MOLECULAR AND CELLULAR BIOLOGY, July 1992, p. 3165–3175
0270-7306/92/073165-11$02.00/0

Vol. 12, No. 7

# Differential Binding of Heterogeneous Nuclear Ribonucleoproteins to mRNA Precursors prior to Spliceosome Assembly In Vitro

MARIA BENNETT, SERAFÍN PIÑOL-ROMA,† DAVID STAKNIS, GIDEON DREYFUSS,† AND ROBIN REED*

*Cellular and Molecular Physiology, Program in Cell and Developmental Biology, Harvard Medical School, Boston, Massachusetts 02115*

Received 3 March 1992/Accepted 23 April 1992

**We have investigated the composition of the earliest detectable complex (H) assembled on pre-mRNA during the in vitro splicing reaction. We show that most of the proteins in this complex correspond to heterogeneous nuclear ribonucleoproteins (hnRNP), a set of abundant RNA-binding proteins that bind nascent RNA polymerase II transcripts in vivo. Thus, these studies establish a direct parallel between the initial events of RNA processing in vitro and in vivo. In contrast to previous studies, in which total hnRNP particles were isolated from mammalian nuclei, we determined the hnRNP composition of complexes assembled on individual RNAs of defined sequence. We found that a unique combination of hnRNP proteins is associated with each RNA. Thus, our data provide direct evidence for transcript-dependent assembly of pre-mRNA in hnRNP complexes. The observation that pre-mRNA is differentially bound by hnRNP proteins prior to spliceosome assembly suggests the possibility that RNA packaging could play a central role in the mechanism of splice site selection, as well as other posttranscriptional events.**

During transcription, nascent heterogeneous nuclear RNAs (hnRNAs) associate with a distinct set of abundant nuclear proteins, known as heterogeneous nuclear ribonucleoproteins (hnRNP) to form hnRNP complexes (see references 10, 12, and 14 for reviews). The major hnRNP proteins detected by sodium dodecyl sulfate (SDS)-gel electrophoresis of isolated mammalian hnRNP complexes (also referred to as hnRNP particles) include a group of 35- to 45-kDa proteins and groups of 68- and 120-kDa proteins (7, 12). However, at least 20 distinct hnRNP proteins have been identified by two-dimensional gel electrophoresis (36). On the basis of the observations that hnRNP proteins bind to nascent pre-mRNA and that splicing can be visualized on these transcripts in the electron microscope (2, 32), hnRNP complexes are thought to be the substrate for posttranscriptional processing events (12).

In vitro studies have shown that pre-mRNA splicing takes place within spliceosomes, large multicomponent complexes containing U1, U2, U4, U5, and U6 small nuclear RNPs (snRNPs) (see references 21 and 28 for reviews) and a number of non-snRNP proteins (26, 27, 38). Spliceosome assembly occurs by means of a stepwise pathway. Pre-mRNA is first assembled into an ATP-independent complex which lacks snRNPs (15, 19, 24, 38). This complex (H) does not appear to be specific to splicing because it assembles as efficiently on natural pre-mRNAs as on RNAs lacking functional splice sites (19, 24, 38). However, H complex assembly precedes that of any splicing-specific complexes (30), indicating that this complex is either a spliceosome precursor or that it disassembles prior to spliceosome formation.

Following H complex assembly, the first complex specific to splicing, E (early) complex, is an ATP-independent complex that commits pre-mRNA to the splicing pathway (30). E complex is a precursor to the first specific ATP-dependent complex (A), which contains U1 and U2 snRNPs (20, 23, 30). U4, U5, and U6 snRNPs then join A complex to form the mature spliceosome, or B complex, which contains U1, U2, U4, U5, and U6 snRNPs (4, 20, 24, 30).

Despite extensive characterization of the structure and composition of hnRNP complexes isolated from mammalian cells, the specific functions of these complexes in nuclear metabolism of pre-mRNAs have not been established (12). However, at least some of the hnRNP proteins have been shown to be associated with the splicing machinery. For example, antibodies against hnRNP C proteins specifically immunoprecipitate spliceosomes assembled in vitro and inhibit the splicing reaction (9, 42). Furthermore, binding studies showed that the hnRNP proteins A, C, and D interact specifically with the 3′ splice site and that mutations in this element disrupt binding (45). In addition to demonstrating an association between hnRNP proteins and pre-mRNA, these and other studies indicated that hnRNP proteins bind RNA in a sequence-dependent manner (6, 45). Moreover, different hnRNP proteins show high binding affinities for different ribohomopolymers, such as poly(U) or poly(G) agarose (46). However, other studies have shown that hnRNP proteins can also bind RNA nonspecifically (10, 11). The conclusions that hnRNP proteins can bind RNA both specifically and nonspecifically are not necessarily contradictory (see reference 14 for discussion). Nevertheless, the actual organization of hnRNP proteins on natural pre-mRNAs remains to be established.

In this study, we investigated the protein composition of H complex assembled on different RNAs in vitro. Previous studies revealed that the predominant proteins in H complex assembled on β-globin pre-mRNA have molecular sizes between 35 and 45 kDa and approximately 65 and 116 kDa (38). On the basis of the similarity in size to hnRNPs and the observation that the proteins in H complex bind to diverse

* Corresponding author.
† Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6148.

RNAs added to splicing extracts, we hypothesized that H complex may contain hnRNP proteins. Here we show that H complex contains the same set of hnRNP proteins as are associated with nascent hnRNA in vivo. We find that distinct combinations of hnRNP proteins associate with pre-mRNAs of different sequences. Thus, our data show that hnRNP complexes assemble on pre-mRNA in a sequence-dependent manner. The implications of transcript-dependent binding of pre-mRNA by hnRNP proteins are discussed.

## MATERIALS AND METHODS

**Plasmids.** pAdML and T7-Hβ have been described by Michaud and Reed (30). pAd3′ was constructed by ligating the *Fsp*I-*Bam*HI fragment from pAdML into the *Pvu*II-*Bam*HI sites of SP72 (Promega). pGC+DX, which encodes a portion of rat α-tropomyosin (44), and pB3P3, which encodes the 3′ portion of the rat α-tropomyosin intron 2 (35), were gifts from B. Nadal Ginard. pAdML, pAd3′, and T7-Hβ were linearized with *Bam*HI for in vitro transcription with T7 polymerase. pAdML was linearized with *Fsp*I to synthesize Ad5′ RNA and with *Eco*RI to synthesize the antisense of AdML RNA (transcription was done with SP6 polymerase). pGC+DX and pB3P3 were linearized with *Bam*HI and *Acc*I, respectively, for transcription with SP6 polymerase.

**Pre-mRNA synthesis and in vitro splicing reactions.** Biotinylated pre-mRNAs (4, 20, 38) were synthesized with SP6 or T7 polymerase in standard transcription reactions (29), which included 15 μM biotinylated UTP (Enzo Biochemicals) and 100 μM UTP. In vitro splicing reactions were carried out as described previously (25). For splicing complex purification, large-scale in vitro splicing reaction mixtures (1.2 ml) contained 4.8 μg of $^{32}$P-labeled biotinylated RNA. For spliceosome purification, reaction mixtures were incubated at 30°C for 15 min. For purification of H complexes, reaction mixtures did not contain ATP, $MgCl_2$, and creatine phosphate and were incubated for 5 min. For Fig. 2, 5, 6, and 7, the HeLa cells used for nuclear extracts were obtained from the Massachusetts Institute of Technology, while HeLa cells from Endotronics were used for Fig. 1 and 3.

**Structure of RNAs.** The AdML pre-mRNA (236 nucleotides [nt]) contains exon 1 (92 nt), intron 1 (97 nt), and exon 2 (47 nt). Ad5′ RNA (128 nt) consists of the 5′ portion of the AdML pre-mRNA, while Ad3′ RNA (108 nts) consists of the 3′ portion of the AdML pre-mRNA. The AdML antisense RNA is 240 nt long and does not contain splicing signals. β-Globin pre-mRNA (505 nt) contains a 130-nt intron, a 170-nt exon 1, and a 205-nt exon 2. α-Tropomyosin pre-mRNA (547 nt) contains a 218-nt intron, and B3P3 RNA consists of 135 nt of the 3′ portion of the α-tropomyosin intron (including the branch-point sequence and a portion of the pyrimidine tract).

**Splicing complex purification.** Large-scale in vitro splicing reaction mixtures (1.2 ml) were loaded directly onto a Sephacryl S-500 column (1.5 by 50 cm) as described previously (30, 38). Fractions (0.5 ml) were collected at a flow rate of ~4 ml/h. An aliquot of each fraction was counted by the Cerenkov method to identify column fractions containing splicing complexes. The peak fractions were then pooled, and 10 to 15 μl of avidin-agarose per ml was added. After being mixed overnight at 4°C, the avidin-agarose was collected, washed once with 20 mM Tris-HCl (pH 7.8)–100 mM NaCl, and washed four times with the same buffer for 15 min each at 4°C. The amount of $^{32}$P-labeled RNA bound to the avidin agarose was determined by counting by the Cerenkov method. Analysis of the protein composition of the purified complexes was carried out as described previously (38).

**Two-dimensional gel electrophoresis.** Two-dimensional gel electrophoresis was carried out essentially as described by O'Farrell et al. (31). The first dimension was separated by nonequilibrium pH gradient gel electrophoresis, and the second dimension was resolved by SDS–12.5% polyacrylamide gel electrophoresis (PAGE) as described by Dreyfuss et al. (13). The separated proteins were visualized by silver staining.

**Immunopurification of hnRNP complexes.** hnRNP complexes were immunopurified from the nucleoplasm of HeLa cells with the anti-C protein monoclonal antibody 4F4 (8) as previously described (7, 36). The immunopurified complexes were eluted from the protein A-Sepharose beads with SDS-PAGE sample buffer or with nonequilibrium pH gradient gel electrophoresis sample buffer for analysis of the protein composition by SDS-PAGE or by two-dimensional gel electrophoresis, respectively.

## RESULTS

Several methods have been employed to fractionate spliceosomes and intermediate complexes in the spliceosome assembly pathway, including density gradient sedimentation (5, 15, 19), native gel electrophoresis (23, 24), and gel filtration (30, 39). With any of these methods, the first complex detected assembles with similar efficiencies on RNAs containing or lacking functional splice sites. This complex has been designated H (24) or nonspecific (48) complex, when detected by native gel electrophoresis, and appears to correspond to a 30S complex on density gradients (3, 15, 19).

H complex forms immediately when pre-mRNA is added to nuclear extracts in either the presence or absence of ATP and at 0° or 30°C (data not shown) (15, 19, 23, 30). Following H complex assembly, the kinetics and efficiency of spliceosome assembly vary significantly between different pre-mRNA substrates and between different preparations of nuclear extracts (38a) (see below). With β-globin pre-mRNA, H complex is detected after a 5-min incubation in the presence or absence of ATP (Fig. 1A). Spliceosomes containing U1, U2, U4, U5, and U6 snRNPs (designated B complex) (30) are detected at the 30-min time point in the presence of ATP, at which time significant levels of H complex are also still present (Fig. 1A).

Although H complexes detected under different conditions and at different times fractionate similarly, it is not known whether their protein composition is the same. To investigate the composition of H complexes formed under various conditions, we used the two-step, large-scale affinity chromatography procedure previously used to determine the protein composition of H complex and other splicing complexes (30, 38). By this method, complexes are assembled on biotinylated RNA, fractionated by gel filtration, and then affinity purified by binding to avidin agarose. In a previous study, H complex was assembled on β-globin pre-mRNA by incubation for 15 min at 30°C in the absence of ATP (38). This complex was found to contain a group of proteins with sizes between 35 and 45 kDa and proteins with sizes of approximately 68 and 116 kDa. Comparison by SDS-gel electrophoresis of H complexes assembled on β-globin pre-mRNA in the absence or presence of ATP (Fig. 1A) shows that they are similar to one another (Fig. 1B, compare lanes 1 to 3) and to H complex previously reported (38). We

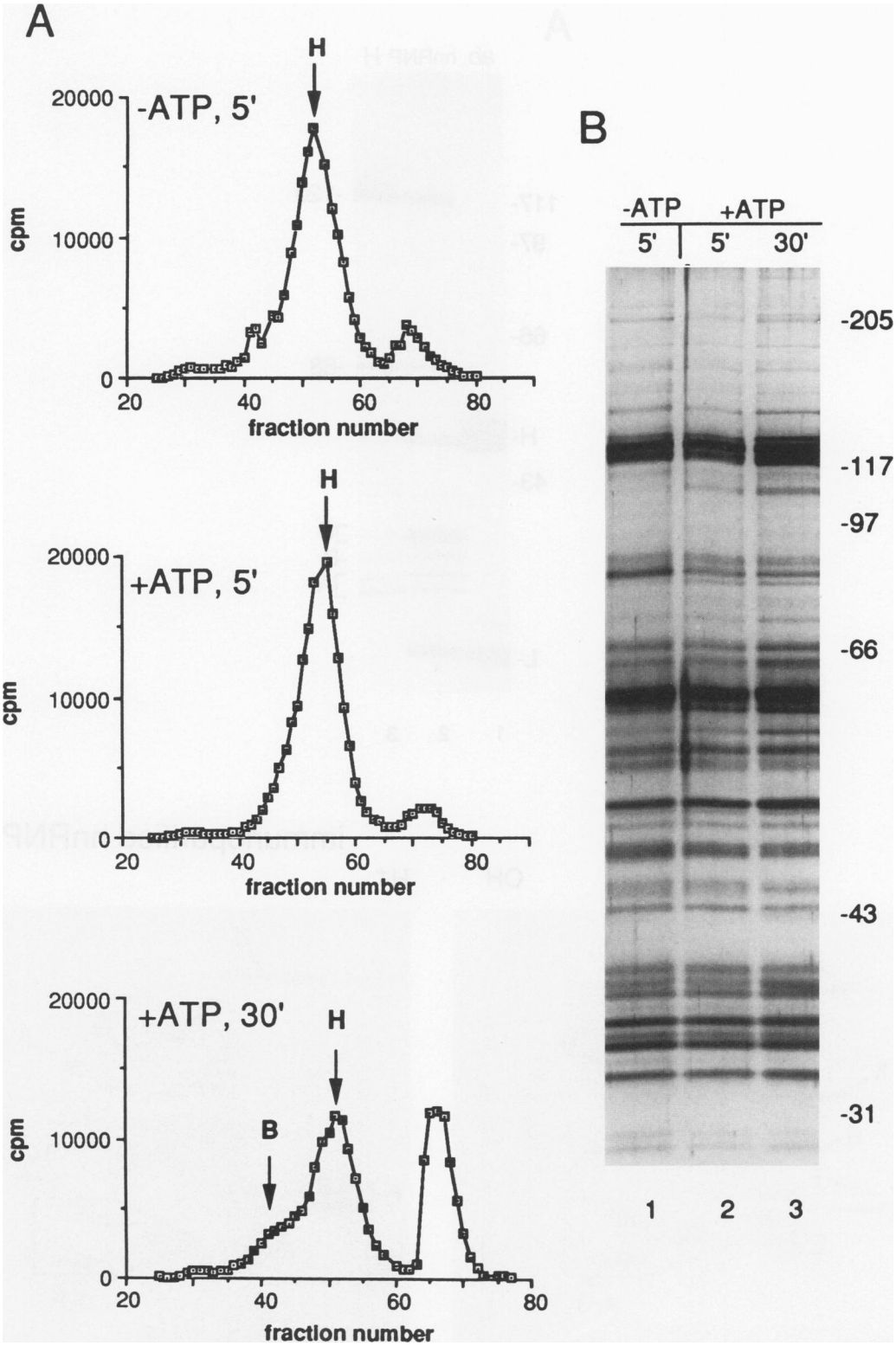


FIG. 1. The protein compositions of H complex assembled in the presence or absence of ATP are similar. (A) $^{32}$P-labeled β-globin pre-mRNA was incubated under splicing conditions in the presence or absence of ATP for the times indicated and then fractionated by gel filtration. The positions of H complex and spliceosome (B complex) are indicated. The peak in fractions 65 to 75 is degraded RNA. (B) Affinity-purified H complexes assembled in the absence (5 min [5′], lane 1) or presence (5 min [5′], lane 2; 30 min [30′], lane 3) of ATP and magnesium were fractionated on an SDS–9% polyacrylamide gel. The sizes (in thousands) of molecular weight markers are indicated on the right. Protein bands were detected by silver staining.

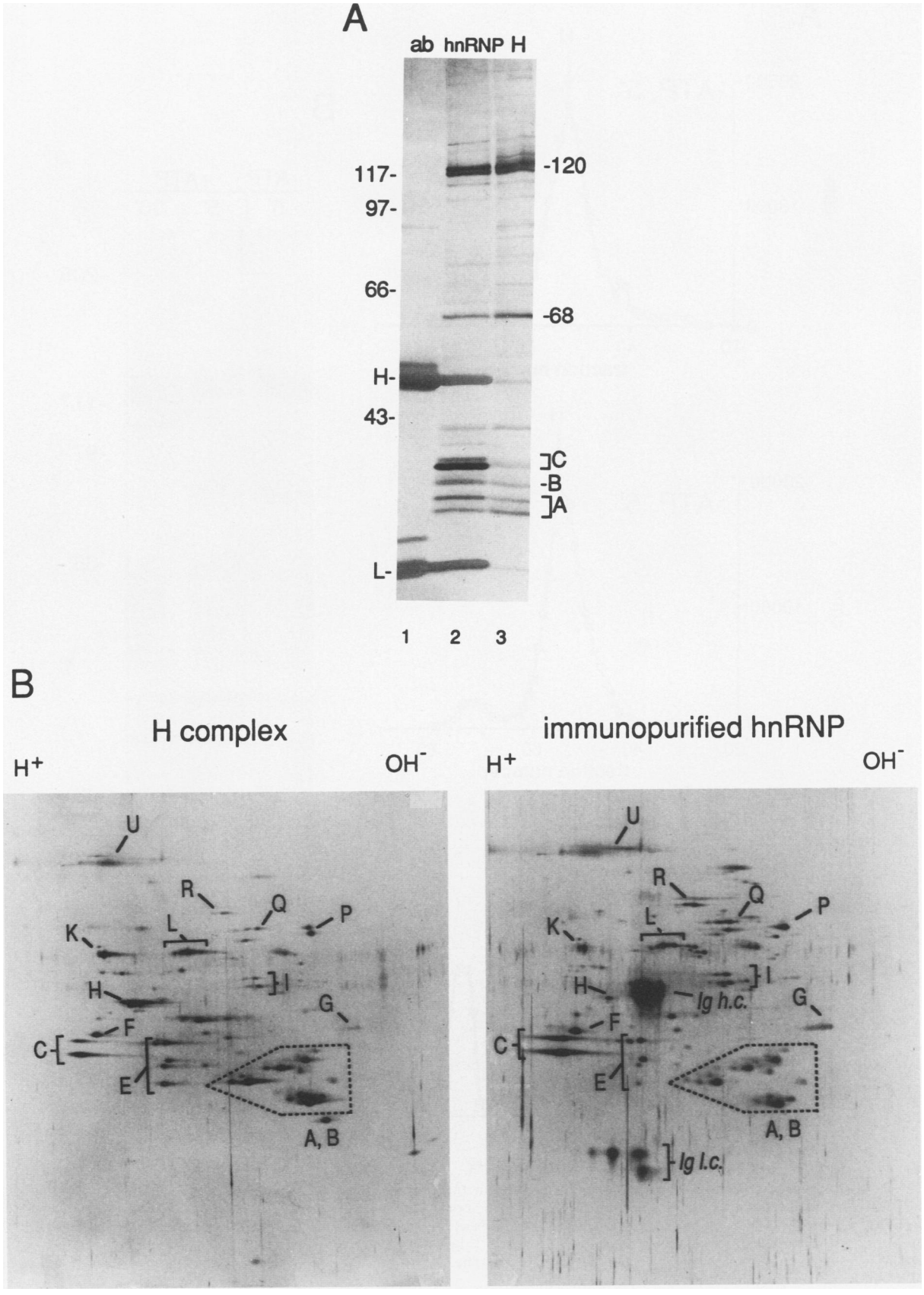
A
ab hnRNP H
117-
97-
66-
H-
43-
L-
-120
-68
]C
-B
]A
1 2 3
B
H complex
immunopurified hnRNP
H+
OH-
U
R
Q
P
K
L
I
H
G
F
C
E
A, B
lg h.c.
lg l.c.

FIG. 2. H complex assembled in vitro has a similar protein composition to hnRNP particles isolated from mammalian nuclei. (A) Affinity-purified H complex (lane 3) and immunoaffinity-purified hnRNP particles (lane 2) were fractionated on an SDS–9% polyacrylamide gel. The 4F4 monoclonal antibody is shown in lane 1. The heavy and light chains of the antibody are indicated by H and L, respectively. The abundant hnRNP proteins are designated on the right, and sizes (in thousands) of the molecular weight markers are indicated on the left. (B) Affinity-purified H complex and immunoaffinity-purified hnRNP particles were fractionated by two-dimensional nonequilibrium gel electrophoresis. The first dimension was nonequilibrium isoelectric focusing, and the second dimension was SDS–12.5% polyacrylamide. The designations of the hnRNP proteins are according to Piñol-Roma et al. (36). The heavy and light chains of the 4F4 monoclonal antibody are indicated as Ig h.c. and Ig l.c., respectively.

conclude that the composition of H complex is not affected significantly by incubation under different conditions.

In order to determine whether H complex contains any hnRNP proteins, the protein composition of H complex assembled on β-globin pre-mRNA was compared with that of immunoaffinity-purified hnRNP complexes (Fig. 2). hnRNP complexes were immunoaffinity purified from nucleoplasm by using a Sepharose-bound monoclonal antibody, 4F4, against hnRNP C proteins (7, 36). Strikingly, the purified H complex contains proteins that comigrate with the previously described hnRNP proteins, including the 35- to 45-, 68-, and 120-kDa protein groups (Fig. 2A, compare lanes 2 and 3). Moreover, the stoichiometry of the proteins appears similar in the two complexes (except for hnRNP C protein, which may be more abundant in the hnRNP complexes because the anti-hnRNP C protein monoclonal antibody was used for their isolation).

Further evidence that the proteins in H complex correspond to hnRNP proteins was obtained by comparison of H complex and hnRNP complexes by two-dimensional gel electrophoresis (Fig. 2B). This analysis revealed a remarkable similarity between the two particles. H complex contains most of the hnRNP proteins previously shown to be present in hnRNP complexes isolated from HeLa cell nucleoplasm (36). Notably, H complex does not contain significant levels of any proteins other than the known hnRNP proteins. We conclude that the protein composition of H complex formed on β-globin pre-mRNA in vitro is virtually identical to that of total hnRNP complexes isolated from mammalian nuclei. These results are also consistent with the general single-stranded nucleic acid-binding properties previously described for most of these hnRNP proteins (33, 36).

To determine whether the full complement of hnRNP proteins binds to all functional splicing substrates, we compared H complex assembly on three pre-mRNAs, AdML, α-tropomyosin, and β-globin (Fig. 3B, lanes 1 to 3). As expected, all three pre-mRNAs efficiently assemble H complexes (Fig. 3A). However, comparison of their protein compositions revealed a number of significant differences (Fig. 3C, lanes 1 to 3). First, the total amount of hnRNP proteins bound to AdML pre-mRNA is markedly lower than the amounts observed for the other RNAs (Fig. 3C, compare lanes 1 to 3). In addition, several proteins, including the 60- to 68-kDa proteins and the 30- to 34-kDa proteins, are specifically reduced in the AdML H complex relative to those in the α-tropomyosin or β-globin H complexes (Fig. 3C, compare lanes 1 to 3). Similarly, analysis of the protein composition of the α-tropomyosin H complex (Fig. 3C, lane 2) indicates that it also is distinct from those of AdML and β-globin.

To determine whether there is any obvious correlation between deficiencies in particular hnRNP proteins and the efficiency of spliceosome formation, we compared complex assembly on AdML, β-globin, and α-tropomyosin pre-mRNAs (Fig. 4). Equimolar amounts of each pre-mRNA were incubated under splicing conditions and then fractionated by gel filtration. U1, U2, U4, U5, and U6 small nuclear RNAs are present in the peak designated B complex in each column profile, confirming that these peaks contain spliceosomes (1a). Comparison of the B:H complex ratios between the columns shows that AdML and α-tropomyosin assemble spliceosomes significantly more efficiently than does β-globin (Fig. 4). Thus, there is not an apparent correlation between the H complex composition (Fig. 3C, compare lanes 1 to 3) and the efficiency of spliceosome assembly.

To determine whether the hnRNP proteins that are deficient in the AdML H complex (Fig. 3C) bind to the pre-mRNA later during the spliceosome assembly pathway, we examined the protein composition of affinity-purified spliceosomes assembled on AdML (Fig. 5A, lanes 1 to 3). H complex assembled on β-globin is shown in lane 6 as a marker for the hnRNP proteins absent in AdML H complex (lane 4). This comparison shows that the hnRNP proteins that are deficient in the AdML H complex are also deficient in the AdML spliceosome (e.g., compare the levels of A and B proteins in Fig. 5A, lanes 1 to 4 with lane 6). Thus, we conclude that each pre-mRNA is bound by hnRNP proteins in a transcript-dependent manner and that this differential binding is, at least in part, maintained in the spliceosome.

Two-dimensional gel analysis of H complex assembled on AdML pre-mRNA indicates that, as was observed for β-globin, most of the proteins present in the AdML H complex correspond to previously described hnRNP proteins (Fig. 5B). As expected from the comparison of AdML and β-globin H complexes by SDS-gel electrophoresis (Fig. 3C, lanes 1 and 3), two-dimensional gel electrophoresis of these H complexes reveals significant differences in the relative levels of particular hnRNP proteins (compare Fig. 2B and 5B). For example, hnRNP proteins A, B, H, and L are more abundant in β-globin H complex than in AdML H complex. Thus, these results show that the protein composition of the H complex is dependent on the RNA substrate.

The observed variations in H complex composition appear to be due to sequence differences between the RNAs. Other variables, such as amounts of RNA added to the nuclear extract or different preparations of H complex with the same nuclear extract, do not affect the protein composition of the complexes (data not shown). In addition, the sizes of the RNAs cannot account for the differences in H complex composition. β-Globin and α-tropomyosin pre-mRNAs are very similar in size (~500 nt, Fig. 3B) yet bind distinct subsets of hnRNP proteins (Fig. 3C). Similarly, other RNAs of the same size assemble distinct H complexes with different protein compositions (see below). However, one variable that does affect H complex composition is the source of HeLa cells used to prepare nuclear extracts (see Materials and Methods). We observed variations in the relative amounts of some hnRNP proteins in H complexes assembled in nuclear extracts prepared from different sources of HeLa cells. This difference is evident by comparing Fig. 3C (lanes

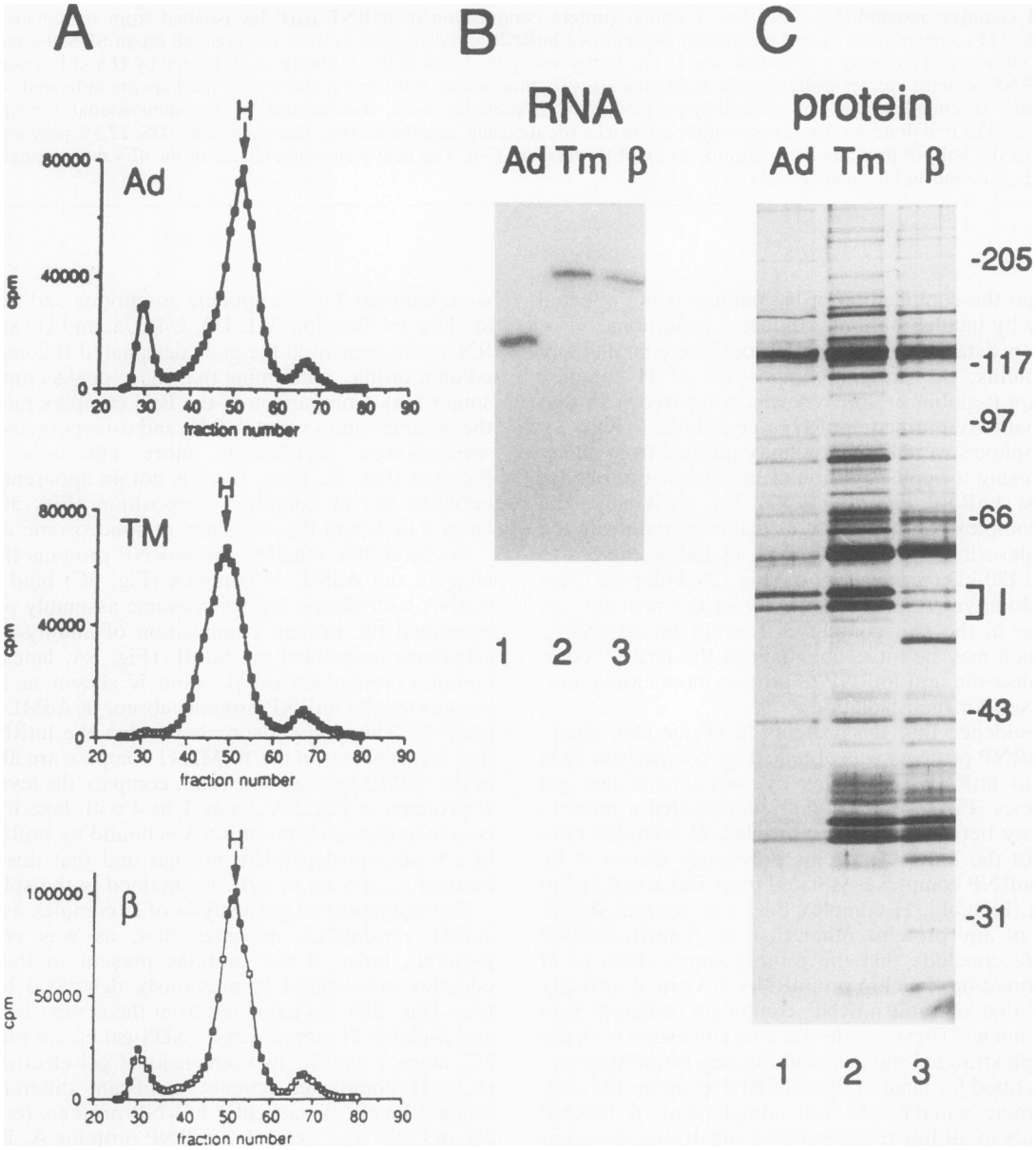


FIG. 3. Transcript-dependent binding of different pre-mRNAs in hnRNP particles. (A) Fractionation of H complexes. AdML, α-tropomyosin, or β-globin pre-mRNAs (3 μg) were incubated for 5 min in a splicing reaction mixture (750 μl) lacking ATP and then fractionated by gel filtration. The position of H complex (H) is indicated. The peak in fractions 25 to 35 is the void volume, and the peak in fractions 65 to 72 is degraded RNA. (B) Fractionation of pre-mRNAs. $^{32}$P-labeled AdML (lane 1), α-tropomyosin (lane 2), or β-globin (lane 3) pre-mRNAs were fractionated on an 8% denaturing polyacrylamide gel. (C) Protein composition of H complexes. AdML, α-tropomyosin, or β-globin H complexes were affinity purified and fractionated on an SDS–9% polyacrylamide gel. The sizes (in thousands) of the molecular weight standards and hnRNP I (I) are indicated on the right.

1 and 3) and 5A (lanes 4 and 6), in which AdML and β-globin H complexes were assembled in two different nuclear extracts. Notably, hnRNP I (identified by two-dimensional gel electrophoresis, e.g., Fig. 5B) is significantly more abundant in the H complexes shown in Fig. 3 (designated by a bracket and labeled I) than in those shown in Fig. 5A. Whether this difference is due to variations in the relative levels of hnRNP proteins or other factors that differ between the HeLa cells is not known. Nevertheless, in several extracts tested from a given Hela cell strain, reproducible transcript-dependent binding of each pre-mRNA was observed. Moreover, a similar set of hnRNP proteins is associated with a particular RNA in the two different HeLa cell strains (see Materials and Methods).

The observation that AdML, β-globin, and α-tropomyosin all formed H complexes containing different combinations of hnRNP proteins provides direct evidence that functional splicing substrates are assembled into hnRNP complexes in a transcript-dependent manner. However, the observed sequence specificity does not appear to be related to the fact that these RNAs are splicing substrates. Distinct patterns of proteins are also observed in H complexes assembled on the antisense transcript of AdML pre-mRNA (Fig. 5A, lane 5) and other RNAs lacking functional splice sites (data not

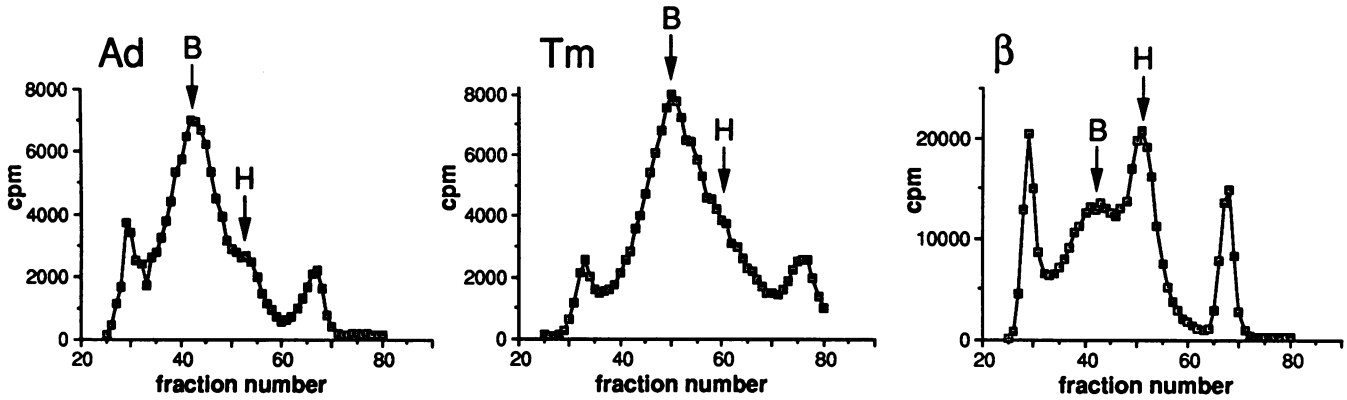


FIG. 4. Comparison of spliceosome assembly on AdML (Ad), α-tropomyosin (Tm), and β-globin (β) pre-mRNAs. Equimolar amounts of the pre-mRNAs indicated were incubated under splicing conditions for 15 min and then fractionated by gel filtration. The positions of the B and H complexes are indicated. The peaks in fractions 25 to 30 are the void volume of the column, and the peaks in fractions 60 to 80 are degraded RNA.

shown). These results indicate that there are sequence-dependent interactions between RNA and hnRNP proteins, regardless of whether the RNA is a splicing substrate.

Not unexpectedly, even greater differences in protein composition could be seen with H complexes assembled on short RNAs (~100 to 150 nt). Figure 6 shows two-dimensional gel analysis of H complexes assembled on Ad5′ (~128 nt) and Ad3′ (~108 nt), RNAs derived from the 5′ and 3′ portions of AdML pre-mRNA, respectively. While hnRNP A, B, C, and K proteins are virtually absent in the Ad5′ H complex, Ad3′ binds most of the proteins present in the AdML H complex (compare Fig. 5B and 6). An even more dramatic example of sequence dependence in hnRNP binding was observed for the H complex assembled on B3P3, a 135-nt pyrimidine-rich RNA consisting of the 3′ portion of the α-tropomyosin intron (35). SDS-gel analysis of H complex assembled on B3P3 shows one predominant group of proteins (Fig. 7A, marked with a bracket and labeled I). Analysis of the B3P3 H complex by two-dimensional electrophoresis shows that these proteins are hnRNP I (Fig. 7B; compare with the hnRNP particles in Fig. 2B and 5B). Thus, hnRNP I appears to bind preferentially to pyrimidine-rich sequences (hnRNP I corresponds to a previously characterized polypyrimidine tract-binding protein designated PTB (see Discussion) (18, 35). However, binding of hnRNP I is not restricted to the polypyrimidine tract at the 3′ end of introns, as hnRNP I is also one of the most abundant protein components of the H complex assembled on Ad5′ (Fig. 6). Similar studies showed that hnRNP A, C, and D also bind preferentially to RNA fragments derived from the 3′ end of introns (45). Thus, these observations indicate that several different hnRNP proteins interact with the 3′ splice site, raising the possibility that these proteins could play a role in the efficiency of splicing or in splice-site selection.

## DISCUSSION

We show here that the first complex that assembles on pre-mRNAs during the in vitro splicing reaction consists of previously described hnRNP proteins, the same class of abundant nuclear RNA-binding proteins that bind nascent pre-mRNAs in vivo. Thus, these studies establish a parallel between the initial events of nuclear pre-mRNA metabolism in vivo and in vitro. In both cases, the substrate for subsequent RNA processing events appears to be an hnRNP particle, in which the pre-mRNA is bound to a specific set of hnRNP proteins. At least 20 distinct hnRNP proteins, ranging in molecular size from 30 to 120 kDa, have been identified and characterized as stable components of total hnRNP complexes isolated from mammalian nuclei (36). This may be a low estimate, as detailed analysis of purified hnRNP complexes by two-dimensional gel electrophoresis revealed multiple additional forms of most of these proteins (36). Surprisingly, the hnRNP complex assembled during the in vitro splicing reaction, designated H complex, contains little other than these known hnRNP proteins.

In the nuclear extract, H complex assembly is followed by the stepwise formation of four distinct splicing complexes: two prespliceosome complexes (E and A) and two spliceosome complexes (B and C). It is not known whether a similar stepwise pathway occurs in vivo. The prespliceosome and spliceosome complexes have each been isolated from the in vitro splicing reaction and chased into spliced products by complementation assays (1, 30, 39). In contrast, H complex was not found to be a functional intermediate by these assays (30). However, the observation that all of the pre-mRNA added to the nuclear extract assembles into H complex prior to the detection of any of the functional intermediate complexes indicates that the pre-mRNA present in H complex is the substrate for the splicing reaction (30). Indeed, many of the proteins present in affinity-purified E, A, B, and C complexes are hnRNP proteins that are also found in H complex (1a, 38). Thus, H complex may not disassemble completely prior to spliceosome formation but instead may serve as the substrate for assembly of prespliceosome complexes. These observations suggest that, although H complex assembly may not be an obligatory step in the splicing reaction, the hnRNP proteins present in this complex could affect the efficiency of the splicing reaction and/or splice-site selection. It is, therefore, essential to understand the structure and composition of the hnRNP complex and to understand how the pre-mRNA is organized within these complexes.

H complex assembly in vitro occurs as efficiently on RNAs lacking splice sites as on functional pre-mRNAs (23, 38). This observation is consistent with the fact that hnRNP proteins bind readily to a variety of single-stranded nucleic acids (11, 36). However, despite the observation that the hnRNP proteins (hnRNP A to U) are general RNA-binding proteins, distinct binding preferences for particular se-

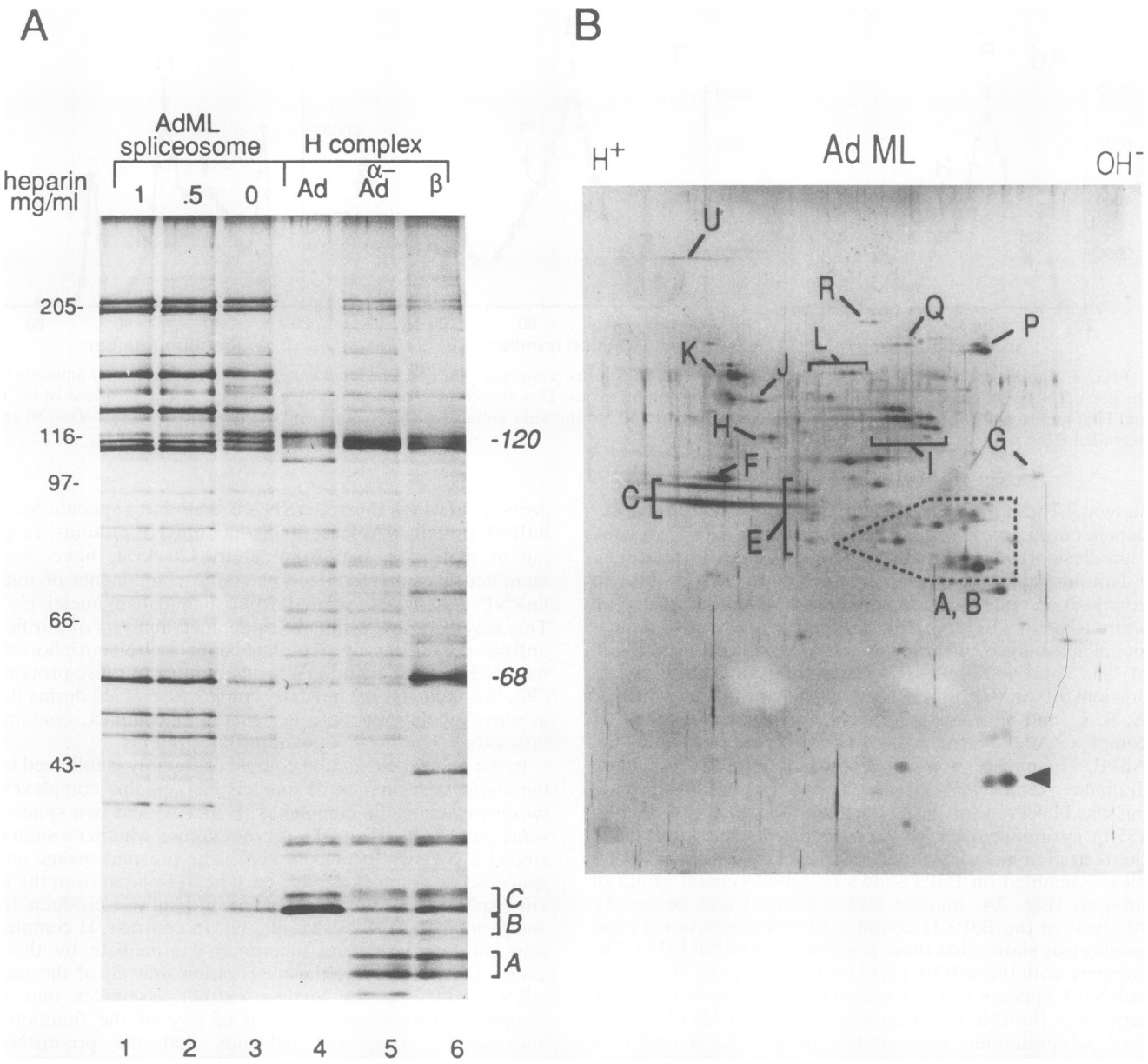


FIG. 5. Analysis of protein composition of H and B complexes. (A) Affinity-purified splicing complexes were fractionated on an SDS–9% polyacrylamide gel. The sizes (in thousands) of the molecular weight standards are shown to the left of the gel, and the A, B, C, 120-, and 68-kDa hnRNP proteins are indicated to the right. Spliceosomes (B complex) assembled on AdML pre-mRNA were fractionated by gel filtration and then either treated with 1 (lane 1), 0.5 (lane 2), or 0 (lane 3) mg of heparin per ml prior to affinity purification. These treatments did not significantly affect the protein composition of the spliceosome (compare lanes 1 to 3; note that all of the proteins bands are relatively fainter in lane 3 because less total protein was loaded in this lane). H complexes assembled on AdML (lane 4), α-AdML (lane 5), or β-globin (lane 6) RNAs were fractionated by gel filtration and then affinity purified under standard conditions. (B) Affinity-purified H complex assembled on AdML pre-mRNA was fractionated by two-dimensional nonequilibrium gel electrophoresis. The first dimension was nonequilibrium isoelectric focusing, and the second dimension was SDS–12.5% polyacrylamide. The designations of the hnRNP proteins are according to Piñol-Roma et al. (36). The arrowhead indicates avidin, which is released from the avidin-agarose during affinity purification.

quences have been observed for some of the proteins (45, 46). For example, hnRNP proteins A, C, and D bind avidly to sequences near the 3′ splice site of the intron, and mutations in these sequences reduce their binding at these sites (6, 45). Similarly, hnRNP I (which is the same as PTB; see below) interacts with the pyrimidine tract at the 3′ splice site, and these interactions are dependent upon the presence of a minimal number of pyrimidine residues (17, 35). Additional evidence that hnRNP proteins have distinct binding preferences comes from the observation that hnRNP proteins bind differentially to different ribohomopolymers (46). Finally, examination of the distribution of hnRNP L protein in amphibian oocyte lampbrush chromosomes showed that, while L binds most nascent transcripts, it is preferentially enriched in the C-rich transcripts generated at the giant loops (22, 37).

In the studies described here, we compared the protein composition of hnRNP complexes that assembled on indi-

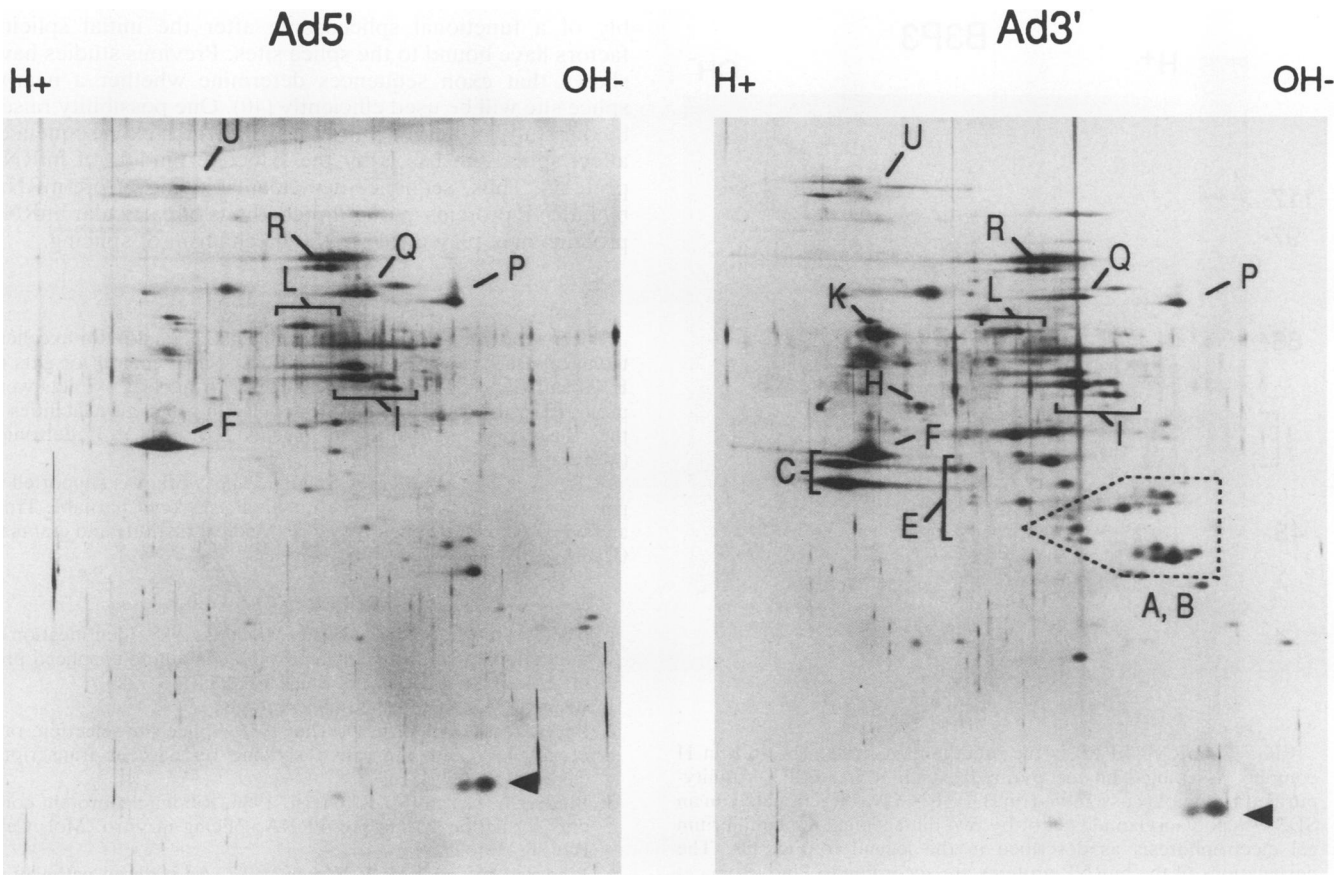


FIG. 6. Two-dimensional gel analysis of H complexes assembled on short RNAs. Affinity-purified H complexes assembled on Ad5′ and Ad3′ RNAs were fractionated by two-dimensional nonequilibrium gel electrophoresis as described in the legend to Fig. 5B. The designations of the hnRNP proteins are according to Piñol-Roma et al (36). The arrowhead indicates avidin, which was released from the avidin-agarose during affinity purification.

vidual RNAs which resemble natural pre-mRNAs. These studies showed that pre-mRNAs are bound with hnRNP proteins in a transcript-dependent manner. Specifically, we observed that different subsets of hnRNP proteins bind to different RNAs. In most cases, the differences were quantitative. For example, significant differences in the levels of hnRNP A, B, C, I, and L proteins were readily apparent in comparisons of H complexes assembled on three functional splicing substrates, AdML, α-tropomyosin, and β-globin pre-mRNAs. These differences in composition between H complexes assembled on different RNAs cannot be explained on the basis of RNA size because very different patterns of hnRNP proteins were observed on RNAs of the same size. Transcript-dependent binding was observed for all RNAs examined, regardless of the presence of splice sites. We conclude that the set of hnRNP proteins that bind to a particular RNA is determined by the sequence and/or structure of that RNA. These studies, together with previous work (45, 46), argue against the proposal that nascent RNA is simply packaged nonspecifically by hnRNP proteins into a repeating regular array of hnRNP particles, or ribonucleosomes, similar to the sequence-independent packaging of DNA in nucleosomes (10, 11).

We also observed significant qualitative differences in hnRNP binding to short RNAs (100 to 200 nt), presumably because smaller RNAs bind relatively fewer total hnRNP proteins. The most dramatic example of sequence specificity of hnRNP protein binding was observed for an RNA derived from the 3′ portion of the α-tropomyosin intron (B3P3 RNA). This 135-nt RNA, which contains a 65-nt pyrimidine-rich region (46% C, 49% U), efficiently binds hnRNP I but shows only very low or undetectable levels of binding to the other hnRNP proteins. Surprisingly, we discovered, initially by comparisons with two-dimensional gels and subsequently by comparing protein sequences (17a, 18, 35), that hnRNP I is the same as the previously characterized polypyrimidine tract-binding protein (PTB) (17, 18, 35). This protein was shown to be a spliceosome component that interacts with the pyrimidine tract at the 3′ splice site of AdML (17) and α-tropomyosin (35) pre-mRNAs.

Although hnRNP I, A, C and D proteins interact avidly and preferentially with RNA segments derived from the 3′ splice site, this sequence element is not the exclusive binding site for these proteins. For example, our studies indicate that hnRNP I also binds to the 5′ portion of AdML pre-mRNA. This portion of the RNA contains a pyrimidine-rich sequence (CCCCgTTCgTCCTCaCTCTCTTCC) consistent with the observed binding preference of hnRNP I/PTB (17, 35, this study). Similarly, our data show that hnRNP A, C, and D bind with sequence preference to different RNAs, whether or not they contain a 3′ splice site. Thus, we conclude that the 3′ splice site is only a subset of the specific binding sites in the RNA for hnRNP I, A, C, and D. Similarly, our data show that many of the other hnRNP

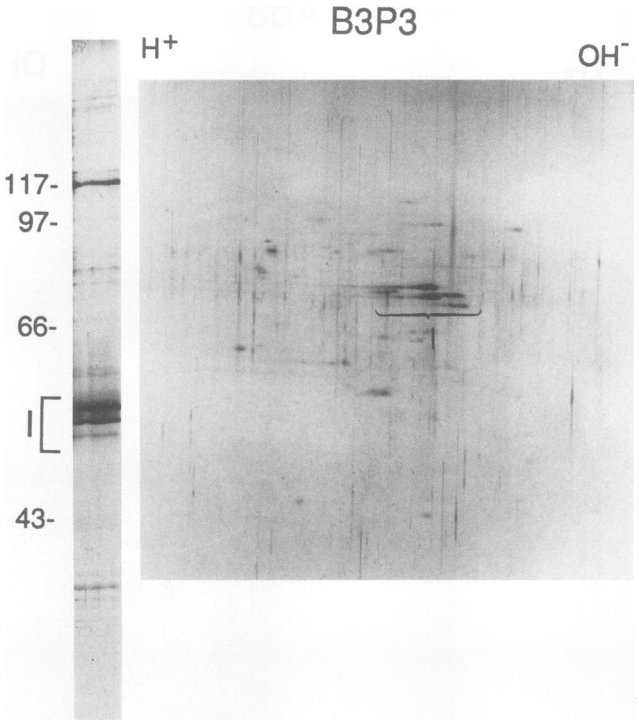


FIG. 7. hnRNP I/PTB is the predominant protein present in H complex assembled on the pyrimidine-rich RNA, B3P3. Affinity-purified H complex assembled on B3P3 RNA was fractionated on an SDS–9% polyacrylamide gel or by two-dimensional nonequilibrium gel electrophoresis as described in the legend to Fig. 5B. The designations of the hnRNP proteins are according to Piñol-Roma et al. (36).

proteins bind RNA with sequence preference. The binding sites may or may not include splicing signals, depending on the sequence preferences of the various hnRNP proteins.

Several essential splicing factors also interact in a sequence-dependent manner with the pre-mRNA. U1 snRNP interacts directly with the 5′ splice site (41, 43, 49) and either indirectly or directly with the 3′ splice site (16, 30a), while U2 snRNP and U2AF interact with the 3′ region of the intron (34, 47, 48, 50). In contrast to these splicing components, there is as yet no direct evidence that any of the hnRNP proteins are essential general splicing factors. However, hnRNP C and I (9, 18, 35, 38a) as well as many of the other hnRNP proteins (38a) are present in the spliceosome. This observation, coupled with the fact that the hnRNP proteins bind RNA in a sequence-dependent manner, raises the possibility that these proteins affect the splicing reaction. For example, hnRNP A, C, D, and I interact with the 3′ splice site, suggesting that these proteins could facilitate or inhibit binding of essential splicing factors.

It is also possible that hnRNP proteins play a role in splice site selection. In particular, the sequence-dependent binding of hnRNP proteins to pre-mRNA could be involved in the mechanism by which cryptic and normal splice sites are distinguished from one another and in the mechanism for determining splice site use in pre-mRNAs containing multiple alternatively or constituitively spliced introns. For example, the pre-mRNA may be bound by hnRNP proteins in such a way that functional and cryptic splice sites are either more or less accessible to splicing factors. Differential binding of hnRNP proteins to RNA could also affect the assembly of a functional spliceosome after the initial splicing factors have bound to the splice sites. Previous studies have shown that exon sequences determine whether a nearby splice site will be used efficiently (40). One possibility raised by our study is that the mechanism by which exon sequences affect splice site use is by the selective binding of hnRNP proteins. Thus, sequence-dependent binding of pre-mRNA by hnRNP proteins and/or direct effects of particular hnRNP proteins may play a role in the mechanism of splicing.

## ACKNOWLEDGMENTS

We are grateful to Joy Kingston and Keiko Kumatori for excellent technical assistance. We thank Bernardo Nadal-Ginard for gifts of B3P3 and GC+DX plasmids. HeLa cells for nuclear extracts were provided by the National Institutes of Health cell culture facilities at the Massachusetts Institute of Technology and at Endotronics (Minneapolis, Minn.).

R.R. is a Lucille P. Markey Scholar. This work was supported in part by a grant to R.R. from the Lucille P. Markey Charitable Trust and NIH and by the Howard Hughes Medical Institute and grants to G.D. from NIH.

## REFERENCES

1. **Abmayr, S. M., R. Reed, and T. Maniatis.** 1988. Identification of a functional mammalian spliceosome containing unspliced pre-mRNA. Proc. Natl. Acad. Sci. USA **85:**7216–7220.
1a. **Bennett, M.** Unpublished observations.
2. **Beyer, A. L., and Y. N. Osheim.** 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. **2:**754–765.
3. **Bindereif, A., and M. R. Green.** 1986. Ribonucleoprotein complex formation during pre-mRNA splicing in vitro. Mol. Cell. Biol. **6:**2511–2519.
4. **Bindereif, A., and M. R. Green.** 1987. An ordered pathway of snRNP binding during mammalian splicing complex assembly. EMBO J. **6:**2415–2424.
5. **Brody, E., and J. Abelson.** 1985. The "spliceosome": yeast premessenger RNA associates with a 40S complex in a splicing-dependent reaction. Science **228:**963–967.
6. **Buvoli, M., F. Cobianchi, G. Biamonti, and S. Riva.** 1990. Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. Nucleic Acids Res. **18:**6595–6600.
7. **Choi, Y. D., and G. Dreyfuss.** 1984. Isolation of the heterogeneous nuclear ribonucleoprotein complex (hnRNP): a unique supramolecular assembly. Proc. Natl. Acad. Sci. USA **81:**7471–7475.
8. **Choi, Y. D., and G. Dreyfuss.** 1984. Monoclonal antibody characterization of the C proteins of heterogeneous nuclear ribonucleoprotein complexes in vertebrate cells. J. Cell Biol. **99:**1997–2004.
9. **Choi, Y. D., P. J. Grabowski, P. A. Sharp, and G. Dreyfuss.** 1986. Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. Science **231:**1534–1539.
10. **Chung, S. Y., and J. Wooley.** 1986. Set of novel, conserved proteins fold premessenger RNA into ribonucleosomes. Proteins **1:**195–210.
11. **Conway, G., J. Wooley, T. Bibring, and W. M. LeStourgeon.** 1988. Ribonucleoproteins package 700 nucleotides of pre-mRNA into a repeating array of regular particles. Mol. Cell. Biol. **8:**2884–2895.
12. **Dreyfuss, G.** 1986. Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. Annu. Rev. Cell Biol. **2:**459–498.
13. **Dreyfuss, G., S. A. Adam, and Y. D. Choi.** 1984. Physical change in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. Mol. Cell. Biol. **4:**415–423.
14. **Dreyfuss, G., M. S. Swanson, and S. Piñol-Roma.** 1990. The composition, structure and organization of proteins in hetero-

geneous nuclear ribonucleoprotein complexes, p. 503–517. *In* P. Strauss and S. Wilson (ed.), The eukaryotic nucleus: structure and function, vol. 2. Telforf Press, Caldwell, N.J.
15. **Frendewey, D., and W. Keller.** 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell **42:**355–367.
16. **Fu, X.-D., and T. Maniatis.** 1992. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3′ splice site. Proc. Natl. Acad. Sci. USA **89:**1725–1729.
17. **García-Blanco, M. A., S. Jamison, and P. A. Sharp.** 1989. Identification and purification of a 62,000 dalton protein that binds specifically to the polypyrimidine tract of introns. Genes Dev. **3:**1874–1886.
17a. **Ghetti, A., et al.** Submitted for publication.
18. **Gil, A., P. A. Sharp, S. F. Jamison, and M. García-Blanco.** 1991. Characterization of cDNAs encoding the polypyrimidine tract-binding protein. Genes Dev. **5:**1224–1236.
19. **Grabowski, P. J., S. R. Seiler, and P. A. Sharp.** 1985. A multicompenent complex is involved in the splicing of messenger RNA precursors. Cell **42:**345–353.
20. **Grabowski, P. J., and P. A. Sharp.** 1986. Affinity chromatography of splicing complexes: U2, U5, and U4+U6 small nuclear ribonucleoprotein particles in the spliceosome. Science **233:** 1294–1299.
21. **Green, M. R.** 1991. Biochemical mechanisms of constituitive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. **7:**559–599.
22. **Hartley, S. E., and H. G. Callan.** 1978. RNA transcription on the giant lateral loops of the lampbrush chromosomes of the American newt Notophthalmus viridescens. J. Cell Sci. **34:**279–288.
23. **Konarska, M. M., and P. A. Sharp.** 1986. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. Cell **46:**845–855.
24. **Konarska, M. M., and P. A. Sharp.** 1987. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell **49:**763–774.
25. **Krainer, A. R., T. Maniatis, B. Ruskin, and M. R. Green.** 1984. Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell **36:**993–1005.
26. **Kramer, A.** 1988. Pre-splicing complex formation requires two proteins and U2 snRNP. Genes Dev. **2:**1155–1167.
27. **Kramer, A., and U. Utans.** 1991. Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. EMBO J. **10:**1503–1509.
28. **Mattaj, I. W.** 1990. Splicing stories and poly (A) tails: an update on RNA processing and transport. Curr. Opin. Cell Biol. **2:**528–538.
29. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:**7035–7056.
30. **Michaud, S., and R. Reed.** 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev. **5:**2534–2546.
30a. **Michaud, S., and R. Reed.** Unpublished data.
31. **O'Farrell, P. Z., H. M. Goodman, and P. H. O'Farrell.** 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell **12:**1133–1142.
32. **Osheim, Y. N., O. L. Miller, Jr., and A. L. Beyer.** 1985. RNP particles at splice junction sequences on *Drosophila* chorion transcripts. Cell **43:**143–151.
33. **Pandolfo, M., O. Valentini, G. Biamonti, P. Rossi, and S. Riva.** 1987. Large-scale purification of hnRNP proteins from Hela cells by affinity chromatography on ssDNA-cellulose. Eur. J. Biochem. **162:**213–220.
34. **Parker, R., P. G. Siliciano, and C. Guthrie.** 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base-pairing to the U2-like snRNA. Cell **49:**229–239.
35. **Patton, J. G., S. A. Mayer, P. Tempst, and B. Nadal-Ginard.** 1991. Characterization and molecular cloning of a polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. Genes Dev. **5:**1237–1251.
36. **Piñol-Roma, S., Y. D. Choi, M. J. Matunis, and G. Dreyfuss.** 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA binding proteins. Genes Dev. **2:**215–227.
37. **Piñol-Roma, S., M. Swanson, J. Gall, and G. Dreyfuss.** 1989. A novel heterogeneous nuclear ribonucleoprotein with a unique distribution. J. Cell Biol. **109:**2575–2587.
38. **Reed, R.** 1990. Analysis of the protein composition of mammalian spliceosomes assembled in vitro. Proc. Natl. Acad. Sci. USA **87:**8031–8035.
38a. **Reed, R., and M. Bennett.** Unpublished observations.
39. **Reed, R., J. Griffith, and T. Maniatis.** 1988. Purification and visualization of native spliceosomes. Cell **53:**949–961.
40. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell **46:**681–690.
41. **Seraphin, B., L. Kretzner, and M. Rosbash.** 1988. A U1 snRNA: pre-mRNA base-pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5′ cleavage site. EMBO J. **7:**2533–2538.
42. **Sierakowska, H., W. Szer, P. J. Furdon, and R. Kole.** 1986. Antibodies to hnRNP core proteins inhibit in vitro splicing of human β globin pre-mRNA. Nucleic Acids Res. **14:**5241–5254.
43. **Siliciano, P. G., and C. Guthrie.** 1988. 5′ splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes Dev. **2:**1258–1267.
44. **Smith, C. W. J., E. B. Porro, J. G. Patton, and B. Nadal-Ginard.** 1989. Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. Nature (London) **342:**243–247.
45. **Swanson, M. S., and G. Dreyfuss.** 1988. RNA binding specificity of hnRNP proteins: a subset bind to the 3′ end of introns. EMBO J. **11:**3519–3529.
46. **Swanson, M. S., and G. Dreyfuss.** 1988. Classification and purification of proteins of heterogeneous nuclear ribonculeoprotein particles by RNA-binding specificities. Mol. Cell. Biol. **8:**2237–2241.
47. **Wu, J., and J. L. Manley.** 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev. **3:**1553–1561.
48. **Zamore, P. D., and M. R. Green.** 1991. Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. EMBO J. **10:**207–214.
49. **Zhuang, Y., and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell **46:**827–835.
50. **Zhuang, Y., and A. M. Weiner.** 1989. A compensatory base change in U2 snRNA suppresses a branch site mutation. Genes Dev. **3:**1545–1552.