RNA (2002), *8*:166–179. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S1355838201010780

Defining a 5′ splice site by functional selection in the presence and absence of U1 snRNA 5′ end

METTE LUND and JØRGEN KJEMS

Department of Molecular and Structural Biology, University of Aarhus, DK-8000 Århus C, Denmark

ABSTRACT

Pre-mRNA splicing in metazoans is mainly specified by sequences at the termini of introns. We have selected functional 5′ splice sites from randomized intron sequences through repetitive rounds of in vitro splicing in HeLa cell nuclear extract. The consensus sequence obtained after one round of selection in normal extract closely resembled the consensus of natural occurring 5′ splice sites, suggesting that the selection pressures in vitro and in vivo are similar. After three rounds of selection under competitive splicing conditions, the base pairing potential to the U1 snRNA increased, yielding a $G^{100\%}U^{100\%}R^{94\%}A^{67\%}G^{89\%}U^{76\%}R^{83\%}$ intronic consensus sequence. Surprisingly, a nearly identical consensus sequence was obtained when the selection was performed in nuclear extract containing U1 snRNA with a deleted 5′ end, suggesting that other factors than the U1 snRNA are involved in 5′ splice site recognition. The importance of a consecutive complementarity between the 5′ splice site and the U1 snRNA was analyzed systematically in the natural range for in vitro splicing efficiency and complex formation. Extended complementarity was inhibitory to splicing at a late step in spliceosome assembly when pre-mRNA substrates were incubated in normal extract, but favorable for splicing under competitive splicing conditions or in the presence of truncated U1 snRNA where transition from complex A to complex B occurred more rapidly. This suggests that stable U1 snRNA binding is advantageous for assembly of commitment complexes, but inhibitory for the entry of the U4/U6.U5 tri-snRNP, probably due to a delayed release of the U1 snRNP.

Keywords: 5′ splice site; in vitro selection; RNA splicing; U1 snRNA; U1 snRNP

INTRODUCTION

Splicing of pre-mRNA is carried out by the spliceosome that consists of five small nuclear ribonucleoprotein particles (snRNPs), U1, U2, U4, U5, and U6, and a number of non-snRNP protein factors. The snRNPs consist of several protein components and a unique small nuclear RNA (snRNA). The sequences of mammalian 5′ splice sites vary considerably but generally conform to the consensus $A_{-2}G_{-1}/G_1U_2R_3A_4G_5U_6$ (R = purine, nucleotides in the pre-mRNA are numbered relative to the exon/intron junction throughout this paper; Mount, 1982; Padgett et al., 1986; Shapiro & Senapathy, 1987). Three of the five spliceosomal snRNAs interact with the 5′ splice site during the course of the splicing reaction (Fig. 1). U1 snRNP associates with the 5′ splice site in an ATP-independent step and forms the earliest detectable splicing complex, the commitment complex (CC), thereby protecting a region spanning positions −3 to +12 relative to the 5′ splice site from RNase digestion (Mount et al., 1983). The 5′ splice site is complementary to the 5′ end of the U1 snRNA and recognition of the 5′ splice site involves base pairing with the U1 snRNA (Zhuang & Weiner, 1986; Séraphin et al., 1988; Siliciano & Guthrie, 1988). The interaction between the pre-mRNA and U1 snRNP is stabilized by other interactions, as in both the *Saccharomyces cerevisiae* and mammalian splicing systems a U1 snRNP particle lacking the 5′ end of the U1 snRNA is still able to interact specifically with an RNA oligonucleotide containing a consensus 5′ splice site, albeit with lower stability (Rossi et al., 1996; Du & Rosbash, 2001). Moreover, in *S. cerevisiae*, several U1 snRNP protein components can be crosslinked to the 5′ splice site region (Zhang & Rosbash, 1999; see Discussion), implying that both U1 snRNP proteins and snRNA associate with the pre-mRNA. In addition to U1 snRNP-associated proteins, other splicing factors may facilitate binding of the U1 snRNP to the 5′ splice site. The best studied example is the splicing factor ASF/SF2, which, upon binding to nearby purine-rich sequences, enhances the interaction of the U1 snRNP with the 5 splice site, prob-

Reprint requests to: Jørgen Kjems, Department of Molecular and Structural Biology, University of Aarhus, C.F. Møllers Allé, Building 130, DK-8000 Århus C, Denmark; e-mail: Kjems@biobase.dk.

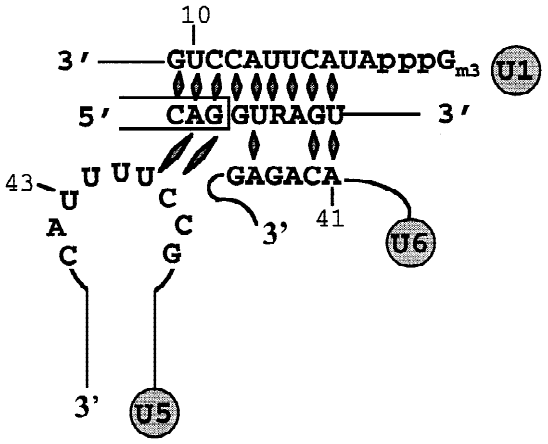


FIGURE 1. Interactions between the 5′ splice site of the pre-mRNA and U1, U5, or U6 snRNAs that are supported by experiments. U1 snRNA is believed to form a transient interaction with the 5′ splice site prior to splicing and is replaced by U6 and U5 snRNAs that interact with the same sequences at the time splicing occurs.

ably through an interaction between the RS domains of ASF/SF2 and the U1 snRNP component U1-70K (Wu & Maniatis, 1993; Kohtz et al., 1994).

U6 snRNA also interacts with the 5′ splice site during splicing. Crosslinking experiments in both mammalian and yeast systems place positions +2 to +6 at the 5′ splice site in close proximity of the conserved $A_{41}CAGAG_{46}$ sequence in U6 snRNA (Wassarman & Steitz, 1992; Kim & Abelson, 1996; numbers refer to the human U6 snRNA; Fig. 1). A site-specific crosslink to position +2 in the intron was only observed in lariat species, suggesting that this base pair only forms during the second step of splicing (Sontheimer & Steitz, 1993). Interactions between positions +1 and +5 in the intron and AGA_{52} and C_{42}, respectively, in the U6 snRNA are also supported by genetic evidence (Kandels-Lewis & Séraphin, 1993; Lesser & Guthrie, 1993; Crispino & Sharp, 1995; Sun & Manley, 1995; Luukkonen & Séraphin, 1998).

The replacement of the U1 snRNA with the U6 snRNA at the 5′ splice is believed to activate the spliceosome, bringing the catalytic site in close proximity of the 5′ splice site. This step is facilitated by a number of spliceosomal components. The RNA helicase Prp28 appears to be involved in the removal of the U1 snRNA from the 5′ splice site. This is mainly based on the observations in *S. cerevisiae* showing that Prp28 is required for the first step of splicing (Strauss & Guthrie, 1994) and that this requirement is bypassed by mutations in the U1 snRNP-specific protein U1C (Chen et al., 2001). Moreover, Prp28 mutants produce a synthetic lethal phenotype with mutations in the U1 snRNA that hyperstabilize the U1 snRNA/5′ splice site interaction in yeast (Staley & Guthrie, 1999). Also the human homolog of Prp28, U5-100kD, is in close contact with the 5′ splice site (Ismaili et al., 2001). Thus, Prp28 probably functions by counteracting the stabilizing effect of U1C on the U1 snRNA/5′ splice site interaction.

Results from the *S. cerevisiae* splicing system indicate that, although base pairing with the U1 snRNA may not be a major determinant in defining the 5′ splice site cleavage site, the U6 snRNA appears to be involved in the definition of the exact scissile bond (Séraphin et al., 1988; Siliciano & Guthrie, 1988; Séraphin & Rosbash, 1990; Kandels-Lewis & Séraphin, 1993; Lesser & Guthrie, 1993). In agreement with this, there appears to be flexibility with respect to the position of U1 snRNA binding in the mammalian system. For instance, U1 snRNAs with an altered 5′ splice site recognition sequence (U1 snRNA nt 3–11), designed to base pair with sequences either upstream or downstream of the 5′ splice site, can activate splicing of otherwise defective 5′ splice sites mutated at positions +3 to +6 (Cohen et al., 1994).

Other mechanisms for early 5′ splice site definition probably exist, as the U1 snRNP is not always required for splicing. Inhibition of splicing upon depletion or inactivation of the U1 snRNP can be abrogated by the addition of high amounts of SR proteins, which appear to replace the U1 snRNP function and restore the splicing efficiency of some but not all pre-mRNA substrates (Crispino et al., 1994; Tarn & Steitz, 1994).

U5 snRNA, which enters the spliceosome as a component of the U4/U6.U5 tri-snRNP to form complex B, also contacts the 5′ splice site (Fig. 1). Phylogenetic comparisons show that the U5 snRNA contains an invariant U-rich sequence of 9 nt displayed in an 11-nt loop structure (Frank et al., 1994). Data obtained from genetic experiments suggests that this sequence interacts with exon nucleotides immediately upstream of the 5′ splice site in the pre-mRNA (Newman & Norman, 1991, 1992) and this is supported by the results from crosslinking experiments (Wassarman & Steitz, 1992; Wyatt et al., 1992; Sontheimer & Steitz, 1993). In addition, the U5 loop sequence contacts the first 2 nt of the intron (Alvi et al., 2001). Because the 5′ exon region is not very well conserved, the interaction is believed to involve nonconventional base pairs. The U5 snRNP may play a direct role in splice site selection based on the observation that, when normal 5′ splice site definition is inhibited by mutation of the first 2 nt of the intron, the U5 loop sequence can influence the choice of the cleavage site (Cortes et al., 1993). A potential protein factor involved in 5′ splice site recognition is the U5 snRNP component Prp8, which has been shown to contact nucleotides on both sites of the 5′ splice site scissile bond, including the first 2 nt of the intron, suggesting that the U1 and U5 snRNPs functionally collaborate in the recognition of the 5′ splice site (Wyatt et al., 1992; Teigelkamp et al., 1995; Reyes et al., 1999; Maroney et al., 2000). Crosslinking studies have identified several other 5′ splice site interacting proteins, although their identities and specific functions

are unknown (Liu et al., 1998; Sha et al., 1998; Ismaili et al., 2001). A picture emerges where a given 5′ splice site region is recognized through a complex network of interactions that changes during the course of splicing.

We have previously used an iterative functional selection approach to study the recognition of the branch point and 3′ splice site (Lund et al., 2000). Here we extend these studies by developing an in vitro strategy for iterative selection of functional 5′ splice site sequences from a pool of pre-mRNAs containing randomized inserts. We provide strong evidence for the hypothesis that factors other than the 5′ end of the U1 snRNA determine the shape of the 5′ splice site. Moreover, based on splicing studies of individual pre-mRNAs with variable complementarity to the U1 snRNA, we conclude that stabilizing U1 snRNA binding to the 5′ splice site in the natural range increases the competitive strength of a splice site, but concurrently inhibits the assembly of the full spliceosome.

RESULTS

Selection of functional 5′ splice sites

A strategy to select functional 5′ splice site sequences from a pool of pre-mRNA containing randomized 5′ splice sites was developed as outlined in Figure 2A (see figure legend for details). The intronic part of the 5′ splice site sequences, which supported the first step of splicing, was recovered from gel-purified lariat intermediate RNA species. To amplify RNA fragments containing unknown 5′ terminal sequences, we exploited the capacity of RNA ligase to form circular molecules from debranched lariat intermediate molecules with high efficiency (20–30%). Moreover, by placing a sequence that was homologous, but nonidentical to the 5′ exon sequence in the end of the 3′ exon, we insured that a functional 5′ exon was reinserted in this ligation step. New templates were generated by PCR using primers that reinserted the extended 3′ exon. The rationale for using different 5′ exon sequences in alternating rounds of selection was to avoid complementarity between the primers in the PCR step, which may lead to formation of primer dimers. Pre-mRNAs having either combination of the 5′ exons were spliced equally well (data not shown).

The changes in the pool composition of selected 5′ splice site sequences were investigated by sequencing individual clones from the unselected pool and from the pools after one and three rounds of selection (Table 1). The conserved sequence motifs obtained from the aligned sequences are illustrated as logos in Figure 3A,B. Note that all sequences in the unselected library contain a fixed exonic CAG sequence that can form potential base pairing with nt 9–11 in the U1 snRNA. The unselected pool contained no significant bias in the nucleotide composition in sequences from 20 random clones, indicating that the sequences are near random (data not shown). After one round of selection, the randomized region of individual clones all conferred to a $G^{100\%}U^{100\%}R^{90\%}A^{59\%}G^{82\%}U^{55\%}$ (R = purine) consensus sequence that resembles the conventional mammalian 5′ splice site (Table 1; Fig. 3E). No significant sequence preference was observed at the positions following this motif. The number of base pairs between the U1 snRNA and the intronic part of the 5′ splice sites ranged from 3 + 4 to 3 + 7 with an average of 3 + 5.9 (3 denoting the number of potential base pairs between the U1 snRNA and the exonic part of the 5′ splice site that is constant).

After three rounds of selection, the motif had evolved to $G^{100\%}U^{100\%}R^{94\%}A^{67\%}G^{89\%}U^{76\%}R^{83\%}$. The most significant difference between the motifs obtained after one and three rounds of selection was that the preference for a U at position +6 had increased from 55% to 76%. Moreover, 83% of the sequences now contained a purine at position +7. The average number of base pairs between the U1 snRNA and the intron part of the 5′ splice site had increased to 3 + 6.4, several 5′ splice sites forming the theoretical maximum of 3 + 8 base pairs with the U1 snRNA.

The splicing efficiencies of the pre-mRNA pools before and after each round of selection were compared with that of a control splicing substrate, PIP85.A (Fig. 4A). The splicing efficiency increased from 22% of the control splicing substrate in the unselected pool to 70% after one round of selection (Fig. 4A, lanes 2 and 4). After two and three rounds of selection, the efficiency was comparable to that of the control splicing substrate (Fig. 4A, lanes 6, 8, and 10). This correlates with the appearance of 5′ splice sites with a high match to the consensus sequence already after one round of selection, and with the slight increase in the homology to the consensus sequence during further rounds of selection. The splicing efficiency of RNA transcripts from arbitrarily selected clones (1, 48, 70, and 75; Table 1) was also analyzed by incubating the transcripts under standard in vitro splicing conditions. All four clones appeared to be excellent substrates for both steps of splicing, exhibiting the same first step and slightly faster second step splicing kinetics than a substrate based on the optimal splicing substrate PIP85.A (data not shown).

Selection of 5′ splice sites in the absence of the 5′ end of the U1 snRNA

To investigate the recognition of the 5′ splice site by other splicing factors than the 5′ end of the U1 snRNA, a new selection experiment was performed using a nuclear extract (NE) where the 5′ end of the U1 snRNA had been degraded by oligonucleotide-directed RNase H cleavage (Δ5′end U1-NE). We estimated from primer extension analysis of total RNA isolated from Δ5′end U1-NE that more than 99% of the U1 snRNA was

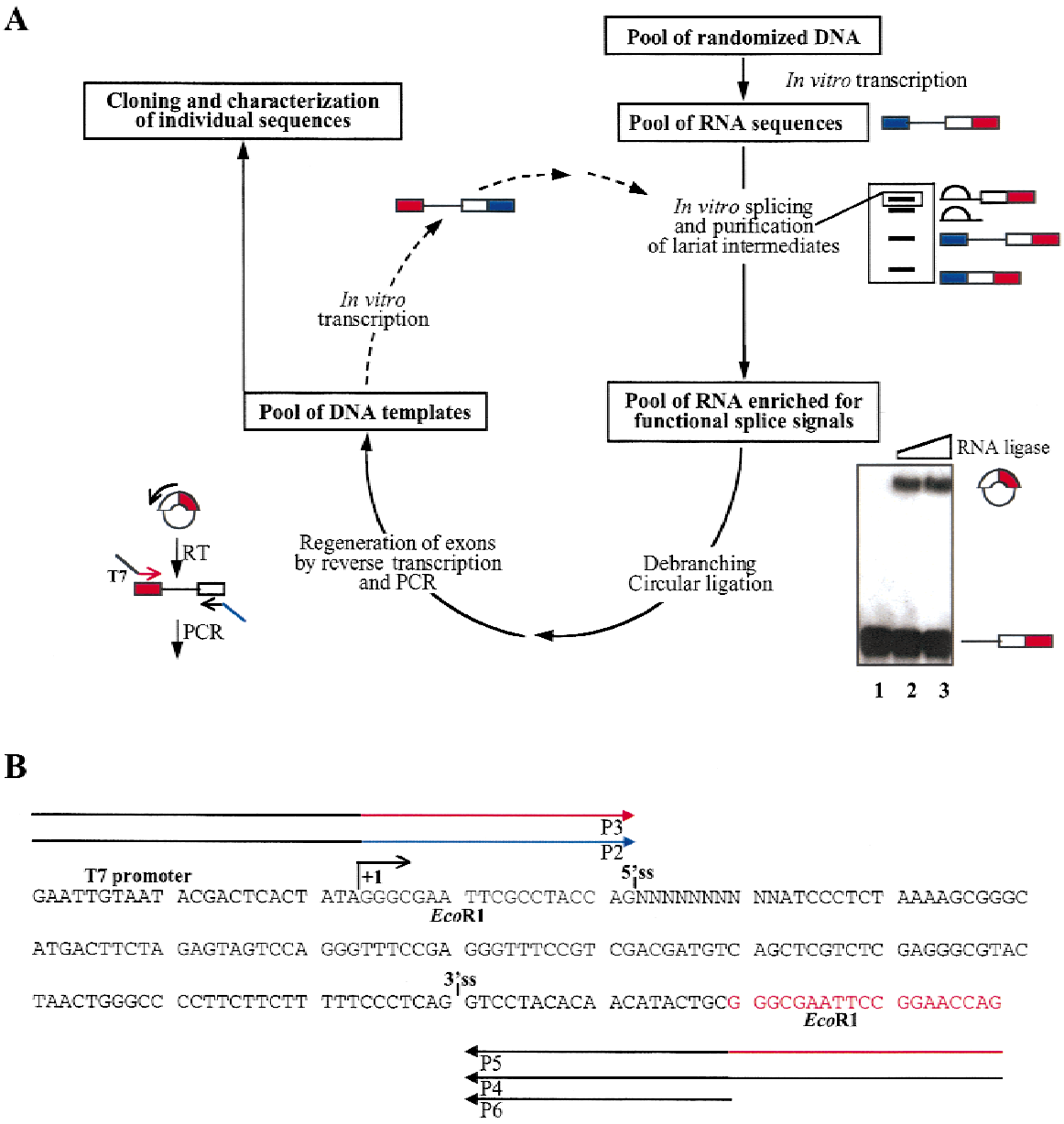


FIGURE 2. Strategy for selection of functional 5′ splice sites from random sequences. **A:** The pre-mRNA pool was generated by in vitro transcription of a DNA template based on the optimal splicing construct pPIP85.A. A randomized 10-nt-long insert replaced the first 10 nt immediately downstream of the 5′ exon (blue box). The 3′ exon (white box) was extended with a 5′ exon sequence variant (red box). The transcripts were incubated under standard in vitro splicing conditions, and lariat intermediate species were isolated by denaturing gel electrophoresis. A schematic drawing of the gel is indicated; lines and boxes represent introns and exons, respectively. After debranching, the lariat intermediates were incubated with RNA ligase, resulting in an intramolecular ligation of the RNAs that joined the downstream 5′ exon with the 5′ end of the intron (see inserted autoradiogram). The circular RNAs were reverse transcribed using a primer annealing to the 3′ exon (P6). The resulting cDNA had the variant 5′ exon (red box) inserted upstream and a shortened 3′ exon. Finally, a T7 promoter was added to the 5′ exon variant and the original 5′ exon (blue box) was inserted downstream from the 3′ exon by PCR, using primers P3 and P4. The subsequent rounds were performed in the same manner except that the two 5′ exons sequences at the termini of the transcript were alternated. **B:** Sequence of the DNA construct used to generate the initial RNA pool. The primers P2–P6 that were used in the reverse transcription and PCR steps are indicated. Primer sets P3/P4 and P2/P5 were used to amplify the pool after unequal and equal numbers of selection rounds, respectively. The site for transcriptional initiation (+1), the splice sites (5′ss and 3′ss), and the *Eco*RI cleavage sites used for cloning are indicated.

cleaved by RNase H in an oligonucleotide dependent fashion (Fig. 5A). Because the RNase H cleavage yielded a 6–10-nt terminal deletion, the truncated U1 snRNA should not be able to form stable base pairs with the pre-mRNA (Fig. 5A).

The pool composition was analyzed after one and three rounds of selection (Table 1) and the consensus sequences are illustrated as logos in Figure 3C,D. Comparing these 5′ splice site motifs with those selected in the presence of intact U1 snRNA revealed only small

TABLE 1. Sequence alignment of 5′ splice site sequences selected under standard conditions of splicing and in the absence of the 5′ end of U1 snRNA (Δ5-end U1-NE) after one or three rounds of selection.[a]

Conditions	Standard conditions, 1 round of selection	Standard conditions, 3 rounds of selection	Δ5′end U1-NE, 1 round of selection	Δ5′end U1-NE, 3 rounds of selection
Sequences of clones	1 **GUA**U**G**CCGUC	23 **GUG**U**G**GUAUU	41 **GU**U**G**CG**G**GUc	60 **GUGAG**GUAau
	2 **GUGa**ucccuc		43 **GUA**U**G**GCGUA	61 **GUAA**CGU**U**CG
	3 **GUGA**ACUAAC	24 **GUAAG**GCAau		62 **GUGA**A**U**CGau
		25 **GUAA**U**U**Uauc	42 **GUAGG**GUCCU	
	4 **GUA**U**GU**CGAa		44 **GUAAG**CUGGU	63 **GUAAG**G**A**Cau
	5 **GU**UC**GUA**AUC	26 **GUA**U**GUA**auc	45 **GUA**C**GU**UCUU	64 **GUG**U**GUG**cau
	6 **GUAGau**cccc	87 **GUAAG**C**G**CUU	65 **GUAAGU**CGUa	
		28 **GUG**U**GUG**CUA	46 **GUAA**A**UG**CGG	66 **GUA**C**GUA**Cau
	7 **GUGAG**GU**U**UA	29 **GUG**C**GUA**CCU	47 **GU**C**GGUA**auc	
	8 **GUAA**U**UA**AUc	30 **GUAAG**C**AAUG**	48 **GUAAG**C**G**CUU	67 **GUA**U**GUau**cc
	9 **GU**U**AGU**U**U**GC		49 **GUG**C**GU**U**U**AC	68 **GUAAGUG**CUa
	10 **GUGAGU**CGAc	31 **GUAA**C**UGU**AU	50 **GUA**U**GUG**CGU	69 **GUA**C**GUGU**Aa
	11 **GUAAG**CU**U**au	32 **GUGAG**C**au**cc	51 **GUAA**UG**AU**CG	
	12 **GUGGGU**CCAC	33 **GUGAGUA**GCU		70 **GUAAGUGU**GU
	13 **GUAAG**C**A**Gau	34 **GUGAGUA**GCG	52 **GUG**C**GUau**cc	71 **GUGAGUGU**UC
	14 **GUA**C**G**A**GU**au	35 **GUGGGUG**auc	53 **GUG**U**GUGU**cc	72 **GUGAGUau**cc
		36 **GUGAGUG**GAU	54 **GU**C**AGUau**cc	73 **GUAAGUau**cc
	15 **GUGAG**G**AU**UU	37 **GU**U**AGUau**cc	55 **GUAAGUA**ACG	
	16 **GUAAGUA**CAU	38 **GUG**C**GUau**cc	56 **GUGGGUG**CCG	74 **G**C**A**U**GUau**cc
	17 **GUAGGUA**Aau		57 **GUGAGUA**Gau	75 **G**C**A**U**GUAU**CG
	18 **GUAAGUG**Cau	39 **GUGAGUau**cc		
	19 **GUAAGUG**auc	40 **GUAAGUau**cc	58 **GUAAGUau**cc	
	20 **GUAGG**C**au**cc			
	21 **GUG**U**GUau**cc		59 **G**A**A**U**G**A**A**ACU	
	22 **GUAAG**A**GU**GA			
Consensus (%)	**G_{+1} U R A G U** 100 100 90 59 82 55	**G_{+1} U R A G U R** 100 100 94 67 89 76 83	**G_{+1} U R A G U R** 100 95 84 42 84 63 73	**G_{+1} U R A G U R U** 100 88 100 62 88 81 75 56
Intron/U1 base pairs	5.95	6.44	5.94	6.50

[a]The consensus sequence is given the cloned sequences where the percentage of conservation is indicated for each position (R: purine). Specific nucleotides are included in the consensus sequence when the degree of conservation is at least 50% for A, G, C, or U and 70% for R. Lowercase letters correspond to nucleotides originating from the constant sequence following the randomized region, and nucleotides predicted to interact with the conserved 5′ end of U1 snRNA are denoted in bold letters. The sequences in each pool are grouped according to the number of bases complementary to the 5′ end of the U1 snRNA, lowest number at the top. Sequences that diverge from the 5′ GU dinucleotide were observed in clones 59, 74, and 75.

differences (Table 1; compare Figs. 3A and 3C, and Figs. 3B and 3D). Three selected clones did not contain the otherwise invariant GU dinucleotide (Table 1, clones 59, 74, and 75), and these clones were tested individually in a splicing assay. Pre-mRNA generated from clone 59, with a noncanonical GA dinucleotide, yielded no detectable splicing in either type of extracts and may therefore be regarded as a contaminant carried through the first round of selection. However, pre-mRNAs generated from clones 74 and 75, containing a GC sequence, were efficiently spliced, which is consistent with previous in vitro studies (Aebi et al., 1986), and the observation that this dinucleotide is the only splice site variant found with a relatively high frequency (1 in 200) in GenBank-annotated mammalian genes (Burset et al., 2000). The appearance of GC-containing clones specifically in the selection performed in Δ5′end U1-NE is probably not significant, as transcripts from these clones were not spliced more efficiently in this extract than GU-containing clones from the same pool (data not shown).

Comparing the efficiency of the splicing of the pre-mRNA pools selected in the absence or presence of the 5′ end of the U1 snRNA revealed almost no difference in untreated nuclear extract (Fig. 4A,B, even lane numbers). Moreover, no significant increase in RNA splicing in Δ5′end U1-NE was observed for the pools that were selected under these conditions (Figs. 4A,B; odd lane numbers). One concern may be that the splicing observed in Δ5′end U1-NE proceeded via a splicing pathway dependent on the 1% intact U1 snRNA remaining in the RNase H-treated extract (Fig. 5A). To address this problem, we purified the A complexes formed on PIP85.A pre-mRNA transcripts in Δ5′end U1-NE on a native gel, extracted the RNA, and, by primer extension, we estimated that at least 95% of the U1 snRNA in the complexes lacked the 5′ end (Fig. 5B). Based on this result and discussions below, we con-

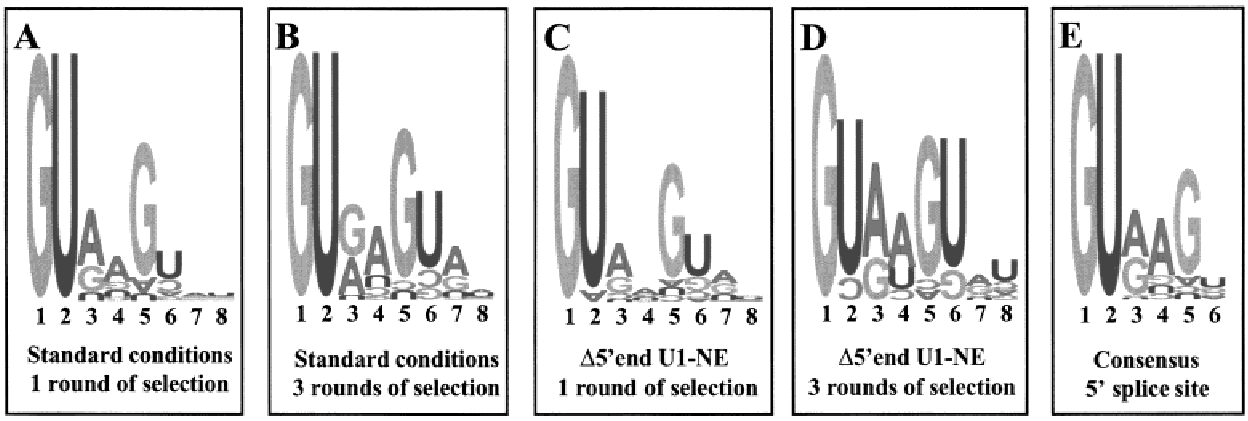
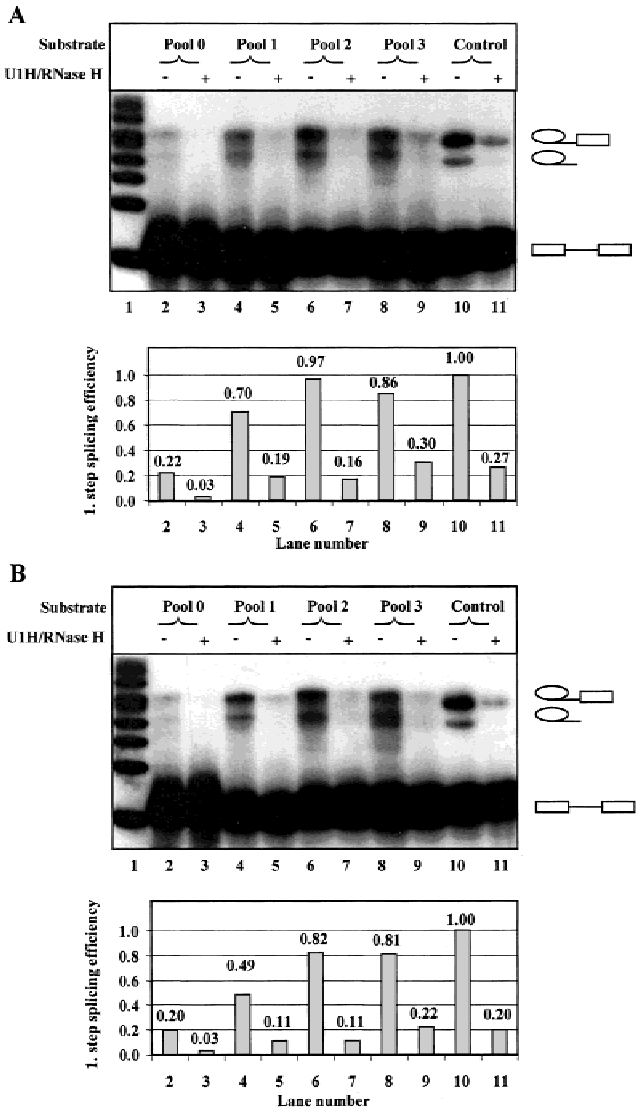


FIGURE 3. Consensus sequence logos generated from the alignments shown in Table 1 (**A**–**D**) and from approximately 22,000 splice site pairs from mammalian GenBank annotated genes (**E**; Burset et al., 2000). The relative size of the letters at a given position reflects their relative frequencies in the alignment, and the height of the entire stack of letters is adjusted to indicate the information content of the position (Schneider & Stephens, 1990). Nucleotides that are underrepresented at a given position are shown upside down.



clude that these complexes are precursors for functional spliceosomes, and that the 5′ end of the U1 snRNA is dispensable for the sequence-specific recognition of the 5′ splice sites.

Effect of 5′ splice site complementarity to U1 snRNA on splicing kinetics

The selected 5′ splice sites all have the potential to form from 3 + 4 to 3 + 8 base pairs with the U1 snRNA. However, because the sequences vary nonsystematically among individual members of the selected pools, conclusions concerning the influence of 5′ splice site/U1 snRNA complementarity on splicing kinetics could not be drawn. To analyze this parameter more systematically, RNA substrates having 3 + 4, 3 + 5, 3 + 6, 3 + 7, and 3 + 8 consecutive nucleotides complementary to the 5′ end of the U1 snRNA in a constant background were constructed. The 5′ end of the intron was gradu-

FIGURE 4. In vitro splicing of pre-mRNA pools selected under standard conditions of splicing (**A**) and in the absence of the 5′ end of the U1 snRNA (**B**). The RNA substrates represent the unselected library (lanes 2–3), the RNA pool after one (lanes 4–5), two (lanes 6–7), and three (lanes 8–9) rounds of selection. The splicing of a control substrate PIP85.A, which is the parental substrate for the library, is included for comparison (lanes 10–11). The RNA substrates were spliced for 45 min, using untreated NE (−) or U1H/RNase H treated NE (+). The RNA products were quantified by PhosphorImager and normalized to the number of uridines in each of the RNA species. The efficiency of the first step of splicing was estimated from the amount of lariat species (lariat and lariat intermediate) divided by the sum of unspliced pre-mRNA, lariat, and lariat intermediate. The splicing efficiencies were normalized to that of the control substrate and are illustrated in the diagrams below the gels; numbers correspond to lane numbers.

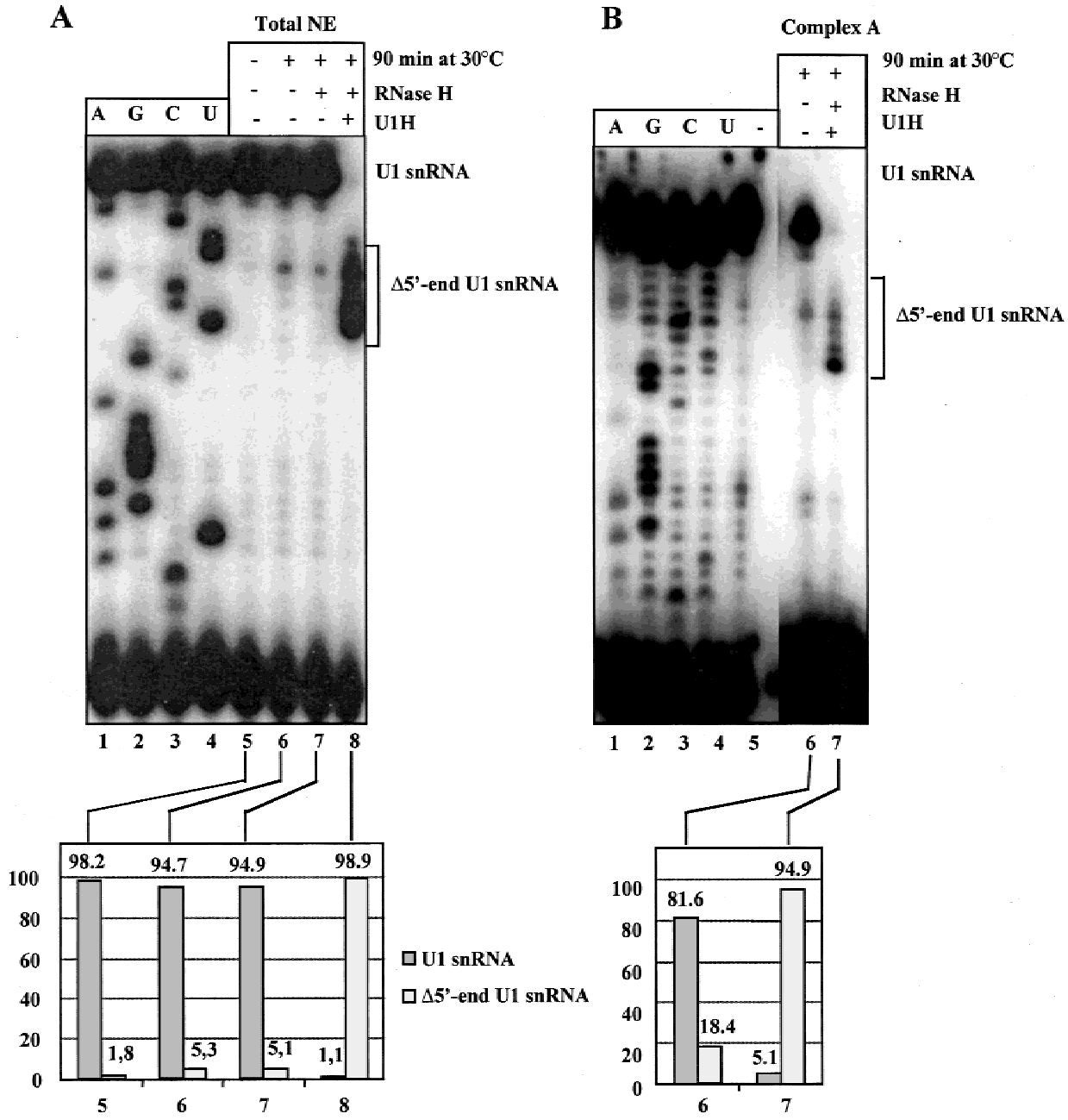


FIGURE 5. Mapping the 5′ end of the U1 snRNA in nuclear extract (**A**) and in purified A complexes (**B**) by primer extension. **A:** The template RNA was isolated from nuclear extract before (lane 5) and after incubation under standard in vitro splicing conditions (lane 6), in the presence of RNase H alone (lane 7), or in the presence of RNase H and oligonucleotide U1H complementary to the 5′ end of the U1 snRNA (lane 8). **B:** Template RNA was extracted from gel-purified A complexes formed in NE (lane 6) and in Δ5′end U1-NE (lane 7). All RNA samples were subjected to primer extension using reverse transcriptase and a primer annealing to U1 snRNA position 28–47. The sequencing marker indicates the corresponding RNA sequence and was obtained by primer extension on RNA from untreated nuclear extract (lanes 1–4 in both panels). The positions of the cDNA corresponding to full-length and cleaved U1 snRNA are indicated, and the yields are indicated in lower panels by gray and white bars, respectively. The numbers below the diagrams correspond to the lane numbers in the upper panel.

ally changed from **G_{+1}UAA**UCCG to **G_{+1}UAAGUAU**, in steps of one nucleotide (Fig. 6A; bold letters denote nucleotides that potentially can interact with the conserved 5′ end of the U1 snRNA). The splicing activities of all RNA substrates were compared after 20, 30, and 45 min of incubation under splicing conditions at low RNA concentration (<1 fmol/μL; Fig. 6B). Based on the results from the selection experiment, we expected that an extension of the 5′ splice site/U1 snRNA complementarity would result in increased splicing efficiency. Surprisingly, extension of the potential base pairing interaction with the U1 snRNA resulted in a

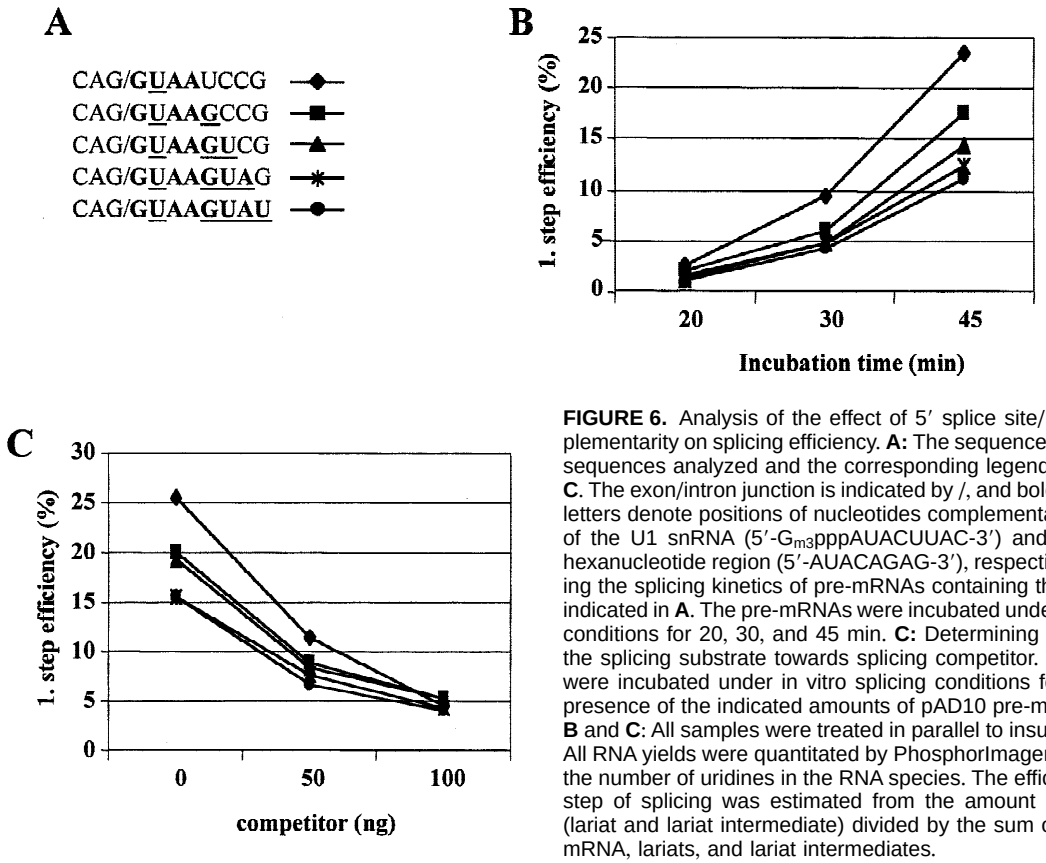


FIGURE 6. Analysis of the effect of 5′ splice site/U1 snRNA complementarity on splicing efficiency. **A:** The sequences of 5′ splice site sequences analyzed and the corresponding legends used in **B** and **C**. The exon/intron junction is indicated by /, and bold and underlined letters denote positions of nucleotides complementary to the 5′ end of the U1 snRNA (5′-G_{m3}pppAUACUUAC-3′) and the U6 snRNA hexanucleotide region (5′-AUACAGAG-3′), respectively. **B:** Measuring the splicing kinetics of pre-mRNAs containing the 5′ splice sites indicated in **A**. The pre-mRNAs were incubated under in vitro splicing conditions for 20, 30, and 45 min. **C:** Determining the sensitivity of the splicing substrate towards splicing competitor. The pre-mRNAs were incubated under in vitro splicing conditions for 45 min in the presence of the indicated amounts of pAD10 pre-mRNA competitor. **B** and **C**: All samples were treated in parallel to insure comparability. All RNA yields were quantitated by PhosphorImager and adjusted to the number of uridines in the RNA species. The efficiency of the first step of splicing was estimated from the amount of lariat species (lariat and lariat intermediate) divided by the sum of unspliced pre-mRNA, lariats, and lariat intermediates.

reduction of the efficiency of the first step of splicing at all incubation times analyzed. An explanation for this observation could be that to commit a 5′ splice site to splicing, it must compete with the other splice sites in the reaction mixture for splicing factors, including U1 snRNA. In the presence of high concentrations of splicing competent RNA (~30 fmol/μL) during the selection experiment, the individual RNA molecules may compete to a higher extent for U1 snRNP binding. An extended 5′ splice site/U1 snRNA complementarity could probably increase the competitive strength of a given 5′ splice site sequence. In contrast, at low RNA concentrations, which is the condition in the experiment shown in Figure 6B, splicing factors may not be limiting, lowering the requirement for stable U1 snRNA binding. This interpretation was supported by studying the splicing kinetics of the transcripts with a variable 5′ splice site/U1 snRNA complementarity in the presence of increasing amounts of pAd10 splicing substrate (containing GGG/GUGAGUAC at the 5′ splice site) as a competitor for splicing factors (Fig. 6C). In the absence of competitor, a short complementary stretch of only 3 + 4 base pairs was most favorable for splicing. However, in the presence of increasing amounts of competitor substrate, the efficiencies of splicing approached each other and at the highest competitor concentration assayed (70 fmol/μL), the splicing efficiencies were basically identical. This implies that extended base pairing between a pre-mRNA and the U1 snRNA, a characteristic that is often considered to be favorable for splicing, is actually inhibitory at low RNA concentrations.

Extended base pairing between the 5′ end of the U1 snRNA and the 5′ splice site stimulates complex A formation, but inhibits complex B assembly

A likely explanation for the observed decrease in splicing efficiency for the 3 + 8 pre-mRNA substrate is that the removal of the U1 snRNA from the 5′ splice site is inhibited. This model implies that the inhibition would be relieved in extracts containing truncated U1 snRNA that is unable to interact with the 5′ splice site. To test this hypothesis, we investigated the kinetics of splicing of the 3 + 4 and 3 + 8 pre-mRNA substrates both in NE and Δ5′end U1-NE (Fig. 7A, quantifications are shown in Fig. 7C). Although the appearance of splicing products was clearly delayed in the absence of the 5′ end of the U1 snRNA, the 3 + 8 construct reached the same level of splicing after 90 min in both types of extract (Fig. 7A, lanes 11 and 22; note that incubation time in Fig. 4 only was 45 min, yielding a relatively small amount of splicing products in Δ5′end U1-NE). Notably, the relative efficiencies of 3 + 4 and 3 + 8

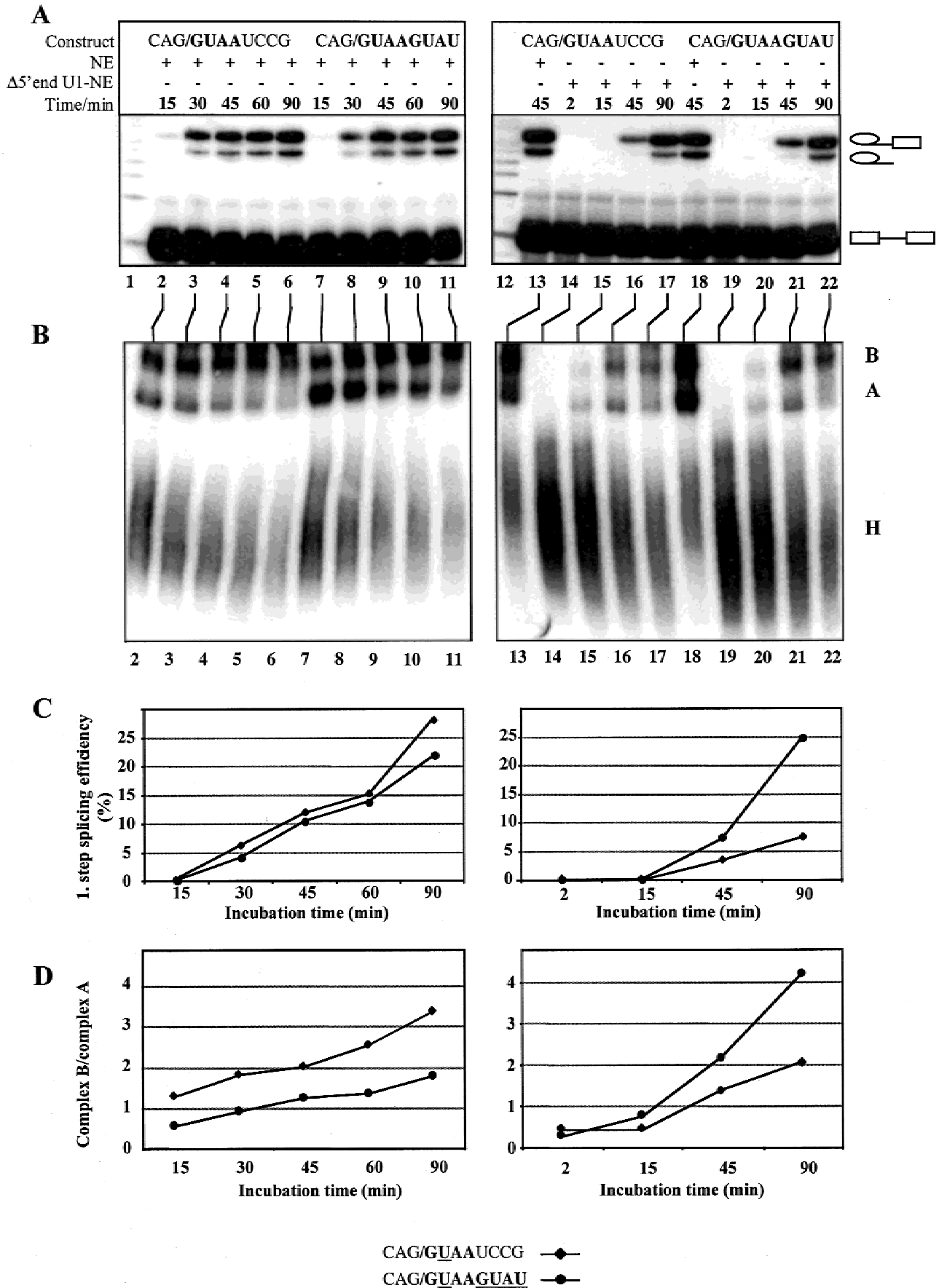


FIGURE 7. *See caption on facing page.*

pre-mRNA splicing were reversed when the pre-mRNA substrates were incubated in Δ5′end U1-NE. The 3 + 8 sequence with a high match to the 5′ splice site consensus was no longer inhibitory, but favorable, for splicing in the absence of the 5′ end of the U1 snRNA (Fig. 7C). To assess the kinetics of spliceosome assembly, the same samples were analyzed by native gel electrophoresis (Fig. 7B, quantifications are shown in Fig. 7D). The formation of splicing complexes was clearly delayed for both the 3 + 4 and 3 + 8 pre-mRNAs, indicating that the 5′ end of the U1 snRNA stimulates the kinetics of splicing complex assembly. However, interesting differences were observed in the kinetics for the A to B complex transition. In normal extract, the 3 + 8 pre-mRNA exhibited a strong accumulation of complex A (Fig. 7B, lanes 7–11, Fig. 7D) whereas the transition into B complexes was significantly more efficient for the 3 + 4 pre-mRNA when comparing the complex B/complex A ratio (Fig. 7B, compare lanes 2–6 and 7–11; Fig. 7D, left panel). The situation was reversed in the Δ5′end U1-NE, where the complex B/complex A ratio was significantly higher for the 3 + 8 pre-mRNA than for the 3 + 4 pre-mRNA at later time points (compare lanes 16–17 and 21–22; Fig. 7D, right panel). Primer extension experiments on RNA purified from the A complexes formed in NE and Δ5′end U1-NE showed that they contained almost exclusively U1 snRNA and truncated U1 snRNA, respectively (Fig. 5B). We conclude that extended base pairing between the U1 snRNA and the 5′ splice site inhibits the transition of complex A to complex B and that this effect can be reversed by removing the 5′ end of the U1 snRNA. However, eliminating the base pairing between the U1 snRNA and the 5′ splice site also reduces the rate of complex A assembly.

DISCUSSION

We have developed a procedure for selecting functional 5′ splice sites in vitro from a pool of pre-mRNA containing a randomized insert at the 5′ end of the intron. To detect subtle sequence preferences, multiple selection rounds were performed. The degree of conservation of individual positions of the 5′ splice site after one round of selection was $G^{100\%}U^{100\%}R^{90\%}A^{59\%}G^{82\%}U^{55\%}$, which is strikingly similar to the level of conservation found in a comparison of approximately 22,000 splice site pairs from mammalian GenBank annotated genes, $G^{100\%}U^{100\%}R^{94\%}A^{71\%}G^{81\%}U^{46\%}$ (compare Figs. 3A and 3E; Burset et al., 2000). This may reflect that the in vitro conditions that we applied in the first round of selection resemble the selection pressure opposed upon 5′ splice sites in the cell. In the subsequent selection rounds, the competition for splicing factors probably increases due to a higher proportion of optimized 5′ splice sites in the RNA library, which may explain the observed increase in the level of U1 snRNA complementarity to $G^{100\%}U^{100\%}R^{94\%}A^{67\%}G^{89\%}U^{76\%}R^{83\%}$. Because only lariat intermediates were amplified in the RT-PCR step, there was a potential risk that our selection procedure may counterselect for 5′ splice sites functional in the second step of splicing. However, in vitro splicing of the selected pools and individual clones all appeared fully functional in both steps of splicing, suggesting that the splice site sequence preferences for the first and second steps of splicing are indistinguishable.

On the basis of sequences in Table 1, we can conclude that the minimum requirement for splicing appears to be a 3 + 4 base pair match to the U1 snRNA (including G-U pairs), the functional 5′ splice site must encompass the dinucleotide GU or GC at the 5′ end of the intron, and the guanosine at position +5 is not obligatory if the two preceding positions are purines. The main difference between sequences in pools 1 and 3 occurred at position +7, which was highly variable after the first round of selection, but selected as a purine after three rounds. This suggests that extending base pairing to the U1 snRNA to include this position is favorable for splicing under competitive conditions. The extent of U1 snRNA base pairing did not have any differential influence on the first and second step of splicing (data not shown). This is consistent with the notion that the GU dinucleotide is the only part of the intron 5′ splice site sequence that has been experimentally implicated in the second step of splicing: Before the second step of splicing, the U6 snRNA is in close contact with U_{+2} of the intron (Sontheimer & Steitz, 1993) and it has been proposed that a direct, non-Watson–Crick interaction between the G_{+1} and the last adenosine of the intron is established at the same time (Parker & Siliciano, 1993; Deirdre et al., 1995).

FIGURE 7. (*facing page*) Analysis of the splicing kinetics (**A**) and complex formation (**B**) of the 3 + 4 and 3 + 8 pre-mRNA substrates. The samples were incubated in normal extract (NE +) or Δ5′end U1-NE (NE Δ5′end U1 +) for the time indicated. The positions of lariat species and pre-mRNA are indicated in **A** and the positions of the spliceosomal complexes A and B and the nonspecific complex H are indicated in **B**. RNA products and spliceosomal complexes were quantified by PhosphorImager and normalized to the number of uridines in each of the RNA species. **C:** The efficiency of the first step of splicing was calculated from the amount of lariat species (lariat and lariat intermediate) divided by the sum of unspliced pre-mRNA, lariat, and lariat intermediate based on the result shown in **A**. **D:** The ratio of complex B to complex A was quantitated based on the result shown in **B**. The legend is indicated below. The experiments were repeated with various preparations of RNA and NE yielding essentially the same result.

The multiple recognition events of the 5′ splice site make it difficult to explore the contribution from individual factors to the sequence specificity. To investigate the contribution from other factors than the 5′ end of the U1 snRNA, we repeated the selection experiment in the presence of 5′ truncated U1 snRNA that could not engage in base pairing with the 5′ splice site. Surprisingly, the consensus 5′ splice site sequences after one and three rounds of selection were highly similar to the motifs obtained in the presence of intact U1 snRNA. It is unlikely that the trace of intact U1 snRNA in the Δ5′end U1-NE contributes significantly to the splicing, as the complex A formed in this extract contained less than 5% intact U1 snRNA. Moreover, the observed complex A most likely constitutes a functional prespliceosome complex because it was converted into a fully assembled spliceosome (complex B) at the same rate as the appearance of splicing products (Fig. 7B, right panel). Notably, splicing of the 3 + 8 clone in Δ5′end U1-NE reached a level of splicing efficiency after 90 min that was almost indistinguishable from splicing in normal extract. This suggests that the 5′ end of the U1 snRNA is dispensable for splicing and that other factors recognize the entire 5′ splice consensus sequence. Because we, in this study, select for turnover of splicing substrate, the consensus 5′ splice site sequence may be shaped both by functional interactions with *trans*-acting snRNA and protein factors and as a *cis*acting sequence with importance for the catalytic step. One candidate protein factor that potentially could contribute to the sequence specificity is the U1 snRNP specific protein U1C. Complementation studies with purified U1 snRNP particles lacking subsets of U1-specific proteins show that U1C, but not U1-70K and U1A, is important for formation of early spliceosome complexes in mammalian systems (Heinrichs et al., 1990; Jamison et al., 1995; Will et al., 1996). Moreover, U1C can be crosslinked to the 5′ splice site both in the mammalian (Rossi et al., 1996) and yeast systems (Zhang & Rosbash, 1999). In yeast, the yU1-70K, ySmD1, ySmD3, ySmB, Nam8, and Snu56 proteins also crosslink to the 5′ splice site (Zhang & Rosbash, 1999) and the Sm proteins have been shown to stabilize U1 binding (Zhang et al., 2001). Another candidate factor is the U5 snRNP-specific protein Prp8 based on numerous reports of Prp8 crosslinks to the 5′ splice site (Wyatt et al., 1992; Teigelkamp et al., 1995; Reyes et al., 1999; Siatecka et al., 1999; Collins & Guthrie, 1999; Maroney et al., 2000). The binding of Prp8 to the 5′ splice site is probably responsible for the recruitment of U4/U6.U5 tri-snRNP to the spliceosome (Konforti & Konarska, 1994) and more recent data suggests that this represents an important ATP-dependent step in early spliceosome assembly (Maroney et al., 2000).

During multiple rounds of selection, the average number of nucleotides complementary to the U1 snRNA rose from 3 + 5.9 after the first round to 3 + 6.4 after the third round, which correlated with increased splicing efficiency. However, extending the complementarity systematically from 3 + 4 to 3 + 8 base pairs in a constant background showed that at low RNA concentration (<1 fmol/μL) splicing kinetics were inversely correlated with the extent of U1 snRNA complementarity, the shortest base pairing tested (3 + 4) being most optimal for splicing. Thus, a short base pairing to the U1 snRNA may only be favorable for splicing at low RNA concentrations. In agreement with this, short base pairing to the U1 snRNA was no longer advantageous in terms of splicing efficiency in the presence of increased levels of competitor pre-mRNA. The analysis of splicing complexes revealed that substrates with extended U1 snRNA base pairing potential formed complex A more rapidly, suggesting that the advantage in splicing under competitive conditions may be ascribed to more efficient recruitment of splicing factor(s) that are limiting for formation of commitment complexes. This factor is most likely the U1 snRNP, since the removal of the 5′ end of the U1 snRNA eliminated the effect. It has previously been shown that hyper stabilization of the U1 snRNA/5′ splice site interaction to include from 2 + 8 to 2 + 10 base pairs is detrimental to splicing and this effect can be neutralized by a concomitant increase in the U6 snRNA/5′ splice site interaction (Staley & Guthrie, 1999). Our observations show that stabilizing the interactions between the U1 snRNA and pre-mRNA in a shorter and more natural range of 3 + 4 to 3 + 8 base pairs also has a negative effect on the association of U4/U6.U5 tri-snRNP in vitro, probably due to a more inefficient release of U1 snRNP. Based on these data, we conclude that the extent of base pairing between U1 snRNA and the 5′ splice site has opposite constraints on the competitive strength and splicing efficiency of the splice site. This balance is likely to play an important role in the regulation of 5′ splice site selection in alternatively spliced genes.

MATERIALS AND METHODS

Constructs

In the following, an asterisk (*) denotes that the restriction site has been blunt-ended with Klenow enzyme. The pPIP85.A contains an artificial transcription unit optimized for splicing (Moore & Sharp, 1992). pPIP85.AΔ5′ss was constructed by deletion of the 59 bp *Eco*RI**Bam*HI* fragment from pPIP85.A, thus deleting the exon and 9 nt of the intron. New 5′ exons were incorporated by PCR. The following primers were used in the 5′ splice site selection procedure:

P1: 5′-GAATTGTAATACGACTCACTATAGGGCGAATTCG CCTACCAGNNNNNNNNNNATCCCCTAAAAGCGGGC ATG-3′, encompassing sequences for T7 promoter, 5′ exon, and 31 nt of intron sequences including a 10-nt random insert.

P2: 5′-GAATTGTAATACGACTCACTATAGGGCGAATTCG CCTACCAG-3′, encompassing sequences for T7 promoter and 5′exon sequences corresponding to the wild type 5′ exon.
P3: 5′-GAATTGTAATACGACTCACTATAGGGCGAATTCC GGAACCAG-3′, encompassing sequences for T7 promoter and 5′exon sequences corresponding to a mutated 5′ exon.
P4: 5′-CTGGTAGGCGAATTCGCCCGCAGTATGTTGTGTA GGAC-3′, encompassing sequences for 3′ exon and sequences corresponding to the wild-type 5′ exon.
P5: 5′-CTGGTTCCGGAATTCGCCCGCAGTATGTTGTGT AGGAC-3′, encompassing sequences for 3′ exon and sequences corresponding to the mutated 5′ exon.
P6: 5′-GCAGTATGTTGTGTAGGAC-3′, encompassing sequences for 3′exon (Fig. 2B).

To generate constructs with variable complementarity to the U1 snRNA the following primers, encompassing sequences for 5′ exon, and 31 nt of intron sequences, were used. Bold letters are used in the 8-nt region where the primers differ:

P7: 5′-GGGCGAATTCCGGAACCAG**GTAATCCG**TCATCC CTCTAAAAGCGGGCATG.
P8: 5′-GGGCGAATTCCGGAACCAG**GTAAGCCG**TCATCCC TCTAAAAGCGGGCATG.
P9: 5′-GGGCGAATTCCGGAACCAG**GTAAGTCG**TCATC CCTCTAAAAGCGGGCATG.
P10: 5′-GGGCGAATTCCGGAACCAG**GTAAGTA**GTCAT CCCTCTAAAAGCGGGCATG.
P11: 5′-GGGCGAATTCCGGAACCAG**GTAAGTAT**TCATC CCTCTAAAAGCGGGCATG.

Templates were generated by a two-step PCR procedure. In the first step, constructs containing no promoter were amplified using pPIP85.A and primer P5 in combination with primers P7, P8, P9, P10, or P11. In the second step, 500 ng of the product from the first step were used as template in combination with primers P2 and P5, adding a T7 promoter to the templates.

Library construction and selection procedure

DNA templates, with 10 randomized intron nucleotides, were prepared by PCR in a 50-μL reaction containing 100 ng template (pPIP85.AΔ5′ss), 25 pmol upstream primer (P1), 25 pmol downstream primer (P5), and standard reaction conditions [200 μM each dATP, dGTP, dCTP, dTTP; 20 mM Tris-HCl, pH 8.4; 50 mM KCl; 1.5 mM $MgCl_2$; 2 U Taq DNA polymerase (GIBCO BRL)]. Thermal cycling was performed at 95 °C for 5 min, 2 times (95 °C for 45 s, 54 °C for 45 s, and 72 °C for 20 s), 25 times (95 °C for 45 s, 54 °C for 45 s, and 72 °C for 20 s), final extension for 10 min.

The selection procedure is outlined in Figure 2A. The unselected pool, pool 0, of randomized pre-mRNA used for in vitro selection was prepared by in vitro transcription of 500 ng (3.5 pmol) of the randomized DNA template using Ambion's MEGAshortscript™ kit. After transcription, the DNA template was removed by the addition of 2 U RNase-free DNase I followed by incubation for 15 min at 37 °C. The RNAs were purified in 6% polyacrylamide denaturing gels. Splicing-competent RNA molecules were selected from approximately 100 ng (1.5 pmol) of the RNA library under in vitro splicing conditions [10 mM HEPES/KOH, pH 7.5, 1.7 mM $MgCl_2$, 1 mM ATP, 5 mM creatine phosphate, 50 mM KCl, 1 mM DTT, 1 U/μL RNasin (Promega), 25% HeLa cell nuclear extract (Computer Cell Culture Belgium; Dignam et al., 1983)] in a 50-μL reaction for 45 min at 30 °C. To perform selections in Δ5′end U1-NE, all components of the standard splicing reaction, except for RNA substrate, ATP, and CP, were mixed together with 150 pmol of oligonucleotide U1H (5′-CCAGGTAAGTAT-3′; complementary to U1 snRNA nt 1–12), and 12.5 U of RNase H (Amersham Pharmacia). After preincubating the mixture for 90 min at 30 °C, RNA substrate, ATP, and CP were added to the reaction that was further incubated for 45 min at 30 °C. The splicing products were separated in denaturing gels, containing 12% polyacrylamide, 8 M urea, 0.75 mM EDTA, and 75 mM Tris-borate, pH 8.3. Splicing and denaturing gel analysis of a marker transcript of the same length and with same positions of the splicing signals identified the expected gel position of the lariat intermediates containing randomized inserts. The lariat species were eluted in elution buffer (0.3 M NaAc, pH 6.0, 1 mM EDTA). After phenol/chloroform extraction and ethanol precipitation, the lariats were debranched by incubation in a 25-μL reaction mixture, containing 20 μL of debranching buffer (20 mM HEPES/KOH, pH 7.6; 125 mM KCl; 0.5 mM $MgCl_2$; 10% glycerol; 1 mM DTT) and 2 μL of purified yeast debranching enzyme (Nam et al., 1994) for 60 min at 30 °C followed by phenol/chloroform extraction and ethanol precipitation. To add a 5′ exon to the selected sequences, the linearized lariat intermediate species were denatured at 95 °C for 30 s and incubated in a 20-μL ligation reaction [50 mM Tris-HCl, pH 7.8; 10 mM $MgCl_2$; 10 mM DTT; 1 mM AMP; 10 U T4 RNA ligase (New England BioLabs)] for approximately 16 h at 4 °C, thereby undergoing intramolecular ligation to produce circular RNA species. To amplify the selected sequences, the circularized RNAs and 1.25 pmol of primer P6 were mixed, denatured at 95 °C for 30 s, and annealed by cooling slowly to 41 °C in 7.5 μL annealing mixture (10 mM Tris/HCl, pH 6,9; 40 mM KCl; 0.5 mM EDTA). Reverse transcription was done by transferring 6 μL annealing mixture to 2 μL 5× first strand buffer (3 mM of each dNTP; 250 mM Tris HCl, pH 8.5; 40 mM $MgCl_2$; 150 mM KCl; 5 mM DTT) and 5 U AMV reverse transcriptase (Boehringer Mannheim) in a total of 10 μL followed by incubation for 1 h at 41 °C. The single-stranded DNA was amplified in a PCR reaction using primer set P3/P4 and P2/P5 after unequal and equal numbers of rounds of selection, respectively. To characterize the selected sequences, part of the DNA pool was digested with *Eco*RI and the fragment was cloned into the *Eco*RI site of pBS+ (Stratagene). Individual clones were sequenced using primer P6, and the homepage www.cbs.dtu.dk/gorodkin/appl/slogo.html was used to generate the sequence logos.

Analysis of pools of RNA and individual RNAs

To prepare library RNA for splicing analysis, DNA pools were used as templates in in vitro transcription reactions generat-

ing uniformly radioactively labeled GpppG-capped pre-mRNAs. Transcriptions were performed in 25 μL transcription mixture (40 mM Tris-HCL, pH 7.5; 10 mM $MgCl_2$; 5 mM DTT; 0.05 mg/mL BSA; 40 U RNasin (Promega); 200 ng template; 20 μCi [α-^{32}P]-UTP (Amersham: 800 Ci/mmol); 1 mM GpppG; 0.4 mM ATP; 0.4 mM CTP; 0.1 mM GTP; 0.04 mM UTP; 20 U T7 RNA polymerase (Amersham Pharmacia), for 3 h at 37 °C. After transcription, the DNA template was removed by the addition of 2 U RNase-free DNase I followed by incubation for 15 min at 37 °C. The RNAs were purified in 6% polyacrylamide denaturing gels. In splicing reactions, 20–40 fmol of pre-mRNA were incubated under in vitro splicing conditions (described above) in a 20-μL reaction for 20–90 min at 30 °C. In some reactions, 50–100 ng (700–1,400 fmol) of in vitro-transcribed pBSAd10 RNA (Garcia-Blanco et al., 1989) were included in the reaction as a competitor of splicing. Splicing complexes were analyzed by running 5 μL of the splicing reaction on native gels containing 3.95% acrylamide, 0.05% bisacrylamide, and 50 mM Tris/Glycine. The gel was run at room temperature at 15 V/cm. To analyze the extent of RNase H cleavage of U1 snRNA, total RNA was isolated from nuclear extract by incubating the extract with Proteinase K (250 ng/mL prot K (Sigma); 12.5 mM EDTA; 0.25% SDS; 12.5% NE) at 37 °C for 30 min followed by phenol/chloroform extraction and ethanol precipitation. The 5′ end of the U1 snRNA was mapped by primer extension using the primer U1RT (5′-GGAAAACCACCTTCGTGATC-3′, complementary to U1 snRNA nt 28–47). The sequencing marker was prepared by primer extension using the same primer in the presence of 0.8 mM ddATP, ddTTP, ddCTP, ddGTP, or no ddNTP. The primer extension reactions were analyzed in denaturing gels containing 6% polyacrylamide, 8 M urea, 1 mM EDTA, and 100 mM Tris-borate, pH 8.3. To map the 5′ end of U1 snRNA present in A complexes, 100 ng of nonradioactive PIP85.AΔ5′ss pre-mRNA was spliced for 90 min at 30 °C in NE or in Δ5′end U1-NE in a 100-μL splicing mixture as described above and loaded on native gels prepared as described above. A splicing reaction containing radioactively labeled PIP85.AΔ5′ss pre-mRNA was comigrated in order to identify the gel position of the nonradioactive A complexes. The RNA from the A complexes were eluted in elution buffer (0.3 M NaAc, pH 6.0; 1 mM EDTA), and after phenol/chloroform extraction and ethanol precipitation, the 5′ end of the U1 snRNA was mapped by primer extension as described above. RNA yields were quantitated by PhosphorImager.

ACKNOWLEDGMENTS

We thank Jef D. Boeke for a gift of yeast debranching enzyme, Rita Rosendahl for excellent technical assistance, and Thomas Ø. Tange and Christian K. Damgaard for discussions and critical reading of the manuscript. The work was supported, in part, by grants from the Danish Natural Science and Medical Research Councils, The Carlsberg Foundation, and The Novo Nordic Foundation. M.L. was supported by Aarhus University.

Received July 11, 2001; returned for revision August 3, 2001; revised manuscript received November 21, 2001

REFERENCES

Aebi M, Hornig H, Padgett RA, Reiser J, Weissmann C. 1986. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell 47*:555–565.

Alvi RK, Lund M, O'Keefe RT. 2001. ATP-dependent interaction of yeast U5 snRNA loop 1 with the 5′ splice site. *RNA 7*:1013–1023.

Burset M, Seledtsov IA, Solovyev VV. 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res 28*:4364–4375.

Chen JY, Stands L, Staley JP, Jackups RR Jr, Latus LJ, Chang TH. 2001. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol Cell 7*:227–232.

Cohen JB, Snow JE, Spencer SD, Levinson AD. 1994. Suppression of mammalian 5′ splice-site defects by U1 small nuclear RNAs from a distance. *Proc Natl Acad Sci USA 91*:10470–10474.

Collins CA, Guthrie C. 1999. Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes & Dev 13*:1970–1982.

Cortes JJ, Sontheimer EJ, Seiwert SD, Steitz JA. 1993. Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5′ splice sites in vivo. *EMBO J 12*:5181–5189.

Crispino JD, Blencowe BJ, Sharp PA. 1994. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science 265*:1866–1869.

Crispino JD, Sharp PA. 1995. A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes & Dev 9*:2314–2323.

Deirdre A, Scadden J, Smith CW. 1995. Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J 14*:3236–3246.

Dignam JD, Martin PL, Shastry BS, Roeder RG. 1983. Eukaryotic gene transcription with purified components. *Methods Enzymol 101*:582–598.

Du H, Rosbash M. 2001. Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA–pre-mRNA base pairing. *RNA 7*:133–142.

Frank DN, Roiha H, Guthrie C. 1994. Architecture of the U5 small nuclear RNA. *Mol Cell Biol 14*:2180–2190.

Garcia-Blanco MA, Jamison SF, Sharp PA. 1989. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes & Dev 3*:1874–1886.

Heinrichs V, Bach M, Winkelmann G, Lührmann R. 1990. U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5′ splice site. *Science 247*:69–72.

Ismaili N, Sha M, Gustafson EH, Konarska MM. 2001. The 100-kda U5 snRNP protein (hPrp28p) contacts the 5′ splice site through its ATPase site. *RNA 7*:182–193.

Jamison SF, Pasman Z, Wang J, Will C, Lührmann R, Manley JL, Garcia-Blanco MA. 1995. U1 snRNP-ASF/SF2 interaction and 5′ splice site recognition: Characterization of required elements. *Nucleic Acids 23*:3260–3267.

Kandels-Lewis S, Séraphin B. 1993. Involvement of U6 snRNA in 5′ splice site selection. *Science 262*:2035–2039.

Kim CH, J Abelson. 1996. Site-specific crosslinks of yeast U6 snRNA to the pre-mRNA near the 5′ splice site. *RNA 2*:995–1010.

Kohtz JD, Jamison SF, Will CL, Zuo P, Lührmann R, Garcia-Blanco MA, Manley JL. 1994. Protein–protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature 368*:119–124.

Konforti BB, Konarska MM. 1994. U4/U5/U6 snRNP recognizes the 5′ splice site in the absence of U2 snRNP. *Genes & Dev 8*:1962–1973.

Lesser CF, Guthrie C. 1993. Mutations in U6 snRNA that alter splice site specificity: Implications for the active site. *Science 262*:1982–1988.

Liu ZR, Sargueil B, Smith CW. 1998. Detection of a novel ATP-dependent cross-linked protein at the 5′ splice site-U1 small nuclear RNA duplex by methylene blue-mediated photo-cross-linking. *Mol Cell Biol 18*:6910–6920.

Lund M, Tange TO, Dyhr-Mikkelsen H, Hansen J, Kjems J. 2000. Characterization of human RNA splice signals by iterative functional selection of splice sites. *RNA 6*:528–544.

Luukkonen BGM, Séraphin B. 1998. Genetic interaction between U6 snRNA and the first intron nucleotide in *S. cerevisiae*. *RNA 4*:167–180.

Maroney PA, Romfo CM, Nilsen TW. 2000. Functional recognition of the 5′ splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol Cell 6*:317–328.

Moore MJ, Sharp PA. 1992. Site-specific modification of pre-mRNA: The 2′-hydroxyl groups at the splice sites. *Science 256*:992–997.

Mount SM. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res 10*:459–472.

Mount SM, Pettersson I, Hinterberger M, Karmas A, Steitz JA. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5′ splice site in vitro. *Cell 33*:509–518.

Nam K, Hudson RH, Chapman KB, Ganeshan K, Damha MJ, Boeke JD. 1994. Yeast lariat debranching enzyme. Substrate and sequence specificity. *J Biol Chem 269*:20613–20621.

Newman A, Norman C. 1991. Mutations in yeast U5 snRNA alter the specificity of 5′ splice-site cleavage. *Cell 65*:115–123.

Newman AJ, Norman C. 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell 68*:743–754.

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA. 1986. Splicing of messenger RNA precursors. *Annu Rev Biochem 55*:1119–1150.

Parker R, Siliciano PG. 1993. Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature 361*:660–662.

Reyes JL, Gustafson EH, Luo HR, Moore MJ, Konarska MM. 1999. The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5′ splice site. *RNA 5*:167–179.

Rossi F, Forne T, Antoine E, Tazi J, Brunel C, Cathala G. 1996. Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5′ splice site-U1 snRNP interaction. *J Biol Chem 271*:23985–23991.

Schneider TD, Stephens RM. 1990. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res 18*:6097–6100.

Séraphin B, Kretzner L, Rosbash M. 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.

Séraphin B, Rosbash M. 1990. Exon mutations uncouple 5′ splice site selection from U1 snRNA pairing. *Cell 63*:619–629.

Sha M, Levy T, Kois P, Konarska MM. 1998. Probing of the spliceosome with site-specifically derivatized 5′ splice site RNA oligonucleotides. *RNA 4*:1069–1082.

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. *Nucleic Acids Res 15*:7155–7174.

Siatecka M, Reyes JL, Konarska MM. 1999. Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes & Dev 13*:1983–1993.

Siliciano PG, Guthrie C. 1988. 5′ splice site selection in yeast: Genetic alterations in base-pairing with U1 reveal additional requirements. *Genes & Dev 2*:1258–1267.

Sontheimer EJ, Steitz JA. 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science 262*:1989–1996.

Staley JP, Guthrie C. 1999. An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell 3*:55–64.

Strauss EJ, Guthrie C. 1994. PRP28, a “DEAD-box” protein, is required for the first step of mRNA splicing in vitro. *Nucleic Acids Res 22*:3187–3193.

Sun JS, Manley JL. 1995. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes & Dev 9*:843–854.

Tarn WY, Steitz JA. 1994. SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes & Dev 8*:2704–2717.

Teigelkamp S, Newman AJ, Beggs JD. 1995. Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J 14*:2602–2612.

Wassarman DA, Steitz JA. 1992. Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. *Science 257*:1918–1925.

Will CL, Rumpler S, Klein GJ, van Venrooij WJ, Lührmann R. 1996. In vitro reconstitution of mammalian U1 snRNPs active in splicing: The U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res 24*:4614–4623.

Wu JY, Maniatis T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell 75*:1061–1070.

Wyatt JR, Sontheimer EJ, Steitz JA. 1992. Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. *Genes & Dev 6*:2542–2553.

Zhang D, Rosbash M. 1999. Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes & Dev 13*:581–592.

Zhang D, Abovich N, Rosbash M. 2001. A biochemical function for the Sm complex. *Mol Cell 7*:319–329.

Zhuang Y, Weiner AM. 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. *Cell 46*:827–835.