The U1 small nuclear ribonucleoprotein particle (snRNP)/5' splice site (5'SS) interaction in yeast is essential for the splicing process and depends on the formation of a short RNA duplex between the 5' arm of U1 snRNA and the 1st intronic nucleotides. This RNA/RNA interaction is characterized by the presence of a mismatch that occurs with all yeast introns and concerns nucleotides 4 on the pre-mRNA (a U) and 5 on U1 snRNA (a Ψ). The latter nucleotide is well conserved from yeast to vertebrates, but its role in yeast and the significance of the associated mismatch in the U1 snRNA/5'SS interaction have never been fully explained. We report here that the presence of this mismatch is a determinant of stability that mainly affects the off rate of the interaction. To our knowledge this is the first report assigning a function to this noncanonical interaction. We also performed SELEX (systematic evolution of ligands by exponential enrichment) experiments by immunoprecipitating U1 snRNP and the associated RNA. The artificial phylogeny derived from these experiments allows the isolation of the selective pressure due to U1 snRNP binding on the 5'SS of yeast introns.

Pre-mRNA splicing is the post-transcriptional maturation step that removes nuclear introns from primary transcripts of split genes. Introns are recognized by inspection of conserved sequences that define the sites of cleavage at the 5' and the 3' border. These sequences promote the assembly of a large ribonucleoprotein complex, called the spliceosome, that catalyzes the two transestereification steps leading to intron removal and exon joining.

Spliceosome assembly is traditionally believed to occur in a temporally ordered and sequential manner, although a more recent view holds that a preformed spliceosomal entity exists independently of the precursor RNA (1, 2). Five small nuclear ribonucleoprotein particles (snRNPs),1 each composed of one RNA (snRNA), various specific proteins, and a set of common proteins (Sm proteins), play a major role in the process. Several non-snRNP proteins also intervene in spliceosome assembly (3, 4).

Splice sites are inspected multiple times during the process, and a number of proofreading mechanisms ensure accuracy in the cleavage and ligation steps. One of the first steps is the recognition of the 5' splice site (5' SS) by the U1 snRNP particle. The 3' region of the intron is subsequently recognized by a complex formed by the branch point binding protein (yBBP/hSF1) and the associated yMud2p/hU2AF factor that, at least in metaazoans, binds to the pyrimidine-rich region downstream of the branch point (5, 6). These steps define the first complexes, which can be isolated by biochemical means, called the commitment or E complex (respectively in yeast and metazoans). The first ATP-requiring step is the formation of the pre-spliceosome. In this complex, the U2 snRNP interacts with the branch point sequence by base pairing, thereby replacing the yBBP-Mud2p complex. A number of protein factors, among them two ATPases, Sub2p and Prp5p, are involved in this step (7–10). After the interaction of the preformed tri-snRNP particle (U4/U6-U5 snRNP) the spliceosome enters a series of structural rearrangements during which the U1 and U4 snRNPs are displaced and the spliceosome is activated (11). The U6 snRNA is a focal point in these rearrangements: it replaces the U1 snRNA in a mutually exclusive interaction with the 5' splice site, while base pairing with U4 snRNA is disrupted to allow interaction with the U2 snRNP. Disruption and formation of RNA/RNA, RNA/protein, and protein/protein interactions that are often mutually exclusive is carefully controlled by a family of RNA-dependent ATPases (or RNA helicases) to ensure the folding, positioning, and activation of the catalytic center in a timely fashion (11).

Recognition of the 5' splice site is paradigmatic in this regard. In yeast, the 5'SS sequence is first recognized by base pairing with the 5' arm of U1 snRNA. Based on genetic and cross-linking studies, it is believed that U1 snRNP-associated proteins, among them U1Cp, Prp40p, Nam6p, and the Sm complex, stabilize this interaction (12–15). More recently (16) it has been shown that the protein component of U1 snRNP can recognize to some extent the 5' splice site sequence even in the absence of base pairing interaction. The U1 snRNP is displaced from the 5'SS later in the process, allowing base pairing between the U5'GU3' portion of the 5'SS and the U6 snRNA. This step is ATP-dependent and is somehow controlled by the Prp28p DEAD box helicase (17), although the exact mechanism of action of this protein is still unknown. It has been suggested that this transition at the 5'SS is operated by unwinding the U1 snRNA/5'SS duplex (17) and/or by actively displacing the

* This work was supported by the CNRS and the Fondation pour la Recherche Medicale. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed. Tel.: 33-1-69823809; Fax: 33-1-69823877; E-mail: libri@cgm.cnrs-gif.fr.

1 The abbreviations used are: snRNP, small nuclear ribonucleoprotein particle; SELEX, systematic evolution of ligands by exponential enrichment; 5' SS, 5' splice site; HA, hemagglutinin; nt, nucleotide(s); BP, branch point.

Received for publication, December 30, 2001, and in revised form, February 26, 2002 Published, JBC Papers in Press, March 4, 2002, DOI 10.1074/jbc.M112460200
U1Cp protein (18). However, it remains possible that Prp28p somehow favors more directly the association of U6 snRNA.

Finally, two additional factors have been shown by genetic and biochemical experiments to interact with the 5′SS region: the U5 snRNA and Prp8p, which interact with the last nucleotides of the 5′ exon and the first nucleotides of the intron (19–25).

The sequence of the 5′ splice site is therefore the result of multiple selective constraints, which is generally difficult to deconvolute. In yeast, the large majority of introns contain the sequence GUAGUA at the 5′ border. This sequence is only partially complementary to the 5′ sequence GUAUGU at the 5′/H11032 (=GUAU) region of the first nucleotide and is generally an A that can base pair with 5′. The conservation of U4 in the yeast 5′SS might be explained by its interaction with the U6 snRNA. However, the presence of 5′ in the U1 snRNA (and of the 5′-U4 mismatch) remains unexplained. Although it was previously suggested that it might constitute a determinant of stability (27), this hypothesis was never demonstrated. We provide here evidence that the presence of the mismatch induces a stabilization of the interaction between the U1 snRNP and the 5′SS. Surprisingly we found that the stability of the mismatch-containing complex is comparable to the stability of a complex containing a fully paired U1 snRNA/pre-mRNA duplex. By performing SELEX experiments we also obtained an artificial phylogeny, which allows the isolation of the selective constraint on the 5′SS solely due to U1 snRNP binding.

EXPERIMENTAL PROCEDURES

RNA Synthesis and Preparation of Extracts—RNA substrates for U1 snRNA binding and immunoprecipitation experiments were synthesized from PCR products carrying a T7 RNA polymerase promoter. The full sequence is shown in Fig. 1. Biotinylated BP-U and BP-A RNAs were prepared as described previously (17). Oligonucleotide 5′-U19 was purchased from Dharmacon Research and is identical to the first 11 nt of U1 snRNA. Extracts were prepared with the Umen and Guthrie method (28) with minor modifications. The U170kHA yeast strain was a gift from J. Tang. The name 5D, U170kHA strain was obtained from the U170kHA strain by deleting NAMS with a PCR-based gene disruption method (29).

Immunoprecipitation Experiments—Immunoprecipitation experiments were carried out essentially as described in Avivotch et al. (30). Briefly, 0.5–2 fmol of radioactive RNA were incubated at 25 °C in ATP-depleted extracts or U2 snRNA-inactivated extracts for the times indicated. The 30-μl reactions were diluted into 500 μl of NET100 containing 20 μl of GammaBind Plus beads (Amersham Biosciences) preincubated with anti-HA antibody (Roche Molecular Biochemicals) and incubated at 4 °C for 30 min. After extensive washings, the radioactive RNA was recovered by phenol extraction and analyzed by PAGE. Alternatively the amount of complex formed was evaluated by counting the radioactivity associated with the beads. As a control for immunoprecipitation efficiency, U1 snRNA from a fraction of the recovered RNA was primer-extended. This method was very sensitive and reliable, although the absolute amount of immunoprecipitated material depended somewhat on the particular extract preparation. In this respect, comparisons between different RNAs for U1 snRNP binding were always performed in side by side experiments.

Complex stability was assessed by adding a 100-fold excess of cold RNA after a 20-min incubation of radioactive RNA in the extract. Aliquots of the reaction were immunoprecipitated at various time points as described.

Oligonucleotide 5′-U19 was incubated for 20 min with biotinylated BP-U or BP-A in splicing salts before addition of streptavidin-agarose beads and incubation for 15 min. The retained radioactivity was measured after extensive washings.

SELEX Experiments—The starting pool of sequences was constructed by PCR as described previously (31). In the first SELEX experiment (8N selection) 8 nt were randomized 2 nt downstream of the branch point sequence. The short random window was specifically chosen to preclude selection of an aptamer to one of the U1 snRNP components. In the second experiment (4N + 3N selection), the U1 snRNP canonical binding site (GUAGUA) was included as a constant region, and 3 nt upstream and 3 nt downstream were the randomized portion. Selection was performed by incubating the pool in the tagged extract for 15 min and immunoprecipitating the complexes as described above. The process was reiterated after reverse transcription-PCR amplification and T7 transcription and stopped when the selected pools bound U1 snRNP better than the BP-U RNA (respectively six and four times). Selections were performed in side by side experiments.

A comparison between different RNAs for U1 snRNP binding were allpend somewhat on the particular extract preparation. In this respect, although the absolute amount of immunoprecipitated material decreased complementarity of the enhancer sequence with the U1 snRNA (26). This mismatch is the temperature in Kelvin.

We also obtained an artificial phylogeny, which allows the isolation of the selective constraint on the 5′SS solely due to U1 snRNP binding. The 8N selection was aligned according to two criteria: first, the GUAGUA motif that was almost universally present (only three sequences did not contain the first G); and second, complementarity to the U1 snRNA 5′ arm was used to align the remaining three sequences (sequences 3, 12, and 19) and to define the register of interaction with the U1 snRNP (and thereby the family assignment based on the identity of the 4th nucleotide).

For three sequences (sequences 5, 11, and 16) the pattern of alignment according to the U1 snRNA complementarity is ambiguous, and they could be classified in theΨ-G family provided that the GUAGUA pattern is misaligned in these sequences. One likely possibility is that even though these sequences have evolved a “double” and an “U1” U1 snRNA binding site (e.g. in group 5, U(GUAGUA) and U(GUAGUA)Au, which is likely to confer a kinetic advantage by lowering the entropic cost for binding. Interestingly six sequences of 32 winners of the 4N + 3N selection have also independently evolved a motif (GUAGUA in the 4 nt that precede the canonical GUAGUA) that could allow a double and overlapping U1 snRNP binding site.

The expected AG for the formation of the RNA duplex with the U1 snRNA arm was calculated with the help of the MFOLD server (bioinfo.math.rpi.edu/~mfold) using version 3.1 of the MFOLD program and the latest version of the free energy parameters (version 3.0) (32–34). A free energy term of +4.1 kcal/mol for duplex initiation and a penalty of +0.45 kcal/mol for terminal AU base pairs have been included in the calculation (33). In the absence of measured free energy changes, the contribution of the Ψ-U noncanonical pair has been approximated as a U-U mismatch. Kd values have been calculated with the equation: Kd = exp(AG/RT), where R is the gas constant (1.987 cal K -1 mol -1) and T is the temperature in Kelvin.

A χ2 test (3 degrees of freedom) was used to assess the statistical significance of base conservation in the 4N + 3N selection experiment. The composition of the initial pool was estimated from sequencing 15 randomly chosen unselected clones (A: 0.2; C: 0.15; U: 0.35; G: 0.30). Conservations significantly different from the composition of the initial pool (p < 0.001) are shown in uppercase and underlined in the consensus.

RESULTS

We have previously demonstrated using in vivo randomiza-
gen-selection experiments (31) that a sequence identical to the intronic portion of a 5′ splice site can act as a splicing enhancer in yeast when located immediately downstream of the branch point sequence. Our experiments also suggested that U1 snRNP was the mediator of this enhancer effect. The conservation in our selected sequences of a U at position 4 was, however, somewhat surprising. The presence of this nucleotide decreases complementarity of the enhancer sequence with the U1 snRNA. Even more puzzling was the observation that substitution of an A for Ψ (GUAGUA), which promotes uninterrupted base pairing with the 5′ arm of U1, led to almost complete loss of splicing enhancement (31).

The Presence of a Mismatch Stabilizes the Interaction of the RNA Substrate with the U1 snRNP—These experiments suggested that the presence of the Ψ-U noncanonical interaction might somehow favor the interaction of U1 snRNP with the enhancer sequence (and as a consequence, with the bona fide 5′SS). To test this hypothesis, we set up in vitro pull-down experiments with a small variety of 54-nt synthetic RNAs containing the relevant region. Splicing extracts were prepared from yeast strains containing a HA-tagged form of the U1 snRNP-associated U170K (or Prp40p, data not shown) protein.

Role for a Ψ-U Mismatch in 5′ Splice Site Recognition
containing, respectively, a U and an A at the fourth position. To estimate the amount of complex formed in the various conditions described below, we simply measured the amount of radioactive RNA retained on the beads with the U1 snRNP after incubation in extract, immunoprecipitation, and extensive washings. As before, the appropriate controls were performed with nontagged extracts and with primer extension to verify that equivalent amounts of U1 snRNA were retained on the beads. Finally, an aliquot of RNA-containing supernatant was analyzed by denaturing PAGE to ensure that the stability of BP-A and BP-U RNAs in the extract was equivalent (data not shown). Since the absolute amount of complex formation was somewhat dependent on different extract preparations, comparisons between different RNAs was always performed in side by side experiments.

Formation of the complex with U1 snRNP might occur faster for BP-U than for BP-A. Alternatively, the former complex might be intrinsically more stable than the latter. In a first attempt to answer this question, we analyzed the time course of complex formation for both substrates. At early time points both complexes were formed at equivalent rates, although in some experiments BP-A complex formation was faster (Fig. 3A and data not shown). At later time points BP-U reproducibly formed more complex than BP-A, confirming the data shown in Fig. 2.

The Presence of a BP Sequence Upstream of the 5′SS Sequence Destabilizes the Fully Paired Complex to a Larger Extent Compared with the Mismatch-containing Complex—We reasoned that the presence of a branch point sequence upstream of the U1 snRNP binding site might somehow affect complex formation or stability. This sequence is a binding site for the BBPp-Mud2p heterodimer (5, 6): the interaction of these proteins 2 nucleotides upstream of the U1 snRNP interaction site might either favor or hinder formation of the complex. In both cases, this might affect to different extents a mismatched and a fully paired U1 snRNA-RNA complex. We then constructed two additional variants of the BP-U and BP-A RNAs in which the branch point sequence was mutated (UACUAc to UACuGgC, constructs GG-U and GG-A) or deleted (Δbp-U and Δbp-A). Mutation of the two As is known to affect binding of BBPp to the branch point (5). Since the results were identical for the mutation and deletion of the branch point sequence we will only discuss the experiments performed with the former constructs. The four RNAs (BP-A, BP-U, GG-A, and GG-U) were incubated in parallel experiments with tagged and nontagged extracts and were immunoprecipitated as described above (Fig. 3B). Interestingly mutation of the BP sequence led to an increase in the amount of complex formed (compare constructs GG with constructs BP), suggesting that the branch point sequence located upstream of the 5′ splice site sequence is a destabilizing factor. Most importantly, and surprisingly, this destabilizing effect was stronger when the complex relied on a fully paired U1 snRNA/5′ splice site interaction than when the RNA/RNA interaction contained a Ψ-U mismatch. GG-A formed the highest amount of complex, but this amount was strongly reduced in the presence of a nonmutated branch point sequence (construct BP-A); on the contrary, GG-U and BP-U formed roughly the same amount of complex, leading to the observed order GG-A > GG-U ≈ BP-U > BP-A.

The Ψ-U Mismatch Affects the Stability of the Complex Even in the Absence of the BP Sequence—Whatever the mechanistic reasons for the destabilizing effect of the branch point might be, the outcome of these experiments is compatible with a simple model: once formed, the U1-RNA complex is stabilized depending on the presence of the Ψ-U mismatch in the duplex. The mismatch containing the U1-RNA complex would then be

---

**Fig. 1.** A, schematic drawing showing the RNA-RNA duplex formed between U1 snRNA and the 5′ splice site. Nucleotides are numbered according to start of transcription (U1 snRNA) or to the 5′ cleavage site. B, sequences of the various RNA transcripts used in this study. Nucleotides in the 5′SS sequence are numbered as in a *bona fide* pre-mRNA complex. Nucleotides are numbered

...
more resistant to the challenging effect of the destabilizing BP sequence.

One important implication of this hypothesis is that the higher amount of complex formed with a fully paired BP-less sequence (GG-A) should be less stable than the corresponding, mismatch-containing GG-U snRNP complex. To address this question we performed the following experiments. We incubated in parallel reactions radiolabeled GG-A and GG-U with U1-tagged extracts for 20 min to allow formation of the U1 snRNP complex. We then added a large excess of cold competitor RNA to isolate the radioactive complex and immunoprecipitated aliquots of the reactions over time to measure the decay rates of the radioactive complexes. We used cold GG-A, cold GG-U, and an equimolar mixture of the two: since the results were identical in the three cases, we only show results obtained with the latter experiment. As shown in Fig. 4, the amount of radioactive complex retained on the beads decreased faster for the fully paired GG-A/U1 snRNP complex than for GG-U snRNP, which is consistent with the latter being more stable than the former. In a second experiment, we measured differences in the decay rates of the two purified complexes in the absence of challenging cold competitor by repeatedly washing the beads and measuring the retained radioactivity over time. Although the U1 snRNP-RNA complex decayed more slowly in these conditions, the dissociation rate was again faster for GG-A than for GG-U (data not shown).

These experiments strongly suggest that the Ψ-U mismatch plays a role in the stability of the U1 snRNP-5' SS complex.

**Strengthening the U1 snRNP/BP-A Interaction Allows a Fully Paired Complex to Withstand the Destabilizing Activity of the Branch Point Sequence**—The above experiments suggest that in the presence of a fully paired interaction, a higher amount of complex can be formed (GG-A versus GG-U) that is more "vulnerable" to destabilization and generally less stable. This model predicts that strengthening the U1 snRNP/BP-A interaction should allow the complex to counteract the destabilization due to the branch point sequence. In this case, it is expected that even in the presence of a BP sequence, a hyper-stabilized fully paired duplex allows more complex formation than a mismatch-containing counterpart. We then constructed BP-A*, in which an additional base pair was added at position 7 of the donor site, and compared this sequence with BP-U* (which is identical but contains the mismatched position) in our U1 snRNP binding assay. As shown in Fig. 5A, and consistent with expectations, a higher amount of U1 snRNP-containing complex was formed with BP-A* than with BP-U*.

**The Integrity of the U1 snRNP Complex Is Required for theMismatch-dependent Stabilization of U1 snRNP-RNA Interaction**—It is possible that the stabilizing nature of the Ψ-U mismatch is linked to the presence of a protein that recognizes the
unpaired region of the RNA/RNA interaction. A number of proteins, among them Nam8p, Prp40p, U1Cp, and SmD3p, have been directly involved in the stabilization of the U1 snRNA/pre-mRNA interaction (12–15). Our attempts to repeat our in vitro U1 snRNP binding experiments in a U1Cp genetically depleted or heat-inactivated U1Cp thermosensitive strain failed because of the very low amount of complexes formed in these conditions (data not shown). Deletion of Nam8p has a no growth phenotype and only a very modest biochemical phenotype (13). We then constructed a nam8Δ/H9004, U170kHA strain and repeated our binding assays in this U1 snRNP-defective environment. Interestingly (Fig. 5B), while deletion of Nam8p decreased the amount of complex formed with both BP-U and BP-A, the U1 snRNP-BP-U complexes were affected to a larger extent compared with the U1 snRNP-BP-A complexes. Since a number of additional proteins are absent or loosely associated with the U1 snRNP complex in the absence of Nam8p (13), it cannot be concluded from this experiment that Nam8p is the stabilizing factor. However, these data strongly suggest that the integrity of the U1 snRNP complex is required for the mismatch-dependent stabilization of U1 snRNP-RNA interaction.

The Ψ-U Mismatch Destabilizes Formation of a Naked RNA-RNA Duplex—In the absence of measured thermodynamic parameters for the stability of Ψ-U mismatches in short RNA duplexes, it is possible that this mismatch is intrinsically stable in the sequence context of the U1 snRNA/5’SS interaction. To take this possibility into account, we synthesized an 11-mer RNA oligonucleotide (5’-U19023) with the same sequence as the U1 snRNA 5’ arm (containing two Ψs at positions 5 and 6). We then performed pull-down experiments with biotinylated BP-U and BP-A and 5’-radiolabeled 5’-U19023. As a control, we used a second radiolabeled oligonucleotide against a region shared by BP-U and BP-A (i.e. outside the mismatch region). As shown in Fig. 6, only in the absence of a mismatch (BP-A), 5’-U19023 was efficiently retained on streptavidin beads, while the control oligonucleotide was retained equally well by biotinylated BP-U and BP-A. This experiment shows that the Ψ-U mismatch is not intrinsically stable at least in this sequence context. Also it suggests that the protein component (or the overall integrity) of the U1 snRNP is required for the mismatch-dependent stabilization of the complex.

Sequences Forming a Mismatched Complex Are Efficiently Recovered in SELEX Experiments—The 5’ splice site sequence is recognized multiple times during the splicing process. Therefore its conservation depends on the existence of multiple, overlapping, selective pressures, one of which is related to U1 snRNP binding.

To substantiate the data presented above and to isolate the U1 snRNP interaction from the other factors that contribute to the definition of the 5’ splice site sequence, we set up a SELEX experiment. RNAs containing eight random positions were immunoselected by tagged U1 snRNP. After extensive washings and reverse transcription-PCR amplification, the procedure was reiterated for additional rounds of selection. We included
Role for a \(\Psi\)-U Mismatch in 5’ Splice Site Recognition

In this report we first used a sensitive U1 snRNP binding complex formation assay in extracts missing the U1 snRNP protein Nam8p. The absence of Nam8p has a stronger effect on the mismatch containing complex formation assay in extracts missing the U1 snRNP protein BP-A, BP-\(U^*\), and BP-A*) was performed in side by side experiments. Binding to U1 snRNP of the four RNAs (BP-U, BP-A, BP-\(U^*\), and BP-A*) was predicted for every sequence the expected change in free energy for additional base pairs to compensate for the mismatch and be able to compete with fully paired sequences. We then calculated for every sequence the expected change in free energy for the formation of the most stable RNA duplexes with U1 snRNA (the \(\Psi\)-U mismatch was assigned the same \(\Delta G\) penalty as a U-U mismatch). In striking contrast to the fact that sequences of the three classes were present at roughly equivalent frequencies, the calculated changes in \(\Delta G\) were significantly different for the U and A or G classes. Sequences of the U classes had expected \(\Delta G\) values ranging from \(-3.6\) to \(-7\) (average, \(-5.5\)) (corresponding to \(K_d\) values ranging from \(3 \times 10^{-3}\) to \(1.6 \times 10^{-5}\)), while calculated \(\Delta G\) values for the two other classes were significantly larger (averages of \(-9.6\) and \(-9.4\), respectively, for the A and G classes), and the corresponding dissociation constants were in the \(10^{-5}\)–\(10^{-8}\) range. These experiments indicate that RNA-U1 snRNP complexes containing a \(\Psi\)-U mismatch can efficiently compete with complexes containing a fully paired RNA duplex.

The SELEX experiment provided some interesting information concerning the interaction of nucleotides surrounding the GGUAGUG sequence with the 5’ arm of U1 snRNA. Notably it suggested that the interaction site on the RNA extends from nucleotides \(-2\) to \(+8\) (facing U1 nucleotides 1–10). To substantiate this result we performed a second SELEX experiment by keeping constant the GGUAGUG sequence and randomizing 4 nucleotides upstream and 3 nucleotides downstream (Fig. 6B). Analysis of the data indicate the existence of a strong selective pressure on nucleotides \(-2\) (A), \(+7\) (A), and \(+8\) (U), which define the U1 interaction sequence on the RNA as AG-GUAUGUAU.

**DISCUSSION**

The consensus sequence for the 5’ splice site of vertebrate introns contains a predominant A at position +4, which forms a base pair with the 5th nucleotide of the U1 snRNA, a pseudouridine (\(\Psi\)) (26). In contrast to vertebrates, in yeast the sequence of the 5’ splice site is almost invariant, which testifies to the existence in this organism of a strong selective pressure on the identity of the first 5 intronic nucleotides. Four of the first 5 nucleotides in yeast introns are compatible with the consensus sequence of vertebrates. The most significant difference between the two systems is the presence of a U at the fourth position in yeast introns, which precludes the possibility of forming a canonical base pair with the corresponding \(\Psi\) of yeast U1 snRNA. In the spliceosome assembly process, the U6 snRNA replaces U1 snRNA at the 5’ splice site, and the intronic UG^\(\Psi\)G\(^U\) base pairs with A^T\(G^C^A^\) of U6 snRNA. It is possible that conservation of U^\(A\) in yeast is rather relevant to the latter interaction, which would only be essential in yeast. However, the conservation of a pseudouridine at position 5 of U1 snRNA remains to be explained. Given the high free energy price paid for the maintenance of a mismatch in this region (with calculated dissociation constants that can increase up to 3 orders of magnitude, Fig. 7, compare sequences 1 and 15; see also Fig. 6) one would expect the existence of a strong selective pressure to maintain a fully paired helix, leading to the U5A mutation in yeast U1 snRNA (as, for instance, might be the case in *Euglena gracilis* (35)).

It is possible that conservation of \(\Psi\) is relevant to U1 snRNP integrity, e.g., biogenesis or stability of the particle. However, we have been unable to detect alterations in the mature levels of a mutated U1 snRNA bearing the \(\Psi^6\)A mutation.\(^2\) We favor the hypothesis that the \(\Psi\)-U mismatch plays a role in splicing, which is essentially related to the regulation of the interaction of the U1 snRNP with the 5’ splice site.

In this report we first used a sensitive U1 snRNP binding and immunoprecipitation assay to compare the U1 snRNP binding efficiencies of 5’ splice sites containing an A or a U at position 4, leading, respectively, to the formation of a fully paired or a mismatch-containing complex. The RNAs we used

---

\(^2\) D. Libri, F. Duconge, L. Levy, and M. Vinauger, unpublished results.
contain a branch point sequence upstream of the 5’ splice site sequence, which leads to the partial destabilization of the complex formed with U1 snRNP as shown by mutation of the two A residues of the BP sequence (Fig. 3). This destabilization is likely to be linked to the binding of the branch point-recognizing protein factors, which might either directly challenge the formation/stability of the 5’ splice site complex, impede the stabilizing activity of the cap binding complex, or somehow alter the structural integrity of the U1 snRNP particle. Surprisingly we found that a complex containing the presumably destabilizing -U mismatch was formed at equivalent or higher levels than a complex that contains a -A base pair at the same position. This was also confirmed by directly comparing the decay rates of the two complexes in the absence of the destabilizing BP sequence: although the absolute dissociation rate constants cannot be reliably measured by these assays, they were significantly different and higher for the complex that contains a fully paired helix. These data are also particularly significant in light of the parallel failure of biotinylated BP-U to efficiently pull down an 11-mer RNA oligonucleotide bearing the same sequence as the 5’ arm of U1 snRNP. Note that the outcome of the latter experiment is compatible with the expected 3 kcal/mol free energy change that accompanies in mismatched duplexes the disruption of two stacking interactions and the associated hydrogen bonds (A–A and A–G), which implies an increase in the dissociation constant of more than 2 orders of magnitude.

We also performed a SELEX experiment and asked whether U-containing sequences could efficiently compete with sequences that could form theoretically more stable helices. As expected, sequences containing a U-mismatched position 4 were efficiently recovered in this experiment.

Interestingly no sequences containing a C in this position were selected, which strongly suggests that U and C mismatches are not equivalent in this context. This result, which was also directly confirmed by immunoselection assays with individual sequences containing a C at position 4 (data not shown), raises the question of why C is the second most frequent nucleotide in natural 5’ splice sites, whileGs are never found (see below).

The outcome of the SELEX experiment might be considered somewhat different from our pull-down assays in that sequences with A (and G) at the mismatch position were not out-competed by sequences containing Us. This is likely due to the fact that the selection protocol is not able to discriminate between complexes with different off rates but similar on rates. It is possible that the U-dependent stabilization follows duplex formation that is expected to be faster for fully complementary helices (which we observed in some experiments, data not shown). Since the U1 snRNP target is limiting, the composition of the winning pool might essentially reflect the competition between sequences for target accession. In the absence of such a competition (i.e. in pull-down experiments with individual sequences) the differences in the off rates would become apparent.

We favor the hypothesis that a protein factor plays an important role in “locking” the U1 snRNP/RNA interaction by recognizing the helix distortion induced by the unpaired nucleotides. The observation that the absence of Nam8p has a more dramatic effect on a mismatch-containing than on a fully paired complex is consistent with this hypothesis, although it does not directly implicate Nam8p in the mismatch-dependent
stabilization. The absence of Nam8p has in fact been reported (36) to loosen the association of other protein factors with the U1 snRNP, like Snu56p, Snu71p, and Prp40p, which might be involved indirectly or directly (as suggested for Prp40p (12)) in the process. U1Cp has also been implicated in stabilizing the U1 snRNP/pre-mRNA interaction in vertebrates (37) and shown to cross-link, albeit weakly, to U4 in yeast introns (14). More recently Chen et al. (18) reported the isolation of multiple U1C alleles (mutation of the same amino acid, Leu15H) that allow bypass of the requirement for Prp28p, a DEAD box ATPase implicated in the dissociation of U1 snRNP from the 5′ splice site. Finally, recombinant U1Cp was shown to directly bypass the requirement for Prp28p, a DEAD box ATPase implicated in the dissociation of U1 snRNP from the 5′ splice site.}

Although these results suggest a role for U1Cp in the recognition of the mismatch, it has to be stressed that they do not necessarily implicate a direct role in the mismatch-dependent stabilization of the interaction. All our attempts to demonstrate a role for this protein using U1Cp thermosensitive extracts failed, essentially due to the very low amount of complexes formed with our RNA substrates.

Staley and Guthrie (17) have recently shown that artificially extending the base pairing between U1 snRNA and the 5′SS decreases the efficiency of U1 snRNP displacement and inhibits splicing at low temperatures. It is possible that a U1 snRNA containing an A in the fifth position (and therefore able to form a fully paired duplex) might not dissociate efficiently from the 5′SS, which would be the rate-limiting step for growth. In apparent accordance with this hypothesis is the observed cold sensitivity of a U1 ¥5A mutant strain.2 However, we have shown4 that the cold-sensitive phenotype of this strain is mainly due to the defective splicing of one single intron contained in a quasi-essential gene that we temporarily called LMD1. Replacement of the gene with its cDNA copy efficiently suppresses the cold sensitivity of the strain, indicating that the growth defect does not reside in a slow U1 snRNP dissociation step from one or more introns containing canonical 5′ splice sites. Rather it is likely that binding to the LMD1 intron is limiting. In fact, the 5′SS of LMD1 contains an unusual A at position 5 that leads to the formation of an A-A mismatch with the mutant U1 snRNA and to inefficient splicing of this particular intron.

It has to be stressed, however, that the role played by the ¥-U mismatch in the stabilization of the U1 snRNP/5′SS interaction does not preclude an additional requirement for an efficient dissociation step. Indeed our favorite hypothesis is that its presence is required to “transfer” (or most) of the interaction energy of the U1 snRNP/5′SS complex to one (or more) protein factors. Modulation of the activities of the latter by phosphorylation or the action of other proteins during the splicing cycle would in turn favor association or dissociation of the particle from the 5′SS. This strategy might be more economical in that it would still exploit the RNA moiety for accurate positioning but would rely on the protein component for modulation of the affinity, which is presumably more efficient than disrupting a fully paired RNA duplex.

Why is the ¥-U mismatch not conserved in vertebrates where the majority of 5′ splice sites contain an A at the fourth position? One likely explanation might be that the yeast U1 snRNP contains a set of specific proteins that are not associated with the vertebrate particle (36). Interestingly at least one of these proteins (Nam8p) has a vertebrate homologue (38) that is not part of the U1 snRNP. It is likely that protein-mediated stabilization of the U1 snRNP/5′SS interaction in vertebrates is not universally mediated by a U1 snRNP component but occurs through the intervention of a set of specific, non-U1 snRNP-associated factors (e.g. SR proteins) playing major roles both in the selection of the less well defined 5′SS and the regulation of its use (for instance, see Refs. 39 and 40).

The 5′SS in yeast introns is inspected multiple times during the splicing process. Its sequence is therefore the result of several selective constraints, only one of which is U1 snRNP binding. Although a number of studies provided essentially genetic evidence for the interaction between individual nucleotides of the U1 snRNA 5′ arm and the 5′SS (27, 41–43), there is very limited biochemical evidence that directly addresses the question of how many intronic nucleotides can actually be inspected by U1 snRNP. The outcome of our SELEX experiments allows the isolation of the latter evolutionary constraint by providing an artificial phylogeny solely based on U1 snRNP binding. Incidentally the strategy we used, i.e. selection based on complex pull-down assays, might be generally more inform-
ative than classical SELEX to identify protein binding sites on the RNA. Contrary to selections based on binding to single polypeptides, complex pull-down selection is more likely to reveal physiological situations when the protein of interest interacts with a complex on the RNA. In this case, indeed, the protein-RNA-interacting surface might be constituted by more than one polypeptide that potentially contributes to the affinity and/or the specificity of the interaction.

The consensus sequence of our artificial phylogeny parallels data issued from statistical analysis of the splice sites of (almost) all yeast introns. In complete accordance with the latter is conservation of nucleotides −2 (A), 1–3 (GUA), and 5 and 6 (GU). Nucleotide G1 is strongly conserved in our analysis but is conservation of nucleotides most) all yeast introns. In complete accordance with the latter than one polypeptide that potentially contributes to the affinity protein/RNA-interacting surface might be constituted by more

---

**REFERENCES**

1. Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D., and Abelson, J. (2002) *Mol. Cell* 9, 31–44

2. Nilsen, T. W. (2002) *Mol. Cell* 9, 8–9

3. Moore, M. J., Query, C. C., and Sharp, P. A. (1993) in *The RNA World* (Gesteland, R. F., and Atkins, J. F., eds) pp. 303–357, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

4. Burge, C. B., Tuschi, T. H., and Sharp, P. A. (1998) in *RNA World II* (R. P. Gesteland, T. R. Cech, and Atkins, J. F., eds) pp. 525–560, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

5. Berglund, J. A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997) *Cell* 89, 781–787

6. Berglund, J. A., Abovich, N., and Rosbash, M. (1998) *Genes Dev.* 12, 858–867

7. Kistler, A. L., and Guthrie, C. (2001) *Genes Dev.* 15, 42–49

8. Libri, D., Graziuni, N., Zapp, C., and Galand, J. (2001) *Genes Dev.* 15, 36–41

9. O’Day, C. L., Dalbande-McFarland, G., and Abelson, J. (1996) *J. Biol. Chem.* 271, 35261–35267

10. Zheng, M., and Green, M. R. (2001) *Genes Dev.* 15, 30–35

11. Staley, J. P., and Guthrie, C. (1998) *Cell* 92, 315–326

12. Kao, H. Y., and Siliciano, P. G. (1996) *Mol. Cell. Biol.* 16, 960–967

13. Puig, O., Gottschalk, A., Fabrizio, P., and Seraphin, B. (1999) *Genes Dev.* 13, 569–580

14. Zhang, D., and Rosbash, M. (1999) *Genes Dev.* 13, 581–592

15. Zhang, D., Abovich, N., and Rosbash, M. (2001) *Mol. Cell* 7, 319–329

16. Du, H., and Rosbash, M. (2001) *RNA* 7, 153–162

17. Staley, J. P., and Guthrie, C. (1999) *Mol. Cell* 3, 55–64

18. Chen, J. Y., Stands, L., Staley, J. P., Jackups, R. R., Jr., Latus, L. J., and Chang, T. H. (2001) *Mol. Cell* 7, 227–232

19. Siestacka, M., Reyes, J. L., and Kenarska, M. M. (1999) *Genes Dev.* 13, 1983–1993

20. Collins, C. A., and Guthrie, C. (1999) *Genes Dev.* 13, 1970–1982

21. Rose, J. L., Kriss, P., Konforti, B. B., and Kenarska, M. M. (1996) *RNA* 2, 213–225

22. Teigelkamp, S., Newman, A. J., and Begg, J. D. (1995) *EMBO J.* 14, 2692–2612

23. McConnell, T. S., and Steitz, J. A. (2001) *EMBO J.* 20, 3577–3586

24. Wyatt, J. R., Sontheimer, E. J., and Steitz, J. A. (1992) *Genes Dev.* 6, 2542–2553

25. Newman, A. J., and Norman, C. (1992) *Cell* 68, 743–754

26. Massenet, S., Motorin, Y., Lafontaine, D. L., Hurt, E. C., Grosjean, H., and Branlant, C. (1999) *Mol. Cell. Biol.* 19, 2142–2154

27. Seraphin, B., and Rosbash, M. (1989) *Genes Dev.* 11, 145–151

28. Umen, J. G., and Guthrie, C. (1995) *Genes Dev.* 9, 855–868

29. Wach, A., Brachat, A., Pohlmann, R., and Philippens, P. (1994) *Yeast* 10, 1793–1808

30. Abovich, N., Liao, X. C., and Rosbash, M. (1994) *Genes Dev.* 8, 843–854

31. Libri, D., Lesure, A., and Rosbash, M. (2000) *RNA* 6, 352–368

32. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol.* 290, 911–940

33. Xia, T., Santa Lucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) *Biochemistry* 37, 14719–14735

34. Zuker, M., Mathews, D. H., and Turner, D. H. (1997) *Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Bio/Technology*, NATO ASI Series (Barciszewski, J., and Clark, B. F., C. eds) pp. 11–43, Kluwer Academic Publishers, Dordrecht, The Netherlands

35. Breckenridge, D. G., Watanabe, Y., Greenwood, S. J., Gray, M. W., and Schnare, M. N. (1999) *Proc. Natl. Acad. Sci. U. S. A.* 96, 852–856

36. Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neuhauser, G., Colot, H. V., Mann, M., Seraphin, B., Rosbash, M., Luhrmann, R., and Fabrizio, P. (1998) *RNA* 4, 374–393

37. Heinrichs, V., Bach, M., Winkelmann, G., and Luhrmann, R. (1990) *Science* 247, 69–72

38. Forch, P., Puig, O., Kedersha, N., Martinez, C., Granneman, S., Seraphin, B., Anderson, P., and Valcarcel, J. (2000) *Mol. Cell* 6, 1089–1098

39. Jamison, S. F., Passman, Z., Wang, J., Will, C., Luhrmann, R., Manley, J. L., and Garcia-Blanco, M. A. (1995) *Nucleic Acids Res.* 23, 3260–3267

40. Kozhit, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., and Manley, J. L. (1994) *Nature* 368, 119–124

41. Seraphin, B., Kretzner, L., and Rosbash, M. (1988) *EMBO J.* 7, 2533–2538

42. Seraphin, B., and Kandels-Lewis, S. (1993) *Cell* 73, 803–812

43. Siliciano, P. G., and Guthrie, C. (1988) *Genes Dev.* 2, 1258–1267

44. Long, M., de Souza, S. J., and Gilbert, W. (1997) *Cell* 81, 739–740

45. Lopez, P. J., and Seraphin, B. (1999) *RNA* 5, 1135–1137

46. Kretzner, L., Krol, A., and Rosbash, M. (1999) *Proc. Natl. Acad. Sci. U. S. A.* 97, 851–855
A Role for the Ψ-U Mismatch in the Recognition of the 5′ Splice Site of Yeast Introns by the U1 Small Nuclear Ribonucleoprotein Particle
Domenico Libri, Frédéric Ducongé, Laurence Levy and Marion Vinauger

J. Biol. Chem. 2002, 277:18173-18181.
doi: 10.1074/jbc.M112460200 originally published online March 4, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M112460200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 43 references, 23 of which can be accessed free at http://www.jbc.org/content/277/20/18173.full.html#ref-list-1