BME 42-620 Engineering Molecular Cell Biology

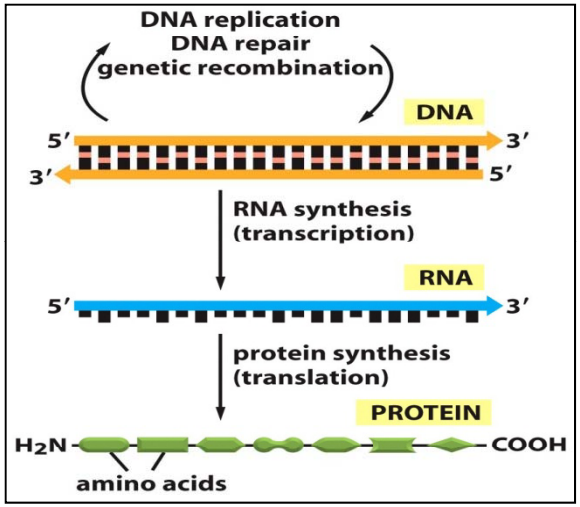
# Lecture 17:

# Gene Expression I: From DNA to RNA

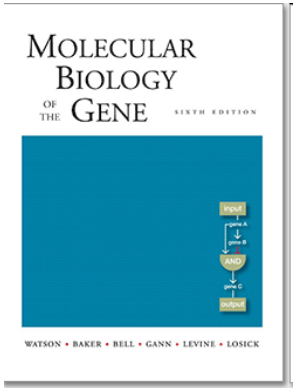
*Chapters 5 & 6*

# Lectures on Gene Expression

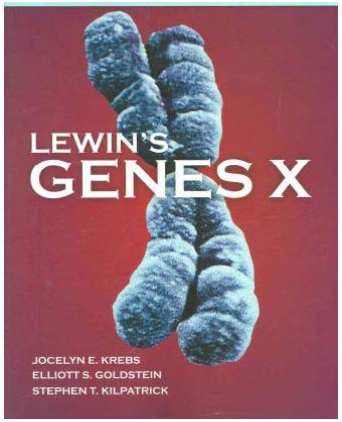
- From DNA to RNA (lecture 17)
- From RNA to protein (lecture 18)
- Regulation of gene expression (lecture 19)
- Quantitative analysis and modeling of gene expression (lecture 19)



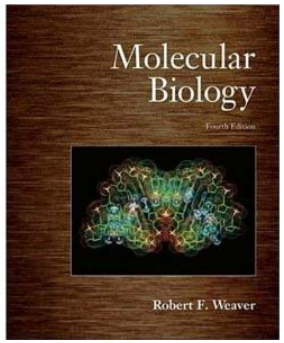
# Some References




Watson et al, Cold Spring Harbor Lab Press, 2008




Krebs et al, Jones & Bartlett, 2009




Weaver, McGraw-Hill, 2007

# Outline

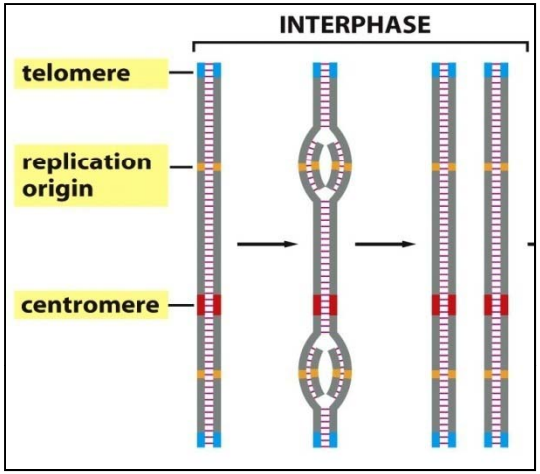
- DNA replication and repair
- Overview of transcription
- RNA
- Transcription Process
- Nuclear export of mRNA

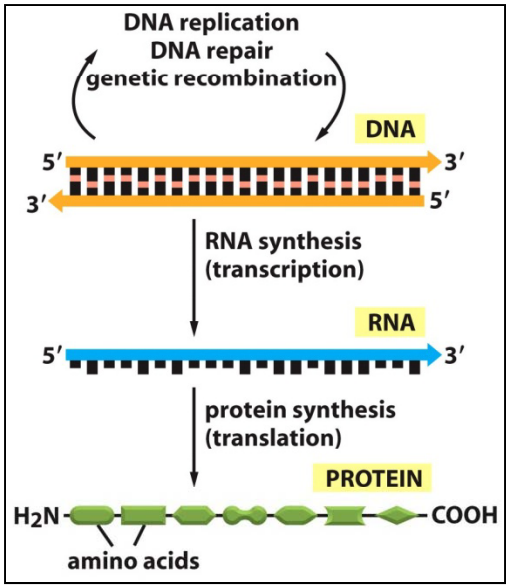
# Outline

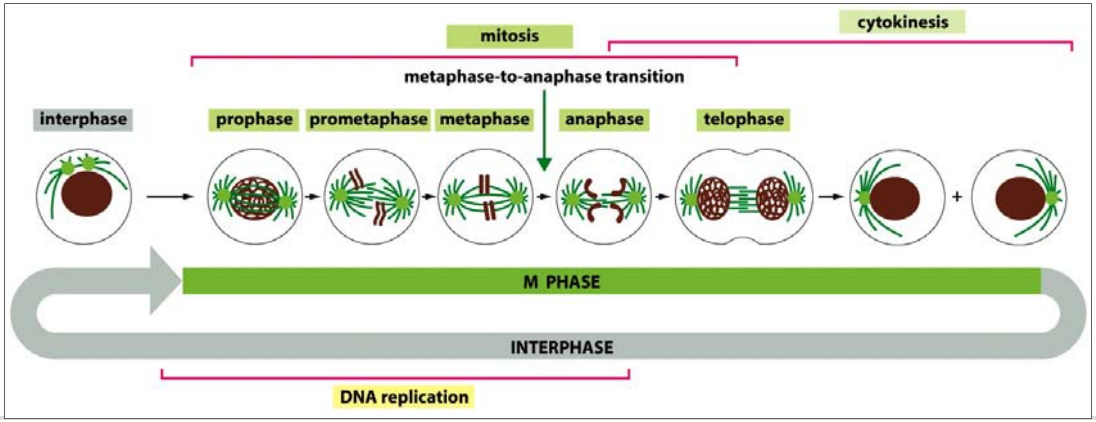
- DNA replication and repair
- Overview of transcription
- RNA
- Transcription Process
- Nuclear export of mRNA

# Replication and Processing of Genetic Information

- In S phase, cells copy genetic information through DNA replication.
- Cells read and process genetic information through transcription and translation.

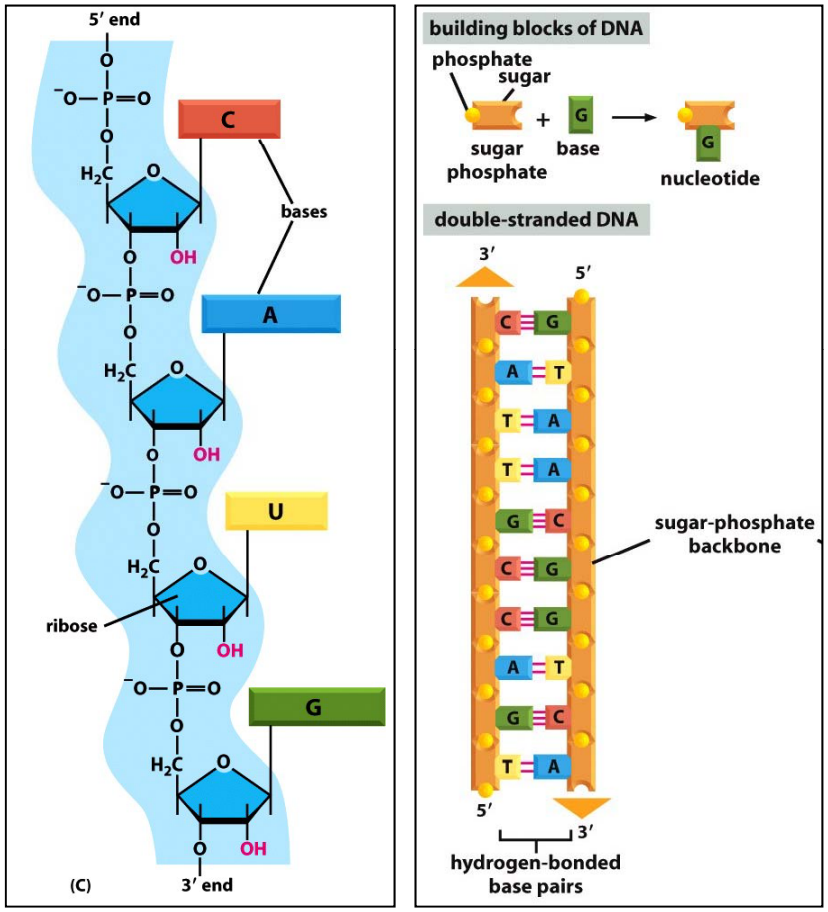


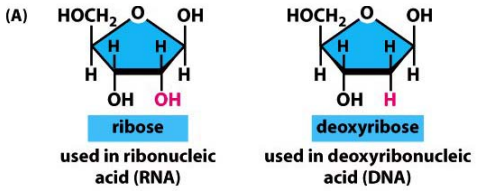



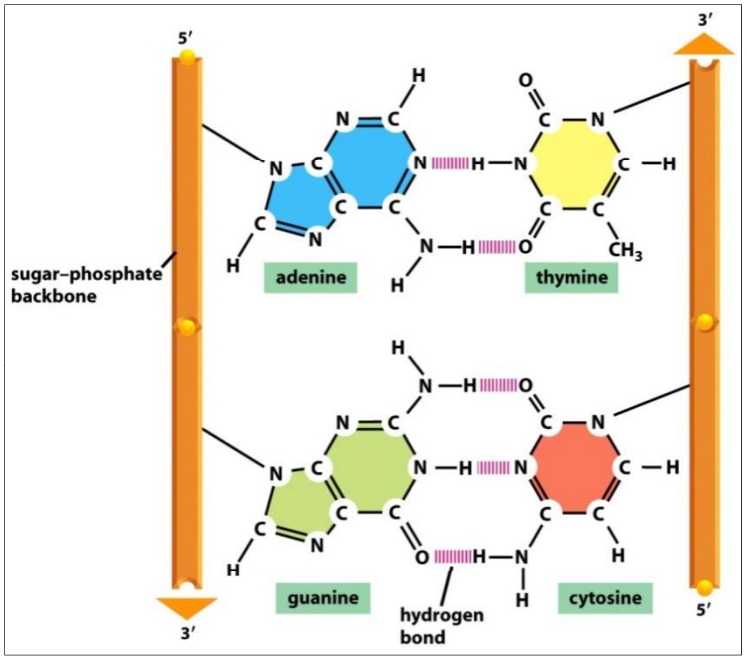
# Polarity of DNA and RNA

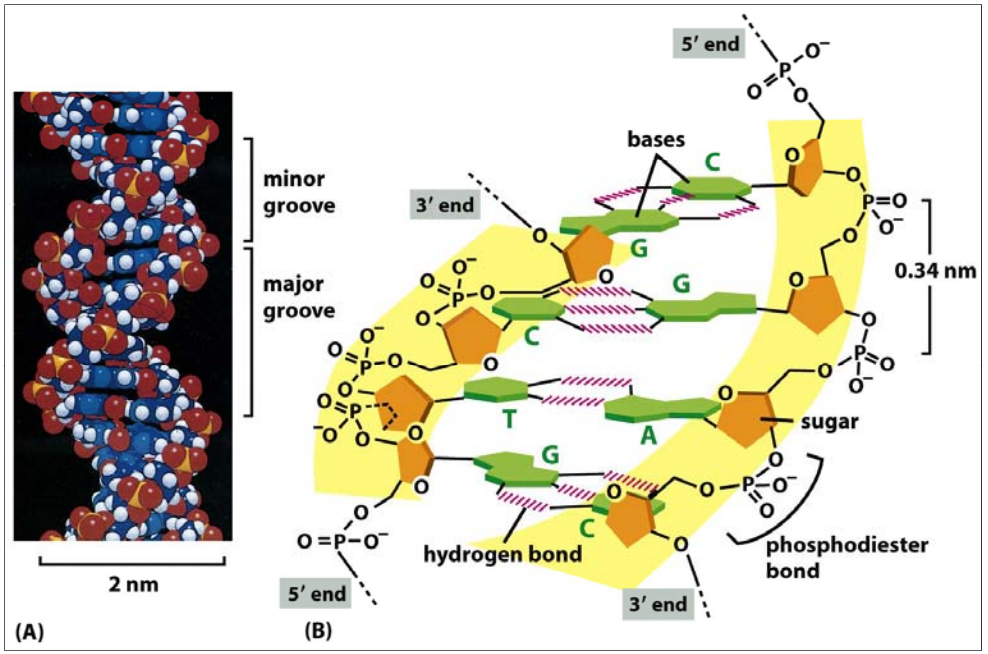
- A nucleotide consists of a base, a five-carbon sugar, and one or more phosphate groups.
- 5' end: the end with the 5' phosphate group
- 3' end: the end with the 3' hydroxyl group
- DNA: A(adenine), T(thymine), G (guanine), C(cytosine)
- RNA: A, U (uracil), G, C

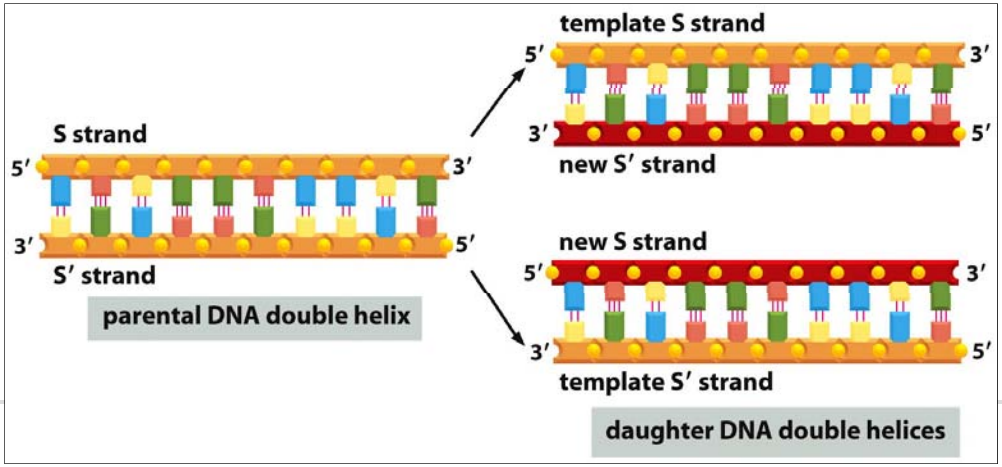


# DNA Structure



# DNA Replication

- DNA replication must ensure high fidelity in the short-term while allowing genetic variations in the long-term.

- High fidelity of DNA replication:
  - One nucleotide error per $10^9$ nucleotides per cell generation.
    → Limits the number of essential genes to ~50000.

  - One amino acid alteration every 200,000 years

  - If we are to model DNA replication in short-term, noise in this process can be largely ignored.

- The high fidelity of DNA replication is achieved using multiple error checking and correction mechanisms.

# DNA Damage Repair (I)

- DNA damage can be caused by many factors
  - Environmental factors: heat, radiation, chemicals
  - DNA of each human cell loses 5000 purine (A,G) bases spontaneously every day due to a process called depurination

- Eventually, less than 1 of 1000 of base changes result in permanent mutation thanks to DNA repair.

- Transcription stalls at DNA damages.

- DNA repair is coupled to transcription so that urgently needed DNA sequences get repaired quickly.

# DNA Damage Repair (II)

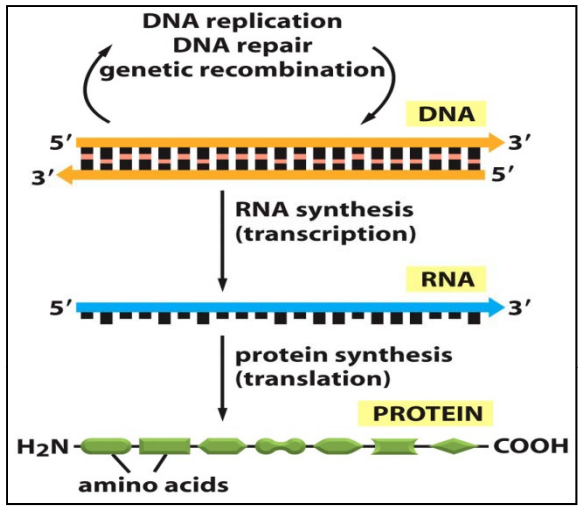
**Table 5–2** Some Inherited Syndromes with Defects in DNA Repair

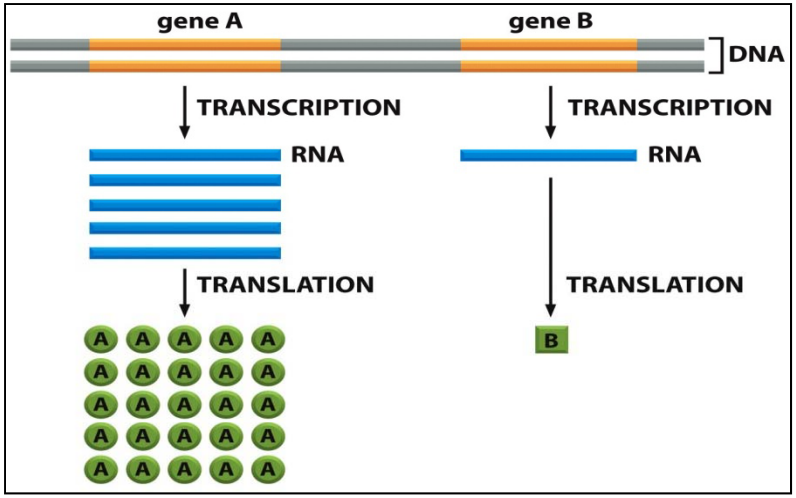
| NAME | PHENOTYPE | ENZYME OR PROCESS AFFECTED |
|---|---|---|
| MSH2, 3, 6, MLH1, PMS2 | colon cancer | mismatch repair |
| Xeroderma pigmentosum (XP) groups A–G | skin cancer, UV sensitivity, neurological abnormalities | nucleotide excision–repair |
| XP variant | UV sensitivity, skin cancer | translesion synthesis by DNA polymerase $\eta$ |
| Ataxia telangiectasia (AT) | leukemia, lymphoma, $\gamma$-ray sensitivity, genome instability | ATM protein, a protein kinase activated by double-strand breaks |
| BRCA2 | breast, ovarian, and prostate cancer | repair by homologous recombination |
| Werner syndrome | premature aging, cancer at several sites, genome instability | accessory 3′-exonuclease and DNA helicase |
| Bloom syndrome | cancer at several sites, stunted growth, genome instability | accessory DNA helicase for replication |
| Fanconi anemia groups A–G | congenital abnormalities, leukemia, genome instability | DNA interstrand cross-link repair |
| 46 BR patient | hypersensitivity to DNA-damaging agents, genome instability | DNA ligase I |

# Outline

- DNA replication and repair
- Overview of transcription
- RNA
- Transcription Process
- Nuclear export of mRNA

# The Central Dogma of Molecular Biology




- Gene expression consists of multiple stepts that are dynamically and closely regulated.
- RNA plays important roles in regulation of gene expression.
- For many genes, RNA is the end product, which fold in 3D and serves structural, catalytic, and regulatory roles.

# Complexity of Genomes

- Genome contains not only protein coding information but also regulatory information that controls when, where, and how genes are expressed.

- Although genomic sequences of many living organisms are known, much less is known about regulation of gene expression.

- It is difficult to decode regulatory information purely based on sequences because information distribution on genome often is not orderly.

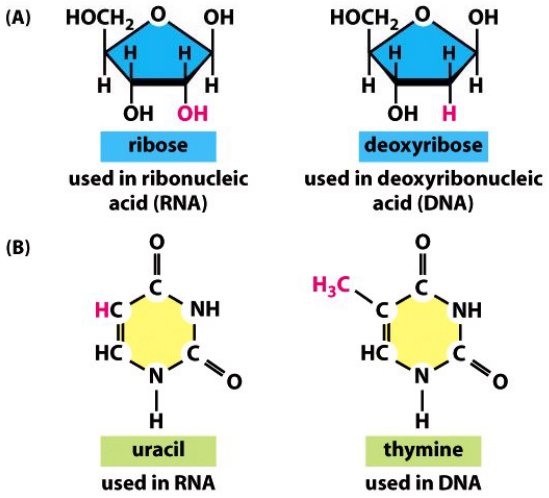
  - Example 1. genes encoding proteins that interact closely with each other often locate on different chromosomes.
  - Example 2. adjacent genes may encode proteins that are uncorrelated in the cell.

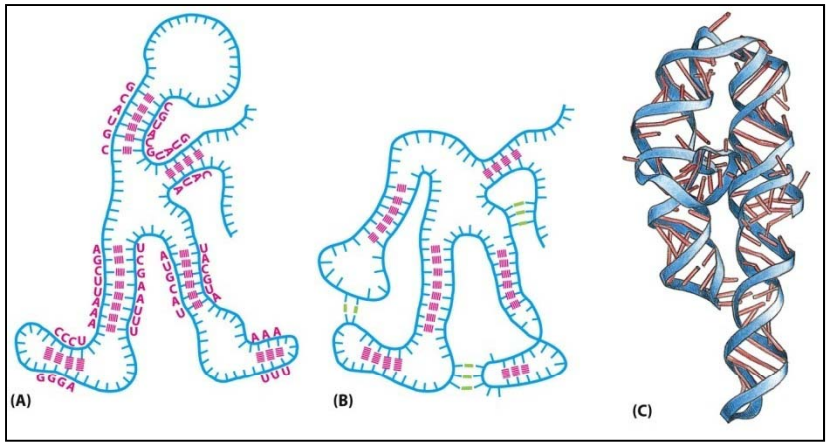
# Outline

- DNA replication and repair
- Overview of transcription
- RNA
- Transcription Process
- Nuclear export of mRNA

# Differences Between DNA and RNA

- DNA
  - Purine: adenine (A) guanosine (G)
  - Pyrimidines: thymine (T) cytosine (C)
- RNA
  - Purine: adenine (A) guanosine (G)
  - Pyrimidines: uracil (U) cytosine (C)
- RNA is single-stranded.
- RNA transcripts are much shorter than DNA.
- RNA can fold into 3D structures.



# Different Types of RNA

**Table 6–1 Principal Types of RNAs Produced in Cells**

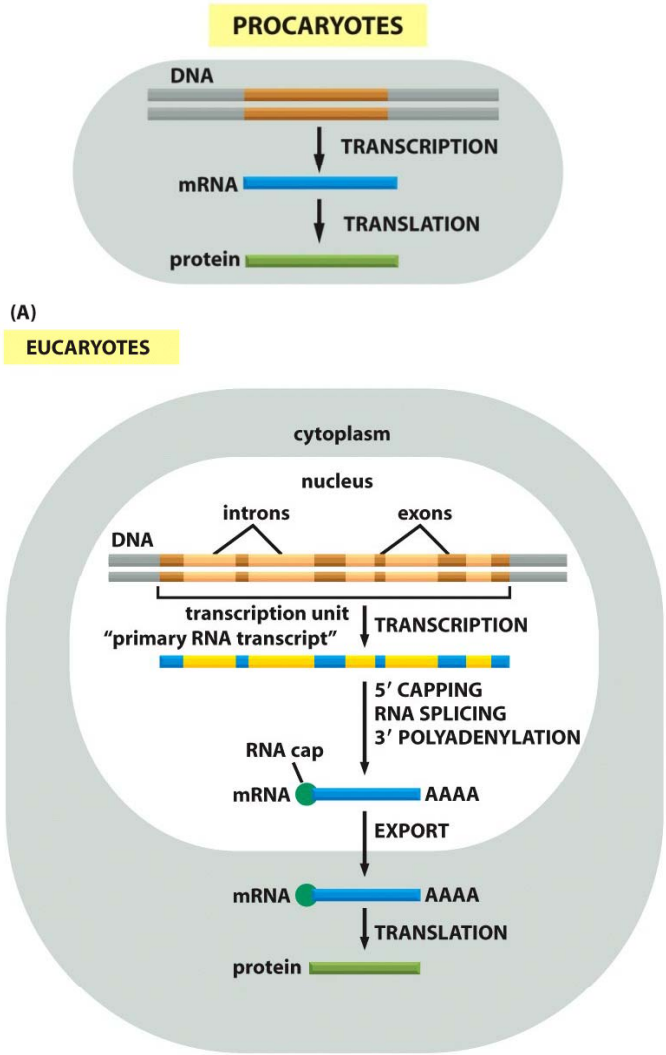
| TYPE OF RNA | FUNCTION |
|---|---|
| mRNAs | messenger RNAs, code for proteins |
| rRNAs | ribosomal RNAs, form the basic structure of the ribosome and catalyze protein synthesis |
| tRNAs | transfer RNAs, central to protein synthesis as adaptors between mRNA and amino acids |
| snRNAs | small nuclear RNAs, function in a variety of nuclear processes, including the splicing of pre-mRNA |
| snoRNAs | small nucleolar RNAs, used to process and chemically modify rRNAs |
| scaRNAs | small cajal RNAs, used to modify snoRNAs and snRNAs |
| miRNAs | microRNAs, regulate gene expression typically by blocking translation of selective mRNAs |
| siRNAs | small interfering RNAs, turn off gene expression by directing degradation of selective mRNAs and the establishment of compact chromatin structures |
| Other noncoding RNAs | function in diverse cell processes, including telomere synthesis, X-chromosome inactivation, and the transport of proteins into the ER |

# Outline

- DNA replication and repair
- Overview of transcription
- RNA
- Transcription Process
- Nuclear export of mRNA

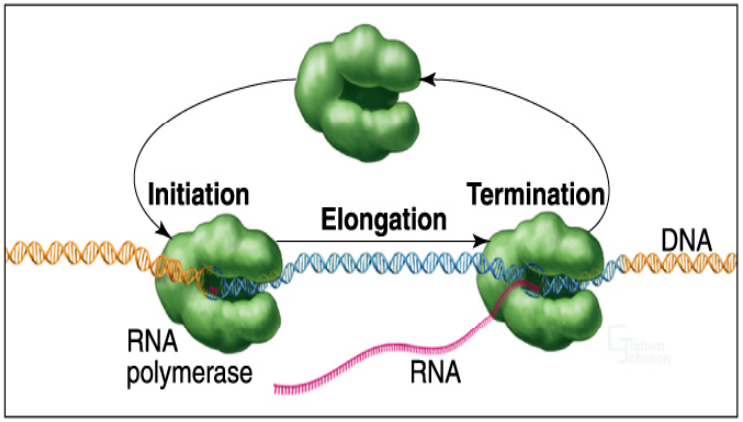
# Steps From DNA to RNA

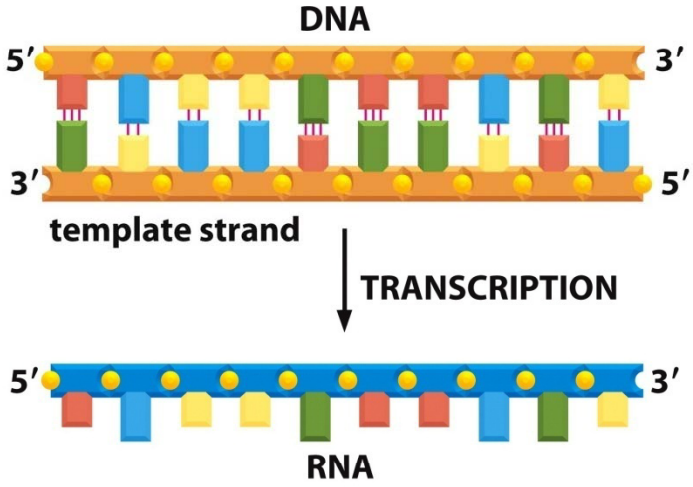
- Transcription
- 5' end capping
- Splicing of pre-mRNA
- 3' end capping
- Nuclear export of mature mRNA
- These steps can occur concurrently.



# The Transcription Cycle

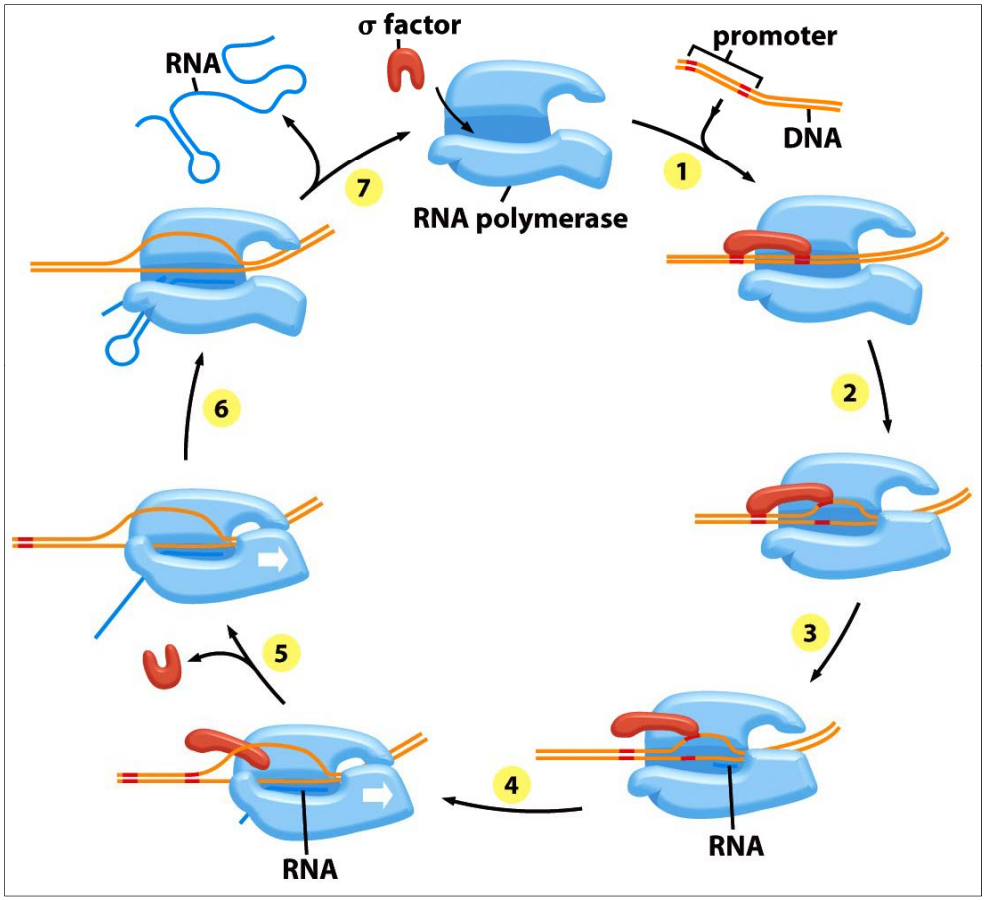
- Transcription is performed by RNA polymerase.
- Three steps: initiation, elongation, termination
- Transcription goes from 3' end to 5' end.
- RNA polymerase makes 1 mistake every $10^4$ bases.



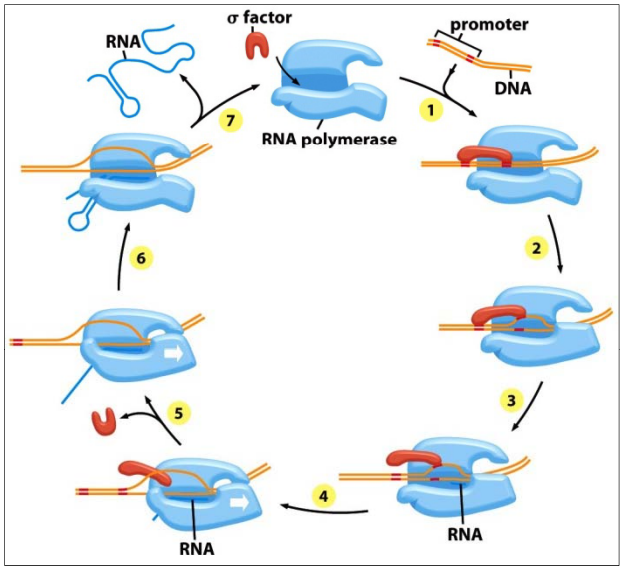
# Transcription Initiation and Termination in Bacteria (I)

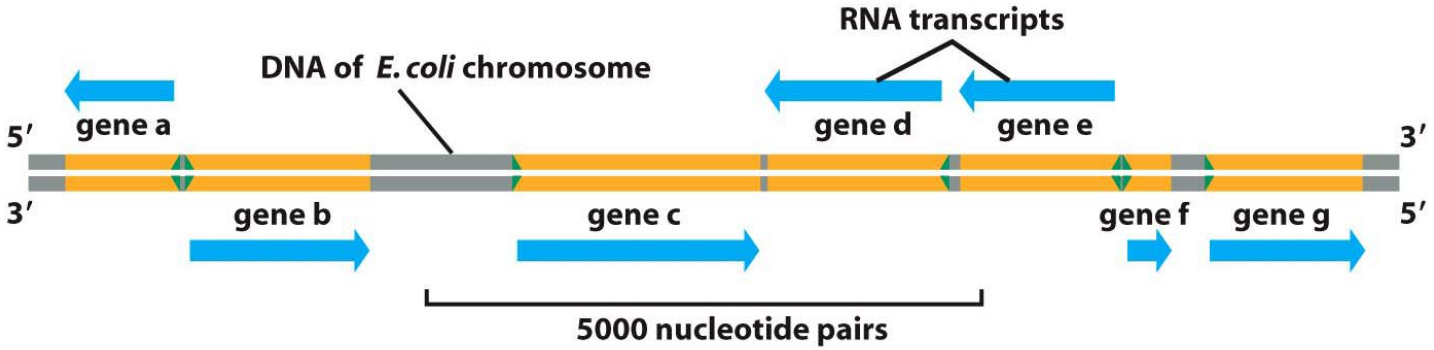
- In bacteria, initiation of transcription requires the σ-factor.
- Transcription starts downstream of the promoter.



# Transcription Initiation and Termination in Bacteria (II)

- Transcription stops at the terminators, which form a structure that destabilizes polymerase's hold on RNA .
- Direction of transcription is determined by the promoter of each gene.



# Three RNA Polymerases in Eukaryotic Cells

- One RNA polymerase in bacteria.
- Three RNA polymerases in eukaryotic cells.

**Table 6–2 The Three RNA Polymerases in Eucaryotic Cells**

| TYPE OF POLYMERASE | GENES TRANSCRIBED |
|---|---|
| RNA polymerase I | 5.8S, 18S, and 28S rRNA genes |
| RNA polymerase II | all protein-coding genes, plus snoRNA genes, miRNA genes, siRNA genes, and most snRNA genes |
| RNA polymerase III | tRNA genes, 5S rRNA genes, some snRNA genes and genes for other small RNAs |

The rRNAs are named according to their "S" values, which refer to their rate of sedimentation in an ultracentrifuge. The larger the S value, the larger the rRNA.

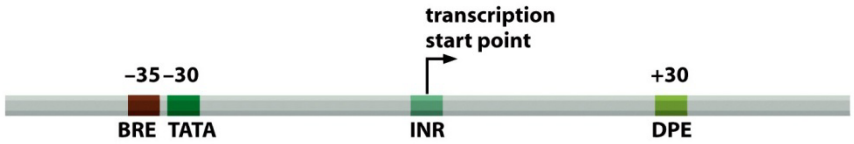
# Transcription Initiation and Termination in Eukaryotic Cells (I)

- RNA polymerase II requires many general transcription factors (TFII's) to initiate.

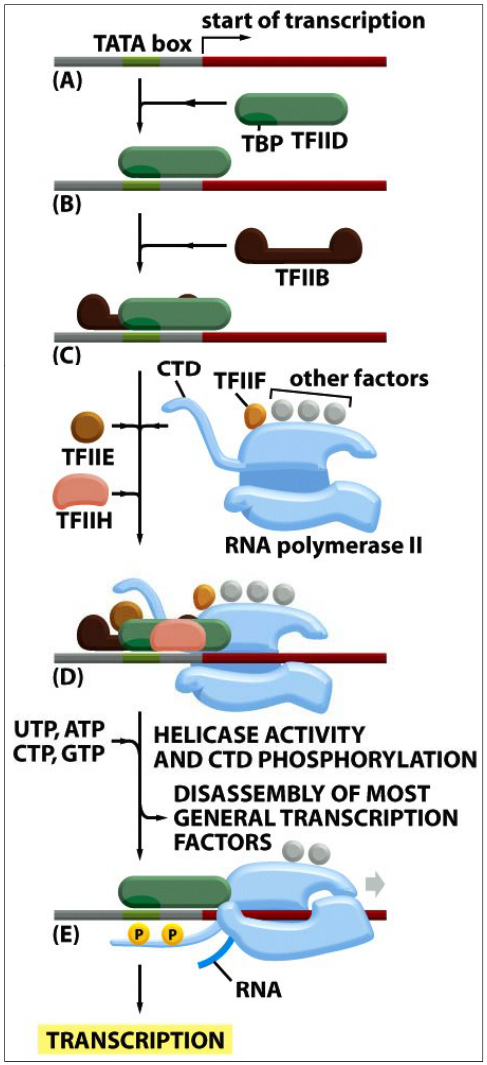
**Table 6–3** The General Transcription Factors Needed for Transcription Initiation by Eucaryotic RNA Polymerase II

| NAME | NUMBER OF SUBUNITS | ROLES IN TRANSITION INITIATION |
|---|---|---|
| TFIID | | |
| TBP subunit | 1 | recognizes TATA box |
| TAF subunits | ~11 | recognizes other DNA sequences near the transcription start point; regulates DNA-binding by TBP |
| TFIIB | 1 | recognizes BRE element in promoters; accurately positions RNA polymerase at the start site of transcription |
| TFIIF | 3 | stabilizes RNA polymerase interaction with TBP and TFIIB; helps attract TFIIE and TFIIH |
| TFIIE | 2 | attracts and regulates TFIIH |
| TFIIH | 9 | unwinds DNA at the transcription start point, phosphorylates Ser5 of the RNA polymerase CTD; releases RNA polymerase from the promoter |

TFIID is composed of TBP and ~11 additional subunits called TAFs (TBP-associated factors); CTD, C-terminal domain.




| element | consensus sequence | general transcription factor |
|---|---|---|
| BRE | G/C G/C G/A C G C C | TFIIB |
| TATA | T A T A A/T A A/T | TBP |
| INR | C/T C/T A N T/A C/T C/T | TFIID |
| DPE | A/G G A/T C G T G | TFIID |



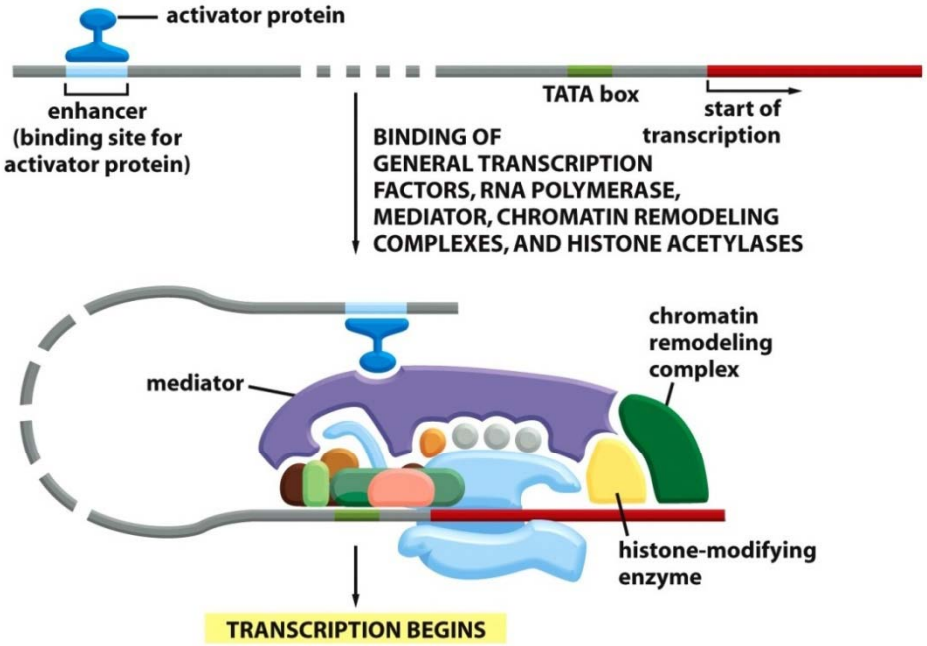

UTP, ATP, CTP, GTP: ribonucleoside triphosphate

# Transcription Initiation and Termination in Eukaryotic Cells (II)

- RNA polymerase II also requires activator, mediator, and chromatin-modifying proteins.
- >100 protein subunits must be assembled to initiate transcripton.
- Why:

  - Must deal with nucleosomes and higher order DNA structures.

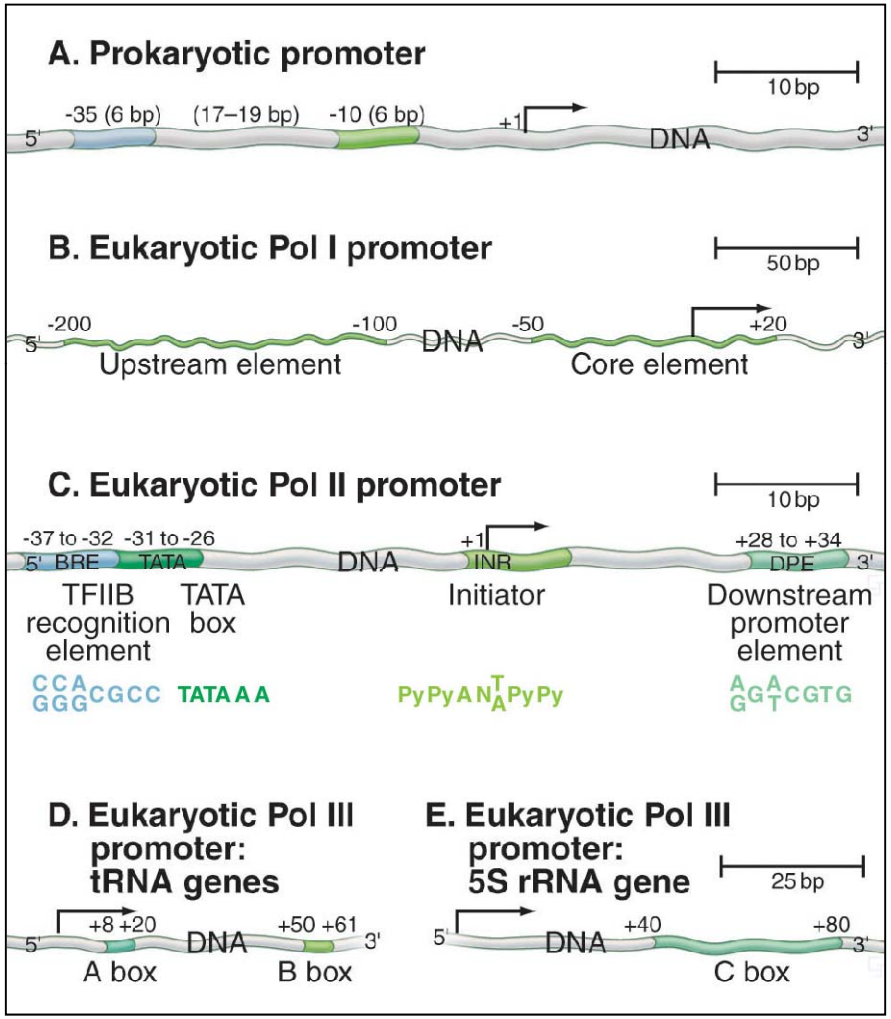
  - Advanced regulatory mechanisms.




**Alberts *MBoC* 5e**

# Transcription Start and Stop Signals

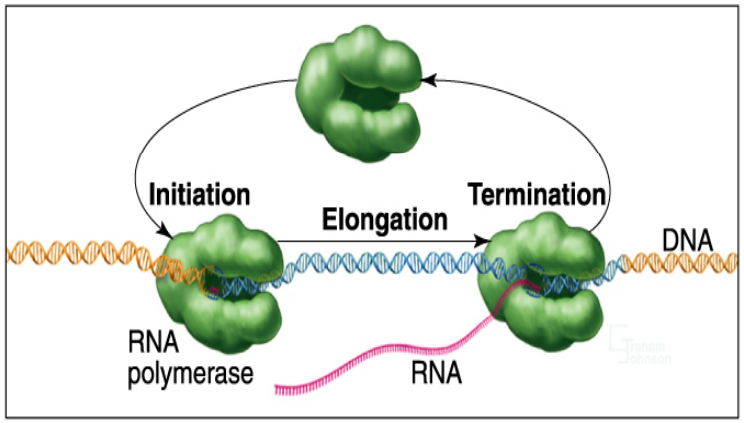
- For RNA polymerase II
  - TATA box
  - Initiators
- Terminator sequences are not always well-specified.
- These signals are heterogeneous in different sequences.



# Transcription Elongation

- Elongation phase begins after around 10 bases are synthesized.

- RNA polymerase conducts multiple processes simultaneously
  - Unwinds DNA in front
  - Reanneals DNA behind
  - Disassociate growing RNA chain from template
  - Perform proofreading




- Elongation factors (proteins) assist movement of RNA polymerases and prevent them from falling off prematurely.

# RNA Capping (I)

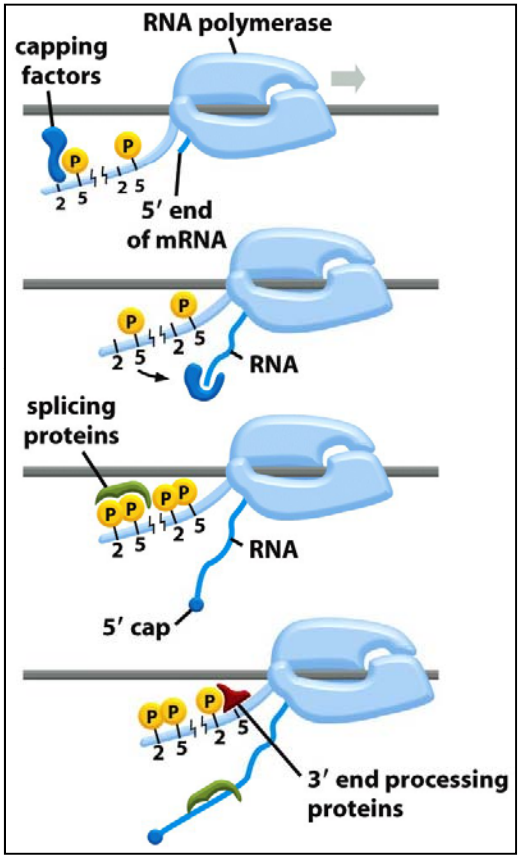
- 5' end capping → splicing → 3' end capping.

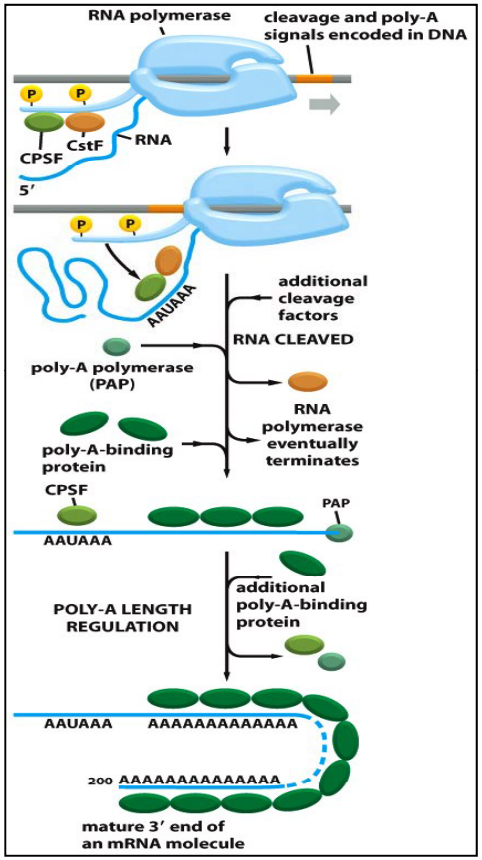
- Main purposes of capping

  - To identify and differentiate mRNA from other RNA's.

  - To protect mRNA from degradation.

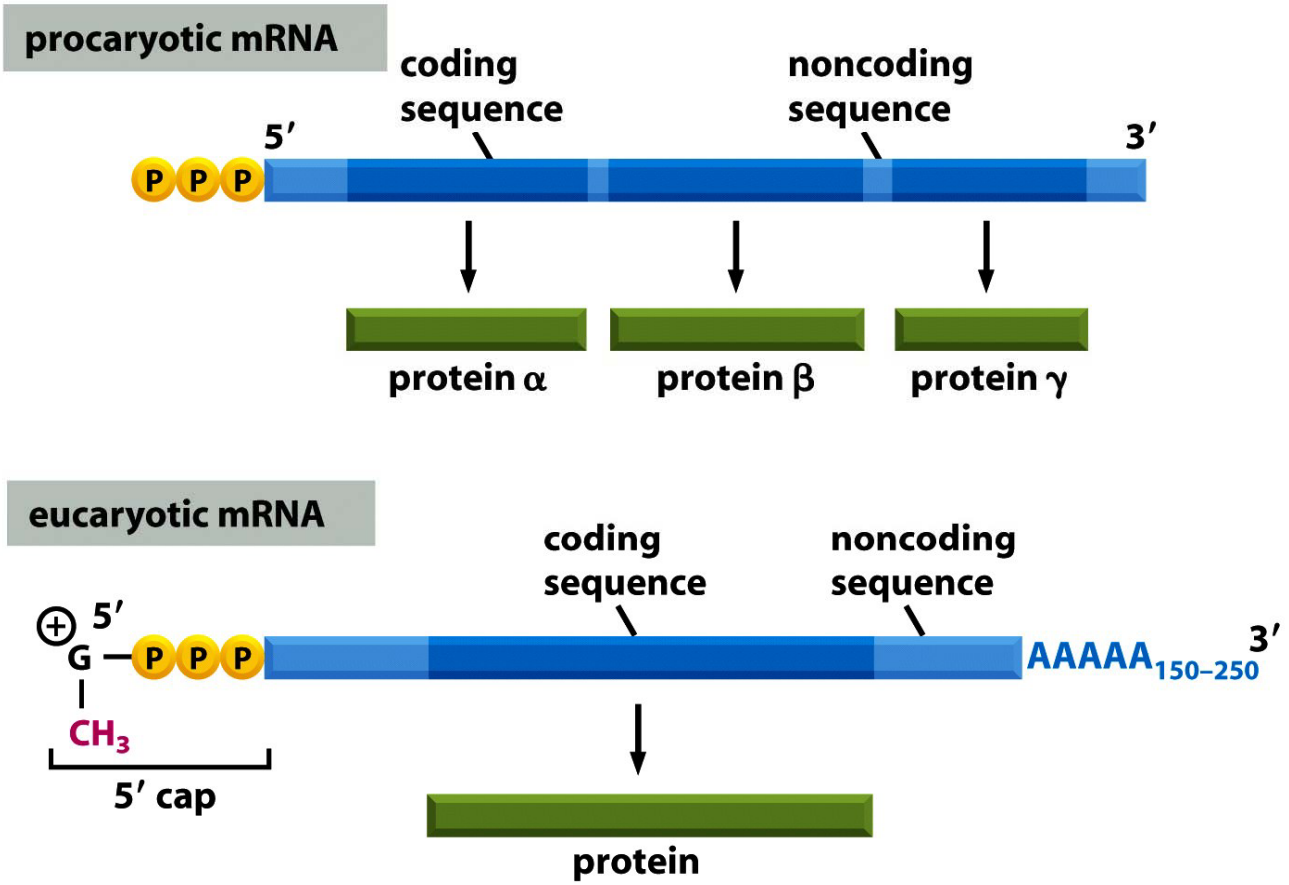
  - To check whether the transcriptions is complete.




CstF: cleavage stimulation factor
CPSF: cleavage and polyadenylation specificity factor

# RNA Capping (II)

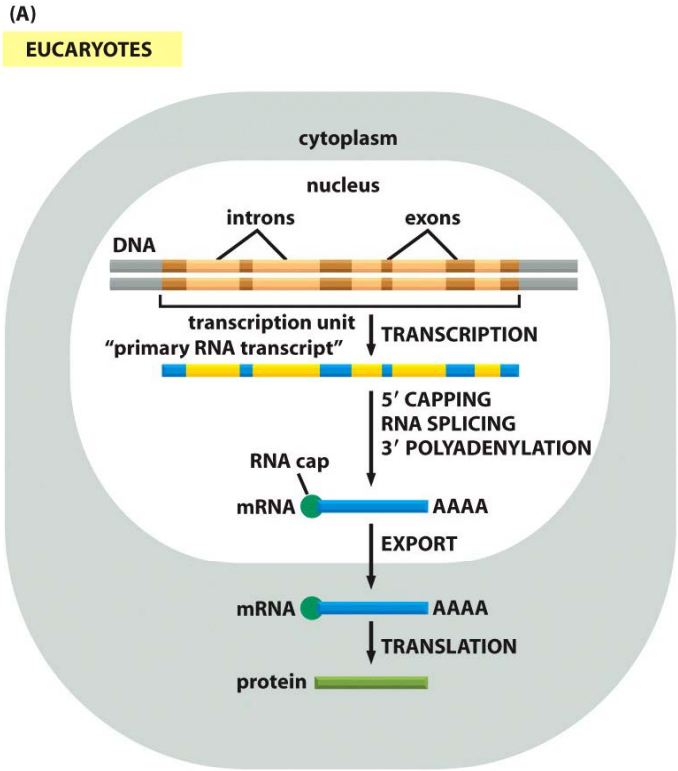


# RNA Splicing in Eukaryotic Cells (I)

- RNA splicing removes introns sequences from newly transcribed pre-mRNA.

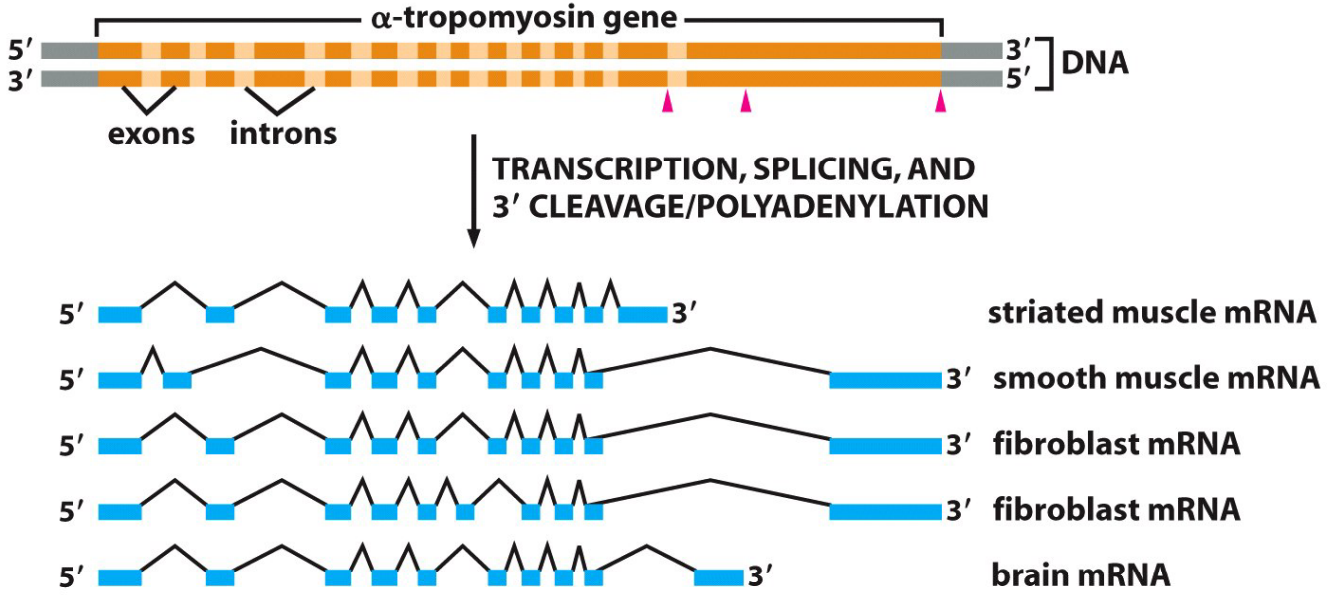
- Splicing is performed by RNA molecules.

- snRNA (small nuclear RNA)
  ↓
  snRNP (small nuclear ribonucleoproteins)
  ↓
  spliceosome



# Alternative Splicing

- Alternative splicing generates different protein isoforms.
- It significantly increases the number of proteins encoded by the genome.



# RNA Splicing in Eukaryotic Cells (II)

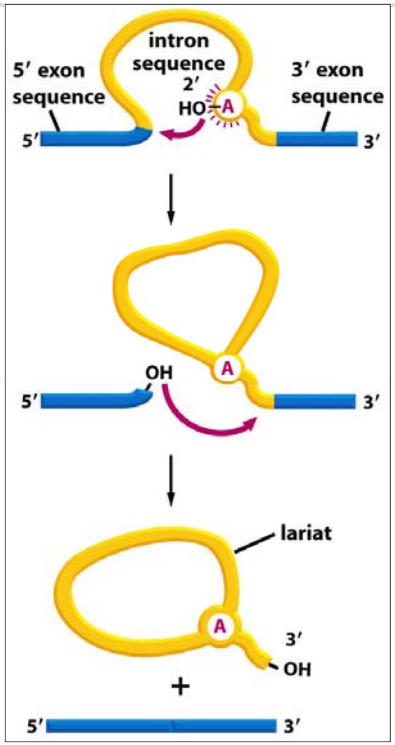
Primary splicing mechanism

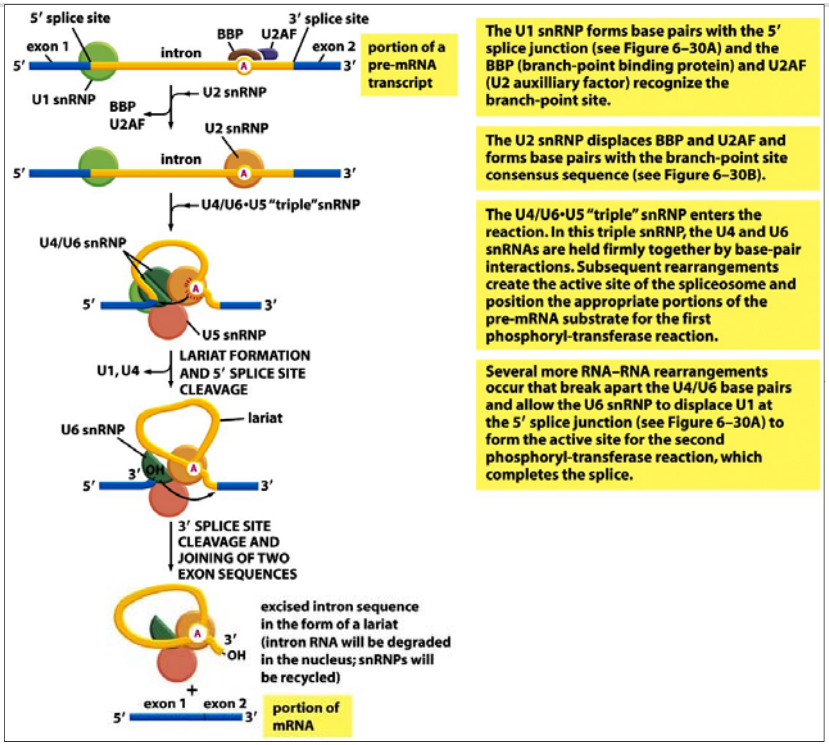
→A: adenine

→5’ splice site: U1 snRNP

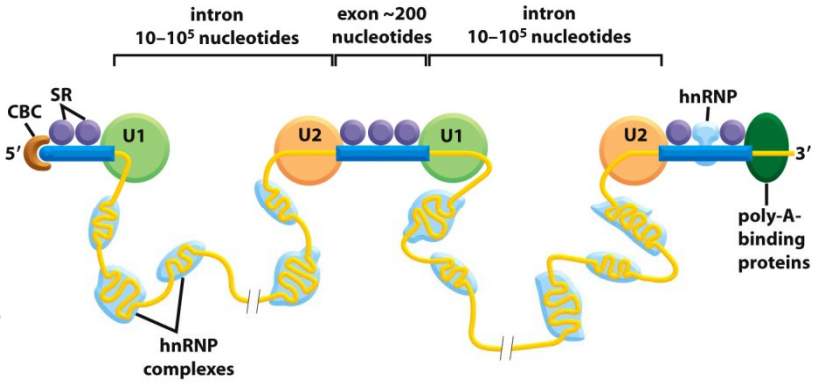
→Branch site: U2 snRNP

→ 3’ spinde site: U6 snRNP




Secondary splicing mechanism in complex eukaryotic cells

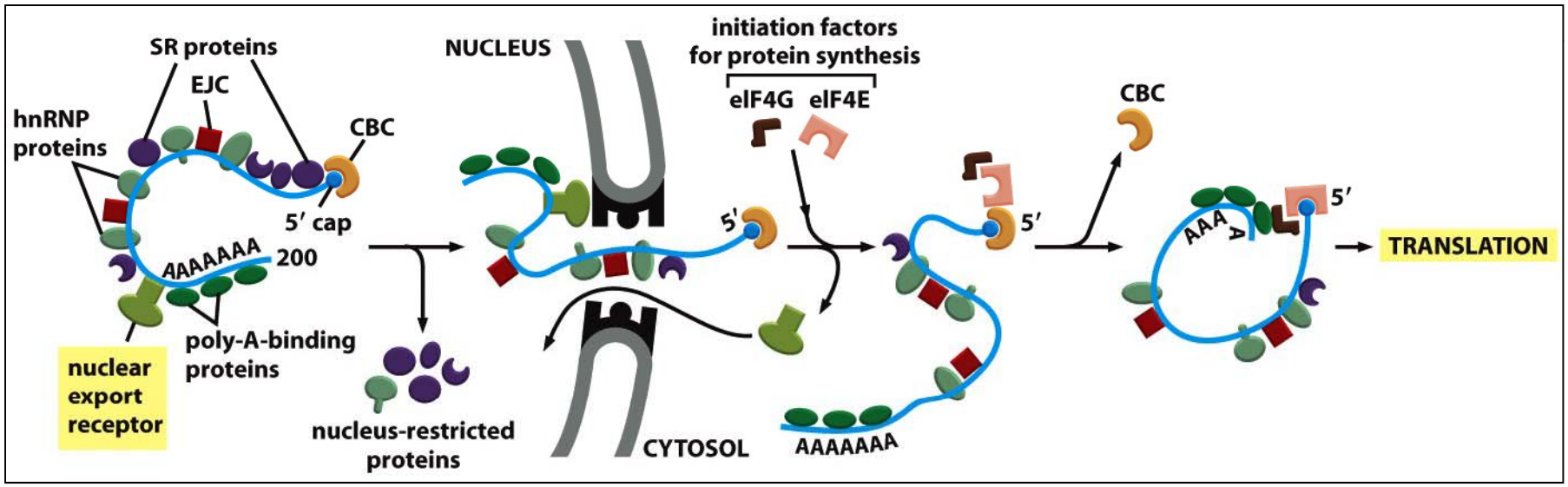


# Outline

- DNA replication and repair
- Overview of transcription
- RNA
- Transcription Process
- Nuclear export of mRNA

# Nuclear Export of RNA

- Export by forming a mRNA-protein complex.
- Nucleus restricted proteins need to unbind.
- Export receptors are re-imported and reused.




Erkmann & Kutay, Nuclear export of mRNA
*Exp. Cell Res.* 296:12, 2004

CBC: cap binding complex
hnRNP: heterogeneous ribonucleoprotein
EJC: exon junction complex
SR: serine-arginine rich protein

# Questions ?