The EMBO Journal (2013) 32, 2804–2818
www.embojournal.org

SOURCE DATA | TRANSPARENT PROCESS

THE EMBO JOURNAL

RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core

Maria Anokhina[1], Sergey Bessonov[1], Zhichao Miao[2], Eric Westhof[2], Klaus Hartmuth[1,*] and Reinhard Lührmann[1,*]

[1]Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany and [2]Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire du CNRS, Université de Strasbourg, Strasbourg, France

Although U snRNAs play essential roles in splicing, little is known about the 3D arrangement of U2, U6, and U5 snRNAs and the pre-mRNA in active spliceosomes. To elucidate their relative spatial organization and dynamic rearrangement, we examined the RNA structure of affinity-purified, human spliceosomes before and after catalytic step 1 by chemical RNA structure probing. We found a stable 3-way junction of the U2/U6 snRNA duplex in active spliceosomes that persists minimally through step 1. Moreover, the formation of alternating, mutually exclusive, U2 snRNA conformations, as observed in yeast, was not detected in different assembly stages of human spliceosomal complexes (that is, B, B^{act}, or C complexes). Psoralen crosslinking revealed an interaction during/after step 1 between internal loop 1 of the U5 snRNA, and intron nucleotides immediately downstream of the branchpoint. Using the experimentally derived structural constraints, we generated a model of the RNA network of the step 1 spliceosome, based on the crystal structure of a group II intron through homology modelling. The model is topologically consistent with current genetic, biochemical, and structural data.

The EMBO Journal (2013) **32,** 2804–2818. doi:10.1038/emboj.2013.198; Published online 3 September 2013
Subject Categories: RNA
Keywords: RNA modelling; RNA structure; spliceosome; U2–U6 duplex; 3-way junction

Introduction

Nuclear pre-mRNA splicing is catalysed by the spliceosome, which assembles *de novo* onto each intron from preformed subcomplexes and single factors (Wahl *et al*, 2009). Central to assembly and catalysis are the U1, U2, U4/U6, and U5 snRNPs, with their respective U snRNA components (denoted as U1, U2, U4, U5, and U6). The snRNAs recognize short, degenerate sequences in the intron that define the 5′ splice site (ss), the so-called branchpoint (bp), and the 3′ ss, which is typically preceded by a pyrimidine-rich track (polypyrimidine tract: PPY). DExD/H-type RNA helicases, as well as a GTPase (Snu114), sequentially rearrange the RNAs (Staley and Guthrie, 1998) to create a catalytic centre and facilitate the two successive transesterification steps of splicing. Initially, the 2′ OH of the branchpoint adenosine attacks the 5′ phosphate of the 5′ ss, resulting in cleavage of the pre-mRNA at this site and the concomitant ligation of the 5′ end of the intron to the branchpoint adenosine, generating a lariat-like structure (step 1). The liberated 3′ OH of the 5′ exon then attacks the 3′ phosphate at the 3′ ss, leading to excision of the intron lariat and the ligation of the 5′ and 3′ exons (step 2).

Spliceosome assembly (Wahl *et al*, 2009) starts with an interaction between U1 snRNP and the 5′ ss to form the E complex (Michaud and Reed, 1993; Das *et al*, 2000). Then, U2 snRNP interacts with the branchpoint, forming the pre-spliceosomal A complex. Next, the U4/U6.U5 tri-snRNP joins to form the pre-catalytic B complex. In this complex, the 3′ end of U6 interacts with the 5′ end of U2 to form U2/U6 helix II, and U4/U6 remains base paired (Figure 1). The activated B complex (B^{act}) then forms by transfer of the 5′ ss from U1 to the ACAGA box of U6 (Staley and Guthrie, 1999) and unwinding of the U4/U6 base-pairing interaction (Figure 1C), leading to release of both U1 and U4 snRNPs. The latter unwinding event allows U6 sequences from U4/U6 stem I to pair with the 5′ end of U2 to form U2/U6 helix I (Madhani and Guthrie, 1992), and U6 sequences from U4/U6 stem II to form the U6 internal stem-loop (ISL) (Figure 1D). The emergent catalytic centre, encompassing U2/U6 helix I and the U6 ISL, is flanked by the first-step reactants, that is, the 5′ ss, via interaction with the U6 ACAGA box (Wassarman and Steitz, 1992; Lesser and Guthrie, 1993), and the branchpoint adenosine, via the connection of the branchpoint helix to U2/U6 helix I. After the final activation step is mediated by the PRP2 RNA helicase (Kim and Lin, 1996) (see Supplementary Table S1 for protein names), the catalytically competent B* complex forms, and it catalyses step 1, leading to C complex formation. U5 loop 1 (Figure 1B) is thought to align the exons for step 2 by contacting both splice sites (Newman and Norman, 1992; Sontheimer and Steitz, 1993; Dix *et al*, 1998; O'Keefe and Newman, 1998; McConnell and Steitz, 2001); but its function may be redundant as it is not essential for step 1 of splicing (O'Keefe *et al*, 1996; Ségault *et al*, 1999). After a remodelling step by the PRP16 RNA helicase, step 2 is catalysed. The mRNA is then released and the spliceosome is disassembled.

Although the splicing apparatus is evolutionarily highly conserved (Fabrizio *et al*, 2009), some of the RNA–RNA interactions differ between yeast and man. For example, a shortened U2/U6 helix I, consisting only of helix Ia (Figure 1D), was proposed for the mammalian spliceosome

*Corresponding authors. K Hartmuth or R Lührmann, Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. Tel.: +49 551 201 1405; Fax: +49 551 201 1197; E-mail: reinhard.luehrmann@mpi-bpc.mpg.de or klaus.hartmuth@mpi-bpc.mpg.de

Received: 13 November 2012; accepted: 24 July 2013; published online: 3 September 2013

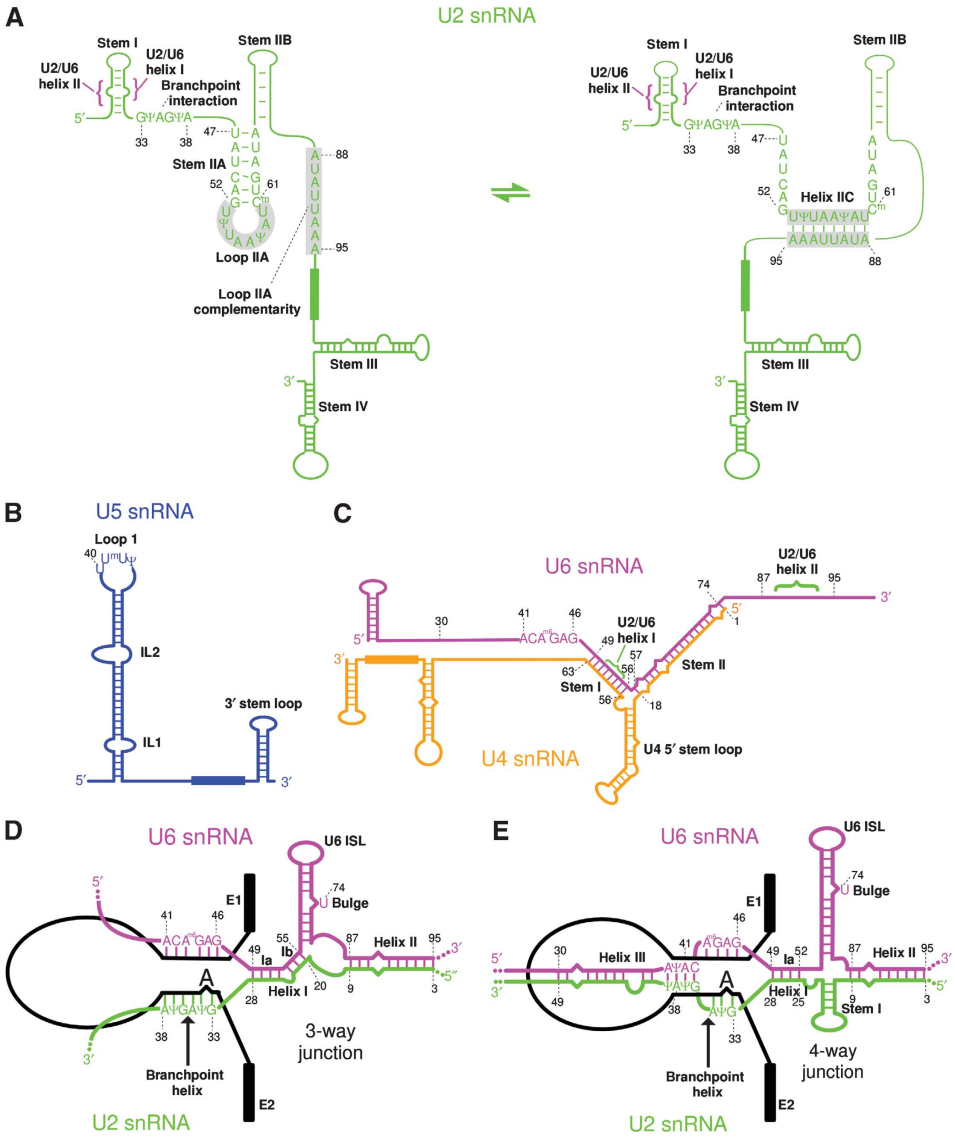

Figure 1 RNA structures in the spliceosome. (**A**) Proposed alternative structures of the human U2 snRNA based on the models of the yeast U2 snRNA. (**B**) Structure of U5 snRNA. (**C**) Structure of the U4/U6 di-snRNA duplex. (**D**, **E**) Models for the step 1 catalytic centre according to Madhani and Guthrie (1992) and Sun and Manley (1995), respectively. The following colour code for the U snRNAs is used in all diagrams: green U2, magenta U6, black pre-mRNA, and blue U5. The solid boxes denote the snRNA Sm binding sites. Black boxes E1 and E2 are exon 1 and exon 2. Pre-mRNA intron is depicted by a black line. All sequences and nucleotide positions shown are from human and abbreviations for natural modifications are in Supplementary Figure S2.

(Sun and Manley, 1995), which would create a 4-way junction between U2 and U6 (Sashital *et al*, 2004), as opposed to the 3-way junction proposed for yeast (Madhani and Guthrie, 1992). A U2/U6 helix III (Figure 1D), upstream of the U6 ACAGA box, was proposed to form in both spliceosomes (Figure 1), but with minor differences in the nature of the helix (Sun and Manley, 1995; Yan and Ares, 1996). Two mutually exclusive U2 structures were proposed to toggle iteratively in the yeast spliceosome (i.e., the U2 stem IIA/IIC toggle; see Figure 1A) (Hilliker *et al*, 2007; Perriman and Ares, 2007). It is not known whether this U2 toggling also occurs in the human spliceosome. Thus, a reliable picture of the RNA–RNA network in the mammalian spliceosome is currently lacking.

The idea that self-splicing group II introns are the evolutionary ancestors of nuclear pre-mRNA introns (Sharp, 1985; Cech, 1986) suggested early on that the function of RNA elements of group II introns might be provided *in trans* by spliceosomal U snRNAs during nuclear pre-mRNA splicing (Sharp, 1988), and that they might catalyse the latter. This suggested that the spliceosome, like group II introns, might be a ribozyme, with its catalytic core embedded in a protein–RNA rather than an RNA-only scaffold. The reaction mechanisms and stereochemistry are identical in the two systems (Moore and Sharp, 1993; Padgett *et al*, 1994; Sontheimer *et al*, 1999). Structurally, domain V of group II introns and the U6 ISL in the spliceosome are homologous (Madhani and Guthrie, 1992; Shukla and Padgett, 2002; Keating *et al*, 2010), and both bind an essential magnesium at identical positions: at C377 in a group IIC intron in *Oceanobacillus iheyensis* (Baker *et al*, 2001), and at U80 in the U6 ISL in yeast (Fabrizio and Abelson, 1992; Yean *et al*, 2000), which is equivalent to

U74 in human (Figure 1D). Both systems have a so-called catalytic triad of three consecutive conserved nucleotides (nts) (AGC at nts 53–55 in human U6, CGC in group IIC, and AGC in group IIA) that also coordinates the essential magnesium. These elements are positioned in the structure of an excised group IIC intron by triple interactions, with the bulged nucleotide of domain V and the so-called J2/3 junction interacting with the catalytic triad (Baker *et al*, 2001; Toor *et al*, 2008). Similar interactions have been proposed for the spliceosome, with the GA immediately following the U6 ACAGA box mimicking the function of the J2/3 junction (Michel *et al*, 2009; Keating *et al*, 2010). Finally, U5 loop 1 (see above) is thought to be functionally analogous to the exon-binding site (EBS1) in group II introns (Newman and Norman, 1992).

To date, RNA structures have never been probed directly in purified spliceosomes. Here, we have characterized the RNA–RNA network in highly purified, human B, B^{act}, and C spliceosomal complexes by RNA structure probing and psoralen crosslinking. Using the proposed similarities to group II introns (Michel *et al*, 2009; Keating *et al*, 2010) and experimentally derived structural constraints, we constructed a model of the RNA–RNA network of the human spliceosomal C complex. This model was then docked to the crystal structure of the Prp8 protein (Galej *et al*, 2013), thought to be the central binding platform for the functional RNA core of the spliceosome.

Results

RNA structure probing of purified human spliceosomes

To probe their RNA structure, human B, B^{act}, and C spliceosomal complexes, assembled in HeLa cell nuclear extracts *in vitro*, were purified via MS2 affinity selection. C complexes were formed on the PM5 pre-mRNA, which lacks exon 2 (Supplementary Figure S1), thereby stalling splicing after step 1. C complexes assembled on PM5 are functional, as they catalyse step 2 when exon 2 is provided *in trans* (Bessonov *et al*, 2008). B complexes were assembled on PM5 by incubating for 8 min, and B^{act} complexes were formed on the PM5–20 pre-mRNA, which contains a truncated PPY tract that stalls splicing at the B^{act} stage (Bessonov *et al*, 2010). We treated the isolated complexes with dimethyl sulphate (DMS; A- and C-specific), 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulphonate (CMCT; U-specific and low reactivity towards G), or β-ethoxy-α-ketobutyaldehyde (kethoxal; G-specific) (Ehresmann *et al*, 1987). Reactivity of a nucleotide towards any of these reagents indicates that it is unpaired and accessible. A lack of reactivity indicates that it is involved in a base-pairing interaction or interacts with a protein. RNA base modifications were detected by primer extension with reverse transcriptase (RT), which stops after transcribing the nucleotide immediately preceding the modified base (Ehresmann *et al*, 1987).

We initially analysed the U5 protection pattern in purified U4/U6-U5 tri-snRNPs and B, B^{act}, and C complexes (Figure 2; Supplementary Figure S2). By quantitating the intensity of the reverse transcription stops, we distinguished three levels of modification: weak, medium, and strong, as indicated by the size of the dots (Figure 2B, inset; see Supplementary Figure S2B for an example of quantification). All known double-stranded regions of U5 were inaccessible in all complexes analysed. Major changes in accessibility of the internal loops (ILs) 1 and 2 were not detected, suggesting that U5 does not undergo major structural rearrangements during splicing. Only minor changes were observed in IL 1 and IL 2 during the B to B^{act} transition, which likely reflect subtle changes in U5 snRNA structure and/or protein–RNA interactions at this stage. In contrast, nearly all U5 loop 1 nucleotides (C38–U41) were accessible in the tri-snRNP, but their accessibility decreased during the transition from the B to C complex (Figure 2A, CMCT for B and C complexes). As U5 loop 1 contacts the 5′ and 3′ exons prior to step 2 of splicing (Sontheimer and Steitz, 1993), shielding of U41 and U42 in

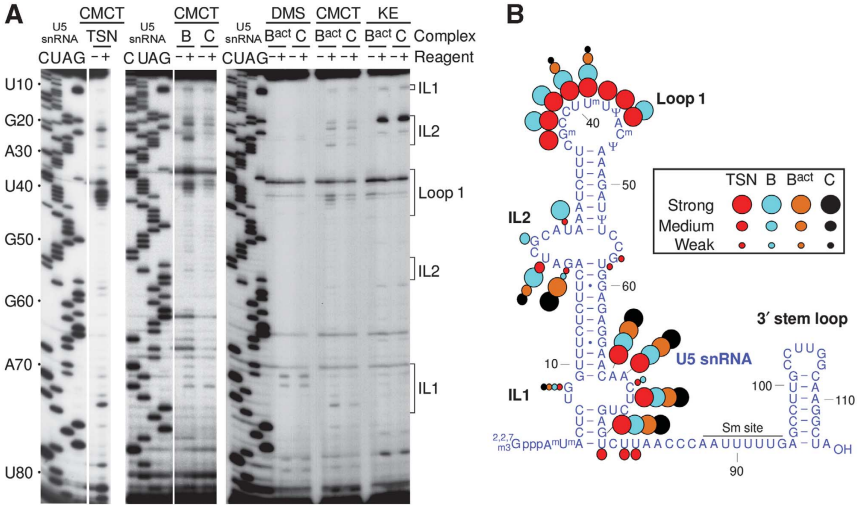

Figure 2 Nucleotide accessibility of U5 snRNA in spliceosomal complexes. (**A**) Chemical modification of the tri-snRNP (TSN) and B, B^{act}, and C complexes with the reagents indicated above the lanes (see text for details; KE: kethoxal). Modified bases were detected by primer extension analysis with an oligonucleotide complementary to U5 nts 83–103. See Supplementary Figure S2 for complete data set. Lanes to the right of a particular sequence were from a single gel; structural features of U5 are indicated on the right. (**B**) Summary of U5 base accessibilities observed in spliceosomal complexes. The colour code and the correlation of the dot sizes with reactivity are shown in the inset. Nucleotides upstream from the primer annealing site that do not bear labels were unreactive. Superscript m: 2′O-methyl. Source data for this figure is available on the online supplementary information page.

U5 is likely due to their interaction with the sole exon present in the PM5 and PM5–20 substrates.

A U2/U6 3-way junction at the core of the RNA network of the B^{act} and C complexes

To determine whether the U2/U6 duplex forms a 3- or a 4-way junction in active human spliceosomes, we investigated U2 and U6 base accessibilities in B, B^{act}, and C complexes (Figure 3; Supplementary Figures S3 and S4). As shown by protection of U2 nucleotides 3–11 (Supplementary Figure S3, hII) and U6 nucleotides 87–95 (Figure 3B; Supplementary Figure S4C), U2/U6 helix II, consistent with both the 3- and 4-way junction model, was present in all three complexes. In the 4-way junction model (Sun and Manley, 1995), the CGC sequence of U6 (nucleotides 79–81) is paired with U6 nucleotides 54–56 (Figure 3A, inset). Significantly, U6 nucleotides 79–86 immediately adjacent to U2/U6 helix II were accessible in B^{act} and C complexes (Figure 3B) and thus not engaged in base-pairing interactions. This is not compatible with the formation of a 4-way junction, but rather supports a 3-way junction similar to that thought to form in yeast (Madhani and Guthrie, 1992).

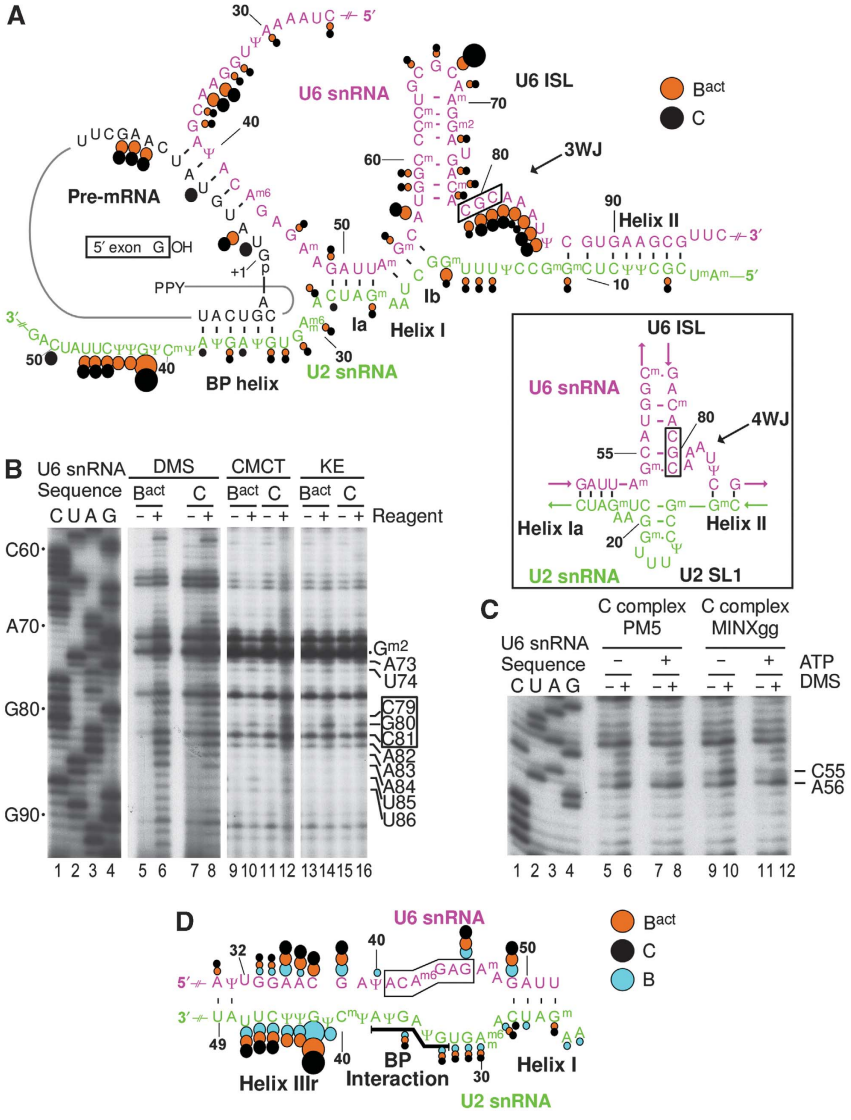

Figure 3 A 3-way junction is at the core of the human spliceosome. (**A**) Summary of nucleotide accessibilities determined by RNA structure probing of U2, U6, and the pre-mRNA in the core region of the B^{act} and C spliceosomal complexes. The degree of accessibility is indicated as in Figure 2B, and is superimposed on the RNA–RNA network of the C complex, drawn according to Madhani and Guthrie (1992); 3WJ: three-way junction. Inset: the four-way junction (4WJ) model (Sun and Manley, 1995); the critical CGC sequence that would be sequestered by base pairing according to this model is boxed. (**B**) Chemical modification of the 3′ end of U6 snRNA, analysed by primer extension (primer complementary to extra nucleotides added to the 3′ end of U6). Labels are as in Figure 2A. The DMS lanes are derived from an independent experiment, which was performed under identical conditions. RT stops corresponding to modified nucleotides in the single-stranded region between the 3′ end of U6 ISL and the beginning of U2/U6 helix II are indicated on the right. The CGC sequence (see inset in **A**) is boxed. Nucleotides U74 and A73 are labelled, as is the strong background stop corresponding to G^{m2}. (**C**) ATP-dependent DMS protection of C55 in U6. Purified C complexes formed on PM5 or MINXgg pre-mRNA were analysed as indicated. Complexes were treated with ATP (+) or left untreated (−) while still bound to the affinity column. After elution, they were immediately treated with DMS (+) or mock-treated (−). RNA was analysed by primer extension with an oligonucleotide complementary to U6 nts 80–100. Only the relevant portion of the sequencing gel is shown. (**D**) Summary of U2 and U6 protection in the proposed helix III region in the B, B^{act}, and C complexes. The original helix III spans U6 nts 30–42 and U2 nts 36–49. The ACAGA box of U6 is boxed. Source data for this figure is available on the online supplementary information page.

The determined accessibility pattern of nucleotides in the U6 ISL in both B^{act} and C complexes is consistent with its previously proposed structure (Epstein *et al*, 1980; Fortner *et al*, 1994). Specifically, nucleotides C62–U64 and A70–G72 were protected from chemical modification, suggesting that the upper stem forms, while nucleotides C66–C68 were accessible, indicating that the upper loop forms (Figure 3B). The lower part of the stem showed some partial reactivity, which is due either to helix breathing or to the presence of the non-canonical G·A base pair. The bulged nucleotide U74 was protected in the B^{act} and C complexes, and A73, proposed to pair with C61, showed consistently low reactivity with DMS. While an alternative structure is possible (see below), the U6 ISL structure that forms upon spliceosome activation persists through step 1 of splicing, as only minor differences were observed in the loop in the B^{act} versus C complex.

RNA structure probing showed that both B^{act} and C complexes likely contain U2/U6 helix Ia and Ib, as nts 49–55 of U6 (Supplementary Figures S4A and B) and 20–28 of U2 (Supplementary Figure S3, hI) were protected from modification, except for possible breathing of the G-C pair at the beginning of U2/U6 helix Ia (summarized in Figure 3A). The only change observed upon C complex formation was DMS reactivity at U6–C55 (Supplementary Figures S4A and B). However, when ATP (which is normally absent during MS2 affinity selection of the C complex) was present throughout the purification steps, U6–C55 was unreactive towards DMS (Figure 3C, cf lanes 6 and 8). No additional ATP-dependent changes in RNA accessibility were found, and identical results were obtained with a MINX pre-mRNA substrate containing a mutated 3′ ss (MINXgg; Supplementary Figure S1) (Figure 3C).

Although all bases in the U6 ACAGA box (nts 41–45) were protected from chemical modification, intron position A + 3 was invariably accessible in both B^{act} and C complexes, and U + 2 and U + 6 became accessible in the C complex (Supplementary Figures S4A–C). This indicates that initially only ACA of the ACAGA box base pairs with intron nucleotides near the 5′ ss (Figure 3A). For the U2 branchpoint binding region (nts 33–38), the protection pattern revealed that the U2/pre-mRNA branchpoint helix is formed in both the B^{act} and C complexes (Figure 3A; Supplementary Figure S3). Finally, most U6 nucleotides proposed to base pair with U2 to form U2/U6 helix III (i.e., nts 30–42) were accessible to chemical modification, with the exception of U31 and U32 at the end of the proposed U2/U6 helix III (Figure 3D; Supplementary Figure S4). Correspondingly, U2 nts 42 and 45–47, which are proposed to base pair with U6 in helix III, were accessible in all complexes analysed (Figures 3D and 4A–C; Supplementary Figure S3); notably U2–A48 (which is thought to base pair with U6-U31), as well as U2–U49 were protected. Taken together, these data indicate that the entire U2/U6 helix III does not form in B, B^{act}, or C complexes.

Mutually exclusive U2 snRNA conformations are not detected in B, B^{act}, or C complexes

We next investigated whether, as observed in yeast, U2 snRNA undergoes a conformational change during the B-to-C complex transition, in which U2 stem IIA opens allowing formation of U2 helix IIC (Figure 1A) (Hilliker *et al*, 2007; Perriman and Ares, 2007). Formation of U2 helix IIC in human spliceosomes would lead to protection of U2 nts 89–95 (Figure 4C). However, in B, B^{act}, and C complexes, these nucleotides, as well as their neighbours, were fully accessible (Figures 4A–C; Supplementary Figure S3). We thus conclude that U2 helix IIC does not form in human B, B^{act}, and C spliceosomes. However, during the B^{act} to C transition, we observed increased accessibility of nts U55 and U58 in the loop, in the 3′ strand of the stem (C61–G63), and at C50 (Figures 4D–F). Thus, some restructuring of U2 stem-loop IIA occurs during or after step 1 of splicing. Finally, the first two base pairs of U2 stem IIA (U47–A66 and A48–U65) did not form in any of the spliceosomes analysed, as A66 and U65 were accessible in B, B^{act}, and C complexes (Figures 4D and E). Although A48 was protected, it does not base pair with U65, but appears to be involved in other interactions.

Psoralen crosslinking reveals a novel intron–U5 interaction

Psoralen crosslinking is a powerful method to determine higher order interactions in RNA assemblies (Hausner *et al*, 1990; Wassarman and Steitz, 1992). Upon UV irradiation at 364 nm, psoralen will primarily form mono-adducts with favourably exposed bases and, to a lesser extent, di-adducts between bases if they are in the correct orientation relative to one another, for example, on opposite strands of a helix facing each other. To analyse such interactions between spliceosomal RNAs, we performed psoralen crosslinking with B, B^{act}, and C complexes, and looked for crosslinked RNAs by northern blotting, probing sequentially for U2, U5, U6 (Figures 5A–C) and U4 (Supplementary Figure S5A), or by direct autoradiography of the blot, to reveal the position of radiolabelled pre-mRNA-containing crosslinks (Figures 5A–C, pre-mRNA panels). The identities of the crosslinked RNAs can be deduced by comparing the migration behaviour of the various bands detected with each probe; a band detected by two (or three in the case of a double crosslink) different probes indicates the formation of a crosslink between each of the probed RNA species.

Within our purified C complexes, there are two different forms of the intron lariat, designated as long (I_L) (i.e., essentially full-length ending at nt 478), and short (I_S), which apparently arises via 3′ end trimming of the intron. The latter is estimated to lack ~50 intron nucleotides (i.e., to end at ca nt 430), based on the relative gel migration of the two intron species and comigrates with exon 1, as revealed by RNase H digestion assays (data not shown). The relative abundance of I_L and I_S varies among different extracts and is probably due to a nucleolytic activity (Ruskin *et al*, 1984; Ruskin and Green, 1985). Crosslinks containing both forms are most apparent if the crosslinked species is small (i.e., with the U5 snRNA, see below) and, due to their poor separation on the gel, both forms are less apparent with longer crosslinked partners (i.e., the U2 snRNA).

The pattern of psoralen crosslinks was independent of ATP during the spliceosome affinity purification steps (Figure 5, compare + ATP with − ATP), suggesting that energy is not required to maintain the RNA architecture once it is established. A U2/U6 crosslink mapping to U2/U6 helix II (Hausner *et al*, 1990) was detected in all complexes (Figure 5, U2/U6 hII). In addition, a crosslink containing the pre-mRNA and U6 (pre/U6; B and B^{act} complexes) or U6 and the lariat intron (I/U6; C complex), corresponding to the intron/U6 ACAGA-box interaction (Wassarman and Steitz, 1992), was also

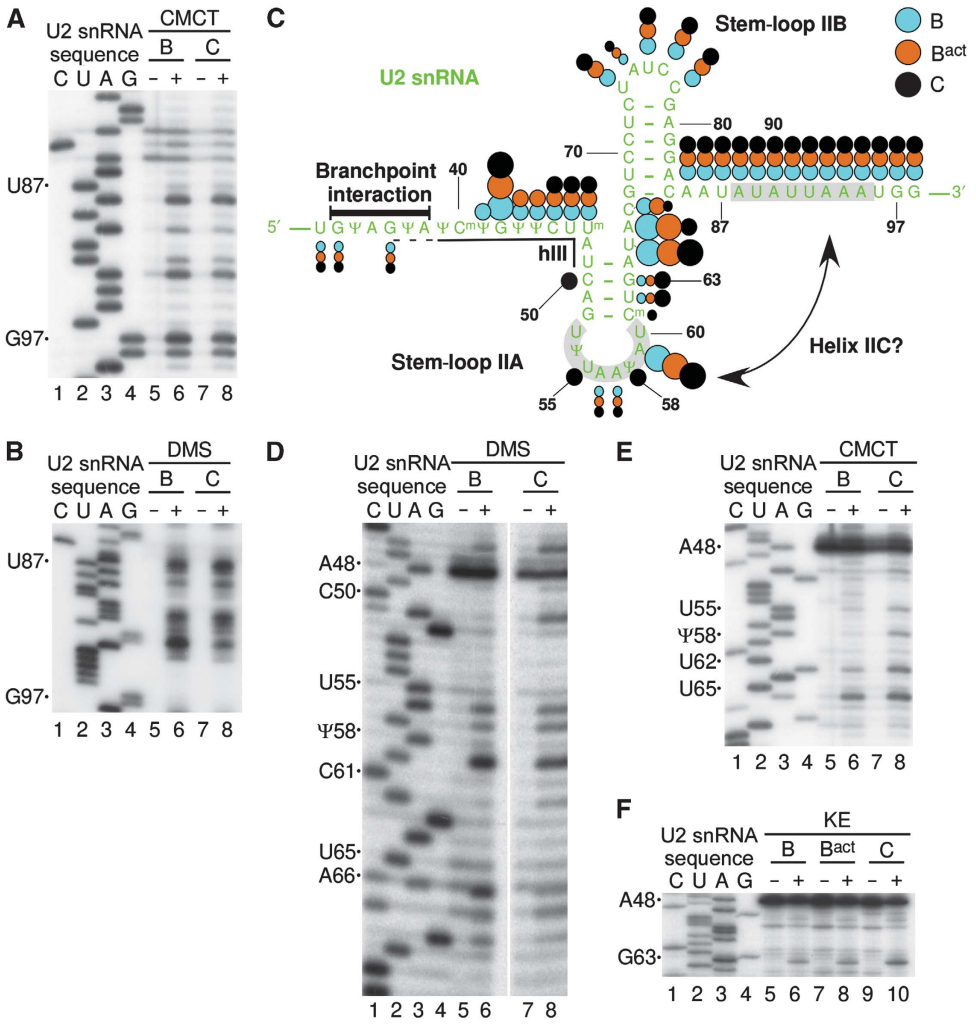

Figure 4 Remodelling of U2 stem-loop IIA between steps 1 and 2 of splicing. Base accessibility to CMCT (**A**) and DMS (**B**) of U2 nts U87–G97 in B and C complexes, analysed by primer extension with oligonucleotide K31 (complementary to U2 nts 149–169). (**C**) Summary of base accessibility of the U2 snRNA from the branchpoint binding region (thick bar) to the Sm binding site (starting at G98). Colour code as in Figure 2B. Solid line: sequences that would be involved in U2 helix III (hIII); shading: nucleotides that would be paired if helix IIC was formed (double-headed arrow). (**D**–**F**) DMS, CMCT, and kethoxal (KE) modifications of the stem IIA region of U2 in B and C complexes. Critical nucleotides are shown on the right. The strong DMS modification of A66 that leads to stops at positions C67 and A66 (compare lane 6 with lanes 1 and 3) is likely due to gel compression that results in the A66 signal being split into two bands. A similar effect was observed at this position in the ddG sequencing lane 1 (see also Supplementary Figure S3B). Source data for this figure is available on the online supplementary information page.

detected. The I/U6 crosslinked RNA species comigrates with the U2/U6 hII crosslink (as revealed by RNase H digestion, Supplementary Figure S6A). A second U2/U6 crosslink (U2/U6 hIIIr) migrating above U2/U6 hII, and a double crosslink containing the intron, U2, and U6 (I/U2/U6) (Figure 5, see also below), were first observed in the C complex. An intron/U2 crosslink (I/U2) was very abundant in the C complex, but almost undetectable in the B and B^{act} complexes. Two crosslinks containing U5 and either the long or short form of the PM5 intron (I_L/U5 and I_S/U5; Figure 5, see also below) were also visible. A previously undetected double crosslink containing U2, the intron, and U5 (U2/I/U5, Figures 5B and C; see also below) was also observed in the C complex. For each double crosslinked species, both single crosslinked species must also be present. As both I/U2 and I/U5 crosslinks are detected (Figure 5C; see also Figure 6), the data suggest that in the U2/I/U5 double crosslinked species, both snRNAs are crosslinked to the intron. Similar psoralen-crosslinked species were also detected in C complexes formed on the MINXgg pre-mRNA (Supplementary Figure S5B), demonstrating that these RNA–RNA interactions are not substrate specific, but rather occur generally within the spliceosome.

To confirm the identity of the crosslinked RNAs and map the crosslinking sites, we performed DNA oligonucleotide-directed RNase H digestions (Hausner *et al*, 1990) with crosslinked RNA from purified C complexes. Oligonucleotide (oligo) induced RNase H hydrolysis will induce a change in mobility of the target depending on the position of the oligonucleotide binding site. We first analysed U5 and U2-containing crosslinks in total crosslinked RNA using oligos complementary to different portions of the intron (Figure 6C), the PM5 exon (Ex) or the U6 snRNA. Cleavage products were visualized by northern blotting using U5 and U2 probes (Figure 6A). The crosslinks designated as I/U5, I/U2, and U2/I/U5 were not cleaved by the exon (lanes Ex) or U6 oligos (which were, however, effective in cleaving the exon or U2/U6 crosslinks; Figure 6B), confirming that these crosslinks do not contain the exon or U6 snRNA. In contrast, the U5-containing

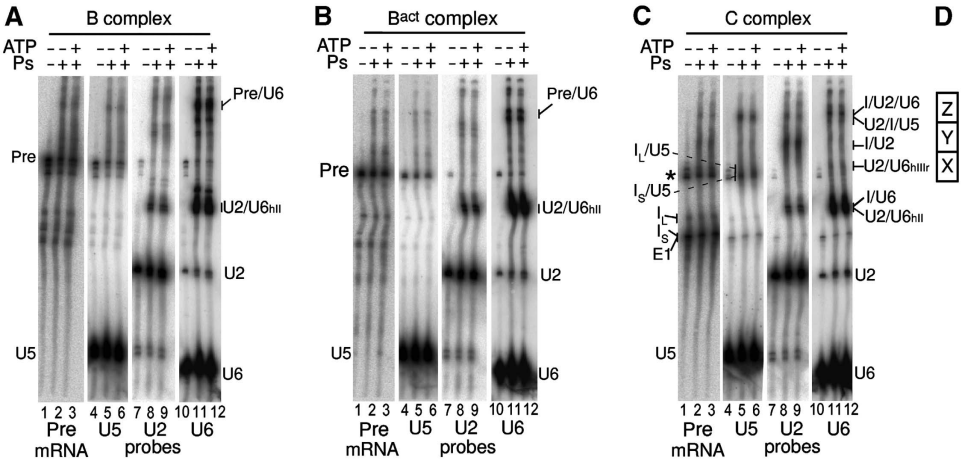

Figure 5 Identification of RNA–RNA interactions via psoralen crosslinking. Northern blot analysis of psoralen-crosslinked RNA obtained from the B (**A**), B^{act} (**B**), and C (**C**) complexes. The panels with signals obtained from the ^{32}P-labelled pre-mRNA (visualized directly), or probes against U5, U2, or U6 (indicated below) are displayed next to one another for each complex. All panels were derived from a single blot. Ps: psoralen; Pre: pre-mRNA; I, intron. Crosslinks are labelled to the right or left of the panels. Pre-mRNAs and first step intermediates are indicated on the left. The PM5–20 pre-mRNA used for B^{act} complex isolation is 68 nts shorter than the PM5 pre-mRNA used for the B and C complexes. The band labelled with an asterisk corresponds to a truncated pre-mRNA (~7% of total) remaining in the C complex preparation that could not undergo the first step of splicing. (**D**) X, Y, and Z are regions excised from an equivalent gel and from which crosslinked species were extracted for further characterization (Figure 7). Source data for this figure is available on the online supplementary information page.

crosslinks were cleaved with oligos complementary to the intron (Figure 6, U5 panel, lanes 2–4), confirming that the major U5 crosslinked species contain the latter.

Close inspection of the cleavage products containing U5 that are generated upon RNase H digestion with the intron oligonucleotides indicates that U5 is crosslinked downstream of the branchpoint in the linear 3′ tail of the lariat intron for the following reasons. With intron oligo C, which spans the branchpoint and could theoretically cleave on either side of it (Figure 6C), two U5-containing products (I and III) reflecting U5 crosslinked to the two forms of the intron lariat (I_L and I_S) are observed in the C complex (Figure 6A, lane 4). These cleavage products were not further cleaved with oligo A, which is complementary to nucleotides in the loop of the lariat (lanes 3 and 4); that oligo A is generally active is evident by the fact that the oligo C cleavage products of the I/U2 crosslink were further cleaved when oligo A was additionally present (Figure 6A, U2 probe, cf lanes 3 and 4). This indicates that the U5-containing cleavage products obtained with oligo C do not contain an intact or opened loop, and that U5 is crosslinked to the intron downstream of the branchpoint. The latter is also supported by the small size of the oligo C cleavage products containing U5. As the most downstream cleavage position of oligo C is theoretically at intron nt 401, the U5 crosslink must be located downstream of this position, but it cannot be further downstream than *ca*. nucleotide 430, the approximate 3′ end of the shorter intron. Consistent with this, oligo B, which binds a few nucleotides further downstream of oligo C, produces even shorter fragments (II and IV) crosslinked to U5 (lane 2), albeit with low efficiency (cleavage with oligo B alone is generally inefficient). In summary, these data indicate that the U5 crosslink is located somewhere between intron positions 401 and ca 430, which places it ~6–34 nucleotides downstream of the branchpoint.

As the U2/I/U5 crosslink was cleaved by the C oligo, but no additional U5-containing cleavage products were detectable (Figure 6A), the same two fragments appear to be generated from both the single and double crosslinks. This was corroborated by RNase H studies with purified U2/I/U5 crosslinked RNA, where identical, short U5-containing cleavage products were observed (Supplementary Figure S6C). Thus, U5 likely crosslinks to the same intron position in both I/U5 and U2/I/U5. This is also consistent with the fact that the double crosslink must arise from two, more abundant, single crosslinks, namely I/U5 and I/U2.

The crosslinks designated as I/U2 and U2/I/U5 were also targeted by the intron oligos, as well as by an oligo complementary to U2 (Figure 6), confirming the presence of the intron and U2 snRNA (Figure 6A, U2 panel); the material still detected at the position of U2 alone after cleavage with the U2 oligo is a cleavage product of the U2/U6 helix II crosslink (compare Figure 6B, lanes 3 and 4, U2 and U6 probe panels). The intron/U2 crosslink was cleaved with intron oligo C that spans the branchpoint region (Figure 6A, U2 panel, lane 4). When oligo A was present in addition to C a smaller (compared with oligo C alone) cleavage product containing U2 (band VII) was observed (compare lanes 3 and 4, bands VII and VI). This indicates that the U2 crosslink is associated with the lariat structure, consistent with it being crosslinked while in a standard branchpoint helix interaction. Accordingly, the U2 crosslinked fragment generated by RNase H digestion with oligo B (band V) was slightly larger than that obtained with oligo C alone (band VI), indicating that U2 is crosslinked upstream of the oligo B binding site.

The identity of the U2/U6 crosslinks in the C complex was similarly confirmed using oligos directed against U2 or U6 (Figure 6D, grey lines labelled U2 and U6), in the case of I/U2/U6 directed against the intron (Supplementary Figure S6), and subsequent northern blotting using U2 and U6 probes (Figure 6B; Supplementary Figure S6A). All crosslinks containing U2 and U6 were cleaved with the U2 and U6 oligos (Figure 6B, lanes 3 and 4, U2 probe and U6 probe panels).

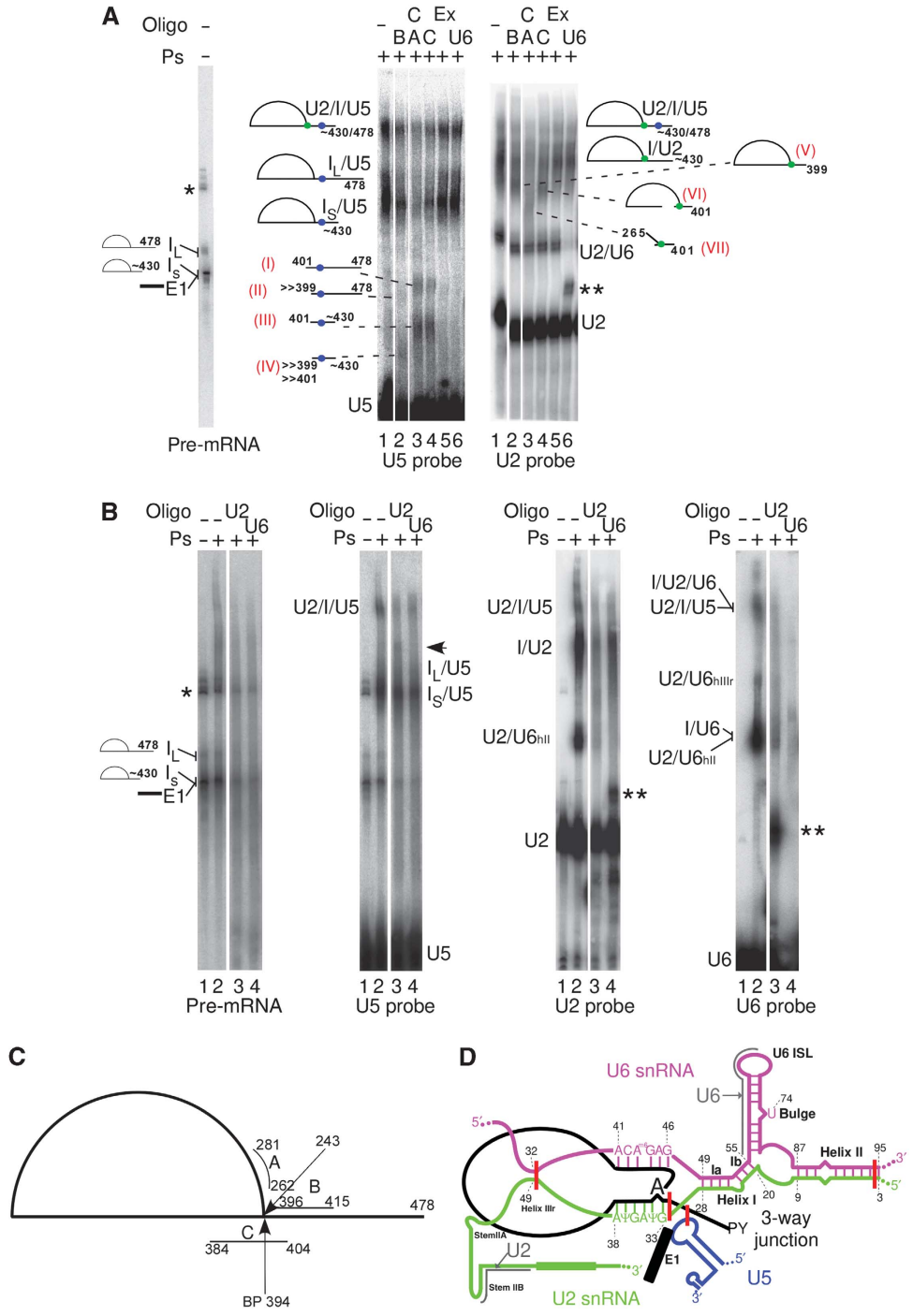

Figure 6 Mapping of C complex-specific psoralen crosslinks by site-specific RNase H cleavage. Total psoralen-crosslinked RNA from the C complex was incubated with RNase H and the oligonucleotides indicated above each lane (see **C** and **D** for their position), and the reaction products were subsequently identified by northern blotting. The origin of the signals is noted below the panels in (**A**) and (**B**), with 'Pre-mRNA' indicating the signal derived from the ^{32}P-labelled pre-mRNA. (**A**) Mapping of crosslinks of U2 and U5 to the intron. The two intron species (intron long: I_L and intron short: I_S) that are present in the C complex are indicated. The likely structures of the main cleavage products are shown schematically, where a blue dot indicates crosslinked U5 snRNA and a green dot indicates crosslinked U2 snRNA. The location of the cleavage sites is indicated on the cleavage products, assuming that RNase H will cleave at most 3 nts away from the ends of a DNA/RNA duplex (Nowotny *et al*, 2005). (**B**) Mapping of snRNA–snRNA crosslinks. The asterisk denotes truncated pre-mRNA (see Figure 5). The double asterisks denote cleavage products from the U2/U6 helix II duplex and the arrow denotes the cleavage product from the U2/I/U5 crosslink. (**C**, **D**) The annealing positions of DNA oligonucleotides used for RNAse H mapping are indicated by thin lines. In (**D**), the position of the crosslinks, as determined by primer extension analysis (see Figure 7), is indicated by red bars. Source data for this figure is available on the online supplementary information page.

As I/U2 and U2/U6 hII (as opposed to U2/U6 hIIIr) are the major, single crosslinks, the double crosslinked I/U2/U6 species most likely consists of U2 crosslinked both to the intron, and to U6 via U2/U6 helix II. Given that the crosslink species designated as U2/U6 IIIr migrates slower than U2/U6 hII, it should contain a crosslink at a position other than U2/U6 helix II; indeed, as shown below, primer extension analyses suggest a crosslink within a region of U2/U6 helix III.

Next, we determined the precise crosslinking sites by primer extension analysis (Figure 7). As RT is unable to extend through the site of crosslinking, the last nucleotide incorporated by RT should directly precede the crosslinked nucleotide. Additional psoralen-induced RT stops observed downstream of the very last stop, can be attributed to the formation of psoralen mono-adducts. As it was not possible to purify individual crosslinks due to their complex comigration pattern (Figure 5), we purified RNA from three regions of the gel (X, Y, and Z in Figure 5D, where the U2/I/U5 and I/U2/U6 crosslinks should be mainly in Z, the I/U2 crosslink mainly in Y, and the I/U5 and U2/U6 hIIIr crosslinks mainly in X). Primer extension with a primer complementary to U5 (Figure 7A) revealed a strong last RT stop at the nucleotide preceding U40 with RNA from the gel regions X and Z, indicating that U5–U40 crosslinks with the intron in both the I/U5 and the U2/I/U5 crosslink species. In the case of U2, the presence of mono-adducts resulted in multiple strong RT stops (reducing the intensity of more upstream stops and making their quantitative comparison difficult), with the last RT stop observed at the nucleotide preceding U2–U32 with RNA isolated from all three regions. These data indicate that in the I/U2 (region Y) and the U2/I/U5 and I/U2/U6 crosslinks (region Z), the intron is crosslinked to U2–U32. As this U2 nucleotide is involved in the U2/pre-mRNA branchpoint helix interaction, these data further indicate that this U2/pre-mRNA interaction is still present in the C complex. The signal observed with RNA from region X is likely due to contamination with the I/U2 crosslink due to smearing.

When U6 crosslinks in regions X, Y, and Z were analysed using a primer that anneals upstream of U2/U6 helix II, the last RT stop observed with RNA from region Z was the full-length U6 snRNA, consistent with the idea that U6 in the I/U2/U6 double crosslink species is only crosslinked to U2 via helix II (which is not detected by the primer used in this experiment). In contrast, the last stop in regions X and Y was U6–U32, which was previously proposed to be involved in a base-pairing interaction in the U2/U6 helix III (see Figure 3D; Sun and Manley, 1995). This suggests that in the crosslink species designated as U2/U6 hIIIr (which may also contaminate the Y region of the gel), U6–U32 is crosslinked to the U2 snRNA, the only other U6 crosslink partner in this region of the gel. Our RNA structure probing showed that U6 nucleotides surrounding the crosslinked U6–U32 are protected (Figure 3A), and that U2 nucleotides surrounding U2–U49 are also protected (Figure 3A), suggesting an RNA–RNA interaction between U2 and U6 that maps to a region of U2/U6 helix III (Figure 3D; Sun and Manley, 1995). However, due to the extensive formation of psoralen mono-adducts on U2, clear evidence for the formation of a U2 crosslink at U2–U49 or neighbouring nucleotides could not be obtained. Nonetheless, both RNA structure probing and psoralen crosslinking suggest that there is an interaction between U6 and U2 in the region previously proposed to form U2/U6 helix III, which does not appear to form entirely in the human C complex. Importantly, similar results were obtained when the crosslinked species from C complexes formed on the MINXgg substrate were also analysed by primer extension (Supplementary Figure S7).

A three-dimensional model of the RNA network in the spliceosomal C complex

The extensive parallels between nuclear pre-mRNA splicing and group II intron autocatalytic excision suggested that the core in both systems may be structurally homologous. As our purified C complexes contain a functional catalytic centre, and as the excised group IIC intron from *O. iheyensis* (Toor *et al*, 2008) is the catalytically relevant structure (Baker *et al*, 2001), we constructed a model of the RNA network of the C complex (i.e., post-step 1) by RNA homology modelling (Massire and Westhof, 1998; Jossinet *et al*, 2010; Westhof *et al*, 2011). To this end, we modelled the respective U2/U6

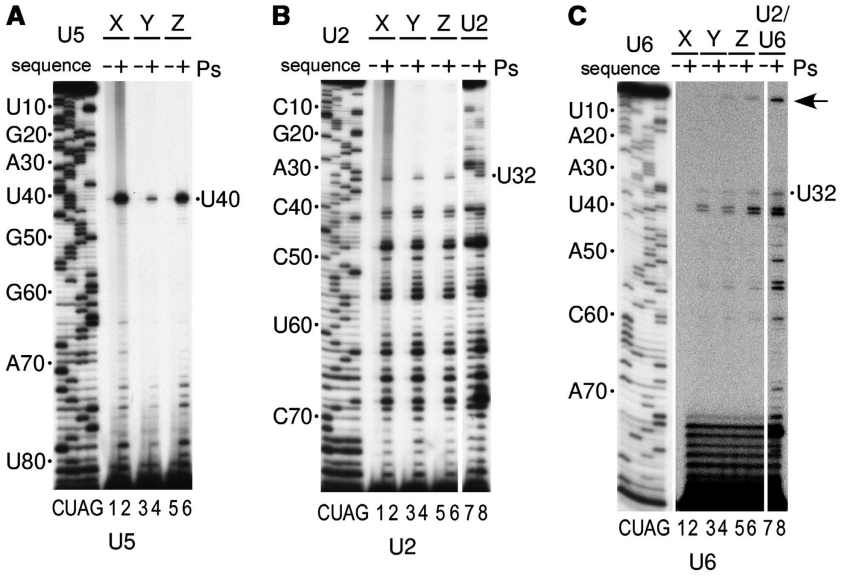

Figure 7 Mapping of psoralen-crosslinked nucleotides by reverse transcription. (**A**–**C**) RNAs crosslinked in C complexes formed on the PM5 substrate were isolated from gel regions X, Y, and Z (see Figure 5D) as indicated above each lane, and analysed by reverse transcription with primers complementary to U5 nts 83–103 (**A**), U2 nts 77–97 (**B**), or U6 nts 80–100 (**C**). Ps: psoralen; the control was material isolated from an equivalent gel region from a non-crosslinked sample. Primer extension of RNA isolated from the U2 (in **B**) or U2/U6 (in **C**) region of a preparative gel (compare Figure 5). Labelled stops correspond to the crosslinked nucleotides. Source data for this figure is available on the online supplementary information page.

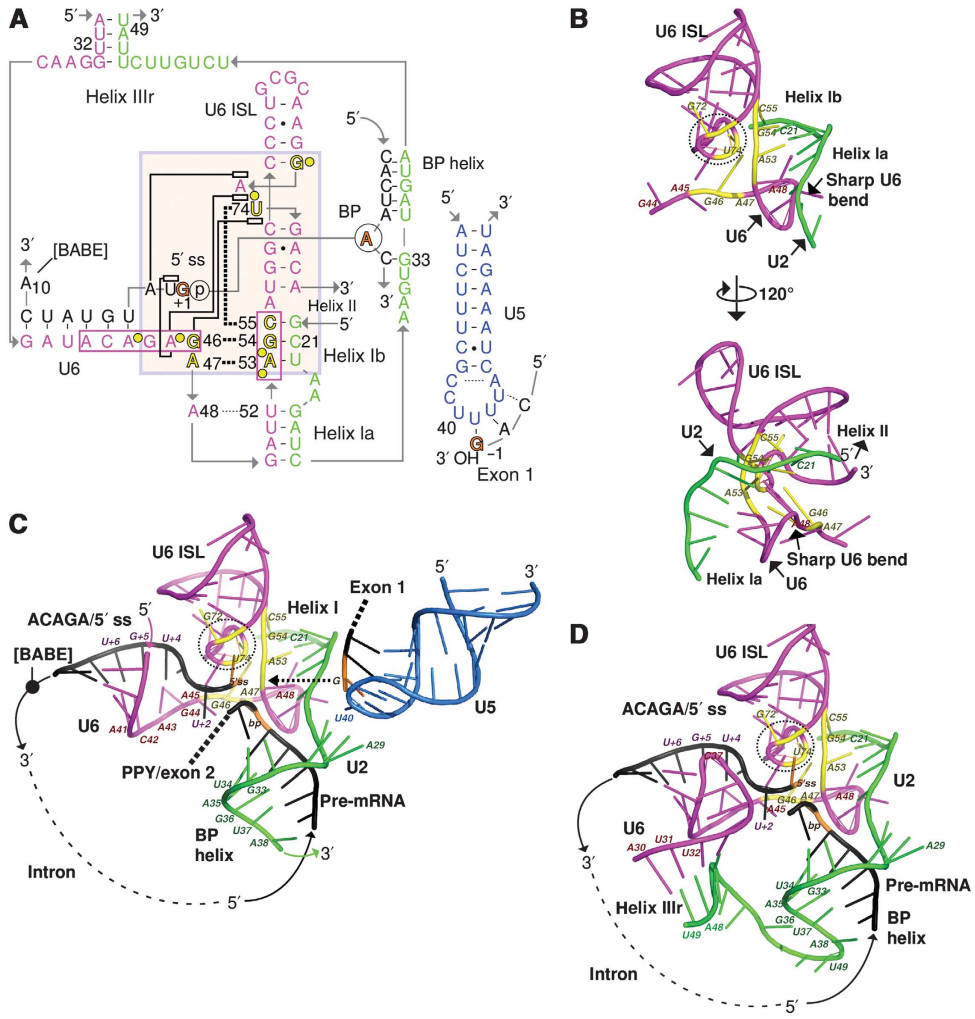

Figure 8 A model of the RNA network in the spliceosomal C complex. (**A**) RNA structural elements used in modelling. Intron positions start with +1. Thick-dotted lines: tertiary interactions taken from a group II intron (Michel *et al*, 2009; Keating *et al*, 2010); thin-dotted lines: modelled interactions; solid lines ending in boxes: stacking interactions. Yellow: the catalytic triad and residues implicated in magnesium binding (Fabrizio and Abelson, 1992; Yean *et al*, 2000); orange: 5′ ss, branchpoint, and last nucleotide of exon 1; arrows connect strands in a 5′ to 3′ direction. BABE: the +10 position used for site-directed hydroxyl-radical probing (Rhode *et al*, 2006). Rectangular light red box: nucleotides modelled on the core of the group II intron by homology. (**B**) Two views of the RNA spliceosomal core. Face-on view of the catalytic centre (top), and a rotation clockwise by 120° (bottom) to show the position of helix II. The boxed region from (**A**) is shown, together with the attached helix Ia. An arbitrary tetraloop sequence was used to close the U6 ISL. (**C**) Positioning of the ACAGA/5′ ss and the branchpoint helix (BP) onto the core structure. See text for details. The most reasonable spatial positioning of U5 with the bound exon 1 is indicated by the arrow. The position of the PPY tract leading to exon 2 is schematically shown by the thick-dotted line (PPY/exon 2). (**D**) Positioning of the U2/U6 helix III region onto the core structure.

sequences (Figure 8A, nucleotides in shaded area; see Supplementary Figure S8B for a conventional representation) into the catalytic core from the group IIC intron. In this core, the bulged nucleotides of domain V, the catalytic triad, and the J2/3 junction form a triple helix (Supplementary Figure S8A) and similar interactions are proposed to be present in the spliceosome (Michel *et al*, 2009; Keating *et al*, 2010). Accordingly, we modelled the respective U6 RNA elements (bulged nucleotides of ISL, the catalytic triad AGC as part of the U6/U2 helix Ib, and the G46–A47 immediately following the ACAGA box) into the triple helix structure of the group II intron (Toor *et al*, 2008) (Figure 8A, thick-dotted lines). This fixed the orientation of the ACAGA box with respect to U6 in the catalytic centre.

The two-nucleotide bulge of domain V is structurally essential for positioning a magnesium ion in the group II intron. To allow a similar structure in the U6 ISL, we shifted the register of the upper stem by one nucleotide (Figure 8A). The resulting alternative pairing agreed much better with our RNA structure probing data (medium DMS reactivity at A73 and none at U74). A two-nucleotide bulge for U6 ISL was also suggested earlier (Toor *et al*, 2008; Michel *et al*, 2009) and the U6atac ISL from the minor spliceosome could successfully be replaced with a group II domain V element that had a two nucleotide bulge (Shukla and Padgett, 2002; Supplementary Figure S8B, inset). We next modelled U6/U2-helix Ia and the two bulging As that connect helix Ib (part of the triple helix). A very similar arrangement of two short helices was found in a 5S rRNA structure (Supplementary Figure S8C), and we thus used it as a template. To reach the upstream U6 dinucleotide G46–A47 (also part of the triple helix), a sharp bend at the connecting A48 had to be introduced (Figure 8B).

We then added the spliceosomal RNA elements to the homology core. To find the correct spatial positions, we were guided by RNA structural constraints (this work; Sontheimer and Steitz, 1993; McConnell and Steitz, 2001; Rhode *et al*, 2006) and the suggested natural continuations of the RNA strands from the homology core. To model the attachment of the ACAGA box, we assumed that only the ACA sequence paired with nucleotides +4 to +6 of the intron. Further, we added an idealized helix with three base pairs to the 5' end of the ACAGA box (Figure 8A) to facilitate spatial positioning. At the 3′ end, the thio-U crosslink of intron U+2 to the ultimate A45 of the ACAGA box (Sontheimer and Steitz, 1993) was modelled as a stacking interaction, in agreement with the crystal structure of a thio-dT/adenosine crosslink (Saintomé *et al*, 1996). The neighbouring U6–A43-intron–U+4 base pair fixed the orientation of the ACA mini-helix of the ACAGA box (Figure 8C). This in turn determined the orientation of the U6 and intron strands that join to form the ACAGA box, and also the spatial position of the 5′ end of the intron on the other side that points towards the catalytic centre. To attach the branchpoint helix, we first modelled the helix according to Berglund *et al* (2001) and connected it to U2/U6 helix Ia with the AAGU connector sequence (nts 29–32) of U2. This initial extended configuration (not shown) was adjusted manually to position the branchpoint. A bend was introduced at U2–A29, so that the branchpoint 2′ OH came close to the 5′ ss phosphate and additionally faced the catalytic centre. The final relative orientation of the branchpoint helix to the ACAGA box is such that the U2 strand faces the ACAGA box and the intron strand is on the opposite side of it. This is in agreement with previous site-directed hydroxyl-radical probing from the +10 intron position of the pre-mRNA (adjacent to the ACAGA box; Figure 8A); in these studies, the U2 backbone of the branchpoint helix was hit by radicals, whereas the intron with the branchpoint was not (Rhode *et al*, 2006).

We next attempted to fit U5 interacting with exon 1. To this end, we constructed an idealized U5 loop 1 with an attached exon 1 (Figure 8C). We modelled the U5 loop 1 on a tRNA anticodon loop because of their close structural and functional similarities (also, the EBS1 site of the *O. iheyensis* group II intron has a similar fold; see Supplementary Figure S8D). The tRNA wobble position corresponded to the U5–U40 position and exon 1 paired to the neighbouring loop 1 nucleotides corresponding to the tRNA anticodon. The model is idealized as the previous crosslinks (Sontheimer and Steitz, 1993) do not imply pairing. However, using the 3′ end of exon 1 as a guide for positioning, U5 with the exon 1 could be fit to the right of the catalytic centre (Figure 8C). This general position is homologous in space to the exon-binding site in the group II intron. Further, U5–U40 is then in the correct spatial proximity to interact with nucleotides of the PPY tract. This would satisfy our constraints from the I/U5 crosslink.

Interestingly, the nucleotides of U2 and U6 that interact in the suggested U2/U6 helix III region are located in the same spatial section of the model (Figure 8C, magenta and green arrows). To model this situation, we first created a small helix containing the interacting nucleotides of U6 and U2 (Figure 8A). By adjusting the U2 connection to the BP helix and the U6 connection to the ACAGA box, this idealized helix III region could be fit to the left of the catalytic centre by two sharp bends at the respective connecting nucleotides (Figure 8D). While some parts of the resulting RNA model are idealized situations, overall it represents a clash-free, homology model of the RNA core of the spliceosome that takes into account all available experimental constraints.

Discussion

Our characterization of the RNA–RNA network in purified spliceosomes solved a number of long-standing issues pertaining to the spliceosomal RNA network and supplied sufficient constraints for homology modelling of the RNA core of the spliceosomal C complex. RNA structure probing demonstrated that a 3-way junction of the U2/U6 duplex is present during step 1 and is actively maintained for step 2 (Figure 3). Transient structures may form during spliceosome transitions, however, the intact helix Ib that we detect during and after step 1 is also required for both steps of splicing in yeast (Hilliker and Staley, 2004). As our C complexes are stalled after the PRP16-mediated rearrangement required for step 2 catalysis (Bessonov *et al*, 2008), our data suggest that the core containing all essential metal-binding sites (Fabrizio and Abelson, 1992; Yean *et al*, 2000) does not change between the two catalytic steps, and thus that the spliceosome has only one catalytic centre, as also proposed previously by others (Steitz and Steitz, 1993). Our results further underscore the similarities between the spliceosome and group II introns, which also have only one catalytic centre (Chanfreau and Jacquier, 1994; Baker *et al*, 2001; Chan *et al*, 2012). For the spliceosome, this suggests that the main function of the structural transition between catalytic steps 1 and 2 is to remove the first-step phosphate, that is, the branchpoint branched to the 5′ end of the intron, from the catalytic centre and to replace it by the second-step phosphate in a concerted structural transition. During this transition, exon 1 is held in place by the interaction of loop 1 of U5, suggesting a placement of step 1 and step 2 phosphates similar to that found in group II introns (Chan *et al*, 2012).

We cannot conclusively determine the conformation of the U6 ISL at A73 and U74. A pH-dependent cycle of protonation and deprotonation of A73 (Venditti *et al*, 2009) is unlikely, as we did not detect any variability of DMS reactivity at A73. We adopted the two nucleotide bulge as the most likely active conformation as a basis for our modelling, as it is the conformation of the domain V transplant that functioned in the minor spliceosome (Shukla and Padgett, 2002). Essential proteins that contact the U6 ISL, such as Prp8 or Cwc2/RBM22 (Newman and Norman, 1992; Teigelkamp *et al*, 1994; Reyes *et al*, 1999; Siatecka *et al*, 1999; Turner *et al*, 2006; Rasche *et al*, 2012; Schmitzova *et al*, 2012) may induce such an active conformation. These proteins undoubtedly also may serve to stabilize the weak base-pairing interactions between the ACAGA-box and the pre-mRNA. Not surprisingly, the reactive groups in the protein-free structure of a U2/U6 model duplex are separated by large distances, and it was postulated that essential protein-mediated activation step(s) are necessary to bring them closer (Burke *et al*, 2012). Thus, genuine nuclear pre-mRNA splicing with only U snRNAs is highly unlikely (see Smith and Konarska, 2009 for discussion).

Stem-loop IIA of U2 was the only RNA structural element that showed restructuring dynamics concomitant with step 1 of splicing (Figure 4). The U2-specific splicing factors SF3a/b are known to be destabilized during catalytic activation

(Bessonov *et al*, 2008; Warkocki *et al*, 2009; Lardelli *et al*, 2010; Ohrt *et al*, 2012) and stem-loop IIA interacts with the RNA binding surface of SF3a (Lin and Xu, 2012). This suggests that stem-loop IIA structural dynamics might be linked to destabilization of SF3a, and possibly of SF3b, and may also be linked to the transition between steps 1 and 2. SF3a/b destabilization may also explain the enhanced intensity of the U2/intron crosslink in the C complex (Figure 5). As SF3a/b proteins also contact the pre-mRNA at/near the branchpoint sequence (Gozani *et al*, 1996), these proteins might shield this region from psoralen intercalation in the B and B^{act} complexes, but no longer in the C complex. The yeast equivalent of helix IIC (Hilliker *et al*, 2007; Perriman and Ares, 2007) did not form upon restructuring of stem-loop IIA (Figure 4). As helix IIC cannot form in U12 (Frilander and Steitz, 1999), the U2 equivalent in the minor spliceosome, these results suggest differences between yeast U2 structural dynamics during splicing, and those of both human U2 and U12. Alternatively, helix IIC may have escaped detection because it forms only transiently and cannot be captured in our stalled, purified spliceosomes, or it may only form in the intermediate B* complex that was not analysed here.

During step 1, U5 loop 1 is near the 5′ ss site (Sontheimer and Steitz, 1993) and the 5′ end of the intron (McConnell and Steitz, 2001) and therefore proximal to the branchpoint, consistent with the U5-intron crosslink identified in this study. Localization of the 3′ ss AG by the spliceosome was proposed to occur via a 5′ to 3′ scan of the PPY tract, starting from the branchpoint (Smith *et al*, 1989; Liu *et al*, 1997). The U5 loop 1 crosslink downstream of the branchpoint may be a snapshot of this scanning process and suggests that, by keeping exon 1 in a fixed position relative to the catalytic centre, U5 snRNP may be crucial in guiding the 3′ ss into the catalytic centre.

In contrast to group II introns, the core RNA network of the spliceosome functions within a protein scaffold in which PRP8 is thought to be the central platform. As the crystal structure of the yeast Prp8 protein in complex with the Aar2 assembly factor was recently solved (Galej *et al*, 2013), we investigated whether our model of the spliceosome's RNA core could be docked to the structure of the Prp8 protein. The crystal structure represents an assembly intermediate in which the so-called RNase H and JAB1/MPN domains are held in place mainly via interaction with Aar2. For the purpose of docking, we removed the Aar2 protein, removed the RNase H domain, as its exact position and functional location in the absence of Aar2 is uncertain, and ignored the JAB1/MPN domain which is not thought to contribute to the RNA binding surface (Galej *et al*, 2013). We used only a few constraints: (i) U5 must be on the side where it contacts the Prp8 protein (blue peptide in Figure 9B; Turner *et al*, 2006); (ii) the branchpoint must be close to the region to where it crosslinks to an unordered Prp8 loop (black loop in Figure 9B; Galej *et al*, 2013); and (iii) the ACAGA/pre-mRNA helix must be proximal to where the RNase H-like domain is positioned; a direct contact of the 5′ ss with the RNase H domain was not considered because a corresponding crosslink was only detected up to the B complex stage (Reyes *et al*, 1999). We found one orientation that satisfied these constraints (Figure 5C). In this orientation, the tip of the finger of the RT-like domain is very close to the catalytic centre. Indeed, a peptide (aa 871–970) containing this finger was found crosslinked to both the 5′ ss and the branchpoint

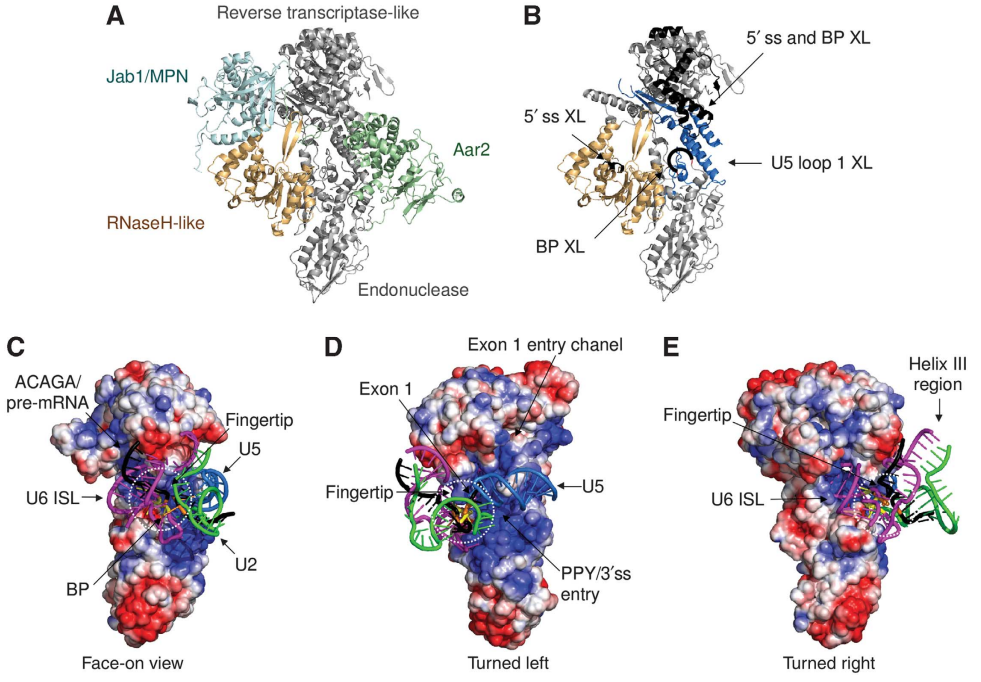

Figure 9 Docking of the RNA model onto the structure of the yeast Prp8 protein. (**A**) Structural overview of the Prp8 protein (Galej *et al*, 2013). (**B**) The landmarks used to dock the RNA model. The Aar2 protein and the Jab1/MPN domain are removed. (**C**–**E**) The docked RNA model is shown on the solvent-accessible surface of the combined reverse transcriptase and endonuclease domains coloured according to their electrostatic potential ($\pm 3\,kTe^{-1}$). Three orientations are shown: a face-on view (**C**) and rotations by 90° of this view to the right (**D**) or left (**E**). A movie animation is in Supplementary Data.

(Turner *et al*, 2006). Exon 1 can be placed in a basic cleft connected to a basic exit channel (Figure 9D). Due to the lack of constraints, the position of U5 is tentative (Figure 9D). The U6 ISL would be oriented into the cavity of the Prp8 protein (Figure 9E), where it is still available for interaction with other proteins. The large basic cavity between the endonuclease domain and the thumb of the RT domain would form a binding surface for helix Ia/b and provide sufficient surface for entry of the PPY tract. Given that we have not attempted to adjust the RNA model to the protein surface, it is interesting that we could dock it in a unique orientation to the Prp8 protein, lending support for validity of our RNA model. Together with future information about RNA–protein interactions, our 3D model and its interaction with Prp8 can now be tested, and thus serve as a basis for further exploration of spliceosomal structure and dynamics.

Materials and methods

Materials

Oligodeoxynucleotides used are complementary to the RNAs listed with beginning and ending nucleotides shown (with name in brackets): U2 snRNA: 75–95 (U2), 77–97 (K91), 149–169 (K31); U4 snRNA: 65–83 (B1); U5 snRNA: 83–103 (B2); U6 snRNA: 56–68 (U6), 80–100 (K34), 90–103 (U689); PM5 pre-mRNA: 281–262 (RTacaga281_262), 171–182 (Ex); MINXgg pre-mRNA: 229–249 (MINXggRT). Universal RNA adaptor is 5′-GACGACUCUAUAGUGU CACCUAAAU-3′, with its complementary oligodeoxynucleotide: 5′-ATTTAGGTGACACTATAG-3′ (URA, positions 8–25). DMS was purchased from Fluka, CMCT from Sigma, and kethoxal from Research Organics.

Purification of spliceosomal complexes

Spliceosomes were assembled *in vitro*, and purified by gradient centrifugation and subsequent MS2-affinity selection using amylose beads (New England Biolabs) essentially as described earlier (Deckert *et al*, 2006; Bessonov *et al*, 2008, 2010; Hartmuth *et al*, 2012), using the PM5, PM5–20, or MINXgg (Golas *et al*, 2010) pre-mRNAs.

Chemical modification and primer extension

Chemical modifications were performed essentially as described elsewhere (Hartmuth *et al*, 1999; Rasche *et al*, 2012) using 1–3 pmol of native complex. To analyse the extreme 3′ end of U6, cyclic 3′ terminal phosphates were removed (Cameron and Uhlenbeck, 1977), the universal RNA adaptor was added using T4 RNA ligase (New England Biolabs) and 3′-extended U6 molecules were gel purified. Sequence-marker templates were also prepared in this way from U6 prepared by transcription *in vitro*. Primer extensions were performed as described (Hartmuth *et al*, 1999) and quantified by phosphorimaging (Typhoon 8600). At least three independent experiments were quantified, using the Quantity One program (Bio-Rad) (see Supplementary Figures S2B for details). Signals were divided into three classes according to the fold increase observed: 1.1- to 2.4-fold increase: weak; 2.5- to 3-fold increase: medium; greater than 3-fold increase: strong.

Psoralen crosslinking and analysis of crosslinks

Purified spliceosomes (2–5 pmol) were crosslinked, and crosslinks were analysed by gel electrophoresis (5% polyacrylamide/8 M urea) and northern blotting as described previously (Behzadnia *et al*, 2007). RNase H digestion of total crosslinked RNA from the C complex was performed essentially as described (Behzadnia *et al*, 2007), except that RNase H was from NEB. Crosslinked species were purified from denaturing gels identical to those shown in Figure 5. RNA from the sections indicated was recovered and analysed by primer extension, with primers B2 (U5), K91 (U2), or K34 (U6).

Model building

Initial trials on the feasibility of building a group II intron homology model of the RNA network of the spliceosome were performed with both the Manip program (Massire and Westhof, 1998) and the newly devised program Assemble (Jossinet *et al*, 2010). Refinement was performed with NUCLIN-NUCLSQ and the model was checked for consistency using built-in routines (Westhof *et al*, 1985). These routines are now part of the Assemble suite (Jossinet *et al*, 2010). The homologous core (red area in Figure 8A) was modelled on coordinates from 3G78 (Toor *et al*, 2008; Wang, 2010), using the following mappings (given as *O. iheyensis* (Oi) to spliceosome): Oi 3–5 to pre-mRNA 1–3; Oi 286–289 to U6 44–47; Oi 358–366 to U6 53–61; Oi 375–381 to U6 72–78; and Oi 383–385 to U2 20–21. Helix Ia and Ib was modelled on 2AWB (Schuwirth *et al*, 2005): 2AWB 29–33 to U6 51–55 and 2AWB 49–57 to U2 20–29. Figures were prepared with Pymol (Schrödinger, LLC). Docking was performed using Hex 6.3 (http://hex.loria.fr/; Ritchie and Kemp (2000)). Initial orientations satisfying crosslink constraints were chosen to orient the receptor (Prp8 from pdb code 4I43) and the ligand (RNA model). The twist angle range was set to 30° with a step size of 7.5° and distance range was set to 5 Å. The electrostatic surface potential was calculated using the adaptive Poisson-Boltzmann solver (Baker *et al*, 2001) implemented in Pymol.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Olexandr Dybkov for help and for discussions, Fabrice Jossinet for on-site insight into Assemble, the members of the Lührmann laboratory for discussions, and Cindy L Will and Berthold Kastner for many helpful and critical comments on the revision. We thank Thomas Conrad for expert HeLa cell culture maintenance and Kami Kohansal for reliable nuclear extract preparation. This work was supported by the Deutsche Forschungsgemeinschaft (grant SFB 860 to RL) and L'Agence nationale de la recherche (grant ANR-10-BLAN-1502-02 'GRP2CONF' to EW).

Author contributions: MA and SB prepared the spliceosomes and conducted the RNA analysis. MA, KH, and RL designed the experiments. MA, KH, and RL analysed the data. KH, ZM, and EW created the model and performed the docking. KH, MA, and RL wrote the manuscript with input from all authors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98:** 10037–10041

Behzadnia N, Golas MM, Hartmuth K, Sander B, Kastner B, Deckert J, Dube P, Will CL, Urlaub H, Stark H, Lührmann R (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J* **26:** 1737–1748

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

Bessonov S, Anokhina M, Krasauskas A, Golas MM, Sander B, Will CL, Urlaub H, Stark H, Lührmann R (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16:** 2384–2403

Bessonov S, Anokhina M, Will CL, Urlaub H, Lührmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452:** 846–850

Burke JE, Sashital DG, Zuo X, Wang YX, Butcher SE (2012) Structure of the yeast U2/U6 snRNA complex. *RNA* **18:** 673–683

Cameron V, Uhlenbeck OC (1977) 3′-Phosphatase activity in T4 polynucleotide kinase. *Biochemistry* **16:** 5120–5126

Cech TR (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* **44:** 207–210

Chan RT, Robart AR, Rajashankar KR, Pyle AM, Toor N (2012) Crystal structure of a group II intron in the pre-catalytic state. *Nat Struct Mol Biol* **19:** 555–557

Chanfreau G, Jacquier A (1994) Catalytic site components common to both splicing steps of a group II intron. *Science* **266:** 1383–1387

Das R, Zhou Z, Reed R (2000) Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell* **5:** 779–787

Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, Stark H, Urlaub H, Lührmann R (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* **26:** 5528–5543

Dix I, Russell CS, O'Keefe RT, Newman AJ, Beggs JD (1998) Protein-RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *RNA* **4:** 1675–1686

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res* **15:** 9109–9128

Epstein P, Reddy R, Henning D, Busch H (1980) The nucleotide sequence of nuclear U6 (4.7 S) RNA. *J Biol Chem* **255:** 8901–8906

Fabrizio P, Abelson J (1992) Thiophosphates in yeast U6 snRNA specifically affect pre-mRNA splicing *in vitro*. *Nucleic Acids Res* **20:** 3659–3664

Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, Lührmann R (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36:** 593–608

Fortner DM, Troy RG, Brow DA (1994) A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev* **8:** 221–233

Frilander MJ, Steitz JA (1999) Initial recognition of U12-dependent introns requires both U11/5′ splice-site and U12/branchpoint interactions. *Genes Dev* **13:** 851–863

Galej WP, Oubridge C, Newman AJ, Nagai K (2013) Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493:** 638–643

Golas MM, Sander B, Bessonov S, Grote M, Wolf E, Kastner B, Stark H, Lührmann R (2010) 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Mol Cell* **40:** 927–938

Gozani O, Feld R, Reed R (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* **10:** 233–243

Hartmuth K, Raker VA, Huber J, Branlant C, Lührmann R (1999) An unusual chemical reactivity of Sm site adenosines strongly correlates with proper assembly of core U snRNP particles. *J Mol Biol* **285:** 133–147

Hartmuth K, van Santen MA, Odenwälder P, Lürhmann R (2012) Assembly and isolation of spliceosomal complexes *in vitro*. In *Alternative pre-mRNA Splicing: Theory and Protocols*, Stamm S, Smith C, Lührmann R (eds), 1st edn, pp 331–340. Weinheim, Germany: Wiley-VCH

Hausner TP, Giglio LM, Weiner AM (1990) Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev* **4:** 2146–2156

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

Hilliker AK, Staley JP (2004) Multiple functions for the invariant AGC triad of U6 snRNA. *RNA* **10:** 921–928

Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26:** 2057–2059

Keating KS, Toor N, Perlman PS, Pyle AM (2010) A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16:** 1–9

Kim SH, Lin RJ (1996) Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol Cell Biol* **16:** 6810–6819

Lardelli RM, Thompson JX, Yates 3rd JR, Stevens SW (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16:** 516–528

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

Lin PC, Xu RM (2012) Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO J* **31:** 1579–1590

Liu ZR, Laggerbauer B, Lührmann R, Smith CW (1997) Crosslinking of the U5 snRNP-specific 116-kDa protein to RNA hairpins that block step 2 of splicing. *RNA* **3:** 1207–1219

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

Massire C, Westhof E (1998) MANIP: an interactive tool for modelling RNA. *J Mol Graph Model* **16:** 197–205 255–257

McConnell TS, Steitz JA (2001) Proximity of the invariant loop of U5 snRNA to the second intron residue during pre-mRNA splicing. *EMBO J* **20:** 3577–3586

Michaud S, Reed R (1993) A functional association between the 5′ and 3′ splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev* **7:** 1008–1020

Michel F, Costa M, Westhof E (2009) The ribozyme core of group II introns: a structure in want of partners. *Trends Biochem Sci* **34:** 189–199

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

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

Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121:** 1005–1016

O'Keefe RT, Newman AJ (1998) Functional analysis of the U5 snRNA loop 1 in the second catalytic step of yeast pre-mRNA splicing. *EMBO J* **17:** 565–574

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

Ohrt T, Prior M, Dannenberg J, Odenwälder P, Dybkov O, Rasche N, Schmitzová J, Gregor I, Fabrizio P, Enderlein J, Lührmann R (2012) Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* **18:** 1244–1256

Padgett RA, Podar M, Boulanger SC, Perlman PS (1994) The stereochemical course of group II intron self-splicing. *Science* **266:** 1685–1688

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

Rasche N, Dybkov O, Schmitzova J, Akyildiz B, Fabrizio P, Lührmann R (2012) Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. *EMBO J* **31:** 1591–1604

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

Rhode BM, Hartmuth K, Westhof E, Lührmann R (2006) Proximity of conserved U6 and U2 snRNA elements to the 5′ splice site region in activated spliceosomes. *EMBO J* **25:** 2475–2486

Ritchie DW, Kemp GJ (2000) Protein docking using spherical polar Fourier correlations. *Proteins* **39:** 178–194

Ruskin B, Green MR (1985) An RNA processing activity that debranches RNA lariats. *Science* **229:** 135–140

Ruskin B, Krainer AR, Maniatis T, Green MR (1984) Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. *Cell* **38:** 317–331

Saintomé C, Clivio P, Favre A, Fourrey J-L, Riche C (1996) RNA photolabeling mechanistic studies: X-ray crystal structure of the photoproduct formed between 4-thiothymidine and

adenosine upon near UV irradiation. *J Am Chem Soc* **118:** 8142–8143

Sashital DG, Cornilescu G, McManus CJ, Brow DA, Butcher SE (2004) U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat Struct Mol Biol* **11:** 1237–1242

Schmitzova J, Rasche N, Dybkov O, Kramer K, Fabrizio P, Urlaub H, Lührmann R, Pena V (2012) Crystal structure of Cwc2 reveals a novel architecture of a multipartite RNA-binding protein. *EMBO J* **31:** 2222–2234

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science* **310:** 827–834

Ségault V, Will CL, Polycarpou-Schwarz M, Mattaj IW, Branlant C, Lührmann R (1999) 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* **19:** 2782–2790

Sharp PA (1985) On the origin of RNA splicing and introns. *Cell* **42:** 397–400

Sharp PA (1988) RNA splicing and genes. *JAMA* **260:** 3035–3041

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

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

Smith CW, Porro EB, Patton JG, Nadal-Ginard B (1989) Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. *Nature* **342:** 243–247

Smith DJ, Konarska MM (2009) A critical assessment of the utility of protein-free splicing systems. *RNA* **15:** 1–3

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

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

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

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

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

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

Teigelkamp S, McGarvey M, Plumpton M, Beggs JD (1994) The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. *EMBO J* **13:** 888–897

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

Turner IA, Norman CM, Churcher MJ, Newman AJ (2006) Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12:** 375–386

Venditti V, Clos 2nd L, Niccolai N, Butcher SE (2009) Minimum-energy path for a U6 RNA conformational change involving protonation, base-pair rearrangement and base flipping. *J Mol Biol* **391:** 894–905

Wahl MC, Will CL, Lührmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701–718

Wang J (2010) Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Acta Crystallogr D Biol Crystallogr* **66:** 988–1000

Warkocki Z, Odenwalder P, Schmitzova J, Platzmann F, Stark H, Urlaub H, Ficner R, Fabrizio P, Lührmann R (2009) Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat Struct Mol Biol* **16:** 1237–1243

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

Westhof E, Dumas P, Moras D (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J Mol Biol* **184:** 119–145

Westhof E, Masquida B, Jossinet F (2011) Predicting and modeling RNA architecture. *Cold Spring Harb Perspect Biol* **3:** pii: a003

Yan D, Ares Jr M (1996) Invariant U2 RNA sequences bordering the branchpoint recognition region are essential for interaction with yeast SF3a and SF3b subunits. *Mol Cell Biol* **16:** 818–828

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

