

# Towards understanding the catalytic core structure of the spliceosome

S.E. Butcher<sup>1</sup> and D.A. Brow

Departments of Biochemistry and Biomolecular Chemistry, University of Wisconsin–Madison, Madison, WI 53706, U.S.A.

## Abstract

The spliceosome catalyses the splicing of nuclear pre-mRNA (precursor mRNA) in eukaryotes. Pre-mRNA splicing is essential to remove internal non-coding regions of pre-mRNA (introns) and to join the remaining segments (exons) into mRNA before translation. The spliceosome is a complex assembly of five RNAs (U1, U2, U4, U5 and U6) and many dozens of associated proteins. Although a high-resolution structure of the spliceosome is not yet available, inroads have been made towards understanding its structure and function. There is growing evidence suggesting that U2 and U6 RNAs, of the five, may contribute to the catalysis of pre-mRNA splicing. In this review, recent progress towards understanding the structure and function of U2 and U6 RNAs is summarized.

## Components of the catalytic core

Most of the eukaryotic protein-coding genes have introns. Before the translation of the RNA transcript, these introns must be accurately removed and the exons spliced together. This process proceeds through two steps: (i) attack of the 2'-hydroxyl of an intronic branch point adenosine at the 5'-splice site, resulting in formation of an intermediate 2'-5' branched lariat intron-3'-exon, (ii) attack of the 3'-hydroxyl of the 5'-exon at the 3'-splice site, producing the ligated exons and liberating the 2'-5' branched lariat intron. The macromolecular machine that catalyses these steps is the spliceosome, a mega-Dalton assembly of RNAs and proteins. The spliceosome undergoes a cascade of assembly events and dynamic rearrangements before forming an active complex on pre-mRNA (precursor-mRNA) [1], during which time the U1 and U4 RNAs are either released or destabilized to the extent that they are only weakly associated with the complex [1]. The U5 RNA, which positions the exons for ligation [2], is probably not involved directly in catalysis, because much of its sequence is dispensable *in vitro* [3]. Therefore, despite the complexity of the spliceosome, which can involve more than a hundred different proteins [4], there are few components known to interact directly with the pre-mRNA substrate that are ideal candidates for catalysing the splicing reaction. These include the Prp8 protein that has no recognizable structural motifs within its sequence and is remarkably large (2400 amino acids) and highly conserved (62% identity between yeast and humans) [2] and the base-paired complex of U2 and U6 RNAs that directly base-pair to the intron [3]. This review will focus on the structure and function of the U2 and U6 RNAs of the spliceosomal catalytic core [3].

## RNA and metal-ion requirements

By directly contacting the pre-mRNA, the U2 and U6 RNAs are certain to reside near the active site of the spliceosome. However, are the RNAs themselves responsible for chemistry? In other words, "is the spliceosome a ribozyme?" [5]. Intriguing parallels exist between the spliceosome and group II self-splicing introns, which are true ribozymes. Valadkhan and Manley [6] investigated the catalytic potential of a protein-free preparation of U2–U6 RNA and found that indeed the human U2–U6 RNA complex is capable of stimulating a slow, inefficient reaction that mimics the first step of splicing [6]. The protein-free reaction is further stimulated by a conserved pseudouridine that serves to remodel the structure of the branch site helix [7,8]. The slow protein-free reaction is reminiscent of the peptidyl transfer reaction rate observed with deproteinized ribosomes [9]. It was not until the ribosome crystal structure was solved, however, that it was realized that ribosomes utilize an RNA core to catalyse the chemical steps of translation [10].

RNA is also likely to dominate the catalytic core of the spliceosome, and the similarities between the group II ribozymes and the spliceosome reinforce this idea. Mechanistic analyses indicate that both group II introns and the spliceosome employ identical reaction pathways and stereochemistry [11]. Additionally, both require metal ion as a cofactor and utilize the same catalytic strategies, in which magnesium ion co-ordinates with the 3'-oxyanion leaving groups to stabilize the build-up of negative charge in the transition state for both steps of splicing [12,13]. These mechanistic similarities have led to the hypothesis that the spliceosome and group II introns evolved from a common molecular ancestor [12].

Lin and co-workers [14] have elucidated an essential metal-binding site within the U6 RNA at the U80 pro-S phosphate oxygen of a highly conserved ISL (intramolecular

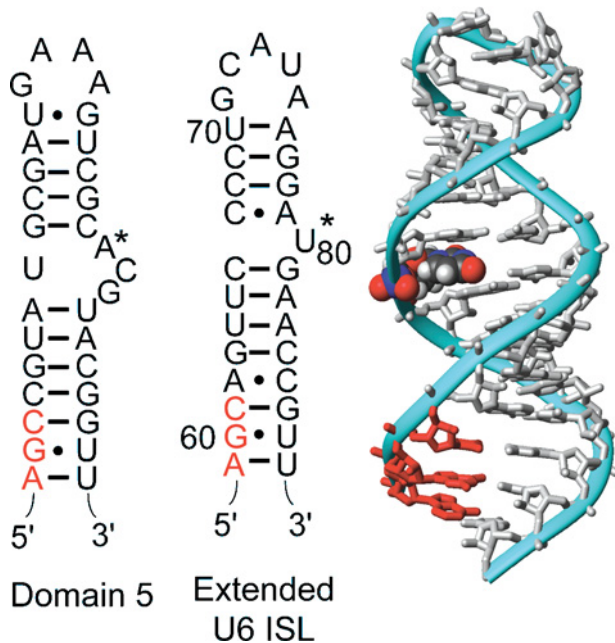
**Key words:** catalytic core, NMR, structure, spliceosome, U2 RNA, U6 RNA.

**Abbreviations used:** ISL, intramolecular stem-loop; pre-mRNA, precursor mRNA.

<sup>1</sup>To whom correspondence should be made (email [butcher@nmrfam.wisc.edu](mailto:butcher@nmrfam.wisc.edu)).

### Figure 1 | Similarities between domain 5 of group II self-splicing introns and the U6 ISL

Metal-binding sites are indicated with an asterisk. The AGC triad is shown in red for both sequences and also in the U6 ISL NMR structure (Protein Data Bank code 1XHP). The U80 nucleotide is represented as a space-filling model.



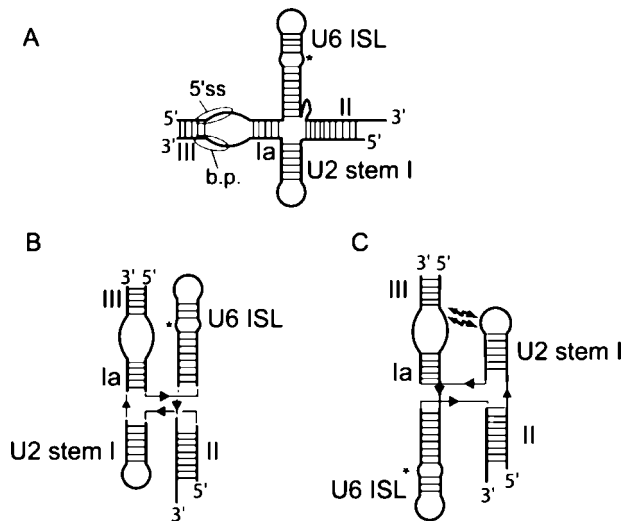
stem-loop) structure (Figure 1). It is not known if this metal-ion-binding site also helps to co-ordinate the same metal ion that is involved in stabilizing the 3'-oxyanion. However, sulphur substitution of the U80 pro-S phosphate oxygen atom is sufficient to alter splicing chemistry to such an extent that it becomes entirely dependent on the addition of metal ions that can co-ordinate with sulphur, such as cadmium [14]. NMR experiments indicate that metal ions readily co-ordinate with the U80 pro-S phosphate oxygen in the isolated U6 ISL domain [15]. Furthermore, the structure of the U6 ISL is not altered when the U80 pro-S phosphate oxygen is substituted with sulphur [16].

### Structural analysis of U2-U6 RNA

We have determined the NMR structure of the U6 ISL (Figure 1) [15,16]. Interesting features of this RNA include a pentaloop that makes a GNRA-type fold and these folds often mediate RNA tertiary interactions [17] or may function as protein-recognition sites [18]. Adjacent to the metal-binding site at U80 is a conserved, protonated C67 + A79 wobble pair with a  $pK_a$  near neutrality [15]. The unprotonated state of A79 favours metal ion binding, and metal ion binding, in turn, lowers the  $pK_a$  [15]. This observation raises the interesting possibility that proton uptake could regulate splicing by influencing the binding of a required metal ion. Additionally, proton uptake results in a significant conformational change in the U6 ISL. At higher

### Figure 2 | Secondary-structure representations of possible folds for the U2-U6 four-helix junction

(A) Secondary structure of the U2-U6 complex as determined by NMR. U2-U6 helices Ia, II and III are shown, as is the U6 sequence that pairs with the 5'-splice site (5'-ss) and the U2 sequence that pairs with the branch point (b.p.) of the intron. (B, C) Possible patterns of coaxial stacking for the four observed helices. Sites of UV-induced cross-linking are shown with jagged arrows. The metal-binding site is marked with an asterisk in all Figure parts.



pH values, U80 stacks above the unprotonated A79 base. At lower pH values, U80 is flipped out of the helix and the protonated A79 base stacks upon G81 [19]. These two conformational states exist in equilibrium and inter-convert on the micro- to millisecond timescale [19].

An atomic-level model of the U2-U6 RNA structure is required to understand how it could participate in the splicing reaction. We therefore used NMR to analyse the hydrogen-bonding patterns for a number of protein-free U2-U6 RNA complexes of up to 110 nt total length [20]. All the complexes studied formed a four-helix junction (Figure 2A). The observed four-helix junction forms an extended U6 ISL structure [20] (Figure 1). This was unexpected, since the extended U6 ISL sequesters a highly conserved AGC sequence that has been shown to participate in the formation of an intermolecular U2-U6 helix, helix 1b [21,22]. Helix 1b is essential for splicing; however, mutagenesis experiments suggest an additional role for the AGC triad beyond helix 1b formation [21,22]. Mechanistic studies suggest that a conformational change in the spliceosome may be rate-limiting for the second catalytic step of splicing [12]. We hypothesize that the four-helix junction may play a role in the first but not second step of splicing and that proteins will be required for helix 1b formation and remodelling of the intrinsic U2-U6 structure [20].

There are several interesting implications of the observed four-way junction fold. First, the extended U6 ISL in this structure closely resembles domain 5, the catalytic core of

group II self-splicing introns, particularly with respect to the conserved AGC sequence (Figure 1). This observation lends further support to the hypothesis that the spliceosome and group II introns share a common molecular ancestor. Secondly, four-way junctions form coaxial helical stacks that could juxtapose catalytically essential elements, as observed in the hairpin ribozyme [23]. We utilized the crystal structure of a hairpin ribozyme four-way junction [24] to model such an interaction [20], and the resulting model was satisfying in that it predicted a close proximity between the U6 ISL metal-binding site and the intron-binding regions of U2 and U6 (Figure 2B). However, UV cross-linking studies [25,26] report tertiary interactions between U2 and U6 that are consistent with a previously identified genetic interaction [27] but inconsistent with a model that juxtaposes the U6 ISL and the intron-binding region of the U2–U6 complex. A different pattern of coaxial stacking probably explains these results (Figure 2C). It is possible that these two coaxial stacking patterns are formed at different points during the splicing reaction, and the switch between them is one of the conformational changes required for spliceosome activation.

## References

- 1 Brow, D.A. (2002) *Annu. Rev. Genet.* **36**, 333–360
- 2 Turner, I.A., Norman, C.M., Churcher, M.J. and Newman, A.J. (2004) *Biochem. Soc. Trans.* **32**, 928–931
- 3 Villa, T., Pleiss, J.A. and Guthrie, C. (2002) *Cell (Cambridge, Mass.)* **109**, 149–152
- 4 Jurica, M.S. and Moore, M.J. (2003) *Mol. Cell* **12**, 5–14
- 5 Collins, C.A. and Guthrie, C. (2000) *Nat. Struct. Biol.* **7**, 850–854
- 6 Valadkhan, S. and Manley, J.L. (2001) *Nature (London)* **413**, 701–707
- 7 Valadkhan, S. and Manley, J.L. (2003) *RNA* **9**, 892–904
- 8 Newby, M.I. and Greenbaum, N.L. (2002) *Nat. Struct. Biol.* **9**, 958–965
- 9 Noller, H.F., Hoffarth, V. and Zimniak, L. (1992) *Science* **256**, 1416–1419
- 10 Cech, T.R. (2000) *Science* **289**, 878–879
- 11 Moore, M.J. and Sharp, P.A. (1993) *Nature (London)* **365**, 364–368
- 12 Gordon, P.M., Sontheimer, E.J. and Piccirilli, J.A. (2000) *RNA* **6**, 199–205
- 13 Sontheimer, E.J. (2001) *Nat. Struct. Biol.* **8**, 11–13
- 14 Yean, S.L., Wuenschell, G., Termini, J. and Lin, R.J. (2000) *Nature (London)* **408**, 881–884
- 15 Huppler, A., Nikstad, L.J., Allmann, A.M., Brow, D.A. and Butcher, S.E. (2002) *Nat. Struct. Biol.* **9**, 431–435
- 16 Reiter, N.J., Nikstad, L.J., Allmann, A.M., Johnson, R.J. and Butcher, S.E. (2003) *RNA* **9**, 533–542
- 17 Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) *Science* **273**, 1696–1699
- 18 Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. (1998) *Cell (Cambridge, Mass.)* **93**, 289–299
- 19 Reiter, N.J., Blad, H., Abildgaard, F. and Butcher, S.E. (2004) *Biochemistry* **43**, 13739–13747
- 20 Sashital, D.G., Cornilescu, G., McManus, C.J., Brow, D.A. and Butcher, S.E. (2004) *Nat. Struct. Mol. Biol.* **11**, 1237–1242
- 21 Hilliker, A.K. and Staley, J.P. (2004) *RNA* **10**, 921–928
- 22 Madhani, H.D. and Guthrie, C. (1992) *Cell (Cambridge, Mass.)* **71**, 803–817
- 23 McKinney, S.A., Tan, E., Wilson, T.J., Nahas, M.K., Declais, A.C., Clegg, R.M., Lilley, D.M. and Ha, T. (2004) *Biochem. Soc. Trans.* **32**, 41–45
- 24 Rupert, P.B. and Ferre-D'Amare, A.R. (2001) *Nature (London)* **410**, 780–786
- 25 Ryan, D.E., Kim, C.H., Murray, J.B., Adams, C.J., Stockley, P.G. and Abelson, J. (2004) *RNA* **10**, 1251–1265
- 26 Valadkhan, S. and Manley, J.L. (2000) *RNA* **6**, 206–219
- 27 Madhani, H.D. and Guthrie, C. (1994) *Genes Dev.* **8**, 1071–1086

Received 13 January 2005