MAJOR MESSAGES

Nuclear pre-mRNA splicing occurs through two sequential phosphodiester trans-esterification reactions that do not result in the net loss or gain of phosphodiester bonds.

Group II ribozymes, a type of transposable element, and the spliceosome share an evolutionary ancestor.

The spliceosome, the enzyme responsible for nuclear pre-mRNA splicing, is assembled from sn-RNPs (small nuclear ribonucleoproteins) and non-snRNP proteins through a pathway characterized by massive changes in protein and RNA content prior to formation of the active enzyme.

Dramatic RNA rearrangements characterize spliceosome formation -- to a degree beyond that described for any other biological system.

The U2-U6 RNA complex coordinates two Mg++ ions to catalyze both chemical steps of pre-RNA splicing via a single active site. The second catalytic step is a pseudo-reversal of the first.

The most conserved protein in the spliceosome, Prp8, cradles the snRNAs and is evolutionarily related to Group II intron-encoded proteins which function to stabilize the Group II catalytic core.

CryoEM analysis of the spliceosomal active site is fully consistent with prior inferences from genetic and biochemical analysis. The core active site, once produced, is static. The branchpoint helix moves out after step 1 enabling the 3’ splice site to move in for step 2 catalysis.
Transcription is only the first step in gene expression

Unlike bacterial transcripts, eukaryotic mRNAs are capped, spliced, and polyadenylated. Nearly all human genes are alternatively spliced, producing more than one mRNA from a single gene. Some genes produce hundreds or thousands of isoforms. The smallest human intron is 44 bp. The largest is over 100 kb. 10% of tumors contain mutations in the enzyme required for splicing, the ‘spliceosome.’ Splicing has recently been shown to become defective during *C. elegans* aging and its improvement increases lifespan.

---

**Figure 6-21** A comparison of the structures of prokaryotic and eukaryotic mRNA molecules. (A) The 5’ and 3’ ends of a bacterial mRNA are the unmodified ends of the chain synthesized by the RNA polymerase, which initiates and terminates transcription at those points, respectively. The corresponding ends of an eukaryotic mRNA are formed by adding a 5’ cap and by cleavage of the pre-mRNA transcript near the 3’ end and the addition of a poly-A tail, respectively. The figure also illustrates another difference between the prokaryotic and eukaryotic mRNAs: bacterial mRNAs can contain the instructions for several different proteins, whereas eukaryotic mRNAs nearly always contain the information for only a single protein. (B) The structure of the cap at the 5’ end of eukaryotic mRNA molecules. Note the unusual 5’-to-5’ linkage of the 7-methyl G to the remainder of the RNA. Many eukaryotic mRNAs carry an additional modification: methylation of the 2’-hydroxyl group of the ribose sugar at the 5’ end of the primary transcript (see Figure 6-23).
RNA biochemistry covered in the Macromolecules Course:
Nuclear pre-mRNA splicing

Nuclear introns were discovered in adenovirus by hybridization of specific RNAs from infected cells to adenovirus DNA from virions. It was quickly realized that there is not a great deal of sequence information in mammalian introns: a GU dinucleotide was present at the beginning of introns and an AG at the end. *S. cerevisiae* introns (about 2% of genes have introns in this yeast) were found to have more information: a GUAUGU consensus at the 5’ splice site, a UACUAAC consensus near the 3’ end of the intron and a PyAG at immediately precedent the 3’ splice site. These elements are nearly invariant in *S. cerevisiae*.

Early work produced crude extract systems from HeLa cell nuclear and *S. cerevisiae* capable of splicing synthetic $^{32}$P-radiolabelled pre-mRNA produced by *in vitro* transcription using phage RNA polymerase. Analysis of the structure of the apparent intermediates and products indicated a two-step reaction which first produces a lariat-intermediate and a cleaved 5’ exon followed by formation of the ligated exons and release of the intron as a lariat-intron.
Origins
Why are there introns? Where did they come from?

Some bacteria (and organelles derived from bacteria) have rare introns called “Group II introns.” Remarkably they splice through the **EXACT SAME CHEMICAL MECHANISM** as eukaryotic introns.

Group II introns are RIBOZYMES (RNA enzymes) have five or six phylogenetically conserved domains. Domain I, the most complex domain organizes the folding of the intron and engages in numerous tertiary interactions. Domain IV plays no role in activity, but is sometimes interrupted by an open reading frame that encodes protein (IEP for ‘intron encoded protein’).
IEP-containing introns require the IEP for splicing. The IEP binds and stabilizes the active conformation of the RNA.

The RNA catalyzes splicing by positioning two Mg\(^{++}\) ions that stabilize the transition state of each trans-esterification reaction. A single active site catalyzes both steps. Step 2 is a pseudo-reversal of Step 1.
The spliceosome

Five facts about the spliceosome:

• It is large: The S. cerevisiae spliceosome contains 90 polypeptide chains and 5 snRNAs. The human spliceosome contains ~170 proteins and 5 snRNAs. Why so many pieces for a such a simple reaction?
• It undergoes a ridiculously elaborate assembly process for every single catalytic event. Eight ATPases are required. Why some much ATP?
• It is disassembled after each catalytic event. A single turnover ‘enzyme.’ Why so wasteful?
• At its core, it’s a ribozyme. What do the proteins do?
• It evolved from a Group II intron and it’s IEP.
• Each substrate is different -- there is no single invariant nucleotide in introns -- yet single nucleotide accuracy is needed.
• The spliceosome depots a protein complex after splicing called the Exon Junction Complex, at each exon junction. This protein “mark” is carried from the nucleus to the cytoplasm. Why mark exon junctions?

The snRNPs

Five evolutionarily-conserved small nuclear RNAs (snRNAs) are part of the spliceosome and are required for splicing in vitro and in vivo. It was hypothesized that they evolved from an errant Group II intron that invaded the nuclear genome near the time of the birth of the eukaryotic lineage. Each snRNA is complexed with a set of proteins to form a ribonucleoprotein (snRNP). While U1 and U2 exist as individual snRNPs, U4/U5/U6 is found as a single triple-snRNP. U4 and U6 can be separated from U5, but are tightly associated with each other even after deproteinization (heat can be used to separate the two).

Figure 2. Protein composition and snRNA secondary structures of the major human spliceosomal snRNPs. All seven Sm proteins (B/B’, D3, D2, D1, E, F, and G) or LSm proteins (Lsm2-8) are indicated by “Sm” or “LSm” at the top of the boxes showing the proteins associated with each snRNP. The U4/U6.U5 tri-snRNP contains two sets of Sm proteins and one set of LSm proteins.
Spliceosome assembly: the snRNP-o-centric view

Analysis of the time course of splicing complex assembly \textit{in vitro} by native gel electrophoresis suggested that the snRNA composition of the spliceosome is not static, but rather assembled de novo via a pathway: $\text{U1} \rightarrow \text{U1+U2} \rightarrow \text{U1+U2+U4/U5/U6} \rightarrow \text{U2/U5/U6} \rightarrow \text{catalysis}$. Inactivation of individual snRNAs indicated that the prior assembly of U1 and U2 snRNAs onto a pre-mRNA is required for the association of the U4/5/6 triple snRNP with the pre-mRNA.

Intronic sequences are recognized at least twice during spliceosome assembly
Summary of what recognizes which sequences within the intron before the first step of splicing

U6: the most conserved RNA in the spliceosome

The structure of the U4-U6 interaction was revealed by the cloning of the *S. cerevisiae* U6 snRNA gene and using phylogenetic comparison to identify compensatory base changes.
Genetic analysis showed that a region of U6 snRNA that base-pairs with U4 was much more sensitive to point mutations than the complementary residues in U4. This result and the apparent loss of U1 and U4 from splicing complexes prior to the appearance of intermediates and products suggested that U6 might engage in new interactions that are mutually-exclusive with the U4-U6 interaction.

Guesswork and genetic analysis revealed that this region of U6 initially base-paired to U4 forms a novel interaction with U2 adjacent to where U2 base-pairs with the intron branch point.

Combined with an intramolecular interaction found in the free U6 snRNP (the intramolecular stem-loop or ISL) and another U2-U6 interaction (helix II), a picture emerges of a presumptive RNA active site.
The structure bears a strong resemblance to Domains V and VI of Group II introns including the AGC triad.

These results imply at least three RNA helices were disrupted during the catalytic activation of the spliceosome: U4-U6 stem I, U5-U6 stem II, and the 5' stem-loop of U2 snRNA.
Rescuing the spliceosome

The ability to reconstitute splicing after degradation of endogenous U6 snRNA by the addition of synthetic U6 snRNA combined with phosphothioate/phosphorothiolate metal rescue experiments demonstrated that the U2-U6 complex functions to position two catalytic Mg++ ions. Moreover, the phosphates that ligand Mg++ occur in positions analogous to those that ligand Mg++ in Domain V of the Group II crystal structured.
2015-present: high-resolution structures of assembled spliceosomes
The first high resolution (3-4 Å) cryoEM structure of spliceosome (the *S. pombe* intron-lariat spliceosome or “ILS”) confirms that the U2-U6 complex coordinates two divalent cations to produce a Group II-like active site.
Structures of spliceosomal intermediates show that the U2-U6 active site is essentially static. They show a rotation of the branch helix after step 1 that moves it out of the way for the 3’ splice site to enter the active site. The U-rich U5 loop binds the 5’ exon and 3’ exons during step 2.

Origin of the most conserved spliceosomal protein component
Crosslinking and genetic suppression studies point to Prp8, the most conserved protein in the spliceosome as being a component of the spliceosome closely associated with the snRNAs and the pre-mRNA substrate. The crystal structure of Prp8, which is part of U5 snRNP indicates that it is highly related to IEPs. IEPs, in addition to promoting splicing of Group II introns are also involved in their mobility. Indeed, Group II introns are retrotransposons. Thus IEPs encode enzymes involved in retrotransposon replication such as a reverse transcriptase and DNA endonuclease domains. Prp8 contains a reverse transcriptase domain, an endonuclease domain and (unlike known IEPs) an RNase H domain. All are catalytically inactive. This finding supports the hypothesis that the spliceosome evolved from a once-mobile Group II intron, presumably near the time of the birth of the eukaryotic lineage when an archaeal cell engulfed a prokaryotic cell like the ones that became present-day mitochondria and chloroplasts.