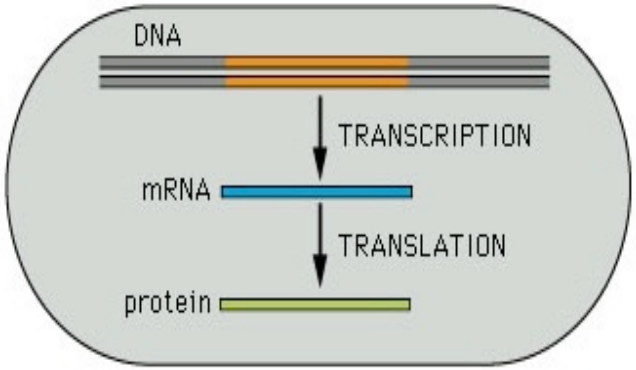
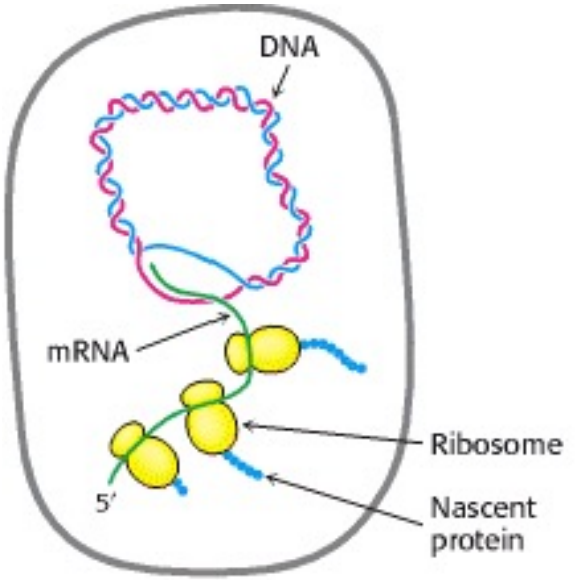
Messenger RNA maturation

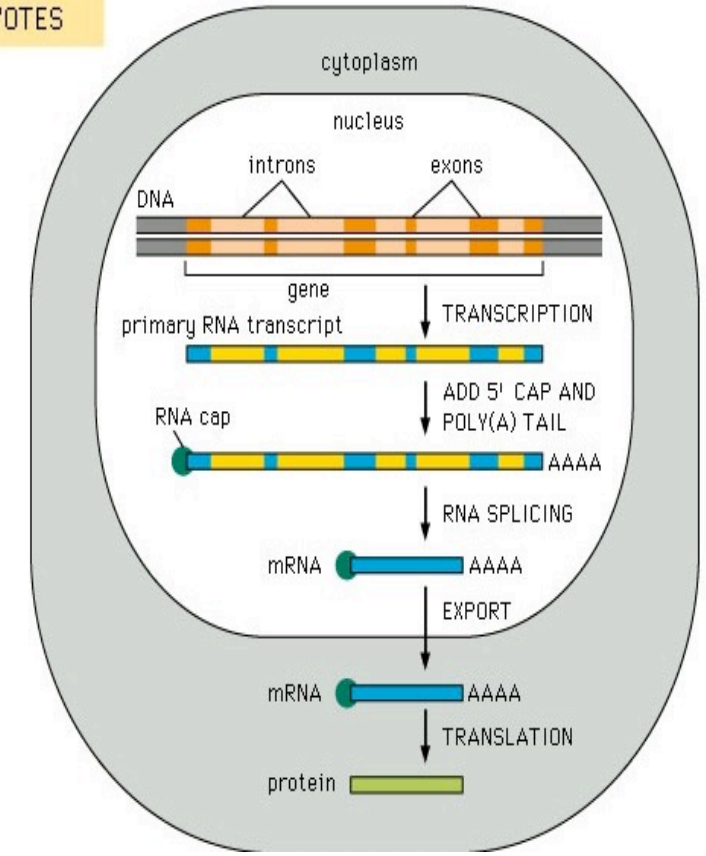
PROCARYOTES

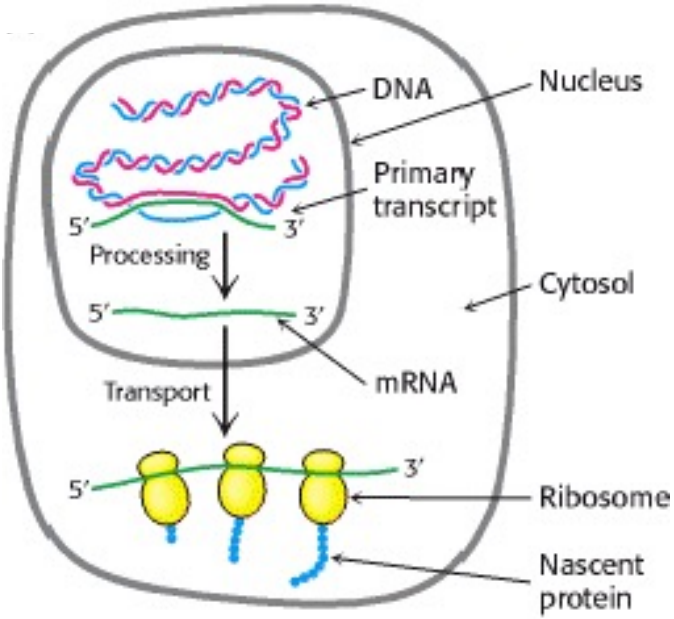




PROKARYOTE

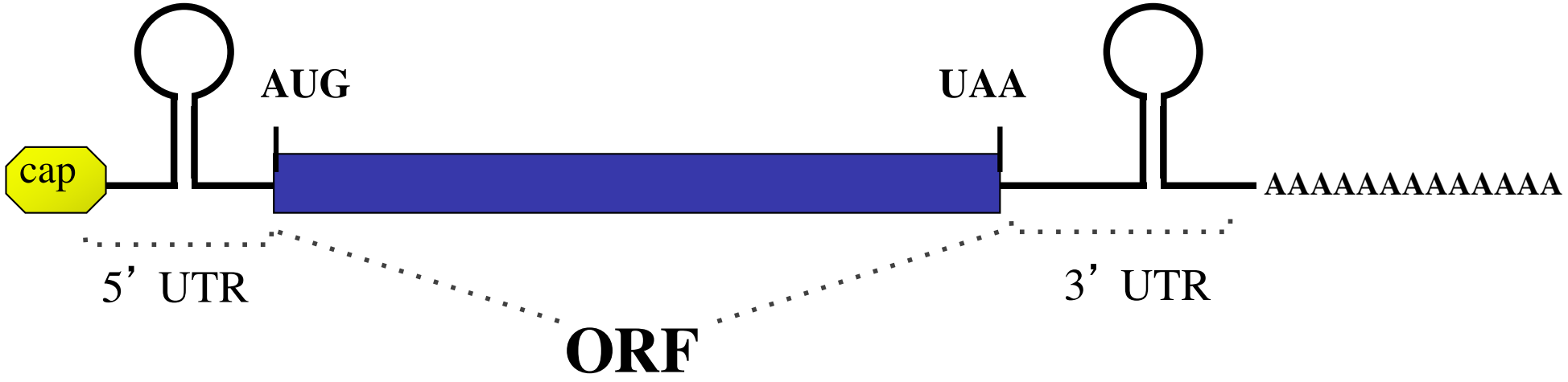
EUCARYOTES





EUKARYOTE

mRNA structures



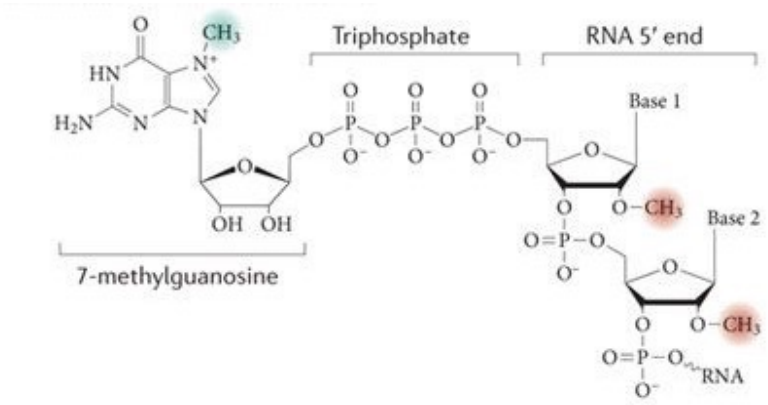
mRNA Capping

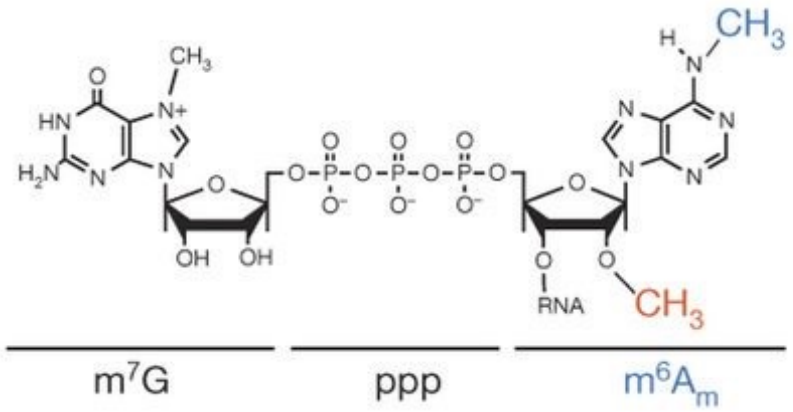
Why the Cap structure is important?

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

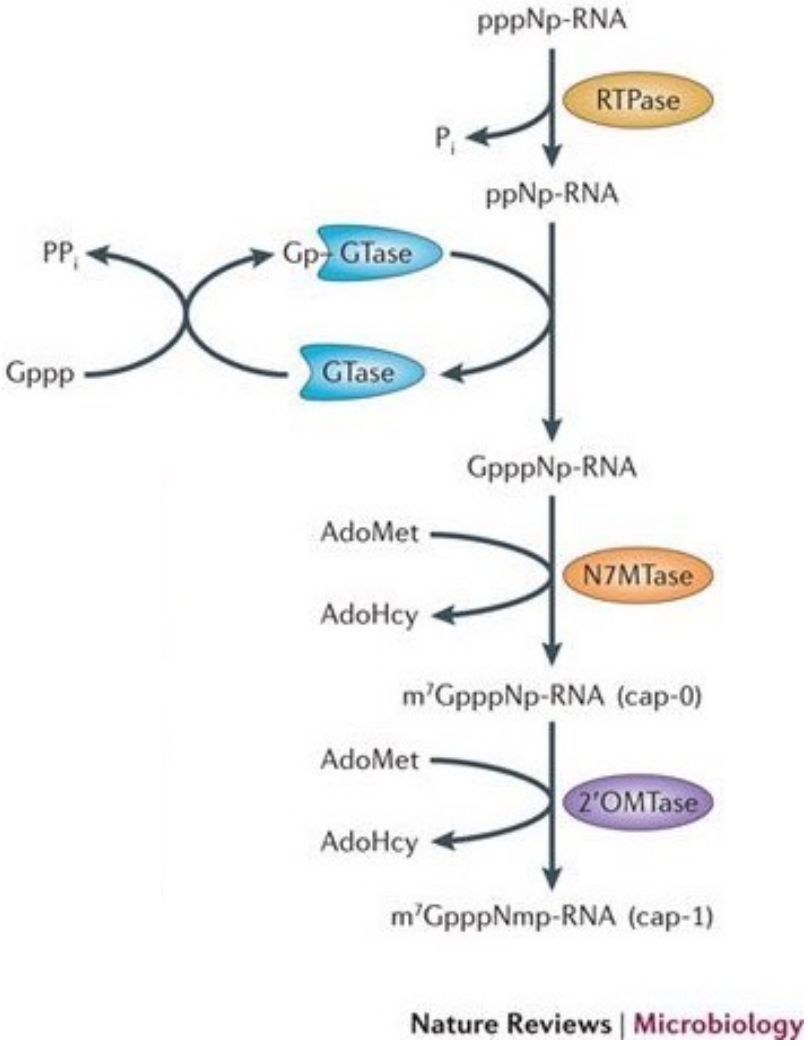
The mRNA cap consists of a **7-methylguanosine** linked to the 5′ nucleoside of the mRNA chain through a 5′–5′ triphosphate bridge. The methyl group at the N^7 position of the guanosine is shaded green, and the **2′-*O*-methyl groups** of the first and second nucleotide residues, forming the **cap-0**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





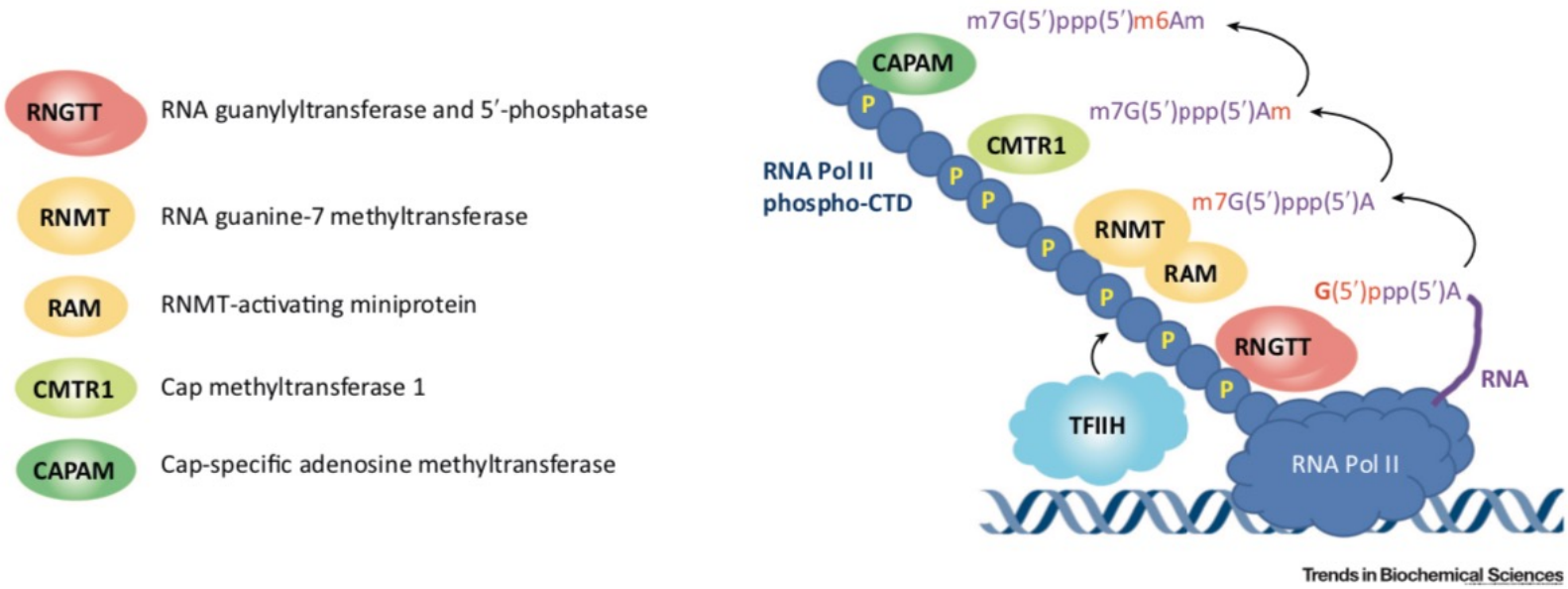
Capping reaction

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-N7)-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)**, generating the cap-1 structure ($m^7GpppNm_{2'\text{-}O}p$).



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 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. RNA is synthesised with a 5′ triphosphate. RNGTT is a triphosphatase and guanylyltransferase that adds the inverted guanosine cap to nascent RNA. The other enzymes, RNMT-RAM, CMTR1 and CAPAM, are methyltransferases.



Capping of viral RNA

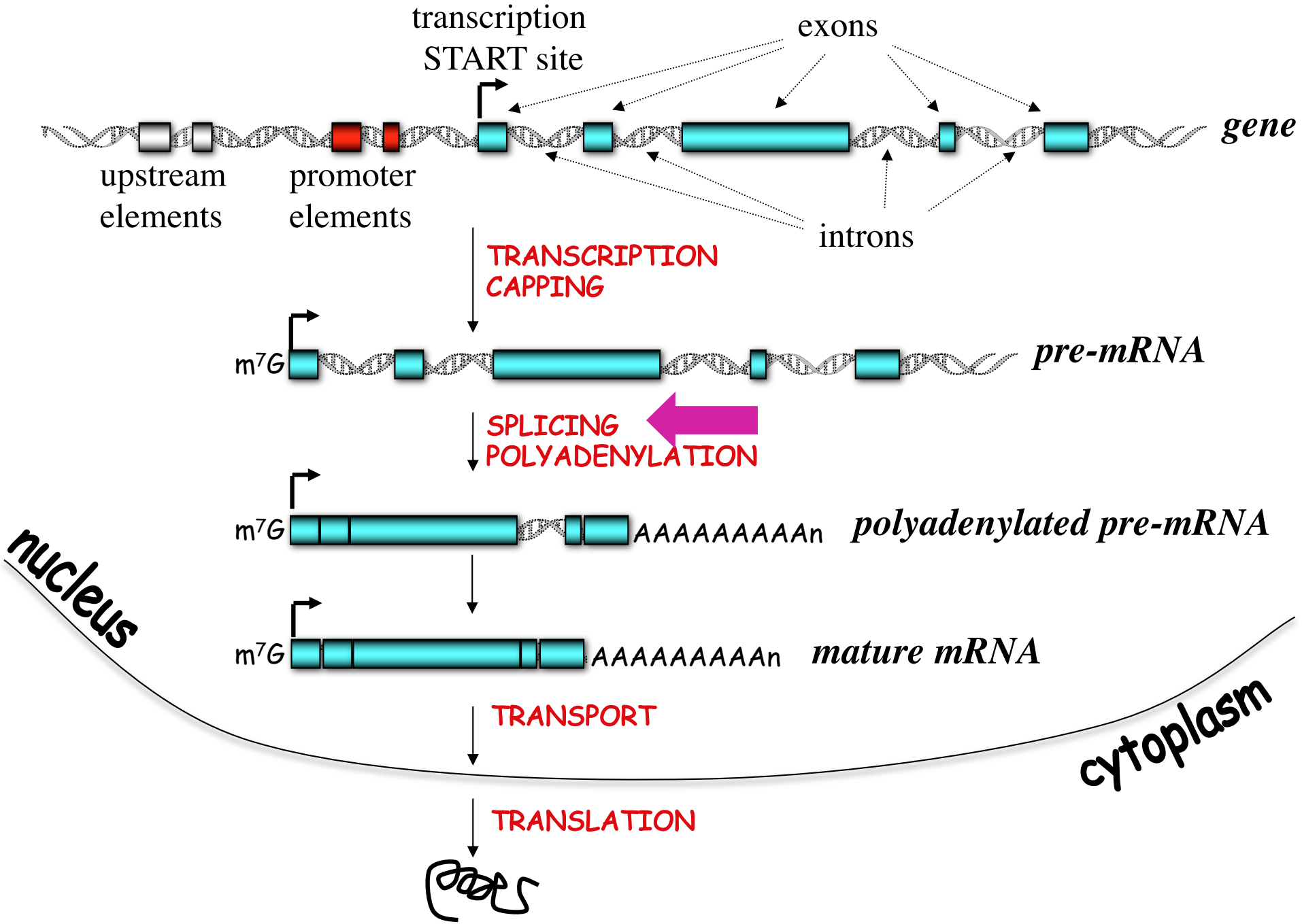
Viruses can acquire their cap structures using the cellular capping machinery or encode their own viral capping machineries that adopt the canonical pathway. The RNAs capped by viral enzymes are indistinguishable from cellular mRNA and can thus be translated into proteins by the cellular ribosomal machinery.

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

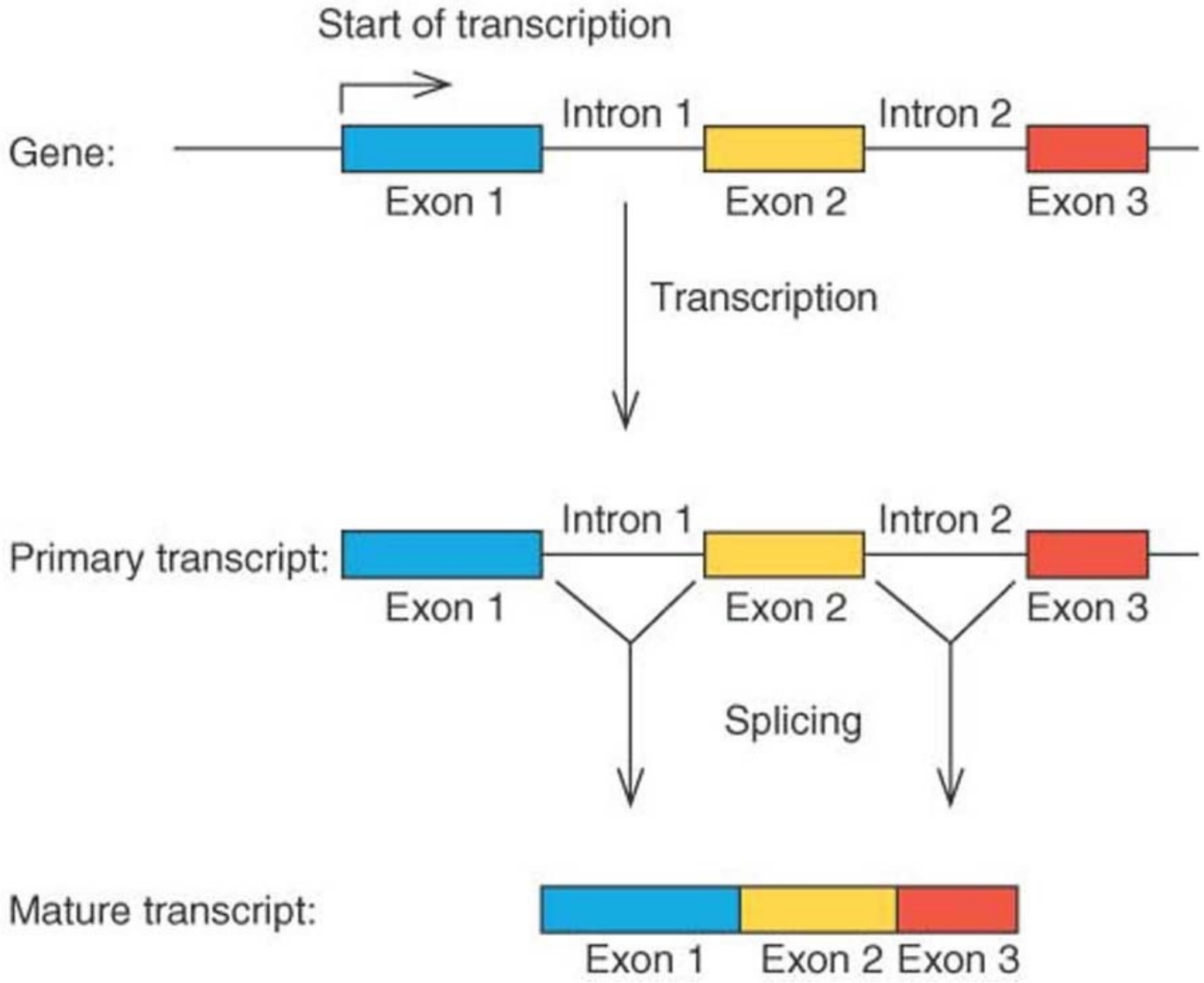
HIV
BTV, bluetongue virus;
HBV, hepatitis B virus;
HSV, herpes simplex viruses;
SARS CoV(2), severe acute respiratory syndrome coronavirus.
Smallpox virus
Papilloma Virus

Eucaryotic gene expression

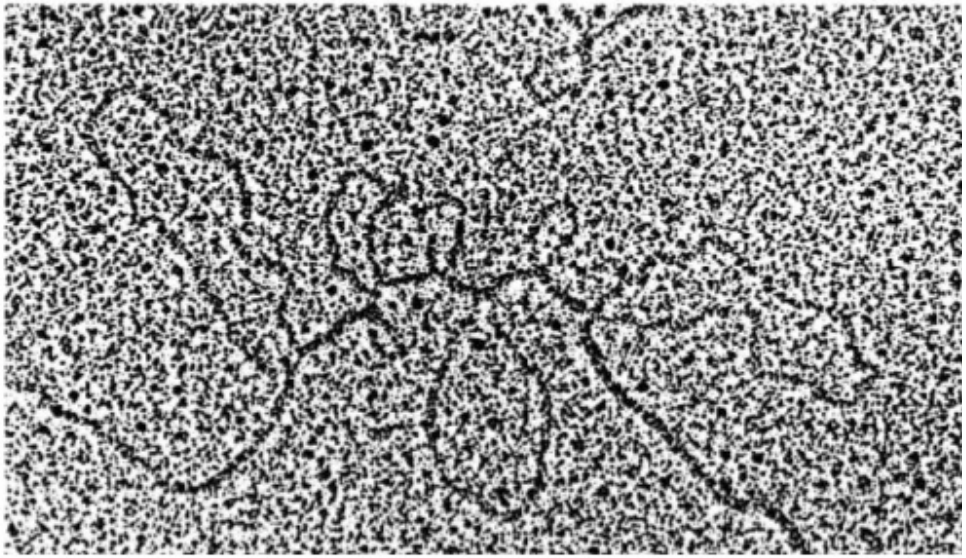


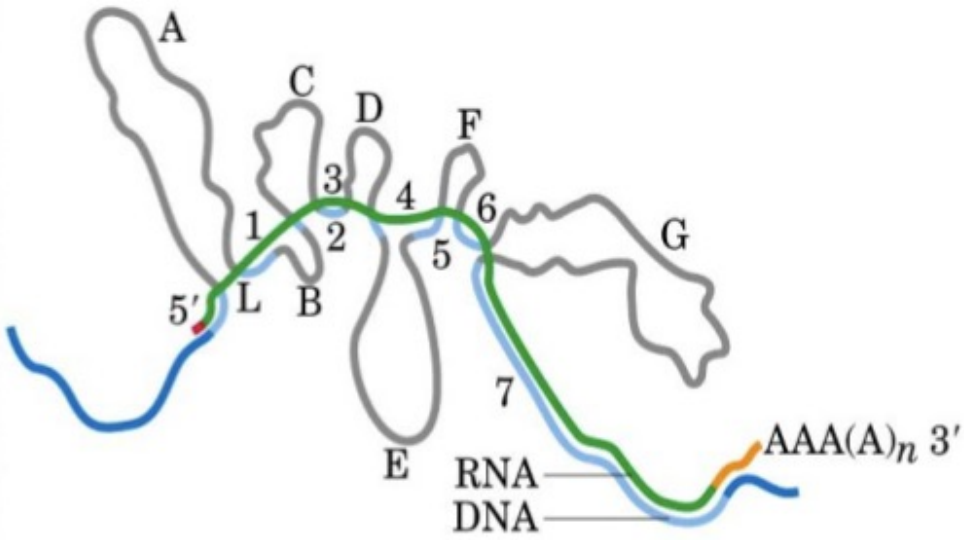
“Splicing”

Eukaryotic genes contain introns



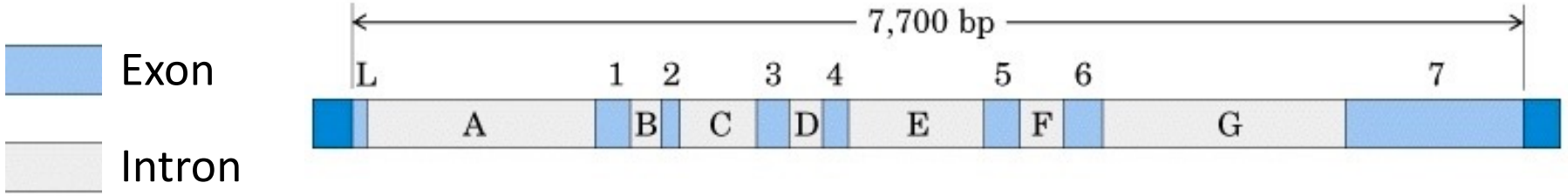
Identification of introns by R-looping



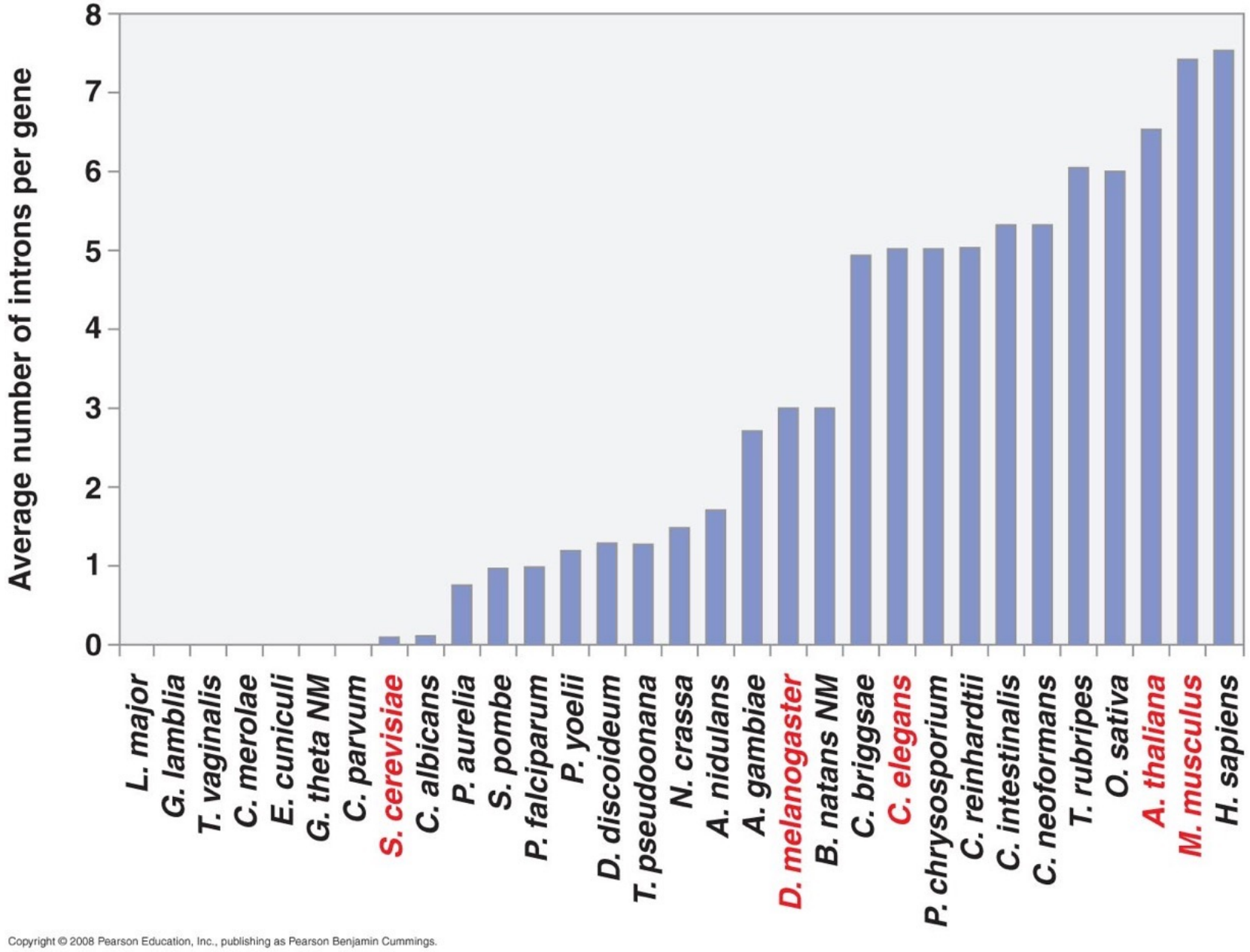


Chicken ovalbumin gene hybridized with its mRNA

Exon:intron organization of the ovalbumin gene

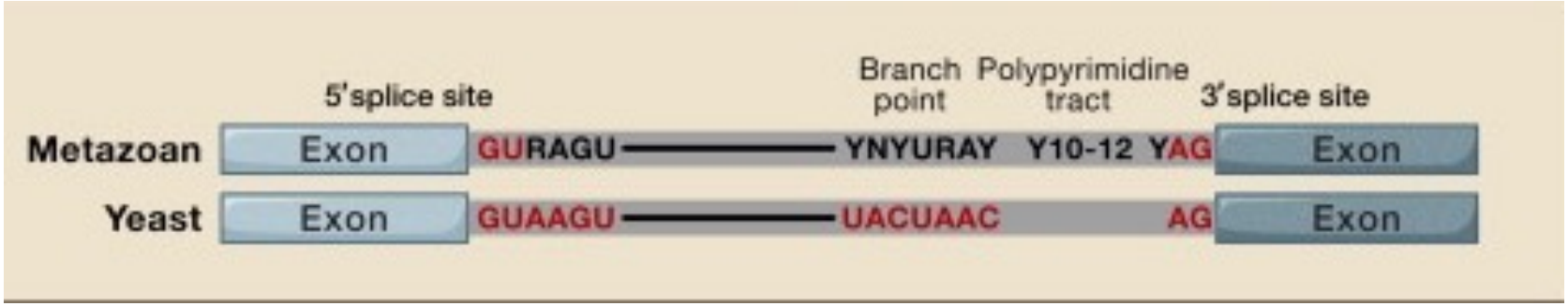


Number of introns per gene in various eukaryotic species

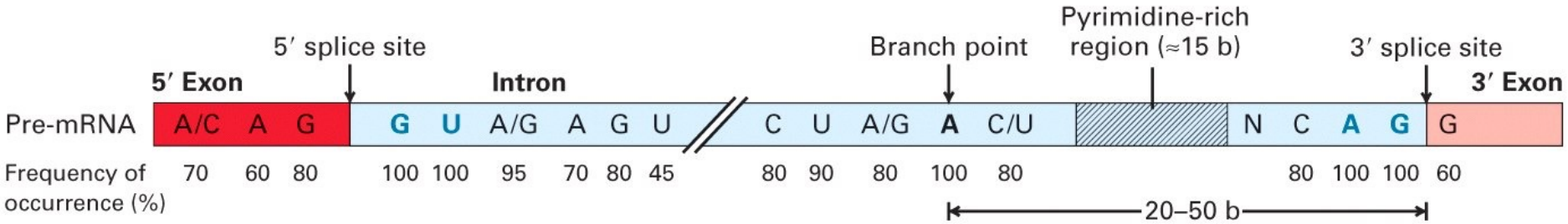


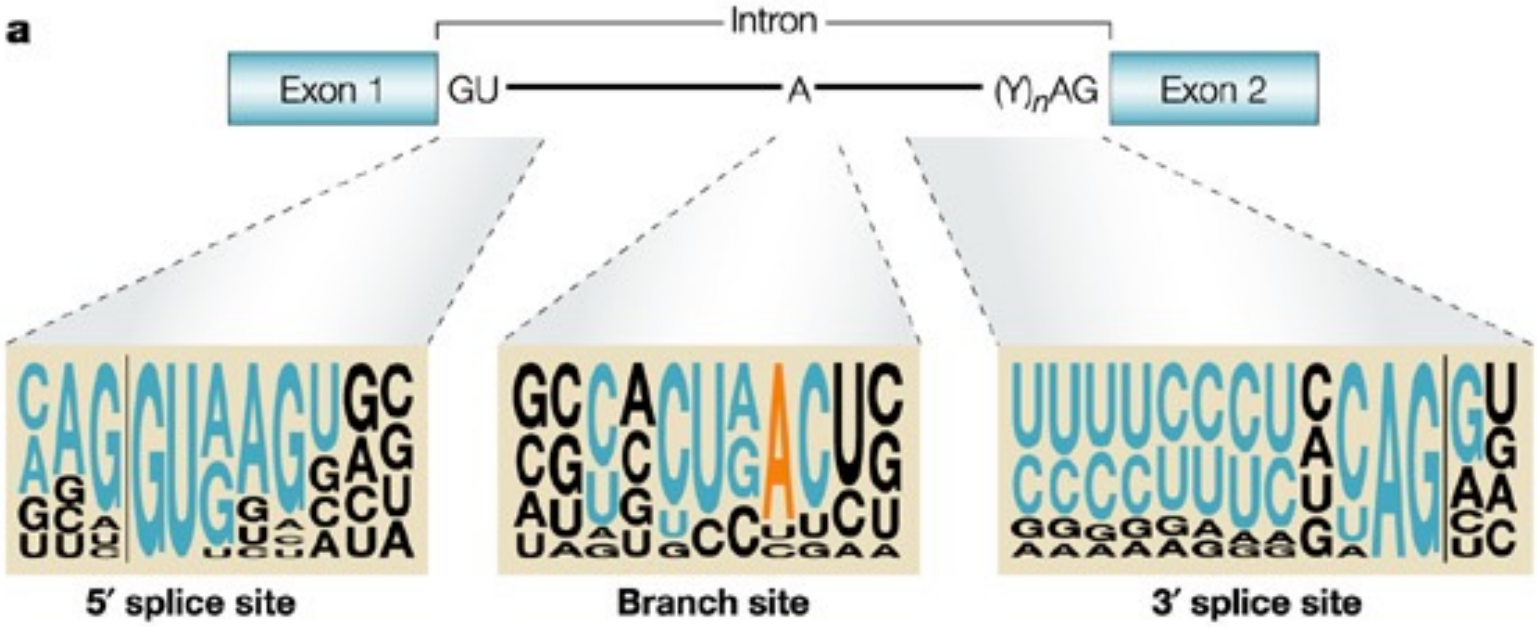
Specific sequences inside the pre-mRNA indicate where the splicing event has to take place.

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

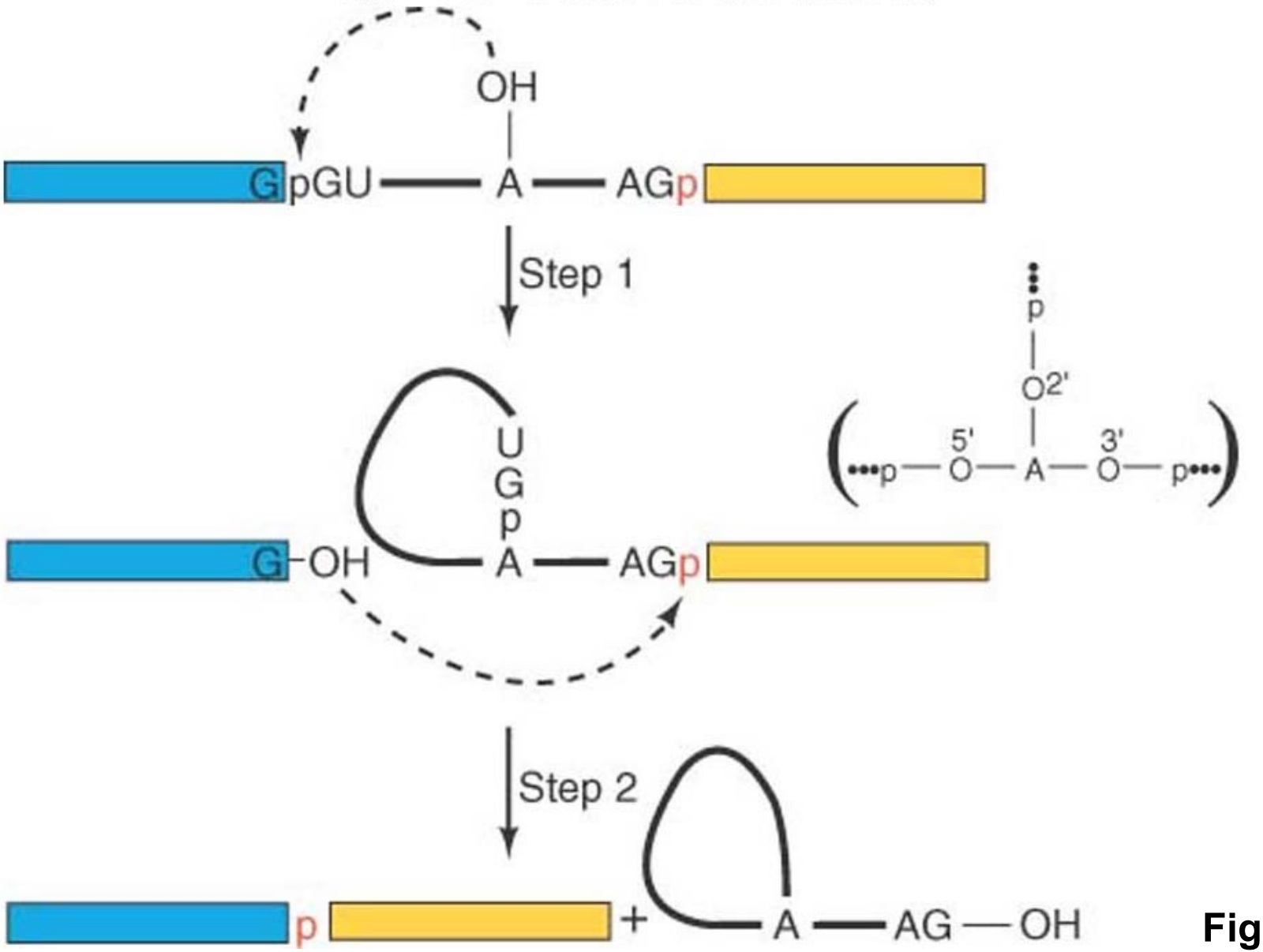


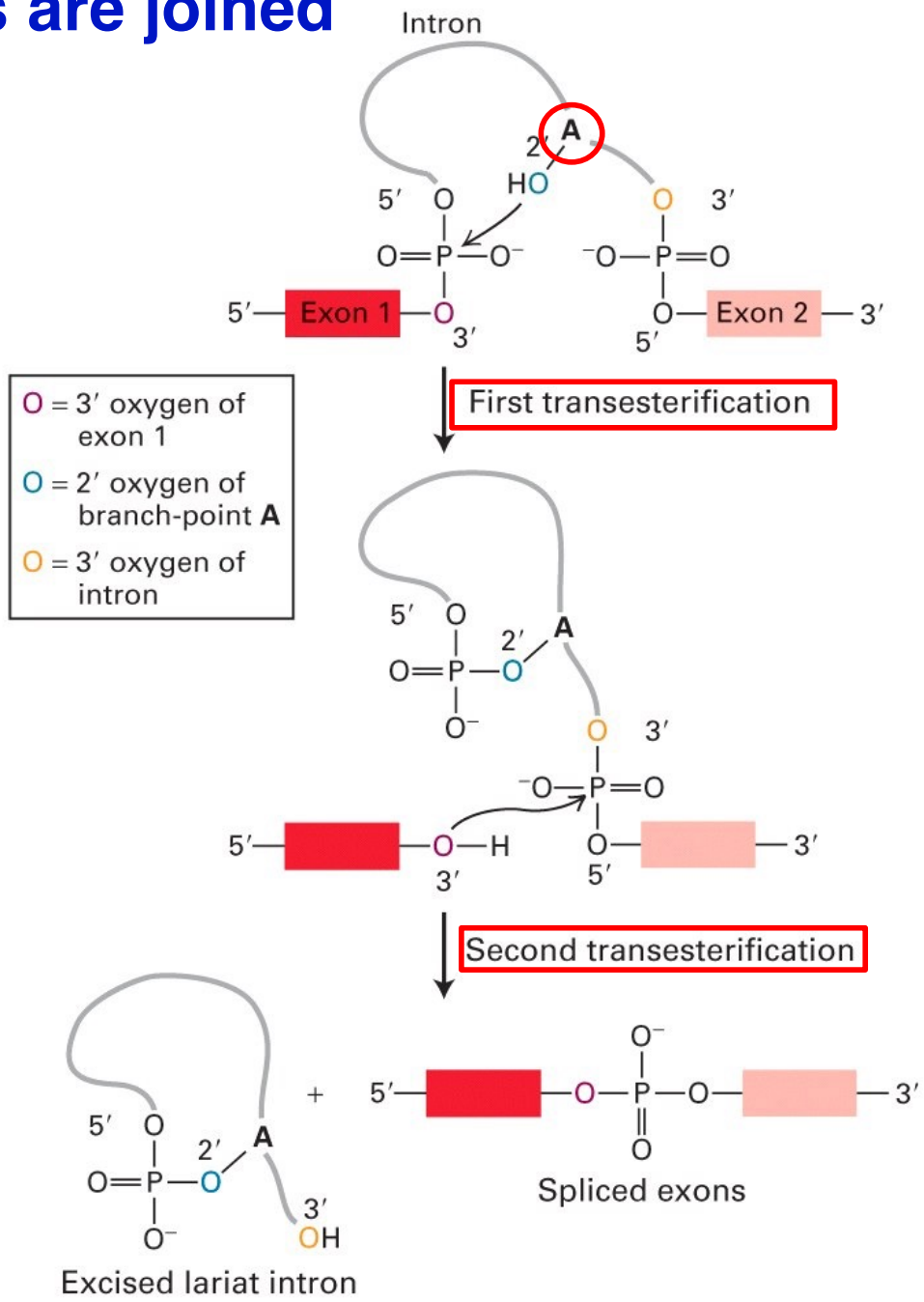
Fig. 14.4

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

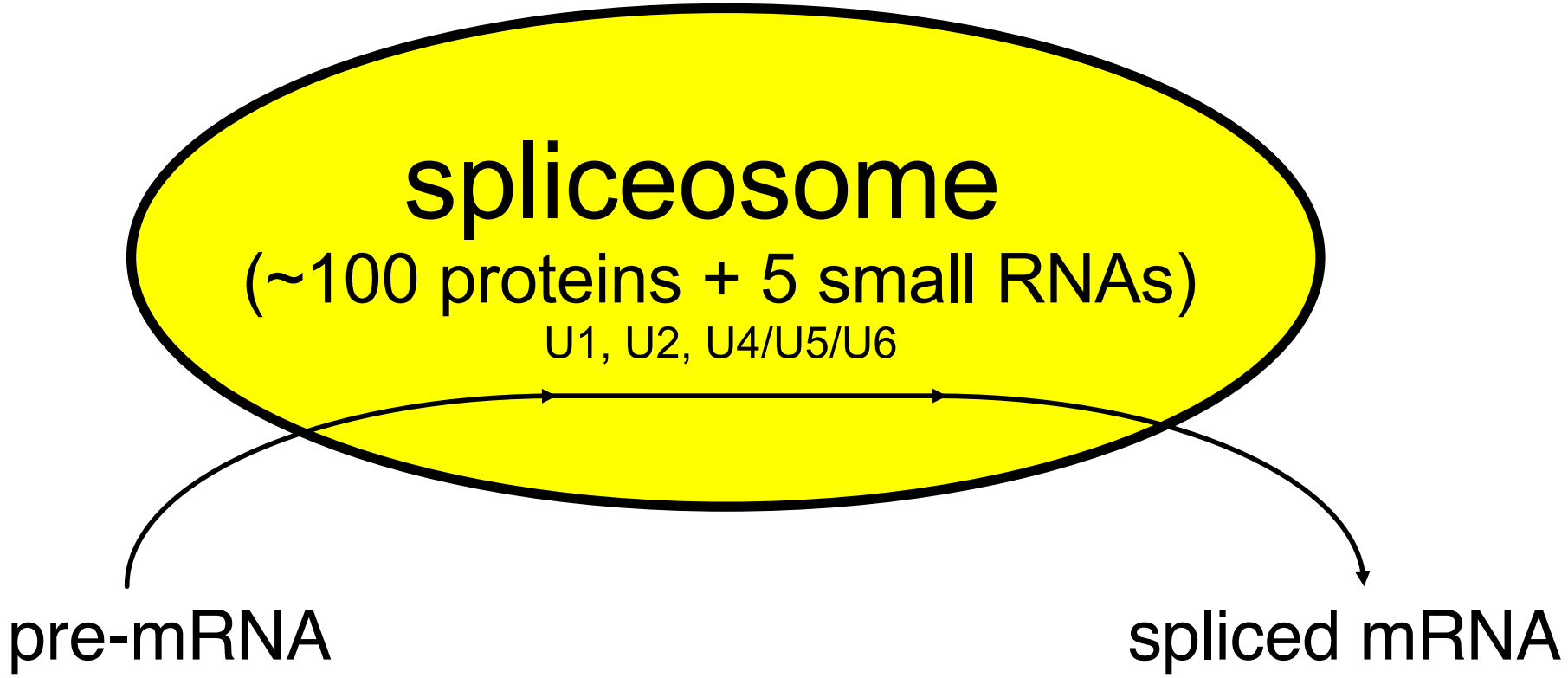
Two **trans-esterifications:**

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

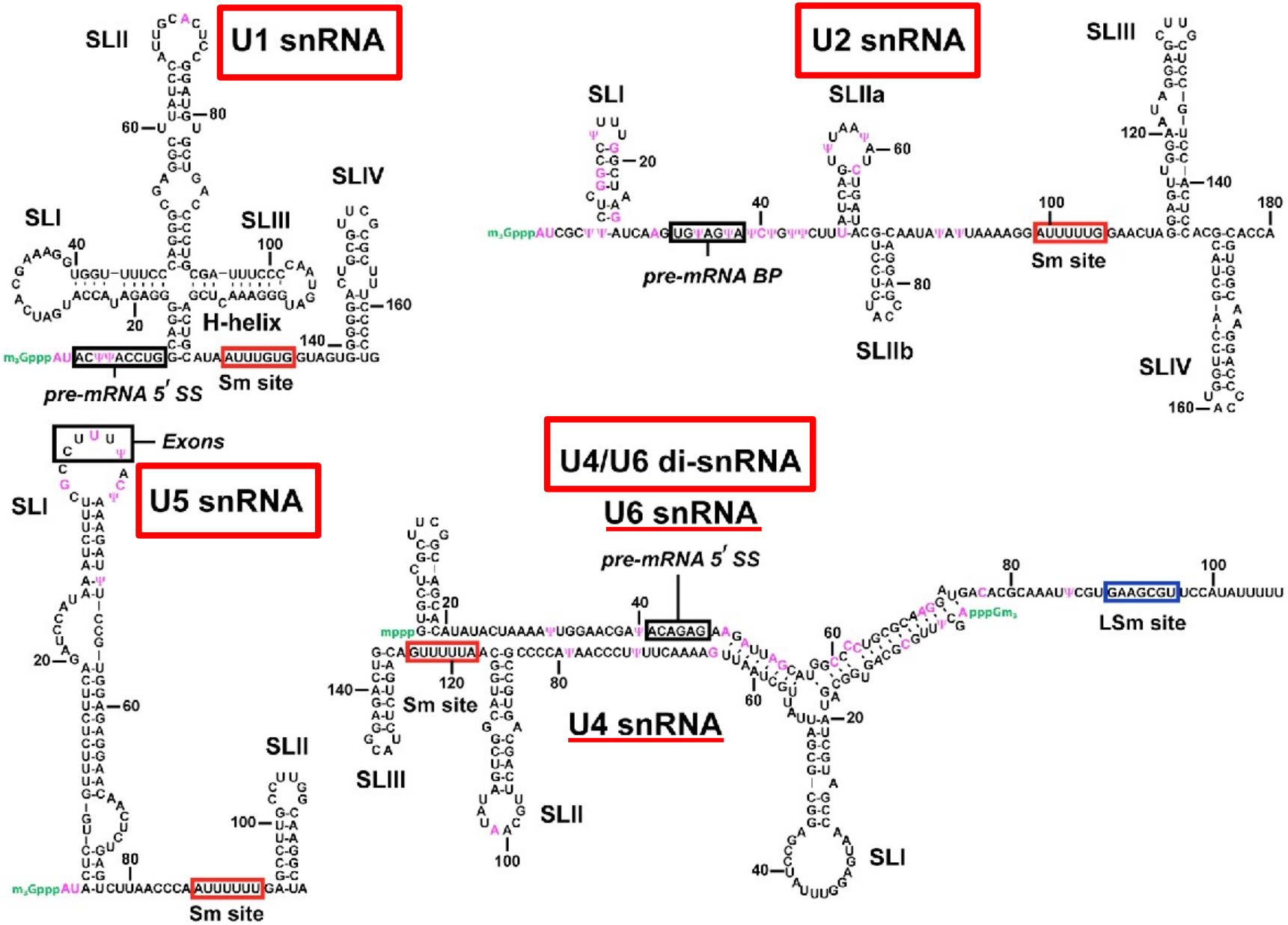


Splicing takes place on a complex called spliceosome



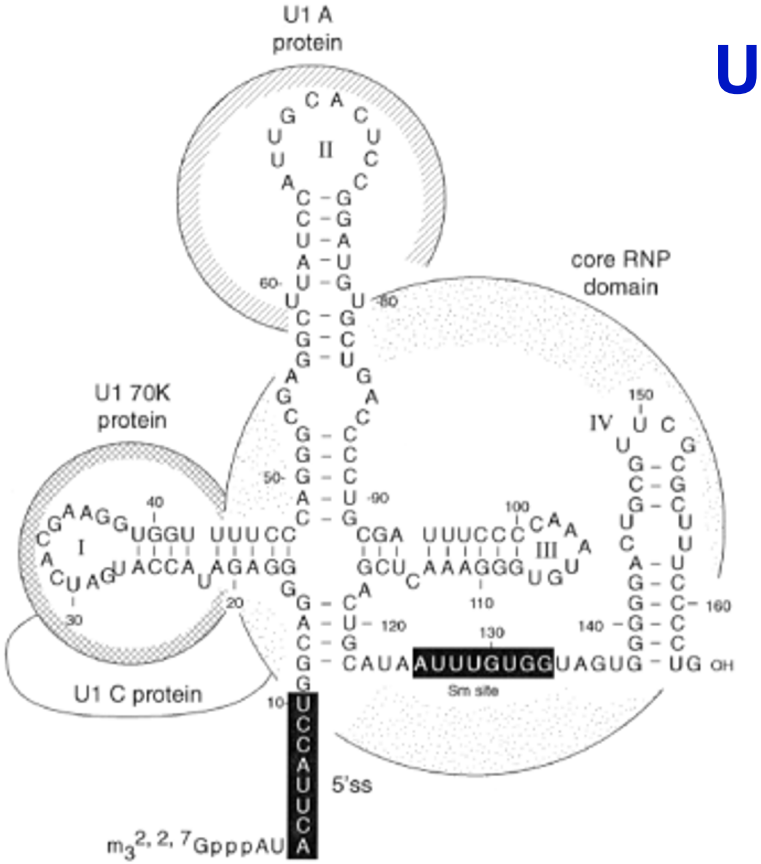
Splicing works similarly in different organisms, for example in yeast, flies, worms, plants and animals.

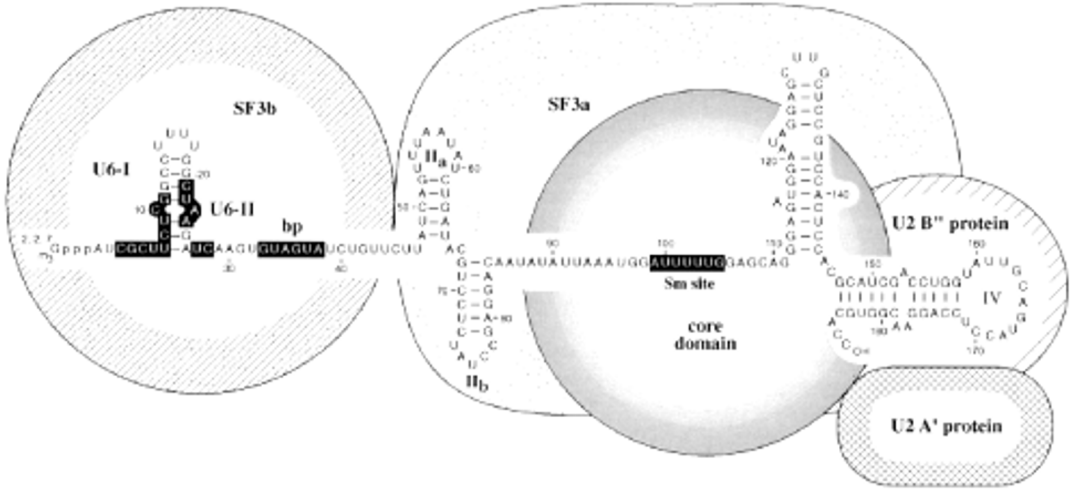
Five snRNAs are involved in pre-mRNA splicing



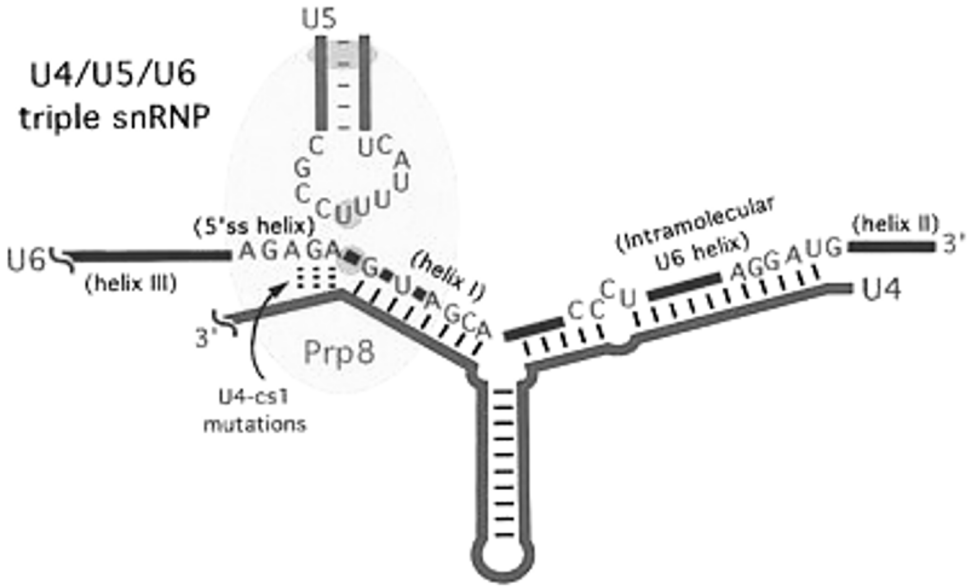
The black boxes indicate the nucleotides participating in RNA-RNA interactions or involved in catalysis during pre-mRNA splicing. The red boxes highlight the Sm binding sites.

U1 and U2 snRNPs



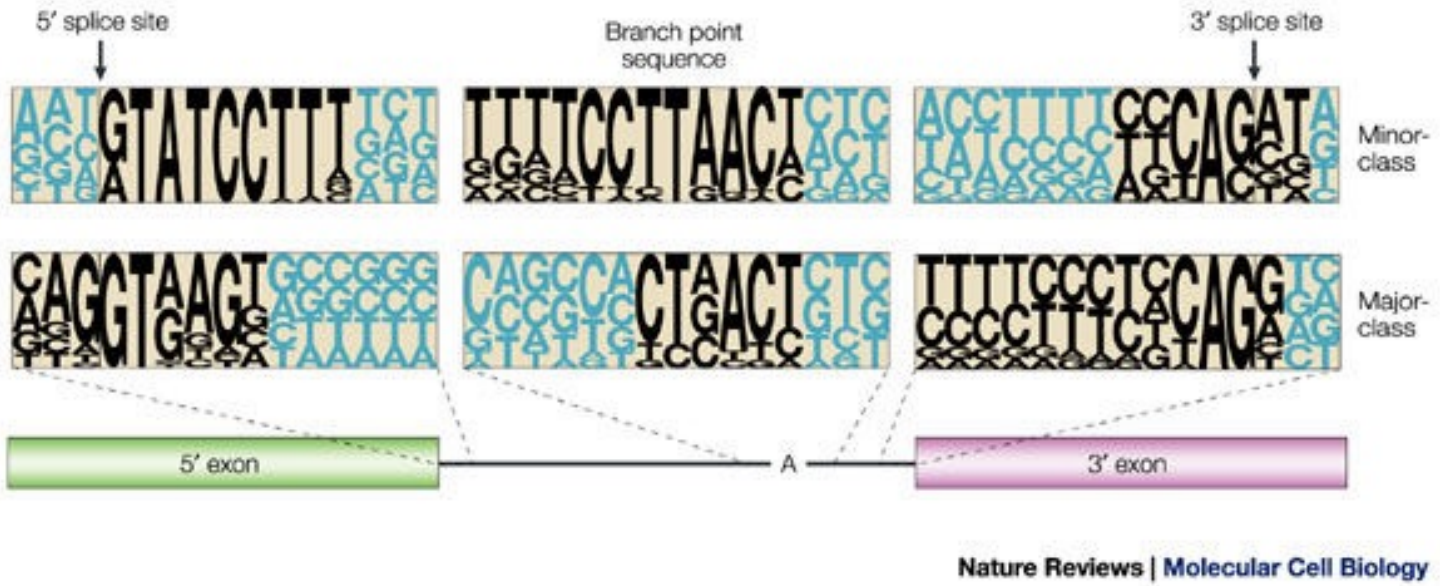


U4/U5/U6 tri-snRNPs



Minor-class introns

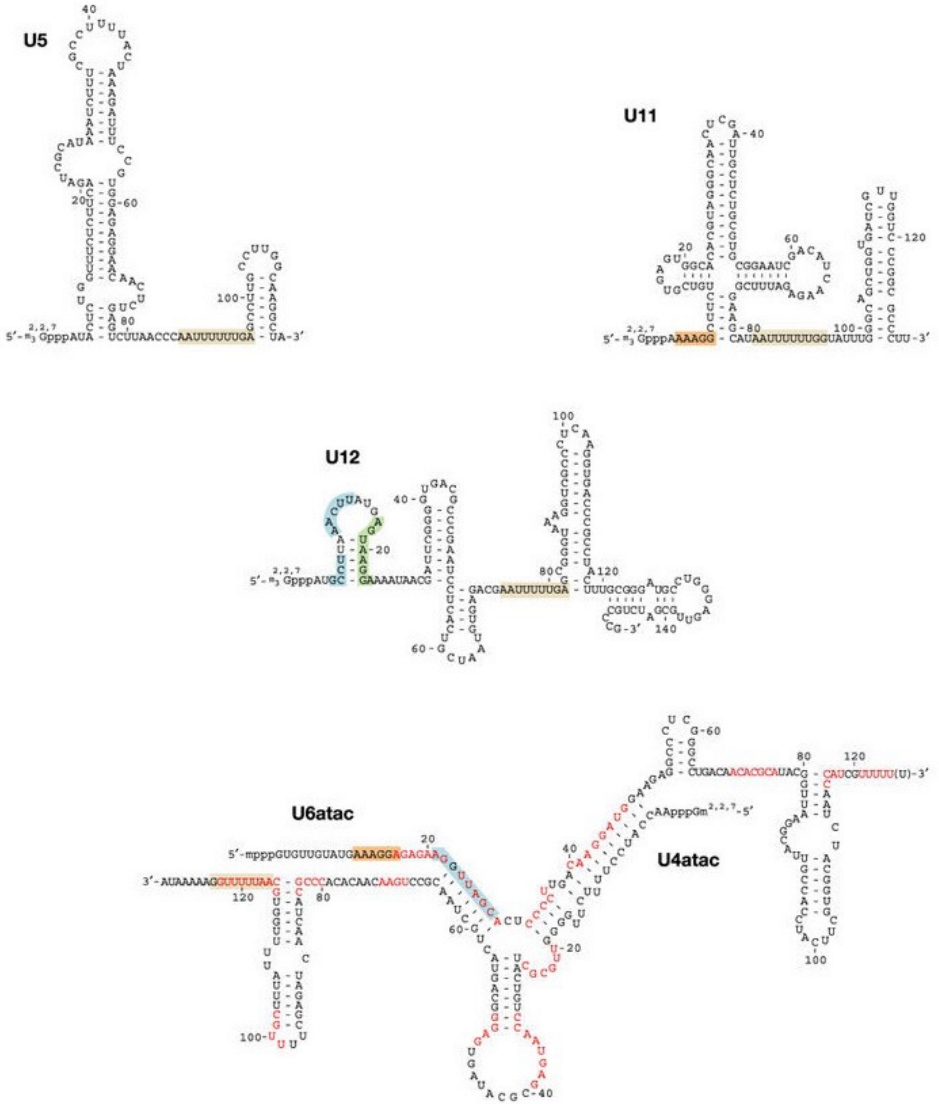
Some rare introns shared unusual consensus sequences and their excision is mediated by a distinct spliceosome, **U12-dependent spliceosome**, that involved low-abundance snRNPs (**U11, U12, U4atac–U6atac) and U5**. U12-dependent splicing is determined by the longer and more tightly constrained consensus 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.

U12-dependent spliceosome

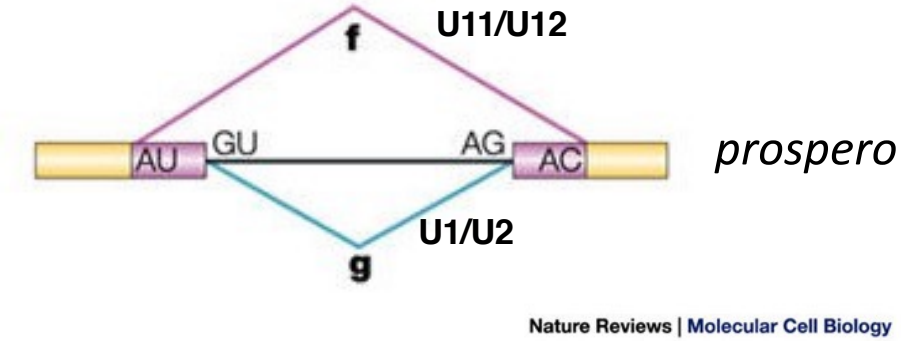
U11 and U12 have base-pairing potential with the 5’-splice-site and branch-site sequences, whereas their secondary structures mimic those of U1 and U2, respectively. While U6atac/U4atac are the analogues of U6 and U4. The U5 snRNP is a component of both spliceosomes.



Nature Reviews | Molecular Cell Biology

Major- vs Minor- spliceosome

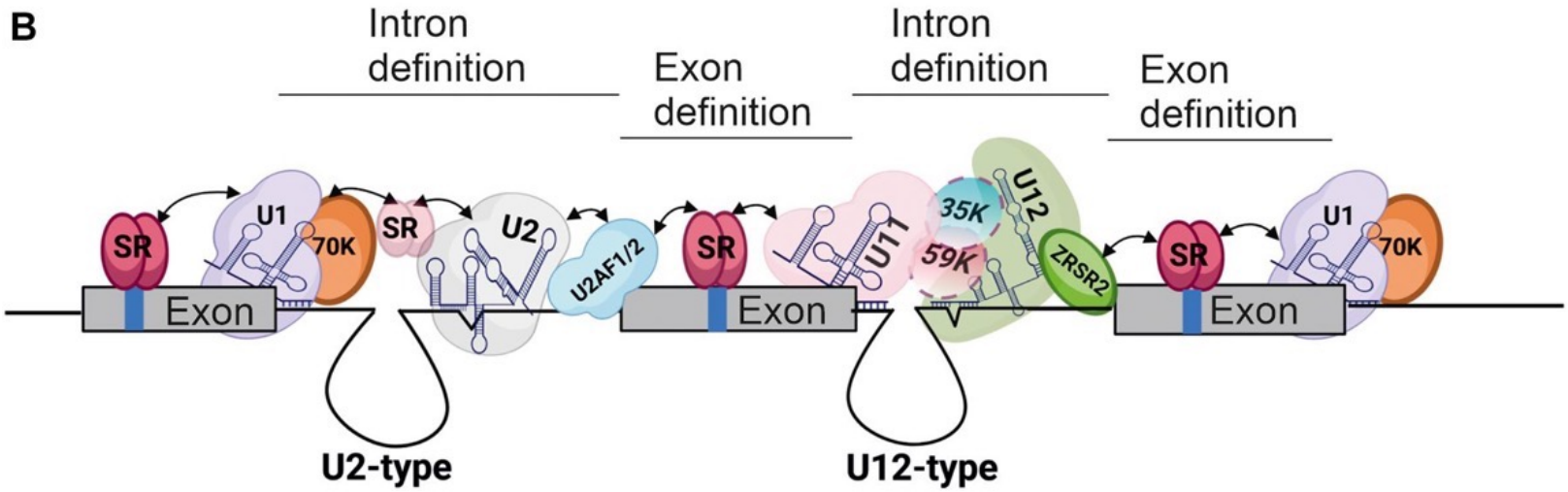
Major-class and minor-class splice sites are incompatible with each other, and as yet no hybrid introns have been identified. Therefore, unique patterns of alternative splice-site choice can arise. One such example is the 'intron-within-an-intron' architecture of the *prospero* gene of *Drosophila melanogaster*, 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.



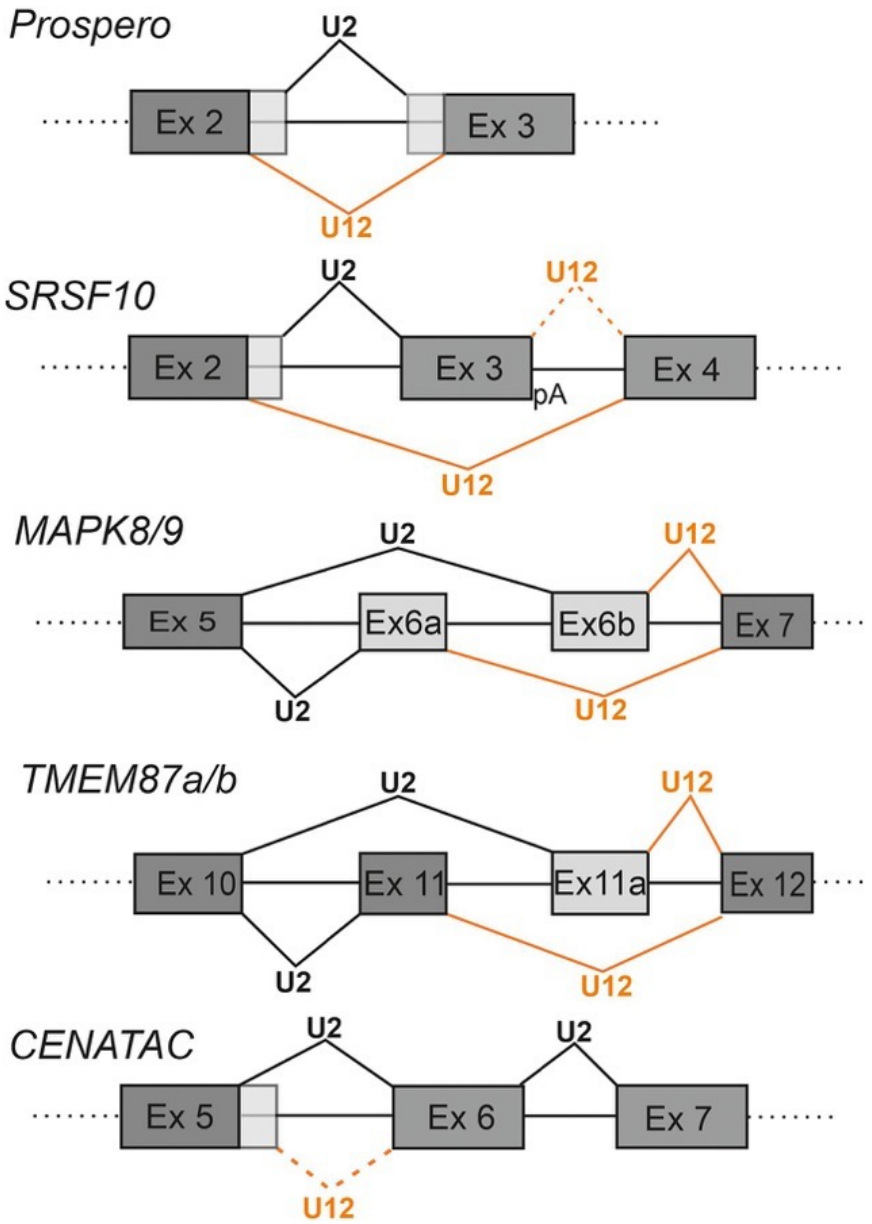
The choice of U12-type versus U2-type splicing in this transcript is modulated during embryogenesis, and regulated by the binding of protein factors to an intronic element.

Major- vs Minor- spliceosome

Both exon and intron definition mechanisms rely on protein-protein interactions to connect spliceosomal complexes assembled on the 5’ss or PPT/3’ss, to enable cross-exon and cross-intron communication. Proteins containing arginine and serine rich domains (RS domains) are the main facilitators of these interactions. These include the SR protein family of splicing regulators and several integral spliceosome components present in both the major and minor spliceosomes



Competition Between Minor and Major Spliceosomes



Competition between minor and major spliceosomes. Examples of known genes utilizing adjacent U12-and U2-type introns in regulating gene expression

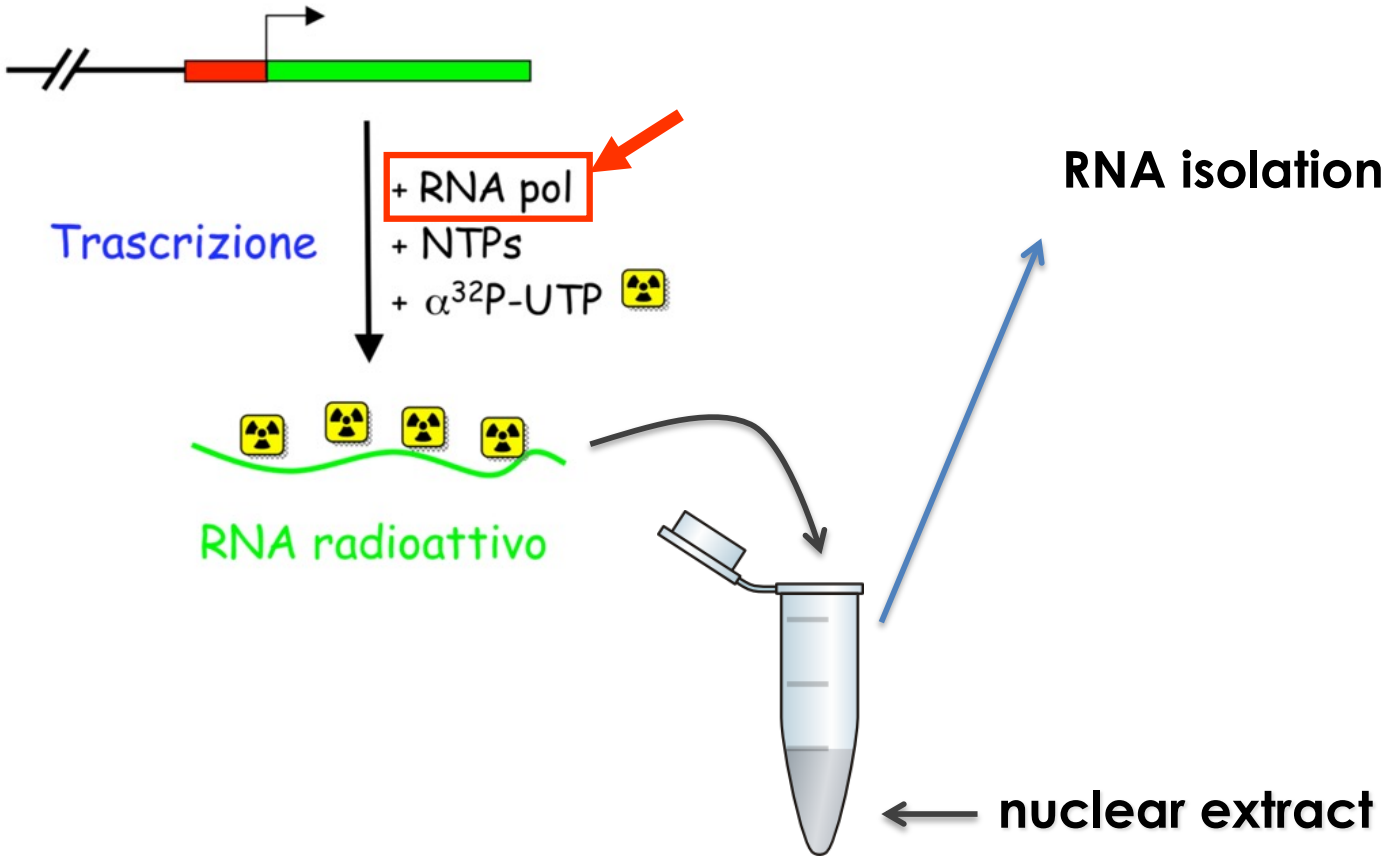
Assembly, rearrangement, and catalysis within the *spliceosome*: the splicing pathway

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 .

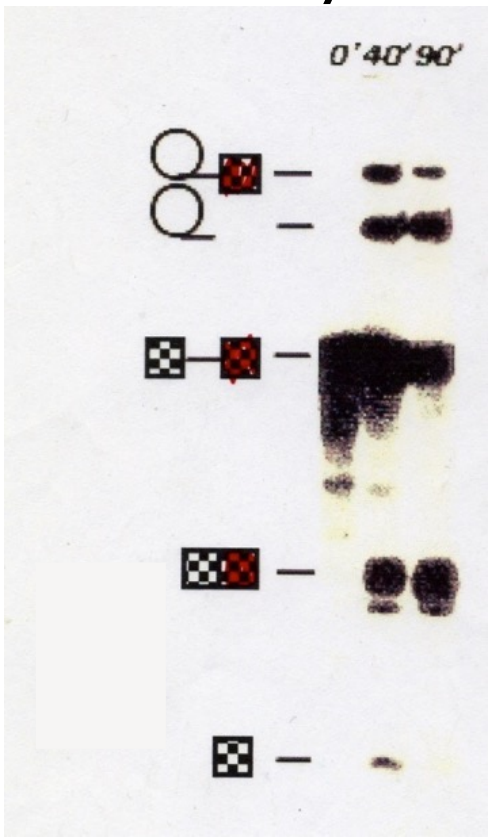
Spliceosome components and assembly are generally conserved in eukaryotes. The ~90 core proteins are conserved between the budding yeast *Saccharomyces cerevisiae* and human spliceosomes, although the number of human spliceosome proteins (~175) is two fold higher compared with yeast .

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

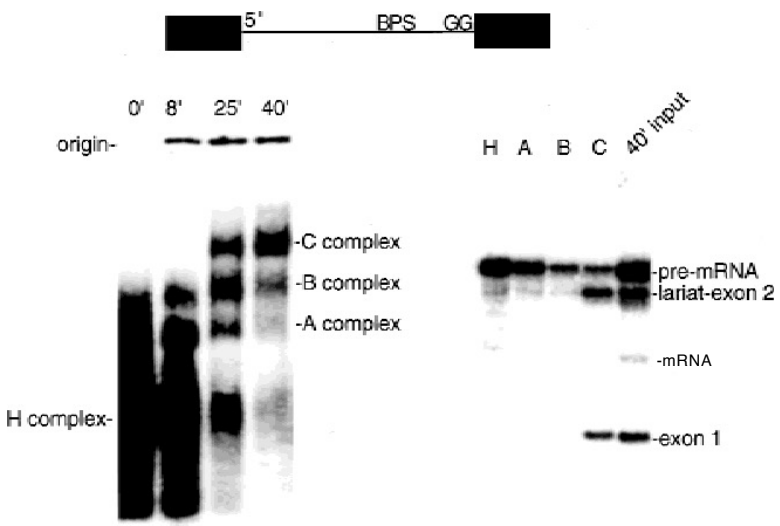


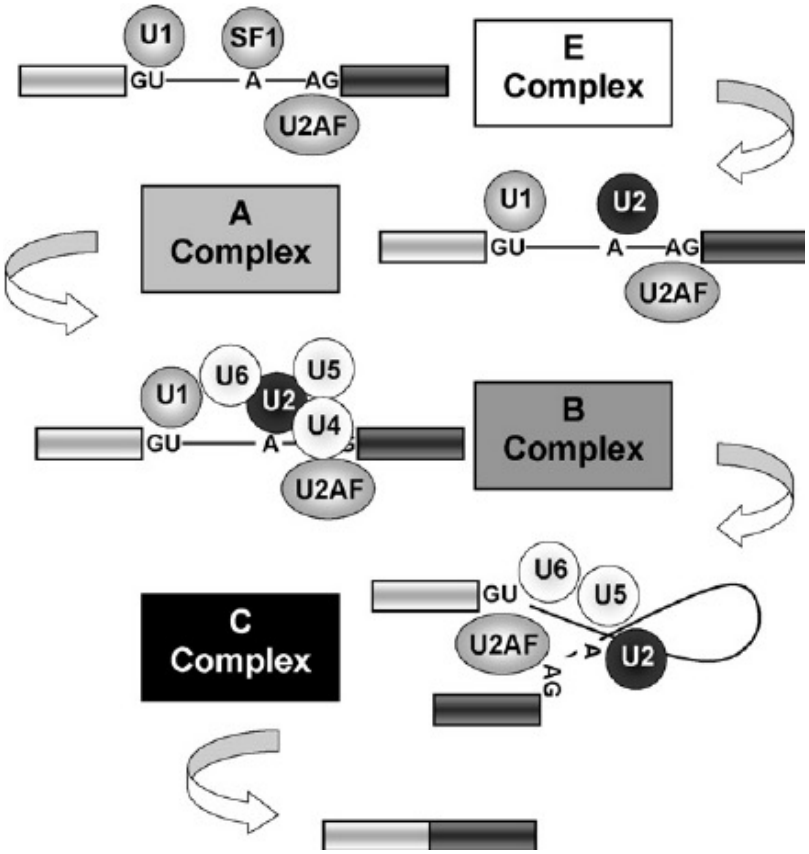
In vitro analysis of splicing complexes formation

In vitro transcribed RNA + HeLa Nuclear Extract (HNE)

EMSA

RNA analysis

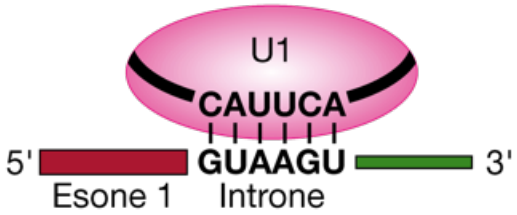




DNA oligonucleotides and RNase H can be used to deplete specific snRNA from HNE

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

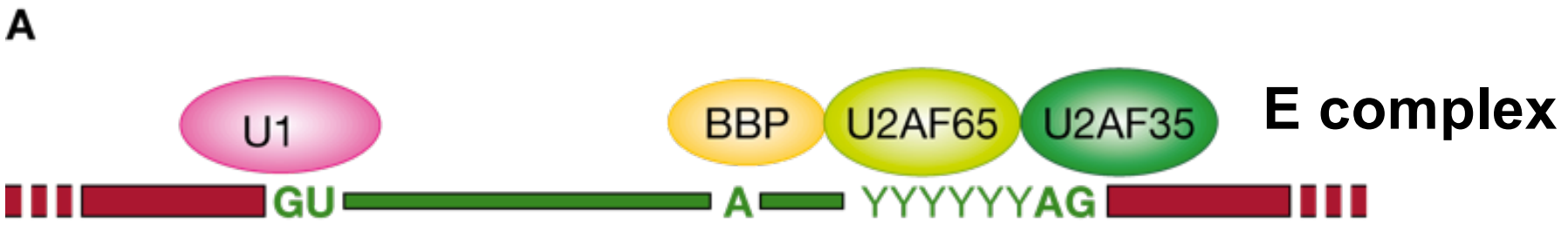
Assembly step 1



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.

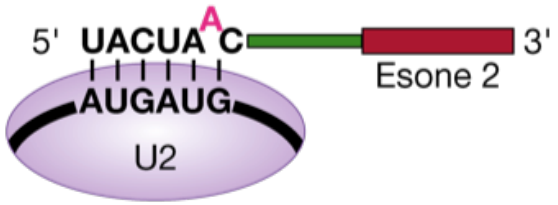
3. **Early (E) complex** is formed



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.

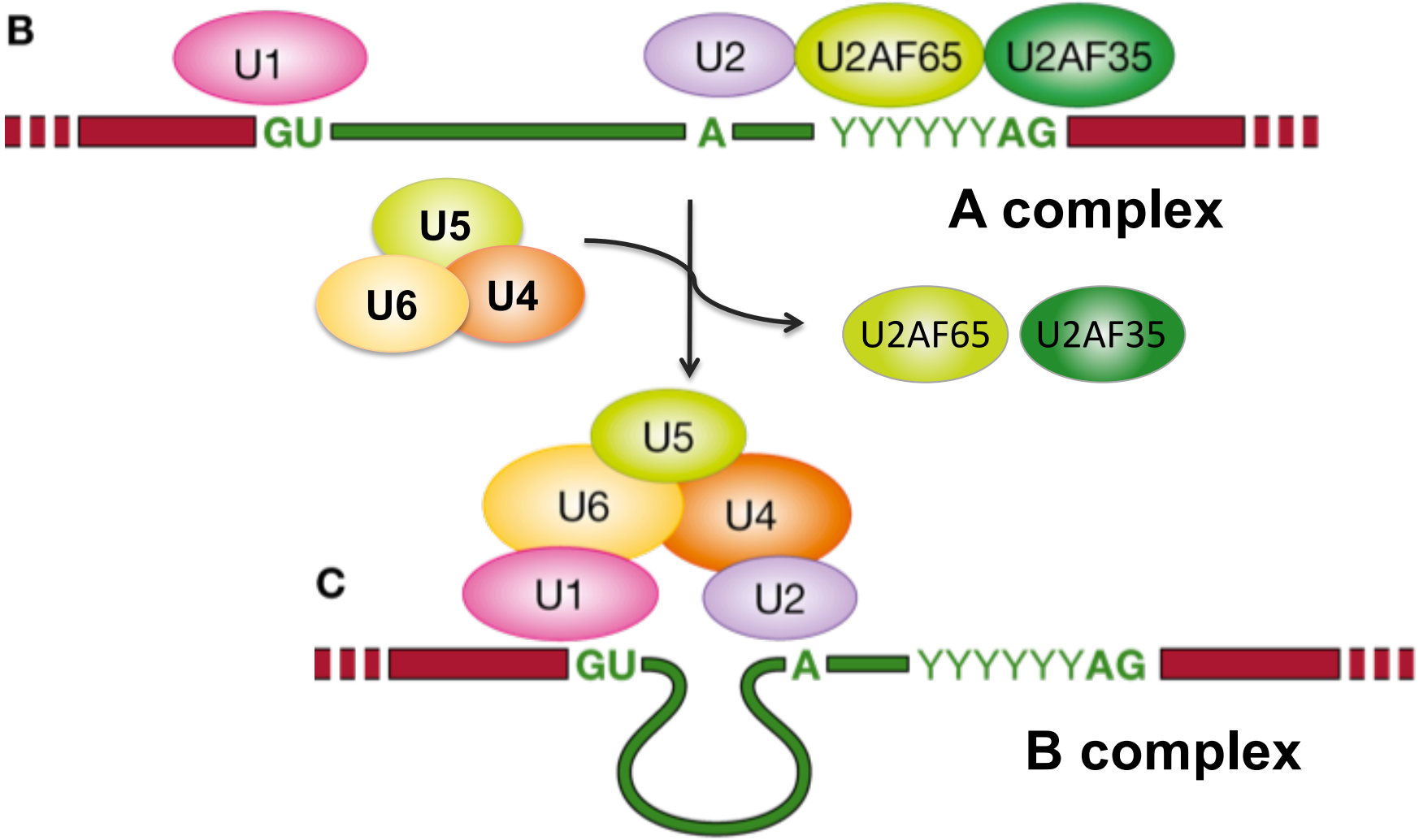


A complex

Assembly step 3

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

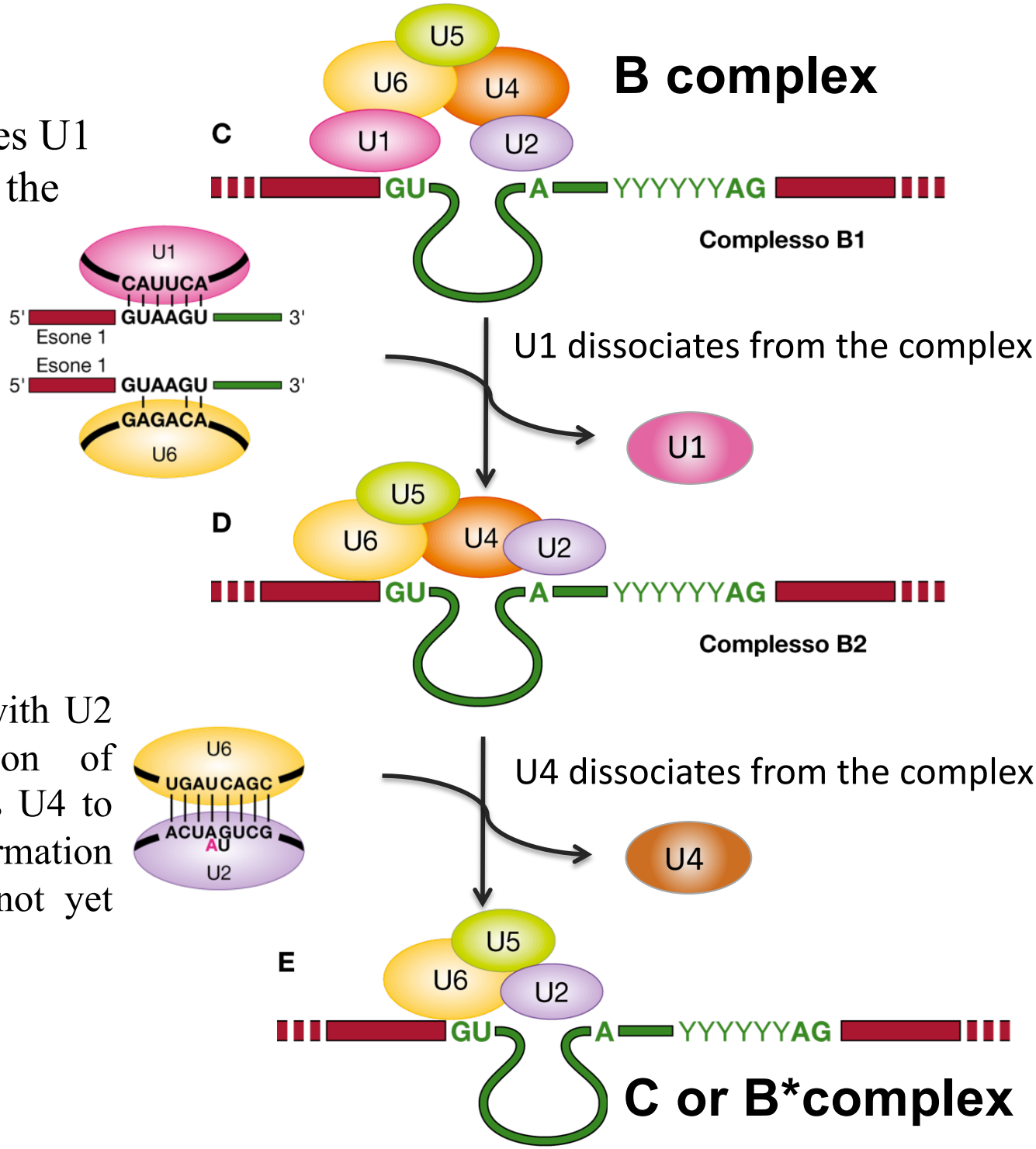
2. With the entry of the tri-snRNP, the A complex is converted into the **B complex**.



Assembly step 4

In the B complex, U6 displaces U1 at the 5' splice site, leading to the formation of the B2 complex.

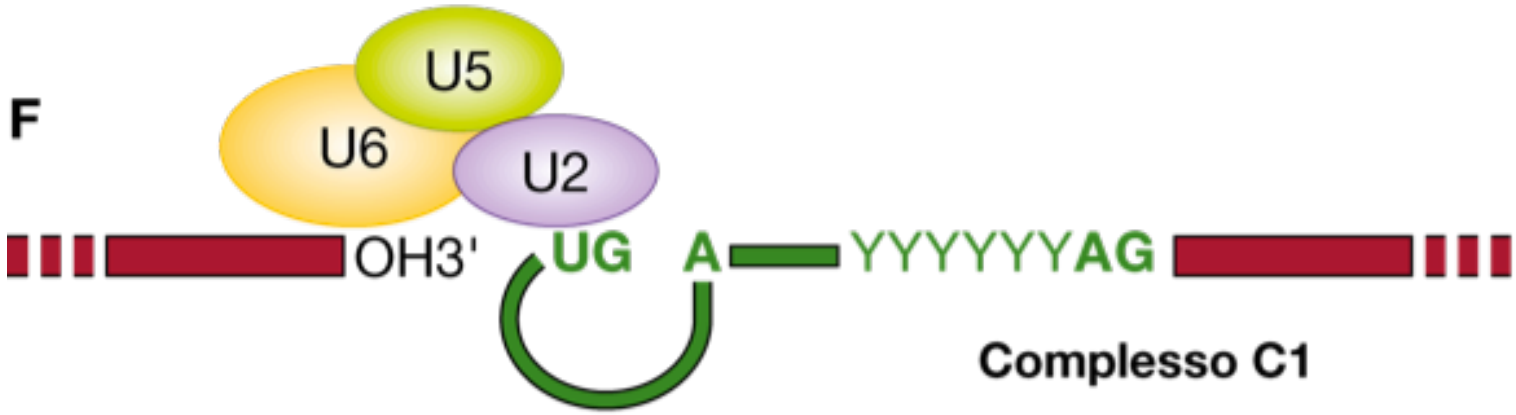
U6 has a region of homology with U2 that overlaps with its region of homology with U4. This allows U4 to be displaced, leading to the formation of the B* complex, which is not yet catalytically active.



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.

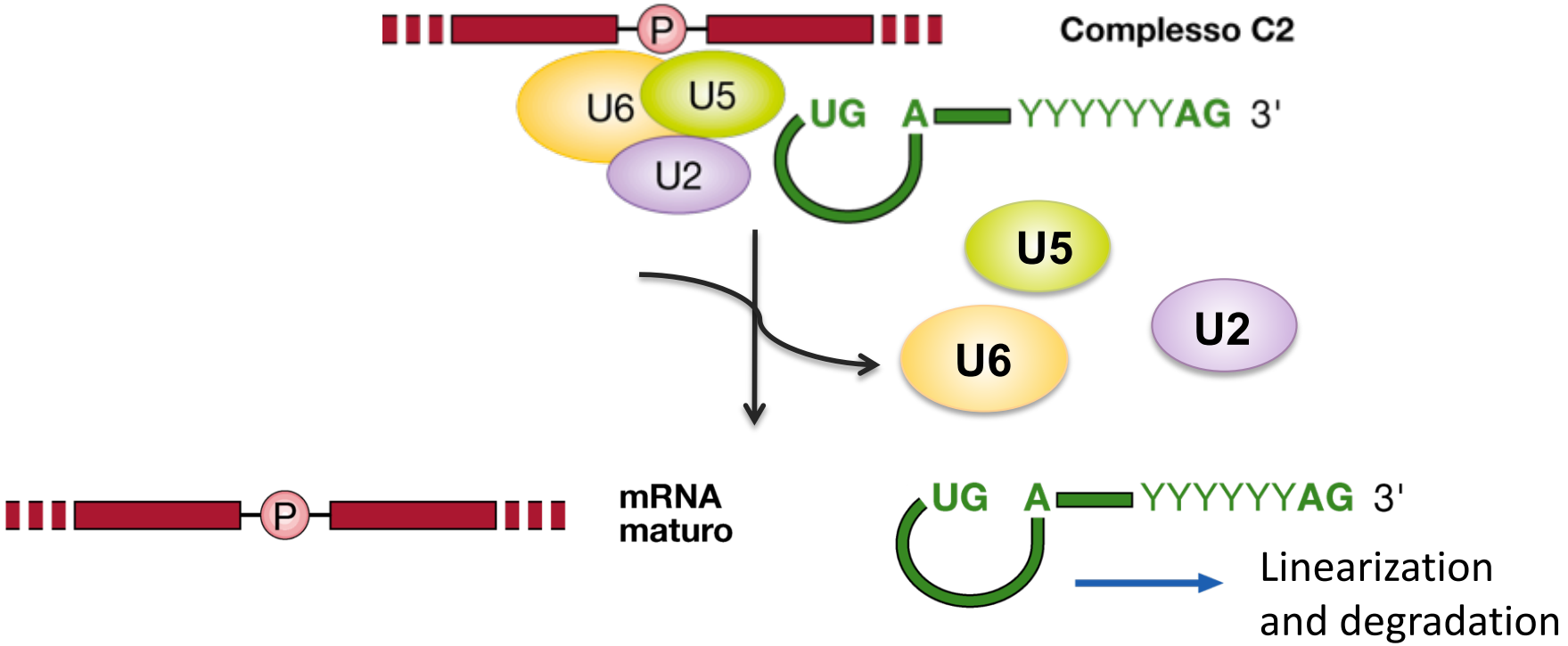


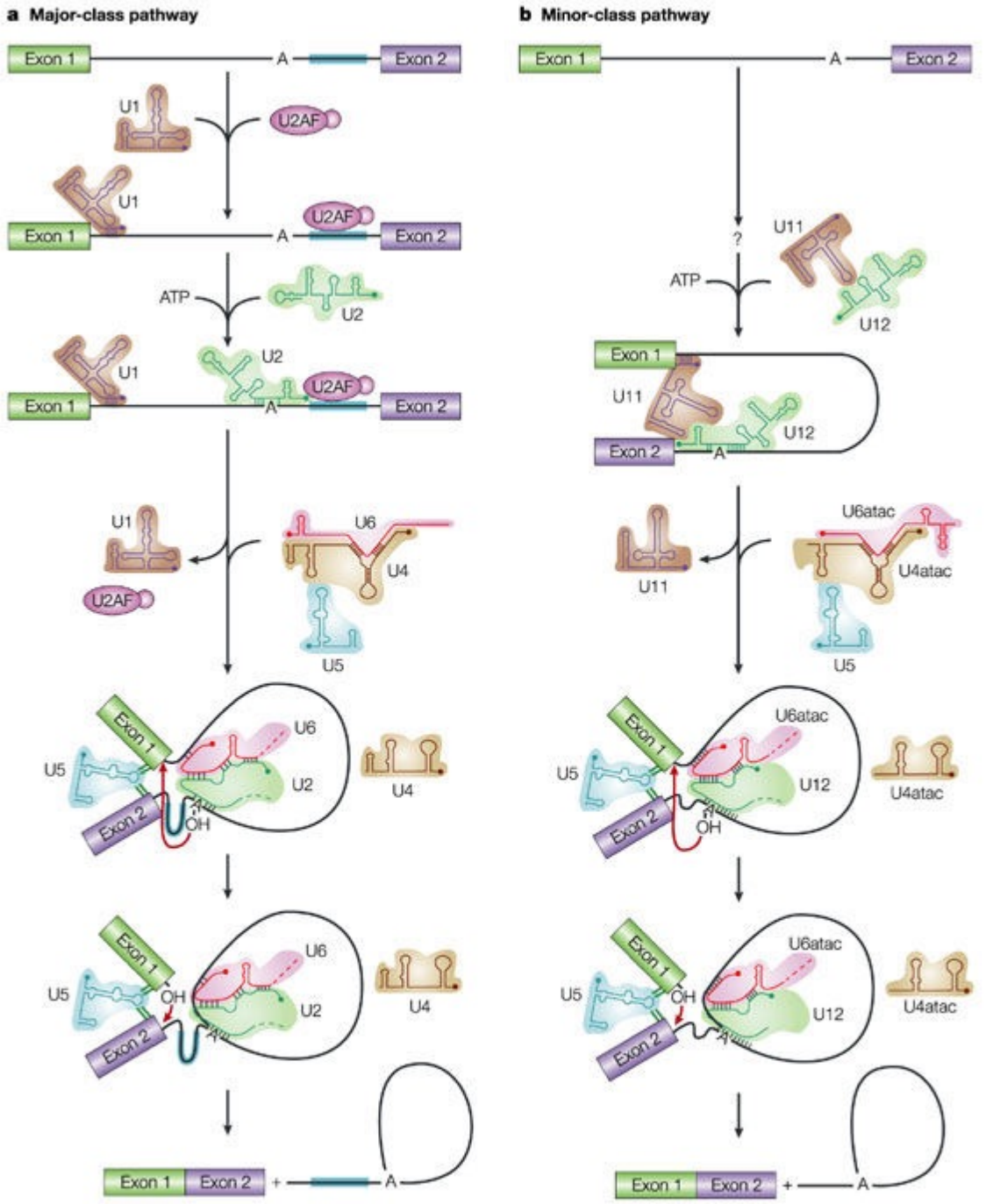
Catalysis Step 2

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.



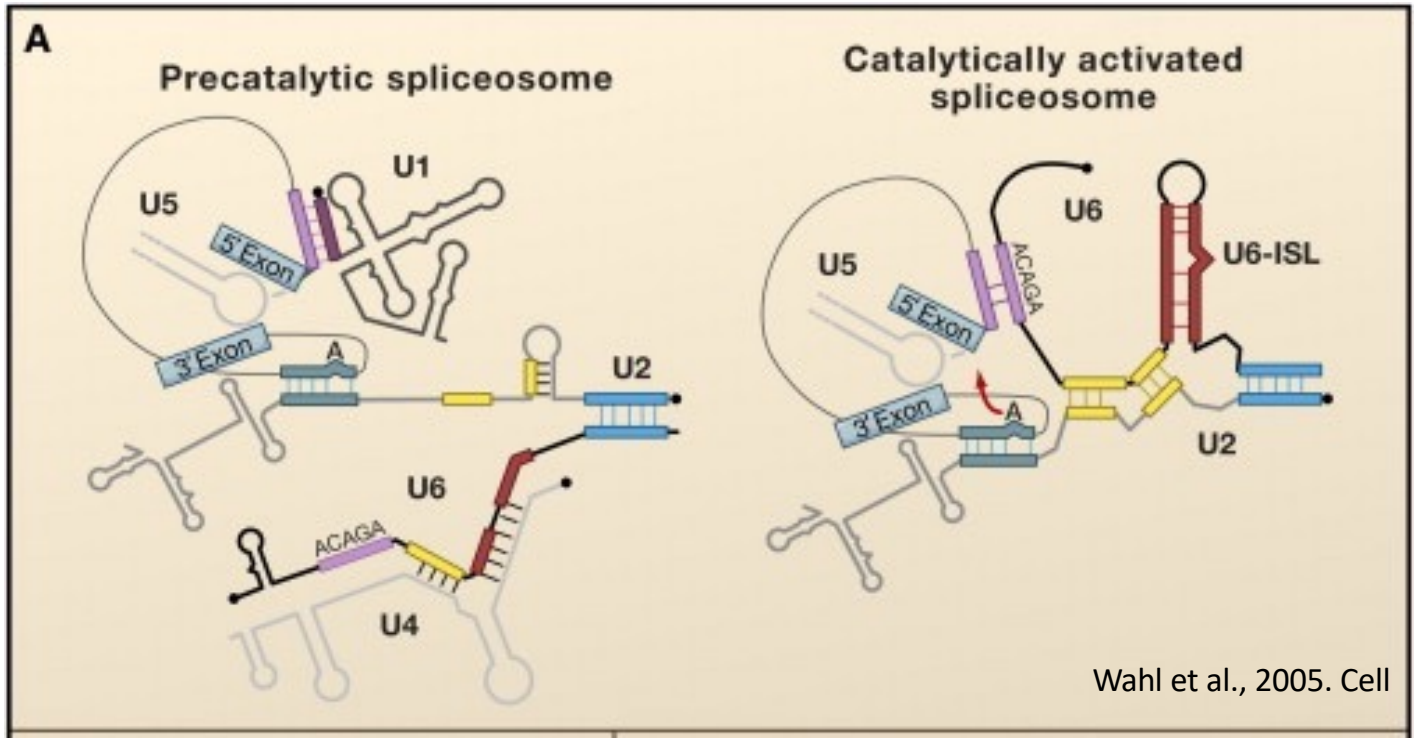


The minor-class splicing reaction *in vitro* proceeds through the same two-step pathway as the major-class reaction.

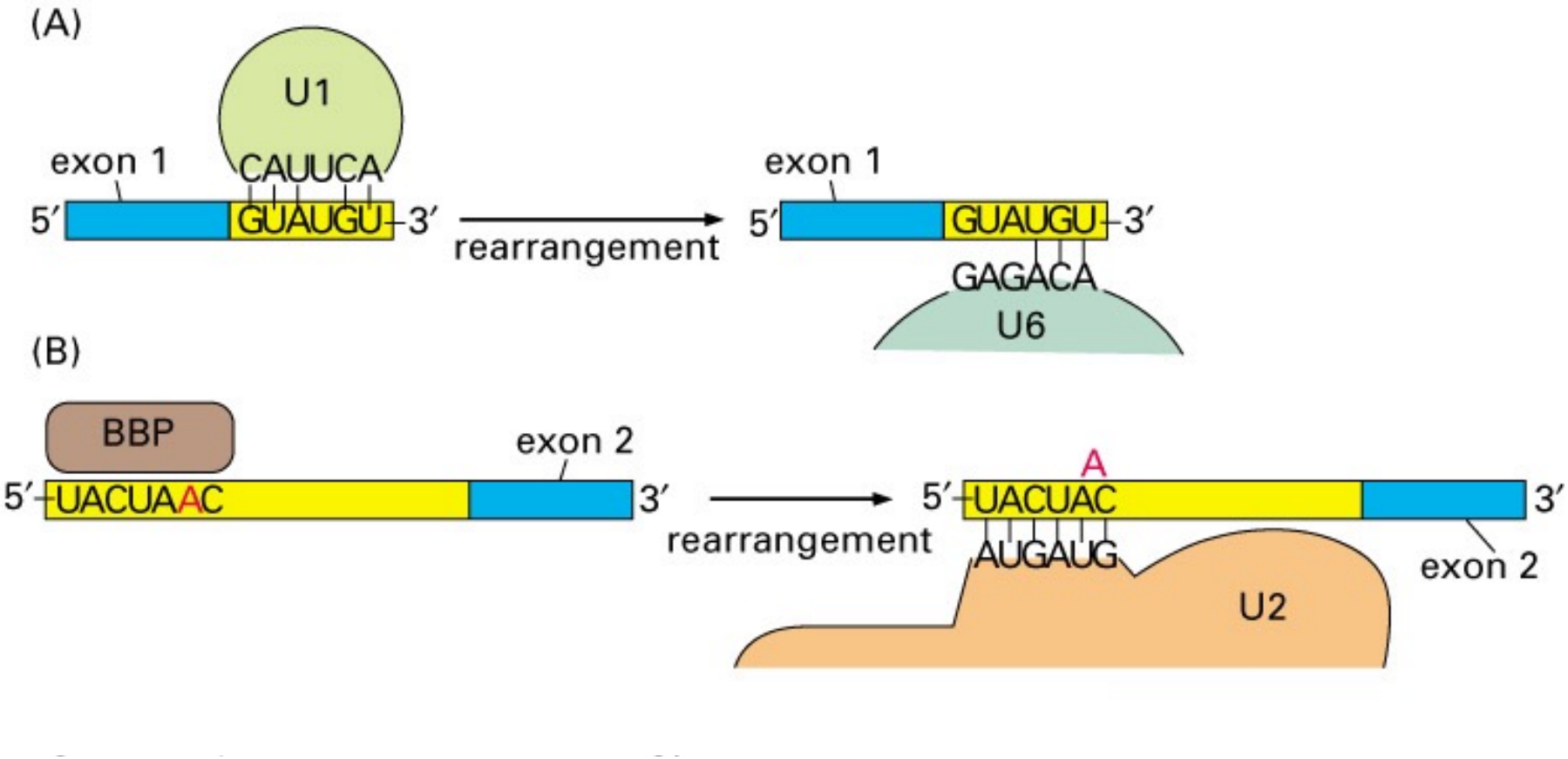
Spliceosomal RNA Network

- 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.

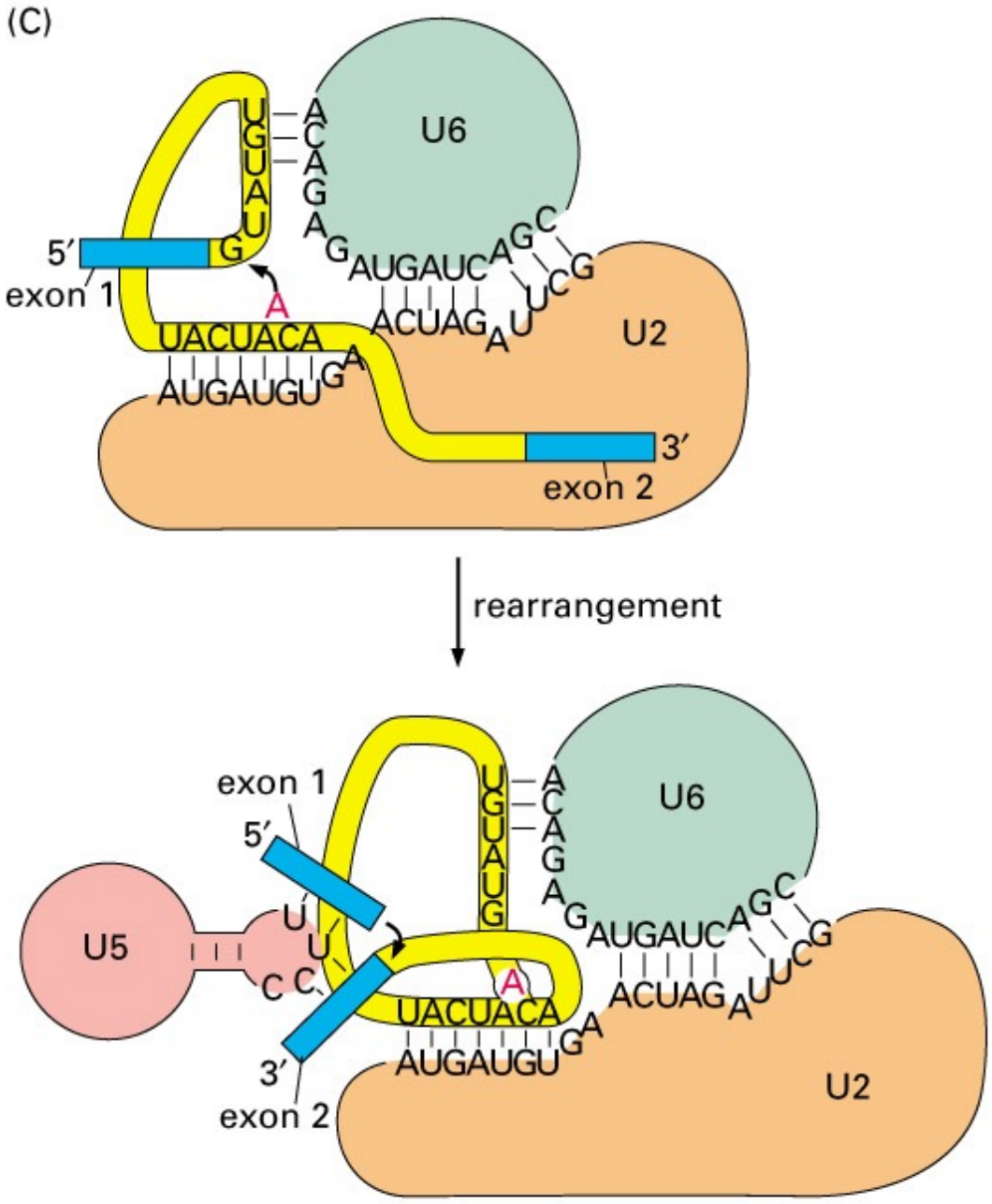
- U4 and U1 are destabilized or dissociate from the spliceosome at the time of activation and no longer are part of the spliceosome's RNA interaction network.



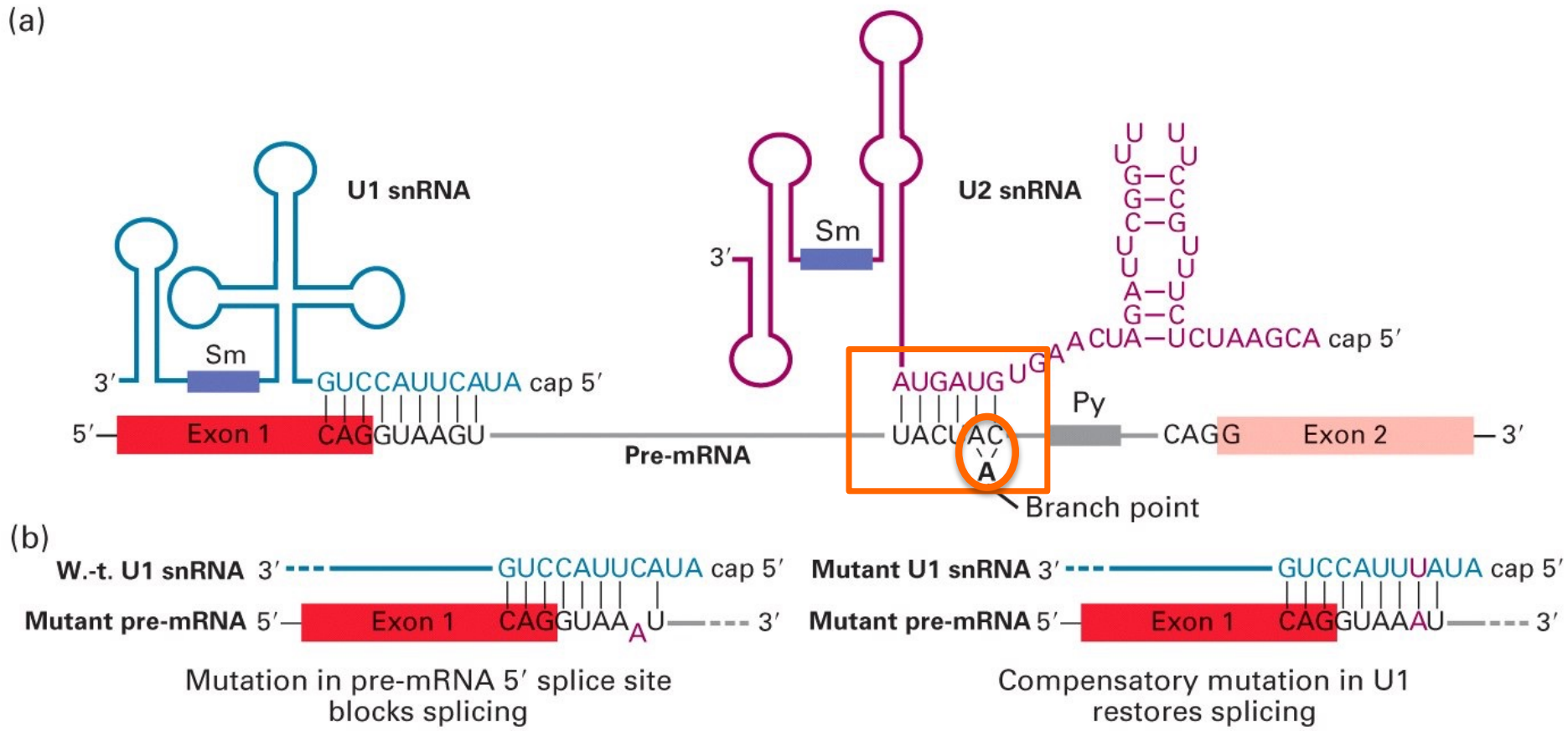
RNA-RNA Rearrangements - I



RNA-RNA Rearrangements - II

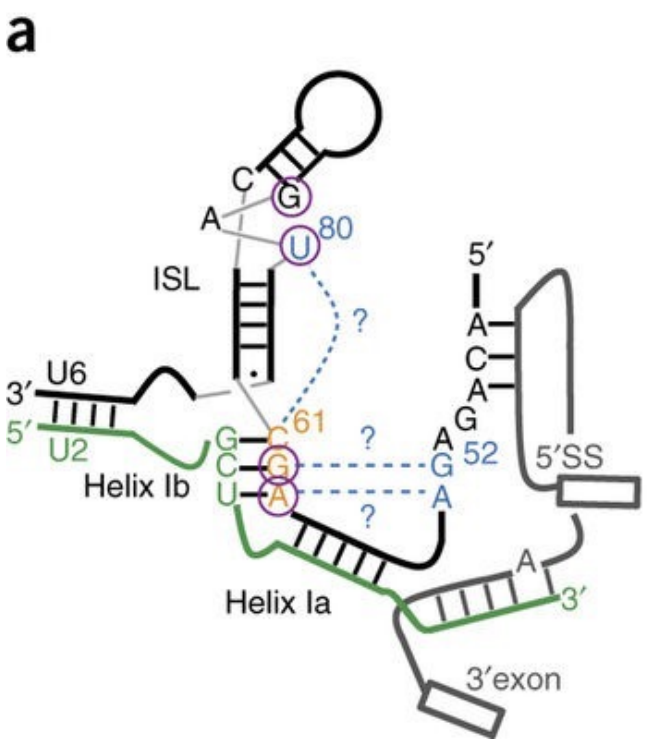


Pre-mRNA Splicing is Accelerated by RNA-RNA Basepairing

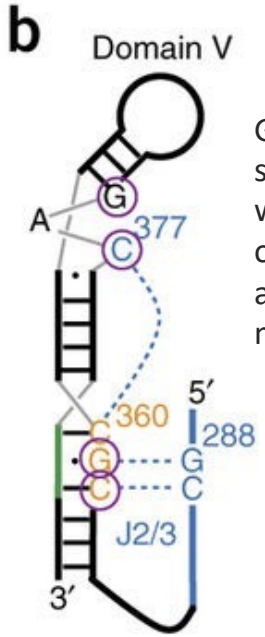


The spliceosome is a protein-assisted ribozyme

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

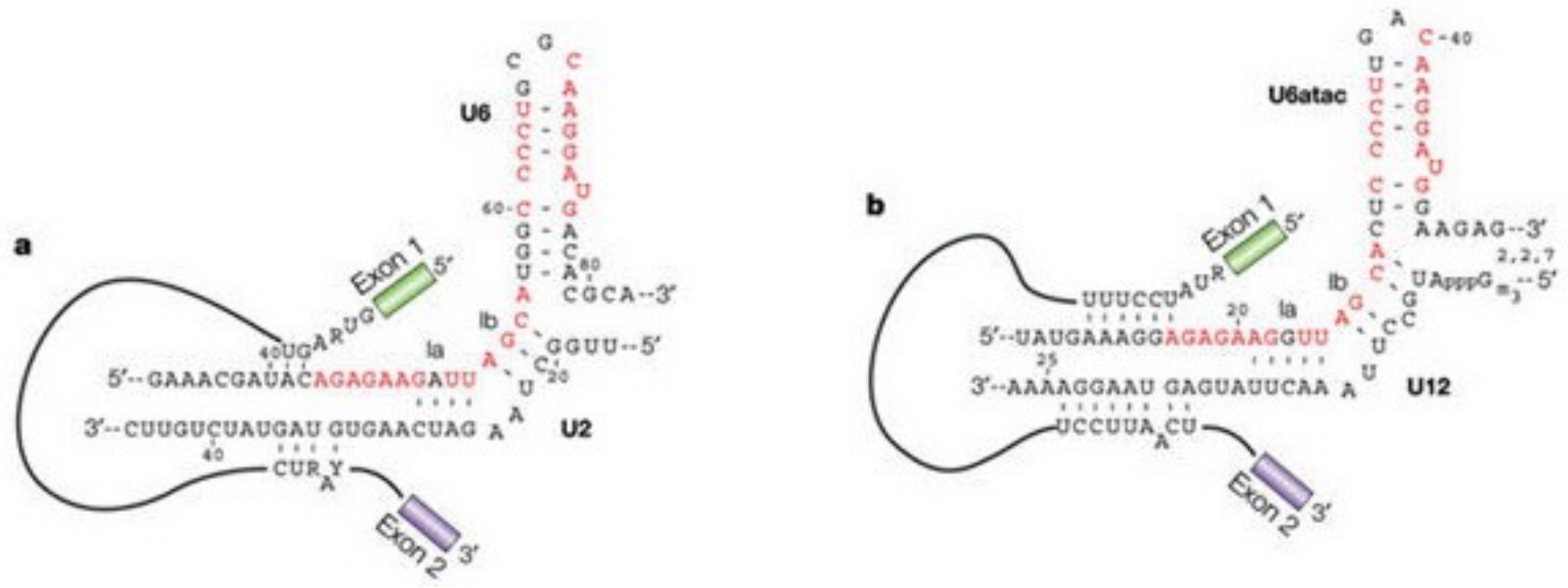


Group II introns

Group II introns are extraordinarily versatile self-splicing ribozymes and retrotransposable elements widespread in bacteria and in bacterial-derived organelles (mitochondria and chloroplasts) of fungi, algae, plants, and of some early-branching metazoans

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.

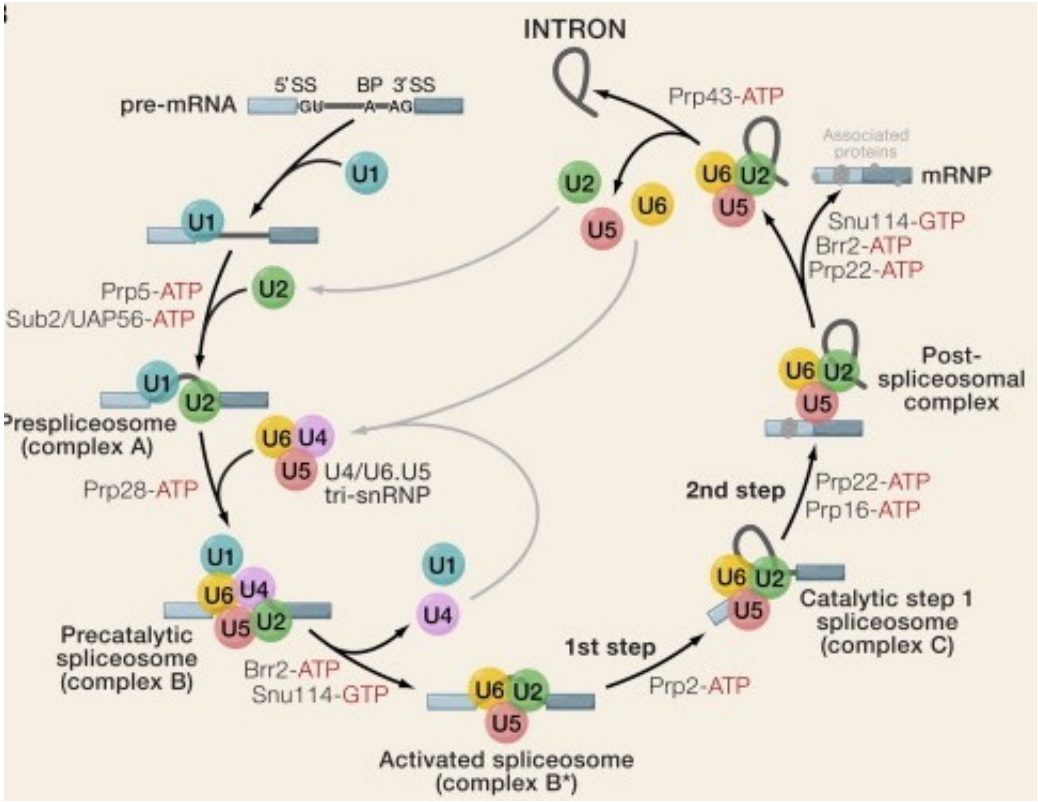


The main mechanistic differences between major- and minor-class splicing are likely to occur at the stage of intron recognition rather than during catalysis.

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 not only necessary for the progression of the spliceosome through the splicing cycle, 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.

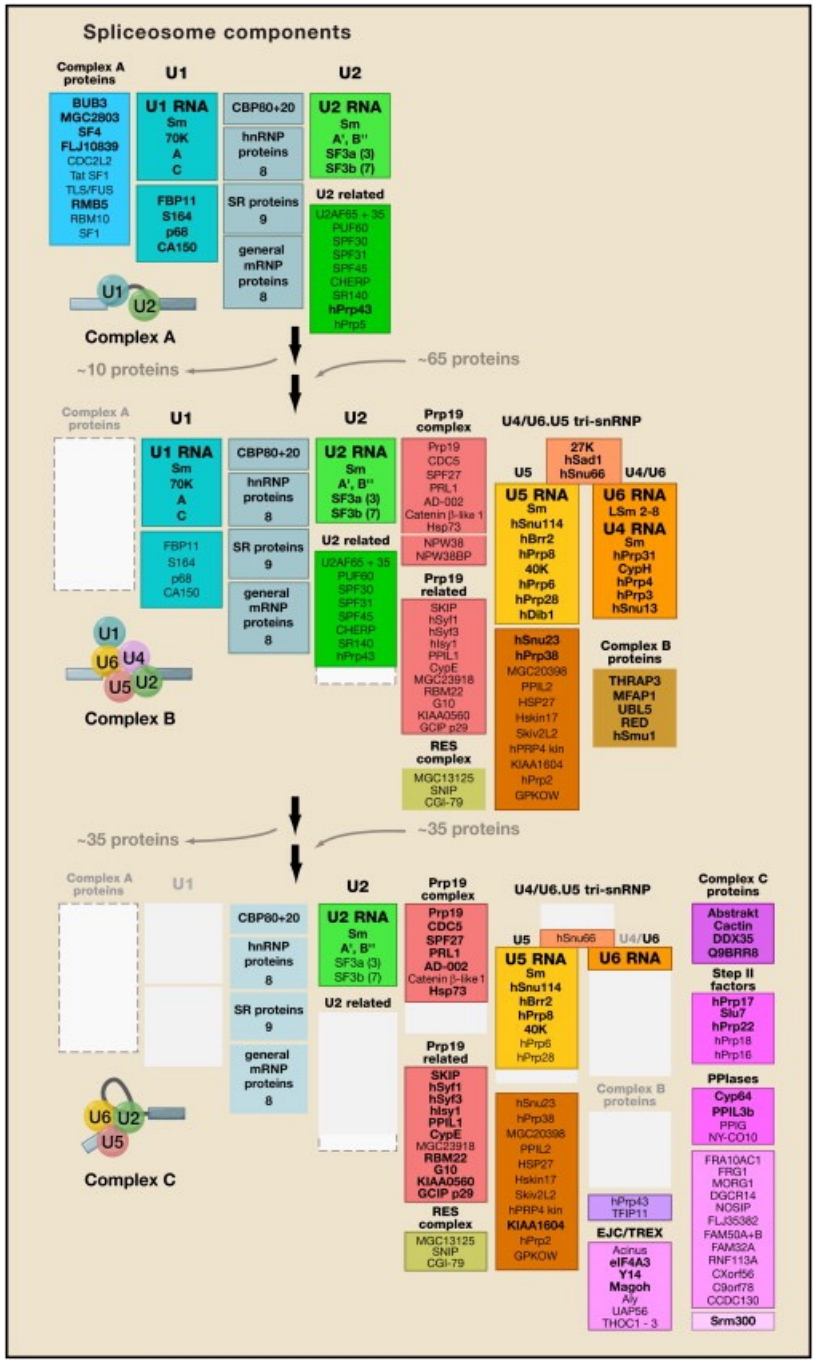


Wahl et al., 2005. Cell

The Spliceosome Is a Protein-Rich Machine

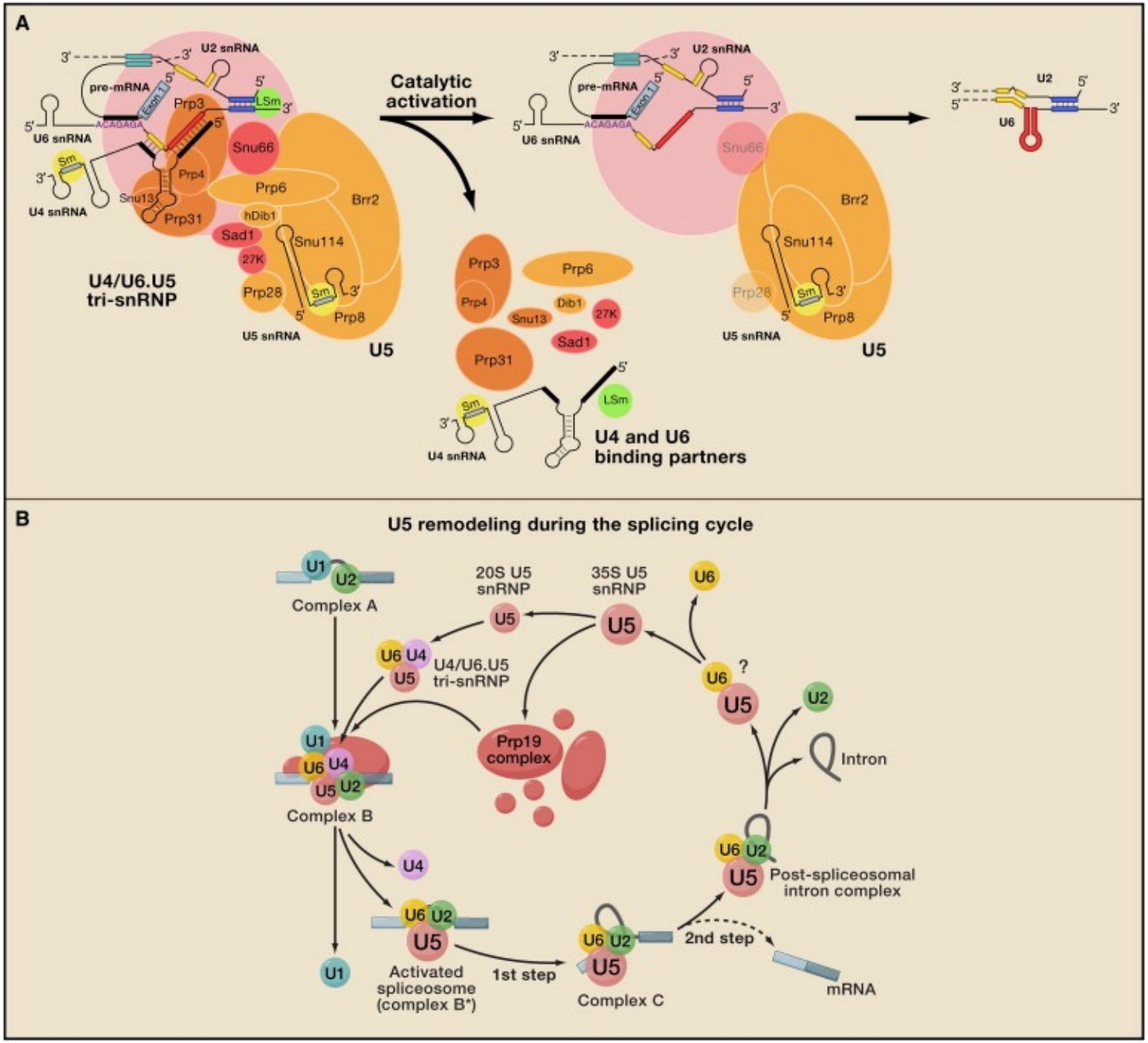
Human spliceosomes contain ~45 distinct snRNP-associated proteins and there is a dramatic exchange of proteins during spliceosome assembly and activation.

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.

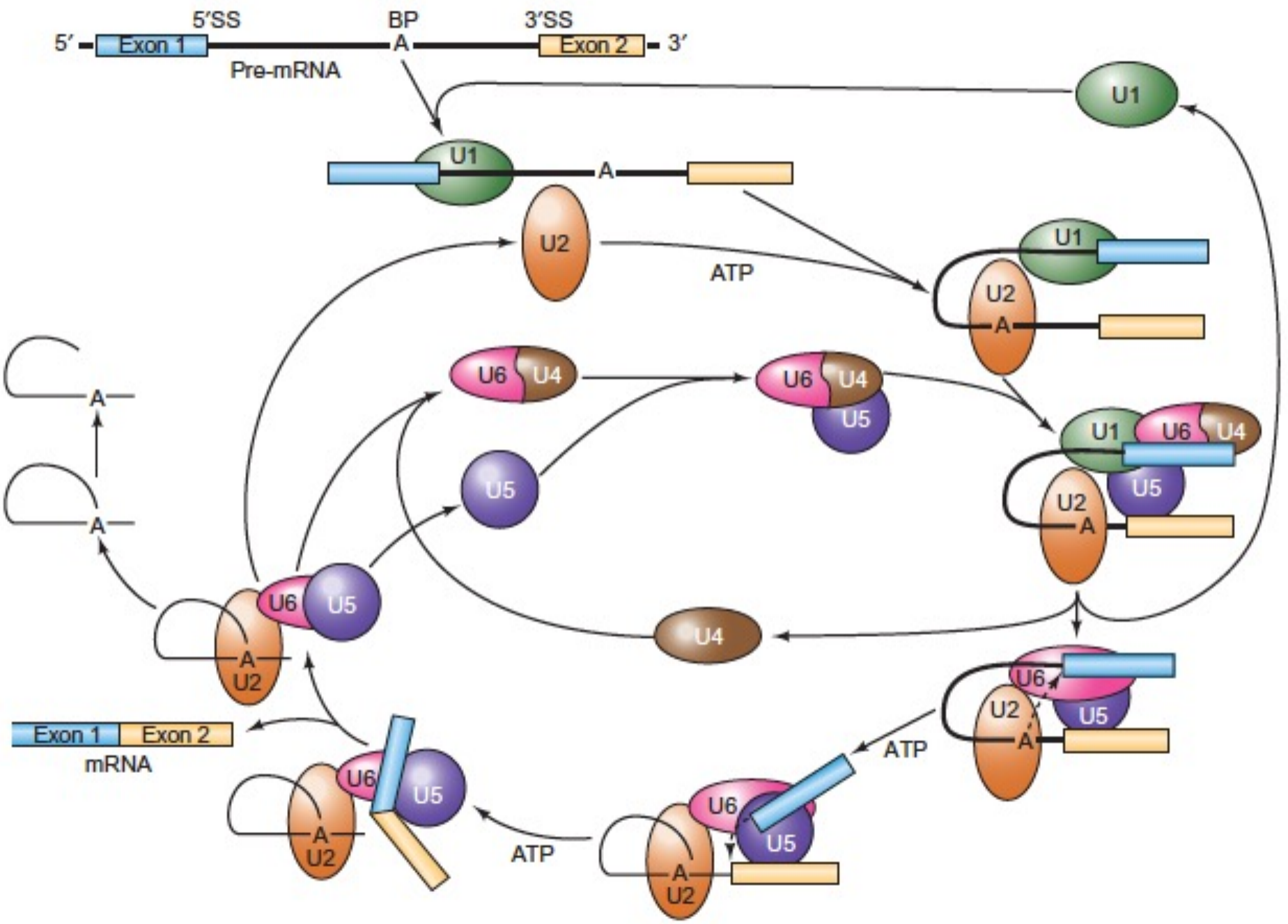


Wahl et al., 2005. Cell

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



The spliceosome cycle



Splicing is mainly co-transcriptional

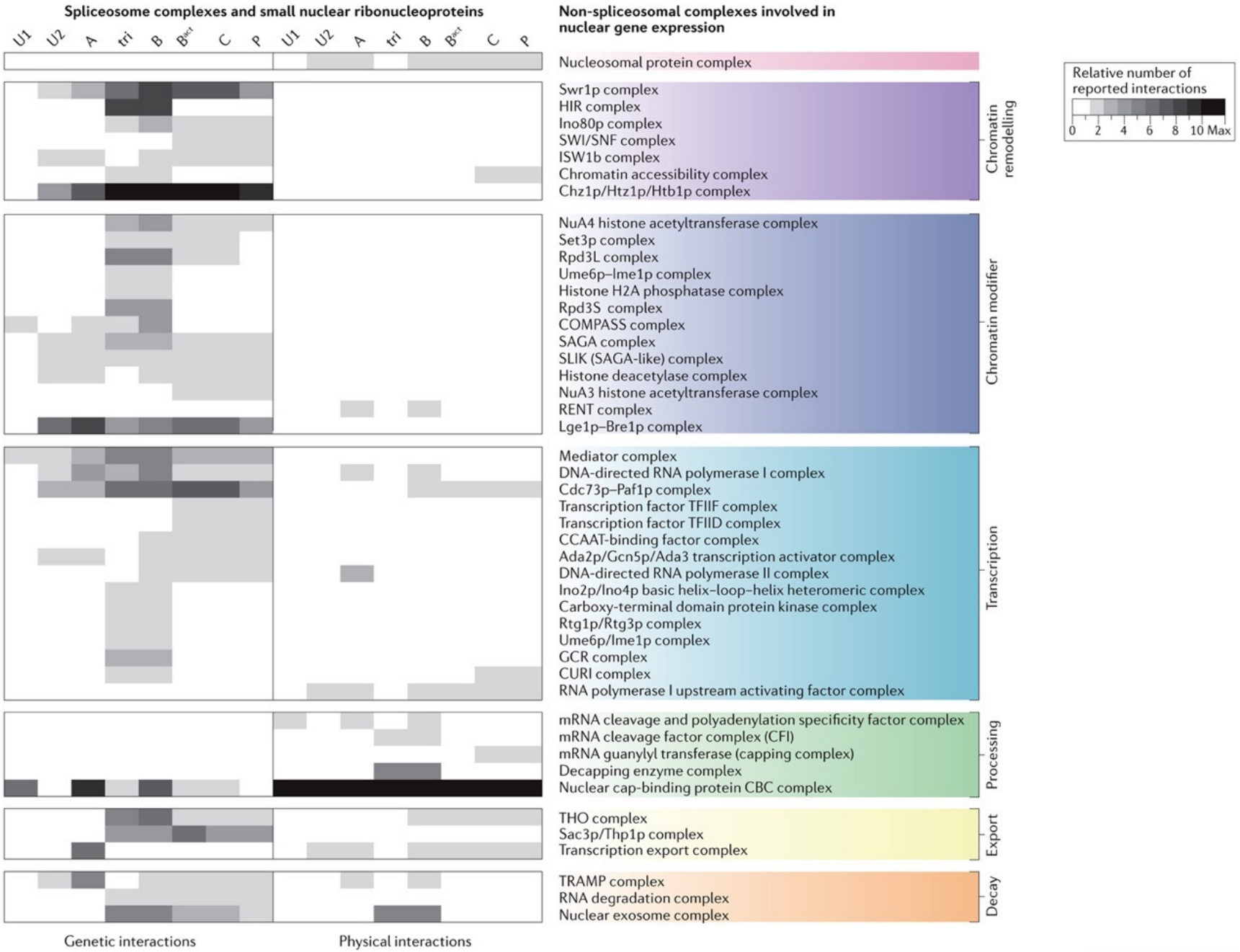
- Chromatin immunoprecipitation (ChIP)-based experiments showed that spliceosomal components accumulate around splice sites following their transcription.

- Single-molecule sequencing of nascent RNA from yeast showed that splicing catalysis occurs when Pol II has transcribed 26–129 nucleotides downstream of the 3′SS.

- Changes in post-translational modifications of the Pol II CTD mirror and influence the different phases of transcription and nascent RNA processing.

- In addition to the Pol II CTD, the entire transcription elongation machinery and the nascent RNA itself interact with proteins, forming RNP complexes. Many of these proteins belong to complexes that are involved in mRNA 5′ end capping, splicing, 3′ end processing, editing, folding, nuclear export and decay, and bind to specific transcript regions, such as untranslated regions, introns and exons

The crosstalk of the spliceosome with nuclear gene expression factors



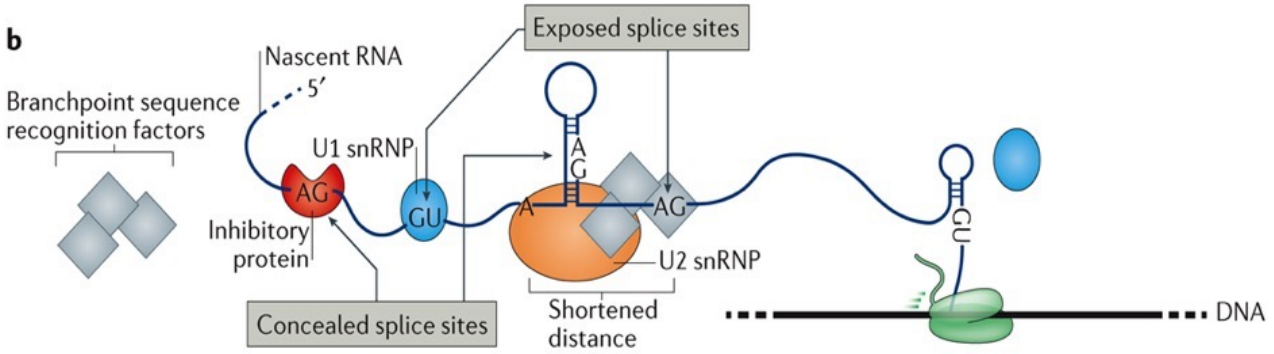
Nature Reviews | Molecular Cell Biology

Transcription and splicing interactions

Spliceosome assembly and catalysis require the recognition of the splice sites and BPS on the nascent transcript. Two main aspects define the co-transcriptional selection:

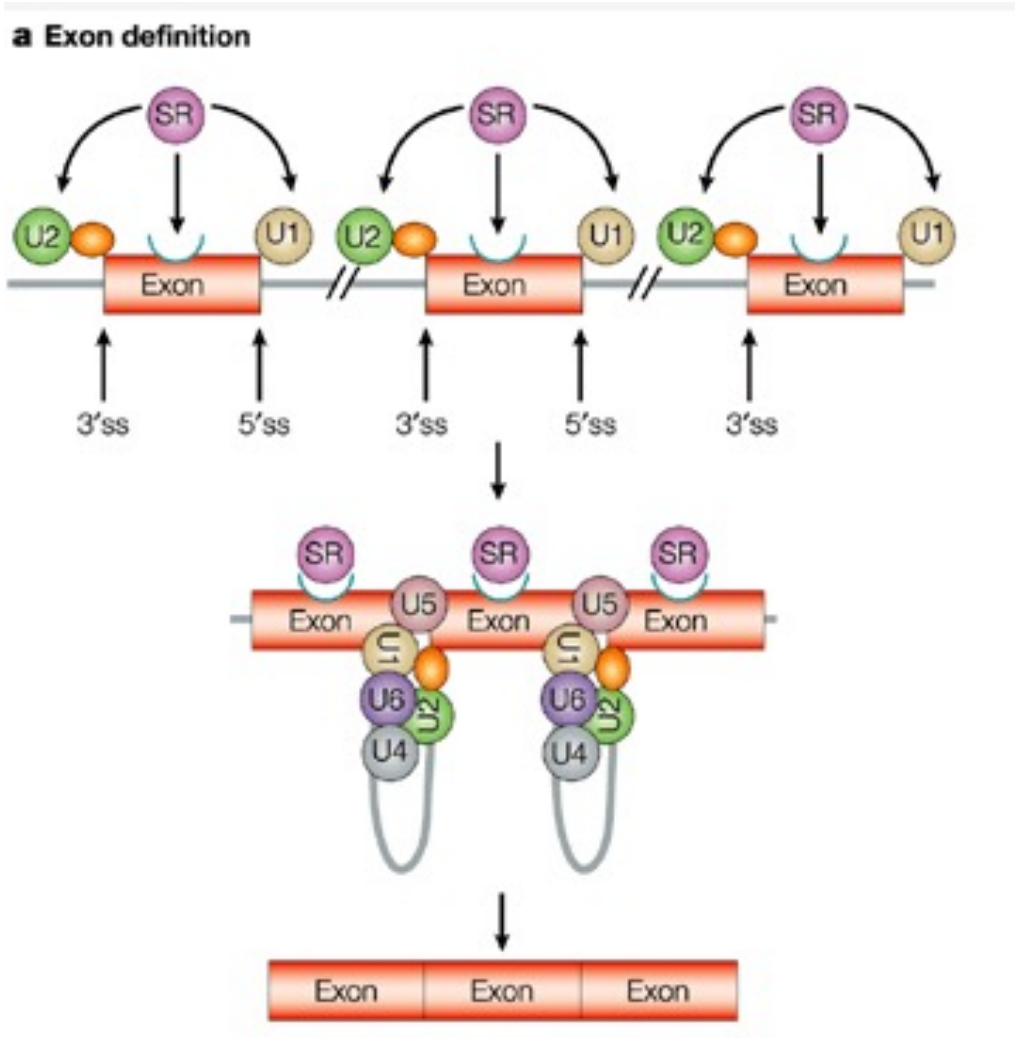
1. only a subset of these sequences is present on the nascent RNA at a given time because of ongoing transcription. Transcription rate determines the portion of the transcript that is available for inspection by the spliceosome.

2. both DNA and RNA undergo different folding states and modifications, and they exhibit dynamic protein-binding profiles that influence transcription rate, accessibility of splice sites and BPS, as well as splicing factor recruitment. Local differences in post-translational modifications of the Pol II-CTD and chromatin environment influence transcription rates and nascent RNA processing dynamics .



How can short exons be identified within very long introns?

Intron > 250 nt

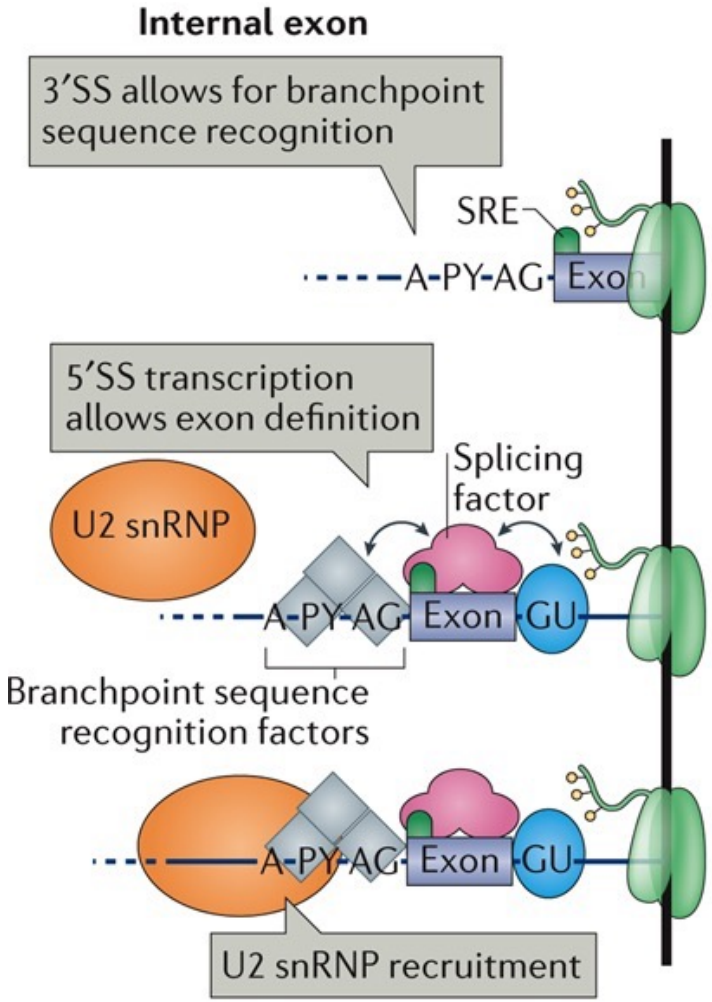


Strategies for splice site identification

The complexity of gene architecture varies between phyla requiring different mechanisms to identify splice sites. Metazoan splice sites are short and poorly conserved, in contrast to budding yeast splice sites. This implies that the spliceosome might encounter frequent 'incorrect' splice sites along the transcripts.

In metazoa, accurate splice site selection depends on short conserved sequences, known as **splicing regulatory elements (SREs)**, that reside in introns or exons.

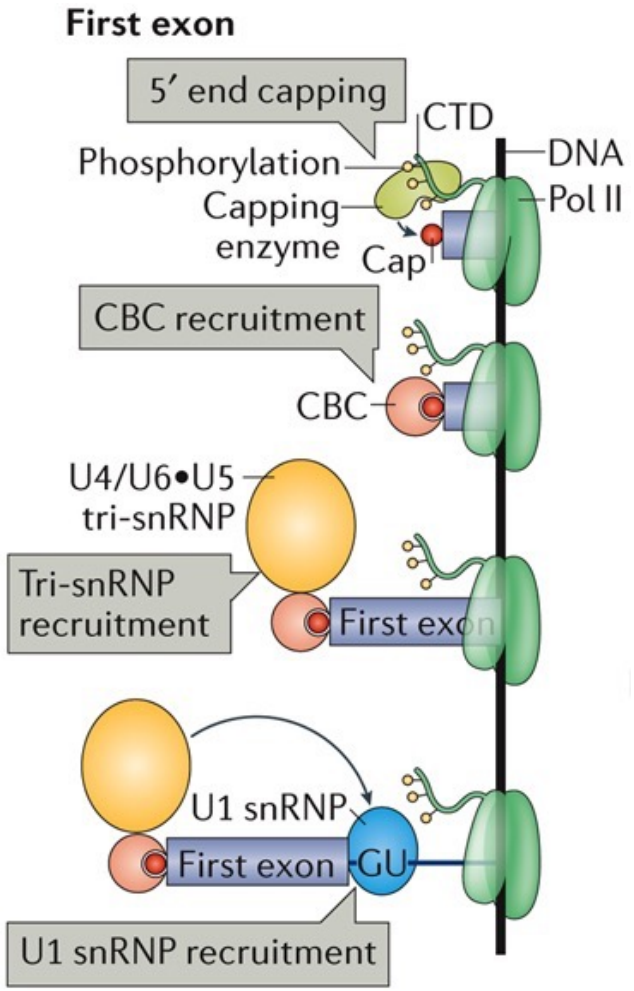
Regulatory proteins, such as **SR proteins** and **hnRNPs**, specifically bind to SREs and influence splice site recognition and/or spliceosome assembly.



First exon definition

First exon: Splicing of the first intron depends on first exon definition . First exon boundaries consist of the 7-methylguanosine (m7G) cap structure at the 5′ end of the transcript and the 5′SS of the first intron. The nuclear **cap-binding complex (CBC)** serves as a platform for interacting with factors that are involved in RNA processing.

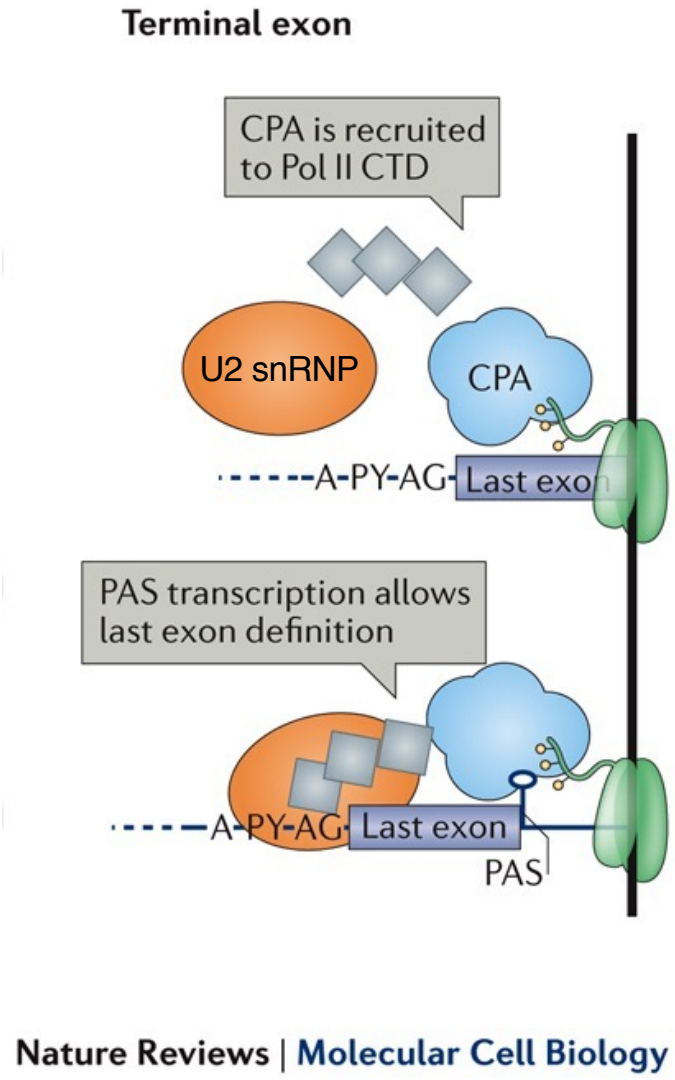
CBC directly interacts with tri-snRNP protein components and its depletion impairs both U1 snRNP and tri-snRNP recruitment to the pre-mRNA .



Terminal exon definition

Terminal exon: The 3′SS of the last intron and the poly(A) site (PAS) set the terminal exon boundaries. PAS elimination results in the specific inhibition of last intron splicing, indicating that 3′ end processing contributes to terminal exon definition. The U2 snRNP, U2AF65 and cleavage and polyadenylation specificity factor (CPSF) functionally and physically interact. Splicing and the regulation of 3′ end processing is reciprocal, as the inactivation of the terminal 3′SS inhibits 3′ end processing and transcription termination.

The mammalian U1 snRNP component U1 70k directly interacts with poly(A) polymerase and inhibits polyadenylation. This may reflect its role in protecting nascent RNA from premature 3′ end processing. In such a model, the U1 snRNP binds to nascent RNA and suppresses the activity of adjacent cryptic PASs.



Nature Reviews | Molecular Cell Biology

Transcription and splicing interactions

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.

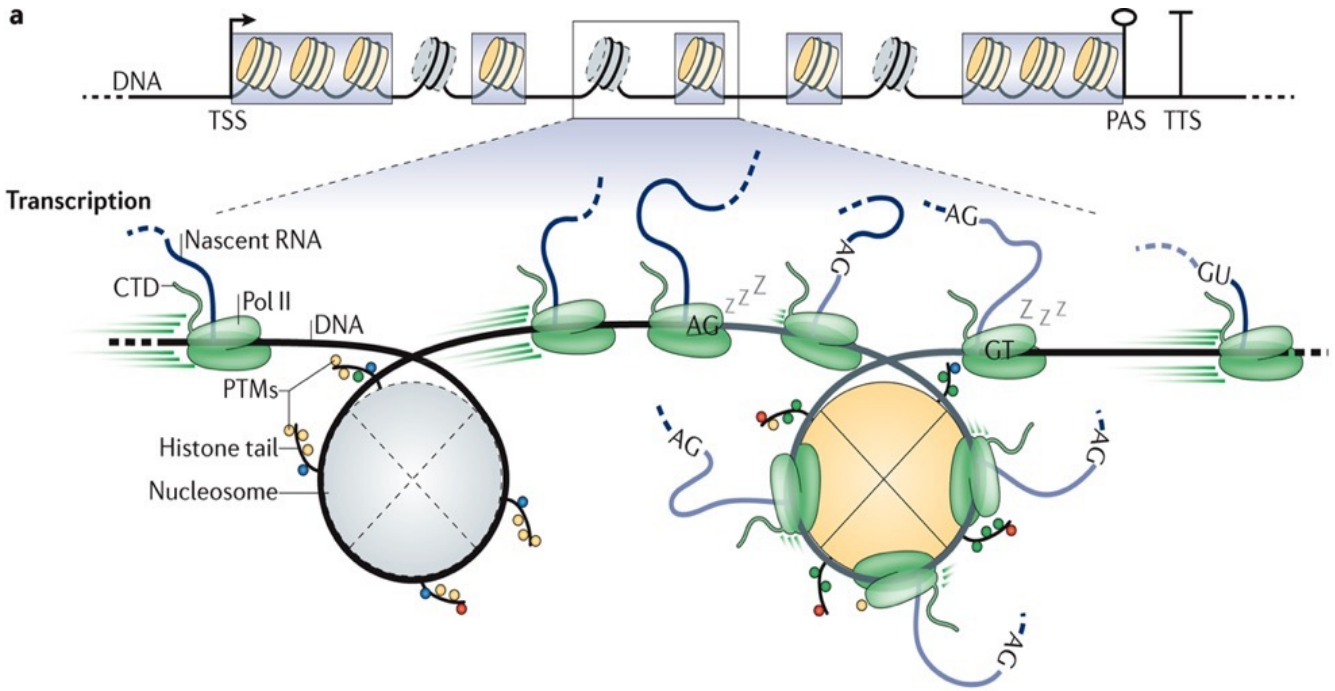
Two models explain how pre-mRNA splicing is controlled by the chromatin state

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

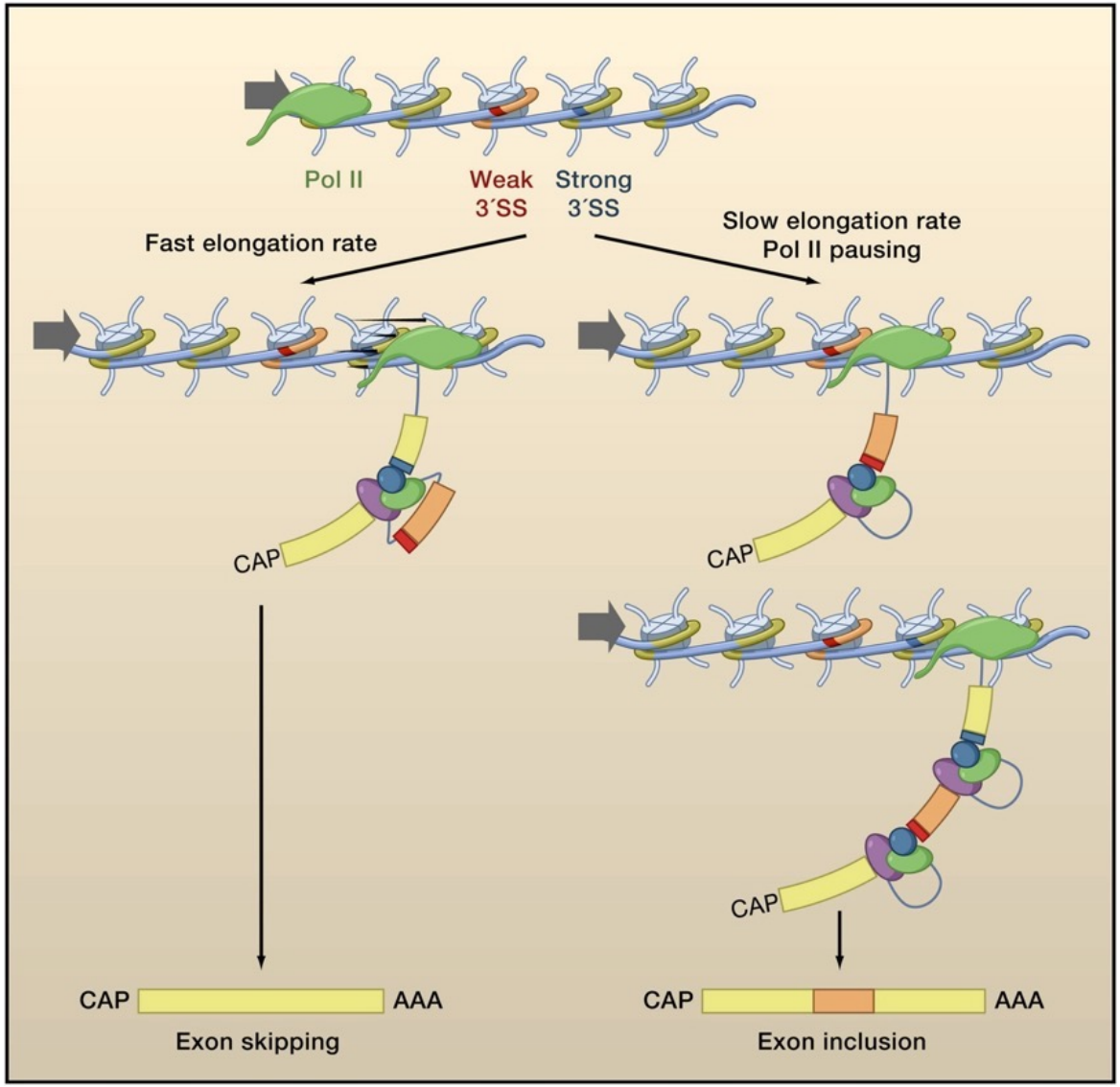
Gene architecture, chromatin features and nascent RNA properties influence co-transcriptional splicing

Pol II elongation rate is influenced by a multitude of factors, such as the underlying DNA sequence, nucleosome position and histone modifications, which affect local chromatin structure, the activity of elongation factors, and the folding and processing of the nascent RNA.



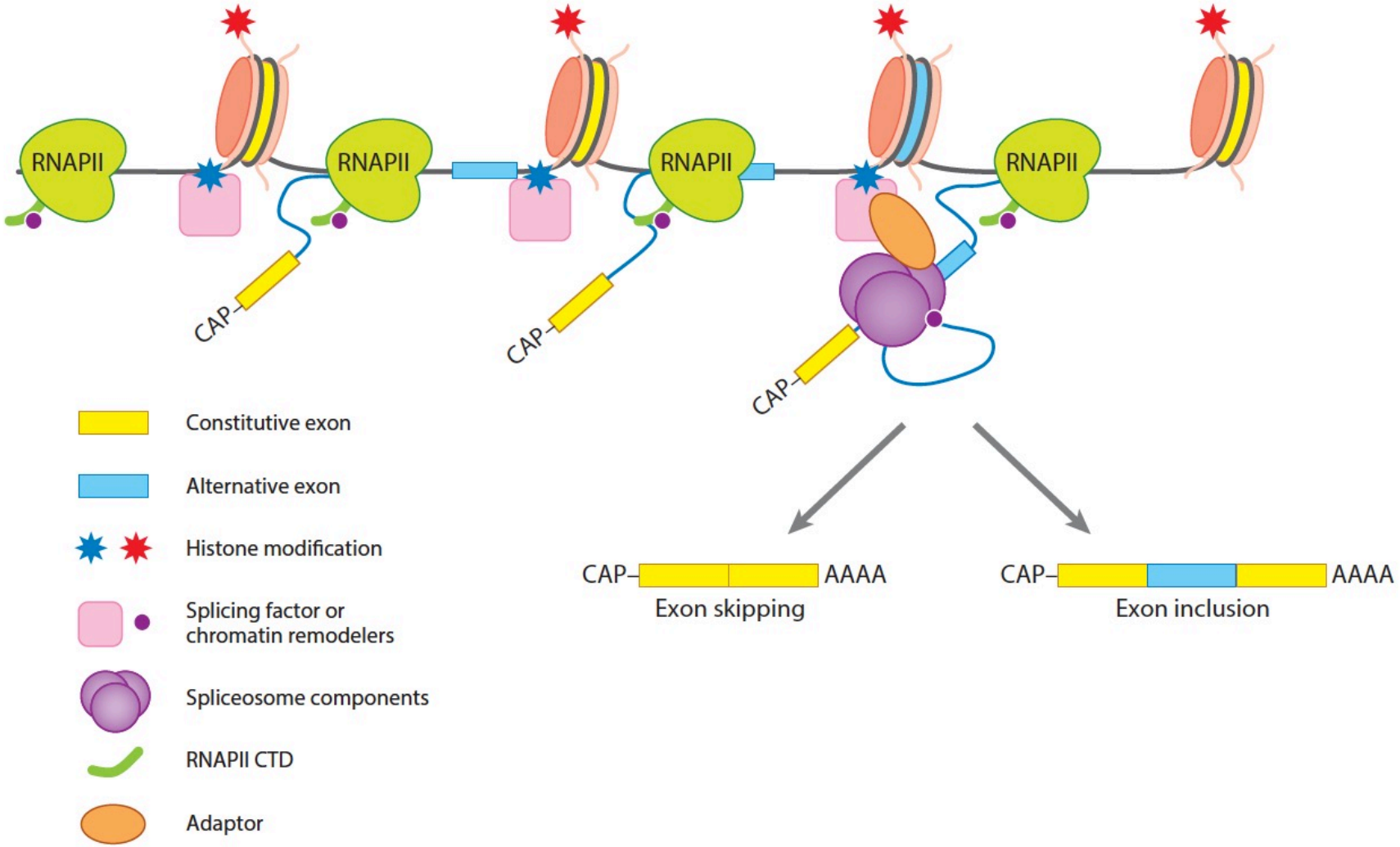
The rate of RNAPII elongation influences splicing

The rate of RNAPII elongation influences alternative splicing by affecting the step at which splice sites and regulatory sequences emerge in the nascent pre-mRNA during transcription. High elongation rates favor skipping, whereas low transcriptional elongation rates favor exon inclusion. Slow elongation causes preferential excision of the upstream intron (first served = first excised).



Kinetic Coupling

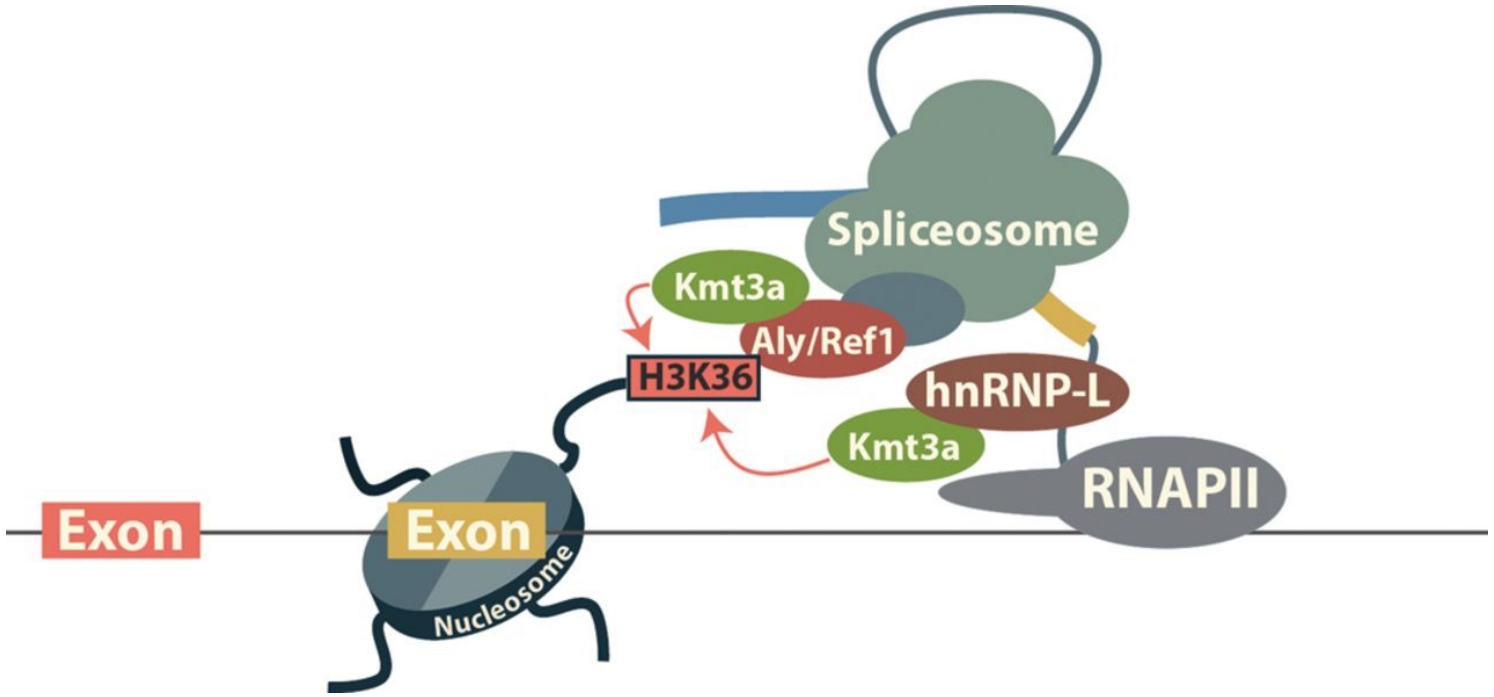
When chromatin meets splicing: 'recruitment' model



H3K36me3 is strongly associated with exonic sequences and correlates with the level of gene expression and the rate of exon inclusion during alternative splicing. Splicing and H3K36me3 modification **can stimulate each other.**

Splicing stimulates H3K36me3 modification

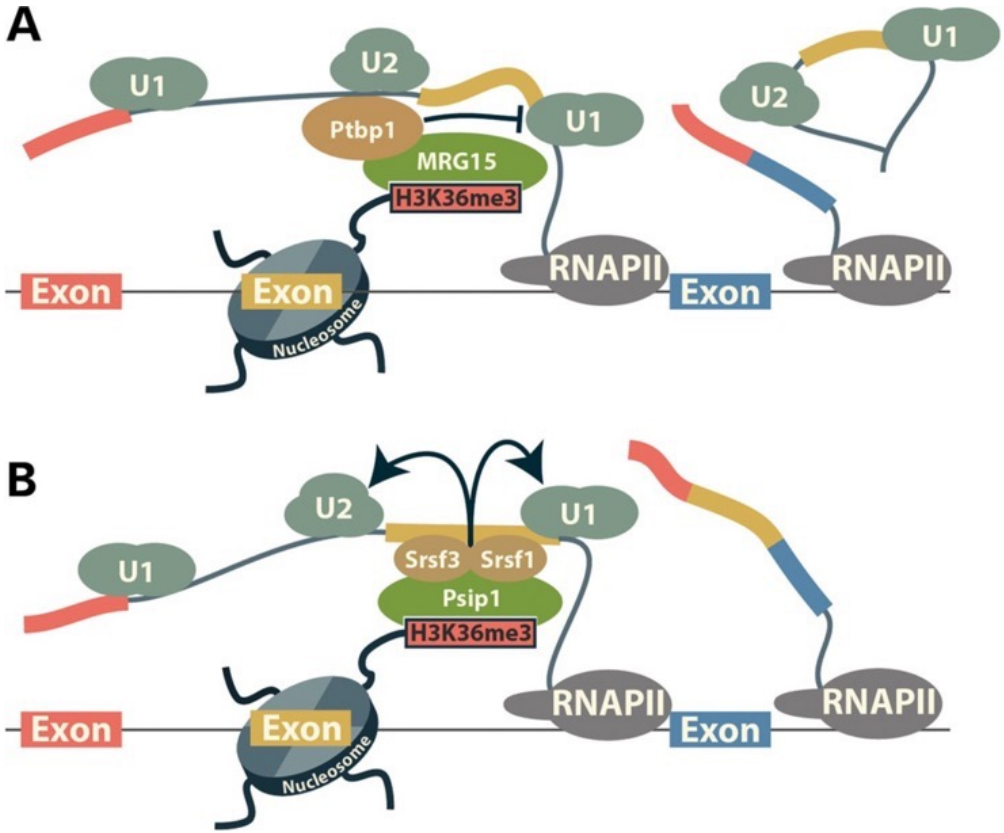
There is a physical association between components of the splicing machinery and the **histone methyltransferases Kmt3,** which installs H3K36me3 modification **.**



In agreement with a mechanism that involves the recruitment of Kmt3a by the spliceosome, inhibition of splicing by splice site mutagenesis, pharmacological agents or depletion of Sf3b3 (SAP130), a component of the core spliceosome, decrease Kmt3a recruitment to the chromatin and H3K36me3 levels.

Seth J. Brown et al. Hum. Mol. Genet. 2012;21:R90-R96

Recruitment model for the control of splicing by H3K36me3

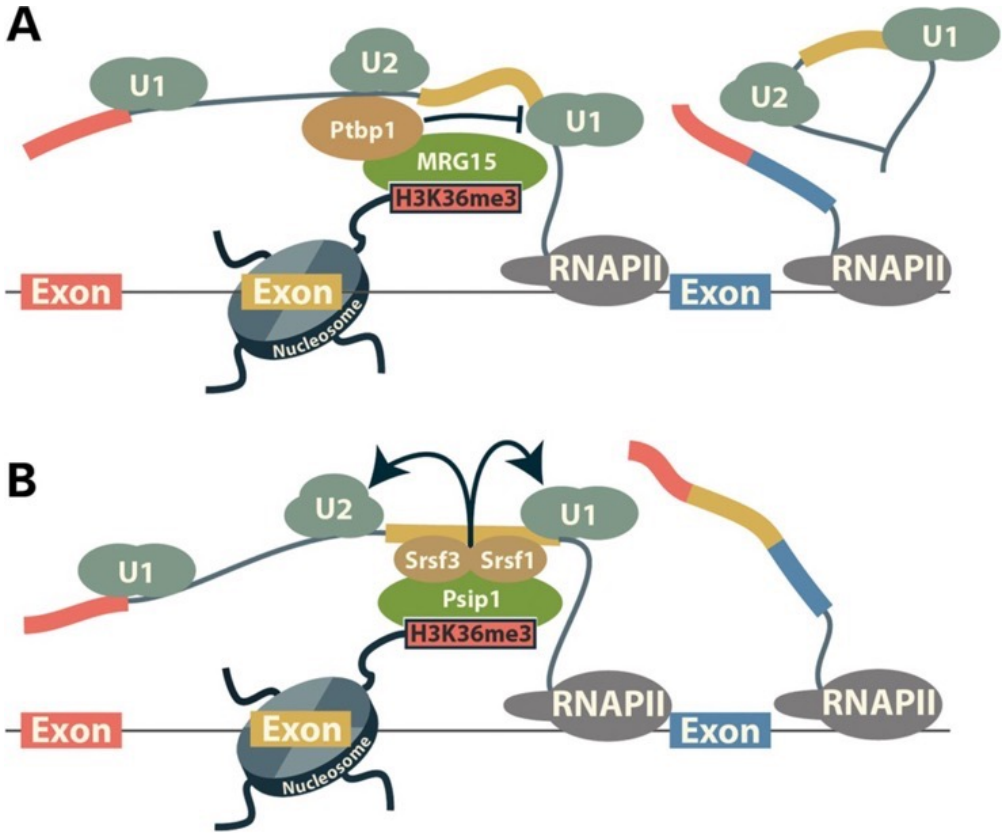


The MRG15 and Psip1 proteins 'read' the H3K36me3 mark by binding and can regulate alternative pre-mRNA splicing by recruiting splicing factors to the chromatin.

MRG15 recognizes H3K36me3 and brings the **Ptbp1** splicing factor to the nascent transcripts .

Psip1 interacts directly with H3K36me3 and associates with a large number of factors involved in pre-mRNA splicing that include SR and hnRNP proteins, DEAD-box helicases and core spliceosome components.

Recruitment model for the control of splicing by H3K36me3



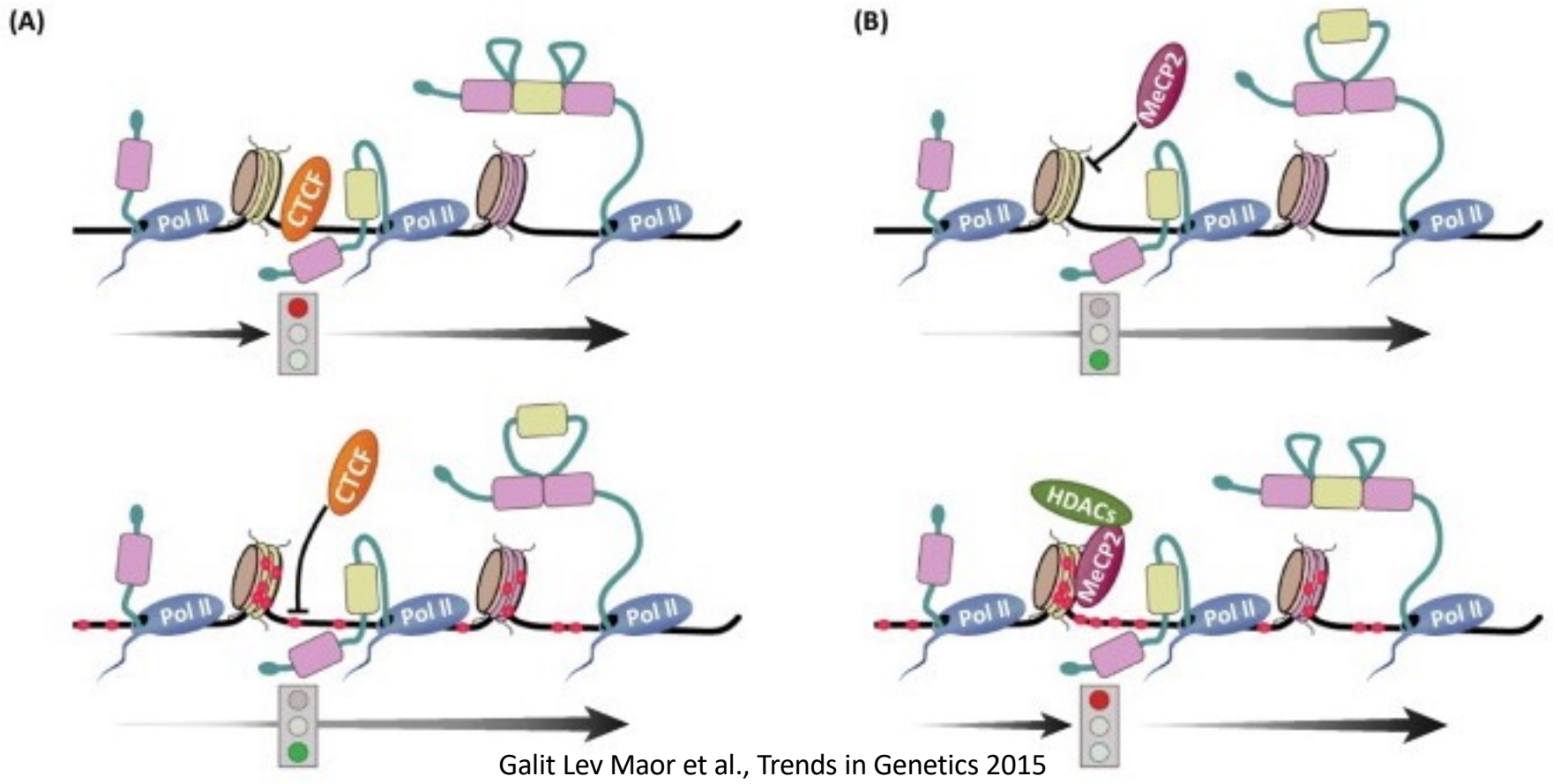
H3K36me3 can **repress** the inclusion of a subset of exons by recruiting Ptbp1 via its interaction with MRG15.

At the same time, the splicing of other exons is **facilitated** by the recruitment of Srsf1 and Srsf3 to H3K36me3 by a different reader, Psip1.

An intriguing consequence of the involvement of H3K36me3 in pre-mRNA splicing is that it completes a regulatory circuit, where pre-mRNA splicing stimulates H3K36me3 over the exons, which in turn promotes exon recognition by recruiting pre-mRNA splicing factors.

Seth J. Brown et al. Hum. Mol. Genet. 2012;

DNA methylation and splicing



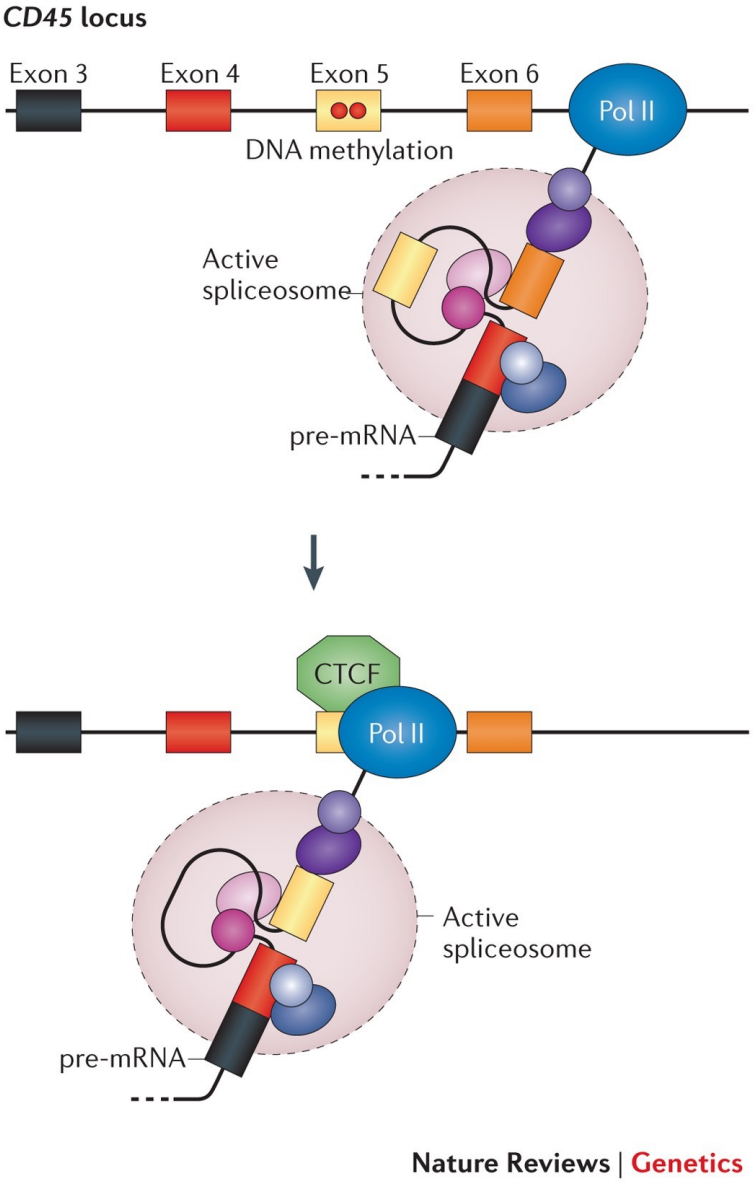
Galit Lev Maor et al., Trends in Genetics 2015

CTCF binds downstream to the alternative exon and creates a roadblock for Pol II elongation which results in exon inclusion on the mature mRNA.
DNA methylation downstream of the alternative exon prevents CTCF binding and promotes fast Pol II elongation and exon skipping on the mature mRNA

In the absence of DNA methylation, MeCP2 does not bind to the alternative exon, which enables fast Pol II elongation and leads to exon skipping in the mature mRNA.
Methylation of the alternative exon causes MeCP2 to bind to the exon and recruit enzymes with HDAC activity, which slows Pol II elongation and results in exon inclusion in the mature mRNA.

DNA methylation and splicing

CD45 is a trans-membrane protein tyrosine phosphatase that initiates signalling through antigen receptors by dephosphorylating the inhibitory tyrosine on Src family kinases. CD45 gene expresses alternatively spliced transcripts during lymphocyte differentiation.



In the absence of CpG methylation CTCF binds to its binding site and facilitates the inclusion of CD45 exon 5 by inducing RNA pol II pausing at this exon.

This result provides evidence that DNA methylation influences splicing. Therefore, it is conceivable that changes in DNA methylation patterns during development or in disease, particularly in cancer, could determine alternative splicing outcome and subsequently affect transcriptome makeup.

Chin-Tong Ong & Victor G. Corces
Nature Reviews Genetics (2014)