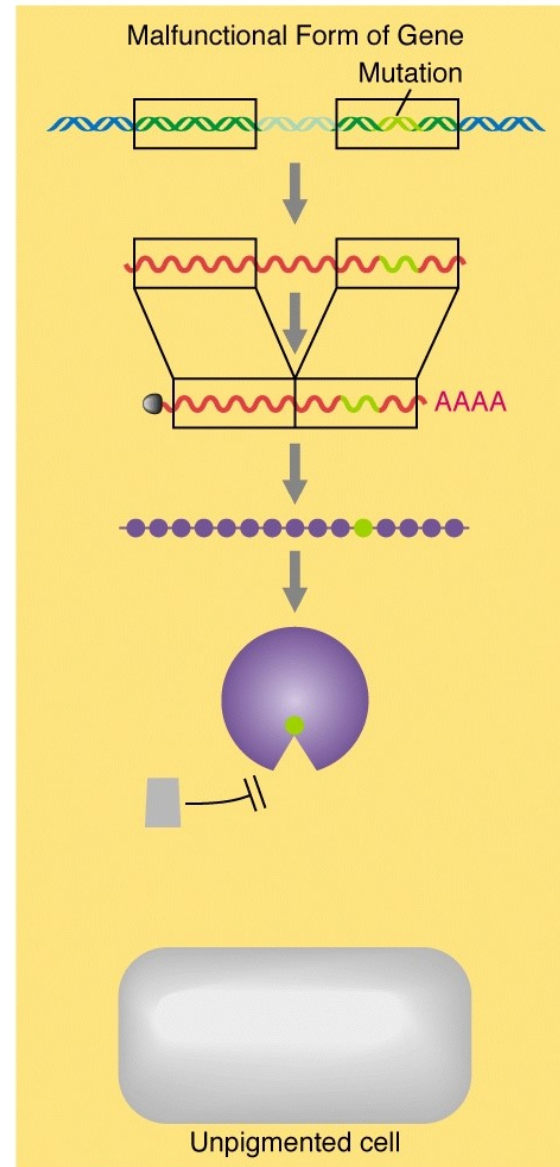
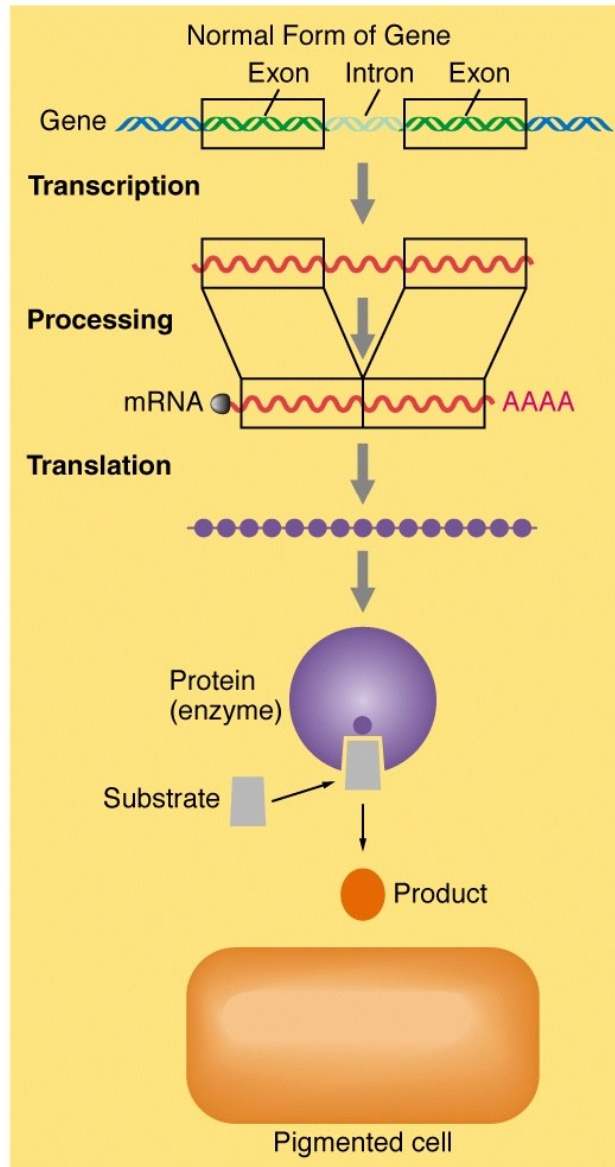
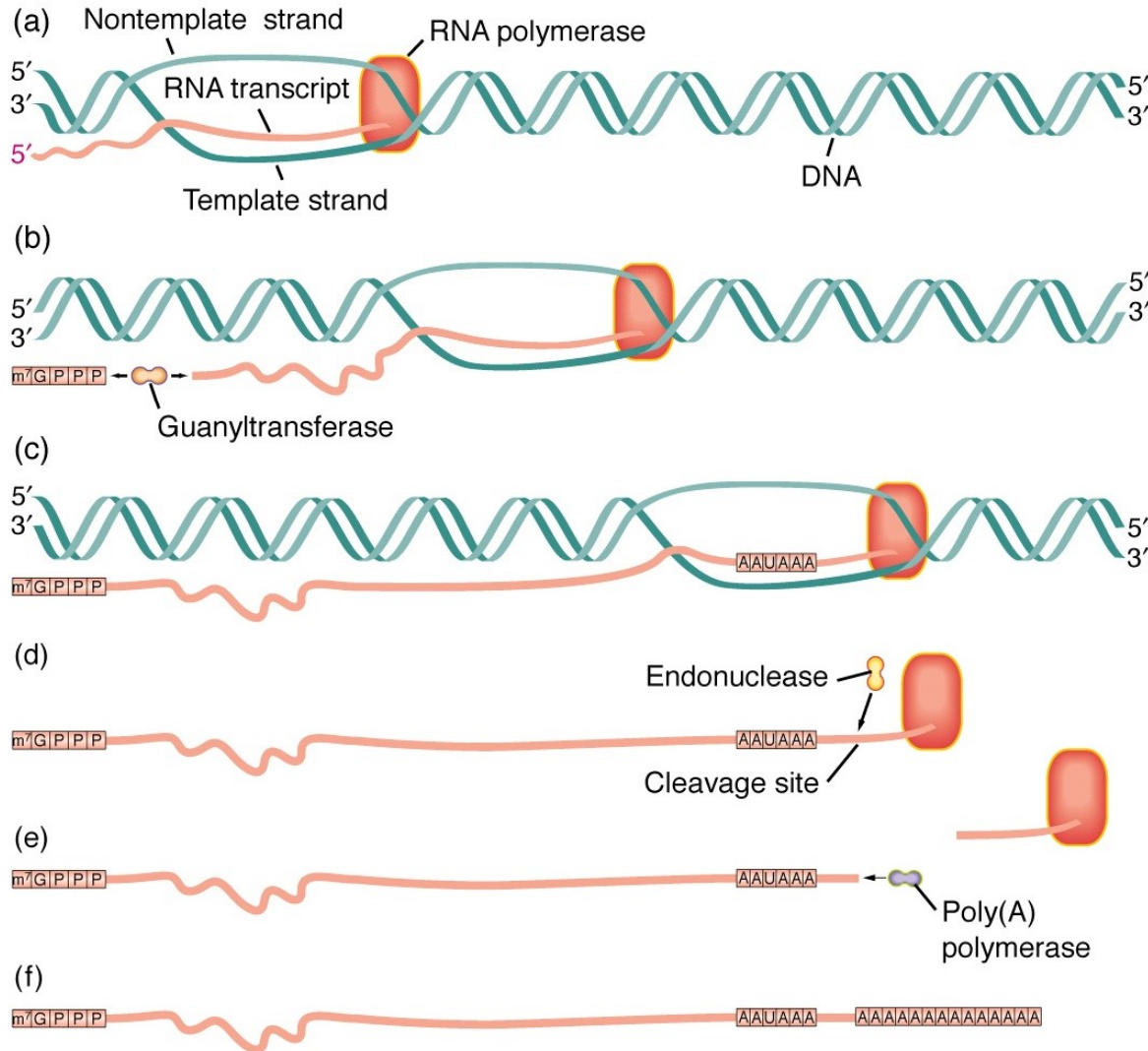


Splice site Identification

Gene Function & Gene Expression



Transcription Termination and mRNA 5'-End and 3'-End-Formation



Steps of Transcription

RNA processing during transcription

- 7mG capping 5' end
- 3' end formation

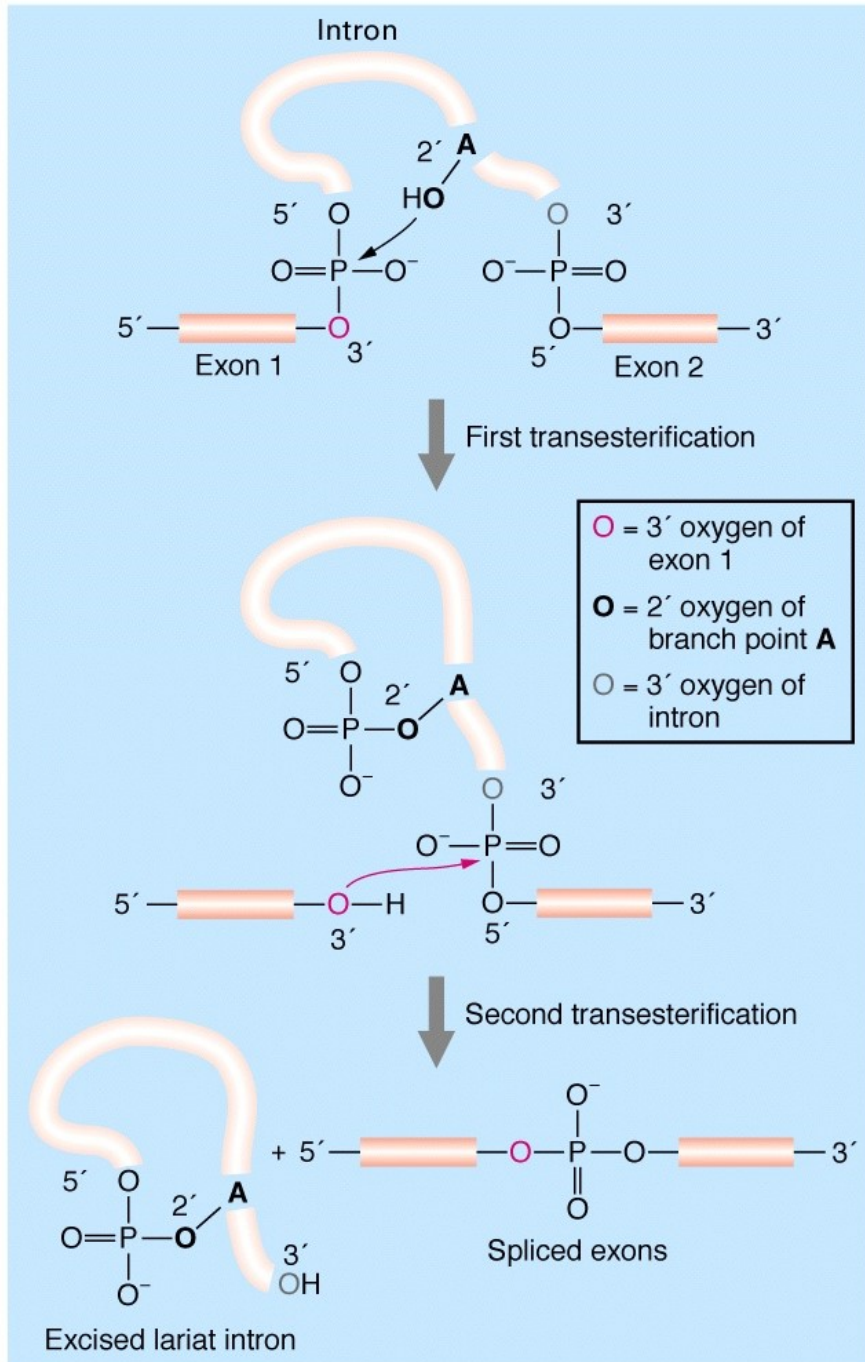
5' CAP

3' Poly A Tail

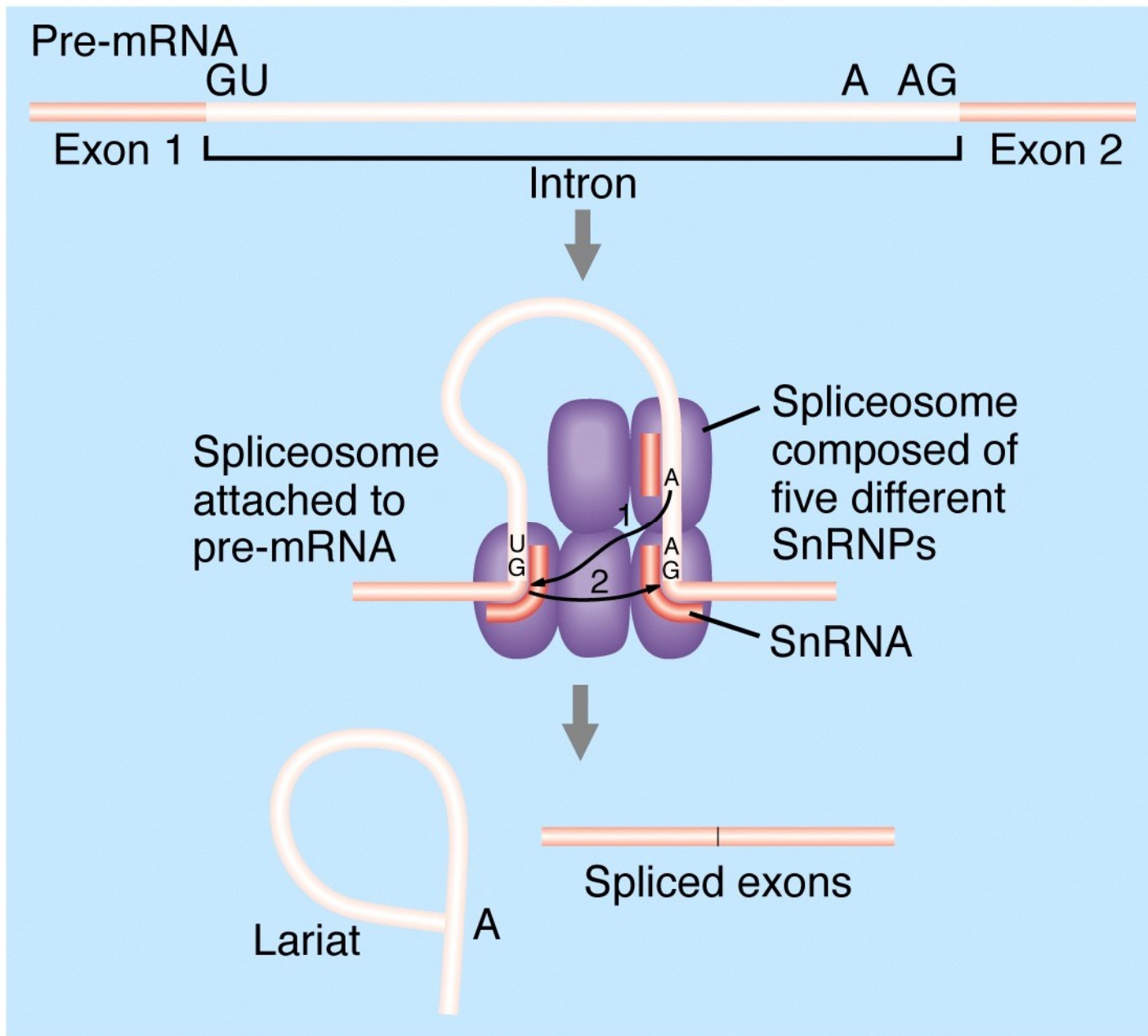
RNA Splicing

- RNA splicing reactions
 - Two transesterification reactions
 - first cleaves at 5' end of intron (5' splice site)
 - second cleaves at 3' end of intron (3' splice site)
- Self-Splicing RNAs (discovery of ribozymes)
 - Group I and Group II introns remove themselves
 - difference is in the first transesterification (see later)
 - Group I introns use a free G nucleotide to catalyze reaction
 - Group II splicing is similar reaction to that in pre-mRNA splicing
- pre-mRNA Splicing
 - catalyzed by “spliceosome”
 - multiple complexes of proteins and RNA (ribonucleoprotein particles called snRNPs
 - snRNAs called U RNAs (U1,2,4,5 and 6)
 - splicing proteins
 - snRNPs bind to pre-mRNA in a specific sequence to catalyze splicing

Transesterification Reactions in RNA Splicing



- Attack of the sugar-phosphate bond at 5' splice site
 - Group 1 = 3' OH of free G nucleotide
 - Group 2 and pre-mRNA = 2' OH group of branch point A nucleotide
- Attack of the sugar phosphate bond at the 3' splice site
 - 3' OH of nucleotide at 3' end of exon 1
- Intron released
 - Group 1 introns = linear
 - Group 2 and pre-mRNA introns = lariat
- Base pairing between U snRNAs and intron sequences brings intron sequences together and catalyzes the transesterification reactions



Identification of Intron Sequences important for Splicing

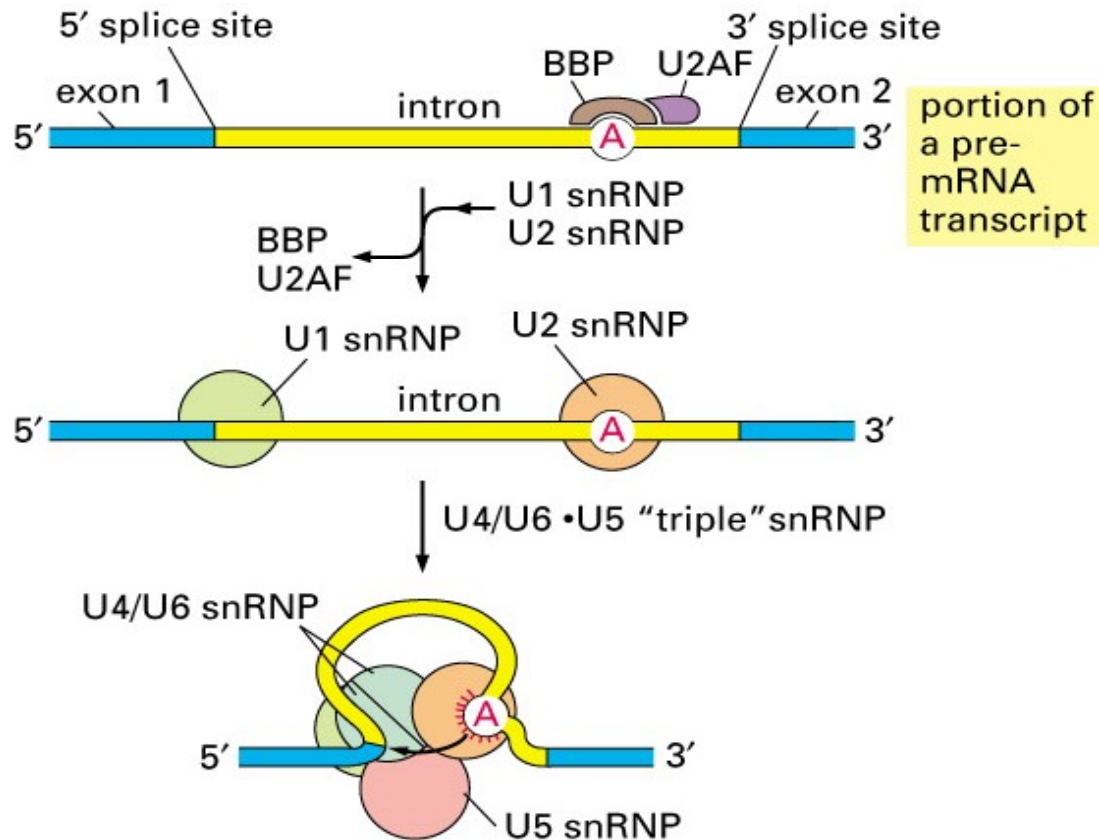
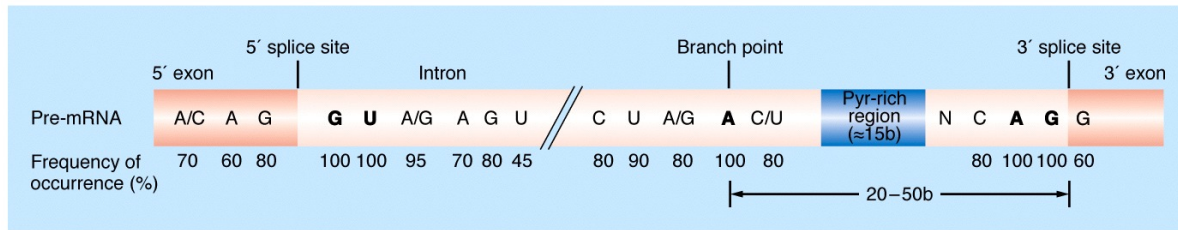
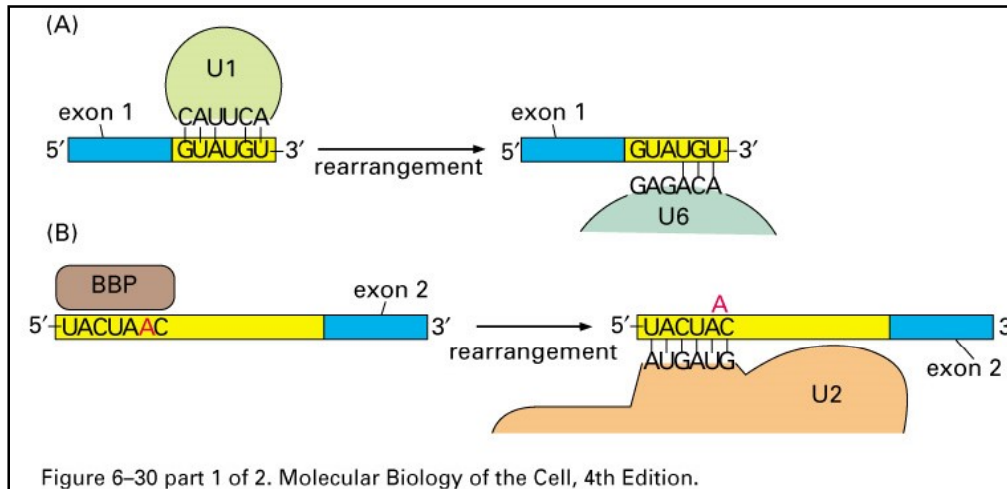


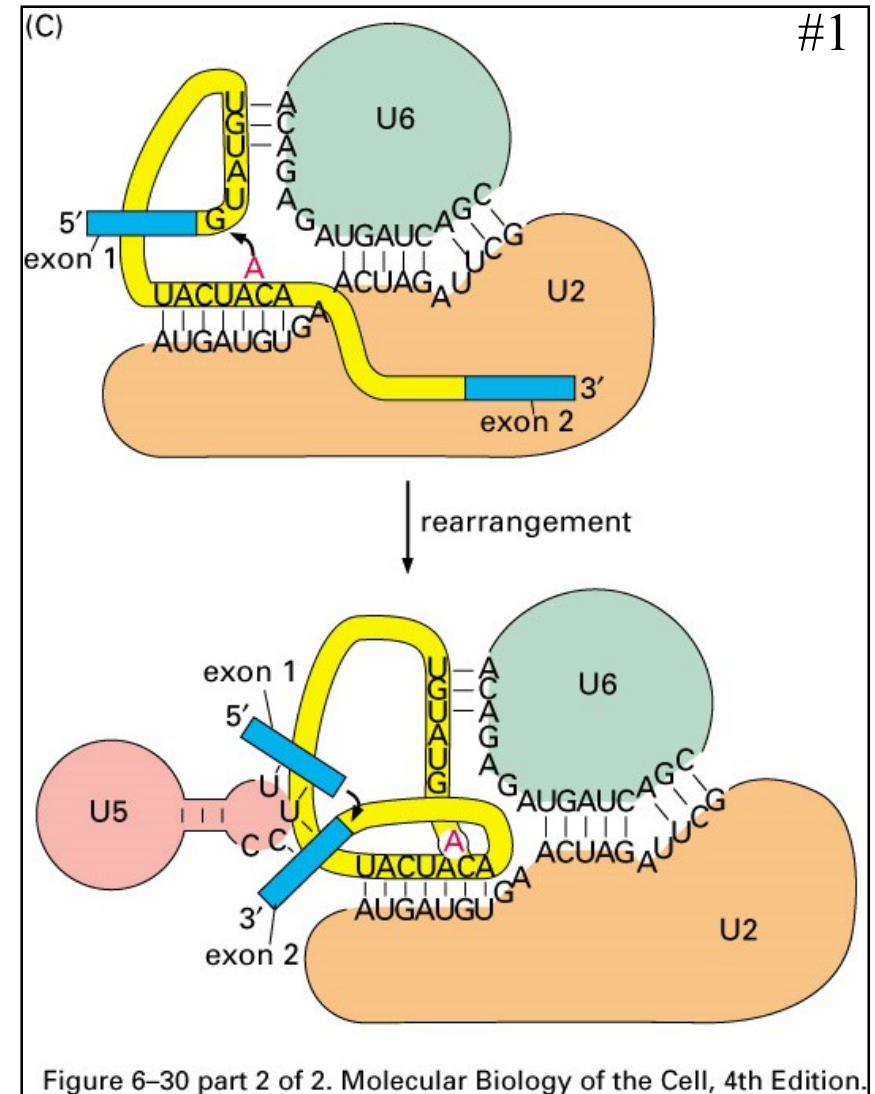
Figure 6-29 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

U snRNA:pre-mRNA base pairing

1. Identification of 5' splice site and Branch Point

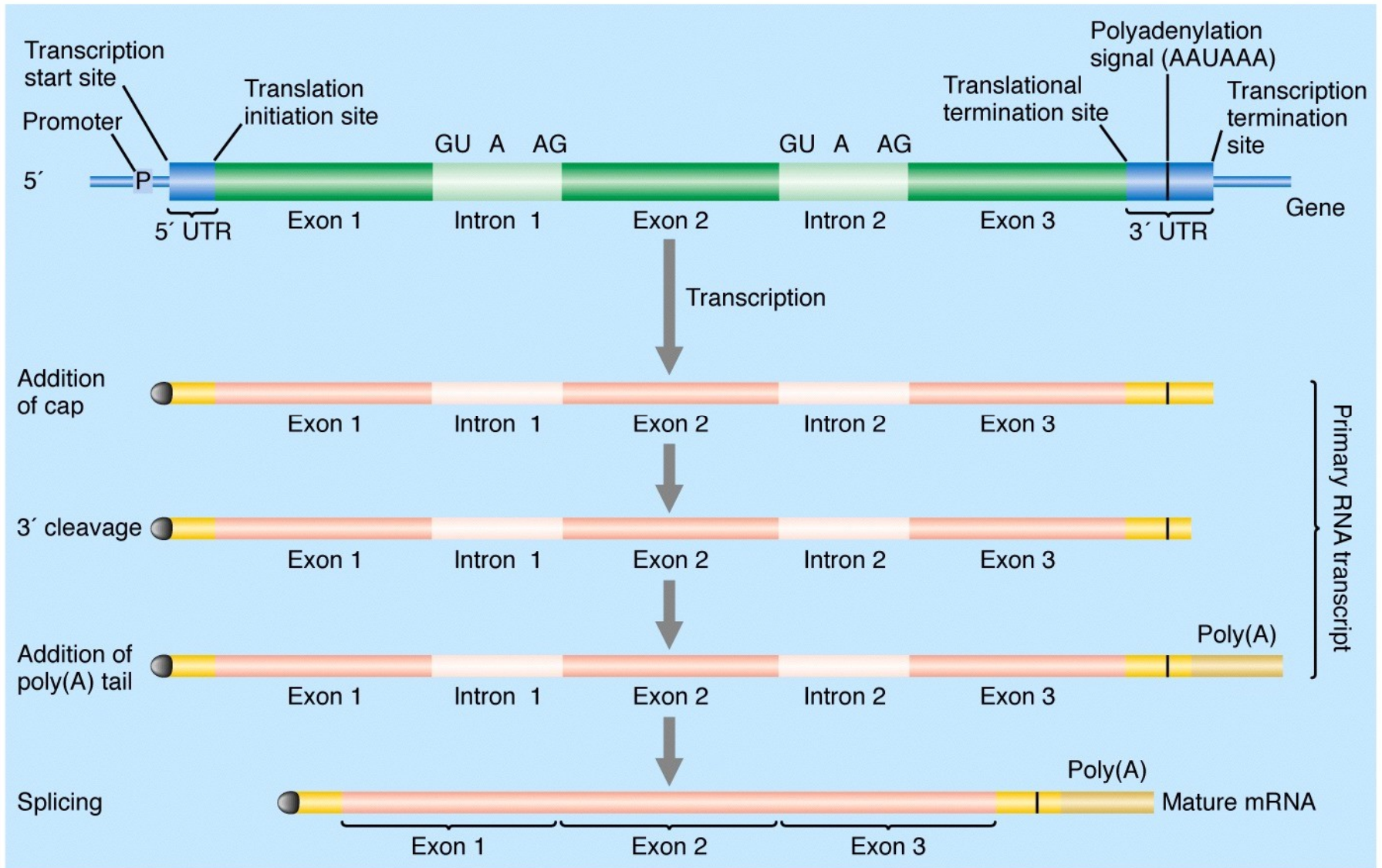


2. Positioning of A nucleotide for transesterification reaction



3. Positioning of 3' splice site near exon1 for transesterification reaction #2

Summary of terms and processes



Splice prediction tools

You can find here a collection of various splice prediction tools. You might want to check out as well our documents and guideline section as it contains reports about the use of those tools. You are also welcome to propose to add tools to below list, just contact us!

[Fruit Fly Splice Predictor](#)

Fruit fly and human splice predictor.
Directly callable from our sequencing software Gensearch.

10203

[Human Splicing Finder](#)

In order to better understand intronic and exonic mutations leading to splicing defects, we decided to create the Human Splicing Finder website. This tool is aimed to help study of the pre-mRNA splicing.

25945

To calculate the consensus values of potential splice sites and search for branch points, new algorithms were developed. Furthermore, we have integrated all available matrices to identify exonic and intronic motifs, as well as new matrices to identify hnRNP A1, Tra2- β and 9G8.

[RegRNA: A Regulatory RNA Motifs and Elements Finder](#)

Integrated web server, namely RegRNA, to identify the homologs of Regulatory RNA motifs and elements against an input mRNA sequence. Both sequence homologs or structural homologs of regulatory RNA motifs can be identified. The regulatory RNA motifs supported in RegRNA are categorized into several classes:

- Motifs in mRNA 5'-UTR and 3'-UTR.
- Motifs involved in mRNA splicing.
- Motifs involved in transcriptional regulation.
- Other motifs in mRNA, such as riboswitches.
- Prediction of the splice sites, such as splicing donor/acceptor sites.

9361

Gene prediction methods

- Automated sequencing of genomes require automated gene assignment
- Includes detection of open reading frames (ORFs)
- Identification of the introns and exons
- Gene prediction a very difficult problem in pattern recognition
- Coding regions generally do not have conserved sequences
- Much progress made with prokaryotic gene prediction
- Eukaryotic genes more difficult to predict correctly

Gene prediction methods

Ab initio methods

- Predict genes on given sequence alone

Uses gene signals

Start/stop codon

Intron splice sites

Transcription factor binding sites ribosomal binding sites

Poly-A sites

Codon demand multiple of three nucleotides

- Gene content

Nucleotide composition – use HMMs

Homology based methods

Matches to known genes

Matches to cDNA

Consensus based

Uses output from more than one program

Prokaryotic gene structure

- ATG (GTG or TTG less frequent) is start codon
- Ribosome binding site (Shine-Dalgarno sequence) complementary to 16S rRNA of ribosome
- AGGAGGT
- TAG stop codon
- Transcription termination site (ρ -independent termination)
- Stem-loop secondary structure followed by string of Ts
- Translate sequence into 6 reading frames
- Stop codon randomly every 20 codons
- Look for frame longer than 30 codons (normally 50-60 codons)

- Presence of start codon and Shine-Dalgarno sequence
- Translate putative ORF into protein, and search databases
- Non-randomness of 3rd base of codon, more frequently G/C
- Plotting wobble base GC% can identify ORFs
- 3rd base also repeats, thus repetition gives clue on gene location

Tools to predict genes

- **GeneMark** (<http://exon.gatech.edu/genemark/>) Trained on complete microbial genomes Most closely related organism used for predictions
- **Glimmer** (Gene Locator and Interpolation Markov Model) (<http://www.cbcb.umd.edu/software/glimmer/>)
- **FGENESB** (<http://linux1.softberry.com/>) 5th-order HMM Trained with bacterial sequences Linear discriminant analysis (LDA)
- **RBSFinder** (<ftp://ftp.tigr.org>) Takes output from Glimmer and searches for S-D sequences close to start sites