Combinatorial Splicing of Exon Pairs by Two-Site Binding of U1 Small Nuclear Ribonucleoprotein Particle

PAULA J. GRABOWSKI,†* FAIZ-UL H. NASIM,‡ HAI-CHIEN KUO,† AND ROBERT BURCH

Section of Biochemistry, Brown University, Providence, Rhode Island 02912

Received 26 March 1991/Accepted 12 September 1991

A two-site model for the binding of U1 small nuclear ribonucleoprotein particle (U1 snRNP) was tested in order to understand how exon partners are selected in complex pre-mRNAs containing alternative exons. In this model, it is proposed that two U1 snRNPs define a functional unit of splicing by base pairing to the 3′ boundary of the downstream exon as well as the 5′ boundary of the intron to be spliced. Three-exon substrates contained the alternatively spliced exon 4 (E4) region of the preprotachykinin gene. Combined 5′ splice site mutations at neighboring exons demonstrate that weakened binding of U1 snRNP at the downstream site and improved U1 snRNP binding at the upstream site result in the failure to rescue splicing of the intron between the mutations. These results indicate the stringency of the requirement for binding a second U1 snRNP to the downstream 5′ splice site for these substrates as opposed to an alternative model in which a certain threshold level of U1 snRNP can be provided at either site. Further support for the two-site model is provided by single-site mutations in the 5′ splice site of the third exon, E5, that weaken base complementarity to U1 RNA. These mutations block E5 branchpoint formation and, surprisingly, generate novel branchpoints that are specified chiefly by their proximity to a cryptic 5′ splice site located at the 3′ terminus of the pre-mRNA. The experiments shown here demonstrate a true stimulation of 3′ splice site activity by the downstream binding of U1 snRNP and suggest a possible mechanism by which combinatorial patterns of exon selection are achieved for alternatively spliced pre-mRNAs.

This work was undertaken because the biochemical mechanisms that specify alternative splicing of complex pre-mRNAs are poorly understood (7, 18, 33). In the basic splicing process, selection of 5′ and 3′ splice sites at the boundaries of an intron occurs by assembly of a spliceosome, the RNA-protein complex in which splicing takes place (8, 10, 12). This process involves interactions of the pre-mRNA with the small nuclear ribonucleoprotein particles (snRNPs) U1, U2, and U5 and the particle containing both U4 and U6 RNAs (4, 13, 16, 34). Recognition of the 5′ splice site occurs by base pairing with the 5′ end of U1 RNA (21, 29, 31, 38). Recognition of the 3′ splice site occurs in an ATP-dependent step by base pairing of U2 RNA to the branchpoint region, normally located 17 to 40 nucleotides upstream of the 3′ splice site (5, 23, 24, 28). Subsequent to U2 snRNP binding, the complex containing U4, U5, and U6 RNAs associates as an integral component of the spliceosome.

Sequence elements known to play a role in specifying the selection of alternatively spliced exons have been shown to be associated with 3′ splice sites. Unusually positioned branchpoints accompany the mutually exclusive splicing of exons 2 and 3 of rat α-tropomyosin (32), exons 6A and 6B of chicken β-tropomyosin (11), and exons 6 and 7 of rat β-tropomyosin (14). In the case of α-tropomyosin, the close proximity of the branchpoint of exon 3 to exon 2 precludes splicing together of these two exons. In the case of chicken β-tropomyosin, the secondary structure of the pre-mRNA surrounding exon 6B masks its use in most cell types (20). Furthermore, tissue-specific blocks to splicing have been revealed by 3′ splice site mutations in the production of alternatively spliced mRNAs from the calcitonin/calcitonin gene-related peptide gene (9).

Recent work has pointed to the importance of 5′ splice sites in specifying exon selection. Both engineered and naturally occurring 5′ splice site mutations that decrease base complementarity to U1 RNA result in skipping of the exon adjacent to the mutated site (1, 6, 35–37). These results lend support to the exon definition model, in which internal exons are selected as a result of coordinate recognition of the 5′ and 3′ splice sites of the same exon (25).

Evidence that U1 snRNP directs exon selection was obtained by using the alternative fourth exon (E4) of the rat preprotachykinin gene (17). In this case, improved binding of U1 snRNP to the E4 5′ splice site was shown to drive branchpoint formation at the 3′ splice site of the same exon. These results suggest that U1 snRNP binding modulates the activity of the E4 3′ splice site in accordance with the general features of the exon definition model. A direct prediction of these results is that for a pre-mRNA containing three or more exons, the selection of any pair of exons will, in general, depend on the binding of U1 snRNP to two distinct sites: to the 5′ splice site of the downstream exon as well as to the conventional 5′ splice site of the intron to be spliced.

In this work, we have tested the two-site model for U1 snRNP binding in order to understand how combinatorial patterns of exon selection are achieved in pre-mRNAs containing alternatively spliced exons. The experimental strategy used was to generate combined 5′ splice site mutations adjacent to neighboring exons such that the first mutation weakened U1 snRNP binding to the downstream site whereas the second mutation improved U1 snRNP binding to the upstream site. In this way, the stringency of the requirement for binding U1 snRNP to the downstream 5′ splice site could be assessed in comparison with an alternative model in which a certain threshold level of U1 snRNP is needed in the vicinity of the intron to be spliced. Since the

* Corresponding author.
† Present address: Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260.
first mutation at the downstream 5' splice site would be expected to block splicing of the upstream intron, this assay would determine whether splicing could be rescued by restoring the total level of U1 snRNP binding at the upstream site.

MATERIALS AND METHODS

Construction of plasmids. For plasmids RP93 and RP94, DNA fragments A and B were synthesized from wild-type plasmid RP23, using the polymerase chain reaction (Fig. 1C). Fragment A, 1,090 nucleotides, was synthesized by using oligodeoxynucleotides (oligos) 5'-GGACACAAAAAGCT TGCATGCTCT and 3'-GAGAGATCTGACCAGCC. Fragment B, 146 nucleotides, was synthesized by using oligos (5') containing the desired mutations (underlined) (RP93, AAAAAACAAAAATCTCAACAATCATTTTG; RP94, AAAAAAGTAAAGCTCAAAATCATTTTG) and 3'-AGGCGAATT CGAGCGGATGACC. A three-part ligation reaction joined the Bluescript vector (Stratagene) to the HindIII site of DNA fragment A and the EcoRI site of fragment B; fragments A and B were joined by blunt-end ligation. Substrates RP123 and RP124 were constructed from parent plasmids RP93 and RP94, respectively, by replacing a central 822-bp BstEII fragment with the analogous fragment from plasmid RP57, which contains a mutation in the E4 5' splice site that improves base pairing to U1 RNA (Table 1). Substrates RP93km (killer mutation) and RP93um (up mutation) were constructed from parent plasmid RP93 by replacing a 27-bp XbaI-EcoRI restriction fragment with a synthetic duplex containing the desired mutation (see Fig. 3A). Plasmid RP125 was constructed by a strategy similar to that dia-
TABLE 1. Summary of 5' splice site sequences

<table>
<thead>
<tr>
<th>Substrates</th>
<th>Sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E3:IVS3</td>
</tr>
<tr>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>RP23 (wild type)</td>
<td>CUG:GUGAAGUG</td>
</tr>
<tr>
<td>RP91</td>
<td>AUG:GUAAACAU</td>
</tr>
<tr>
<td>IVS4 test</td>
<td></td>
</tr>
<tr>
<td>RP94</td>
<td>CUG:GUGAAGUG</td>
</tr>
<tr>
<td>RP93</td>
<td>AUG:GUAAACAU</td>
</tr>
<tr>
<td>RP124</td>
<td>CUG:GUGAAGUG</td>
</tr>
<tr>
<td>RP123</td>
<td></td>
</tr>
<tr>
<td>IVS3 test</td>
<td></td>
</tr>
<tr>
<td>RP25</td>
<td>CUG:GUAAAGAU</td>
</tr>
<tr>
<td>RP57</td>
<td>CUG:GUGAAGUG</td>
</tr>
<tr>
<td>RP126</td>
<td>CUG:GUAAAGAU</td>
</tr>
</tbody>
</table>

* Shown below each sequence are Watson-Crick or G-C base pair interactions with U1 RNA (●); sequence of the 5' end of U1 RNA available for base-pairing interactions 3'-GUCAUCAUA-5'); mutations (overlines); 5' splice site (colons). Blank spaces indicate identity with sequence immediately above.

grammed in Fig. 1C, using an oligo containing the E3 5' splice site mutation, ATTGCTACTTATCCCGATCC CG-5', an abutting oligo, AGTACTAGGCTCATGCT ACC-3', and oligos overlapping the HindIII and EcoRI sites, with plasmid RP23 as the template. Plasmid RP126 was constructed by using plasmid RP57 as the template. Correct clones were confirmed by DNA sequence analysis of all splice site regions.

Synthesis of pre-mRNA and in vitro splicing. Pre-mRNA substrates for in vitro splicing reactions were synthesized, using T3 RNA polymerase (Stratagene), from corresponding plasmids linearized with EcoRI unless otherwise indicated. Standard splicing conditions were 44% (vol/vol) HeLa nuclear extract, 2.0 mM MgCl₂, 1.5 mM ATP, 5 mM creatine phosphate, and 1 to 10 μg of substrate RNA per ml in a total volume of 25 μl (22).

Branchpoint analysis. Substrates for splicing containing unlabeled nucleotides. RNA isolated from total (25 μl) splicing reactions was analyzed by primer extension. A 5'-end-labeled DNA primer was hybridized 30 to 40 nucleotides downstream of the predicted branchpoint region and extended with reverse transcriptase and deoxynucleotides. Hybridization was carried out at 65°C for 10 min. Standard extension conditions were 42°C for 20 min in a mixture containing 86 mM NaCl, 14 mM Tris-HCl (pH 8.1), 1 mM MgCl₂, 72 μg of actinomycin D per ml, 13 mM dithiothreitol, 1.5 mM deoxynucleoside triphosphates, and 5 μl of reverse transcriptase (Bio-Rad). Product cDNA strands were resolved on 10% polyacrylamide–7 M urea sequencing gels. Debranching reactions were carried out as described previously (27), using RNA from total splicing reactions.

Affinity selection. Substrate RNAs were prepared by transcription in the presence of biotinylated UTP (13). One-hour splicing reaction mixtures (50 μl) were adjusted to binding buffer (0.5 M KCl, 3.0 mM MgCl₂, 34 mM N,N'-ethenylsulfonic acid [HEPES; pH 7.5], 0.14 mM EDTA, 9.4% [vol/vol] glycerol, 0.14 mg each of glycogen and bovine serum albumin per ml, in a final volume of 100 μl). Samples were centrifuged briefly (12,000 rpm, 5 min, 4°C) and bound to streptavidin columns for 2 h at 4°C. After four 1 ml washes in binding buffer, selected components were eluted with 1% sodium dodecyl sulfate (90°C, 5 min). Released snRNAs were resolved on a 10% polyacrylamide–7 M urea gel and electropholetive onto a GeneScreen membrane. Northern (RNA) blot hybridization was performed with 5 × 10⁶ cpm each of U1, U2, U4, U5, and U6 DNA probes.

RESULTS

Splicing substrates used in this study are variants of the three-exon substrate containing E3, intervening sequence 3 (IVS3), E4, IVS4, E5, and a portion of IVS5 of the rat preprotachykinin gene, in which E4 is an alternatively spliced exon. All substrates are identical in length and sequence to the original gene segment except for a nucleotide change(s) in the 5' splice site adjacent to E3, E4, or E5. The wild-type gene segment characteristically splices by an E4-skipping mechanism, both in vitro and in vivo (22). To test the two-site model for U1 snRNP binding, single-site mutations were first introduced into the 5' splice site of E5, resulting in reduced base complementarity to U1 RNA (substrates RP94 and RP93; Table 1).

Analysis of splicing reactions for the mutant substrates demonstrated that both splicing events involving E5 (alternative and constitutive splicing) were blocked, as indicated by the absence of the characteristic lariat product RNAs, IVS4 and an RNA containing IVS3, E4, and IVS4 (RP93 and RP94; Fig. 1A and B). Control splicing reactions containing the previously characterized RP23 and RP91 substrates serve as markers for the lariat products and intermediates assayed in these experiments (17, 22). Primer extension analysis, a sensitive measure of E5 selection, confirmed that formation of the E5 branchpoint was blocked for each of the mutant substrates (RP93 [Fig. 2A] and RP94 [Fig. 4B]). These single-site mutants were subsequently used to construct substrates with combined 5' splice site mutations flanking IVS4 (see below).

Novel branchpoints activated by 5' splice site mutations. Although the single-site mutations blocked E5 selection,
aberrant splicing was indicated by the formation of slowly migrating, ATP-dependent products (substrates RP93 and RP94; Fig. 1A and B, filled arrowheads). Concomitant production of E3 (data not shown) suggested that aberrant splicing might involve joining E3 to cryptic 3' splice sites in IVS5, across a distance of approximately 1,000 nucleotides. To address the possibility that a 5' splice site mutation was responsible for directing 3' splice site formation to novel sites, we set out to map the location of the putative branchpoints formed for these substrates. Having predicted that the putative branchpoints were located in IVS5, splicing reactions were subjected to primer extension analysis using a radiolabeled primer positioned near the 3' end of IVS5 (primer 428; Fig. 2B). Three novel branchpoints were indeed

![Diagram](http://mcb.asm.org/)

**FIG. 2.** Analysis of E5 and novel branchpoints. (A) Primer extension analysis carried out with E5-specific primer 426 (lanes 1 to 21) and IVS5-specific primer 428 (lanes 22 to 28). Splicing reactions were incubated for 0 (lanes 2, 9, 16, and 23), 60 (lanes 3, 10, 17, and 24), 120 (lanes 4, 11, 18, and 25), and 180 (lanes 5, 12, 19, and 26) min. Control reactions were carried out for 180 min without ATP (lanes 6, 13, 20, and 27) or with ATP followed by debranching (lanes 7, 14, 21, and 28). Sequencing ladders of corresponding plasmid DNAs, containing ddTTP, show the positions of adenosine residues (lanes 1, 8, 15, and 22). Arrows indicate branchpoints (see below). A schematic of substrates and pattern of splicing is shown at the bottom. Shown are positions of E5 5' splice site (ss) mutations (asterisk) and primers (filled boxes). (B) Sequence of E5 (bold type) and adjacent IVS4 and IVS5. Sequence at the E5 5' splice site (ss) is that of substrate RP93. Primers specific for E5 (426) and IVS5 (428) are indicated. Shown are the E5 branchpoint (BP), cryptic branchpoints (a, b, c, d, and e), poly pyrimidine tract (Py), splice sites (colons); nonspecific plasmid sequences (NS); XbaI restriction site (Xba I), and 3' end of substrate RNA (EcoRI [Eco R1]). The cryptic 5' splice site and base complementarity with nucleotides 6 to 11 of U1 RNA are also shown (filled circles).

<table>
<thead>
<tr>
<th>Branchpoint sequence&lt;sup&gt;*&lt;/sup&gt;</th>
<th>Approx relative use</th>
<th>Match to consensus</th>
<th>Distance (nt) to AG Cryptic 5' splice site</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) UGCUAU</td>
<td>1×</td>
<td>7/7</td>
<td>44 75</td>
</tr>
<tr>
<td>(b) UGUUAU</td>
<td>3×</td>
<td>7/7</td>
<td>38 69</td>
</tr>
<tr>
<td>(c) GAUGUG</td>
<td>2×</td>
<td>3/7</td>
<td>31 62</td>
</tr>
<tr>
<td>(d) UAGUGA</td>
<td>20×</td>
<td>5/7</td>
<td>29 60</td>
</tr>
<tr>
<td>(e) UGAUAU</td>
<td>5×</td>
<td>4/7</td>
<td>26 57</td>
</tr>
</tbody>
</table>

<sup>*</sup> A, branchpoint.
found in IVS5, at positions 88 (b), 97 (d), and 100 (e) nucleotides downstream of the normal E5 branchpoint (Fig. 2A, lanes 25 and 26).

Inspection of the IVS5 sequence revealed that the novel branchpoints were situated in sequences that matched the mammalian branchpoint consensus sequence to various degrees (Table 2) followed by an imperfect polypurine/polypyrimidine stretch and AG dinucleotide (Fig. 2B). Thirty-four nucleotides downstream from the AG dinucleotide, a 5′ splice site-like sequence (cryptic 5′ splice site) was also observed with which nucleotides 6 to 11 of U1 RNA can be aligned (Fig. 2B, filled circles). This is the normal alignment, in which nucleotide C8 of U1 RNA is positioned opposite G+1 of the intron.

Interestingly, formation of the novel branchpoints was abolished for both the RP93 and RP94 substrates when the XbaI runoff site was used to truncate the cryptic 5′ splice site (Fig. 2B; data not shown). Thus, it appeared likely that as a consequence of the E5 5′ splice site mutation, the (downstream) cryptic 5′ splice site was in fact the sequence element responsible for directing formation of the novel branchpoints in IVS5. To rigorously test this idea, second-site mutations were created in the cryptic 5′ splice site that improved (RP93um) or destroyed (RP93km) base complementarity to U1 RNA (Fig. 3A, bottom). These results demonstrated that formation of the novel branchpoints is abolished when base complementarity to U1 RNA is destroyed (RP93km; Fig. 3A). In contrast, when base complementarity was increased, five novel branchpoints were observed, three of which corresponded to those previously mapped for the single-site mutant, substrate RP93 (RP93um in Fig. 3A; Table 2). Furthermore, analysis of total splicing reactions for these substrates demonstrated the involvement of E3 in the aberrant splicing reaction as well as formation of the corresponding spliced product RNA (Fig. 3B).

**Substrates that test the model: 5′ splice site mutations flanking IVS4.** Starting with substrates containing a mutation in the downstream 5′ splice site (5′ splice site of E5) that reduced base complementarity to U1 RNA, second-site mutations were introduced in the upstream 5′ splice site (E4 5′ splice site) in order to restore the total level of U1 snRNP binding (RP123 and RP124; Table 1). The splicing efficiency of the intron between the mutations, IVS4, was then assayed. If the bindings of U1 snRNP to the upstream and downstream 5′ splice sites are independent requirements for splicing, then the second-site mutations would not be ex-
expected to rescue IVS4 splicing activity. Primer extension analysis of splicing reactions containing the two-site mutants demonstrated that the second-site mutation does not rescue selection of E5, as indicated by the absence of the E5 branchpoint (Fig. 4B, RP123 and RP124). Control reactions demonstrated that the E4 branchpoint was, as expected from previous work (17), greatly improved for substrate RP123 compared with RP93 as a consequence of the second-site mutation (Fig. 4A). Furthermore, the similarity in levels of the E4 branchpoint and the E3E4 spliced product RNA for substrates RP57 and RP123 demonstrates that the distal mutation in the 5′ splice site of E5, in the latter substrate, does not influence IVS3 splicing. Analysis of total splicing reactions containing substrates RP123 and RP124 confirmed that IVS3 but not IVS4 was produced (Fig. 4C, lanes 7 to 12).

**Mutations flanking IVS3.** The two-site model was tested again by mutating the 5′ splice sites flanking IVS3 (RP125, RP57, and RP126; Table 1). Starting with the wild-type substrate, RP23, which contains a naturally weak, downstream 5′ splice site (E4 5′ splice site), a single mutation was introduced to increase the total level of U1 snRNP binding to the upstream 5′ splice site (E3 5′ splice site) to create substrate RP125. As a control, we used the previously characterized substrate RP57, which contains the identical mutation in the downstream 5′ splice site (E4 5′ splice site). This mutation substantially improves U1 snRNP binding to the E4 5′ splice site (17). An additional substrate, RP126, which contains the same mutation at both 5′ splice sites flanking IVS3, was made.

(i) Three-exon substrates. Splicing reactions were performed to assess the effect of these mutations on the test intron, IVS3, as well as the downstream intron, IVS4, and the intron produced by alternative splicing, which contains IVS3, E4, and IVS4. When only the upstream 5′ splice site was improved, no IVS3 splicing was observed, and the profile of RNA products generated was identical to that observed for the wild-type substrate, RP23 (RP125; Fig. 5A); that is, by itself, the mutation in the upstream 5′ splice site had no effect on splicing. In contrast, when the downstream 5′ splice site was improved singly or in combination with the upstream site, a strong increase in IVS3 and IVS4 splicing...
resulted, and there was no detectable level of alternative splicing (RP57 and RP126; Fig. 5A). IVS3 splicing was driven further to completion for substrate RP126, as judged by a decrease in the level of the intermediate RNA containing IVS3 in a lariat form joined to E4, IVS4, E5, and IVS5 (Fig. 5A, lanes 10 to 12).

(ii) Two-exon substrates. IVS3 splicing was similarly tested by using substrates containing E3, IVS3, E4, and 74 nucleotides of IVS4. The mutation at the upstream site (E3 5' splice site) gave only a trace of IVS3 splicing (Fig. 5B, lanes 5 to 8), the same level observed for the analogous wild-type

substrate, RP23 Bgl II (22). Thus, the mutation at the upstream site, alone, has no effect on IVS3 splicing, in accord with the results obtained with three-exon substrates (Fig. 5A). In contrast, IVS3 splicing was strongly enhanced by the downstream mutation (E4 5' splice site), and the combination of the two mutations in substrate RP126 drove IVS3 splicing further to completion (Fig. 5B; compare lanes 1 to 4 and 9 to 12).

U1 snRNP binding and spliceosome assembly. As an important test of the proposed model, the IVS3 test substrates were used to verify the predicted binding of U1 snRNP. For this purpose, biotinylated substrate RNAs were incubated under splicing conditions and then subjected to affinity selection on immobilized streptavidin at high salt (0.5 M KCl). Under these conditions, U1 snRNP binding is retained at 5' splice sites that have a nine- but not six-nucleotide base complementarity to U1 RNA (17). These results show that binding of U1 snRNP to either 5' splice site flanking IVS3 is approximately equivalent, which would be expected for two sites having the identical mutation (U1 RNA; Fig. 6A and B, lanes 4, and 5). Furthermore, when both 5' splice sites are improved, binding is approximately doubled, and this binding is ATP independent (Fig. 6A and B, lanes 6 and 7). Background levels of U1 snRNP binding are indicated by

FIG. 5. Single and combined mutations flanking IVS3. Shown is analysis of total splicing reactions resolved on 4% (A) or 7% (B) polyacrylamide-7 M urea gels. The substrate and time of incubation under splicing conditions are indicated at the top. (A) Three-exon substrates; (B) two-exon substrates. Two-exon substrates contained E3, IVS3, E4, and 74 nucleotides of IVS4 and were produced by transcription with corresponding plasmids cleaved at the BglII site.

FIG. 6. Northern blot analysis of U1 snRNP binding and spliceosome assembly. Splicing reactions (60 min) were performed as for Fig. 5 with biotinylated substrate RNAs followed by affinity selection on streptavidin columns. (A) Three-exon substrates; (B) two-exon substrates. It is important to note that in panel B, control reactions shown in lanes 2 and 3 represent the normal background level of the snRNAs. Reactions lacking ATP reproducibly show higher background levels of these RNAs (lanes 6 and 7). Lanes M, total HeLa nuclear RNA as markers for snRNAs. Control substrate RP58 is identical to substrate RP23 except for the sequence of the E4 5' splice site, AUG:CAAAUCAU (overlines indicate mutations; see Table 1).
control reactions lacking substrate RNA (Fig. 6A and B, lanes 2).

Spliceosome formation can also be assessed in the affinity selection experiments by noting the level of U2, U4, U5, and U6 RNAs compared with levels in control reactions lacking substrate RNA (Fig. 6A and B, lanes 2). For two-exon substrates, spliceosome formation is evident for substrates RP57 and RP126 Bgl II, which contain the improved downstream 5’ splice site, and which show strong IVS3 splicing (Fig. 6B, lanes 4 and 6; Fig. 7B). These results are in contrast to a marginal level of spliceosome formation for the substrate containing the improved upstream site in combination with the weak downstream site (RP125 Bgl II; Fig. 6B, lane 5). In the latter case, marginal spliceosome assembly is congruous with the marginal splicing activity characteristic of substrate RP125 Bgl II (Fig. 5B). One substrate, RP58 Bgl II, is devoid of splicing activity, as previously characterized (17), and serves as an additional negative control for spliceosome formation (Fig. 6B, lane 3).

Strong spliceosome assembly occurs for the three-exon substrates, consistent with their strong splicing efficiency (Fig. 5A and 6A). This is due to the presence of at least one downstream 5’ splice site with a six-nucleotide or greater base complementarity to U1 RNA (Fig. 6A, lanes 3 to 6; Fig. 7A). Some variation in spliceosome assembly for these long substrates may be due to differences in splicing partners. At present, we cannot explain why spliceosome assembly is routinely less pronounced for two-exon than for three-exon substrates with comparable splicing efficiencies. Taken together, these experiments show that a substantial improvement in U1 snRNP binding to the upstream 5’ splice site enhances spliceosome assembly not by itself but only in conjunction with improved binding to the downstream 5’ splice site.

**DISCUSSION**

In this work, experiments were designed to test the model whereby exon partners are specified by the binding of U1 snRNP to the 5’ splice site of the downstream exon as well as the conventional 5’ splice site of the intron to be spliced. The strategy used was to determine the effect of single and combined 5’ splice site mutations at neighboring exons by measuring splicing of the intron in between. We show, using IVS3 and IVS4 as test introns derived from the preprotrachykinin gene, that improved binding of U1 snRNP to the upstream 5’ splice site fails to rescue splicing when the downstream 5’ splice site is naturally weak (IVS3) or impaired by mutation (IVS4) (substrates RP125 and RP123/ RP124; Fig. 7). These findings rule against a model that would explain exon skipping by lowering the total level of U1 snRNP bound in the vicinity of the intron to be spliced and instead support the hypothesis that U1 snRNP binding to the 5’ splice site of the downstream exon is a site-specific and independent requirement for splicing. This hypothesis is congruous with the general features of the exon definition model proposed previously by Robberson et al. (25).

Additional support for the proposed model is indicated by the surprising effect of mutations in the 5’ splice site of E5 that weaken base complementarity to U1 RNA. Although 5’ splice site mutations typically activate nearby cryptic 5’ splice sites, this mutation had two effects on 3’ splice site selection: the 3’ splice site of E5 was inactivated upstream of the mutation, and, unexpectedly, novel branchpoints were activated downstream of the mutation. Formation of the novel branchpoints was shown by site-specific mutagenesis to be dependent on the integrity of a cryptic 5’ splice site, present just downstream of these sites. These data also show that selection of the novel branchpoints is specified chiefly by their proximity to the cryptic 5’ splice site (Table 2).

The results described here demonstrate a true stimulation of 3’ splice site activity by the binding of U1 snRNP directly across the exon (substrates RP123 and RP124; Fig. 4). An effect across the exon is indicated because the mutation that improves the 5’ splice site of E4 strongly stimulates E4 branchpoint formation without having any effect on E5 branchpoint formation, which is abolished for these substrates. The opposite result would be expected if simple splice site competition (in the absence of effects across the exon) were the factor determining exon pair selection. Based
on comparison with consensus sequences, the 5' and 3' splice sites of IVS4 would be expected to be the optimal pair of splice sites used, yet no IVS4 splicing is detected in these substrates. Thus, these results are compatible with a mechanism in which U1 snRNP first base pairs to a 5' splice site and interacts in some fashion with the upstream 3' splice site to modulate branchpoint formation.

The results shown here are complementary to previous studies that have provided evidence for the interaction of U1 snRNP at the 3' splice site of pre-mRNAs, across intron sequences. In yeast cells, the binding of U1 snRNP to the conventional 5' splice site is dependent on the sequence of the branchpoint region located within the same intron, in support of a model in which U1 snRNP interacts with both the 5' splice site and branchpoint regions at an early stage in the splicing process (26). Furthermore, yeast pre-mRNAs form an early U1 snRNP-containing complex, which requires an intact branchpoint region, but not U2 snRNP (19, 30). Thus, these previous studies are distinct in that they address the mechanism of how 5' and 3' splice sites are juxtaposed.

Evidence to indicate a direct interaction of U1 snRNP at the 3' splice site has been reported in mammalian systems. Barabino et al. have shown, by using extracts depleted in specific snRNPs, that U1 snRNP is required for the stable binding of U2 snRNP to the branch site of pre-mRNA (2). Interestingly, this requirement was found to be independent of the 5' splice site or the 5' terminus of U1 RNA. These findings are not necessarily at variance with those reported here, since differences in the methods and substrates used would have precluded detecting the effect of a downstream 5' splice site in the previous study. It is also relevant to consider a possible variation of the exon definition model in which some internal exon sequences might contain binding sites for U1 snRNP and thus stimulate 3' splice site activity in the absence of an authentic downstream 5' splice site. It is interesting to note that in the case of the rat K and T kininogen genes, the last exons contain internal sequences that influence alternative splicing and have complementarity to the 5' end of U1 RNA (15).

This work supports and extends the conclusions of a previous study in which splicing of an alternative exon was shown to be highly sensitive to the adjacent binding of U1 snRNP (17). This sensitivity is not an idiosyncratic feature of an alternative exon, since quite similar results are obtained with a normal exon (E5) of similar size (this work). For this reason, we suspect that the model addressed in this work is of general importance, not only as a way of explaining why some exons are subject to alternative splicing but also as a way of ensuring that normal exons are not mistakenly skipped.

U1 snRNP binding may serve as a simple, structural framework for the recognition of (generally) small exon sequences in long pre-mRNAs. Ultrastructural analysis of actively transcribing genes from Drosophila embryos demonstrates that 25-nm particles are formed first at the 3' and then at the 5' splice site region of the pre-mRNA, followed by a merger of these particles and intron loop removal, interpreted as splicing (3). The size of the 25-nm particles could easily accommodate a complex the size of U1 snRNP. For the case in which exon skipping was observed, no 25-nm particle was present to demarcate the exon. These data are harmonious with the two-site model for U1 snRNP binding and may represent the visualization, in progress, of exon pair selection on nascent pre-mRNA transcripts.

ACKNOWLEDGMENTS

We thank H. Hoffman and B. Hoffmann for critical comments on the manuscript.

This work was supported by the National Institutes of Health (GM39695) and the National Science Foundation (DMB 8858492).

REFERENCES


