Mayeda and Krainer, 1992) was modified to express variant forms of the hnRNP A1 protein. The plasmid expressing A1-M(RRM1), which contains Asp residues in place of Phe57 and Phe59 in RRM1, was constructed by replacing a restriction fragment in the wild-type gene with a corresponding PCR-generated fragment bearing the desired mutations. PCR amplification of the N-terminal half of hnRNP A1 cDNA was carried out with the hnRNP A1 N-terminal oligonucleotide, oligo #1 (5′-GTCACATGTCTAAGTCAGAGTCTCCTAAAGAG-3′; Mayeda and Krainer, 1992) and oligo #2 (5′-TGGCATATGTGACA**T**-CCCCA**TC**GCCCCTAGAG-3′). Oligo #2 spans the RNP-1 sequence in RRM1 and extends to a unique *Nde*I site in the cDNA. The 96 nucleotide *Bsu*36I−*Nde*I fragment of hnRNP A1 cDNA was replaced with the same PCR-generated fragment containing the bold substitutions in oligo #2, converting two TTT codons for Phe to GAT codons for Asp. A plasmid expressing A1-M(RRM2), which contains Asp residues in place of Phe148 and Phe150 in RRM2, was similarly constructed. Sequences corresponding to RRM2 and the C-terminal half of hnRNP A1 were amplified with oligo #3 (5′-AAATCATGACTGACCG**C**GGCAGTGGC-AAGAAACGTGGC**GA**TGCC**GA**TGTAACCTTTG-3′) and the hnRNP A1 C-terminal oligo #4 (5′-GCTGGATCCTTAAAATCTTCTGCCAC-TGCCAT-3′; Mayeda and Krainer, 1992). The *Bsp*HI−*Bam*HI fragment (561 nt) of wild-type cDNA was replaced by the corresponding PCR-generated fragment containing the bold substitutions. In addition to the Phe to Asp double substitution, oligo #3 introduces a silent A to C transversion, creating a *Sac*II site that was employed in screening.

Plasmids expressing UP1 were constructed by inserting a *Bcl*I linker (5′-CTGATCAG-3′) at the unique *Bam*HI site at the end of the coding region in pET9d-hnRNP A1, and at a *Taq*I site spanning codons 195 and 196. The *Bcl*I linkers were added after filling the 5′ overhangs at the *Bam*HI and *Taq*I sites. After deleting the intervening *Taq*I−*Bam*HI fragment, and religating the new *Bcl*I linkered ends, the linker provides the third nucleotide (C) of the Arg196 codon immediately followed by an in-frame TGA termination codon. The resulting plasmid codes for a recombinant UP1 protein with exactly the same sequence as that reported for the proteolytic hnRNP A1 fragment UP1 (Williams *et al.*, 1985). Three UP1 variants were constructed by incorporating the same substitutions as in hnRNP A1. The plasmid expressing the UP1-M(RRM1) mutant was generated by replacing the 3190 bp *Xho*I−*Nde*I fragments of UP1 with the corresponding fragment from A1-M(RRM1). UP1-M(RRM2) was constructed by replacing the *Xba*I−*Bcl*I fragment of the UP1 construct with the *Xba*I−*Taq*I fragment (627 nt) of A1-M(RRM2), modified by addition of the *Bcl*I linker to the filled *Taq*I site. A1-M(RRM1,2) and UP1-M(RRM1,2), which include the Phe to Asp double substitutions in

both RRMs, were constructed by combining the 3190 bp *Xho*I–*Nde*I fragments from the M(RRM1) mutant with the 2080 or 1710 bp *Nde*I–*Xho*I fragment representing the opposite half of the M(RRM2) plasmid expressing hnRNP A1 or UP1, respectively. The plasmid encoding the C-terminal fragment of hnRNP A1, A1-CT, was constructed by inserting the *Taq*I–*Bam*HI fragment from the 3′ end of the cDNA sequence into the pET9d expression vector (Novagen) cut with *Nco*I and *Bam*HI, after filling in *Taq*I and *Nco*I overhangs with Klenow fragment. This procedure restored the Met1 initiation codon, joining it to the Arg196 codon, which is the second of 126 residues in the Gly-rich A1-CT polypeptide. The entire length of all PCR-generated inserts and the junction regions of the UP1 and A1-CT constructs were sequenced to verify their structure. The sequences of the coding regions were as predicted in all cases. A single G residue downstream of the termination codon in the *Bcl*I linker was missing at the junction between this linker and the filled *Bam*HI site.

Two steps of PCR amplification were employed to generate the A1/RS chimeric protein expression plasmid. First, pET9d-hnRNP A1 (Mayeda and Krainer, 1992) and pET19b-SF2 (Cáceres and Krainer, 1993) were independently amplified with oligo pairs #5 (5′-TTTGT-CACATATGCCACTGT-3′)/#6 (5′-TACGGGGCCCCTCTTGCTTTG-ACAG-3′) and #7 (5′-AAAGCAAGAGGGGCCCCGTAGCC-CG-3′)/#8 (5′-CTTGGATCCTTAGGTACGAGA-3′), respectively. 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 #5 and #8. The resulting PCR product was purified, digested with *Nde*I and *Bam*HI, and subcloned into the corresponding sites of pET9d-hnRNP A1 (Mayeda and Krainer, 1992). The resulting plasmid pET9d-A1/RS codes for a chimeric protein of 238 amino acids, comprising amino acids 1–185 of hnRNP A1 and 195–248 of SF2/ASF.

The hnRNP $A1^B$ expression plasmid was generated by RT-PCR from HeLa poly(A)$^+$ RNA as described (Krainer *et al.*, 1991). Two separate fragments were generated with the N-terminal A1 oligo #1 (see above) and with oligo #9 (5′-CTCCGCCTCCGTTGTTATAGCTGTCATAGCT-3′) or with oligo #10 (5′-CGGCCCTGGTTACTCTGGAGGAAGC-AGA-3′) and with the C-terminal A1 oligo #4 (see above). The first product was purified and digested with *Aff*III and *Bst*XI, while the second product was digested with *Bst*XI and *Bam*HI. These fragments were ligated together into pET9d digested with *Nco*I and *Bam*HI. The cDNA inserts of hnRNP A1 and $A1^B$ were sequenced in their entirety. They were identical to the published sequences containing the Tyr128 and Arg146 polymorphisms (Buvoli *et al.*, 1988, 1990a), except for two silent mutations within the codons for Ser54 (TC**C**) and Gly248 (GG**T**).

The hnRNP A2 and B1 expression plasmids were also generated from two separate RT-PCR fragments. A common C-terminal fragment was amplified with oligo #11 (5′-GCAGAAATACCATACCATCAATGGTC-3′) and oligo #12 (5′-GCTGGATCCATGGCAAATAGGAAGAAGC-TCA-3′), followed by digestion of the gel-purified product with *Xba*I and *Bam*HI. N-terminal fragments were amplified with oligo #13 (5′-CTTCCACTCCTAGAACTCTGAACTTC-3′) and either oligo #14 (5′-CACGAGCATATGGAGAGAGAAAAGGAACAG-3′) in the case of hnRNP A2, or oligo #15 (5′-CACGAGCATATGGAGAAAACTTTA-GAAACT-3′) in the case of hnRNP B1. Both purified products were digested with *Nde*I and *Xba*I. N- and C-terminal fragments for either protein were ligated together into pET9c digested with *Nde*I and *Bam*HI.

Bacterial expression and purification of recombinant proteins

Expression of all 13 recombinant proteins in the *E.coli* strain BL21(DE3)-pLysS (Novagen) and purification of recombinant hnRNP A1, A1-M(RRM1) and hnRNP $A1^B$ proteins were carried out as described (Mayeda and Krainer, 1992). Pure hnRNP A1, A1-M(RRM1) and $A1^B$ proteins were respectively recovered in the 0.84–0.46, 1.0–0.6 and 0.24–0 M NaCl gradient fractions of the final phenyl-Superose column (Pharmacia).

hnRNP A2, hnRNP B1, UP1, UP1-M(RRM1) and UP1-M(RRM2) proteins were recovered from the supernatant of the sonicated *E.coli* lysates as described (Mayeda and Krainer, 1992). hnRNP A2 and B1 proteins were prepared by the same method as hnRNP A1, except that different conditions were used for elution from the phenyl-Superose column. Since these proteins were highly hydrophobic, the bound material was eluted with a 30 ml linear gradient from 2 M NaCl to 0 M NaCl plus 50% ethylene glycol in buffer A [20 mM HEPES-Na^+, pH 8.0, 0.2 mM EDTA, 5% (v/v) glycerol, 1 mM DTT]. Pure hnRNP A2 and B1 proteins were recovered in the 40–50% ethylene glycol fractions. UP1 and UP1-M(RRM1) proteins were dialyzed in buffer B (buffer A containing 0.5 mM PMSF) after fractionation by CsCl density gradient centrifugation, and applied to a heparin–Sepharose column (1 ml HiTrap affinity column, Pharmacia). The column was washed with 10 ml of buffer A and eluted with a 22 ml linear gradient from 0 to 0.25 M KCl in buffer A. UP1-M(RRM2) protein became insoluble during CsCl centrifugation; the insoluble material was subjected to the following denaturation/renaturation procedure. The insoluble pellet was dissolved in 6 ml of buffer C (buffer B plus 0.1 M KCl) containing 6 M guanidine hydrochloride by brief sonication, rocked for 30 min at 4°C, and dialyzed in 1 liter of buffer B. The dialyzed UP1-M(RRM2) protein was loaded on a heparin–Sepharose column as described above. The pure UP1, UP1-M(RRM1) and UP1-M(RRM2) proteins eluted at 0.17–0.2, 0.13–0.17 and 0.08–0.13 M KCl, respectively.

A1-M(RRM2), A1-M(RRM1,2), UP1-M(RRM1,2) and A1-CT proteins were recovered from the pelleted inclusion bodies of sonicated *E.coli* lysates, as described (Cáceres and Krainer, 1993). The pellet was dispersed by sonication in 30 ml buffer C containing 1% sodium deoxycholate. After centrifugation at 16 500 r.p.m. for 20 min (SS-34 rotor, Sorvall), the pellet was washed with 30 ml of buffer B containing 1 M KCl and then with buffer C by the same procedure. The washed pellet was then dissolved by brief sonication in 14 ml of buffer C containing 6 M guanidine hydrochloride and rocked for 30 min at 4°C. A1-M(RRM2) and A1-M(RRM1,2) were purified by phenyl-Superose chromatography after dialysis in buffer B containing 2 M NaCl as described (Mayeda and Krainer, 1992) and the pure proteins eluted between 0.6 and 0.3 M NaCl. UP1-M(RRM1,2) was purified by heparin–Sepharose chromatography after dialysis in buffer B and the pure protein eluted at ~0.03–0.07 M KCl. Since A1-CT protein remained highly insoluble, the guanidine-denatured inclusion bodies containing A1-CT were dialyzed in buffer B containing 6 M urea. The dialyzed sample was loaded onto a 1 ml Mono S column (Pharmacia), washed with 20 ml buffer A containing 6 M urea, and eluted with a 30 ml linear gradient from 0 to 0.5 M NaCl in buffer A containing 6 M urea. Pure A1-CT protein eluted between 0.1 and 0.16 M NaCl.

After purification, all recombinant proteins except A1-CT were finally dialyzed in 1 liter of buffer C. The pure fraction of A1-CT was first dialyzed in 0.5 l buffer C containing 3 M urea, and then gradually dialyzed into 0.75 M urea by slow addition of buffer C, and finally dialyzed in 1 liter of buffer C. This slow dialysis procedure helped prevent protein aggregation and precipitation. Occasionally some proteins became partially insoluble during dialysis; however, substantial amounts of soluble protein were recovered in the supernatant after a final centrifugation step. To test the effect of the denaturation/renaturation treatment, wild-type hnRNP A1 protein was also denatured by dialysis into buffer C containing 6 M guanidine hydrochloride. After dialysis into buffer C, the renatured hnRNP A1 protein was as active as native hnRNP A1, as determined by serial dilutions of the proteins in the standard alternative splicing assay (unpublished data). Therefore, proteins subjected to the denaturation and renaturation process were assumed to have the correct tertiary structure. The final protein concentrations were estimated by the dye-binding method (Bio-Rad), standardized with BSA. Each protein was adjusted to the desired concentration in the splicing reaction by dilution with buffer A containing 0.1 M KCl.

In vitro splicing assays

Template plasmids pSP64-HβΔ6 for β-globin pre-mRNA, pSP64-HβΔ6-IVS1,2 for β-globin mRNA, and pSP64–5′D16X for a β-globin pre-mRNA derivative containing a duplicated 5′ splice site (Krainer *et al.*, 1984; Reed and Maniatis, 1986; Krainer *et al.*, 1990b) were linearized at the *Bam*HI site near the end of exon 2.

^{7m}GpppG-capped ^{32}P-labeled pre-mRNA substrates were transcribed from each linearized template DNA with SP6 RNA polymerase essentially as described (Krainer and Maniatis, 1985; Mayeda and Ohshima, 1988). *In vitro* splicing reactions were carried out at 30°C for 4 h in 25 µl (Mayeda and Ohshima, 1990) with the indicated amounts of S100 extract, HeLa cell purified SF2/ASF and each wild-type or mutant recombinant hnRNP protein (see legends for Figures 5 and 7). RNA products were analyzed by electrophoresis on a 5.5% polyacrylamide/7 M urea gel followed by autoradiography. The spliced products were quantified using a Fujix Bio-Imaging Analyzer BAS2000 (Fuji Medical Systems, Inc.).

Nitrocellulose filter binding assay

The indicated amount of recombinant protein was mixed with BSA (final concentration 50 µg/ml) and 11 fmol of [α-^{32}P]UTP-labeled β-globin pre-mRNA or mRNA (both truncated at the *Bam*HI site) in a total volume of 110 µl, under standard splicing buffer conditions, but lacking ATP, creatine phosphate and polyvinyl alcohol (PVA) for 20 min

at 30°C (Krainer *et al.*, 1984). A portion of each reaction mixture (100 µl) was filtered through Millipore HAWP filters (mixed cellulose acetate and cellulose nitrate), using a 96-well microfiltration assembly (Schleicher & Schuell) connected to a vacuum pump (Mayeda *et al.*, 1986). Each sample well was washed with 200 µl of buffer A containing 60 mM KCl and 3.2 mM $MgCl_2$ warmed to 30°C. The filter was dried and the radioactivity present in each well was measured with toluene-based cocktail in a scintillation counter (Beckman). Background radioactivity obtained in the absence of protein (usually <2% of input radioactivity) was subtracted from each data point.

UV crosslinking assay

The indicated amounts of recombinant protein plus BSA (final concentration 50 µg/ml) and 20 fmol of β-globin pre-mRNA or mRNA labeled with all four [α-^{32}P]rNTPs, were incubated for 20 min at 30°C under splicing conditions (15 µl), in the absence of PVA (Mayeda and Ohshima, 1990). The reaction mixtures were irradiated in open microcentrifuge tubes with 254 nm UV light at 1.8 J/cm^2 using a UV crosslinker (XL1000, Spectronics Corporation). After digestion with 20 µg of RNase A at 37°C for 30 min, the samples were loaded on a 12% SDS–polyacrylamide gel. The gel was stained with Coomassie Blue R250 and dried, and the crosslinked products were detected by autoradiography.

RNA annealing assay

RNA annealing was assayed by RNase T1 protection using previously described substrates (Munroe and Dong, 1992). HB500, a 500 nt β-globin pre-mRNA truncated at the second exon, and EI200, a 260 nt transcript complementary to 208 nt at the 5′ end of β-globin pre-mRNA, were prepared by transcription *in vitro* as described (Munroe, 1988). Assays were carried out in 25 µl reactions for 16 min at 30°C with 0.5 fmol of HB500 RNA (0.25 pmol nucleotides) and 2.5 fmol of EI-200 RNA under standard splicing conditions, except for omission of ATP, creatine phosphate and PVA (Krainer *et al.*, 1984). Reactions were started by addition of 15 µl protein, diluted in buffer A, to RNA in 10 µl water with RNasin (Promega) and $MgCl_2$. RNA annealing was terminated by addition of 12.5 U of RNase T1 and incubation for an additional 16 min at 30°C. Digestions were stopped by addition of proteinase K, incubated for 30 min, extracted with phenol/chloroform and analyzed by electrophoresis on 5.5% polyacrylamide–7 M urea gels. Dried gels were scanned using a Radioanalytical Imager (Ambis 100).

Acknowledgements

We are grateful to L.Vigue, I.H.Watakabe, A.Disdier and D.Kozak for expert technical assistance. We thank D.S.Horowitz and I.C.Eperon for helpful comments on the manuscript. A.M. was supported in part by a fellowship from the International Human Frontier Science Program Organization. A.R.K. is a Pew Scholar in the Biomedical Sciences. This work was supported by National Institutes of Health grants CA13106 to A.R.K. and GM44044 to S.H.M.

References

Ben-David,Y., Bani,M.R., Chabot,B., De Koven,A. and Bernstein,A. (1992) *Mol. Cell. Biol.*, **12**, 4449–4455.
Bennett,M., Piñol-Roma,S., Staknis,D., Dreyfuss,G. and Reed,R. (1992a) *Mol. Cell. Biol.*, **12**, 3165–3175.
Bennett,M., Michaud,S., Kingston,J. and Reed,R. (1992b) *Genes Dev.*, **6**, 1986–2000.
Beyer,A.L., Christensen,M.E., Walker,B.W. and LeStourgeon,W.M. (1977) *Cell*, **11**, 127–138.
Biamonti,G., Buvoli,M., Bassi,M.T., Morandi,C., Cobianchi,F. and Riva,S. (1989) *J. Mol. Biol.*, **207**, 491–503.
Biamonti,G., Ruggiu,M., Saccone,S., Della Valle,G. and Riva,S. (1994) *Nucleic Acids Res.*, **22**, 1996–2002.
Birney,E., Kumar,S. and Krainer,A.R. (1993) *Nucleic Acids Res.*, **21**, 5803–5816.
Burd,C.G. and Dreyfuss,G. (1994) *EMBO J.*, **13**, 1197–1204.
Burd,C.G., Swanson,M.S., Görlach,M. and Dreyfuss,G. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9788–9792.
Buvoli,M., Biamonti,G., Tsoulfas,P., Bassi,M.T., Ghetti,G., Riva,S. and Morandi,C. (1988) *Nucleic Acid Res.*, **16**, 3751–3770.
Buvoli,M., Cobianchi,F., Bastagno,M.G., Mangiarotti,A., Bassi,M.T., Biamonti,G. and Riva,S. (1990a) *EMBO J.*, **9**, 1229–1235.
Buvoli,M., Cobianchi,F., Biamonti,G. and Riva,S. (1990b) *Nucleic Acids Res.*, **18**, 6595–6600.
Buvoli,M., Cobianchi,F. and Riva,S. (1992) *Nucleic Acids Res.*, **20**, 5017–5025.
Cáceres,J.F. and Krainer,A.R. (1993) *EMBO J.*, **12**, 4715–4726.
Cáceres,J.F., Stamm,S., Helfman,D.M. and Krainer,A.R. (1994) *Science*, **265**, 1706–1709.
Carey,J., Cameron,V., de Haseth,P.L. and Uhlenbeck,O.C. (1983) *Biochemistry*, **22**, 2601–2610.
Casas-Finet,J.R., Karpel,R.L., Maki,A.H., Kumar,A. and Wilson,S.H. (1991) *J. Mol. Biol.*, **221**, 693–709.
Casas-Finet,J.R., Smith,J.D.,Jr, Kumar,A., Kim,J.G., Wilson,S.H. and Karpel,R.L. (1993) *J. Mol. Biol.*, **229**, 873–889.
Choi,Y.D., Grabowski,P.J., Sharp,P.A. and Dreyfuss,G. (1986) *Science*, **231**, 1534–1539.
Cobianchi,F., Karpel,R.L., Williams,K.R., Notario,V. and Wilson,S.H. (1988) *J. Biol. Chem.*, **263**, 1063–1071.
Dreyfuss,G., Matunis,M.J., Piñol-Roma,S. and Burd,C.G. (1993) *Annu. Rev. Biochem.*, **62**, 289–321.
Fu,X.-D. (1993) *Nature*, **365**, 82–85.
Fu,X.-D., Mayeda,A., Maniatis,T. and Krainer,A.R. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 11224–11228.
Garrett,D.S., Lodi,P.J., Shamoo,Y., Williams,K.R., Clore,G.M. and Gronenborn,A.M. (1994) *Biochemistry*, **33**, 2852–2858.
Ge,H. and Manley,J.L. (1990) *Cell*, **62**, 25–34.
Ge,H., Zuo,P. and Manley,J.L. (1991) *Cell*, **66**, 373–382.
Ghetti,A., Piñol-Roma,S., Michael,W.M., Morandi,C. and Dreyfuss,G. (1992) *Nucleic Acids Res.*, **20**, 3671–3678.
Gil,A., Sharp,P.A., Jamison,S.F. and Garcia-Blanco,M.A. (1991) *Genes Dev.*, **5**, 1224–1236.
Görlach,M., Wittekind,M., Bechman,R.A., Mueller,L. and Dreyfuss,G. (1992) *EMBO J.*, **11**, 3289–3295.
Hamilton,B.J., Nagy,E., Malter,J.S., Arrick,B.A. and Rigby,W.F.C. (1993) *J. Biol. Chem.*, **12**, 8881–8887.
Harper,J.E. and Manley,J.L. (1991) *Mol. Cell. Biol.*, **11**, 5945–5953.
Herrick,G. and Alberts,B. (1976) *J. Biol. Chem.*, **251**, 2133–2141.
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.
Horowitz,D.S. and Krainer,A.R. (1994) *Trends Genet.*, **10**, 100–106.
Huang,M., Rech,J.E., Northington,S.J., Flicker,P.F., Mayeda,A., Krainer,A.R. and LeStourgeon,W.M. (1994) *Mol. Cell. Biol.*, **14**, 518–533.
Ishikawa,F., Matunis,M.J., Dreyfuss,G. and Cech,T.R. (1993) *Mol. Cell. Biol.*, **13**, 4301–4310.
Karsch-Mizrachi,I. and Haynes,S.R. (1993) *Nucleic Acids Res.*, **21**, 2229–2235.
Kenan,D.J., Query,C.C. and Keene,J.D. (1991) *Trends Biochem. Sci.*, **3**, 431–437.
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) *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.
Kumar,A. and Wilson,S.H. (1990) *Biochemistry*, **29**, 10717–10722.
Kumar,A., Williams,K.R. and Szer,W. (1986) *J. Biol. Chem.*, **261**, 11266–11273.
Kumar,A., Sierakowska,H. and Szer,W. (1987) *J. Biol. Chem.*, **262**, 17126–17137.
Kumar,A., Casas-Finet,J.R., Luneau,C.J., Karpel,R.L., Merrill,B.M., Williams,K.R. and Wilson,S.H. (1990) *J. Biol. Chem.*, **265**, 17094–17100.
Lee,C.-G., Zamore,P.D, Green,M.R. and Hurwitz,J. (1993) *J. Biol. Chem.*, **268**, 13472–13478.
LeStourgeon,W.M., Barnett,S.F. and Northington,S.J. (1990) In Strauss,P.R. and Wilson,S.H. (eds), *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies*. The Telford Press, Caldwell, NJ, Vol. 2, pp. 477–502.
McKay,S.J. and Cooke,H. (1992) *Nucleic Acids Res.*, **20**, 6461–6464.
Matunis,E.L., Kelley,R. and Dreyfuss,G. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 2781–2784.
Mayeda,A. and Krainer,A.R. (1992) *Cell*, **68**, 365–375.
Mayeda,A. and Ohshima,Y. (1988) *Mol. Cell. Biol.*, **8**, 4484–4491.

Mayeda,A. and Ohshima,Y. (1990) *Nucleic Acids Res.*, **18**, 4671–4676.
Mayeda,A., Tatei,K., Kitayama,H., Takemura,K. and Ohshima,Y. (1986) *Nucleic Acid 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.
Mayeda,A., Helfman,D.M. and Krainer,A.R. (1993) *Mol. Cell. Biol.*, **13**, 2993–3001.
Mayrand,S.H. and Pederson,T. (1990) *Nucleic Acids Res.*, **18**, 3307–3318.
Merrill,B.M., Stone,K.L., Cobianchi,F., Wilson,S.H. and Williams,K.R. (1988) *J. Biol. Chem.*, **263**, 3307–3313.
Moore,M.J., Query,C.C. and Sharp,P.A. (1993) In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.
Mulligan,G.J., Guo,W., Wormsley,S. and Helfman,D.M. (1992) *J. Biol. Chem.*, **267**, 25480–25487.
Munroe,S.H. (1988) *EMBO J.*, **7**, 2523–2532.
Munroe,S.H. and Dong,X. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 895–899.
Nadler,S.G., Merrill,B.M., Roberts,W.J., Keating,K.M., Lisbin,M.J., Barnett,S.F., Wilson,S.H. and Williams,K.R. (1991) *Biochemistry*, **30**, 2968–2976.
Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) *Nature*, **348**, 515–520.
Patton,J.G., Mayer,S.A., Tempst,P. and Nadal-Ginard,B. (1991) *Genes Dev.*, **5**, 1237–1251.
Piñol-Roma,S., Choi,Y.D., Matunis,M.J. and Dreyfuss,G. (1988) *Genes Dev.*, **2**, 215–227.
Pontius,B.W. (1993) *Trends Biochem. Sci.*, **18**, 181–186.
Pontius,B.W. and Berg,P. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8403–8407.
Pontius,B.W. and Berg,P. (1992) *J. Biol. Chem.*, **267**, 13815–13818.
Portman,D.S. and Dreyfuss,G. (1994) *EMBO J.*, **13**, 213–221.
Reed,R. and Maniatis,T. (1986) *Cell*, **46**, 681–690.
Sierakowska,H., Szer,W., Furdon,P.J. and Kole,R. (1986) *Nucleic Acids Res.*, **14**, 5241–5254.
Smith,K.C. (1976) In Wang,S.Y. (ed.), *Photochemistry and Photobiology of Nucleic Acids*. Academic Press, New York, NY, Vol. 2, pp. 187–218.
Stolow,D.T. (1992) Ph.D. thesis. Baylor College of Medicine, Houston, TX.
Stolow,D.T. and Berget,S.M. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 320–324.
Sun,Q., Mayeda,A., Hampson,R.K., Krainer,A.R. and Rottman,F.M. (1993) *Genes Dev.*, **7**, 2598–2608.
Swanson,M.S. and Dreyfuss,G. (1988) *EMBO J.*, **7**, 3519–3529.
Tian,M. and Maniatis,T. (1993) *Cell*, **74**, 105–114.
Williams,K.R., Stone,K.L., LoPresti,M.B., Merrill,B.M. and Planck,S.R. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 5666–5670.
Wyatt,J.R., Sontheimer,E.J. and Steitz,J.A. (1992) *Genes Dev.*, **6**, 2542–2553.
Yang,X., Bani,M.-R., Lu,S.-J., Rowan,S., Ben-David,Y. and Chabot,B. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 6924–6928.
Zahler,A.M., Lane,W.S., Stolk,J.A. and Roth,M.B. (1992) *Genes Dev.*, **6**, 837–848.
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.
Zuo,P. and Manley,J.L. (1993) *EMBO J.*, **12**, 4727–4737.

Received on June 23, 1994; revised on August 29, 1994