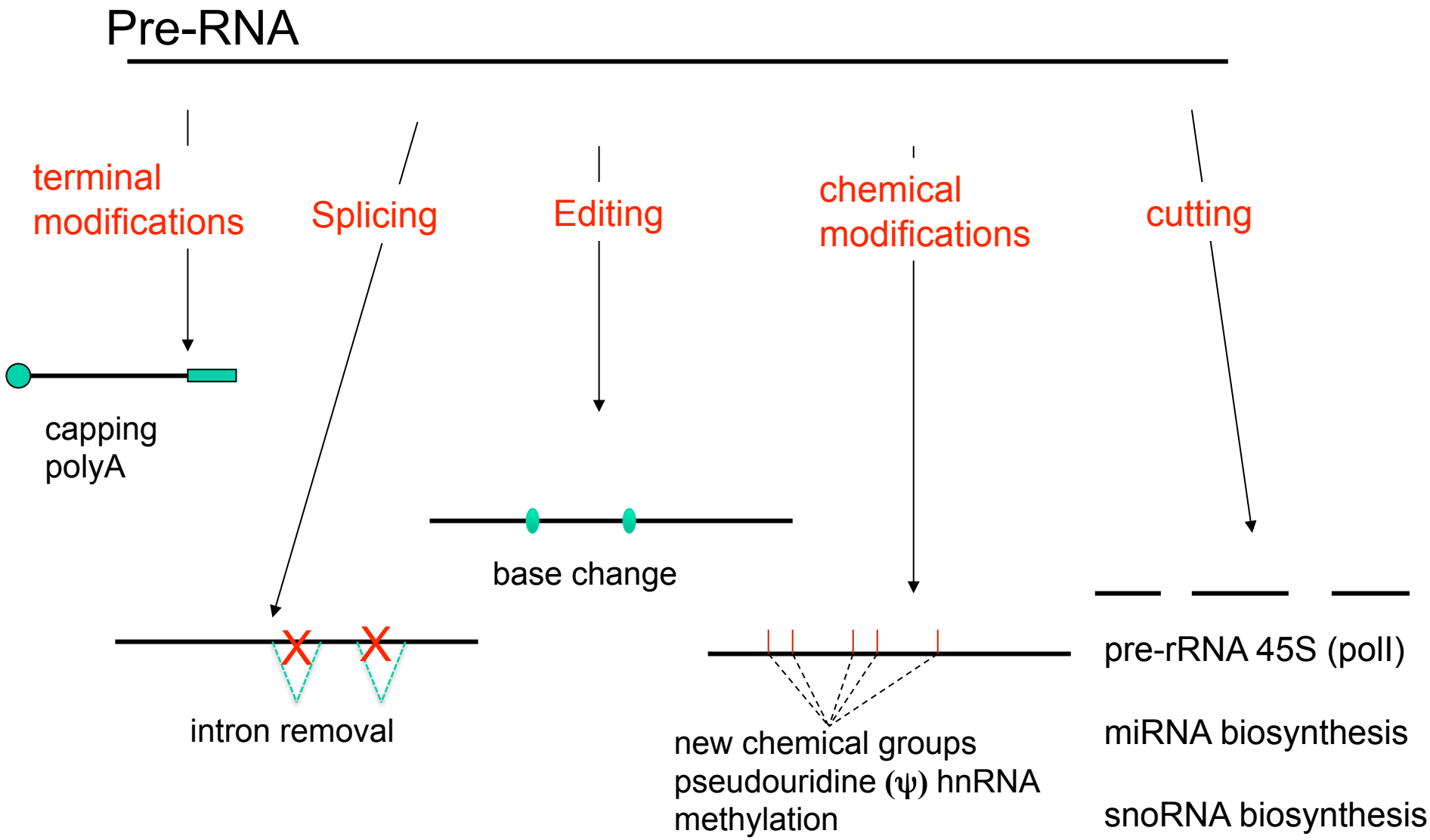
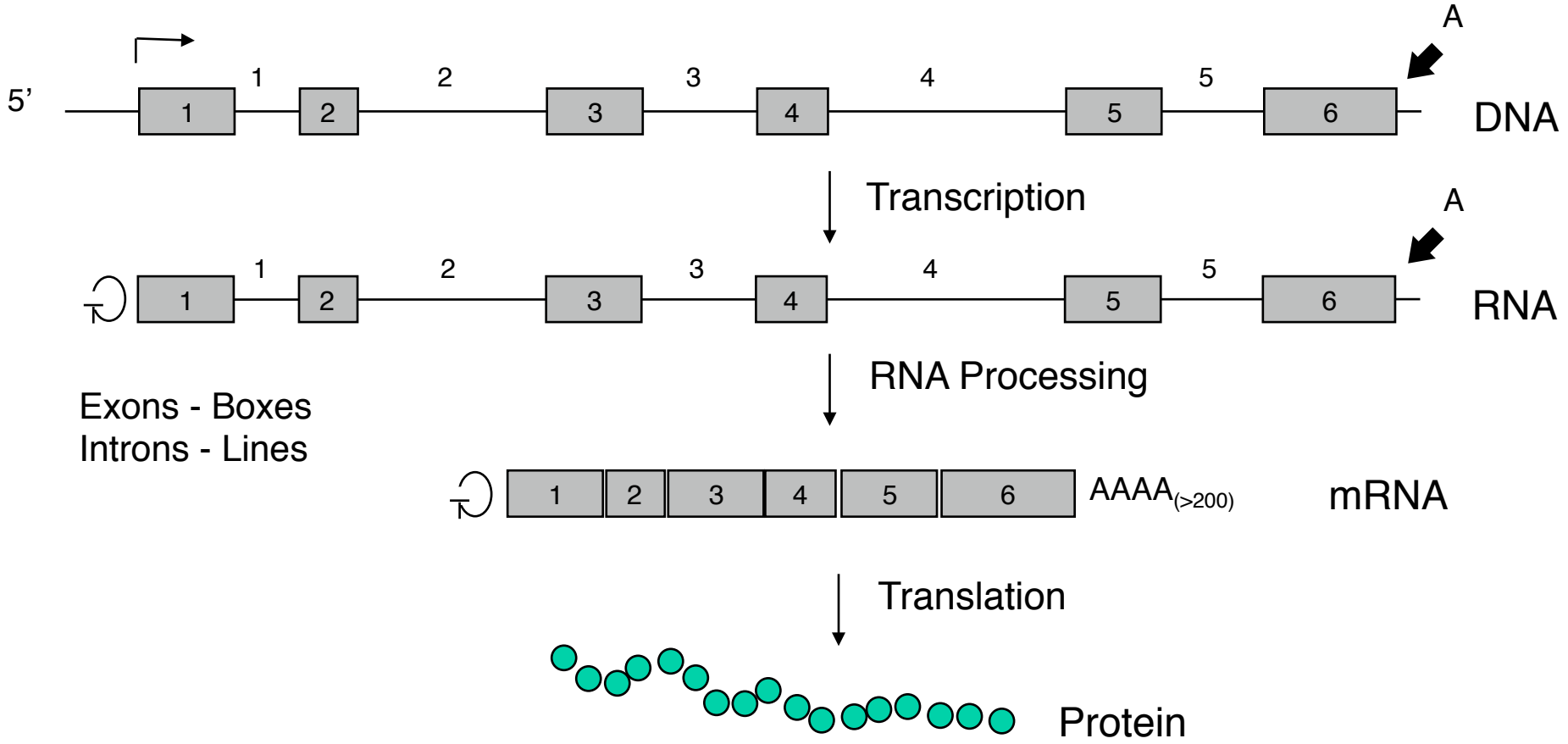
Post-Transcriptional Regulation: Splicing

pre-RNA processing



Genes Exist in Pieces



RNA Splicing Puts the Pieces Together

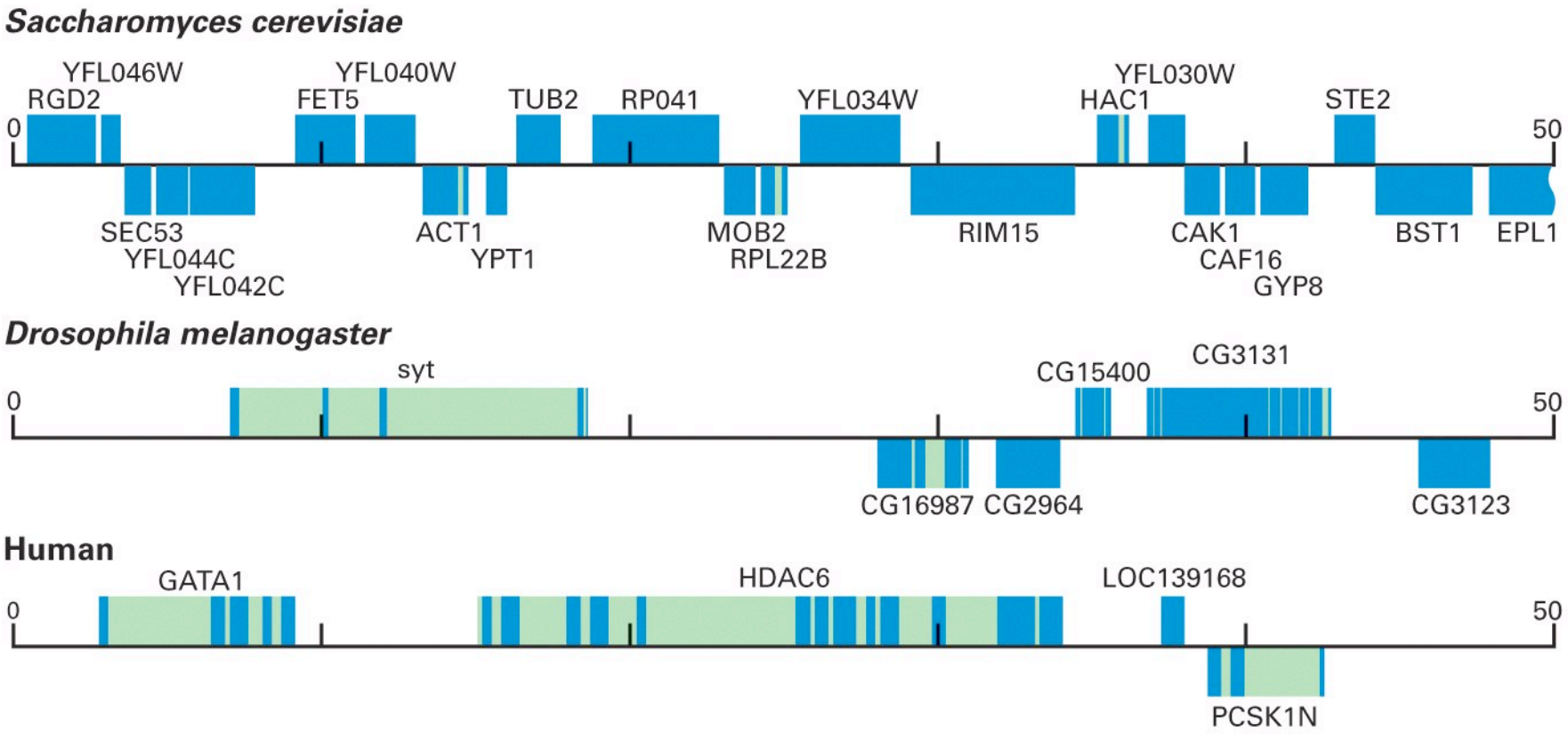
Exon - Ex pressed Sequences
Intron - In tervening Sequences
The terms were coined by Walter Gilbert in 1978

Definizione di esone

Segmento di un gene eucariotico che consiste di una sequenza di nucleotidi che sarà rappresentata nell'RNA messaggero o nell'RNA transfer o ribosomiale finale. Nei geni che codificano per proteine, gli esoni codificano aminoacidi della proteina.
(Alberts & al. Molecular Biology of the Cell)

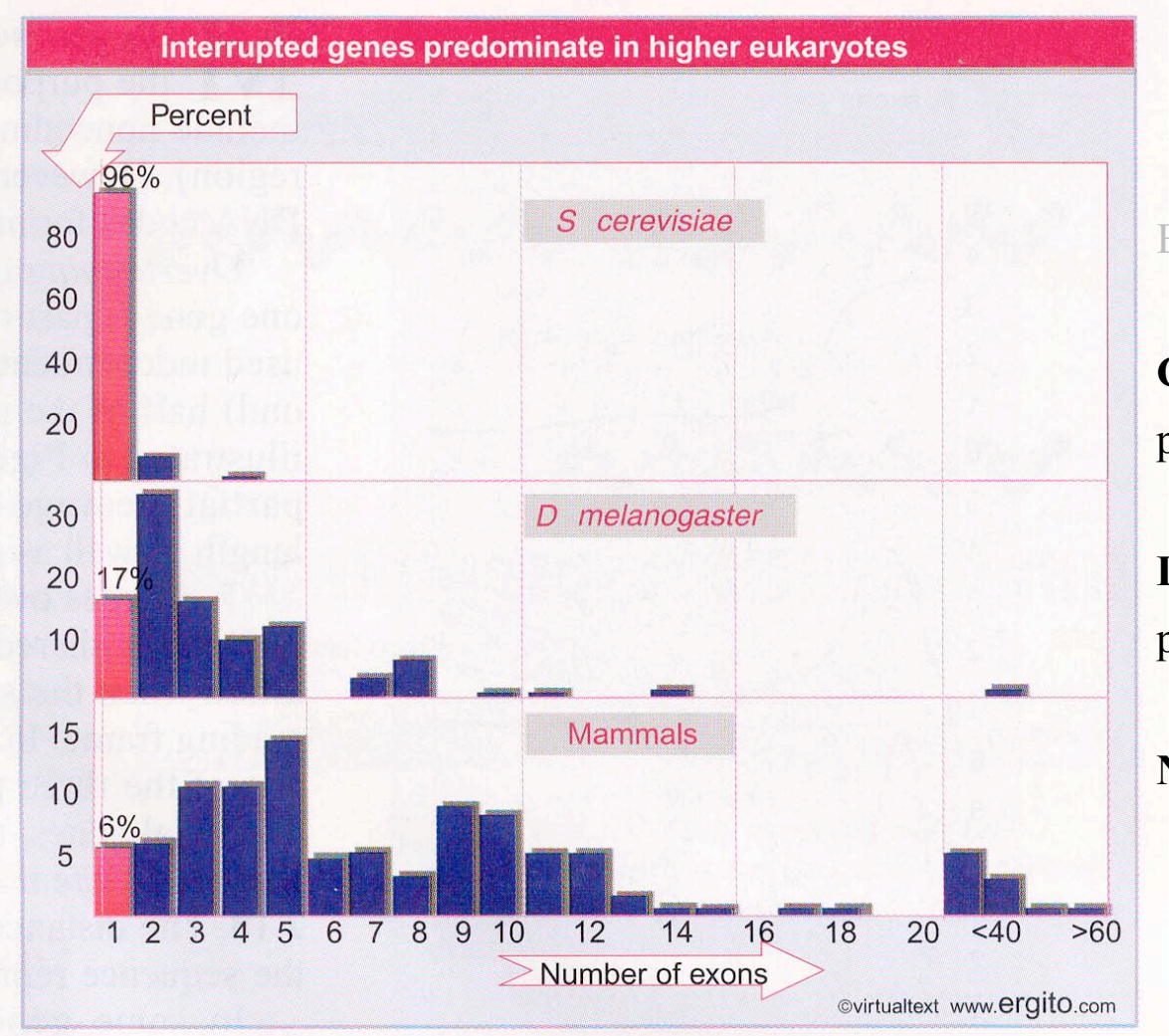
Segmento di un gene che viene rappresentato nel prodotto di RNA maturo. I singoli esoni possono contenere DNA codificante e/o non codificante (sequenze non tradotte).
(Strachan & Read, Human Molecular Genetics)

Arrangement of Eukaryotic Genes



Yeast genes have long exons and infrequent short introns.
Human genes have short exons and frequent long introns.

Most mammalian genes contain more than one intron



Extreme Examples:

Collagen Gene - 50 exons and a 40 kb precursor RNA

DMD Gene - 79 exons, 2.3 mb precursor, intron 20 is 180 kb

Neurexin - has a 0.5 mb intron!

Most genes are uninterrupted in yeast, but most genes are interrupted in flies and mammals

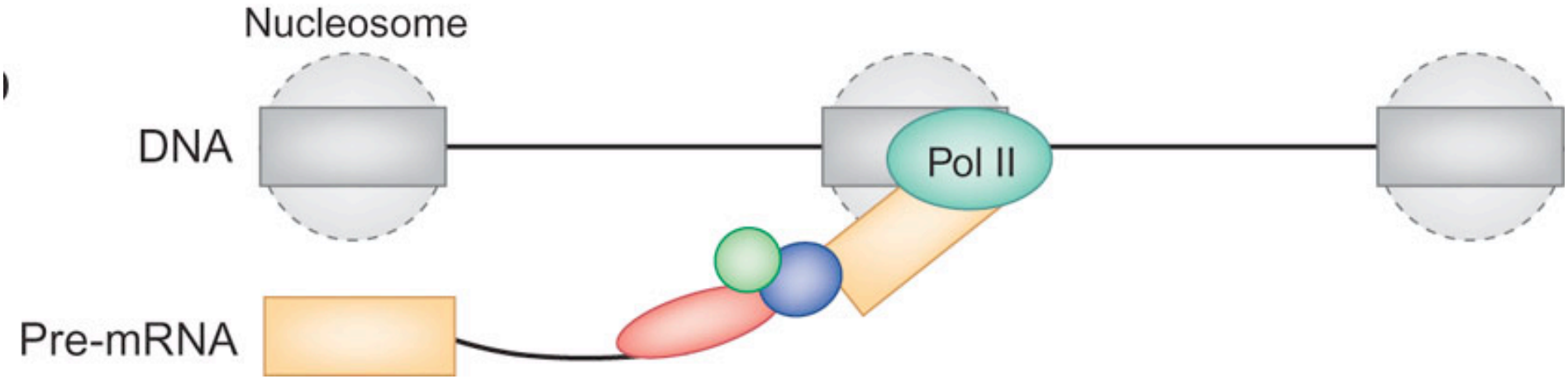
Intron statistics

Species	Average exon No.	Average intron No.	Average length(kb)	Average kb mRNA	% exon per gene
Yeast	1	0	1.6	1.6	100
Nematode	4	3	4.0	3.0	75
Fruit fly	4	3	11.3	2.7	24
Chicken	9	8	13.9	2.4	17
Mammals	7	6	16.6	2.2	13

Human genes

	Median	Mean
Size of internal exons	122 bp	145 bp
Number of exons	7	8.8
Size of introns	1023 bp	3356 bp

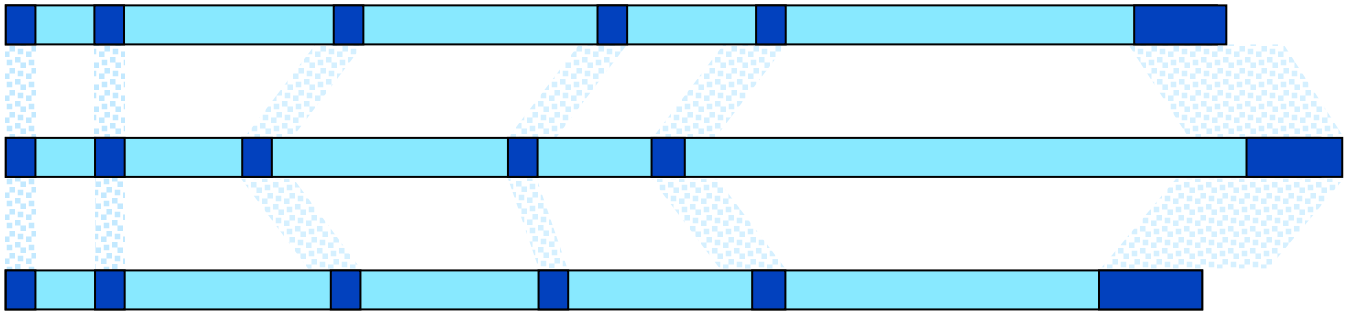
Mean exon size corresponds to nucleosome accommodation on DNA

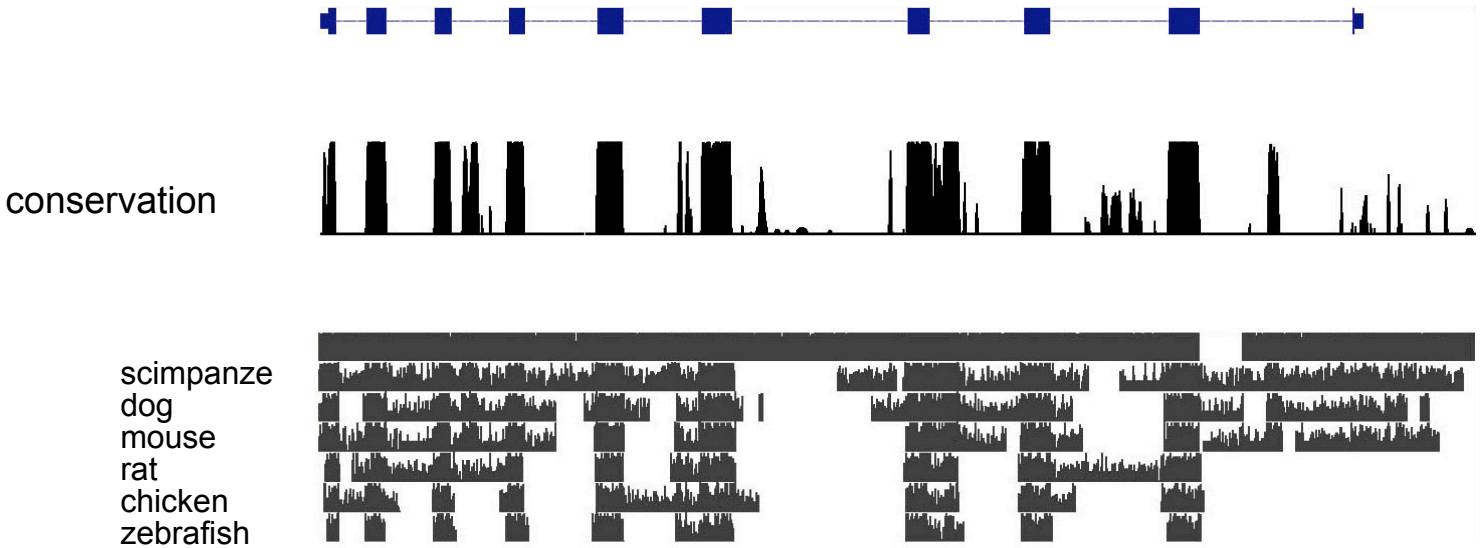


Nucleosomes (broken circles) are preferentially bound to exons, whereas introns are mostly devoid of nucleosomes. Exons are therefore marked at the DNA level by nucleosome positioning, which may act as 'speed bumps' for RNA polymerase II, helping in the co-transcriptional recruitment of splicing factors to the nascent pre-mRNA and improving exon definition.
As nucleosomes accommodate DNA stretches of approximately 147 nt, their preferential location on exons (mean size 145bp) may act as the selective pressure factor for the conservation in exon length.

Nature Structural & Molecular Biology 16, 902 - 903 (2009)
When chromatin meets splicing Kornblihtt1 et al

Orthologous genes





Intron origin

There are currently two opposing theories of intron origin.
The introns-early theory proposes that introns already existed at the progenote (i.e., the last common ancestor of prokaryotes and eukaryotes) to facilitate the construction of the first genes.
The introns-late theory, on the other hand, holds that genes at the progenote were intronless, similar to those in present-day prokaryotes, and introns were gained late, after the emergence of eukaryotes.
There has been no decisive resolution to the debate, and each of these theories has supporting arguments that have not been satisfactorily disproved.

Intron advantages

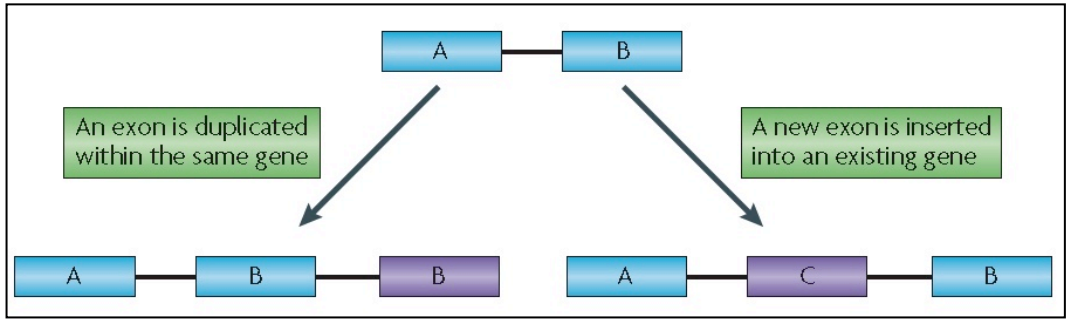
a) Alternative splicing
b) Create new exons (exon shuffling)
c) Create new exons (intron exonization, *Alu* repeats)

exon shuffling

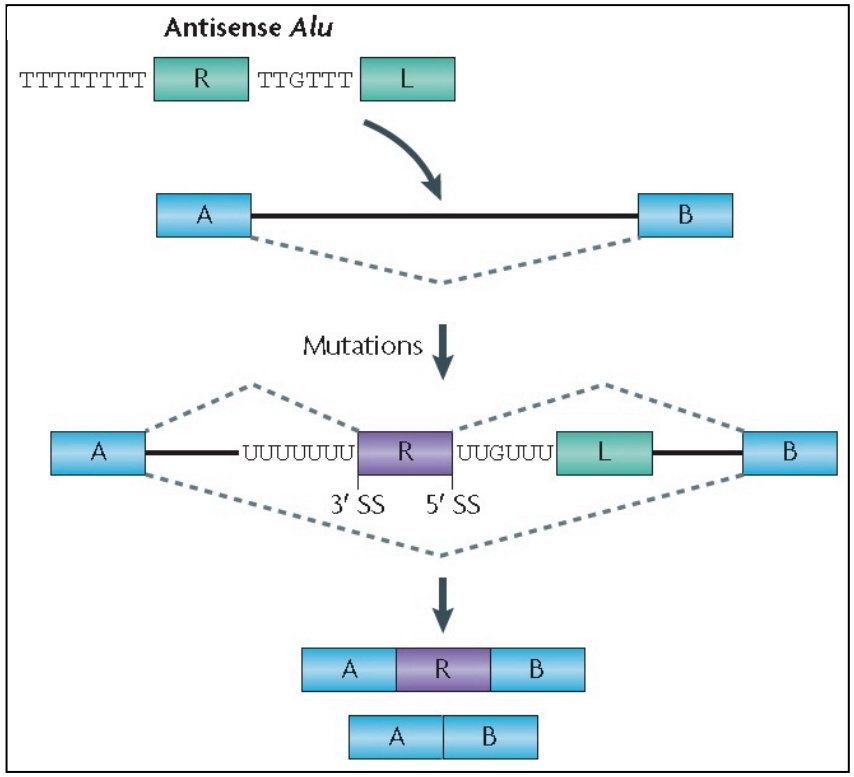
1) exons-> proteins domains
2) many genes have evolved with exon duplication or shuffling
3) similar exons are shared by different genes

Ways of exon expansions

Exon shuffling

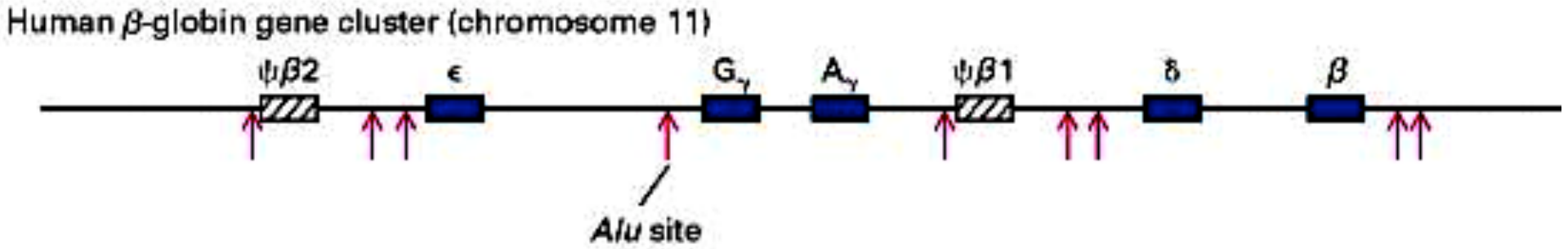


Alu exonization



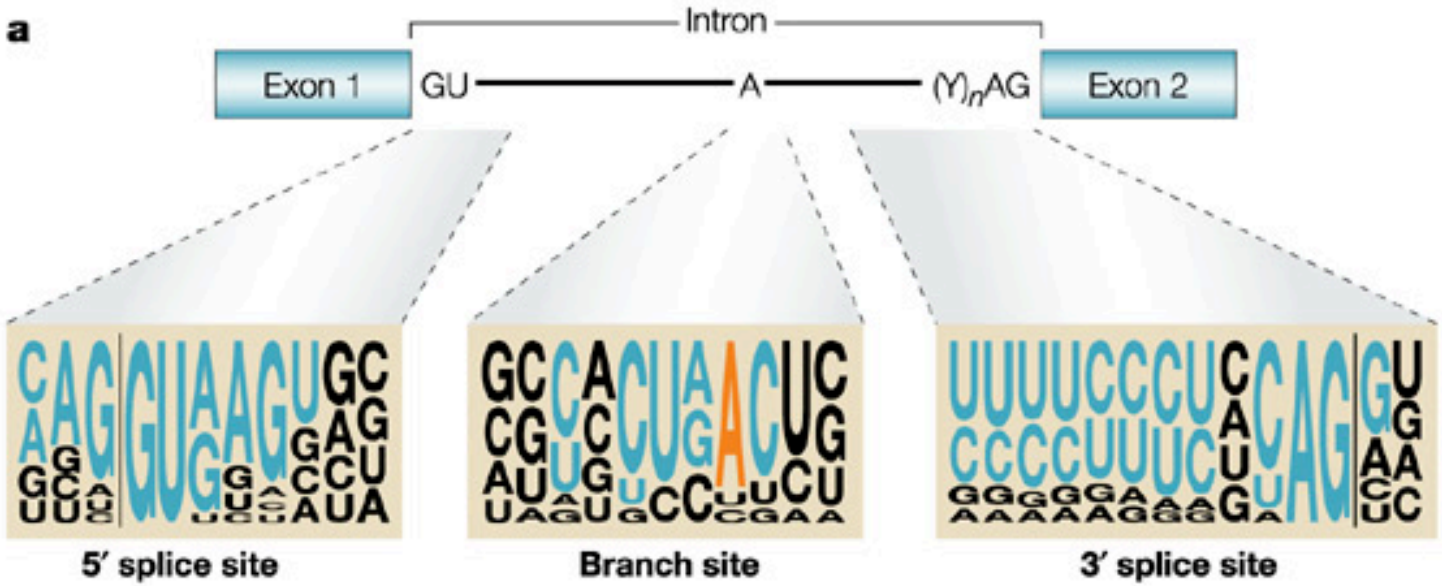
Alu elements

- Length = ~300 bp
- Repetitive: > 1,400,000 times in the human genome
- Constitute >10% of the human genome
- Found mostly in intergenic regions and introns
- Propagate in the genome through retroposition (RNA intermediates).

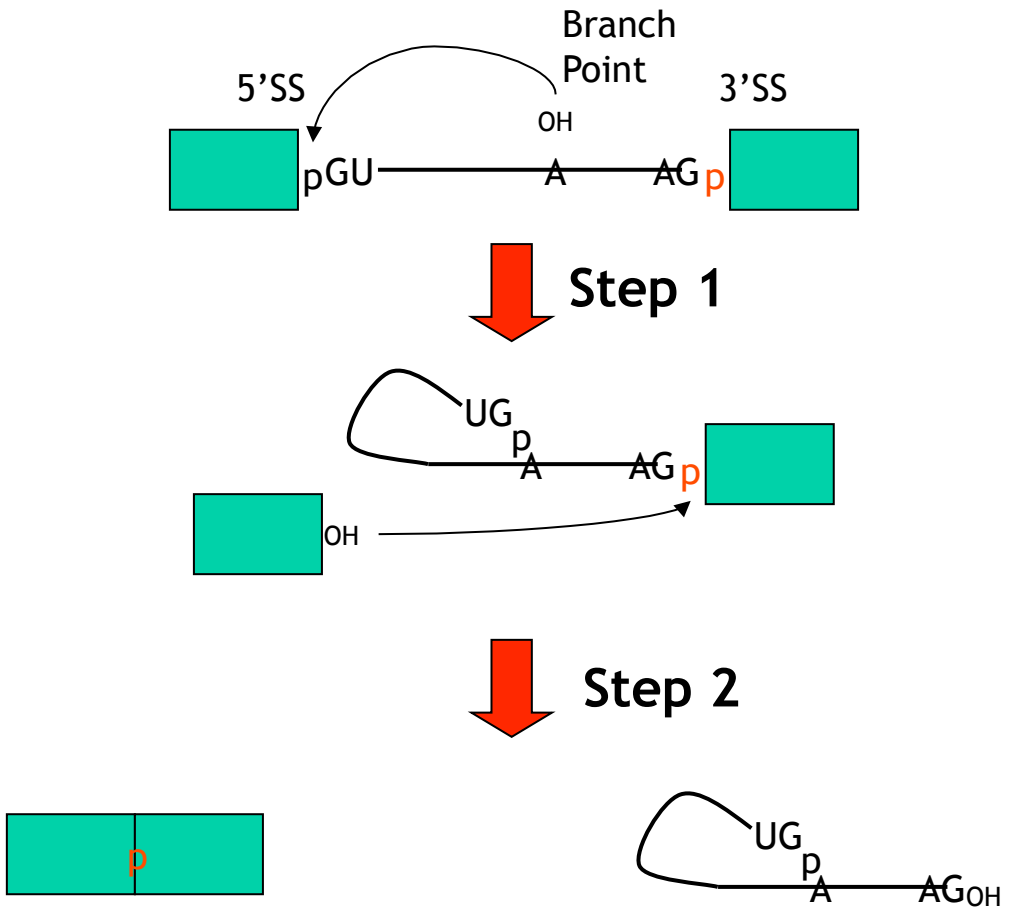


Basic of splicing

Canonical consensus sequences



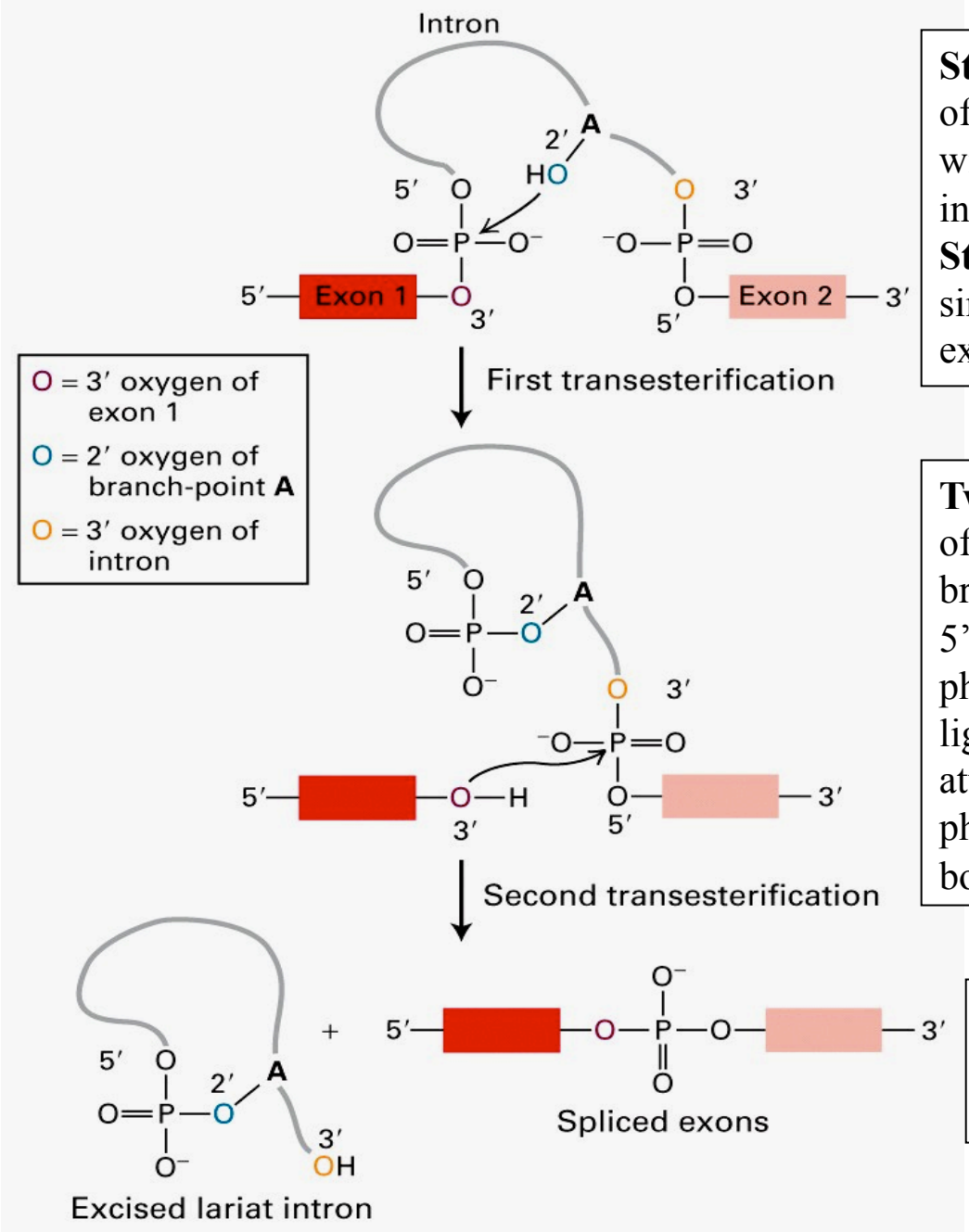
Two steps of pre-mRNA splicing



Spliced mRNA

Intron excised as "lariat"

The splicing reaction proceeds in two steps

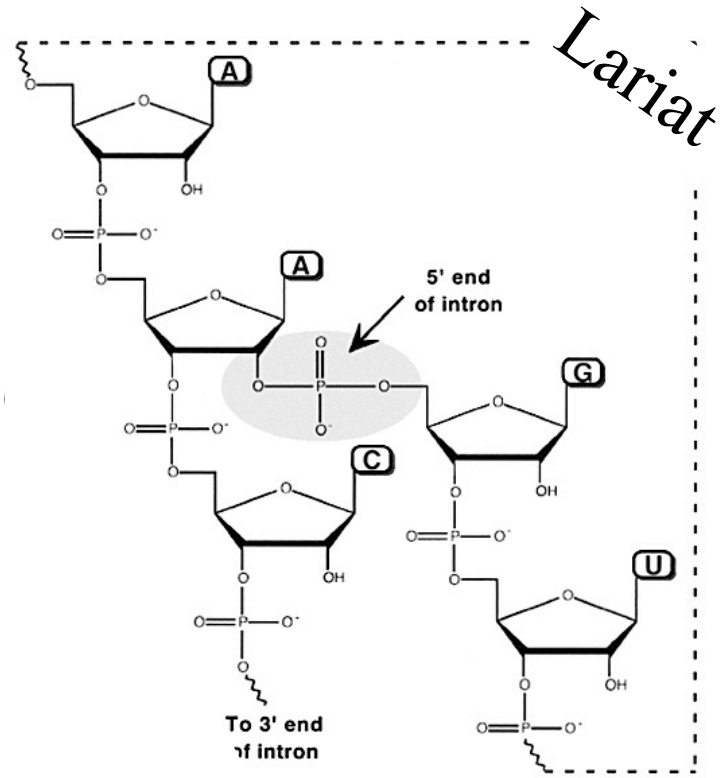


Step 1. Cleavage at the 5’ splice site and joining of the 5’ end of the intron to the branch point A within the intron, producing a lariat-like intermediate.
Step 2. Cleavage at the 3’ splice site and simultaneous ligation of the exons, resulting in excision of the intron as a lariat-like structure.

Two transesterification reactions: (1) The 5’ P of the intron is attacked by the 2’-OH of the branch site Adenosine, causing cleavage of a 3’, 5’-phosphodiester bond and formation of a 2, 5’-phosphodiester bond (not hydrolysis followed by ligation). (2) The newly formed 3’-OH of exon 1 attacks the 5’ P of exon 2, causing cleavage of a phosphodiester bond and formation of a new bond.

Two transesterification reactions: the number of phosphodiester bonds remains unchanged in either reaction.

The Adenosine branch site



legame diesterico 5’-2’ fra il primo nucleotide dell’introne (G) e l’adenosina interna

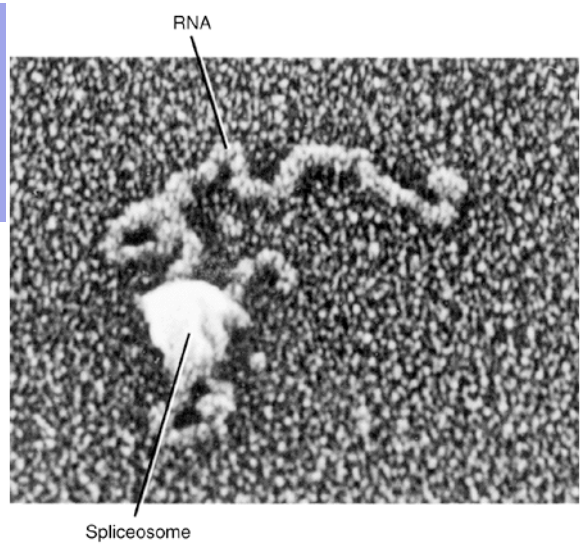
spliceosome

The chemistry is simple, but the *in vivo* process in not

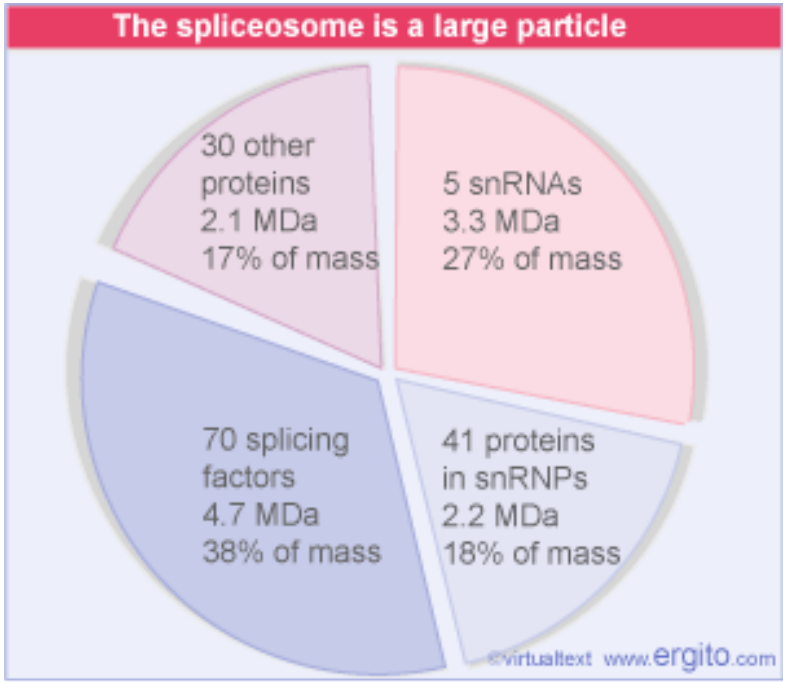
The problem is that you need to accurately define where exons end and introns begin. This requires a huge RNA splicing complex:

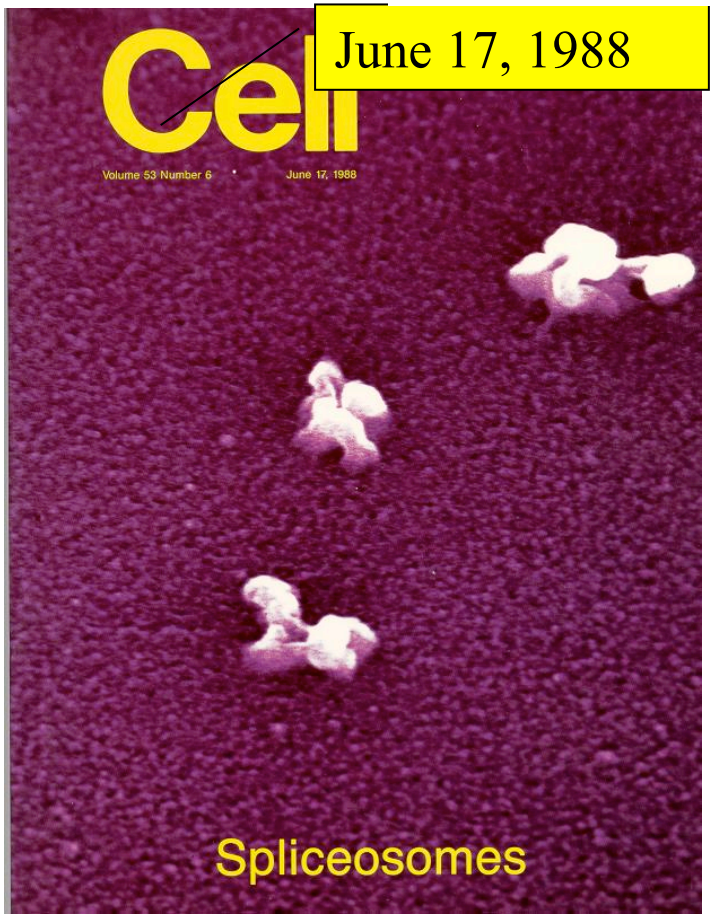
Spliceosome

Splicing Requires RNA and Protein Regulators



Spliceosome - large assembly of 5 **RNA** & over 150 **proteins** that performs pre-mRNA splicing in eukaryotic cells (50-60S)



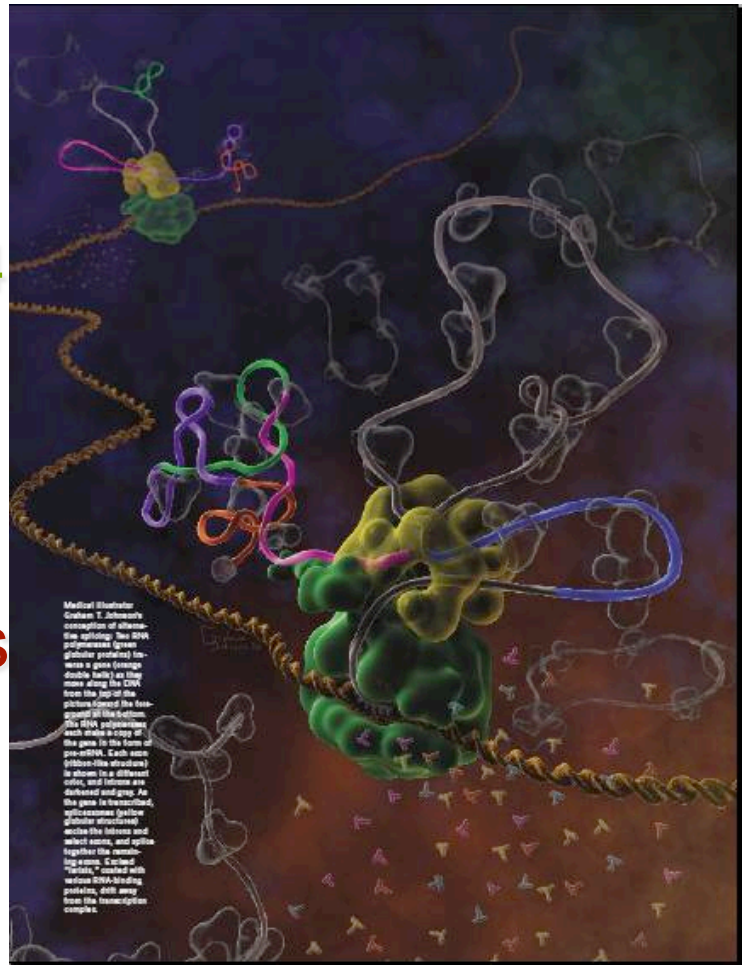


Splicing & the spliceosome

Structure

- 60S dynamic structure – a large complex consisting of ~ 150 proteins
- Five small nuclear RNAs (U1, U2, U4 U5 & U6)
- RNAs assemble with proteins to form snRNPs ("snurps")
- Protein splicing factors

Assembly of spliceosome requires ATP



Green globule: RNA pol

Yellow globule: spliceosome

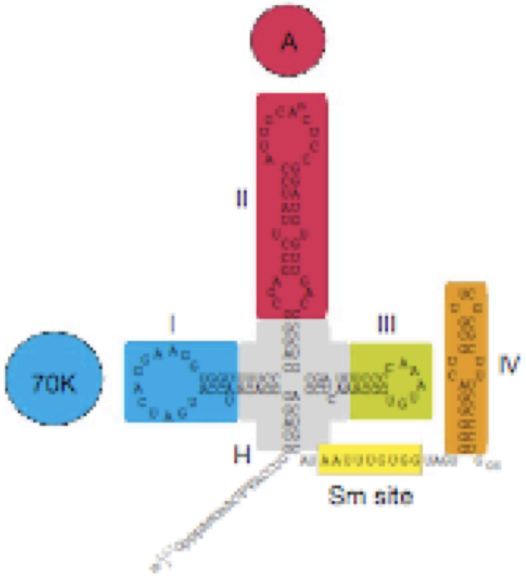
Spliceosome

- Spliceosome contains snRNAs, snRNPs and many other proteins, totally about 300 subunits.
- This makes it the most complicted macromolecular machine known to date.
- But why is spliceosome so extremely complicated if it only catalyzes such a straightforward reaction as an intron deletion? Even more, it seems that some introns are capable to excise themselves without aid of any protein, so why have all those 300 subunits?

- No one knows for sure, but there might be at least 4 reasons:
- 1. Defective mRNAs cause a lot of problems for cells, so some subunits might assure correct splicing and error correction
- 2. Splicing is coupled to nuclear transport, this requires accessory proteins
- 3. Splicing is coupled to transcription and this might require more additional accessory proteins
- 4. Many genes can be spliced in several alternative ways, which also might require additional factors

- 5 small nuclear RNAs (**U1, U2, U4, U5 e U6**)
- 100-300 nt
- bind to proteins to form snRNP (*small nuclear ribonucleoparticles*).

A typical small nuclear RNA : U1 snRNP

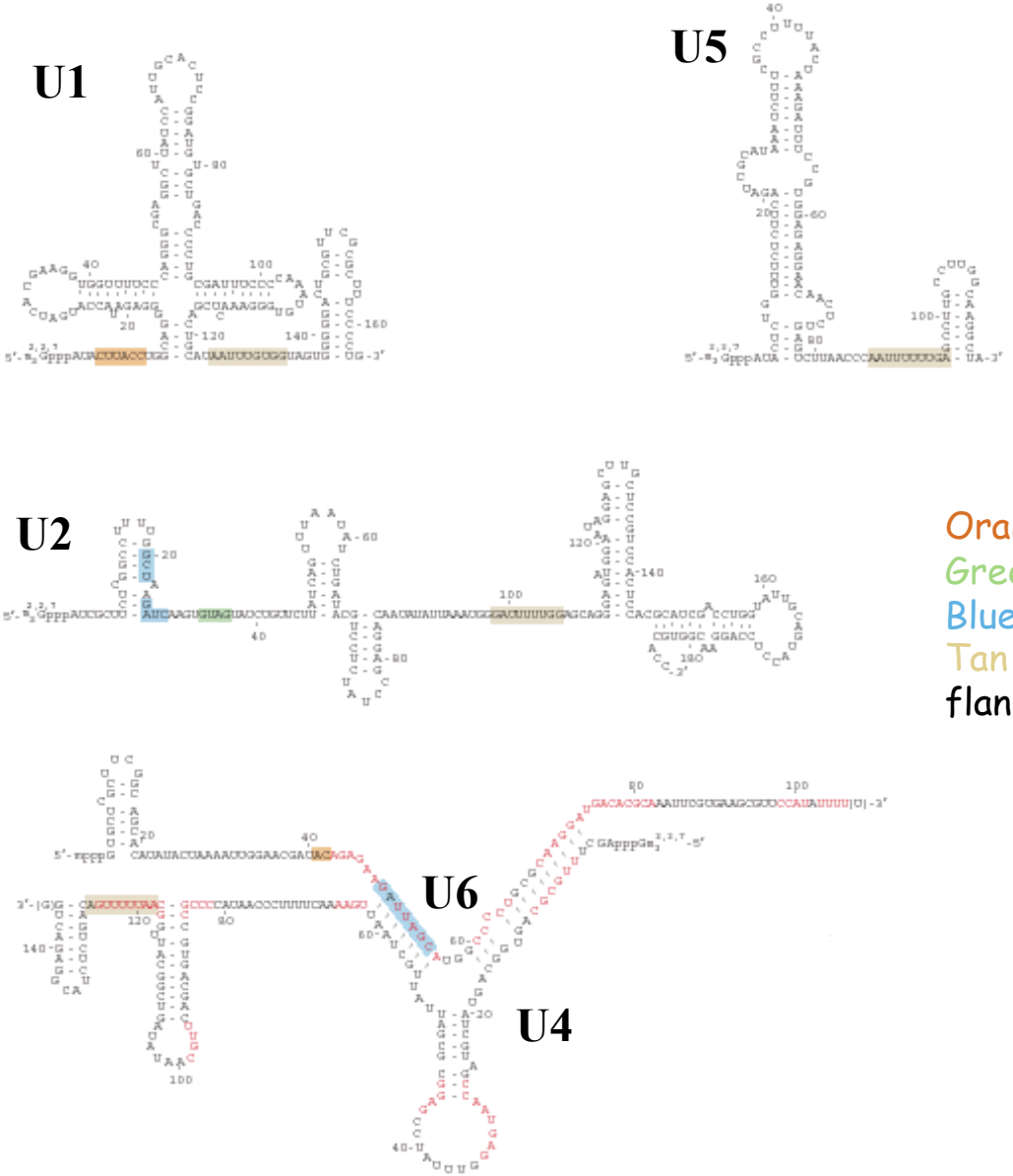


Mammalian U1-snRNP consists of ten different proteins and a 165-nucleotide RNA molecule. Seven of the proteins, the Sm proteins (B, D1, D2, D3, E, F and G), are common to all U snRNPs, whereas the other three (U1-70K, U1-A and U1-C) are specific to U1 snRNP.

Contains conserved sequence complementary to 5' splice site of nuclear mRNA introns

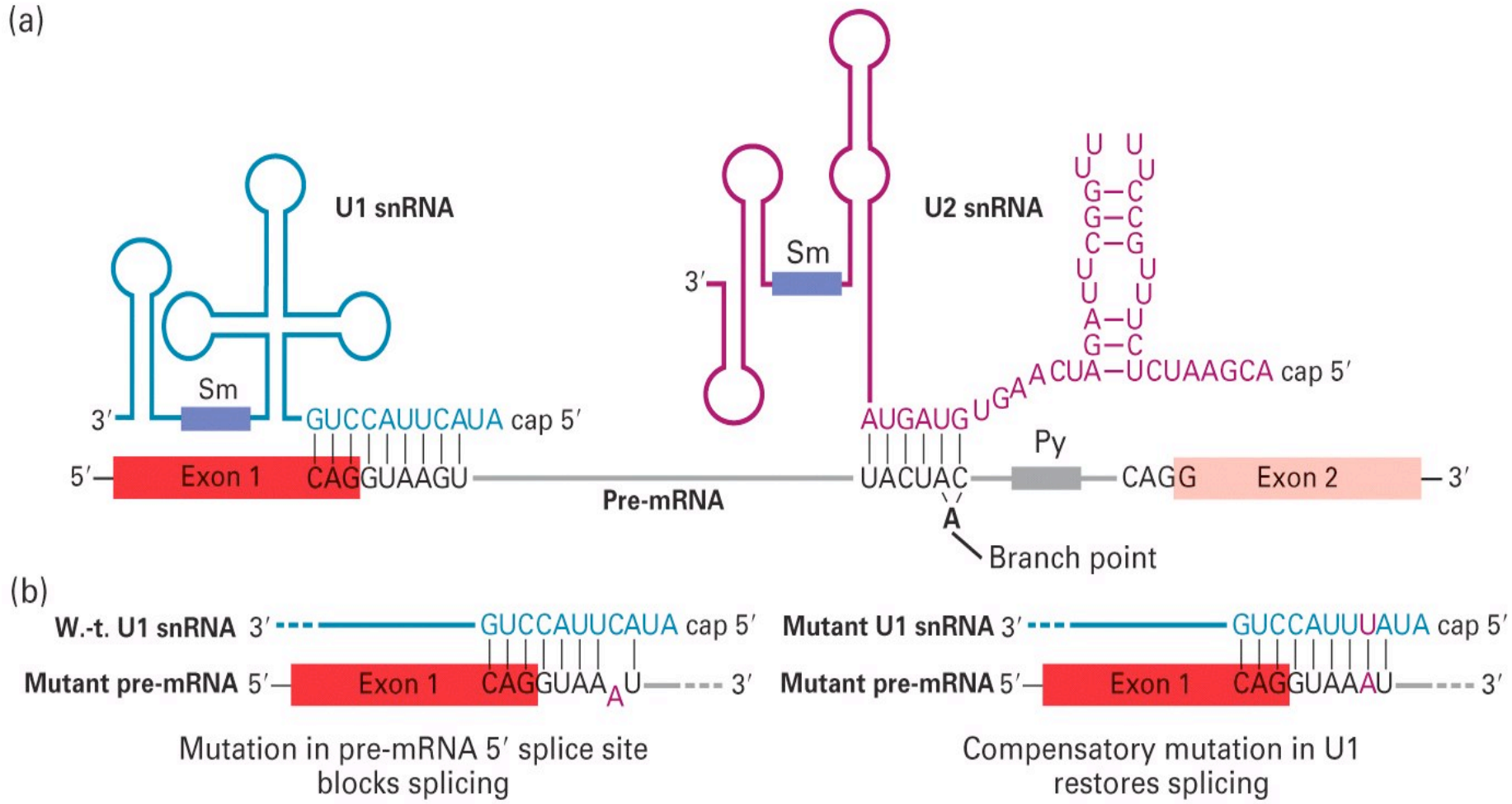
Contains pseudouridine (ψ)

Secondary structure of snRNAs

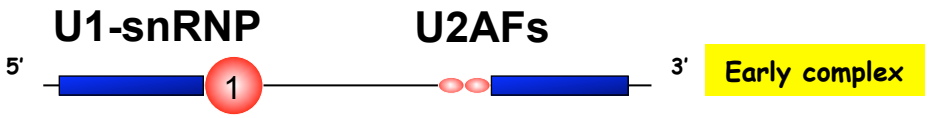


Orange - interaction with 5′ splice site
Green – Interaction with branch site
Blue - interaction between U2 and U6
Tan - Sm-binding site ($PuAU_{4\text{-}6}GPu$) flanked by two stem-loop structures

Base pairing between pre-mRNA, U1 snRNA, and U2 snRNA early in the splicing cycle and experimental demonstration that the base pairing between U1 and the 5’ splice site in pre-mRNA is important



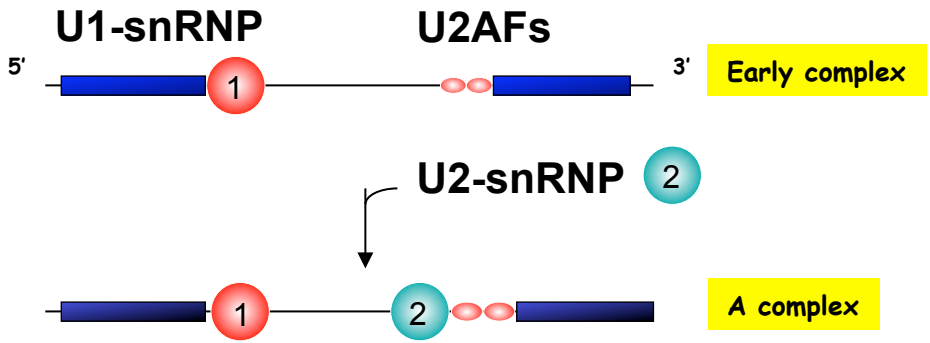
Model of spliceosome-mediated splicing of pre-mRNA



L'U1-snRNP si lega al sito di splicing al 5'(in parte mediante appaiamento RNA-RNA), mentre $U2AF^{35}$ e $U2AF^{65}$ si legano al tratto polipirimidinico

Model of spliceosome-mediated splicing of pre-mRNA

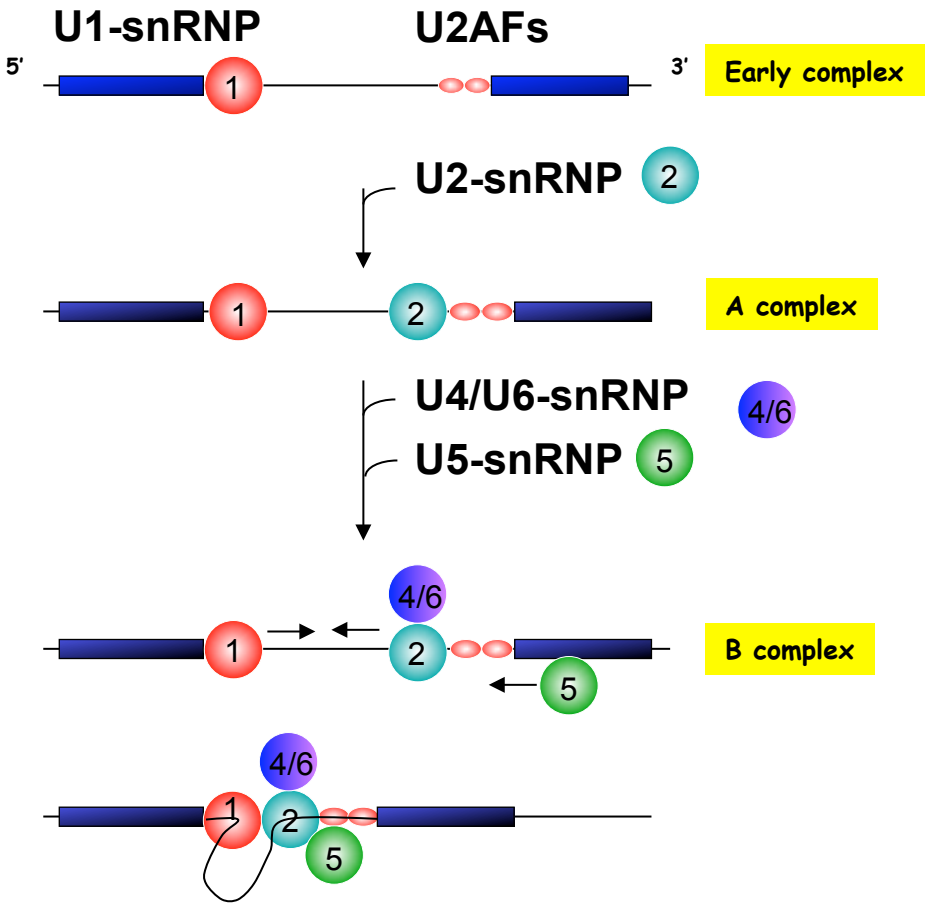
L'U2-snRNP si lega al sito di ramificazione formando il complesso A



Model of spliceosome-mediated splicing of pre-mRNA

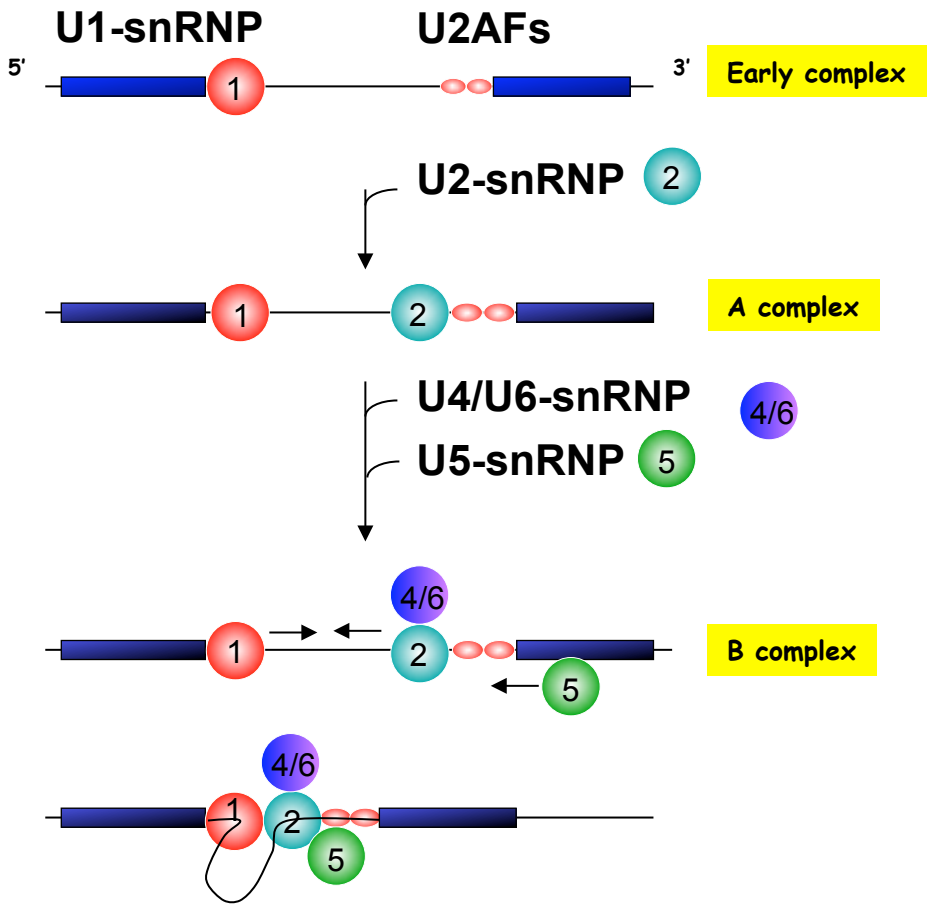
Il legame di U4/U6-snRNP e U5-snRNP al complesso A genera il complesso B

U4/U6-snRNP si lega a U2-snRNP mentre U5-snRNP si lega inizialmente all'esone a valle e poi migra verso il 3' dell'introne



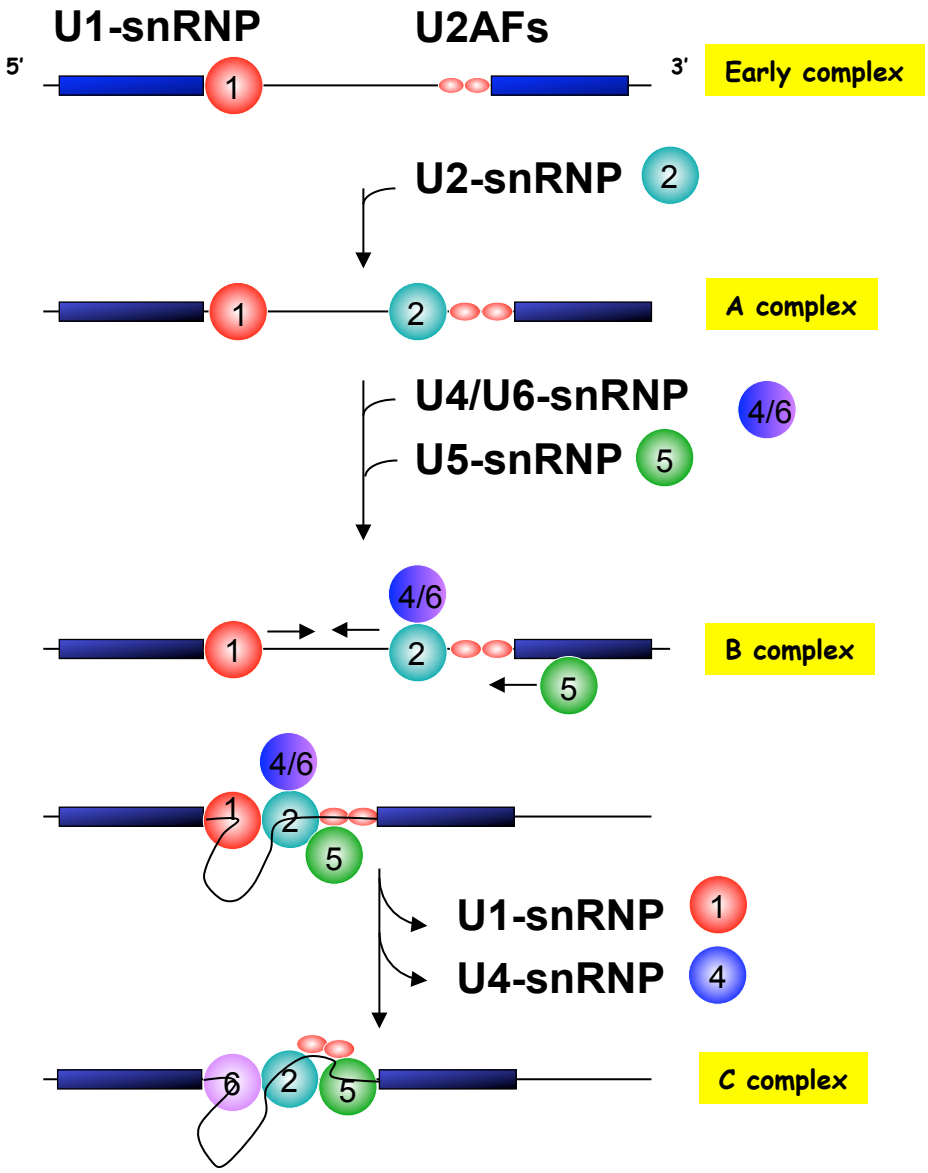
Model of spliceosome-mediated splicing of pre-mRNA

U1-snRNP e U2-snRNP si associano portando il sito di splicing al 5’ in vicinanza con il sito di ramificazione U1-snRNP si dissocia e viene sostituito da U6-snRNP



Model of spliceosome-mediated splicing of pre-mRNA

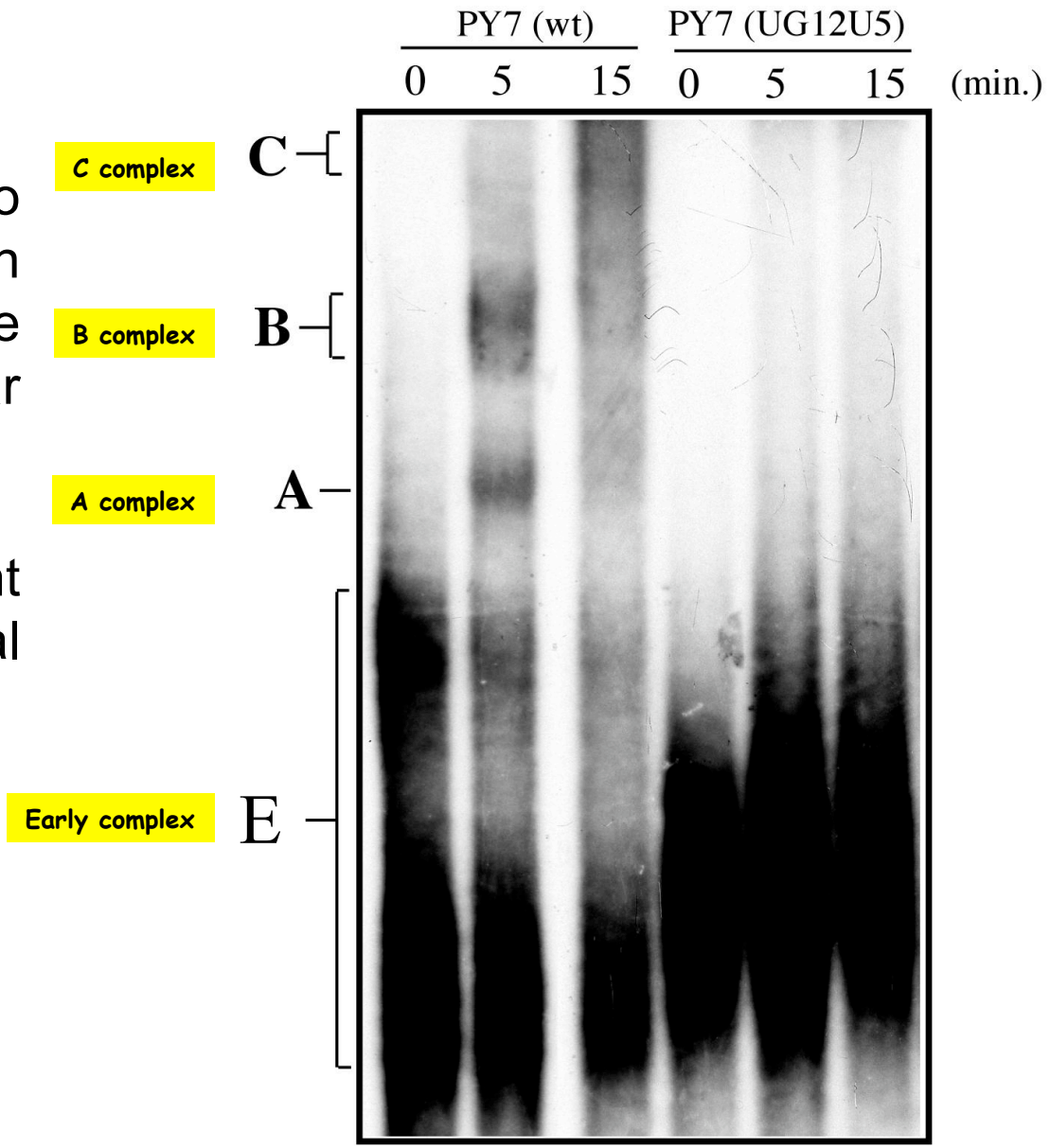
U4-snRNP si distacca dal complesso mentre U5-snRNP si riposiziona sul sito di splicing 3'. A questo punto i due siti di splicing e il punto di ramificazione sono in stretta vicinanza e U6-snRNP catalizza le due reazioni di transesterificazione



Spliceosomal complex formation

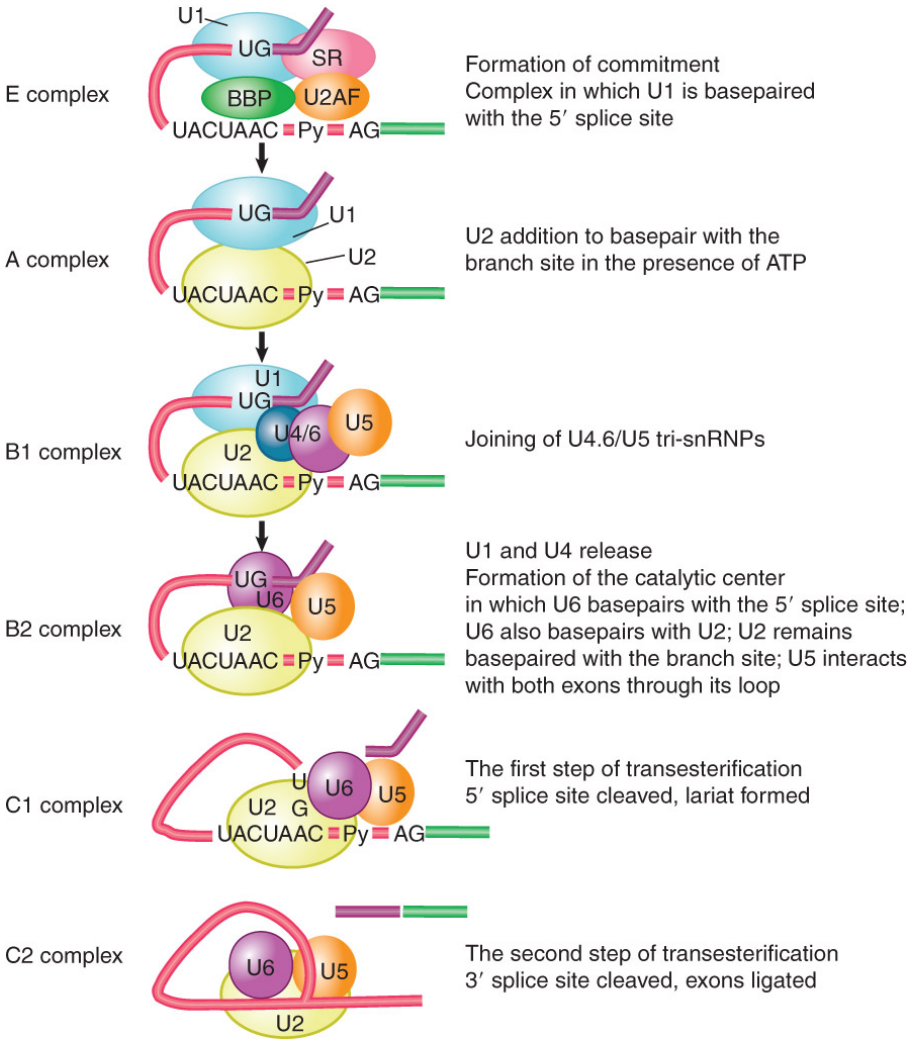
The same reactions for in vitro splicing are loaded on a non denaturing gel to resolve the size of the molecular complexes formed.

Each complex contains different sets of snRNPs plus several other splicing factors.



The Spliceosome Assembly Pathway

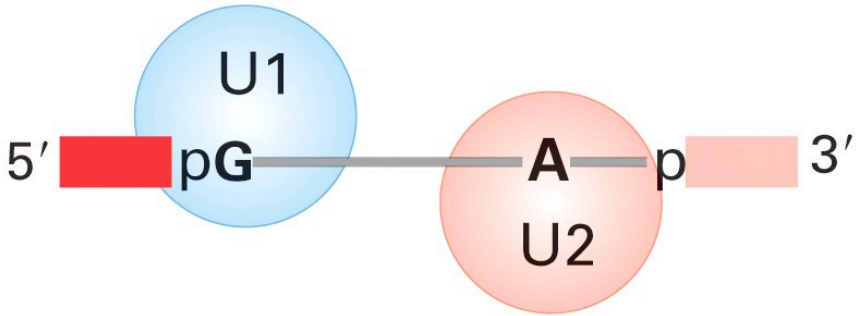
Splicing reaction proceeds through discrete stages

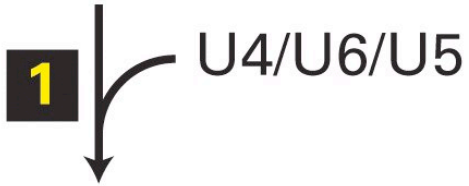


- U4 dissociates from U6 snRNP to allow U6 snRNA to pair with U2 snRNA to form the catalytic center for splicing.
- Both transesterification reactions take place in the activated spliceosome (the C complex).
- The splicing reaction is reversible at all steps.

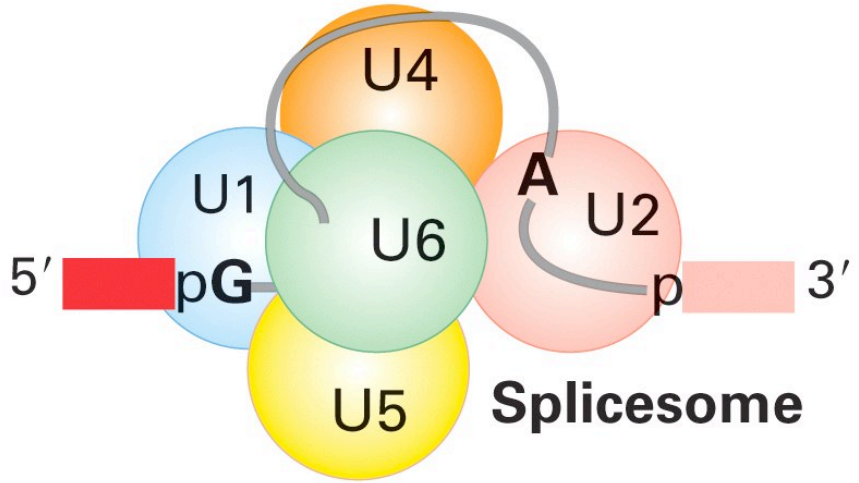
The essential steps in splicing

Binding of U1 and U2 snRNPs

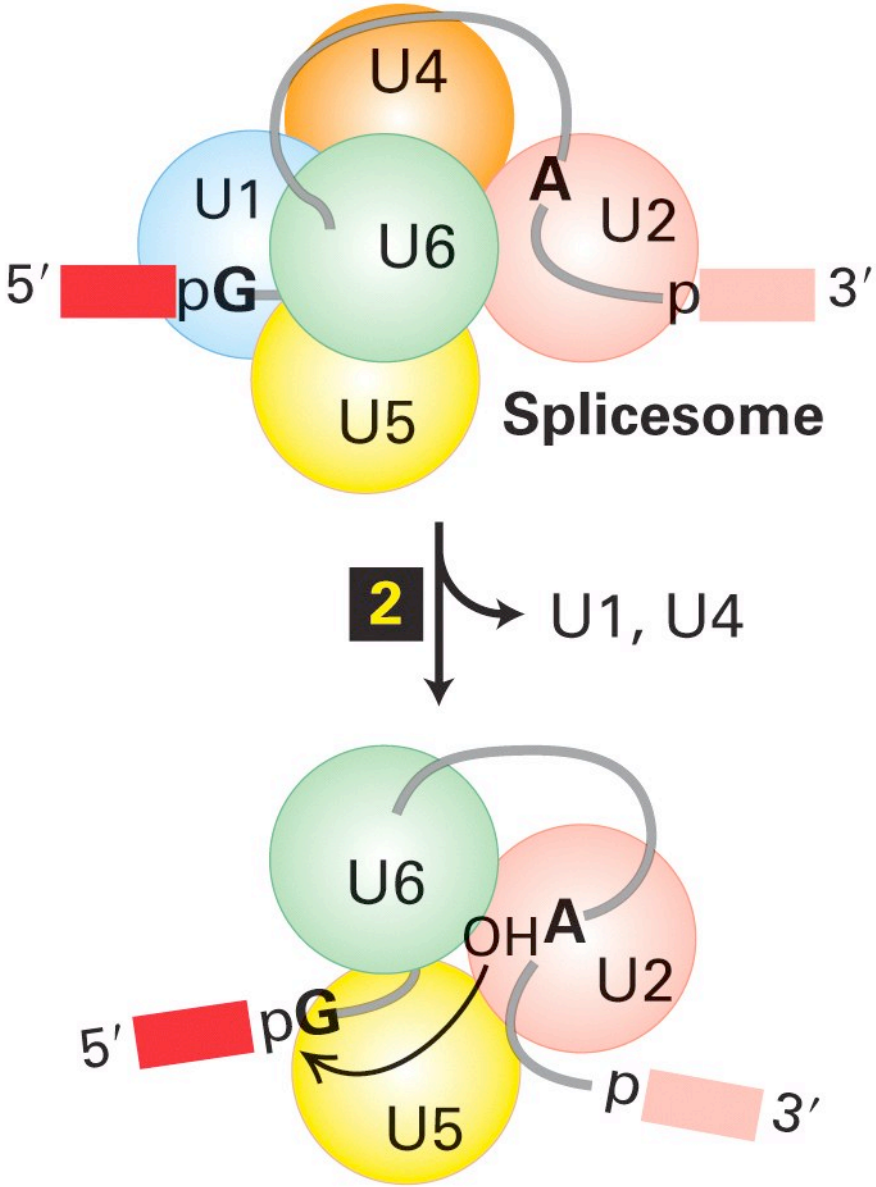




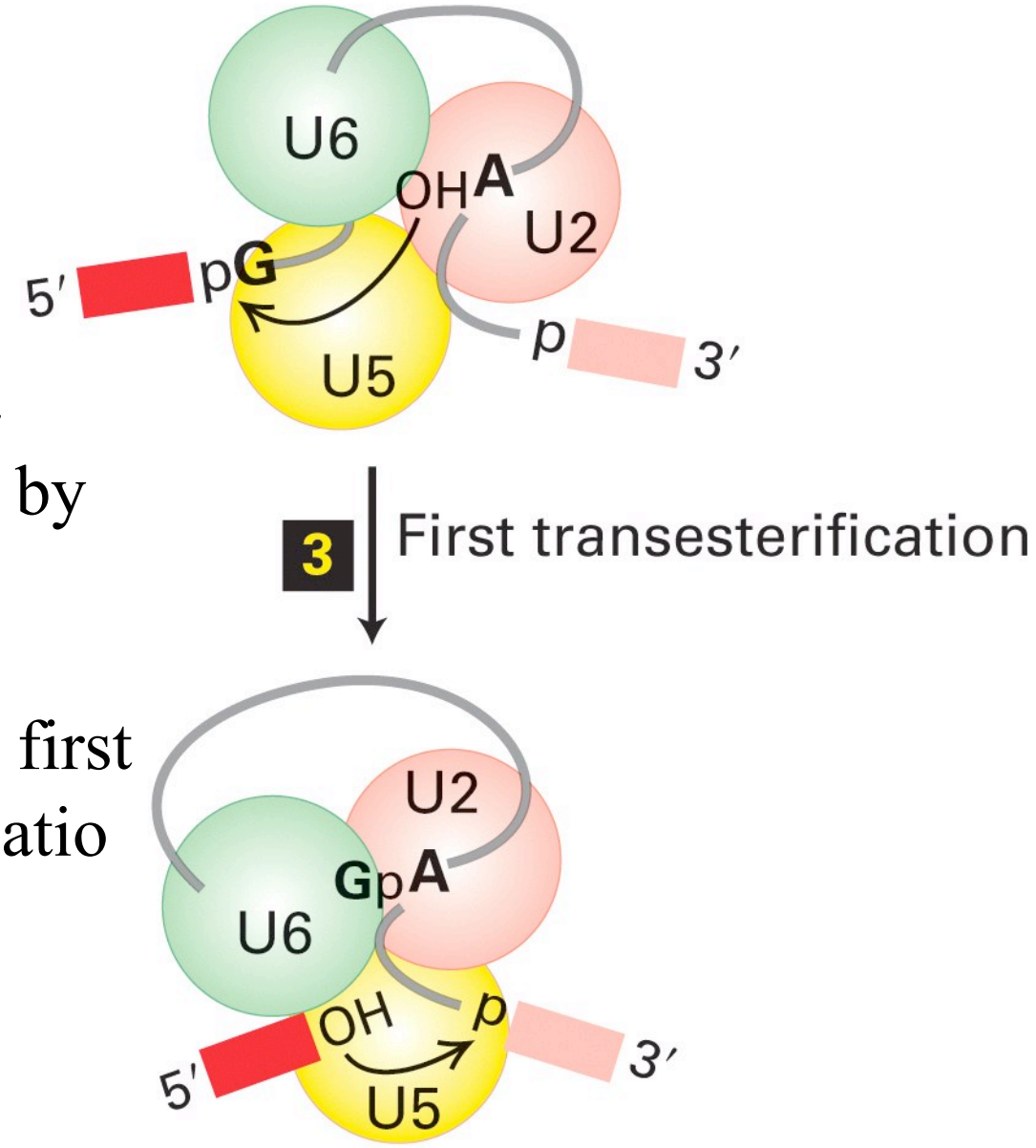
Binding of U4, U5 and U6 snRNPs



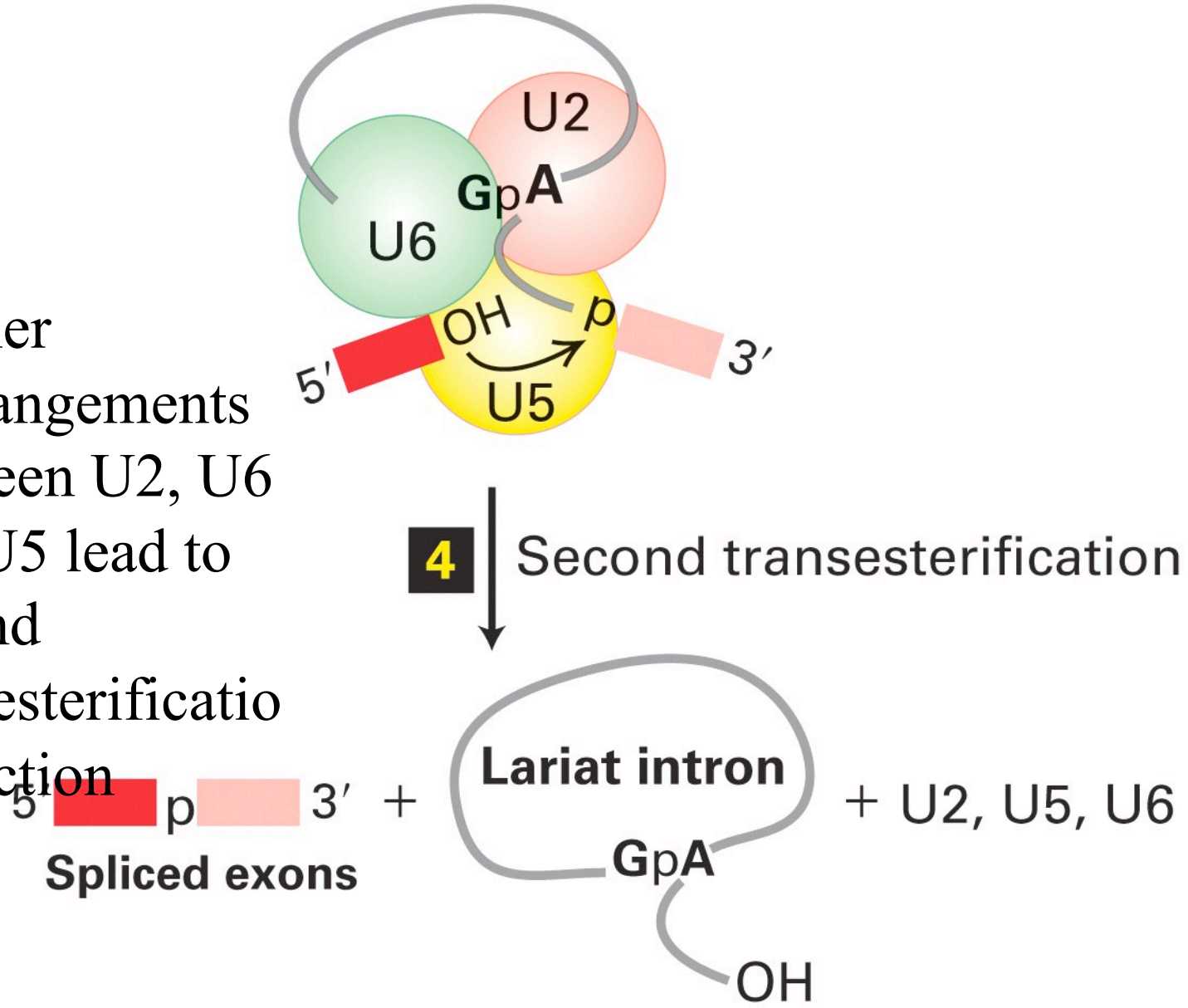
Rearrangement of base-pair interactions between snRNAs, release of U1 and U4 snRNPs

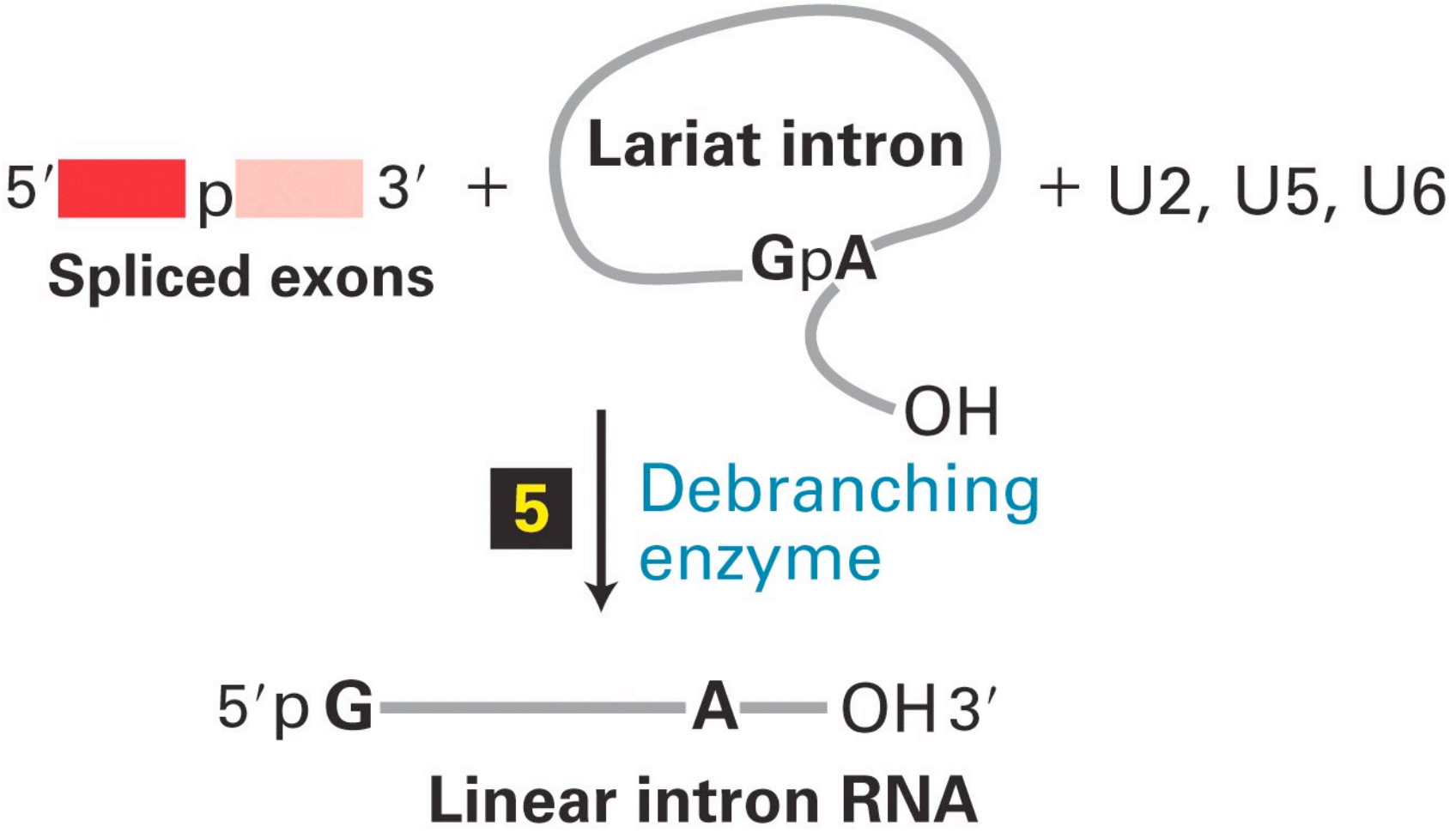


The catalytic core, formed by U2 and U6 snRNPs catalyzes the first transesterificatio n reaction



Further rearrangements between U2, U6 and U5 lead to second transesterificatio n reaction

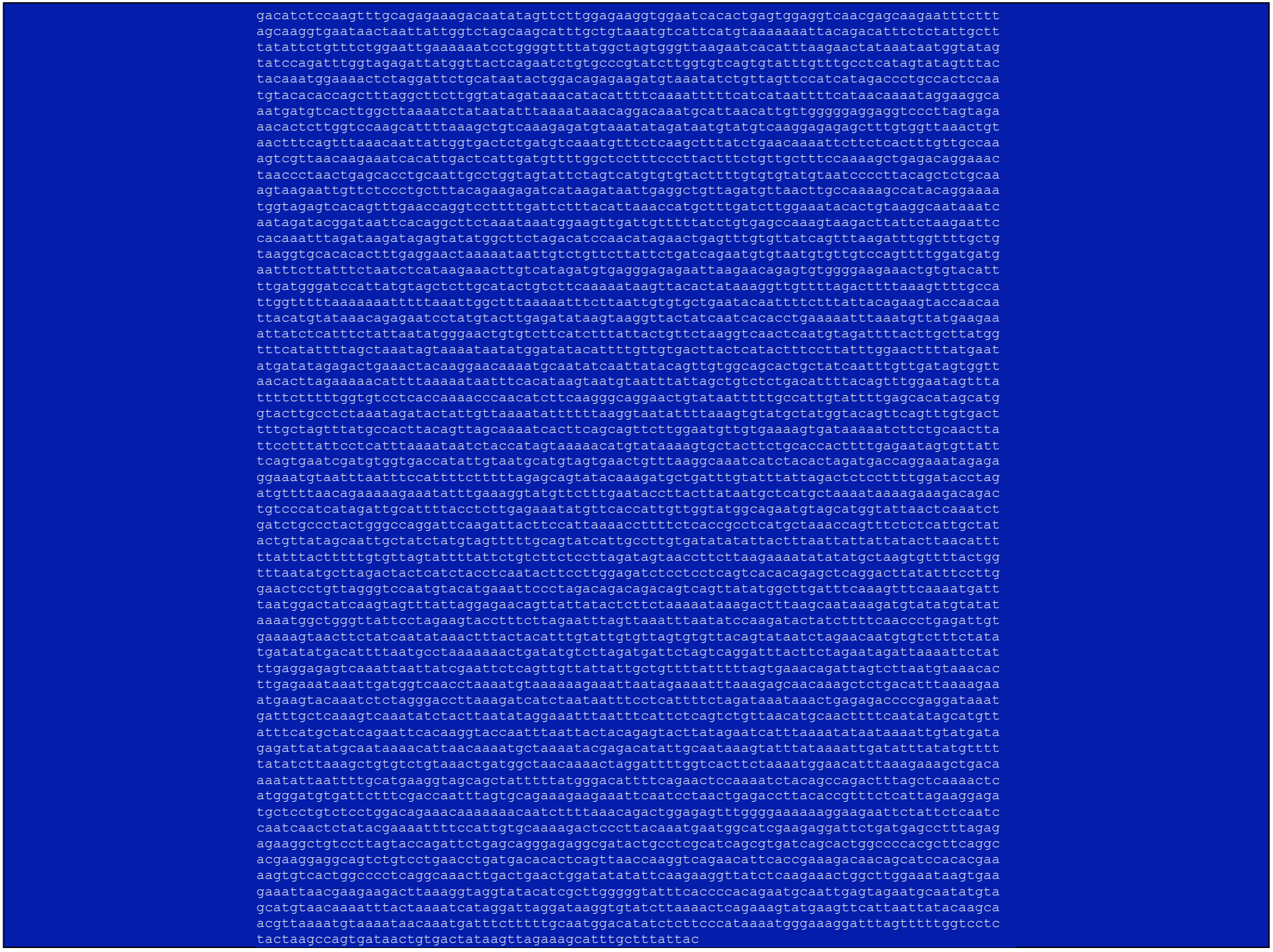


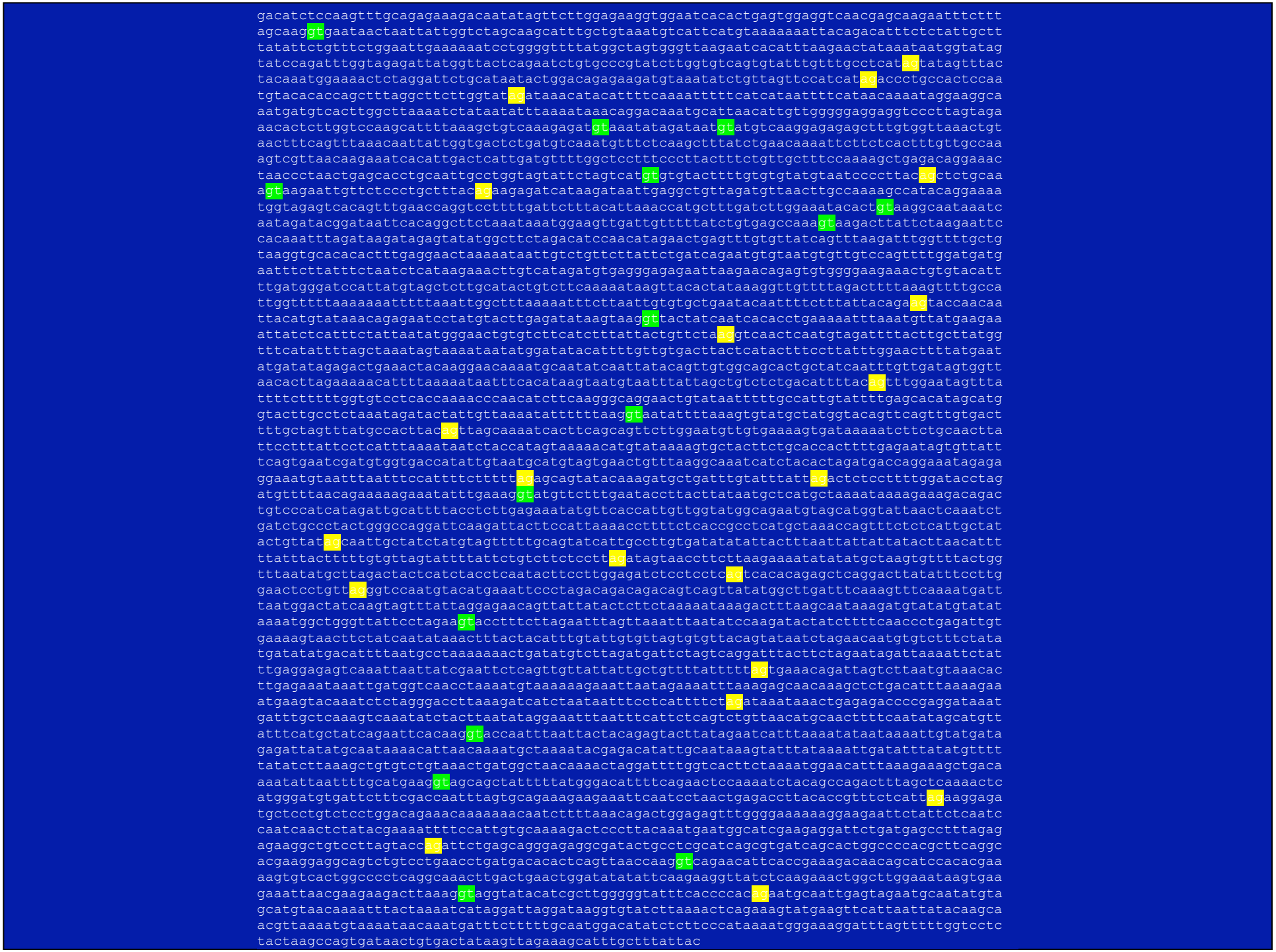


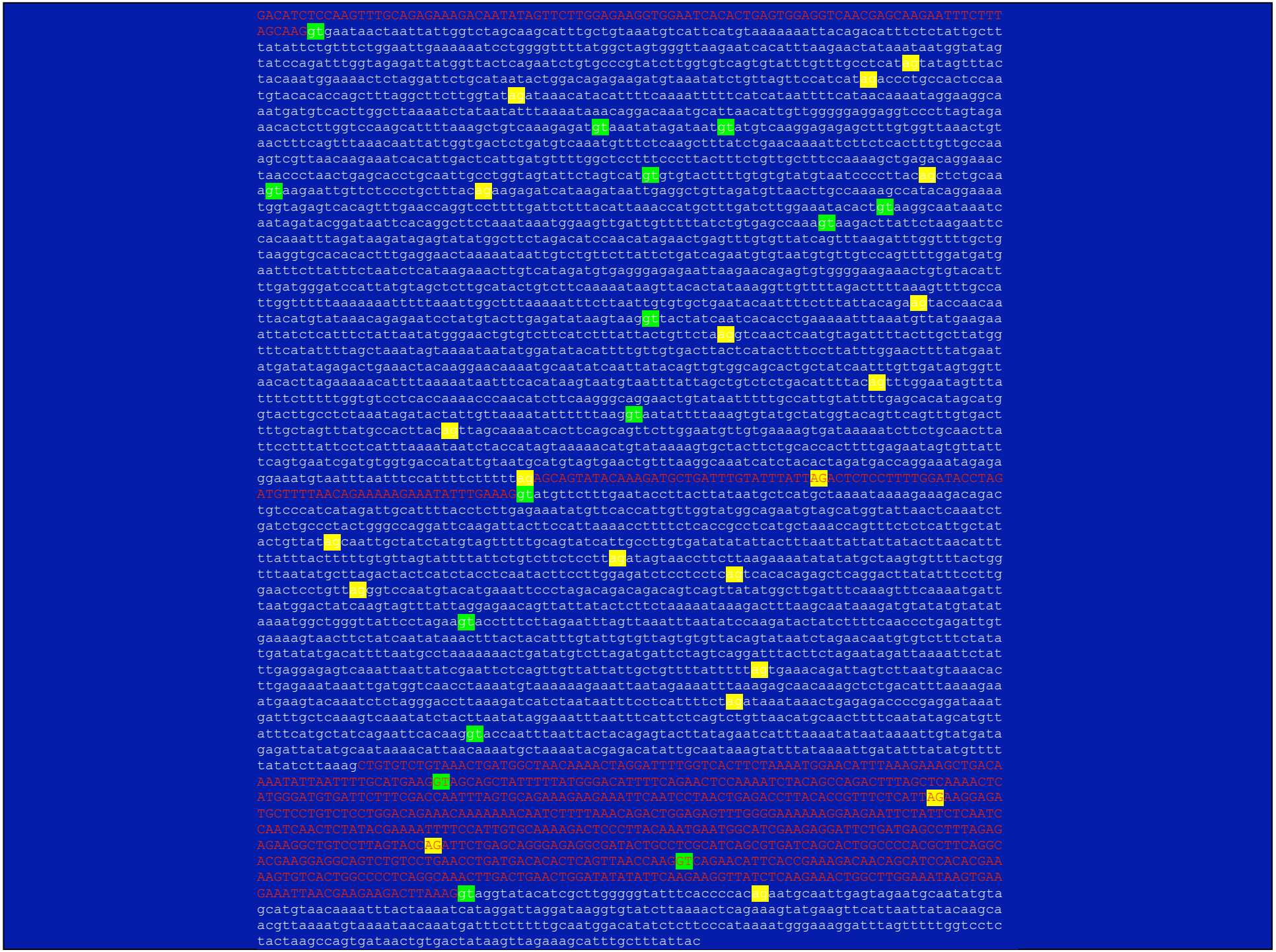
The spliced lariat is linearized by debranching enzyme and further degraded in exosomes

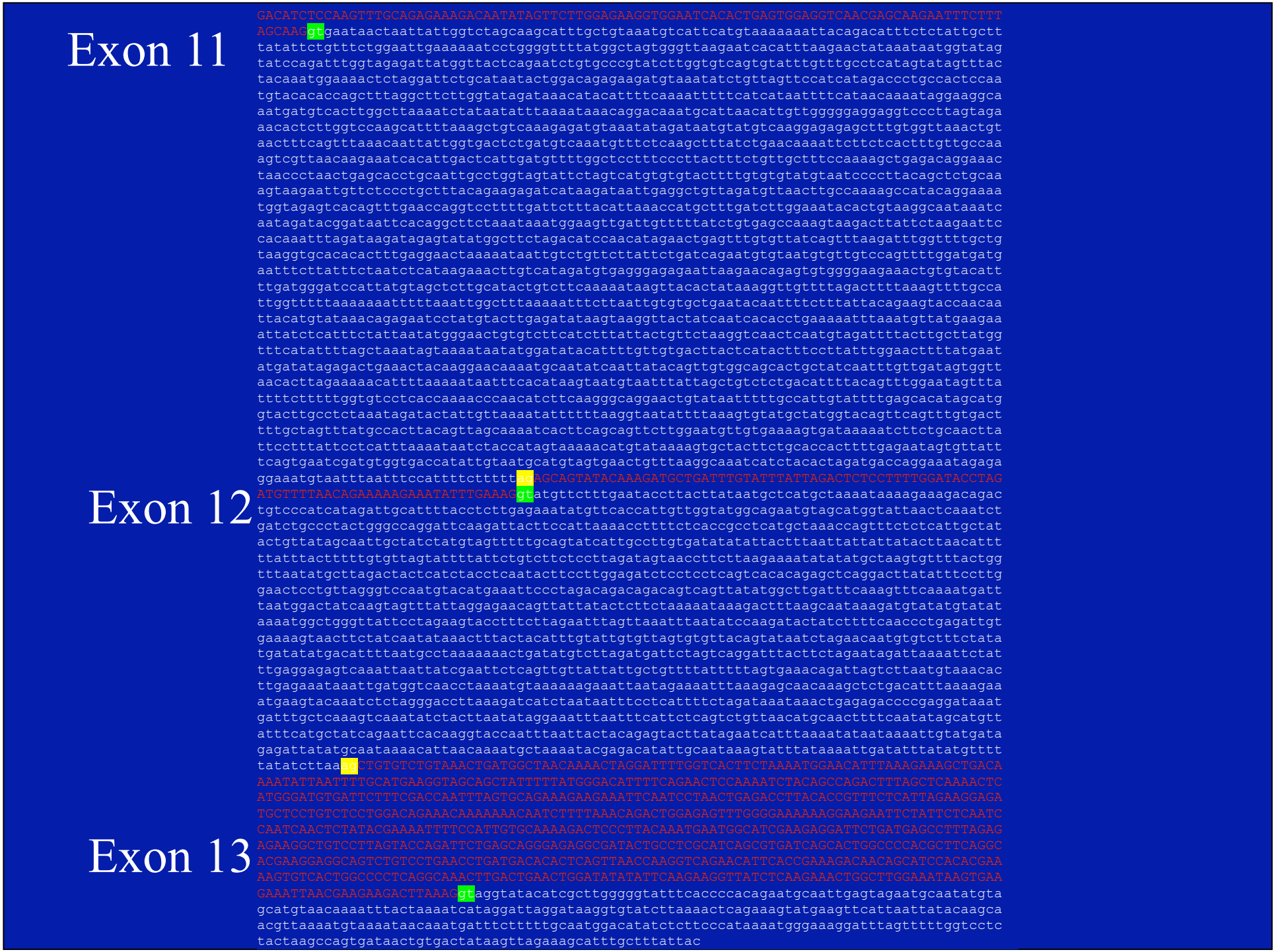
Not all introns are completely degraded. Some end up as functional RNAs, different from mRNA (mi RNA sno RNA)

Non canonical splicing regulatory elements

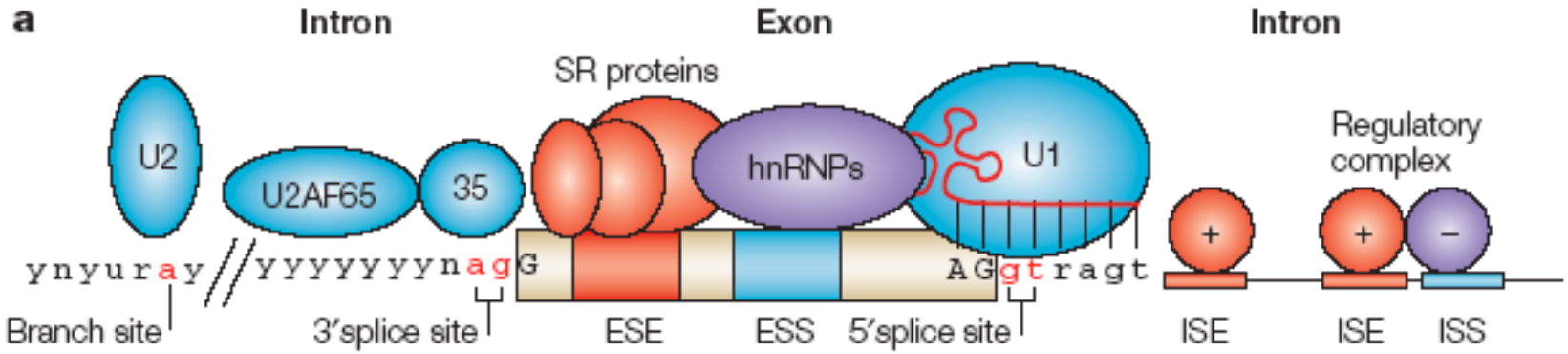

gacatctccaagtttgcagagaaagacaatatagttcttggagaaggtggaatcacactgagtggaggtcaacgagcaagaatttctttagcaaggtgaataactaattattggtctagcaagcatttgctgtaaatgtcattcatgtaaaaaaattacagacatttctctattgcttatattctgtttctggaattgaaaaaatcctggggttttatggctagtgggttaagaatcacatttaagaactataaataatggtatagtatccagatttggtagagattatggttactcagaatctgtgcccgtatcttggtgtcagtgtatttgtttgcctcatagtatagtttactacaaatggaaaactctaggattctgcataatactggacagagaagatgtaaatatctgttagttccatcatagaccctgccactccaatgtacacaccagctttaggcttcttggtatagataaacatacattttcaaaatttttcatcataatttttcataacaaaataggaaggcaaatgatgtcacttggcttaaaatctataatatttaaaataaacaggacaaatgcattaacattgttgggggaggaggtcccttagtagaaacactcttggtccaagcattttaaagctgtcaaagagatgtaaatatagataatgtatgtcaaggagagagctttgtggttaaactgtaactttcagtttaaacaattattggtgactctgatgtcaaatgtttctcaagctttatctgaacaaaattcttctcactttgttgccaaagtcgttaacaagaaatcacattgactcattgatgttttggctcctttcccttactttctgttgctttccaaaagctgagacaggaaacttaaccctaactgagcacctgcaattgcctggtagtattctagtcatgtgtgtacttttgtgtgtatgtaatccccttacagctctgcaaagtaagaattgttctccctgctttacagaagagatcataagataattgaggctgttagatgttaacttgccaaaagccatacaggaaaatggtagagtcacagtttgaaccaggtccttttgattctttacattaaaccatgctttgatcttggaaatacactgtaaggcaataaatcaatagatacggataattcacaggcttctaaataaatggaagttgattgtttttatctgtgagccaaagtaagacttattctaagaattccacaaatttagataagatagagtatatggcttctagacatccaacatagaactgagtttgtgttatcagtttaagatttggttttgctgtaaggtgcacacactttgaggaactaaaaataattgtctgttcttattctgatcagaatgtgtaatgtgttgtccagttttggatgatgaatttcttatttctaatctcataagaaacttgtcatagatgtgagggagagaattaagaacagagtgtggggaagaaactgtgtacattttgatgggatccattatgtagctcttgcatactgtcttcaaaaataagttacactataaaggttgttttagacttttaaagttttgccattggttttaaaaaaatttttaaattggctttaaaaatttcttaattgtgtgctgaatacaattttctttattacagaagtaccaacaattacatgtataaacagagaatcctatgtacttgagatataagtaaggttactatcaatcacacctgaaaaatttaaatgttatgaagaaattatctcatttctattaatatgggaactgtgtcttcatctttattactgttctaaggtcaactcaatgtagattttacttgcttatggtttcatattttagctaaatagtaaaataatatggatatacattttgttgtgacttactcatactttccttatttggaacttttatgaatatgatatagagactgaaactacaaggaacaaaatgcaatatcaattatacagttgtggcagcactgctatcaatttgttgatagtggttaacacttagaaaaacatttaaaaataatttcacataagtaatgtaatttattagctgtctctgacattttacagtttggaatagttttattttcttttggtgtcctcaccaaaacccaacatcttcaagggcaggaactgtataatttttgccattgtattttgagcacatagcatggtacttgcctctaaatagatactattgttaaaatatttttttaaggtaatattttaaagtgtatgctatggtacagttcagtttgtgacttttgctagtttatgccacttacagttagcaaaatcacttcagcagttcttggaatgttgtgaaaagtgataaaaatcttctgcaacttattccttttattcctcatttaaaataatctaccatagtaaaaacatgtataaaagtgctacttctgcaccacttttgagaatagtgttattttcagtgaatcgatgtggtgaccatattgtaatgcatgtagtgaactgtttaaggcaaatcatctacactagatgaccaggaaatagagaggaaatgtaatttaatttccatttttcttttttagagcagtatacaaagatgctgatttgtatttattagactctccttttggatacctagatgttttaacagaaaaagaaatatttgaaaggtatgttcttttgaatacctttacttataatgctcatgctaaaataaaagaaagacagactgtcccatcatagattgcattttacctcttgagaaatatgttcaccattgttggtatggcagaatgtagcatggtattaactcaaatctgatctgccctactgggccaggattcaagattacttccattaaaaccttttctcaccgcctcatgctaaaccagtttctctcattgctatactgttatagcaattgctatctatgtagtttttgcagtatcattgccttgtgatatatattactttaattattattatacttaacatttttatttacttttgtgttagtattttattctgtcttctccttagatagtaaccttcttaagaaaatatatatgctaagtgttttactggtttaatatgcttagactactcatctacctcaatacttccttggagatctcctcctcagtcacacagagctcaggacttatatttccttggaactcctgttagggtccaatgtacatgaaattccctagacagacagacagtcagttatatggcttgatttcaaagtttcaaaatgatttaatggactatcaagtagtttattaggagaacagttattatactcttctaaaaataaagacttaagcaataaagatgtatatgtatataaaatggctgggttattcctagaagtacctttcttagaatttagttaaatttaatatccaagatactatcttttcaaccctgagattgtgaaaagtaacttctatcaatataaactttactacatttgtattgtgttagtgtgttacagtataatctagaacaatgtgtctttctatatgatatatgacatttttaatgcctaaaaaaactgatatgtcttagatgattctagtcaggatttacttctagaatagattaaaattctattgaggagagtcaaattaattatcgaattctcagttgttattattgctgttttatttttagtgaaacagattagtcttaatgtaaacactgagaaataaattgatggtcaacctaaaatgtaaaaaagaaattaatagaaaatttaaagagcaacaaagctctgacatttaaaagaaatgaagtacaaatctctagggaccttaaagatcatctaataatttcctcattttctagataaataaactgagagaccccgaggataaatgatttgctcaaagtcaaatatctacttaatataggaaatttaatttcattctcagtctgttaacatgcaacttttcaatatagcatgttatttcatgctatcagaattcacaaggtaccaatttaattactacagagtacttatagaatcatttaaaatataataaaattgtatgatagagattatatgcaataaaacattaacaaaatgctaaaatacgagacatattgcaataaagtatttataaaattgatatttatatgttttatatcttaaagctgtgtctgtaaactgatggctaacaaaactaggattttggtcacttctaaaatggaacatttaaagaaagctgacaaaatattaattttgcatgaaggtagcagctatttttatgggacattttcagaactccaaaatctacagccagactttagctcaaaactcatgggatgtgattctttcgaccaatttagtgcagaaagaagaaattcaatcctaactgagaccttacaccgtttctcattagaaggagatgctcctgtctcctggacagaaacaaaaaaacaatcttttaaacagactggagagtttgggggaaaaaaggaagaattctattctcaatccaatcaactctatacgaaaattttccattgtgcaaaagactcccttacaaatgaatggcatcgaagaggattctgatgagcctttagagagaaggctgtccttagtaccagattctgagcagggagaggcgatactgcctcgcatcagcgtgatcagcactggccccacgcttcaggcacgaaggaggcagtctgtcctgaacctgatgacacactcagttaaccaaggtcagaacattcaccgaaagacaacagcatccacacgaaaagtgtcactggcccctcaggcaaacttgactgaactggatatatattcaagaaggttatctcaagaaactggcttggaaataagtgaagaaattaacgaagaagacttaaaggtaggtatacatcgcttgggggtatttcaccccacagaatgcaattgagtagaatgcaatatgtagcatgtaacaaaatttactaaaatcataggattaggataaggtgtatcttaaaactcagaaagtatgaagttcattaattatacaagcaacgttaaaatgtaaaataacaaatgatttcttttgcaatggacatatctcttcccataaaatgggaaaggatttagtttttggtcctctactaagccagtgataactgtgactataagttagaaagcatttgctttattac


gacatctccaagtttgcagagaaagacaatatagttcttggagaaggtggaatcacactgagtggaggtcaacgagcaagaatttctttagcaaggtgaataactaattattggtctagcaagcatttgctgtaaatgtcattcatgtaaaaaaattacagacatttctctattgcttatattctgtttctggaattgaaaaaatcctggggttttatggctagtgggttaagaatcacatttaagaactataaataatggtatagtatccagatttggtagagattatggttactcagaatctgtgcccgtatcttggtgtcagtgtatttgtttgcctcatagtatagtttactacaaatggaaaactctaggattctgcataatactggacagagaagatgtaaatatctgttagttccatcatagaccctgccactccaatgtacacaccagctttaggcttcttggtatagataaacatacattttcaaaatttttcatcataatttttcataacaaaataggaaggcaaatgatgtcacttggcttaaaatctataatatttaaaataaacaggacaaatgcattaacattgttgggggaggaggtcccttagtagaaacactcttggtccaagcattttaaagctgtcaaagagatgtaaatatagataatgtatgtcaaggagagagctttgtggttaaactgtaactttcagtttaaacaattattggtgactctgatgtcaaatgtttctcaagctttatctgaacaaaattcttctcactttgttgccaaagtcgttaacaagaaatcacattgactcattgatgttttggctcctttcccttactttctgttgctttccaaaagctgagacaggaaacttaaccctaactgagcacctgcaattgcctggtagtattctagtcatgtgtgtacttttgtgtgtatgtaatccccttacagctctgcaaagtaagaattgttctccctgctttacagaagagatcataagataattgaggctgttagatgttaacttgccaaaagccatacaggaaaatggtagagtcacagtttgaaccaggtccttttgattctttacattaaaccatgctttgatcttggaaatacactgtaaggcaataaatcaatagatacggataattcacaggcttctaaataaatggaagttgattgtttttatctgtgagccaaagtcaagacttattctaagaattccacaaatttagataagatagagtatatggcttctagacatccaacatagaactgagtttgtgttatcagtttaagatttggttttgctgtaaggtgcacacactttgaggaactaaaaataattgtctgttcttattctgatcagaatgtgtaatgtgttgtccagttttggatgatgaatttcttatttctaatctcataagaaacttgtcatagatgtgagggagagaattaagaacagagtgtggggaagaaactgtgtacattttgatgggatccattatgtagctcttgcatactgtcttcaaaaataagttacactataaaggttgttttagacttttaaagttttgccattggttttaaaaaaatttttaaattggctttaaaaatttcttaattgtgtgctgaatacaattttctttattacagaagtaccaacaattacatgtataaacagagaatcctatgtacttgagatataagtaaggttactatcaatcacacctgaaaaatttaaatgttatgaagaaattatctcatttctattaatatgggaactgtgtcttcatctttattactgttctaagtcaactcaatgtagattttacttgcttatggtttcatattttagctaaatagtaaaataatatggatatacattttgttgtgacttactcatactttccttatttggaacttttatgaatatgatatagagactgaaactacaaggaacaaaatgcaatatcaattatacagttgtggcagcactgctatcaatttgttgatagtggttaacacttagaaaaacatttaaaaataatttcacataagtaatgtaatttattagctgtctctgacatttttacagtttggaatagttttttctttttggtgtcctcaccaaaacccaacatcttcaagggcaggaactgtataatttttgccattgtattttgagcacatagcatggtacttgcctctaaatagatactattgttaaaatatttttttaaggtaatatttaaagtgtatgctatggtacagttcagtttgtgacttttgctagtttatgccacttacagttagcaaaatcacttcagcagttcttggaatgttgtgaaaagtgataaaaatcttctgcaacttattccttttattcctcatttaaaataatctaccatagtaaaaacatgtataaaagtgctacttctgcaccacttttgagaatagtgttattttcagtgaatcgatgtggtgaccatattgtaatgcatgtagtgaactgtttaaggcaaatcatctacactagatgaccaggaaatagagaggaaatgtaatttaatttccatttttctttttagagcagtatacaaagatgctgatttgtatttattagactctccttttggatacctagatgttttaacagaaaaagaaatatttgaaaggtatgttcttgaataccttacttataatgctcatgctaaaatataaagaaagacagactgtcccatcatagattgcattttacctcttgagaaatatgttcaccattgttggtatggcagaatgtagcatggtattaactcaaatctgatctgccctactgggccaggattcaagattacttccattaaaaccttttctcaccgcctcatgctaaaccagtttctctcattgctatactgttatagcaattgctatctatgtagtttttgcagtatcattgccttgtgatatatattactttaattattattatacttaacatttttatttacttttgtgttagtatttattctgtcttctccttagatagtaaccttcttaagaaaatatatatgctaagtgttttactggtttaatatgcttagactactcatctacctcaatacttccttggagatctcctcctcagtcacacagagctcaggacttatatttccttggaactcctgttagggtccaatgtacatgaaattccctagacagacagacagtcagttatatggcttgatttcaaagtttcaaaatgatttaatggactatcaagtagtttattaggagaacagttattatactcttctaaaaataaagacttaagcaataaagatgtatatgtatataaaatggctgggttattcctagaagtaccttttcttagaatttagttaaatttaatatccaagatactatcttttcaaccctgagattgtgaaaagtaacttctatcaatataaactttactacatttgtattgtgttagtgtgttacagtataatctagaacaatgtgtctttctatatgatatatgacatttttaatgcctaaaaaaactgatatgtcttagatgattctagtcaggatttacttctagaatagattaaaattctattgaggagagtcaaattaattatcgaattctcagttgttattattgctgttttatttttagctgaaacagattagtcttaatgtaaacactgagaaataaattgatggtcaacctaaaatgtaaaaaagaaattaatagaaaatttaaagagcaacaaagctctgacatttaaaagaaatgaagtacaaatctctagggaccttaaagatcatctaataatttcctcatttttctagataaataaactgagagaccccgaggataaatgatttgctcaaagtcaaatatctacttaatataggaaatttaatttcattctcagtctgttaacatgcaacttttcaatatagcatgttatttcatgctatcagaattcacaaggtaccaatttaattactacagagtacttatagaatcatttaaaatataataaaattgtatgatagagattatatgcaataaaacattaacaaaatgctaaaatacgagacatattgcaataaagtatttataaaattgatatttatatgttttatatcttaaagctgtgtctgtaaactgatggctaacaaaactaggattttggtcacttctaaaatggaacatttaaagaaagctgacaaaatattaattttgcatgaaggtagcagctatttttatgggacattttcagaactccaaaatctacagccagactttagctcaaaactcatgggatgtgattctttcgaccaatttagtgcagaaagaagaaattcaatcctaactgagaccttacaccgtttctcattagaaggagatgctcctgtctcctggacagaaacaaaaaaacaatcttttaaacagactggagagtttggggaaaaaaggaagaattctattctcaatccaatcaactctatacgaaaatttccattgtgcaaaagactcccttacaaatgaatggcatcgaagaggattctgatgagcctttagagagaaggctgtccttagtaccagattctgagcagggagaggcgatactgcctcgcatcagcgtgatcagcactggccccacgcttcaggcacgaaggaggcagtctgtcctgaacctgatgacacactcagttaaccaaggtcagaacattcaccgaaagacaacagcatccacacgaaaagtgtcactggcccctcaggcaaacttgactgaactggatatatattcaagaaggttatctcaagaaactggcttggaaataagtgaagaaattaacgaagaagacttaaaggtaggtatacatcgcttgggggtatttcaccccacacaatgcaattgagtagaatgcaatatgtagcatgtaacaaaatttactaaaatcataggattaggataaggtgtatcttaaaactcagaaagtatgaagttcattaattatacaagcaacgttaaaatgtaaaataacaaatgatttcttttgcaatggacatatctcttcccataaaatgggaaaggatttagtttttggtcctctactaagccagtgataactgtgactataagttagaaagcatttgctttattac


GACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTGAGTGGAGGTCAACGAGCAAGAATTTCTTT
AGCAACgtgaataactaattattggtctagcaagcatttgctgtaaatgtcattcatgtaaaaaaattacagacatttctctattgctt
tatattctgtttctggaattgaaaaaatcctgggggttttatggctagtgggttaagaatcacatttaagaactataaataatggtatag
tatccagatttggtagagattatggttactcagaatctgtgcccgtatcttggtgtcagtgtatttgtttgcctcatagtatagtttac
tacaaatggaaaactctaggattctgcataatactggacagagaagatgtaaatatctgttagttccatcatagaccctgccactccaa
tgtacacaccagctttaggcttcttggtatagataaacatacattttcaaaatttttcatcataatttttcataacaaaatagggaaggca
aatgatgtcacttggcttaaaatctataatatttaaaataaacaggacaaatgcattaacattgttgggggaggaggtcccttagtaga
aacactcttggtccaagcattttaaagctgtcaaagagatgtaaatatagataatgtatgtcaaggagagagctttgtggttaaactgt
aactttcagtttaaacaattattggtgactctgatgtcaaatgtttctcaagctttatctgaacaaaattcttctcactttgttgccaa
agtcgttaacaagaaatcacattgactcattgatgttttggctcctttcccttactttctgttgctttccaaaagctgagacaggaaac
taaccctaactgagcacctgcaattgcctggtagtattctagtcatgtgtgtacttttgtgtgtatgtaatccccttacagctctgcaa
agtaagaattgttctccctgctttacacgaagagatcataagataattgaggctgttagatgttaacttgccaaaagccatacaggaaaa
tggtagagtcacagtttgaaccaggtccttttgattctttacattaaaccatgctttgatcttggaaatacactgtaaggcaataaatc
aatagatacggataattcacaggcttctaaataaatggaagttgattgtttttatctgtgagccaaagtcaagacttattctaagaattc
cacaaatttagataagatagagtatatggcttctagacatccaacatagaactgagtttgtgttatcagtttaagatttggttttgctg
taaggtgcacacactttgaggaactaaaaataattgtctgttcttattctgatcagaatgtgtaatgtgttgtccagttttggatgatg
aatttcttatttctaatctcataagaaacttgtcatagatgtgagggagagaattaagaacagagtgtggggaagaaactgtgtacatt
ttgatgggatccattatgtagctcttgcatactgtcttcaaaaataagttacactataaaggttgttttagacttttaaagttttgcca
ttggttttaaaaaaatttttaaattggctttaaaaatttcttaattgtgtgctgaatacaattttctttattacagaagtaccaacaa
ttacatgtataaacagagaatcctatgtacttgagatataagtaaggttactatcaatcacacctgaaaaatttaaatgttatgaagaa
attatctcatttctattaatatgggaactgtgtcttcatctttattactgttctaagtcaactcaatgtagattttacttgcttatgg
tttcatattttagctaaatagtaaaataatatggatatacattttgttgtgacttactcatactttccttatttggaacttttatgaat
atgatatagagactgaaactacaaggaacaaaatgcaatatcaattatacagttgtggcagcactgctatcaatttgttgatagtggtt
aacacttagaaaaacatttaaaaataatttcacataagtaatgtaatttattagctgtctctgacatttttacagtttggaatagttta
ttttcttttggtgtcctcaccaaaacccaacatcttcaagggcaggaactgtataatttttgccattgtattttgagcacatagcatg
gtacttgcctctaaatagatactattgttaaaatatttttttaaggtaatatttaaagtgtatgctatggtacagttcagtttgtgact
tttgctagtttatgccacttacagttagcaaaatcacttcagcagttcttggaatgttgtgaaaagtgataaaaatcttctgcaactta
ttcctttattcctcatttaaaataatctaccatagtaaaaacatgtataaaagtgctacttctgcaccactttgagaatagtgttatt
tcagtgaatcgatgtggtgaccatattgtaatgcatgtagtgaactgtttaaggcaaatcatctacactagatgaccaggaaatagaga
ggaaatgtaatttaatttccatttttcttttttagAGCAGTATACAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGATACCTAG
ATGTTTTAACAGAAAAAGAAATATTTGAAAGgtatgttcttgaatacctactatataatgctcatgctaaaataaaagaaagacagac
tgtcccatcatagattgcattttacctcttgagaaatatgttcaccattgttggtatggcagaatgtagcatggtattaactcaaatct
gatctgccctactgggccaggattcaagattacttccattaaaaccttttctcaccgcctcatgctaaaccagtttctctcattgctat
actgttatagcaattgctatctatgtagtttttgcagtatcattgccttgtgatatatattactttaattattattatacttaacattt
ttatttacttttgtgttagtatttattctgtcttctccttagatagtaaccttcttaagaaaatatatatgctaagtgttttactgg
tttaatatgcttagactactcatctacctcaatacttccttggagatctcctcctcagtcacacagagctcaggacttatatttccttg
gaactcctgttaggtccaatgtacatgaaattccctagacagacagacagtcagttatatggcttgatttcaaagtttcaaaatgatt
taatggactatcaagtagtttattaggagaacagttattatactcttctaaaaataaagacttaagcaataaagatgtatatgtatat
aaaatggctgggttattcctagaagtaccttcttagaatttagttaaatttaatatccaagatactatcttttcaaccctgagattgt
gaaaagtaacttctatcaatataaactttactacatttgtattgtgttagtgtgttacagtataatctagaacaatgtgtctttctata
tgatatatgacattttaatgcctaaaaaaactgatatgtcttagatgattctagtcaggatttacttctagaatagattaaaattctat
ttgaggagagtcaaattaattatcgaattctcagttgttattattgctgttttatttttagctgaaacagattagtcttaatgtaaacac
ttgagaaataaattgatggtcaacctaaaatgtaaaaaagaaattaatagaaaatttaaagagcaacaaagctctgacatttaaaagaa
atgaagtacaaatctctagggaccttaaagatcatctaataatttcctcattttctagataaataaactgagagaccccgaggataaat
gatttgctcaaagtcaaatatctacttaatataggaaatttaatttcattctcagtctgttaacatgcaacttttcaatatagcatgtt
atttcatgctatcagaattcacaaggtaccaatttaattactacagagtacttatagaatcatttaaaatataataaaattgtatgata
gagattatatgcaataaaacattaacaaaatgctaaaatacgagacatattgcaataaagtatttataaaattgatatttatatgtttt
tatatcttaaagCTGTGTCTGTAAACTGATGGCTAACAAAACTAGGATTTTGGTCACTTCTAAAATGGAACATTTAAAGAAAGCTGACA
AAATATTAATTTTGCATGAAGGTAGCAGCTATTTTTATGGGACATTTTCAGAACTCCAAAATCTACAGCCAGACTTTAGCTCAAAACTC
ATGGGATGTGATTCTTTCGACCAATTTAGTGCAGAAAGAAGAAATTCAATCCTAACTGAGACCTTACACCGTTTCTCATTAGAAGGAGA
TGCTCCTGTCTCCTGGACAGAAACAAAAAAACAATCTTTTAAACAGACTGGAGAGTTTGGGGAAAAAAGGAAGAATTCTATTCTCAATC
CAATCAACTCTATACGAAAATTTTCCATTGTGCAAAAGACTCCCTTACAAATGAATGGCATCGAAGAGGATTCTGATGAGCCTTTAGAG
AGAAGGCTGTCCTTAGTACCAGATTCTGAGCAGGGAGAGGCGATACTGCCTCGCATCAGCACTGGCCCCACGCTTCAGGC
ACGAAGGAGGCAGTCTGTCCTGAACCTGATGACACACTCAGTTAACCAAGTTCAGAACATTCACCGAAAGACAACAGCATCCACACGAA
AAGTGTCACTGGCCCCTCAGGCAAACTTGACTGAACTGGATATATATTCAAGAAGGTTATCTCAAGAAACTGGCTTGGAAATAAGTGAA
GAAATTAACGAAGAAGACTTAAAGgtaggtatacatcgcttgggggtatttcaccccacagaatgcaattgagtagaatgcaatatgta
gcatgtaacaaaatttactaaaatcataggattaggataaggtgtatcttaaaactcagaaagtatgaagttcattaattatacaagca
acgttaaaatgtaaaataacaaatgatttcttttgcaatggacatatctcttcccataaaatgggaaaggatttagtttttggtcctc
tactaagccagtgataactgtgactataagttagaaagcatttgctttattac


Exon 11
Exon 12
Exon 13

Regulatory elements in pre-mRNA splicing



Cis acting regulatory elements

ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
ISS	Intronic Splicing Enhancer
ISE	Intronic Splicing Silencer
CERES	Composite Exonic Regulatory Elements of Splicing

Trans-acting factors

SR proteins	Serine arginine rich proteins (SF2/ASF)
hnRNPs	heterogeneous nuclear RiboNucleoprotein Particles (hnRNPA1)
snRNPs	small nuclear RiboNucleoProteins (U1 snRNP, U2 snRNP)

Exon splicing signals: *enhancers* e *silencers*

Questi elementi *in cis* di regolazione dello *splicing* possono svolgere un ruolo come:

STIMOLATORI (***EXONIC SPLICING ENHANCERS***, **ESE**) o come

INIBITORI (***EXONIC*** o ***INTRONIC SPLICING SILENCERS***, **ESS** o **ISS**).

Gli ESE in particolare sembrano essere molto frequenti e se ne ipotizza la presenza nella maggior parte, se non in tutti gli esoni.

EXONIC SPLICING ENHANCERS

Legano principalmente specifiche **proteine ricche in Serina/ Arginina (SR)** che fanno parte di un gruppo di fattori di *splicing* altamente conservati caratterizzati dalla presenza di:

- **1-2 motivi di riconoscimento dell'RNA** (**RRM**)
- da un dominio C-terminale ricco in **dipeptidi Arg/Ser** (il **dominio RS**) coinvolto in interazioni proteina-proteina

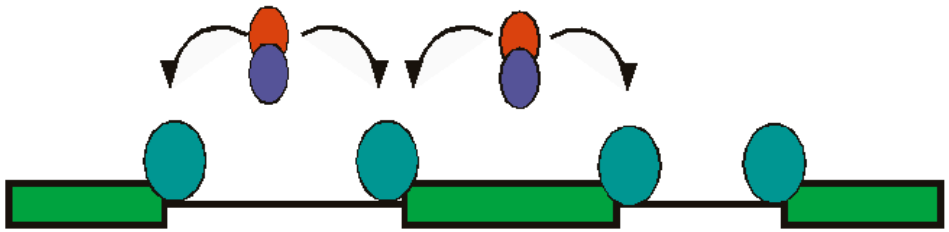
SR Proteins

Labs of Fu, Krainer, Manley, Roth and others.

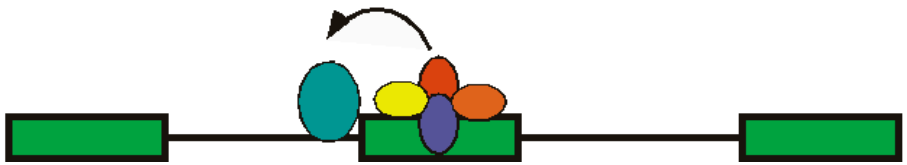
Required splicing factors and effectors of splice site choice.

1 or 2 RNP Domains

C-terminal SR Domain containing multiple SR Dipeptides



And / Or

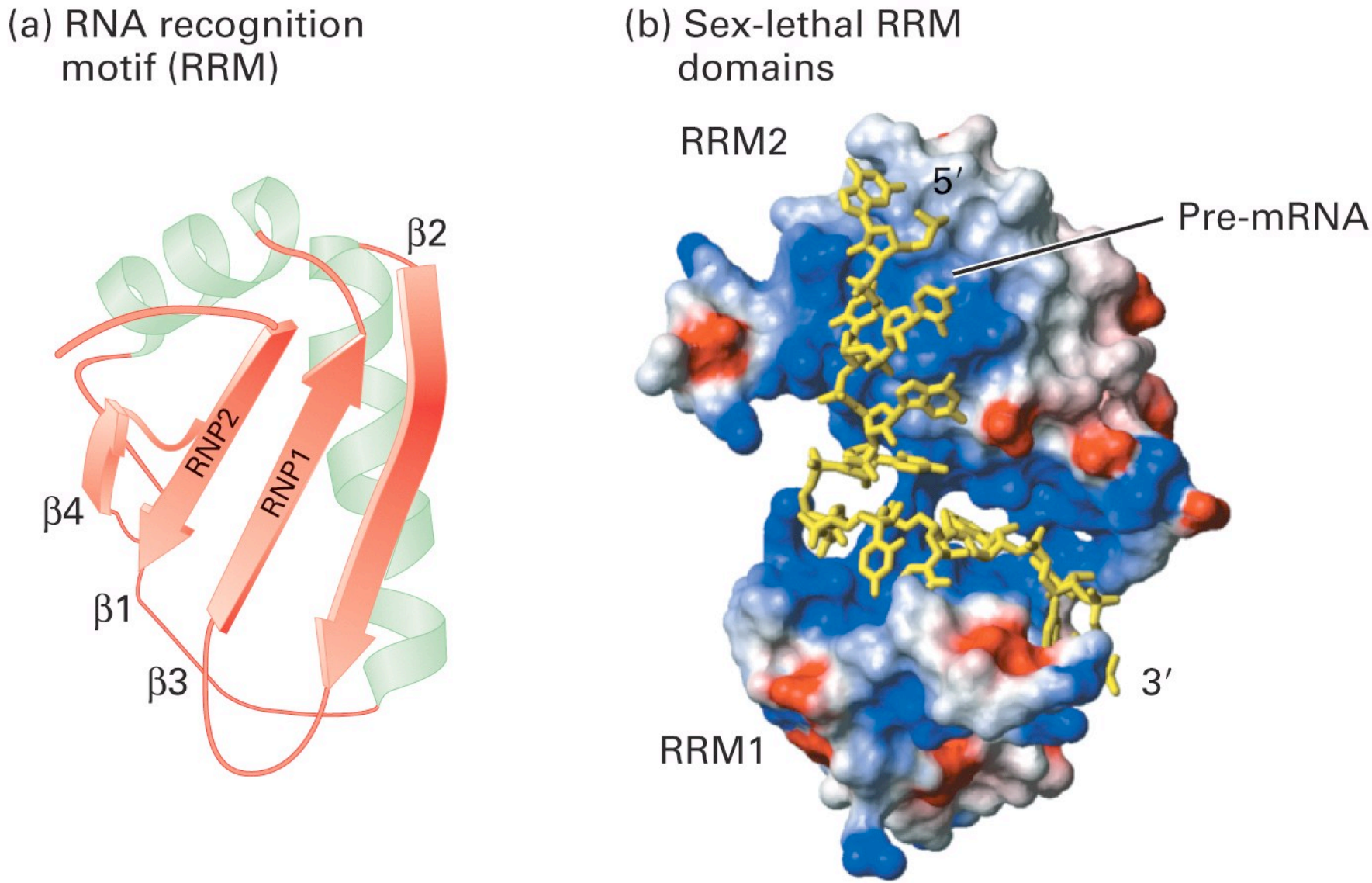


The SR Proteins are a family of proteins with a common domain structure of 1 or 2 RNP RNA binding domains (also called RRMs) and a C-terminal domain rich in SR dipeptides.

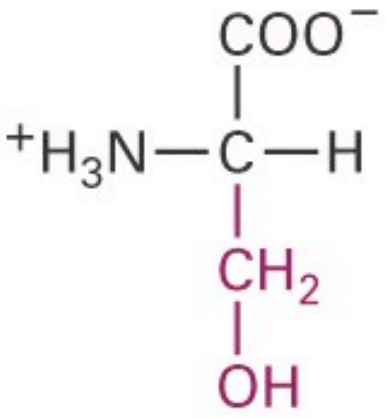
These proteins are involved in many aspects of splicing, but most significantly they bind to Exonic Splicing Enhancers (ESEs) and stimulate spliceosome assembly at the adjacent sights.

It is thought that most exons carry ESE's and require SR proteins for exon recognition.

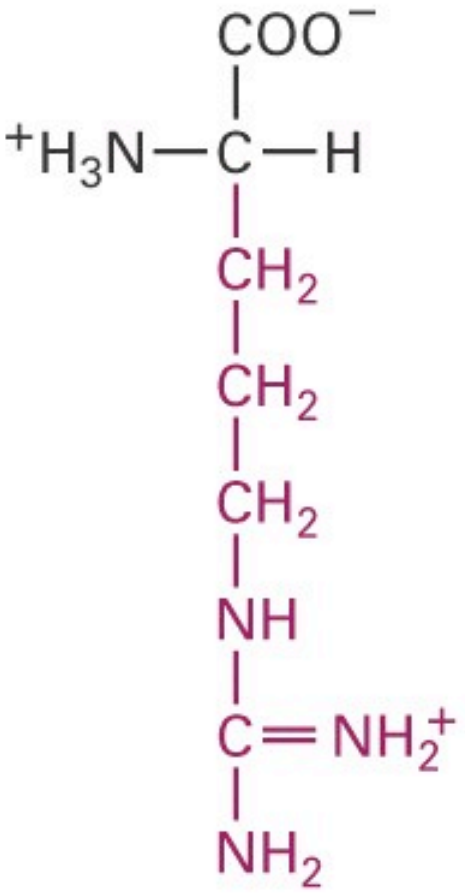
SR Proteins bind to specific RNA elements using their RNA binding domains similar to those in the Sex-Lethal protein.



The SR domain is an effector domain needed for splicing activation. There is evidence that it interacts both with other proteins and with RNA. It can be highly phosphorylated on Serine and SR protein activity is thought to be modulated by specific kinases and phosphatases.



Serine
(Ser or S)



Arginine
(Arg or R)

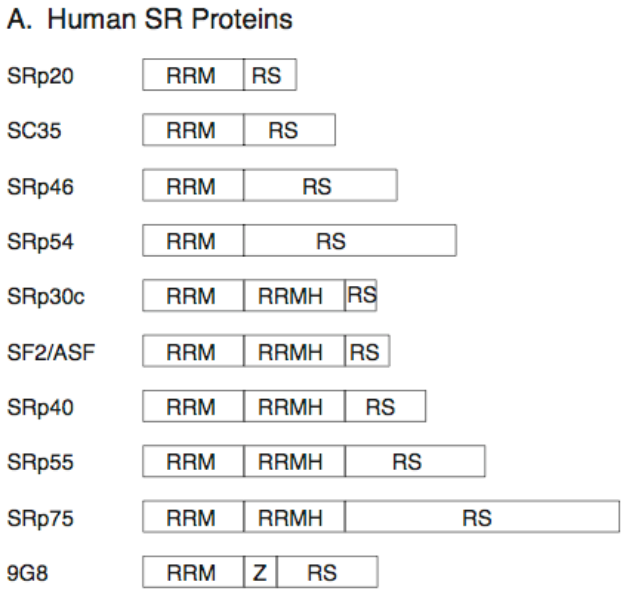
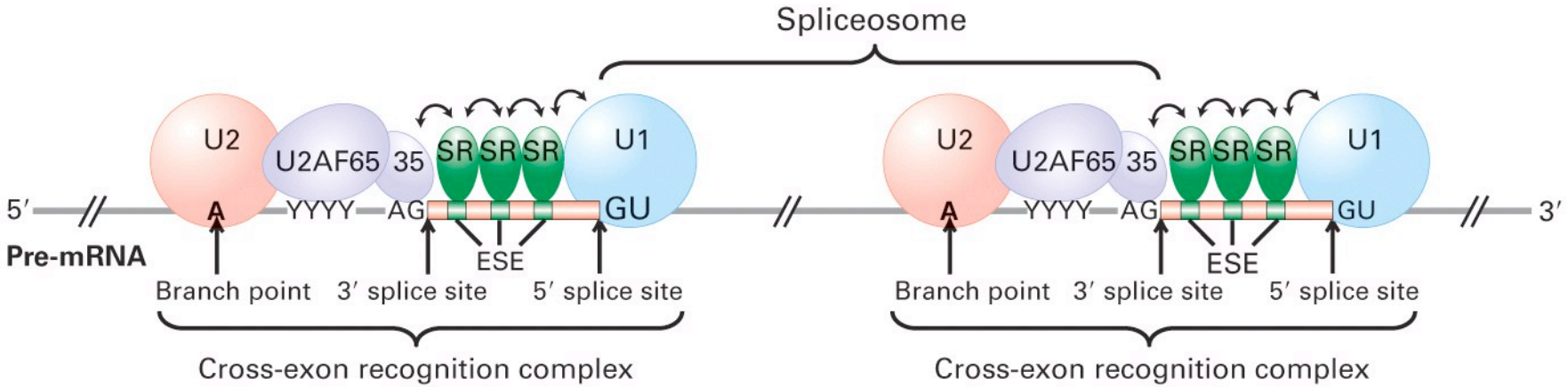


FIGURE 2. Schematic diagram of human SR proteins and SR related proteins. **A**: The domain structures of the known members of the human SR protein family are depicted. RRM: RNA recognition motif; RRMH: RRM homology; Z: zinc knuckle, RS: arginine/serine-rich domain. **B**: The domain struc-

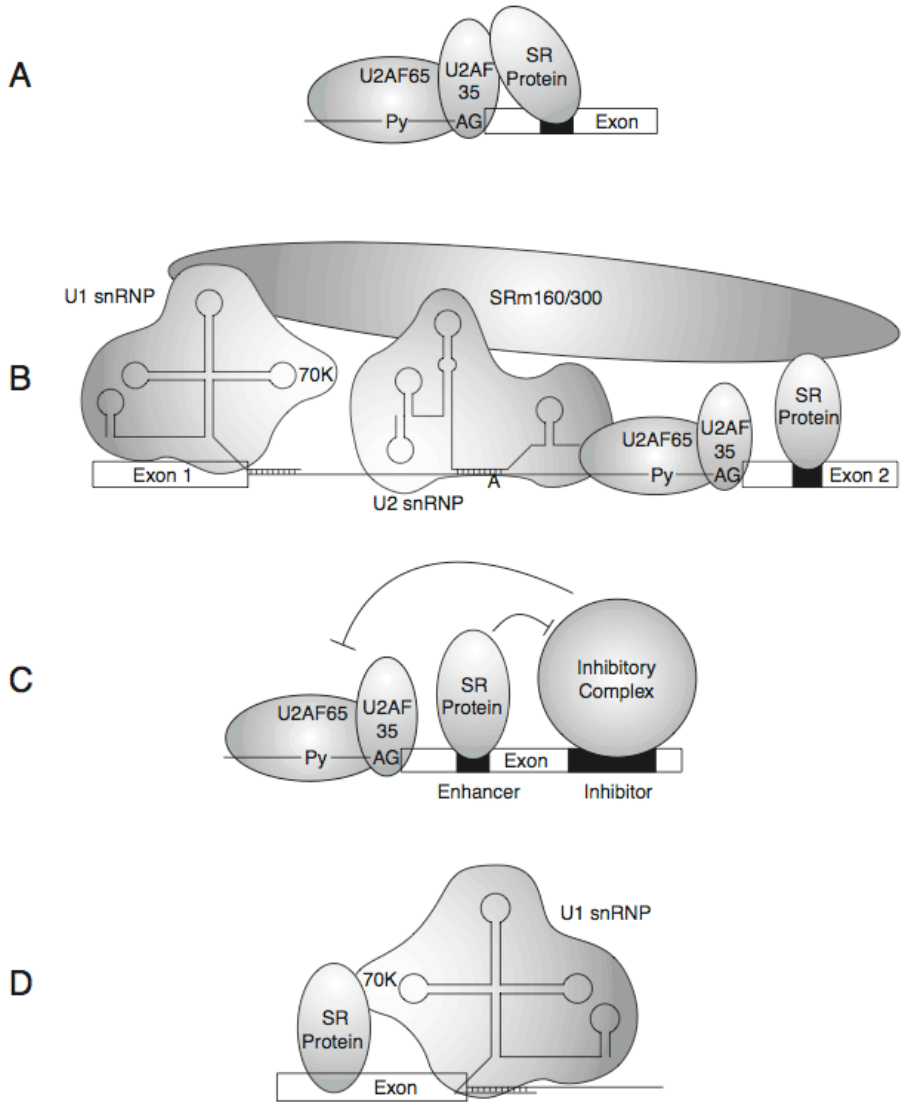
SR Proteins are important in stimulating assembly of cross exon complexes. These complexes must subsequently rearrange to form the spliceosome across the intron.

SR proteins are also important targets for the regulation of splicing.

Because of the need to form a cross exon complex prior to spliceosome assembly, mutations in single splice sites, or in ESE's, frequently cause skipping of the entire exon.



Exon-dependent functions of SR proteins



assist 3’ss definition

assist intron definition
cofactor model

inhibitor model
antagonistic function

assist 5’ss definition

FIGURE 3. The regulated exon-dependent functions of SR proteins. **A**: U2AF recruitment model. An enhancer-bound SR protein interacts with the RS domain of $U2AF^{35}$, thereby recruiting $U2AF^{65}$ to the pre-mRNA. **B**: Coactivator model. In this model, the splicing enhancer functions through interactions with the splicing coactivator SRm160/300, which also interacts with U1 snRNP and U2 snRNP. Some of these interactions may be indirect. **C**: Inhibitor model. In the absence of the enhancer-bound SR protein, a downstream splicing inhibitor functions to prevent the splicing of the upstream intron. The function of the splicing enhancer is to counteract the splicing inhibitor. **D**: Recruitment of U1 snRNP to a 5′ splice site. An SR protein bound to the upstream exon interacts with U1-70K and recruits U1 snRNP to the 5′ splice site.

In vitro splicing

Why to perform an in vitro splicing assay

- it allows the characterization of many fundamental features of the splicing reaction

- modification in reaction conditions can lead to the accumulation, isolation and characterization of reaction intermediates, that can be important to decipher the splicing mechanism

- biochemical manipulation permits to investigate the importance of RNA elements and *trans*-acting factors

Method principles

- The pre-mRNA of interest (usually labelled) is incubated with nuclear extracts (NE) which contain all the elements to perform the splicing reaction (the spliceosome).
- The nuclear extract can be selectively depleted of some splicing factors to understand their importance.
- As a control the assay is performed with the S100 cytoplasmic fraction which is depleted of essential splicing factors (the SR proteins).

Method description

1 NE preparation
S100 preparation

+

2 pre-mRNA in vitro transcription

+

3 splicing assay

1

Normally HeLa cell NE.

The quality is fundamental for the splicing reaction to occur.

To be tested with a substrate of known splicing efficiency.

S100 is the supernatant of the pelletted nuclei, without SR proteins that are depleted with $MgCl_2$.

2

The DNA of interest (<2000bp) with the bacteriophage SP6 or T7 promoters is incubated with the suitable polymerase, cap analogs, [^{32}P-α]UTP and cold ATP, CTP and GTP. Incubation at 37^{o} C, 2h. SDS-PAGE and elution of the pre-mRNA.

3

Combine NE or S100 with the pre-mRNA and incubate at 30^{o} C for different times. Digest with proteinase K, extract with phenol-chloroform and run on denaturing SDS-PAGE. Bands are detected with the radioactive emission.

Advantages:

- experimental flexibility: investigation directed towards specific aspects of the kinetic of the splicing reaction

- biochemical manipulation: possibility to investigate the influence of specific RNA elements and *trans* acting factors on the splicing reaction

Disadvantages:

-rates of intron removal are slower *in vitro* than *in vivo*: lack of compartmentalization and consequent concentration differences. No cotranscriptional processing. PolII transcription is missing

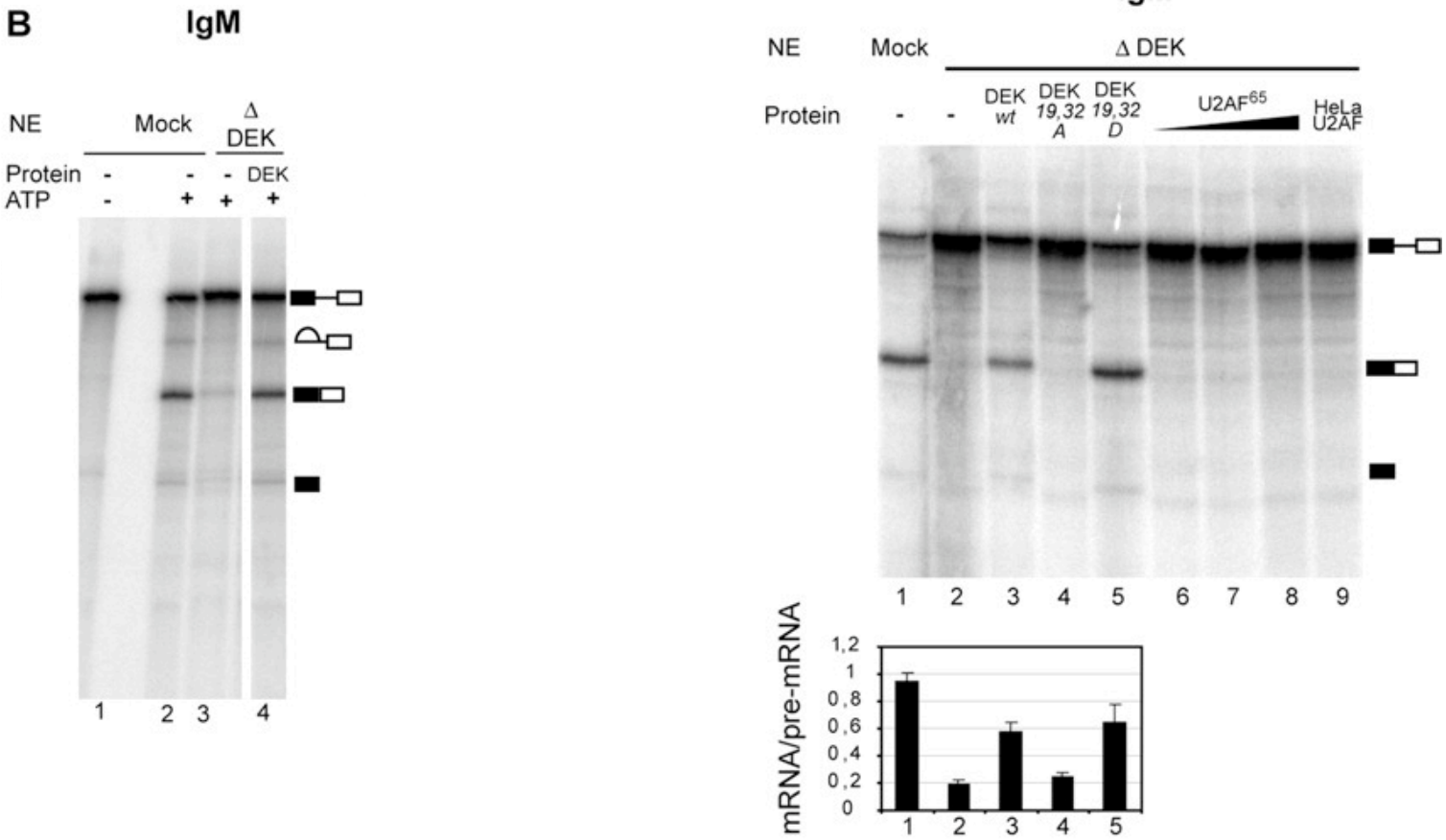
-in vitro transcription, RNA purification and splicing reaction efficiencies limit the size of the test substrate (less than ~2000 bp)

-in the nuclear extract many other processes go on: artifacts should be considered

- impossible to evaluate the relationship with other gene expression events such as transcription, capping, poly(A)

some splicing aspects can be better evaluated using hybrid minigenes (we will explain later)

Example 1 DEK, a chromatin- and RNA-associated protein is required for pre-mRNA splicing

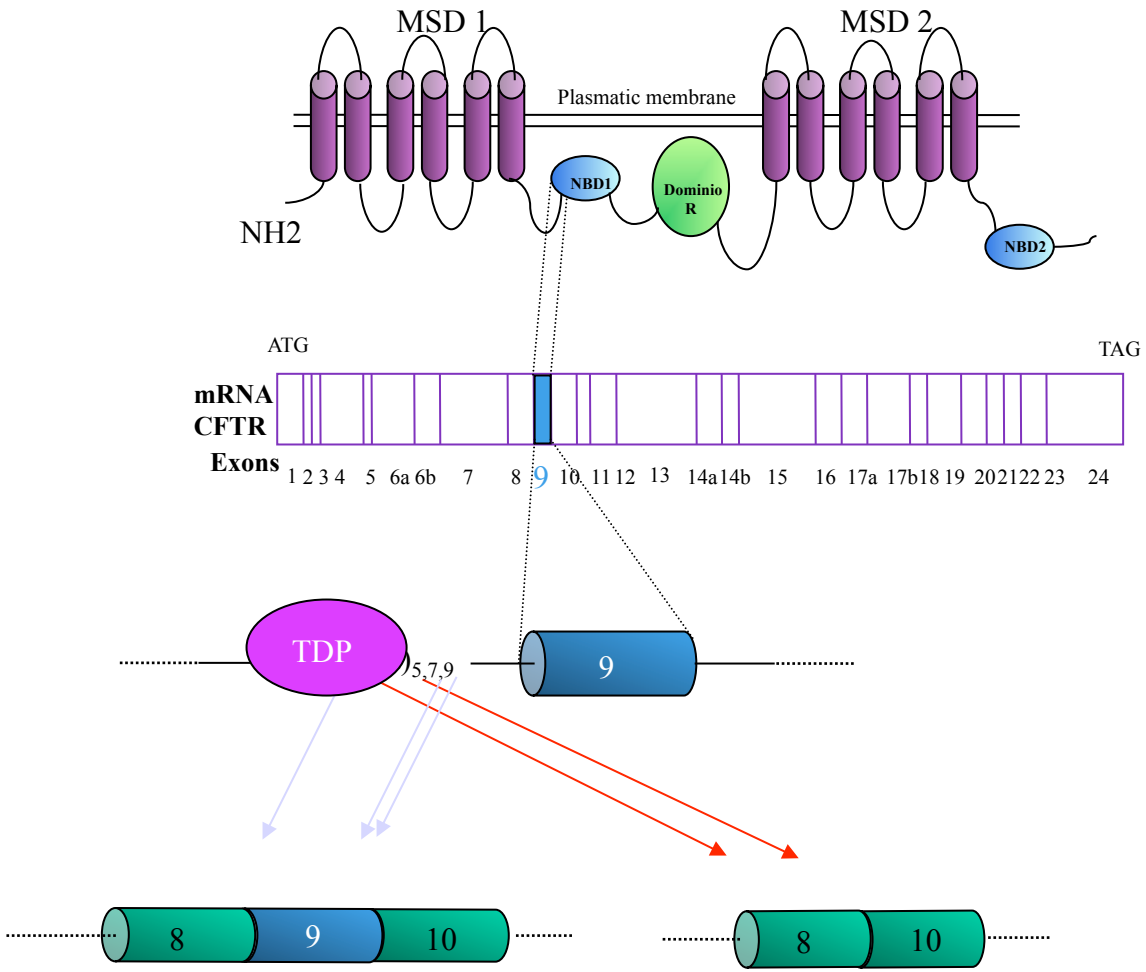


The novelty is this case is that DEK is a chromatin associated protein which has never been implicated in splicing.

(B) Splicing assay using wild-type pre-mRNA, in the presence or absence of recombinant purified DEK.
(C) Splicing complementation assay as in (B) with recombinant DEK; DEK Ala (A) and Asp (D) mutants at serines 19 and 32; recombinant U2AF^{65}; purified HeLa U2AF heterodimer. Quantification of splicing efficiency is shown.

Soarers et al., 2006, Science

Example 2 :
UG tracts bind to TDP43 splicing factor and inhibits CFTR exon 9 inclusion



Example 2: effect of a UG repeat near the 3'ss on splicing efficiency

Variation in the number of GT T repeats near the 3'ss in CFTR intron 8 is associated to Cystic Fibrosis