The EMBO Journal vol.13 no.5 pp.1197–1204, 1994

RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing

Christopher G.Burd and Gideon Dreyfuss[1]

Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6148, USA
[1]Corresponding author

Communicated by J.A.Steitz

Pre-mRNA is processed as a large complex of pre-mRNA, snRNPs and pre-mRNA binding proteins (hnRNP proteins). The significance of hnRNP proteins in mRNA biogenesis is likely to be reflected in their RNA binding properties. We have determined the RNA binding specificity of hnRNP A1 and of each of its two RNA binding domains (RBDs), by selection/amplification from pools of random sequence RNA. Unique RNA molecules were selected by hnRNP A1 and each individual RBD, suggesting that the RNA binding specificity of hnRNP A1 is the result of both RBDs acting as a single RNA binding composite. Interestingly, the consensus high-affinity hnRNP A1 binding site, UAGGGA/U, resembles the consensus sequences of vertebrate 5′ and 3′ splice sites. The highest affinity 'winner' sequence for hnRNP A1 contained a duplication of this sequence separated by two nucleotides, and was bound by hnRNP A1 with an apparent dissociation constant of 1×10^{-9} M. hnRNP A1 also bound other RNA sequences, including pre-mRNA splice sites and an intron-derived sequence, but with reduced affinities, demonstrating that hnRNP A1 binds different RNA sequences with a >100-fold range of affinities. These experiments demonstrate that hnRNP A1 is a sequence-specific RNA binding protein. UV light-induced protein–RNA crosslinking in nuclear extracts demonstrated that an oligoribonucleotide containing the A1 winner sequence can be used as a specific affinity reagent for hnRNP A1 and an unidentified 50 kDa protein. We also show that this oligoribonucleotide, as well as two others containing 5′ and 3′ pre-mRNA splice sites, are potent inhibitors of *in vitro* pre-mRNA splicing.
Key words: hnRNP A1/pre-mRNA splicing/RNA binding specificity

Introduction

As pre-mRNA emerges from the transcription complex, it is bound by a large assortment of proteins (Dreyfuss *et al.*, 1993). All of the proteins that bind pre-mRNA and are not stable components of other ribonucleoproteins (e.g. snRNP complexes), are collectively termed hnRNP proteins (Dreyfuss *et al.*, 1993). There are over 20 abundant hnRNP proteins in human (HeLa) cells, termed A–U, and most, if not all, of them bind directly to RNA (Dreyfuss *et al.*, 1993). The importance of these proteins is underscored by the fact that along the entire pathway of gene expression, the functional form of the RNA is a protein–RNA complex (Dreyfuss, 1986). With few exceptions, the RNA binding properties of individual hnRNP proteins are poorly understood. As the function of hnRNP proteins in mRNA biogenesis is certain to be reflected in their RNA binding properties, it is essential to know their RNA binding specificities and affinities.

The general RNA binding properties of the human hnRNP A1 protein are the best characterized so far. At its N-terminus hnRNP A1 contains two RNP consensus sequence (RNP-CS) RNA binding domains (RBDs), followed by a C-terminal region that is particularly rich in glycine (Merrill *et al.*, 1988; Buvoli *et al.*, 1990). The carboxy region also appears to contribute to RNA binding (Kumar *et al.*, 1990; Nadler *et al.*, 1991; Casas-Finet *et al.*, 1993) and it contains an RGG box (Kiledjian and Dreyfuss, 1992). A large body of evidence has shown that hnRNP A1, as well as several other hnRNP proteins, discriminates between RNA sequences (Swanson and Dreyfuss, 1988a,b; Buvoli *et al.*, 1990; Bennett *et al.*, 1992; Matunis *et al.*, 1993; M.Görlach, C.Burd and G.Dreyfuss, manuscript in preparation). However, other studies employing ribonucleotide homopolymers and synthetic base analogs have been used to argue against sequence-specific RNA binding by hnRNP A1 (Cobianchi *et al.*, 1988; LeStourgeon *et al.*, 1990; Nadler *et al.*, 1991; Casas-Finet *et al.*, 1993).

Several functional properties of hnRNP A1 have been described. Addition of hnRNP A1 to complementary DNA or RNA molecules promotes the formation of base-paired double strands (Kumar and Wilson, 1990; Pontius and Berg, 1990; Munroe and Dong, 1992), and several other hnRNP proteins share this interesting property (Portman and Dreyfuss, 1994). It has been suggested that, by its strand-annealing activity, hnRNP A1 can influence pre-mRNA splicing by modulating the association of snRNPs and other RNA processing factors with the pre-mRNA (Buvoli *et al.*, 1992; Eperon *et al.*, 1993; Portman and Dreyfuss, 1994). Potential roles for hnRNP A1 in the regulation of pre-mRNA splicing have also been described (Buvoli *et al.*, 1992; Mayeda and Krainer, 1992). For example, *in vitro* splicing assays have shown that hnRNP A1 and another RNA binding protein, ASF/SF2, influence 5′ splice site choice in pre-mRNAs containing multiple 5′ splice sites (Mayeda and Krainer, 1992; Mayeda *et al.*, 1993). The specific pre-mRNA binding site, if any, for each of these proteins is presently unknown. In addition to nuclear functions, possible roles for hnRNP A1 in the cytoplasm and in nuclear–cytoplasmic mRNA transport were suggested by the observation that it shuttles between the nucleus and the cytoplasm (Piñol-Roma and Dreyfuss, 1992).

To understand the functional significance of RNA binding by hnRNP A1, it is necessary to know its RNA binding specificity and its affinity for different RNA sequences. We have directly addressed these issues by selection/amplification from pools of random sequence RNA (Tuerk and

Gold, 1990) to determine sequences that hnRNP A1 binds with highest affinity. The affinity of hnRNP A1 for several selected sequences, as well as for several other oligoribonucleotides containing natural RNA sequences, was determined. Finally, we have assessed the interaction of several of these oligoribonucleotides with endogenous proteins of HeLa cell nuclear extract and demonstrated their usefulness as reagents for probing *in vitro* pre-mRNA splicing.

Results

Identification of an hnRNP A1 high-affinity binding site

hnRNP A1 was immobilized on protein A Sepharose with a mAb (9H10) that reacts with an epitope at the extreme C-terminus of the protein (S.Piñol-Roma, H.Siomi and G.Dreyfuss, unpublished observations). Then the protein was presented with a large molar excess of RNA containing a randomized region of 20 bases flanked by two regions of known sequence. After binding, selected RNA molecules were purified and amplified by reverse transcription-polymerase chain reaction (RT-PCR). The resultant DNA was transcribed by T7 RNA polymerase and the process was repeated for a total of eight cycles (Tuerk and Gold, 1990). The products were cloned after the final round of amplification and the sequence of the selected region from 30 independent clones was determined (see Figure 1). All of the selected RNAs contain the sequence UAGGGA/U or, in a small proportion of the clones, a highly related motif. Approximately one-third of the sequences are identical over the entire selected region and these contain two copies of this motif (see Figure 1). As this 20mer sequence was predominant in the collection of selected sequences and exhibited the highest affinity for hnRNP A1 (see below), it was termed 'A1 winner'. Therefore, the greatest selective constraint appears to be on the six base sequence, UAGGGA/U, and we consider this sequence a putative high-affinity binding site for hnRNP A1. In addition to this experiment, we have repeated the selection/amplification under the same conditions and have also repeated it with a randomized region of only 10 bases. The results of all our selection/amplification experiments are consistent with that presented here. Selection/amplification experiments with only protein A Sepharose yielded no consensus selected sequence after eight cycles (unpublished data).

To gain an insight into the determinants of hnRNP A1 RNA binding specificity, we carried out parallel selection/amplification experiments with the individual RBDs of hnRNP A1. Constructs were made by PCR cloning that expressed each RBD fused to the carboxy glycine-rich region which contains the 9H10 epitope (see Figure 2). Each construct (RBD I:GLY and RBD II:GLY) was expressed in *Escherichia coli*, immunopurified from bacterial extract and used for a selection/amplification experiment. After eight cycles of selection/amplification, a unique consensus winner RNA sequence for each RBD construct was apparent and they are shown in Figure 2. It is apparent that while RBD I:GLY and RBD II:GLY could specifically bind unique RNA sequences, the overall binding activity of hnRNP A1 is different, and is not simply the sum of the specificities of the individual RBDs.

To confirm that the selection/amplification enriched for

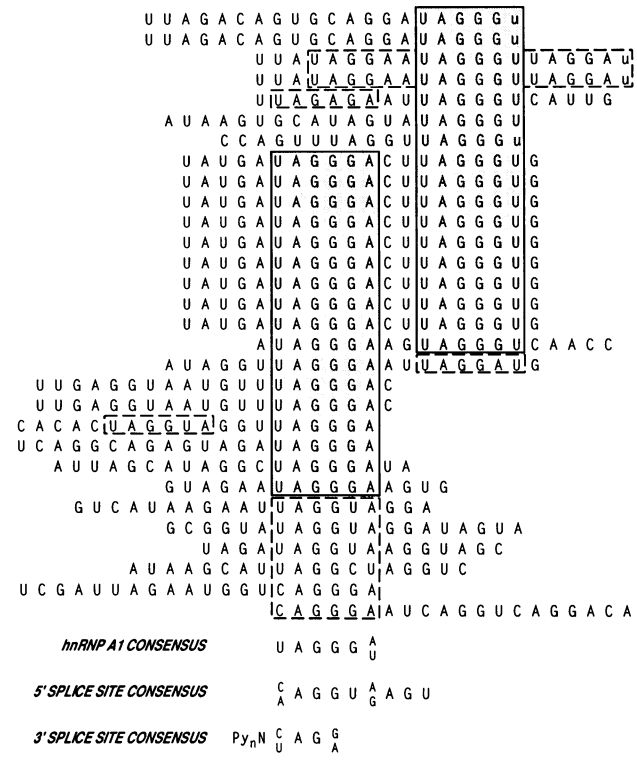

Fig. 1. The sequences of hnRNP A1-selected RNA molecules. A selection/amplification experiment was carried out as described in Materials and methods and the sequence of the selected region of 30 RNA molecules is shown. The collection of sequences was aligned according to the boxed regions that contain a sequence common to all molecules. Sequences contained within shaded boxes exactly match the consensus and sequences in the unshaded dashed boxes have one mismatch to the consensus. Lower case letters represent positions that form part of the putative binding site but were from the flanking constant region of each RNA (see Materials and methods). Below the list are the sequences of the hnRNP A1 consensus high-affinity binding site and the consensus 5′ and 3′ vertebrate splice sites.

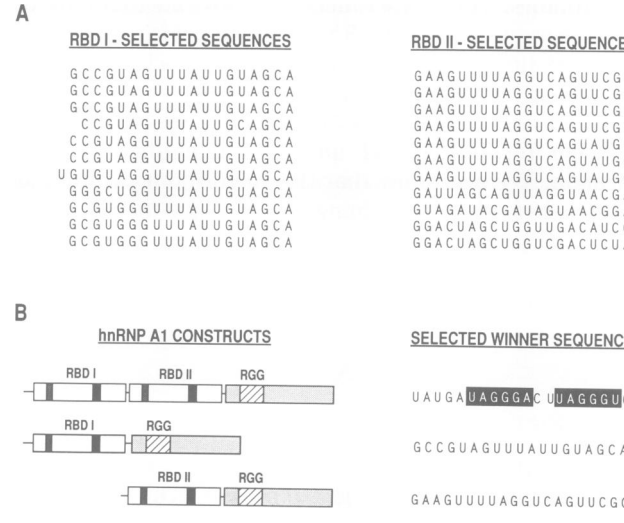

Fig. 2. hnRNP A1 constructs and their winner sequences. (**A**) The sequence of 10 selected RNAs from the eighth round of a selection/amplification experiment employing the individual RBDs of hnRNP A1 are shown. (**B**) On the left are schematic drawings of hnRNP A1 (top), RBD I:GLY (middle) and RBD II:GLY (bottom), as used in the selection/amplification experiments. The eighth round consensus winner sequences for each protein are listed to the right of each construct.

RNA sequences that hnRNP A1 binds with highest affinity, we determined the affinity (apparent equilibrium dissociation constant, K_d) of hnRNP A1 for several different RNA molecules using a nitrocellulose filter binding assay (Carey and Uhlenbeck, 1983). We first sought to determine the affinity of hnRNP A1 for the pools of RNA used in each round of selection, however, for technical reasons this was not possible. These problems could be circumvented by using oligoribonucleotides corresponding only to the selected region of individual clones (see Figure 3 and unpublished data). As a gauge for binding to an unselected RNA sequence, we also measured the affinity for an oligoribonucleotide whose sequence was derived from the first intron of the human β globin gene (β intron). It does not contain any sequence that resembles the high-affinity hnRNP A1 binding site. Figure 3 shows the normalized binding curves and Figure 4 summarizes these results. The oligoribonucleotide bound by hnRNP A1 with highest affinity ($K_d = 1 \times 10^{-9}$ M) was the most prevalent RNA sequence after eight rounds of selection/amplification (A1 winner). hnRNP A1 bound the unrelated β intron RNA with a 300-fold reduced affinity ($K_d = 3 \times 10^{-7}$ M) compared with A1 winner. Thus, the selection/amplification experiment did indeed identify the preferred binding sites for hnRNP A1. A selected sequence (A1R8.2) from a second experiment (unpublished data), containing only one UAGGGA motif, was bound with slightly lower affinity ($K_d = 3 \times 10^{-9}$ M) than A1 winner (see Figure 4).

We compared several hnRNP A1-selected sequences (or portions thereof) with the nucleic acid sequences of the GenBank and EMBL databases. There was no apparent pattern to the location of these sequences in the context of pre-mRNAs. However, by visual inspection we noticed that the high-affinity hnRNP A1 binding site conforms to the vertebrate consensus 5′ and 3′ splice sites (Mount, 1982; Ohshima and Gotoh, 1987). Figure 1 shows a comparison of the sequences of the hnRNP A1 high-affinity binding site and the consensus 5′ and 3′ vertebrate splice sites. We directly tested whether hnRNP A1 has a preference for splice sites by measuring the affinity of hnRNP A1 for oligoribonucleotides (20mer) containing a 5′ or 3′ splice site from the human β globin gene (shown in Figure 4). Both of these RNAs were bound with an intermediate affinity ($K_d = 7 \times 10^{-8}$ M) compared with A1 winner and β intron RNAs. Because each of these ligands were bound with equal affinity, hnRNP A1 could bind preferentially to both 5′ and 3′ splice sites of this pre-mRNA.

hnRNP A1 binds A1 winner in nuclear extract

All of the experiments described thus far employed purified RNA and purified recombinant hnRNP A1. To ask whether hnRNP A1 can interact specifically with A1 winner in a complex mixture of proteins and RNA, UV light-induced crosslinking experiments were carried out. In these experiments, internally ^{32}P-labeled oligoribonucleotides (22mer), containing hnRNP A1-selected sequences or natural pre-mRNA sequences, were incubated in HeLa cell pre-mRNA splicing extract and irradiated with UV light for 8 min. After RNase digestion, the products were separated by SDS−PAGE and proteins containing residual crosslinked radioactive nucleotides were visualized by autoradiography. Figure 5 shows the results of these experiments. A protein

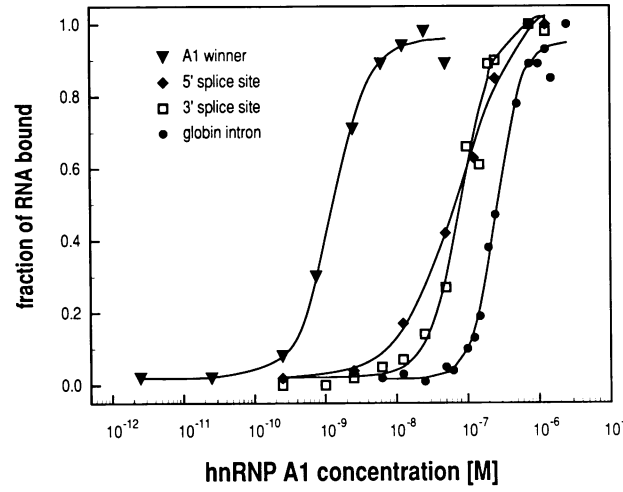

Fig. 3. hnRNP A1 RNA binding curves. The affinity of hnRNP A1 for several different oligoribonucleotides was measured by a filter binding assay. Each point represents the average of either three or six independent binding reactions and the data for each curve are normalized to the saturation point for each oligoribonucleotide, corrected for background RNA binding and for the fraction of hnRNP A1 active in RNA binding. The K_d value is equal to the protein concentration at which 50% of RNA is bound. The legend to the data points is shown in the upper left corner of the plot.

Summary of hnRNP A1 RNA-binding Experiments

	RNA sequence	K_d
A1 winner	U A U G A **U A G G G A** C U **U A G G G U** G	1 nM
A1R8.2	U U U U U G A **U A G G G A** A A U U A	3 nM
β globin 5' splice site	C C C U G G G C A G ▼ G U U G G U A U C A	70 nM
β globin 3' splice site	C C A C C C U U A G ▼ G C U G C U G G U G	70 nM
β globin intron	G A U C A C U U G U G U C A A C A C A G	300 nM

Fig. 4. Summary of hnRNP A1 binding activity. The name of each oligoribonucleotide tested is given on the left, its sequence in the center and the affinity with which hnRNP A1 binds it on the right. The arrows indicate the position of the 5′ and 3′ splice sites.

that comigrates with hnRNP A1 and an unidentified 50 kDa protein crosslink preferentially to A1 winner and, to a lesser extent, to β 3′SS. Surprisingly, the proteins that crosslink to A1R8.7, containing only one hnRNP A1 high-affinity binding site, and to β intron are nearly identical. By this assay, we did not identify any proteins that crosslink specifically to β 5′SS. Finally, several proteins crosslink to all RNAs tested, suggesting that these interactions are relatively non-specific.

To determine if hnRNP A1 crosslinked to A1 winner, we performed a crosslinking reaction followed by immunopurification with a mAb (9H10) specific for hnRNP A1 (see Figure 5). As a control, we performed the same experiment with ^{32}P-labeled β intron RNA. hnRNP A1 crosslinked very well to A1 winner, in contrast to the very small amount of hnRNP A1 crosslinked to β intron RNA. UV crosslinking experiments in the presence of increasing amounts of different cold competitor oligoribonucleotides confirmed the specificity of hnRNP A1, as well as the specificity of the 50 kDa protein, for A1 winner (unpublished data).

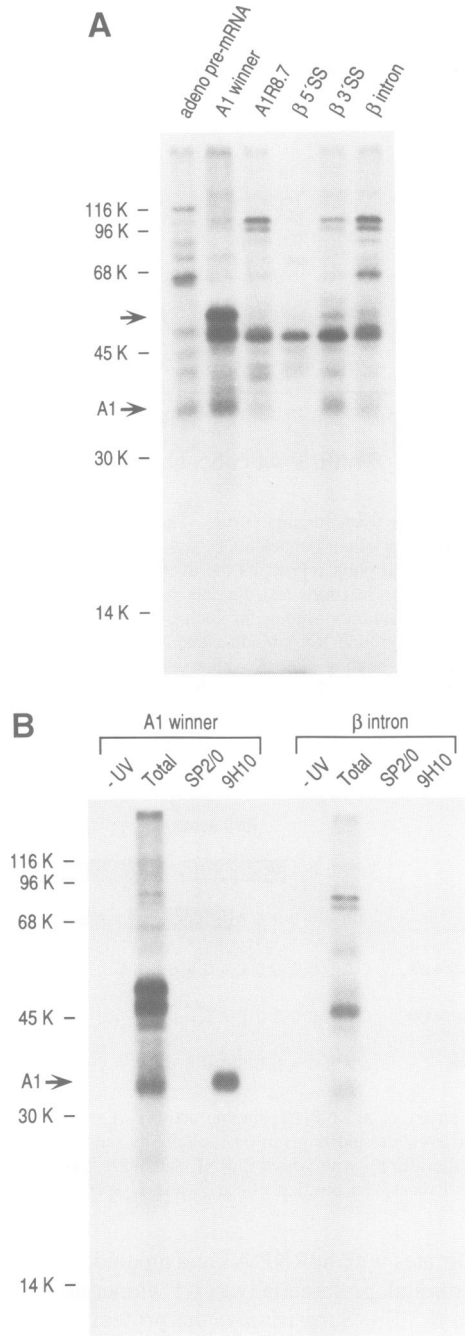

Fig. 5. UV light-induced crosslinking to different RNA molecules. (**A**) ~1 × 10^5 c.p.m. of internally ^{32}P-labeled adenovirus MLP pre-mRNA or oligoribonucleotides (22mer) were incubated in HeLa cell nuclear extract and crosslinked with UV light. Crosslinked products were resolved by SDS–PAGE and autoradiography. The identity of each RNA molecule is indicated at the top of each lane and the migration of molecular weight standards (in kDa) is indicated on the left. Arrows point to hnRNP A1 or to an ~50 kDa protein that crosslinks specifically to A1 winner. (**B**) After crosslinking to A1 winner, proteins were immunopurified with the mAb 9H10 (hnRNP A1) or nonimmune serum (SP2/0). For comparison, the results of an identical analysis of crosslinked products to β intron oligoribonucleotide is shown. Approximately 20% of each crosslinking reaction (before immunoprecipitation) is run in the lane marked 'Total'.

Pre-mRNA splicing inhibition by oligoribonucleotides

The experiments described above demonstrate that hnRNP A1, as well as at least one other protein, specifically interact

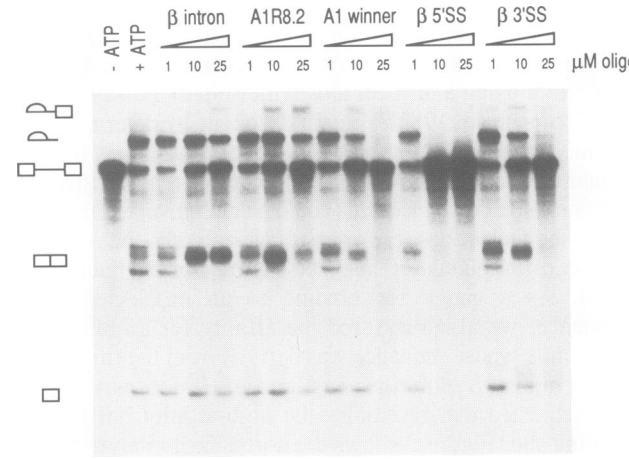

Fig. 6. Inhibition of pre-mRNA splicing by oligoribonucleotides. Internally ^{32}P-labeled adenovirus MLP pre-mRNA was added to splicing reactions that had been preincubated on ice for 5 min with the indicated amount of each oligoribonucleotide. After 90 min at 30°C, reaction products were resolved by electrophoresis. Reaction intermediates and products are indicated schematically to the left. The heterogeneity in the spliced products is probably due to cleavage of a second 5′ splice site downstream of the 3′ splice site (Grabowski *et al.*, 1984).

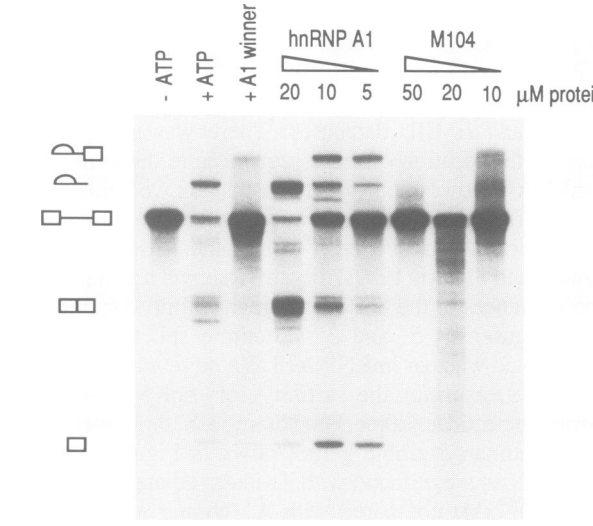

Fig. 7. Restoration of splicing activity by hnRNP A1. A1 winner oligoribonucleotide was added (final concentration, 25 μM) to a splicing reaction, incubated for 5 min on ice, followed by the addition of various amounts of hnRNP A1 or the RNA binding domain of the hnRNP C proteins (M104). After a further 5 min incubation on ice, radiolabeled pre-mRNA was added and the reaction was allowed to proceed at 30°C for 90 min. Products and intermediates were resolved by electrophoresis and are indicated to the left of the autoradiogram.

with A1 winner in nuclear extract. In an effort to explore the functional significance of these interactions, we tested the effect of adding different oligoribonucleotides to *in vitro* splicing reactions of the adenovirus-derived MLP model pre-mRNA (Grabowski *et al.*, 1984). Different amounts of oligoribonucleotide (final concentrations of 5, 10 or 25 μM) were added to splicing extracts and allowed to incubate for 5 min on ice, followed by the addition (~10 pmol) of radiolabeled pre-mRNA (see Figure 6). Oligoribonucleotides containing authentic 5′ or 3′ splice sites and A1 winner inhibited splicing almost completely at a concentration of

25 μM. Inhibition by the β 5′SS oligoribonucleotide was especially striking, partially inhibiting splicing at 1 μM and completely at 10 μM. This observation is consistent with similar experiments using the vertebrate consensus 5′ splice site oligoribonucleotide (Hall and Konarska, 1992). The effect of addition of A1 winner was slightly stronger than inhibition due to the 3′ splice site oligoribonucleotide. The A1R8.2 oligoribonucleotide, derived from a different hnRNP A1-selected sequence (containing only one UAGGGA motif), inhibited splicing slightly at 25 μM and β intron had no effect at any concentration tested.

We also tested the effect of adding increasing amounts of purified hnRNP A1 to splicing reactions that were inhibited by A1 winner. Splicing inhibition was relieved in a manner dependent on the amount of added hnRNP A1 (see Figure 7). As a control for the specificity of this experiment, we added increasing amounts of the RNA binding domain of the hnRNP C proteins. This fragment is the minimal specific RBD of the hnRNP C proteins and has a different RNA binding specificity than hnRNP A1 (M.Görlach, C.Burd and G.Dreyfuss, manuscript in preparation). Addition of this polypeptide at all tested concentrations did not rescue A1 winner-mediated inhibition. Splicing reactions inhibited by β 5′SS or β 3′SS were not rescued by the addition of hnRNP A1 (data not shown).

Discussion

hnRNP A1 is a sequence-specific RNA binding protein

Our experiments demonstrate that hnRNP A1 is a sequence-specific RNA binding protein. Based on what we know about other abundant hnRNP proteins, this is likely to be a general feature of these proteins. We used selection/amplification from pools of random sequence RNA to determine the RNA binding specificity of hnRNP A1. This approach has been used to determine the preferred binding sites for a wide variety of RNA binding proteins, including members of the RNP-CS family (Tuerk and Gold, 1990; Bartel *et al.*, 1991; Tsai *et al.*, 1991; Schneider *et al.*, 1992, 1993; M.Görlach, C.Burd and G.Dreyfuss, manuscript in preparation). After eight rounds of selection/amplification, all of the selected RNA molecules were remarkably similar in that they all contained the six base sequence, UAGGGA/U, or a highly related motif (see Figure 1). These data suggest that this sequence is a high-affinity binding site for hnRNP A1. We directly confirmed this hypothesis by binding experiments in which hnRNP A1 exhibited high affinity for molecules containing this sequence (see Figures 2–4). These experiments definitively prove that hnRNP A1 binds RNA with sequence specificity and a wide range of affinities.

There are several notable features of the hnRNP A1 high-affinity binding site. As shown in Figure 1, it is similar to both the vertebrate 5′ and 3′ splice site consensus sequences. The hnRNP A1 high-affinity binding site matches each consensus equally well and this is reflected in the equivalent and intermediate affinity that hnRNP A1 displays for oligoribonucleotides containing natural 5′ or 3′ splice sites. In other selection/amplification experiments (unpublished data), many selected sequences nearly matched the entire 3′ splice site consensus (e.g. A1R8.2). These findings are consistent with previous experiments that mapped hnRNP A1 binding sites to the 3′ end of introns (Swanson and Dreyfuss, 1988a,b; Buvoli *et al.*, 1990, 1992; Ishikawa *et al.*, 1993). Computer analysis of potential RNA secondary structures of hnRNP A1-selected RNA molecules did not predict a common structural conformation for the selected sequences.

We were surprised to find only one hnRNP A1 high-affinity sequence in the results of the selection/amplification experiments because hnRNP A1 contains two RBDs that appear to have been individually conserved through evolution (Dreyfuss *et al.*, 1988; Matunis *et al.*, 1992). As many of the selected regions contain multiple potential high-affinity binding sites (e.g. A1 winner), it is possible that the specificity of each domain is almost identical. However, parallel experiments with each hnRNP A1 RBD (see Figure 2) revealed that each domain has a unique specificity, so it seems unlikely that this is the case for intact hnRNP A1. Another possibility is that specific binding to UAGGGA/U requires both RBDs acting as a single RNA binding composite. Several observations support this interpretation. First, the selection/amplification experiments with the individual RBDs indicate that both RBDs are necessary for the specificity of the entire protein. Second, in the poly(A) binding protein which contains four RBDs, the strongest poly(A) binding activity of the protein is in the first two RBDs (Nietfeld *et al.*, 1990; Burd *et al.*, 1991). The junction between these two RBDs is the most highly conserved region of all mRNP poly(A) binding proteins, suggesting that this region is important for poly(A) binding (Burd *et al.*, 1991). Third, studies of the hnRNP C RBD bound to a high-affinity site indicate that regions immediately C-terminal to the canonical RBD are important for specific RNA binding (Görlach *et al.*, 1992; M.Görlach, C.Burd and G.Dreyfuss, manuscript in preparation). By analogy with the PABP (Burd *et al.*, 1991), we suggest that in hnRNP A1 important determinants of RNA binding specificity reside within the region spanning the fourth β-sheet of RBD I and the first β-sheet of RBD II, and that this region probably makes sequence-specific contacts with the RNA. Other regions of each RBD must also contribute to overall specificity because this region was included in both of the individual RBD constructs used in the selection/amplification experiments (see Figure 2 and Materials and methods). Furthermore, a selection/amplification experiment with the individual RBD constructs mixed together did not generate hnRNP A1 specificity, indicating that the two RBDs must be linked in *cis* (unpublished data).

An important conclusion from these experiments is that, as a previous analysis of the RNA binding specificity of the PABP suggested (Burd *et al.*, 1991), the overall RNA binding activity of RNP-CS proteins, such as PABP and hnRNP A1, is not strictly the sum of the individual RBD specificities. We have not measured the affinity of the individual RBD winner sequences for hnRNP A1 (i.e. the intact protein), but it can be anticipated that they will be bound with an intermediate affinity because of their resemblance to the two splice site RNAs tested and to the hnRNP A1-selected RNA molecules. In fact, the RBD II:GLY winner sequence resembles the 5′ splice site consensus more closely than it matches the hnRNP A1 high-affinity binding site (see Figures 1 and 2). It is possible that in the absence of highest-affinity binding sites, each RBD could bind a unique RNA sequence. If this is the case, it would indicate that hnRNP A1 can bind RNA in two ways. More experiments are required to test this possibility. It is

important to note that in other multiple RBD proteins, such as the U1 snRNP A protein and PABP, RBDs can act independently (Lutz-Freyermuth *et al.*, 1990; Nietfeld *et al.*, 1990; Burd *et al.*, 1991).

hnRNP A1 binds RNA with a range of affinities

Oligoribonucleotides containing one or two high-affinity binding sites were bound by hnRNP A1 with highest affinity (K_d = 3 and 1 nM, respectively). Oligoribonucleotides containing natural splice site sequences from human β globin pre-mRNA were bound by hnRNP A1 with an intermediate affinity (K_d = 70 nM), and an oligoribonucleotide of unrelated sequence (β intron) was bound with an affinity of 300 nM. This non-specific affinity of hnRNP A1 (i.e. 300 nM) for RNA is consistent with previous measurements for different ribonucleotide homopolymers (Nadler *et al.*, 1991). We conclude that hnRNP A1 binds RNA with a range of affinities and the difference in binding to high-affinity versus low-affinity binding sites spans at least two orders of magnitude. Furthermore, these data predict that binding sites that conform to the UAGGGA/U motif will be bound most tightly and there will be a commensurate loss in affinity for binding sites as they diverge from this sequence. The results of our selection/amplification experiments suggest that the last two positions of the high-affinity binding site are the most flexible and other binding studies are consistent with this observation (Swanson and Dreyfuss, 1988a,b; Ishikawa *et al.*, 1993). A possible function of non-specific RNA binding is that it facilitates the search for high-affinity binding sites by limiting the dimensions of diffusion as we and others have discussed (Swanson and Dreyfuss, 1988a,b; Dreyfuss *et al.*, 1990 and references therein).

We have shown that purified hnRNP A1 binds specifically to RNA molecules that were discovered by selection/amplification. If such ligands are to be useful reagents, it is of interest to determine if hnRNP A1 can bind them in a complex mixture of protein and RNA. By analogy with antibodies, it is important to determine if the high-affinity RNA reagent that was generated exhibits specificity towards hnRNP A1. The experiment shown in Figure 5 demonstrates that hnRNP A1 binds to A1 winner in nuclear extract much better than to an unrelated oligoribonucleotide (β intron). Several other proteins crosslink to A1 winner and all of them, with the notable exception of an unidentified ~50 kDa protein, crosslink to nearly all oligoribonucleotides tested, suggesting that these interactions are relatively non-specific. Members of the hnRNP A/B and D families were recently identified as the major nuclear proteins that bind oligonucleotides containing the vertebrate 3′ splice site, UUAGG (Ishikawa *et al.*, 1993) or, interestingly, the vertebrate telomere sequence, TTAGGG (McKay and Cooke, 1992; Ishikawa *et al.*, 1993). These proteins are among the candidates for proteins that bind the hnRNP A1 high-affinity site.

The characterization of the RNA binding properties of hnRNP A1 allow some predictions to be made concerning where on pre-mRNAs hnRNP A1 will be bound *in vivo*. It is clear that hnRNP A1 is an abundant component of the nucleus, with an intranuclear concentration in the order of 10^{-6} M (D.Portman and G.Dreyfuss, unpublished results). Therefore most, if not all, high-affinity binding sites will be stably bound and a substantial proportion of hnRNP A1 will be bound to intermediate- (e.g. splice sites) and lower-affinity sites. The affinity of other splice site binding factors (e.g. U1 snRNP) are not known but, considering the abundance of hnRNP A1 in the nucleus and its affinity for splice sites (70 nM), hnRNP A1 could influence the binding of other pre-mRNA binding factors to these sites on the pre-mRNA. For comparison, the range of affinities (K_d) with which the splicing factor U2 auxiliary factor (U2AF) binds to its pre-mRNA binding site is from 1×10^{-8} to 2×10^{-6} M (Zamore *et al.*, 1992). Interplay between hnRNP A1 and components of the splicing machinery has been observed previously (Mayrand and Pederson, 1987; Buvoli *et al.*, 1992; Mayeda and Krainer, 1992; Eperon *et al.*, 1993). Finally, as hnRNP A1 has equal and considerable affinity for both β globin 5′ and 3′ splice sites, it could bind both of these sequences of the pre-mRNA, and possibly help bring them together via homotypic protein−protein interactions.

hnRNP A1 shuttles between the nucleus and the cytoplasm and is bound to polyadenylated RNA in both compartments (Piñol-Roma and Dreyfuss, 1992). Since mRNA will not contain pre-mRNA processing signals (e.g. splice sites), it is unclear where on the cytoplasmic mRNA hnRNP A1 will be bound. An interesting feature of all of the hnRNP A1-selected sequences is that they contain at least one termination codon and/or initiation codon. This raises the possibility that hnRNP A1 could be bound to these sites during transport and also may be able to regulate translation. Such an activity for an hnRNP protein was recently suggested for the hnRNP I/polypyrimidine tract binding protein (Hellen *et al.*, 1993).

Pre-mRNA splicing inhibition

The observation that the high-affinity binding site for hnRNP A1 conforms to the vertebrate 5′ and 3′ splice site consensus sequences suggests a role for hnRNP A1 in pre-mRNA splicing. Therefore, we tested the effect of adding oligoribonucleotides to *in vitro* splicing reactions. Similar experiments by Hall and Konarska (1992) demonstrated pre-mRNA splicing inhibition by the addition of a vertebrate 5′ splice site consensus oligoribonucleotide (11mer). We found that A1 winner and the two oligoribonucleotides containing natural splice sites (β globin 5′ and 3′ splice sites) inhibited splicing.

At present, we do not know the mechanism of pre-mRNA splicing inhibition by the three different oligoribonucleotides, but it is likely that there are at least two. First, Hall and Konarska (1992) found that the addition of the consensus 5′ splice site oligoribonucleotide inhibited splicing at concentrations only slightly lower than β 5′SS, so inhibition by β 5′SS is probably due to the same mechanism (formation of pseudospliceosomes). Second, considering that significantly higher concentrations of A1 winner and β 3′SS were necessary to inhibit splicing (compared with β 5′SS), and that a unique set of proteins bind specifically to both A1 winner and β 3′SS (see Figure 5), inhibition by these two oligoribonucleotides is probably by a different mechanism than by β 5′SS. Furthermore, inhibition of splicing correlates with the crosslinking of hnRNP A1 and an unidentified 50 kDa protein to these two oligoribonucleotides. The relatively high concentrations of A1 winner and β 3′SS necessary for inhibition is consistent with a mechanism in which they act as a sink for an essential splicing factor. At this point we do not know if the same factor is titrated out of the splicing reaction by both oligoribonucleotides. Only pre-mRNA splicing reactions

inhibited by A1 winner are rescued by addition of hnRNP A1 (Figure 7 and unpublished data), but the interpretation of this experiment is complicated by our observation that addition of enough hnRNP A1 to bind all β 3′SS (at inhibitory concentrations) does itself inhibit pre-mRNA splicing (unpublished data). Interestingly, there is something unique about the winner sequence, which contains two hnRNP A1 high-affinity binding sites, because an oligoribonucleotide containing only one hnRNP A1 high-affinity binding site (A1R8.2) inhibited splicing only slightly at 25 μM. The results of these experiments are consistent with an essential role for hnRNP A1 in pre-mRNA splicing; however, this is unlikely as a cell line has been produced that lacks a functional hnRNP A1 gene (Ben-David *et al.*, 1992). We consider it likely that an essential splicing factor, with overlapping specificity with hnRNP A1, is titrated out of the splicing reaction by A1 winner (possibly by β 3′SS also) and the results of the crosslinking assay suggest that the 50 kDa protein that specifically crosslinks to A1 winner is a candidate.

Materials and methods

Oligonucleotides

Synthetic DNA oligonucleotides used for the selection/amplification experiments are (name, sequence): T7TOP, TAATACGACTCACTATAGGGATCCAAGATGCCGACT; RTBOT, GCGTCTCGAGCGTAGTTA; and T7TEMPLATE(20), GCGTCTCGAGCGTAGTTAN$_{20}$AGTCGGCATCTTGGATCCCTATAGTGAGTGTATTA (where N$_{20}$ indicates that A, G, T or C was inserted randomly for 20 consecutive bases). The following oligonucleotides were used as PCR primers to amplify coding regions of hnRNP A1 cDNA: A1RBDITOP, GCCCATATGTCTAAGTCAGAGTCT; A1RBDIBOT, GTTAAGTGGGTACCTGGTCT; A1RBDIITOP, GCTGTCTCCATGGAAGATTCTC; A1GLYBOT, TTGTTGGATCCTTAAAATCTTCTGCC; and A1GLYTOP, CGAAGTGGTTCTGGTACCTTTGG. All synthetic DNA and RNA was made on an Applied Biosystems (Foster City, CA) 392 DNA–RNA oligonucleotide synthesizer. The sequences of the oligoribonucleotides are given in Figure 4 and the sequence of A1R8.7 (see Figure 4) is ACUUAACGAUUAGGGACAUA. All synthetic RNAs were purified by ion exchange HPLC as described in Görlach *et al.* (1992).

Expression of hnRNP A1 constructs in bacteria

Full-length hnRNP A1 was expressed in *E.coli* using the pET expression system (Novagen) as described previously (Portman and Dreyfuss, 1994). DNA encoding the individual RBDs or the glycine-rich carboxyl domain (GLY) was produced by PCR (see above for primer sequences). To make RBD I:GLY, a *Kpn*I site was generated at the end of the RBD I coding region and at the beginning of the GLY coding region. These two fragments were cut with *Kpn*I, ligated, cut with *Nde*I and *Bam*HI and then cloned into pET11a. To express RBD II:GLY, an *Nco*I site was generated by PCR at the beginning of the RBD II coding region and a *Bam*HI site was generated at the end of hnRNP A1 coding region. This DNA fragment was cloned into pET11d. All cloned PCR-generated DNAs were sequenced to confirm the correct sequence. The RBD I:GLY construct (see Figure 2) encoded hnRNP A1 amino acids 1–100 fused to amino acids 200–320 and the RBD II:GLY construct encoded amino acids 89–320. Protein expression was induced and protein was harvested according to previously published procedures (Portman and Dreyfuss, 1994).

Selection/amplification from pools of random sequence RNA

We used a strategy based on the SELEX method of Tuerk and Gold (1990). The sequences of all the DNA oligonucleotides used for these experiments are listed above. Bacterially made hnRNP A1 was produced using the pET system and purified to homogeneity as described by Portman and Dreyfuss (1994). Approximately 3 × 10^{-11} mol purified A1 was immobilized on 25 μl protein A Sepharose (Pharmacia) with the mAb 9H10 (unpublished data). Individual RBD constructs were immunopurified from bacterial extracts. The initial round of selection used 3.5 × 10^{-9} mol RNA which corresponds to an RNA:protein ratio of 110:1. In subsequent rounds of selection this ratio was ~180:1. Purified RNA was added to immobilized hnRNP A1 in a final volume of 100 μl binding buffer [10 mM Tris–HCl (pH 7.4), 100 mM KCl, 2.5 mM $MgCl_2$ and 0.1% Triton X-100] and was allowed to incubate at room temperature for 20 min with gentle flicking of the tube every minute. After selection, the beads were washed four times with 1 ml binding buffer and 200 μl 4 M urea was added after the last wash. 50 μg proteinase K was then added and this mixture was incubated for 20 min at 37°C. This mixture was phenol:chloroform extracted and precipitated with ethanol.

The selected RNA was amplified by RT-PCR. To do so, 100 ng RTBOTT primer was annealed to the selected RNA by incubating the RNA and primer in 10 μl 10 mM Tris (pH 7.4) and 50 mM KCl at 65°C for 3 min, allowing to cool to 23°C over 10 min, and placed on ice. The mixture was reverse transcribed in a total of 25 μl according to the reverse transcriptase manufacturer's (Boehringer Mannheim) instructions. 5 μl of this mixture was subsequently used for PCR. Amplification was carried out for 35 cycles and the resultant DNA was cut with *Bam*HI and *Xho*I and cloned into pGEM7Z. Plasmids were sequenced by the dideoxynucleotide chain termination method using Sequenase (US Biochemical). The sequences of 30 hnRNP A1-selected clones, 11 RBD I:GLY-selected clones and 22 RBD II:GLY-selected clones were determined.

In vitro transcriptions

The RNA pool for the first round of the selection/amplification experiments was transcribed from TEMPLATE20 and T7TOP DNA oligonucleotides (Milligan *et al.*, 1987) and was purified by gel electrophoresis (Maniatis *et al.*, 1982). Subsequent transcriptions used PCR-generated DNA in 100 μl reactions. T7 RNA polymerase was purchased from US Biochemical and all nucleotides were purchased from Pharmacia or Boehringer Mannheim. Following transcription, the DNA template was destroyed by addition of 50 U RNase-free DNase (Boehringer Mannheim) and incubation at 37°C for 30 min. The RNA was purified by extracting once with phenol:chloroform and then by G25 spin chromatography (Maniatis *et al.*, 1982). The RNA was precipitated and quantitated spectrophotometrically.

The RNA for the UV crosslinking experiments was produced from synthetic oligonucleotides and contained an additional two guanosine residues at their 5′ ends to increase the efficiency of the transcription (Milligan *et al.*, 1987). Synthesis of capped pre-mRNA for splicing reactions was performed as described (Choi *et al.*, 1986) and purified on 10% polyacrylamide urea gels.

Nitrocellulose filter binding assays

The determination of dissociation constant (K_d) values was done in a two-part method based upon the procedure of Carey and Uhlenbeck (1983). Each RNA tested (listed in Figure 4) was 5′ end-labeled to a specific activity of ~1 × 10^8 c.p.m./μg by label transfer from [γ-^{32}P]ATP (Amersham) by T4 polynucleotide kinase (NEB) (Maniatis *et al.*, 1982). Labeled RNAs were subsequently purified by spin chromatography through G10 Sephadex (Pharmacia) (Maniatis *et al.*, 1982). In the first part of the experiment, increasing amounts of purified hnRNP A1 were added to 50 fmol of ^{32}P-labeled RNA in a total volume of 100 μl binding buffer (without Triton X-100). This mixture was allowed to incubate for 20 min at 23°C, and 90 μl was slowly filtered through a pre-soaked nitrocellulose filter. Each filter was washed with 3 ml ice-cold binding buffer and then dried. The amount of RNA retained on the filter was determined by liquid scintillation counting. Each assay included a control for RNA retention without any added protein and this value was never >2% input RNA. The retention efficiency of RNA–protein complexes averaged ~60%. In the second step, the amount of protein competent for RNA binding was determined by an RNA excess assay. After correcting for the retention efficiency of the filter, the amount of hnRNP A1 active for RNA binding was 44%. The results of the protein excess experiments were corrected and normalized to saturation for each experiment to facilitate direct comparison. Finally, to determine the K_d value, the results of the protein excess assays were expressed as the fraction of RNA bound at saturation and plotted as a function of hnRNP A1 concentration using the SigmaPlot (Jandel Scientific, Corte Madera, CA) software package. The protein concentration at which half of the amount of RNA bound at saturation is retained on the filter is equal to the dissociation constant (K_d) (Carey and Uhlenbeck, 1983; Irvine *et al.*, 1991).

Pre-mRNA splicing assays

Assays were performed as described previously (Choi *et al.*, 1986). For oligoribonucleotide inhibition assays all reaction components, except pre-mRNA, were mixed, incubated on ice for 5 min, followed by the addition of pre-mRNA. In cases where purified protein was added to the reactions, the protein was added 5 min after the oligoribonucleotide, then incubated for 5 min on ice, followed by the addition of the pre-mRNA.

UV light-induced crosslinking assays

The ^{32}P-labeled RNA used for these assays was produced by transcription from synthetic DNA oligonucleotides and was gel purified on 18%

polyacrylamide urea gels. For crosslinking, ~1 × 10^5 c.p.m. RNA was added to a 20 μl splicing extract (without any added ATP or pre-mRNA), incubated on ice for 5 min and then irradiated for 8 min as described. After irradiation, the samples were treated with a cocktail of RNase A (5 μg/sample) and micrococcal nuclease (10 U/sample) and were subjected to SDS–PAGE. To immunopurify proteins from crosslinking reactions, 5 μl of mAb 9H10 ascites, or nonimmune serum (SP2/0), was used to immunopurify from two crosslinking reactions (total of 40 μl).

Acknowledgements

We would like to thank Douglas Portman for generously providing recombinant hnRNP A1 and nuclear extract, Matthias Görlach for M104 protein, Serafín Piñol-Roma for mAb 9H10, Vartkes Apkarian for assistance with oligonucleotide synthesis and Gina Daly for help with the manuscript. We also thank members of our laboratory for critical reading of the manuscript. This work was supported with grants from the National Institutes of Health and the Howard Hughes Medical Institute.

References

Bartel,D.P., Zapp,M.L., Green,M.R. and Szostak,J.W. (1991) *Cell*, **67**, 529–536.

Ben-David,Y., Bani,M.R., Chabot,B., De Koven,A. and Berstein,A. (1992) *Mol. Cell. Biol.*, **12**, 4449–4455.

Bennett,M., Piñol-Roma,S., Staknis,D., Dreyfuss,G. and Reed,R. (1992) *Mol. Cell. Biol.*, **12**, 3165–3175.

Burd,C.G., Swanson,M.S., Görlach,M. and Dreyfuss,G. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9788–9792.

Burd,C.G., Matunis,E.L. and Dreyfuss,G. (1991) *Mol. Cell. Biol.*, **11**, 3419–3424.

Buvoli,M., Cobianchi,F., Biamonti,G. and Riva,S. (1990) *Nucleic Acids Res.*, **18**, 6595–6600.

Buvoli,M., Cobianchi,F. and Riva,S. (1992) *Nucleic Acids Res.*, **20**, 5017–5025.

Carey,J. and Uhlenbeck,O.C. (1983) *Biochemistry*, **22**, 2601–2610.

Casas-Finet,J.R., Karpel,R., 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., SenGupta,D.N., Zmudzka,B.Z. and Wilson,S.H. (1986) *J. Biol. Chem.*, **261**, 3536–3549.

Cobianchi,F., Karpel,R.L., Williams,K.R., Notario and Wilson,S.H. (1988) *J. Biol. Chem.*, **263**, 1063–1071.

Cobianchi,F., Calvio,C., Stoppini,M., Buvoli,M. and Riva,S. (1993) *Nucleic Acids Res.*, **21**, 949–955.

Dreyfuss,G. (1986) *Annu. Rev. Biochem.*, **2**, 459–498.

Dreyfuss,G., Swanson,M.S. and Piñol-Roma,S. (1988) *Trends Biochem. Sci.*, **13**, 86–91.

Dreyfuss,G., Swanson,M.S. and Piñol-Roma,S. (1990) In Strauss,P. and Wilson,S. (eds), *The Eukaryotic Nucleus: Structure and Function*. The Telford Press, Caldwell, NJ, Vol. 2, pp. 503–517.

Dreyfuss,G., Matunis,M.J., Piñol-Roma,S. and Burd,C. (1993) *Annu. Rev. Biochem.*, **62**, 289–321.

Eperon,I.C., Ireland,D.C., Smith,R.A., Mayeda,A. and Krainer,A.R. (1993) *EMBO J.*, **12**, 3607–3617.

Görlach,M., Wittekind,M., Beckman,R.A., Mueller,L. and Dreyfuss,G. (1992) *EMBO J.*, **11**, 3289–3295.

Grabowski,P.J., Padgett,R.A. and Sharp,P.A. (1984) *Cell*, **37**, 415–427.

Hall,K.B. and Konarska,M.M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 10969–10973.

Hellen,C.U.T., Witherell,G.W., Schmid,M., Shin,S.H., Pestova,T.V., Gil,A. and Wimmer,E. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 7642–7646.

Irvine,D., Tuerk,C. and Gold,L. (1991) *J. Mol. Biol.*, **222**, 739–761.

Ishikawa,F., Matunis,M.J., Dreyfuss,G. and Cech,T.R. (1993) *Mol. Cell. Biol.*, **13**, 4301–4310.

Kiledjian,M. and Dreyfuss,G. (1992) *EMBO J.*, **11**, 2655–2664.

Kumar,A. and Wilson,S.H. (1990) *Biochemistry*, **29**, 10717–10722.

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.

LeStourgeon,W.M., Barnett,S.F. and Northington,S.J. (1990) In Strauss,P. and Wilson,S. (eds), *The Eukaryotic Nucleus: Structure and Function*. The Telford Press, Caldwell, NJ, Vol. 2, pp. 477–502.

Lutz-Freyermuth,C., Query,C.C. and Keene,J.D. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 6393–6397.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Matunis,E.L., Matunis,M.J. and Dreyfuss,G. (1992) *J. Cell Biol.*, **116**, 257–269.

Matunis,E.L., Matunis,M.J. and Dreyfuss,G. (1993) *J. Cell Biol.*, **121**, 219–228.

Mayeda,A. and Krainer,A.R. (1992) *Cell*, **68**, 365–375.

Mayeda,A., Elfman,D.M. and Krainer,A.R. (1993) *Mol. Cell. Biol.*, **13**, 2993–3001.

Mayrand,S.H. and Pederson,T. (1987) *Mol. Biol. Rep.*, **12**, 182.

McKay,S.J. and Cooke,H. (1992) *Nucleic Acids Res.*, **20**, 6461–6464.

Merrill,B.M., Stone,K.L., Cobianchi,F., Wilson,S.H. and Williams,K. (1988) *J. Biol. Chem.*, **263**, 3307–3313.

Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhelenbeck,O.C. (1987) *Nucleic Acids Res.*, **15**, 8783–8798.

Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.

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.

Nietfeld,W., Mentzel,H. and Pieler,T. (1990) *EMBO J.*, **9**, 3699–3705.

Ohshima,Y. and Gotoh,Y. (1987) *J. Mol. Biol.*, **195**, 247–259.

Piñol-Roma,S. and Dreyfuss,G. (1992) *Nature*, **355**, 730–732.

Pontius,B.W. and Berg,P. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8403–8407.

Portman,D. and Dreyfuss,G. (1994) *EMBO J.*, **13**, 213–221.

Schnieder,D., Tuerk,C. and Gold,L. (1992) *J. Mol. Biol.*, **228**, 862–869.

Schneider,D., Gold,L. and Platt,T. (1993) *FASEB J.*, **7**, 201–207.

Swanson,M.S. and Dreyfuss,G. (1988a) *EMBO J.*, **11**, 3519–3529.

Swanson,M.S. and Dreyfuss,G. (1988b) *Mol. Cell. Biol.*, **8**, 2237–2241.

Tsai,D.E., Harper,D.S. and Keene,J. (1991) *Nucleic Acids Res.*, **19**, 4931–4936.

Tuerk,C. and Gold,L. (1990) *Science*, **249**, 505–510.

Zamore,P.D., Patton,J.G. and Green,M.R. (1992) *Nature*, **355**, 609–614.

Received on November 11, 1993; revised on December 16, 1993