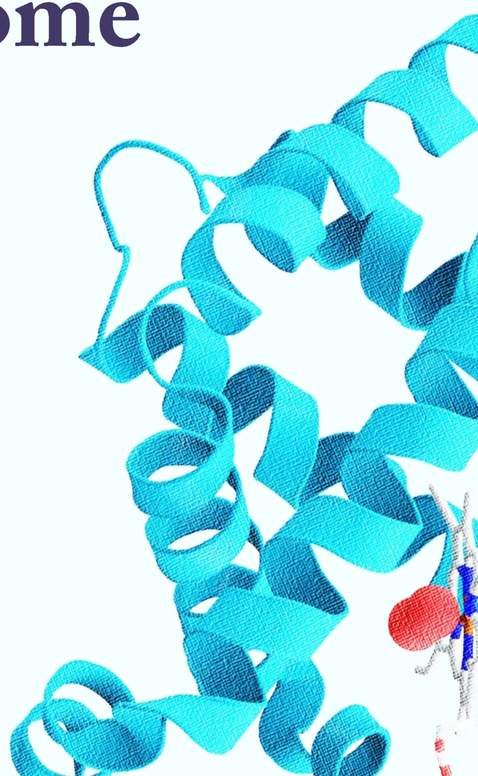
From DNA to Protein
How Cells Read the Genome

Introduction to Bioinformatics Course

Niloufar Razani
Department of Computer Engineering
Sharif University of Technology

Fall 2023

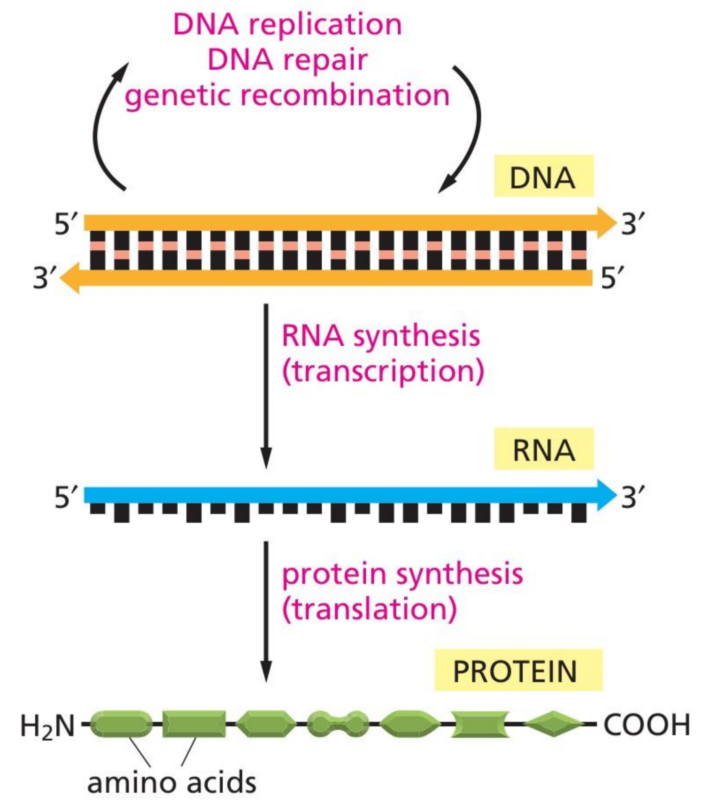


- How does the cell decode and use the information in DNA?

- How genetic instructions carried by DNA specify the amino acid sequences of **proteins**?

- DNA does not direct protein synthesis itself, but acts rather like a manager, delegating the various tasks to a team of workers.

- All cells, from bacteria to humans, express their genetic information in the same way—a principle so fundamental that it has been termed the **Central Dogma** of molecular biology.

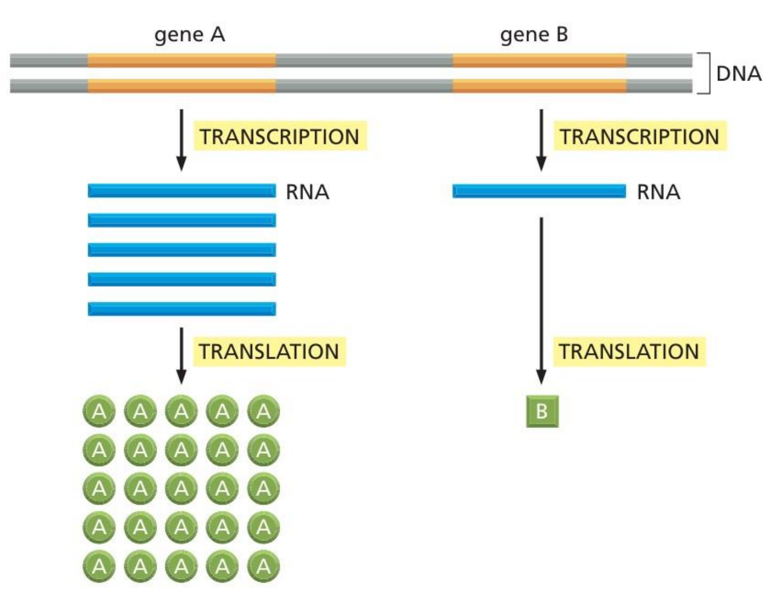


- Transcription and Translation are the means by which cells read out, or express, their genetic instructions—their genes.

- Many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules.

- Because each cell contains only one or two copies of any particular gene, this successive amplification enables cells to rapidly synthesize large amounts of protein whenever necessary.

- At the same time, each gene can be transcribed and translated with a different efficiency, providing the cell with a way to make vast quantities of some proteins and tiny quantities of others.



Differences between DNA and RNA

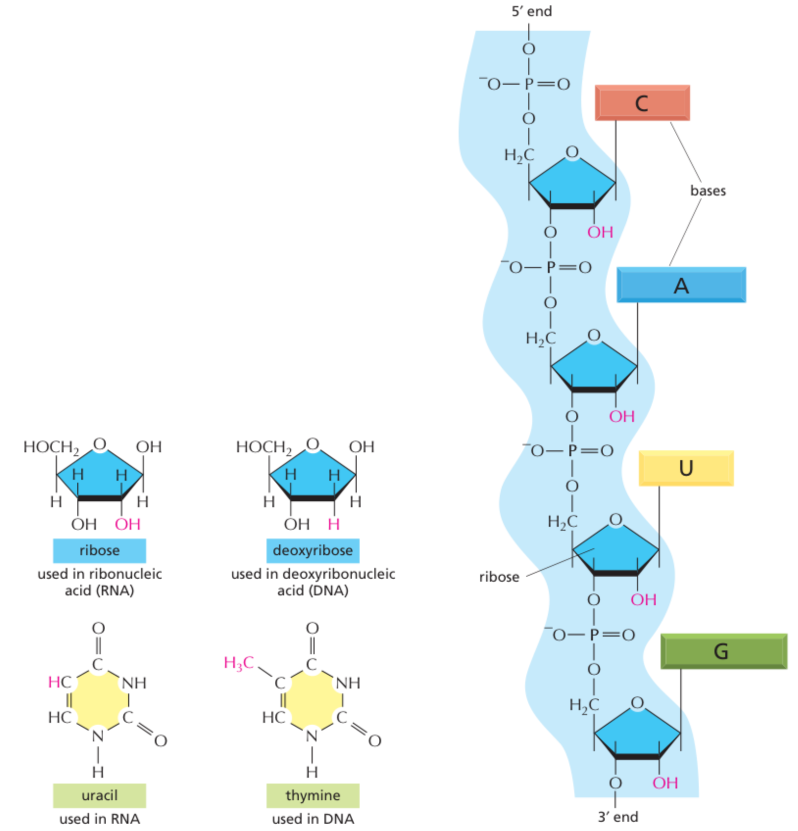
- The first step a cell takes in reading out one of its many thousands of genes is to copy the nucleotide sequence of that gene into RNA.

- There are several important differences between DNA and RNA:

  1) The sugar of RNA is always a **Ribose** instead of a deoxyribose.

  2) The RNA molecule does not contain thymine bases; instead they are replaced by **Uracil**.

  3) The DNA molecule normally appears as a Double strand, while The RNA Molecule appears as a **Single Strand**.



RNA Structure

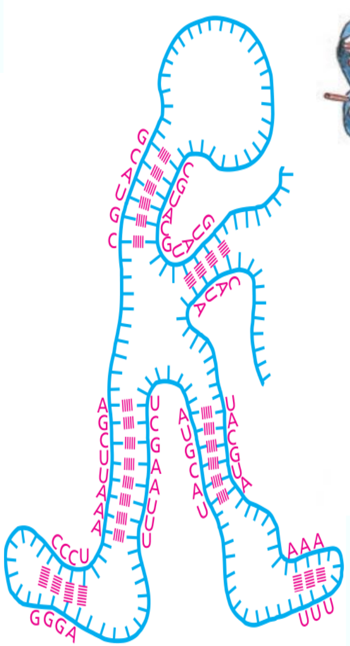
- Because an RNA chain is single-stranded, it can fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein; double-stranded DNA cannot fold in this fashion.

- the ability to fold into a complex 3D shape allows RNA to carry out functions in cells in addition to conveying information between DNA and protein.

- Whereas DNA functions solely as an information store, RNA comes in different varieties, some having structural and even catalytic functions.

Primary

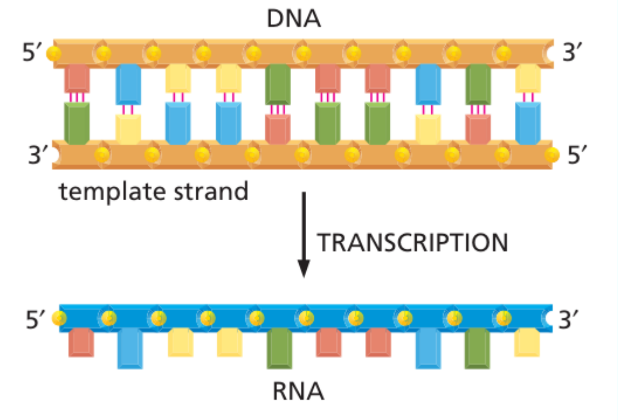
Secondary



Tertiary

From DNA to RNA

- Transcription begins with the opening and unwinding of a small portion of the DNA double helix to expose the bases on each DNA strand.
- One of the two strands of the DNA double helix then acts as a template for the synthesis of RNA.
- Ribonucleotides are added, one by one, to the growing RNA chain, and the nucleotide sequence of the RNA chain is determined by complementary base-pairing with the DNA template.



Transcription, however, differs from DNA replication in several crucial features:

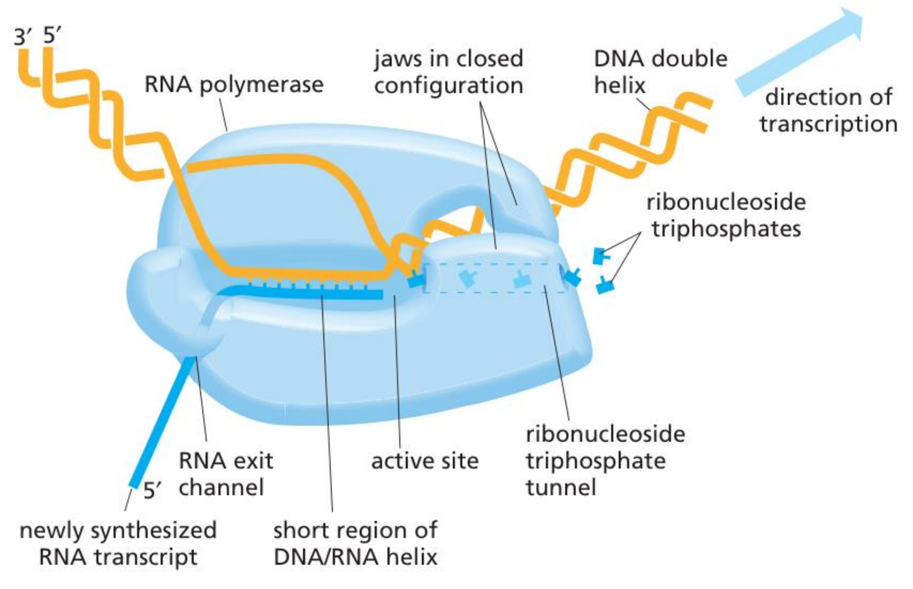
- Unlike a newly formed DNA strand, the RNA strand **does not remain hydrogen-bonded to the DNA template strand**. Instead, just behind the region where the ribonucleotides are being added, the RNA chain is displaced and the DNA helix reforms.
- As RNAs are copied from only a limited region of DNA, these molecules are **much shorter than DNA** molecules.

From DNA to RNA

- The enzymes that carry out transcription are called **RNA polymerases.**

- RNA polymerases **catalyze the formation of the phosphodiester bonds that link the nucleotides together** and form the sugar–phosphate backbone of the RNA chain.

- The RNA polymerase moves stepwise along the DNA, unwinding the DNA helix just ahead to expose a new region of the template strand.



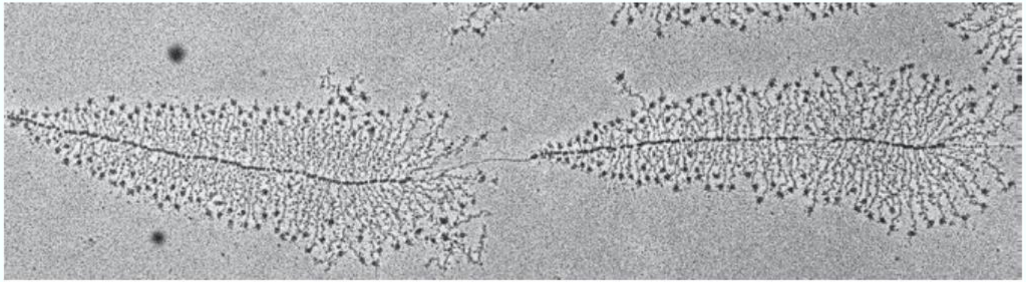
- The growing RNA chain is extended by one nucleotide at a time in the 5’-to-3’ direction.

- The enzyme uses **ribonucleoside triphosphates** (ATP, CTP, UTP, and GTP), whose high-energy bonds provide the energy that drives the reaction forward.

From DNA to RNA

- The almost immediate release of the RNA strand from the DNA as it is synthesized means that **many RNA copies can be made from the same gene in a relatively short time.**

- the synthesis of the next RNA is usually started before the first RNA has been completed.



- Although RNA polymerase catalyzes essentially the same chemical reaction as DNA polymerase, there are some important differences between the two enzymes:

First, and most obviously, RNA polymerase catalyzesthe linkage of ribonucleotides, not deoxyribonucleotides.

Second, unlike the DNA polymerase involved in DNA replication, RNA polymerases can start an RNA chain **without a primer**.

Several types of RNA are produced in Cells

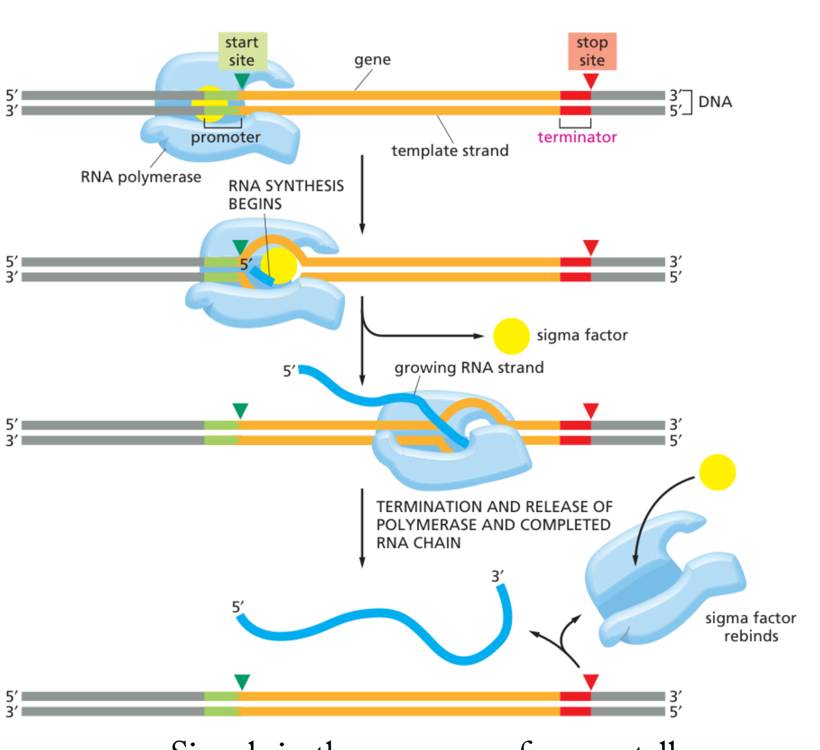
- The vast majority of genes carried in a cell's DNA specify the amino acid sequence of proteins, and the RNA molecules that are copied from these genes are collectively called **messenger RNA (mRNA).**

- The final product of other genes, however, is the RNA itself.

TYPES OF RNA PRODUCED IN CELLS	
TYPE OF RNA	FUNCTION
mRNAs	code for proteins
rRNAs	form the core of the ribosome and catalyze protein synthesis
miRNAs	regulate gene expression
tRNAs	serve as adaptors between mRNA and amino acids during protein synthesis
Other small RNAs	used in RNA splicing, telomere maintenance, and many other processes

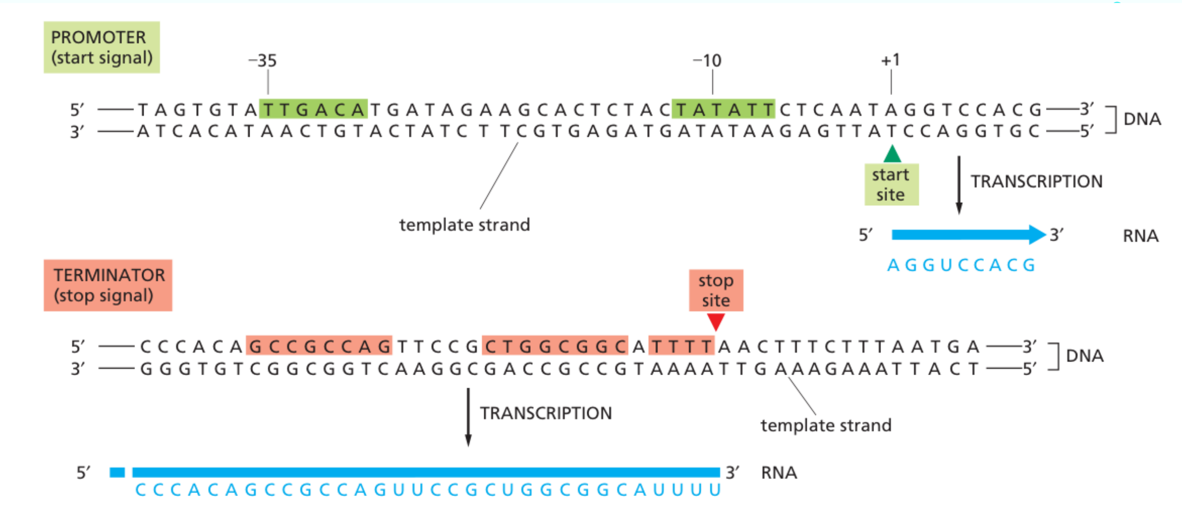
How do signals in DNA tell RNA polymerase where to start and finish?

- To begin transcription, **RNA polymerase** must be able to recognize the start of a gene and bind firmly to the DNA at this site.
- The way in which RNA polymerases recognize the transcription start site **differs somewhat between bacteria and eukaryotes.**
- **Promoter** contains a specific sequence of nucleotides indicating the starting point for RNA synthesis.
- A subunit of the bacterial polymerase, called **sigma factor,** is primarily responsible for recognizing the promoter sequence on DNA.

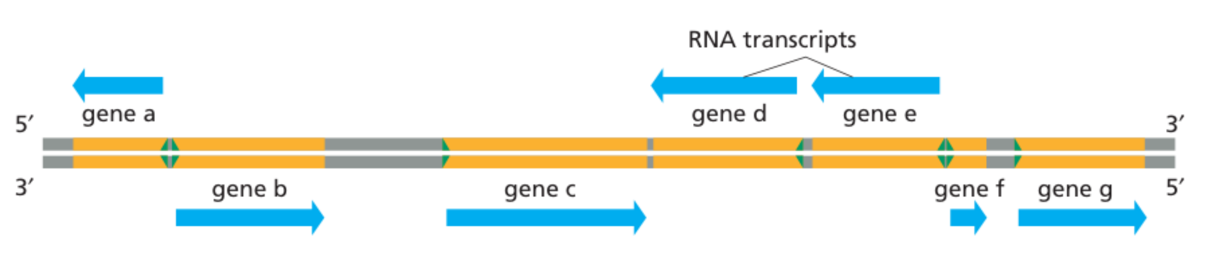


Signals in the sequence of a gene tell bacterial RNA polymerase where to start and stop transcription.

Bacterial promoters and terminators have specific nucleotide sequences that are recognized by RNA polymerase.



Some genes are transcribed using one DNA strand as a template, whereas others are transcribed using the other DNA strand.



Eukaryotic Transcription

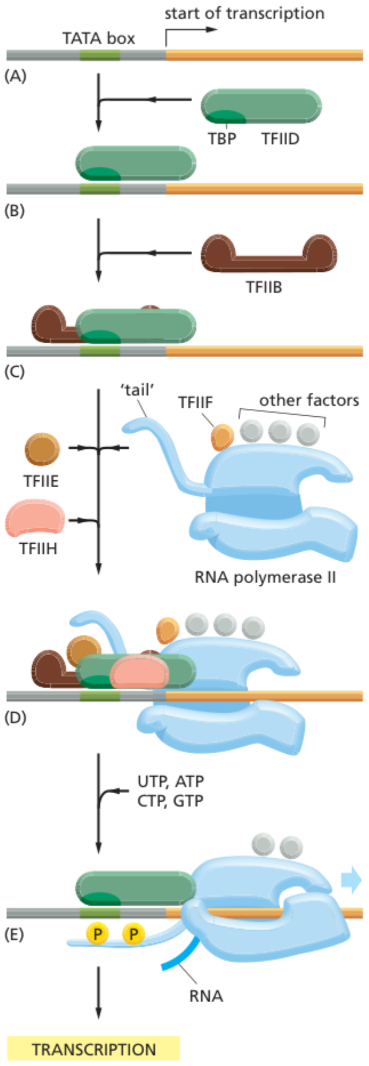
- Transcription initiation in eukaryotes differs in from that in bacteria:

  Eukaryotic cells have three- **RNA polymerases I and III** transcribe the genes encoding transfer RNA, ribosomal RNA, and small RNAs. **RNA polymerase II** transcribes the vast majority of eukaryotic genes, including all those that encode proteins.

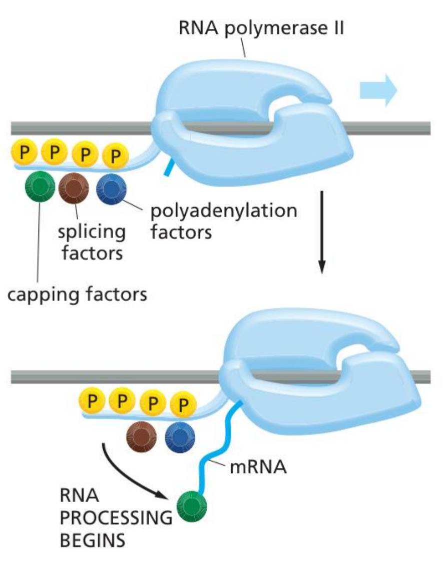
  Eukaryotic RNA polymerases require the assistance of a large set of accessory proteins. Principal among these are the **General Transcription Factors.**

- These Transcription Factors **assemble on the promoter** , where they position the RNA polymerase.

- The opposite figure shows how the general transcription factors assemble at a promoter used by RNA polymerase II.



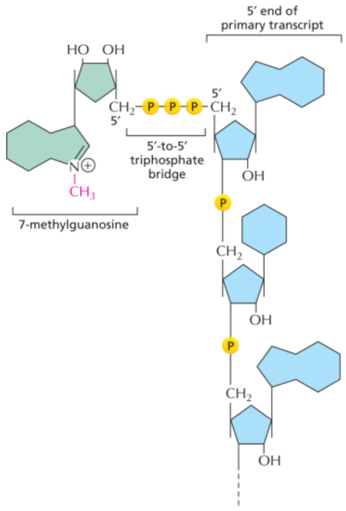
Eukaryotic RNA Processing

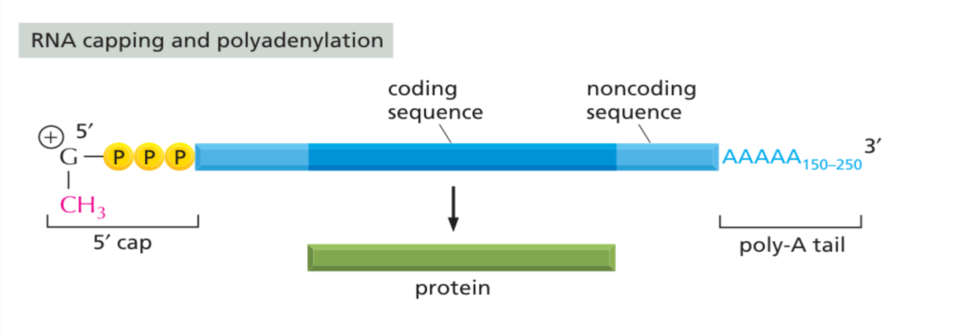


- Before a eukaryotic RNA exits the nucleus, it must go through several different **RNA processing steps.**

- The enzymes responsible for RNA processing ride on the 'tail' of the eukaryotic RNA polymerase as it transcribes an RNA, processing the transcript as it emerges from the RNA polymerase.

- Two processing steps that occur only on transcripts destined to become mRNA molecules are **Capping** and **Polyadenylation**.



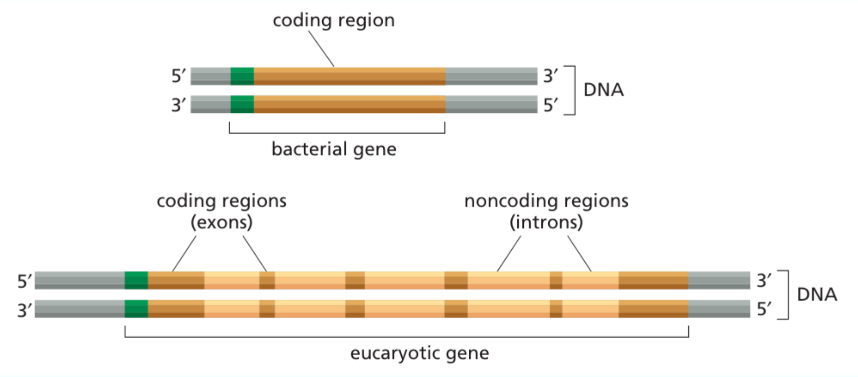


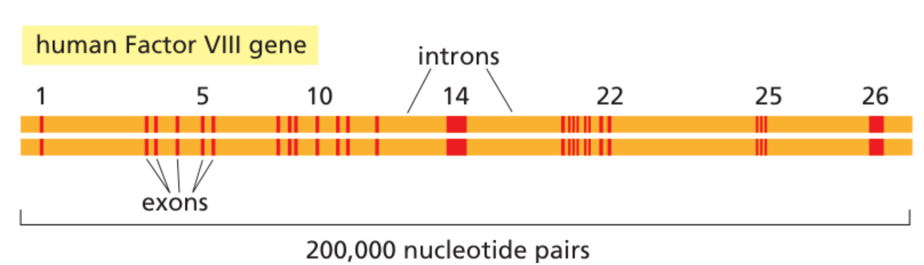
Eukaryotic genes Are interrupted by Noncoding sequences

- Most eukaryotic genes, in contrast to bacteria cells, have their coding sequences interrupted by long, noncoding intervening sequences, called **Introns.**

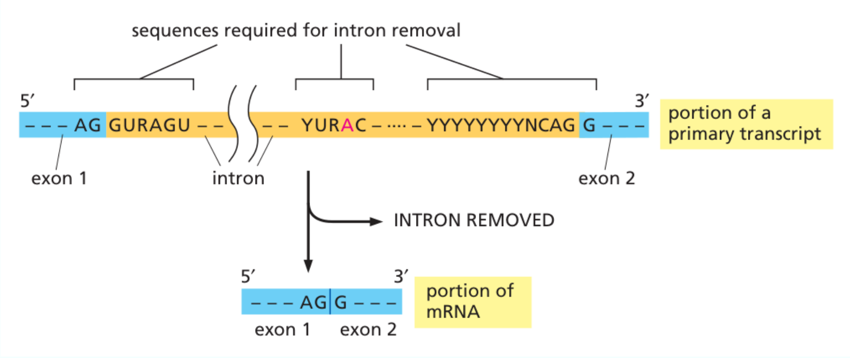
- The scattered pieces of coding sequences, or expressed sequences, called **Exons**, are usually shorter than the introns, and the coding portion of a eukaryotic gene is often only a small fraction of the total length of the gene.

- Introns range in length from a single nucleotide to more than 10,000 nucleotides





Introns are removed by RNA splicing

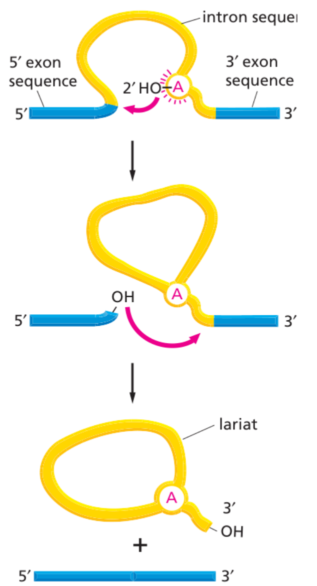


- In **RNA splicing**, the intron sequences are removed from the newly synthesized RNA and the exons are stitched together.

- Once a transcript has been spliced and its 5’ and 3’ ends have been modified, the RNA is a functional mRNA molecule that can now leave the nucleus and be translated into protein.

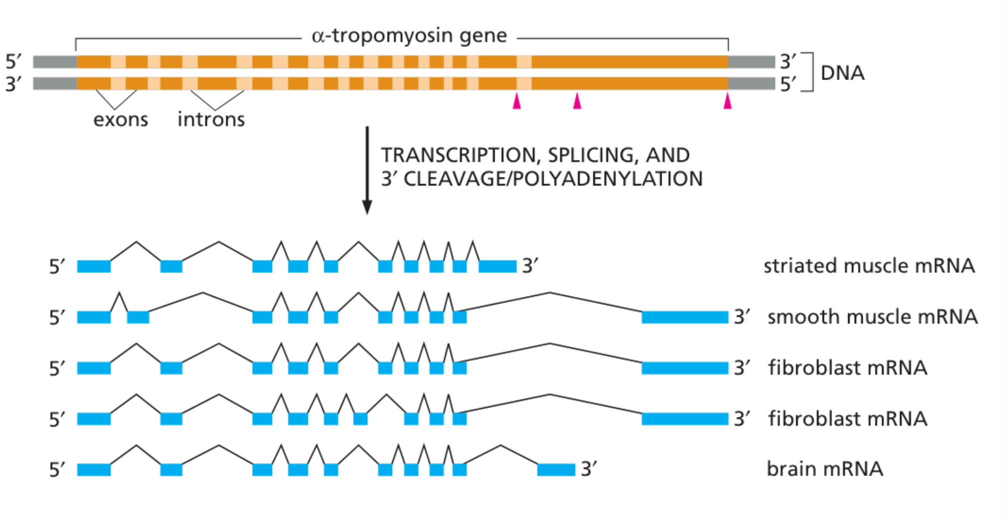
- **Each intron contains a few short nucleotide sequences that act as cues for its removal.**

- Guided by these sequences, an elaborate splicing machine cuts out the intron sequence in the form of a ‘lariat’ structure.



Alternative Splicing

- the transcripts of many eukaryotic genes can be spliced in different ways, each of which can produce a distinct protein.
- Such alternative splicing thereby allows many different proteins to be produced from the same gene.
- An estimated 60% of human genes probably undergo alternative splicing.
- Thus RNA splicing enables eukaryotes to increase the already enormous coding potential of their genomes.

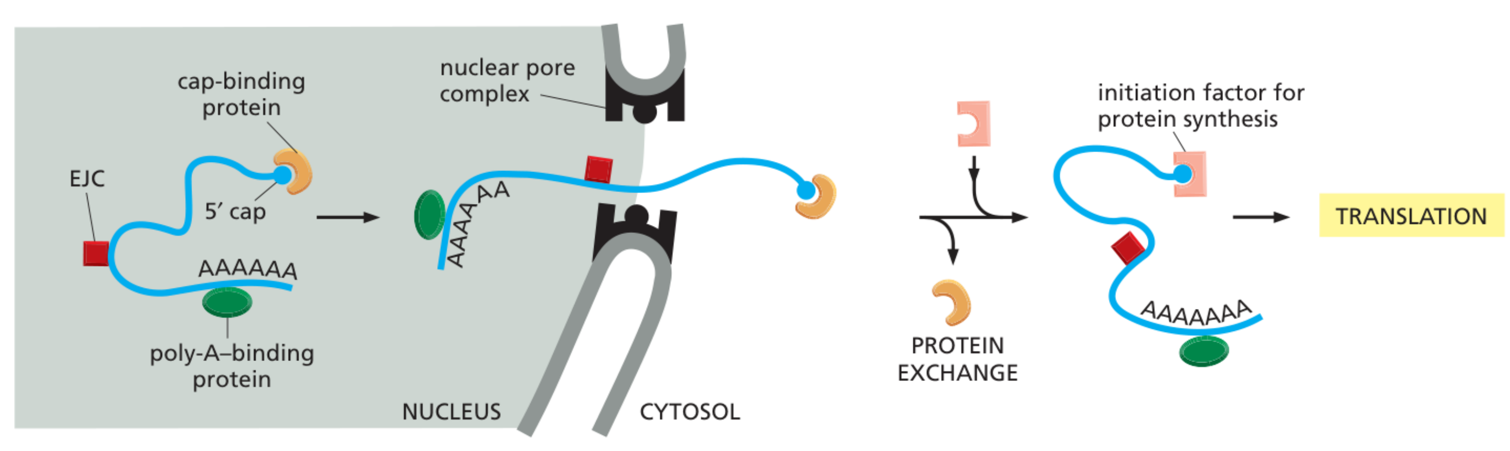


Mature eukaryotic mRNAs Are selectively exported from the Nucleus

- How does the cell distinguish between the relatively rare mature mRNA molecules it needs to keep and the overwhelming amount of debris generated by RNA processing?

- the transport of mRNA from the nucleus to the cytoplasm, where it is translated into protein, is highly **Selective.**

- Only correctly processed RNAs are allowed to pass.



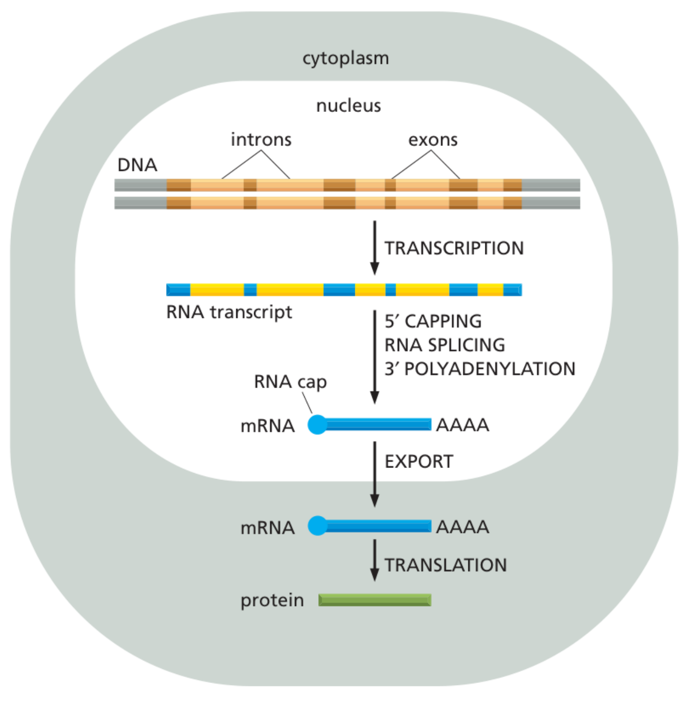
From RNA to Protein

- By the end of the 1950s, biologists had demonstrated that the information encoded in DNA is copied first into RNA and then into protein.

- Transcription as a means of information transfer is simple to understand, since DNA and RNA are chemically and structurally similar.

- In contrast, the conversion of the information in RNA into protein represents a translation of the information into another language that uses quite different symbols.

- **But how is the information in a linear sequence of nucleotides in RNA translated into the linear sequence of a chemically quite different set of subunits—the amino acids in proteins?**



mRNA sequence is Decoded in sets of Three Nucleotides

- Because there are only 4 different nucleotides in mRNA but 20 different types of amino acids in a protein, this translation cannot be accounted for by a direct one-to-one correspondence between a nucleotide in RNA and an amino acid in protein.

- The sequence of nucleotides in the mRNA molecule is read consecutively in groups of three.

- The sequence of nucleotides in the mRNA molecule is read consecutively in groups of three.

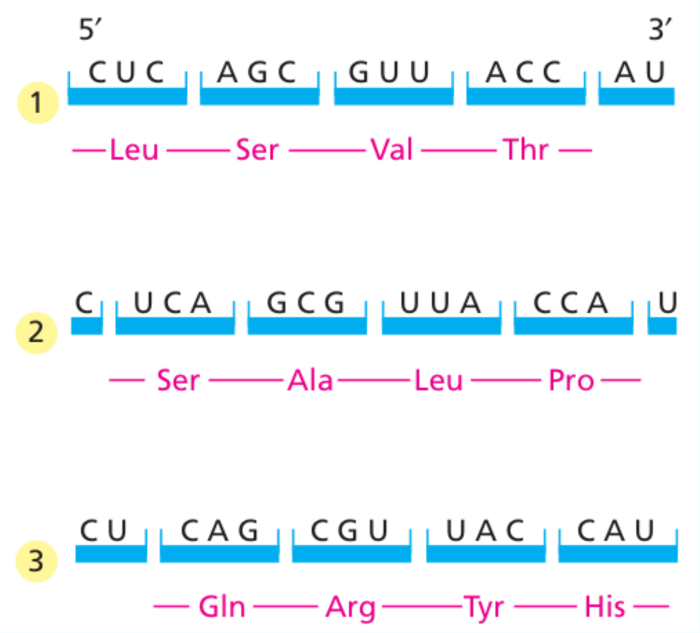
- there are thus 4 * 4 * 4 = 64 possible combinations of three nucleotides: AAA, AUA, AUG, and so on. However, only 20 different amino acids are commonly found in proteins.

- Each group of three consecutive nucleotides in RNA is called a **Codon**, and each specifies one amino acid.

	AGA									UUA					AGC					
	AGG									UUG					AGU					
GCA	CGA						GGA			CUA				CCA	UCA	ACA			GUA	
GCC	CGC						GGC		AUA	CUC				CCC	UCC	ACC			GUC	UAA
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	CUG	AAA		UUC	CCG	UCG	ACG		UAC	GUG	UAG
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	UGG	UAU	GUU	UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

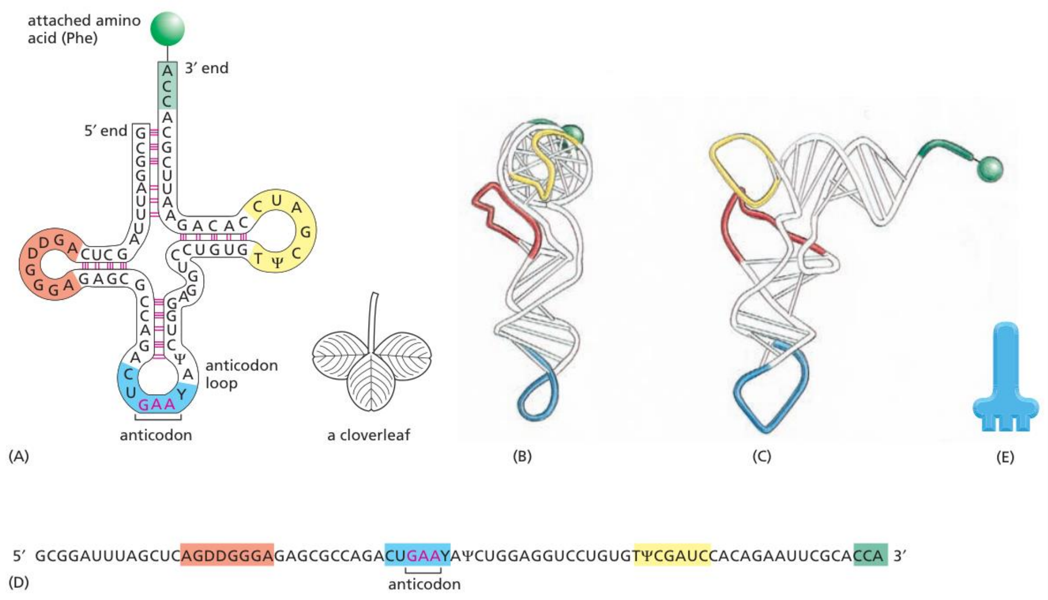
mRNA sequence is Decoded in sets of Three Nucleotides

- In principle, an RNA sequence can be translated in any one of three different **Reading Frames** depending on where the decoding process begins.
- Only one of the three possible reading frames in an mRNA specifies the correct protein.
- The translation of an mRNA begins with the codon **AUG**.
- This codon adds the amino acid methionine to amino acid chain so that newly made proteins all have methionine as the first amino acid at their N-terminal end.



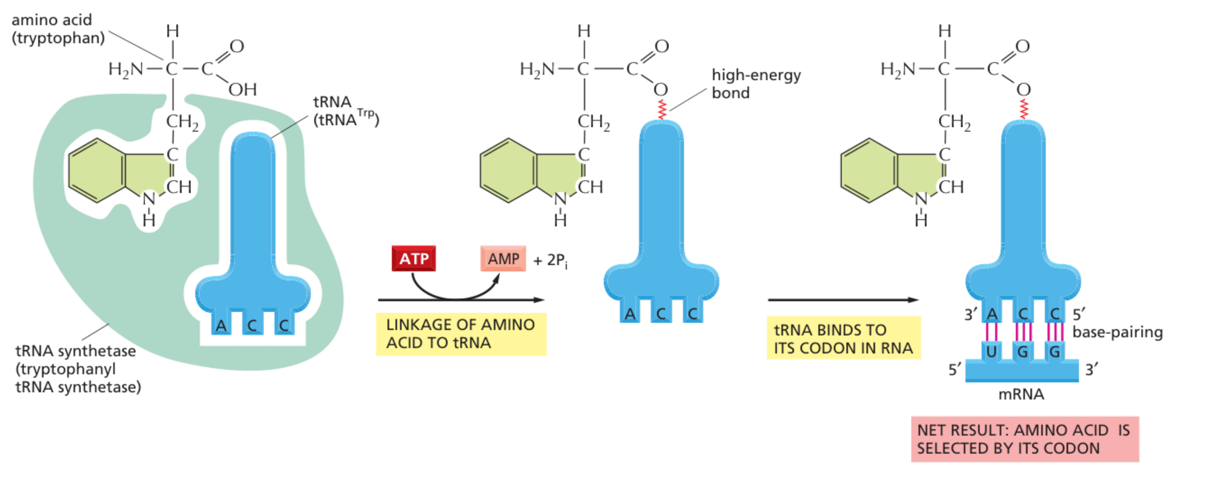
tRNA Molecules Match Amino Acids to Codons in mRNA

- The codons in an mRNA molecule do not directly recognize the amino acids they specify.
- These adaptors consist of a set of small RNA molecules known as transfer RNAs (**tRNAs**), each about 80 nucleotides in length.
- **Anticodon** is a set of three consecutive nucleotides that through base pairing bind the complementary codon in an mRNA molecule.
- A short single-stranded region at the 3’ end of the molecule; this is the site where the amino acid that matches the codon is attached to the tRNA.
- Some amino acids have more than one tRNA and some tRNAs are constructed so that they require accurate base-pairing



Specific Enzymes couple tRNAs to the Correct Amino Acid

- How each tRNA molecule becomes charged—linked to the one amino acid in 20 that is its right partner.
- Recognition and attachment of the correct amino acid depend on enzymes called **aminoacyl-tRNA synthetases.**
- These enzymes covalently couple each amino acid to its appropriate set of tRNA molecules.
- In most organisms, there is a different synthetase enzyme for each amino acid.



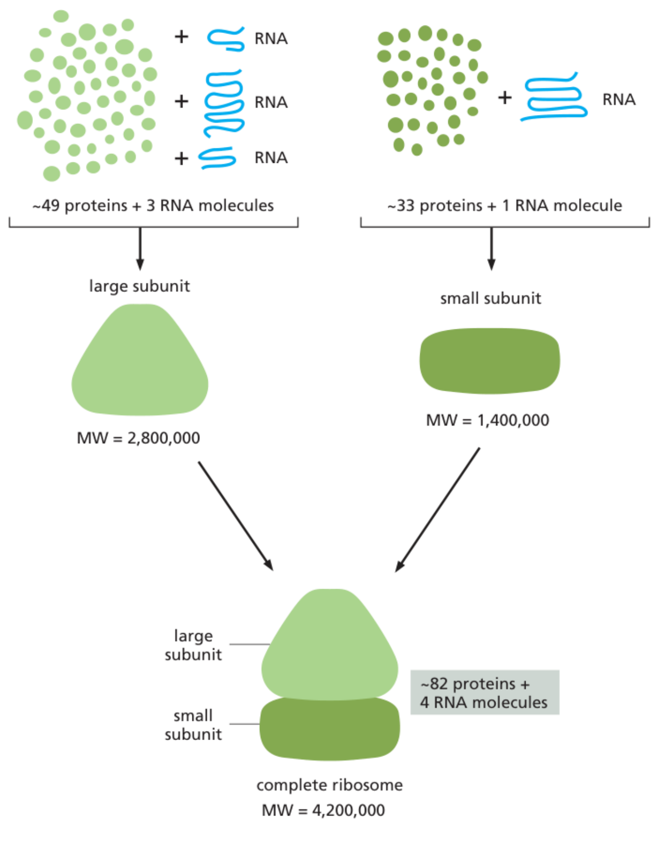
The RNA Message is Decoded on Ribosomes

- Accurate and rapid translation of mRNA into protein requires a large molecular machine that moves along the mRNA, captures complementary tRNA molecules, holds them in position, and covalently links the amino acids that they carry so as to form a protein chain.

- This protein-manufacturing machine is the **Ribosome**—a large complex made from more than 50 different proteins.

- Ribosomes are composed of one **large** and one **small subunit** that fit together to form a complete ribosome.

- The **small subunit** matches the tRNAs to the codons of the mRNA, while the **large subunit** catalyzes the formation of the peptide bonds that covalently link the amino acids together into a polypeptide chain.



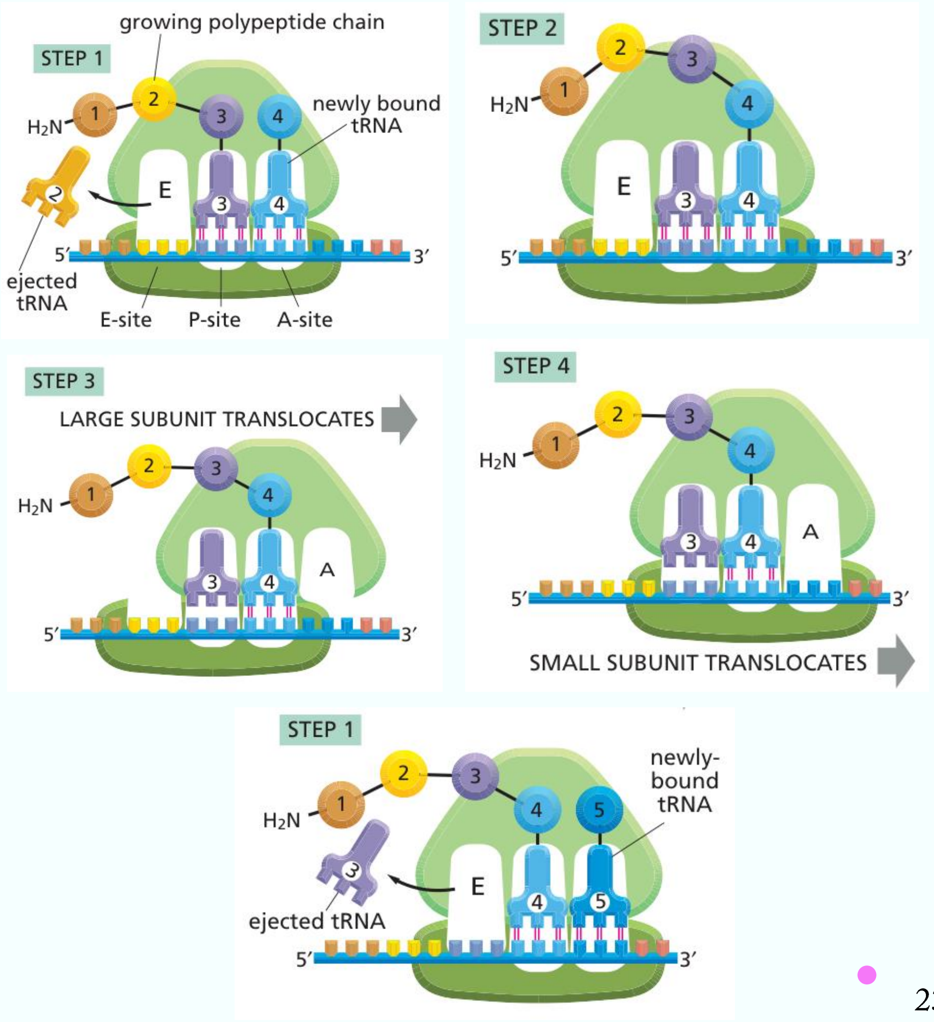
How Ribosomes work?

- Each ribosome contains a binding site for an mRNA molecule and three binding sites for tRNA molecules, called the **A-site**, the **P-site**, and the **E-site**.

- To add an amino acid to a growing peptide chain, the appropriate charged tRNA enters the A-site.

- Its amino acid is then linked to the peptide chain held by the tRNA in the neighboring P-site.

- Next, the ribosome shifts, and the spent tRNA is moved to the E-site before being ejected



RESOURCES

- Alberts, Bruce, et al. *Essential cell biology*, Chapter 7, From DNA to protein: how Cells read the Genome.