MOLECULAR AND CELLULAR BIOLOGY, May 1997, p. 2774–2780 Vol. 17, No. 5
0270-7306/97/$04.00+0

Pyrimidine Tracts between the 5′ Splice Site and Branch Point Facilitate Splicing and Recognition of a Small *Drosophila* Intron

CATHARINE F. KENNEDY AND SUSAN M. BERGET*

Program in Cell and Molecular Biology and Verna and Marrs McLean Department of Biochemistry, Houston, Texas 77030

Received 26 August 1996/Returned for modification 24 October 1996/Accepted 10 February 1997

The minimum size for splicing of a vertebrate intron is approximately 70 nucleotides. In *Drosophila melanogaster*, more than half of the introns are significantly below this minimum yet function well. Such short introns often lack the pyrimidine tract located between the branch point and 3′ splice site common to metazoan introns. To investigate if small introns contain special sequences that facilitate their recognition, the sequences and factors required for the splicing of a 59-nucleotide intron from the *D. melanogaster mle* gene have been examined. This intron contains only a minimal region of interrupted pyrimidines downstream of the branch point. Instead, two longer, uninterrupted C-rich tracts are located between the 5′ splice site and branch point. Both of these sequences are required for maximal in vivo and in vitro splicing. The upstream sequences are also required for maximal binding of factors to the 5′ splice site, cross-linking of U2AF to precursor RNA, and assembly of the active spliceosome, suggesting that sequences upstream of the branch point influence events at both ends of the small *mle* intron. Thus, a very short intron lacking a classical pyrimidine tract between the branch point and 3′ splice site requires accessory pyrimidine sequences in the short region between the 5′ splice site and branch point.

Intron/exon architecture varies considerably across the eucaryotic kingdom (8). In vertebrates, introns are larger than exons, with an average size over 1 kb. In lower eucaryotes, introns are often smaller than exons. In fact, lower eucaryotes often have introns smaller than the vertebrate minimum. *Drosophila melanogaster* contains a mixture of intron sizes, with both very small introns characteristic of lower eucaryotes and larger introns similar to those found in vertebrates. A number of small introns in *Drosophila* cluster in the 51- to 80-nucleotide (nt) range, with a median of 79 nt (12), a length below the vertebrate limit. It is unclear how such small *Drosophila* introns function in splicing given the fact that identified splicing factors in *Drosophila* are similar in size to their vertebrate counterparts.

One of the prominent vertebrate splicing consensus sequences is the pyrimidine tract located between the branch point and 3′ splice site (for a recent review of splicing, see reference 11). This sequence is the binding site for the required splicing factor U2 snRNP auxiliary splicing factor, U2AF (22). In vertebrates, a functional pyrimidine tract has been defined as at least five consecutive uridines or nine consecutive pyrimidines (15), although weaker tracts do exist in known constitutively spliced genes. Large *Drosophila* introns often possess similar pyrimidine tracts. It has been noted, however, that short *Drosophila* introns statistically lack such pyrimidine tracts (12), and in this characteristic they resemble introns from *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (13). Specifically, 51% of large (81- to 5,392-nt) *Drosophila* introns have a pyrimidine tract downstream of the branch point consisting of at least eight consecutive pyrimidines, while 72% have a pyrimidine-rich region between the branch point and 3′ splice site with at least 10 of 12 nt being pyrimidines. In contrast, only 25% of small (51- to 80-nt) *Drosophila* introns have a stretch of eight uninterrupted pyrimidines downstream of the branch point, and only 51% are pyrimidine rich (12). This difference suggests the existence of two distinct classes of introns in *D. melanogaster*: large, vertebrate-like introns that possess a 3′ pyrimidine tract, and small, yeast-like introns that lack this consensus sequence. These two classes of introns can be functionally distinguished by the inability of the small introns to tolerate internal expansion (7, 19).

To investigate the possibility that small introns contain special sequences that facilitate their recognition and splicing, the 59-nt first intron from the *D. melanogaster mle* gene has been studied. The *mle* intron is spliced in *Drosophila* splicing extract but is neither spliced nor assembled into active spliceosomes in mammalian extract, reflective of its small size (19). This intron lacks a significant pyrimidine tract between the branch point and the 3′ splice site. Instead, it contains two C-rich sequences between the 5′ splice site and branch point. In the *mle* intron, the distance between the 5′ splice site and any potential branch point is quite short, between 40 and 43 nt, close to the theoretical minimum (12). Here we show that two C-rich tracts from +19 to +38 within the intron just upstream of the first possible branch point are required for efficient in vitro and in vivo splicing of the *mle* intron. Interestingly, the upstream pyrimidine tracts are not necessary when the *mle* intron is converted into a more classical intron via the introduction of a pyrimidine tract between the branch point and 3′ splice site. In fact, the C-rich sequences and a classical 3′ pyrimidine tract appear to be antagonistic. In the wild-type *mle* intron, the upstream C-rich tract sequences are necessary for maximal binding of U2AF to pre-mRNA, interaction of proteins with the 5′ splice site, and proper assembly of the active spliceosome, suggesting that these sequences affect events at both ends of the intron. Because of the small size of the *mle* intron,

* Corresponding author. Mailing address: Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. Phone: (713) 798-5758. Fax: (713) 795-5487. E-mail: sberget@bcm.tmc.edu.

the C-rich tracts are located close to both the 5′ splice site and the branch point. We suggest that the small *mle* intron is recognized as a unit during spliceosome assembly and that sequences between the 5′ splice site and branch point have a function in intron recognition similar to the role of exon sequences in exon recognition.

MATERIALS AND METHODS

Plasmids. *mle* minigenes for in vivo and in vitro splicing contained the 59-nt first intron of the *D. melanogaster mle* gene, along with 56 nt of the first exon and 96 nt of the second exon. Mutant pre-mRNAs were identical to the wild-type *mle* pre-mRNA except for the introduced mutations indicated in Fig. 1 and 2. Mutants were constructed by using mutagenic PCR. The sequences of all plasmids were verified by using a Sequenase sequencing kit (United States Biochemical Corp.).

In vitro splicing and assembly. ^{32}P-labeled precursor mRNA splicing substrates were synthesized by using SP6 polymerase in an in vitro transcription reaction as previously described (1). Reaction mixtures consisting of 6.5 fmol of radiolabeled substrate, 50% *D. melanogaster* Schneider 2 (S2) nuclear extract, 1.6 mM $MgCl_2$, 20 mM phospho-L-arginine, 1.2 mM dithiothreitol (DTT), 1.2% polyethylene glycol, and 2 mM ATP were incubated at 22°C. Aliquots were taken at the indicated time points. RNA products and intermediates were analyzed on 5% denaturing polyacrylamide gels. Spliceosome complexes were analyzed by the addition of heparin to a final concentration of 2 mg/ml and electrophoresis on native RNP gels (19). Splicing reactions using a U2AF-specific or polypyrimidine tract binding protein (PTB)-specific competitor oligonucleotide, as determined by reiterative selection (5′-UUUUCCCUUUUUUUUC-3′ or 5′-GCCU GCUGCUCCUCUUCUGUC-3′, respectively [16]), were added to the reaction mixture at time zero.

Quantification of reactions. Polyacrylamide gels of in vitro splicing reactions and 5′ splice site protection studies were quantified by scanning in a Betagen Betascope 603 blot analyzer. Splicing was expressed as the percentage of initial precursor appearing as product (wild type normalized to 100%). For protection experiments, total counts in the band resulting from protection were calculated as a percentage of total counts.

UV cross-linking. In vitro splicing reaction mixtures were incubated at 22°C for 7 min before the addition of heparin to a final concentration of 2 mg/ml. Mixtures were transferred to ice and UV irradiated immediately for 10 min. Reactions were subsequently digested with RNase A at 37°C for 30 min. Labeled proteins were displayed on sodium dodecyl sulfate–9% polyacrylamide gels and visualized by autoradiography. Immunoprecipitation of cross-linked proteins was performed with an antibody against the 50-kDa subunit of *Drosophila* U2AF (dU2AF^{50}), provided by D. Rio (9). Competition of cross-linking used unlabeled competitor RNAs, either wild-type or mutant *mle* splicing substrates, added at time zero. Competitors were added at 300, 500, 700, or 900 times the amount of radiolabeled wild-type *mle* pre-RNA.

5′ splice site protection. To detect protein interactions at the *mle* 5′ splice site, an oligonucleotide complementary to this region was added in excess along with RNase H after splicing complex assembly had been allowed to proceed for the indicated amount of time. Antisense inhibition of U1 snRNA or U3 snRNA used reaction mixtures pretreated at 30°C for 10 min with oligonucleotides complementary to nt 1 to 14 of U1 snRNA or nt 1 to 11 of U3 snRNA, along with RNase H.

Depletion and purification of U2AF. Depletion of S2 nuclear splicing extract for U2AF was achieved by a modification of the protocol for depletion of HeLa extracts (23). S2 nuclear extract was dialyzed against buffer A (20 mM HEPES [pH 7.9], 1 M KCl, 3 mM $MgCl_2$, 0.05% Nonidet P-40, 1 mM DTT, 20% glycerol) for 2 h. Dialyzed extract was applied to a 3-ml poly(U)-Sepharose column (Pharmacia) equilibrated in the same binding buffer. Depleted extract was concentrated twofold by dialysis against dry polyethylene glycol compound for 1 h at 4°C. Concentrated extract was dialyzed twice against 500 ml of buffer D (20 mM HEPES [pH 7.9], 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol) for 2 h. For mock-depleted extract, chromatography on plain Sepharose beads was used. U2AF was purified from 10 ml of HeLa nuclear extract as described by Cote et al. (4).

Immunoblotting. Transfer and detection procedures were performed as described in the PolyScreen instruction manual (Dupont). A polyclonal antibody raised against the large subunit of U2AF was provided by J. G. Patton (13a); arginine/serine (SR) protein-specific monoclonal antibodies 16H3 and 104 were provided by M. Roth (21); and PTB-associated splicing factor (PSF)-specific antibody was provided by J. G. Patton (13a).

RESULTS

Pyrimidine tracts upstream of the branch point are required for *mle* splicing. Examination of the 59-nt first intron of the *mle* gene revealed that the longest uninterrupted pyrimidine stretch in this short intron resided between the 5′ splice site and any potential branch point. Two C-rich pyrimidine sequences (CUCCCCC and UUUCCC) are located between +19 and +38 of the intron (Fig. 1A), upstream of any potential branch point (the minimum distance observed for *Drosophila* between the 5′ splice site and branch point is 38 nt [12]). To determine if these nucleotides are functionally involved in precursor mRNA processing, we constructed RNA substrates that contained mutations within either pyrimidine tract 1 (CUCC CCC→CUGGCCC), pyrimidine tract 2 (UUUCCC→UUAG CC), or both pyrimidine tracts 1 and 2 (Fig. 1A). Care was taken not to create any obvious splicing signals or to significantly change the A-U content of the intron, which has been shown to be important for splicing in *D. melanogaster* (12). These constructs were tested for the ability to be spliced in Schneider S2 nuclear extract. Both mutations lowered splicing efficiency compared to *mle* wild type (Fig. 1B). Also, mutation of both sequences was more inhibitory than mutation of either sequence alone, suggesting that the sequences act additively. Mutation of either pyrimidine stretch alone reduced splicing efficiency to 67 to 70% of the wild-type level, while disrupting both sequences further inhibited splicing to 47% of wild-type levels. These results suggest that the pyrimidine tracts upstream of the branch point are necessary for maximal splicing efficiency of the *mle* intron.

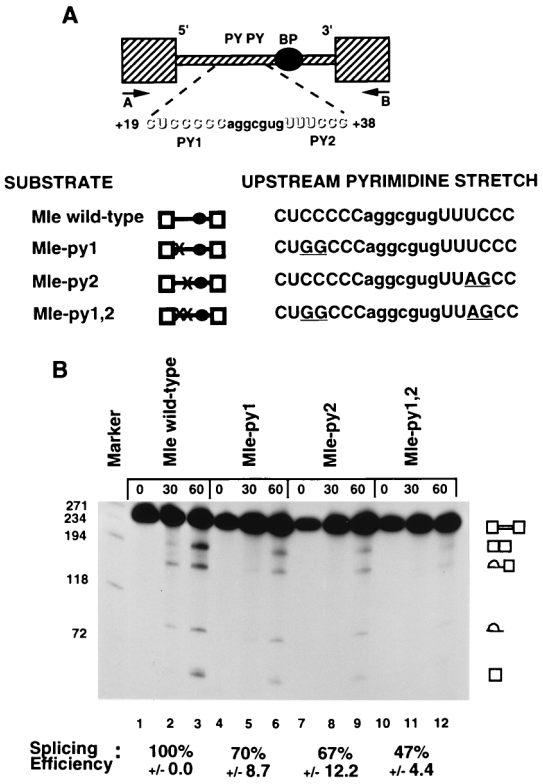

FIG. 1. Two C-rich pyrimidine tracts (PY) located between the 5′ splice site and the branch point (BP) are required for maximal in vitro *mle* splicing. (A) Sequences of wild-type and mutant *mle* introns. The two pyrimidine tracts located between the 5′ splice site and branch point in the wild-type *mle* intron are indicated. Introduced mutations are underlined. (B) In vitro splicing of wild-type and mutant *mle* precursor mRNAs. Radiolabeled RNA substrates, indicated above the gel, were spliced in *Drosophila* S2 nuclear extract. Reaction aliquots were taken at the indicated times (minutes) for display of RNA products on a 5% denaturing polyacrylamide gel. Reaction substrates, products, and intermediates are as previously identified (19) and are indicated. The efficiency of splicing, expressed as a percentage of the wild-type value, is indicated at the bottom. Sizes are indicated in nucleotides.

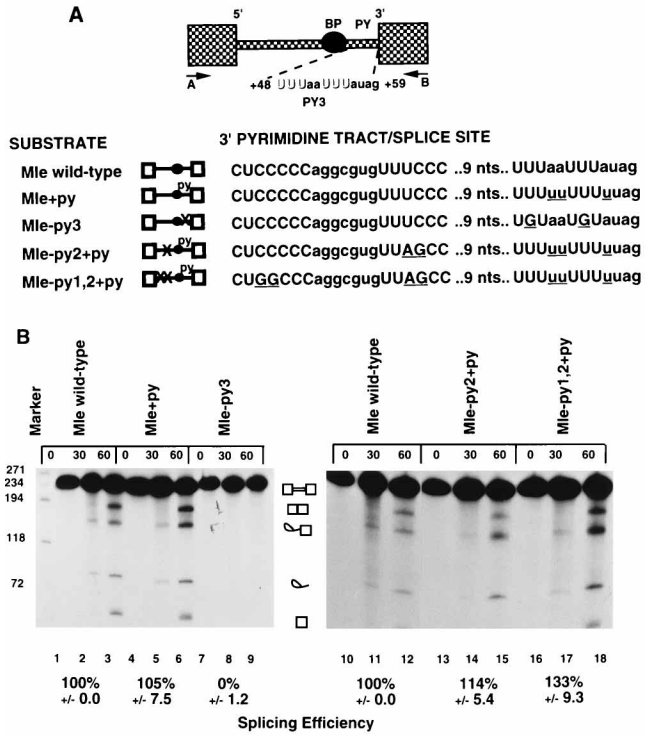

FIG. 2. The pyrimidine-poor region (PY) between the branch point (BP) and 3′ splice site is required for *mle* splicing. (A) Sequence of the *mle* intron highlighting the pyrimidine-poor 3′ splice site (capital letters). Constructed mutants (introduced mutations are underlined) are shown under the wild-type sequence. Mutants *mle* −py2+py and *mle* −py1,2+py contain the mutations py2 and py1,2 diagrammed in Fig. 1 combined with the +py mutation. (B) Single-intron *mle* splicing substrates possessing either wild-type or mutant pyrimidine tracts were assayed for in vitro splicing. Conditions were identical to those described in Fig. 1. Sizes are indicated in nucleotides.

A short, pyrimidine-containing region downstream of the branch point also functions in *mle* splicing. In addition to the two pyrimidine tracts in the first half of the *mle* intron, the region between the branch point and the 3′ splice site contains two UUU sequences separated by purines (Fig. 2A). To ascertain the role of the downstream uridines in splicing of the *mle* intron, the region between the branch point and 3′ splice site was mutated to either create a consensus pyrimidine tract (*mle* +py; UUUAAUUUAUAG→UUUUUUUUUUAG) or interrupt the few pyrimidines present (*mle* −py3; UUUAAUU UAUAG→UGUAAUGUAUAG) (Fig. 2A). In vitro splicing assays were conducted to test the effects of these changes on splicing efficiency (Fig. 2B, left). With *mle* wild-type splicing normalized to 100%, the addition of a pyrimidine tract to create *mle* +py was found to not significantly change the level of splicing. Mutation of the 3′ pyrimidine tract in *mle* −py3, on the other hand, inhibited splicing. Other mutations of this pyrimidine tract (UUUAAUUUAUAG changed to UAUAA UGUAUAG or UUUAAUAUAUAG) also inhibited processing (data not shown). Thus, the presence of the sequence UUU between the branch point and the 3′ splice site appears to be required for in vitro splicing of the *mle* intron.

To see if the *mle* upstream pyrimidine tracts would be necessary in the presence of a downstream pyrimidine tract, the upstream tracts were mutated in the +py variant *mle* intron containing an uninterrupted pyrimidine tract between the branch point and 3′ splice site (the *mle* +py construct diagrammed in Fig. 2A). As shown in Fig. 2B (right), mutation of one or both upstream pyrimidine tracts did not inhibit splicing of the variant containing the downstream improved pyrimidine tract. In fact, if anything, mutation of the upstream tracts increased splicing (1.3-fold improvement) in this background. Thus, the upstream pyrimidine tracts are necessary for maximal splicing of a wild-type *mle* intron lacking a classical pyrimidine tract between the branch point and 3′ splice site but are inhibitory for splicing of the *mle* intron when such pyrimidine tracts are added to the RNA.

The upstream pyrimidine tracts are required for binding of U2AF. To assess the requirement for the upstream pyrimidine tracts in binding of splicing factors to the *mle* intron, in vitro UV cross-linking experiments were performed with wild-type and mutant *mle* precursor RNAs (Fig. 3). Cross-linking of some proteins decreased when the upstream or downstream pyrimidine tracts were mutated. Of particular interest was a protein of approximately 53 kDa (p53). This factor was cross-linked to the wild-type *mle* intron (Fig. 3, lane 1). Mutation of

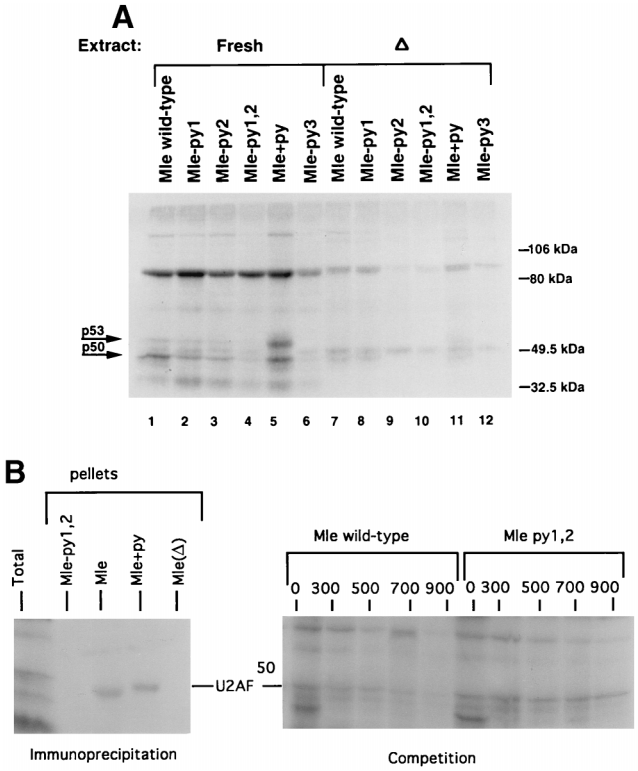

FIG. 3. Pyrimidine tracts both upstream and downstream of the branch point are required for maximal UV cross-linking of U2AF to the *mle* intron. (A) The indicated radiolabeled precursor RNAs (mutants are defined in Fig. 1 and 2) were incubated for 7 min in either complete S2 extract (Fresh) or extract depleted for U2AF by chromatography on poly(U)-Sepharose (Δ) and then subjected to UV cross-linking. Following digestion with RNase A, cross-linked proteins were visualized on a sodium dodecyl sulfate–9% polyacrylamide gel. The bands corresponding to proteins of interest (53 kDa [p53] and 50 kDa [p50]) are indicated. The protein marked p50 had characteristics of a U2 snRNP protein in that its cross-linking was ATP, U2 RNA, and U2AF dependent (data not shown) and is tentatively identified as spliceosome-associated protein SAP49 or SF3b[53] (2, 3, 18). Markers are as shown. (B) Identification of p53 as U2AF and competition of U2AF cross-linking to *mle* precursor RNA by wild-type but not mutant competitor RNAs. UV-cross-linking reactions with wild-type *mle* or mutant precursor RNAs were immunoprecipitated from normal or U2AF-depleted (Δ) extract with anti-dU2AF[50] antibodies (9) (left). Cross-linking of U2AF to wild-type *mle* precursor RNA was monitored in the presence of increasing amounts of wild-type *mle* or mutant *mle* py1,2 RNA (containing mutant upstream pyrimidine tracts) (right). The amount of competitor used is indicated as fold increase over substrate.

upstream pyrimidine tract 1 or pyrimidine tract 2 reduced the level of p53 cross-linking (Fig. 3; compare lanes 2 and 3 with lane 1). The binding of p53 to the mle intron was severely depressed when both of the upstream sequences were disrupted (Fig. 3, lane 4). In contrast, the level of p53 cross-linking was greatly increased when a pyrimidine tract was added between the branch point and 3′ splice site in the *mle* +py mutant (Fig. 3, lane 5). Elimination of pyrimidines downstream of the branch point eliminated cross-linking (Fig. 3, lane 6). Thus, it appears that there are three short pyrimidine tracts present in the 59 nt *mle* intron that are required for both splicing and binding of p53; two are located upstream of the branch point, and one is located downstream of the branch point.

The molecular weight of p53, its dependence on pyrimidine tracts for association with precursor RNA, and other binding characteristics of this protein (ATP-independent binding [data not shown]) suggest that it might be dU2AF^{50}. U2AF-specific antibodies directed against dU2AF^{50} (9) were used to immunoprecipitate UV-cross-linked proteins from a splicing reaction of *mle* (Fig. 3B, left). A cross-linked protein of 53 kDa, identical in molecular mass to the p53 identified in Fig. 3A, was immunoprecipitated with the antibody. Furthermore, immunoprecipitation depleted the reaction of cross-linked p53 (data not shown). Therefore, p53 is dU2AF^{50}. Immunoprecipitation was also used to confirm that the upstream pyrimidine tracts were necessary for U2AF binding. No detectable band was immunoprecipitated when *mle* py1,2 was used as the precursor RNA for UV cross-linking.

Mutation of the upstream pyrimidine tracts could alter the ability to detect U2AF binding by UV cross-linking either because a binding site for a protein(s) had been destroyed or because of loss of uridines necessary for chemical cross-linking of a protein(s) whose binding was unaffected by the mutations. To differentiate these possibilities, we performed competition experiments in which UV cross-linking to wild-type *mle* RNA was monitored in the presence of increasing amounts of wild-type or mutant *mle* competitor RNAs (Fig. 3B, right). UV cross-linking of U2AF to wild-type *mle* precursor RNA was effectively competed by wild-type *mle* competitor RNA but not by equivalent concentrations of a competitor RNA containing the py1,2 mutations. Therefore, the upstream pyrimidine tracts bind a *trans*-acting factor(s) necessary for maximal U2AF binding.

Splicing of the *mle* intron is dependent on dU2AF^{50}. A role for pyrimidine tracts upstream of the branch point for the binding of U2AF has not been noticed for other precursor RNAs. Because the *mle* wild-type construct did not contain the defined binding site for U2AF downstream of the branch point, it was possible that the association of dU2AF^{50} with precursor RNA was nonspecific and of no functional relevance. To test the requirement for U2AF in *mle* splicing, S2 nuclear extract was biochemically depleted of this splicing factor via chromatography on poly(U)-Sepharose. Chromatography was performed in 1 M KCl-phosphate-based buffer to minimize nonspecific binding to the column (essentially as described in reference 23). Flowthrough fractions from the column were pooled and dialyzed against extract buffer (see Materials and Methods) for use as an U2AF-depleted splicing extract.

Figure 4A shows that dU2AF^{50} was specifically depleted from S2 extract by this procedure, whereas the levels of other identifiable splicing factors that could potentially bind to the poly(U) column were unaffected. Western blot analysis of flowthrough fractions indicated essentially complete removal of dU2AF^{50} (compare lanes 1 and 2). Other important splicing

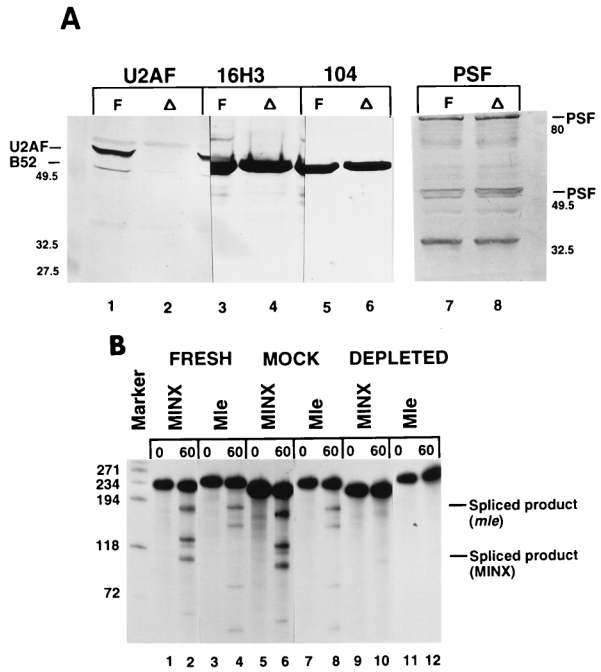

FIG. 4. dU2AF^{50} is required for splicing and assembly of the *mle* intron. (A) dU2AF^{50} was specifically depleted from S2 nuclear extract by chromatography on poly(U)-Sepharose (see Materials and Methods). Shown is a Western blot of fresh extract (F) or poly(U)-depleted extract (Δ) probed with antibodies specific for dU2AF^{50}, S/R proteins (monoclonal antibodies 16H3 and 104), or PSF. Sizes are indicated in kilodaltons. (B) In vitro splicing assays of dU2AF^{50}-depleted extract. Radiolabeled adenovirus (MINX) and *mle* precursor RNAs were tested for the ability to be spliced in fresh (FRESH), mock-depleted (chromatography on plain Sepharose; MOCK), or poly(U)-depleted (DEPLETED) S2 extracts. Aliquots were taken at 0 and 60 min. Splicing of the MINX intron has been shown to be U2AF dependent (23) and thus was included as a control. The positions of fully spliced product RNAs are shown. Sizes are indicated in nucleotides.

factors, including various members of the SR family of proteins, such as B52 (compare lanes 3 and 4 and lanes 5 and 6), as well as PSF (compare lanes 7 and 8) were present in the depleted extract. Levels of PTB in the depleted extract could not be assessed due to the lack of cross-reactivity of available antibodies, which were raised against human PTB. Results in our laboratory using HeLa nuclear extract, however, demonstrated that only human U2AF^{65} (hU2AF^{65}), not PSF or PTB, binds to a poly(U) column in 1 M KCl (data not shown). Therefore, the depleted extract contains SR proteins, PSF, and PTB but is severely depleted for dU2AF^{50}.

UV cross-linking was used as a more sensitive determinant for the absence of U2AF in the depleted extract. Labeled *mle* splicing substrates were incubated with the dU2AF^{50}-depleted S2 extract. A band of 53 kDa corresponding to U2AF was missing from the cross-linking of wild-type and mutant precursor RNAs in depleted extract (Fig. 3A, lanes 7 to 12). Furthermore, no cross-linked U2AF could be immunoprecipitated from depleted extract with the anti-dU2AF^{50} antibody (Fig. 3B). By these criteria, the depleted extract was judged to be effectively depleted of U2AF.

The existence of a requirement for U2AF in *mle* splicing was assessed by splicing assays using the depleted extract (Fig. 4B). Splicing of an adenovirus-based substrate RNA (MINX [14]) was used as a control because splicing of this substrate has been demonstrated to be U2AF dependent (23). Mock depletion inhibited the splicing of neither MINX nor *mle* wild-type

precursor RNA (Fig. 4B, lanes 5 to 8). In fact, splicing levels and phenotypes were equivalent in fresh and mock-depleted extracts. In contrast, processing of both precursors was completely inhibited in the dU2AF^{50}-depleted extract (Fig. 4B, lanes 9 to 12). This result implied that the *mle* intron not only binds U2AF but also requires this factor for splicing.

To confirm the requirement for U2AF in *mle* splicing, purified U2AF was added back to depleted extracts to reconstitute various activities. Mammalian U2AF was purified from HeLa nuclear extract (4). The purified U2AF used for this experiment is shown in the Coomassie blue-stained gel in Fig. 5A (left). Experiments in other laboratories have previously demonstrated that the large subunit of U2AF from *Drosophila* will complement HeLa extract depleted of human U2AF (9). As shown in Fig. 5B (right), addition of HeLa U2AF to *Drosophila* extract depleted of U2AF resulted in UV cross-linking of a 65-kDa protein to the *mle* precursor RNA that is the same molecular weight as the purified hU2AF^{65} subunit detected by Western blotting. Furthermore, cross-linking to the complementing protein paralleled cross-linking to *Drosophila* U2AF. Mutation of the upstream pyrimidine tracts depressed the ability of complementing hU2AF^{65} to cross-link to precursor RNA, and improving the downstream pyrimidine tract increased cross-linking. Taken together, the preceding results strongly argue that both the upstream and downstream pyrimidine tracts in the *mle* intron play a role in binding U2AF to the *mle* intron.

Analysis of spliceosome assembly in depleted extracts indicated a functional requirement for U2AF in *mle* splicing. Neither an *mle* substrate nor an adenovirus-based substrate was capable of assembly into complex A in extract depleted of U2AF (Fig. 5B). Addition of purified human U2AF restored assembly of complex A to both substrates to approximately the same level. Therefore, U2AF is required for assembly of the *mle* intron.

To examine the U2AF dependence of the *mle* intron in a different way, the effect of hU2AF^{65}-specific competitor RNAs on *mle* precursor mRNA splicing was examined (Fig. 5C). As a control, a similar but functionally different U-rich sequence specific for PTB binding was used as another competitor RNA. The hU2AF^{65}-specific and PTB-specific oligonucleotides were optimized sequences selected by the respective proteins via reiterative selection (16). In vitro splicing assays were carried out as usual except that various amounts of competitor RNA were included in the reaction mixtures and single time points were taken at 60 min. The hU2AF^{65}-specific oligonucleotide began to inhibit *mle* wild-type splicing at 2.5 pmol and abolished processing at 25 pmol. The PTB-specific oligonucleotide, on the other hand, had negligible effects at any tested concentration. This experiment supports the finding that the *mle* intron, containing three short pyrimidine stretches, is dependent on dU2AF^{50} for splicing.

Correct spliceosome assembly of the *mle* intron is dependent on the upstream pyrimidine tracts. The preceding results suggested that the upstream pyrimidine tracts were necessary for both splicing and UV cross-linking of U2AF to the *mle* intron. U2AF is normally required quite early in the spliceosome assembly pathway. Depletion of U2AF or mutation of the 3′ pyrimidine tract normally eliminates assembly of the first ATP-dependent spliceosome complex, complex A. To assess the requirement for the upstream pyrimidine tracts on assembly of the *mle* intron, splicing reactions of wild-type and mutant precursor RNAs were examined by native gel electrophoresis (Fig. 6A). Wild-type *mle* precursor RNA assembles multiple ATP-dependent complexes (19), similar to the assembly patterns of vertebrate precursor RNAs and more standard *Drosophila* introns containing consensus pyrimidine tracts, such as the single intron from the *furishi taramazu* gene (17). Mutation of the upstream pyrimidine tracts of the *mle* intron permitted assembly of complex A but arrested assembly at that point such that assembly of higher-order complexes did not occur. This result suggests that the upstream pyrimidine tracts are necessary either for formation of active complex A or for its progression through the assembly pathway to the active spliceosome, i.e., that the complex formed in the absence of wild-type upstream pyrimidine tracts is a dead-end complex unable to complete spliceosome assembly. A role for the upstream tracts in early assembly is consistent with a role in maximizing U2AF binding.

The upstream pyrimidine tracts are necessary for maximal binding of factors to the 5′ splice site. Because of the small size of the *mle* intron, the upstream pyrimidine tracts are as close to the 5′ splice site as they are to the branch point. To see if these sequences are required for assembly of factors on the 5′ splice site, protection studies were performed with wild-type and mutant *mle* precursor RNAs (Fig. 6B, top). Binding of factors to the 5′ splice site was assessed by measuring the susceptibility of precursor RNA to cleavage in the presence of oligonucleotides complementary to the 5′ splice site in the presence of RNase H. Wild-type RNA demonstrated protection of the 5′ splice site within 5 min of incubation in S2 nuclear extract. In contrast, *mle* substrate RNAs in which the upstream pyrimidine tracts had been mutated supported protection to only 30% of wild-type levels, indicating that the upstream pyrimidine tracts are necessary for maximal binding of factors to the 5′ splice site. Furthermore, this protection was dependent on the presence of U1 snRNA (Fig. 6B, bottom). Pretreating the extract with an oligonucleotide complementary to nt 1 to 14 of U1 snRNA and RNase H eliminated the appearance of the protection band. Preincubating the extract with an oligonucleotide that hybridizes to nt 1 to 11 of U3 snRNA, on the other hand, had no effect. Thus the upstream pyrimidine tracts are necessary for maximal interactions at both ends of the small *mle* intron.

DISCUSSION

The general process of splicing is highly conserved among eucaryotes. Many *cis*-acting sequences and *trans*-acting factors are shared from humans to yeasts (for reviews, see references 11 and 10). Such conservation of factors is somewhat surprising given the fact that the exon/intron architecture of split genes varies significantly across the eucaryotic kingdom. The large introns of vertebrate genes stand in marked contrast to the small introns present in *S. pombe*. Both classes of introns are found in *D. melanogaster* (8, 12). Large *Drosophila* introns can often be spliced in HeLa nuclear extract, whereas small introns cannot be (7). This restriction agrees with the observation that internal deletion of vertebrate introns to less than 70 nt inhibits their in vitro and in vivo splicing (5, 20). The expansion of some, but not all, small *Drosophila* introns alleviates repression of their splicing in HeLa extract (7). Such expansion sometimes results in the inability of expanded introns to be spliced in *Drosophila* systems (7). These observations raise several interesting questions. One problem centers on how small introns are recognized in *Drosophila*, since most factors required for vertebrate splicing are found in *Drosophila* and are of approximately the same molecular weight (7). To study the mechanisms whereby small introns are recognized and subsequently spliced, the first intron from the *mle* gene was examined.

This pre-mRNA splicing substrate was chosen because it is a small intron lacking a consensus uninterrupted pyrimidine

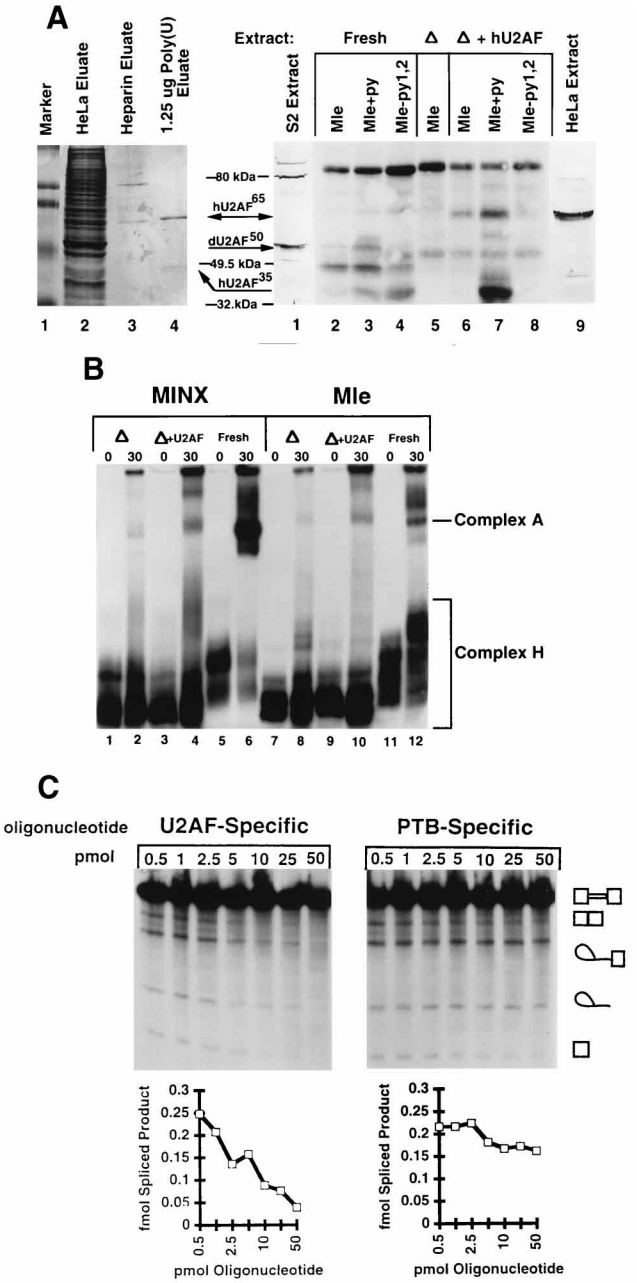

FIG. 5. Addition of U2AF restores UV cross-linking and assembly of the *mle* intron. (A) A Coomassie blue-stained gel of the purified human U2AF fraction used is shown at the left. The positions of the 65- and 35-kDa subunits of human U2AF are shown. UV cross-linking of wild-type and mutant *mle* precursor mRNAs in fresh and dU2AF^{50}-depleted S2 extracts is shown at the right. Radiolabeled substrates (as diagrammed) were UV cross-linked in fresh extract (Fresh), depleted extract (Δ), or depleted extract complemented with 0.5 μg of purified human U2AF (Δ + hU2AF). The positions of dU2AF^{50} and hU2AF^{65} are indicated on flanking lanes, which are Western blots of S2 and HeLa nuclear extracts, respectively. (B) In vitro assembly reaction. Assembly reactions were performed with normal extract (Fresh), extract depleted of dU2AF^{50} (Δ), and extract depleted of U2AF to which 0.5 μg of purified human U2AF had been added (Δ+U2AF). The position of spliceosome complex A is indicated. (C) Effects of competitor RNAs specific for either U2AF or PTB on *mle* precursor mRNA splicing. In vitro splicing assays using wild-type *mle* were performed in the presence of one of two competitor RNAs (5′-UUUUCCCUUUUUUUUC-3′, containing a sequence among those selected by human U2AF via reiterative selection [left] or 5′-GCCUGCUGCUCC UCUUCUGUC-3′, containing a sequence among those selected by PTB via reiterative selection [right]) (16) (top). Amounts of oligonucleotides used are shown above the lanes. All samples were taken at a single 60-min time point. Splicing products and intermediates are indicated. Quantification of competition of splicing with U2AF-specific ribo-oligonucleotides is shown at the bottom.

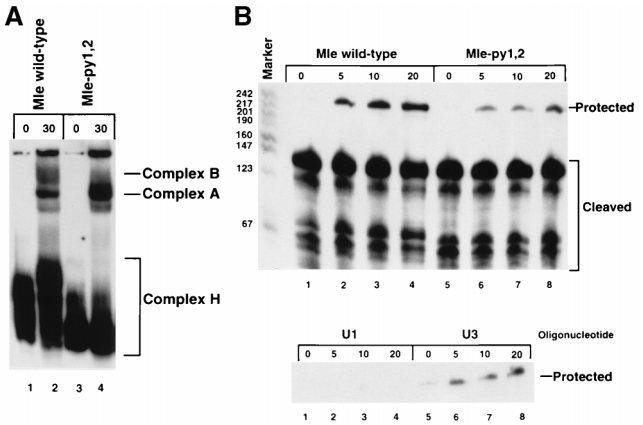

FIG. 6. Pyrimidine tracts between the 5′ splice site and the branch point are required for active spliceosome assembly. (A) Spliceosome assembly reactions using wild-type or mutant *mle* precursor RNAs were performed as previously described (19). ATP-dependent complexes A and B are indicated. (B) Protection of the 5′ splice site in wild-type or mutant *mle* precursor RNAs. To detect binding at the 5′ splice site, an oligonucleotide complementary to the 5′ splice site was added during the splicing reaction along with RNase H (top). Time points indicate time (minutes) of addition of the complementary oligonucleotide. Cleavage and protection products are indicated. A similar protection experiment was performed with extract in which U1 RNA had been precleaved by RNase H and a complementary oligonucleotide (bottom).

tract between the branch point and the 3′ splice site and also because it fails to function in *Drosophila* if internally expanded (19). Examination of the sequence of the *mle* intron revealed that it contained two uninterrupted C-rich pyrimidine stretches residing between +19 and +38 downstream of the 5′ splice site and upstream of any putative branch point occurring beyond the minimal acceptable distance from the 5′ splice site of 38 nt (22). Mutational analysis demonstrated that these upstream pyrimidine tracts were required for efficient splicing of this small intron, both in vitro in nuclear extracts from *Drosophila* tissue culture cells and in transfected S2 cells (data not shown). The pyrimidines between the 5′ splice site and branch point were also required for maximal U2AF binding to the RNA as measured by UV cross-linking. The contribution to splicing of the minimal U sequences between the branch point and 3′ splice site was also studied. While improvement of this tract to fit the defined sequence did not significantly boost splicing levels in vitro, it did increase the level of U2AF cross-linking, suggesting that the ability to bind U2AF is not the limiting factor in the splicing of this intron. Disruption of the minimal 3′ pyrimidine-containing region also inhibited processing as well as U2AF cross-linking. Interestingly, the upstream pyrimidine tracts were not necessary when the downstream pyrimidine tract was strengthened. In fact, an improvement in splicing efficiency was observed when the upstream sequences were mutated in an improved 3′ pyrimidine tract background. This incompatibility could suggest the presence of two competing mechanisms for assembly of *Drosophila* introns. Taken together, these data indicate that multiple pyrimidine-containing regions flanking the branch point are required for *mle* splicing and binding of U2AF.

The upstream pyrimidine tracts were also required for spliceosome assembly on the *mle* intron. In their absence, factors did not bind normally to the 5′ splice site and the complex A formed did not assemble further to form complex B, the active spliceosome. A requirement for pyrimidine sequences upstream of the branch point for either U2AF binding or early spliceosome assembly has not been observed in mammalian systems, although various U2 snRNP proteins have been observed to UV

cross-link to pre-mRNA in this region. In a recent study (6), sequences upstream of the branch point were referred to as sequence-independent anchoring sequences for U2 snRNPs because addition of an RNA oligonucleotide complementary to the upstream region to an in vitro mammalian splicing reaction inhibited complex A assembly and UV cross-linking of U2 snRNP proteins. In contrast to our results, mutation of the upstream sequences of the adenovirus-derived precursor RNA used for this previous study had no effect on splicing activity or spliceosome assembly, suggesting that there was no sequence requirement in this region. Even mutations that introduced considerable purine content to the upstream sequences had no effect. In the *mle* intron, even addition of a few purines into this region affected processing activity. In agreement with our results, a study of another small *Drosophila* intron found that sequences proximal to the branch point were important for maximal activity (7). Thus, small *Drosophila* introns reveal a requirement for sequences between the 5′ splice site and branch point not obvious for vertebrate introns.

Other differences between requirements for mammalian introns and the *mle* intron were revealed in this study. Mutation of pyrimidine tracts upstream of the *mle* branch point also affected binding of factors to the 5′ splice site as well as binding of factors to the 3′ splice site. Effects of sequences between the 5′ splice site and branch point on 5′ splice site recognition have not been seen before in mammalian systems. Interestingly, the sequences also affected spliceosome assembly, although not in a manner that might have been predicted. As revealed by native gel electrophoresis, mutation of the upstream tracts depressed conversion of the initial ATP-dependent complex to the active spliceosome. This result suggests that the upstream tracts are necessary either for the correct formation of an active complex A or for the addition of the U4-U5-U6 tri-snRNP to the spliceosome and the corresponding conformational changes required during this addition.

The fact that U2AF binding to the *mle* pre-mRNA appears to be dependent on pyrimidine tracts between the 5′ splice site and branch point suggests a mechanistic variation between introns in the nature of their interaction with U2AF. One possible model predicts that the function of the two upstream pyrimidine tracts in the *mle* intron is to facilitate U2AF binding to the weak pyrimidine tract near the 3′ splice site. In this view, the upstream sequences in *mle* act as accessory U2AF binding sites to stabilize an otherwise weak interaction between U2AF and the minimal U-containing sequences downstream of the branch point. The pyrimidine tracts upstream of the branch point could either directly bind U2AF or interact with it via another factor.

An attractive alternative model is that the upstream tracts are part of the mechanism whereby small introns bypass the minimal length restriction operating on larger introns with consensus pyrimidine tracts between the branch point and 3′ splice site. Binding of proteins to the upstream tracts could alter interactions of factors with either the 5′ or 3′ splice site so as to create sufficient space in the intron for the binding of other splicing factors.

It should be noted that in introns as small as the *mle* intron, sequences upstream of the branch point are necessarily proximal to the 5′ splice site. In fact, the 5′-most upstream pyrimidine tract in the *mle* intron begins at intron nt +19, just 13 nt downstream of the last nucleotide of the 5′ splice site. Thus, the upstream sequences could be considered as part of either the 5′ or 3′ splice site. Given the inability of small *Drosophila* introns to function when expanded (7, 19), our results suggest that small introns function as a unit. From this perspective, interior intronic sequences play a role in splice site recognition in small introns much like interior exon sequences play a role in pre-mRNAs recognized via exon definition.

ACKNOWLEDGMENTS

We thank M. Green, D. Rio, and J. Patton for generous gifts of antibodies.

C.F.K. is a graduate fellow of the Robert A. Welch Foundation. This work was supported by PHS grant RO1 GM38526 and the Robert A. Welch Foundation.

REFERENCES

1. **Berget, S. M., and B. L. Robberson.** 1986. U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for *in vitro* splicing but not polyadenylation. Cell **46:**691–696.
2. **Brosi, R., K. Groning, S.-E. Behrens, R. Luhrmann, and A. Kramer.** 1993. Interaction of mammalian splicing factor SF3a with U2 snRNP and relationship of its 60-kD subunit to yeast PRP9. Science **262:**102–105.
3. **Champion-Arnaud, P., and R. Reed.** 1994. The prespliceosome components SAP 49 and SAP 145 interact in a complex implicating in tethering U2 snRNP to the branch site. Genes Dev. **8:**1974–1983.
4. **Cote, J., J. Beaudoin, R. Tacke, and B. Chabot.** 1995. The U1 small nuclear ribonucleoprotein/5′ splice site interaction affects $U2AF^{65}$ binding to the downstream 3′ splice site. J. Biol. Chem. **270:**4031–4036.
5. **Fu, X.-Y., J. D. Colgan, and J. L. Manley.** 1988. Multiple *cis*-acting sequence elements are required for efficient splicing of simian virus 40 small-t antigen pre-mRNA. Mol. Cell. Biol. **8:**3582–3590.
6. **Gozani, O., R. Feld, and R. Reed.** 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev. **10:**233–243.
7. **Guo, M., P. C. H. Lo, and S. M. Mount.** 1993. Species-specific signals for the splicing of a short *Drosophila* intron in vitro. Mol. Cell. Biol. **13:**1104–1118.
8. **Hawkins, J. D.** 1988. A survey on intron and exon lengths. Nucleic Acids Res. **16:**9893–9908.
9. **Kanaar, R., S. E. Roche, E. L. Beall, M. R. Green, and D. C. Rio.** 1993. The conserved pre-mRNA splicing factor U2AF from *Drosophila*: requirement for viability. Science **262:**569–572.
10. **Kramer, A.** 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. Annu. Rev. Biochem. **65:**367–409.
11. **Moore, M. J., C. C. Query, and P. A. Sharp.** 1993. Splicing of precursors to messenger RNAs by the spliceosome, p. 303–357. *In* R. F. Gesterland and J. F. Atkins (ed.), The RNA world. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
12. **Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White, and C. Fields.** 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Nucleic Acids Res. **20:**4255–4262.
13. **Parker, R., and B. Patterson.** 1987. Architecture of fungal introns: implications for spliceosome assembly, p. 451. *In* M. Inouye and B. S. Dudock (ed.), Molecular biology of RNA: new perspectives. Academic Press, San Diego, Calif.

13a. **Patton, J. G.** Personal communication.

14. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10:**84–94.
15. **Roscigno, R., M. Weiner, and M. A. Garcia-Blanco.** 1993. A mutational analysis of the polypyrimidine tract of introns. J. Biol. Chem. **268:**11222–11229.
16. **Singh, R., J. Valcárcel, and M. R. Green.** 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine-tract binding proteins. Science **268:**1173–1176.
17. **Spikes, D., and P. M. Bingham.** 1992. Analysis of spliceosome assembly and the structure of a regulated intron in *Drosophila in vitro* splicing extracts. Nucleic Acids Res. **20:**5719–5727.
18. **Staknis, D., and R. Reed.** 1994. Direct interactions between the pre-mRNA and six U2 snRNP proteins during spliceosome assembly. Mol. Cell. Biol. **14:**2994–3005.
19. **Talerico, M., and S. M. Berget.** 1994. Intron definition in splicing of small *Drosophila* introns. Mol. Cell. Biol. **14:**3434–3445.
20. **Wieringa, B., E. Hofer, and C. Weissmann.** 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β-globin intron. Cell **37:**915–925.
21. **Zahler, A. M., W. S. Lane, J. A. Stolk, and M. B. Roth.** 1992. SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev. **6:**837–847.
22. **Zamore, P. D., and M. R. Green.** 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc. Natl. Acad. Sci. USA **86:**9243–9247.
23. **Zamore, P. D., and M. R. Green.** 1991. Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intramolecular distribution. EMBO J. **10:**207–214.