Spliceosome structure: Piece by piece

Dustin B. Ritchie, Matthew J. Schellenberg, Andrew M. MacMillan*

Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

ARTICLE INFO

Article history:
Received 12 May 2009
Received in revised form 22 August 2009
Accepted 27 August 2009
Available online 3 September 2009

Keywords:
Spliceosome
Structure
X-ray
NMR
EM

ABSTRACT

Processing of pre-mRNAs by RNA splicing is an essential step in the maturation of protein coding RNAs in eukaryotes. Structural studies of the cellular splicing machinery, the spliceosome, are a major challenge in structural biology due to the size and complexity of the splicing ensemble. Specifically, the structural details of splice site recognition and the architecture of the spliceosome active site are poorly understood. X-ray and NMR techniques have been successfully used to address these questions defining the structure of individual domains, isolated splicing proteins, spliceosomal RNA fragments and recently the U1 snRNP multiple-protein-RNA complex. These results combined with extant biochemical and genetic data have yielded important insights as well as posing fresh questions with respect to the regulation and mechanism of this critical gene regulatory process.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The split gene structure of eukaryotes, in which protein-coding exon sequences are separated by non-coding intron sequences, is replicated at the level of the transcribed pre-messenger RNA (pre-mRNA). Intron excision and exon ligation occur concurrently in a process referred to as pre-mRNA splicing which is both a fundamental gene regulatory mechanism and a source of proteome diversity in higher eukaryotes [1–3].

The chemistry of pre-mRNA splicing involves two sequential transesterification reactions (Fig. 1). In the first step, the 2′ hydroxyl of a conserved adenosine within the intron carries out a nucleophilic attack at the 5′ splice site to generate a free 5′ exon and a cyclic (lariat) intermediate containing a 2′–5′ phosphodiester branch. Attack of the free 5′ exon at the 3′ splice site then yields ligated exons and the lariat intron product (Fig. 1A and B). These two steps of splicing are catalyzed by a massive ribonucleoprotein (RNP) complex referred to as the spliceosome that is reminiscent in terms of its large size and RNA-protein composition with the ribosome [1–3]. Indeed, the advancement of our understanding of the ribosome by high resolution structure determination [4,5] suggests that a similar structural analysis of the spliceosome is a worthwhile goal. Such an achievement would greatly enhance our understanding of splicing on multiple levels including both the regulation and basic chemical mechanism of pre-mRNA splicing. Complicating this endeavor is the complexity of the splicing machinery not only in terms of the sheer number of its constituents but also the nature of their interaction and dynamic association with each other and pre-mRNA substrate.

2. Spliceosome assembly and rearrangement

The spliceosome includes the U1, U2, and U4/U6–U5 snRNPs (small nuclear ribonucleoprotein particles) each containing a unique RNA and associated proteins. A central organizing feature of each snRNP is the set of seven Sm core proteins (LSm for U6 snRNP) that recognize the Sm binding site on each snRNA [6]. Estimates of the number of spliceosomal proteins vary but mass spectrometric analysis of affinity purified spliceosomes have suggested that upwards of 300 polypeptides may be associated with the splicing machinery [7–9]. Assembly of the spliceosome from smaller subunits is an ATP dependent process templated by the pre-mRNA substrate and is directed by conserved sequences at and proximal to the splice sites within the intron (Fig. 1C).

In the canonical pathway, discrete steps of spliceosome assembly through E (CC in yeast), A, B, and C complexes have been characterized [10]. Commitment of a pre-mRNA to splicing involves the ATP independent formation of the early (E) or commitment complex on the RNA substrate. This complex includes U1 snRNP, tightly associated with the 5′ splice site, as well as non-snRNP protein factors. In metazoans, these include the heterodimer U2AF, containing large and small subunits, which binds to the polypyrimidine tract and 3′ splice site and SF1 which recognizes the branch region [11,12]. The A complex is formed in part by the stable, ATP dependent, association of U2 snRNP with the pre-mRNA; a duplex formed between U2 snRNA and the pre-mRNA branch region bulges out the branch adenosine specifying it as the nucleophile for the first transesterification [13]. Association of the U4/U6–U5 tri-snRNP with the A complex produces B complex which undergoes a series of rearrangements to yield C complex, the mature spliceosome. These rearrangements include displacement of U1 snRNP at the 5′ splice site by U6 snRNP, the disruption of U4/U6 snRNA base-pairing, and the formation of a U2/U6 snRNA structure which is believed to form the active site of the spliceosome [2,3].

* Corresponding author.
E-mail address: andrew.macmillan@ualberta.ca (A.M. MacMillan).

1874-9399/– see front matter © 2009 Elsevier B.V. All rights reserved.
doI:10.1016/j.bbagrm.2009.08.010
3. Approaches to characterization of spliceosomal structure

Over and above non-trivial issues of abundance and purification, the dynamic nature of the spliceosome and the substrate dependence of its assembly have posed major difficulties in approaching structural studies. Current high resolution structural analysis of the spliceosome has relied primarily on a dissection of the complex into functionally important subunits amenable to analysis by X-ray or NMR (Table 1). With one exception, structural descriptions of higher order complexes have been restricted to EM studies as outlined here (Table 2).

This review highlights advances in our understanding of spliceosomal structure with an emphasis on what has been learned with respect to splice site recognition during the dynamic process of spliceosome assembly as well as the catalytic components at the heart of the splicing machinery.

4. Early recognition of pre-mRNA by spliceosomal factors: the E complex

4.1. Polypyrimidine tract recognition by U2AF

The U2 auxiliary factor (U2AF) is a heterodimer consisting of 55 and 35 kDa subunits. The large subunit (U2AF65) binds the polypyrimidine tract at the 3′ splice site, interacts with other E complex spliceosomal proteins [14] and, later, the U2 snRNP component SF3b155 [15]. U2AF65 contains an N-terminal RS domain followed by three regions originally described as RRMs (RNA recognition motifs). However, only the first two of these domains are involved in RNA binding; the third represents a general protein interaction domain referred to as a UHM (U2AF homology motif; see below).

An RRM is a nucleic acid binding domain found in all organisms and typically binds single-stranded RNA, usually by recognition of a specific nucleotide sequence. RRMs are a subclass of the ferredoxin fold [16] that contain a four-stranded β-sheet buttressed by two α-helices in a β-α-β-α-β-α arrangement; they are distinguished by the presence of two amino acid motifs, RNP1 and RNP2, featuring conserved aromatic residues. High resolution structures of RRM domains both alone and bound to RNA have been described [17–21]. These reveal that single-stranded RNA typically binds on the face of the β-sheet, and that the two RNP motifs within the β-sheet are important for this interaction [22]. In particular, a tyrosine or phenylalanine within the RNP motif makes a stacking interaction with bound nucleotidyl bases while the identity of the nucleotide is determined by hydrogen bonding interactions. Structural analyses of RNA binding by the polypyrimidine tract binding protein (PTB) and the alternative splicing factor Fox-1 reveal further complexity in RNA recognition by RRM domains [23,24]. These include the substitution of RNP aromatic·RNA contacts with a separate set of hydrophobic interactions [23] and distinct features with respect to the participation of RRM loops in RNA binding [24].
Proper recognition of the polypyrimidine tract is essential for correct identification of the 3' splice site; however, this sequence is of a variable length, and often interrupted by purine nucleotides. Analysis of the recent crystal structure of U2AF65 RRM1.2 bound to a polypyrimidine RNA as well as the high resolution structure of RRM1 alone suggests how U2AF65 is able to bind these disparate sequences [25,26]. Each RRM interacts with three or four uridine nucleotides; a combination of hydrogen bonding through waters and conformationally flexible amino acid residues would allow hydrogen bonding interactions with the occasional non-uridine nucleotide in a target sequence. The crystal lattice of U2AF65 bound to RNA has an unexpected structure wherein the RNA is bound by both RRM1 and RRM2 but with each domain contributed from a different protein molecule. Based on this, Kielkopf and coworkers propose a model for the U2AF65 pyrimidine tract interaction in which the RRMs are in close proximity in a relatively condensed structure (Fig. 2A), Solution data of the protein alone obtained by SAXS [27] and NMR (Kent, Sprycopoulos, and MacMillan unpublished) are more consistent with an extended structure for RRM1–RRM2. A model of RNA binding consistent with these observations may also be derived from the crystal structure (Fig. 2B). An intriguing possibility is that the two models represent distinct binding modes to accommodate pyrimidine tracts of varying length. This may be an important aspect of U2AF65's function in bending the pre-mRNA to bring the branch point sequence and 3' splice site in close proximity to each other as part of an early organization of the pre-mRNA substrate during spliceosome assembly [28].

Early recognition of the 3' splice site AG dinucleotide is mediated by U2AF35, the small subunit of the U2AF heterodimer [29–31], but there is no high resolution structural data with respect to this interaction. This splice site recognition is coupled to pyrimidine tract binding by virtue of the U2AF35–U2AF65 pairing which is characterized by the tight interaction of a peptide from U2AF65 with the U2AF35 UHM.

4.2. Protein recognition mediated by U2AF homology motifs

In contrast to classical RRMs, U2AF homology motifs (UHM) have been shown to function as a protein-binding module [32]. The RNP1 and RNP2 sequence motifs are not conserved in these domains and the observation that they do not bind RNA is further explained by the occlusion of the canonical RNA binding surface by a C-terminal extension of the RRM positioned over the β-sheet as well as the overall negative charge of the resulting surface.

Table 1

<table>
<thead>
<tr>
<th>PDB</th>
<th>Structure</th>
<th>Function</th>
<th>Method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2G4B</td>
<td>U2AF65/PPT complex</td>
<td>early 3' splice site recognition/choice</td>
<td>X-ray</td>
<td>[25]</td>
</tr>
<tr>
<td>100P</td>
<td>U2AF65/RM3/N-terminal SF1 peptide complex</td>
<td>early 3' splice site recognition/choice</td>
<td>NMR</td>
<td>[14]</td>
</tr>
<tr>
<td>1JMT</td>
<td>core U2AF65/U2AF35 heterodimer</td>
<td>coupling PPT-3' splice site recognition</td>
<td>X-ray</td>
<td>[33]</td>
</tr>
<tr>
<td>1KIG</td>
<td>SF1/branch site complex</td>
<td>branch point sequence recognition</td>
<td>NMR</td>
<td>[38]</td>
</tr>
<tr>
<td>3CW1</td>
<td>U1 snRNP</td>
<td>5' splice site recognition</td>
<td>X-ray</td>
<td>[44]</td>
</tr>
<tr>
<td>1H0K, 1LMV, 3CGQ</td>
<td>branchpoint/U2 snRNA duplex</td>
<td>1st step nucleophile selection</td>
<td>X-ray, NMR</td>
<td>[54–56]</td>
</tr>
<tr>
<td>2Z9D</td>
<td>SF3b14-SF3b155 peptide complex</td>
<td>recognition branch point adenosine</td>
<td>X-ray</td>
<td>[60]</td>
</tr>
<tr>
<td>1XHP</td>
<td>U2AF65 snRNA extended DSL</td>
<td>Mg2+ coordination; 1st step</td>
<td>NMR</td>
<td>[70]</td>
</tr>
<tr>
<td>3EE6, 3ENB, 3E9I, 3E9O, 3E9P</td>
<td>PRP8 domain IV core</td>
<td>structuring spliceosome active site</td>
<td>X-ray</td>
<td>[80-82]</td>
</tr>
</tbody>
</table>

PPT: polypyrimidine tract.

* See text for additional structures.

Table 2

<table>
<thead>
<tr>
<th>Accession*</th>
<th>Structure</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>U1 snRNP</td>
<td>[42]</td>
</tr>
<tr>
<td>1325</td>
<td>A complex</td>
<td>[48]</td>
</tr>
<tr>
<td>1096</td>
<td>U11/U12 di-snRNP</td>
<td>[49]</td>
</tr>
<tr>
<td>1043</td>
<td>SF3b155</td>
<td>[51]</td>
</tr>
<tr>
<td>1257</td>
<td>U4/U6-US tri-snRNP</td>
<td>[61]</td>
</tr>
<tr>
<td>1258</td>
<td>U5 snRNP</td>
<td>[61]</td>
</tr>
<tr>
<td>1259</td>
<td>U6 di-snRNP</td>
<td>[61]</td>
</tr>
<tr>
<td>1066</td>
<td>B complex</td>
<td>[62,63]</td>
</tr>
<tr>
<td>1062</td>
<td>C complex</td>
<td>[87]</td>
</tr>
</tbody>
</table>

* Indicated structures are available at The Electron Microscopy Data Bank (EMDB) at EBI: http://www.ebi.ac.uk/pdbe/emdb/.

4.3. SF1 and branch point sequence recognition

The branch region of the pre-mRNA substrate, like the splice sites, is recognized several times during the course of spliceosome assembly. In the commitment complex, the association of SF1 with the branch point sequence is mediated by the protein's KH domain.

The hnRNP K homology (KH) domain is a ~70 amino acid single-stranded nucleic acid binding domain. Akin to the RRM, it is found in a diverse variety of organisms and is typically present in one or more copies within a protein. The KH domain family is further diversified by the type I eukaryotic and type II prokaryotic variants [37]. The type I KH domain folds into a three-stranded antiparallel β-sheet abutted by three α-helices. This structure represents a binding cleft which typically interacts with four nucleotides in a single-stranded extended conformation.

The NMR structure of SF1 bound to a 10 nucleotide branch sequence RNA shows that it associates with the pre-mRNA via an extended type 1 KH domain in such a way that it identifies the branch sequence by hydrogen bonding with the nucleotides — especially with

D.B. Ritchie et al. / Biochimica et Biophysica Acta 1789 (2009) 624–633
4.3.2. U1 snRNA

In the crystal structure of human U1 snRNP, the U1 snRNA is clearly defined with an overall structure similar to that proposed based on the cryo-EM study [42]. A four-helix junction composed of helices I–III and H of U1 snRNA, the threading of the RNA through the Sm core, and helix IV are all apparent in the structure (Fig. 4A). Interestingly, within the crystal lattice, the 5′ end of U1 snRNA, which base pairs with the 5′ splice site of the pre-mRNA during spliceosome assembly, interacts with the equivalent RNA of an adjacent monomer. This forms a duplex which is directly analogous to the U1 snRNA-splice site pairing and is useful for interpretation of the function of the U1C protein.

4.3.3. The Sm core proteins

The Sm core proteins form a heptameric ring structure and the seven nucleotides of the Sm binding site are threaded through this ring. The mounting of the U1 snRNA four-helix junction on one side of the Sm core is mediated by interactions between two of the helices from the junction with the N-terminal helical extensions of Sm D2 and Sm B (Fig. 4B). Thus, the structure suggests that the Sm core acts as a platform for complex RNA structures; this is likely to be a common feature of the other spliceosomal snRNPs as well because complex RNA structures are always found at the 5′ side of the Sm core binding site [44].

4.3.4. UIC protein

The human UIC protein is known to contain a zinc finger structure [40], and yeast UIC has been proposed to directly interact with the 5′ splice site [45]. Integration of U1C into the U1 snRNP particle is known to be dependent on the N-terminal region of U1-70K and the Sm core domain [46]. The structure shows an extended segment, helix B, of U1C is responsible for interaction with these proteins (Fig. 4B). As described above, U1 snRNA-splice site pairing is mimicked in the crystal lattice and the zinc finger of U1C can be seen interacting with this duplex. Due to the limited resolution of this structure, specific details of this interaction are not visible. Nonetheless, U1C is positioned along the minor groove of the RNA duplex, including the location corresponding to the base pairs with the invariant GU dinucleotide which defines the 5′ splice site. This suggests that the function of U1C may be to communicate to the rest of the snRNP that the correct 5′ splice site interaction has been formed.

4.3.5. U1-70K

The U1-70K protein contains a central RRM domain which binds the end of SL1 in U1 snRNP and density corresponding to this interaction is seen in the crystal structure. The N-terminal 100 amino acids of this protein has no predicted domain structure and can be seen to extend along SL1, around the Sm core to where the N-terminus of U1-70K forms the binding site for U1C. This striking encirclement of the U1 snRNP likely functions to stabilize the snRNP structure like wrapping a string around a package (Fig. 4B).
U1C in isolation, although similar, is not identical and difficult to interpret in the absence of the U1 snRNA duplex observed in the U1 snRNP structure, and of course no meaningful structural information for the N-terminus of U1-70K could be attained because its conformation is dependent on its context within the larger particle.

The crystal structure of the U1 snRNP also allows re-interpretation of older biochemical data such as hydroxyl radical footprinting data showing that the 5′ end of U2 snRNA is in close proximity to U1 snRNP [47]. Mapping of strong cleavage sites from that study suggests that the 5′ end of U2 snRNA is very near the center of the four-helix junction of U1 snRNA which is considerably less splayed than proposed in the original model.

5. Structures related to the A complex

5.1. EM structures

An EM structure of affinity purified spliceosomal A complex has been described [48]. Owing to the low resolution of the model (~40 Å) features corresponding to known proteins cannot be identified; the overall shape is slightly elongated, reminiscent of the cryo-EM structure of the U11/U12 di-snRNP [49] (the U1 and U2 homologs from the minor spliceosome the cellular machinery responsible for splicing of the rare (less than 1%) class of U12 introns with atypical splice sites) [50].

A salt-dissociable component of the 17S U2 snRNP, SF3b, can be purified from HeLa cell nuclear extracts and has been shown to be necessary for pre-mRNA splicing. SF3b is a ~450-kDa multi-protein complex containing seven polypeptides ranging in molecular weight from 10 to 150 kDa; this includes the factors SF3b10, SF3b14, SF3b14b, SF3b49, SF3b130, SF3b145, and SF3b155. The cryo-EM structure of SF3b has also been described [51]. The locations of SF3b49 and SF3b14 could be identified based on the fact that they contain RRM domains. The C-terminal region of SF3b155 contains α-helical HEAT repeats, visible in the EM structure, which form a ladder structure and may represent a scaffold for assembly of other proteins within the particle.

5.2. U2 snRNA-branch point sequence duplex structures

Progression of spliceosome assembly to the A complex involves recruitment of U2 snRNP to the branch point sequence where the U2
snRNA forms an imperfect duplex that bulges out the branch adenosine and selects it as the nucleophile for the first transesterification of splicing [13]. An analogous bulged duplex structure is formed in the self-splicing group II introns [52,53] and is believed to be essential for catalysis. Several structures of a branch point sequence with a bulged adenosine have been described [54–56]. They differ slightly with respect to whether the preferred branch adenosine or the adjacent adenosine is bulged out as well as in the conformation of the bulged residue, which ranges from flipped out in an extended conformation to slightly extruded from the helix. These differing conformations imply a flexibility that is presumably fixed within the context of the spliceosome. However, it has also been shown that the conserved pseudouridine within U2 snRNA that lies opposite the branch nucleotide stabilizes the extruded conformation of the bulged adenosine, perhaps by extra hydrogen bonding to the N5 position of the pseudouridine [55,56].

5.3. Late branch adenosine recognition: SF3b14–SF3b155 structure

Formation of the bulged pre-mRNA·U2 snRNA duplex occurs within the context of an SF3b branch region interaction the function of which is poorly understood. A 14-kDa protein subunit of the SF3b particle, SF3b14 directly contacts the branch adenosine in the A complex [57,58] and this interaction persists into the mature spliceosome although recent work has shown that the intimate association of SF3b with the pre-mRNA is disrupted at or after the first step of splicing [59]. The 2.5–Å crystal structure of SF3b14 bound to a peptide from SF3b155 revealed the occlusion of the canonical RNA binding surface of the SF3b14 RRM by a C-terminal α-helical extension tightly associated with the peptide (Fig. 5A) [60]. Cross-linking suggests that the bulged branch adenosine is bound within a tight pocket on this surface stacking on the conserved tyrosine of RNP2 (Fig. 5B). The surface surrounding this pocket includes four basic residues (R24, R57, R96, and K100) which could plausibly interact with the phosphate backbone of the RNA duplex surrounding the bulged adenosine. The identities of R24 and R57 are highly conserved among SF3b14 orthologs but not between SF3b14 and other RMRs consistent with them being associated with SF3b14’s specific mode of RNA binding. Thus the SF3b14–SF3b155 interface establishes a unique surface for the recognition of the bulged duplex during spliceosome assembly.

6. Higher order spliceosomal complexes

6.1. Multi-snRNP complexes

Complex B formation is marked by the arrival of the U4/U6·U5 tri-snRNP to the pre-catalytic A complex. Similar to the U1 and U2 snRNPs, insights about interactions between tri-snRNP components have relied on high resolution structures of the individual components. EM structures of the U4/U6·U5 tri-snRNP, the individual U5 snRNP and the U4/U6 di-snRNP [61] as well as human spliceosomal B complex lacking U1 snRNP [62] and Drosophila B complexes [63] have been described. A full appreciation of the contribution of these structures to an understanding of spliceosome assembly dynamics awaits high resolution structures of the individual components or larger assemblies.

6.2. High resolution structures of RNA active site components

Once assembled the spliceosome active site is believed to be composed of a complex network of snRNA interactions. Speculation regarding the individual roles of RNA and protein in splicing catalysis has centered on the high evolutionary conservation between the group II self-splicing introns and the spliceosome. The chemistry of intron removal from group II RNAs is identical to that of pre-mRNA splicing indicating that the snRNAs of the spliceosome may be directly involved in catalysis. Indeed, similarities in the two systems extend to the reversibility of both transesterification reactions [64–66].

The catalytic strategy of group II introns most likely involves a two-metal ion mechanism as proposed for both the group II system as well as the spliceosomal reaction on the basis of the mechanism of DNA and RNA polymerases in phosphoryl-transfer reactions [67]. Domain 5 of group II introns is a metal-binding platform that coordinates Mg2+ ions that are likely responsible for activating the first step nucleophile as well as for stabilizing the oxanion leaving group during the first step of the splicing reaction [68]. In the 3.1 Å crystal structure of a group II intron from Oceanobacillus iheyensis, tertiary RNA contacts stabilize the arrangement of the juxtaposed exon-binding sequences 1 and 3 which interact with the 5′ and 3′ exons, respectively, over the bulge of domain 5 which binds the essential Mg2+ ions [68]. The two metal ions are 3.9 Å apart matching the ideal distance noted for the active sites of analogous protein enzymes [67].

The NMR data with respect to the yeast U2/U6 RNA duplex which mimics the snRNA arrangement at the spliceosome active site reveals

Fig. 5. Late branch adenosine recognition by SF3b14. (A) X-ray structure of complex between SF3b14 (grey) and SF3b155 amino acids 373–415 (green) [60] which occludes much of the canonical RRM RNA binding surface including Y61 of the RNP1 motif. The RNP2 Y22 forms the base of a pocket that crosslinking shows accommodates the bulged adenosine of the branch sequence·U2 snRNA duplex. (B) Surface representation of the SF3b14–SF3b155 complex showing Y22 exposed within the surface pocket surrounded by conserved basic residues; SF3b14 (light grey), SF3b155 (dark grey), Y22 (yellow), and R24, R57, R96, and K100 (blue).
a complex fold resembling the four-helix junction adopted by other RNAs such as the Hairpin ribozyme [69,70]. The intermolecular base-pairing between U2 and U6 forms helices I–III; a key feature specific to the U2/U6 complex is the U6 ISL, which forms through intramolecular base-pairing when U6 snRNA is unwound from the U4/U6 duplex. The identification of the U2/U6 complex as a four-helix junction has important implications for how several crucial RNA elements in the U2/U6 duplex are arranged for catalysis. Models of four-helix junction folds predict that the relative orientation of the U6 ISL with respect to the 5’ splice site in the U2/U6 structure is expected to juxtapose essential sequence elements within the ISL with the 5’ splice site and branchpoint sequence [70] analogous to the situation observed in the crystal structure of the O. iheyensis group II intron [68]. An unpaired residue, U80 in the yeast ISL, serves as a binding site for a catalytically essential metal [71] analogous to the metal–binding site reported for domain 5 of group II introns. Domain 5 and the U6 ISL are believed to be structurally and mechanistically analogous [72].

The solution structure of a U2/U6 model duplex includes an extension of the U6 ISL involving the invariant AGC triad of U6 snRNA which was previously believed to make base-pairing interactions with U2 snRNA residues to form U2/U6 helix Ib. The U2 snRNA residues instead participate in intramolecular base-pairing interactions to form part of U2 helix stem I. The extension of the U6 ISL to include the AGC triad increases its similarity with domain 5, as well as the similarity between the yeast U2/U6 structure and the structure previously predicted for the mammalian U2/U6 duplex [70].

Recent work challenges the relevance of the four-helix junction model of U2/U6 structure at the spliceosome active site [73]. As noted above, the four-helix model predicts the formation of an extended U6 ISL as opposed to the mutually exclusive U2/U6 helix Ib. Studies aimed at elucidating a role for helix Ib during splicing catalysis utilized mutant pre-mRNA substrates rate-limiting for either the first or second step of splicing. Mutations in U2/U6 that weaken either helix Ib or the four-helix junction conformation exacerbated the effect of mutant substrates rate-limiting for either step. Intriguingly, only restoration of the helix Ib conformation but not the four-helix junction eliminated this exacerbation of the splicing defect, providing compelling evidence that a U2/U6 structure that includes helix Ib is important at both steps of splicing catalysis [73]. This work does not rule out the possibility that a four-helix junction conformation exists at other stages of spliceosome assembly, including between the two steps of splicing.

In the crystal structure of the group II intron, tertiary interactions cause a bend in the domain 5 helix, as well as an unusual kink in the bulge region responsible for coordination of the magnesium ions. As well, a tertiary interaction between domain 5 and the linker between domains 2 and 3, J2/3, which is analogous to the highly conserved ACAGAGA sequence of U6 snRNA, brings together catalytically essential sequences (Fig. 6) [68]. These features are not present in domain 5 studied in isolation [52,74], so it seems likely that analogous structural features might be relevant to the U6 ISL in the context of the spliceosome perhaps mediated in this case by interactions with protein components.

6.3. High resolution structures of a protein active site component

The essential and very highly conserved U5 snRNP protein Prp8 has long been known to be intimately associated with key components of the spliceosome including the pre-mRNA substrate and snRNAs present at the active site [75]. A large number of mutant alleles of Prp8 have been characterized in yeast and the variety of associated phenotypes attest to the central role of this factor in spliceosome assembly and possibly catalysis [76]. These include suppression of the effects of hyper-stabilized U4/U6 as well as suppression of splicing defects due to mutations at both splice sites and the branch point sequence [77–79]. The clustering of these alleles, including those with distinct phenotypes, combined with interspecies sequence conservation has suggested the presence of two domains (3 and IV) that by virtue of these characteristics are believed to interact with the spliceosomal core.

Recently, three groups independently solved the crystal structure of the Prp8 domain IV core [80–82] a region previously demonstrated

![Group II intron domain 5](image1)

![U6 snRNA ISL](image2)

Fig. 6. Comparison of group II intron and spliceosomal active site RNA structures. The crystal structure of domain 5 from the Oenococcus iheyensis group II intron [68] (left) alongside the NMR structure of yeast U6 snRNA ISL (right) [70]. The location of two catalytic magnesium ions are shown as green spheres in the domain 5 structure and the U80 nucleotide implicated in magnesium binding by the U6 ISL is highlighted (green). Tertiary RNA contacts between domain 5 and the J2/3 region (red) important for catalysis are shown. The catalytic triad is shown for both domain 5 and U6 ISL (yellow).
by crosslinking to interact with the 5′ splice site in the spliceosome active site [83]. The structure of the core domain is bipartite consisting of an N-terminal sub-domain (amino acids 1769–1887) with an RNase H fold and a tightly packed C-terminal cluster of five helices (amino acids 1900–1990). The RNase H fold exhibits a characteristic five-stranded parallel/anti-parallel β-sheet, buttressed by two α-helices. This structural homology was not predicted by analysis of primary sequence because of a seventeen amino acid β-finger insertion (amino acids 1787–1803) that in itself is a striking feature of the core domain (Fig. 7A). The Prp8–5′ splice site interaction referred to above occurs before the first step of splicing in the B complex, and has been mapped to a short peptide sequence corresponding to amino acids 1894–1898 within domain IV by proteolytic methods [83]. The structure of the core domain shows that these amino acids correspond to the helix (α5) that forms the base of the channel separating the RNase H homology domain and the C-terminal helical cluster (Fig. 7A) [83].

The presence of an RNase H fold in the spliceosome active site has several important implications. The crystal structure of Bacillus halodurans RNase H (Fig. 7A) reveals an active site containing two metal-binding sites which coordinate functional Mg2+ ions ~4 Å apart with distinct mechanistic roles in RNA hydrolysis [84]; three putative metal-binding ligands corresponding to one of the two RNase H metal binding sites are conserved spatially, though not in primary sequence, within Prp8 domain IV. Although no metal was observed in the Prp8 structures, mutating the residues at this site has deleterious effects in yeast suggesting that a conserved and functionally important metal-binding site may exist in Prp8 (Fig. 7B). The magnesium bound at the analogous site in the RNase H crystal structure is responsible for stabilization of the leaving group in the cleavage reaction [84] and one could envision a similar functional role for a metal bound at this site of Prp8 in the spliceosome. Interestingly, the Prp8 site is blocked by a capping interacting with a conserved arginine side chain. This exactly mirrors the P. furiosus Argonaute PIWI domain in which an arginine residue is positioned in a similar fashion [85]; another example of this structural feature is found in the Tn5 transposase effectively blocking access to the site [86]. In the context of the spliceosome, interactions with other components, for example a phosphate in the backbone of a double-stranded RNA, could possibly reorient the arginine essentially freeing the site to allow a metal to bind. It is still an open question whether a Prp8 metal-binding site could have a functional role in the splicing reaction.

---

**Fig. 7.** Structure of a protein component of the spliceosomal active site. (A) Comparison of human Prp8 domain IV core with RNase H. Left: The X-ray structure of the domain IV core [81] contains an RNase H fold (yellow) and features a two strand β-finger (blue). The site of a crosslink to the 5′ splice site [83] is highlighted in red. Right: X-ray structure of an A-form duplex bound to B. halodurans RNase H [84]. (B) Alignment of Prp8 domain IV core (cyan) with RNase H (gold) reveals the conservation of a putative metal-binding site in Prp8 that corresponds to a functionally critical site in RNase H. The RNase H DDE triad consisting of residues D71, D132, and E109 binds Mg2+ (green) and corresponds with a DDT triad in Prp8 composed of residues D1781, D1782, and T1864 although no metal is observed in the Prp8 structure.
Apart from a possible functional metal-binding site, the PrP8 RNAseH fold provides a platform for interacting with RNA [81]. Gel mobility shift assays are strongly supportive of RNA binding by this domain and a modest though significant hierarchy of affinities for different RNAs was observed with the strongest binding interaction occurring with a model of the U2/U6 snRNA pairing. The demonstration that PrP8 interacts specifically with a U2/U6 RNA structure suggests an intimate interaction at different steps of spliceosome activation and catalysis between PrP8 and other key active site components. Perhaps PrP8 is mediating conformational changes of the U6 ISL similar to the tertiary RNA contacts in the group II intron. It is tempting to speculate that the U6 ISL can adopt a similar kinked and bent conformation to domain 5 mediated by PrP8 in the context of the activated spliceosome, and that similar to domain 5 these structural features are not observed when the U6 ISL is in isolation.

7. Structural studies of the catalytic spliceosome

While detailed pictures of the components behind spliceosome activation and catalysis are provided by high resolution structural techniques, EM has provided useful information about the relative locations of sub-complexes in the larger catalytic C complex which contains the U2, U5 and U6 snRNPs [87]. An EM structure of the spliceosome at ~30 Å resolution reveals a complex with dimensions of 270 × 240 Å arranged in three distinct domains. The C complex, therefore, is roughly the same size as purified tri-snRNPs, a surprising observation considering how many proteins have been identified as associated with the catalytic spliceosome [88]. In the C complex EM structure a roughly cylindrical domain is separated from two additional, distinct domains creating a large cavity in the centre of the structure. The open arrangement of the three domains has been interpreted to represent structural information and validation for X-ray based structure determination.

While detailed pictures of the components behind spliceosome activation and catalysis are provided by high resolution structural techniques, EM has provided useful information about the relative locations of sub-complexes in the larger catalytic C complex which contains the U2, U5 and U6 snRNPs [87]. An EM structure of the spliceosome at ~30 Å resolution reveals a complex with dimensions of 270 × 240 Å arranged in three distinct domains. The C complex, therefore, is roughly the same size as purified tri-snRNPs, a surprising observation considering how many proteins have been identified as associated with the catalytic spliceosome [88]. In the C complex EM structure a roughly cylindrical domain is separated from two additional, distinct domains creating a large cavity in the centre of the structure. The open arrangement of the three domains has been interpreted to represent structural information and validation for X-ray based structure determination.

8. Summary and perspectives

There has been considerable recent progress with respect to our understanding of components of spliceosome structure. EM techniques combined with immuno-labeling have generated envelopes that may be fitted with structures of RNA or protein fragments determined by X-ray or NMR. Ultimately these may serve to provide phasing information and validation for X-ray based structure determination. The culmination of efforts to understand higher order spliceosomal structure in the context of an RNP is represented in the landmark X-ray analysis of the U1 snRNP. Several features of this structure, especially the ordering of the N-terminus of U1-70K, highlight a lesson familiar from structural studies of the ribosome: while structural analyses of individual protein or RNA fragments represent an important advance in vitro, EMBO J. 15 (1996) 2565–2569.


