

SPLICING AND ALTERNATIVE SPLICING

Spliceosome Assembly, the Lariat Mechanism, Regulation, and Biological Significance

1. Introduction

Most eukaryotic genes are interrupted by non-coding sequences called **introns**, which separate the coding **exons**. **Splicing** is the process by which introns are precisely removed from the primary transcript and the flanking exons are joined together to produce a mature mRNA. Splicing was discovered in 1977 by Phillip Sharp and Richard Roberts (independently), who found that adenovirus mRNA was shorter than the DNA sequence that encoded it — work recognised with the 1993 Nobel Prize.

- Introns are typically far more numerous, and often much longer, than exons — in humans, introns make up roughly a quarter of the genome, while exons account for only about 1-1.5%.
 - Splicing must be extremely precise: removing or retaining even a single incorrect nucleotide shifts the reading frame of all downstream exons, generally destroying the protein.
 - Splicing is carried out by a large, dynamic ribonucleoprotein machine called the **spliceosome**, and, in most cases, proceeds through a two-step chemical mechanism that generates a characteristic branched "lariat" intron intermediate.
-

2. Splice Sites and Spliceosome Components

Three short, conserved sequence elements mark every intron and are recognised, directly or indirectly, by the spliceosome.

- **5' splice site (donor site)** — the intron almost always begins with the dinucleotide **GU**, within the broader consensus (A/C)AG|GURAGU (the "GU-AG rule").
- **Branch point** — a conserved adenosine, usually 18-40 nucleotides upstream of the 3' end of the intron, within the consensus sequence YNYURAC; this adenosine's 2'-OH group is the catalytic nucleophile in the first step of splicing.
- **Polypyrimidine tract and 3' splice site (acceptor site)** — a pyrimidine-rich stretch just upstream of the intron's end, which itself ends in the dinucleotide **AG**.

Spliceosome Components

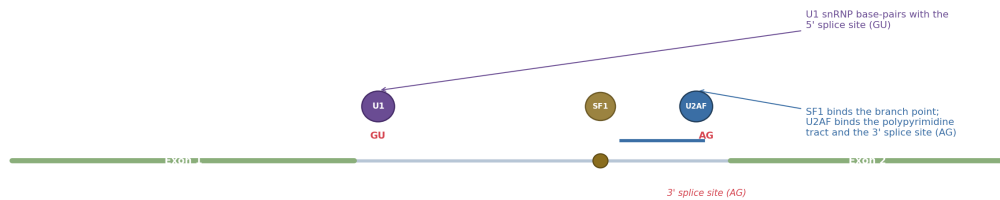
- The major spliceosome is built from five **small nuclear ribonucleoproteins (snRNPs)** — U1, U2, U4, U5, and U6 — each consisting of a small nuclear RNA (snRNA) bound to a set of proteins.
 - Additional non-snRNP factors are essential for early recognition: **SF1** (splicing factor 1) and the heterodimer **U2AF** (U2 auxiliary factor, made of U2AF65 and U2AF35 subunits).
 - Over 150 additional proteins associate with the spliceosome at various stages, making it comparable in complexity to the ribosome.
-

3. Spliceosome Assembly

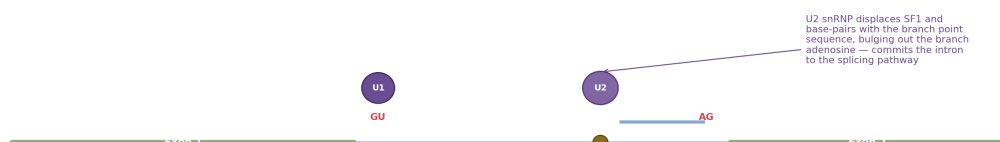
The spliceosome does not exist as a pre-formed particle; it assembles anew on each intron through an ordered series of complexes, each named for historical reasons by a letter.

Spliceosome Assembly: the E, A, B, and C Complexes

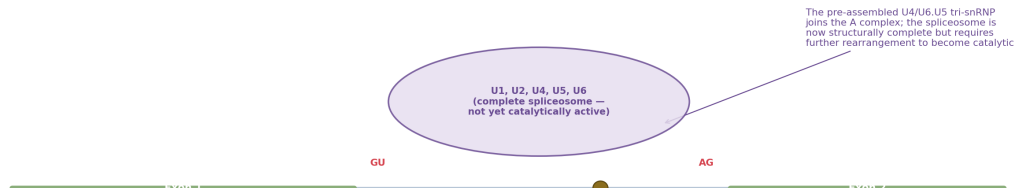
E complex — U1 snRNP, SF1, and U2AF mark the splice sites (commitment complex)



A complex (pre-spliceosome) — U2 snRNP replaces SF1 at the branch point (ATP-dependent)



B complex — the U4/U6.U5 tri-snRNP joins to form the complete, pre-catalytic spliceosome



Activation → C complex — U1 and U4 are released; U6 takes over to form the catalytic core

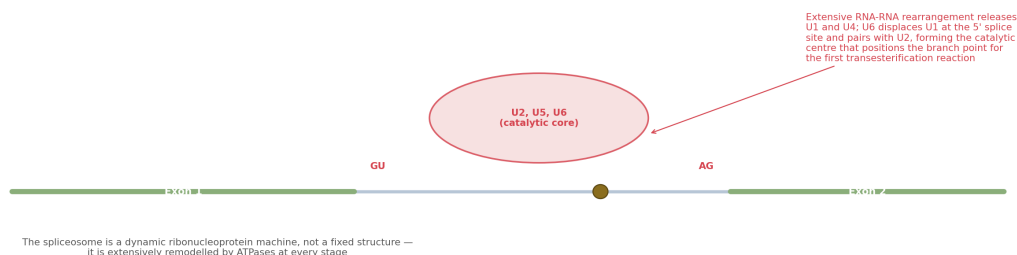


Fig. 1 — Spliceosome assembly proceeds through the E, A, B, and (after activation) C complexes, with snRNPs joining and rearranging in a defined order.

- **E complex (early/commitment complex)** — U1 snRNP base-pairs with the 5' splice site; SF1 binds the branch point; U2AF binds the polypyrimidine tract and 3' splice site, committing the intron to splicing.

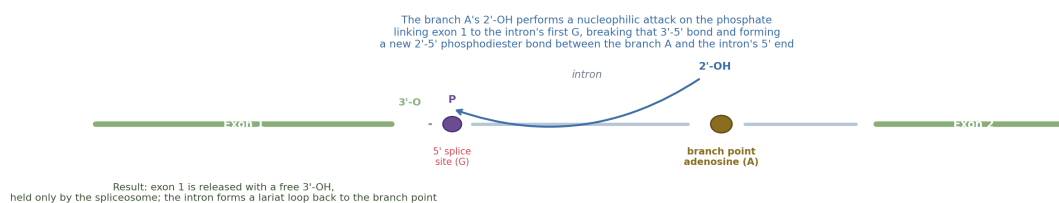
- **A complex (pre-spliceosome)** — in an ATP-dependent step, U2 snRNP displaces SF1 and base-pairs stably with the branch point sequence, bulging out the branch adenosine so it is positioned for catalysis.
- **B complex** — the pre-assembled U4/U6.U5 tri-snRNP joins, producing a structurally complete but still catalytically inactive spliceosome.
- **Activation to the B^{act}/B* and C complexes** — extensive ATP-dependent RNA-RNA and RNA-protein rearrangement releases U1 and U4; U6 snRNA displaces U1 at the 5' splice site and base-pairs with U2, forming the catalytic core. This activated spliceosome (C complex, after the first catalytic step) carries out both transesterification reactions before disassembling and releasing the spliced mRNA and the intron lariat.

4. The Catalytic Mechanism: Two Transesterification Reactions

Splicing chemistry is unusual: it proceeds through exactly two sequential **transesterification** reactions, each of which breaks one phosphodiester bond while simultaneously forming another, so that the number of phosphodiester bonds in the system never changes and no external energy source (such as ATP hydrolysis) is consumed by the chemistry itself — only by the extensive spliceosome remodelling that surrounds it.

The Chemistry of Splicing: Two Sequential Transesterification Reactions

Step 1 — the branch-point adenosine's 2'-OH attacks the phosphate at the 5' splice site



Step 2 — exon 1's newly freed 3'-OH attacks the phosphate at the 3' splice site

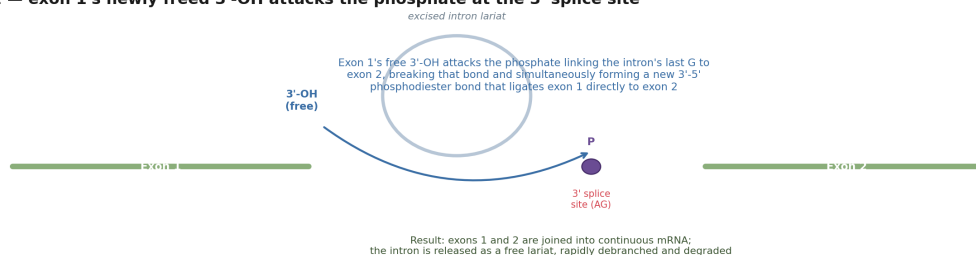


Fig. 2 — The branch-point adenosine's 2'-OH first attacks the 5' splice site to form a lariat; the freed exon 1 then attacks the 3' splice site to complete ligation.

- **Step 1** — the branch point adenosine's 2'-OH group performs a nucleophilic attack on the phosphate at the 5' splice site, cleaving the bond between exon 1 and the intron and forming a new, unusual **2'-5' phosphodiester bond** between the branch adenosine and the intron's first nucleotide. This

creates the branched **lariat** structure, and exon 1 is now held in place only by the spliceosome, not by a covalent bond.

- **Step 2** — the newly exposed 3'-OH of exon 1 attacks the phosphate at the 3' splice site, simultaneously cleaving the bond between the intron and exon 2 and forming a new phosphodiester bond that ligates exon 1 directly to exon 2.
- The excised intron, now a free lariat with a short 3' tail, is released still bound to certain spliceosomal proteins; a debranching enzyme opens the 2'-5' bond, converting it to a linear molecule that is rapidly degraded by exonucleases, and the released snRNPs are recycled for further rounds of splicing.

5. Types of Introns

Intron type	Removed by	Where found
Major (U2-type) spliceosomal	The major spliceosome (U1, U2, U4, U5, U6)	>99% of human introns; GU-AG boundaries
Minor (U12-type) spliceosomal	The minor spliceosome (U11, U12, U4atac, U6atac, U5)	A small, evolutionarily conserved subset of introns (~0.5% or fewer); AT-AC boundaries in some
Group I self-splicing	No protein enzyme — the intron's own RNA structure is a ribozyme; uses a free guanosine nucleotide	Some rRNA genes in protists, fungi, and organelles; a few bacteriophage genes
Group II self-splicing	Also a ribozyme, using a bulged adenosine — mechanistically similar to the spliceosome, and thought to be its evolutionary ancestor	Organellar genes in fungi, plants, and protists; some bacterial genes
tRNA introns	A dedicated splicing endonuclease and RNA ligase, unrelated to the spliceosome or to self-splicing	A minority of nuclear and archaeal tRNA genes

An evolutionary link

The spliceosome's catalytic mechanism — including the branch-point chemistry and the central role of RNA (U6 snRNA is thought to form the catalytic centre) — closely parallels self-splicing group II introns, leading to the widely held hypothesis that the spliceosome evolved from an ancient, mobile group II intron whose catalytic RNA components were gradually replaced or supplemented by protein snRNPs.

6. Alternative Splicing

Because splice sites are chosen based on relatively short, degenerate sequence elements rather than a single unambiguous signal, the same pre-mRNA can often be spliced in more than one way. This **alternative splicing** is estimated to affect over 90-95% of human multi-exon genes, and is a major reason the proteome is far more diverse than the roughly 20,000 protein-coding genes would suggest.

Pattern	Description
Exon skipping (cassette exon)	An internal exon is sometimes included, sometimes spliced out entirely — the most common pattern in mammals

Pattern	Description
Alternative 5'/3' splice-site selection	A different splice site is used at one end of an exon, shortening or lengthening it
Intron retention	An intron is left unspliced and remains part of the mature mRNA, often introducing a premature stop codon
Mutually exclusive exons	Of two or more alternative exons, exactly one is always included and the rest excluded
Alternative promoter usage	Transcription starts at different promoters, changing the first exon included
Alternative polyadenylation	A different poly(A) site is used, changing the last exon or 3' UTR length

7. Regulation of Alternative Splicing

Which splice sites are used in a given cell type or condition is controlled by short **cis-acting regulatory sequences** in the pre-mRNA and the **trans-acting proteins** that bind them.

Regulation of Alternative Splicing: Enhancers vs Silencers

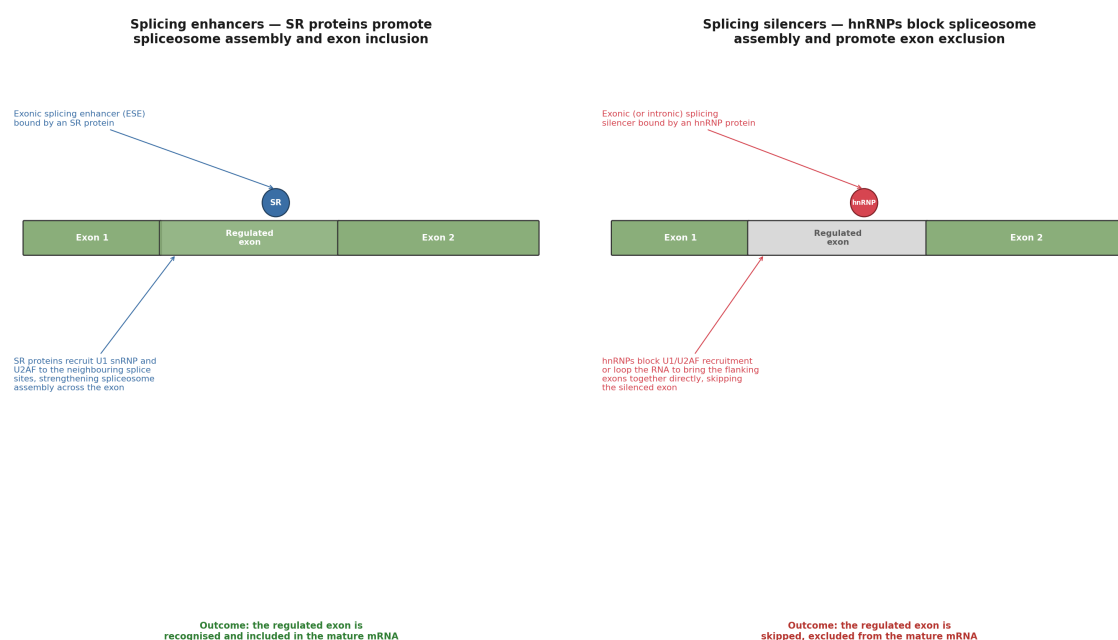


Fig. 3 — SR proteins bound to splicing enhancers promote spliceosome assembly and exon inclusion, while hnRNPs bound to silencers block assembly and promote exon skipping.

- **Exonic and intronic splicing enhancers (ESEs, ISEs)** are typically bound by **SR proteins** (named for a domain rich in serine-arginine repeats), which recruit U1 snRNP and U2AF to strengthen recognition of the neighbouring splice sites and promote exon inclusion.
- **Exonic and intronic splicing silencers (ESSs, ISSs)** are typically bound by **heterogeneous nuclear ribonucleoproteins (hnRNPs)**, which interfere with spliceosome assembly or loop the RNA to bring flanking exons together directly, promoting exon skipping.

- The relative concentration and activity of different SR proteins and hnRNPs — which vary between cell types, developmental stages, and in response to signalling — allows the same gene to be spliced differently in different contexts, a major mechanism of combinatorial gene regulation.
-

8. Biological Significance and Disease

(a) Classic Examples

- *Drosophila* sex determination is controlled by a cascade of alternative splicing decisions: the Sex-lethal (Sxl) protein regulates splicing of transformer (tra), whose product in turn regulates splicing of doublesex (dsx) — producing sex-specific Dsx protein isoforms that direct male or female development from the same genes.
- The *Drosophila* **Dscam** gene can in principle generate over 38,000 distinct mRNA isoforms through mutually exclusive alternative splicing of clusters of variable exons, contributing to the molecular diversity needed for neuronal wiring — one of the most extreme known examples of alternative splicing.

(b) Splicing in Disease

- Mutations that disrupt a splice site, branch point, or a splicing regulatory element are a common cause of genetic disease — estimated to account for a substantial fraction of all disease-causing mutations, not just those in obviously "coding" positions.
- In **spinal muscular atrophy (SMA)**, loss of the SMN1 gene leaves patients dependent on the nearly identical SMN2 gene, which — due to a single nucleotide difference affecting an exonic splicing enhancer — predominantly skips exon 7 and produces mostly non-functional protein. The drug **nusinersen**, an antisense oligonucleotide, works by binding pre-mRNA to block a splicing silencer and promote exon 7 inclusion from SMN2, restoring functional protein — a leading example of splice-switching therapy.
- Dysregulated alternative splicing (through mutated splicing factors or splice sites) is a recognised hallmark of many cancers, altering the balance of pro- and anti-apoptotic or growth-regulatory protein isoforms.

9. Quick Revision Summary

Topic	Key point
Splice site sequences	5' site (GU), branch point (bulged A), polypyrimidine tract, 3' site (AG) — the "GU-AG rule"
Spliceosome assembly	E → A → B → activated (Bact/C) complexes; U1, U2, U4, U5, U6 snRNPs join and rearrange in order
Catalytic mechanism	Two transesterifications: branch A's 2'-OH attacks the 5' site (forms lariat); exon 1's 3'-OH attacks the 3' site (ligates exons)
Intron types	Major (U2) and minor (U12) spliceosomal introns; self-splicing group I/II ribozyme introns; tRNA introns
Alternative splicing	Exon skipping, alternative splice sites, intron retention, mutually exclusive exons, alternative promoters/poly(A) sites
Regulation	SR proteins on enhancers promote inclusion; hnRNPs on silencers promote exon skipping
Disease relevance	SMN2 exon-7 skipping in SMA (treated with the splice-switching drug nusinersen); widespread splicing dysregulation in cancer

Compiled study note — molecular biology reference. Figures are original schematic diagrams prepared for this note and are not to scale.