Modification and processing of eukaryotic pre-mRNAs

RNA Splicing: Removal of Introns From Primary Transcripts
Pre-mRNA splicing

- most eukaryotic protein-coding genes are interrupted with introns
- Intron (intervening sequence-IVS) does not code for protein
- Exon – protein coding sequence
- Exons relatively short (1 nt)
- Introns can be up to several 1,000 nt
- Primary transcripts (pre-mRNAs) up to 100,000 nt
Cis elements required for splicing

Yeast

$5\text{'}ss$  
\begin{center}
\begin{tabular}{cc}
AG & GUAAGU
\end{tabular}
\end{center}

$3\text{'}ss$

\begin{center}
\begin{tabular}{cc}
* & UACUAAC
\end{tabular}
\end{center}

Vertebrates

\begin{center}
\begin{tabular}{cc}
ESE & AG
\end{tabular}
\end{center}

$5\text{'}ss$

\begin{center}
\begin{tabular}{cc}
AG & GUAAGU
\end{tabular}
\end{center}

$BP$

\begin{center}
\begin{tabular}{cc}
* & CURAY
\end{tabular}
\end{center}

$3\text{'}ss$

\begin{center}
\begin{tabular}{cc}
YYY & YAG
\end{tabular}
\end{center}

Plants

\begin{center}
\begin{tabular}{cc}
ESE? & AG
\end{tabular}
\end{center}

$5\text{'}ss$

\begin{center}
\begin{tabular}{cc}
AG & GUAAGU
\end{tabular}
\end{center}

$UA\text{-rich}$

\begin{center}
\begin{tabular}{cc}
62100 70 49 & 64 95100 44
79 99 58 53 & 42100 57
\end{tabular}
\end{center}

$BP$

\begin{center}
\begin{tabular}{cc}
* & CURAY
\end{tabular}
\end{center}

$3\text{'}ss$

\begin{center}
\begin{tabular}{cc}
YYY & UGYAG
\end{tabular}
\end{center}

5’ss – 5’ splice site (donor site)
3’ss – 3’ splice site (acceptor site)
BP – branch point (A is branch point base)
YYYY$^{10-15}$ – polypyrimidine track

Y – pyrimidine
R – purine
N – any base
Frequency of bases in each position of the splice sites

Donor sequences: 5’ splice site

<table>
<thead>
<tr>
<th></th>
<th>exon</th>
<th>intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>%A</td>
<td>30</td>
<td>40 64 9 0 0 62 68 9 17 39 24</td>
</tr>
<tr>
<td>%U</td>
<td>20</td>
<td>7 13 12 0 100 6 12 5 63 22 26</td>
</tr>
<tr>
<td>%C</td>
<td>30</td>
<td>43 12 6 0 0 2 9 2 12 21 29</td>
</tr>
<tr>
<td>%G</td>
<td>19</td>
<td>9 12 73 100 0 29 12 84 9 18 20</td>
</tr>
</tbody>
</table>

Polypyrimidine track (Y = U or C;  N = any nucleotide)

Acceptor sequences: 3’ splice site

<table>
<thead>
<tr>
<th></th>
<th>exon</th>
<th>intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>%A</td>
<td>15</td>
<td>10 10 15 6 15 11 19 12 3 10 25 4 100 0 22 17</td>
</tr>
<tr>
<td>%U</td>
<td>51</td>
<td>44 50 53 60 49 49 45 45 57 58 29 31 0 0 8 37</td>
</tr>
<tr>
<td>%C</td>
<td>19</td>
<td>25 31 21 24 30 33 28 36 36 28 22 65 0 0 18 22</td>
</tr>
<tr>
<td>%G</td>
<td>15</td>
<td>21 10 10 10 6 7 9 7 7 5 24 1 0 100 52 25</td>
</tr>
</tbody>
</table>

Polypyrimidine track (Y = U or C;  N = any nucleotide)
Chemistry of pre-mRNA splicing

two cleavage-ligation reactions
  • transesterification reactions - exchange of one phosphodiester bond for another - not catalyzed by traditional enzymes
    • branch site adenosine forms 2’, 5’ phosphodiester bond with guanosine at 5’ end of intron

Pre-mRNA

5’ G-p-G-U A-G-p-G 3’

branch site adenosine

First cleavage-ligation (transesterification) reaction
- ligation of exons releases **lariat** RNA (intron)
Spliceosome

- large ribonucleoprotein complex
  - five snRNPs and approx. 200 additional proteins
  - assembly at each intron

- snRNP (small nuclear ribonucleoprotein)
  - snRNA and seven core Sm (LSM – U6 snRNA) proteins
  - snRNP-specific proteins
  - snRNAs contain unique 5' terminal cap 2,2,7-trimethylguanosine (3mG)
snRNPs

U₁, U₂, U₄, U₅, U₆: Nukleoplasma, U₃: Nucleolus (bis ~ U₃₀).
U₁, U₂, U₄, and U₅ snRNA: 5‘-Ende cap:m₃GpppN
U₆ snRNA: 5‘ Ende pppN
Proteine: „core“ Proteine: jedes snRNP (Sm – Proteine: B/B’, D1, D2, D3, E, F und G), und Proteine, die spezifisch für jede RNA sind.

U₄/U₆ komplexiert, alle anderen einzeln
Recognition of splice sites

invariant GU and AG dinucleotides at intron ends
  - donor (upstream) and acceptor (downstream) splice sites are within conserved consensus sequences

- small nuclear RNA (snRNA) U1 recognizes the donor splice site sequence (base-pairing interaction)
- U2 snRNA binds to the branch site (base-pairing interaction)

Y= U or C for pyrimidine; N= any nucleotide
Spliceosome - assembly of the splicing apparatus

- splicing snRNAs - U1, U2, U4, U5, U6
- snRNAs are associated with proteins (snRNPs or “snurps”)
- antibodies to snRNPs are seen in the autoimmune disease **systemic lupus erythematosus** (SLE)
U2 snRNA Base Pairs With Intron Branch Point
Step 2: binding of U4/U6.U5 tri snRNP

Step 3: U1 is released, then U4 is released
Step 4: U6 binds the 5’ splice site and the two splicing reactions occur, catalyzed by U2 and U6 snRNPs.
Spliceosome assembly

A complex

U1
GU
U2
U2AF
A
YAG

B complex

U4
U6
U5

U4
U1
U5
U2
YAG

C complex

U2
U6
U5

+ ~200 non-snRNP proteins

hnRNP
SR proteins
RNA helicases
Cyclophilins

kinases and phosphatases
U5, U6 Interactions in Splicing
Roles of snRNPs in Splicing

- **U1 snRNA** binds to 5’ splice site
- **U2 snRNP** binds to:
  - branch-point sequence within intron
  - U6 snRNP
- **U5 snRNP**
  - Not complementary to splicing substrate or other snRNPs
  - Associates with last nucleotide of one exon and first nucleotide of next
  - Aligns two exons for splicing reaction
- **U4 snRNP**
  - Binds U6 snRNP
  - No evidence for direct role in splicing reaction
  - May sequester U6 snRNP until appropriate time for U6 to bind to 5’ splice site
- **U6 snRNA** binds to:
  - 5’ splice site
  - U2 snRNP
Spliceosome & ATP -> RNA-RNA Rearrangements - I

(A) U1
exon 1
CAUUCA
GUAUGU
3’
rearrangement
exon 1
GUAUGU
3’

(B) BBP
exon 2
UACUAAC
3’
rearrangement
exon 2
UACUAAC
3’

Figure 6–30 part 1 of 2. Molecular Biology of the Cell, 4th Edition.
Spliceosome & ATP -
RNA-RNA
Rearrangements - II

Figure 6-30 part 2 of 2. Molecular Biology of the Cell, 4th Edition.
Spliceosome cycle
The Exon Definition Hypothesis

Figure 6–33. Molecular Biology of the Cell, 4th Edition.
5` and 3` splice site selection

Intron definition model

Exon definition model
Human Genome

3.2 million DNA base pairs

1.5% encode proteins $\leq 98.5\%$ not protein encoding

~ 30,000 genes encoding 100,000 - 200,000 proteins

How are 100,000 to 200,000 proteins produced from 30,000 genes?

Alternative splicing
Alternative pre-mRNA splicing

- Frequent event in mammalian cells

- Genes coding for tens to hundreds of isoforms are common.

- For ex. it is estimated that ~60% of genes on chromosome 22 encode >2 mRNAs

- ~50% of human genes are alternatively spliced

- Regulation of alternative splicing imposes requirement for signals that modulate splicing

-Enhancers and silencers of splicing:
  Enhancers: Exonic Splicing Enhancers: SR proteins
  Silencers: Exonic Splicing Silencers: not well characterized.
  Intronic Splicing Silencers: hnRNP family

An amazing example of splicing complexity- how many variants???
What is the largest number of possible spliced mRNAs derived from a Drosophila gene?
A. 300 spliced variants
B. 3,000 spliced variants
C. 30,000 spliced variants
D. 300,000 spliced variants

38,016 different spliced forms in Dscam gene (cell surface protein involved in neuronal connectivity)
Alternative pre-mRNA Splicing
Patterns of alternative exon usage

- one gene can produce several (or numerous) different but related protein species (isoforms)

Cassette

Mutually exclusive

Internal acceptor site

Alternative promoters
Alternative Pre-mRNA Splicing Can Create Enormous Diversity - I

Figure 6-27. Molecular Biology of the Cell, 4th Edition.
The Troponin T (muscle protein) pre-mRNA is alternatively spliced to give rise to 64 different isoforms of the protein.

- Constitutively spliced exons (exons 1-3, 9-15, and 18)
- Mutually exclusive exons (exons 16 and 17)
- Alternatively spliced exons (exons 4-8)

Exons 4-8 are spliced in every possible way giving rise to 32 different possibilities.

Exons 16 and 17, which are mutually exclusive, double the possibilities; hence 64 isoforms.
How is alternative splicing achieved?

Alternative exons often have suboptimal splice sites and/or length

Splicing of regulated exons is modulated:
1. Proteins – SR proteins and hnRNPs
2. cis elements in introns and exons – splicing enhancers and silencers

Differences in the activities and/or amounts of general splicing factors and/or gene-specific splicing regulators during development or in different tissues can cause alternative splicing
SR proteins

- nuclear phosphoproteins, localised in speckles
- phosphorylation status regulates their subcellular localisation and protein-protein interactions
- shuttling proteins (h9G8, hSRp20, hSF2/ASF)

- constitutive splicing
- alternative 5` splice site selection
- alternative 3` splice site selection
  exon-(in)dependent

- found in all eukaryotes except in *S. cerevisiae*
5` and 3` splice site selection – role for SR proteins

Specific sequence independent – over both intron and exon

Specific sequence dependent - over both intron and exon
Negative and Positive Control of Alternative Pre-mRNA Splicing

(A) NEGATIVE CONTROL

TISSUE 1

primary transcript

splicing

mRNA

TISSUE 2

repressor R

no splicing

mRNA

(B) POSITIVE CONTROL

primary transcript

no splicing

mRNA

activator A

splicing

mRNA

Figure 7–90. Molecular Biology of the Cell, 4th Edition.
U2AF recruitment model

Specific sequence required

SR protein binds to ESE and promote binding of U2AF to Py tract, which results in activation of adjacent 3' ss

This is mediated by interaction of RS domain of SR protein with the small subunit (U2AF35) of U2AF
Functional antagonism of SF2/ASF (SR protein) and hnRNP A1 in splice site selection

Excess of hnRNP A1 results in usage of distal 5′ss

Mechanism:
SF2/ASF interferes with hnRNP A1 binding and enhances U1 snRNP binding at both duplicated 5′ss.

Simultaneous occupancy of both 5′ss results in selection of proximal 5′ss

hnRNP A1 binds cooperatively to pre-mRNA and interferes with U1 snRNP binding at both sites. This results in a shift to the distal 5′ss

No specific target sequences required
Functional antagonism of SF2/ASF (SR protein) and hnRNP A1 in splice site selection

Specific sequence required – splicing enhancers can antagonize the negative activity of hnRNP bound to ESS

SR protein binds to ESE and hnRNP A1 binds to silencer
Initial binding of hnRNP A1 to silencer causes further binding of hnRNP A1 upstream in the exon, but this is prevented by binding of SF2/ASF to ESE.
SC35 does not affect hnRNP A1 binding

ESS suppresses SC35, but not SF2/ASF-dependent splicing

HIV-1 tat exon 3
Negative regulation of alternative splicing by hnRNP I (PTB)

PTB – pyrimidine tract binding protein
- 4 RRM
- three alternative forms
  - Differential expression of isoforms in neural cell lines and in rat brain

PTB represses several neuron-specific exons in non-neuronal cells. In β-tropomyosin exon 7 is repressed in non-muscle tissue, but in α-tropomyosin PTB represses exon 3 in smooth muscle. How is repression achieved?

PTB binds to intronic splicing repressor (black lines; UC-rich; 80-124 nt long), and prevents binding of U2AF to the Py tract.
Alternative splicing in sex determination of *Drosophila*

- **Sxl** gene:
  - Transcribed; pre-mRNA spliced in one way.
  - *Sxl* protein

- **tra** gene:
  - *Sxl* protein causes *tra* pre-mRNA to undergo specific processing.
  - *tra* protein

- **dsx** gene:
  - *tra* protein with *tra-2* protein directs ♀ splicing of *dsx* pre-mRNA.
  - ♀-specific protein
  - ♀-specific protein with *ix* protein represses ♀ differentiation genes

- **Somatic cell differentiation**:
  - ♀-somatic cell differentiation
  - ♀-somatic cell differentiation

- **X:A ratio**:
  - 1.0 (XX female)
    - ♀-somatic cell differentiation
  - 0.5 (XY male)
    - ♀-somatic cell differentiation

- **No functional Sxl protein**
  - Transcribed; alternative splicing of pre-mRNA; mRNA does not encode functional protein
  - No functional *Sxl* protein
  - No functional *tra* protein
  - No functional ♀-specific protein
  - ♀-specific protein with *ix* protein represses ♀ differentiation genes
  - ♀-somatic cell differentiation

- **♂-specific protein**
  - Transcribed; alternative splicing of pre-mRNA; mRNA does not encode functional protein
  - No functional ♀-specific protein
  - No functional ♀-somatic cell differentiation
  - ♀-somatic cell differentiation

The Cascade that Determines Sex in Drosophila - I

**GENE**

**MALE primary RNA transcript**
\[ X : A = 0.5 \]

regulated 3’ splice site

regulated 3’ splice site

regulated 3’ splice site

regulated 3’ splice site is inactive

**X chromosome/autosome ratio**

**Sex-lethal (Sxl)**

5’

3’

nonfunctional protein produced

nonfunctional protein produced

nonfunctional protein produced

Dsx protein

400 aa

150 aa that are male-specific

REPRESSES FEMALE DIFFERENTIATION GENES

MALE DEVELOPMENT

---

Figure 7–91. Molecular Biology of the Cell, 4th Edition.

Figure 7–92 part 1 of 2. Molecular Biology of the Cell, 4th Edition.
The Cascade that Determines Sex in Drosophila - II

Figure 7–91. Molecular Biology of the Cell, 4th Edition.

Figure 7–92 part 2 of 2. Molecular Biology of the Cell, 4th Edition.
Alternative RNA Splicing in *Drosophila* Sex Determination

Female

Female-specific splicing

Pre-mRNA

Default splicing

Male

*tra*

*dsx*
Alternative polyadenylation and splicing of the human CACL gene in thyroid and neuronal cells.
Other examples of splicing regulation

• **CELF (CUG-BP and ETR3-like factors) proteins are involved in cell-specific and developmentally regulated alternative splicing**
  – Three RRM

  – CELF4, CUG-BP, and ETR3 expression is developmentally regulated in striated muscle and brain

  – There they bind to muscle specific enhancers in the cardiac troponin-T gene (cTNT) and promote inclusion of the dev. regulated exon 5 (role in the pathogenesis of myotonic distrophy)

  – Myotonic distrophy type 1 (DM1) is caused by a CTG trinucleotide expansion in the 3'-UTR of the DM protein kinase gene. These repeats bind CUG-BP (CELF protein), which results in elevated level of CUG-BP expresion, leading to aberrantly regualted splicing of cardiac troponin T and insuline receptor in DM1 skeletal muscle

• **NOVA-1 is a neuron –specific RNA binding protein**
  – One KH domain

  – NOVA-1 null mice show splicing defects in pre-mRNAs for glycine α2 exon 3A and in the GABA<sub>A</sub> exon γ2L

  – It recognises intronic site adjacent to the alternative exon 3A and promotes ist inclusion
Mutations that disrupt splicing

- $\beta^o$-thalassemia - no $\beta$-chain synthesis
- $\beta^+$-thalassemia - some $\beta$-chain synthesis

Normal splice pattern:

Exon 1 \hspace{1cm} Intron 1 \hspace{1cm} Exon 2 \hspace{1cm} Intron 2 \hspace{1cm} Exon 3

Donor site: /GU
Acceptor site: AG/

Intron 2 acceptor site $\beta^o$ mutation: no use of mutant site; use of cryptic splice site in intron 2

Exon 1 \hspace{1cm} Intron 1 \hspace{1cm} Exon 2

Intron 2 cryptic acceptor site: UUUCUUUCAG/G

Translation of the retained portion of intron 2 results in premature termination of translation due to a stop codon within the intron, 15 codons from the cryptic splice site
Intron 1 $\beta^+$ mutation creates a new acceptor splice site: use of both sites

Donor site: /GU AG/: Normal acceptor site (used 10% of the time in $\beta^+$ mutant)

CCUAUUAG/U: $\beta^+$ mutant site (used 90% of the time)
CCUAUUUGG U: Normal intron sequence (never used because it does not conform to a splice site)

Translation of the retained portion of intron 1 results in termination at a stop codon in intron 1

Exon 1 $\beta^+$ mutation creates a new donor splice site: use of both sites

/GU: Normal donor site (used 60% of the time when exon 1 site is mutated)

GGUG/GUAAGGCC: $\beta^+$ mutant site (used 40% of the time)
GGUG GUGAGGCC: Normal sequence (never used because it does not conform to a splice site)

The GAG glutamate codon is mutated to an AAG lysine codon in Hb E

The incorrect splicing results in a frameshift and translation terminates at a stop codon in exon 2
AT-AC introns I

A minor class of nuclear pre-mRNA introns

Referred to as AT-AC or U12-type introns (they frequently start with AT and terminate with AC)

Contain different splice site and BP sequences and are excised by an alternative U12-type spliceosome

Their splicing also requires five snRNAs

Only U5 is common to both spliceosome types, while U11, U12, U4atac, and U6atac carry out the functions of U1, U2, U4, and U6 snRNAs, respectively

Other components of the splicing machinery appear to be shared by both spliceosomes

But some snRNP specific proteins are different
AT-AC introns II

Of note is that introns with GT-AG borders, but which are spliced by the U12 spliceosome, and introns with AT-AC borders, spliced by the classical U2 spliceosome also occur, at a frequency comparable to that of the U12-type with AT-AC termini.

Hence, residues other than terminal dinucleotides determine which of the two spliceosomes will be utilised.

U12 class introns represent approximately 0.1% of all introns.

They are found in organisms ranging from higher plants to mammals, and their positions within equivalent genes are frequently phylogenetically conserved.

The genomes of *Saccharomyces cerevisiae* and *Caenorhabditis elegans* contain no U12-type introns.

Since U12 introns clearly originated prior to the divergence of the plant and animal kingdoms, their absence in *C. elegans* is most easily explained by their conversion to U2-type introns or by intron loss, rather than by intron gain in plants and vertebrates.
Major U2 spliceosome

- SRp34
- U1-70K
- U1
- GU
- U2
- U2AF
- SRp30
- YAG

Minor U12 spliceosome

- U11-35K
- U11
- SRp
- U12
- YAC
- AU
- A
- YAC
Types of RNA Splicing

• Splicing of nuclear RNA encoding proteins ($cis$-splicing)
  – Requires conserved sequences in introns, spliceosomes

• $Trans$-splicing of nuclear RNA

• Self-splicing introns
  – Type I, Type II
    • Classification depends on cleavage mechanism
  – Yeast tRNA
  – Ribosomal RNAs in lower eukaryotes
  – Fungal mitochondrial genes
  – Bacteriophage T4 (3 genes); bacteria (rare)
Self-Splicing Introns

- **Group I introns**
  - *Tetrahymena* rRNA, others
  - Requires added GTP

- **Group II introns**
  - Fungal mitochondrial genes, others
  - Lariat intermediate for splicing
  - Reaction mechanism similar to spliceosomes
Self-Splicing Introns - I

Group I self-splicing intron sequences

precursor RNA molecule

transient intermediate

excised intron sequence

ligated exon sequences

Figure 6–36 part 1 of 2. Molecular Biology of the Cell, 4th Edition.
Self-Splicing Introns - II

Figure 6–6 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

Figure 6–36 part 2 of 2. Molecular Biology of the Cell, 4th Edition.
Trans-splicing

Generates 5' ends of mRNAs

All mRNAs in Trypanosomes are generated by trans-splicing

In *C. elegans* and *Ascaris lumbricoides* mixed situation

Tightly coupled with polyadenylation

Transcript #1 SLRNA (spliced leader RNA)

Transcript #2 mRNA

Hybrid mRNA
Organisms With *Trans*-Splicing

Trypanosomes: only trans splicing
Euglena, Nematoden, Fachwürmer: cis - und trans splicing
Trans splicing in Drosophila

Drosophila: vor kurzem gefunden, Mod(mdg4) Gen, codiert für 26 verschiedenen nuklearen Proteine, die verschiedene Aktivitäten im Kern ausführen. Ein Gen, die ersten 4 Exons sind gleich, das letzte Exon wird durch trans-splicing angefügt. Die 26 terminalen Exons sind teilweise am gleichen DNA Strang, aber teilweise am Gegenstrang des Genlocus codiert und werden separat transkribiert.
**Trans-splicing**

- Splicing does not require U1 snRNP
- Trypanosomes do not contain U5 snRNP: each mRNA: 35 nt same at the 5' end
- 35 nt come from 140 nt SL RNA (200 copies in tandem array)
- SL RNA takes place of U1 RNA
  - Contains, like other snRNAs, trimethylguanosine cap at the 5' end
  - Exists as a RNP particle
  - Contains Sm core proteins
- Complementarity between SL RNA and U6 snRNA, which does not appear between U1 and U6 snRNAs
- Otherwise, splicing is almost identical to cis-splicing and requires U2, U4, and U6 snRNP
- What is the function of the 35 nt leader?
- No one knows--it doesn’t code for anything (amino acids)
Trans-Splicing of *Trypanosome* RNAs

Unlike other snRNPs, which can be repeatedly utilised, the SL snRNP is consumed during the trans-splicing reaction.

Unpaired strand RNA — SL RNA

Paired strand RNA — mRNA

Y-shaped molecule (no lariat)

Hybrid RNA

Unpaired strand RNA — SL RNA

Paired strand RNA — mRNA

Unlike other snRNPs, which can be repeatedly utilised, the SL snRNP is consumed during the trans-splicing reaction.
Trans splicing of polycistronic pre-mRNAs in *C. elegans*