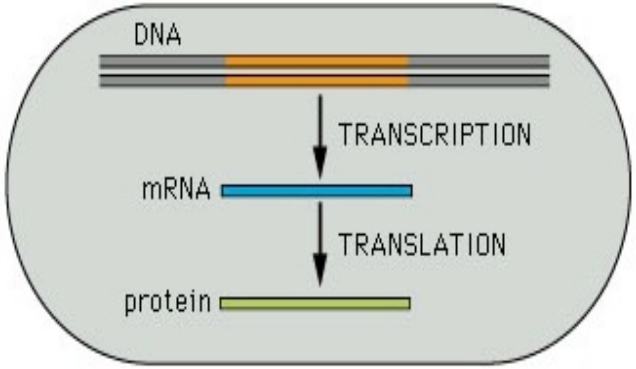
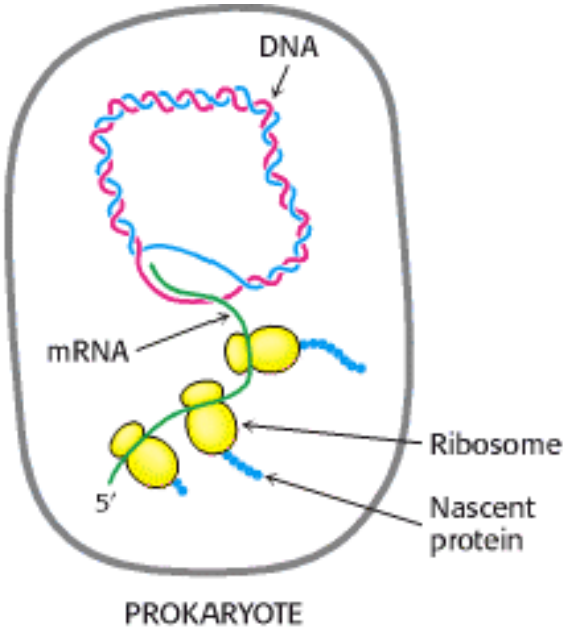
Messenger RNA maturation

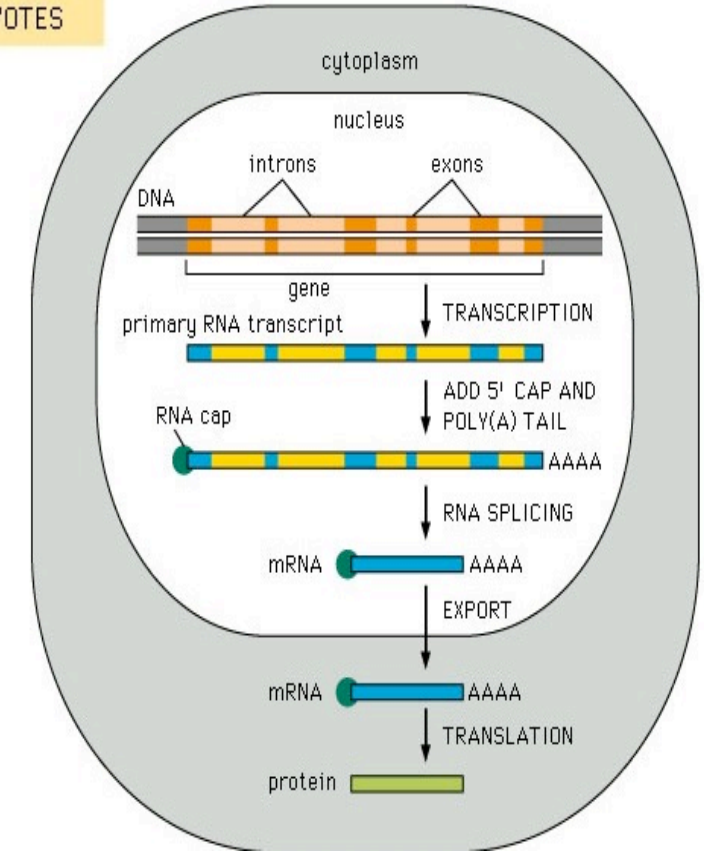
PROCARYOTES

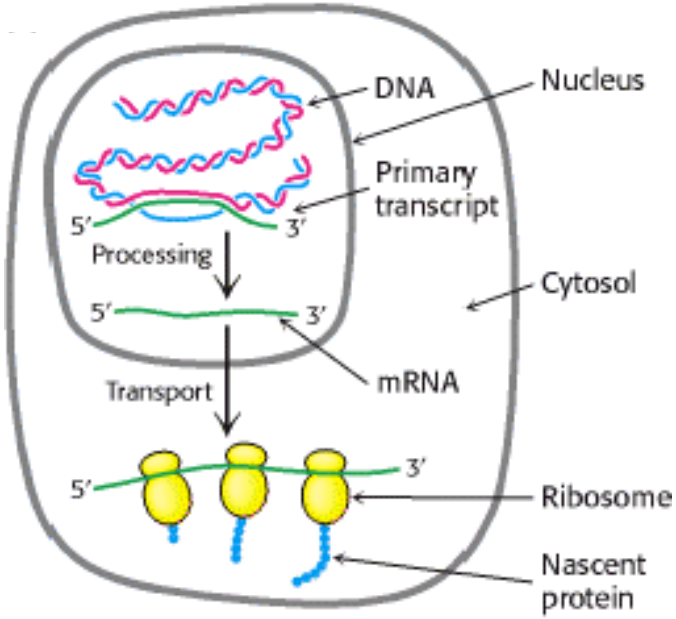




PROKARYOTE

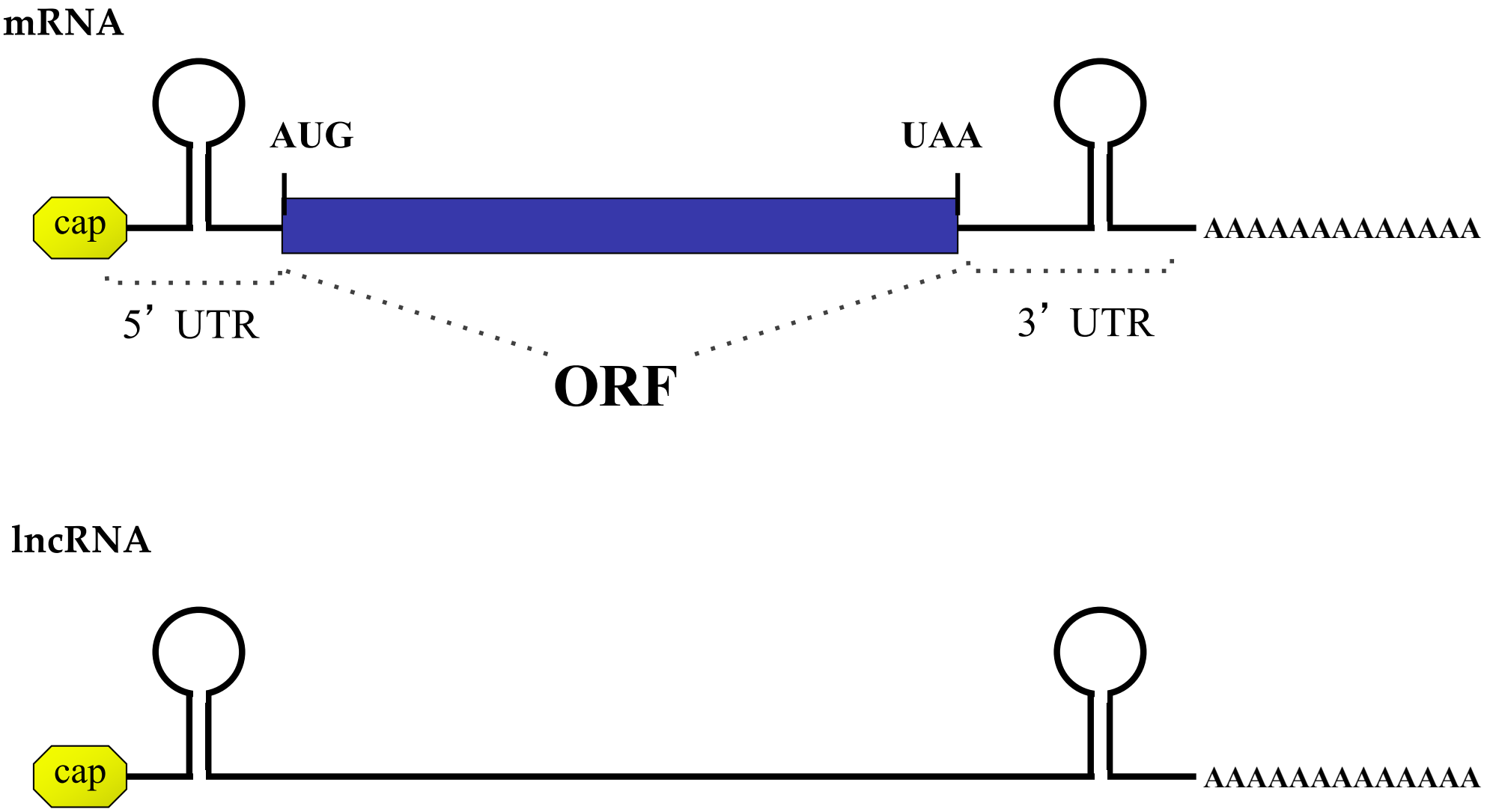
EUCARYOTES





EUKARYOTE

Mature mRNA and lncRNA



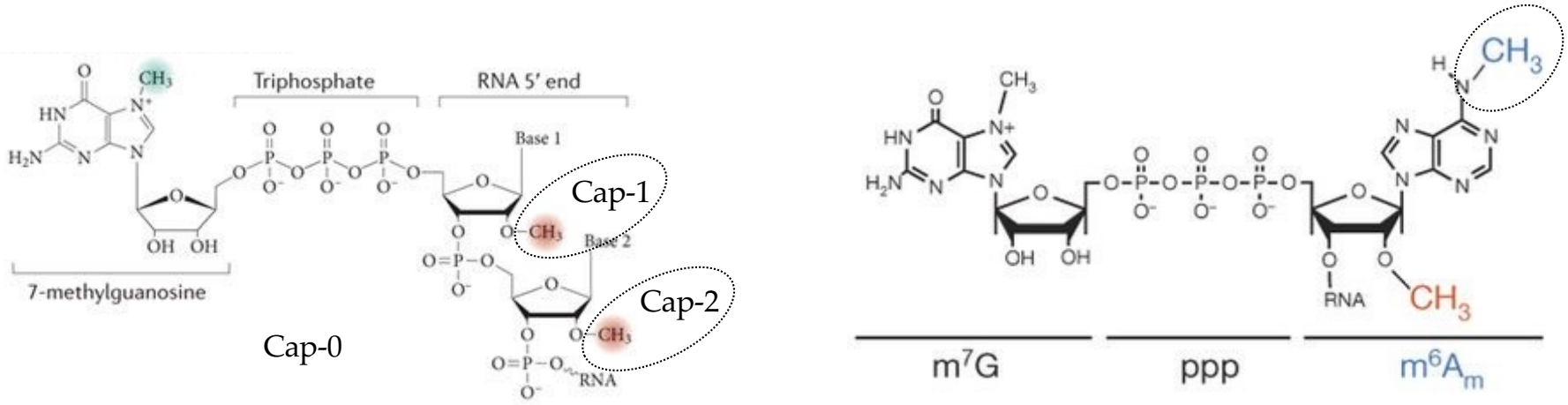
mRNA Capping

Why the Cap structure is important?

1) Increases RNA stability
2) Favours mRNA transport to the cytoplasm
3) Increases translation
4) Avoids RNA-mediated antiviral immunity

5′ CAP: m7G(5′)ppp

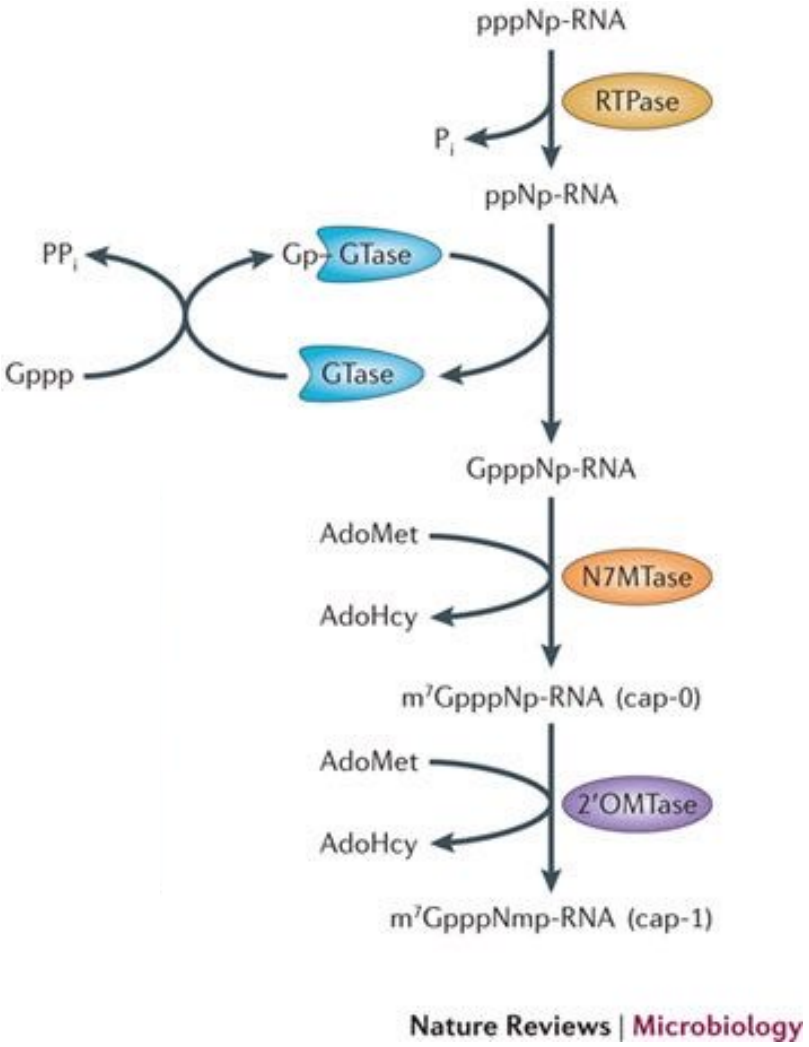
The mRNA cap consists of a **7-methylguanosine** linked to the 5′ nucleoside of the mRNA chain through a 5′–5′ triphosphate bridge **(Cap-0).** The methyl group at the N^7 position of the guanosine is shaded green, and the **2′-*O*-methyl groups** of the first and second nucleotide residues, forming the **Cap-1** and the **Cap-2** structures, respectively, are shaded red. In mammals, about 90% of mRNA initiating with adenosine has an additional modification, **N^6,2′-*O*-dimethyladenosine (m^6A_m).**



The cap protects the RNA from 5′–>3′ exonuclease decay and its removal (**decapping**) is a critical regulatory event that influences transcript expression and affects overall protein synthesis within the cell. Cap 1 and 2 are important for authenticate mRNAs as "self RNAs".In the cytosol, the cap 0 is specifically recognized by the translation initiation factor eIF4E and required for efficient mRNA translation.

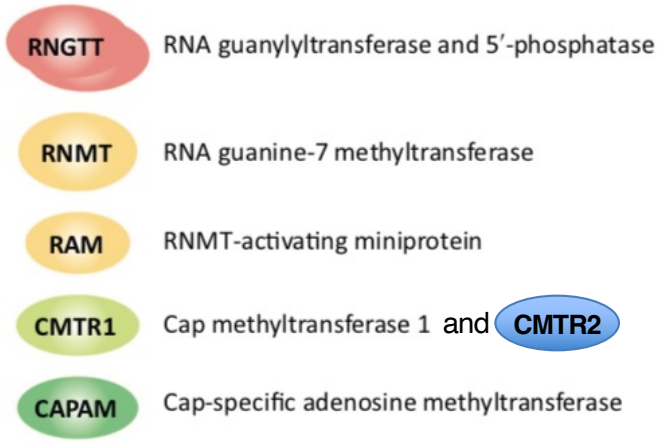
Capping reaction

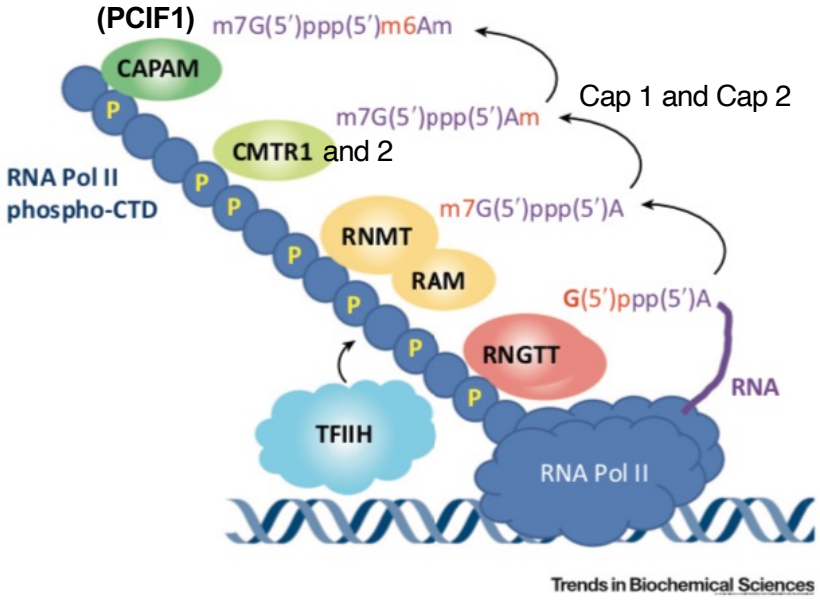
The CAP structure is formed on nascent RNA chains by the sequential action of three enzymes. First, the **RNA triphosphatase (RTPase)** hydrolyses the γ-phosphate of the nascent RNA to yield a diphosphate RNA (ppNp-RNA) and inorganic phosphate (P_i). Then, **guanylyltransferase (GTase)** reacts with the α-phosphate of GTP (Gppp), releasing pyrophosphate (PP_i) and forming a covalent enzyme–guanylate intermediate (Gp–GTase). The GTase then transfers the GMP molecule (Gp) to the 5′-diphosphate RNA to create GpppNp-RNA. In the final step, **(guanine-N^7)-methyltransferase (N7MTase)** transfers the methyl group from *S*-adenosyl-L-methionine (AdoMet) to the cap guanine to form the cap-0 structure, 7-methyl-GpppNp (m^7GpppNp). The capping reaction is completed by methylation of the ribose-2′-*O* position of the first nucleotide by the **AdoMet-dependent (nucleoside-2′-*O*)-methyltransferase (2′OMTase, CRMT1)**, generating the cap-1 structure ($m^7GpppNm_{2'-O}p$).

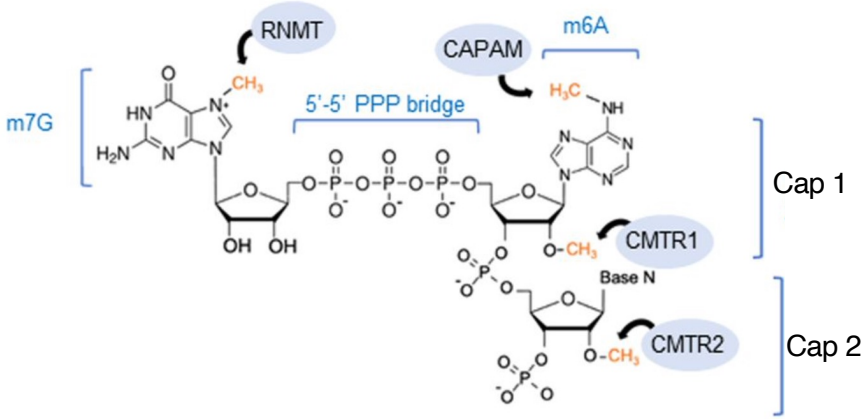


mRNA Capping Enzymes are recruited to RNA Pol II During Transcription

During the initial phase of transcription, TFIIH phosphorylates the Pol II CTD on serine 5. The capping enzymes RNGTT, RNMT–RAM, CMTR1, and CAPAM (PCIF1) are then recruited to the phosphorylated RNA Pol II. Next to each capping enzyme is a cap structure that they produce, with the specific component added marked in red.

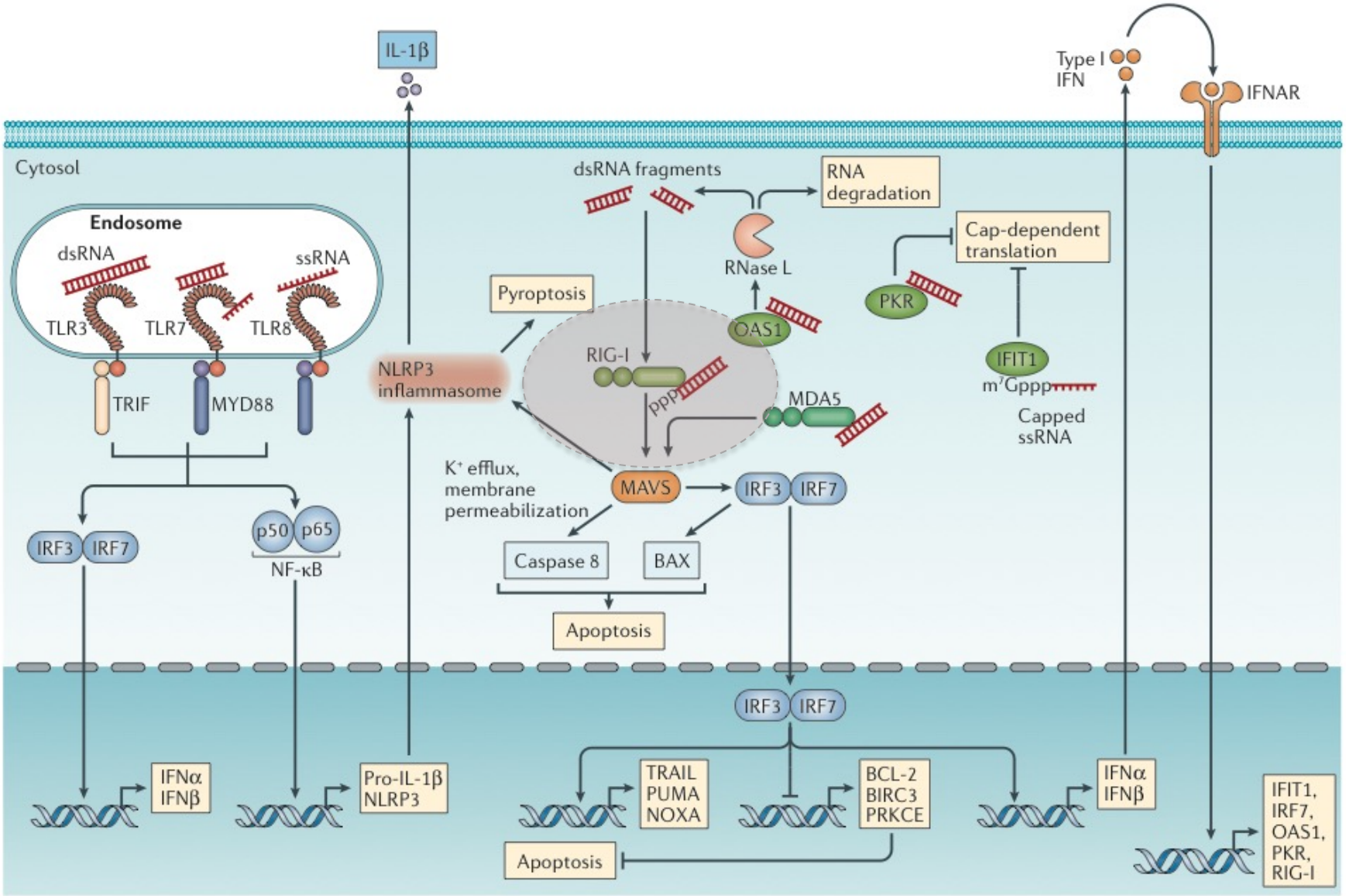




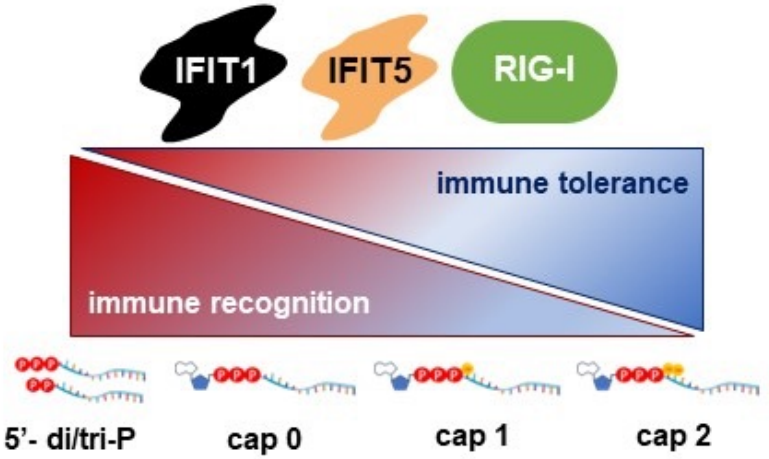


In yeast is only present Cap0

Antiviral innate immunity pathways



5′ end of RNA recognized by antiviral innate immune sensor



RNA 5′ ends	RIG-I	IFIT1	IFIT5
5′-P	–	–	++
5′-PPP/PP	++	++	++
m7G (cap 0)	++	++	+
m7G N1 (cap 1)	–	+	–
m7G N2 (cap 2)	–	–	–

(++) Interaction.
(+) Weak interaction.
(–) No interaction.

Capping of viral RNA

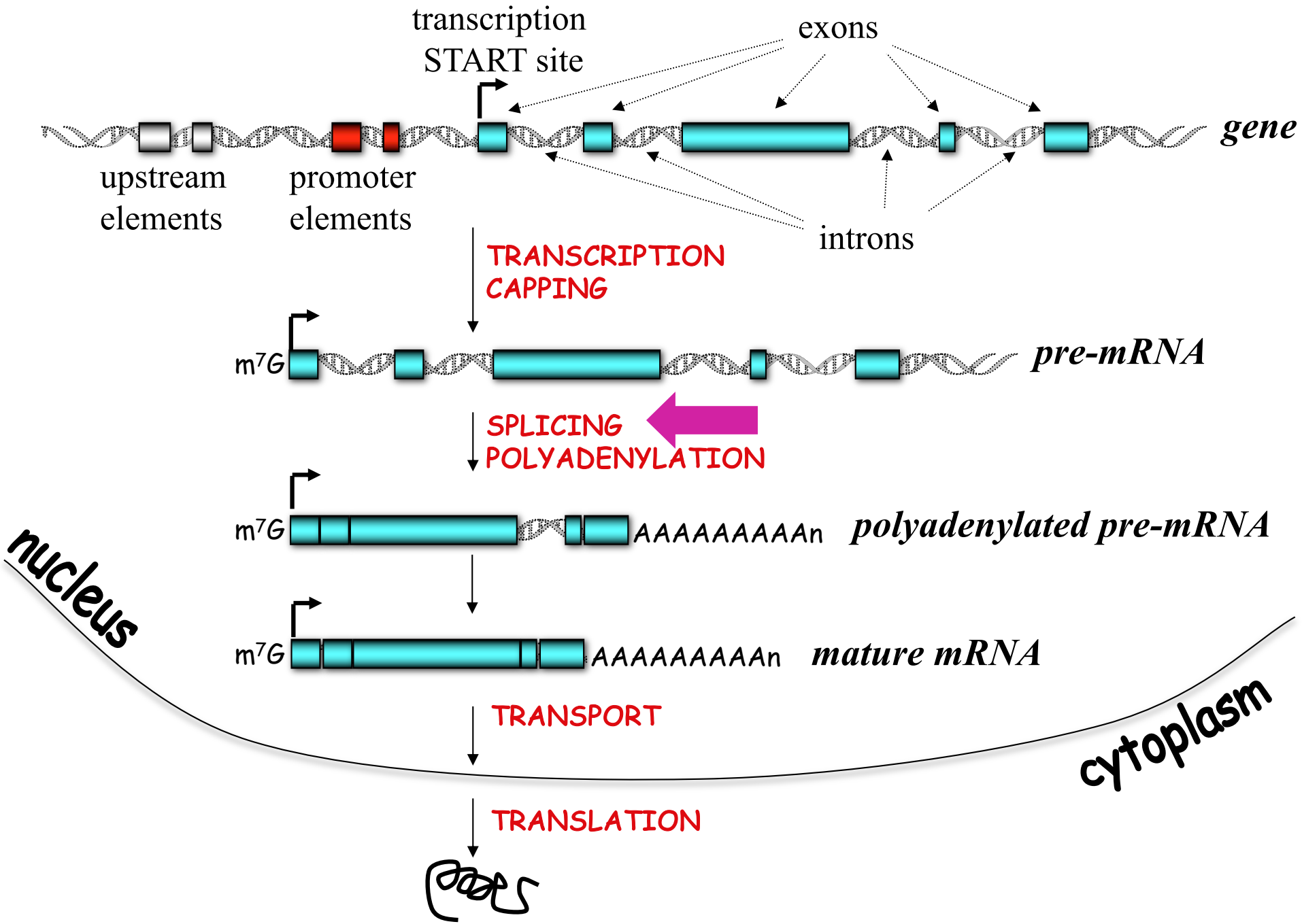
Viruses can acquire their cap structures using the cellular capping machinery or encode their own viral capping machineries that adopt the canonical pathway. The RNAs capped by viral enzymes are indistinguishable from cellular mRNA and can thus be translated into proteins by the cellular ribosomal machinery.

The capping of viral mRNA can be classified as either 'conventional', when it follows the mRNA-capping pathways used by eukaryotes and DNA viruses or 'unconventional', when it does not .

Capped viral RNA

HIV
BTV, bluetongue virus;
HBV, hepatitis B virus;
HSV, herpes simplex viruses;
SARS CoV(2), severe acute respiratory syndrome coronavirus.
Smallpox virus
Papilloma Virus

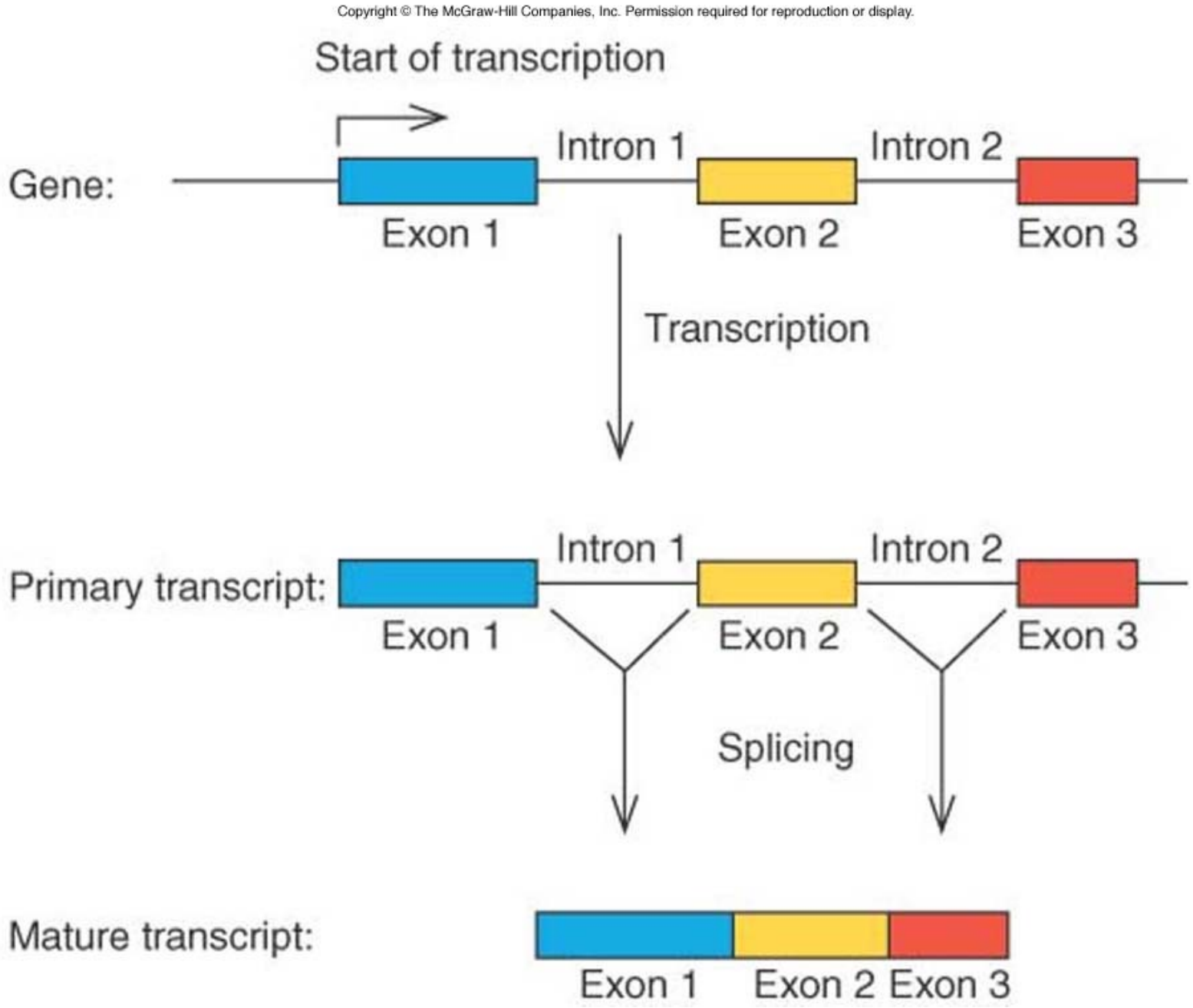
Eucaryotic gene expression



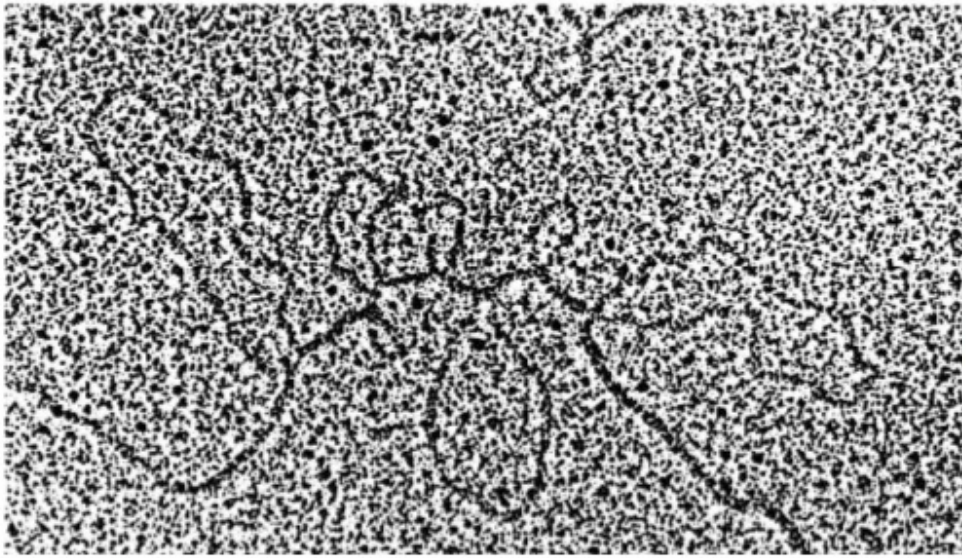
“Splicing”

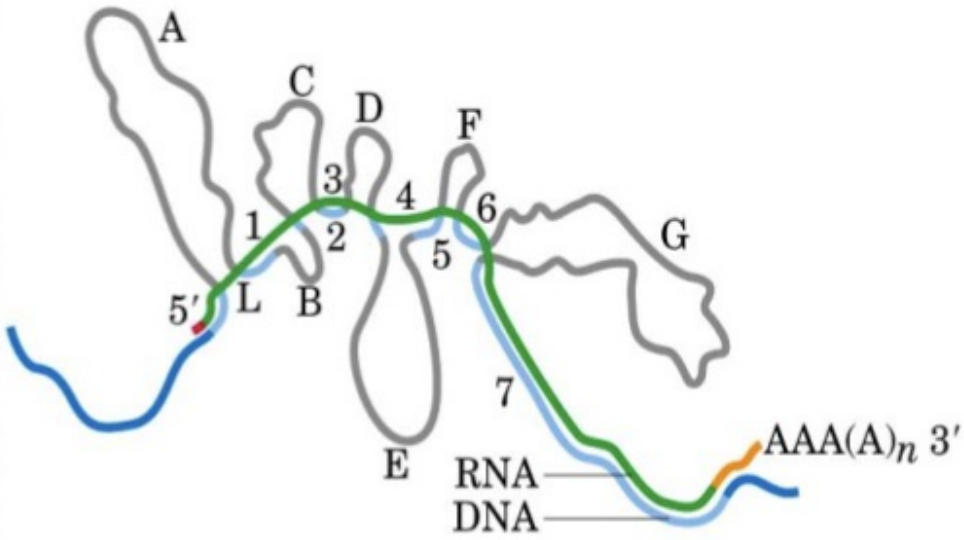
Eukaryotic genes contain introns

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



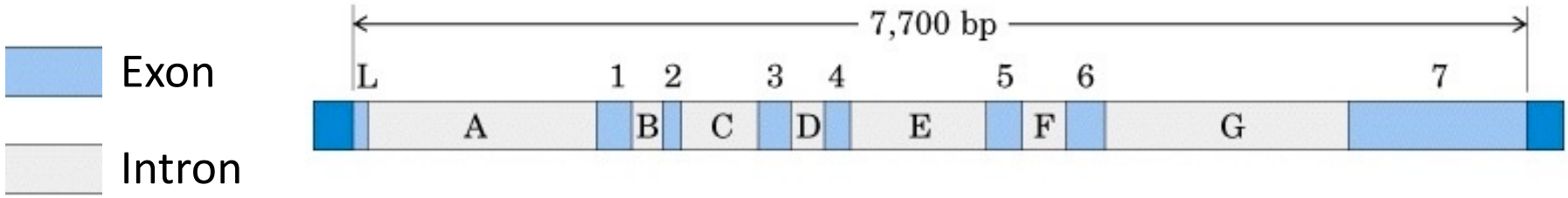
Identification of introns by R-looping



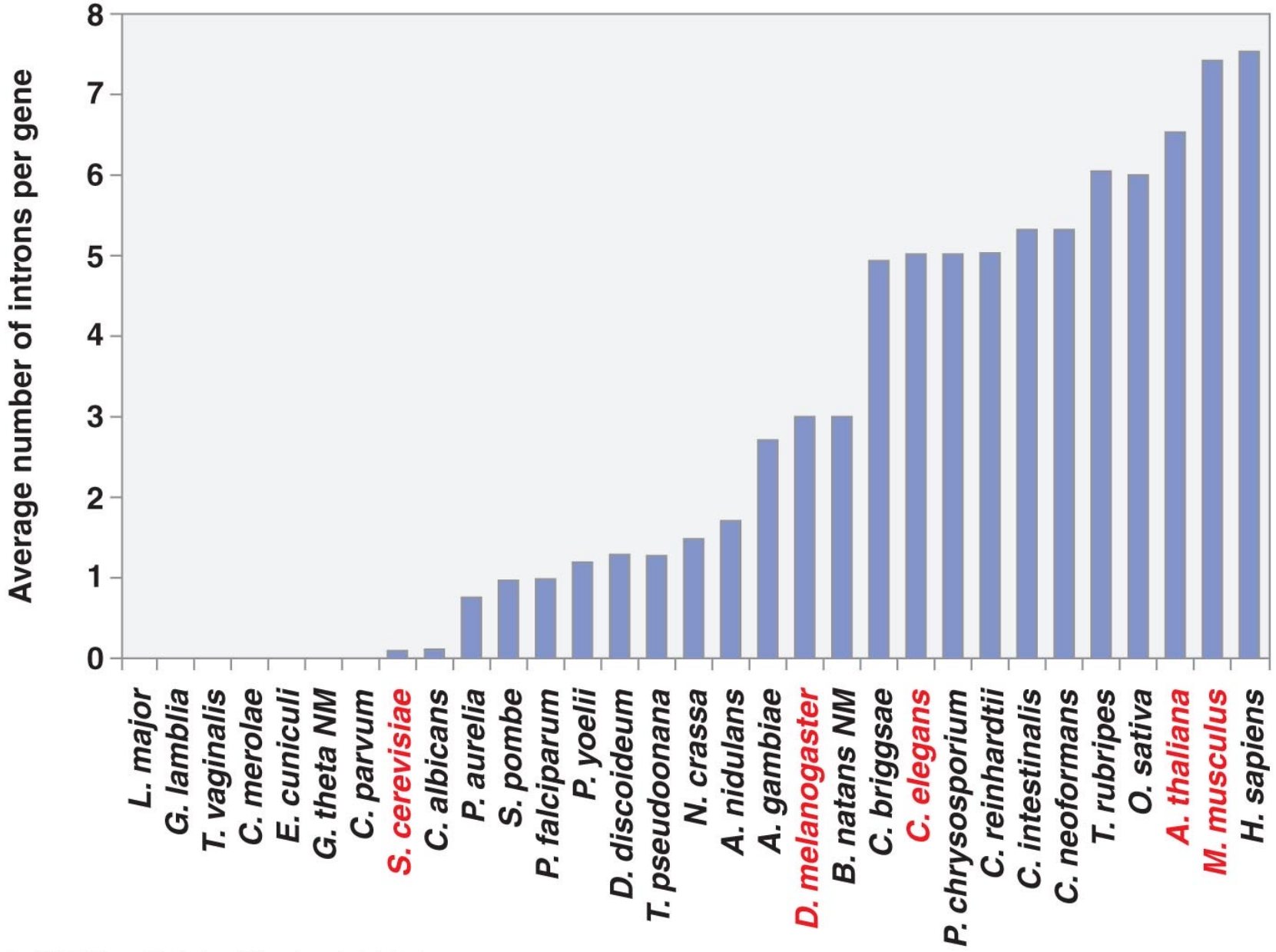


Chicken ovalbumin gene hybridized with its mRNA

Exon:intron organization of the ovalbumin gene



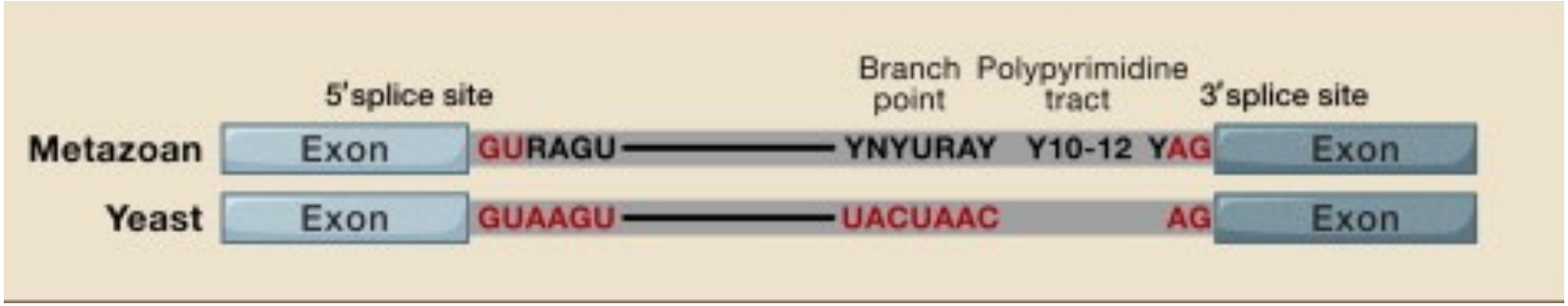
Number of introns per gene in various eukaryotic species



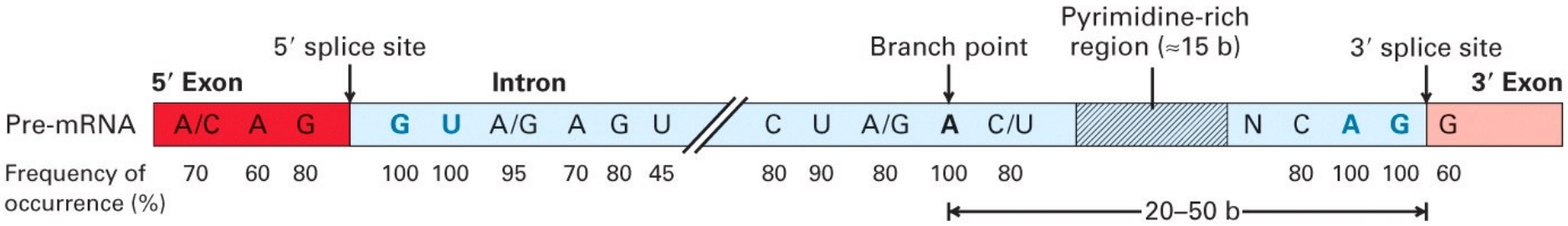
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

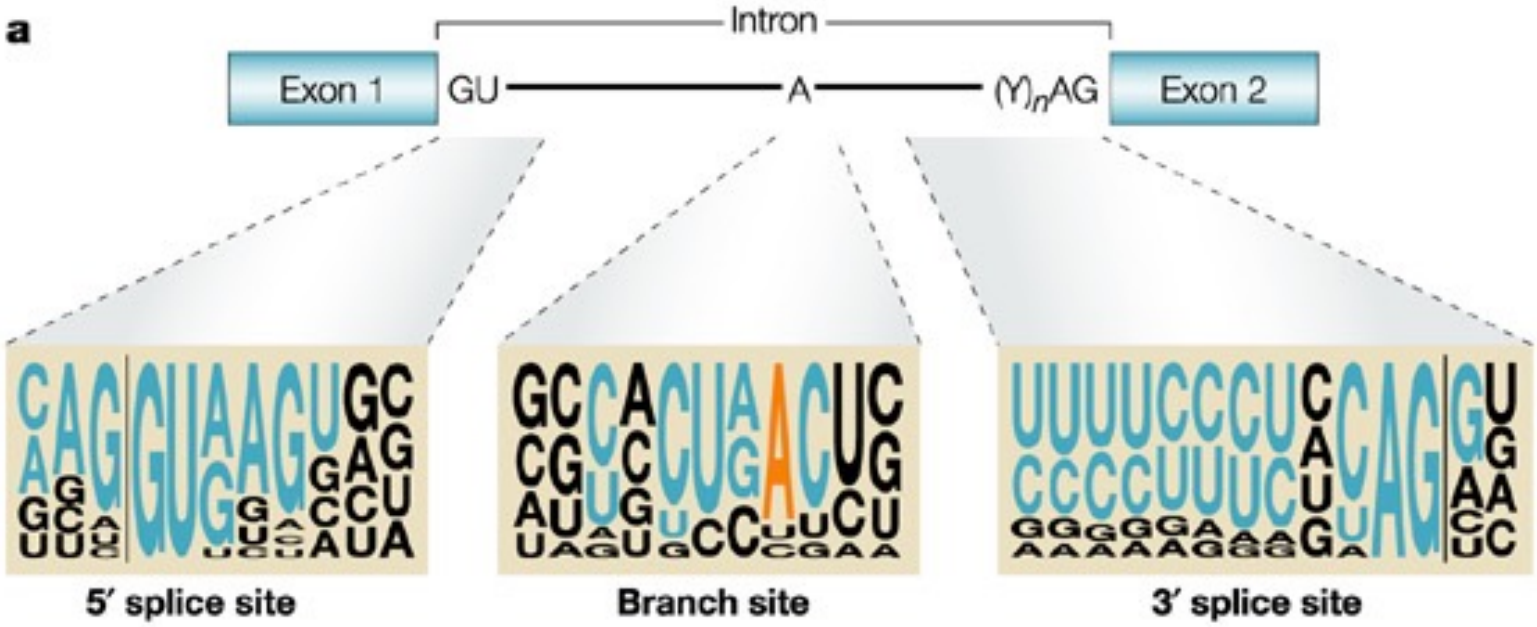
Specific sequences inside the pre-mRNA indicate where the splicing event has to take place.

These sequences indicate the boundary between introns and exons.



Consensus Sequences Surrounding the 5’ and 3’ Splice Sites





Despite the nucleotide sequence conservation of these sites, intron ends and their differential usage are difficult to define computationally in metazoa. Therefore, most intron ends are annotated from empirical data, such as cDNA sequencing.

Nuclear pre-mRNA splicing proceeds through two trans-esterification reactions

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

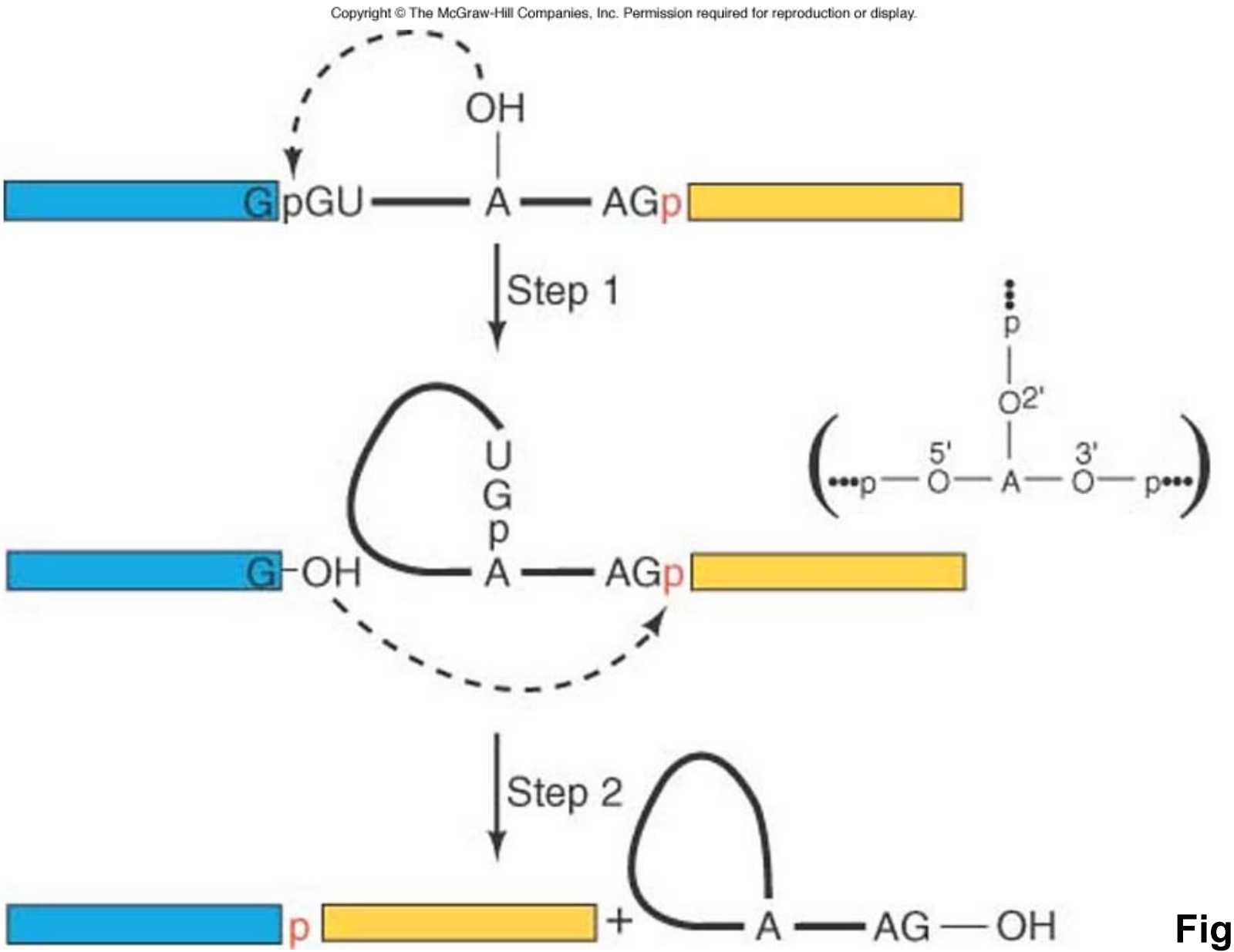


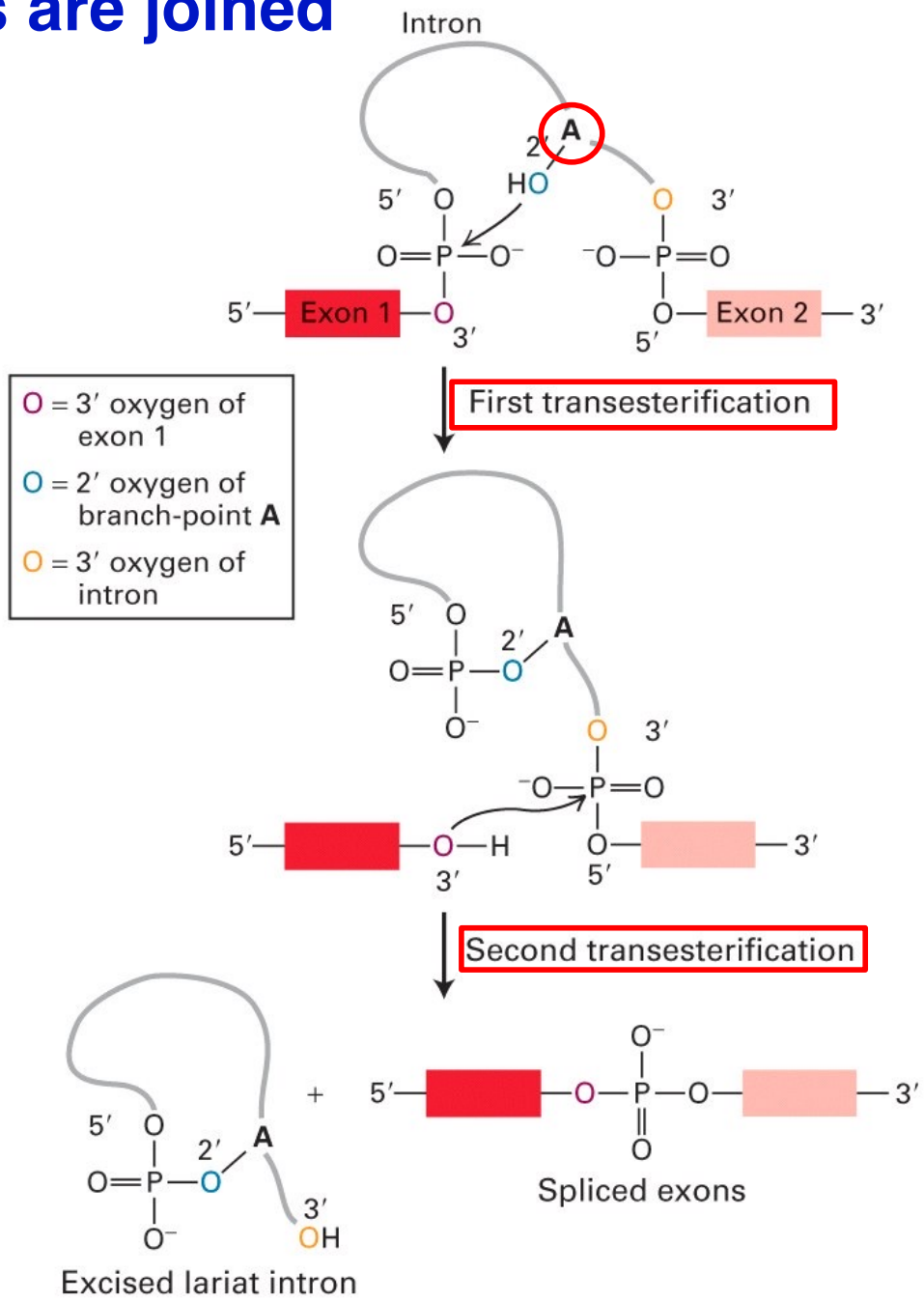
Fig. 14.4

The intron is removed in a Form Called Lariat and the Flanking Exons are joined

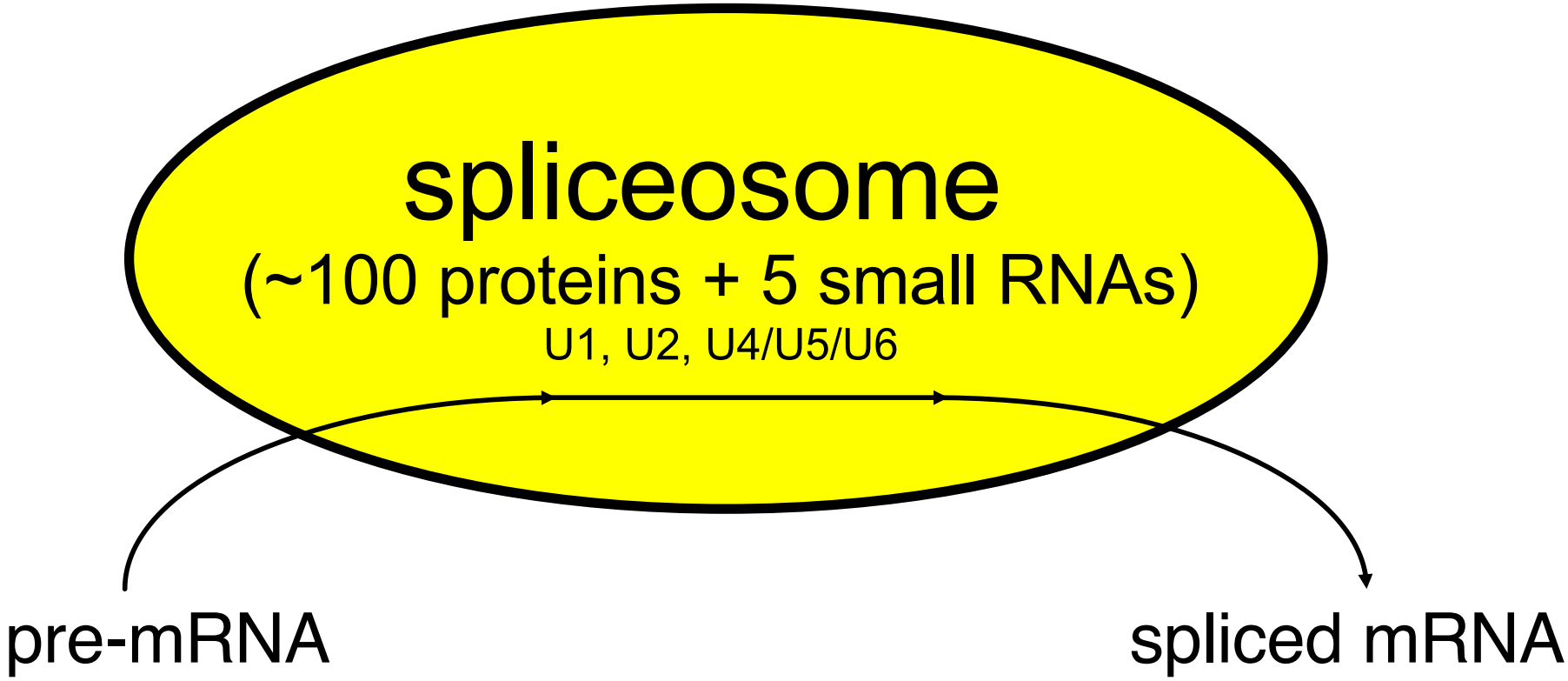
Two **trans-esterifications:**

Step 1: The OH of the conserved **A at the branch site** attacks the phosphoryl group of the conserved G in the 5' splice site. As result, the 5' exon is released, and the 5'-end of the intron forms a three-way junction structure.

Step 2: The OH of the 5' exon attacks the phosphoryl group at the 3' splice site. As a consequence, the 5' and 3' exons are joined, and the intron is released in a lariat shape

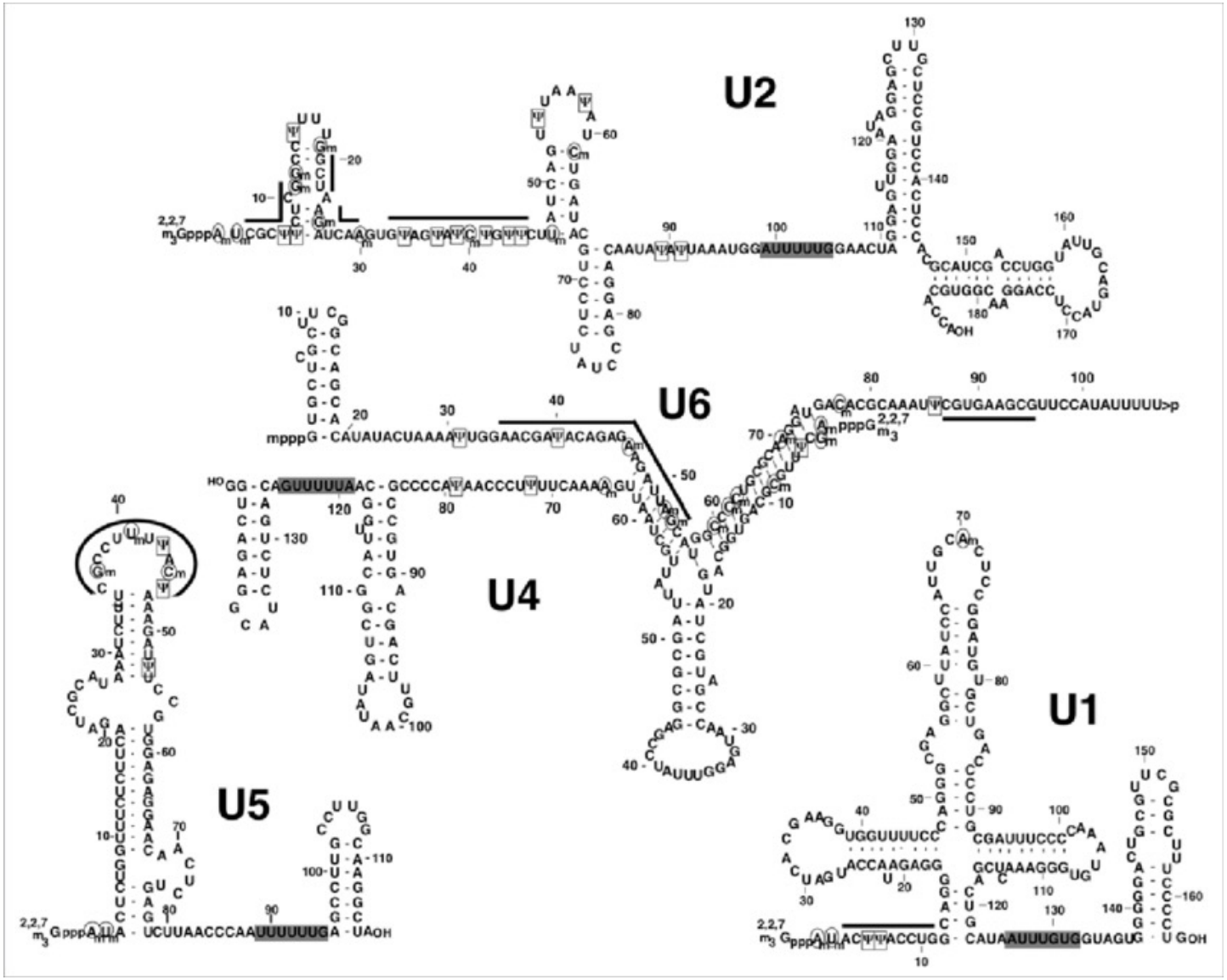


Splicing takes place on a complex called spliceosome



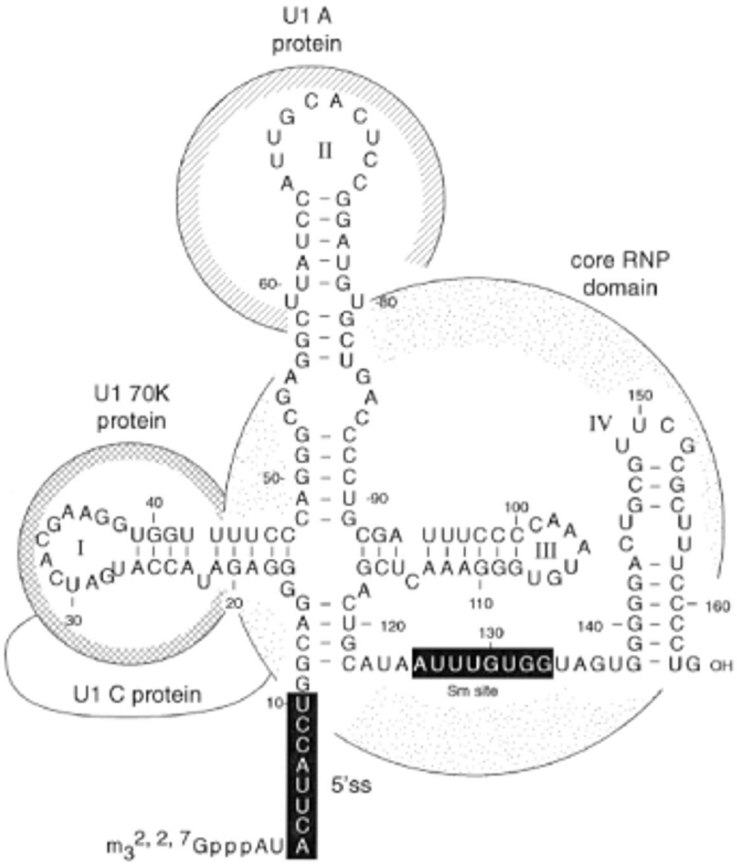
Splicing works similarly in different organisms, for example in yeast, flies, worms, plants and animals.

Five snRNAs are involved in pre-mRNA splicing

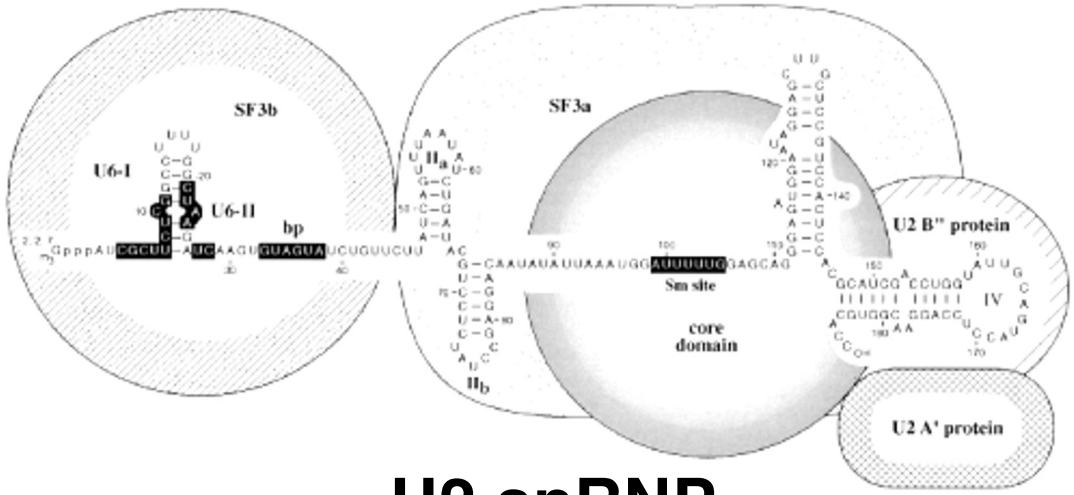


The thick lines indicate the nucleotides participating in RNA-RNA interactions or involved in catalysis during pre-mRNA splicing. The gray boxes highlight the Sm binding sites.

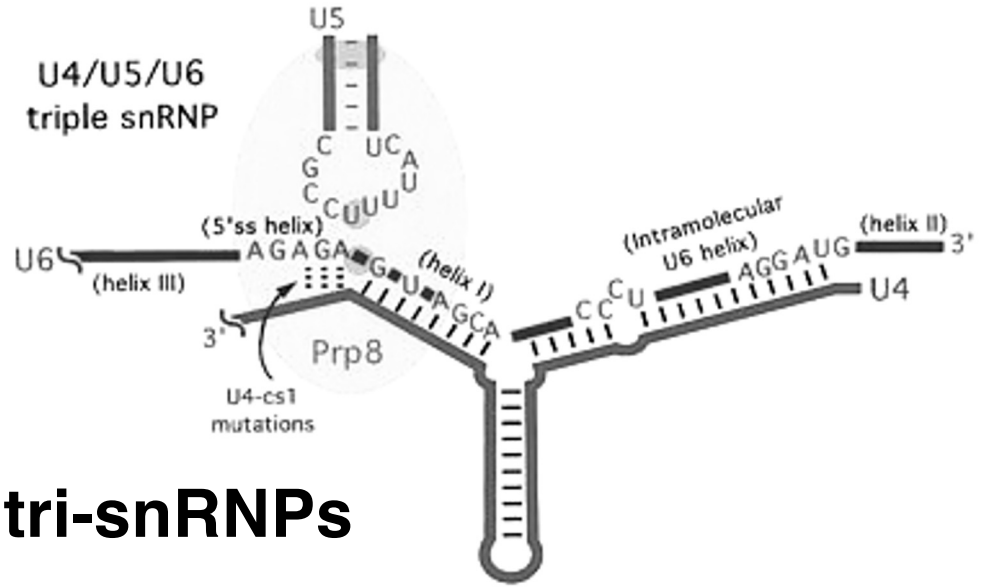
Small nuclear ribonucleoproteins (snRNPs)



U1 snRNP



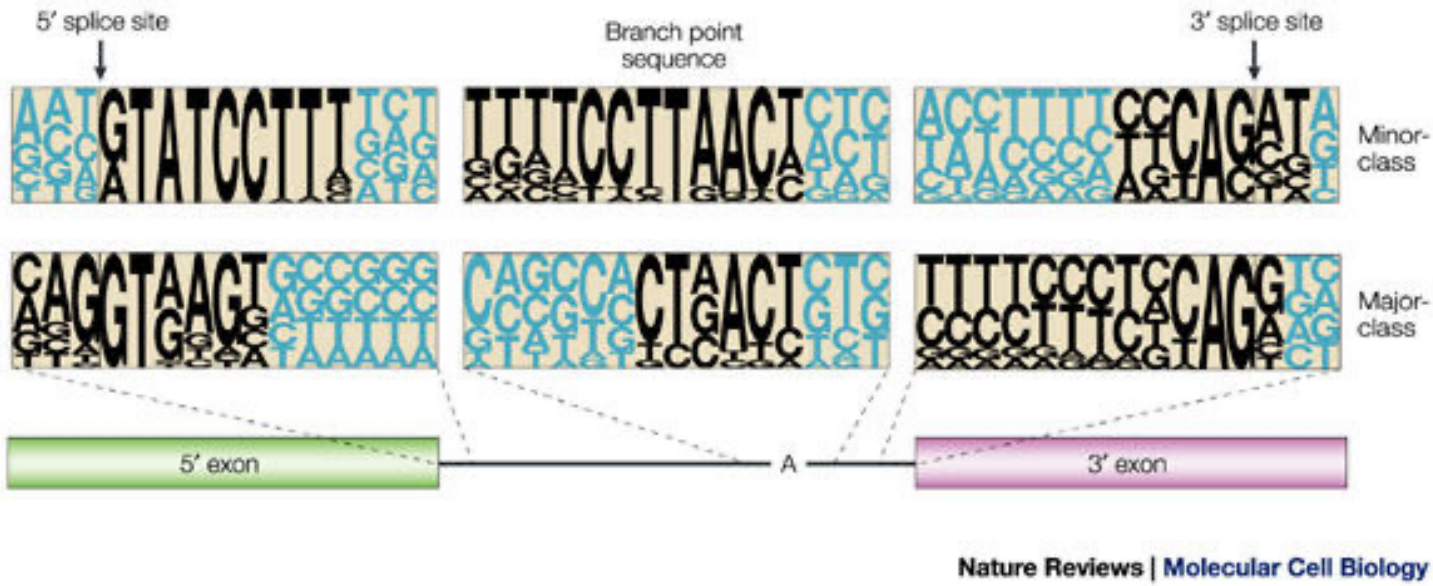
U2 snRNP



U4/U5/U6 tri-snRNPs

Minor-class introns

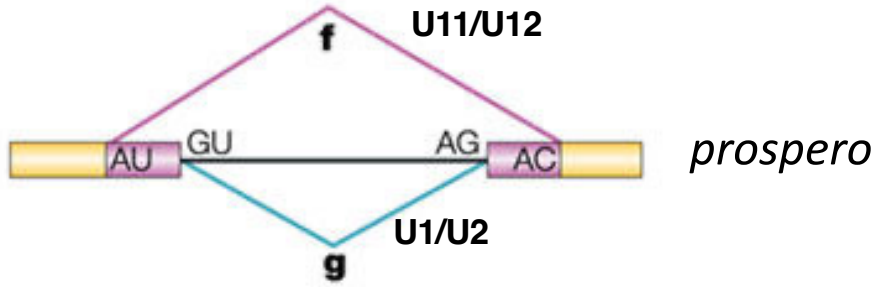
Some rare introns shared unusual consensus sequences, and their excision is mediated by a distinct spliceosome, **U12-dependent spliceosome**, that involved low-abundance snRNPs (U11, U12, U4atac–U6atac). U12-dependent splicing is determined by the longer and more tightly constrained consensus sequences at the 5' splice site and branch site of minor-class introns, as well as by the lack of a polypyrimidine tract upstream of the 3' splice site.



human U12-type introns are found in only 1 out of approximately 200 transcripts.

Major- vs Minor- spliceosome

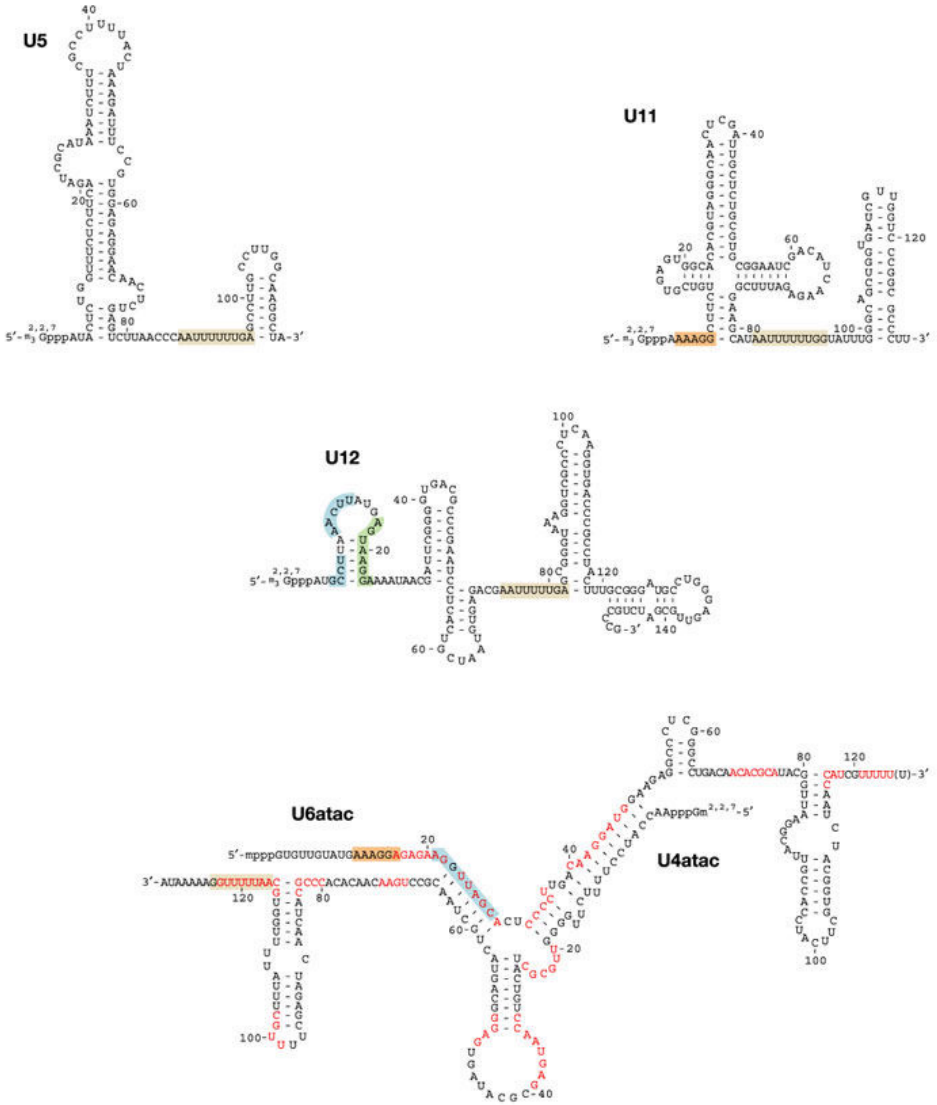
Major-class and minor-class splice sites are incompatible with each other, and as yet no hybrid introns have been identified. Therefore, unique patterns of alternative splice-site choice can arise. One such example is the 'intron-within-an-intron' architecture of the *prospero* gene of *Drosophila melanogaster*, in which minor-class splice sites (**f**) flank an internal major-class intron (**g**). In this case, splicing occurs by one or the other pathway, but not both, which results in a change of five amino acids at the amino terminus of the protein's functionally important homeodomain.



The choice of U12-type versus U2-type splicing in this transcript is modulated during embryogenesis and regulated by the binding of protein factors to an intronic element.

U12-dependent spliceosome

U11 and U12 have base-pairing potential with the 5’-splice-site and branch-site sequences, whereas their secondary structures mimic those of U1 and U2, respectively. While U6atac/U4atac are the analogues of U6 and U4. The U5 snRNP is a component of both spliceosomes.



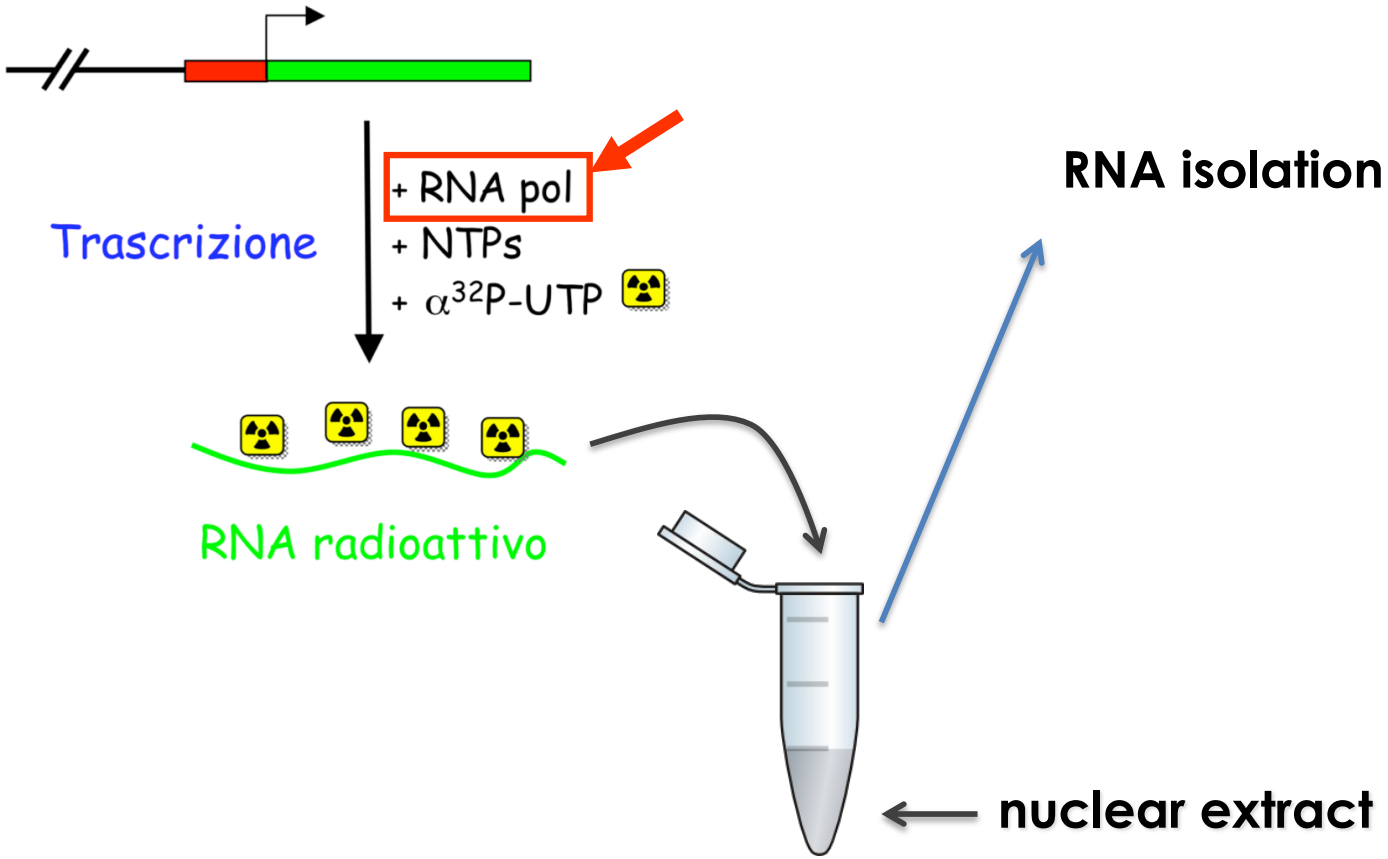
Assembly, rearrangement, and catalysis within the *spliceosome*: the splicing pathway

In vitro splicing in nuclear extracts has revealed how snRNPs associate and are released from pre-synthesized pre-mRNAs. This has led to the formulation of a spliceosome assembly model in which the spliceosome transitions through sequential assembly stages .

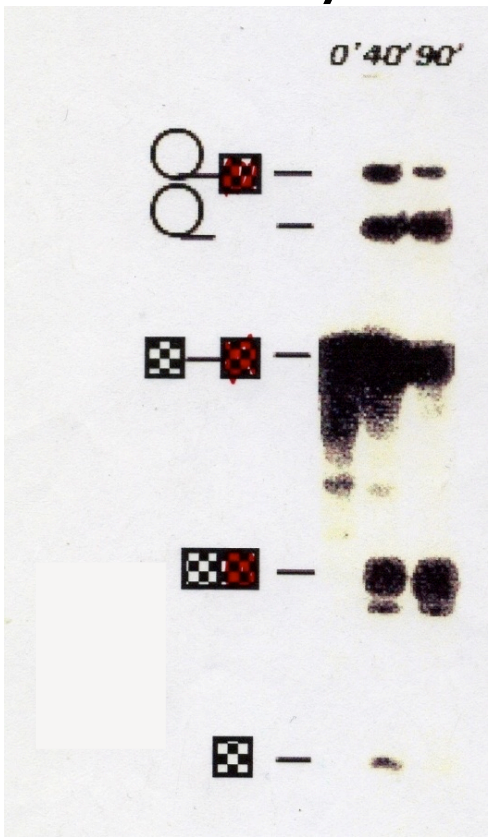
Spliceosome components and assembly are generally conserved in eukaryotes. The ~90 core proteins are conserved between the budding yeast *Saccharomyces cerevisiae* and human spliceosomes, although the number of human spliceosome proteins (~175) is two fold higher compared with yeast .

How to study splicing *in vitro*

- **procedure**
  - -*in vitro* transcribed RNA
  - -RNA plus nuclear extract
  - -RNA isolation ot different time points
  - -elettrophoresis for the analysis of the reaction products

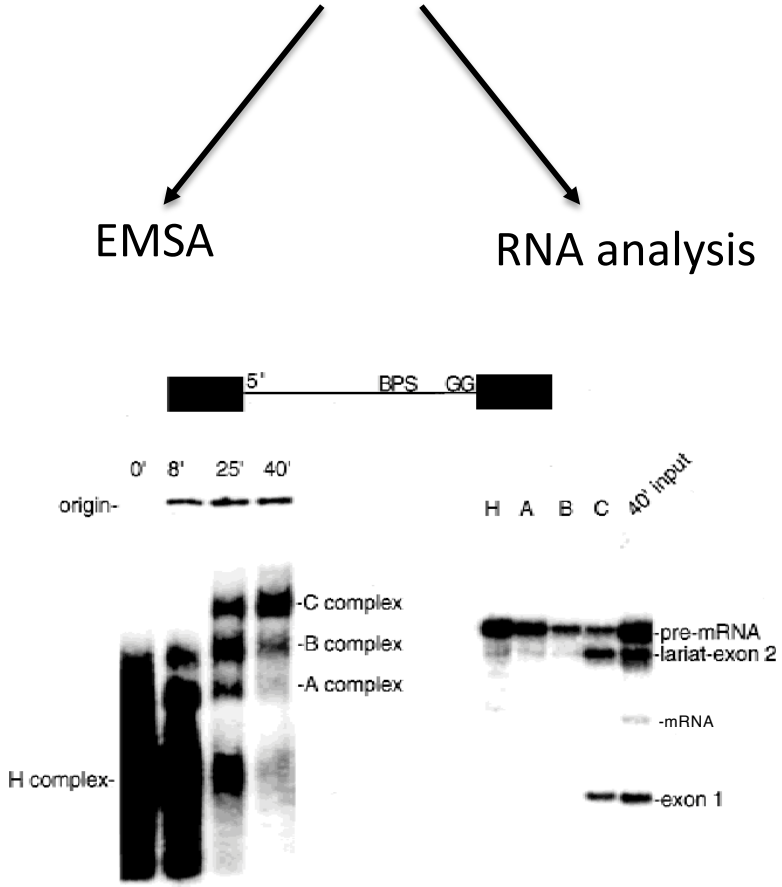


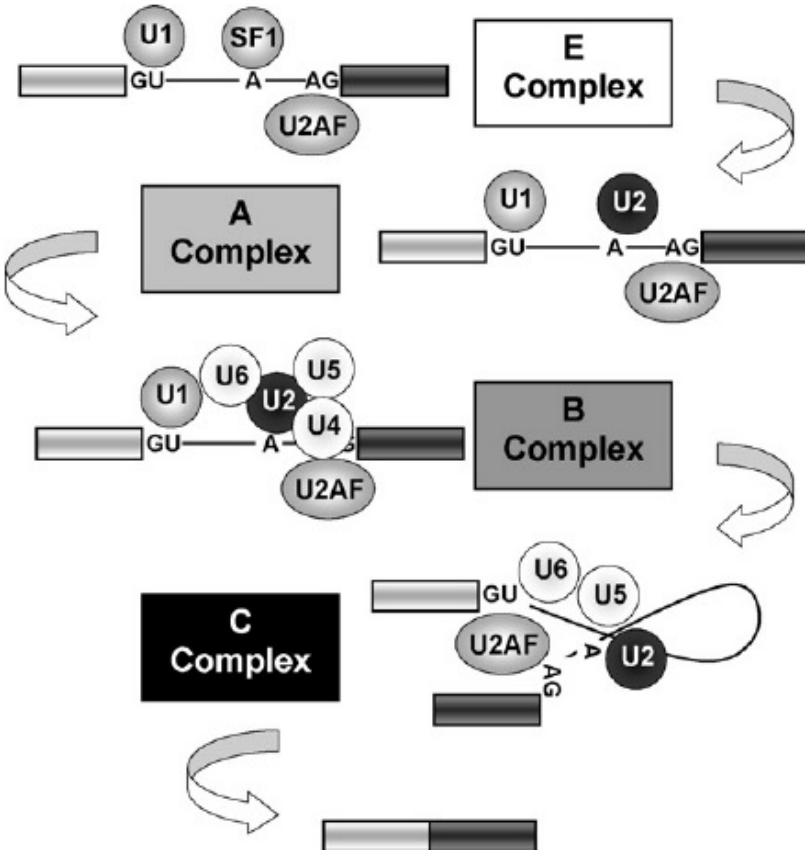
RNA analysis



In vitro analysis of splicing complexes formation

In vitro transcribed RNA + HeLa Nuclear Extract (HNE)





DNA oligonucleotides and RNase H can be used to deplete specific snRNA from HNE

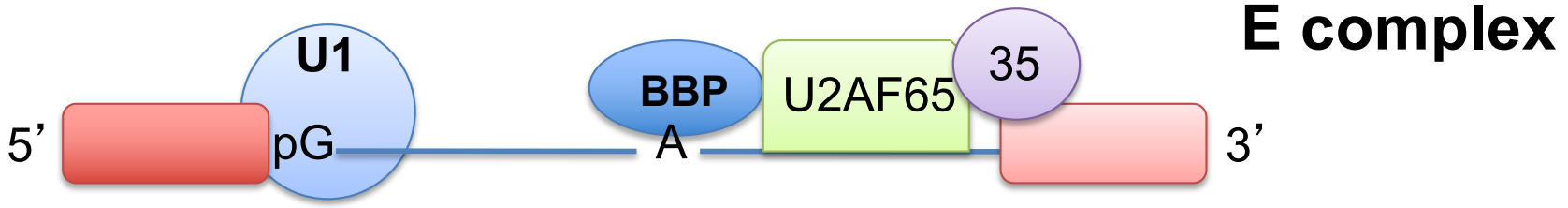
Assembly, rearrangement, and catalysis within the *spliceosome*: the splicing pathway

Assembly step 1

1. U1 recognizes 5’ splice site.

2. One subunit of U2AF binds to Py tract and the other to the 3’ splice site. The former subunits interacts with BBP and helps it to bind to the branch point.

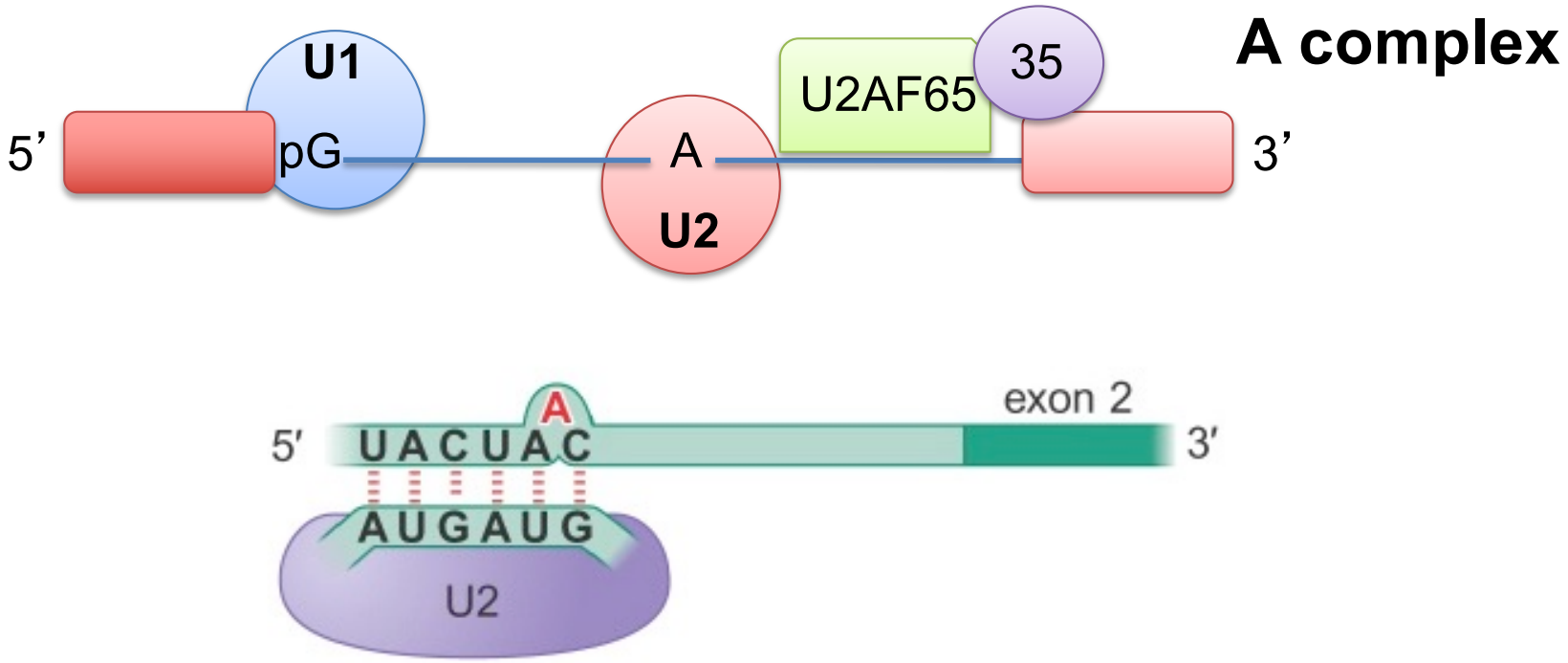
3. **Early (E) complex** is formed



Assembly step 2

1. U2 binds to the branch site, and then **A complex** is formed.

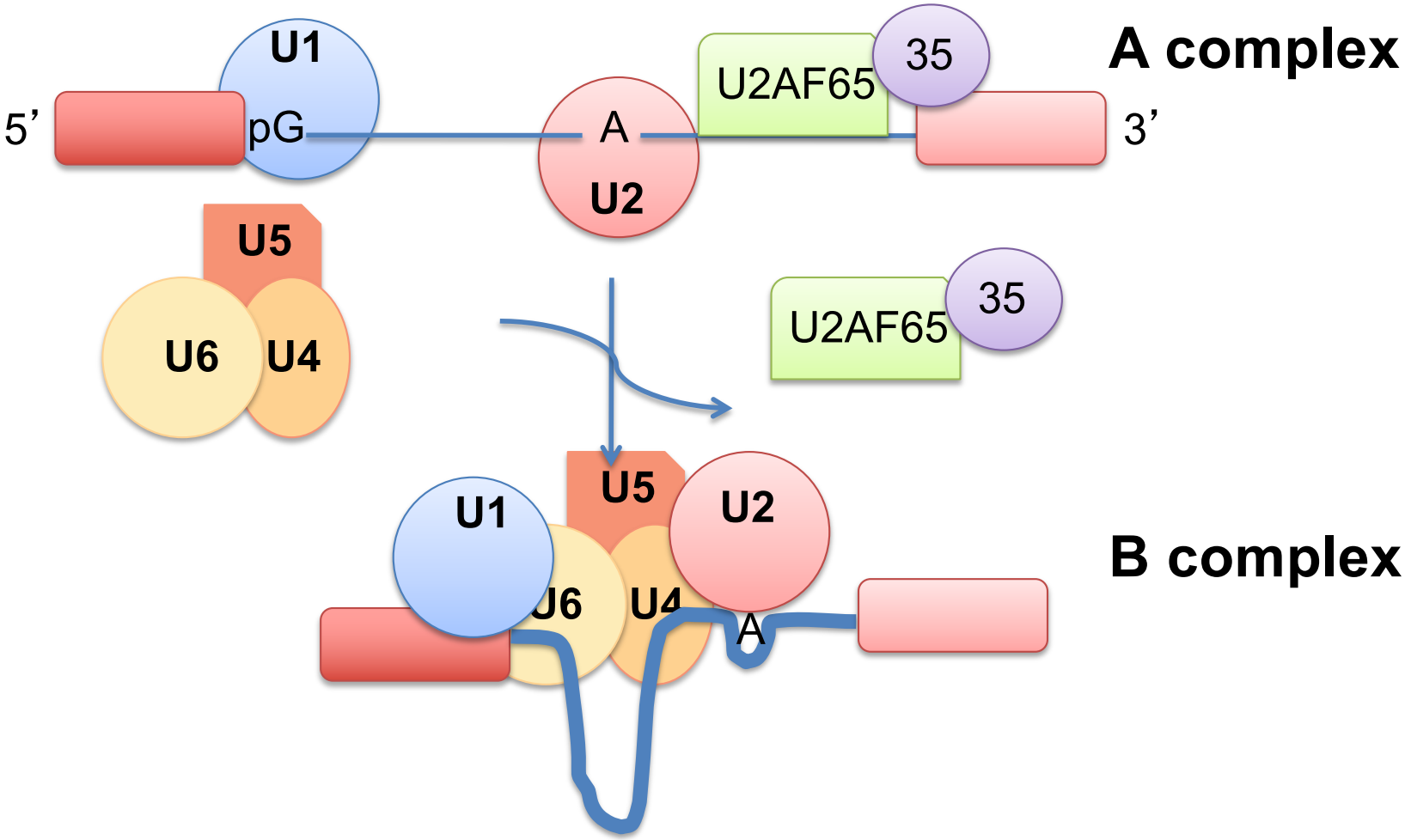
2. The base-pairing between the U2 and the branch site is such that the A is extruded. This A residue is available to react with the 5’ splice site.

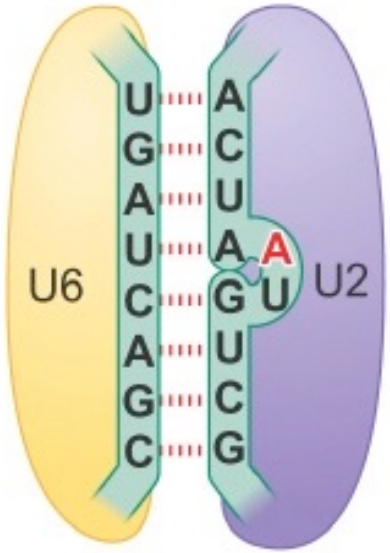


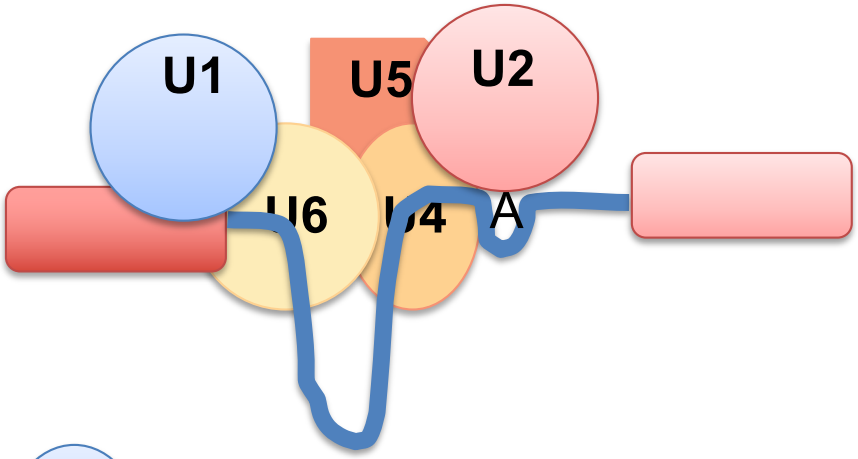
Assembly step 3

1. U4, U5 and U6 form the tri-snRNP Particle.

2. With the entry of the tri-snRNP, the A complex is converted into the **B complex**.

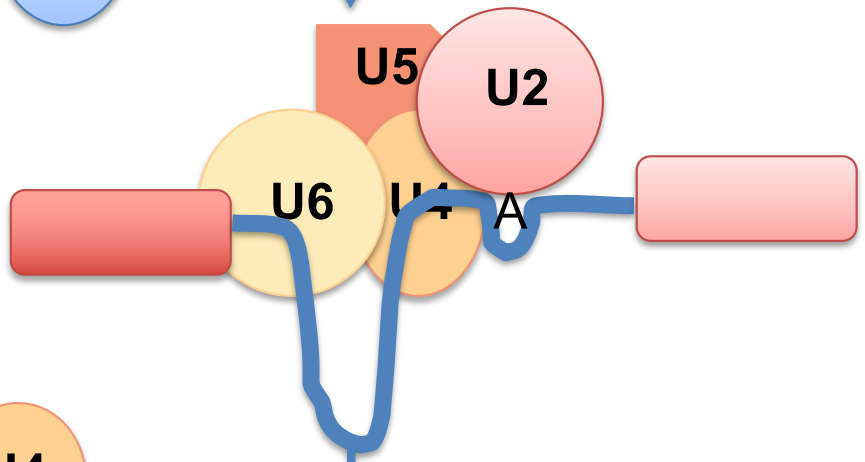






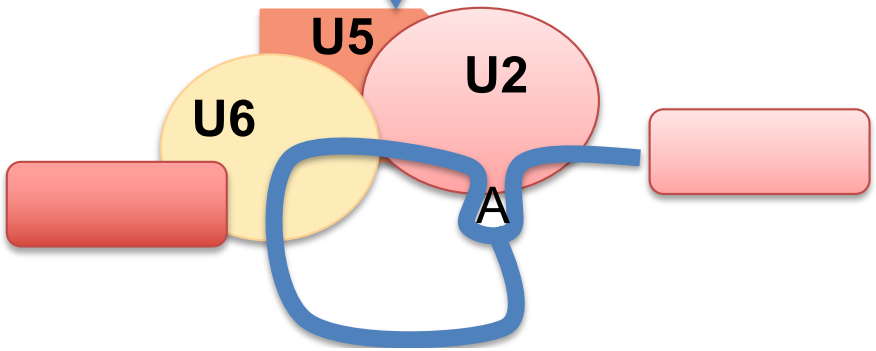
B complex

U1 dissociates from the complex



C complex in which the catalysis has not occurred yet

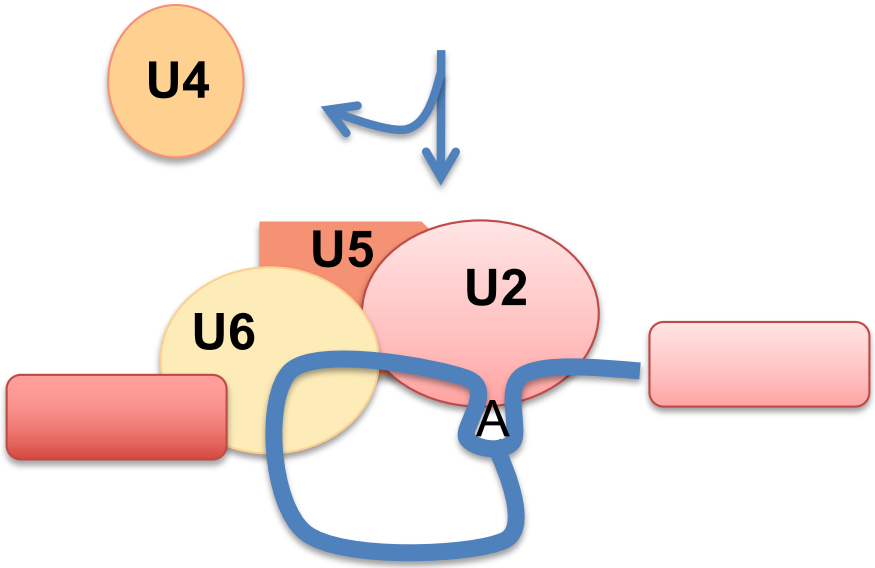
U4 dissociates from the complex



Catalysis Step 1

- Formation of the C complex produces the **active site**, with **U2 and U6 RNAs** being brought together.

- Formation of the active site juxtaposes the 5’ splice site of the pre-mRNA and the branch site, allowing the **branched A residue** to **attack** the 5’ splice site to accomplish the first transesterification reaction.

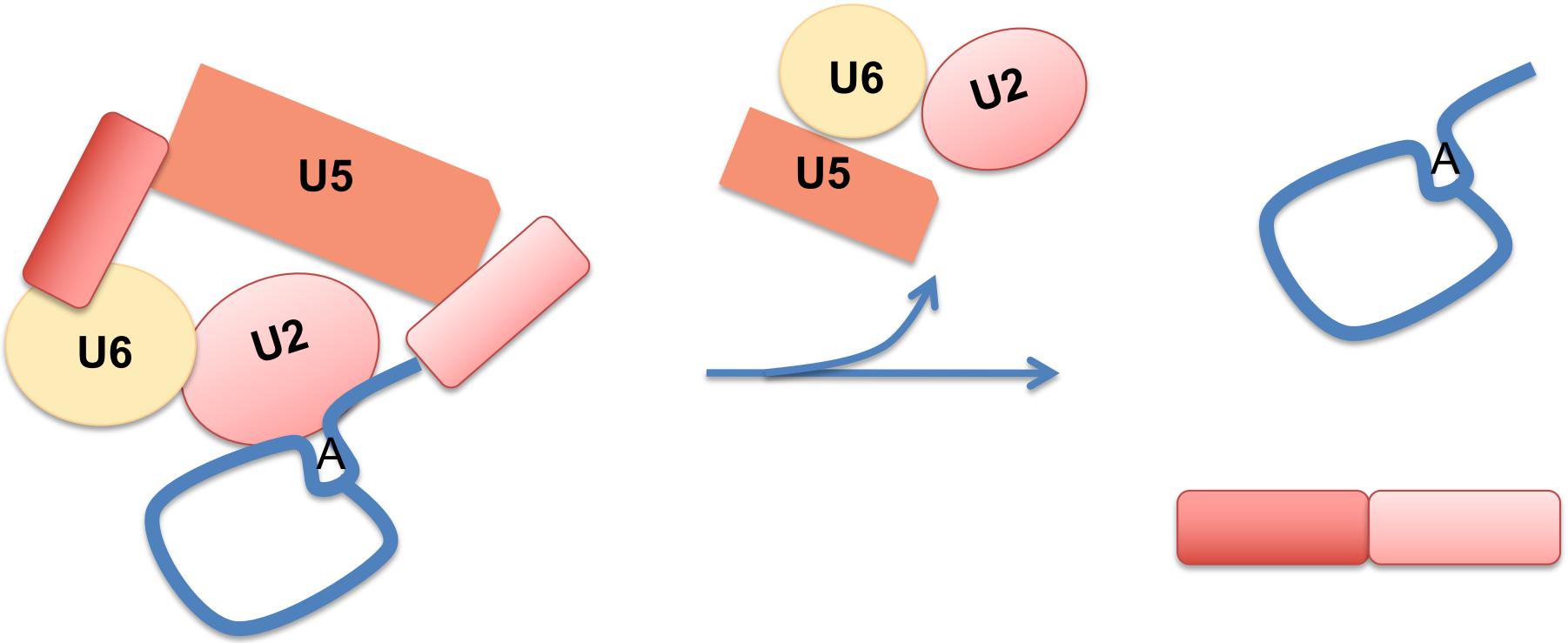


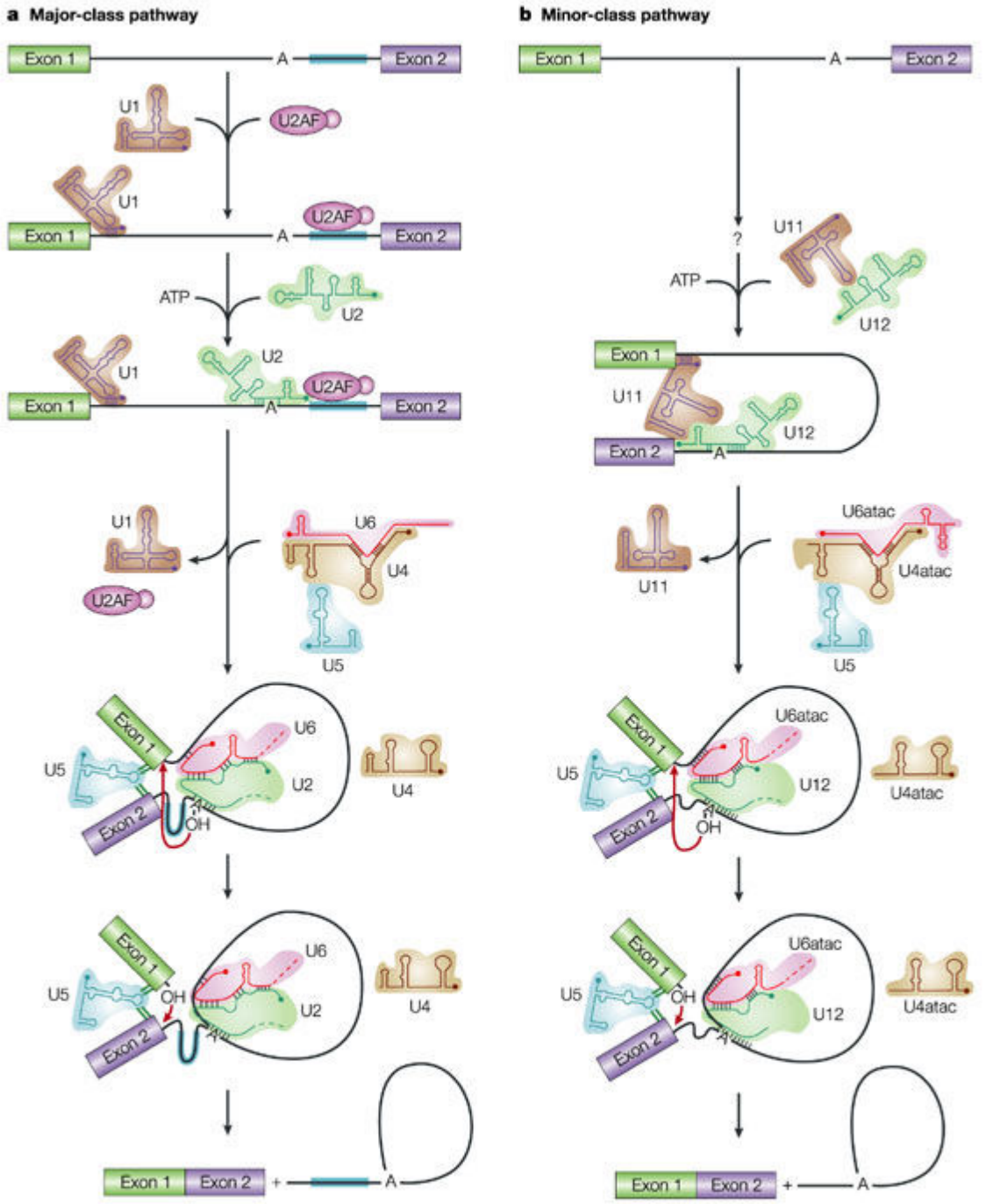
Catalysis Step 2

U5 snRNP helps to bring the two exons together, and aids the second transesterification reaction, in which the 3'-OH of the 5' exon attacks the 3' splice site.

Final Step

Release of the mRNA product and the snRNPs.





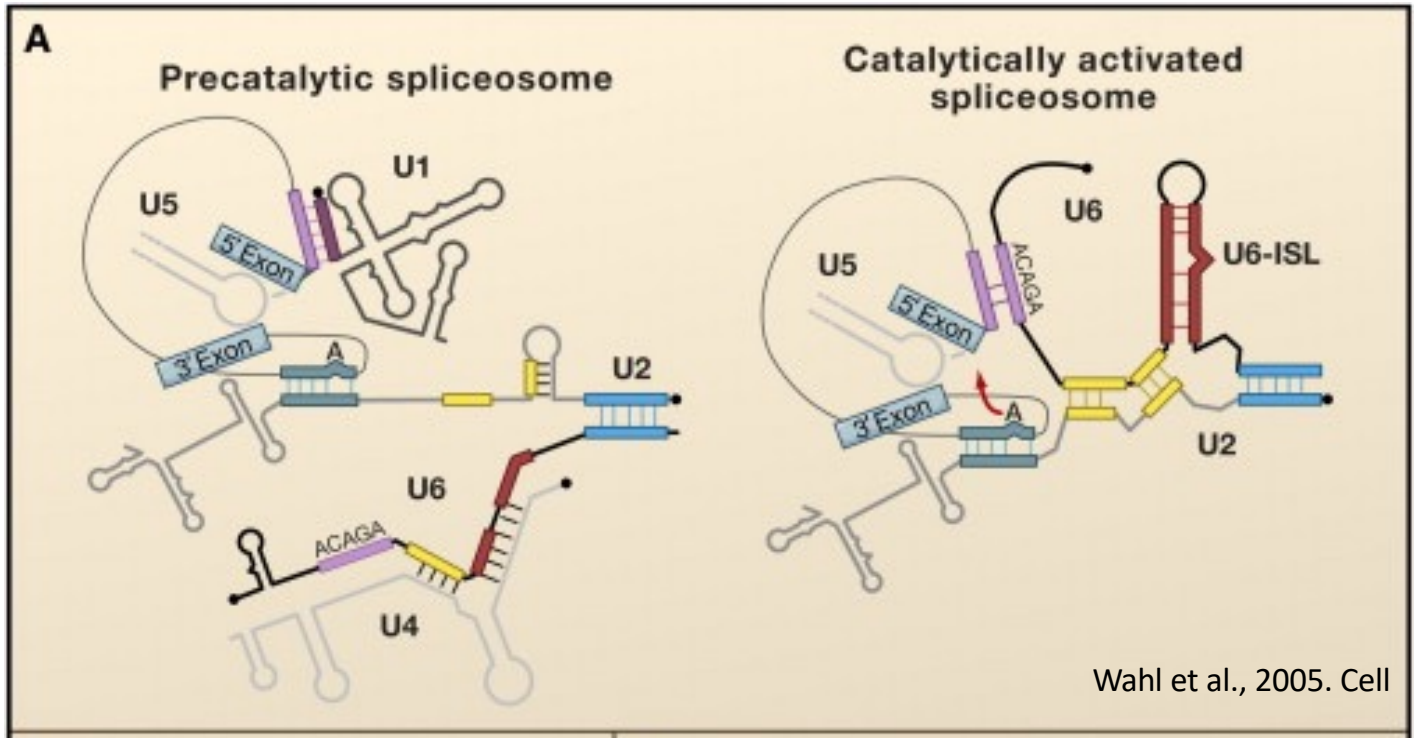
Nature Reviews | Molecular Cell Biology

The minor-class splicing reaction *in vitro* proceeds through the same two-step pathway as the major-class reaction.

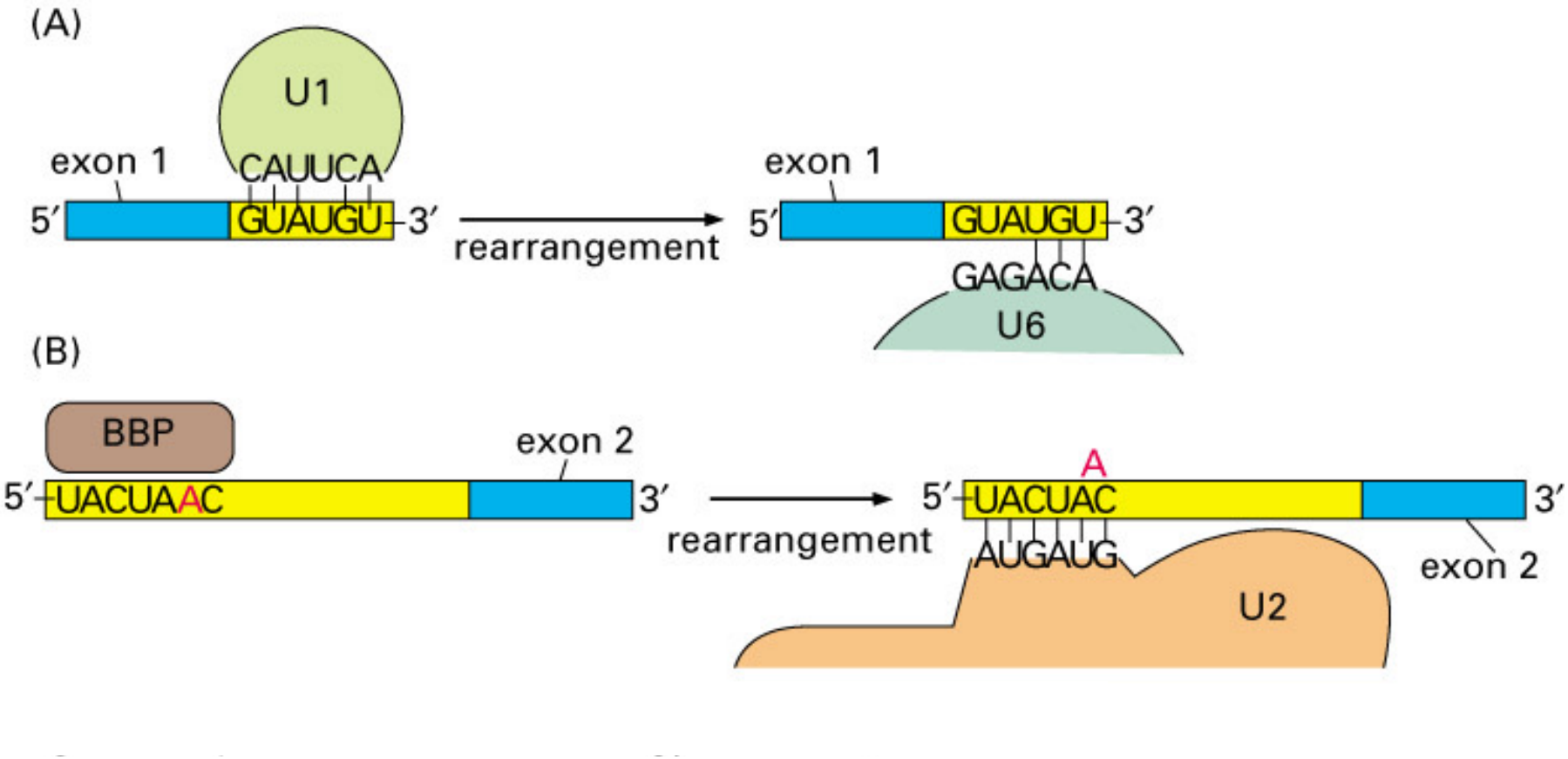
Spliceosomal RNA Network

- During activation of the spliceosome, regions of U6 and U2 (yellow or red) undergo major rearrangements. The 5′-end (black ball) of the U6 snRNA base pairs through its highly conserved ACAGAG motif to the 5′-splice site (SS), displacing U1.

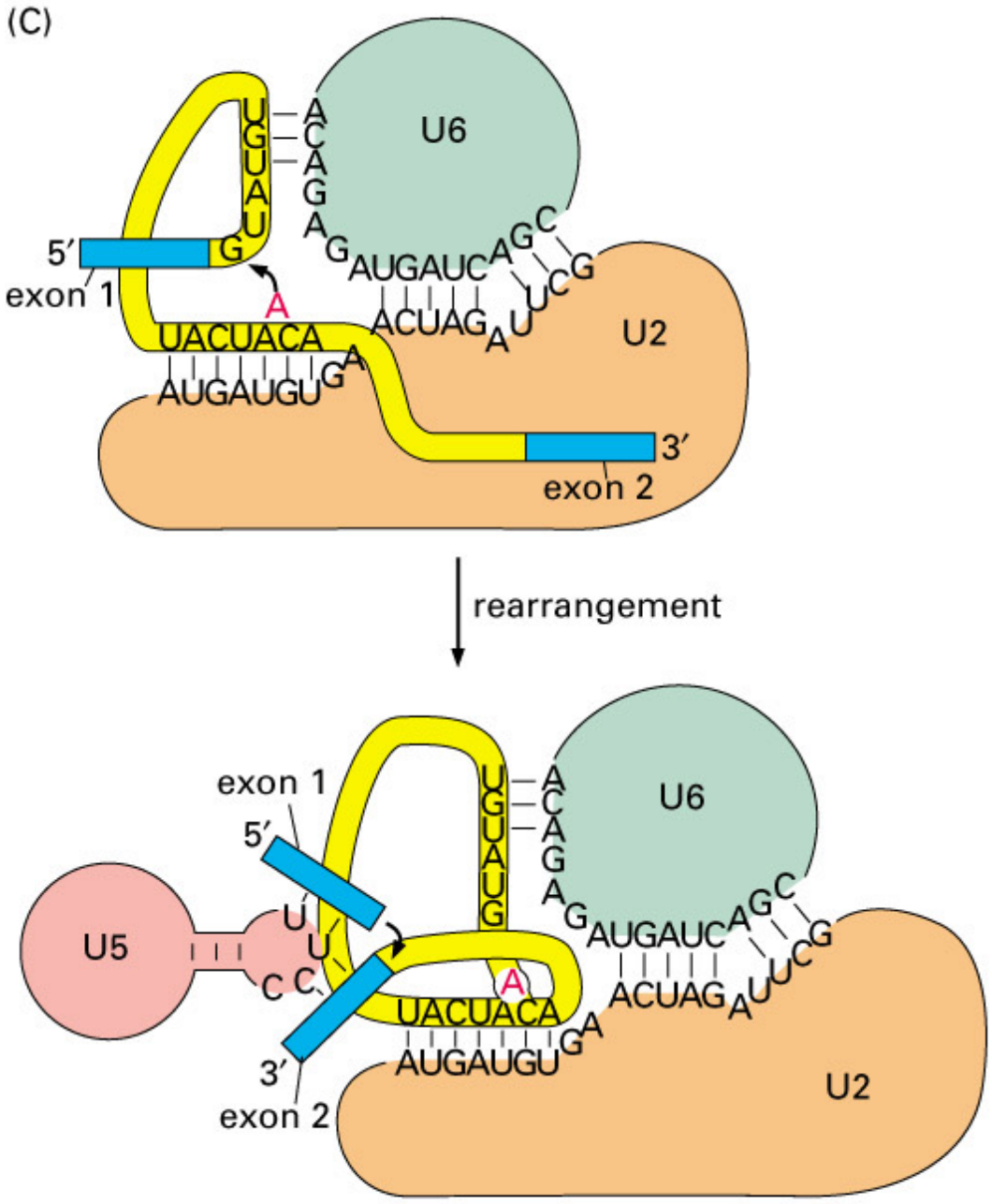
- U4 and U1 are destabilized or dissociate from the spliceosome at the time of activation and no longer are part of the spliceosome's RNA interaction network.



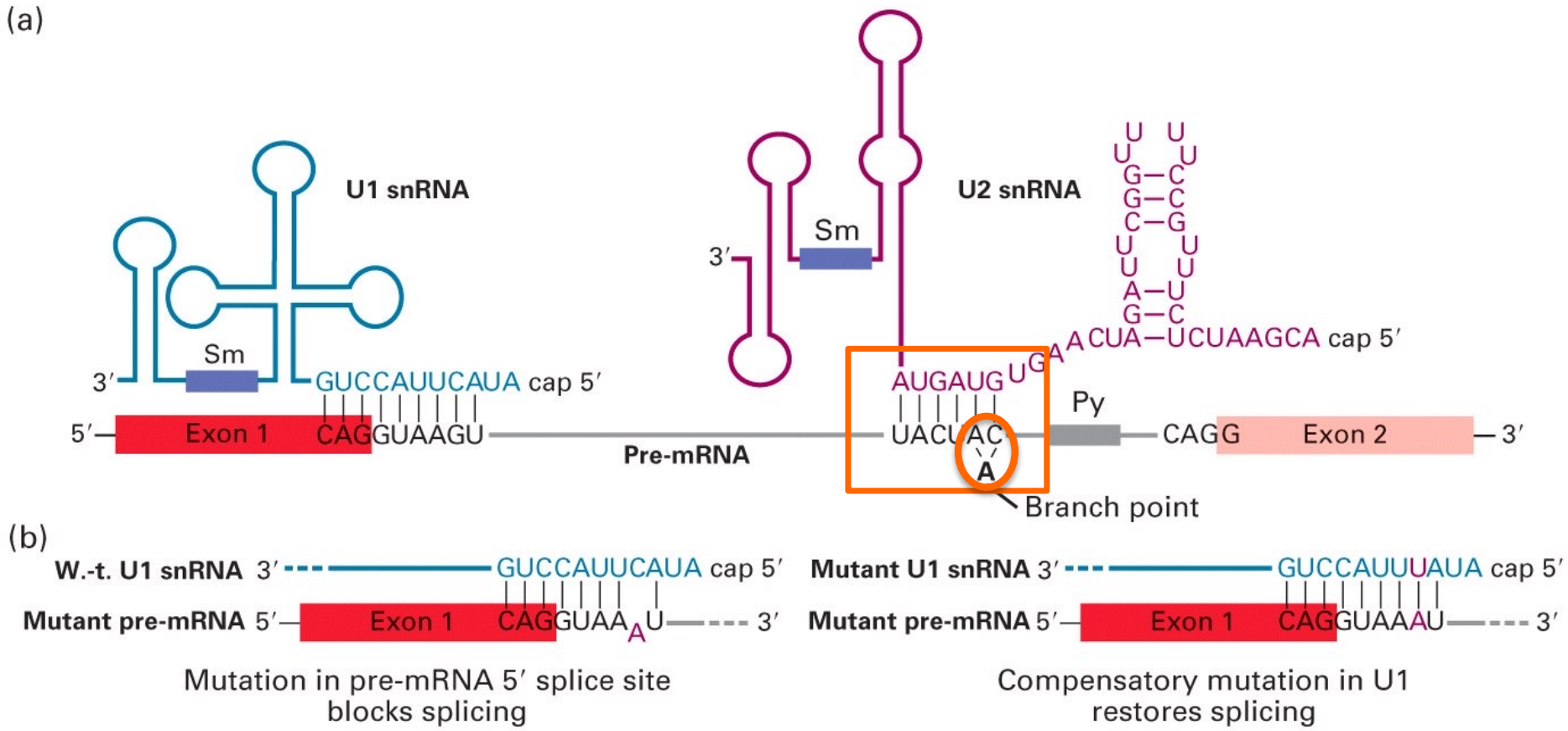
RNA-RNA Rearrangements - I



RNA-RNA Rearrangements - II

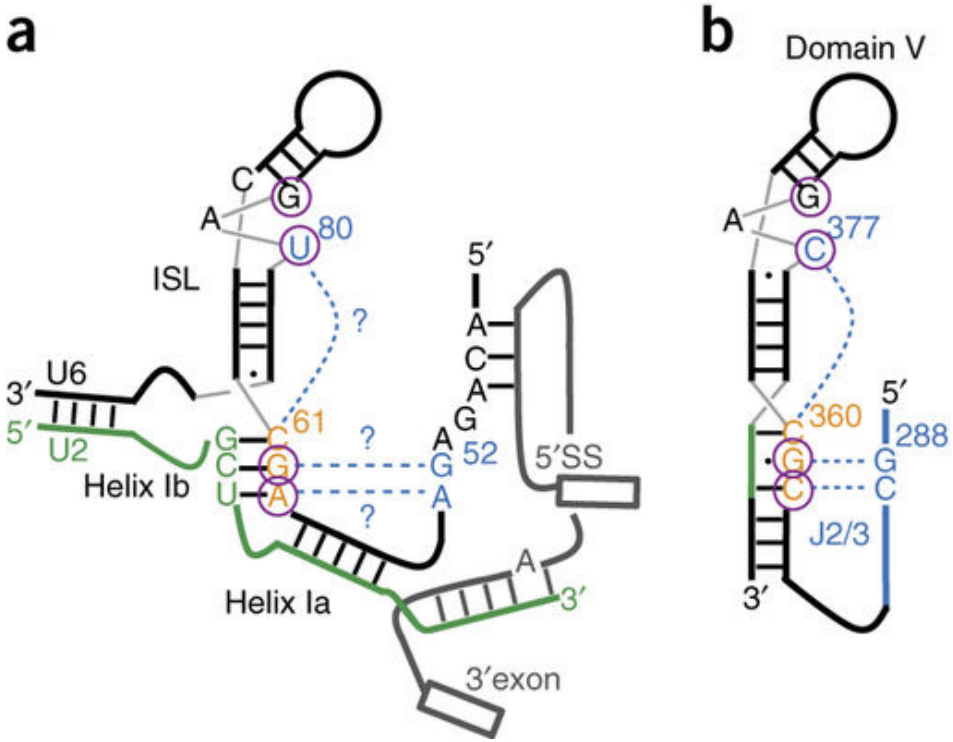


Pre-mRNA Splicing is Accelerated by RNA-RNA Basepairing



The spliceosome is a protein-assisted ribozyme

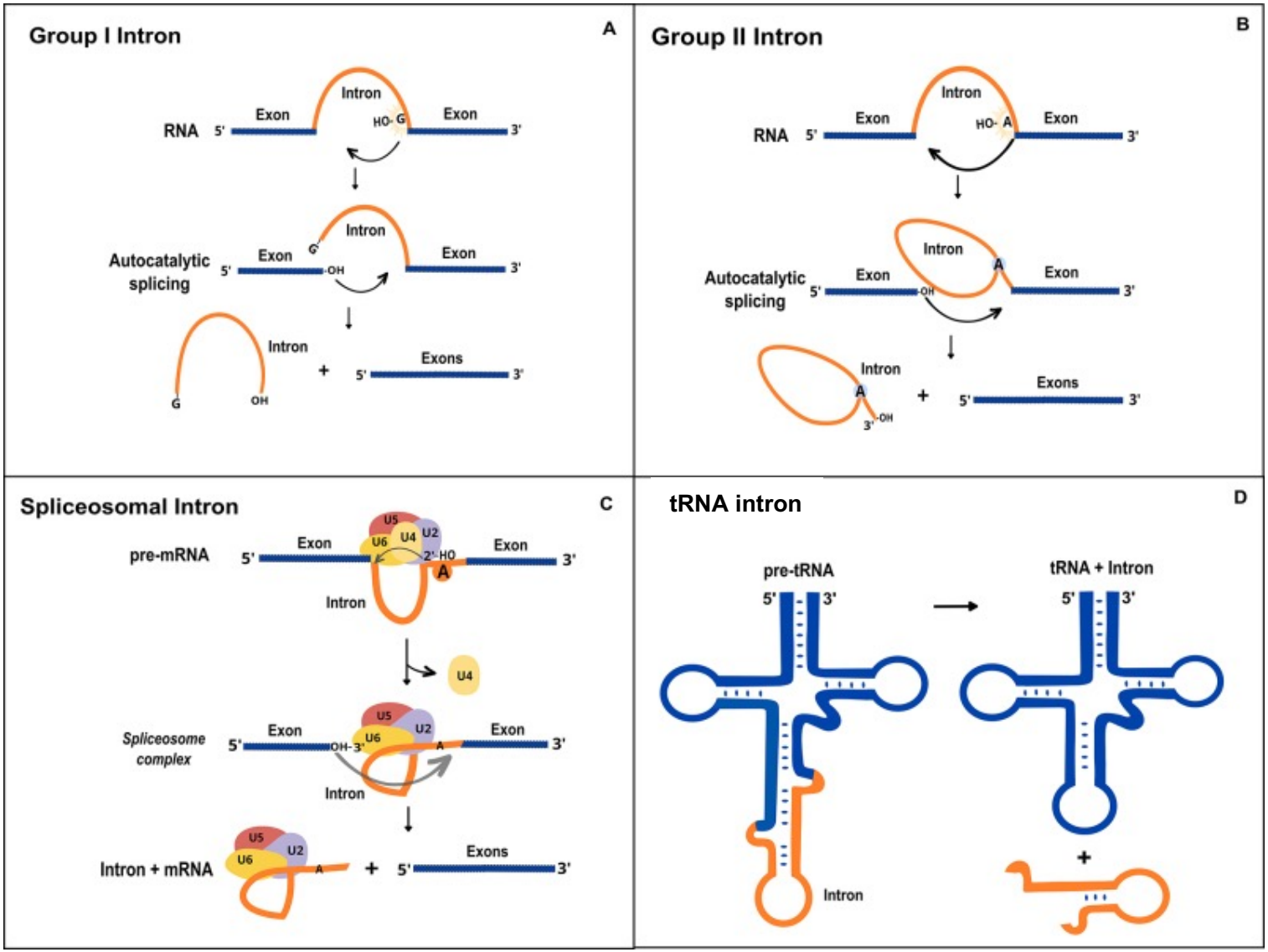
It has been proposed that spliceosomal introns and spliceosomal snRNAs arose as trans-acting fragments that were derived from ancestors of the self-splicing Group II Introns. The case for common ancestry is substantiated by the mechanistic similarities of spliceosomal and group-II self-splicing systems and they share numerous RNA structural motifs.



spliceosome

Group II introns

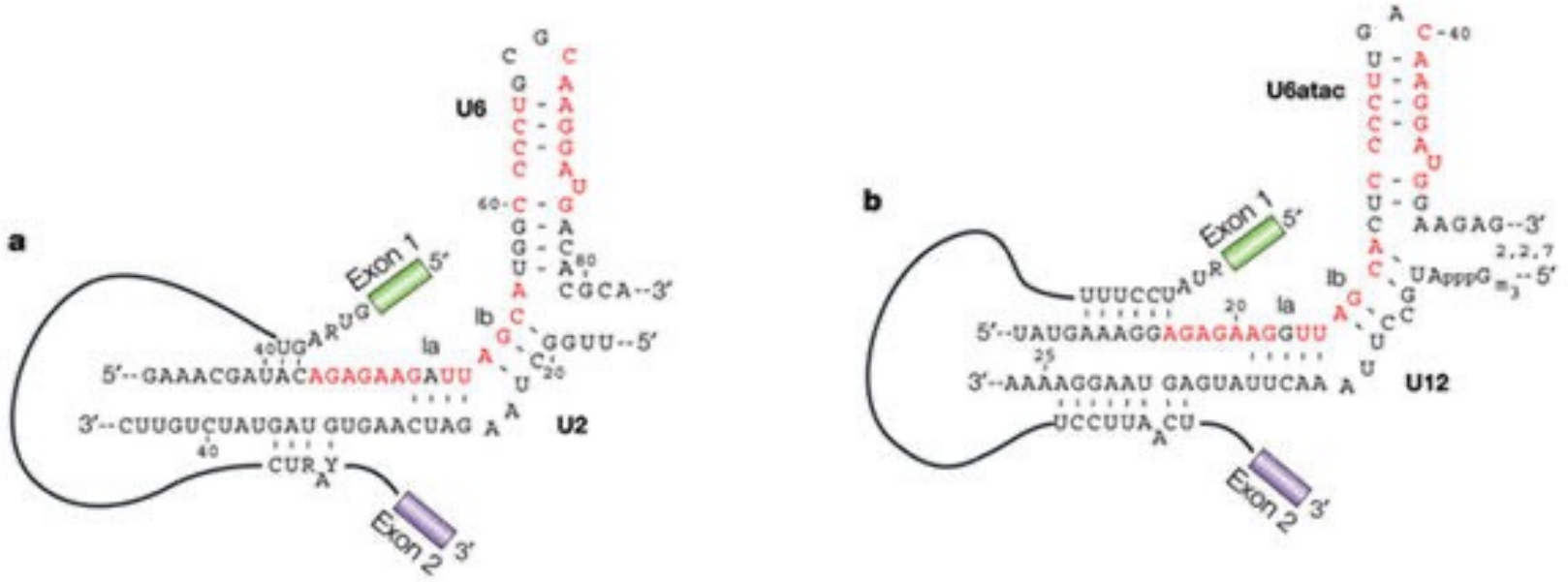
The four different classes of introns



group I and II introns are autocatalytic

The RNA catalytic core of the spliceosome

Major- and minor- spliceosomes contain an intramolecular stem-loop element (present in U6 and U6atac) that corresponds to a structure in the self-splicing group II intron, which is known as domain V. The importance of this stem-loop element in the catalytic core of both spliceosomes is underscored by the similarities in structure and sequence of this U6 region in organisms as diverse as humans, plants and fission yeast, and in human U6atac.

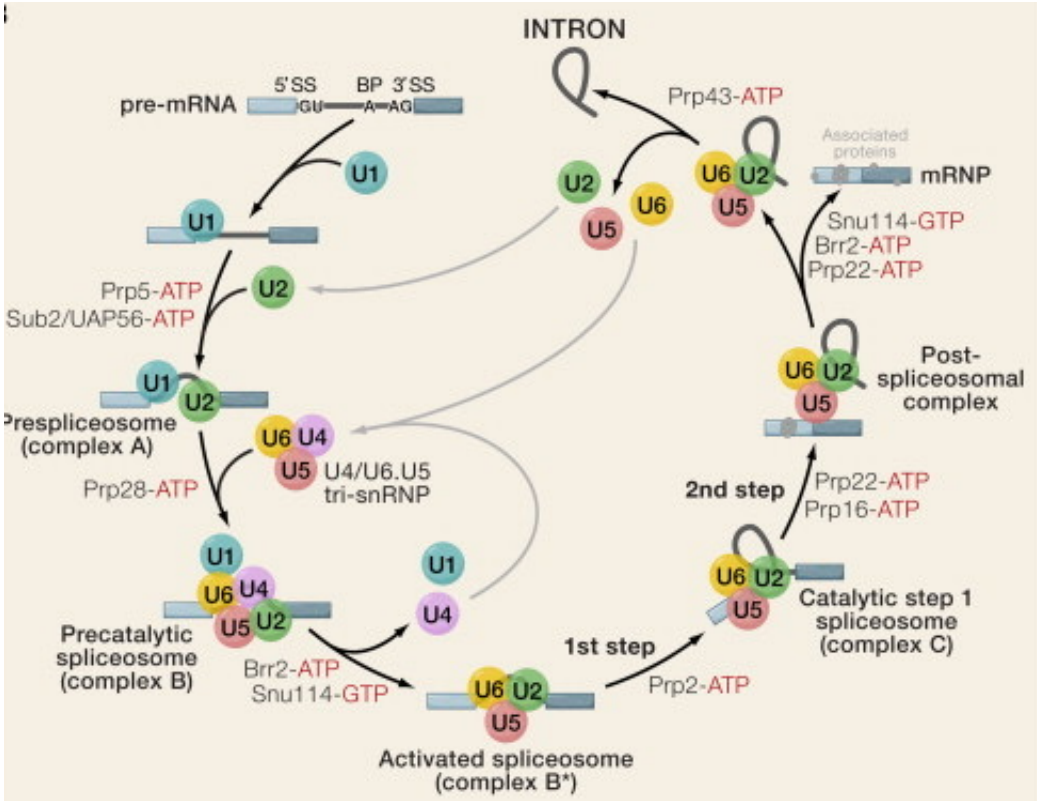


The main mechanistic differences between major- and minor-class splicing are likely to occur at the stage of intron recognition rather than during catalysis.

RNA helicases and splicing

Although splicing catalysis does not require ATP *per se*, essential **ATP-dependent helicases** assist the conformational transitions during assembly by establishing and rearranging RNA-RNA, RNA-protein and protein-protein interactions.

RNA helicases have been functionally linked to splicing. RNA helicases are necessary for the progression of the spliceosome through the splicing cycle, and they also function as proofreaders of the splicing process. Some of these factors increase the specificity of splicing and therefore play a part in splicing fidelity by rejecting suboptimal splicing substrates.



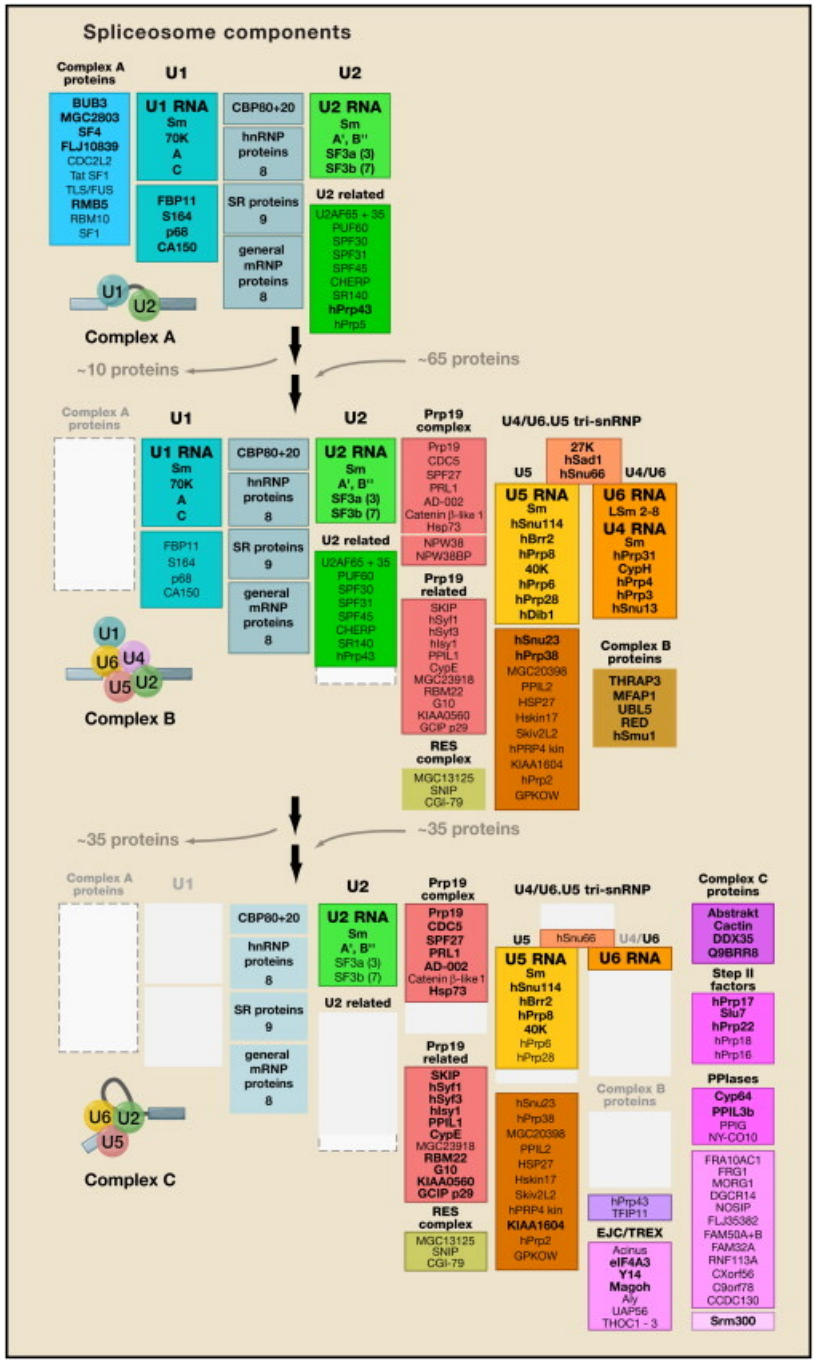
Wahl et al., 2005. Cell

The Spliceosome Is a Protein-Rich Machine

Human spliceosomes contain ~45 distinct snRNP-associated proteins and there is a dramatic exchange of proteins during spliceosome assembly and activation.

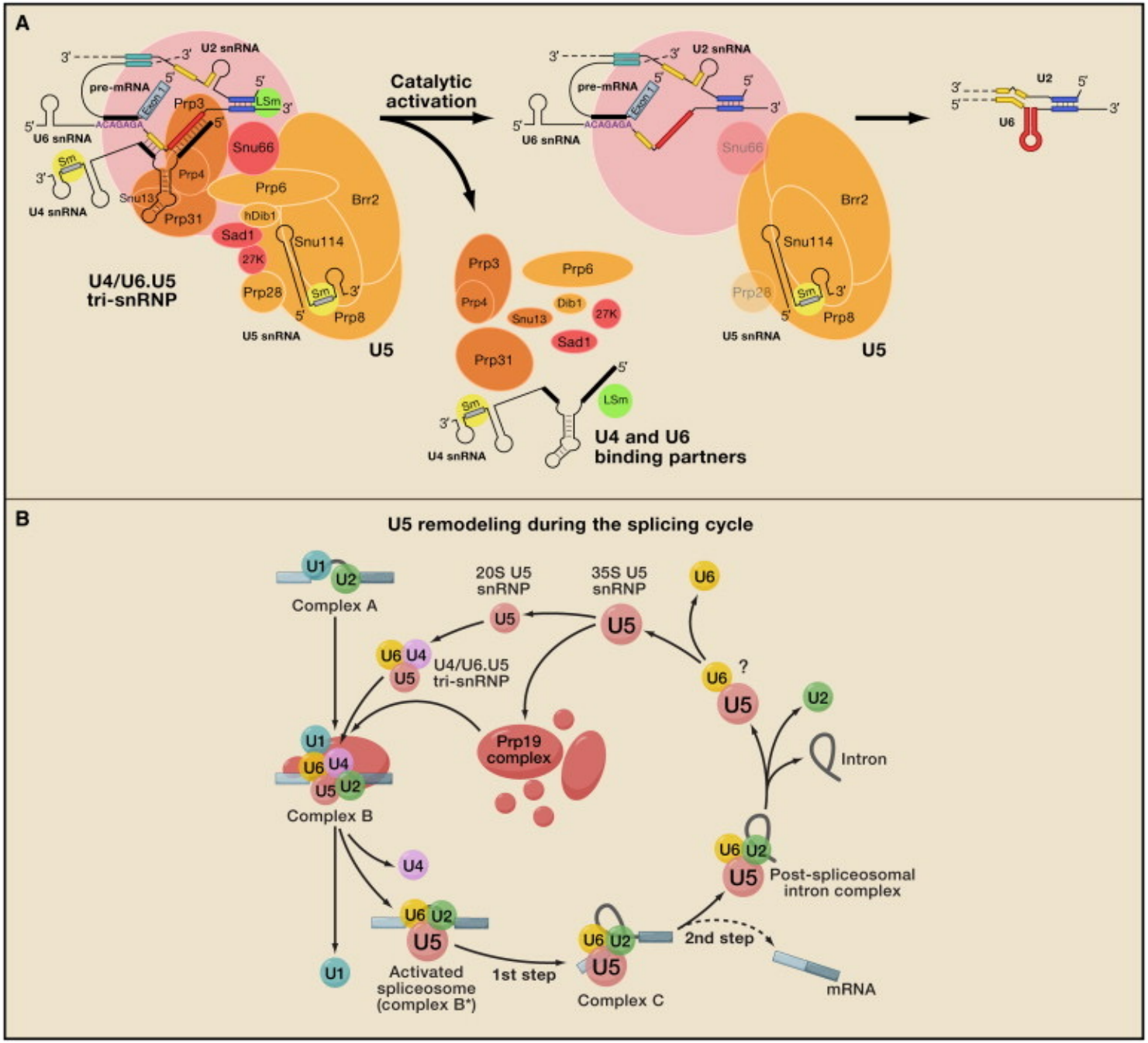
Human spliceosomes also contain numerous non-snRNP proteins, approximately 170!

A number of spliceosome-associated proteins are likely involved in coupling the splicing machinery to other molecular machines in the cell such as those responsible for transcription, 3'-end processing, or quality control.

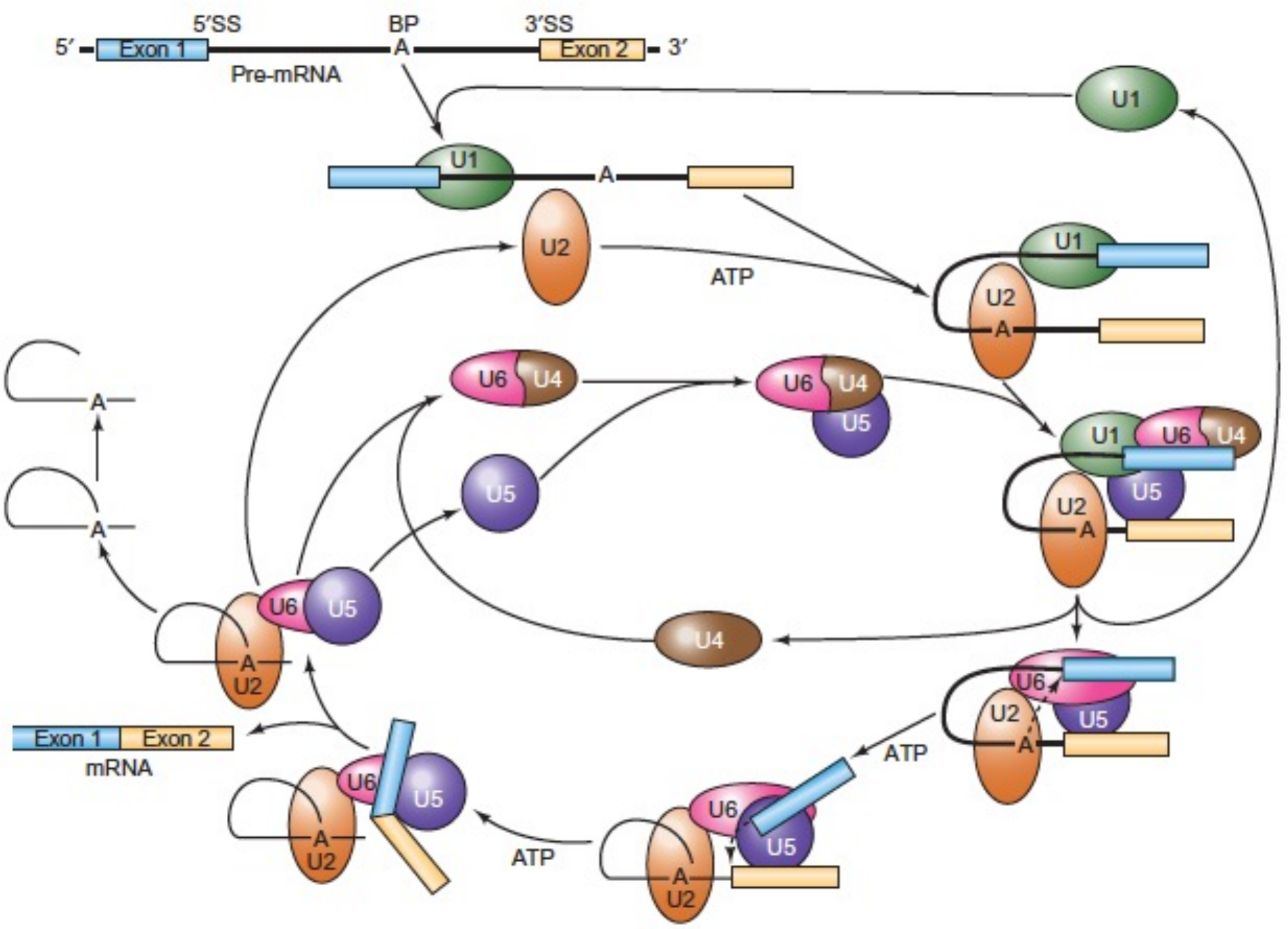


Wahl et al., 2005. Cell

The U4/U6/U5 tri-snRNP Is Extensively Remodeled during Splicing



The spliceosome cycle



Experimental Evidence for Co-Transcriptional Splicing

- **ChIP & Physical Interaction with RNAPII CTD:** Core splicing machinery (U1/U2 snRNPs, SR proteins) directly binds the phosphorylated **CTD** (Ser5-P/Ser2-PSer5-P/Ser2-P) of elongating RNAPII.

- **Nascent RNA Sequencing:** Detection of fully spliced exon-exon junctions within chromatin-bound, nascent RNA fractions prior to cleavage and polyadenylation.

- **RNAPII Mutant Experiments (Kinetic Model of Alternative Splicing)**

Control of Co-Transcriptional Splicing

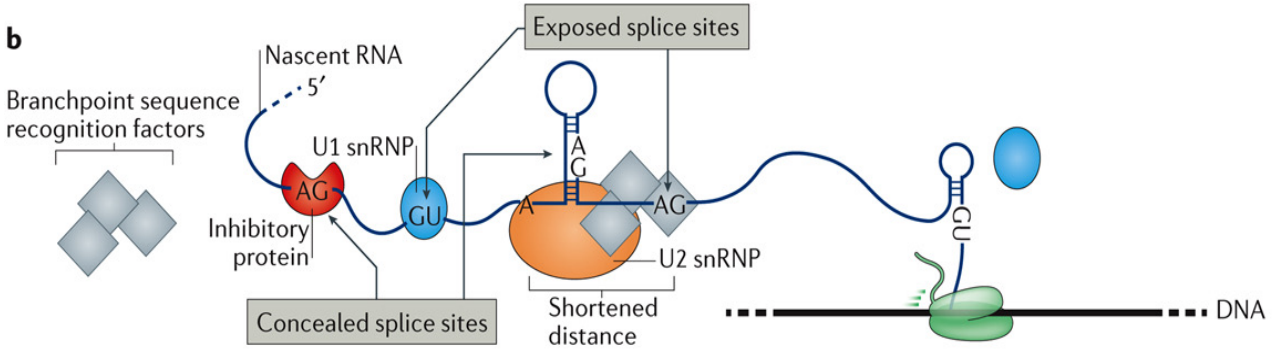
•**Transcription Rate :** The speed of RNAPII dictates which portion of the nascent pre-mRNA is exposed for **spliceosomal inspection** at any given moment.

•**Factor Access:**

- •**DNA/RNA Folding & Marks:** Dynamic folding states and chemical modifications alter protein-binding profiles and transcription speed.
- •**Target Accessibility:** Controls spatial exposure of splice sites and branch point sequences dictating splicing factor recruitment.

•**Pol II-CTD PTMs:** Local variations in CTD post-translational modifications coordinate the spatial-temporal loading of RNA processing factors.

•**Chromatin Architecture:** Epigenetic marks and nucleosome positioning directly modulate RNAPII elongation rates to fine-tune co-transcriptional processing.



Control of Co-Transcriptional Splicing

The close association between chromatin and splicing leads to a significant cross-talk in which pre-mRNA splicing influences the chromatin state, while the chromatin state facilitates pre-mRNA splicing and controls the rate of alternative exon inclusion.

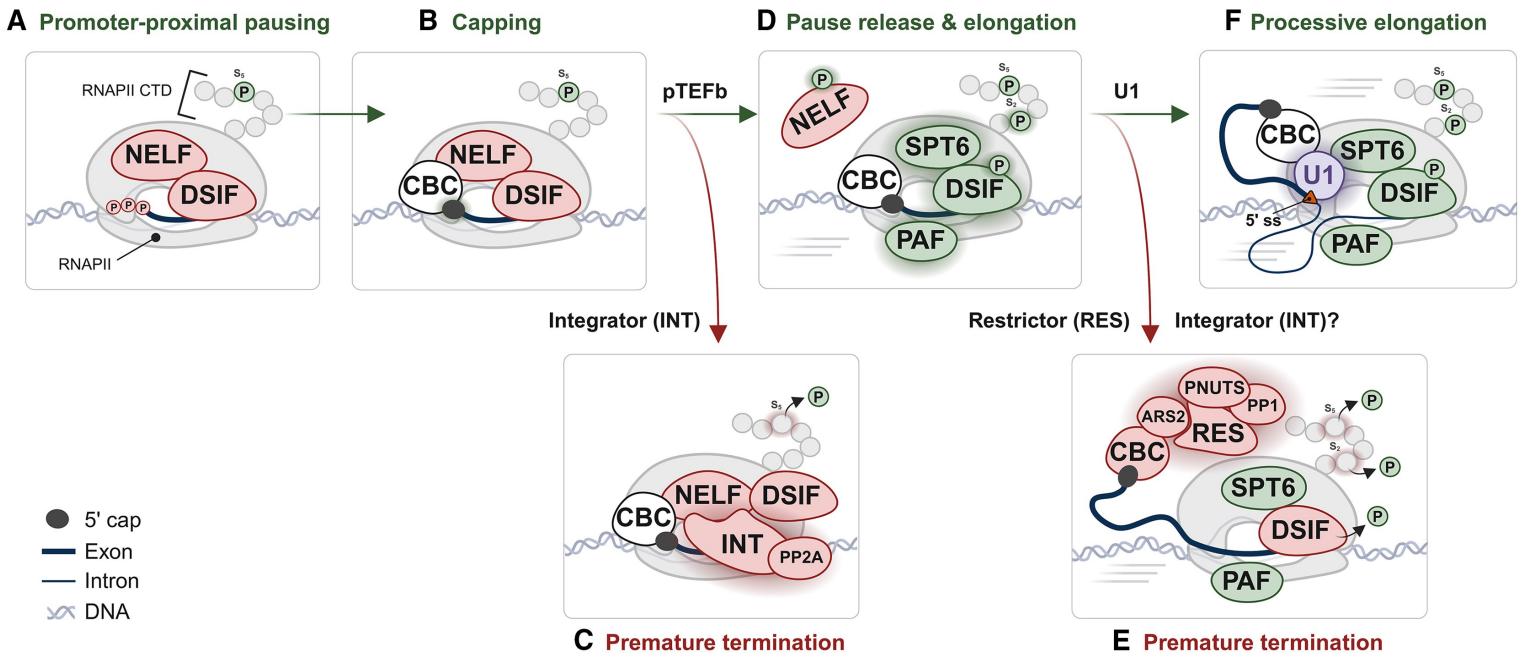
Two models explain how pre-mRNA splicing is controlled by the chromatin state

1- **'kinetic' model**: the chromatin state controls exon inclusion by modulating the rate of transcription elongation

2- **'recruitment' model**: histone modifications associated with the exons in cooperation with the C-terminal domain of RNAPII function as a 'landing pad' to load the spliceosome components necessary for splice site recognition onto the emerging nascent transcript

U1 snRNP promotes RNA Pol II elongation

Capping defects or termination cues recruit the INTAC complex (Integrator + PP2A); INTS11 cleaves the nascent transcript while PP2A dephosphorylates the CTD to enforce premature termination. In early elongation, non-processive RNAPII producing short non-coding RNAs (e.g., PROMPTs) is terminated by the Restrictor complex (ZC3H4/WDR82/ARS2) via PP1/PNUTS dephosphorylation. To transition into processive gene elongation, recognition of the first 5′ splice site by U1 snRNP - tethered to Pol II (RPB2/12) and SPT6 - prevents premature termination and locks RNAPII into a highly processive state.



Yoon et al., 2026 Mol Cell