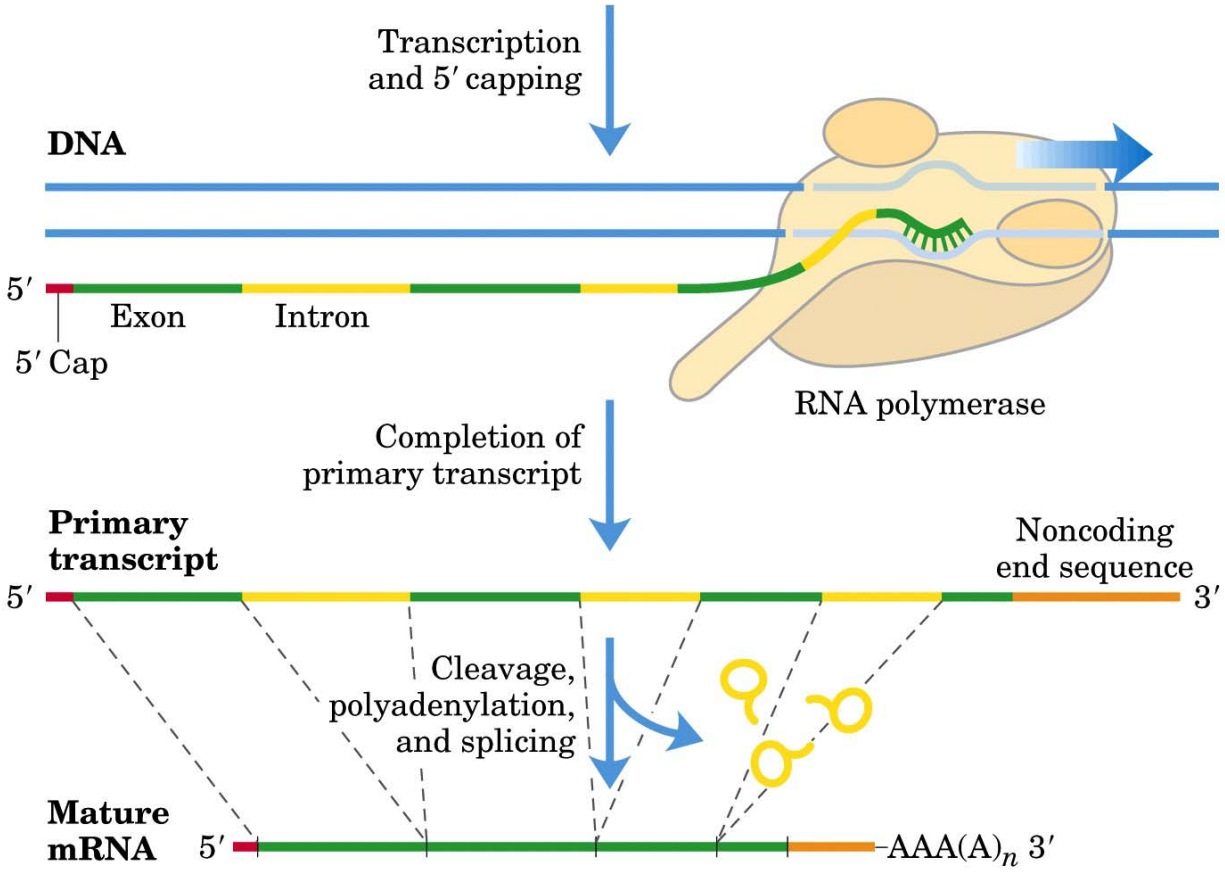
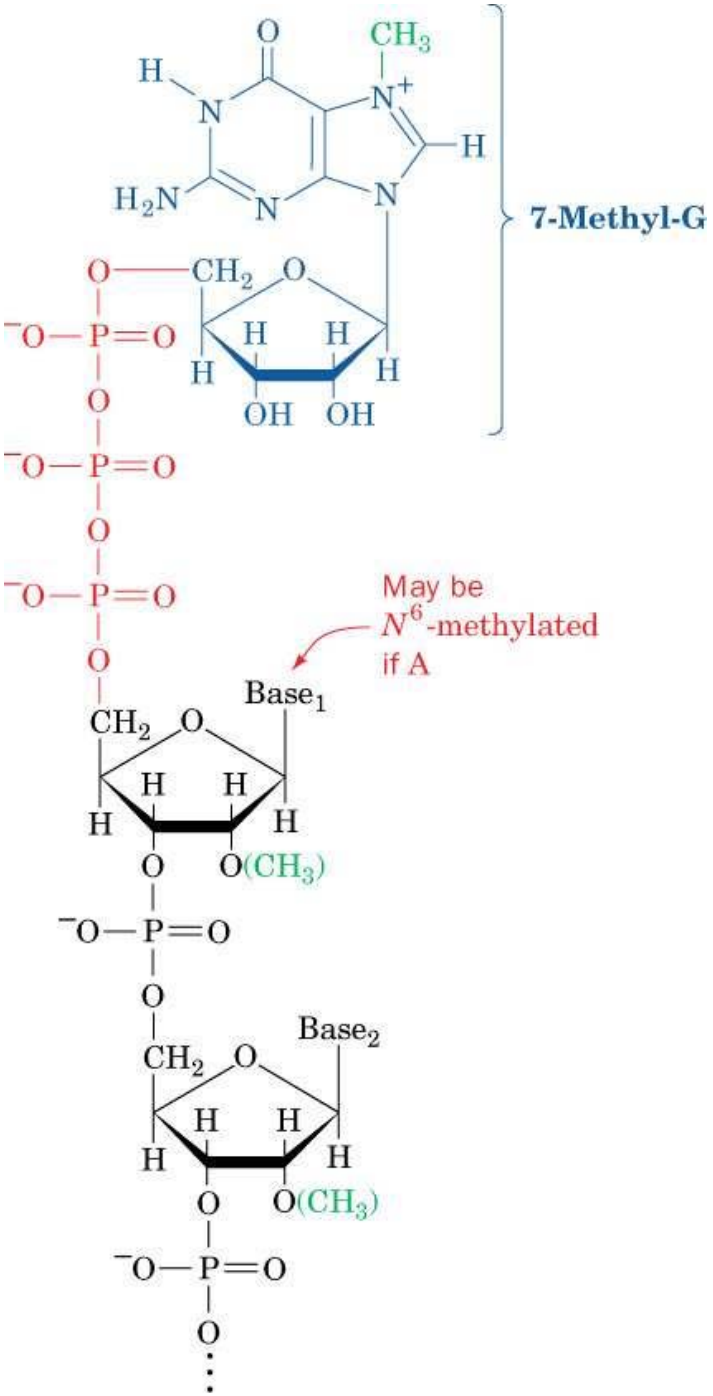
RNA Processing

- Voet-Voet 1254
- RNA molecules
  - Post-transcriptional processing
    - Most bacterial RNA molecules
    - All eukaryotic RNA molecules
  - Enzymes
    - Catalytic RNA
      - Ribozymes
    - Catalytic protein

Summary RNA processing of eukaryotic mRNA

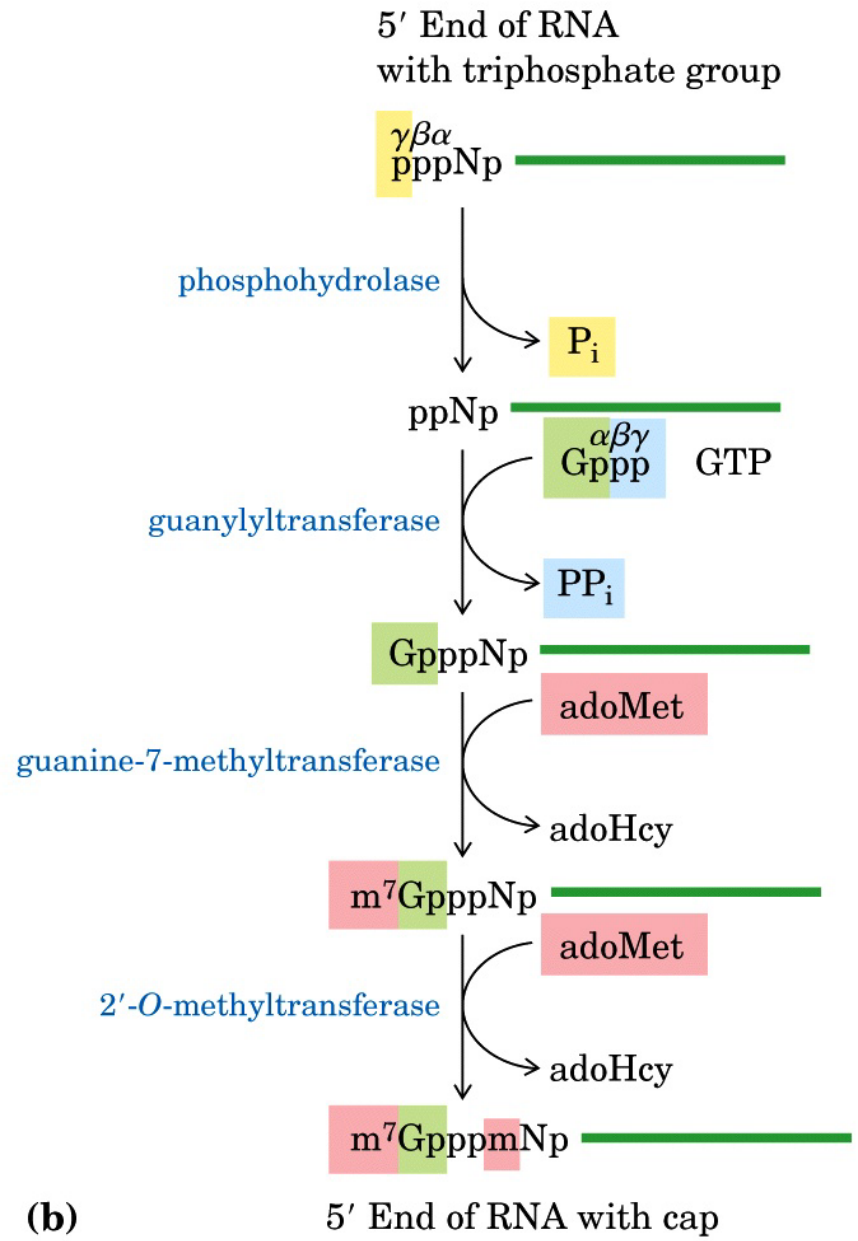


- Primary transcript
  - Newly synthesized RNA
- 5’ end
  - Capping,
  - 5’ cap
- 3’ end
  - Cleaved
  - Polyadenylation,
    - 80-250 adenylate residues added
  - Poly (A) tail
- Splicing
  - Introns removed
  - Exons joined



Capping

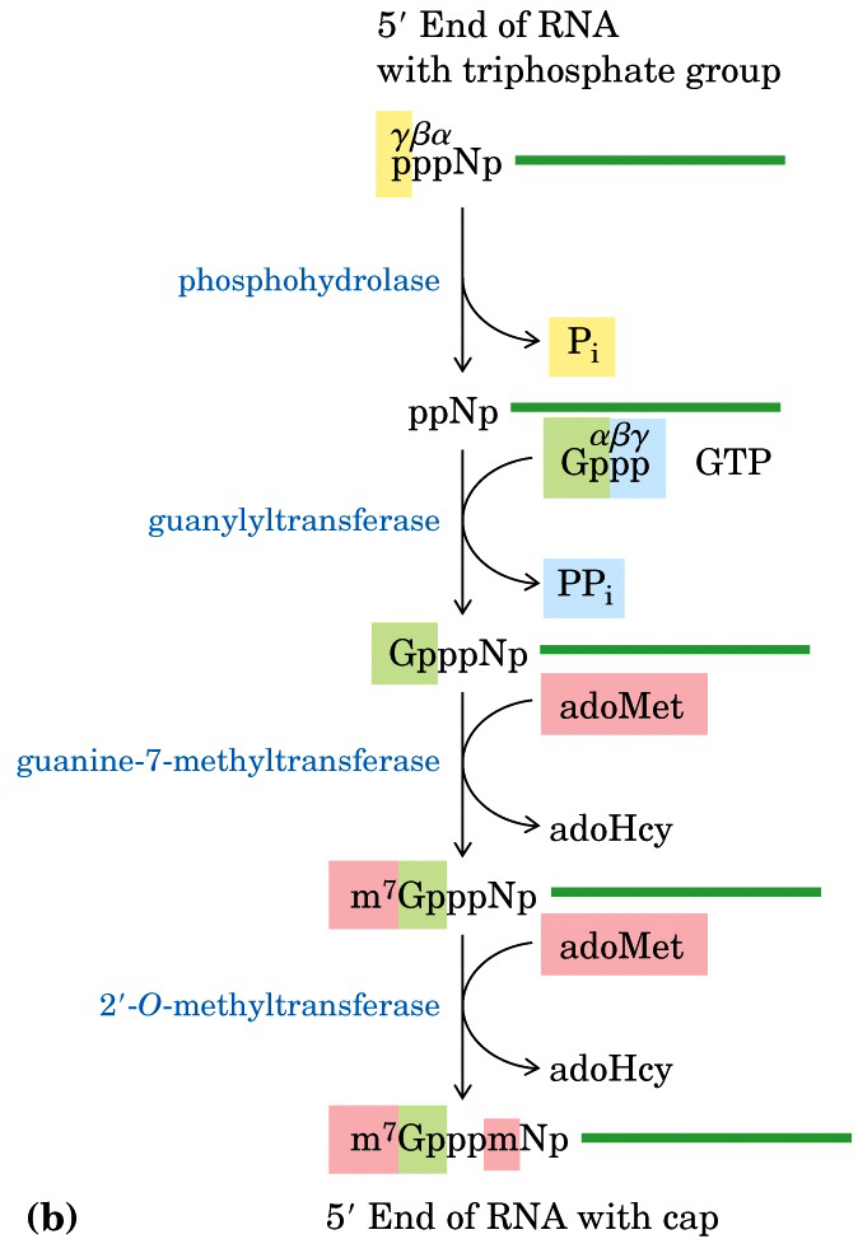
- Most eukaryotic mRNAs have 5’ cap
  - 7-methylguanosine linked to the 5’-terminal residue
  - 5’—5’ triphosphate bridge
- A cap may be $O^{2'}$ methylated
  - at the transcripts leading nucleoside, **cap-1**
    - Predominant cap in multicellular organisms
  - At the first two nucleosides, **cap-2**
  - At neither, **cap-0**
    - The predominant cap in unicellular eukaryotes
- Has role in translation
  - Initiation



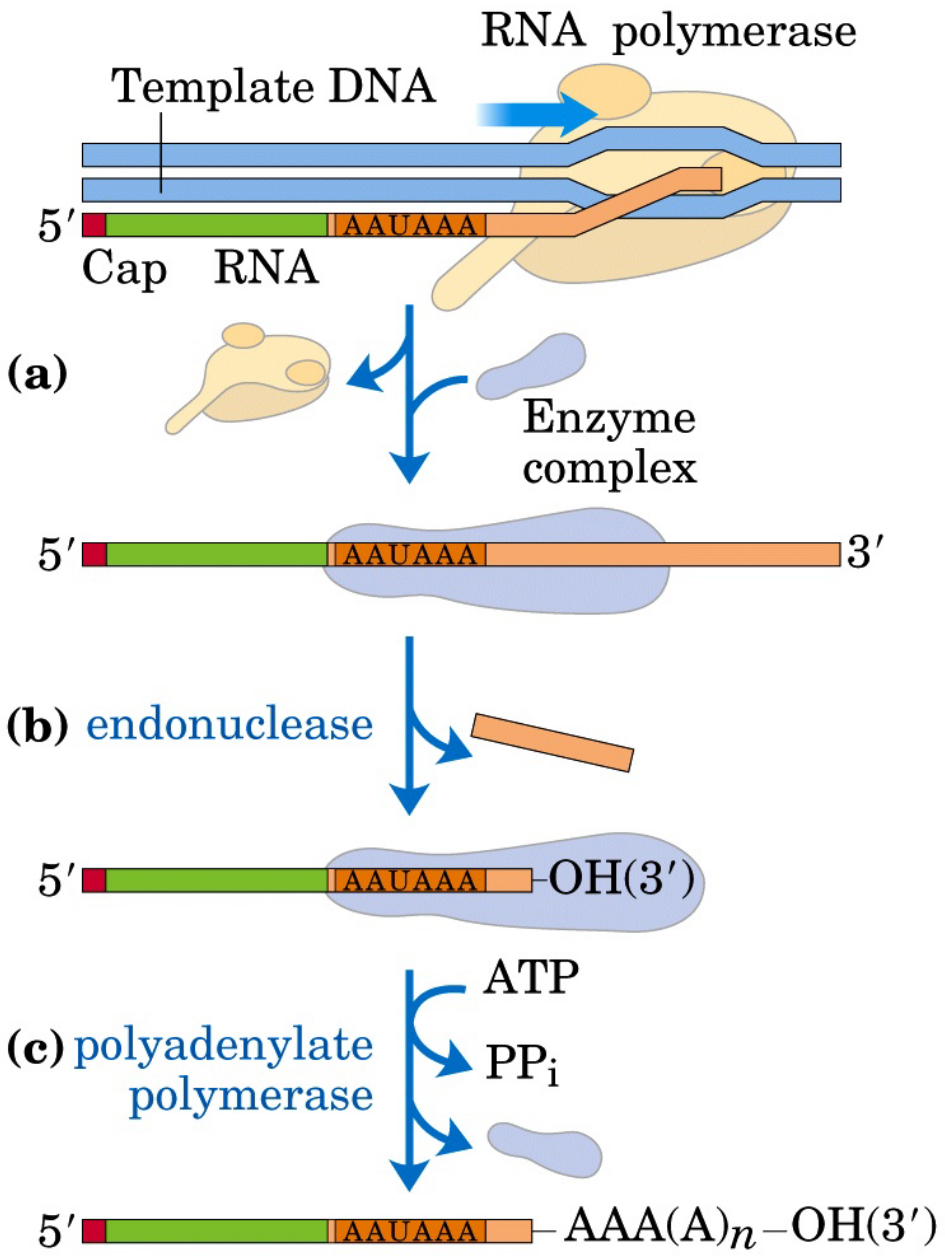
Steps in capping

- Cap added when transcript is about 30 nucleotides long
- 1. Removal of the leading phosphate group from the mRNA's 5' terminal triphosphate group
  - Phosphohydrolase (also called RNA triphosphatase)
- 2. Capping enzme
  - A guanylyltransferase

Steps in capping



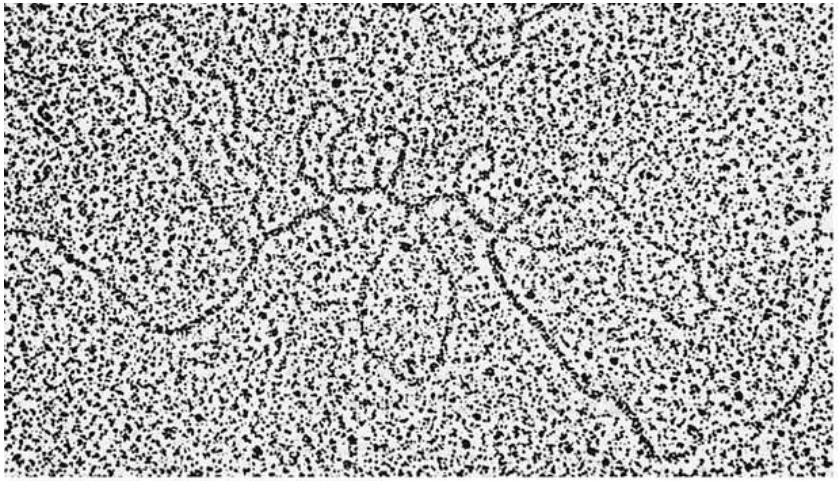
- 3. Methylation of guanine
  - Guanosine-7-methyltransferase
    - Uses S-adenosylmethionine (SAM or adoMet), product is S-adenosylhomocysteine (adoHcy)
- 4. 2'-O-methyltransferase
  - SAM (cap-1, cap-2, cap-0)
- Both the capping enzyme and guanosine-7-methyltransferase bind to RNA-Pol II's phosphorylated CTD (c-terminal domain)



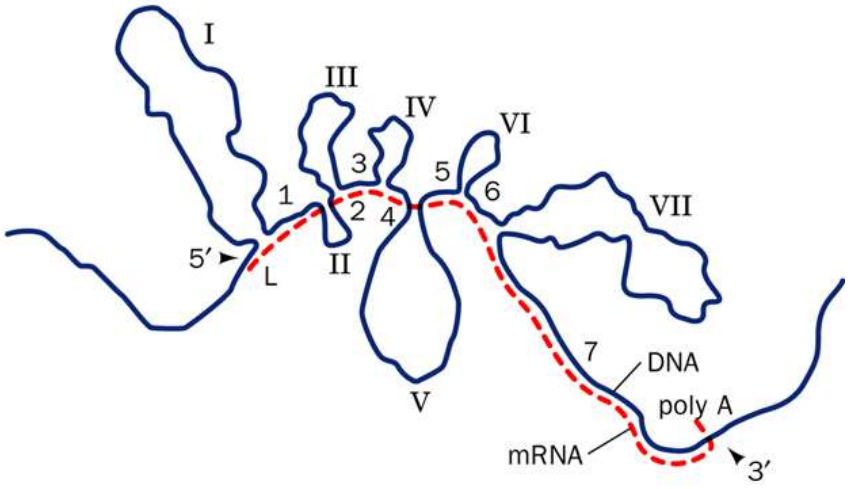
Tailing

- Poly(A) tails added to primary transcripts of mRNA
  - Eukaryotic mRNA invariably mono-cistronic
- 1. Transcript extends beyond site of poly(A) addition
  - Large complex binds
  - Endonuclease component cleaves 15 to 25 nucleotides on 3' side of
    - AAUAAA
- 2. Poly(A) Polymerase (PAP)
  - (Polyadenylate polymerase in diagram)
  - No template
  - Needs a primer
  - Adds 80-250 A

RNA-DNA hybridization, demonstrates introns

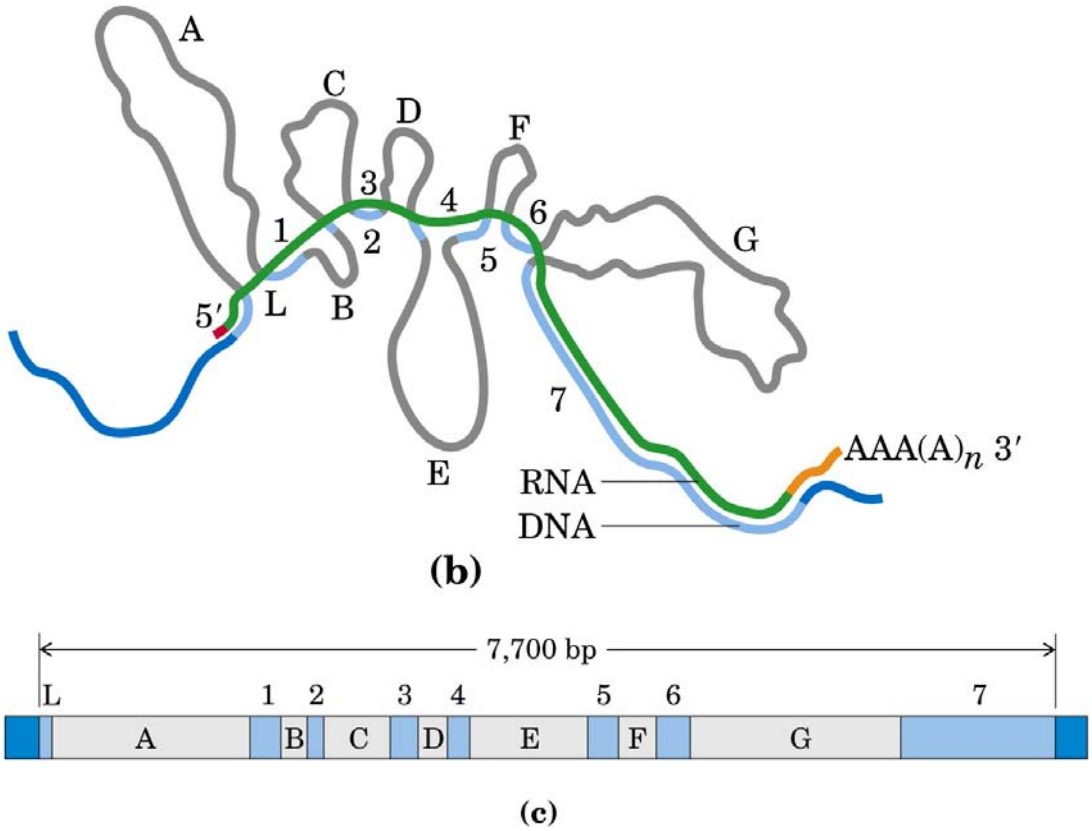


From Chambon, P., Sci. Am. 244(5), 61 (1981).]



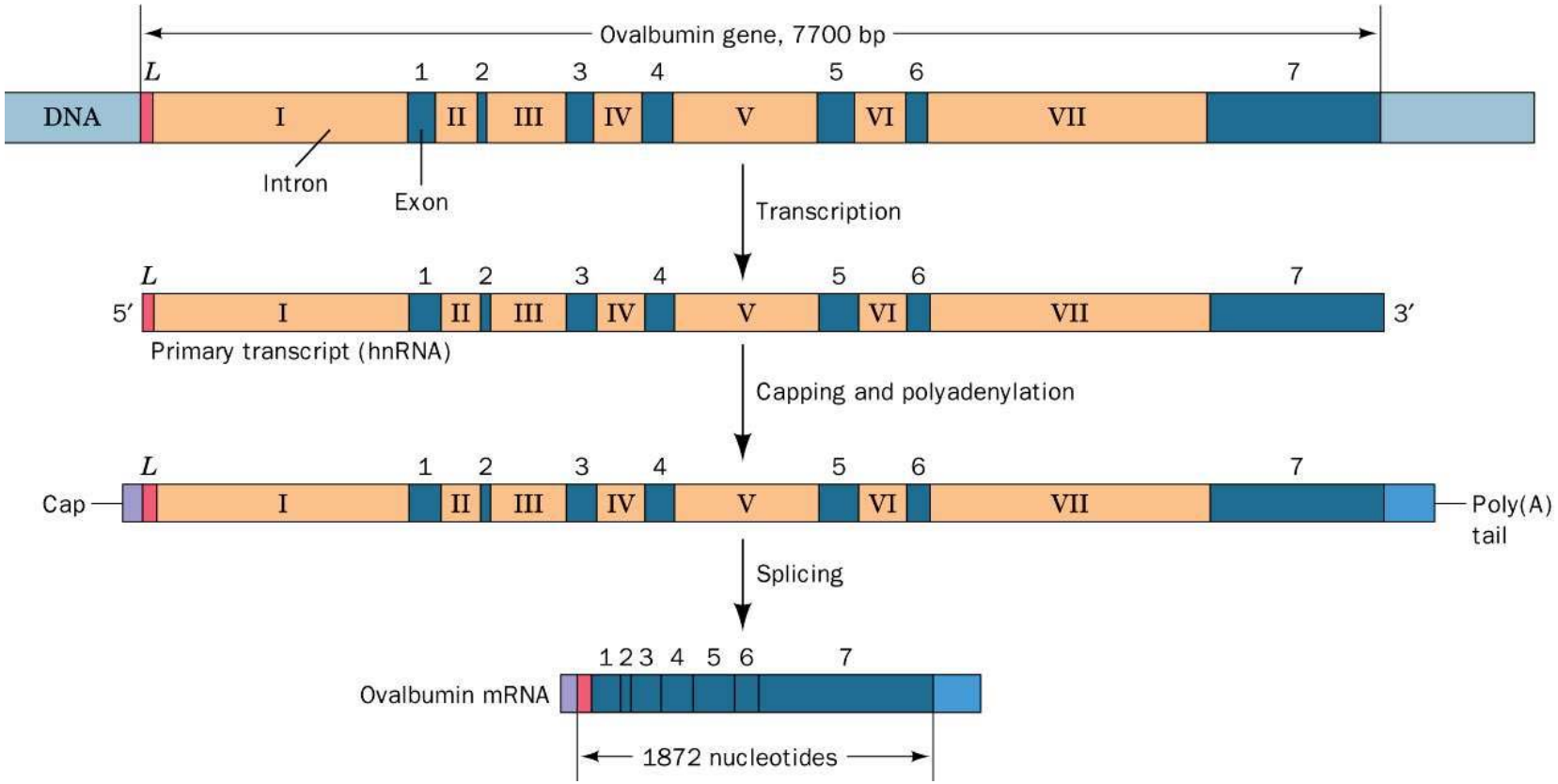
- Chicken ovalbumin gene
- EM view (a), diagram (b)
  - Hybrid between anti-sense strand of ovalbumin gene and its corresponding mRNA

Diagram



- A - G, Introns
- L, Signal sequence, ovalbumin targeted for secretion
- 1 - 7, Exons

Steps to ovalbumin mRNA



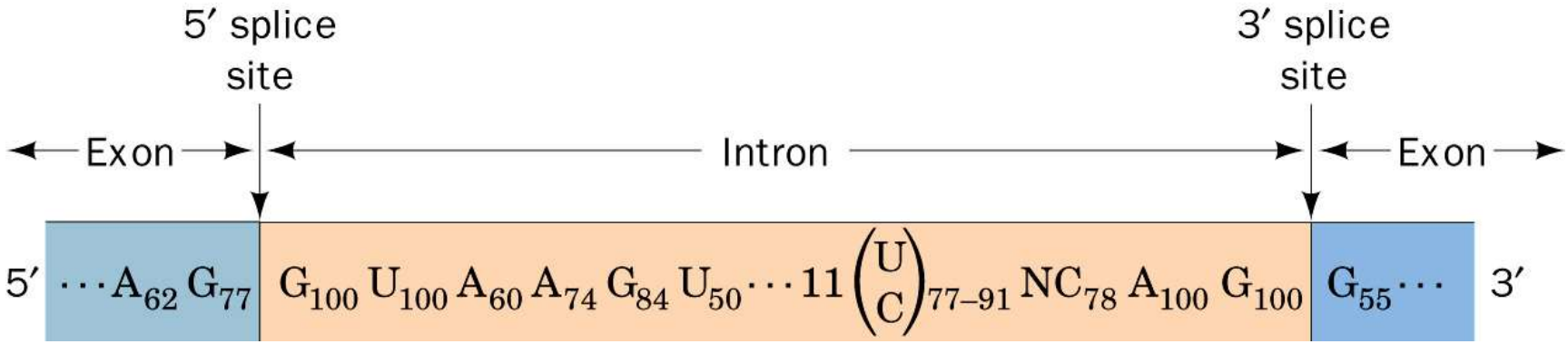
Genes/introns

- Eukaryotes
  - Vertebrates
    - Most genes contain **introns**
      - The aggregate length of introns averages 4-10 times that of the exons
    - **Exons** (expressed sequences) have lengths up to 17,106 nt (in the coding region of the 29,926-residue muscle protein **titin**, the largest known single-chain protein)
      - Most exons less than 300 nt and average 150 nt in humans
  - Other eukaryotes
    - Variable occurrence of genes with introns
    - *Saccharomyces cerevisiae*
      - Many genes do not have introns
- Prokaryotes
  - Introns found in a few bacterial and archaebacterial genes

Eukaryotic RNA

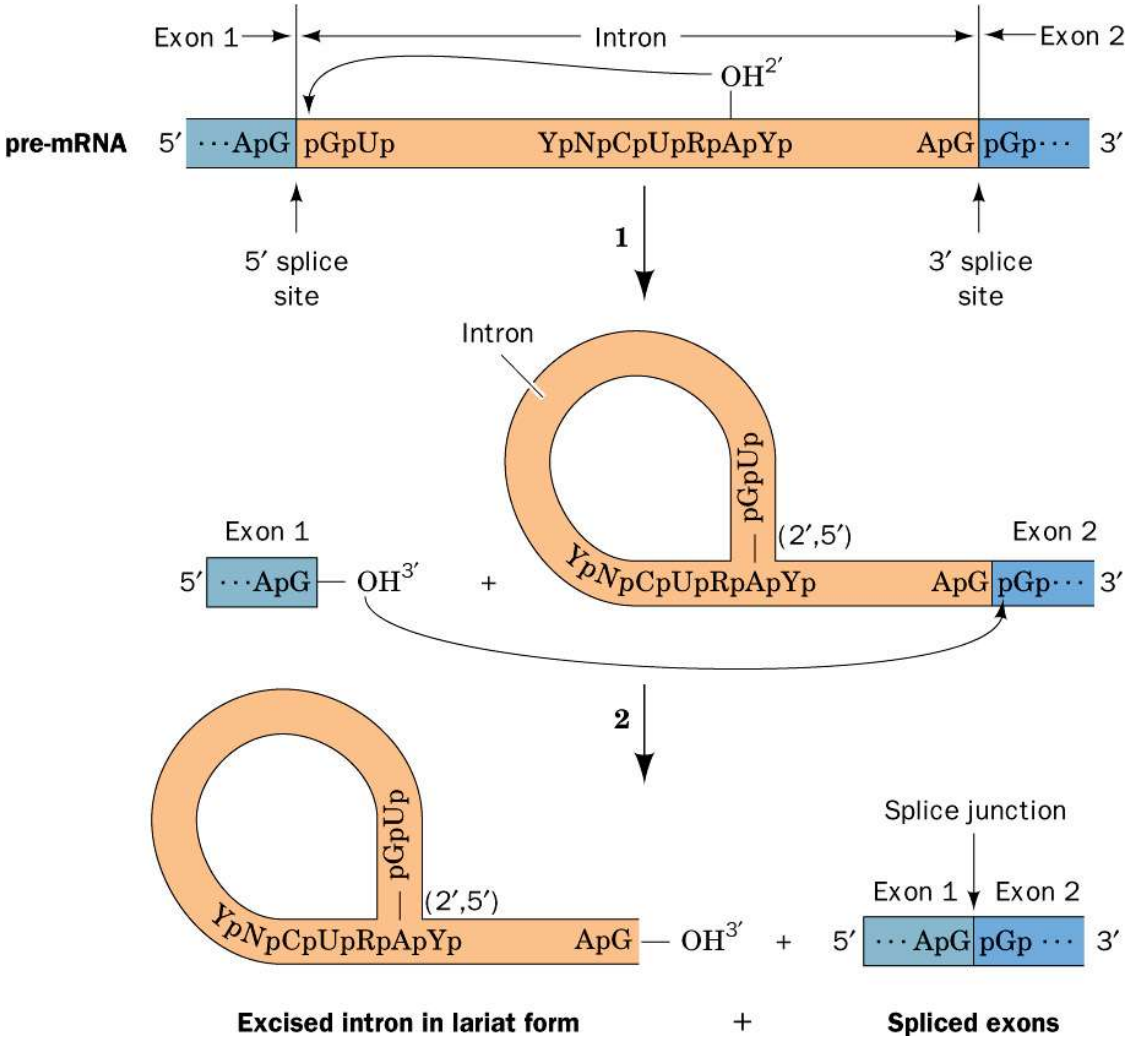
- Exons
  - Most less than 300 nucleotides long
  - Many 100 to 200 nucleotides long
    - Encode stretches of 30 to 60 amino acids
- Introns
  - Vary from 50 to 20,000 nucleotides
  - Human genes typically have more DNA devoted to introns then to exons
  - Genes can have dozens of introns

Consensus sequence



- The consensus sequence at the exon–intron junctions of vertebrate pre-mRNAs.
  - Subscripts in diagram- % of pre-mRNA with that specified base
  - Invariant GU at the intron's 5' boundary
  - Invariant AG at the intron's 3' boundary

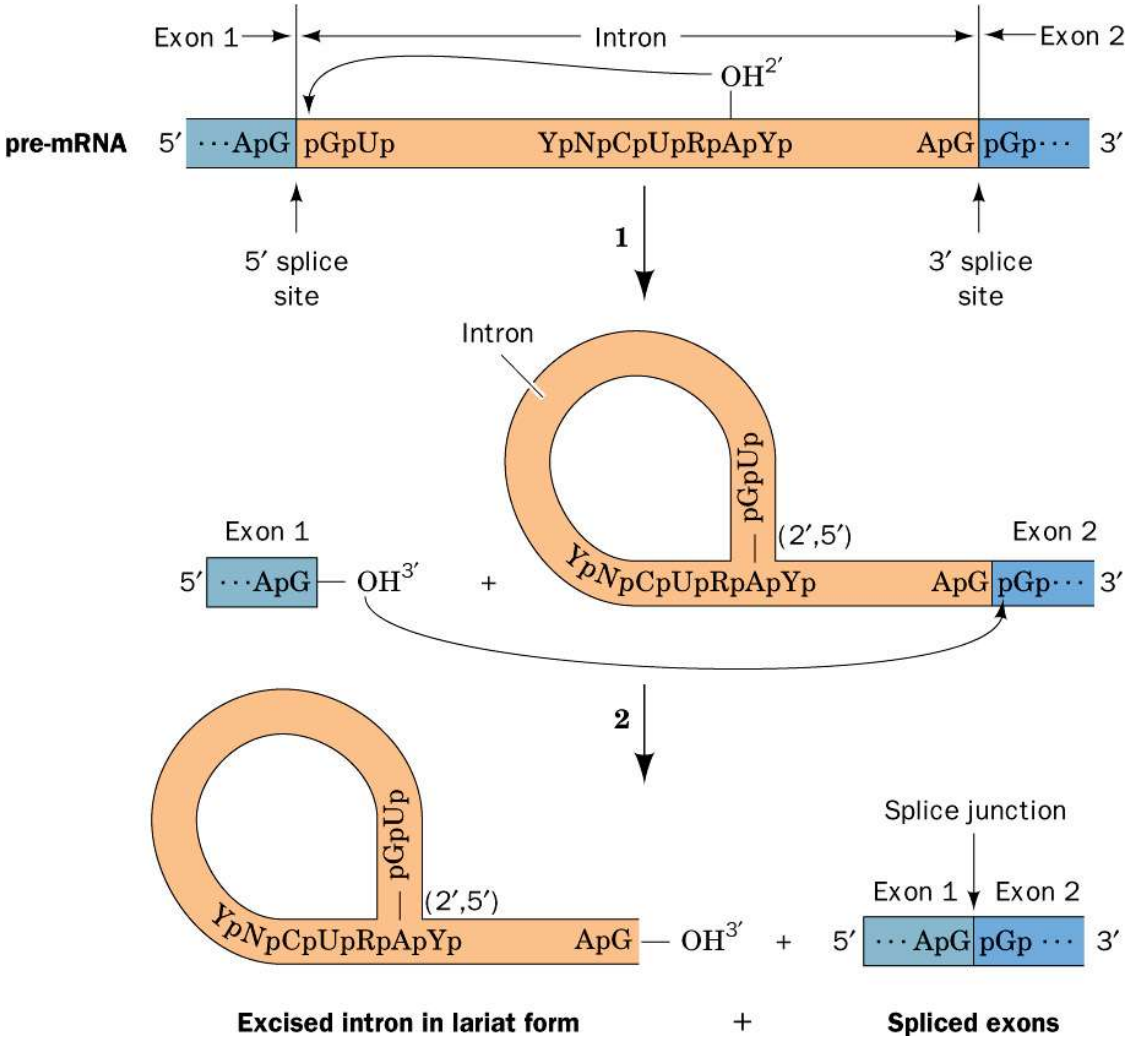
Exon splicing



- Two transesterification reactions

- (1) Formation of a 2', 5' phosphodiester bond between an intron adenosine residue and its 5'-terminal phosphate group with the concomitant liberation of the 5' exon's 3'-OH group.
  - Lariat structure
  - YNCURAP

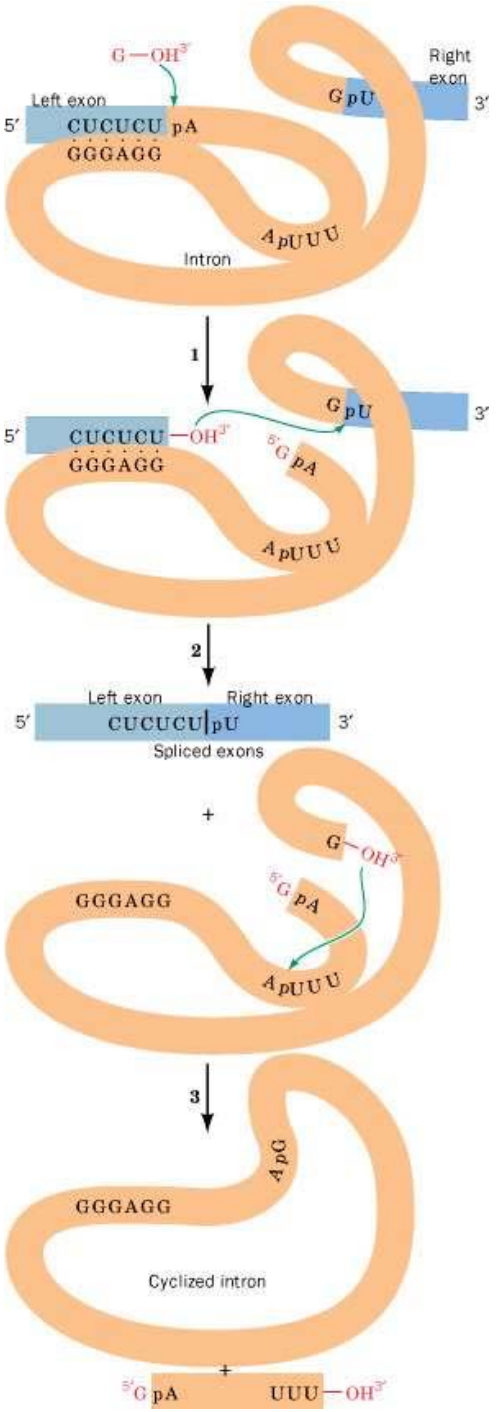
Exon splicing



- Two transesterification reactions
- (2) 3'-OH group of the 5' exon forms a phosphodiester bond with the 5'-terminal phosphate of the 3' exon
  - Spliced product
  - Lariat
- Exonic sequence enhancers (ESEs)
  - Help identify spice sites

Types of introns

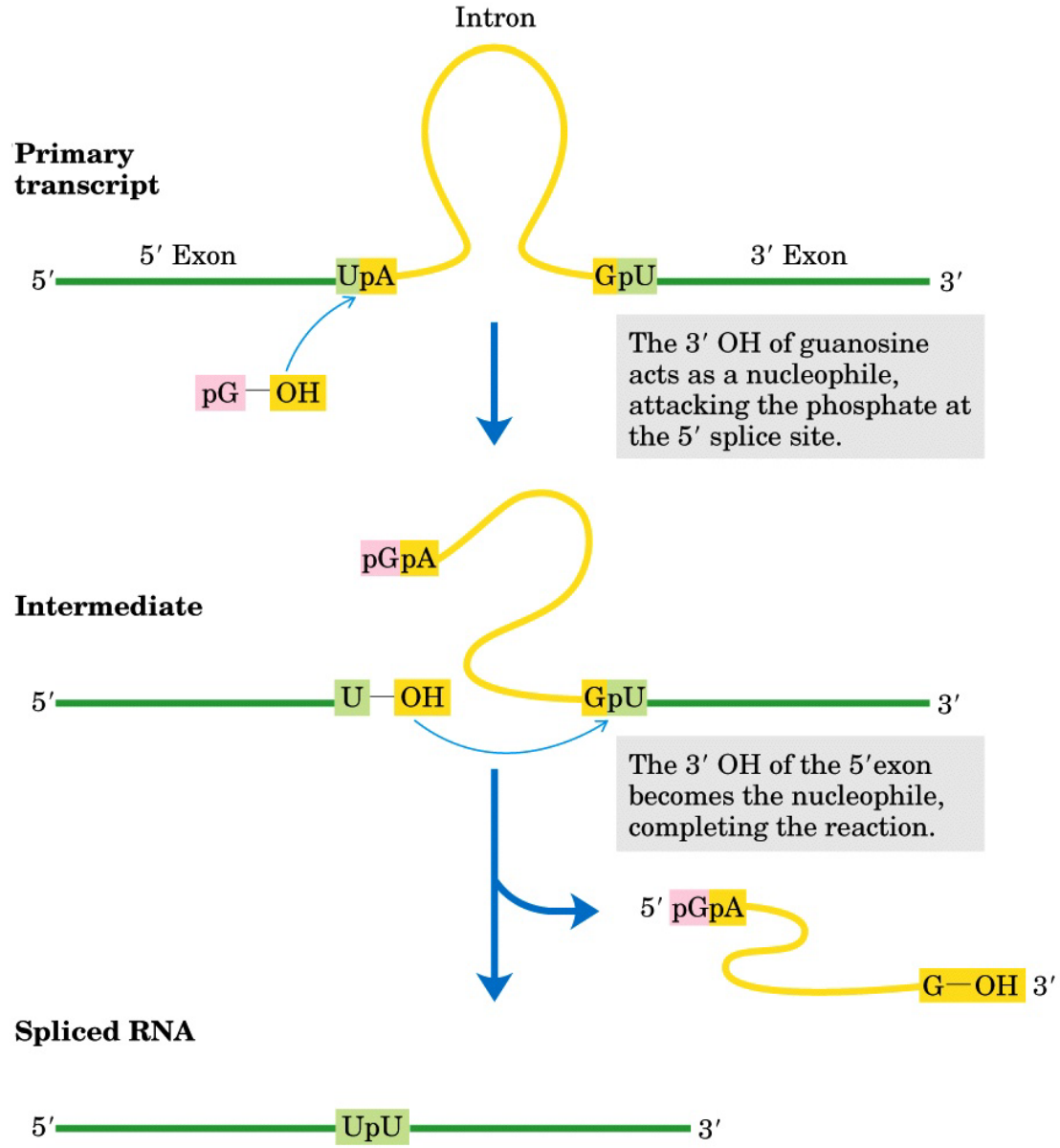
Intron Type	Where Found
GU-AG introns	Eukaryotic nuclear pre-mRNA (spliceosome)
AU-AC introns	Eukaryotic nuclear pre-mRNA
Group I	Eukaryotic nuclear pre-mRNA, Organelle RNAs, a few bacterial RNAs (self-splicing, not found in vertebrates)
Group II	Organelle RNAs, a few prokaryotic RNAs (self-splicing)
Group III	Organelle RNAs
Twintrons (composites of two and/or more goup II or III introns)	Organelle RNAs
Pre-tRNA introns	Eukaryotic nuclear pre-tRNAs
Archaeal introns	Various RNAs



Self splicing

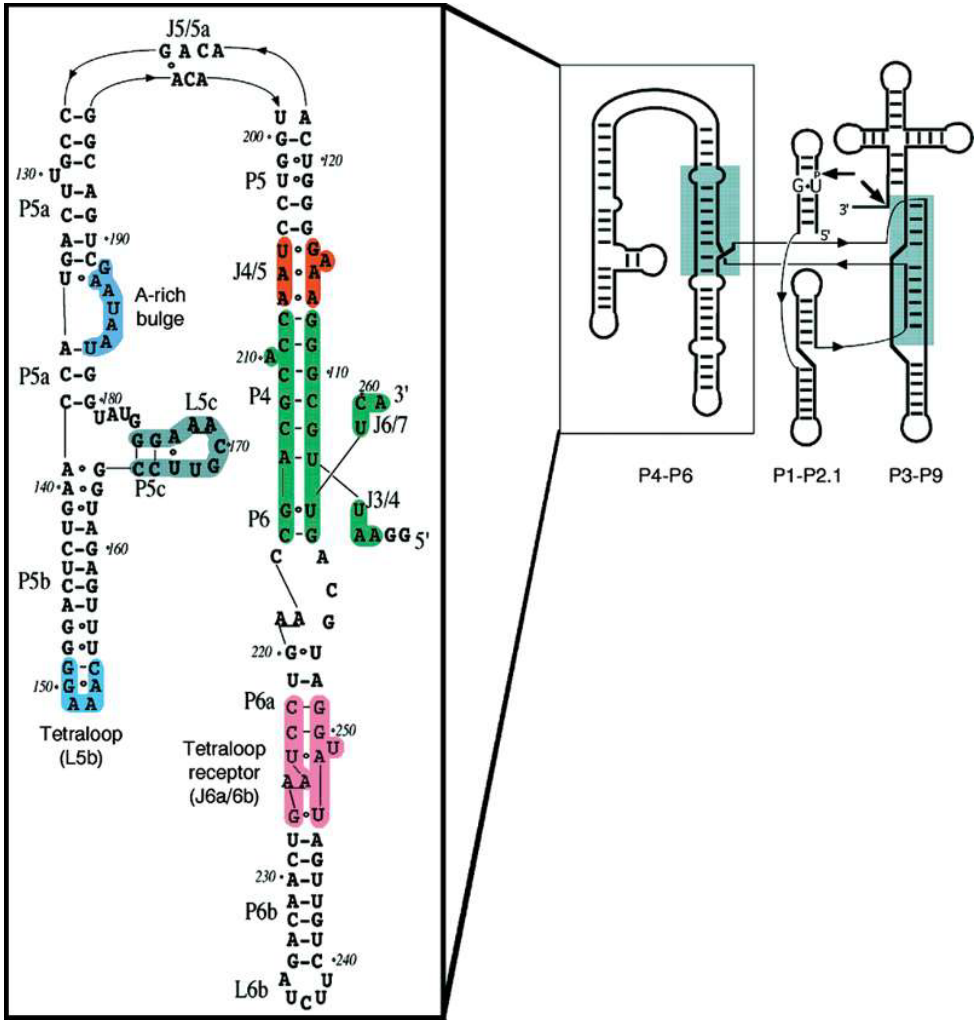
- Group I introns
  - Nuclei, mitochondria and chloroplasts of diverse eukaryotes (not in vertebrates)
- *Tetrahymena thermophila* pre-rRNA, (Tom Cech 1981)
  - Isolated pre-rRNA incubated with free guanosine or guansine nucleotides, spiced without any protein
  - Ribozyme
    - Catalytic RNA

Splicing of Group I introns



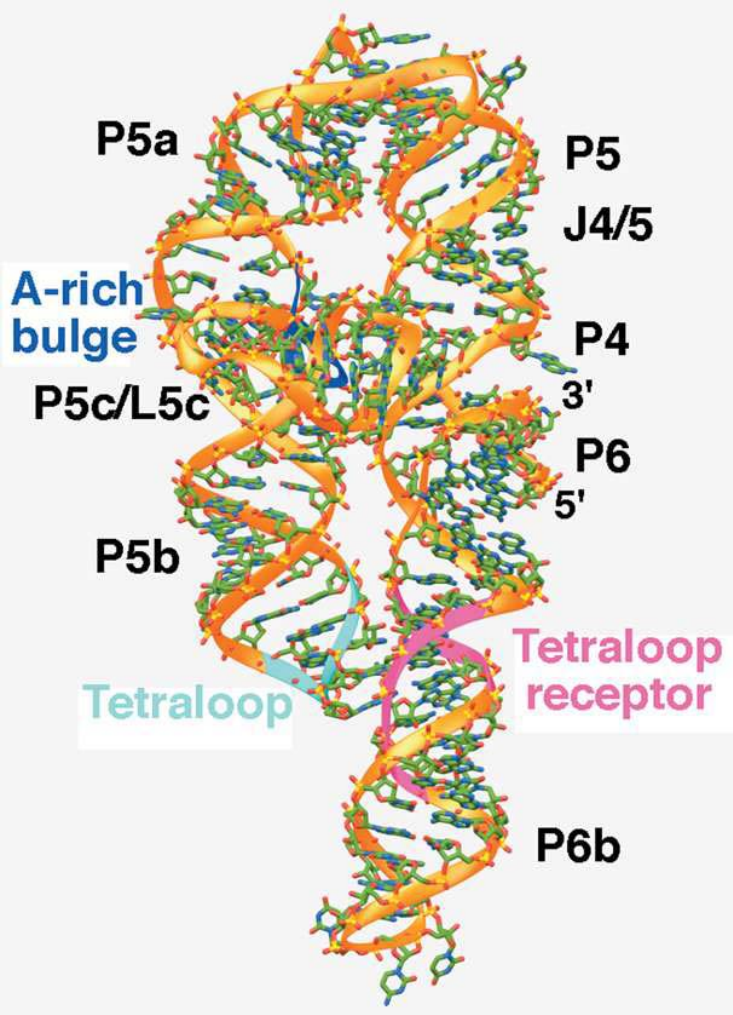
- Another diagram
- Requires guanine nucleoside or nucleotide
  - (not for energy)
  - May be guanosine, GMP, GDP or GTP
- No protein required
- Self-splicing
- *Tetrahymena thermophila*
  - Tom Cech
- Ribozyme
  - Catalytic RNA

Self-splicing group I



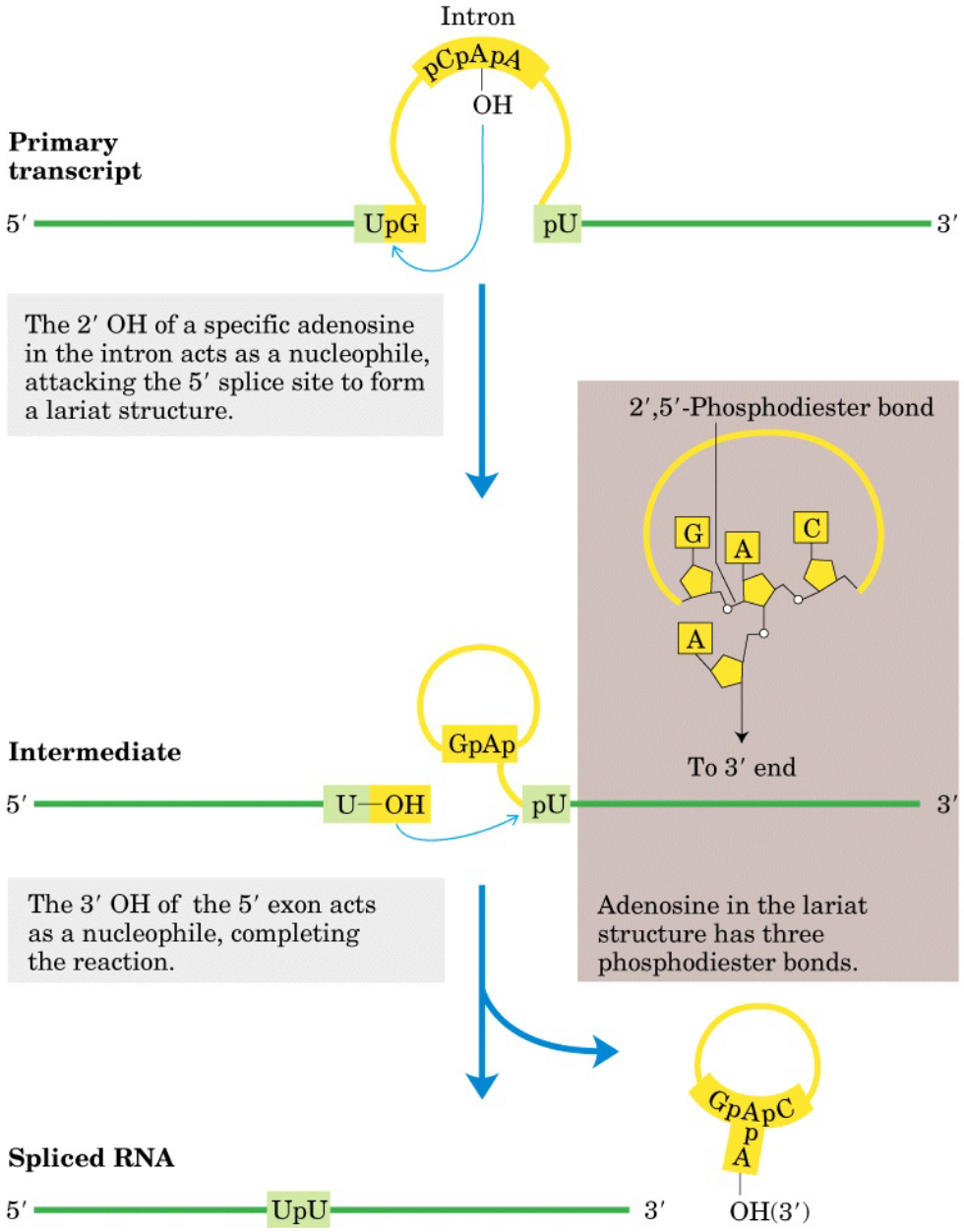
- *Tetrahymena thermophila*
- Secondary structure

X-ray structure



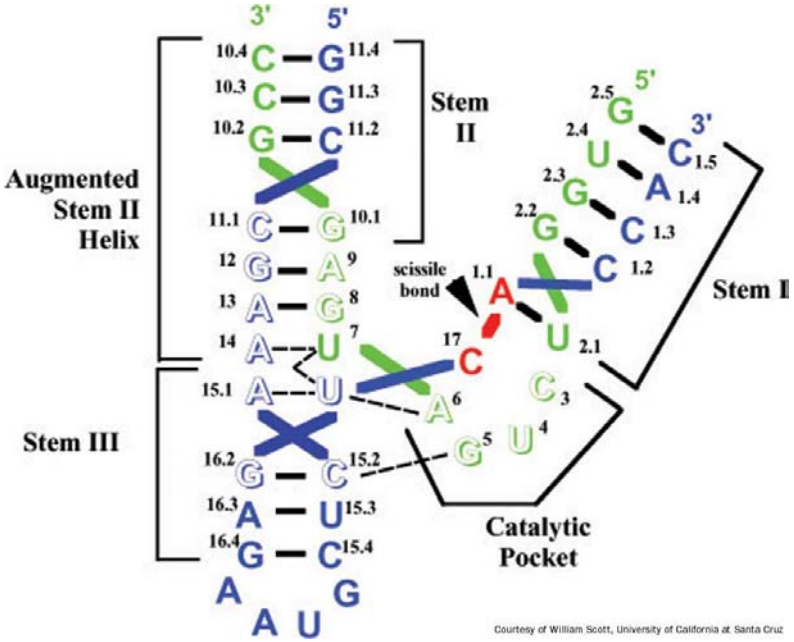
- Group I Ribozyme
- Have catalytic activity

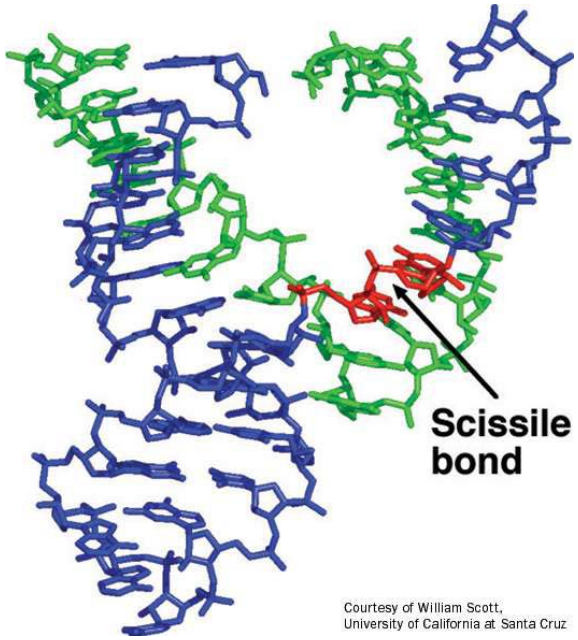
Splicing Group II introns



- Self splicing
- No extra guanosine or guanosine nucleotides needed
- Nucleophile
  - Adenosine within the intron
  - 2’ hydroxyl group
- Branched lariat structure formed as an intermediate
- Similar to spliceosome
- Ribozyme?

Hammerhead ribozyme



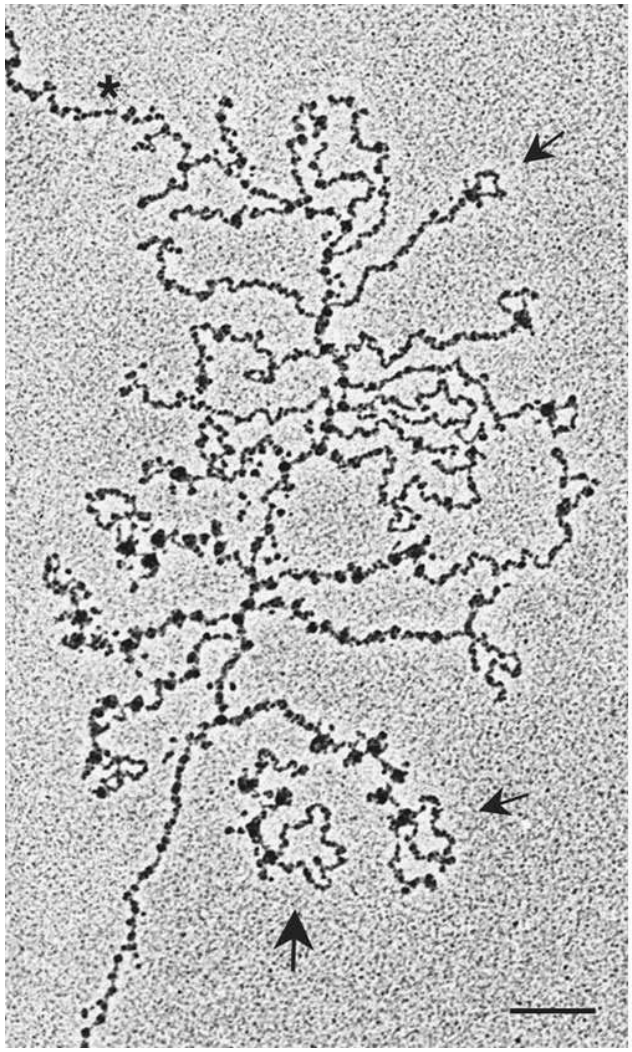


- Virusoids, virus like elements, need another virus to assist with replication and/or packaging (Small RNAs associated with plant RNA viruses.)
- Segments of their RNA genome promote site-specific RNA cleavage reactions associated with replication
- Hammerhead ribozyme
- Substrate RNA

Examples of ribozymes

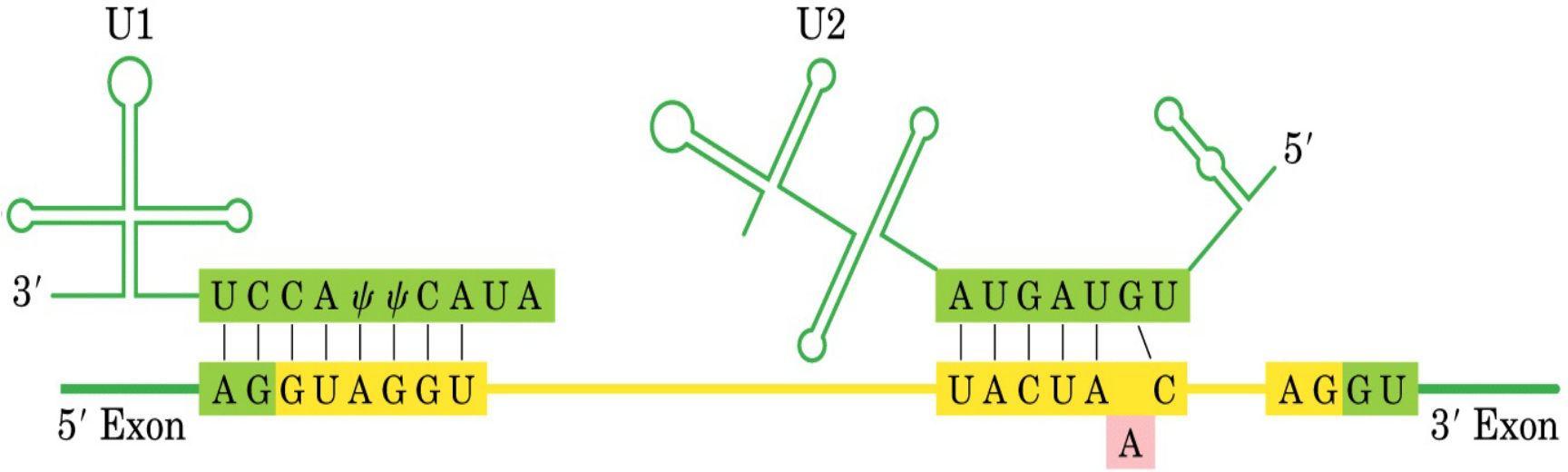
- Group1 and Group2 introns
- Some virusoids, small RNAs associated with plant RNA viruses
  - Example, hammerhead ribozyme
- *E. coli* RNase P
  - M1 RNA, 377 nucleotides
    - Alone is capable of cleaving tRNA precursors at the correct position
  - Protein: 17,500 daltons
    - Functions to stabilize the structure
- Peptide bond formation in ribosomes
- Spliceosome catalytic center formed by U2, U5 and U6 snRNAs

25. Spliceosome in action

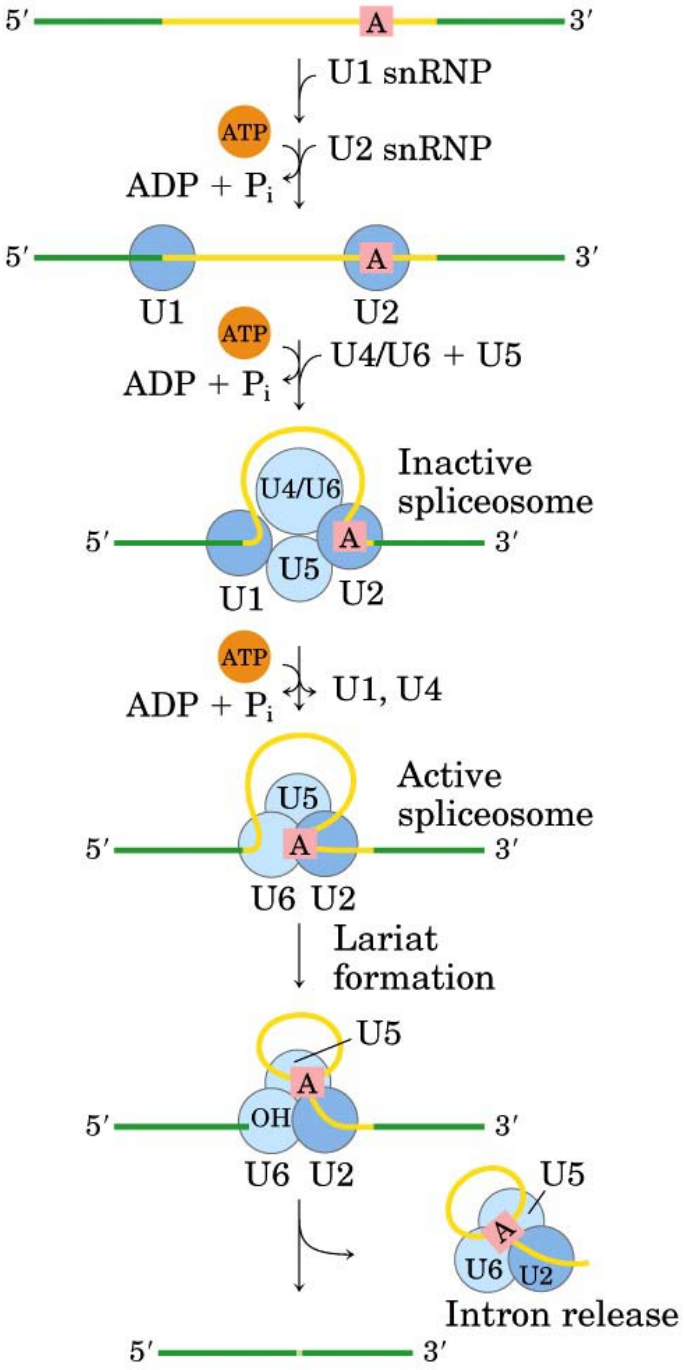


- Introns looped out (small arrows)
- Separated intron (large arrow)

26. snRNPs (small nuclear ribonucleoproteins, snurps)



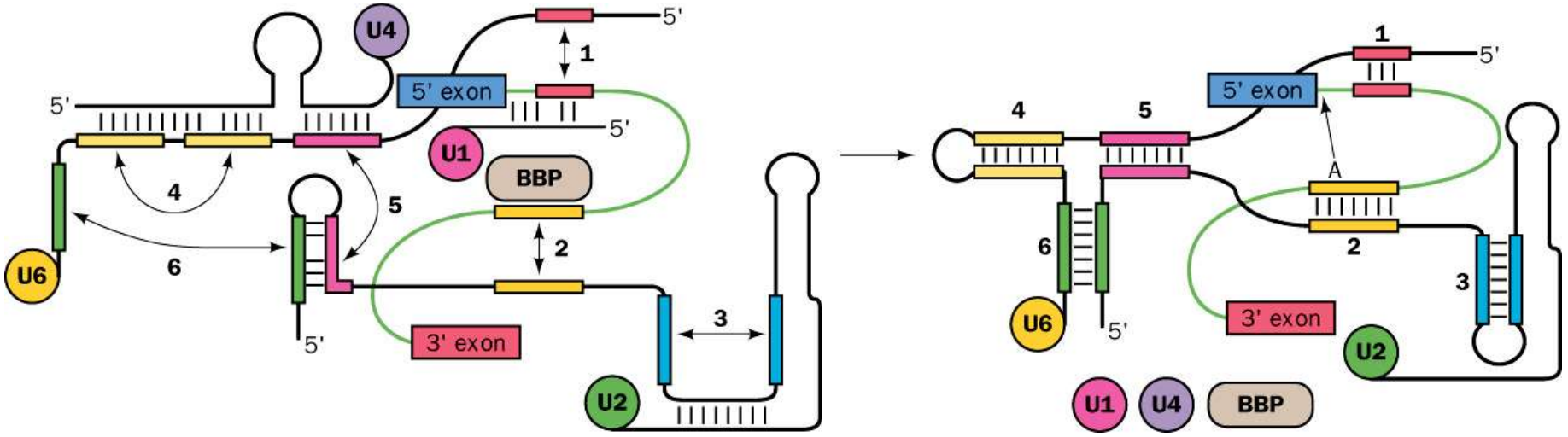
- snRNPs (pronounced snurps)
  - Small nuclear ribonucleoproteins (snRNPs)
    - U1, U2, U4, U6, U5
      - snRNA, 100 to 200 nucleotides long
    - 5 snRNAs (U1, U2, U4, U5, and U6) involved in splicing
- U1 snRNA, complementary to sequences near 5' splice site
- U2 snRNA, paired to region containing A that participates as a nucleophile during splicing



27. Spliceosome

- Complex reaction
- U1 and U2 snRNPs bind
- U4/U6 and U5 bind forming an inactive spliceosome
- U1 and U4 displaced, active spliceosome
- Catalytic steps
  - Lariat formation
  - Intron released
  - Spliced RNA
- Spliceosome: about 50 proteins, the size of a ribosome

Another view of splicing



- Steps and rearrangements leading to the first transesterification

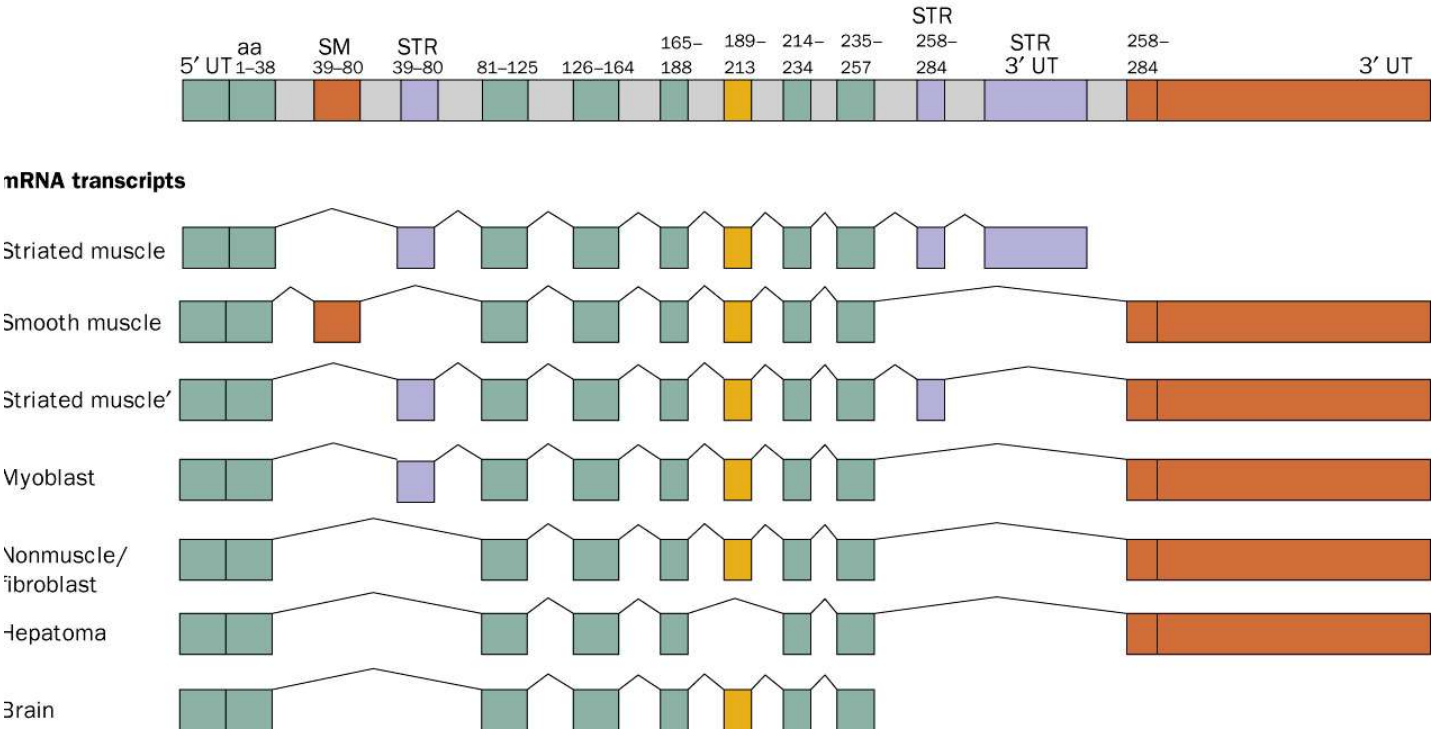
Significance of gene splicing

- Observations:
  - 1. Introns are rare in prokaryotic structural genes
  - 2. Introns are uncommon in lower eukaryotes such as yeast
    - 239 introns in its ~6000 genes
  - 3. Abundant in higher eukaryotes
    - Unexpressed sequences constitute ~80% of a typical vertebrate structural gene and >90% of a few of them
- Roles of gene splicing
  - 1. Agent for rapid evolution
  - 2. Alternative splicing
    - One gene; several proteins

Modular structure of proteins

- Example: LDL receptor (839 residues)
  - 45-kb gene containing 18 exons
  - 13 exons specify polypeptide segments that are homologous to segments in other proteins.
  - 1. Five exons encode a 7-fold repeat of a 40-residue sequence that occurs once in complement C9
  - 2. Three exons each encode a 40-residue repeat similar to that occurring four times in epidermal growth factor as well as in other proteins
  - 3. Five exons encode a 400-residue sequence that 33% identical with a polypeptide segment shared with EGF

Rat cell specific α-tropomyosin variants



- Seven alternately spliced variants in rat.
  - Brown: Smooth only
  - Purple: Striated only
  - Yellow: Variable
  - 3’ end varies

How to get 100,000 gene products from 30,000 genes

- Humans have about 30,000 genes
  - Generate 50-140,000 gene products
  - Estimated on average each structural gene produces three proteins
  - Much of these variants generated by alternate splicing, others by alternate 3' cleavage, alternate start sites and RNA editing.
- Alternate splicing can add a high degree of variability

Types of changes produced by alternate splicing

- Types of changes in the proteins generated by splice variants
  - Soluble or membrane bound
  - Phosphorylation by a specific kinase
  - Subcellular localization
  - Allosteric effector binding
  - Rate of transcription
  - Rate of degradation
  - Affinity for substrate
- ~15% of human genetic diseases are caused by point mutations that result in pre-mRNA splicing defects

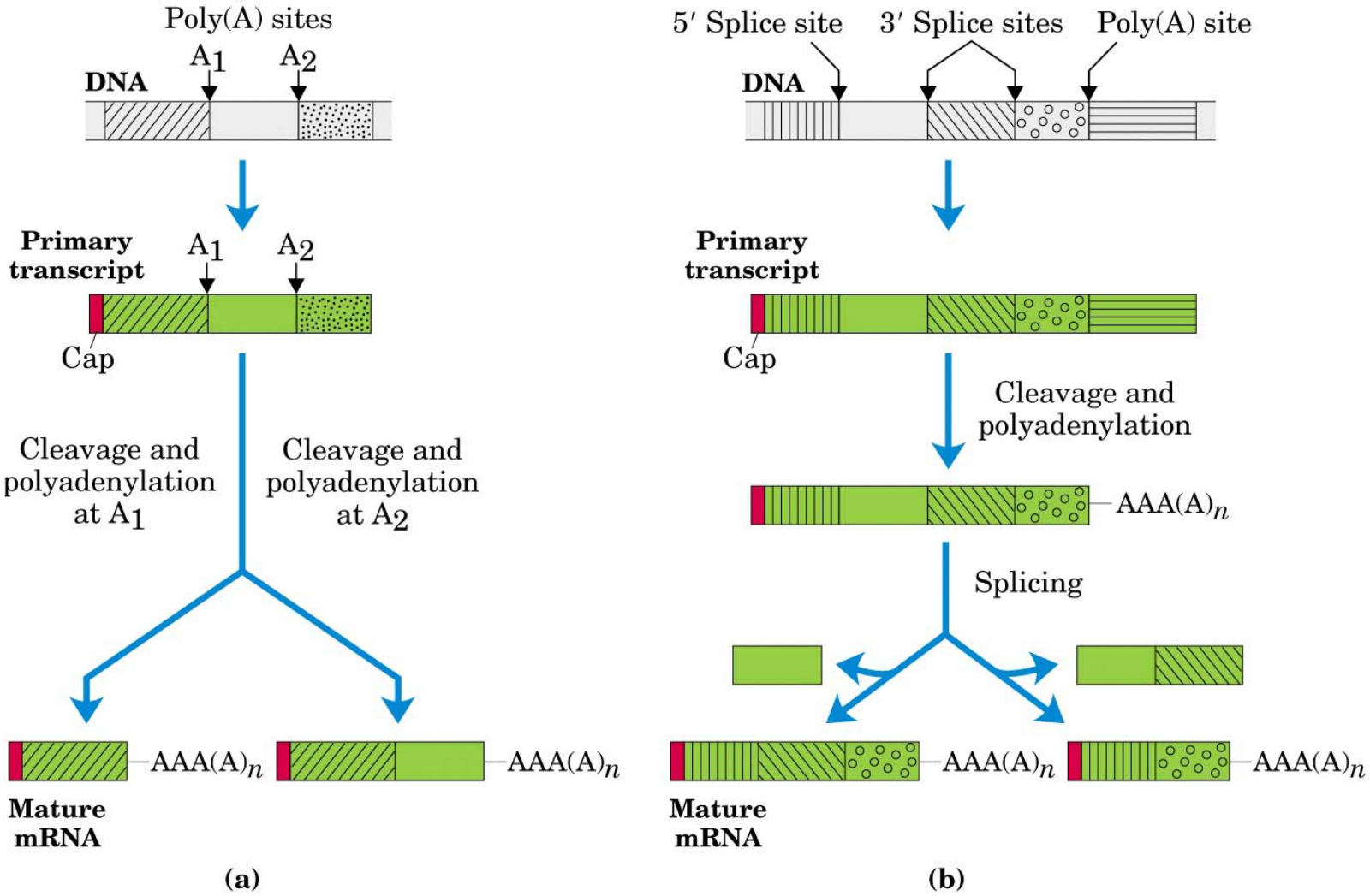
RNA enzymes, ribozymes

- Best characterized catalytic RNAs
  - Group I introns
  - RNase P
    http://www.mbio.ncsu.edu/RNaseP/rna/threed/threed.html
  - Hammerhead ribozymes
- Ribozymes vary in size
  - Self-splicing group I introns have over 400 nucleotides
  - Hammerhead ribozyme consists of two RNA strands with a total of 41 nucleotides

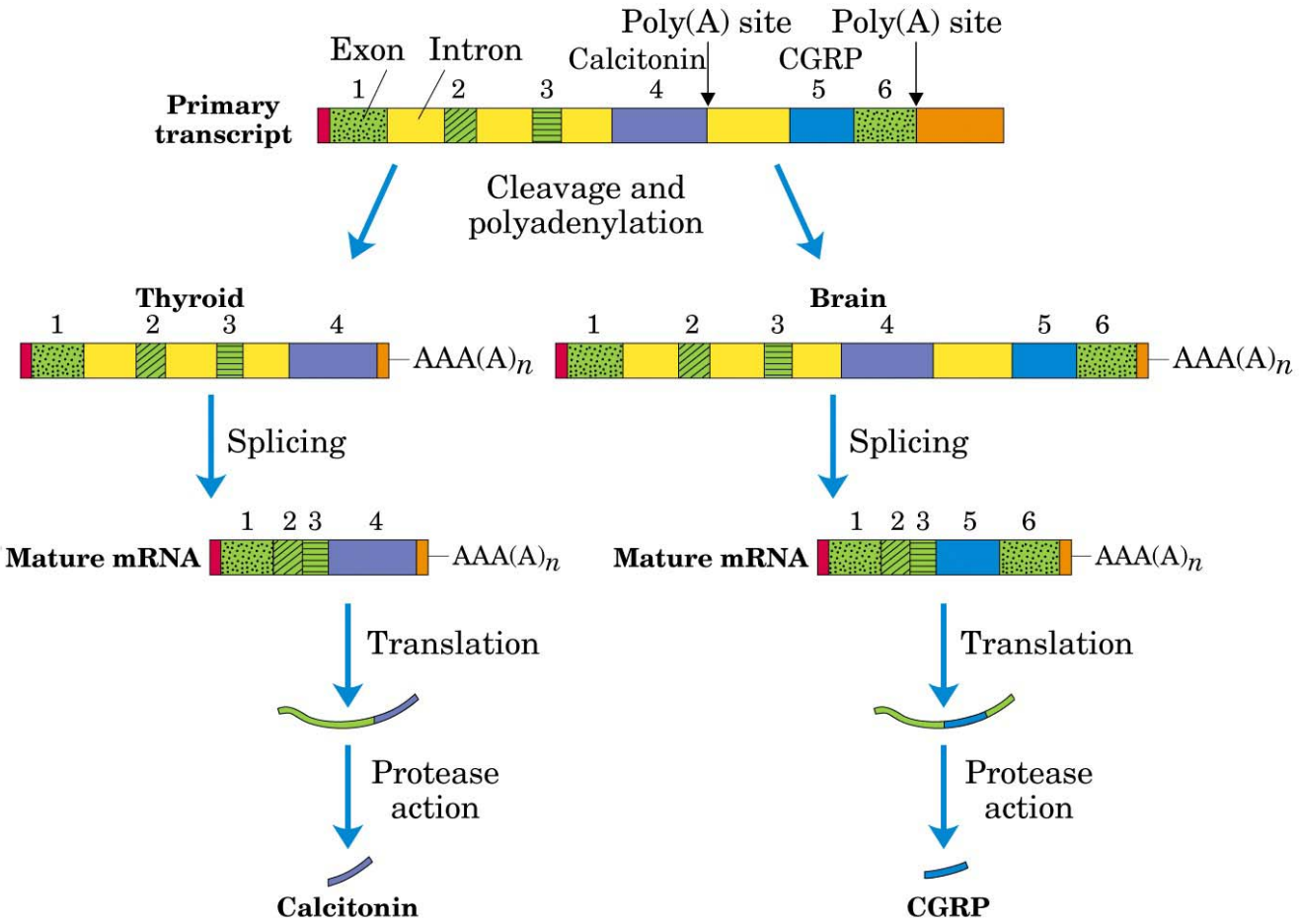
Mechanisms for alternative processing of eukayotic mRNA

- How can one gene produce more than one gene product?
- At the level of the mRNA
  - Alternative Splicing
  - Alternative Poly(A) sites
  - Trans-splicing
  - RNA editing; change the mRNA after it is made
    - Insertions or deletions
    - Base deamination

6. Alternative Poly (A) sites and alternative Splicing



7. Alternative Poly (A) sites

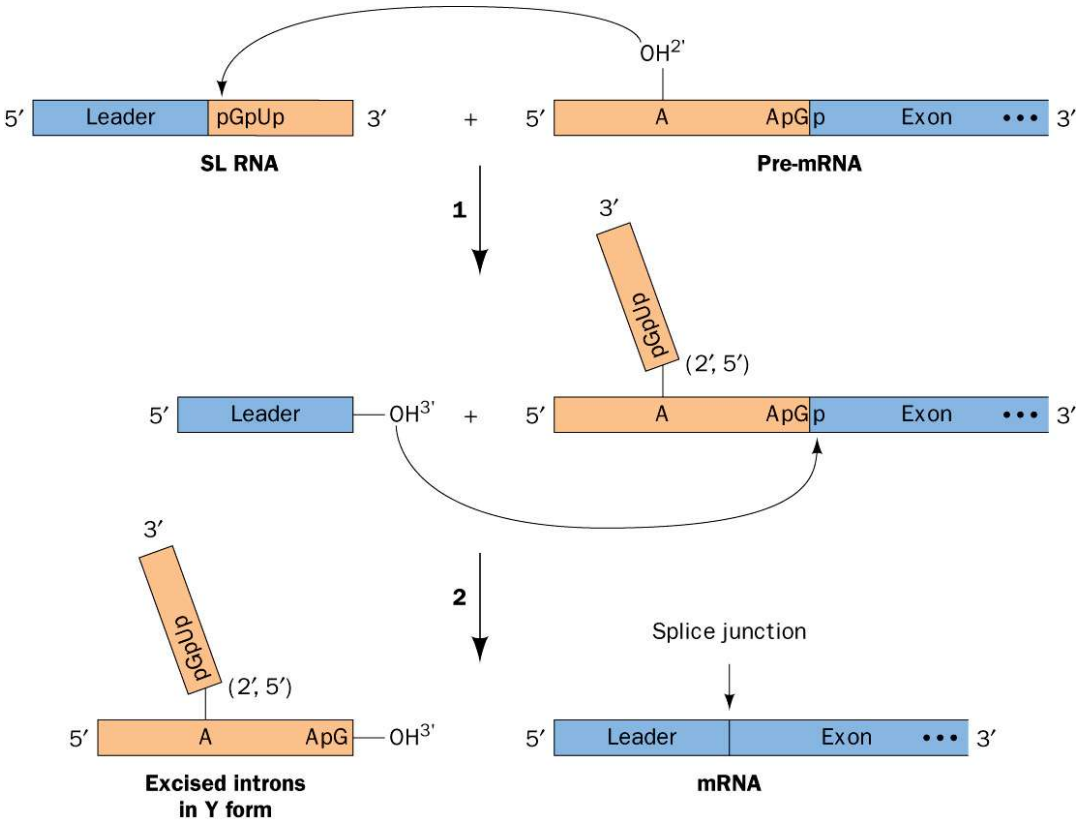


- Thyroid
  - First poly(A) site
  - Exon 4 retained
  - Produces
    - Calcitonin
- Brain
  - Second poly(A) site predominates
  - Exon 4 spliced out
  - Produces:
    - CGRP
    - (calcitonin-gene-related peptide)

8. Trans-Splicing

- As opposed to cis-splicing
- Trypanosomes
  - (Cause of African sleeping sickness)
  - Their mRNA all have the same 35-nt noncoding leader sequence
  - Leader is not present in the corresponding genes
  - Spliced leader

9. Trans-Splicing

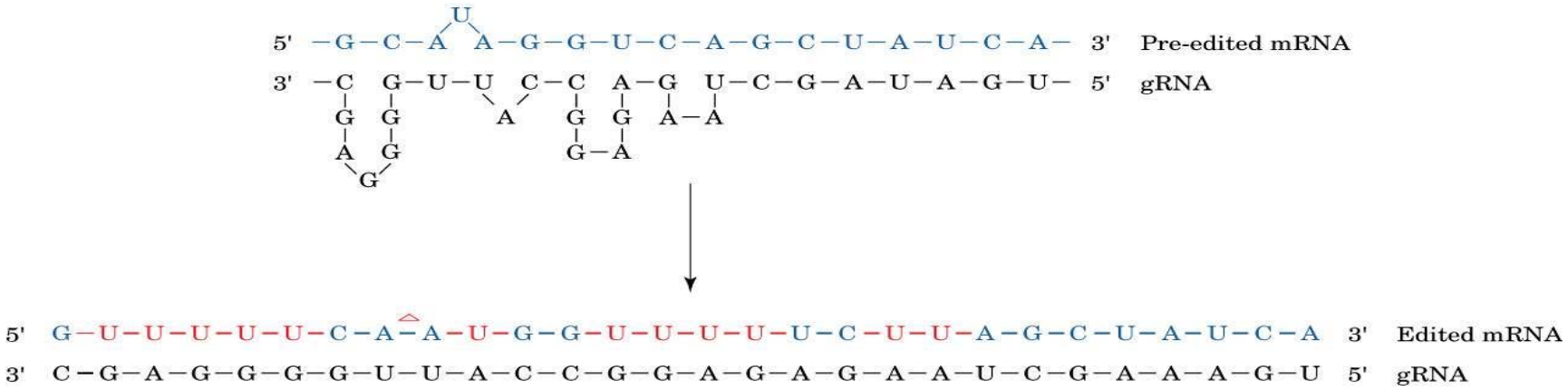


- Reaction resembles spliceosomal cis-splicing
- First trans-esterification reaction generates Y-shaped rather than lariat-shaped product
- Also found in nematodes (*C. elegans*) and flatworms and perhaps *Drosophila* and vertebrates.

10. RNA editing

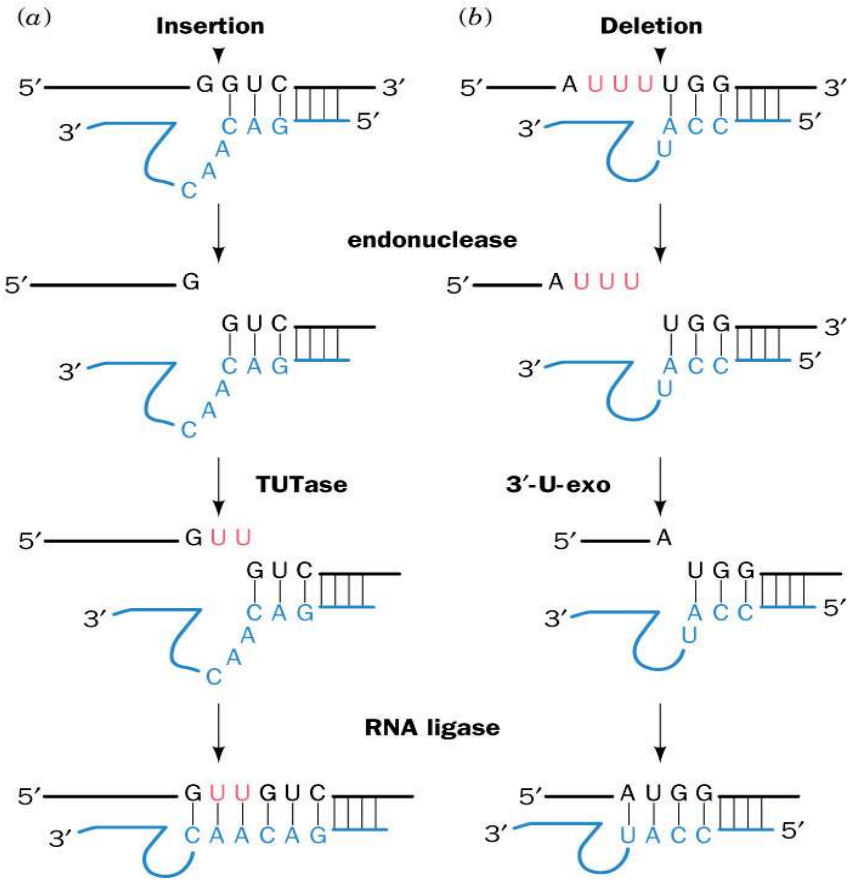
- mRNA sequences do not correspond to the corresponding gene
  - C to U
  - U to C
  - Insertion or deletion of U
  - Insertion of multiple G or C residues

11. RNA editing, Insertions or Deletions



- mRNA can be altered after it is made, RNA editing
  - In Trypanosomes, addition and removal of up to hundreds of U's to and from 12 untranslatable mitochondrial mRNAs
- Guide RNAs (gRNAs)
- Editosome
  - ~20S RNP

12. Trypanosomal RNA Editing Pathway



- Editosome
  - Catalyzes insertions and deletiions
  - Endonuclease
  - Insertions
    - Terminal uridylyltransferase (TUTase)
  - Deletions
    - 3’-U-Exonuclease (3’-U-exo)
  - RNA ligase

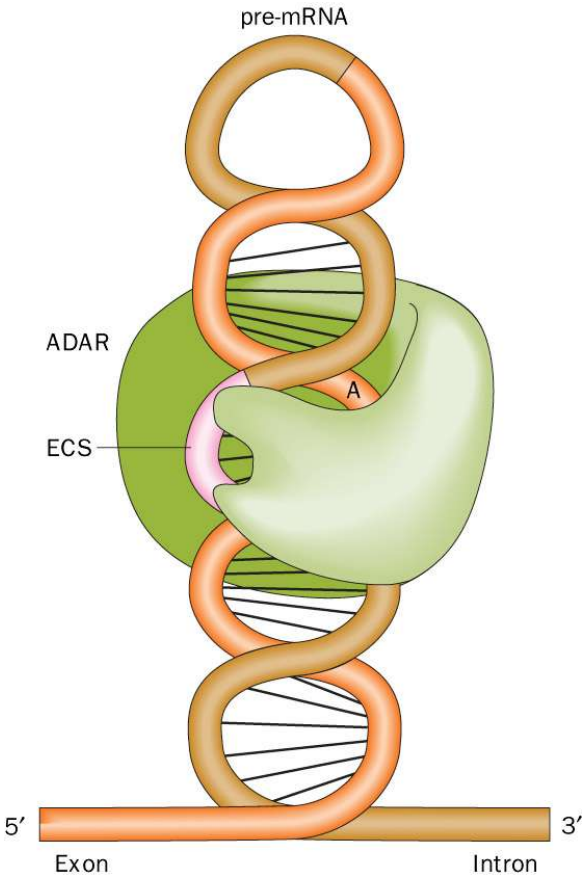
13. RNA editing by Base Deamination

- Humans
- mRNA once made, modified, by cytosine or adenosine deamination
  - C becomes a U
  - A becomes I; i.e. Adenosine becomes Inosine
    - Reads like G in translation

14. ApoB100 & ApoB48

- ApoB100- large 4536-residue protein, (512 kD)
  - Made in the liver
  - Functions in VLDL, IDL and LDL
  - C-terminal domain mediates LDL receptor binding
- ApoB48- 2152-residue protein (250 kD)
  - Made in the intestine
  - Functions in Chylomicrons
  - N-terminal residues of ApoB100
- Codon for Gln 2153 (CAA) is changed to UAA, a stop codon
- Site specific cytidine deaminase (substitutional editing)

15. Glutamate receptor



- Brain: "Memory receptor"
  - Glutamate: Brain stimulatory neurotransmitter
  - Ligand-gated ion channel
- Pre-mRNA undergoes an A–I deamination
  - Transforms a Gln codon (CAG) to (CIG) (read as CGG, Arg)
- ADAR1 and ADAR2
  - Adenosine deaminases acting on RNA
- Substrate RNA double helix
  - Involves exon-intron junction Thus must precede splicing

16. Another mechanism contributing to diversity

- *Drosophila cacophony*
  - mRNA that encodes a voltage-gated Ca^{+2} channel has 10 different substitutional editing sites
    - C to U
    - A to I
  - Thus potential of generating 1000 different isoforms in the absence of alternative splicing

17. RNAi Nobel Prize

The Nobel Prize in Physiology or Medicine 2006

"for their discovery of RNA interference - gene silencing by double-stranded RNA"

Photo: L. Cicero/Stanford

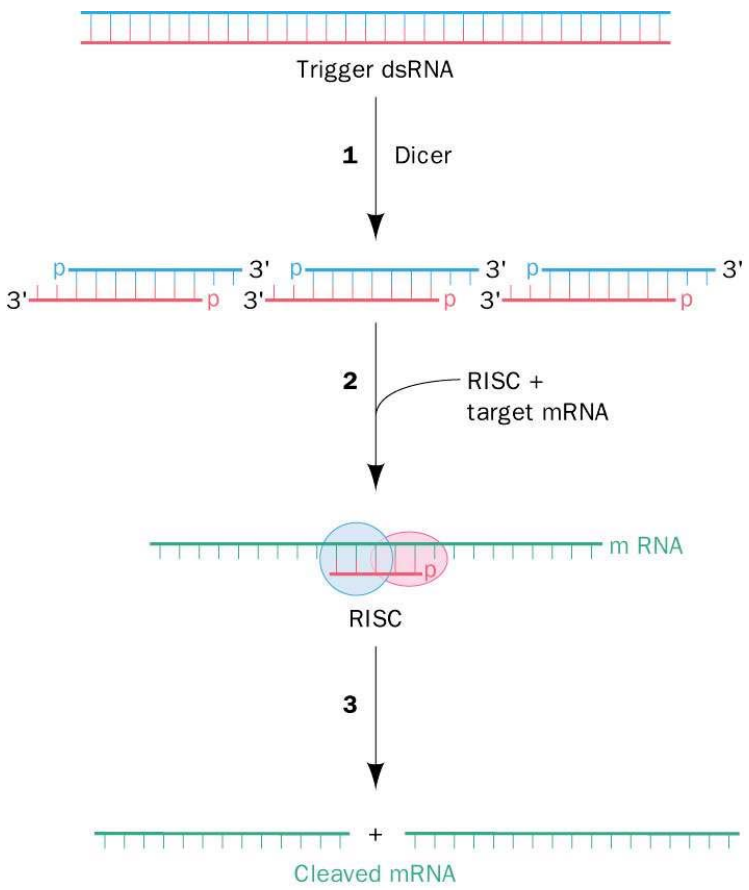
Photo: R. Carlin/UMMAS

- Andrew Fire
  - Stanford
- Craig Mello
  - U. Mass. Med. Center

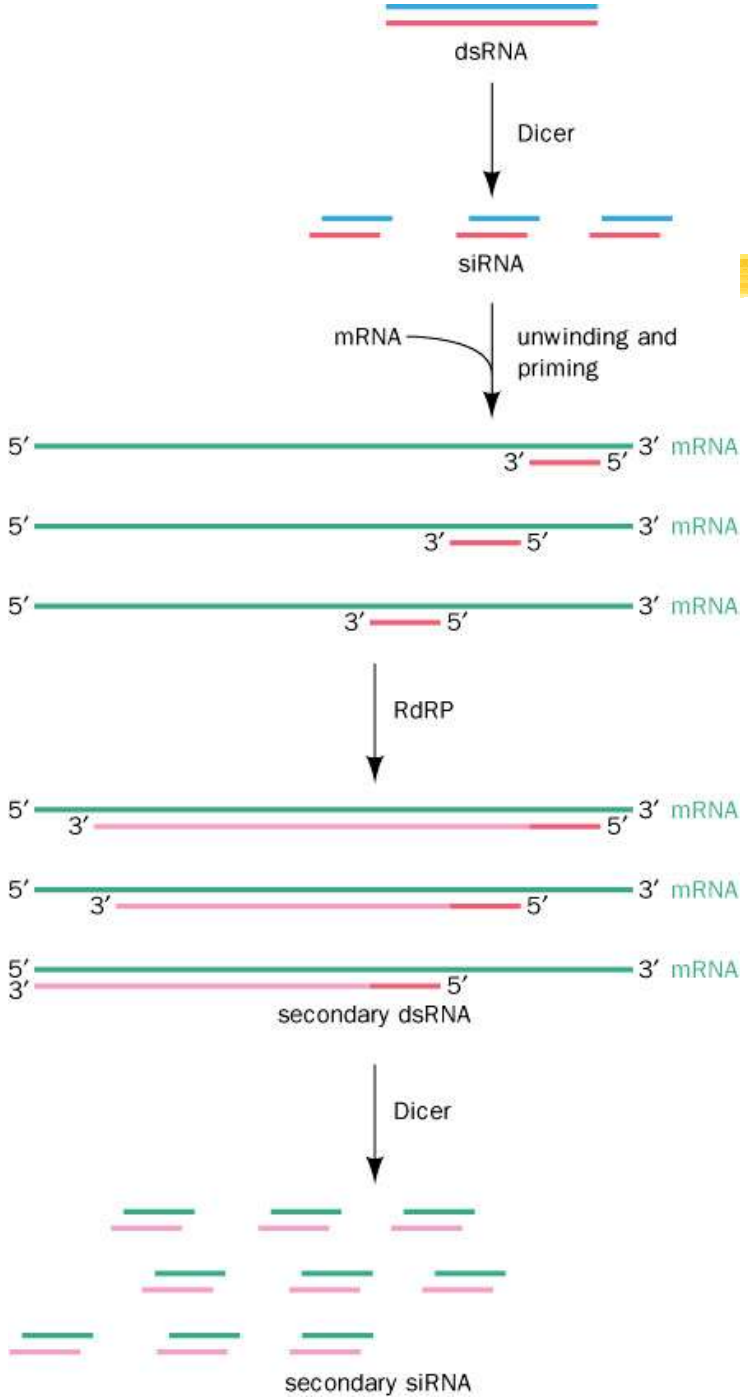
19. RNA interference, RNAi

- RNA interference (RNAi)
  - Post-transcriptional gene silencing (PTGS)
    - Induced by injection of
      - Antisense RNA
      - Sense RNA
      - Double stranded RNA (dsRNA)
        - Most effective
    - Only catalytic amounts needed
      - Amplification
- RNAi/PTGS method of choice for generating null mutants (knockout) in plants and non-vertebrates
- Defense against viral infection (most eukaryotic viruses store and replicate their genomes as RNA
- Potential mechanism to silence disease-causing mutant genes such as oncogenes.

20. Model RNAi



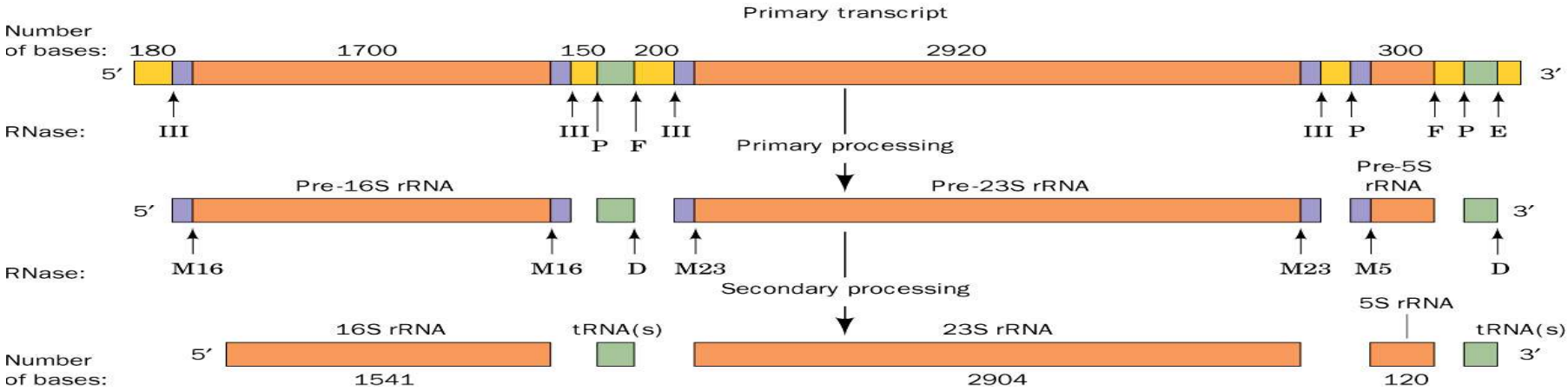
- dsRNA cleaved by Dicer
  - Member of RNase III family of double-strand specific RNA endonucleases
  - Generates siRNA
    - Small interfering RNAs
    - 2nt overhang at its 3'end and a 5' phosphate
- RISC
  - RNA-induced silencing complex
  - Mediates unwinding of siRNA (ATP)
  - Antisense strand of siRNA guides RISC complex to an mRNA with the complementary sequence
  - Slicer
    - Endonuclease activity in RISC
- mRNA degraded, gene silenced



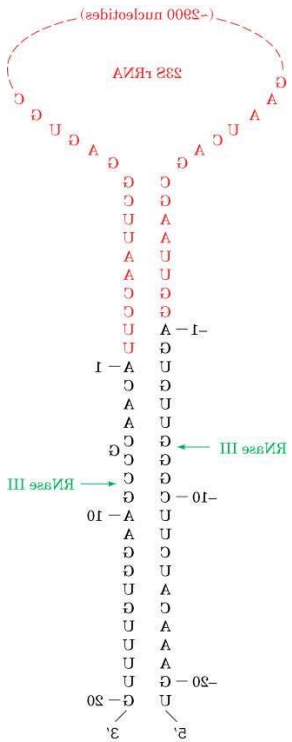
21. Transitive RNAi

- Genes can be silenced that are not complementary to the original trigger dsRNA
- Amplification process
  - Trigger dsRNA copied
  - RNA-dependent RNA polymerase (RdRP)
  - Secondary dsRNA cleaved by Dicer
    - Resulting siRNAs (small interfering RNAs) may extend beyond the sequence complementary to the original trigger dsRNA
- Amplification apparently does not occur in mammals
  - Thus the effect of RNAi in mammals is transient

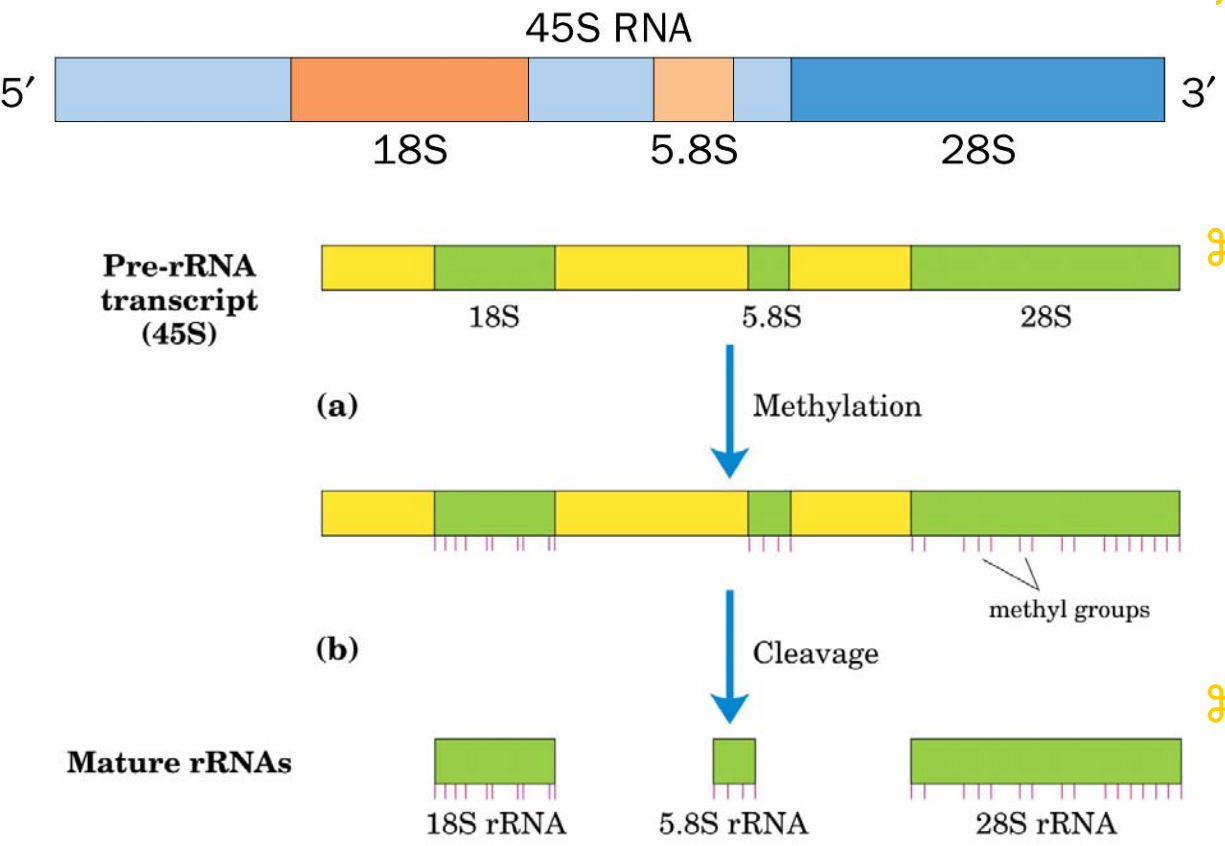
23. Bacterial pre-rRNA processing



- 30S transcript (pre-rRNA)
- Endonucleolytic Cleavage
  - RNase III, RNase P,RNase E, RNase F
  - Rnase III cleavage occurs in a stem loop structure
  - RNase P is a ribozyme
- 5’ and 3’ ends of pre-rRNAs trimmed away
  - RNase D, M16, M23 and M5
- Products:
  - 16S rRNA, tRNA, 23S rRNA, 5S rRNA
- *E. coli* genome encodes 7 pre-rRNA molecules



24. Processing pre-rRNA transcripts in vertebrates



- 45S pre-rRNA transcript
  - Nucleoli
  - RNA polymerase I
- Methylation
  - (Eukayotic and prokaryotic rRNA methylated)
- rRNA processing guided by snoRNA (Small nucleolar RNAs)
  - "Guide RNAs"
    - Complementary to methylation sites
  - 200 in mammals
- In intron rich organisms, most snoRNAs are encoded in the introns of structural genes

25. SnoRNA

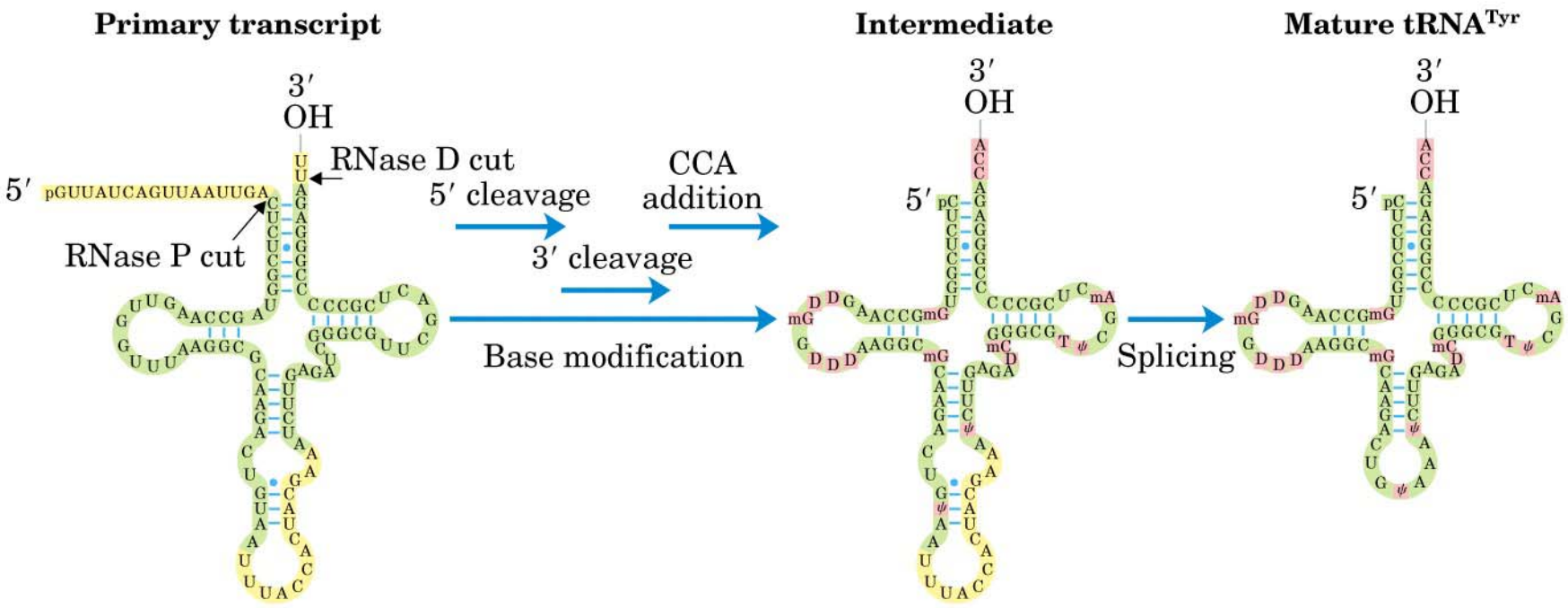
- Small nucleolar RNAs
- Also involved in
  - 2’ O-methylation
  - Pseudouridylation U to Ψ
  - Nucleolytic processing
  - Synthesis of Telomeric DNA

- Review articles posted
  - Exosomes and snoRNA
  - RNA trans-splicing

26. tRNA processing

- Many cells have 40-50 distinct tRNAs
  - Eukaryotes have multiple copies of many of the tRNA genes
    - Several hundred to several thousand tRNA genes
    - Many eukaryotic primary tRNA transcripts contain introns
- tRNAs derived from longer precursors
  - 5’ end cleavage, RNase P (all organisms)
    - Contains catalytic RNA
  - 3’ end cleavage, various enzymes

27. tRNA processing

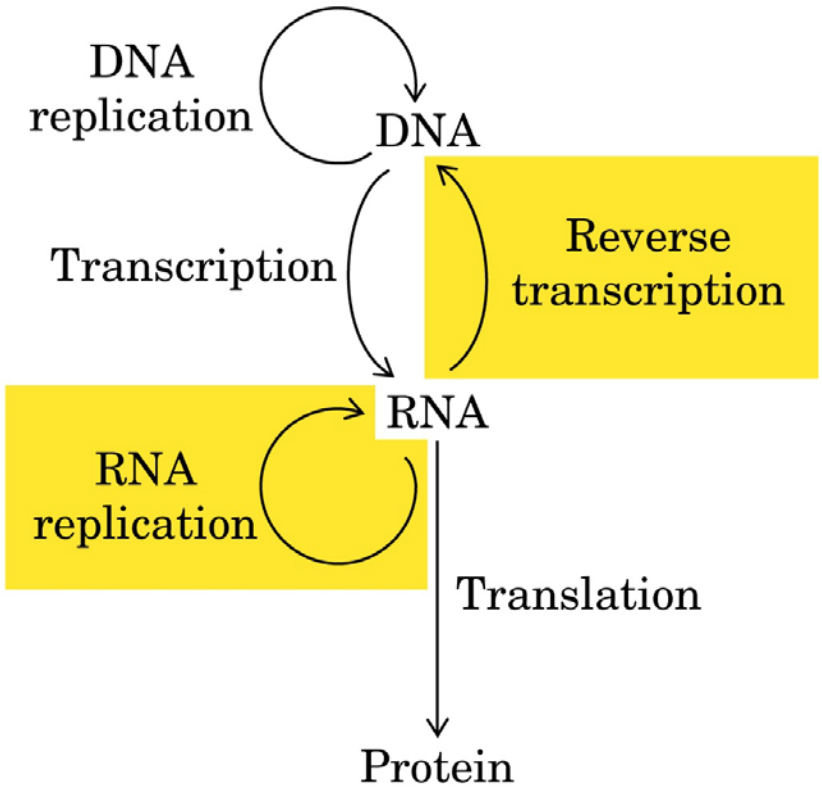


- RNAse P, catalytic RNA, ribozyme
- CCA, absent from some bacterial and from all eukaryotic tRNA precursors
  - tRNA nucleotidyltransferase adds CCA(3') sequence

28. mRNA degradation

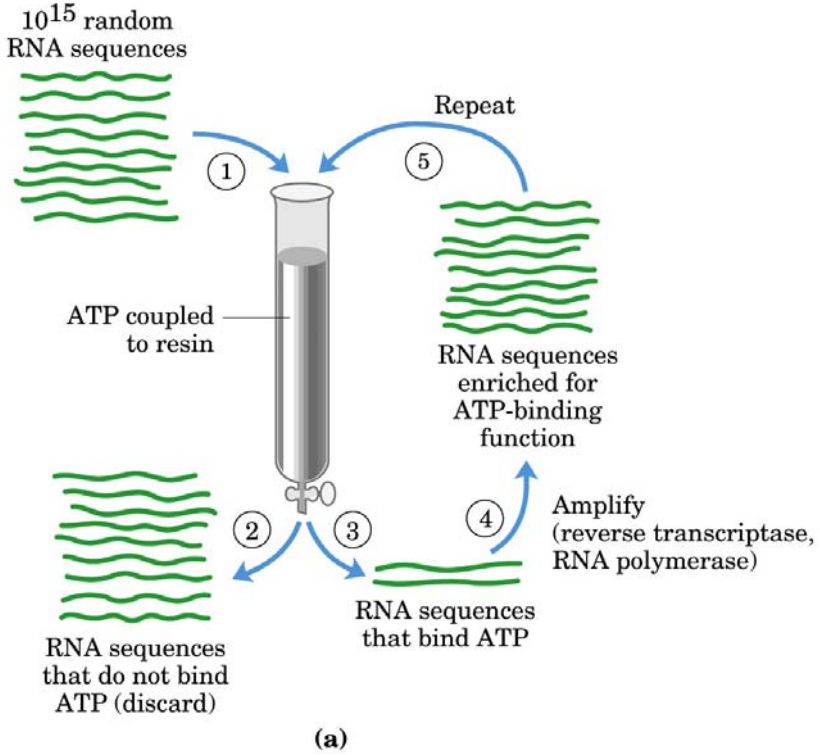
- Steady state level
  - Rate of synthesis
  - Rate of degradation
- Average half-life of vertebrate mRNA about 3 hours (varies from seconds to days)
- Half-life of bacterial mRNA, 1.5 min.
- Eukaryotes:
  - Shortening of Poly(A) tail
  - Decapping
  - Degradation in 5'-3' direction (usually, cells also have 3'-5' exoribonucleases)

29. Central dogma



- Includes RNA dependent synthesis of RNA and DNA

39. SELEX



- Systematic evolution of ligands by exponential enrichment
- RNAs with new functions
- Binding of ATP