2009

The Role and Fate of Branch Site-U2 snRNA Interaction During Pre-mRNA Splicing

Duncan Smith

The Role and Fate of Branch Site-U2 snRNA Interaction During Pre-mRNA Splicing

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Duncan Smith

June 2009

The Role and Fate of Branch Site-U2 snRNA Interaction During Pre-mRNA Splicing

Duncan Smith, Ph.D.

The Rockefeller University 2009

Many details of the conformational dynamics underlying transitions during pre-mRNA splicing, particularly during the catalytic phase, remain unclear despite our deep knowledge of spliceosome composition. The complex assembly phase that precedes catalysis, and the involvement of several motifs in multiple interactions along the splicing pathway, exacerbate the difficulty of accessing splicing catalysis experimentally.

snRNAs likely form some or all of the spliceosomal active site. We attempted to establish an snRNA-only splicing system as a means to bypass assembly and recapitulate splicing catalysis. Our approach, however, yielded a 2'-3' bond-forming ribozyme activity in a linker region in our fused snRNA construct; this highlights the uncertainty inherent in minimised systems and, by contrast, the strength of experimental approaches based on complete spliceosomes.

We describe the *in vivo* analysis of the interaction between the intron branch site (BS) and its cognate region of U2 snRNA; the absolute requirement for this base-pairing interaction during spliceosome assembly has previously impeded the investigation of its role and fate during splicing catalysis. By substituting the entire duplex, we produced nonessential second-copy spliceosomes in which assembly and catalysis each partially limit the splicing reaction. Using such spliceosomes, we established unconfirmed details of branch site-U2 interaction, including a requirement for this duplex in bulging the branch nucleophile for first step splicing catalysis, and similar spatial organisation of key components of the spliceosome and group II self-splicing introns. Our second-copy spliceosomes should represent a system that can be extended to include other spliceosomal components, allowing the *in vivo* investigation of normally lethal, and even dominant lethal, alleles.

In another line of experiments, we used RT-PCR to identify the products of reactions in which multiple sites in the BS-binding region of U2 were used in *trans* as 5' splice sites during the splicing of a reporter gene. Characterisation of *trans*-splicing products revealed previously unknown dynamics in BS-U2 pairing, suggesting that this interaction may be disrupted between early assembly and first step catalysis, and again before the second step of the reaction. Our results also indicate a surprising degree of flexibility in the active site of the normally stringent *S. cerevisiae* spliceosome.

ACKNOWLEDGEMENTS

First and foremost, thanks to Magda Konarska and Charles Query for their constant faith and support in all my endeavours, scientific and otherwise, and for providing a constantly stimulating and challenging environment. I also wish to thank Tom Tuschl for excellent advice, particularly relating to snRNA reactivity, and for serving as the chairman of my thesis committee. I would like to express my gratitude to Fred Cross and Mike O'Donnell, who also served on my thesis committee and provided excellent insight and perspective, and to Anna Pyle for serving as my external examiner. Thanks to past and present members of the Konarska lab for technical and intellectual support, to Stew Barnes and Lola McRae for I.T. assistance, and to the Rockefeller University for their generous financial and administrative support. Finally, to family and friends who have helped me in ways too numerous to detail, thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

3'SS	3' splice site
5'SS	5' splice site
BS	Branch site
CTD	C-terminal domain
ISL	Intramolecular stem loop
LI	Lariat intermediate
NTC	Prp19 complex (a.k.a. CDC5L complex)
nt	Nucleotide
SELEX	Systematic evolution of ligands by exponential enrichment
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein particle
TBDMS	t-butyl dimethylsiloxy
TLC	Thin layer chromatography
TPP	Thiamine pyrophosphate
Ψ	Pseudouridine
U26	Fused U2/U6 snRNA construct
wt	Wild-type

CHAPTER 1. INTRODUCTION.

A number of reviews have been published over the past decade (1-11), addressing numerous aspects of pre-mRNA splicing and its coordination with other nuclear events. This introduction will focus on selected themes, highlighting the dynamic nature of both the assembly and catalytic phases. Multiple transitions in spliceosome assembly and catalysis can be modulated by alterations in the identity or activity of spliceosomal components, or by modulation of the stability of interactions between the pre-mRNA substrate and the spliceosome. For a given intron and set of cellular conditions, one or more of these transitions will limit splicing and thus be available as a potential point of regulation. Changes in the efficiency of any transition in the entire splicing pathway can therefore result in regulated – and thus alternative – splicing.

THE DYNAMICS OF SPLICEOSOME ASSEMBLY

The classical, sequential view of spliceosome assembly (2), shown schematically in Fig. 1-1, holds that the 5' splice site (5'SS) is first bound by U1 snRNP, then the branch site (BS) and 3'SS by U2 snRNP and associated protein factors to form a pre-spliceosome, also known as complex A. The [U4/U6.U5] tri-snRNP joins the complex to form complex B, and a series of conformational and compositional changes result, including the loss of U1 and U4 snRNPs to leave U2/5/6. Recruitment of the CDC5L complex (known as the NTC in *S. cerevisiae* – this nomenclature will be used throughout) follows (12), giving rise to an

activated spliceosome. Assembly can be stimulated or repressed by the binding of general or specific splicing factors to snRNPs and pre-mRNA; snRNPs can also interact both with pre-mRNA and with each other. Spliceosome assembly is thus highly cooperative, and the fact that many interactions can occur independently of one another in fact results in an assembly cascade that does not follow a single obligatory trajectory, but instead can occur via multiple pathways.

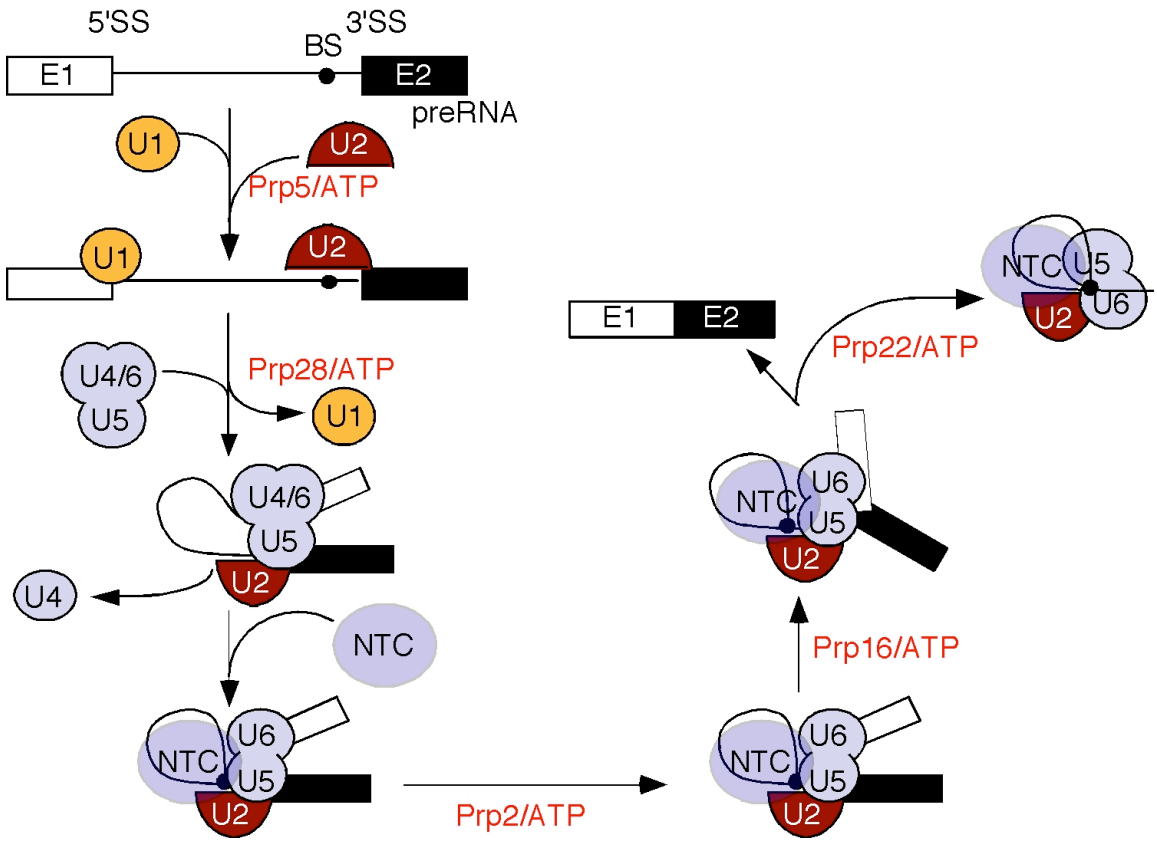


Figure 1-1. The classical, sequential view of spliceosome assembly.

An extreme example of cooperative assembly is the penta-snRNP, which contains all five snRNPs in the absence of pre-mRNA and can be isolated from

S. cerevisiae. Although the penta-snRNP alone is not competent to catalyse splicing, addition of nuclease-treated cell extract restores its activity without requiring disassembly and reassembly (13). Together with the observation of splicing in the absence of U1 recruitment in multiple systems (2), this finding suggests that there is unlikely to be any essential compositionally defined assembly intermediate prior to the activated spliceosome. Rather, a number of potential assembly pathways can lead to this same point. For a given substrate, the stabilities of intermediates along a given pathway might be such that this mode of assembly is observed to the exclusion of all others. In such cases, stable intermediates may be isolated (14), but this does not mean that the assembly pathway in question is an obligate one for any substrate.

Spliceosome assembly appears to be driven by the stochastic association of snRNPs with pre-mRNA. Data from photobleaching experiments using various GFP-tagged spliceosomal components are consistent with free diffusion of snRNPs within the nucleus and their transient, random association with pre-mRNA (15). Two major roles of factors stimulating or repressing assembly in such a system would be to increase or decrease the local concentration of snRNPs on or near a transcript, and to modulate the stability of snRNP-pre-mRNA and inter-snRNP interactions. Given that the CTD of RNA polymerase II is one of the multiple regulators that can interact with snRNPs to modulate assembly (3), future studies must address how assembly is modulated in response to transcriptional events and chromatin structure. In addition, it is

possible that future work will identify more entry points into the spliceosome assembly pathway – for example, “re-initiation” by postcatalytic U2/5/6 complexes.

Another convergence point during spliceosome assembly exists due to the two possible orientations of interaction between the U1 and U2 snRNPs. U1-U2 interaction can occur across the intron to be removed by the spliceosome that will ultimately contain these two snRNPs (intron definition); in multi-intron genes, however, this interaction can also occur across an exon such that U1 bound at the 5’SS of the downstream intron interacts with U2 bound to the BS-3’SS of the upstream intron (exon definition). A reporter gene with long exons is spliced efficiently only if the flanking introns are short (16), and increasing intron length in a model substrate favours exon definition (17), consistent with the interpretation that simple binding kinetics determine the predominant assembly pathway for a given pre-mRNA substrate. It is not known whether multiple U1 snRNPs can interact with a single U2, or vice versa. If such polymeric interactions are not possible, inter-snRNP interactions promoting exon definition at a given splice site will preclude intron definition. An exon-defined assembly intermediate must therefore make the transition to an intron-defined state in order for functional splicing to proceed. Indeed, splicing can be inhibited by preventing the establishment of intron definition, for example by hnRNP L binding to an exonic splicing silencer and stabilising exon definition interactions between U1 and U2 (18). hnRNP L has a stimulatory effect when bound within an intron, likely due to

facilitation of a cross-intron U1-U2 interaction (19). Similarly, the binding of SR proteins within an intron can inhibit splicing (20), whereas SR binding within exons is generally stimulatory (4). These observations are consistent with the existence of mutually exclusive interactions during exon- and intron-defined states.

Circular exons, the predicted products of splicing from an exon-defined state, have been detected in several systems (21-23). The formation of such products, even at a slow rate, suggests that the maintenance of U1-U2 interaction across an exon (i.e. with the wrong polarity) is not sufficient to prevent the formation of a catalytically competent spliceosome. It is therefore likely that a polarity sensing mechanism normally exists to distinguish between exon- and intron-defined complexes. U1-U2 interactions, when intron-defined, may provide a binding surface for the [U4/U6.U5] tri-snRNP, with exon-defined complexes normally lacking such a surface. The recruitment of non-tri-snRNP proteins is also likely to play a role in this transition, with candidates for proteins involved in polarity sensing expected to bind only to intron-defined complexes. Although no mechanism or factors responsible for such polarity sensing are known, DEK, a chromatin-associated protein not required for early assembly but important for 3'SS definition and splicing catalysis (24), could represent one such factor. Proteomic analysis of human spliceosome assembly intermediates suggests that DEK joins the complex during or after the exon definition-intron definition

transition (25). In addition, there is no DEK homologue in *S. cerevisiae*, whose almost exclusively single-intron genes presumably lack an exon-defined stage.

THE ROLE OF RNA STRUCTURE

Consistent with many alternative splicing factors having a role in increasing or decreasing the local concentration of snRNPs on transcripts, a large number of sequence-specific RNA binding proteins have been shown to modulate spliceosome assembly (26). Several protein motifs that bind single stranded RNA have been characterised, and these are commonly found in splicing factors. Consistent with their action in a single-stranded state, a set of splicing enhancers and silencers has been confirmed bioinformatically to be more single stranded than bulk sequence, and to function more effectively when placed in the loop than the stem of a hairpin structure (27). The stabilisation or destabilisation of secondary structure around such regulatory elements could therefore serve as a potential mechanism to reduce or enhance, respectively, their effects on splicing.

All evidence suggests that splice sites themselves must be single stranded in order to allow spliceosome assembly, with secondary structure inhibiting U1 and U2 snRNP binding. Inclusion of the 3'SS in a hairpin is inhibitory for splicing, although this can be overcome by the presence of a single stranded 'helper' downstream 3'SS, likely recognised during assembly (28). A particularly elegant example of alternative splicing regulation by direct modulation of secondary structure around splice sites is the control of alternative splicing by a thiamine

pyrophosphate (TPP) riboswitch in *Neurospora crassa* (29). When TPP concentration is low, the pre-mRNA adopts a structure such that an otherwise favoured downstream 5'SS is occluded, and the branch region is flexible. Splicing proceeds using a suboptimal upstream 5'SS to produce mRNA encoding a functional NMT1 protein. When TPP concentration is high, however, conformational changes in the riboswitch cause structure around the favoured downstream 5'SS to be disrupted, leading to the predominant production of a longer mRNA containing uORFs that prevent NMT1 translation; in addition, the branch region is partially occluded, yielding an overall decrease in splicing efficiency. It is likely that similar examples of splicing regulation, mediated by proteins or small molecules, will be discovered in other systems: how common such mechanisms of splicing regulation will prove to be remains an open question.

Secondary structure is not always inhibitory to splicing; for example, the *S. cerevisiae RP51B* intron contains complementary sequences close to the 5'SS and BS that bring the ends of the intron together and aid spliceosome assembly (30), and it is possible that this is a common way to increase the efficiency of U1-U2 binding and intron definition. The splicing of exon clusters 4 and 6 in the well-characterised *Dscam* gene in *Drosophila* provides two further examples of stimulatory secondary structures. Disruption of the iStem – a large hairpin loop downstream of exon 3 – interferes with the splicing of all twelve exon 4 variants (31), although the mechanism by which this stem stimulates exon 4 splicing

remains unclear. The basis of maintenance of mutually exclusive splicing in the exon 6 cluster, however, is better understood. Each exon 6 variant is preceded by a selector sequence complementary to a docking site downstream of exon 5. Interaction between a given selector sequence and the docking site leads to splicing of the following exon and, as the docking site is thus removed from the transcript, the inclusion of further exon 6 variants is suppressed under normal conditions (32). Knockdown of the hnRNP protein hrp36 leads to the inclusion of multiple exon 6 variants, suggesting that this protein mediates the repression of splicing across the cluster (33).

hnRNP proteins normally act as general inhibitors of splicing: they are antagonised by the generally activating SR proteins (4). There is increasing evidence that SR proteins exert at least some of their stimulatory effect via the stabilisation of RNA-RNA interactions during both spliceosome assembly and splicing catalysis. A pre-mRNA with a 5' exon as short as one nucleotide can undergo SR protein-dependent splicing in HeLa extract, suggesting a post-assembly role for these proteins (34). The arginine-serine rich (RS) domain of a natural SR protein, or an artificial domain comprising multiple RS repeats, when tethered to pre-mRNA, directly contacts the branch site and facilitates prespliceosome formation (35); it is thought that the BS is already base paired to U2 snRNA in such assembly intermediates (36). Defects due to SR protein depletion can be suppressed by increasing the strength of the interaction between the 5'SS and U6 snRNA that is required for the first catalytic step (37,

38). Subsequent work has demonstrated that an RS domain can be crosslinked to the intronic 5'SS region bound to U6 snRNA during the first step of splicing, and subsequently to the exonic 5'SS region bound to U5 snRNA during the second step (39). This argues for the involvement of enhancer-recruited SR proteins not only in assembly but also during the catalytic phase of splicing.

Data concerning the role of SR proteins in alternative splicing have generally been interpreted with assembly in mind: such data might now need to be reconsidered, as the stabilisation of duplexes could produce diverse phenotypes during the dynamic and structurally complex catalytic phase. It is also possible that hnRNP proteins might exert some of their effect via the disruption of RNA-RNA interactions – most simply by sequence sequestration. The mechanistic basis for duplex stabilisation by SR proteins remains unclear, as does the issue of whether this stabilisation is general or protein-duplex specific, with defined SR proteins stabilising only certain duplexes.

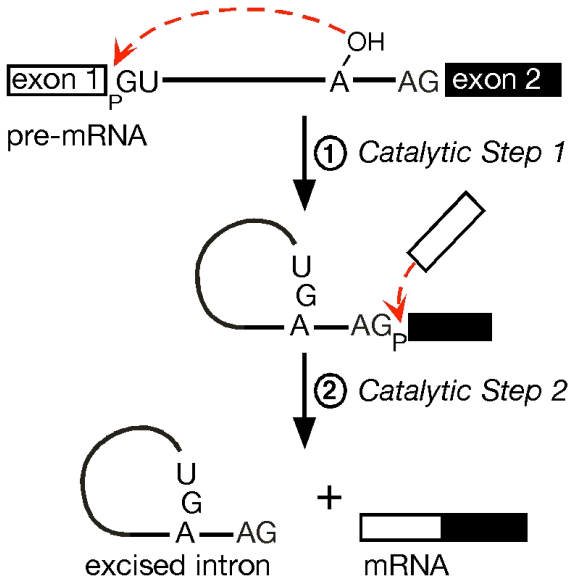


Figure 1-2. The two steps of splicing catalysis.

THE TRANSITION BETWEEN THE TWO CHEMICAL STEPS

Splicing catalysis consists of two successive transesterification reactions: in the first step, the 2' hydroxyl of the BS nucleotide nucleophilically attacks the 5'SS to yield a lariat intermediate and a free 5' exon; in the second step, this free exon nucleophilically attacks the 3'SS, producing mRNA and an excised lariat intron (Fig. 1-2). The 3'SS remains sensitive to nuclease degradation until after the first step; this suggests that it enters the active site after first step catalysis (40) – a repositioning event that would require removal of the newly formed branch structure of the lariat intermediate from the catalytic centre. Indeed, multiple lines of evidence suggest that the 3'SS replaces the branch structure, with the 5' exon remaining in a fixed position relative to loop 1 of U5 snRNA. Crosslinks between U5 loop 1 and the terminal nucleotide of the 5' exon can be chased

through both steps of splicing: those between loop 1 and position +2 of the intron, however, can be chased into a lariat intermediate but not a lariat product (41), suggesting disruption of U5-intron 5'SS interaction following the first step and consistent with genetic and biochemical interactions between loop 1 and the 3'SS during the second step (42) and references therein). Further crosslinking studies show juxtaposition of the 5'SS and U6 snRNA positions U47-A51 during the first step (proposed RNA-RNA interactions in the first step catalytic centre are shown in Fig. 1-3), and positions G39-A44 when in the lariat intermediate branch structure (43). Genetic evidence also supports the disruption of the 5'SS-U6 pairing required for the first step of splicing: the 5'SS A3C mutation (/GUAUGU to /GUCUGU) hyperstabilises pairing between the 5'SS and U6; this mutation inhibits the second step, but can be suppressed by mutations that destabilise 5'SS-U6 pairing, presumably restoring the duplex to near wild-type stability (44).

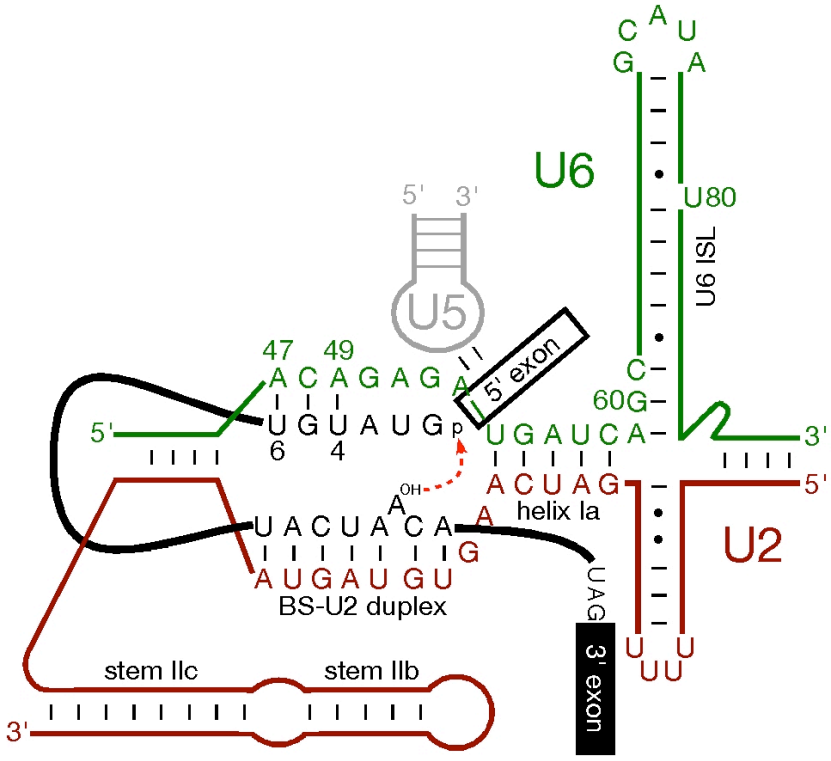


Figure 1-3. Schematic of RNA:RNA interactions during the first step of splicing.

Pre-mRNA is shown in black, U2 snRNA in red, U5 snRNA in grey, and U6 snRNA in green; numbering corresponds to *S. cerevisiae* snRNAs and indicates U6 nucleotides discussed in this review. Nucleophilic attack of the 5'SS by BS (the first catalytic step of the splicing reaction) is shown.

The splicing defect due to a hyperstabilised 5'SS-U6 helix can be suppressed not only directly by duplex destabilisation, but also by the U6 U57A mutation, a variety of mutations in Prp8 (44) or deletion of the NTC component Isy1 (45). Mutations that suppress the second step defect of an A3C intron also increase the efficiency of the second step for a variety of other intron mutations, including those at the branch site adenosine and 3'SS; commensurate with their stimulation of the second step, these suppressors inhibit the first step of splicing. There also exists the opposite class of spliceosomal mutants: those that increase

the efficiency of the first step at the expense of the second (46). The existence of two opposing classes of suppressor allele, each capable of suppressing a wide range of intron mutations, suggests that suppression is not necessarily via direct contact between mutated bases/amino acids, but that the spliceosomal conformations competent to carry out the first and second steps are in competition with one another. Mutations that stabilise the first step conformation relative to the second will stimulate the first step and inhibit the second, while the opposite will be true for those that cause relative stabilisation of the second step conformation. As a mutation is more likely to disrupt an interaction than to form a new one, it is likely that relative stabilisation takes the form of destabilisation of the competing conformation, such that most first step suppressors would destabilise the second step conformation and vice versa. At present, however, the molecular basis of the action of these general suppressor mutations is unknown. This two-state model provides a mechanism by which modulation of the stabilities of conformational states of the catalytic spliceosome can modulate the efficiency of splicing of suboptimal substrates. Such tolerance of suboptimal splice sites is manifested in metazoa as alternative splicing, meaning that local or global modulation of conformational stability during catalysis could impact the splicing pattern of individual transcripts, or of classes thereof, respectively.

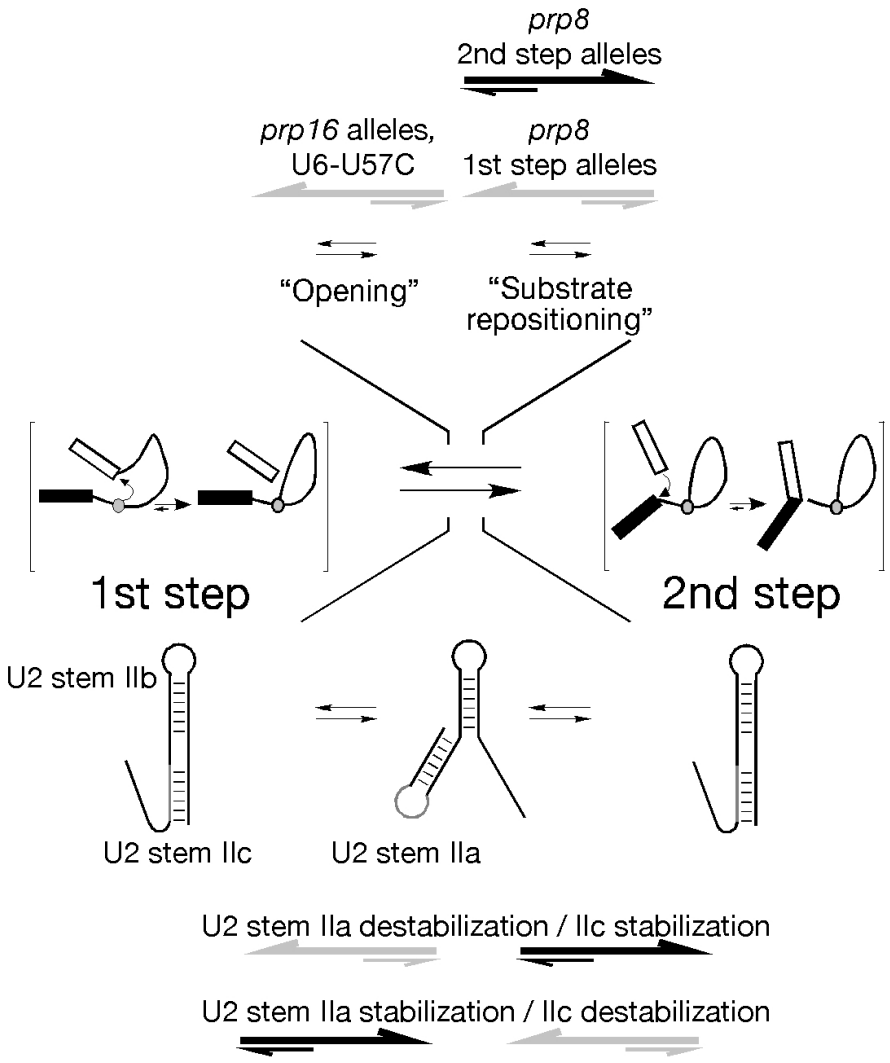


Figure 1-4. Two-state conformational model of the catalytic spliceosome.

Spliceosomal conformations competent to carry out the first and second steps are in competition with one another. In addition, two opposing classes of *prp8* alleles modulate an event in the first-to-second step transition distinct from that modulated by alleles of *prp16* and U6 snRNA (upper) (47); thus, two distinct events ('opening' and 'substrate repositioning') can be distinguished. Accompanying conformational changes in U2 snRNA stem II (lower) parallel these two events (48, 49).

Recent detailed analysis of the interplay between global suppressor mutations has led to a refinement of the two-state model (Fig. 1-4). First and second step suppressor point mutations combined in the same molecule of Prp8, a large and

exceptionally well-conserved U5 snRNP protein that makes contacts with the 5'SS, BS and 3'SS (50), cancel one another's effects and produce a phenotype resembling wild-type Prp8. A different effect is observed, however, when Prp8 second step suppressor mutations are combined with first step suppressors in U6 snRNA or Prp16, the ATPase that modulates the first-to-second step transition (51); in this instance, the suppressors act in concert such that both the first and second steps are improved. Cancellation by the opposing classes of *prp8* allele indicates that these *prp8* alleles act at the same kinetic step as one another, whereas the additive nature of the Prp16/U6-Prp8 suppressor pairs requires that they be affecting different kinetic steps in the transition. These observations have led to a model in which the transition between the two steps has multiple phases, requiring an 'opening' step (affected by *prp16* and U6 mutants) and a repositioning step (affected by *prp8* mutants), followed by 'closure' into the second step conformation (47).

The suggestion that transitions traditionally considered as one step actually comprise multiple phases has important implications for future proteomic, biochemical and structural studies of the spliceosome. A complex containing a lariat intermediate and a free 5' exon could plausibly be in one of at least four conformations that are currently compositionally and conformationally undefined: post-first step but pre-opening, open but un-repositioned, open and repositioned, or closed pre-second step. In the first or fourth case the purified complex would be competent to carry out first or second step catalysis, respectively, whereas

the open complexes are presumably not catalytically competent. Accurately ascertaining the state of purified complexes is therefore essential in order to allow coherent and reliable insights into the mechanism of splicing. As will be discussed below, there is an emerging view that multiple transitions along the splicing pathway resemble one another, so it is possible that the transition between the catalytic steps is not the only one composed of several smaller remodelling events.

CONFORMATIONAL TOGGLING, ASYMMETRY, AND THE RE-USE OF MOTIFS

Alleles of *prp22*, the ATPase involved in transitions during and after the second step of splicing, produce a cold-sensitive phenotype due to an mRNA release defect; a screen for suppressors of this phenotype identified a *prp8* allele (52) subsequently shown also to act as a general suppressor of first step splicing defects (47). Indeed, all known first step suppressor alleles of *prp8* also suppress *prp22* defects. This observation is consistent with the hypothesis that these alleles destabilise the second step conformation of the spliceosome, thereby stimulating the surrounding steps on the splicing pathway (first step catalysis and mRNA release), but could also suggest some degree of similarity between these flanking states. Consistent with the existence of such similarity, a growing number of interactions appear to be disrupted and re-form at defined stages of the splicing reaction – i.e. to toggle.

Many RNA-RNA interactions between snRNAs, as well as between snRNAs and pre-mRNA, have been identified, and take the form of (generally short) intra- and intermolecular helices (5). Some of these interactions are mutually exclusive with others, which suggests that they might exist only transiently, or may toggle between competing conformations. As previously noted, during the first catalytic step the UGU trinucleotide at positions 4-7 of the 5'SS base pairs with a conserved ACA in U6 snRNA (positions 47-49) (37). During the second step, when the 5'SS is in the branch of the lariat intermediate, it is in proximity to U6 positions 42-44 (43): interestingly, this region of U6 has also been shown to bind the 5'SS in early complexes (53, 54). Thus, the same binding site may be used for the 5'SS before and after its involvement in first step catalysis. Indeed, spliceosomal states even further apart on the splicing pathway display surprising similarities: the ATPase Brr2 disrupts the interaction between U4 and U6 snRNAs during spliceosome assembly, allowing U6 to interact with U2 and the 5'SS (55). It has recently been demonstrated that Brr2 is activated by the GTP-bound form of the U5 snRNP protein Snu114 and repressed by its GDP-bound form, and that Brr2 activation is required for spliceosome disassembly as well as U4/U6 unwinding (56). GTP hydrolysis by Snu114 is not required for Brr2 activation, suggesting a mechanism of action resembling that of classical G proteins. Although the GAP and GEF acting on Snu114 have not been identified, it is tempting to speculate that the relevant conformational states of the spliceosome might perform these roles, akin to the action of the signal recognition particle itself as the GEF for SR-β (57).

Genetic work in *S. cerevisiae* provides another example of conformational toggling (48, 49). The dynamic stem II region of U2 snRNA can form two mutually exclusive interactions, known as stem IIa and stem IIc. Stabilisation of stem IIa stimulates prespliceosome formation but inhibits the first catalytic step, which can be promoted by stem IIc stabilisation or IIa destabilisation. Disruption of stem IIc suppresses the splicing defect due to *prp16* mutation, whereas disruption of IIa (and therefore relative IIc stabilisation) suppresses second step splicing defects. These data are consistent with a model whereby stems IIa and IIc toggle, coexisting with the previously discussed open and closed forms of the spliceosome, respectively, with stem IIc therefore present during catalysis and IIa during repositioning (Fig. 1-4).

It has recently been demonstrated that the spliceosome, like group II introns (58), can be induced to catalyse both the forward and the reverse reactions of splicing. Purified postcatalytic spliceosome/product complexes prepared using a release-defective *prp22* mutant, when incubated under mild buffer conditions, undergo remarkably efficient reversal of both catalytic steps (59). Consistent with disparate states on the splicing pathway sharing conformational similarities as discussed above, F2, the forward reaction of the second step, and R1, the reverse of the first, are both favoured by the presence of KCl, while R2 is favoured by its absence. Thus, KCl appears to stabilise both pre-first and post-second step conformations of the spliceosome relative to the post-first, pre-

second step conformation (it is possible that states in which the 5'SS is bound to its upstream U6 snRNA binding site are stabilised). This observation provides further evidence for the global conformational similarity of the states before the first and after the second catalytic step of splicing. We anticipate that further analysis will highlight the similarities and differences between other spliceosomal states, and clarify the extent to which conformational toggling is a general phenomenon in splicing.

Although more examples of conformational toggling will likely be identified, many spliceosomal interactions are unlikely to reoccur once disrupted – for example 5'SS binding by U1, which is replaced by a 5'SS-U6 interaction via the action of Prp28 (60). In fact, a general theme in intermolecular interactions involving U6 snRNA is that of asymmetry. U6 mutations that disrupt a structure often have a more severe phenotype than corresponding mutations in the interacting partner, and incomplete suppression by compensatory changes is common. For example, mutations on the U6 side of U2/U6 helix Ia are substantially more severe than those on the U2 side (61); this may suggest that helix Ia does not remain intact throughout the entire splicing reaction, and that the interactions of the U6 component when not engaged in this helix are more critical for splicing than those of the U2 component. An alternative possibility is that involvement of helix Ia in a triple-helical interaction could give rise to the observed asymmetric phenotypes. Similar asymmetry is observed for the conserved AGC triad of U6, which can interact with sequences in U4 snRNA, with U2 snRNA (to form helix

Ib), and within U6 to extend the intramolecular stem loop (ISL) (5). Most mutations in AGC are viable if accompanied by a compensatory mutation in U2 that restores helix Ib, but some substitutions such as G60Y cannot be suppressed in this manner (62). A similar AGC motif in domain 5 of group II self-splicing introns acts as a metal binding site crucial for catalysis (63) and, as noted by Hilliker and Staley, spliceosomal AGC mutations that cannot be suppressed by U2 are expected to interfere with metal binding.

A second metal bound in domain 5 of group II introns is thought to mediate a docking interaction (63), and again an analogous metal binding motif exists in the ISL of U6. U80 (Throughout all sections of this work, *S. cerevisiae* numbering will be used unless otherwise specified) is bulged from the ISL and binds magnesium (64). The formation of this U80 bulge occurs after U4/U6 unpairing (5), and suppression data indicate the delicate balance of relative stabilities required to allow both structures to form, and thus permit splicing (65). However, the importance of specifying not only the nature but also the precise timing of interactions within the spliceosome is illustrated by the complex behaviour of this nucleotide. When substituted by 4-thio-uridine, U80 forms a site-specific crosslink with a nucleotide well upstream of the branch site of actin pre-mRNA (66); in addition, an Fe-BABE group tethered at the presumably fairly innocuous +10 position of the 5'SS stimulates cleavage at the human equivalent of U80 (67). Although it is possible that these biochemical data correspond to off-pathway intermediates, it is also plausible that they indicate that at least some of

the groups responsible for catalysis are sequestered by other interactions until the immediate pre-catalytic state of the spliceosome.

Spliceosome conformations can also be affected by transient protein modifications. For example, in addition to the known effects of the phosphorylation state of SR and other proteins on spliceosome assembly, Snu114 and the U2 snRNP protein SF3b155 appear to be dephosphorylated for the second step of splicing (68) and references therein). Similar effects are likely to be uncovered for many splicing factors, and indeed Prp8 ubiquitinylation has recently been shown to affect spliceosome assembly (69)

ATPASES AND FIDELITY

Although the short duplexes involved in spliceosome assembly and catalysis may be able to unwind naturally to facilitate conformational changes, perhaps in concert with other remodelling events, DExD/H ATPases represent a major class of spliceosomal proteins; by stimulating conformational transitions within the spliceosome, these ATPases play an integral role in the maintenance of splicing fidelity. Prp16, which as previously noted facilitates opening following the first step, was isolated as a suppressor of a branch site mutation in *S. cerevisiae*, and the relaxation of the requirement for adenosine at the branch site was subsequently shown to be due to ATPase impairment, resulting in the kinetic proofreading model of splicing fidelity (70). According to this model, functional progression into the second step occurs if catalysis precedes Prp16 ATP

hydrolysis, whereas substrate discard results if catalysis has not occurred before ATP hydrolysis. ATPase-deficient *prp16* alleles reduce fidelity and suppress splicing defects by allowing more time for catalysis to occur, resulting in the progression of suboptimal substrates, which would otherwise be discarded, through the first step.

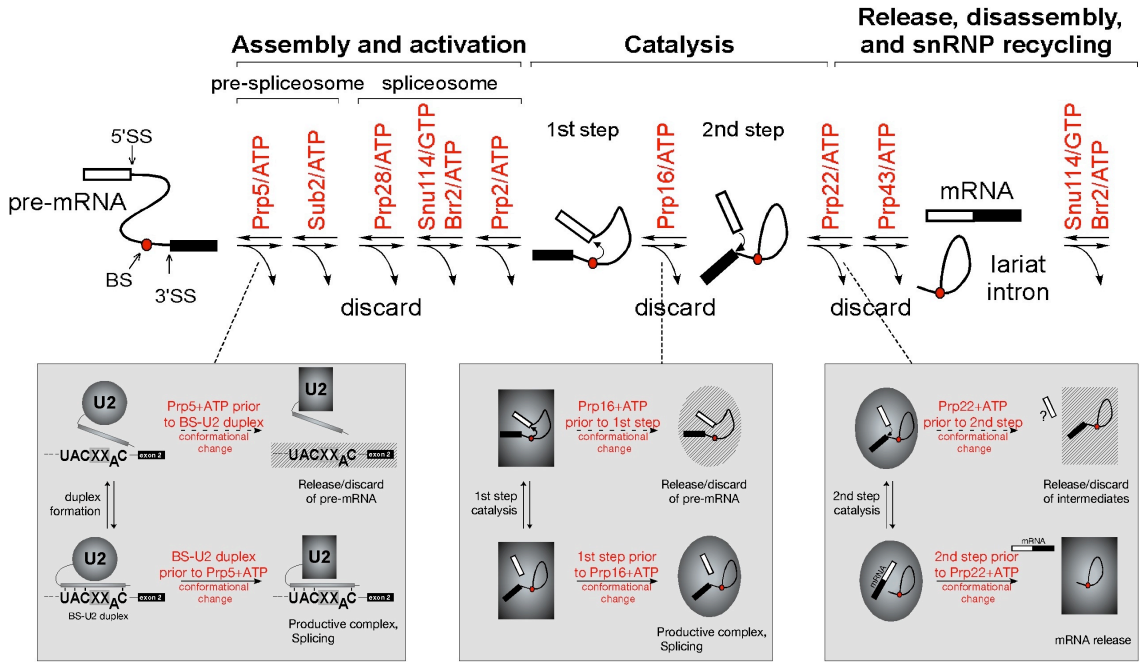


Figure 1-5. NTPase-associated steps during splicing as opportunities for kinetic discrimination.

(Upper) Schematic of transitions facilitated by DExD/H-box ATPases and the Snu114 GTPase during pre-mRNA splicing. SS, splice site; BS, branch site. (lower) Characterised examples of kinetic proofreading mediated by spliceosomal ATPases: (left) Altered competition between BS-U2 pairing and the conformational change mediated by the Prp5 ATPase changes the fidelity of BS selection. (center) Altered competition between the first catalytic step and Prp16 ATPase activity affects the fidelity of splice site usage in this step. (right) Altered competition between the second catalytic step and Prp22 ATPase affects second step splice site fidelity.

An important prediction of the kinetic proofreading model is that each ATPase-mediated conformational change affords opportunity for such a progression/discard branch in the splicing pathway: recent work has indeed demonstrated analogous behaviour for two more spliceosomal ATPases, consistent with the generality of this mechanism. *S. cerevisiae* spliceosomes assembled on 3'SS mutant substrates and purified after the first step proceed through the second in the presence of mutant Prp22 protein deficient for ATPase and/or unwindase activity, or in the absence of ATP (Fig. 1-5) (71). Genetic work in *S. cerevisiae* has also shown kinetic proofreading of branch site-U2 snRNA interaction by Prp5 (Fig. 1-5). ATPase-deficient *prp5* alleles suppress mutations flanking the BS that destabilise its pairing to U2, but suppression can be superseded by re-stabilising this interaction, either by compensatory mutations in U2 or intron mutations that generate extra upstream base pairs. The level of suppression by *prp5* alleles correlates inversely with their ATPase activity. Prp5 proteins from organisms in which the branch site is less highly conserved than in *S. cerevisiae* have lower ATPase activity, thus providing a mechanism by which the fidelity of branch site selection is reduced in these organisms (36). This work identifies the first structure to be in direct competition with the activity of a spliceosomal DExD/H ATPase, but the precise molecular consequences of ATPase activity remain to be elucidated for this and other ATPase-mediated transitions.

The kinetic proofreading model holds that the conformational change mediated by each ATP hydrolysis event is in competition with the preceding step on the splicing pathway. ATP hydrolysis results in progression through the pathway if the preceding step has been completed, and otherwise in substrate discard (36, 51, 71). Optimal substrates, which interact most favourably with the catalytic centre and therefore undergo each step of the splicing reaction more quickly than suboptimal substrates, are thus selectively driven through the reaction. Kinetic proofreading, if the catalytic rates of the ATPases in question are sufficiently high, allows the extent of spliceosome assembly and catalysis on suboptimal substrates to be minimised, as only optimal substrates will react quickly enough to compete with ATP hydrolysis. As the rate of ATP hydrolysis decreases, the ratio of discard to progression (predominantly on suboptimal-, but also on optimal substrates) decreases as substrates are afforded more time to assemble/react.

In the extreme case of zero ATPase activity, the reversibility of splicing (59) will likely allow equilibrium to be reached prior to the spontaneous occurrence of spliceosomal transitions. Because spliceosomal complexes containing optimal (and optimally reacted) substrates will be more stable than those containing aberrant substrates/products, the most populated intermediates at equilibrium will be those corresponding to assembly and catalysis on such optimal substrates. The equilibrium scheme will therefore favour the use of intrinsically strong splice sites to an extent determined by the difference in binding energy between the relevant wild-type and cryptic sites. Although it is therefore theoretically possible

to use reversibility and the resulting equilibrium as a slow proofreading mechanism, this is normally precluded by ATPase-mediated introduction of forward ratcheting and discard into the splicing pathway. In certain situations, however, equilibrium-based maintenance of splicing fidelity may be advantageous. Metazoan genes can be extremely long: a well-known example is the human dystrophin gene, which gives rise to a 14kb mRNA, is 2.3 Mb long and has 79 exons that are spliced during the 16 hours required for its transcription (72). In the context of such a long timeframe and energetically costly transcript, it is conceivable that rapid splicing with the concomitant risk of discard is disfavoured, and that local ATPase concentration/activity may be downregulated such that splicing events lie closer to equilibrium and fidelity is maintained by reversibility. It is worth noting that upregulation of ATPase activity could act as an effective and rapidly activatable transcriptional attenuator in this system, as discard between the catalytic steps would cleave the nascent transcript. At the other end of the spectrum lie ribosomal protein and other highly transcribed genes, especially those in rapidly dividing cells: for such genes, the energetic waste associated with substrate discard is likely outweighed by the greatly increased rate of splicing facilitated by ATPases, and thus the more rapid production of functional pre-mRNA.

Most spliceosomal ATPases are currently thought to facilitate a single transition along the splicing pathway, although the example of Brr2 demonstrates the possibility of a single ATPase acting multiple times (Fig. 1-5). The question of

how ATPase activity is limited to the correct stage(s) remains an open one. Binding of one ATPase to the spliceosome is not necessarily mutually exclusive with the presence of others, as illustrated by the persistent presence of Prp43 from early complexes until disassembly (6) during which time many other ATPases act. It is, however, possible that the binding sites for some ATPases share common elements, such that mutually exclusive subsets exist. A requirement for cofactors to stimulate ATPase activity is one mechanism by which activity could be temporally regulated: the helicase activity of Prp43 is stimulated by Ntr1, and this stimulation is required for Prp43's role in spliceosome (73). The recruitment of a cofactor or, in the event of an ATPase interacting with multiple spliceosomal components, the conformation of the spliceosome itself, could therefore activate an individual ATPase among several simultaneously bound, and repress others such that inappropriate conformational changes are not induced. The recent demonstration that Brr2's helicase activity is activated by the C-terminal domain of Prp8 (74) supports the hypothesis that ATPases may be activated by specific spliceosomal conformations, as Prp8 is without doubt associated with the spliceosome from tri-snRNP recruitment onwards.

Reversal of both steps of splicing catalysis by the spliceosome demonstrates the possibility of undergoing the transition between the two chemical steps of splicing (albeit in reverse) in the absence of Prp16, the ATPase normally responsible for this transition. There are two potential mechanistic interpretations of this Prp16-

less reversal. It is possible that Prp16 normally promotes transition into the second step by destabilising the first step conformation of the spliceosome. Because the reverse splicing reactions described by Tseng and Chen (59) contain neither Prp16 nor ATP, such reversal would be favoured and would therefore represent the truly spontaneous transition from the second to the first step. It is also possible, however, that ATPase binding and/or ATP hydrolysis may act to facilitate spliceosomal transitions by the stabilisation of transition states along the splicing pathway. In the event that Prp16 stabilised a transition state between the first and second step conformations, it is possible that its absence for the second-to-first-step transition may be partially complemented by the presence of Prp22, the ATPase normally required for the second catalytic step and mRNA release (75). The existence of at least some shared binding site elements for Prp16 and Prp22, as would be required for such complementation, is supported by the lack of temporal overlap in the binding of these two ATPases (76).

Substrate discard, while detrimental to the generation of mature mRNA, can play important physiological roles. The spliceosome is ideally suited to generate precise 3' ends possessing terminal hydroxyl groups, as such RNAs are a normal product of the first catalytic step. Recent work from Baumann and co-workers (77) indicates that the 3' end of telomerase RNA, TER1, in *S. pombe* is generated by spliceosomal processing: mature TER1 in this organism is in effect a free 5' exon that has been discarded by the spliceosome prior to the second

catalytic step. The precursor RNA from which mature TER1 is produced contains an intron with an inefficiently used 3'SS; strengthening this 3'SS reduces the level of mature TER1 as a direct result of an improved second step of splicing, and leads to a decrease in telomere length. Although in this instance the dominant effect appears to be due to weak splice sites, it is possible that other cis-acting sequences are present that, directly or via other trans-acting factors, enhance substrate discard. The generation of functional cellular RNAs via first step catalysis and discard by the spliceosome may not be limited to *S. pombe* telomerase: virtually any RNA requiring a defined 3' end could be produced via such a mechanism. We note that splicing, with or without discard, could also serve as a means to generate lariat RNAs; although evidence for biological roles for such RNAs is lacking, they could for example serve to protect transcripts from exonuclease digestion, or to generate RNAs lacking 5' ends, which could presumably remain stably un-capped.

Many important mechanistic questions regarding the role of DExD/H-box ATPases in the maintenance of splicing fidelity remain to be addressed. As will be discussed in chapter 5, the direct targets of these ATPases remain to be identified. Another open question is the nature of the discard possible at each ATPase-mediated step, which remains enigmatic. It is possible that the conformational change resulting from ATPase activity is not compatible with binding of the substrate for the previous step, such that it would cause the spliceosome to fall apart: alternatively, an active disassembly cascade could be

triggered by such a conformational change. It is even possible that the spliceosome may need to undergo several conformational changes resembling functional progression along the splicing pathway in order to allow discard. A role in discard for Prp43 and Ntr1, which cooperate in spliceosome disassembly (78), might suggest that discard is mechanistically similar to progression. Some support for this model is provided by the observation that discarded intermediates are degraded in the cytoplasm (79); this finding implies that discard, like mRNA release, is coupled to nuclear export.

THE SECOND CATALYTIC STEP

The second step of splicing remains substantially less well-characterised than the first. In addition to early recognition at the stage of complex A formation, the 3'SS has been proposed to be selected after the first step via a simple scanning mechanism as the first AG dinucleotide, or the second if sufficiently close to the first AG, downstream of the branch site (yeast) or polypyrimidine tract (metazoa) (80-82). Distance from the branch site is an important determinant of 3'SS strength: Prp22, Slu7 and Prp18 are all dispensable *in vitro* for introns with short BS-3'SS distances (75) and references therein), and Prp22 mutants stimulate the use of non-AG splice sites closer to the branch (71). In addition, the splicing of genes with short BS-3'SS spacing is unaffected by knockout of the second step factor Prp17 in *S. cerevisiae* (83). Aside from proximity to the branch site or polypyrimidine tract, what constitutes a favoured 3'SS remains unclear. The YAG 3'SS consensus sequence has not been extended by bioinformatic

investigation, although it seems possible that local RNA structure may play a role in determining the quality of a 3'SS for the second step; such bioinformatics, along with genetic screens to search for possible determinants of YAG strength, may reveal higher complexity.

Experimental investigation of the second step is hindered by several factors: the apparent ability of the spliceosome to assemble on one 3'SS and use another on the same pre-mRNA is one such hindrance. Previously noted experiments in plants demonstrating that a 3'SS sequestered in a hairpin could act as a splice acceptor only in the presence of a downstream 3'SS suggested the possibility of re-specifying the 3'SS after early assembly (28). Work on autoregulatory splicing by *sex-lethal* in *Drosophila* has provided evidence that a 3'SS can play a critical role in exon definition but not be preferred for catalysis (84). It is therefore possible that 3'SS requirements for assembly and catalysis are, from an experimental point of view, at best not necessarily identical and at worst obligately distinct. An 'ideal' 3'SS could thus represent a balance between assembly-competent and catalytically favoured states; this, and potential effects due to the almost unavoidable presence of nearby AG dinucleotides in natural genes, must be taken into account in any systematic analysis of 3'SS quality.

A second obstacle to the investigation of the second step is that any active site component required for both steps will presumably exert its effects at the stage of first step catalysis, thus rendering the investigation of its role in the second step

technically difficult. Although many spliceosomal components involved in the second step are dispensable for the first, such as the 3'SS itself (85), Slu7 (86) and loop 1 of U5 snRNA (87), several shared active site components might exist. Indirect information about the components of the second step active site could be derived from knowledge about the second step binding site of the lariat branch. The branch structure of the lariat intermediate must be repositioned and bound during the second step, consistent with the second step defect of 5'SS and BS mutants being suppressed by spliceosomal alleles that improve the second step (44, 46, 71). Although the 5'SS portion of the branch structure appears to reposition relative to U6 as previously discussed, the nature of this interaction, as well as the fate of branch site pairing to U2 snRNA after first step catalysis, is unknown.

In fact, temporal epistasis – that is, the manifestation of mutations that cause multiple sequential defects in a pathway only at their earliest point of action – can also impede the study of first step catalysis due to the long preceding assembly phase. Systems in which assembly can be bypassed, at least in part (71, 81, 88, 89), could represent possible ways to circumvent this problem.

Many proteins join the spliceosome between the two catalytic steps (6). *In vitro* depletion/reconstitution studies, together with genetic work *in vivo,* have provided clues as to the function of many second step factors. Loop 1 of U5 snRNA appears to act as a 'platform' on which the 5' and 3' exons are juxtaposed for

ligation (42), and interacts functionally with Prp18 – a protein that also binds Slu7 (90). Slu7 depletion leads to a loss of 3'SS fidelity, and also appears to destabilise free 5' exon binding to the pre-second step spliceosome (91). Prp22 crosslinks directly to the 3'SS following the action of Prp16 (92) and is also involved in mRNA release (75, 93). Interestingly, although all of these factors are essential *in vivo*, the requirement for each can be at least partially bypassed *in vitro* (42, 75, 91, 94), suggesting both that the second step spliceosome is fairly robust, and that these factors do not form interactions strictly necessary for catalytic events. Instead, a likely possibility is that they all contribute to stabilisation of the second step conformation relative to the first. Elucidation of the interactions made by these factors will be necessary to understand how they can impact 3'SS use and therefore alternative splicing.

ALTERNATIVE SPLICING

In the absence of repression, strong splice sites give rise to constitutive splicing; alternative splicing therefore represents the suppression of optimal splice sites and/or the use of those that are suboptimal. Most alternatively spliced introns are thought to be controlled by multiple splicing enhancer and silencer elements whose activity depends on their location relative to splice sites; these regulatory elements are thought to affect splicing predominantly through corresponding RNA binding proteins that exert their effects by altering a specific step in the spliceosome assembly pathway (95). However, modulation of splicing efficiency is theoretically possible at any stage of the splicing reaction. Each splicing event,

depending on the introns and exons involved, will be limited by one or more transitions, and will as such be sensitive to modulation of their efficiency while being insensitive to all but the most major changes in that of non-limiting steps. Given the diversity of pre-mRNA substrates, the multi-step nature of the spliceosome assembly and catalysis pathway, and the enormous number of factors involved in the splicing of every transcript, it is almost certain that examples exist in nature of splicing regulation at every possible stage. The existence of natural splicing events at least partially limited by post-assembly transitions has already been demonstrated: the splicing of overlapping but non-identical sets of endogenous introns is sensitive to the knockdown or mutation of core spliceosomal proteins important for the catalytic phase of the reaction (96, 97); this sensitivity affords opportunities for splicing regulation during catalysis. The ability to regulate the splicing of an individual intron by modulating the local activity or concentration of a given protein, or that of a class of introns via more global changes, together with the combinatorial effects of regulation of multiple transitions, can allow the robust and specific regulation of splicing events without a requirement for large numbers of individual splicing factors to exert this control. Thus, in order to understand alternative splicing, specific factors acting on each transcript need not necessarily be sought, and the entire reaction through assembly, catalysis and disassembly must be considered.

The stimulation or repression of spliceosome assembly naturally represents a common mechanism of splicing regulation, and may indeed be the most

prevalent in higher eukaryotes. Such assembly-based regulation might, however, be more subtle and complex than commonly thought. Virtually all metazoan transcripts contain multiple sequences capable of being recognised and used as splice sites, and much silencer-based repression may therefore require kinetic competition between these sites, serving predominantly to redirect rather than strictly to repress spliceosome assembly and/or catalysis.

Alternative splicing has not conventionally been considered to occur in *S. cerevisiae*. Data showing that the splicing of meiosis-specific genes is activated during meiosis (98) through the action of the Mer1 and Snu56 proteins (99), however, identify a system in which true alternative splicing can indeed be observed in this organism. We also note with interest the recent demonstration that the SR-like protein Npl3 activates the splicing of ribosomal protein genes, apparently via the recruitment of U1 and/or U2 snRNPs to sites of transcription (100); this behaviour is reminiscent of aspects of mammalian SR protein action. In spite of its relative paucity of natural alternative splicing events, *S. cerevisiae* represents the ideal model organism in which to carry out mechanistic splicing research *in vivo*; its ease of genetic manipulation and strong splice site consensus requirements allow many different events in splicing to artificially be made limiting for gene expression, facilitating detailed mechanistic analysis of the suboptimal splice site use that underlies alternative splicing.

CONCLUDING REMARKS

Enormous amounts of data exist regarding regulated and cell-type-specific splicing patterns (101). Specific splicing factors that affect the splicing of small numbers of transcripts must exert their effects via spliceosomal transitions and through core components, both of which are finite in number. This imposes a limit on the number of possible unique mechanisms of splicing regulation. In addition to the discovery of more specific splicing factors, we anticipate that more widespread and varied regulatory roles will be discovered for core spliceosomal components themselves. A clear, mechanistic description of the splicing process is necessary to explain regulated splicing, and the detailed analysis of various alternative splicing systems is also likely to identify additional transitions in the splicing pathway.

CHAPTER 2. AN ATTEMPT TO RECAPITULATE *S. CEREVISIAE* SNRNA-ONLY REACTIVITY YIELDS A 2'-3' BOND-FORMING RIBOZYME

INTRODUCTION

The question of whether the active site of the spliceosome is composed partly or entirely of RNA, like that of the ribosome (102), remains unresolved (103, 104). Prompted by the demonstrated ribozyme activity of U6 snRNA-derived sequences (105, 106) and the observation that RNAs with sequences based on those of human snRNAs and splicing substrates can form covalent bonds *in vitro* (89, 107, 108), we sought to investigate whether *S. cerevisiae* snRNAs would show such reactivity. *S. cerevisiae* is deficient in SR proteins, which stimulate splicing in metazoan systems at least in part via the stabilisation of RNA-RNA interactions (38, 39), and its spliceosome therefore relies more heavily on the intrinsic stability of these interactions than do those of higher eukaryotes (2). We reasoned that yeast snRNAs and splicing substrates may represent the optimal basis for a protein-free system to study pre-mRNA splicing.

One common issue with RNA-based systems, however, is the intrinsically reactive nature of RNA: reactions such as self-cleavage (109) and uncatalysed RNA ligation (110) are commonplace and may have simple, or even no, structural requirements. This chapter describes the positionally specific and relatively efficient formation of a 2'-3' bond between nucleotides in linker regions

in our original snRNA- and branch-site-based constructs. The formation of this non-physiological bond does not require snRNA sequence beyond that required for substrate binding, and can be observed using two short synthetic RNA oligonucleotides containing a minimal hairpin motif. These results reiterate the reactivity of RNA towards phosphate transfer reactions and suggest that a large amount of potentially reactive RNA sequence space is likely to be explored in cells. The manner in which this ribozyme activity was identified also highlights both the need for more stringent characterisation and validation of RNA-only reactions currently interpreted as indicative of ribozyme catalysis by the spliceosome, and the strength of fully spliceosomal approaches to the study of splicing catalysis.

RESULTS

AN RNA BASED ON FUSED *S. CEREVISIAE* U2 AND U6 SNRNA SEQUENCES SLOWLY FORMS A COVALENT BOND WITH A BRANCH SITE OLIGONUCLEOTIDE *IN VITRO*

In order to investigate the possibility that *S. cerevisiae* snRNAs may be able to catalyse reactions similar to those observed with human snRNAs (9, 107, 108), we generated a construct comprising the 5' terminal 46 nucleotides of U2 snRNA fused to the 3' portion of U6 snRNA beginning at position G39, hereafter referred to as U26. The two snRNAs were joined via a loop of four uridine residues, and two extra G-C base pairs were added around this loop to encourage intramolecular folding (Fig. 2-1A). This construct was mixed at ~1 μM with a

labelled oligonucleotide corresponding to the canonical *S. cerevisiae* UACUAAC branch site sequence with additional nucleotides upstream to stabilise binding to U2 snRNA, and an unlabelled 5' splice site (5'SS) oligonucleotide with similarly hyperstabilised binding to U6 snRNA in the register required for the first step of splicing (Fig. 2-1A) (37, 111, 112). The RNA mixture was heated to 95˚C then slowly cooled to room temperature in 50 mM Tris-HCl pH 7.2 containing 40 mM $MgCl_2$ and 200 mM KCl, left to react at room temperature, ethanol precipitated and analysed in a denaturing 20% polyacrylamide gel. A time course showed the slow generation of a specific low-mobility band over the course of many hours (Fig. 2-1B); the formation of this band was unaffected by the location of the label (5' phosphate or 3' pCp) on the BS RNA and by the presence or absence of the 5'SS RNA (Fig. 2-1C), indicating that it contained the entire BS and no 5'SS sequence. The slow mobility of the product suggested that it likely contained some or all of U26. We therefore carried out RNase H degradation on the purified product using oligonucleotides directed against either end of U26, and observed an increase in mobility with an oligo against the 5', but not the 3' end (Fig. 2-1D). The product therefore retains the 5' end of U26 but has lost the 3' end. Taken together, these observations are suggestive of a reaction between an internal nucleophile in BS and an internal phosphodiester bond in U26.

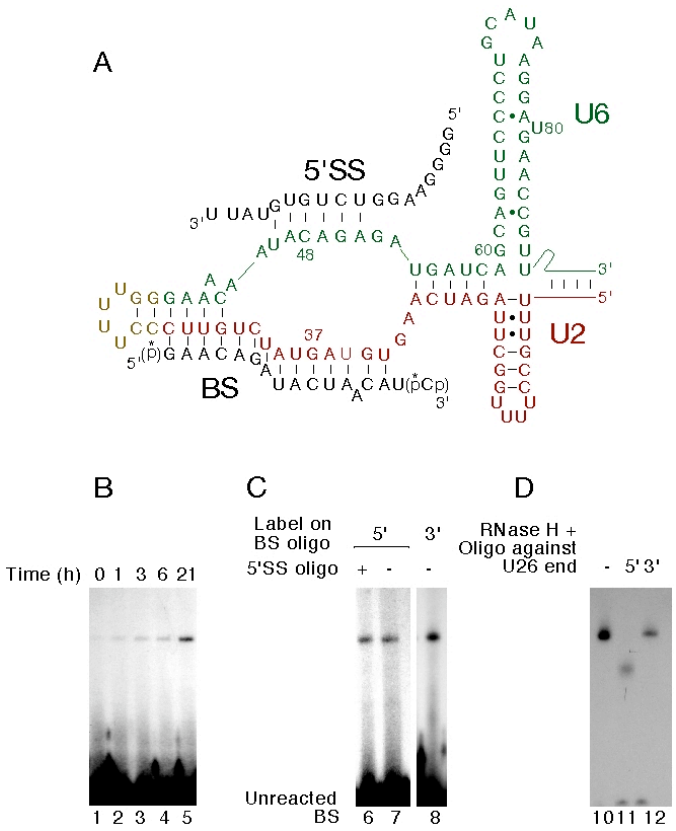


Figure 2-1. Slow formation of a covalently linked product containing the 5' end of U26 and both ends of a branch site oligonucleotide.

(A) Schematic of constructs based on *S. cerevisiae* snRNAs and splicing substrates: U6 snRNA sequence is shown in green, U2 snRNA sequence in red, and splicing substrates in black. Potential base pairs, even when mutually exclusive, are indicated. (B) The substrates depicted in Fig. 1a form a slowly migrating product over the course of several hours. (C) Product formation does not depend on the presence of the 5'SS oligonucleotide, and the product contains both ends of the BS oligonucleotide. (D) RNase H digestion of purified product with oligonucleotides complementary to the 5' or 3' ends of the U26 construct indicates that the product contains the 5'-, but not the 3' end, of U26.

CONSERVED U6 SNRNA SEQUENCES ARE NOT REQUIRED FOR REACTIVITY TOWARDS BS

The single sharp band observed following reaction of BS and U26 suggested a positionally specific reaction. In addition, the observed internal cleavage of U26 is

suggestive of a transesterification; we therefore wished to investigate the dependence of our observed reactivity on snRNA sequences known to be important for pre-mRNA splicing. We generated constructs in which either or both of the ACAGAGA and AGC motifs in U6, which are universally conserved and essential for viability (2), were mutated to runs of adenosine residues (Fig. 2-2A). The reactivity of these constructs was indistinguishable from that of U26 containing wild-type ACAGAGA and AGC (Fig. 2-2B). In addition, we made a series of deletion constructs in which U26 was truncated from its 5'- (Fig. 2-2C Δ1-3) or 3' ends (Fig. 2-2C, Δ4-6) and expressed from a cloning vector (pDrive, Qiagen). U26 RNAs truncated at their 5' ends showed strong reactivity until most of the BS binding site had been removed (Fig. 2-2D lanes 2-4), with further truncation presumably precluding binding of the BS oligonucleotide. 3' end truncations/substitutions removed U6 sequence upstream of a constant vector sequence preceding a T7 terminator. Truncations beginning at and downstream of ACAGAGA retained reactivity (Fig. 2-2D cf. lanes 1 & 5-6), and a construct containing only two nucleotides of U6 sequence still produced a trace amount of product (Fig. 2-2D lane 7). The mobilities of products derived from truncated U26 molecules are consistent with retention of only the 5' end of the molecule; this rules out a 2'-5' linkage and, consistent with this, the mobility of the product was not affected by treatment with recombinant debranching enzyme (data not shown). The reactivity of these truncation constructs also suggests (conservatively) that the scissile bond in U26 lies between positions G37 of U2 and C48 of U6. Subsequent experiments were therefore performed with a

synthetic RNA oligonucleotide corresponding to the region of U26 from U2 position G37 to U6 position C48 (Fig. 2-3A).

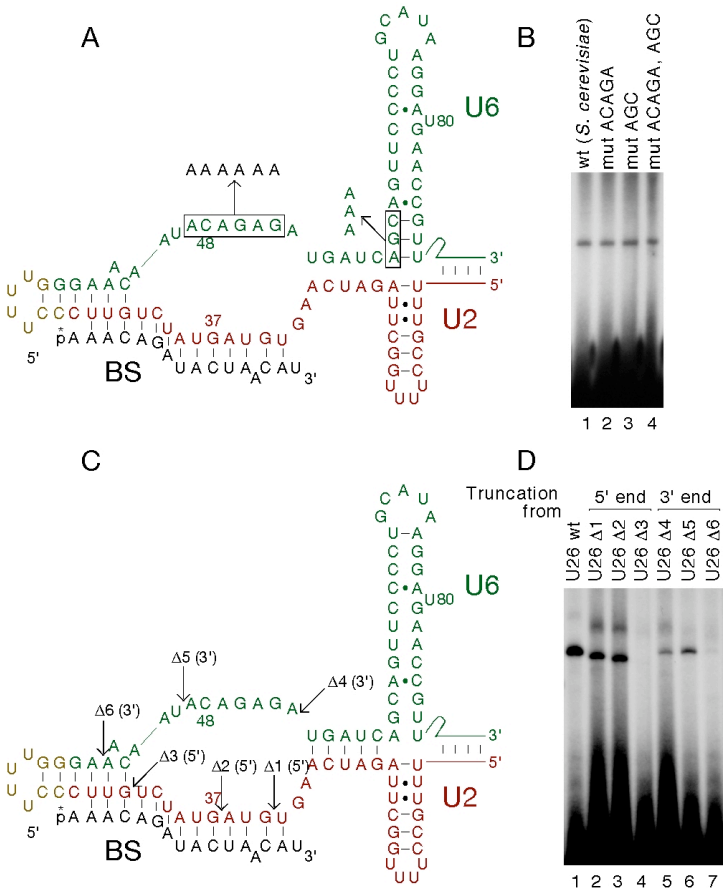


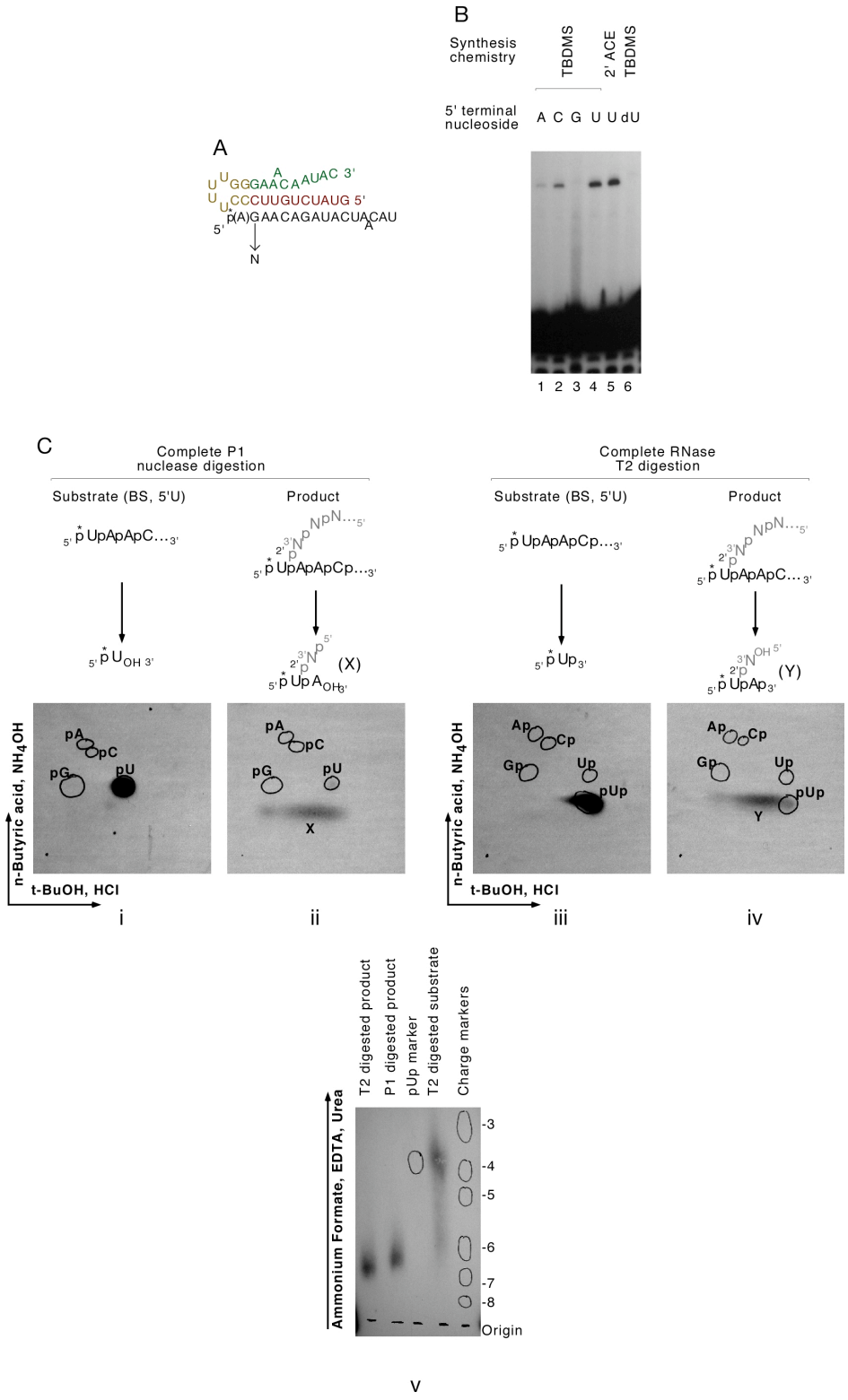
Figure 2-2. Virtually no snRNA sequence is required for U26 reactivity.

(A). Schematic indicating mutations in the ACAGAGA and AGC motifs of the U6 portion of the U26 construct: these motifs are universally conserved and essential for viability. (B) Mutation of either or both of the ACAGAGA and AGC motifs to runs of As has no effect on the formation of the U26-BS product. (C) Schematic indicating truncations made from the 5' end (1-3) or 3' end (4-6) of the U26 construct. These truncations were made in the context of constant surrounding vector (pDrive, Qiagen) sequence. (D) Some truncated U26 constructs retain reactivity towards BS. Deletion construct 6 shows little reactivity, and construct 3 none, the latter likely due to an inability to bind the BS oligonucleotide. The observed variation in product size is consistent with retention of only the 5' end of U26 in the product.

Figure 2-3. Identification of the reactive site on the BS oligonucleotide.

(A) Schematic indicating the sequences of constructs used in this figure. (B) The 2' hydroxyl of the 5'-terminal nucleotide of BS is strictly required for reactivity towards U26. BS oligonucleotides with diverse 5' ends can react; reaction efficiency varies with the identity of this nucleotide, but not in a manner consistent with a functional group from the base acting as a nucleophile. Reactivity is not due to a synthesis artefact, as oligonucleotides made via different synthesis routes (TBDMS and 2' ACE) react with identical efficiency. (C) TLC characterisation of the product: (i-iv) 2-D TLC analysis of P1 nuclease/RNase T2 digestions of unreacted BS substrate (i, iii) or reaction product (ii, iv) with a labelled 5' terminal U. The migration of unlabelled markers is indicated, and expected labelled digestion products shown. BS-derived sequence is shown in black, and unmapped U26 sequence in grey. (v) DEAE TLC analysis of P1 nuclease/RNase T2 digestions of the 5'U reaction products to assay net charge; the positions of an unlabelled pUp marker and markers generated by partial alkaline hydrolysis of unreacted substrate are indicated.

Figure 2-3


A
B
Synthesis chemistry
TBDMS
2' ACE
TBDMS
5' terminal nucleoside
A C G U U dU
1 2 3 4 5 6
C
Complete P1 nuclease digestion
Substrate (BS, 5'U)
Product
Complete RNase T2 digestion
Substrate (BS, 5'U)
Product
(X)
(Y)
n-Butyric acid, NH4OH
t-BuOH, HCl
i
ii
iii
iv
T2 digested product
P1 digested product
pUp marker
T2 digested substrate
Charge markers
Ammonium Formate, EDTA, Urea
-3
-4
-5
-6
-7
-8
Origin
v

THE PRODUCT COMPRISES THE ENTIRE BS RNA JOINED VIA ITS 5'-TERMINAL NUCLEOTIDE TO A 5' FRAGMENT OF U26 VIA A 2'-3' PHOSPHODIESTER BOND

Rough mapping by RNase digestion using internally labelled BS oligonucleotides suggested that the reactive group in BS was close to the 5' end of the molecule (data not shown). 5' end-labelled BS with a uridine as its 5'-terminal nucleoside was reacted with the minimal U26 oligonucleotide depicted in Fig. 2-3A, and this product gel-purified. Purified BS-U2 product was digested to completion with P1 nuclease, which cuts 5' of each nucleotide to yield pN, or RNase T2, which cuts 3' of each nucleotide to yield Np, and pNp in the case of a phosphorylated 5'-terminal nucleotide; digestion products were separated by two-dimensional thin layer chromatography as previously described (113). The labelled phosphate from unreacted BS purified from the same gel as the product comigrated as expected with pU and pUp markers after P1 and T2 digestions, respectively (Fig. 2-3C, i&iii). P1 and T2 digestions of the BS-U2 product, however, produced species that did not comigrate with any of our mononucleotide markers, but which migrated similarly to one another (Fig. 2-3C, ii&iv). Neither P1 nor T2 can cleave at a branch nucleotide (113), so the failure of these nucleases to produce labelled pU and pUp, respectively, from BS-U2 is indicative of branching from the 5'-terminal nucleotide of BS. Branching from further downstream nucleotides would not result in aberrant labelled P1/T2 digestion products. While this initial TLC analysis did not clarify the nature of the covalent bond between BS and U26, we reasoned that the most parsimonious chemistry was a 2'-3' phosphodiester bond between the 2' hydroxyl of the 5'-terminal uridine of BS and

a phosphate group from U26. Such 2'-3' branched molecules would be expected to produce similar products on digestion with P1 nuclease/RNase T2 (these products are shown schematically in Fig. 2-3C, ii&iv). Because the P1 and T2 digestion products of such a BS-U26 product are expected to differ only in the location of a single phosphate group (Fig. 2-3C), they should carry an identical net charge of -6. Further TLC analysis of the P1 and T2 digestion products using DEAE plates indicated that both did carry a -6 charge (Fig. 2-3C, v – the slight mobility difference is due to the 5' or 3' location of the unlabelled terminal phosphate); only a branched RNA molecule could carry more than four charges after P1 or T2 digestion, and only a 2'-3' branch would generate P1 and T2 digestion products with the same charge. The bond between BS and U26 can therefore be assigned as a 2'-3' phosphodiester. We obtained analogous results for all mapping experiments using full-length U26 and a BS oligonucleotide with a 5'-terminal G, indicating the generality of this product (data not shown).

In addition to our TLC analysis, further data support the assignment of this reaction as a 2'-3' transesterification. The reaction proceeds regardless of the 5' end phosphorylation state of the BS oligonucleotide (Fig. 2-1C), and a 5'-terminal phosphate on BS remains sensitive to phosphatase treatment in the product (data not shown); this rules out an involvement for the 5' phosphate itself meaning that the nucleophile must either be a group on the 5'-terminal base or its associated ribose/phosphate. Changing the 5'-terminal base of BS has a profound effect on reactivity (Fig. 2-3B lanes 1-4), with the 5'U being the most

reactive and the 5'G the least. However, the relative efficiencies of reaction that we observe are inconsistent with attack by a functional group on the base, and we believe them to be related to a basic structural requirement for this reaction. Reactivity due to a contaminant or incompletely deprotected group arising from chemical oligonucleotide synthesis was ruled out by the identical reactivity of oligonucleotides produced via TBDMS and 2'ACE synthesis (Fig. 2-3B lanes 4-5). The 2' hydroxyl of the 5'-terminal nucleotide therefore remains as the only possible nucleophile for this reaction. Indeed, substitution of the normally reactive 5' uridine with deoxyuridine abolished reactivity (Fig. 2-3B lanes 4 and 6).

Transesterification reactions are strongly dependent on metal ions, requiring divalent cations for nucleophile activation and leaving group stabilisation (114). Consistent with the BS-U26 reaction being a transesterification, no product is observed in the absence of divalent metal ions, but magnesium, manganese or calcium can all support reactivity (Fig. 2-4 lanes 10-13). Two metal ion catalysis requires inner shell coordination of reactive groups by the divalent metal ion; such coordination is not possible with cobalt hexamine – a structural mimic of a fully hydrated Mg^{2+} - and the absence of reactivity in the presence of cobalt hexamine (Fig. 2-4 lane 14) therefore further supports the assignment of this reaction as a transesterification. Finally, the increase in efficiency observed with increasing pH (Fig. 2-4 lanes 1-5) is suggestive of the expected requirement for nucleophile deprotonation.

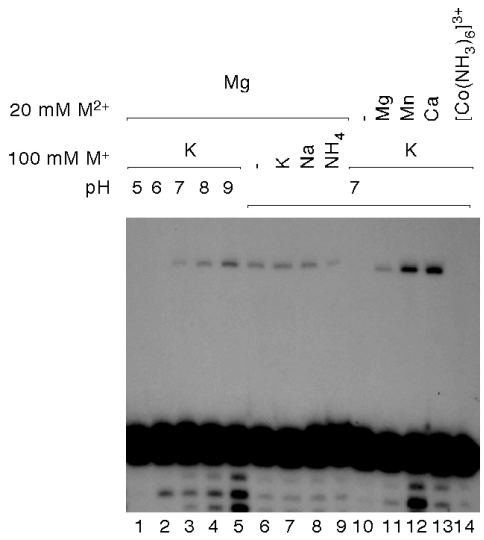


Figure 2-4. pH and metal ion requirements of the reaction.

pH and metal ion dependence of the reaction between [5'U] BS and the minimal U26 oligonucleotide.

MAPPING THE SCISSILE PHOSPHATE WITHIN U26

Together, our data demonstrate the formation of a 2'-3' phosphodiester bond between BS and U26. Such a bond is currently unprecedented in any physiological system, but has previously been observed in the products of artificial ribozymes (105). Incompatible strand polarities precluded mapping the site of reactivity in U26 by primer extension, and the U26 concentrations required to generate detectable levels of product rendered specific labelling experiments impractical. Indirect identification of the scissile bond was, however, possible via RNase digestion of products containing end-labelled BS.

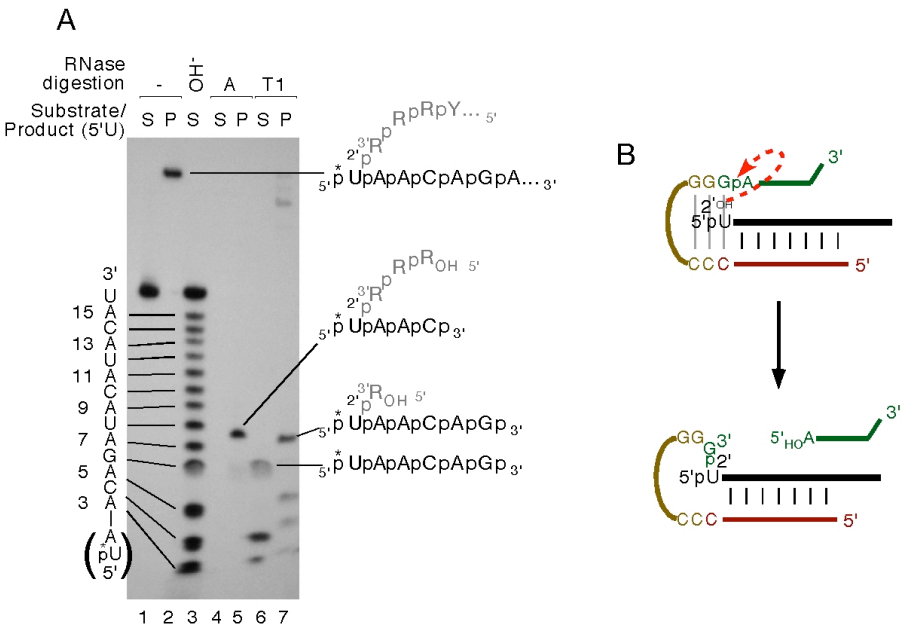


Figure 2-5. Mapping the scissile phosphate in U26.

RNase digestion of [5'U] BS (S) and its product with U26 (P) identifies 5'YRRRN3' as the 3'-most nucleotides of U26 appended to BS; the sequence of BS and the sizes of hydrolysis fragments are indicated beside the ladder in lane 3. BS sequence is shown in black, and U26 sequence in grey. (B) Schematic of the reaction with sites of reactivity and polarity of phosphodiester linkages detailed.

In order to allow mapping of the scissile bond within U26, purified product resulting from reaction between a 5' uridine end-labelled BS and the minimal U26 oligonucleotide was digested with RNases A or T1 (Fig. 2-5A). Typically RNase T1 cuts 3' of guanosines, but under our conditions of high enzyme concentration (presumably required due to secondary structure in the products) it cuts 3' of all purines with only a slight preference for guanosine. T1 digestion of the product indicated that that U26 was joined to BS via a purine, as each band resulting from digestion of the product showed a decrease in mobility corresponding to 1nt relative to the analogous band resulting from the digestion of unreacted substrate (Fig. 2-5A, lanes 6&7 – digestion products are shown schematically beside the

gel). Digestion of the product with RNase A, which cuts 3' of pyrimidines, generated a species that migrated as a 7-mer (Fig. 2-5A lane 5). Branching from the terminal U of BS would inhibit RNase A cleavage at this position, so the first cut in BS must be made 3' of C4; therefore the 7-mer digestion product indicates that the last four most 3' nucleotides of the retained U26 fragment must be (with the RNase A cleavage site indicated by an asterisk) 5'Y*RRR3'—2'BS (shown schematically in Fig. 2-5A). The minimal U26 oligonucleotide contains only one YRRRN sequence – the UGGGA running from the terminal U of the U_4 loop to U6 A40, indicating that the scissile phosphate is that between the third G and the A in this motif. We therefore conclude that the U26-derived strand of the product comprises the entire U2 portion and the CCCUUUUGGG stem loop (reactants/products are shown schematically in Fig. 2-5B). The released 3' end of U26 is expected to migrate among degradation products, and can therefore not be followed.

Despite the simple ionic requirements of this reaction (Fig. 2-4), and the ability of BS oligonucleotides with diverse 5' ends to react (Fig. 2-3B), at least some degree of higher-order structure appears to be important. Mutations in the hairpin region of the U26 oligonucleotide, including compensatory changes as well as those that disrupt the helix, strongly inhibit or abolish reactivity towards all tested BS substrates (data not shown), suggesting that the juxtaposition of nucleophile and electrophile requires structure beyond simple base pairing. The sensitivity of the U26 hairpin to mutation precludes detailed analysis of such

structural requirements. However, given that the 5'-terminal nucleotide (G) of BS is capable of pairing to the 3'-terminal nucleotide (C) of U2 sequence in U26 (Fig. 2-1A), and that the reactivity of BS oligonucleotides correlates inversely with their ability to form a base pair at this position (Fig. 2-3B, lanes 1-4), we hypothesise that nucleophilic attack occurs when the 5' end of BS is single-stranded.

DISCUSSION

IMPLICATIONS FOR THE POTENTIAL REACTIVITY OF CELLULAR RNAS

The ease with which ribozymes, particularly those that catalyse RNA hydrolysis or transesterification reactions, can be generated by SELEX (115), suggests that a large proportion of RNA sequence space is at least minimally catalytically active. Our data, which show the relatively efficient and positionally specific catalysis of a transesterification reaction, apparently by a small and simple structural motif, support this idea.

The use of tiling arrays and deep sequencing has recently indicated that a greater proportion of many organisms' genomes is transcribed than was previously thought (116); given the diversity of cellular RNA sequence that such widespread expression would generate, high intracellular RNA concentrations, and the potential for ribozyme catalysis by extremely simple structures, such intrinsic RNA catalysis may occur *in vivo*. Cells have likely evolved mechanisms to prevent and/or mitigate the effects of inappropriate RNA reactivity *in vivo.*

These mechanisms may include evolutionary selection against catalytically active sequences or structures, the rapid turnover of RNA such that sufficient time is not normally allowed for catalysis, and the coating of RNA molecules with proteins such as hnRNPs. It is, however, also possible that such ribozyme-generated RNAs may play physiological roles. The analysis of such potential products, as well as of enzymes involved in their further processing or degradation, may be complicated in eukaryotes by the background of lariats arising due to pre-mRNA splicing, as these will comprise the majority of non-linear RNAs. In addition, ribozyme-derived products may contain bonds (such as 2'-3' phosphodiesters) that do not lend themselves to conventionally used mapping techniques.

IMPLICATIONS FOR THE UTILITY OF PROTEIN-FREE SYSTEMS AS TOOLS TO STUDY PRE-MRNA SPLICING

The ability of snRNA-derived RNA sequences to catalyse chemical reactions including 2'-5' phosphodiester bond formation has previously been interpreted as strong evidence for ribozyme catalysis by the spliceosome (89, 106, 107). The data in this chapter highlight the feasibility of direct and unambiguous characterisation of the products of such *in vitro* reactions: in addition, they caused us to reflect on the requirement for stringent validation of such systems, if such validation is indeed possible. In order for experimental systems based entirely on RNA to live up to their stated promise as the basis for analysis of splicing catalysis, they must be fully validated. Given that catalytic mechanisms cannot be proven but only disproven, such validation requires that protein-free systems be shown, as far as possible, to catalyse reactions identical to those of

splicing both in terms of chemistry and catalytic mechanism. Both stringent characterisation of intermediates and products, and the concordance of results between protein-free and complete spliceosomal systems, are therefore required. The need for recapitulation of results from RNA-only reactions highlights in turn the pressing need for model systems in which the catalytic phase of splicing can be studied in a truly spliceosomal context.

DEFINING AND VALIDATING RELATEDNESS TO SPLICING

RNA is a structurally promiscuous polymer that contains diverse functional groups: as such, a single RNA sequence may have the potential to catalyse and/or participate in many reactions (117). Because sensitive detection methods allow for the investigation of products formed in low yields and at slow rates, multiple products may appear on extended incubation of diverse RNA molecules (including but not limited to RNAs that are snRNA-derived). Thus, early in the development of a minimised model system for pre-mRNA splicing, it will almost certainly be necessary to decide which reaction product one defines to be derived from a splicing-related reaction. RNA-only systems that generate products with the same covalent linkages as those that result from spliceosomal splicing may merit study, and could serve as interesting models for RNA catalysis. A shared chemical identity, however, does not necessarily indicate that the precise mechanisms by which two bonds are formed are identical. The present definition of relatedness to splicing, and our incomplete knowledge of

contributors to spliceosomal catalysis, make it unclear whether conclusions drawn from protein-free splicing are truly applicable to the spliceosome.

Relatedness of a ribozyme reaction to pre-mRNA splicing has conventionally been concluded from its sensitivity to mutation of the ACAGAGA and/or AGC motifs in the U6 snRNA (or its derivative) component of the reaction (89, 106, 107). Sensitivity to such mutations is, however, not necessarily indicative of direct similarity to either reaction of splicing. Although these core motifs, which are shared between U6 snRNA and group II introns, have recently been demonstrated to form the basis of the group II intron active site (118), similar sequences are also found in other ribozymes believed to be unrelated to snRNAs, and the full details of the role(s) of ACAGAGA and AGC in both these ribozymes and the spliceosome remain unclear. For example, in spite of its universal conservation, the AGC triad is remarkably mutable in *S. cerevisiae*, with all substitutions except A59U and G60Y viable either alone or in the presence of compensatory U2 snRNA mutations that restore a helical interaction (helix Ia) (62). Backbone groups, rather than bases, coordinate metal ions; the results of sequence-based mutagenesis could reflect structural defects common to various RNA folds and must therefore be treated with caution.

Group II self-splicing introns provide a natural precedent for RNA catalysis of the two chemical reactions of pre-mRNA splicing, and their well-established similarities and likely evolutionary relatedness to snRNAs (63), especially in light

of the recent group II intron crystal structure (118) provide strong support for the hypothesis that the spliceosome active site is at least partially RNA-based. We note that, by the present definition of relatedness to splicing, group II introns qualify as a protein-free splicing system. We therefore suggest that either group II introns provide a pre-existing and ideal protein-free system in which mechanistic studies of pre-mRNA splicing can be carried out, or that more stringent requirements must be applied for a reaction to be considered as sufficiently related to spliceosomal splicing to serve as a model: one requirement is similar dependence on divalent metal ions. Magnesium has been directly demonstrated to be important for spliceosomal catalysis (119, 120), and further investigation should clarify the coordination and precise mechanistic roles of divalent ions in the spliceosome. Similar metal ion dependence in protein-free systems must be considered a prerequisite for splicing-relatedness.

Before sensitivity to the mutation of core motifs in U6 snRNA can be used to define a splicing-like reaction, the nature of the defects arising from such mutations within a truly spliceosomal context must be elucidated. Subsequent work demonstrating that defects are mechanistically identical in spliceosomal and protein-free contexts would, we believe, provide evidence that the observed reaction resembled splicing. Again, however, we note that this mechanistic identity may already exist between group II introns and the spliceosome.

THE SPLICEOSOME AS AN RNP

The above discussion highlights the need to develop parallel, fully spliceosomal model systems for the study of splicing catalysis before the applicability of protein-free reactions (including those of group II introns) to splicing can be reliably assessed. The protein component of the spliceosome should not be ignored despite its daunting complexity (6). A large number of spliceosomal proteins are essential for cellular viability and splicing *in vitro*: while many play important roles in the conformational rearrangements required for productive splicing (reviewed in chapter 1) a growing body of evidence suggests that some spliceosomal proteins may have a function beyond simple scaffolding of a catalytic RNA core. Several recent studies show that Prp8 has an RNase H fold, also found in integrases and transposases (121-123), that contacts the 5'SS at at least one stage of spliceosome function: this further highlights the need for caution in interpreting results from RNA-only systems in the context of spliceosomal splicing: a fuller treatment is given to this issue in chapter 5.

Although snRNA-only splicing cannot demonstrate that the spliceosome is a ribozyme, a validated system might rapidly produce a number of hypotheses regarding spliceosomal catalysis. Given the importance of spliceosomal proteins, the parallel systems able to validate protein-free splicing and test these hypotheses would by necessity be based on at least predominantly intact spliceosomes. Thus, while protein-free splicing could conceivably raise interesting questions, the hypotheses thus generated must necessarily remain

un-validated until they are supported by data from a model that supersedes the snRNA-only reaction. For this reason, the establishment of more complete spliceosomal models likely represents the most constructive avenue for future research.

CHAPTER 3. *TRANS*-SPLICING OF A REPORTER GENE TO U2 SNRNA *IN VIVO*

INTRODUCTION

Pre-mRNA splicing can be considered as three broad phases, each comprising multiple steps: these phases are spliceosome assembly, splicing catalysis, and spliceosome disassembly/mRNA release. Although the majority of work in the splicing field is focused on assembly, there is growing consensus that transitions during the catalytic phase can limit the rate and fidelity of the overall splicing reaction and act as potential points of regulation (96, 97).

The nature of the two chemical reactions of splicing (113) and a model for the mechanism by which fidelity is maintained (51) were proposed twenty and fifteen years ago, respectively, yet, despite the more recent advances outlined in chapter 1, many details of splicing catalysis remain enigmatic. The relative neglect of the catalytic phase is in part due to the complex assembly phase that precedes it, as many defects will manifest themselves during assembly and thus preclude the investigation of catalysis; this highlights the need for new experimental systems for the study of the catalytic phase of splicing. The duplex between the BS and the highly conserved cognate region of U2 snRNA (124-126), shown schematically in Fig. 3-1A, exemplifies a structure believed to be important in multiple aspects of the splicing reaction, whose investigation has been hindered by the limitations of current experimental models.

During spliceosome assembly, the BS is transiently bound by SF1/BBP before being transferred to U2 snRNA (127, 128). The stability of the resulting BS-U2 duplex is proofread by kinetic competition between duplex formation and ATP hydrolysis by the DExD/H box ATPase Prp5 (36), with sufficiently stable BS-U2 interaction allowing progression into the productive splicing pathway as opposed to pre-mRNA discard. Mutations that weaken the interaction between the BS and U2 therefore strongly inhibit spliceosome assembly.

The BS-U2 pairing established during spliceosome assembly has generally been thought to persist unaltered through the first catalytic step of splicing. As will be discussed in detail in the introduction to chapter 4, the catalytically relevant form of the branch nucleophile is thought to be that in which it is bulged from the BS-U2 duplex (Fig. 3-1A), with this helix therefore playing a role in branch site specification and activation for catalysis. No data exists regarding the fate of the BS-U2 duplex after branching catalysis has been completed: although, as discussed in chapter 1, the 5'SS portion of the newly formed lariat moves relative to U6 snRNA following the first step (44), the BS-U2 duplex has generally been assumed to persist unaltered through this transition even though this would presumably require movement of U2 relative to U6, likely resulting in the disruption of some intermolecular interactions formed between these snRNAs.

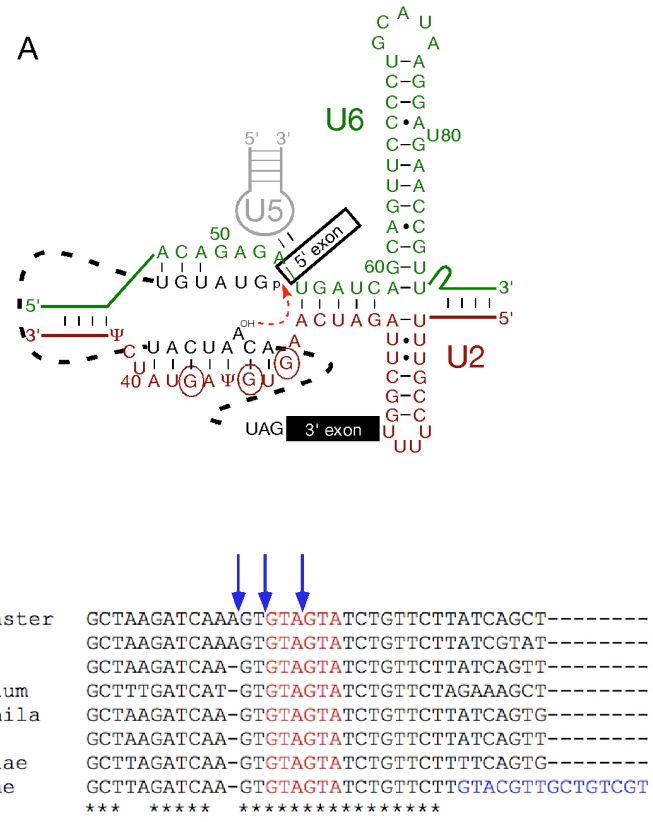


Figure 3-1. The BS-binding region of U2 snRNA contains multiple 5'SS-like sequences, and is the site of an intron in *R. hasegawae.*

(A) Schematic of RNA-RNA interactions that contribute to the first step of splicing, with the initial G of each of the three 5' splice site-like sequences in the BS-binding region of U2 snRNA indicated by circles. (B) Alignment of U2 snRNA sequences from various species. The BS-binding sequence is indicated in red, and intronic sequence in blue: arrows point to potential 5' splice sites

As discussed in chapter 2, U2 and U6 snRNAs do appear to have some intrinsic catalytic potential, although it remains unclear whether this is related to splicing. *In vitro*-transcribed snRNAs, when incubated with a transcript containing a branch site sequence, induce the formation of a covalent bond proposed to be a phosphotriester between the canonical BS-A and the guanosine in the universally conserved U6 AGC triad (108). This reaction may have some biological

relevance, as the site of reactivity in U6 is adjacent to the location of an intron in the *S. pombe* U6 snRNA gene (129), generally thought to have arisen due to an aberrant splicing event followed by retrotransposition of the resulting product into the genome.

U6 reactivity in a spliceosomal context has in fact been observed directly (130). In an *in vitro* system designed to reconstitute nematode cis- and trans-splicing using *Ascaris lumbricoides* extracts, Nilsen and co-workers detected branched and linear RNA products indicating the use of U6 as both a branch acceptor and a 5' exon in a 5'SS-independent splicing reaction. The reaction appears to represent an aberrant first step followed by a quasi-normal second step, as the canonical BS and 3'SS nucleotides are used, and U6 mutants that abolish the ability to reconstitute normal cis- and trans-splicing also render U6 trans-splicing undetectable.

Although U6 is the only snRNA that has previously been demonstrated to be reactive, the presence of multiple 5'SS-like sequences in the BS-binding region of U2 (Fig. 3-1A), and the presence of a spliceosomal intron close to this region in *Rhodotorula hasegawae* (131), (Fig. 3-1B) prompted us to search for the products of a trans-splicing reaction involving U2 in *S. cerevisiae*. Described below is the characterisation of an RNA product comprising the highly conserved 5' end of U2 and the 3' exon of the *ACT1-CUP1* reporter gene in intact *S. cerevisiae* cells. The formation of this product requires full spliceosome

assembly, and is only detectable when the reporter has a mutant 5'SS predicted to bind poorly at the spliceosome catalytic centre. The predominant 5'SS used in U2 is in the middle of the predicted BS-U2 duplex, but multiple sites can be used; these data suggest that BS-U2 pairing is unlikely to be required for the second catalytic step of splicing, and that it may be at least partially disrupted prior to the first step.

RESULTS

THE BS-BINDING REGION OF U2 SNRNA CAN BE USED IN TRANS AS A 5' EXON FOR BOTH STEPS OF SPLICING

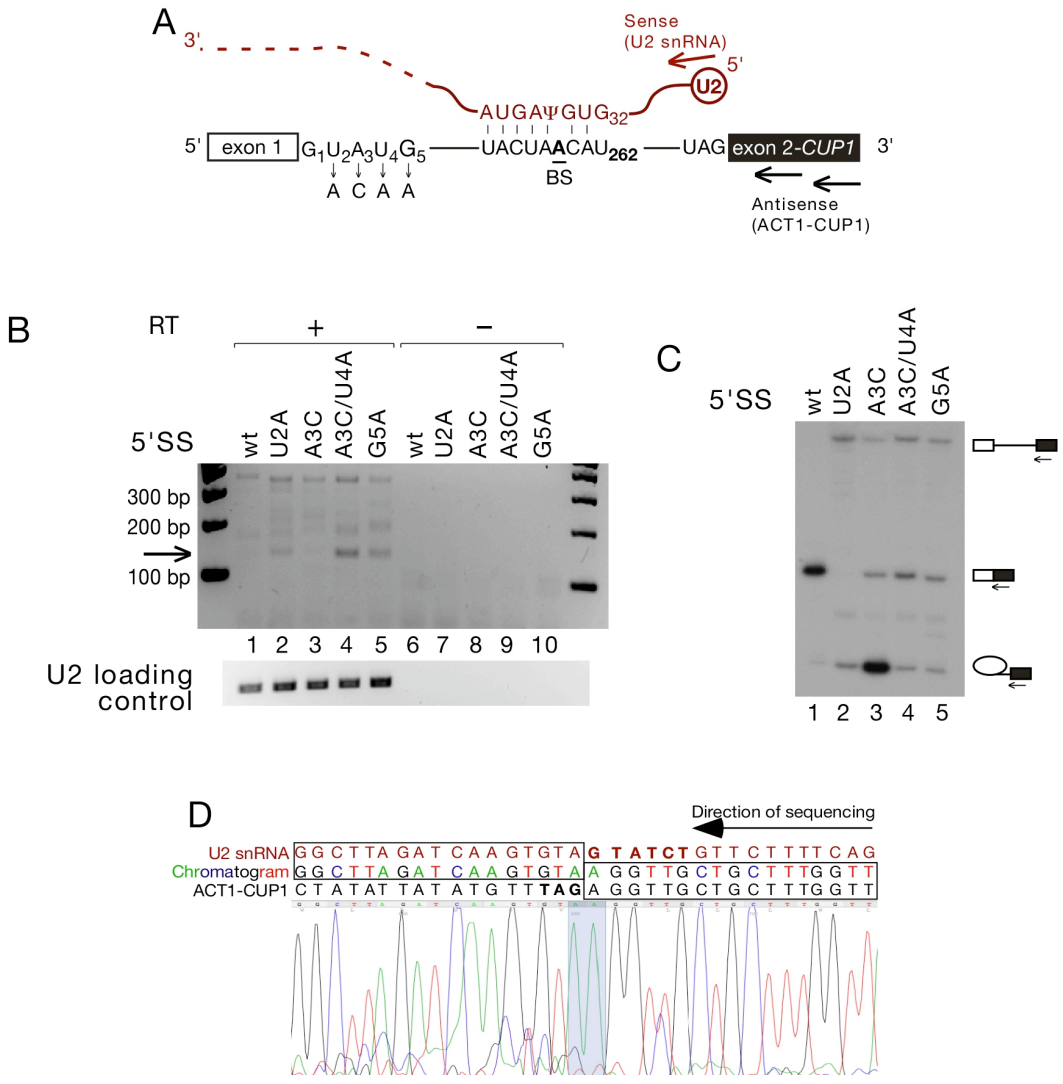
Attack of U2 snRNA by the branch site nucleophile would be expected to generate a truncated U2 molecule covalently linked via its new 5' end to the 2' hydroxyl of the BS, detectable as a strong stop by primer extension. An initial screen for such branched derivatives of U2 snRNA in *S. cerevisiae* total RNA was inconclusive due to unavoidable low levels of degradation (data not shown), which precludes the assignment of stops as branches rather than hydrolysis products. We therefore reasoned that any 'free 5' exon' generated by such aberrant attack may be able to undergo the second step of splicing in trans to the 3'SS of a reporter gene, generating a novel linear RNA detectable by RT-PCR. We performed nested RT-PCR on DNaseI-treated total RNA from strains carrying *ACT1-CUP1* reporter genes (132) with various mutations, using antisense primers complementary to the CUP1 3' exon and a sense primer corresponding to the 5' end of U2 snRNA (Fig. 3-2A). For multiple intron 5'SS

mutants (G1C, G1U, U2A, A3G, [A3C+U4A], G5A, G5C), we were able to detect a ~130 bp PCR product whose size corresponded to that expected for the *ACT1-CUP1* 3' exon appended to approximately 40 nucleotides of U2 (Fig. 3-2B, lanes 2, 4, 5 and data not shown). The 130 bp product was not generated from equivalent amounts of total RNA from strains carrying wild-type or 5'SS A3C *ACT1-CUP1* reporter, both of which show an efficient first step of canonical cis-splicing (Fig. 3-2B&C, lanes 1 and 3).

Figure 3-2. Trans-splicing in *S. cerevisiae* can generate an RNA comprising the 5' end of U2 snRNA and the 3' exon of a reporter gene.

(A). Schematic of *ACT1-CUP1* reporter pre-mRNA and U2 snRNA, indicating the mutations used in panels C and D, and the location of the RT and PCR primers (arrows) used in panel C. (B) RT-PCR using the primers indicated in Fig. 3-2A can amplify a product, of the size expected for a trans-splicing product generated using the BS-binding region of U2 as a 5'SS, from total RNA from *S. cerevisiae* Y04999 cells carrying reporters with 5'SS mutations as indicated. (C) Primer extension analysis of RNA recovered from cells containing the *ACT1-CUP1* reporters as indicated. Primer complimentary to the 3' exon was used to reveal levels of pre-mRNA, mRNA, and lariat intermediate. Strain Y04999 (Δdbr1) was used in order to accurately monitor the efficiency of the first step. (D) The 130 bp RT-PCR product corresponds to trans-spliced U2-ACT1-CUP1. Reverse-complemented sequencing trace from the purified 130 bp product. U2 snRNA and *ACT1-CUP1* sequence, as well as the chromatogram read, are indicated above the trace; concordance between chromatogram and gene sequence is indicated by boxes, and the splice junction is highlighted on the chromatogram trace

Figure 3-2


A
3'
Sense
(U2 snRNA)
5'
U2
AUGAΨGUG32
5'
exon 1
G1U2A3U4G5
UACUAACAU262
UAG
exon 2-CUP1
3'
A C A A
BS
Antisense
(ACT1-CUP1)
B
RT
+
−
5'SS
wt
U2A
A3C
A3C/U4A
G5A
300 bp
200 bp
100 bp
1 2 3 4 5 6 7 8 9 10
U2 loading
control
C
5'SS
wt
U2A
A3C
A3C/U4A
G5A
1 2 3 4 5
D
Direction of sequencing
U2 snRNA GGCTTAGATCAAGTGTAGTATCTGTTCTTTTCAG
Chromatogram GGCTTAGATCAAGTGTAAGGTTGCTGCTTTGGTT
ACT1-CUP1 CTATATTATATGTTTAGAGGTTGCTGCTTTGGTT

Direct sequencing of the putative trans-splicing product indicated that it comprised the 5'-terminal 36 nt of U2 joined to the 3' exon of *ACT1-CUP1* (Fig. 3-2D). The homogeneity of the sequencing trace indicates a strong splice site preference, with the wild-type *ACT1-CUP1* 3'SS and the A/GUAUCΨ site in U2 (the best match in this region of U2 to G/GUAUGU, the *S. cerevisiae* consensus) apparently used substantially more efficiently than other sites.

The trans-splicing reaction appears to be very inefficient, with steady state levels of the U2-ACT1-CUP1 product apparently lower than one molecule per cell based on RT-PCR standard curves (data not shown). We estimate that, in the case of the G5A mutant, trans-splicing represents around one per hundred thousand two-step splicing events, comparable to the error rates of other macromolecular processes such as transcription.

The generation of the U2-ACT1-CUP1 RNA by an aberrant but otherwise mechanistically normal splicing reaction should also produce a second species – a branched RNA comprising the 5' exon and intron of *ACT1-CUP1* joined via the 2' hydroxyl of BS-A to the 3' 1139 nt of U2, and an intermediate consisting of the '5' exon' portion of U2 with a 3' hydroxyl; although we attempted to detect these products using RT-PCR- and Klenow-based methods, respectively, we were unable to do so. This may in part be due to the rapid degradation of internally cleaved U2 snRNAs (133).

TRANS-SPLICING TO U2 SNRNA REQUIRES POOR BINDING OF THE CANONICAL 5'SS AT THE SPLICEOSOMAL CATALYTIC CENTRE

To investigate whether trans-splicing to U2 snRNA is simply a result of inhibition of the first step of the canonical splicing reaction, we analysed the effects of non-5'SS reporter gene mutations on the efficiency of trans-splicing. Mutations at and around BS (C256G, U257A, U257C, BS-C, BS-G – Fig. 3-3A), which virtually abolish productive splicing of the reporter transcript (Fig. 3-3C) did not promote trans-splicing (Fig. 3-3B, lanes 1-7). When combined with a 5'SS G5A mutation, which alone showed easily detectable levels of trans-splicing, diverse mutations at and around BS inhibited the trans-splicing reaction to the extent that the product was no longer detectable (Fig. 3-3B, lanes 15-19). Even mutations that show only mild inhibition of the first step of canonical cis-splicing of G5A mutants, such as an extension of the pairing interaction between BS and U2 snRNA, abolished trans-splicing (Fig. 3-3B&C, lanes 15-16). 3'SS mutants exhibited behaviour identical to that of BS mutants, failing to promote trans-splicing in an otherwise wild-type reporter, and reducing it to undetectable levels in the context of a G1C mutation (data not shown). BS and 3'SS mutations exacerbate the inhibition of canonical mRNA splicing caused by the G5A mutation (Fig. 3-3C, lanes 15-19 and data not shown), so their failure to further stimulate the aberrant trans-splicing reaction indicates that the requirements for this reaction extend beyond simple first step inhibition.

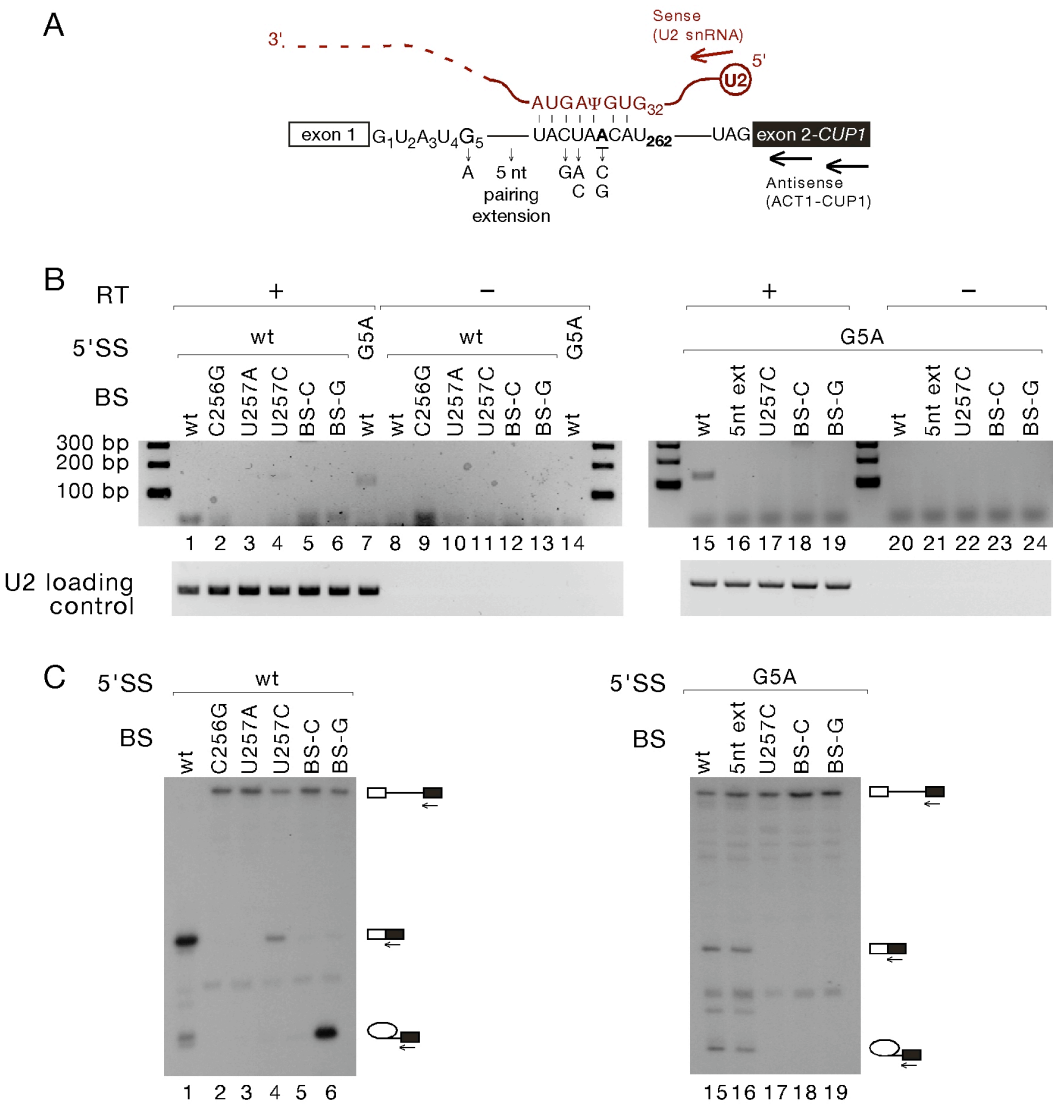


Figure 3-3. Trans-splicing does not occur in reporter genes with mutations outside the 5'SS.

(A) Schematic of *ACT1-CUP1* reporter pre-mRNA and U2 snRNA, indicating the mutations used in panels C and D, and the schematic location of RT and PCR primers (arrows) used in panel B. (B) Mutations at and around the branch site do not stimulate trans-splicing in an otherwise wild-type context, and inhibit it in the context of an accompanying 5'SS mutation. RT-PCR analysis of RNA recovered from Y04999 cells containing the *ACT1-CUP1* reporters as indicated, as in Fig. 3-2B. (C). Primer extension analysis of RNA recovered from Y04999 cells containing the *ACT1-CUP1* reporters as indicated, as in Fig. 3-2C.

Even among 5'SS mutants, the degree of first step inhibition is not a good predictor of trans-splicing efficiency. For example, splicing is much more strongly

inhibited by the 5'SS U2A mutation than by [A3C+U4A] (Fig. 3-2C, lanes 2 and 4), yet the extent of trans-splicing to U2 snRNA is greater in the [A3C+U4A] mutant (Fig. 3-2B, lanes 2 and 4). [A3C+U4A] has been shown to affect splicing predominantly by altering 5'SS binding at the catalytic centre of the spliceosome (44). Formation of the observed U2-ACT1-CUP1 RNA therefore appears to result from the aberrant attack by BS on U2 in the presence of a 5'SS that is poorly bound within the spliceosome catalytic centre, rather than simply from an inefficient first step of splicing.

THE FUNCTIONAL U2 SNRNA COMPONENT OF THE SPLICEOSOME IS THE SUBSTRATE FOR TRANS-SPLICING

To verify that the U2 snRNA portion of the trans-spliced product was derived from the U2 component of a functional spliceosome (rather than, for example, a second molecule of U2 displacing the reporter gene 5'SS) we took advantage of previously characterised compensatory mutations in U2 and U6 snRNAs. Mutation of U6 positions 56 and 57 from AU to UA is lethal except in the context of the compensatory, and likewise lethal in isolation, AU to UA mutations at positions 27 and 28 in U2, predicted to restore pairing in helix Ib (61) (Fig. 3-4A). *S. cerevisiae* U2/U6 double knockout strains were constructed in which plasmid-borne copies of both wild-type and mutant U2, and either wild-type or mutant U6 were present. In these strains, the identity of the U2 snRNA component of functional spliceosomes is expected to depend on the identity of the U6 snRNA gene carried, i.e. spliceosomes in the strain with wild-type U6 should contain wild-type U2, and those in the strain with mutant U6 should contain mutant U2

(Fig. 3-4B&C). In both cases, the pool of 'free U2', as well as early spliceosome assembly intermediates, should contain both wild-type and mutant U2. The 'wild-type spliceosome' and 'mutant spliceosome' strains were transfected with G5A *ACT1-CUP1* reporter, and the U2-ACT1-CUP1 trans-spliced product from each strain amplified by RT-PCR. Sequencing indicated that the U2 snRNA portion of the trans-spliced product differed as predicted. The strain carrying a wild-type U6 snRNA gene generated a trans-spliced product in which U2 sequence was wild-type (Fig. 3-4D), and that carrying the mutant U6 snRNA gene generated product with mutant U2 sequence (Fig. 3-4E), indicating that the trans-splicing reaction involves nucleophilic attack of the U2 snRNA component of a fully assembled, catalytically competent spliceosome, and therefore of the BS-binding region of U2 to which BS was bound.

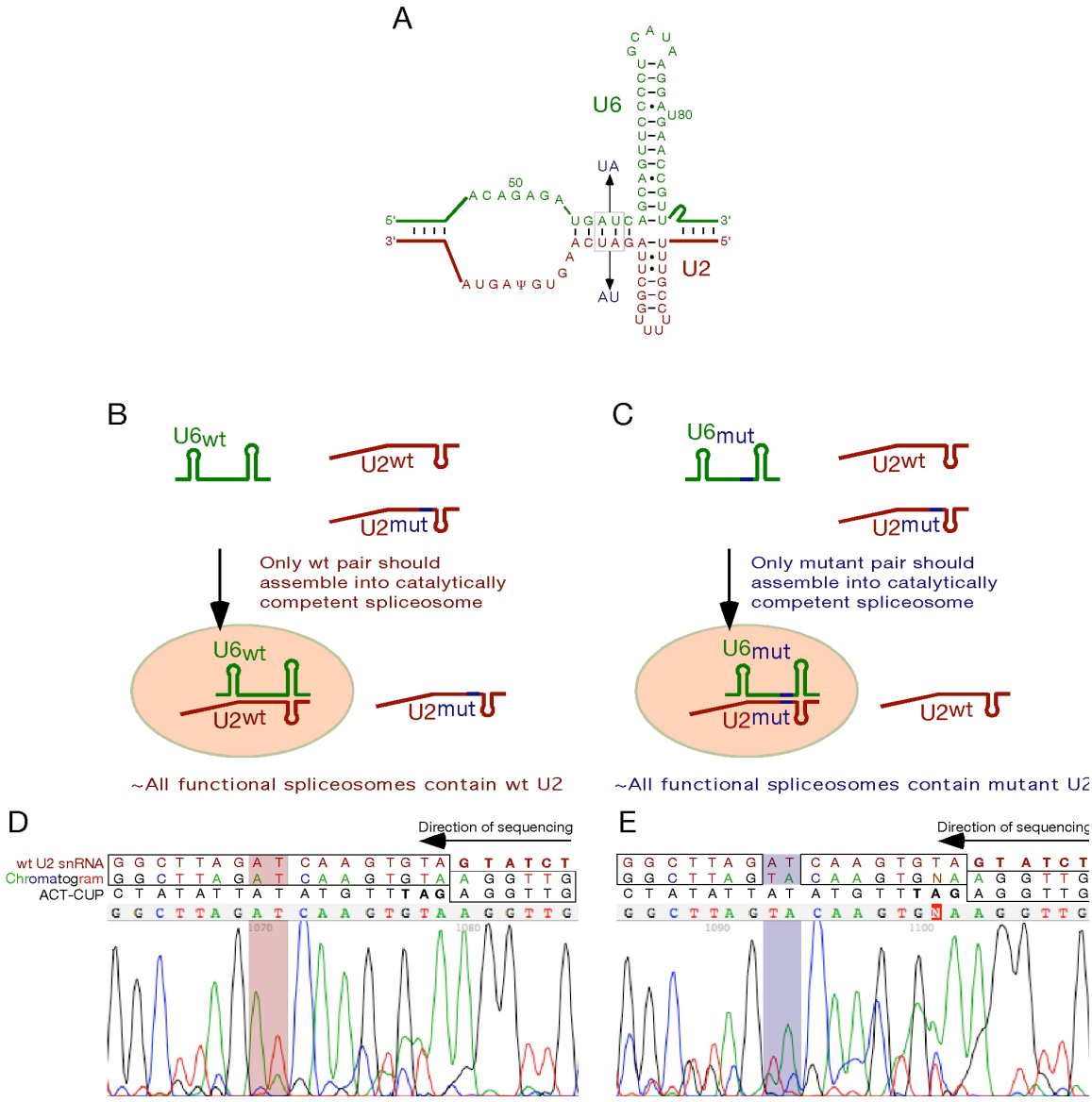


Figure 3-4. U2 snRNA in functional spliceosomes is the substrate for trans-splicing.

(A). Compensatory U2 and U6 snRNA mutations used in panels C and D. (B, C) Predicted behaviour of wild-type and mutant U2 snRNAs in strains carrying a wild-type [B] or mutant [C] U6 snRNA gene. (D, E) The trans-spliced product is generated by attack of U2 snRNA in functional spliceosomes. Reverse-complemented sequencing traces from purified trans-splicing product from yCQ62 cells carrying G5A *ACT1-CUP1* reporter, both wild-type and mutant U2 snRNA genes, and a wild-type [D] or mutant [E] U6 snRNA gene. Concordance between chromatogram and wild-type gene sequence is indicated by boxes, and the mutated/wild-type helix Ib U2 snRNA nucleotides highlighted.

SPLICING CAN OCCUR TO ALL THREE PREDICTED 5'SS IN U2 SNRNA

Although direct sequencing of the purified trans-splicing product had indicated the preferred use of CAAGUGΨA/GUAUCΨ as a 5'SS, and of the wild-type *ACT1-CUP1* 3'SS, we reasoned that the use of less-favoured sites may be apparent if individual DNAs were analysed. We therefore cloned the product into the pDrive sequencing vector (Qiagen), and sequenced multiple clones. As expected, the majority of these clones (9/13) had used the preferred 5'SS in U2; however, one had used the GΨ/AG site one nt upstream, two the GU/GU site a further two nt upstream of this, and one the AA/GU a further two nt upstream (splice site usage in U2 snRNA is shown in Fig. 3-5E). All clones had used the wild-type *ACT1-CUP1* 3'SS. Due to our inability to detect reaction intermediates, we do not know whether the levels of trans-spliced products reflecting the use of distinct 5' splice sites within U2 snRNA indicate differences in the efficiencies of the first step, the second step, or both, for each potential 5'SS. The 5' splice sites observed to be used span half a helical turn of RNA: the use of such spatially disparate 5' splice sites is inconsistent with nucleophilic attack by the canonical BS-A within an intact, unaltered and undisrupted BS-U2 duplex. Given that, even in the context of BS mutations that dramatically reduce splicing efficiency (BS-C, BS-U), *S. cerevisiae* introns show a strong preference for branching from the mutated canonical BS nucleotide rather than the adjacent adenosine (134), at least some transient unwinding of the BS-U2 duplex seems more likely than aberrant attack by a non-canonical BS nucleotide.

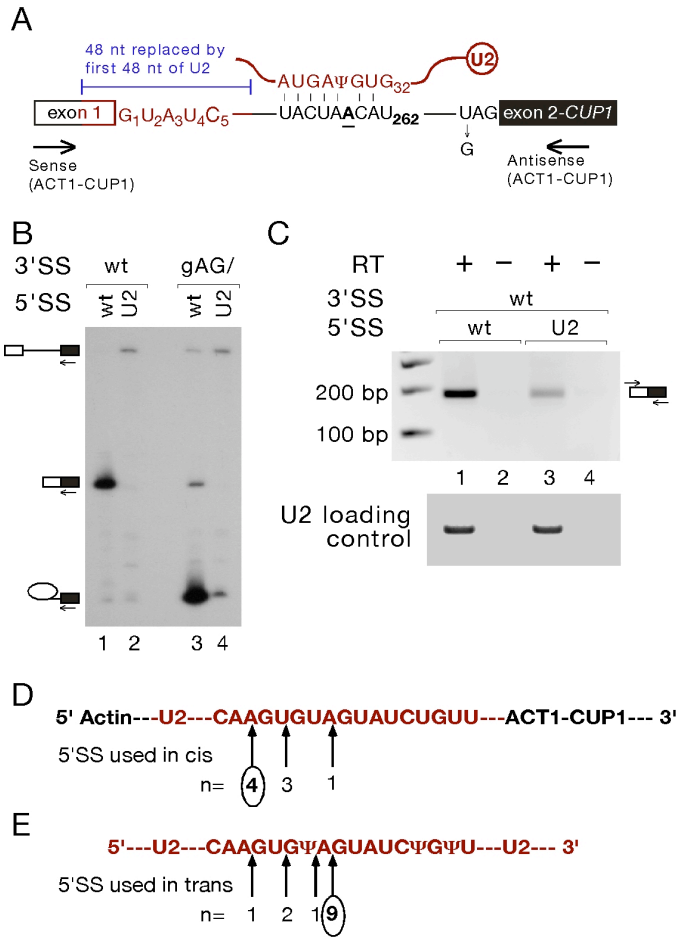


Figure 3-5. U2 snRNA sequence is a suboptimal 5'SS in the context of a reporter gene.

(A). Schematic of *ACT1-CUP1* reporter pre-mRNA and U2 snRNA, indicating the U2 sequence substitution into *ACT1-CUP1* and 3'SS gAG/ mutation used in panels B and C, and the RT-PCR primers used in panel C. (B) Primer extension analysis of RNA recovered from Y04999 cells containing the *ACT1-CUP1* reporters as indicated, as in Fig. 3-2C. (C) Spliced ACT1-U2-ACT1-CUP1 mRNA can be detected by RT-PCR (lane 3). Wild-type mRNA is shown for comparison (lane 1). (D, E) Schematic indicating 5'SS usage in U2 sequence in cis in the context of *ACT1-U2-ACT1-CUP1* [D] and in trans in the context of U2 snRNA [E]. n refers to the number of plasmids sequenced containing trans-spliced product that had used the 5'SS indicated by the arrow above, and circles indicate the predominant 5'SS detected in directly sequenced PCR products.

THE BS-BINDING REGION OF U2, IN THE CONTEXT OF THE *ACT1-CUP1* REPORTER GENE, IS INEFFICIENTLY USED IN CIS AS A 5'SS AND SHOWS A 5'SS PREFERENCE DISTINCT FROM THAT OBSERVED IN TRANS-SPLICING

Given the tandem arrangement of three 5'SS-like sequences in the BS-binding region of U2 snRNA, and their use as 5' splice sites in the context of reporter gene mutations such as [A3C+U4A] and G5A that do not dramatically reduce the overall efficiency of cis-splicing, we investigated the efficiency with which U2 sequence could be used as a 5'SS in the context of the *ACT1-CUP1* reporter. 48 nt of reporter sequence around the 5'SS was replaced with sequence corresponding to the first 48 nt of *S. cerevisiae* U2 snRNA, such that the /GUAUCU 5'SS from U2 replaced the /GUAUGU 5'SS in the reporter (Fig. 3-5A). Primer extension analysis of total RNA from strains expressing these *ACT1-U2-ACT1-CUP1* constructs revealed that this U2-substituted 5'SS was used inefficiently (Fig. 3-5B). A small amount of spliced product could be detected by RT-PCR (Fig. 3-5C); this RT-PCR product was cloned into pDrive (Qiagen). Sequencing of eight clones revealed that 5'SS usage in U2 sequence in the context of the reporter gene (Fig. 3-5D) was different from that in the context of U2 (Fig. 3-5E). 5'SS usage in the U2-substituted reporter gene correlated relatively well with predicted U1 snRNA binding affinity (135), with the upstream-most 5'SS (AA/GUGUA) used in 4/8 clones and the downstream-most 5'SS (UA/GUAUCU) in only 1/8 (cf. 9/13 in the context of trans-splicing to U2). This suggests that, while this region of U2 snRNA sequence can be recognised by U1 snRNA and participate to a low level in a canonical mRNA splicing reaction when placed in the context of an otherwise wild-type reporter gene, it is a suboptimal

substrate for splicing. This rationalises the observation that U2 cannot compete effectively with a wild-type 5'SS. In addition, the different 5'SS preference observed between the cis- and trans-splicing reactions indicates that pairing between U1 and U2 snRNAs does not determine 5'SS usage in U2 snRNA in trans-splicing; the most likely candidate for such a determinant of 5'SS preference in trans-splicing is spatial proximity to the BS nucleophile. The use of multiple 5' splice sites in the trans-splicing reaction is therefore indicative of a high degree of flexibility in the spliceosomal catalytic centre.

DISCUSSION

THE NATURE OF THE TRANS-SPLICING REACTION

The data presented in this chapter demonstrate the formation of a trans-spliced RNA product comprising the 5' end of U2 and the 3' exon of an *ACT1-CUP1* reporter with a 5'SS predicted to bind poorly at the spliceosome catalytic centre. The possibility of this product being generated by an RT artefact is ruled out by its strict dependence on 5'SS mutation. In strains expressing two distinct U2 snRNAs, the identity of the U2 portion of the trans-spliced product was observed to be dependent on the identity of the U6 expressed by the strain, with the potential for helix I formation between U2 and U6 an essential requirement for trans-splicing. Interactions between U2 and U6 are established during the late stages of spliceosome assembly, so the requirement for such interactions

indicates that U2 snRNA within a functional spliceosome is nucleophilically attacked, presumably by the branch site it recruited.

We were able to detect trans-splicing using reporter genes with wild-type 3' splice sites, but not with 3'SS mutants. In addition, all trans-splicing events observed involved the same 3'SS as is used in canonical cis-splicing of the reporter gene. These observations support the hypothesis that the trans-splicing reaction is essentially identical to cis-splicing, but with an aberrant first catalytic step in which U2, rather than the 5'SS, is attacked; the exon generated by this aberrant first step proceeds through the second step in a quasi-normal fashion. As we were unable to detect reaction intermediates, the efficiency of this second step could not be investigated. Our inability to detect a branched intermediate also renders uncertain the identity of the first-step nucleophile, although the inhibition of trans-splicing by mutations at and around BS supports a 2'-5' branching reaction. A hydrolytic first step analogous to that observed in group II introns (136) cannot be ruled out by our data, but in any case the existence of such a first step would not substantially affect the interpretation of these data.

FLEXIBILITY IN THE SPLICEOSOME CATALYTIC CENTRE, AND THE FATE OF THE BS-U2 DUPLEX FOR THE FIRST CATALYTIC STEP OF PRE-MRNA SPLICING

Splicing is a multi-step reaction that necessarily involves conformational rearrangements as well as complex assembly and disassembly phases; global conformational flexibility is clearly required for such a reaction. In addition to this

necessary flexibility, however, splicing is characterised by the need for nucleotide precision; this is especially true in *S. cerevisiae*, an organism with strong splice site consensuses in which functionally important alternative splicing has not been demonstrated. Therefore, while both global flexibility throughout the assembly-reaction-disassembly pathway, and local flexibility at the catalytic centre between the two chemical steps are inevitable, tight substrate binding between rearrangements is expected. As noted by Nilsen and co-workers (130), snRNA reactivity in spliceosomes is indicative of catalytic centre flexibility. The reactivity of an snRNA region intimately involved in substrate binding during assembly, and possibly also catalysis, extends our view of this flexibility to include pairing interactions important for splicing fidelity. The nucleophilic attack of four sites in U2 is inconsistent with reactivity within a BS-U2 duplex that has remained intact from early spliceosome assembly until the catalytic phase, even in the unlikely scenario of a hydrolytic first step. To rationalise the reactivity of multiple sites in U2, it is therefore necessary to invoke at least transient disruption of the interaction between BS and U2. Following this unpairing, three possibilities exist between which our data do not allow us to distinguish: the branch site, the presumed nucleophile for the first step of the trans-splicing reaction, could attack U2 while in a single-stranded state; this seems unlikely, as such circumvention of the need for nucleophile activation would be highly surprising in a reaction as important and precise as splicing. A second possibility is that BS may simply re-pair somewhat inaccurately with U2, and attack from within the resulting re-formed duplex (Fig. 3-6, iia). The third possibility is that BS may pair to another

RNA strand and attack U2 from within this duplex (Fig. 3-6, iib). Given the similarity between the sequences of the yeast consensus 5'SS and the BS-binding region of U2, the 5'SS represents an obvious candidate for such a partner. While a complete strand exchange reaction, in which U2 and the 5'SS are able to swap positions and roles after assembly, does not appear consistent with the wide range of 5'SS mutations that facilitate trans-splicing, partial strand exchange cannot be ruled out.

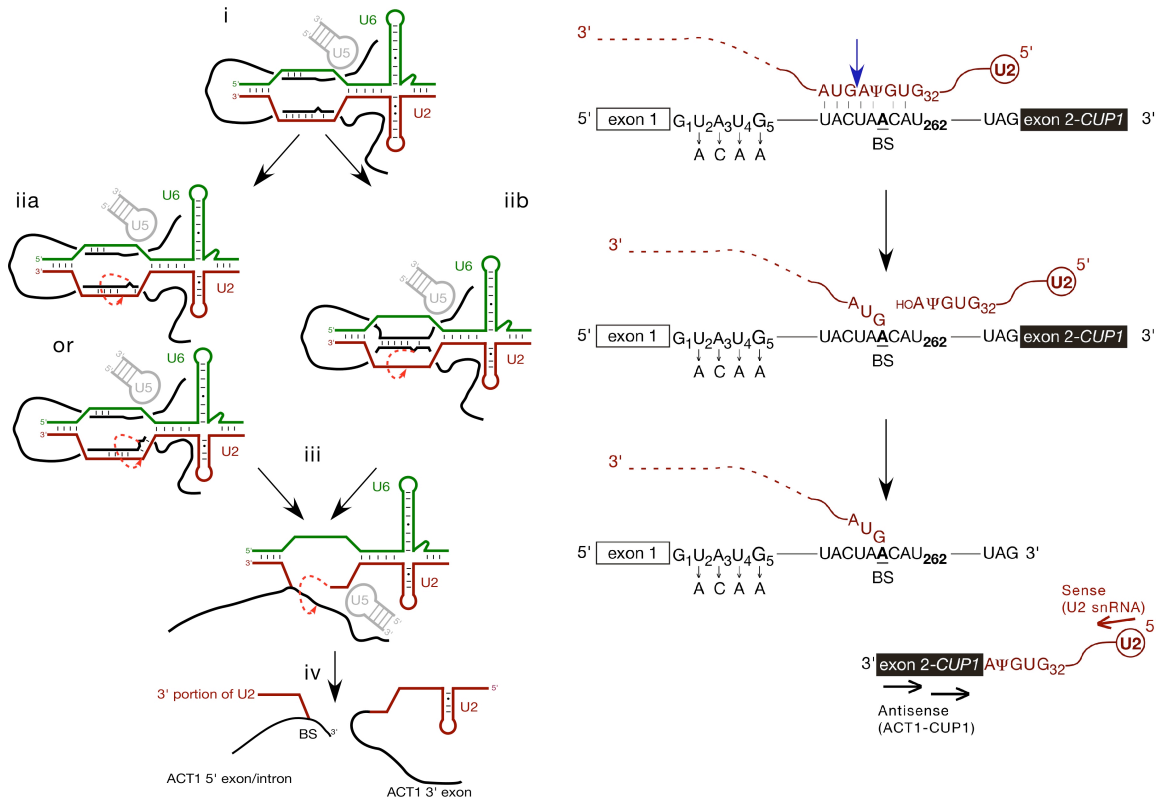


Figure 3-6. Possible trans-splicing mechanisms.

Proposed mechanisms of trans-splicing: (i) RNA-RNA interactions in the fully assembled spliceosome, with BS paired to U2. (iia) BS and U2 unpair and re-pair potentially inaccurately, with BS nucleophilically attacking U2 from within the resulting helix, or (iib) BS and U2 unpair, with new pairing established between BS and the 5'SS; BS, bulged from this new helix, attacks U2. (iii) 5'SS-U6 and BS-U2 pairing both having been disrupted, the 'U2 5' exon' generated in step ii attacks the reporter gene 3'SS, generating (iv) a branched product comprising the 3' end of U2 appended to the 5' exon and intron of the reporter gene, and a linear product comprising the 5' end of U2 appended to the reporter 3' exon.

THE FATE OF THE BS-U2 DUPLEX FOR THE SECOND CATALYTIC STEP OF PRE-MRNA SPLICING

The persistence after the first step of splicing of the entire U2-BS helix, established during the early stages of spliceosome assembly, is not inconsistent with any previously published data. Our data, however, clearly suggest that the

persistence of this helix is at least not required for the second step of splicing. As indicated in figures 3-1A and 3-5E, the predominant 5'SS used in U2 lies in the middle of this helix, which must therefore be at least partially disrupted in order for the 'U2 5' exon' to be free to attack the 3'SS during the second step. Given that the structural consequences of the absence of a helix normally present in the second step conformation of the spliceosome would likely cause a severe, if not terminal, second step block, and that the second step of the trans-splicing reaction appears to be analogous to the second step of canonical cis-splicing, we hypothesise that the U2-BS helix does not normally persist during the second step, and that the removal of the lariat intermediate, generated in the first step, from the catalytic centre prior to the second step may be linked to the unpairing of its BS portion from U2 snRNA as well as the disruption of the 5'SS-U6 snRNA interaction (44). Although these conclusions are based on an infrequent trans-splicing reaction, the capacity of the spliceosome to support such a reaction has important implications for physiologically relevant pre-mRNA splicing. We note that such *trans*-spliced products may have been identified in high-throughput RNA sequencing experiments in both yeast and higher eukaryotes, but that hybrid RNAs containing snRNA sequence would likely be filtered out by existing algorithms. Identification of further such products could provide valuable mechanistic information, and as such may be worth the generation of algorithms to mine the large amounts of raw sequence data that already exist.

CHAPTER 4. INVESTIGATION OF BRANCH NUCLEOPHILE ACTIVATION *IN VIVO* USING ORTHOGONAL BRANCH SITE-U2 SNRNA PAIRS

INTRODUCTION

The BS-U2 duplex has been proposed to play a role in nucleophile specification and activation at the stage of first step splicing catalysis, in which the 2' hydroxyl group of a conserved adenosine at the BS nucleophilically attacks the 5'SS. The conserved 5'-GΨAGUA-3' motif in U2 allows the second A in the conserved 5'-UACUA**A**C-3' branch site (the canonical BS nucleophile is highlighted) to bulge from the helix (Fig. 4-1). Substitution of the first adenosine in the AA pair with arabinosyl adenosine favours bulging of this unreactive nucleotide rather than the adjacent and normally favoured adenosine, and inhibits first step catalysis in HeLa extract (137). In addition, NMR studies on model duplexes suggest that pseudouridylation at position 35 in U2 snRNA favours the conformation of the duplex in which the BS-A is extruded from, rather than intercalated into, the helix (138). Deletion of Pus7, the pseudouridine synthase responsible for this modification in *S. cerevisiae*, confers sickness, presumably due to a splicing defect, on cells with mildly mutated U2 snRNA (139). Finally, abrogation of an analogous bulged adenosine in domain 6 of a group II self-splicing intron virtually abolishes first step catalysis by branching, with little to no negative impact on the hydrolytic first step pathway (140). Together, these studies indirectly suggest

that bulging of the BS nucleophile via interaction with U2 is important for branching catalysis by the spliceosome.

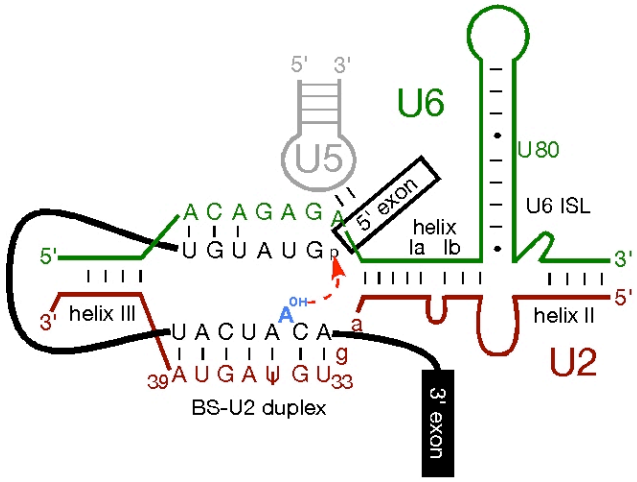


Figure 4-1. Schematic of RNA-RNA interactions during the first step of splicing.

The pre-mRNA is shown in black, with the branch site (BS) adenosine highlighted in blue: U2 snRNA is shown in red, U5 in grey and U6 in green. U2 positions A31 and G32 are shown in lowercase throughout figures in this chapter for positional reference.

The ability of the higher eukaryotic spliceosome to catalyse 2'-5' branch formation using either adenosine in -U**AA**C as a nucleophile (137), and poor conservation of both the BS and 5'SS in higher eukaryotes (2) are indicative of flexibility with respect to the selection and activation of nucleophile and electrophile. Yeast, by contrast, have strong splice site consensus sequences and appear to have strict positional requirements for the BS nucleophile: mutation of the BS-A to a pyrimidine, instead of activating the upstream adenosine, results in inefficient splicing from the mutated nucleotide (141). However, the observation, described in chapter 3, that multiple sites within the BS-binding region of U2 snRNA can be used in *trans* as 5' splice sites for both

steps of splicing *in vivo* suggests a fairly large degree of active site flexibility, and implies that the BS-U2 interaction might be more dynamic and/or transient than previously thought.

The data in this chapter demonstrate the importance of nucleophile bulging for splicing catalysis. From the systematic movement of a bulge motif through the BS-U2 duplex *in vivo* in *S. cerevisiae*, we conclude that adenosines bulged from at least three positions within the BS-U2 duplex are competent to act as the first-step nucleophile. Mutations in the BS-binding region of U2 have unpredictable effects on splicing and cell growth. An orthogonal system is therefore desirable in which a constant second copy U2 snRNA, capable of splicing a reporter transcript but not of interacting with endogenous introns, can be used for the *in vivo* investigation of nucleophile positioning. We report the development and characterisation of such a BS-U2 system. We show unambiguously that three bulge positions can participate in first step catalysis, and confirm the importance of nucleotide identity at the branch site. We also show that the branch site and 5'SS are positioned independently for first step catalysis, and that the favoured position of the branch nucleophile is dependent on distance from the U2 component of U2/U6 helix Ia.

RESULTS

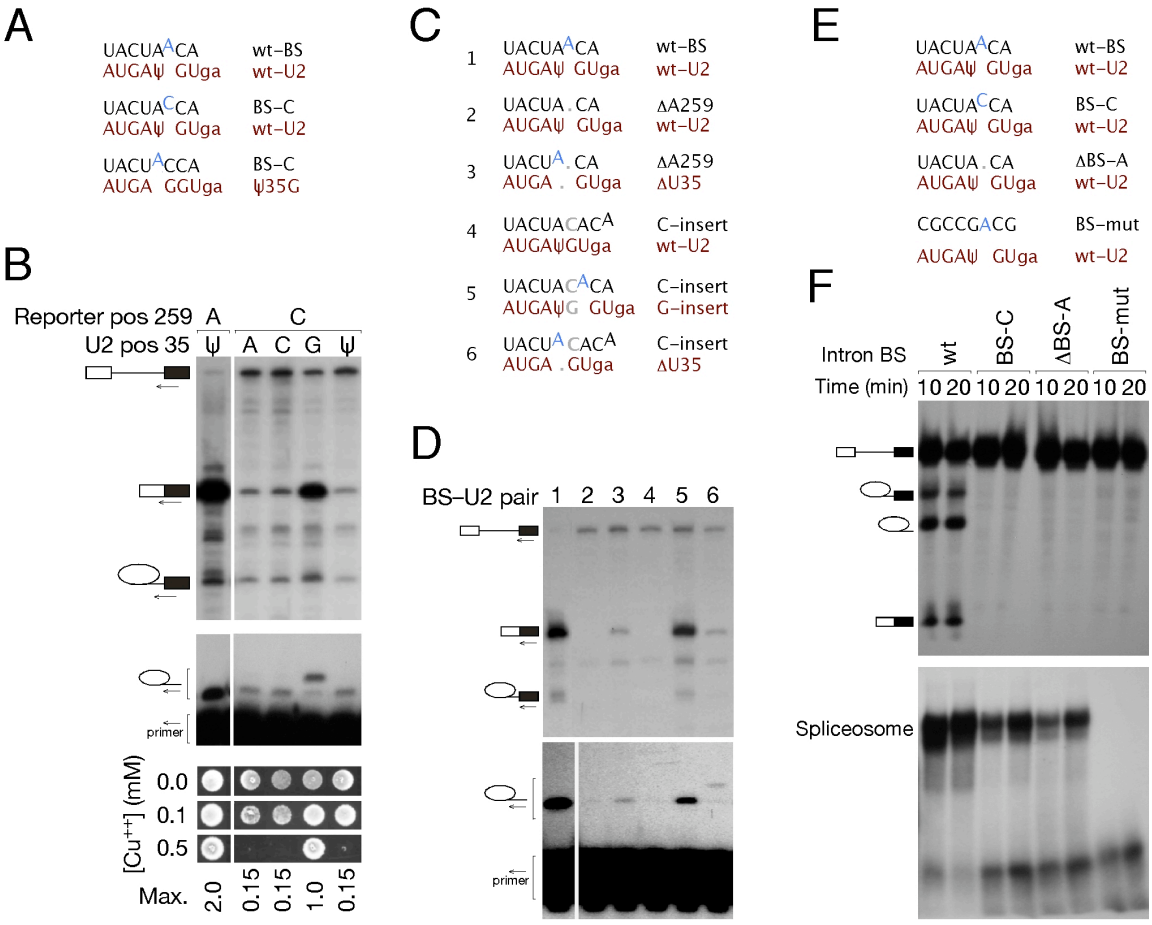
BULGING IS REQUIRED FOR SPLICING, BUT THE POSITION OF THE BULGE WITHIN THE BS-U2 DUPLEX IS FLEXIBLE

The *trans*-splicing reaction described in chapter 3 suggested global flexibility in the *S. cerevisiae* spliceosome: further evidence of flexibility came from the result of a mutagenesis screen for U2-based suppressors of splicing defects due to mutation of the branch site nucleotide. Because U2 is essential and sensitive to mutation in its BS-binding region, this mutagenesis screen and all subsequent experiments involving mutant U2 snRNAs were performed using a strain carrying both a wild-type copy of U2 and a second gene encoding the mutant. One U2 allele capable of suppressing the splicing defect of BS-C (5'-UACUACC-3') was identified as Ψ35G (5'-GGAGUA-3'). Other mutations at this position failed to suppress BS-C (Fig. 4-2A&B), and other BS mutations could be suppressed by analogous mutations of position 35 to the Watson-Crick base-pairing partner of the mutated BS nucleotide (data not shown). This suggested that the splicing defect due to BS mutation was overcome via bulging and activation of the preceding adenosine – a conclusion confirmed by primer extension analysed at single nucleotide resolution (Fig. 4-2B, middle panel). The *ACT1-CUP1* reporter encodes a metallothionein whose expression correlates over a large dynamic range with resistance to copper in the growth medium (132). All primer extension data in this and subsequent experiments were confirmed by copper growth assays (Fig. 4-2B, lower panel, and data not shown); this verifies that mRNA is released from the spliceosome, exported to the cytoplasm and translated.

Figure 4-2. Bulging of the branch site nucleophile is required for splicing in *S. cerevisiae*: the position of the bulge within the BS-U2 duplex is flexible.

(A) Schematic of the BS-U2 pairing regions of the wt and BS-C mutant reporters, and wt and Ψ35G U2 snRNAs used in [B]. (B) An intron BS-C mutation can be suppressed by U2 Ψ35G but not by other mutations at this position: upper panel – splicing efficiency monitored by primer extension with a 3' exon primer on *S. cerevisiae* total RNA from strains carrying the indicated U2 snRNA and reporter constructs; middle panel – primer extension with an intronic primer immediately upstream of BS indicates that the Ψ35G suppressor promotes the use of a non-canonical adenosine as the branch nucleophile; lower panel – copper growth assay to verify that the mRNA observed in the upper panel is functional. (C) Schematic of reporter and U2 constructs used in [D]. (D) Primer extensions with 3' exon (upper panel) and BS-proximal intronic (lower panel) primers on total RNA from the indicated strains indicate that branch nucleotide bulging is required for splicing, but the position of the bulge within the BS-U2 duplex is not fixed. (E) Schematic of reporter and U2 constructs used in [F]. (F) *In vitro* splicing (upper panel) and spliceosome assembly (lower panel) of the *S. cerevisiae* UBC4 intron carrying the indicated BS mutations, as assayed by denaturing and native PAGE, respectively. BS-C and ΔBS-A mutations allow reduced spliceosome assembly but no branching, indicating the importance of the bulged nucleotide for spliceosomal catalysis.

Figure 4-2


A
UACUAACA wt-BS
AUGAψ GUga wt-U2
UACUACCA BS-C
AUGAψ GUga wt-U2
UACUACCA BS-C
AUGA GGUga ψ35G
B
Reporter pos 259 A C
U2 pos 35 ψ A C G ψ
primer
[Cu++] (mM)
0.0
0.1
0.5
Max. 2.0 0.15 0.15 1.0 0.15
C
1 UACUAACA wt-BS
AUGAψ GUga wt-U2
2 UACUA.CA ΔA259
AUGAψ GUga wt-U2
3 UACUA.CA ΔA259
AUGA . GUga ΔU35
4 UACUACACA C-insert
AUGAψGUga wt-U2
5 UACUACACA C-insert
AUGAψG GUga G-insert
6 UACUACACA C-insert
AUGA .GUga ΔU35
D
BS–U2 pair 1 2 3 4 5 6
primer
E
UACUAACA wt-BS
AUGAψ GUga wt-U2
UACUACCA BS-C
AUGAψ GUga wt-U2
UACUA.CA ΔBS-A
AUGAψ GUga wt-U2
CGCCGACG BS-mut
AUGAψ GUga wt-U2
F
Intron BS wt BS-C ΔBS-A BS-mut
Time (min) 10 20 10 20 10 20 10 20
Spliceosome

The use of a positionally non-canonical BS nucleotide does not normally occur in *S. cerevisiae*; our data suggested both that bulging of the BS was likely required for catalysis as well as assembly, and that *S. cerevisiae* could be induced to display the same flexibility as higher eukaryotes with respect to branch nucleophile positioning. We introduced pairs of mutations into the BS and a second copy U2 snRNA such that the bulge was flattened by insertion or deletion of a nucleotide in the BS sequence, and restored by corresponding deletion or insertion in the second copy U2 (Fig. 4-2C). As expected for an essential bulge, introns that generated flat duplexes failed to support splicing (Fig. 4-2D, lanes 2&4), and restoration of the bulge via either insertion or deletion in U2 snRNA restored splicing activity, albeit with reduced efficiency relative to the wt BS-U2 pair (Fig. 4-2D, lanes 3, 5&6). To distinguish between assembly and catalytic phases of splicing, we assayed spliceosome assembly and splicing progression *in vitro* using *S. cerevisiae* whole cell extract and the efficiently spliced UBC4 transcript (John Abelson, personal communication) carrying a variety of mutations at and around the BS. Analysis of splicing reactions in native and denaturing polyacrylamide gels indicated that deletion of the branch nucleophile to produce a flat BS-U2 duplex allowed spliceosome assembly at a level comparable to that shown by an intron carrying a BS-C mutation: neither of these mutant introns, however, showed detectable splicing catalysis, in contrast to the efficient splicing of the wt UBC4 intron (Fig. 4-2E&F). As expected, an intron with a severely mutated BS region failed to elicit detectable spliceosome assembly

(Fig. 4-2E&F). These results indicate that both the presence and the nature of the bulged nucleophile are important for splicing catalysis as well as assembly.

High-resolution primer extension of *in vivo* splicing indicated that the expected branch nucleophile was used for each BS-U2 pair that showed detectable splicing (Fig. 4-2D, lower panel – the background band at the wt position is due to extension on endogenous actin), indicating that changing the length of the BS-U2 duplex, and correspondingly moving the bulged adenosine towards or away from its helix III-proximal end (Fig. 4-1, 4-2B&C), does not abolish the ability of this adenosine to act as the nucleophile for the first step of splicing in this context.

SYSTEMATIC MOVEMENT OF A CAC MOTIF THROUGH THE BS-U2 DUPLEX INDICATES THAT THREE DUPLEX BULGE POSITIONS CAN PARTICIPATE IN CATALYSIS

To further investigate the apparent flexibility of nucleophile location within the BS-U2 duplex, we generated a series of BS-U2 pairs in which a 5'-CAC-3' motif was systematically placed at all positions in the BS-U2 duplex opposite a 5'-GG-3' motif in the second copy U2 to produce a bulge whose nature remained constant while its position within the helix was varied (Fig. 4-3A). Primer extension analysis of total RNA isolated from strains carrying these BS-U2 pairs indicated that a bulged adenosine placed at any of three positions in the duplex could participate in first step splicing catalysis (Fig. 4-3B). If the canonical BS position is defined as 0, these positions are -1 (immediately upstream of the BS in the

pre-mRNA), 0 and +1 (immediately downstream) (Fig. 4-3B, lanes 4-7). Bulged adenosines further up- or downstream showed no reactivity (Fig. 4-3B, lanes 2-3 and data not shown).

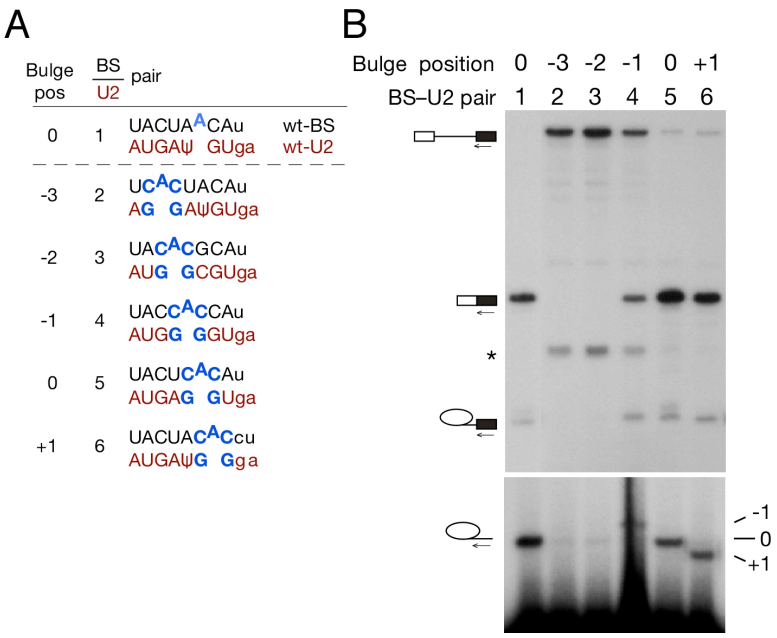


Figure 4-3. Systematic variation of bulge position indicates that at least three positions within the BS-U2 duplex are competent for first step catalysis.

(A) Schematic of reporter and U2 constructs used in [B]. (B) Primer extension with a 3' exon primer indicates that the -1, 0 and +1 BS bulge positions allow splicing catalysis. In this and subsequent figures, a band resulting from degradation of pre-mRNA is marked by an asterisk.

GROSSLY SUBSTITUTED BS-U2 PAIRS FUNCTION IN SPLICING: DEVELOPMENT OF AN ORTHOGONAL BS-U2 SYSTEM

The BS-binding region of U2 snRNA is universally conserved in crown group eukaryotes, although U2s from highly divergent eukaryotes (e.g. kinetoplastids) have different sequences (Fig. 4-4A): minor changes both in this sequence and

throughout the highly conserved 5' end of U2 are generally lethal as a sole copy and can impair cell growth as a second copy (124, 142). These negative effects, together with potentially variable expression levels of mutant U2s, preclude direct, quantitative comparison of branching efficiency from different BS-U2 duplex positions, as a different U2 is required for each reporter gene. We sought to generate BS-U2 pairs to allow the fair comparison of different branch positions and nucleophiles within the context of a constant U2 sequence. These second copy U2 snRNAs were designed to be orthogonal to wt U2 – that is, they would interact with the reporter transcript but not with endogenous introns, while the reporter transcript in turn would react only with the second copy and not with wt U2.

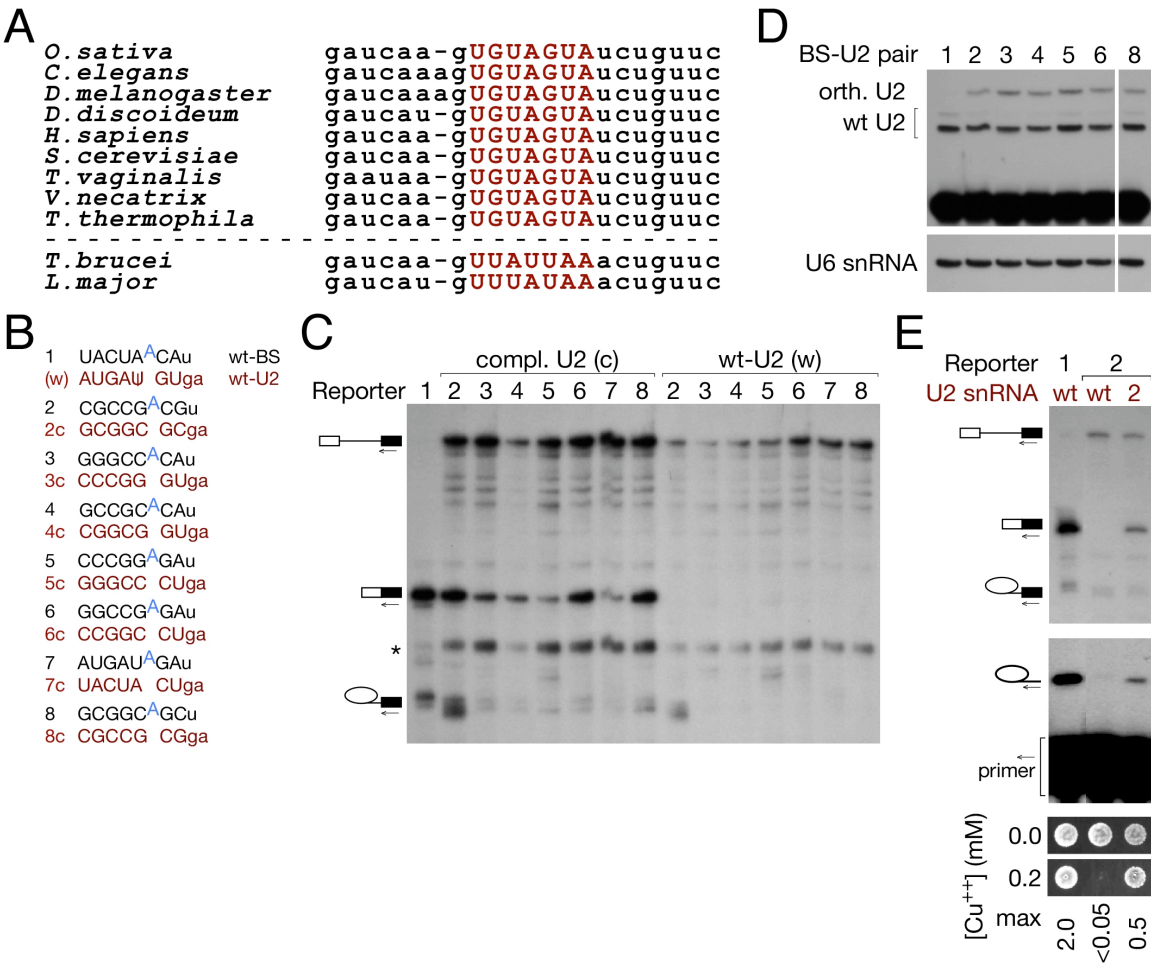


Figure 4-4. A wide variety of grossly substituted BS-U2 pairs are functional for splicing, and are orthogonal to the endogenous splicing machinery.

(A) Alignment of U2 snRNA sequences from (*S. cerevisiae*) positions 26 to 49 from diverse organisms: the BS-binding region, highlighted in red, is 100% conserved in crown group eukaryotes. (B) Schematic of wt and grossly substituted reporter and U2 constructs used in [C]. (C) Primer extension with a 3' exon primer indicates that grossly substituted branch sites allow splicing catalysis in the presence of their cognate U2 (c) (numbers refer to [B]), but not wt U2 (w). (D) Primer extension with ddTTP replacing dTTP to assay the expression of U2 snRNAs with wild-type and GC-substituted (orthogonal) BS-binding regions. (E) Primer extension on total RNA from the indicated strains with a BS-proximal intronic primer (middle panel) and copper growth (lower panel) indicate that the expected branch nucleotide is used and that the mRNA resulting from splicing is functional, for representative grossly substituted BS-U2 duplex 2 from [B]

We generated several U2 snRNAs in which the entire BS-binding region was substituted, typically with only C and G nucleotides, together with corresponding reporter transcripts for each (a selection is shown schematically in Fig. 4-4B). The U2s that we tested included ones in which all nucleotides were mutated to C or G, with purine/pyrimidine identity preserved (Fig. 4-4B, #2) or transverted (#8) at each position, and one in which the sequences of the BS and U2 were effectively swapped such that the mutant U2 sequence corresponded to that of a flattened branch site and the reporter to that of a bulged U2 snRNA (Fig. 4-4B, #7). Virtually all grossly substituted BS-U2 pairs could participate in both steps of splicing, although the efficiency of the first step was variable (Fig. 4-4C, lanes 2-8(c), and data not shown). As expected, the substituted reporter failed to splice in the absence of its corresponding U2 (Fig. 4-4C, lanes 2-8(w), and data not shown). For all BS-U2 pairs, we verified by high-resolution primer extension that the anticipated branch nucleophile was used, and confirmed by copper growth (representative data for BS-U2 pair #2 are shown in Fig. 4-4E; data not shown for other constructs). Thus, despite its extreme conservation, the BS-U2 duplex is remarkably tolerant to substitution.

Using a primer extension assay in which the primer abutted the BS-binding sequence in U2, and ddTTP replaced dTTP, we were able to distinguish between, and thus assay the levels of, wt and most orthogonal U2 snRNAs (Fig. 4-4D); we presume these to reflect the level of U2 snRNP, as we expect unassembled U2 snRNA to be rapidly degraded. Orthogonal U2 expression,

while variable, shows no correlation with splicing efficiency (Fig. 4-4C&D, quantitation in Fig. 4-5C), suggesting that U2 snRNA expression does not normally limit splicing in our system. The first step defect indicated by the accumulation of pre-mRNA (Fig. 4-4C) may instead reflect suboptimal packing of substituted BS-U2 duplexes with other spliceosome components. This is discussed more fully below in the context of Fig. 4-9.

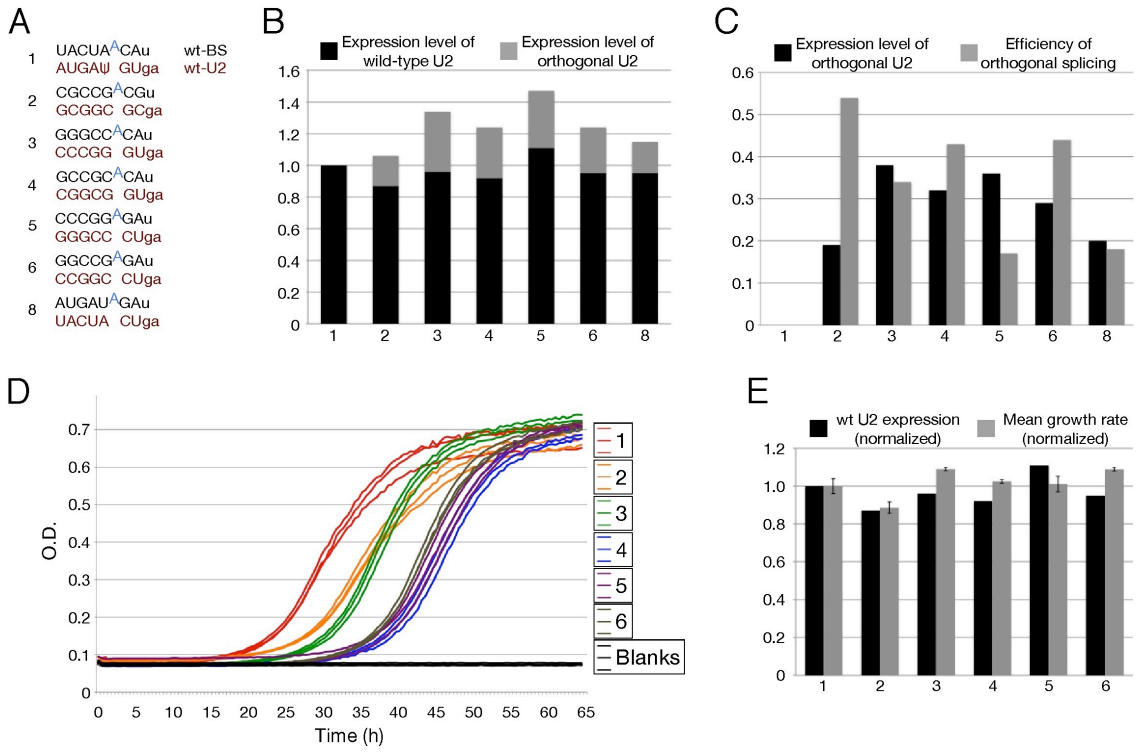


Figure 4-5. Further characterisation of the orthogonal BS-U2 system.

(A) Constructs used in this figure: sequences and numbering as in Fig. 4-4. (B) Total U2 snRNA expression (wild-type and orthogonal U2 snRNAs) in each strain: quantitation refers to Fig. 4-4D. U2 expression was corrected for unequal loading using the internal U6 control, and normalised to the expression of wild-type U2 in the strain carrying no mutant U2 allele. (C) The efficiency of orthogonal splicing does not correlate with the expression of the orthogonal U2 snRNA: quantitation refers to Fig. 4-4C&D. (D) Growth curves of strains carrying the indicated second-copy U2 snRNA. (E) The variable growth rate of strains carrying orthogonal U2 snRNAs is not related to reduced expression of wild-type U2. Relative growth rates were calculated as the gradient of the linear region from O.D. 0.2-0.4 in [D] after log transformation, represent the mean value from the three curves, and were normalised relative to the strain carrying two wild-type copies of U2. Error bars represent ±1 standard deviation. U2 expression levels refer to Fig. 4-4D.

Strains carrying certain mutant second copy U2 alleles exhibited slow growth (Fig. 4-5D&E). We do not know the basis of this growth defect, as it does not correlate with reduced or excessive expression of either the wt or the mutant copy of U2 snRNA (Fig. 4-5B&E and data not shown). We also do not believe that deleterious generation of new, orthogonal BS-containing introns contributes substantially to the observed growth defect: although we identified candidate pre-mRNAs in which new introns may be generated by our orthogonal U2s, we were unable to detect the products of such new splicing events by RT-PCR for either of two strongly predicted introns tested (data not shown).

ESTABLISHMENT OF NUCLEOPHILE REQUIREMENTS FOR THE FIRST STEP OF SPLICING

The orthogonal BS-U2 system minimises both inter-strain growth differences and potential interaction between the mutant BS reporter and wt U2 snRNA, allowing systematic investigation of the effect of branch nucleotide identity, context and positioning in a constant U2 background. Our previous analysis had suggested that bulges from at least three positions within the BS-U2 duplex were competent for first step catalysis. Using reporters in which all BS flanking sequence was mutated to G/C with purine/pyrimidine identity either maintained or transverted (UACUA**A**CA to CGCCG**A**CG or GCGGC**A**GC, respectively – the zero position branch nucleotide is highlighted), we placed a bulged adenosine at all positions from -4 to +1 (Fig. 4-6A&C) and assayed splicing by primer extension and copper growth (Fig. 4-6B&D and data not shown). We observed relatively efficient branching from the 0 or -1 positions in both constructs (Fig. 4-6B&D), and

additionally from the +1 position in the transverted duplex, but not from bulged nucleotides further upstream than the -1 position (Fig. 4-6D and data not shown). This confirms the positional flexibility of branching in the context of a constant U2 snRNA background – bulges at the -1 and +1 positions are expected to be separated by around 8 Å and 90˚ (143).

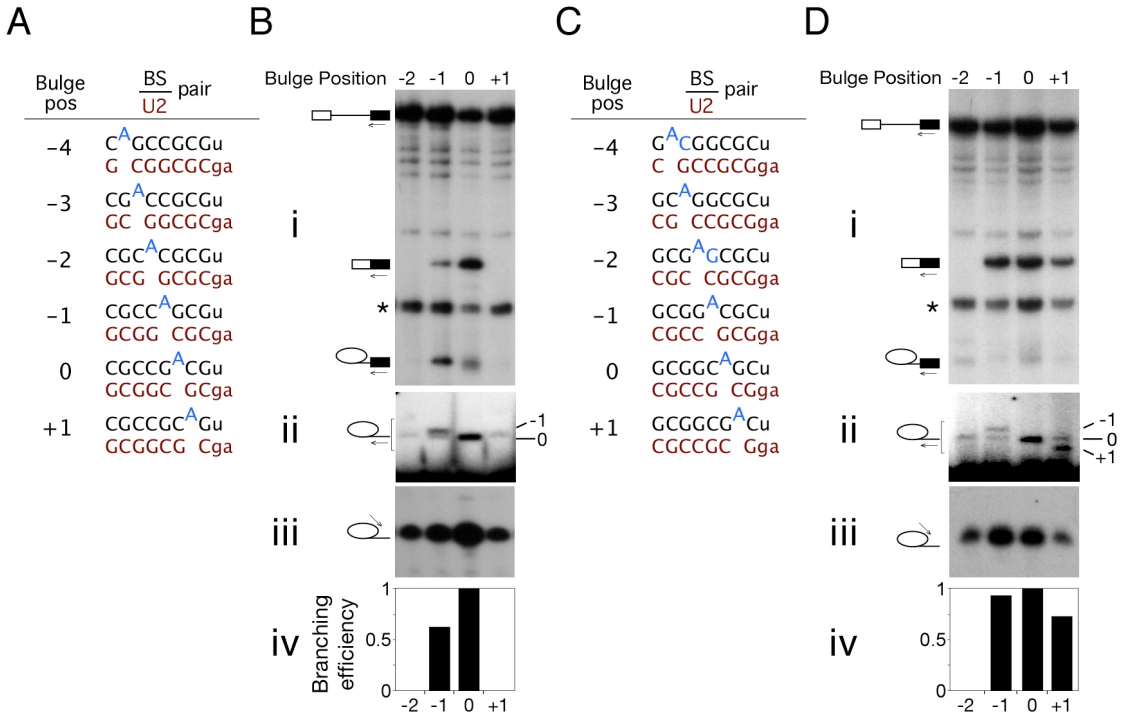


Figure 4-6. Confirming the positional flexibility of the branch nucleophile.

(A) Schematic of reporter and U2 constructs used in [B]. (B) Primer extension on total RNA from the indicated strains with 3' exon (panel i), BS-proximal intronic (panel ii), and 5'SS-proximal intronic (panel iii) primers indicates that two bulge positions in this duplex are catalytically competent, that the expected branch nucleotide is used in each case, and that the location of the branch nucleotide in the BS-U2 duplex does not affect the position of 5'SS cleavage. Panel iv, quantitation of the 3' exon primer extension in (i). (C) Schematic of reporter and U2 constructs used in [D]. (D) Primer extension on total RNA from the indicated strains primers as in [B] indicates that three bulge positions in this duplex are catalytically competent, and, as in [B], that the branch nucleophile can be moved without changing the position of 5'SS cleavage

We believe the disparity between the two constructs to be related to base pair identity flanking the bulge. Base pairing at both flanking positions strongly favours splicing (Fig. 4-7A), and nucleotide identity within the pair impacts splicing efficiency (Fig. 4-7B): consistent with flanking base pair identity underlying the observed difference in the use of the +1 position in the pseudo wild-type CG (CGCCGC**A**G) and transverted GC (GCGGCG**A**C) duplexes (both sequences shown with a highlighted +1 adenosine), splicing is more efficient for branch nucleotides preceded by a purine and followed by a pyrimidine than the converse (Fig. 4-7B). Thus, one would expect the +1 position to be relatively favoured in the context of the transverted GC duplex, and relatively disfavoured in the pseudo wild-type CG context. Moving the branch site A in the context of some orthogonal BS-U2 duplexes imposes a strong block on the second step of splicing (Fig. 4-7B and data not shown): we do not know the mechanistic basis for this defect, but this is discussed more fully in chapter 5, in the context of branch site interactions during the second step of splicing.

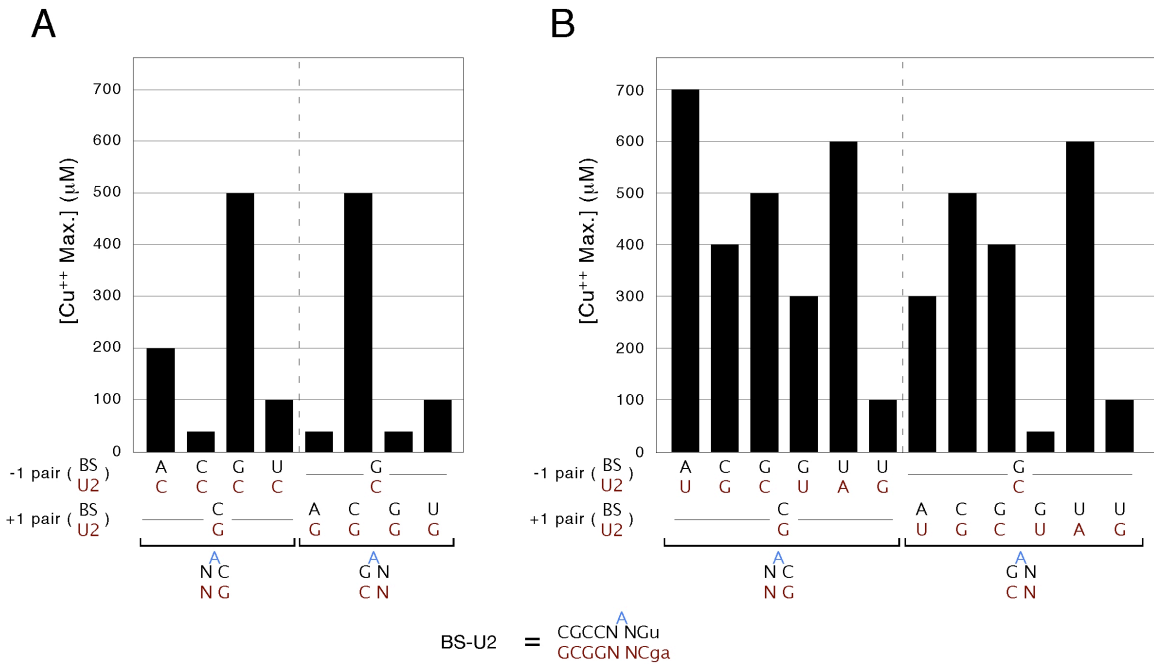


Figure 4-7. Base pairs flanking the branch are required for efficient splicing: base pair identity impacts splicing efficiency.

(A, B) Graphs summarising the copper growth of strains carrying the indicated reporter-U2 pairs. The sequence of the BS used in this analysis is CGCCNANG.

Splicing, both in yeast (134) and in higher eukaryotes (144), is most efficient with an adenosine branch nucleophile: guanosine is mildly suboptimal and either pyrimidine nucleotide even more so. The effects of BS mutation, however, have been studied in a small number of genes, and in an otherwise wt BS-U2 context: alternative pairings for such duplexes can be envisaged in which an 'incorrect' nucleotide is bulged – i.e. the observed splicing defect may arise due to bulging nucleotides suboptimal in terms of identity or of position. We produced duplexes with only one pairing register, and a BS nucleotide flanked by a pair of either guanosine or cytosine nucleotides (Fig. 4-8A), to facilitate the predictable bulging of A/C/U or A/G/U, respectively. The splicing profile of these reporters mirrored

that of BS mutants in a wt context, with guanosine supporting a fairly efficient first step and pyrimidines showing little to no branching (Fig. 4-8B). These data support an inherent preference for adenosine at the branch position, as previously suggested by work showing recognition of multiple groups on the adenosine nucleotide during spliceosome assembly and catalysis ADDIN BEC(145, 145), the branch site specificity of group II introns (146), and an apparently intrinsic preference for adenosine as the nucleophile in 2'-5' branch formation in some other contexts (147). A fuller treatment is given to the issue of the role of adenosine in RNA branching reactions in Chapter 5.

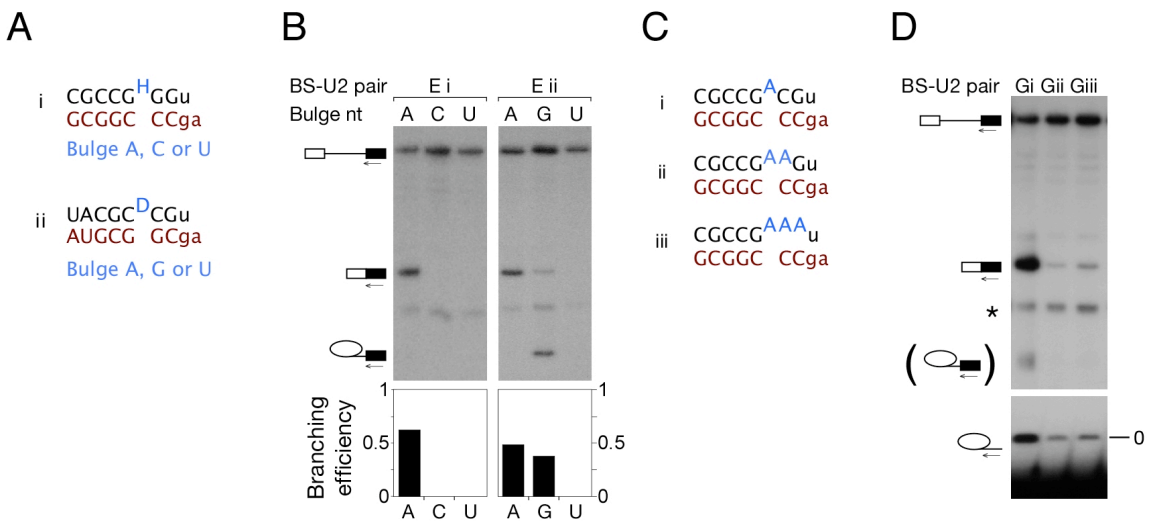


Figure 4-8. Confirming the preferred identity of the branch nucleophile.

(A) Schematic of reporter and U2 constructs used in [F]. (B) Primer extension using a 3' exon primer indicates an intrinsic nucleotide preference at the branch site, with A>G>>C~U. (C) Schematic of reporter and U2 constructs used in [H]. (D) Primer extension indicates that branching occurs from the upstream-most nucleotide in a run of consecutive single-stranded adenosines.

A remaining question was the behaviour of branch sites with multiple consecutive bulged nucleotides. We constructed BS-U2 pairs containing one, two or three consecutive unpaired adenosines (Fig. 4-8C). The presence of additional unpaired adenosines impaired splicing of the reporter (Fig. 4-8D, upper panel), presumably due to the lack of base pairing downstream of the branch site adenosine (cf. Fig. 4-7). Splicing did, however, occur to detectable levels and high-resolution primer extension indicated that the upstream-most adenosine was invariably used as the branch nucleophile (Fig. 4-8D). These results recall data regarding branch site selection in the minor spliceosome (148), and suggest a model whereby the most upstream unpaired adenosine in the BS-U2 or BS-U12 duplex is specified as the branch nucleophile.

GROSSLY SUBSTITUTED BS-U2 DUPLEXES PREDOMINANTLY IMPAIR SPLICEOSOME ASSEMBLY: INAPPROPRIATELY BULGED NUCLEOTIDES LIMIT CATALYSIS

Although virtually all grossly substituted BS-U2 duplexes supported splicing, first-step efficiency was invariably diminished relative to that observed for a wt BS-U2 pair (Fig. 4-4C). Spliceosome assembly precedes branching catalysis, and the observed accumulation of pre-mRNA could be due to a defect in either or both of these steps. The BS-U2 duplex, as a highly conserved element containing a first-step substrate, likely forms several tertiary interactions in the spliceosome core; although our mutant pairs are stably base-paired, we expected that they might be deficient in some such interactions.

To clarify the nature of the defect shown by our constructs, we investigated their splicing in the context of mutant alleles of the spliceosomal ATPases Prp5 and Prp16. Prp5 has been proposed to monitor the stability of BS-U2 interaction during spliceosome assembly via kinetic competition between its ATPase activity and BS-U2 pairing – failure to establish stable interaction prior to ATP hydrolysis by Prp5 leads to pre-mRNA discard (36). Mutant *prp5* alleles, by relaxing the stability requirement for the nascent complex, thus facilitate spliceosome assembly on introns with suboptimal BS-U2 pairing and presumably also those in which a stable duplex is imperfectly packed. Prp16 enhances splicing fidelity by acting in competition with splicing catalysis – failure to complete first step catalysis prior to ATP hydrolysis by Prp16 leads to pre-mRNA discard (51). A suboptimally packed BS-U2 duplex could inhibit first step catalysis by globally destabilising the first step conformation of the spliceosome.

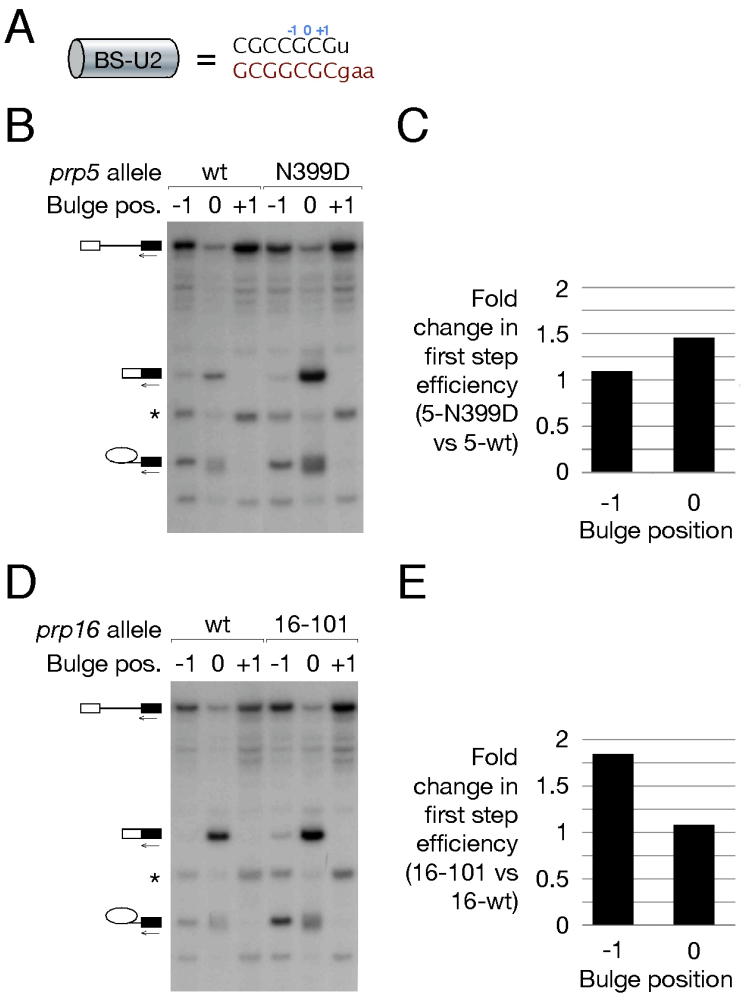


Figure 4-9. Grossly substituted BS-U2 duplexes predominantly impair spliceosome assembly: inappropriately bulged nucleotides predominantly impair splicing catalysis.

(A) Schematic of the BS-U2 duplex used in this figure, with the positions of bulged adenosines indicated. (B) Primer extension using a 3' exon primer on total RNA from strains carrying reporters with adenosines bulged from the indicated positions in a grossly substituted BS-U2 duplex and either wt or mutant Prp5. (C) Graph indicating the fold change in branching efficiency [calculated as (M+LI)/(P+M+LI)] at the -1 and 0 bulge positions as a result of Prp5 mutation. (D, E) As [B, C], but comparing wt and mutant Prp16.

For each ATPase we assayed the splicing efficiency, in a wt and a mutant strain, of substrates with a CGCCG**A**C branch site with an adenosine at either the -1, 0 (highlighted here), or +1 positions (Fig. 4-9A). The *prp5 N399D* allele (36) stimulated splicing from the 0 position ~1.5-fold, yet had little effect on the splicing of a reporter with a -1 position branch adenosine (Fig. 4-9B&C). By contrast, the *prp16-101* allele (51) stimulated the splicing of the -1 position reporter 1.8-fold but had little effect on the 0 position reporter (Fig. 4-9D&E). Taken together, these observations suggest distinct limiting steps in the splicing pathway for the -1 and 0 position reporters. In this context, spliceosome assembly is impaired by duplex substitution: the mutant *prp5* allele suppresses this assembly defect. In a substrate with an optimally positioned 0 position bulge, enhanced assembly leads to enhanced splicing. With the nucleophile bulged from the -1 position, however, increased assembly does little to stimulate splicing because the overall efficiency of the reaction is limited by first step catalysis: in this case, the slowed exit from the first step conformation afforded by *prp16* mutation strongly stimulates branching. These results not only define the limiting steps in the splicing pathway for orthogonal BS-U2 pairs, but also indicate that the position of the bulged nucleophile within the BS-U2 duplex is important for first step catalysis by the spliceosome.

THE PREFERRED NUCLEOPHILE POSITION DEPENDS ON DISTANCE FROM THE U2 PORTION OF U2/U6 HELIX IA, AND DOES NOT DETERMINE 5'SS SELECTION

The analysis described above demonstrates the flexibility of branch site activation within a constantly positioned BS-U2 duplex, without addressing how movement of the duplex itself may affect branch site selection. U2 snRNA is involved in several intra- and intermolecular structures within the catalytic spliceosome (2). We therefore sought, again using a constant BS sequence from which an adenosine could be bulged at multiple positions, to investigate the impact on preferred bulge position of changing the distance between BS-U2 and these spliceosomal structures. Because our second copy U2s do not interfere with the splicing of endogenous introns, we were relatively unconstrained in our choice of mutations, and made a large number of insertion and deletion mutants throughout the 5' end of U2, including into either the stem or loop of the 5' stem loop, up- and downstream of helix I, and upstream of helix III (Fig. 4-10). Several of these mutant U2s failed to show stable expression, presumably due to an inability to assemble into snRNPs (data not shown); all expressed U2s with insertions/deletions anywhere upstream of helix I showed at worst a mild splicing defect with no impact on branch nucleophile selection (data not shown). Of most interest were insertions predicted to alter the distance between helices Ia & III and the BS-U2 duplex, which lies between the U2 components of each (61, 149), as this network of interactions may help to juxtapose key catalytic components and splicing substrates.

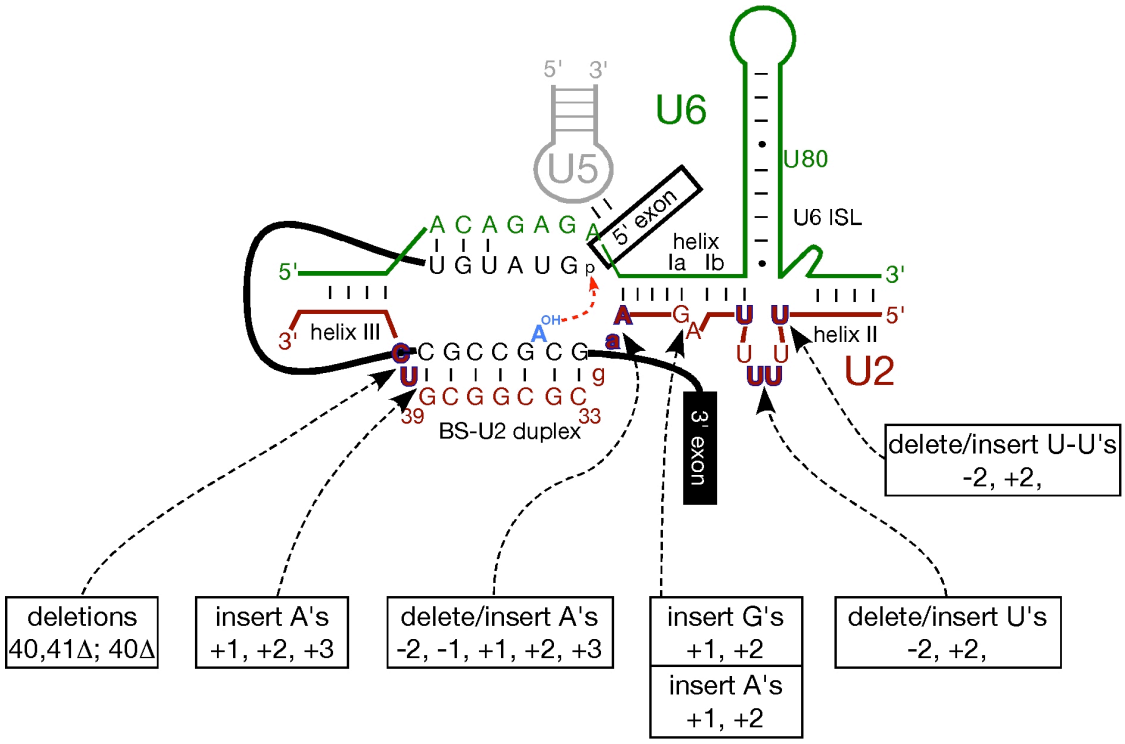


Figure 4-10. Insertions and deletions in the conserved 5' end of U2 tested for effects on branch nucleophile selection.

Arrows indicate the locations of insertions/deletions. Deleted nucleotides are highlighted with blue trim.

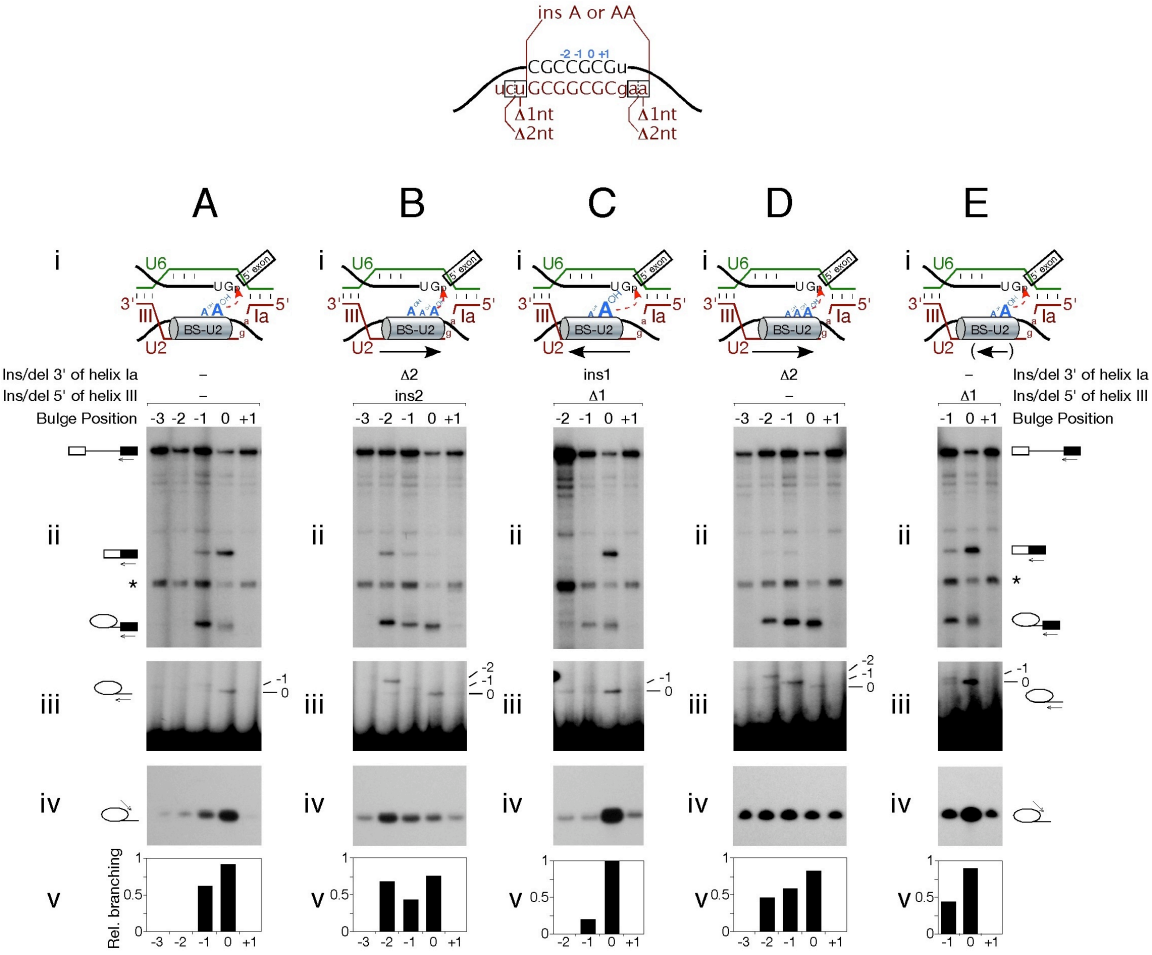
An all-CG branch site with the (0 position branch) sequence CGCCG**A**C shows branching from the -1 and 0 positions in the presence of its cognate U2 snRNA (Fig. 4-11A). Insertion of two nucleotides into U2 upstream of helix III, with accompanying deletion of two nucleotides downstream of helix Ia, predicted to shift the BS-U2 duplex towards helix Ia and away from -III, altered this splicing profile such that the -2 position was now highly active for branching and the 0 and -1 positions less favoured than in the original context (Fig. 4-11B). An analogous 1-nt insertion/deletion pair also shifted the observed branching preference towards upstream bulges (data not shown). Conversely, deleting one

nucleotide upstream of helix III and inserting one downstream of helix Ia, predicted to shift BS-U2 away from helix Ia and towards -III, disfavoured the use of the -1 position while maintaining robust use of the 0 position BS (Fig. 4-11C). Collectively, these data suggest that branch position relative to either or both of these helices impacts the bulge positions within the BS-U2 duplex that are competent for first step catalysis.

Figure 4-11. Distance from the U2 component of U2/U6 helix Ia partially determines the preferred position of the BS nucleophile within the BS-U2 duplex.

For each panel: (i) Simplified schematic of the spliceosome catalytic centre with helices Ia and III indicated; BS adenosines in blue are sized proportionally to their ability to participate in splicing catalysis based on quantitation (v) of the 3' exon primer extension in (ii). Branch site and 5'SS use are assayed by primer extension in (iii) and (iv), respectively: the position of the branch nucleotide both within the BS-U2 duplex and relative to the U2 component of helix Ia does not impact the position of 5'SS cleavage. All panels use an all-CG BS-U2 duplex, indicated below the figure, in which purine-pyrimidine identity is maintained at each position relative to wild-type. (A) Profile of -3 to +1 bulged reporters with CG-substituted but otherwise wt U2. (B) Deletion of two nucleotides in U2 downstream of helix Ia, and insertion of two nucleotides upstream of helix III, predicted to move the BS-U2 duplex towards helix Ia and away from helix III, activates upstream bulge positions for catalysis. (C) Insertion of one nucleotide in U2 downstream of helix Ia, and deletion of one nucleotide upstream of helix III, predicted to move the BS-U2 duplex away from helix Ia and away towards helix III, represses the use of upstream bulge positions for catalysis. (D) Deletion of two nucleotides in U2 downstream of helix Ia recapitulates the effect seen in [B], activating upstream bulge positions for catalysis. (E) Deletion of one nucleotide in U2 upstream of helix III does not strongly impact the catalytic potential of different bulge positions.

Figure 4-11


ins A or AA
-2 -1 0 +1
CGCCGCGu
ucuGCGGCGCGaaa
Δ1nt
Δ2nt
Δ1nt
Δ2nt
A
B
C
D
E
i
ii
iii
iv
v
U6
5' exon
UGR
3'
III
Ia
5'
BS-U2
U2
Ins/del 3' of helix Ia
Ins/del 5' of helix III
Bulge Position
-3 -2 -1 0 +1
–
–
Δ2
ins2
ins1
Δ1
Δ2
–
–
Δ1
-1 0 +1
Rel. branching
1
0.5
0

To investigate the relative importance of BS distance from helices Ia and III, we produced constructs with single-site deletions either downstream of helix Ia or upstream of helix III. Deletion of two nucleotides downstream of helix Ia again activated branching from the -2 position (Fig. 4-11D), recapitulating the result of such a deletion in the context of an accompanying helix III-proximal insertion (cf. Fig. 4-11B&D). By contrast, a 1-nt deletion upstream of helix III produced a splicing profile very similar to that of the original orthogonal U2 (cf. Fig. 4-11E&A). These data show that the distance from the branch nucleophile to the U2 component of helix Ia is an important determinant of nucleophile specification for catalysis. We note that this distance requirement may be relative either to helix Ia or to a rigid higher-order structure containing this helix. Despite the activation of upstream bulge positions for catalysis by helix Ia-proximal deletion in U2, branching remains most efficient from the zero position (Fig. 4-11B&E). We hypothesise that this position within the duplex remains the optimal one, and that the distance between the bulge and helix Ia is an (at least partially) independent factor in nucleophile determination.

To confirm the above analysis, we recapitulated the results in the context of the transverted GC duplex. Again, reducing the distance between the BS-U2 duplex and helix Ia led to relative activation of upstream bulge positions for first step splicing catalysis (Fig. 4-12A&B). We note that the distribution of upstream branch use differs between the two constructs; as previously discussed, we believe that flanking base pair identity underlies this observed difference.

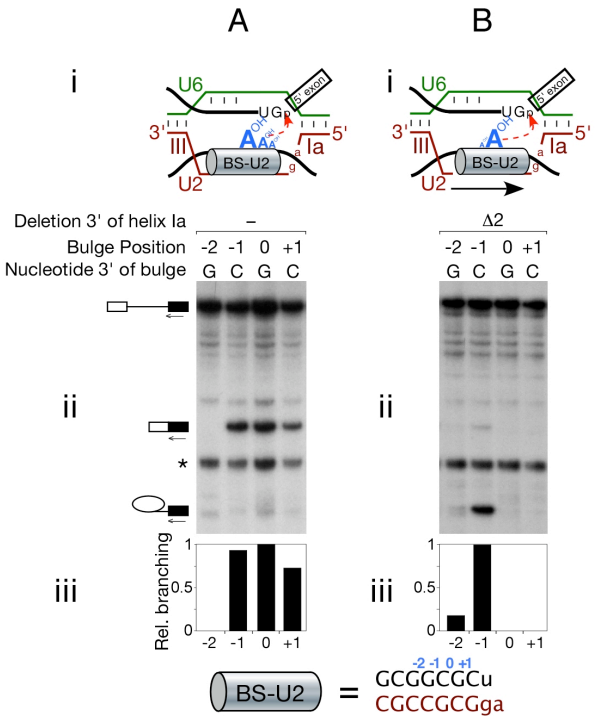


Figure 4-12. Confirmation of the effect of distance between nucleophile and helix Ia in the context of a transverted GC duplex.

(A, B) (i) Simplified schematic as in Fig. 4-11, with BS adenosines sized according to quantitation (iii) of the 3' exon primer extension in (ii). An all-CG BS-U2 duplex, indicated below panel (iii), in which purine-pyrimidine identity is reversed at each position, is used. Relative to the U2 used in [A] (also shown in Fig. 4-6D), the U2 in [B] carries a deletion downstream of helix Ia, predicted to activate bulge positions upstream of 0.

Although either or both of the branch nucleophile and the entire duplex from which it was bulged were moved in the above analysis (Fig. 4-6, 4-11&4-12), and high-resolution primer extension confirmed the potential to use various different branch sites, the 5'SS remained unchanged and no activation of cryptic sites was observed in any of this work (Fig. 4-6B&D, lower panel, and panel iv of Fig. 4-11). If 5'SS positioning were dependent on BS positioning, or if the structures that contained the 5'SS and BS, rather than the reactive groups themselves,

were positioned relative to one another, movement of the branch site nucleophile and/or BS-U2 duplex would be expected to result in an analogous movement of the 5'SS. Therefore, the observation that moving the BS or the duplex containing it fails to impact 5'SS selection suggests that the position of the BS nucleophile does not determine the position of the 5'SS electrophile for the first step of pre-mRNA splicing.

DISCUSSION

In vivo systems incontrovertibly provide the most valid environment for the study of biological processes. Unfortunately, however, *in vivo* mutational analyses are restricted by a requirement for cell viability. Orthogonal systems, in which a nonessential second copy enzyme exclusively targets its cognate substrate (which is in turn not a substrate for the wild-type cellular enzyme) offer the possibility of combining the power of the *in vivo* setting with the mutational freedom of *in vitro* analysis. Essential, highly conserved base pairing interactions, when not complicated by competing pairings, are ideally suited to the generation of such systems: an approach based on the replacement of the Shine-Dalgarno interaction, has been used to produce orthogonal ribosomes in *E. coli* (150). Here we describe a system for the study of pre-mRNA splicing, based on the analogously conserved interaction between the intron branch site and its cognate sequence in U2 snRNA. We use a variety of BS-U2 pairs to investigate the flexibility of nucleophile selection for the first step of splicing in yeast. Our results establish conclusively that bulging of the nucleophile is required for first step catalysis, confirm the importance of branch nucleophile

identity and location, and suggest a model whereby the first unpaired nucleotide in the BS-U2 duplex is specified as the nucleophile. Additionally, our data indicate a high degree of flexibility in the active site of the yeast spliceosome, and strongly argue that 5'SS and branch site specification within the catalytic centre are independent. Finally, they extend the evidence for similarity between the spliceosome and group II self-splicing introns to include a common mechanism of first step nucleophile specification and positioning.

FLEXIBILITY OF FIRST STEP NUCLEOPHILE SPECIFICATION

The yeast spliceosome has stringent substrate requirements for assembly and catalysis; we show that nucleophile bulging is required for splicing catalysis but that, outside of this constraint, there is flexibility with respect to branch nucleophile selection and positioning.

No single feature of the branch site aside from nucleophile identity appears to exert a strongly dominant effect on the efficiency of splicing; instead, positive and negative influences accumulate to produce intermediate phenotypes. This is clear from the distribution of branch site use from all-CG BS-U2 duplexes in which purine-pyrimidine identity is maintained (BS 0 sequence CGCCG**A**C – pseudo wild-type) or reversed (BS 0 sequence GCGGC**A**G - transverted) relative to wild-type (UACUA**A**C) at each position. In the context of an otherwise wild-type U2 (Fig. 4-12A), the efficiencies of branching from the -1, 0 and +1 positions in the reporter are reasonably similar. Although the 0 bulge is positionally

optimal within the duplex and relative to helix Ia, identity of flanking base pairs is not (Fig. 4-7B). Bulges at ±1 are suboptimal for branching, and -1 appears to be less disfavoured than +1 (see for example Fig. 4-6B), but in the transverted duplex this effect is mitigated by the presence of flanking base pairs that support efficient splicing. The net result of these positive and negative factors, in this case, is that -1>0>+1 in terms of branching efficiency. When two nucleotides proximal to helix Ia are deleted in this substituted U2, upstream bulge positions become favoured: of the -2, -1 and 0 bulges, -1 relatively favourable flanking sequence and, as it is also well positioned, becomes the most efficiently used branch position.

Similar considerations to those applied above can explain the preference for branching from the 0 position in the pseudo-wt CG-substituted BS-U2 duplex (Fig. 4-11A&B), and indeed in the truly wild-type UACUA**A**C context. The canonical BS in each context is an optimally positioned adenosine with favourable sequence context, and each of the flanking nucleotides is deficient in at least two of these respects; this allows what appears to be a relatively flexible active site capable of acting on suboptimal substrates to give rise to strong specificity. In systems with less stringent branch site consensus sequences, even greater flexibility or different relative contributions of the various factors that determine branching efficiency could allow more promiscuous branching.

Because the spliceosome is robust, several deleterious changes may be required to generate a visible effect on splicing efficiency *in vivo*, particularly if these changes predominantly affect steps that are not close to being rate-limiting. In the generation of an orthogonal BS-U2 system tailored for the investigation of nucleophile specification we show that, although BS-U2 pairing is required for splicing, the sequence of the BS and its cognate region of U2 are much less important. Complete replacement of the BS-U2 duplex results in both spliceosome assembly and first step catalysis being partially limiting; this allows a systematic analysis of the determinants of branch nucleophile use. We identify three major determinants of branching efficiency: the identity of the branch nucleotide (and to a lesser extent the identity of flanking base pairs), the position of the bulged branch nucleotide within the BS-U2 duplex and, independently, its position relative to the U2 component of U2/U6 helix Ia.

In a case in which multiple unpaired adenosines are present with little potential for downstream pairing to U2, branching occurs at the first unpaired A (Fig. 4-8D). This suggests that, although appropriate positioning of a bulge relative to the catalytic core provides the strongest determinant for efficient branching, a single-stranded nucleotide at the edge of a helix can also act as a branch nucleophile. The distribution of mapped branch positions in the U12 spliceosome (148) obeys similar rules, with the strong features being a well-defined BS-U12 duplex containing an unpaired A nucleophile positioned similarly to that in the BS-U2 duplex.

The branch site adenosine of a pre-mRNA can be bulged from almost any Watson-Crick duplex, as indicated by the observation that virtually all BS-U2 pairs we tested were functional. The extreme conservation of the BS-binding region of U2 (the GUA sequence opposite the bulged branch nucleophile is conserved between U2 and U12) is therefore not explained by a dependence of spliceosome function on a specific sequence. The splicing defect of highly substituted BS-U2 pairs with an optimally positioned bulge appears to be predominantly due to impaired spliceosome assembly, as suggested by partial suppression by *prp5* alleles, and to a lesser impact on splicing catalysis (Fig. 4-9). In contrast, BS-U2 pairs with a sub-optimally positioned bulge are catalytically defective, indicated by strong improvement by ATPase-defective *prp16* alleles. Thus, assembly is primarily affected by alteration of the BS-U2 duplex, whereas catalysis is sensitive to altered positioning of the nucleophile.

INDEPENDENT POSITIONING OF 5'SS AND BRANCH SITE

If higher-order structures containing the branch nucleophile and 5'SS electrophile for the first step of splicing were fixed relative to one another, moving the branch nucleophile should alter the position of 5'SS cleavage. We observe, however, that moving the branch nucleophile relative to the BS-U2 duplex or to helix Ia does not alter 5'SS selection, indicating that this is not governed by BS selection. This observation is in agreement with the ability of higher eukaryotic spliceosomes to cleave the same 5'SS at the same position using more than one

possible branch nucleotide (137), and highlights the commonality of substrate selection in the *S. cerevisiae* and metazoan spliceosomes.

The position of branching within the BS-U2 duplex can impact the second step of splicing. The precise mechanistic basis of the defect sometimes observed with non-zero branch positions and/or helix Ia-proximal insertions and deletions (e.g. Fig. 4-11) is unclear, especially as branching from non-zero positions (e.g. Fig. 4-6D) and insertions into U2 (e.g. Fig. 4-1D) do not in themselves invariably lead to second step defects. The role of U2 in the second step of splicing is given a fuller treatment in chapter 5.

EVOLUTIONARY CONSIDERATIONS

The spliceosome and group II self-splicing introns share a number of features that suggest a common evolutionary ancestor. Support for significant overlap of catalytic mechanism between the two systems has recently been provided by the crystal structure of a spliced group II intron: in this structure, motifs that are conserved between group II introns and U6 snRNA (in which they are close to helix Ia) coordinate magnesium ions (118). Our data extend this catalytic similarity by suggesting a conserved mechanism of branch nucleophile selection and positioning. A branch site adenosine incapable of bulging cannot efficiently participate in first step branching catalysis in either system – Fig 4-1 and (140) – and the identity of flanking base pairs influences branching efficiency – Fig. 4-7; cf. Fig. 4-11&4-12 and (151). In addition, the use of cryptic branch sites in group

II introns can be observed upon alteration of the distance between domains 5 and 6, the equivalents of the U6 ISL and BS-U2 duplex, respectively (151). In our BS-U2 system, the use of upstream bulges is favoured by deletions predicted to decrease the distance between the U6 ISL and the BS-U2 duplex, and disfavoured by insertions predicted to increase it (Fig. 4-11). Thus, our data support a spatially similar organisation of conserved elements in group II introns and the spliceosome.

CHAPTER 5. DISCUSSION

PREAMBLE: THE STRENGTH OF A FULLY SPLICESOMAL EXPERIMENTAL APPROACH TO SPLICING RESEARCH

One of the original aims of my thesis work was to establish an RNA-only system that could be used as a model to bypass spliceosome assembly and investigate the catalytic phase of splicing, in order to generate hypotheses that could be tested in complete spliceosomes. However, as discussed in chapter 2, a great deal of assumption and subjective judgement is inherent in this approach. As will be discussed further below, the current extent of our knowledge of the active site of the spliceosome is such that even its composition remains an open question. Given such a combination of uncertainty and subjectivity, the use of rationally designed minimal experimental systems appears as likely to generate misleading results and hypotheses as informative ones; my subsequent work therefore focused on the use, almost exclusively in an *in vivo* context, of the established and powerful *S. cerevisiae* genetic system as a tool to investigate splicing catalysis. Although *in vivo* and extract-based experimental work are not without their disadvantages, the relevance of results thus obtained is fairly easy to establish, even in the context of large and relatively poorly understood complexes such as the spliceosome.

In one sense, the work described in chapters 3&4 follows well-worn conceptual paths: this work is, however, representative of the suite of approaches that will be required to further our mechanistic understanding of the splicing reaction. In

chapter 3 the identification and characterisation of an aberrant product, in this case a hybrid RNA resulting from *trans*-splicing of U2 snRNA to a reporter gene, allowed us to draw mechanistic conclusions about the normal course of the reaction. In chapter 4, our detailed investigation of the role of branch site-U2 pairing in branching catalysis by the spliceosome was made possible by compensatory suppressor mutations. Rather than being an end in themselves, however, the suppressor pairs allowed us to establish a second-copy spliceosome, dissociating splicing efficiency from cellular viability and thus allowing us to combine a physiologically relevant *in vivo* environment with the mutational freedom of *in vitro* studies. As will be discussed further below, we hope that this system can be extended and built upon in the future.

Various direct conclusions resulting from the work described in this thesis have been discussed in detail in individual chapters. In this chapter I will focus on more global implications of the work as a whole, as well as on open questions to which extensions of this work may provide answers.

THE CHEMISTRY OF RNA BRANCHING

In Chapter 4, we verified that adenosine is favoured as the branch site nucleotide, and that this preference is at least in part manifested during the catalytic phase of splicing. The extraordinary complexity of the spliceosome compared to the simple reactions it catalyses suggests that specific contacts

likely enforce such sequence preference. However, both the broad catalytic repertoire available to RNA implied by the work in Chapter 2 and considerations of spliceosome evolution from simpler progenitors merit consideration of the role of adenosine as a branch nucleophile in splicing and in unrelated branching reactions.

Although, as discussed in Chapter 4, group II introns, the spliceosome and some *in vitro* evolved ribo- and deoxyribozymes display selectivity for adenosine as a branch nucleophile, such a preference is not universal among RNA branching reactions. A natural ribozyme that forms a 2'-5' bond as part of a 3 nt lariat cap structure uses uridine as a branch nucleotide (152), and artificial DNAzymes have been generated that efficiently branch from pyrimidine nucleotides (153). As noted by Silverman and colleagues (147), a commonality between enzymes that show a strong preference for a BS-A is that they are predicted to bulge the branch nucleotide for catalysis. Consistent with this observation is the lack of involvement of an obvious bulge structure in the reaction described in Chapter 2, in which uridine is the preferred branch nucleotide. The adherence of this ribozyme reaction to the observed pattern of behaviour despite the formation of a 2'-3' rather than a 2'-5' bond suggests that the lack of a preference for a BS adenosine is a more general feature of transesterifications that do not involve bulged nucleophiles.

Adenosine is over-represented among bulged or unpaired nucleotides in solved RNA and RNP structures, with over 60% of adenosines unpaired in 16S rRNA structures compared to ~30% for each other nucleotide (154): while tertiary structure formation, for example the ability of unpaired adenosines to read Watson-Crick base-pair geometry via A-minor interactions (155) presumably partially accounts for this observation, it is possible that the intrinsic stability of adenosine-containing extrahelical bulges is also a factor. A systematic analysis of the dependence of extrahelical bulge stability on nucleotide identity has, to our knowledge, not been carried out; such an analysis would have to be based on complex NMR and/or molecular dynamics due to the ability of single unpaired nucleotides to adopt several conformations: bulged (156), intercalated between the adjacent base pairs (157), engaged in a base triple with a flanking base pair (158) and possibly others. Unpredictable effects of sequence identity outside the flanking base pairs on the relative stabilities of these structures (159) may further complicate this issue.

Independently of the stability of RNA bulge structures, adenosine is inherently likely to be the most reactive nucleotide towards branching reactions because the pKa of its 2' hydroxyl group is lower than that of the other three (160), resulting in more facile deprotonation for nucleophilic attack. Enzymatic catalysis may, however, render considerations of intrinsic nucleotide reactivity moot. Many studies have concluded that 2' hydroxyl groups are up to a hundredfold more reactive towards transesterification or condensation reactions than 3' hydroxyls,

both in template-free and template-directed non-enzymatic systems (161, 162), yet enzymatic systems with free nucleotide (for example group I introns) or oligonucleotide-terminal (for example RNA ligases and polymerases) RNA nucleophiles are capable of exquisite 3' hydroxyl regiospecificity via substrate positioning and transition state stabilisation.

RNA branching reactions from non-terminal nucleotides are limited in their choice of nucleophile to the 2' hydroxyl group: such reactions can, however, theoretically generate either 2'-5' or 2'-3' linked products. Although 2'-5' branches should have lower energy transition states due to the greater acidity of the *cis*-diol 3' hydroxyl than the primary alcohol 5' hydroxyl (105), the reasoning applied above argues that this consideration is likely to be virtually irrelevant in an enzymatic context. With this in mind, our knowledge of several natural systems that generate 2'-5' branches and none that generate the 2'-3'-linked form is surprising and, as noted in Chapter 2, could reflect the relative difficulty of identifying and mapping 2'-3' linkages rather than their absence from the range of naturally generated cellular RNAs.

Mapping 2'-5' branches, especially in a high-throughput manner, is not trivial for several reasons. Reverse transcriptase can read through such branches, presumably with variable efficiency (163), and in addition both the 2'-5' lariat debranching enzyme (164), and antibodies that recognise this branch structure (165) show strong sequence specificity: the branch antibody, for example, is

extremely selective for A2'-5'G branches. Finally, nucleases that cleave more than 1 nt away from a branch (such a nuclease would allow the position of such branches, rather than simply their existence, to be inferred) have not been identified or generated. This mapping challenge is even greater for 2'-3' branches: antibodies, debranching enzymes and nucleases that leave sequence tags are all lacking for this structure. In addition, reverse transcriptase, in theory capable of mapping the branch site only on the 2' donor strand, can read through such branches (105) although nothing is known about the efficiency of this readthrough. To date only directed, labour-intensive and imperfect methodologies have been used to identify 2'-3' branches. Direct, positionally specific labelling followed by nuclease digestion, as exemplified in Chapter 2, can reliably map branches but is limited to *in vitro* reactions and is only practical if the locations of the reactive sites have already been narrowed down. The existence of 2'-3' branches in cellular RNAs could be confirmed via sequential overdigestions with lariat debranching enzyme and P1 nuclease, but information regarding the sequences of the RNAs involved in such linkages would be lost. Given that no high-throughput methodology for 2'-5' branch mapping has been forthcoming despite the existence of pressing and long-standing mechanistic questions regarding mammalian branch site use in pre-mRNA splicing, we are not optimistic about the imminent development of such techniques for 2'-3' branches.

ADDING TEMPORAL RESOLUTION TO OUR KNOWLEDGE OF SPLICEOSOMAL INTERACTIONS

Much work in the mechanistic splicing field is justifiably focused on the generation of spliceosome preparations suitable for X-ray crystallography. The high-resolution structures that will hopefully result from this work should resolve many questions, and provide invaluable information about the mechanistic details of the splicing reaction. The preparation of large quantities of sufficiently pure, conformationally homogeneous spliceosomes, however, remains a serious challenge. Further and more detailed knowledge of the dynamic behaviour of the spliceosome will aid such crystallographic efforts, and will also be necessary for the rationalisation and accurate interpretation of their results. The branch site-U2 interaction, as the focus of this work, illustrates many of the challenges posed by the dynamic spliceosome and our imperfect knowledge of the conformational states during which specific interactions are relevant.

This work unequivocally shows that BS-U2 pairing is required to bulge and activate the BS nucleophile for first step catalysis (Chapter 4). The use of sites in the branch site binding region of U2 as 5' splice sites for both steps of splicing (Chapter 3) suggests that the BS-U2 pairing established during spliceosome assembly does not, as previously assumed, persist after the first catalytic step. The fine details of our analysis, however, indicate substantial further complexity. That multiple *trans*-spliced sites exist in U2 (Fig. 3-5) implies that at least one state exists between the establishment of BS-U2 pairing and the first catalytic step in which the branch site is unpaired from U2. This should perhaps not be

surprising due to the long and complex nature of the spliceosome assembly pathway: it is known that U2 toggles between its stem IIa and IIc forms during assembly (48, 49), and the canonical BS-U2 interaction may toggle similarly, either synchronously or asynchronously with the presence of U2 stem IIc (Fig. 5-1).

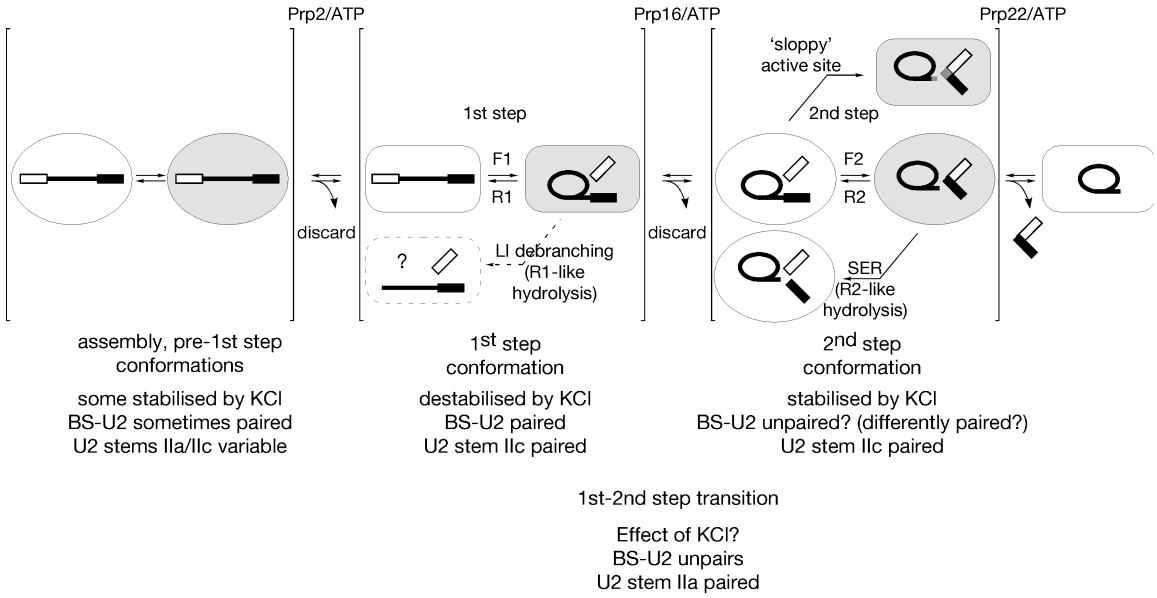


Figure 5-1. Spliceosomal side reactions and conformational toggling.

Schematic depiction of the observed and potential reactions occurring during reversible splicing catalysis by compositionally homogeneous spliceosomes, as described in (59). Interactions that may toggle during the splicing pathway, as well as the effect of potassium ions on certain states, are noted underneath.

Despite the apparent lack of a requirement for BS-U2 base pairing for exon ligation by the spliceosome, U2 must play a role in this step. Crosslinks have been observed between the 3'SS and helix I-proximal nucleotides in U2 (166), and insertions around the U2 component of helix Ia block the second step *in vitro*

(167, 168). In addition, moving the branch nucleophile within the catalytic centre often imposes a second step block (cf. Fig. 4-3, 4-11&4-12). It is possible that this defect is at least partially due to the formation of a non-canonical branch. The formation of the branch structure likely eliminates the positional and rotational flexibility of the BS-U2 duplex; this in turn could influence the repositioning of the branch structure that is required for entry into the second step (44). An aberrant branch may not reposition efficiently, perhaps due to a poor fit with the second step binding site that likely exists for the lariat branch; alternatively, the disruption of the BS-U2 duplex suggested by the data shown in chapter 3 may be inhibited.

The lariat branch is almost certainly specifically bound during the second step of splicing, but the nature of its binding site is unclear. Between the two catalytic steps, the 5'SS is thought to move to a new binding site on U6 snRNA 5 nt upstream of its first step site. Given that the 5'SS and BS are covalently linked at this stage in the reaction, movement of the BS is presumably a necessary corollary of 5'SS repositioning. An appealing possibility is that the BS could move 3nt along U2, such that the 5'-UAC**UAAC**-3' branch sequence paired with the second GUA repeat in the U2 5'-GΨA**GUA**-3' BS-binding sequence rather than the first. Such repositioning would also be consistent with the observed 5'SS use during *trans*-splicing, as the downstream-most 5'SS in U2 is at GΨA/GUA. We have, however, been unable to find genetic evidence in support of a 3 nt repositioning event: BS-U2 pairs incapable of forming the putative

second step interaction display no exon ligation defect (Fig. 4-4C), even in the context of a 3'SS mutation that should make the second step limiting for overall splicing (data not shown). In a less hypothesis-driven line of experiment, we attempted crosslinking-based experiments to identify the second-step binding site of the BS-side of the lariat branch. A UBC4 RNA site-specifically derivatised with 4-thio-uridine (UAC[4SU]A**A**C) did not detectably crosslink to any snRNAs: the synthesis of a similarly derivatised version of actin, for which the second step is less efficient and which may therefore represent a more promising substrate, proved too inefficient to be experimentally feasible (data not shown).

The importance of establishing the temporal details of spliceosomal interactions is clear. Tseng and Chen, referring to the effects of ionic environment on the efficiency of various splicing reactions *in vitro* (59), note that the selective stabilisation of spliceosomal conformations via modulation of salt conditions will facilitate the enrichment of defined states for experiments such as snRNA secondary structure mapping and hopefully crystallography; hyperstabilisation of combinations of interactions unique to a given state would allow further enrichment, and should also help to inhibit the reverse and side reactions (some such reactions are indicated in Fig. 5-1) observed to occur in stalled spliceosomes, which will otherwise severely diminish homogeneity. As a final note, and as alluded to in chapter 1, the large number of conformational transitions that appear to be necessary for splicing may render difficult the interpretation of the high-resolution crystal structures we hope will become

available in the future. Knowledge of the RNA-RNA interactions particular to each state of the spliceosome will help to alleviate these difficulties.

GENERATING NEW AND EXTENDED ORTHOGONAL SPLICING SYSTEMS

Many important mechanistic analyses of splicing have been carried out *in vitro*; the purity and lack of constraint of such biochemical approaches, often offset by technical difficulties and the unnatural environment inherent in all *in vitro* work, can be recapitulated *in vivo* by nonessential spliceosomes that do not impact endogenous splicing. Further such spliceosomes, both derived from that described here and entirely novel, could prove to be invaluable in the continuing analysis of the catalytic phase of splicing.

The analysis of branch site-U2 interaction described in chapter 4 builds on previous genetic analyses of the spliceosome, employing compensatory base pairs and second copy mutants; these approaches have previously identified and validated many RNA-RNA interactions in both the yeast and human spliceosomes (61, 149, 169, 170), and the concept of using compensatory mutations is clearly not novel in itself. However, our system is distinguished by its conception to address a specific mechanistic question regarding splicing catalysis, rather than to identify/validate an intermolecular interaction. In addition, our results show the acceptability of gross sequence substitutions within an extremely conserved element in the core of the spliceosome. The minor spliceosome can be thought of as a natural example of an orthogonal BS-U2

system, specific for a subset of introns. This system relies on several dedicated interactions (e.g., U11-5'SS, U12-BS, U12-U6atac), suggesting that analogous extensions to our system should be feasible. We hope that other highly conserved interacting partners will be similarly amenable to substitution; this will allow the generation and *in vivo* analysis of spliceosomes with more and/or different components separate from the endogenous splicing machinery.

The most obvious candidate for a second component to be incorporated into orthogonal spliceosomes is U6 snRNA. As previously discussed, U6 forms several intermolecular base pairs with U2 snRNA and, although these interactions may be more dynamic than predicted by current models of the spliceosome active site, the requirement of such structures for cellular viability (61), the existence of a dedicated U6 snRNA in the minor spliceosome, and the data presented in chapter 3 (cf. Fig. 3-4), suggest the feasibility of generating rationally designed U2/U6 pairs (Fig. 5-2A&B). Thus far, however, our attempts to incorporate dedicated second-copy U6 snRNAs into our system have been unsuccessful, with mutant U2 molecules appearing to work equally well with wild-type U6 as with their intended mutant partners (two representative helix Ia mutants are shown in Fig. 5-2C; gross substitutions throughout helices I and II have been tried).

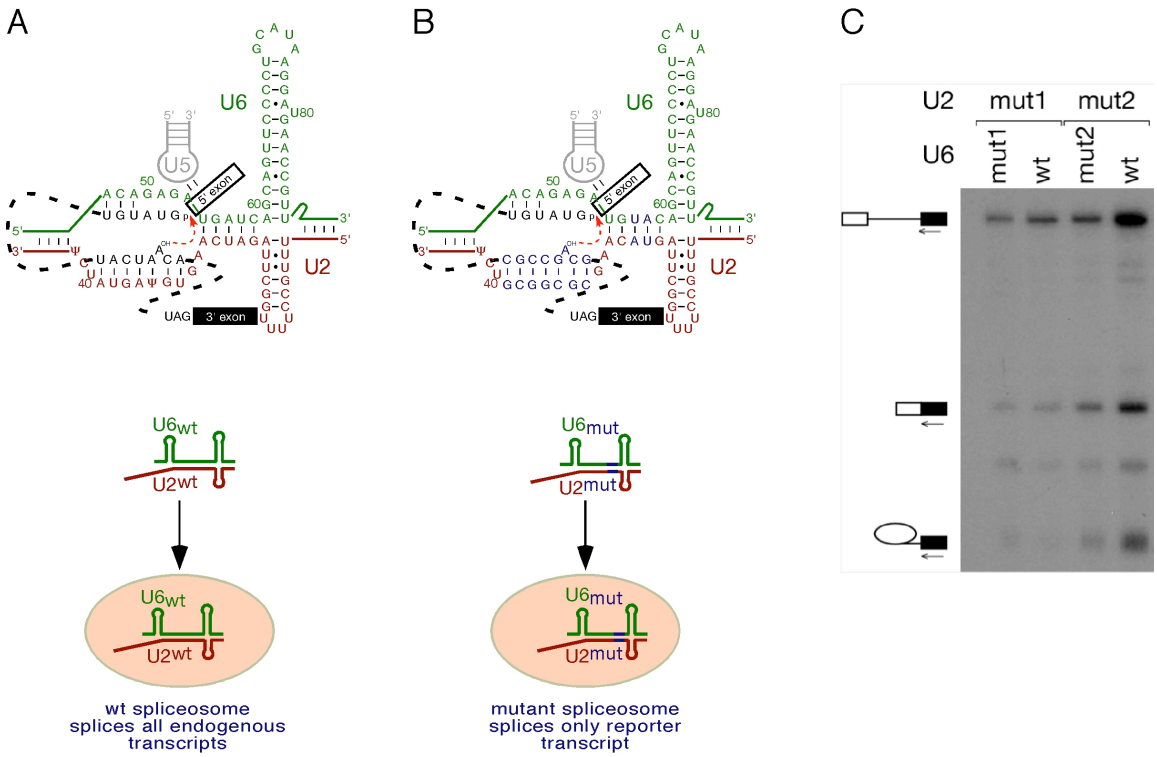


Figure 5-2. Initial attempts to incorporate U6 into the orthogonal BS-U2 system.

Schematics depicting the nature of wild-type (A) and orthogonal (B) spliceosomes. The orthogonal spliceosome in [B] contains U2 snRNA with a substituted BS-binding region and a mutations in helix Ia that are inviable *in vivo* in the absence of compensatory mutations; mutant U6 contains these compensatory changes in helix Ia. Only wtU2-wtU6 and mutU2-mutU6 pairing should be possible. (C) Primer extension on total RNA from strains carrying substrates with all-CG branch sites and U2/U6 pairs with two representative sets of helix Ia mutations. The intended specificity of wild-type and mutant snRNAs for their pairing partners is insufficient for the incorporation of mutant U6 into spliceosomes containing mutant U2.

In general, second-copy mutant U6 snRNAs show much weaker effects than do second-copy U2s, likely due to very low-level expression of the mutant allele (Charles Query and Magda Konarska, unpublished data); this could be a factor underlying our current inability to incorporate U6 into our system. If poor expression does indeed at least partially underlie the problem, it is likely that a

simple solution exists – expression of U6 from a higher copy-number plasmid or different promoter, for example, may be sufficient to allow strong expression of mutant U6 snRNAs.

Weak expression of mutant U6 alleles is not the only impediment to the establishment of an orthogonal U2/U6 system. From Fig. 5-1C it is clear that both wild-type and mutant U6 can be incorporated into functional spliceosomes with mutant U2, possibly suggesting that U2/U6 pairing does not limit the splicing of our mutant reporters and that the spliceosome can therefore tolerate a lack of pairing in helix I or II. It may therefore be necessary to incorporate additional mutations that impair the formation of catalytic spliceosomes containing mismatched U2 and U6 molecules. An intuitive approach, reminiscent of experiments used to identify helix III (149), would be to use a reporter gene carrying a mutated 5'SS in addition to its substituted BS, and for the mutant U6 to carry a complementary substitution (e.g. 5'SS position G5 and U6 position C48) in addition to helix I/helix II mutations. One potential pitfall of this approach may be the detrimental effects of multiple substrate and spliceosome mutations on splicing efficiency – grossly substituted reporters are already spliced at least threefold less efficiently than wild-type reporters (Fig. 4-4).

U5 snRNA, Prp8 and various components of the NTC (for example, Cef1) represent further desirable targets for work in this direction. Our incomplete knowledge of intermolecular interactions within the spliceosome will, however,

make the incorporation of these components more challenging than is the case for U6 snRNA, and will almost certainly preclude rational design. It is possible that non-hypothesis-driven approaches such as genetic screens may allow us to identify, for example, Cef1 alleles that only function in the context of substituted U2 and U6 molecules. If such screens prove to be successful, the utility of our system be greatly extended: additionally, the mutations required for the chosen components to function in the mutant spliceosome should be informative in themselves.

DEXD/H-BOX ATPASES

Although it is known that each ATPase-mediated step in the splicing pathway affords an opportunity for either discard or progression, the molecular details of how these are achieved are unclear. As noted in chapter 1, it is not known whether the precise molecular changes that induce progression are the same as those that lead to discard if the preceding step has not been completed: in addition, while helicase activity is often assumed to underlie the effects of these proteins, the identification and validation of direct targets has proved challenging.

Many DExD/H ATPases, both spliceosomal and non-spliceosomal, have been shown to have RNA unwindase activity correlating with ATPase activity *in vitro*: among the spliceosomal ATPases, Brr2, Prp16 and Prp22 (93, 171, 172) have been directly demonstrated to possess such helicase activity. Brr2, which acts at the stage of U4 dissociation from the spliceosome, can unwind U4/U6 in vitro;

this activity is greatly stimulated by a fragment of Prp8 (74), and appears likely to be physiologically relevant. It is therefore tempting to speculate that each ATPase is associated with the disruption of a single RNA duplex.

Despite the example of Brr2, it is unclear whether ATPase/unwindase activity is the mechanistic basis of all spliceosomal activity of DExD/H ATPases. Although conserved helicase activity is suggestive, the possibility that not all of these proteins rely on this activity to modulate spliceosomal transitions should not be discounted. ATPase-defective *prp28* alleles impose a block on the splicing pathway, inhibiting the release of U1 and U4; this is also inhibited by hyperstabilisation of the duplex formed between the 5'SS and U1 snRNA (60). Such data could be taken to imply that the role of Prp28 is to unwind the U1-5'SS duplex; these observations, however, could be coincidental: inhibition of similar or even identical steps on the splicing pathway by different mutations does not necessarily indicate mechanistic identity. Prp22 has been proposed to stimulate mRNA release by the spliceosome by first binding downstream of the 3'SS, then translocating along the 3' exon and disrupting the base pairing between U5 snRNA and the newly spliced exons (173); this hypothesis, however, is based on genetic interactions between *prp22* alleles and those of other spliceosomal components (U5, Prp8, Slu7) (52, 174) reminiscent of the interaction between *prp22* and *prp8* alleles discussed in the context of chapter 1, whose effects are likely indirect (47).

Data from non-spliceosomal DExD/H ATPases makes matters still more complex. Cyt19, a DExD/H protein that promotes group I intron splicing *in vivo* and *in vitro*, does so by acting as a non-specific chaperone, resolving kinetic traps along the folding pathway (175); similarly, splicing of the ai5γ group II intron in *S. cerevisiae* is stimulated by Mss116, and splicing activity *in vitro* correlates with Mss116 ATPase and unwinding activity (176). However, the same study reported a residual, ATPase-independent unwinding activity for Mss116, and recent data suggest that Dbp5, which is involved in mRNA export, functions to remodel RNPs only in its ADP-bound form, with ATP hydrolysis thus acting as a conformational switch rather than a power stroke (177).

Even if some spliceosomal ATPases exert their effects through helicase activity, it need not necessarily be the case that all do. The ability of these proteins to bind nucleic acids, together with the profusion of RNA in both the spliceosome and its substrate, however, make it almost inevitable that their association with the spliceosome will depend heavily on protein-RNA interactions. An obvious implication is that RNA structures specific to a given stage of the reaction are at least as likely to be recognised by relevant ATPases as they are to be unwound. Indeed, the recognition of stage-specific structures could even help to stimulate ATPase-mediated rearrangement (again, regardless of whether or not this includes active duplex unwinding via helicase activity). Such stimulatory action could theoretically act as a mechanism to increase the rate at which optimal

substrates react, akin to the specific rate acceleration observed for cognate ribosomal substrates (178).

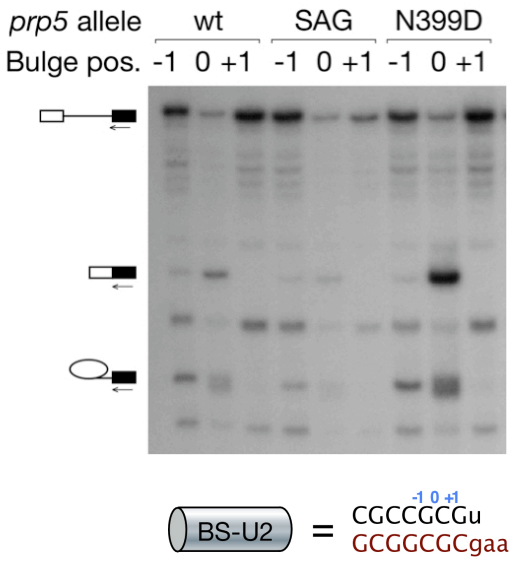


Figure 5-3. Effects of different *prp5* alleles on grossly substituted BS-U2 pairs.

Primer extensions on CGCCG**A**C reporters with BS adenosines bulged from the -1, 0 or 1 positions, grouped by *prp5* allele. The -*T448G* (SAG) mutant impairs splicing of the 0 position reporter, presumably by increasing discard, while the –*N399D* mutant increases splicing efficiency.

The behaviour of the grossly substituted BS-U2 pairs described in chapter 4 in combination with different alleles of *prp5* may offer support for a model in which a transition mediated by ATP hydrolysis by Prp5 is stimulated by a wild-type BS-U2 duplex. The *prp5–SAG* and *–N399D* mutants, both isolated by Xu and Query as suppressors of splicing defects due to single mismatches in the BS-U2 duplex (36), show distinct effects on the splicing of grossly substituted reporters (Fig. 5-3). *prp5-N399D* strongly improves splicing, while *prp5-SAG* inhibits it, with

substrate discard instead appearing to predominate. A crucial difference between these two *prp5* alleles is that –*SAG* is ATPase-defective, while –*N399D* is not (36). The behaviour exhibited in Fig. 5-3 is consistent with the failure of the substituted BS-U2 duplex to stimulate ATPase activity, leaving too little residual activity for efficient progression in the context of the already defective –*SAG* allele, but sufficient activity in the context of –*N399D* (whose mechanism of suppression remains unknown). Of course, while this result is consistent with the hypothesis that Prp5 is stimulated by the wt BS-U2 duplex, the same caution of interpretation stressed in relation to Prp28 and Prp22 is required. It is worth noting, however, that recognition and unwinding of the BS-U2 duplex by Prp5 could give rise to the apparent pre-first step unpairing of the branch site (chapter 3).

THE NATURE OF THE SPLICESOMAL ACTIVE SITE(S)

The data presented in chapter 4 suggest similarity in the spatial organisation of various active site and substrate elements between group II introns and the spliceosome. These splicing systems, which catalyse the same two chemical reactions, are thought to be evolutionarily related. As noted in previous chapters, a great number of similarities beyond simple mechanistic identity stand in support of this hypothesis (these data have already been detailed, and will not be re-stated here). The generally accepted evolutionary link between group II introns and the spliceosome has led to the almost equally global inference that the spliceosome must be a ribozyme – that is, that the divalent metal ions

presumably required (114) for the catalytic steps, as well as the pre-mRNA nucleotides involved in these reactions, are bound directly by snRNA.

A critical difference between the splicing activities of group II introns and the spliceosome is that, while each group II intron catalyses its own removal in a single-turnover reaction, the spliceosome must assemble and perform catalysis in *trans* on diverse intron substrates. Specificity that can be attained by intramolecular folding in group II introns must be established and maintained via other mechanisms during spliceosomal splicing. Given that splicing must occur in *trans*, and that accurate splicing is essential for viability in all known eukaryotes, it is reasonable to speculate that substrate recognition may be more complex in the spliceosome than in self-splicing introns. Consistent with such increased complexity of selection, increasing the distance between domains V and VI in group II introns can activate both ectopically bulged branch adenosines and more conventionally defined cryptic branch sites (151), while in the spliceosome (at least in *S. cerevisiae*), changing the distance between the equivalent structures only gives rise to the former (chapter 4).

Repeated recognition of a motif can, if mediated via different interactions, generate increased substrate selectivity (179). The protracted assembly and activation phase required for spliceosome action thus presumably serves to allow effective discrimination between substrates as well as presenting multiple points of regulation. One of the key mediators of selectivity is likely to be Prp8, the

largest and most highly conserved protein in the spliceosome, which contacts each of the three reactive substrate sites at some stage during the reaction (50). Intriguingly, the C-terminal domain of Prp8 has recently been shown independently by three groups to adopt an RNase H fold (121-123); this domain has previously been shown to contact the GU dinucleotide at the 5'SS (180), and binds a structured RNA reminiscent of proposed intron-snRNA complexes (122). Although the residues corresponding to the RNase H active site do not conform to the consensus or bind metal ions in any of these crystal structures, a similar active site arrangement can be found in catalytically active RNase H domains, as noted by MacMillan and colleagues (122). This raises the possibility that Prp8 may directly participate in one or both steps of splicing catalysis rather than solely contacting splice sites during assembly. It may therefore be the case that splicing catalysis, as well as assembly of the relevant enzyme, is more complex in the spliceosome than in group II introns. Regardless of Prp8's direct contribution to catalysis, the intimate involvement of this and other proteins means that, even if the metal ions required for splicing catalysis are coordinated entirely by groups from snRNAs, the spliceosome might most productively be considered an RNP with an RNA active site, rather than a ribozyme.

Splicing catalysis, comprising two sequential transesterification reactions, almost certainly makes use of two metal ion catalysis (114). As part of their exhaustive analysis of the ionic requirements for forward and reverse splicing, Tseng and Chen have investigated the effects of divalent metal ions on the efficiency of

each step (59). The results indicate that Ca^{2+} or Mn^{2+} can substitute for Mg^{2+} for R1 (first step reversal); Mn^{2+} can support F2 but with relaxed specificity relative to Mg^{2+}, and only Mg^{2+} can support R2. Mn^{2+} or high pH, both of which fail to support R2, instead appear to catalyse spliced exon reopening (SER) – a reaction first identified in self splicing introns (181, 182) and considered to be a reversal of the second step in which water, rather than the 3' end of the excised intron, acts as the nucleophile. The distinct catalytic capabilities of divalent metal ions suggest subtle differences in active site requirements for the two steps of splicing, although the effects of different ions could be either direct – resulting from coordination of these ions in the active site itself, or indirect – resulting from more global structural differences due to coordination of these ions at other sites in the spliceosome. Stereochemical evidence has been used to argue for two spliceosomal active sites (183), even if at least some key elements are shared between the two. The involvement of a given motif in one step does not necessarily indicate a role in the other, and it remains to be clarified how similar the active sites for the two catalytic steps are: although much genetic, biochemical and structural work remains to be carried out, answers are hopefully not too distant.

MATERIALS AND METHODS

IN VITRO TRANSCRIPTION AND TRANSCRIPT PURIFICATION

Transcription templates were prepared either by PCR, adding the [T7 promoter + start] sequence CGCGTAATACGACTCACTATAGGGAAAGCT to the 5' end of the amplified product, or by cloning into the pDrive vector (Qiagen), which contains a T7 promoter. Where homogeneous 3' ends were required, templates were prepared by PCR using an antisense primer in which the 5'-most two nucleotides carried 2'-OMe substitutions, as this prevents the addition of non-templated adenosine nucleotides to the 3' end of the transcript by T7 polymerase. Transcription was carried out using 2 μl of the lab stock of T7 polymerase and 1 μg template DNA per 10 μl reaction for 2-4h at 37 ˚C in 1x transcription buffer (40 mM Tris-HCl, pH 8.0, 8 mM MgCl2, 2 mM spermidine-HCl, 50 mM NaCl, 10 mM DTT). NTPs were present at 1 mM each and supplemented with radiolabelled NTP unless high specific activity was desired, in which case the unlabelled NTP corresponding to the label was at 130 μM. Reactions were stopped by the addition of 5 U DNaseI and incubation at 37 ˚C for a further 30 min. Generally, transcripts were purified from polyacrylamide/urea gels (if larger than 20 μl, they were first extracted and precipitated to reduce their volume). An equal volume of deionized formamide containing bromophenol blue and xylene cyanole was added to the DNase-treated reaction, this was heated at 95 ˚C for 2 min, then loaded into the gel. Gel slices of interest were excised, crushed, and rotated overnight at 4 ˚C with 400 μl

each of Tris-phenol and 0.3 M NaOAc, pH 5.5. The aqueous phase was collected, avoiding gel fragments, and 20 μg glycogen and 40 μl 3M NaOAc, pH 5.5 added before ethanol precipitation.

TRANSCRIPT LABELING AND QUANTITATION

Transcripts were generally body-labeled with either [α^{32}P]-U- or ATP and quantitated via scintillation counting (Cerenkov) Vs the stock nucleotide, which served as a standard of known concentration and activity. Where accurate quantitation of non-radioactive transcripts was required, small aliquots of transcription reactions were generally removed immediately after addition of polymerase, and spiked with a known amount of [α^{32}P]-UTP. These spiked aliquots were processed and purified alongside their parent reactions.

OLIGONUCLEOTIDE LABELLING AND PURIFICATION

5' end-labelling was carried out using T4 Polynucleotide kinase (PNK). 15 μl reactions contained 1 μl Tris-HCl pH 7.5, 1 μl 200 mM $MgCl_2$/34 mM spermidine-HCl, 1 μl 0.3M β-Mercaptoethanol, 1 μl T4 PNK, 10 μl [γ^{32}P]-ATP (normally 3000 Ci/mmol) and 1 μl oligonucleotide diluted to produce a labeled stock of the desired specific activity (generally 100 μM for low, 20 μM for high, and 5 μM for very high). Reactions were normally either 7.5 or 15 μl, and ran at 37 ˚C for 30-120 min depending on the desired yield and the nature of the substrate (RNA or DNA). Short internally labeled substrates, including those with a 3'-appended *pCp, were prepared by ligation with T4 RNA ligase. The downstream ligation

substrate was kinased as normal (7.5 μl reaction, low S.A to use up all the [γ^{32}P]-ATP): this was heated to 65 ˚C for 3 min to inactivate PNK, followed by the addition of 0.5 μl 100 μM stock of the upstream ligation substrate, 1.1 μl 10x RNA ligase buffer (NEB) and 2 μl T4 RNA ligase. Ligation reactions were incubated at 37 ˚C for 40-180 min. End-labeled substrates were generally, and internally labeled substrates always, purified from polyacrylamide/urea gels.

SITE-SPECIFIC LABELLING OF LONG RNAS

Ligation fragments were purified transcripts (prepared with homogeneous 3' ends) and derivatised oligonucleotide, 5' end-labelled at very high specific activity. DNA splint oligonucleotides, capable of forming at least 15 base pairs with each substrate to generate a nicked duplex of length >30 bp were used. Both ligation substrates were quantitated, and were mixed in a 1:1 molar ratio with 0.8 molar equivalents of the DNA splint oligonucleotide. AB buffer (see primer extension) was added to a final concentration of 0.5x, and the mixture heated to 95 ˚C, cooled slowly to 50 ˚C and placed briefly on ice. T4 RNA ligase 2 buffer was added such that its final concentration would be 1x, followed by 20 U T4 RNA ligase 2 (NEB). Reactions were supplemented by the addition of unlabelled ATP to a final concentration of 0.5 mM, and incubated at 37 ˚C for 2 hr. Products were purified from polyacrylamide/urea gels.

RNA OLIGONUCLEOTIDES

RNA oligonucleotides were ordered from Oligos Etc. (TBDMS chemistry) or from Dharmacon (2' ACE chemistry or 4-thio-uridine modifications), and were treated according to the manufacturer's instructions.

PROTEIN-FREE SPLICING REACTIONS

~20-1000 fmol labeled BS oligonucleotide (50k-3M cpm) and >1000 fmol U26 was used for each reaction. A typical reaction was 5-10 ml and contained 50 mM Tris-HCl pH 7.2, 40 mM $MgCl_2$, 200 mM KCl. The mixture was heated to 95 ˚C, cooled slowly to room temperature, left for 1-24h, ethanol precipitated and separated in a 20% polyacrylamide/8M urea gel.

PARTIAL ALKALINE HYDROLYSIS AND DIGESTION OF RNA.

For partial hydrolysis, RNA was heated at 95 ˚C for 5-20 min (somewhat length and sequence-dependent) in 50 mM Na_2CO_3, pH9.2. RNA to be digested was mixed with 1 μl yeast total RNA (corresponding to 1 30^{th} the amount of RNA purified from a saturated 10 ml yeast culture in minimal medium), and digested for 1h according to the conditions below:

RNase	Amount	Buffer	Temp (˚C)
A	1 μg	50 mM Tris-HCl, pH 7.5	37
T1	2U	50 mM Tris-HCl, pH 7.7	37
T2	1 μg	50 mM NaOAc, pH 5.5	50
P1 nuclease	0.5 μg	50 mM NaOAc, pH 5.5	50

THIN LAYER CHROMATOGRAPHY

Solvents for 2-dimensional thin layer chromatography were as follows: dimension 1 – 57.7% n-butyric acid, 0.9% ammonium hydroxide in H_2O; dimension 2 – 70% isopropanol, 15% conc. HCl in H_2O. Plates were loaded 1 μl at a time with 5 μl of the appropriately digested RNA, followed by 1 μl of markers (generally Np or pN) at 1 mM each. Plates were dried completely between dimensions by sitting on a 50 ˚C hotplate for at least 8 hours. The solvent for DEAE TLC was 0.2 M ammonium formate, 1 mM EDTA, 9M urea. All plates were dried thoroughly and wrapped in cling film before autoradiography.

CLONING AND PLASMID GENERATION/SHUTTLING

Plasmids were generally prepared by gap repair recombination in *S. cerevisiae* strain 46ΔCUP. A PCR insert containing the desired mutations was generated, and yeast transformed with this product and a cut vector before auxotrophic selection. Briefly, 30 μl competent cells were incubated in 140 μl transformation buffer (40% PEG-3500, 0.1 M LiOAc supplemented with 10 μg boiled salmon-sperm DNA) with 1 μg cut vector and 1/10 of a phenol:chloroform extracted and ethanol precipitated 100 μl PCR reaction. Plasmids were isolated from yeast and transfected into *E. coli* as follows. 5ml cultures were grown in the appropriate minimal medium and collected by spinning at 3500 rpm for 5 min. Cells were resuspended in 200 μl STET buffer (8% sucrose, 50 mM Tris-HCl, pH 8.0, 50 mM EDTA, 5% Triton X-100), and 200 μl acid-washed glass beads (Sigma) added. Tubes were vortexed at maximum power for 5 minutes, heated to 95 ˚C

for 3 min, cooled on ice and the supernatant (150 μl) from a 10 min 14000 rpm spin at 4 ˚C collected to a tube containing 150 μl 4M NH_4OAc. This was incubated at -20 ˚C for at least 1hr, and the supernatant (300 μl) from a 10 min 14000 rpm spin at 4 ˚C collected. DNA was precipitated for at least 1hr at -80 ˚C with 30 μg glycogen and 450 μl ethanol. DNA pellets were dried thoroughly and resuspended in 21 μl H_2O. 7 μl were used for transformation into chemically competent bacteria, or 2 μl into electrocompetent bacteria. Following a standard miniprep (using a Qiagen kit) and sequencing, plasmids were transfected into the appropriate yeast strain using a protocol identical to that used for gap repair.

YEAST STRAINS USED

Strains used in this work were Y04999 [MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YKL149c::kanMX4] (derived from BY4741), 46DCUP [MATa, ade2 cup1Æ::ura3 his3 leu2 lys2 trp1 ura3, GAL+], yCQ62 [MATa ade2 cup1Δ::ura3 his3 leu2 lys2 trp1 ura3 snr6Δ::loxP snr20Δ::loxP, pU6+U2 (SNR6 SNR20 URA3 CEN ARS)], yCQ64 [MATa ade2 cup1Δ::ura3 his3 leu2 lys2 trp1 ura3 snr6Δ::loxP snr20Δ::loxP YKL149cΔ::loxP, pU6+U2 (SNR6 SNR20 URA3 CEN ARS)], yZX02 [MATa ade2 cup1Δ::ura3 his3 leu2 lys2 trp1 ura3 prp5Δ::loxP] and yCQ08 [MATa ade2 cup1Δ::ura3 his3 leu2 lys2 trp1 ura3 prp16Δ::loxP].

YEAST TOTAL RNA PREPARATION

Yeast cultures were picked to 1 ml cultures of minimal medium lacking the appropriate supplements. These were grown to saturation at 30 ˚C, an aliquot

taken for copper growth assays (see below), and an aliquot used to inoculate a 10ml culture of the same minimal medium. 10 ml cultures were grown to late log/early stationary phase before processing. Cells were collected by centrifugation at 3500 rpm for 5 min at 4 ˚C, washed once with water, transferred to 1.5 ml tubes, and kept on ice before being transferred to -80 ˚C for at least 20 min. To the still-frozen cell pellets were added, in order, 420 μl Tris-phenol, 42 μl 10% SDS, 420 μl AE buffer (50 mM NaOAc, pH 5.5, 10 mM EDTA). The tube was vortexed until the cell pellet was fully resuspended, then placed at 65 ˚C for 4.5-6.5 min (pressure was released from the tube, and the mixture vortexed, after ~1.5 min). After a brief vortex and a 5 min microfuge spin at 14000 rpm, 2x 180 μl of the aqueous phase was collected (in order to clear the aqueous phase, each sample was heated to between 60 and 75 ˚C for ~30 sec before collection) into a tube containing 180 μl 50:50 phenol:chloroform. This was spun at 14000 rpm for 5-10 min in a microfuge, and 2x150 μl of the aqueous phase collected to a tube containing 30 μl 3M NaOAc, pH 5.5 and precipitated with 1 ml EtOH. Nucleic acid pellets were resuspended in 30 μl water before use.

This RNA preparation protocol contains two potential pausing points: cells can be frozen at -80 ˚C for at least 6 months with no detrimental effect on RNA quality, and the phenol/SDS/AE mix, after heating, can be quick-frozen on dry ice and stored at -80 ˚C at least overnight.

PRIMER EXTENSION

1.2 μl of a yeast total RNA prep was added to 5 μl radiolabelled primer (20k-200kcpm undecayed, depending on the specific activity of the primer and the desired sensitivity of the reaction) in 1x AB (annealing buffer: 50 mM Tris-HCl, pH 8.3, 60 mM HCl, 10 mM DTT). Tubes were placed in a 95 ˚C heat block on the benchtop, which cooled slowly to 50 ˚C: at 70 ˚C, tubes were spun in a nanofuge for 5 sec to collect liquid that had evaporated and condensed on the lid. At 50 ˚C, tubes were placed on ice for 1-3 min, followed again by a brief spin. To the 6.2 μl reaction was added 6 μl 2.5x dNTPs (all 4 dNTPs at 200 μM each, in 1x AB) 3 μl 5x RT buffer (250 mM Tris-HCl, pH 8.3, 300 mM NaCl, 150 mM MgOAc, 50 mM DTT) and 3 μl AMV RT (diluted 1:50 from the purchased stock in 0.2 M $KHPO_4$, pH 7.2, 2 mM DTT, 0.2% Triton X-100, 50% glycerol in H_2O). The mixture was incubated at 42 ˚C for 20-80 min then either mixed with an equal volume of formamide + dyes or ethanol precipitated and resuspended in this, then loaded onto a denaturing polyacrylamide gel. Products were visualized by autoradiography. For sequencing primer extensions, four mixes were made in which the dNTP concentrations were halved, and the appropriate dNTP supplemented with 1/5 concentration of the corresponding ddNTP. For primer extensions to assay relative levels of wild-type and orthogonal U2 snRNA, dTTP was replaced by an identical concentration of ddTTP in the reaction mix.

YEAST COPPER GROWTH ASSAYS

Aliquots of saturated 1 ml RNA prep pre-cultures were diluted 1:5 with water, and 8 μl spotted onto SD-Leucine plates containing various concentrations of $CuSO_4$. Typical copper concentrations used were (in μM): 0, 25, 50, 75, 100, 200, 300, 400, 500, 600, 800, 1000, 2000.

QUANTITATION OF SPLICING EFFICIENCY

Primer extension results were visualized by autoradiography: films were photographed the resulting image files converted to grayscale. Lanes were defined and the area under peaks corresponding to bands of interest calculated using ImageJ. Numbers thus obtained were used in Excel to calculate splicing efficiencies. Where pre-mRNA, mRNA and lariat intermediate are P, M and LI, first step, second step, and overall splicing efficiency were defined as (M+LI)/(P+M+LI), M/(M+LI) and M/(P+M+LI), respectively.

RT-PCR

For RT-PCR reactions to amplify *trans*-spliced U2-ACT1-CUP1, yeast total RNA was treated with DNaseI (Sigma) according to the manufacturer's instructions. Reverse transcription was carried out (using the primer extension protocol described above) on the equivalent of 2 μl total RNA, using the primer YAC94 5'-CGTCGCTGTTACACCC-3'. Reverse-transcribed DNA was amplified for 22 cycles using primers YAC94 and yU2-22 5'-ACGAATCTCTTTGCCTTTTGGC (53.5 ˚C annealing temperature), diluted 1:1000 and further amplified using

primers yU2-22 and YAC6 5'-GGCACTCATGACCTTC-3' (30 cycles). Products were separated in 2% agarose gels. RT-PCRs for loading controls were performed with yU2-138-154 5'-AAAGTCTCTTCCCGTCC-3' and yU2-22 as antisense and sense primers, respectively. PCR products were purified for sequencing and cloning using Qiaquick gel extraction kits (Qiagen) according to the manufacturer's instructions. RT-PCR reactions to amplify spliced *ACT1-CUP1* and *ACT1-U2-ACT1-CUP1* products used YAC94 and YAC11 5'-(**SEQ??), and did not require re-amplification with nested primers.

YEAST WHOLE CELL EXTRACT PREPARATION

2 l cultures were grown to O.D. ~3 in YPD (or O.D. 1.5 in SD-His if a second-copy U2 snRNA was to be maintained) and collected by centrifugation at 3500 rpm for 5 minutes. Cells were washed once with 200 ml cold H_2O and with 50 ml then 20 ml of cold buffer AGK (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 200 mM KCl, 10% glycerol, 0.5 mM DTT): if resuspension was difficult, a sterile 25 ml Pasteur pipette was used. A slurry was then made by adding an amount of cold AGK corresponding to around 0.4x the volume of the cell pellet. To this was added 2 μl 1x EDTA-free protease inhibitor cocktail V (Calbiochem). The slurry was dripped into liquid nitrogen using a p1000, expelling around 200 μl per drop, and stored at -80 ˚C before further processing. The frozen cells were ground using a Retsch MM301 ball mill exactly as described in the technical literature (available at http://www.retsch.com/products/milling/ball-mills/mm-400/product-information/dlDetails/1/file/1538/), using 5 cycles of 3 min at 10 Hz. The powder

was thawed at room temperature, stirred gently at 4 ˚C for 20 min, then spun at 35,000g for 30 min (for large volumes, a J-20 rotor was used; for small volumes, a TLA 100-3 rotor was used at 25,000 rpm). The aqueous phase between the top lipoprotein layer and the pellet was carefully collected, and spun again at 100,000g for 1 hr (43,000 rpm, TLA 100-3 rotor). Again, the aqueous phase was collected. This was dialysed in MwCO 10,000 tubing against 5 x 200 ml buffer D (20 mM HEPES, pH 7.9, 0.2 mM EDTA, 50 mM KCl, 20% glycerol, 0.5 mM DTT), changing the dialysis buffer every 30 min. The dialysed preparation was kept on ice and dropped 20 μl at a time into liquid nitrogen before storage at -80 ˚C.

IN VITRO SPLICING

4 μl yeast whole cell extract was added to tubes containing 1 μl 20 mM ATP, 1fmol splicing substrate (in 1 μl), 1 μl 60 mM $KHPO_4$, pH 7.0, 1 μl 25 mM $MgCl_2$, 1 μl 30% PEG-8000 and 1 μl of either H_2O or superasein (Ambion). Placental RNase inhibitor must not be used, as it does not work in the absence of DTT and in fact acts as a source of co-purified nucleases. Reactions were carried out at room temperature. After the appropriate amount of time, these were stopped via the addition of 240 μl H_2O and 200 μl 50:50 phenol:chloroform. Reactions were extracted before ethanol precipitation in the presence of 10 μg glycogen (one extraction was required for clean results with UBC4 substrates and two with actin substrates). Freezing the tubes in dry ice immediately after the initial addition of water and phenol:chloroform helped to produce pellets that could readily be

resuspended. Pellets were resuspended in formamide + dyes, then loaded on denaturing polyacrylamide gels.

SEPARATION OF SPLICING COMPLEXES IN NATIVE ACRYLAMIDE AGAROSE GELS

3 μl aliquots of splicing reactions were taken after an appropriate incubation, and added to ice-cold tubes containing 3 μl 50% glycerol with xylene cyanole and bromophenol blue, and 3 μl Buffer R (2 mM MgOAc, 50 mM HEPES-NaOH, pH 7.5) containing either HeLa or *S. cerevisiae* total RNA. This mixture was kept on ice for 10 minutes then loaded directly onto 3.5% acrylamide/agarose gels (3% acrylamide, 80:1 aa:bis, 0.5% agarose; 50 mM Tris, 50 mM glycine). Gels were prepared with a 10% acrylamide plug at the bottom to prevent the gel from sliding out from between the glass plates, and were run until xylene cyanole was approximately 2/3 of the way through the gel. Gels were run at room temperature at currents low enough to avoid heating, which would dissociate splicing complexes. After running, gels were dried and exposed directly to film.

DETERMINATION OF RELATIVE GROWTH RATES OF *S. CEREVISIAE* STRAINS

Growth rate data were collected using a Bioscreen C™ analyzer: cells were grown in minimal medium lacking histidine in order to ensure retention of U2 snRNA-encoding plasmids, and containing 0.1% NP-40 to prevent cell clumping. Relative growth rates were calculated from the relative gradients of the linear regions of graphs of ln(O.D.) vs time. Data were collected from three

independent cultures of each strain, allowing mean growth rate and standard deviations to be calculated.

BIBLIOGRAPHY

1. Staley, JP, Guthrie, C: Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 1998, 92:315–326

2. Burge, CB, Tuschl, TH, Sharp, PA: Splicing of precursors to mRNAs by the spliceosomes. 1999, The RNA World, Second Edition:525–560

3. Cramer, P, Srebrow, A, Kadener, S, Werbajh, S, de la Mata, M, Melen, G, Nogues, G, Kornblihtt, AR: Coordination between transcription and pre-mRNA processing. FEBS Lett 2001, 498:179-182

4. Hastings, ML, Krainer, AR: Pre-mRNA splicing in the new millennium. Curr Opin Cell Biol 2001, 13:302-309

5. Brow, DA: Allosteric cascade of spliceosome activation. Annu Rev Genet 2002, 36:333-360

6. Jurica, MS, Moore, MJ: Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 2003, 12:5–14

7. Nilsen, TW: The spliceosome: the most complex macromolecular machine in the cell? Bioessays 2003, 25:1147-1149

8. Stark, H, Luhrmann, R: Cryo-electron microscopy of spliceosomal components. Annu Rev Biophys Biomol Struct 2006, 35:435-457

9. Valadkhan, S: The spliceosome: caught in a web of shifting interactions. Curr Opin Struct Biol 2007, 17:310-315

10. Hertel, KJ: Combinatorial control of exon recognition. J Biol Chem 2008, 283:1211-1215

11. House, AE, Lynch, KW: Regulation of alternative splicing: more than just the ABCs. J Biol Chem 2008, 283:1217-1221

12. Makarov, EM, Makarova, OV, Urlaub, H, Gentzel, M, Will, CL, Wilm, M, Luhrmann, R: Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science 2002, 298:2205-2208

13. Stevens, SW, Ryan, DE, Ge, HY, Moore, RE, Young, MK, Lee, TD, Abelson, J: Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell 2002, 9:31-44

14. Behzadnia, N, Hartmuth, K, Will, CL, Luhrmann, R: Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. RNA 2006, 12:1738-1746

15. Rino, J, Carvalho, T, Braga, J, Desterro, JM, Luhrmann, R, Carmo-Fonseca, M: A stochastic view of spliceosome assembly and recycling in the nucleus. PLoS Comput Biol 2007, 3:2019-2031

16. Sterner, DA, Carlo, T, Berget, SM: Architectural limits on split genes. Proc Natl Acad Sci U S A 1996, 93:15081-15085

17. Fox-Walsh, KL, Dou, Y, Lam, BJ, Hung, SP, Baldi, PF, Hertel, KJ: The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A 2005, 102:16176-16181

18. House, AE, Lynch, KW: An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. Nat Struct Mol Biol 2006, 13:937-944

19. Hui, J, Hung, LH, Heiner, M, Schreiner, S, Neumuller, N, Reither, G, Haas, SA, Bindereif, A: Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. Embo J 2005, 24:1988-1998

20. Ibrahim el, C, Schaal, TD, Hertel, KJ, Reed, R, Maniatis, T: Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proc Natl Acad Sci U S A 2005, 102:5002-5007

21. Bailleul, B: During in vivo maturation of eukaryotic nuclear mRNA, splicing yields excised exon circles. Nucleic Acids Res 1996, 24:1015-1019

22. Braun, S, Domdey, H, Wiebauer, K: Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract. Nucleic Acids Res 1996, 24:4152-4157

23. Schindewolf, C, Braun, S, Domdey, H: In vitro generation of a circular exon from a linear pre-mRNA transcript. Nucleic Acids Res 1996, 24:1260-1266

24. Soares, LM, Zanier, K, Mackereth, C, Sattler, M, Valcarcel, J: Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. Science 2006, 312:1961-1965

25. Sharma, S, Kohlstaedt, LA, Damianov, A, Rio, DC, Black, DL: Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat Struct Mol Biol 2008,

26. Jensen, KB, Musunuru, K, Lewis, HA, Burley, SK, Darnell, RB: The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. Proc Natl Acad Sci U S A 2000, 97:5740-5745

27. Hiller, M, Zhang, Z, Backofen, R, Stamm, S: Pre-mRNA Secondary Structures Influence Exon Recognition. PLoS Genet 2007, 3:e204

28. Liu, HX, Goodall, GJ, Kole, R, Filipowicz, W: Effects of secondary structure on pre-mRNA splicing: hairpins sequestering the 5' but not the 3' splice site inhibit intron processing in Nicotiana plumbaginifolia. Embo J 1995, 14:377-388

29. Cheah, MT, Wachter, A, Sudarsan, N, Breaker, RR: Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature 2007, 447:497-500

30. Charpentier, B, Rosbash, M: Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly. Rna 1996, 2:509-522

31. Kreahling, JM, Graveley, BR: The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the Drosophila Dscam pre-mRNA. Mol Cell Biol 2005, 25:10251-10260

32. Graveley, BR: Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell 2005, 123:65-73

33. Olson, S, Blanchette, M, Park, J, Savva, Y, Yeo, GW, Yeakley, JM, Rio, DC, Graveley, BR: A regulator of Dscam mutually exclusive splicing fidelity. Nat Struct Mol Biol 2007,

34. Hertel, KJ, Maniatis, T: Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. Proc Natl Acad Sci U S A 1999, 96:2651-2655

35. Shen, H, Kan, JL, Green, MR: Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell 2004, 13:367-376

36. Xu, YZ, Query, CC: Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. Mol Cell 2007, 28:838-849

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

38. Shen, H, Green, MR: RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. Genes Dev 2006, 20:1755-1765

39. Shen, H, Green, MR: RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. Nat Struct Mol Biol 2007, 14:597-603

40. Schwer, B, Guthrie, C: A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. EMBO J 1992, 11:5033–5039

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

42. Crotti, LB, Bacikova, D, Horowitz, DS: The Prp18 protein stabilizes the interaction of both exons with the U5 snRNA during the second step of pre-mRNA splicing. Genes Dev 2007, 21:1204-1216

43. Sawa, H, Abelson, J: Evidence for a base-pairing interaction between U6 small nuclear RNA and 5' splice site during the splicing reaction in yeast. Proc Natl Acad Sci USA 1992, 89:11269–11273

44. Konarska, MM, Vilardell, J, Query, CC: Repositioning of the reaction intermediate within the catalytic center of the spliceosome. Mol Cell 2006, 21:543-553

45. Villa, T, Guthrie, C: The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. Genes Dev 2005, 19:1894–1904

46. Query, CC, Konarska, MM: Suppression of multiple substrate mutations by spliceosomal *prp8* alleles suggests functional correlations with ribosomal ambiguity mutants. Mol Cell 2004, 14:343–354

47. Liu, L, Query, CC, Konarska, MM: Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. Nat Struct Mol Biol 2007, 14:519-526

48. Hilliker, AK, Mefford, MA, Staley, JP: U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. Genes Dev 2007, 21:821-834

49. Perriman, RJ, Ares, M, Jr.: Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. Genes Dev 2007, 21:811-820

50. Grainger, RJ, Beggs, JD: Prp8 protein: at the heart of the spliceosome. Rna 2005, 11:533-557

51. Burgess, SM, Guthrie, C: A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell 1993, 73:1377-1391

52. Schneider, S, Campodonico, E, Schwer, B: Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-

defective Prp22 mutants are suppressed by Prp8. J Biol Chem 2004, 279:8617-8626

53. Johnson, TL, Abelson, J: Characterization of U4 and U6 interactions with the 5' splice site using a S. cerevisiae in vitro trans-splicing system. Genes Dev 2001, 15:1957–1970

54. Chan, SP, Kao, DI, Tsai, WY, Cheng, SC: The Prp19p-associated complex in spliceosome activation. Science 2003, 302:279–282

55. Raghunathan, PL, Guthrie, C: RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr Biol 1998, 8:847-855

56. Small, EC, Leggett, SR, Winans, AA, Staley, JP: The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. Mol Cell 2006, 23:389-399

57. Helmers, J, Schmidt, D, Glavy, JS, Blobel, G, Schwartz, T: The beta-subunit of the protein-conducting channel of the endoplasmic reticulum functions as the guanine nucleotide exchange factor for the beta-subunit of the signal recognition particle receptor. J Biol Chem 2003, 278:23686-23690

58. Augustin, S, Muller, MW, Schweyen, RJ: Reverse self-splicing of group II intron RNAs in vitro. Nature 1990, 343:383-386

59. Tseng, CK, Cheng, SC: Both Catalytic Steps of Nuclear Pre-mRNA Splicing Are Reversible. Science 2008, 320:1782-1784

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

61. Madhani, HD, Guthrie, C: A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell 1992, 71:803-817

62. Hilliker, AK, Staley, JP: Multiple functions for the invariant AGC triad of U6 snRNA. Rna 2004, 10:921-928

63. Pyle, AM: Group II Introns: Catalysts for Splicing, Genomic Change and Evolution. London, Royal Society of Chemistry, 2008.

64. Yean, SL, Wuenschell, G, Termini, J, Lin, RJ: Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature 2000, 408:881-884

65. McManus, CJ, Schwartz, ML, Butcher, SE, Brow, DA: A dynamic bulge in the U6 RNA internal stem-loop functions in spliceosome assembly and activation. Rna 2007, 13:2252-2265

66. Ryan, DE, Kim, CH, Murray, JB, Adams, CJ, Stockley, PG, Abelson, J: New tertiary constraints between the RNA components of active yeast spliceosomes: a photo-crosslinking study. RNA 2004, 10:1251–1265

67. Rhode, BM, Hartmuth, K, Westhof, E, Luhrmann, R: Proximity of conserved U6 and U2 snRNA elements to the 5' splice site region in activated spliceosomes. Embo J 2006, 25:2475-2486

68. Shi, Y, Reddy, B, Manley, JL: PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. Mol Cell 2006, 23:819-829

69. Bellare, P, Small, EC, Huang, X, Wohlschlegel, JA, Staley, JP, Sontheimer, EJ: A role for ubiquitin in the spliceosome assembly pathway. Nat Struct Mol Biol 2008, 15:444-451

70. Burgess, SM, Guthrie, C: Beat the clock: paradigms for NTPases in the maintenance of biological fidelity. Trends Biochem Sci 1993, 18:381–384

71. Mayas, RM, Maita, H, Staley, JP: Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol 2006, 13:482–490

72. Tennyson, CN, Klamut, HJ, Worton, RG: The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat Genet 1995, 9:184-190

73. Tanaka, N, Aronova, A, Schwer, B: Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. Genes Dev 2007, 21:2312-2325

74. Maeder, C, Kutach, AK, Guthrie, C: ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. Nat Struct Mol Biol 2009, 16:42-48

75. Schwer, B, Gross, CH: Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J 1998, 17:2086-2094

76. Brys, A, Schwer, B: Requirement for SLU7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and the 3' splice site. Rna 1996, 2:707-717

77. Box, JA, Bunch, JT, Tang, W, Baumann, P: Spliceosomal cleavage generates the 3' end of telomerase RNA. Nature 2008, 456:910-914

78. Pandit, S, Lynn, B, Rymond, BC: Inhibition of a spliceosome turnover pathway suppresses splicing defects. Proc Natl Acad Sci U S A 2006, 103:13700-13705

79. Hilleren, PJ, Parker, R: Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates. Mol Cell 2003, 12:1453–1465

80. Smith, CW, Chu, TT, Nadal-Ginard, B: Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Mol Cell Biol 1993, 13:4939-4952

81. Anderson, K, Moore, MJ: Bimolecular exon ligation by the human spliceosome. Science 1997, 276:1712–1716

82. Chua, K, Reed, R: An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. Mol Cell Biol 2001, 21:1509-1514

83. Sapra, AK, Arava, Y, Khandelia, P, Vijayraghavan, U: Genome-wide analysis of pre-mRNA splicing: intron features govern the requirement for the second-step factor, Prp17 in Saccharomyces cerevisiae and Schizosaccharomyces pombe. J Biol Chem 2004, 279:52437-52446

84. Penalva, LO, Lallena, MJ, Valcarcel, J: Switch in 3' splice site recognition between exon definition and splicing catalysis is important for sex-lethal autoregulation. Mol Cell Biol 2001, 21:1986-1996

85. Chiara, MD, Reed, R: A two-step mechanism for 5' and 3' splice-site pairing. Nature 1995, 375:510-513

86. Chua, K, Reed, R: Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. Genes Dev 1999, 13:841-850

87. O'Keefe, RT, Norman, C, Newman, AJ: The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. Cell 1996, 86:679-689

88. Konforti, BB, Konarska, MM: U4/U5/U6 snRNP recognizes the 5' splice site in the absence of U2 snRNP. Genes Dev 1994, 8:1962-1973

89. Valadkhan, S, Mohammadi, A, Wachtel, C, Manley, JL: Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. Rna 2007, 13:2300-2311

90. Bacikova, D, Horowitz, DS: Mutational analysis identifies two separable roles of the Saccharomyces cerevisiae splicing factor Prp18. Rna 2002, 8:1280-1293

91. Chua, K, Reed, R: The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. Nature 1999, 402:207-210

92. McPheeters, DS, Schwer, B, Muhlenkamp, P: Interaction of the yeast DExH-box RNA helicase prp22p with the 3' splice site during the second step of nuclear pre-mRNA splicing. Nucleic Acids Res 2000, 28:1313-1321

93. Wagner, JD, Jankowsky, E, Company, M, Pyle, AM, Abelson, JN: The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. Embo J 1998, 17:2926-2937

94. Segault, V, Will, CL, Polycarpou-Schwarz, M, Mattaj, IW, Branlant, C, Luhrmann, R: Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic steps of pre-mRNA splicing in HeLa nuclear extracts. Mol Cell Biol 1999, 19:2782-2790

95. Black, DL: Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 2003, 72:291–336

96. Park, JW, Parisky, K, Celotto, AM, Reenan, RA, Graveley, BR: Identification of alternative splicing regulators by RNA interference in Drosophila. Proc Natl Acad Sci U S A 2004, 101:15974-15979

97. Pleiss, JA, Whitworth, GB, Bergkessel, M, Guthrie, C: Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. PLoS Biol 2007, 5:e90

98. Juneau, K, Palm, C, Miranda, M, Davis, RW: High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. Proc Natl Acad Sci U S A 2007, 104:1522-1527

99. Balzer, RJ, Henry, MF: Snu56p is required for Mer1p-activated meiotic splicing. Mol Cell Biol 2008, 28:2497-2508

100. Kress, TL, Krogan, NJ, Guthrie, C: A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. Mol Cell 2008, 32:727-734

101. Moore, MJ, Silver, PA: Global analysis of mRNA splicing. Rna 2007,

102. Ban, N, Nissen, P, Hansen, J, Moore, PB, Steitz, TA: The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 2000, 289:905-920

103. Abelson, J: Is the spliceosome a ribonucleoprotein enzyme? Nat Struct Mol Biol 2008, 15:1235-1237

104. Collins, CA, Guthrie, C: The question remains: is the spliceosome a ribozyme? Nat Struct Biol 2000, 7:850–854

105. Tuschl, T, Sharp, PA, Bartel, DP: Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library. Embo J 1998, 17:2637-2650

106. Tuschl, T, Sharp, PA, Bartel, DP: A ribozyme selected from variants of U6 snRNA promotes 2',5'-branch formation. RNA 2001, 7:29-43

107. Valadkhan, S, Manley, JL: Splicing-related catalysis by protein-free snRNAs. Nature 2001, 413:701-707

108. Valadkhan, S, Manley, JL: Characterization of the catalytic activity of U2 and U6 snRNAs. RNA 2003, 9:892-904

109. Husken, D, Goodall, G, Blommers, MJ, Jahnke, W, Hall, J, Haner, R, Moser, HE: Creating RNA bulges: cleavage of RNA in RNA/DNA duplexes by metal ion catalysis. Biochemistry 1996, 35:16591-16600

110. Bartel, DP, Szostak, JW: Isolation of new ribozymes from a large pool of random sequences [see comment]. Science 1993, 261:1411-1418

111. Kandels-Lewis, S, Seraphin, B: Involvement of U6 snRNA in 5' splice site selection. Science 1993, 262:2035–2039

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

113. Padgett, RA, Konarska, MM, Grabowski, PJ, Hardy, SF, Sharp, PA: Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science 1984, 225:898-903

114. Steitz, TA, Steitz, JA: A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci USA 1993, 90:6498–6502

115. Szostak, JW, Bartel, DP, Luisi, PL: Synthesizing life. Nature 2001, 409:387-390

116. Graveley, BR: Molecular biology: power sequencing. Nature 2008, 453:1197-1198

117. Lorsch, JR, Szostak, JW: Chance and necessity in the selection of nucleic acid catalysts. Acc Chem Res 1996, 29:103-110

118. Toor, N, Keating, KS, Taylor, SD, Pyle, AM: Crystal structure of a self-spliced group II intron. Science 2008, 320:77-82

119. Gordon, PM, Sontheimer, EJ, Piccirilli, JA: Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. Rna 2000, 6:199-205

120. Sontheimer, EJ, Gordon, PM, Piccirilli, JA: Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. Genes Dev 1999, 13:1729-1741

121. Pena, V, Rozov, A, Fabrizio, P, Luhrmann, R, Wahl, MC: Structure and function of an RNase H domain at the heart of the spliceosome. EMBO J 2008, 27:2929-2940

122. Ritchie, DB, Schellenberg, MJ, Gesner, EM, Raithatha, SA, Stuart, DT, Macmillan, AM: Structural elucidation of a PRP8 core domain from the heart of the spliceosome. Nat Struct Mol Biol 2008, 15:1199-1205

123. Yang, K, Zhang, L, Xu, T, Heroux, A, Zhao, R: Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. Proc Natl Acad Sci U S A 2008, 105:13817-13822

124. Parker, R, Siliciano, PG, Guthrie, C: Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell 1987, 49:229-239

125. Wu, J, Manley, JL: Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev 1989, 3:1553-1561

126. Zhuang, Y, Weiner, AM: A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev 1989, 3:1545-1552

127. Berglund, JA, Chua, K, Abovich, N, Reed, R, Rosbash, M: The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell 1997, 89:781-787

128. Rutz, B, Seraphin, B: Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. RNA 1999, 5:819-831

129. Tani, T, Ohshima, Y: mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in pre-mRNA splicing. Genes Dev 1991, 5:1022-1031

130. Yu, YT, Maroney, PA, Nilsen, TW: Functional reconstitution of U6 snRNA in nematode cis- and trans-splicing: U6 can serve as both a branch acceptor and a 5' exon. Cell 1993, 75:1049-1059

131. Takahashi, Y, Urushiyama, S, Tani, T, Ohshima, Y: An mRNA-type intron is present in the Rhodotorula hasegawae U2 small nuclear RNA gene. Mol Cell Biol 1993, 13:5613-5619

132. Lesser, CF, Guthrie, C: Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, CUP1. Genetics 1993, 133:851–863

133. McPheeters, DS, Fabrizio, P, Abelson, J: In vitro reconstitution of functional yeast U2 snRNPs. Genes Dev 1989, 3:2124-2136

134. Vijayraghavan, U, Parker, R, Tamm, J, Iimura, Y, Rossi, J, Abelson, J, Guthrie, C: Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. EMBO J 1986, 5:1683–1695

135. Seraphin, B, Rosbash, M: Mutational analysis of the interactions between U1 small nuclear RNA and pre-mRNA of yeast. Gene 1989, 82:145-151

136. Podar, M, Chu, VT, Pyle, AM, Perlman, PS: Group II intron splicing in vivo by first-step hydrolysis. Nature 1998, 391:915-918

137. Query, CC, Moore, MJ, Sharp, PA: Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev 1994, 8:587–597

138. Newby, MI, Greenbaum, NL: Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. Nat Struct Biol 2002, 9:958–965

139. Yang, C, McPheeters, DS, Yu, YT: Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in Saccharomyces cerevisiae. J Biol Chem 2005, 280:6655-6662

140. Chu, VT, Liu, Q, Podar, M, Perlman, PS, Pyle, AM: More than one way to splice an RNA: branching without a bulge and splicing without branching in group II introns. Rna 1998, 4:1186-1202

141. Jacquier, A, Rosbash, M: RNA splicing and intron turnover are greatly diminished by a mutant yeast branch point. Proc Natl Acad Sci U S A 1986, 83:5835-5839

142. Yan, D, Ares, MJ: Invariant U2 RNA sequences bordering the branchpoint recognition region are essential for interaction with yeast SF3a and SF3b subunits. Mol Cell Biol 1996, 16:818-828

143. Berglund, JA, Rosbash, M, Schultz, SC: Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. RNA 2001, 7:682-691

144. Hornig, H, Aebi, M, Weissmann, C: Effect of mutations at the lariat branch acceptor site on beta-globin pre-mRNA splicing in vitro. Nature 1986, 324:589-591

145. Query, CC, Strobel, SA, Sharp, PA: Three recognition events at the branch-site adenine. EMBO J 1996, 15:1392-1402

146. Liu, Q, Green, JB, Khodadadi, A, Haeberli, P, Beigelman, L, Pyle, AM: Branch-site selection in a group II intron mediated by active recognition of the adenine amino group and steric exclusion of non-adenine functionalities. J Mol Biol 1997, 267:163-171

147. Zelin, E, Wang, Y, Silverman, SK: Adenosine is inherently favored as the branch-site RNA nucleotide in a structural context that resembles natural RNA splicing. Biochemistry 2006, 45:2767-2771

148. McConnell, TS, Cho, SJ, Frilander, MJ, Steitz, JA: Branchpoint selection in the splicing of U12-dependent introns in vitro. RNA 2002, 8:579-586

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

150. Rackham, O, Chin, JW: A network of orthogonal ribosome x mRNA pairs. Nat Chem Biol 2005, 1:159-166

151. Chu, VT, Adamidi, C, Liu, Q, Perlman, PS, Pyle, AM: Control of branch-site choice by a group II intron. EMBO J 2001, 20:6866-6876

152. Nielsen, H, Westhof, E, Johansen, S: An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. Science 2005, 309:1584-1587

153. Wang, Y, Silverman, SK: Deoxyribozymes that synthesize branched and lariat RNA. J Am Chem Soc 2003, 125:6880-6881

154. Noller, HF: RNA structure: reading the ribosome. Science 2005, 309:1508-1514

155. Nissen, P, Hansen, J, Ban, N, Moore, PB, Steitz, TA: The structural basis of ribosome activity in peptide bond synthesis. Science 2000, 289:920-930

156. Newby, MI, Greenbaum, NL: A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. RNA 2001, 7:833-845

157. Borer, PN, Lin, Y, Wang, S, Roggenbuck, MW, Gott, JM, Uhlenbeck, OC, Pelczer, I: Proton NMR and structural features of a 24-nucleotide RNA hairpin. Biochemistry 1995, 34:6488-6503

158. Schmitz, M: Change of RNase P RNA function by single base mutation correlates with perturbation of metal ion binding in P4 as determined by NMR spectroscopy. Nucleic Acids Res 2004, 32:6358-6366

159. Popenda, L, Bielecki, L, Gdaniec, Z, Adamiak, R: Structure and dynamics of adenosine bulged RNA duplex reveals formation of the dinucleotide platform in the C:G-A triple. ARKIVOC 2008, 2009:130-144

160. Acharya, S, Foldesi, A, Chattopadhyaya, J: The pK(a) of the internucleotidic 2'-hydroxyl group in diribonucleoside (3'-->5') monophosphates. J Org Chem 2003, 68:1906-1910

161. Lohrmann, R, Orgel, LE: Studies of oligoadenylate formation on a poly (U) template. J Mol Evol 1979, 12:237-257

162. Sawai, H, Totsuka, S, Yamamoto, K, Ozaki, H: Non-enzymatic, template-directed ligation of 2'-5' oligoribonucleotides. Joining of a template and a ligator strand. Nucleic Acids Res 1998, 26:2995-3000

163. Lorsch, JR, Bartel, DP, Szostak, JW: Reverse transcriptase reads through a 2'-5'linkage and a 2'-thiophosphate in a template. Nucleic Acids Res 1995, 23:2811-2814

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

165. Reilly, JD, Freeman, SK, Melhem, RF, Kierzek, R, Caruthers, MH, Edmonds, M, Munns, TW: Antibodies specific for branched ribonucleic acids. Anal Biochem 1990, 185:125-130

166. Newman, AJ, Teigelkamp, S, Beggs, JD: snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. RNA 1995, 1:968–980

167. McGrail, JC, Tatum, EM, O'Keefe, RT: Mutation in the U2 snRNA influences exon interactions of U5 snRNA loop 1 during pre-mRNA splicing. EMBO J 2006, 25:3813-3822

168. McPheeters, DS, Abelson, J: Mutational analysis of the yeast U2 snRNA suggests a structural similarity to the catalytic core of group I introns. Cell 1992, 71:819-831

169. Datta, B, Weiner, AM: Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing. Nature 1991, 352:821-824

170. Shukla, GC, Padgett, RA: A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. Mol Cell 2002, 9:1145-1150

171. Laggerbauer, B, Achsel, T, Luhrmann, R: The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. Proc Natl Acad Sci U S A 1998, 95:4188-4192

172. Wang, Y, Wagner, JD, Guthrie, C: The DEAH-box splicing factor Prp16 unwinds RNA duplexes in vitro. Curr Biol 1998, 8:441–451

173. Schwer, B: A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol Cell 2008, 30:743-754

174. Aronova, A, Bacikova, D, Crotti, LB, Horowitz, DS, Schwer, B: Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing. Rna 2007, 13:1437-1444

175. Mohr, S, Matsuura, M, Perlman, PS, Lambowitz, AM: A DEAD-box protein alone promotes group II intron splicing and reverse splicing by acting as an RNA chaperone. Proc Natl Acad Sci U S A 2006, 103:3569-3574

176. Del Campo, M, Tijerina, P, Bhaskaran, H, Mohr, S, Yang, Q, Jankowsky, E, Russell, R, Lambowitz, AM: Do DEAD-box proteins promote group II intron splicing without unwinding RNA? Mol Cell 2007, 28:159-166

177. Tran, EJ, Zhou, Y, Corbett, AH, Wente, SR: The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. Mol Cell 2007, 28:850-859

178. Rodnina, MV, Wintermeyer, W: Ribosome fidelity: tRNA discrimination, proofreading and induced fit. Trends Biochem Sci 2001, 26:124-130

179. Fersht, A: Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman, 1998.

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

181. Inoue, T, Sullivan, FX, Cech, TR: New reactions of the ribosomal RNA precursor of Tetrahymena and the mechanism of self-splicing. J Mol Biol 1986, 189:143-165

182. Jarrell, KA, Peebles, CL, Dietrich, RC, Romiti, SL, Perlman, PS: Group II intron self-splicing. Alternative reaction conditions yield novel products. J Biol Chem 1988, 263:3432-3439

183. Moore, MJ, Sharp, PA: Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature 1993, 365:364–368