CENTRAL DOGMA: GENE EXPRESSION

By:

Prof. Idowu Akande

CGNPH Summer School July 2025

Learning Objectives

- Meaning of Central Dogma in Molecular Biology
- Definition of Gene
- Overview of gene expression
- Mutations and gene expression
- SNPs -Single Nucleotide Polymorphisms

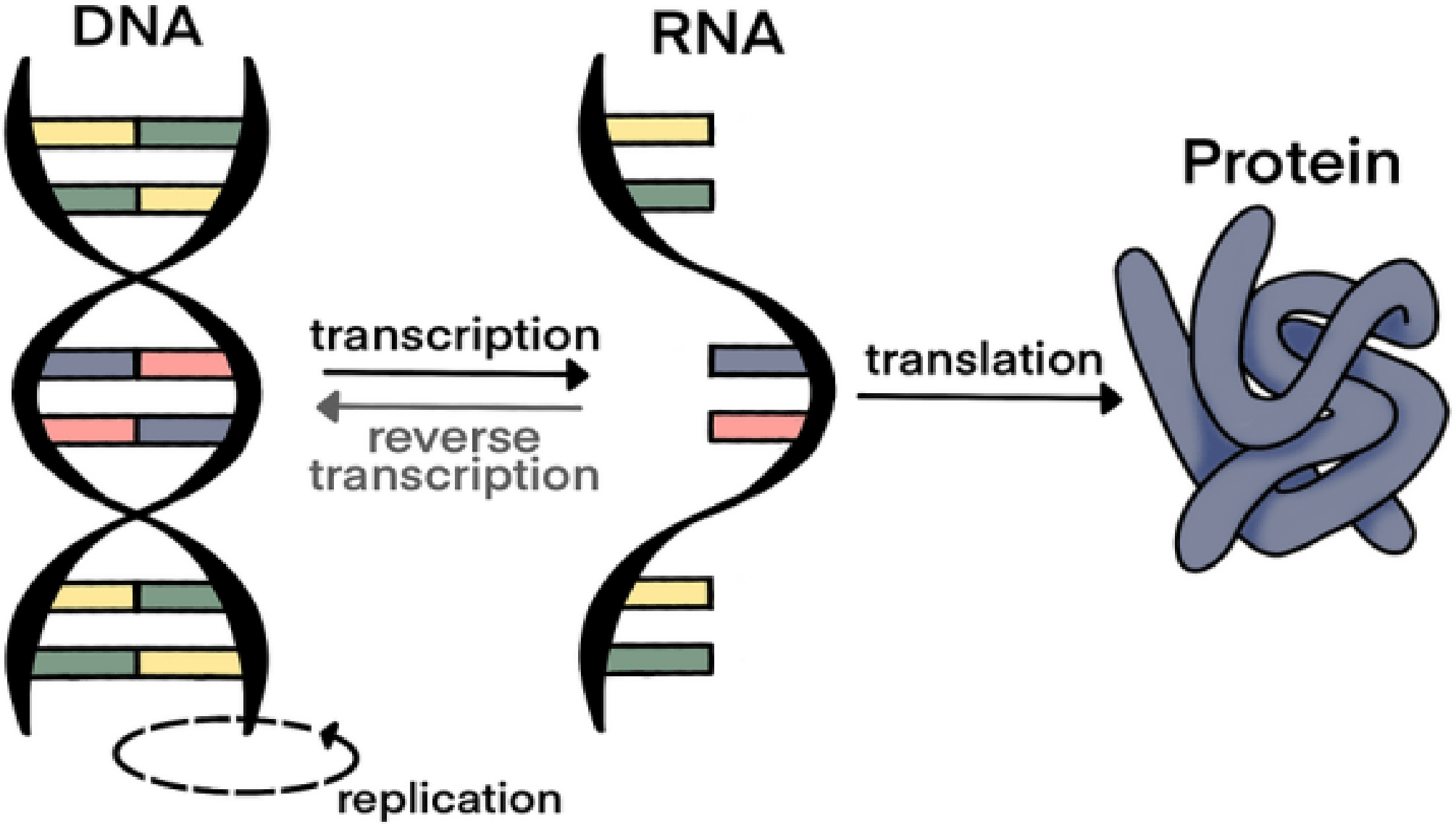
The Central Dogma of Molecular Biology

- **Coined By**: Francis Crick (1958)
- **Definition**: The Central Dogma describes the flow of genetic information in biological systems:

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Protein}$$

- **Significance**: This principle is the foundation of molecular biology, explaining how genetic information stored in DNA is used to create the proteins essential for life.

central dogma

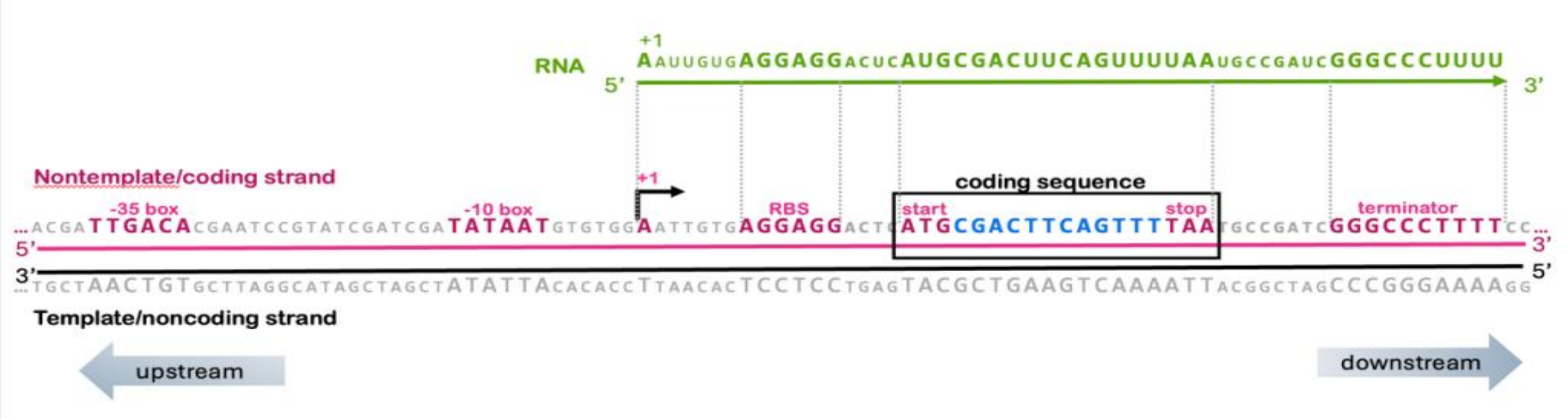


What is a Gene?

- **Definition:** A **gene** is a segment of **DNA** that contains the instructions to produce a functional product, typically a protein, but sometimes RNA (e.g., tRNA, rRNA, miRNA).
- **Structure of a Gene:** Genes are made up of two main parts:
- **Regulatory regions**: Control **when**, **where**, and **how much** a gene is expressed (e.g., **promoter**, **enhancer**)
- **Coding regions**: The actual **DNA sequences** (exons) that are **transcribed and translated** into protein

Furthermore:

- In eukaryotes, genes may also contain **introns** (non-coding) that are spliced out during RNA processing.
- Genes are inherited and determine traits.



Overview of Gene Expression

Gene Expression is the process by which genetic information in DNA is used to produce functional molecules (mainly proteins).

Two Main Steps:

1. Transcription

- DNA is transcribed into RNA
- Catalyzed by RNA polymerase
- **Product**: pre-mRNA (eukaryotes) or mRNA (prokaryotes)

2. Translation

- RNA is translated into protein
- Occurs on ribosomes
- **Product**: Polypeptide/protein

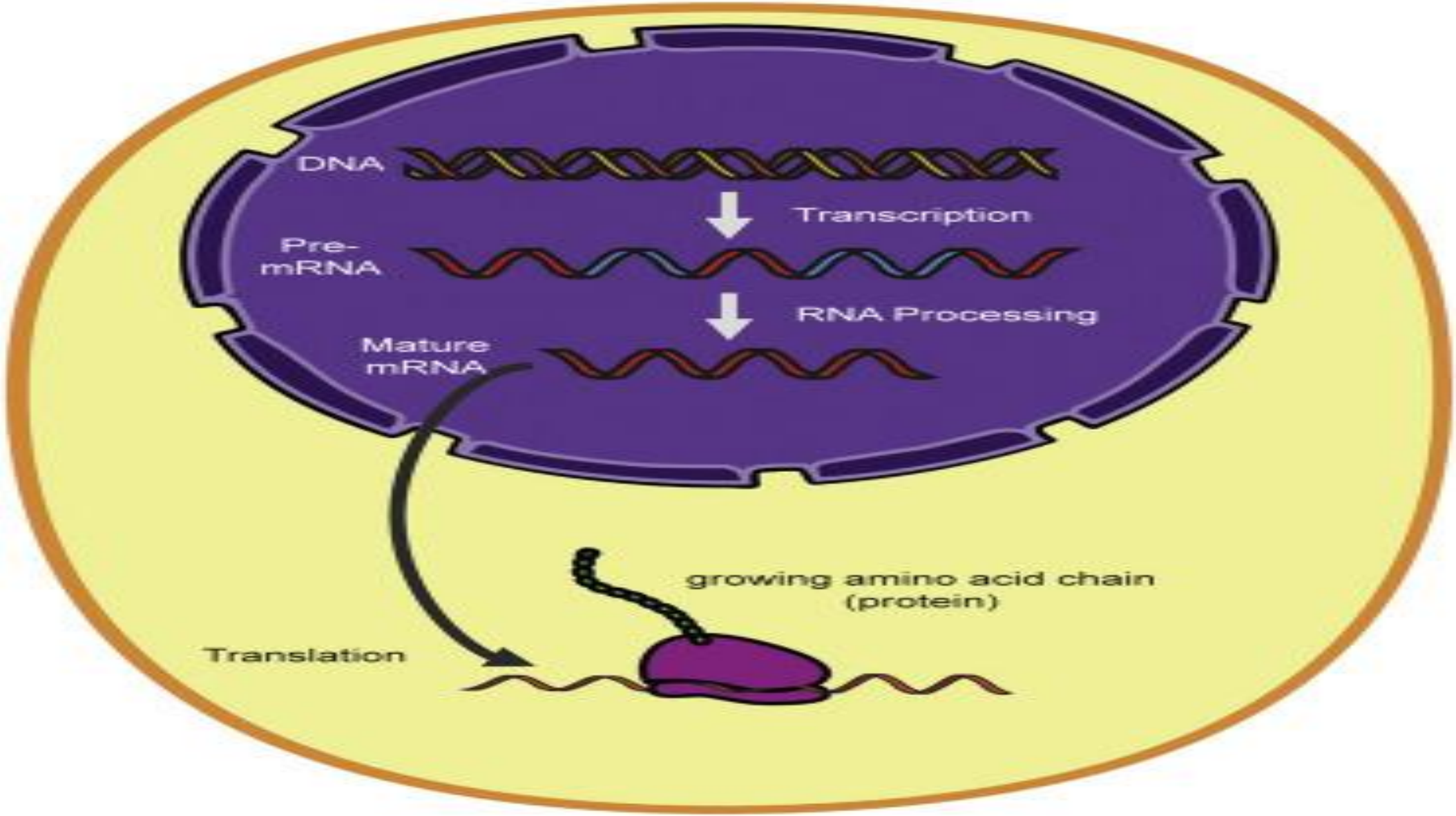


Fig 3: A cell diagram showing: Nucleus with transcription symbol (DNA → mRNA) Cytoplasm with ribosome translating mRNA into protein

DNA Structure Recap

DNA (Deoxyribonucleic Acid)

- A double-helix molecule composed of nucleotides
- Each nucleotide contains: A phosphate group; deoxyribose sugar; and a nitrogenous base (A, T, C, or G)

Base Pairing Rules

- Adenine (A) pairs with Thymine (T)
- Cytosine (C) pairs with Guanine (G)
- Held together by hydrogen bonds

Key Properties

- **Antiparallel strands**: One strand runs 5′→3′, the other 3′→5′
- **Complementary**: Each base has a specific partner, allowing accurate replication and transcription

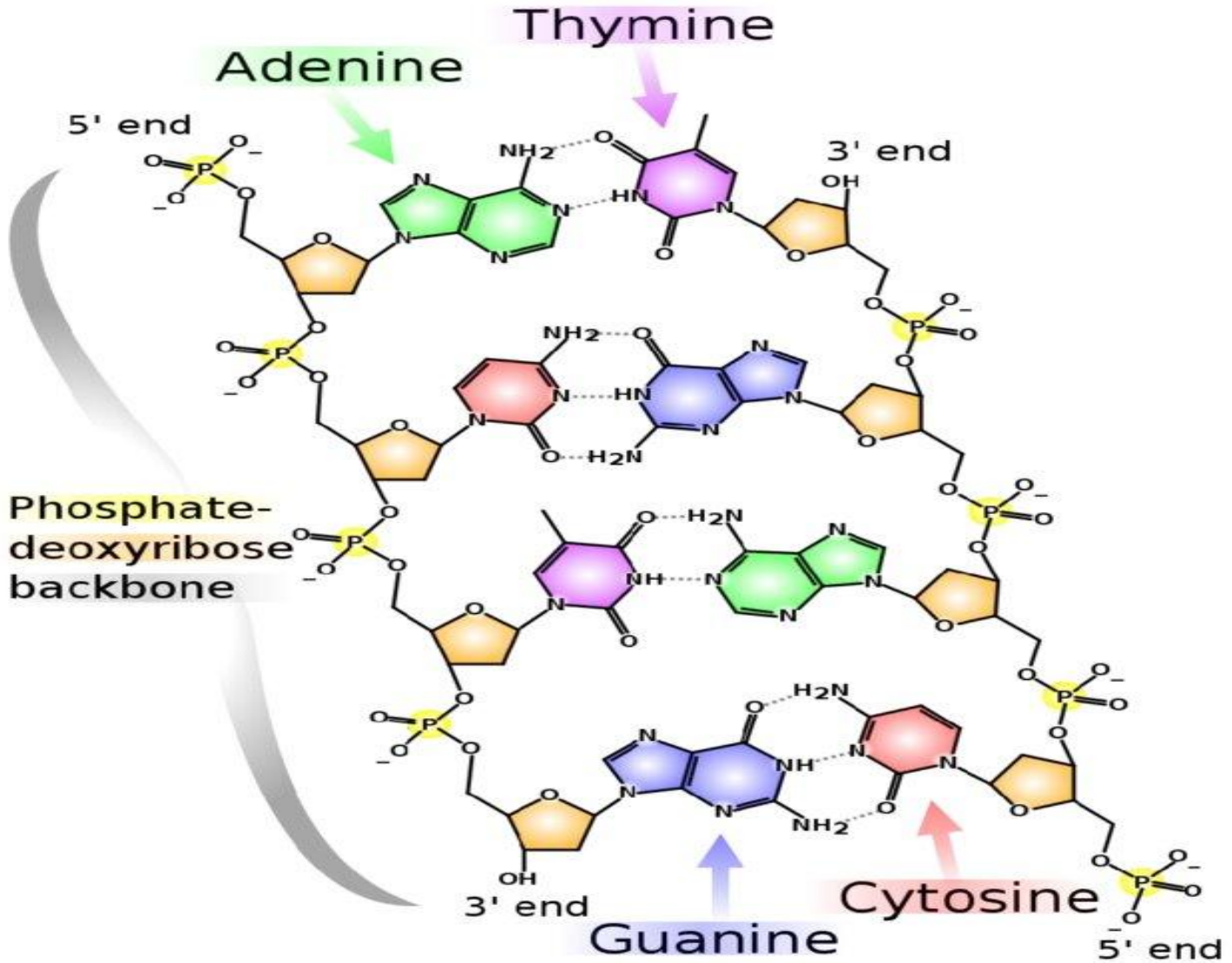


Fig. 4: A schematic view of a DNA double-helix structure. Two single DNA strands comprised of the sugar-phosphate backbone and the alternating bases are linked together via Watson-Crick base pairing.

Transcription

➢ What is Transcription?

Transcription is the process by which RNA is synthesized from a DNA template.

➢ Key Steps in Transcription:

- **Initiation**: RNA Polymerase binds to a promoter region upstream of the gene
- **Elongation**: RNA Polymerase reads the DNA template strand; Synthesizes a complementary strand of pre-mRNA (5′ → 3′ direction)
- **Termination**: Transcription stops at a termination signal, releasing the pre-mRNA

➢ Important Notes:

- Only one strand of DNA is used as a template
- The resulting mRNA sequence is complementary to the template strand and identical (except U for T) to the coding strand

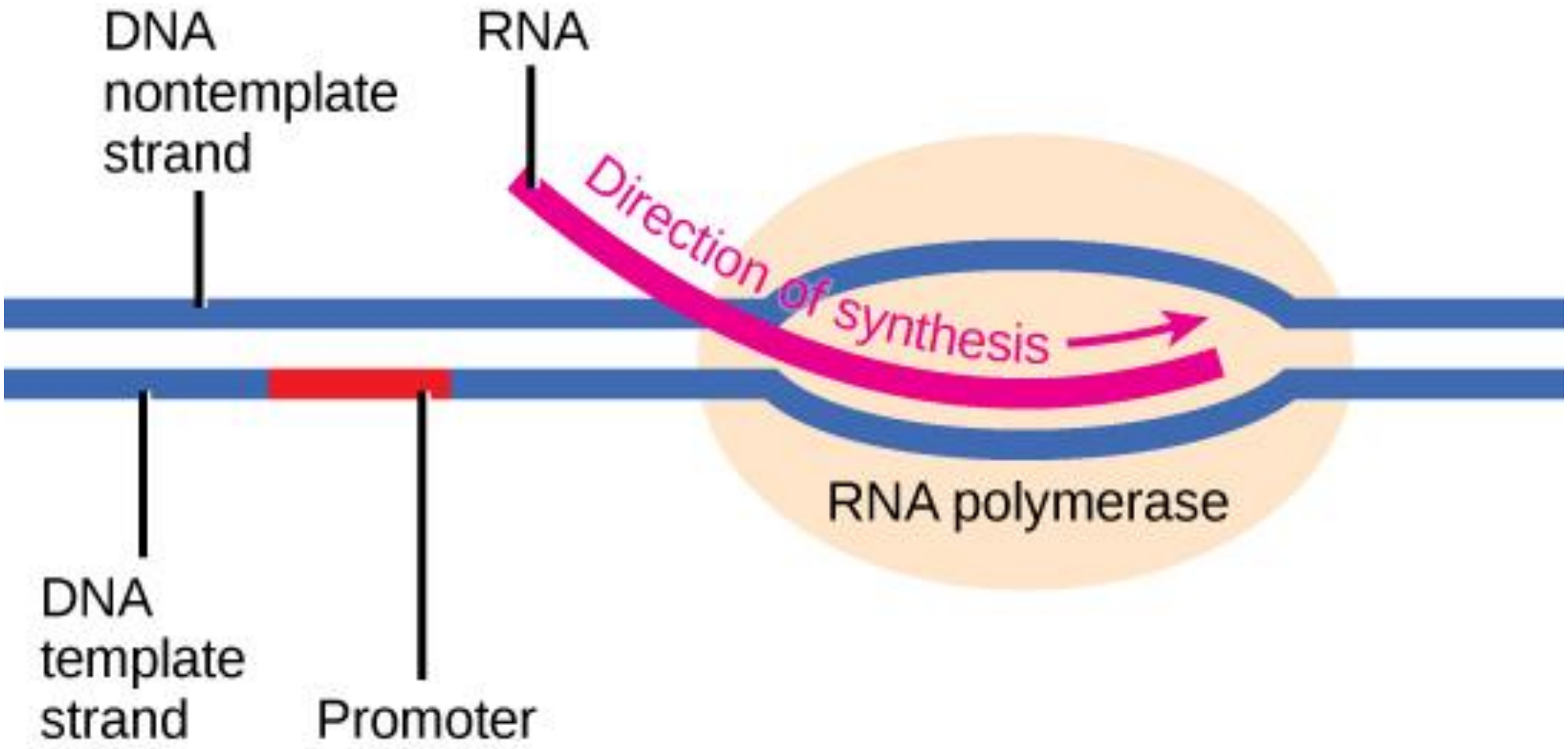


Fig.5: The initiation of transcription begins when DNA is unwound, forming a transcription bubble. Enzymes and other proteins involved in transcription bind at the promoter.

RNA Processing (Eukaryotes)

From Pre-mRNA to Mature mRNA

In eukaryotic cells, the initial transcript (pre-mRNA) undergoes several modifications before becoming mature mRNA ready for translation.

Key RNA Processing Steps:

- **5′ Capping**
  - Addition of a modified guanine nucleotide at the 5′ end
  - Functions: protects mRNA, aids ribosome binding
- **3′ Poly-A Tail**
  - Addition of ~200 adenine nucleotides to the 3′ end
  - Functions: increases mRNA stability, facilitates export from nucleus
- **Splicing**
  - Removal of introns (non-coding regions)
  - Exons (coding regions) are joined together
  - Catalyzed by the spliceosome

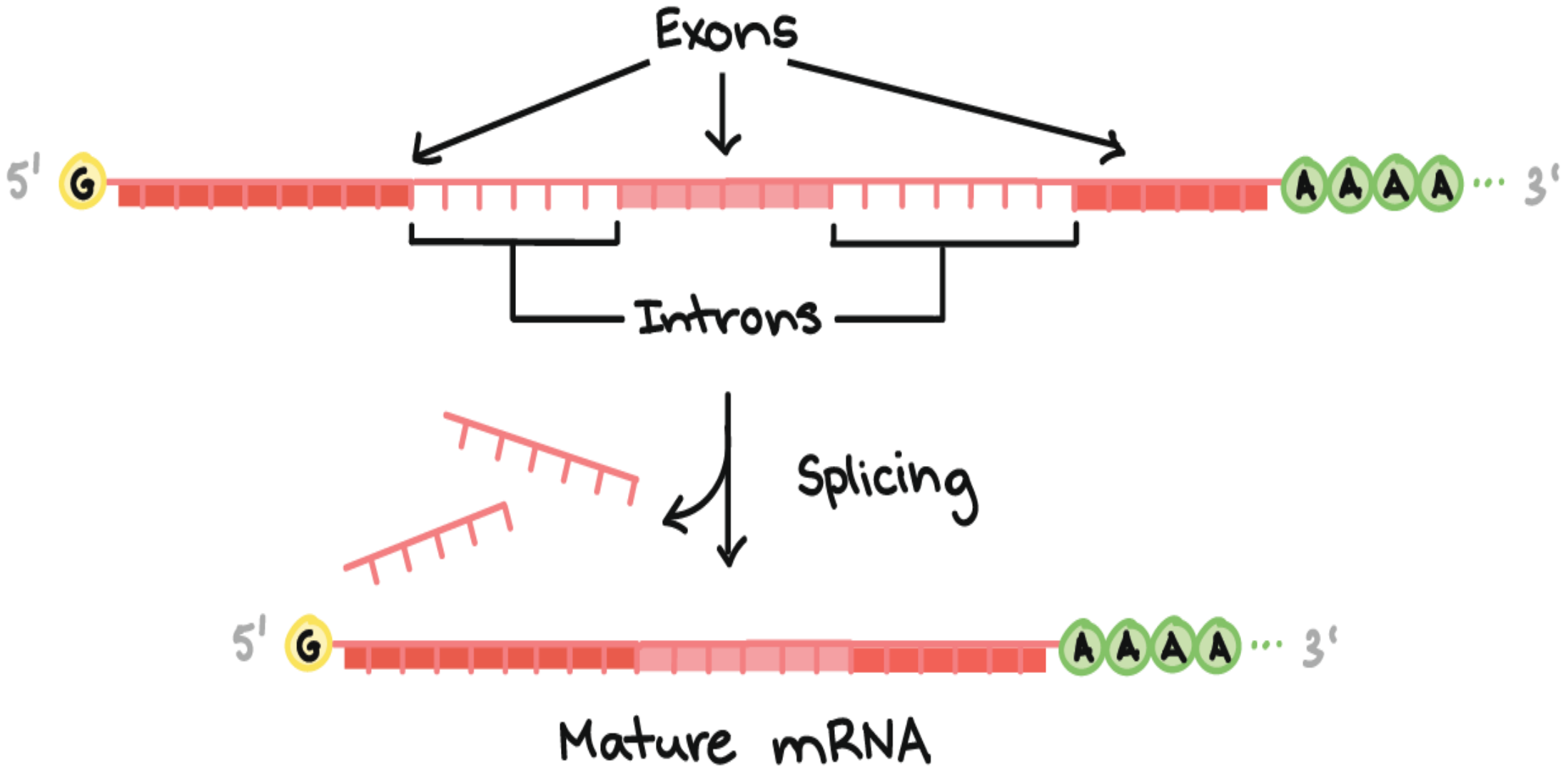


Fig. 6: RNA Processing (Eukaryotes)

Translation

What is Translation?

Translation is the process by which the genetic code carried by mRNA is decoded to build a polypeptide chain (protein).

Where It happens:

Occurs on ribosomes, found in the cytoplasm (or on rough ER)

Key Features of Translation:

➢ **Codons**:

- mRNA is read in triplets of nucleotides (called codons)
- Each codon specifies a particular amino acid

➢ **tRNA (Transfer RNA):**

- Carries the correct amino acid to the ribosome
- Each tRNA has an anticodon complementary to the mRNA codon

➢ **Protein Assembly**:

- Amino acids are joined by peptide bonds to form a polypeptide
- Process continues until a stop codon is reached

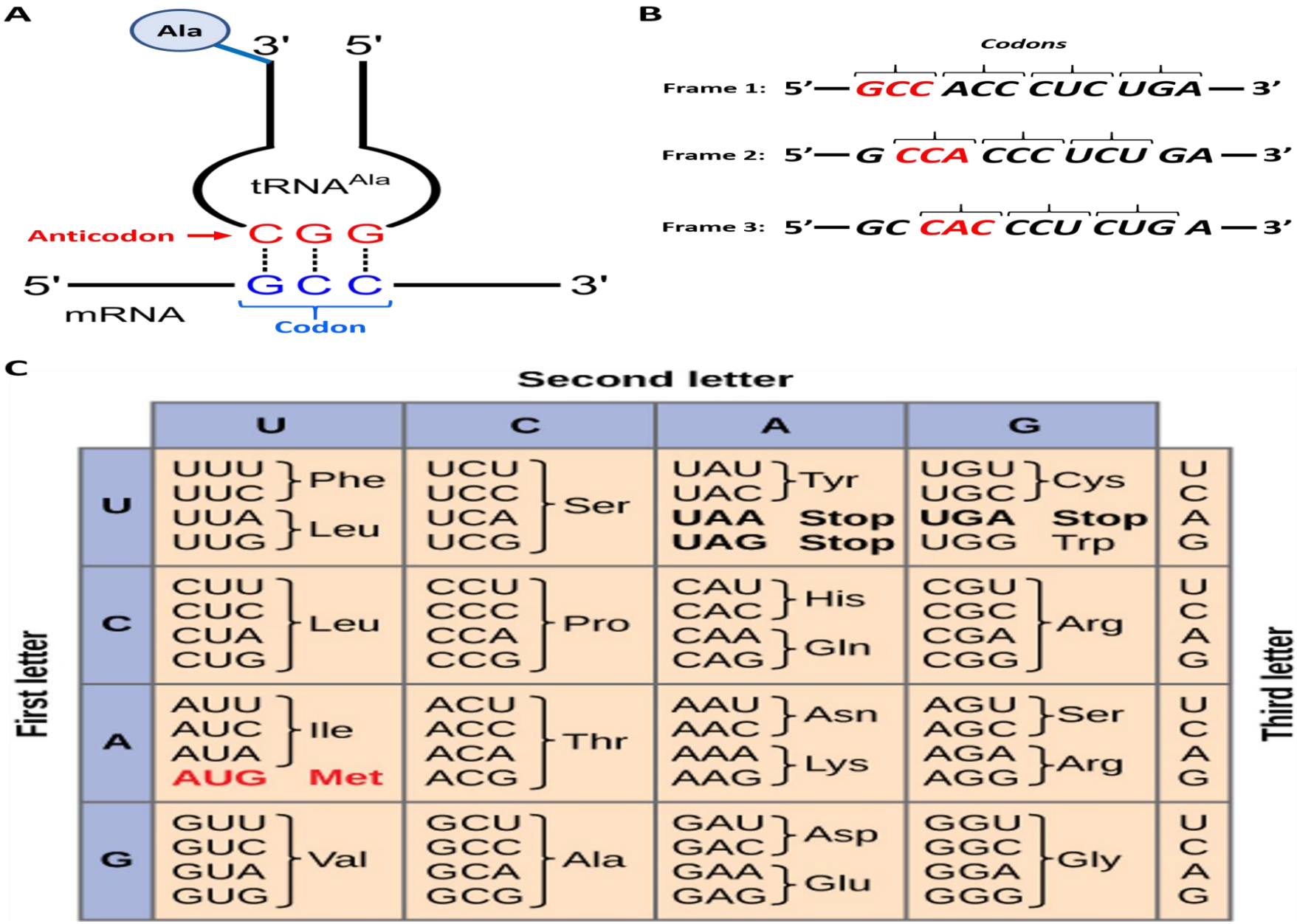


Fig. 7: Reading the mRNA Template

Stages of Translation

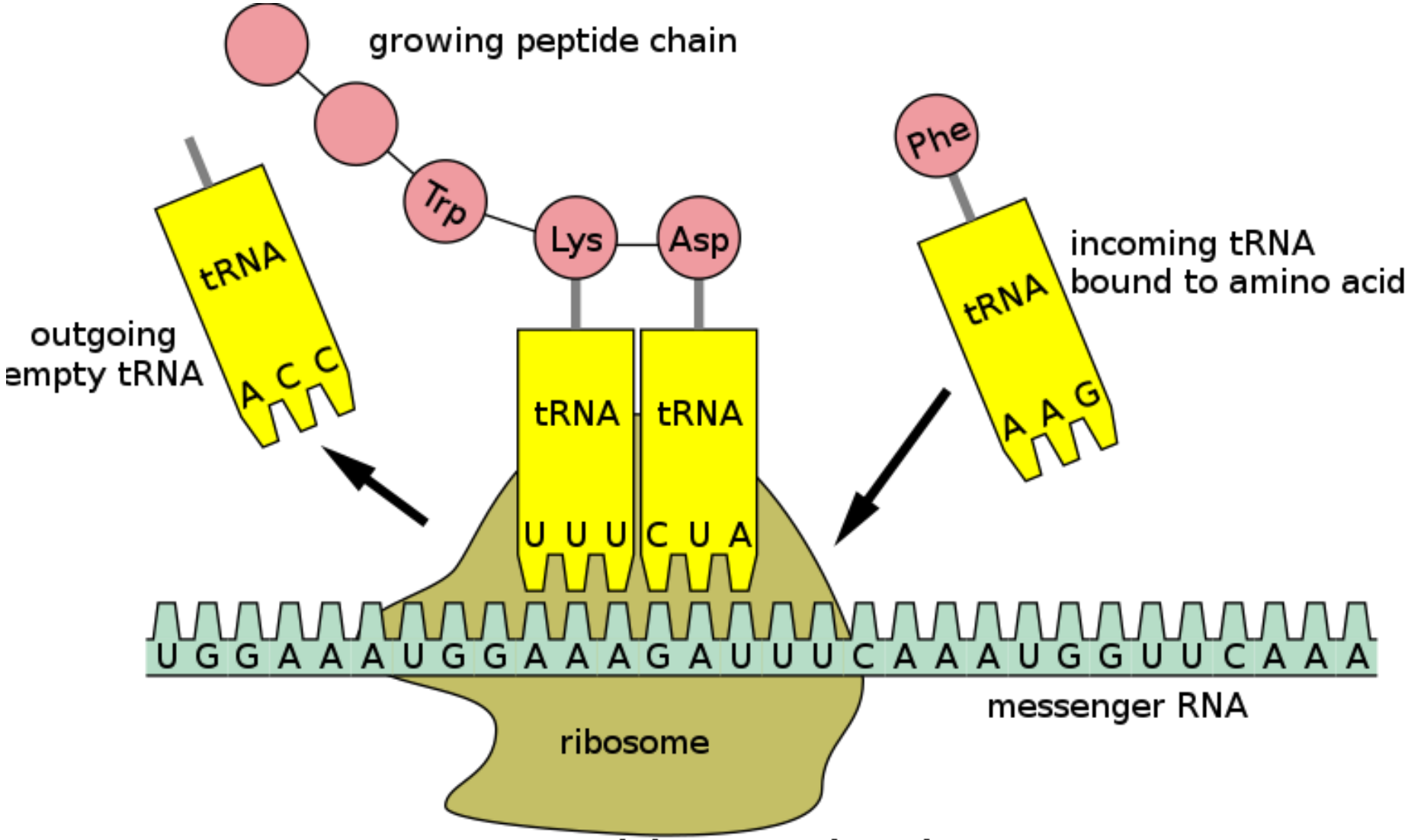
Translation occurs in three major stages, each playing a critical role in converting genetic information into a functional protein.

➢ **Initiation**

- The small ribosomal subunit binds to the mRNA at the start codon (AUG)The initiator tRNA carrying Methionine pairs with AUG
- The large ribosomal subunit joins to form the full translation complex

➢ **Elongation**

- Ribosome moves along mRNA, reading codons
- tRNAs bring the corresponding amino acids
- Amino acids are linked by peptide bonds to form a growing polypeptide chain



Peptide Synthesis

The Genetic Code

What is the Genetic Code?

- The genetic code is the set of rules by which the sequence of nucleotide triplets (codons) in mRNA is translated into amino acids in proteins.

Key Features:

➢ **Triplet Code**

- Each codon is made up of 3 RNA nucleotides
- There are 64 codons in total: 61 code for amino acids and 3 are stop codons (UAA, UAG, UGA)

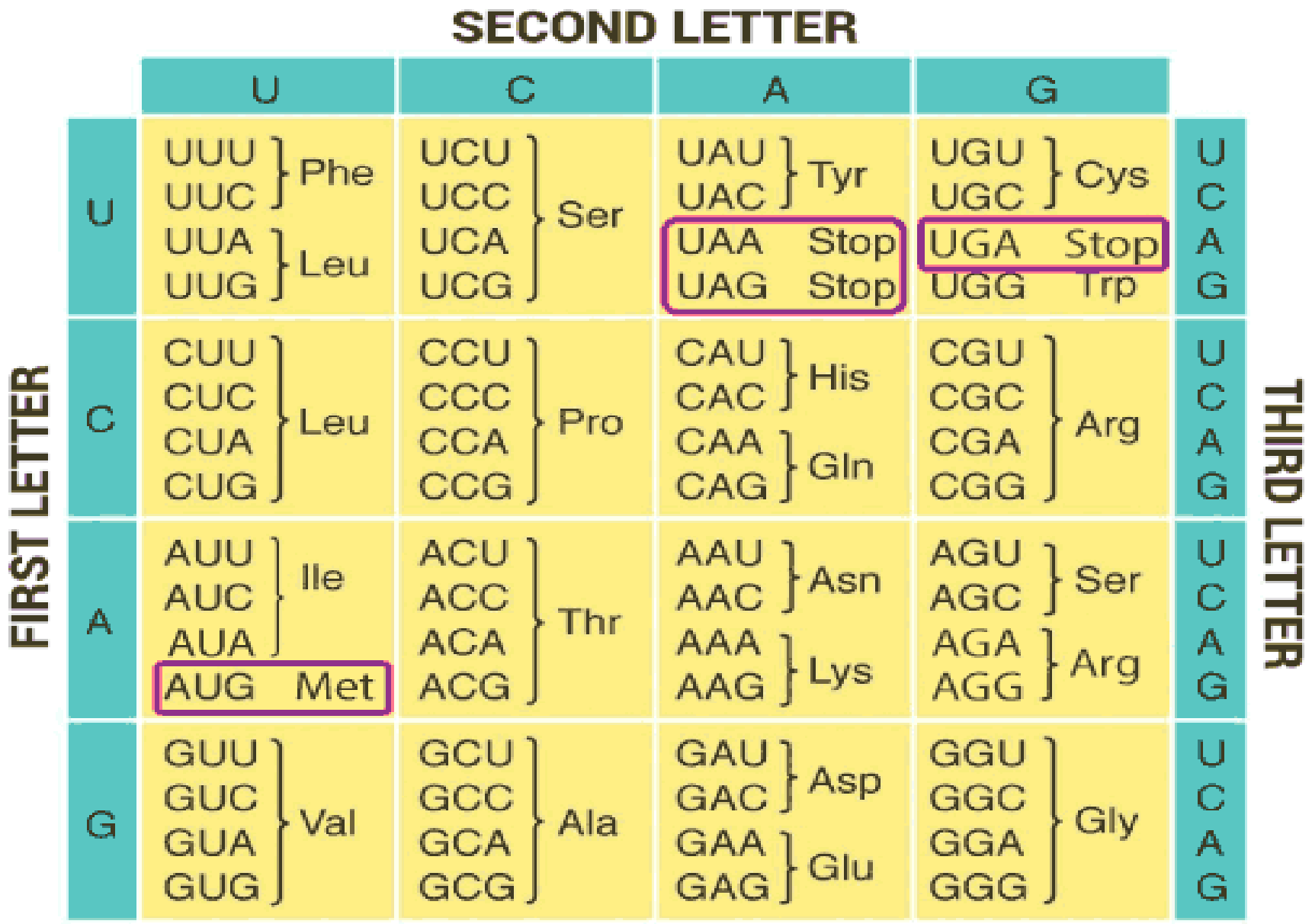
➢ **Redundancy, Not Ambiguity**

- Multiple codons can specify the same amino acid
- But each codon corresponds to only one amino acid
- Example: GGU, GGC, GGA, and GGG all code for Glycine

➢ **Universal Code**

- The genetic code is nearly universal
- Shared by almost all organisms, from bacteria to humans

GENETIC CODE TABLE



Prokaryotic vs. Eukaryotic Gene Expression

- Gene expression differs significantly between **prokaryotic** and **eukaryotic** organisms due to differences in cellular structure and gene organization.

Feature	**Prokaryotes**	**Eukaryotes**
Location	Cytoplasm	Nucleus (Transcription) & Cytoplasm (Translation)
mRNA Processing	None	5′ Capping, 3′ Poly-A Tailing, Splicing
Genes	Polycistronic (multiple proteins per mRNA)	Monocistronic (one gene per mRNA)

Regulation of Gene Expression

- Gene expression is tightly controlled to ensure the right genes are expressed at the right time, in the right cells, and in the right amount.

Levels of Regulation:

- **Transcriptional Regulation**
  ➤ Controls whether a gene is transcribed
  ➤ Involves transcription factors, enhancers, repressors
  ➤ Most energy-efficient control point
- **Post-Transcriptional Regulation**
  ➤ Includes RNA splicing, mRNA stability, RNA interference (RNAi)
  ➤ Modulates mRNA availability
- **Translational Regulation**
  ➤ Controls how efficiently mRNA is translated
  ➤ Influenced by mRNA structure, ribosome availability
- **Post-Translational Regulation**
  ➤ Modifications to proteins (e.g., phosphorylation, ubiquitination)
  ➤ Affects protein function and lifespan

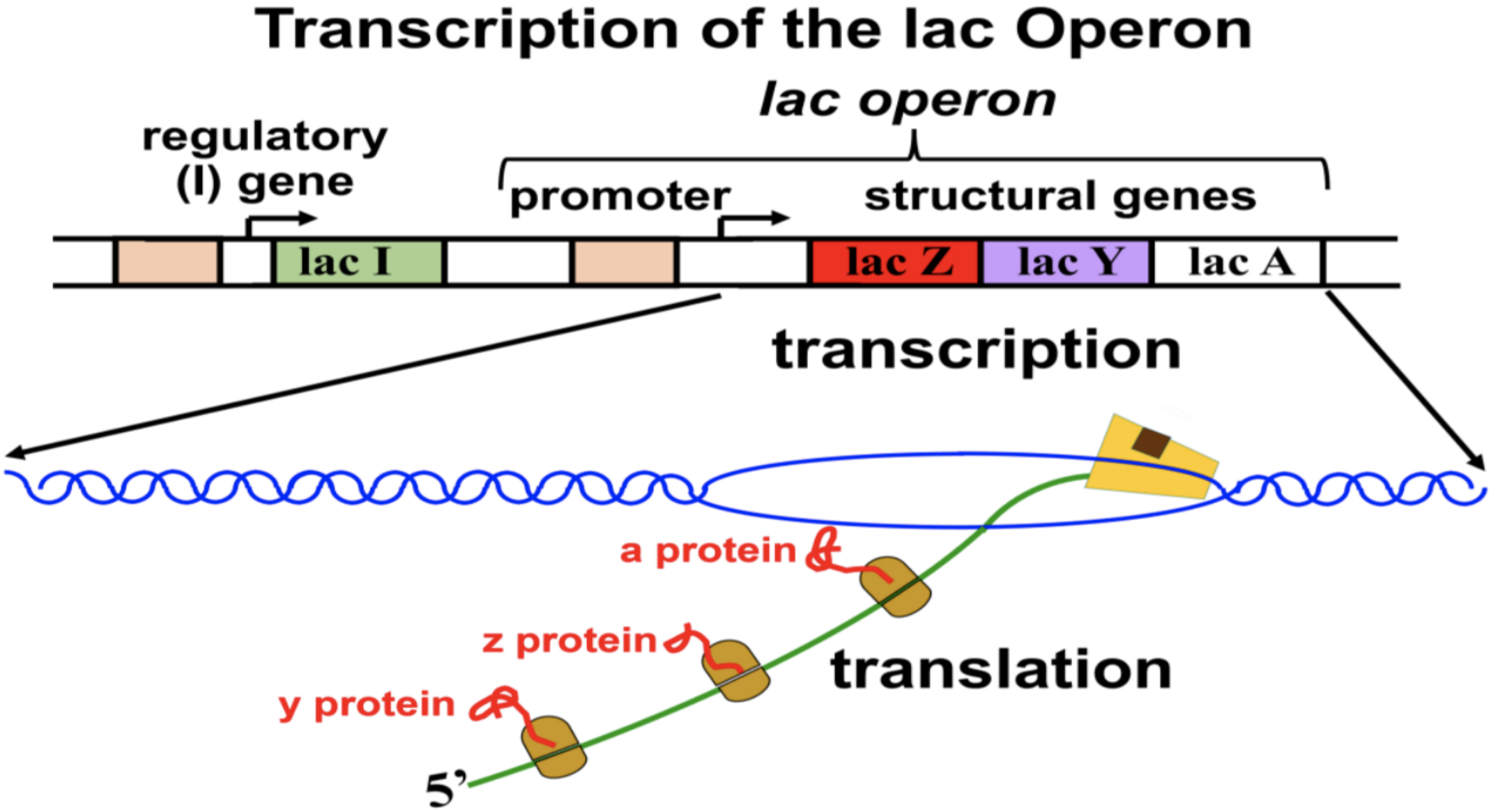


Fig. 10: Overview of Prokaryotic Gene Regulation

Mutations and Gene Expression

Mutations are permanent changes in the DNA sequence that can impact how genes are expressed and how proteins function.

Types of Mutations:

➢**Point Mutation:** A single nucleotide change

- Silent: No change in amino acid
- Missense: Different amino acid
- Nonsense: Early stop codon

Regulatory Region Mutations

- Changes in promoter/enhancer sequences
- Can affect when, where, or how much a gene is expressed

Effects on Gene Expression:

- Loss of function: Protein is nonfunctional or absent
- Gain of function: Protein is overactive or expressed inappropriately
- Neutral: No significant change in function or expression

Example:

Point Mutations

Silent: has no effect on the protein sequence

Missense: results in an amino acid substitution

Nonsense: substitutes a stop codon for an amino acid

Central Dogma Exceptions

- While the Central Dogma explains the general flow of genetic information (DNA → RNA → Protein), there are exceptions observed in certain organisms and molecular processes.

1. Reverse Transcription

- Information flows from RNA back to DNA
- Carried out by reverse transcriptase
- Found in retroviruses like HIV

RNA → DNA → RNA → Protein

2. RNA Replication

- Some RNA viruses replicate their RNA genomes directly

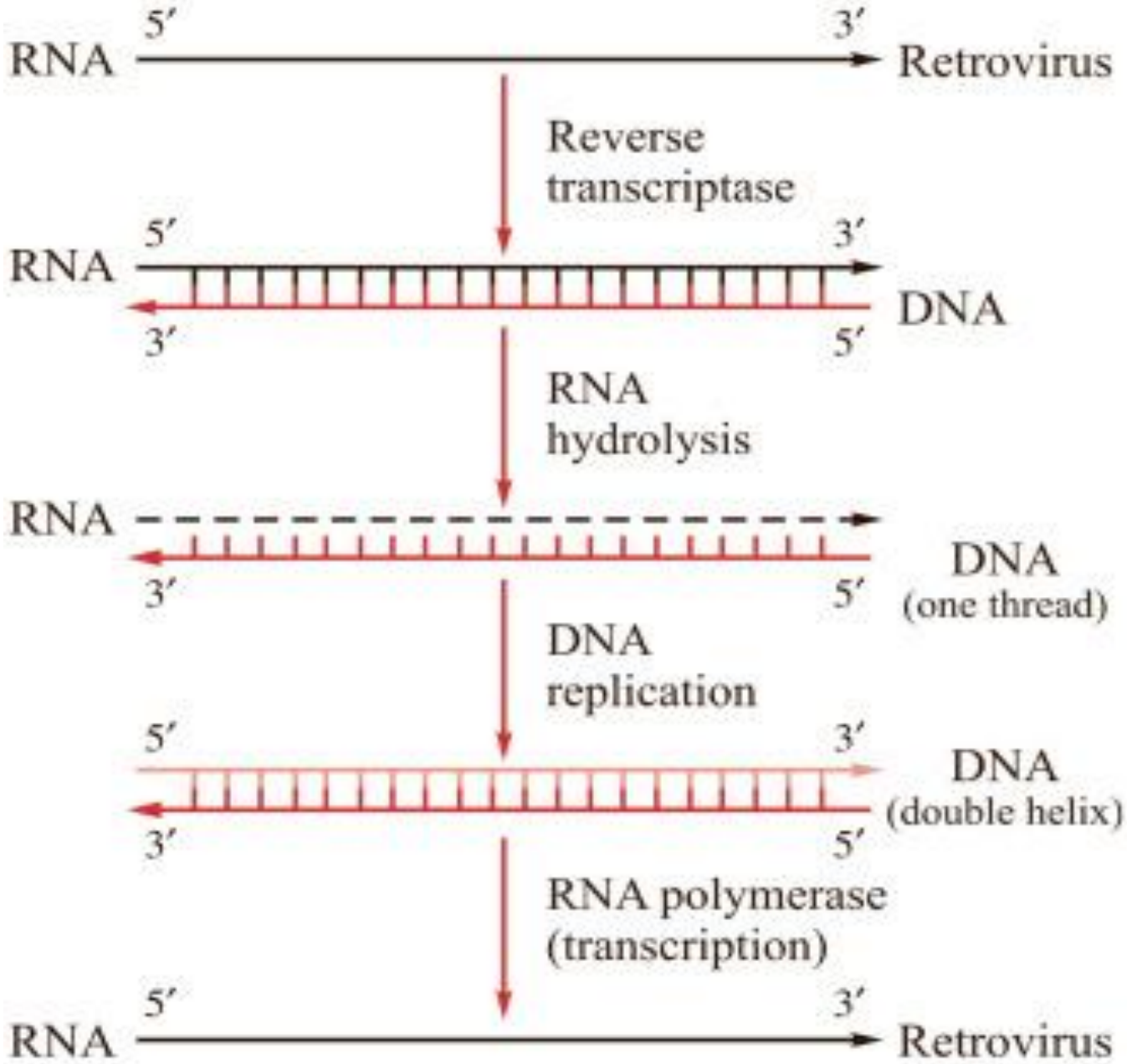
3. Non-Coding RNAs (ncRNAs)

- Not all RNA is translated into protein

Examples:

- rRNA: Forms part of ribosomes
- tRNA: Transfers amino acids
- miRNA, siRNA: Regulate gene expression

DNA → RNA (functional RNA, not protein)


RNA
5′
3′
Retrovirus
Reverse transcriptase
RNA
5′
3′
DNA
3′
5′
RNA hydrolysis
RNA
3′
5′
DNA (one thread)
DNA replication
5′
3′
DNA (double helix)
3′
5′
RNA polymerase (transcription)
RNA
5′
3′
Retrovirus

SNPs – Single Nucleotide Polymorphisms

What are SNPs?

- SNPs (pronounced "snips") are single base-pair variations in the DNA sequence that occur in at least 1% of the population.

Example:

- Normal: AAGCCTA
- SNP Variant: AAGTCTA

Key Characteristics:

- Most common type of **genetic variation**
- Occur every **300–1,000 bases** in the human genome
- Usually found in **non-coding regions**, but can also occur in **coding**

Impact of SNPs on Gene Expression

- May alter promoter or enhancer activity
- Can affect splicing, mRNA stability, or protein function
- Some SNPs are associated with disease susceptibility or drug response

Applications of SNPs on Gene Expression

- **Pharmacogenomics**: Tailoring drug treatments based on SNP profiles
- **Genome-wide association studies (GWAS)**: Linking SNPs to complex traits/diseases
- **Ancestry and population genetics**
- **Biomarker discovery** in personalized medicine

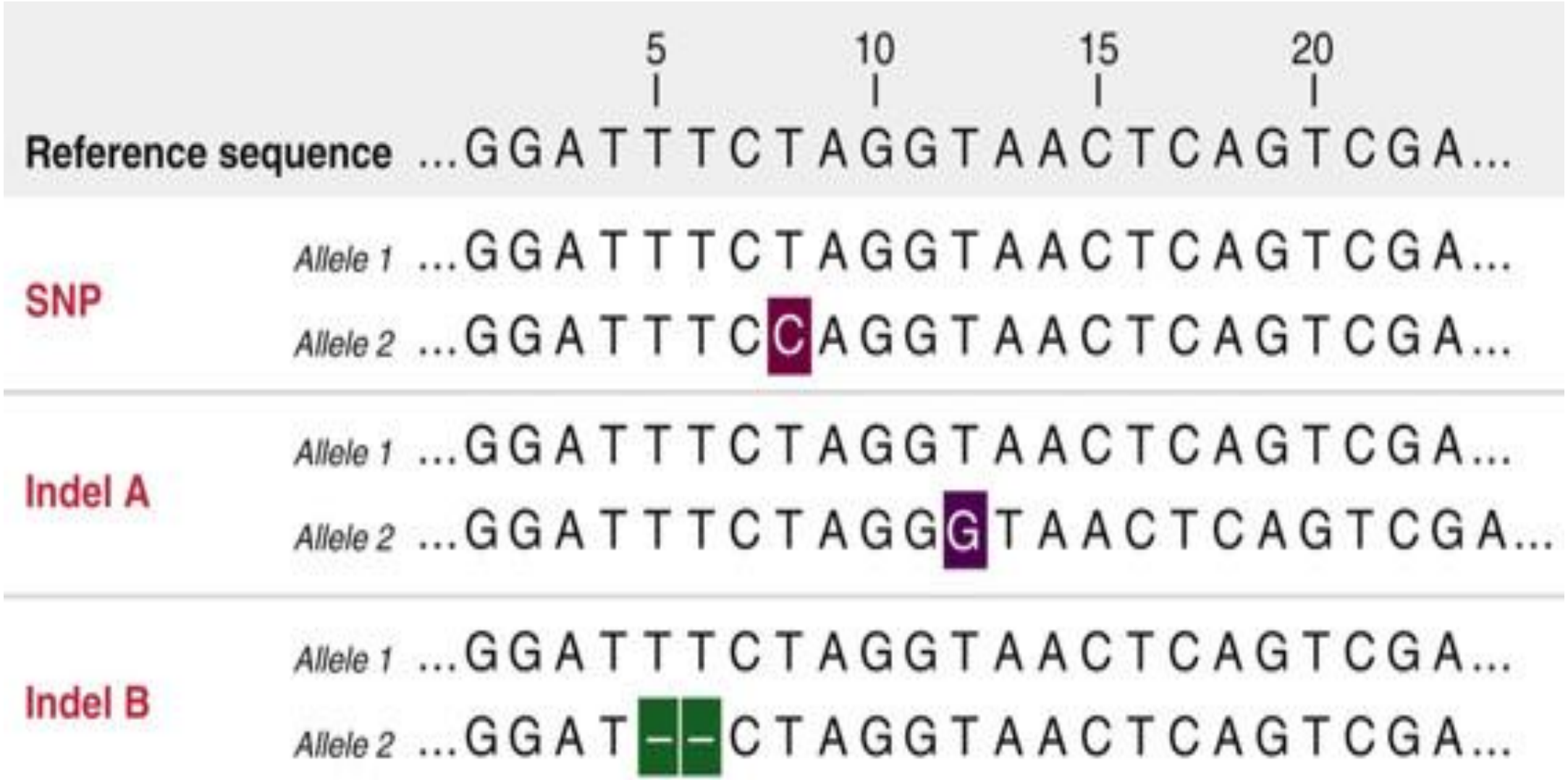


Fig. 11: Three polymorphisms in genomic DNA from the segment of the human genome reference assembly shown at the top