CS123A Bioinformatics

Module 1 – Week 1 – Presentation 2

Leonard Wesley

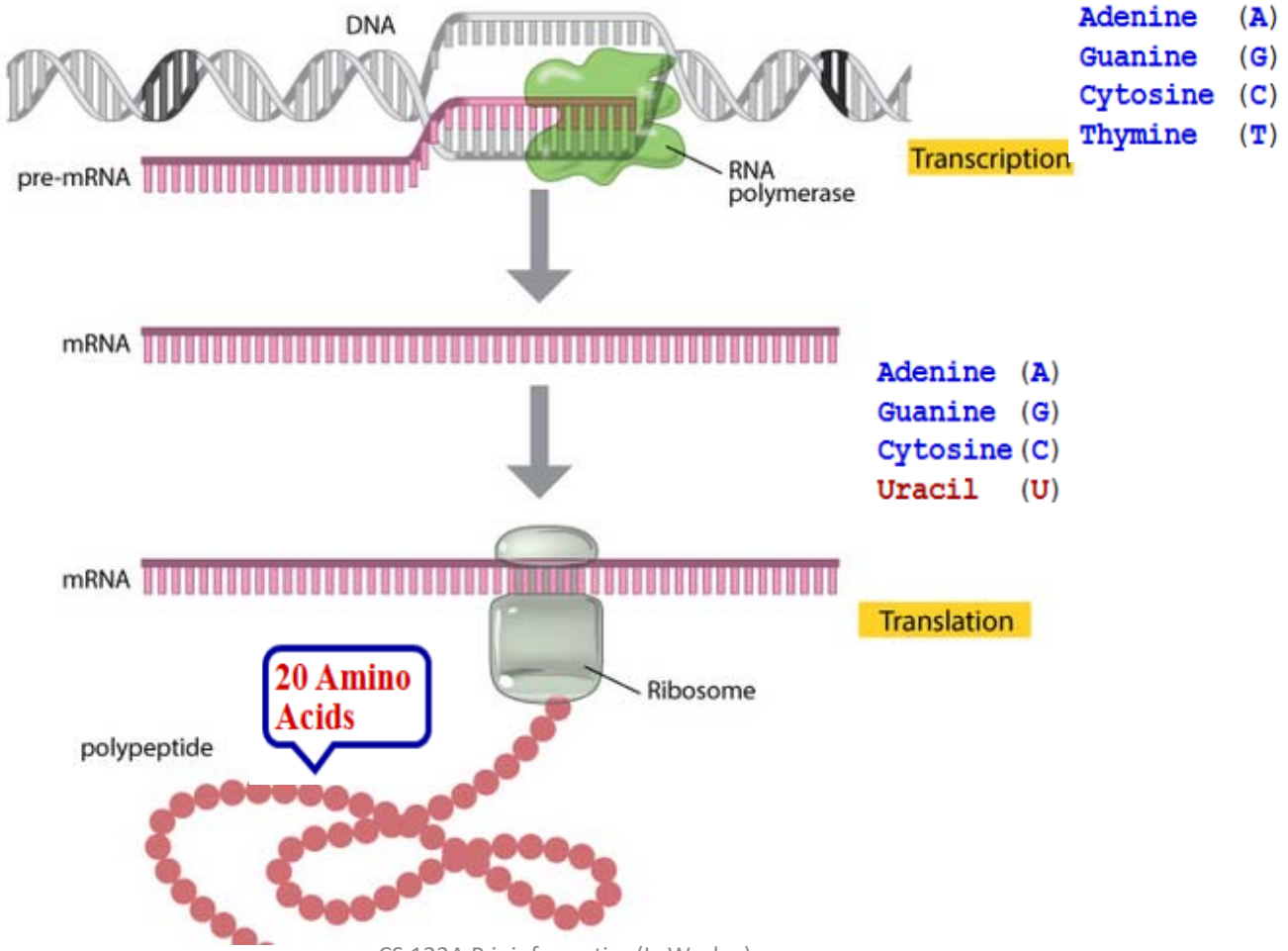
Computer Science Dept

San Jose State Univ

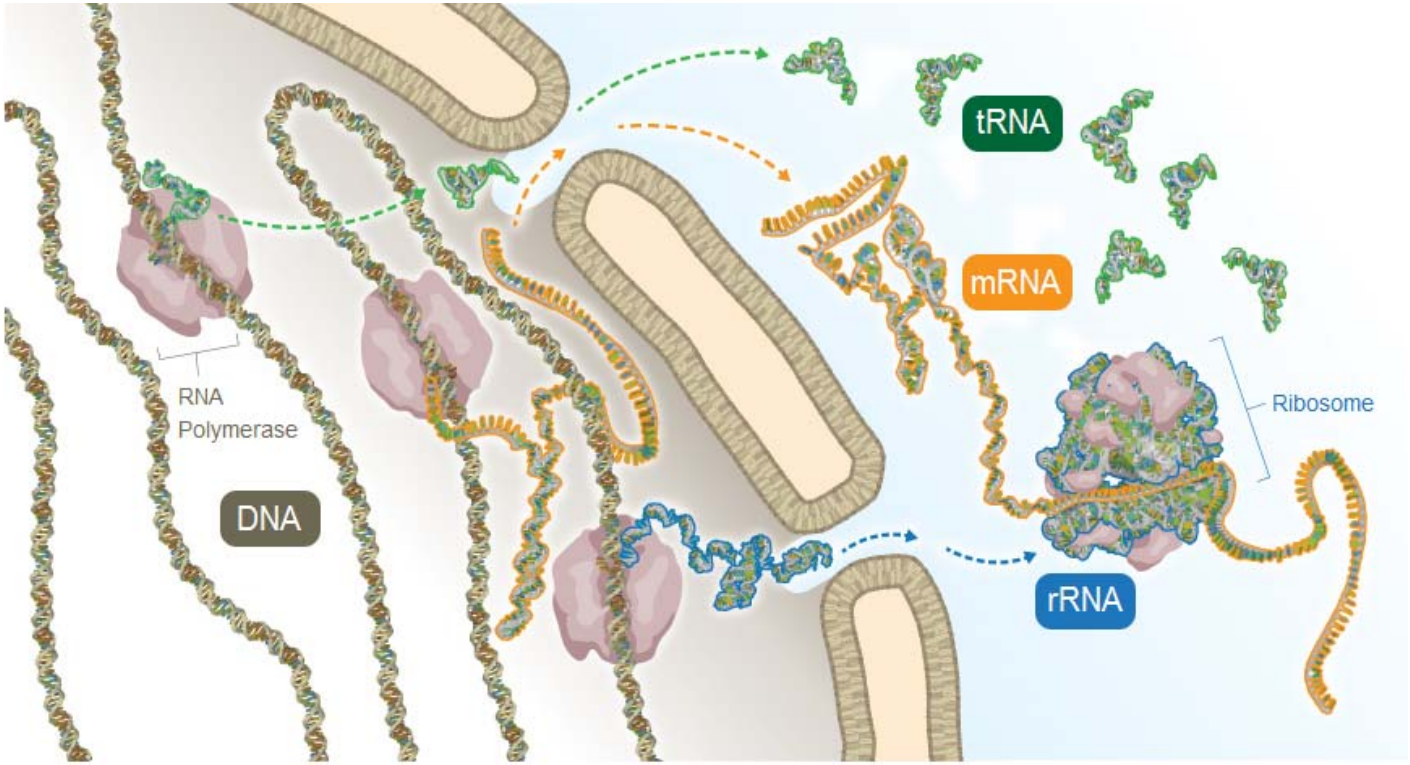
Agenda

- Biology Review
  - Review of Molecular Cell Biology

DNA Is The Blueprint For Life

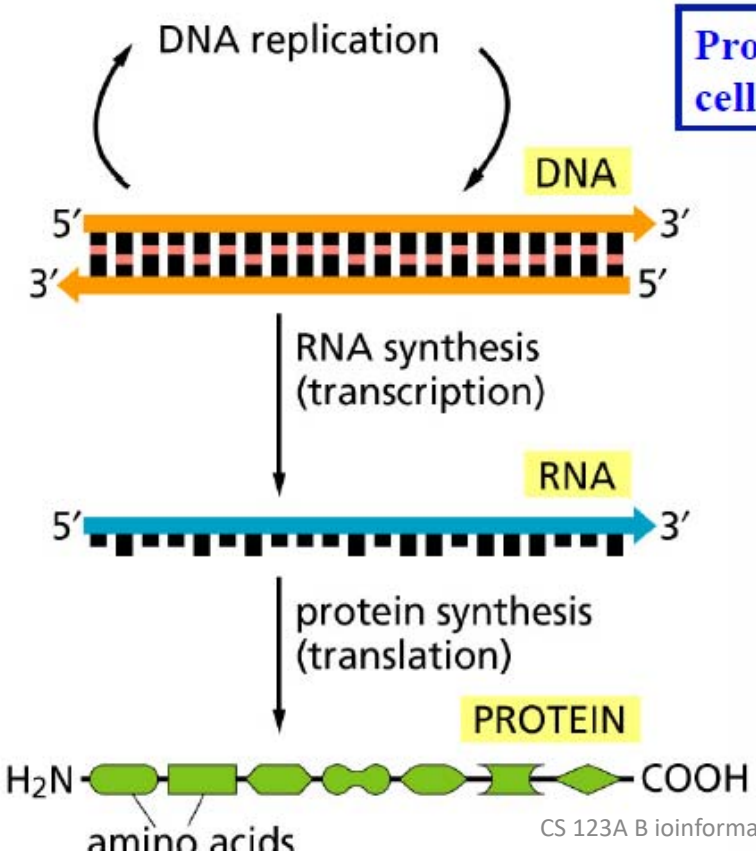


Central Dogma Of Molecular Biology



Proteins are the workhorse of life: structure, regulation, transport, defense, enzymes

Central Dogma of Molecular Biology



Proteins, are the molecules that put the cell's genetic information into action.

- Traits
- Diseases
- Drug Resistance
- Physiology
- Metabolism

Prokaryotes and Eukaryotes

A **cell** is the fundamental working unit of every living organism.

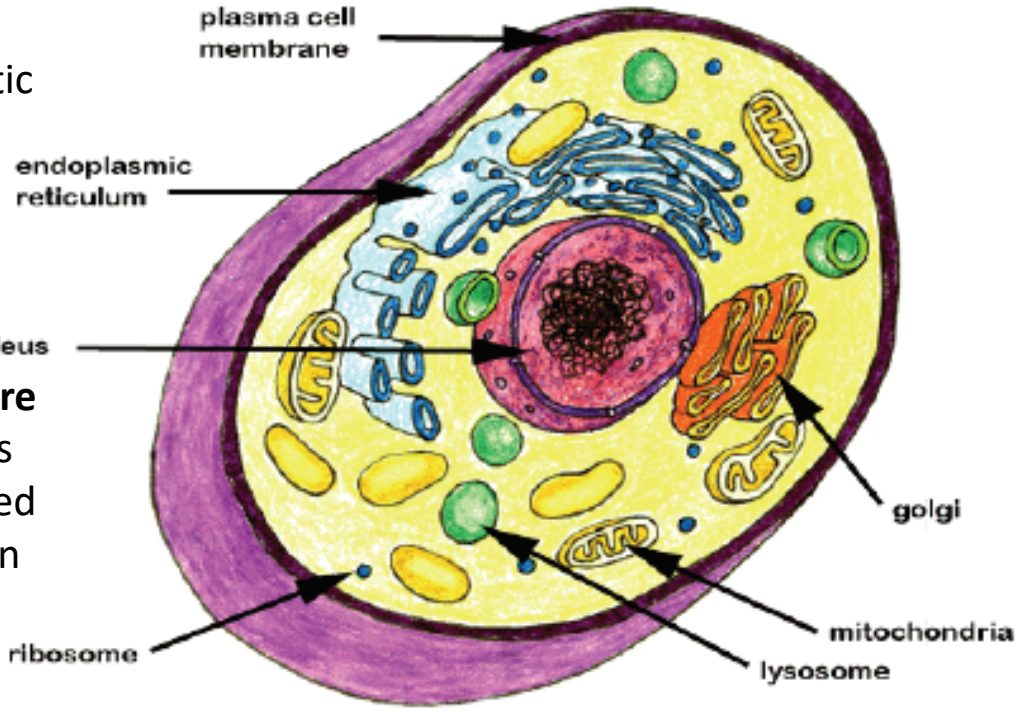
There are two kinds of cells:

- **prokaryotes**, which are single-celled organisms with no cell nucleus: archea and bacteria.
- **eukaryotes**, which are higher level organisms, and their cells have nuclei: animals and plants.

Eukaryote Cell

A **cell carries the** entire set of genetic instructions: the genome, that makes an entire organism.

The **instructions are** encoded in DNA as genes and packaged as chromosomes in the nucleus.



Proteins & Nucleic Acids

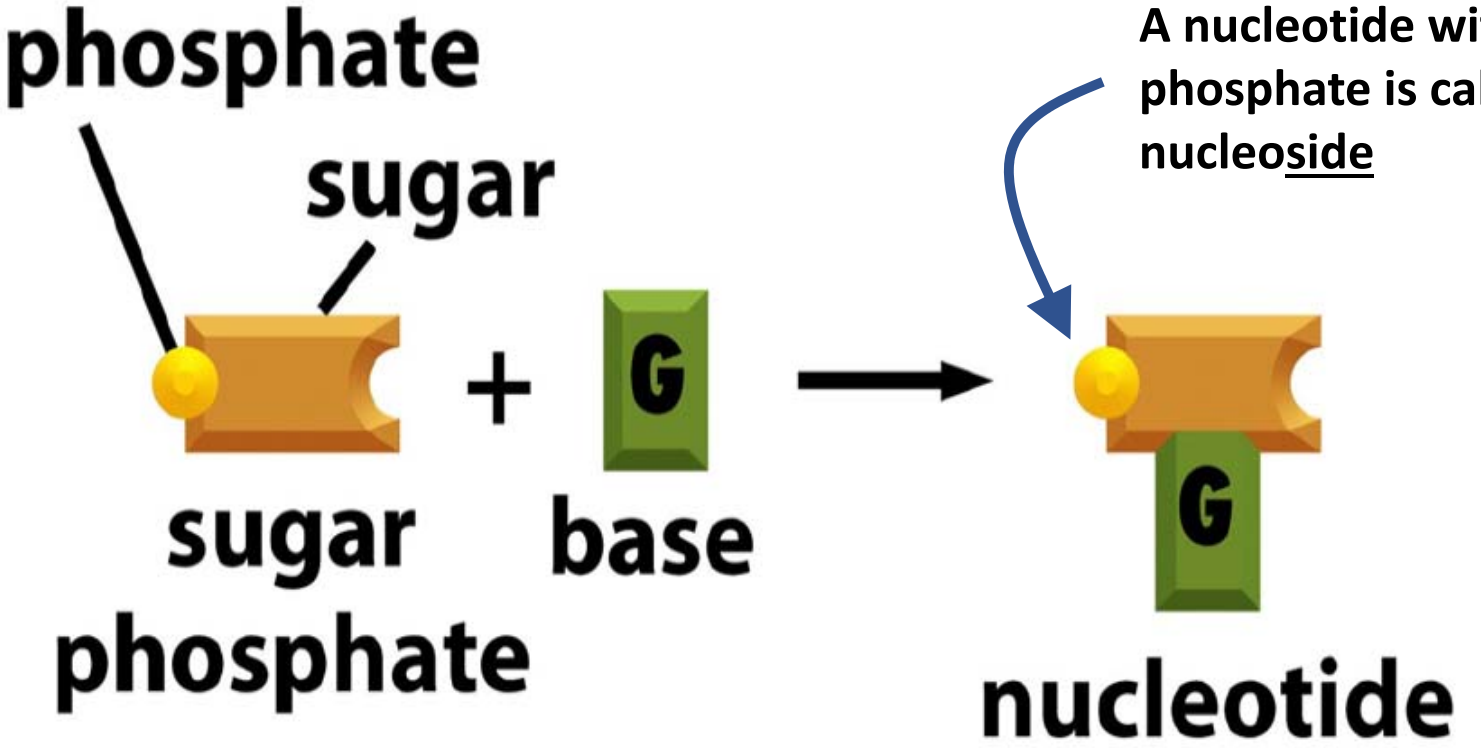
- All living organisms have a similar molecular chemistry.

  The main actors in the chemistry of life are molecules:
  - – **proteins: which are responsible for what a living being is and** does in a physical sense.
  - "We are our proteins" R. Doolittle.

- – **nucleic acids: which encode the information necessary to**
  - produce proteins and are responsible for passing the "recipe" to subsequent generations.
  - Living organisms contain 2 kinds of nucleic acids:

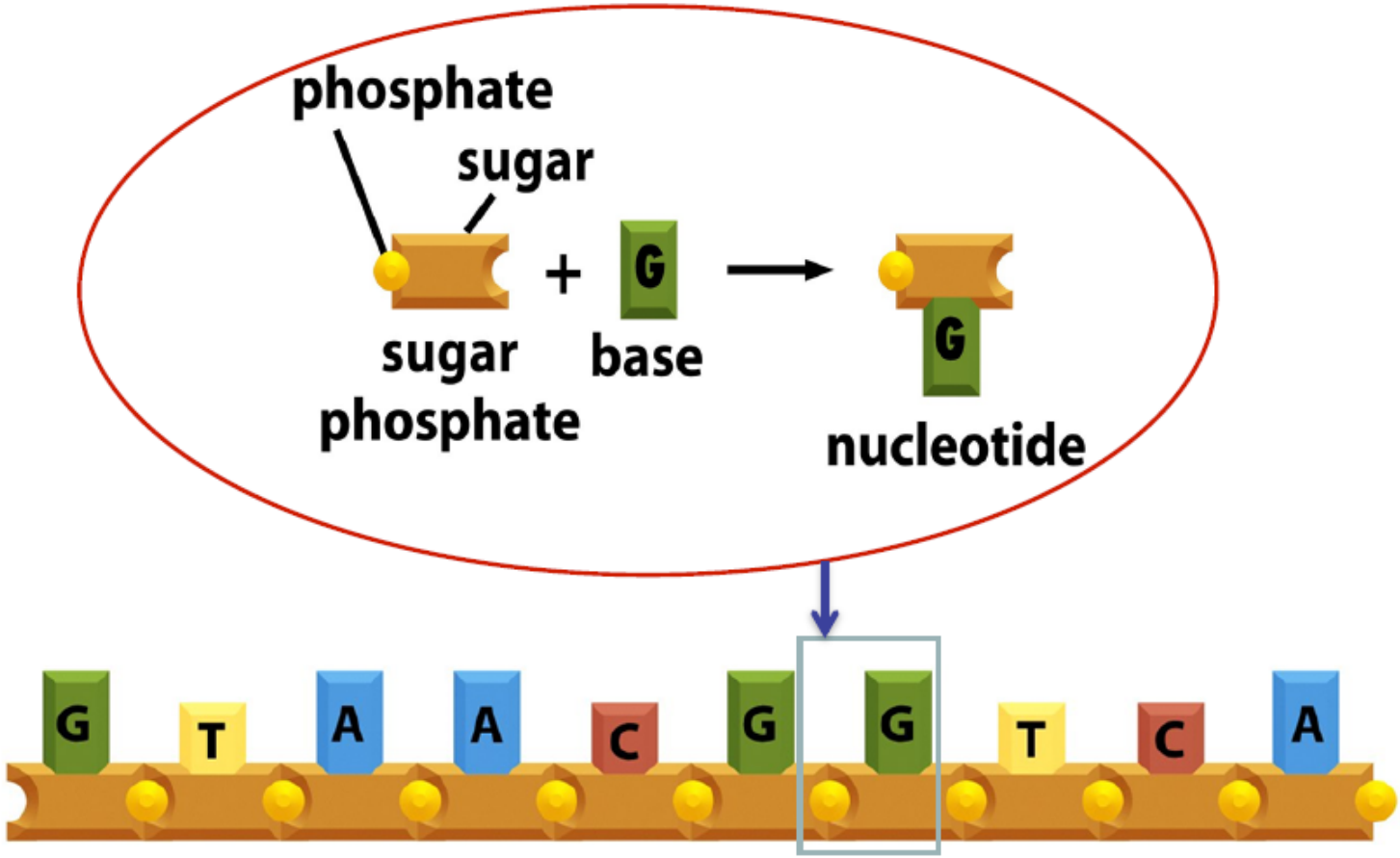
- – **Ribonucleic acid (RNA)**
- – **Deoxyribonucleic acid (DNA)**

Building Block of DNA

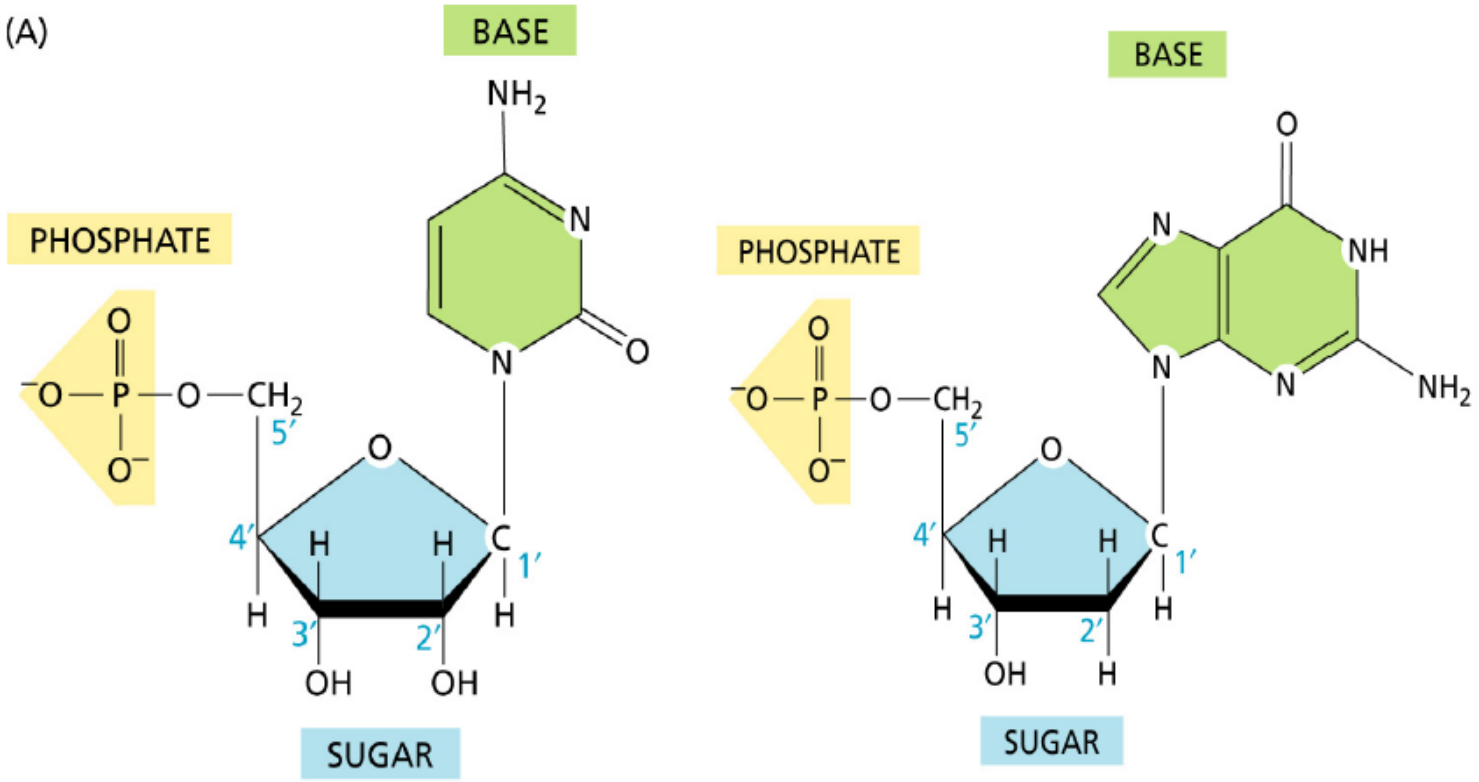


A nucleotide without the phosphate is called a nucleo<u>side</u>

DNA Strand


phosphate
sugar
+
G
sugar
phosphate
base
G
nucleotide
G
T
A
A
C
G
G
T
C
A

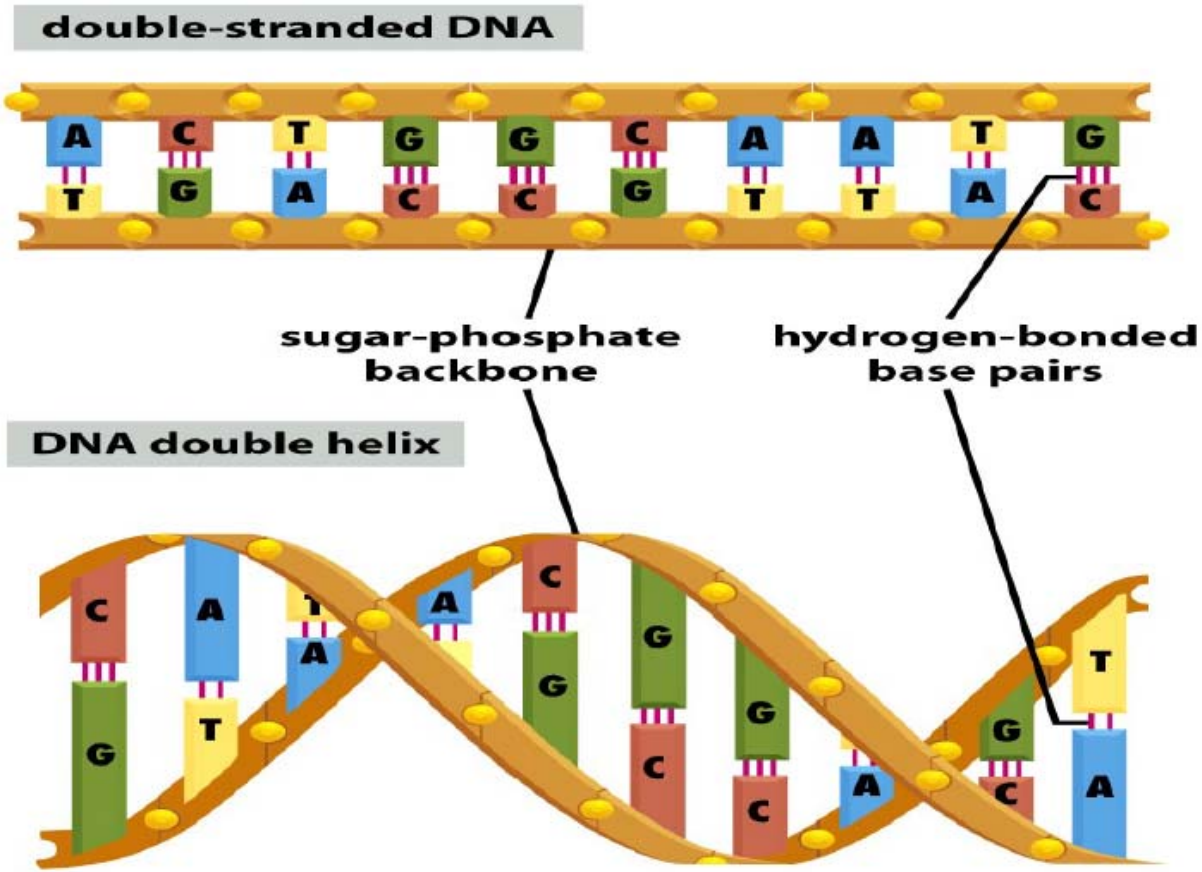
Sugars Found in Nucleic Acids



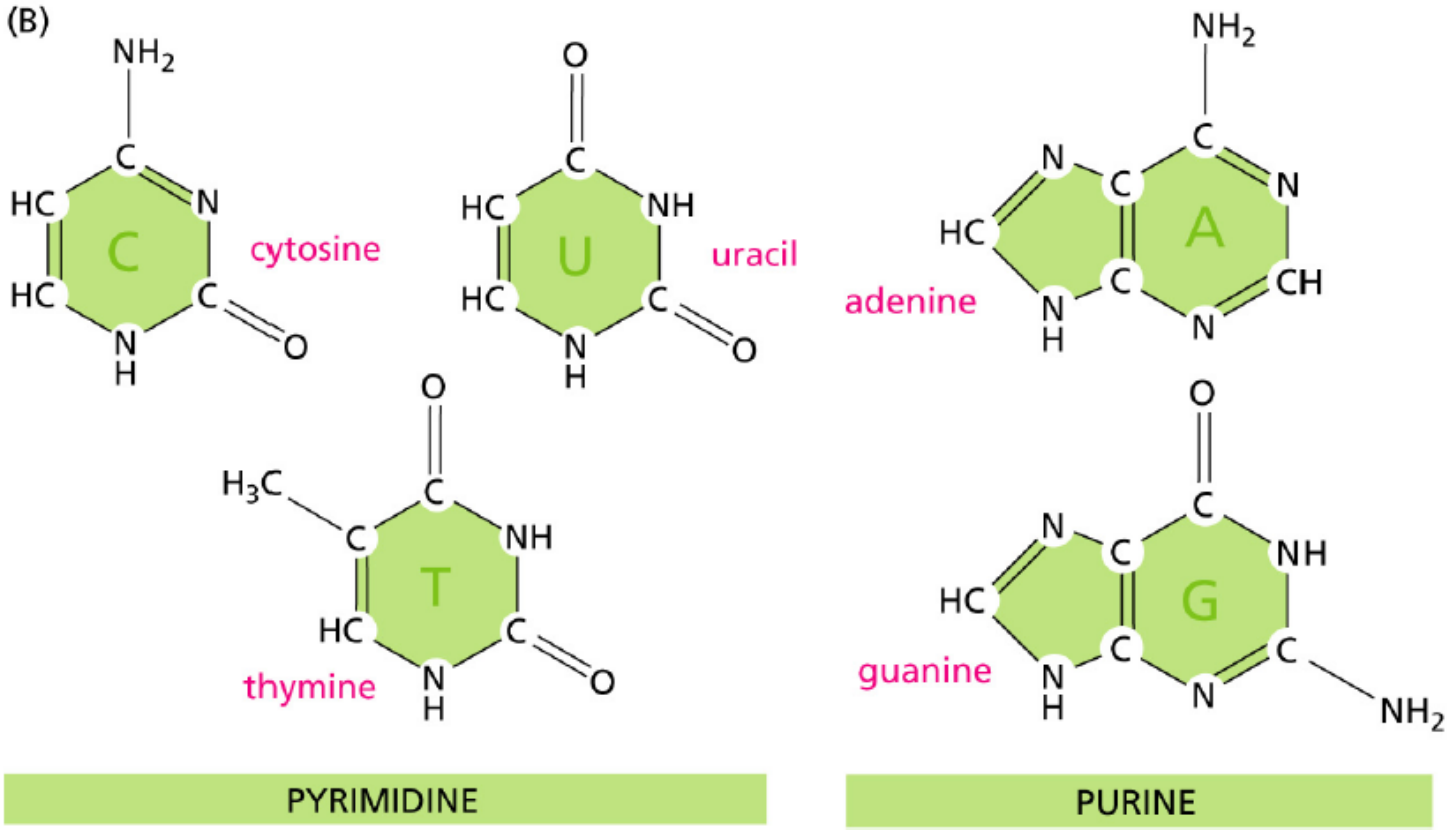
Pentose sugar present in RNA

Pentose sugar present in DNA

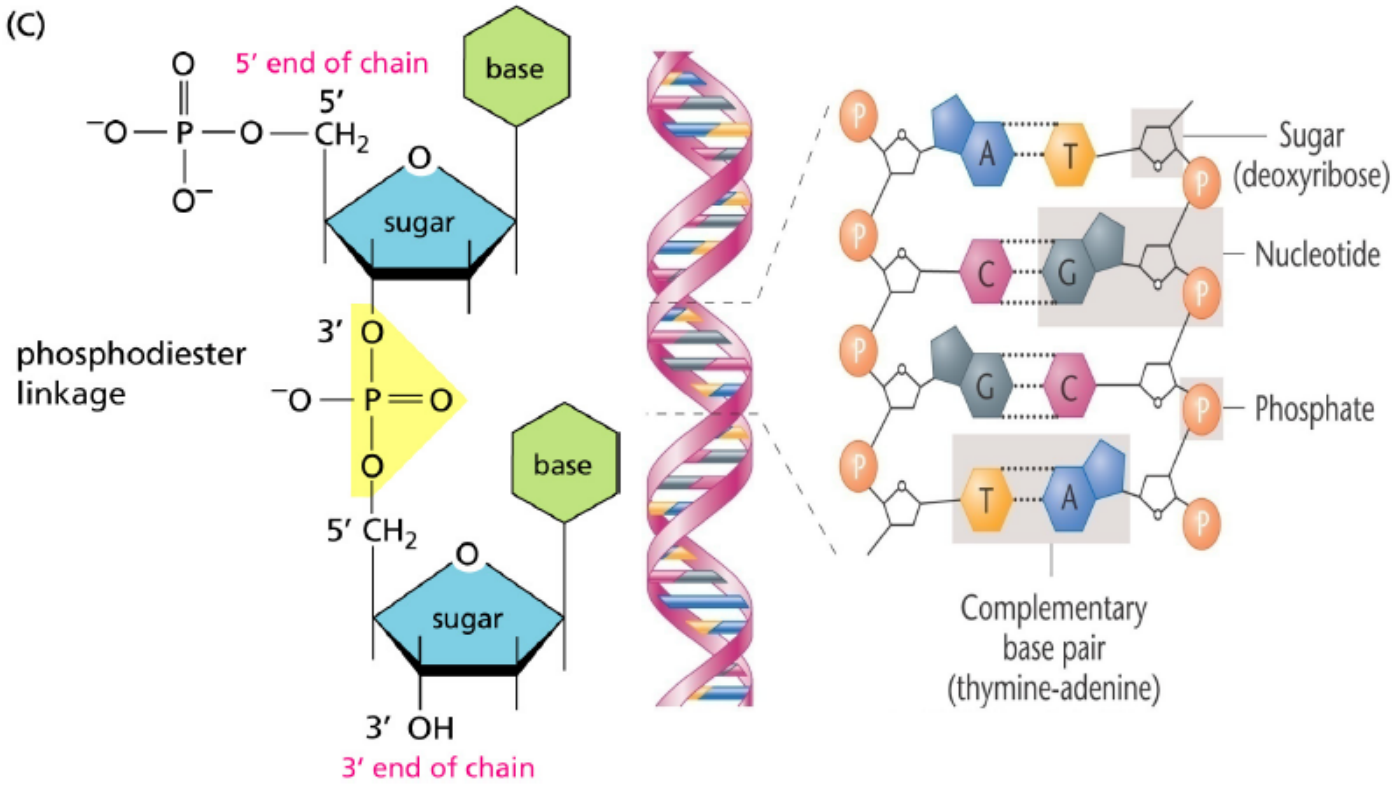
Double-Stranded DNA



Purines and Pyrimidines

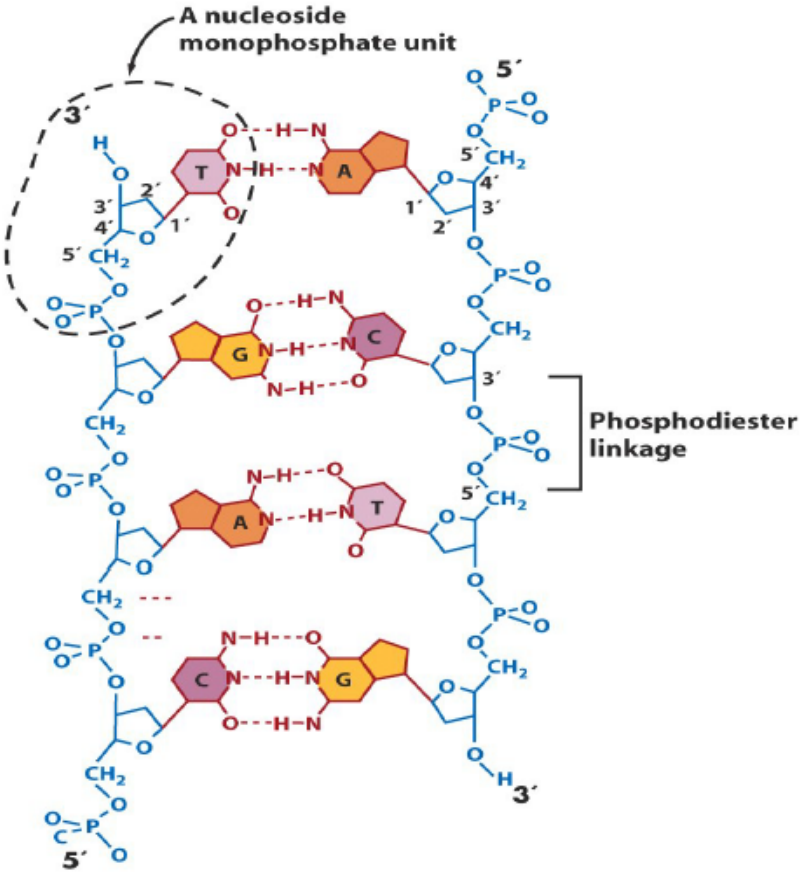


DNA Phosphodiester Backbone



The Two Backbones of DNA

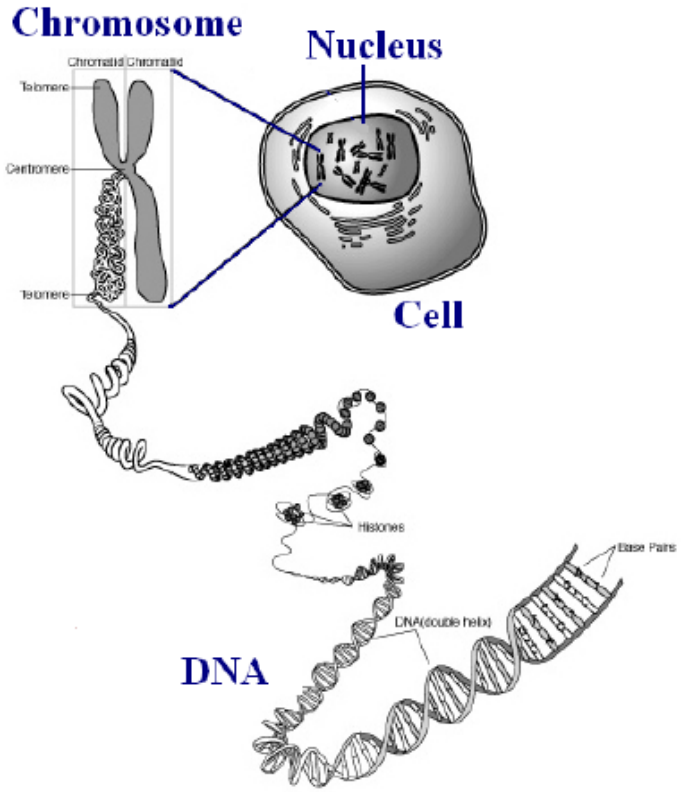
The DNA backbones have alternating sugar-phosphate components. The backbones run in opposite directions.

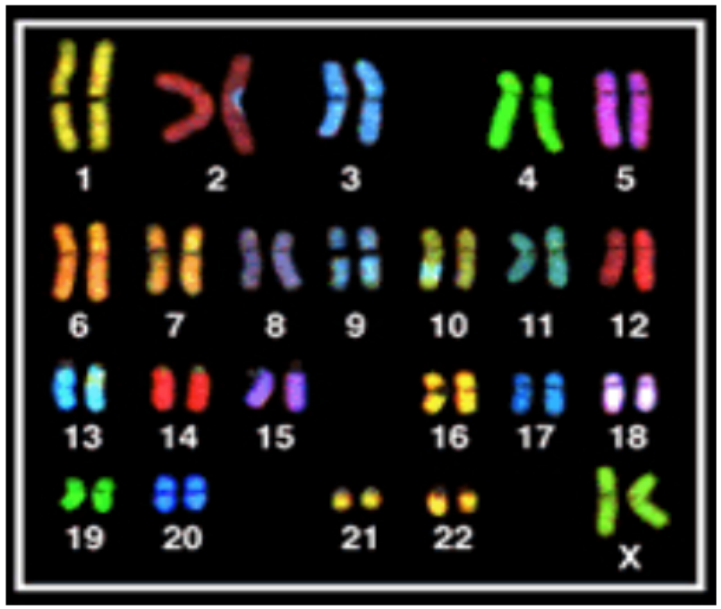


DNA and Chromosomes

- The **genome** is a complete set of instructions for making an organism, consists of tightly coiled threads of **DNA** organized into structures called **chromosomes**.
- Besides the reproductive cell and red blood cell, every single cell in the human body contains the human genome.

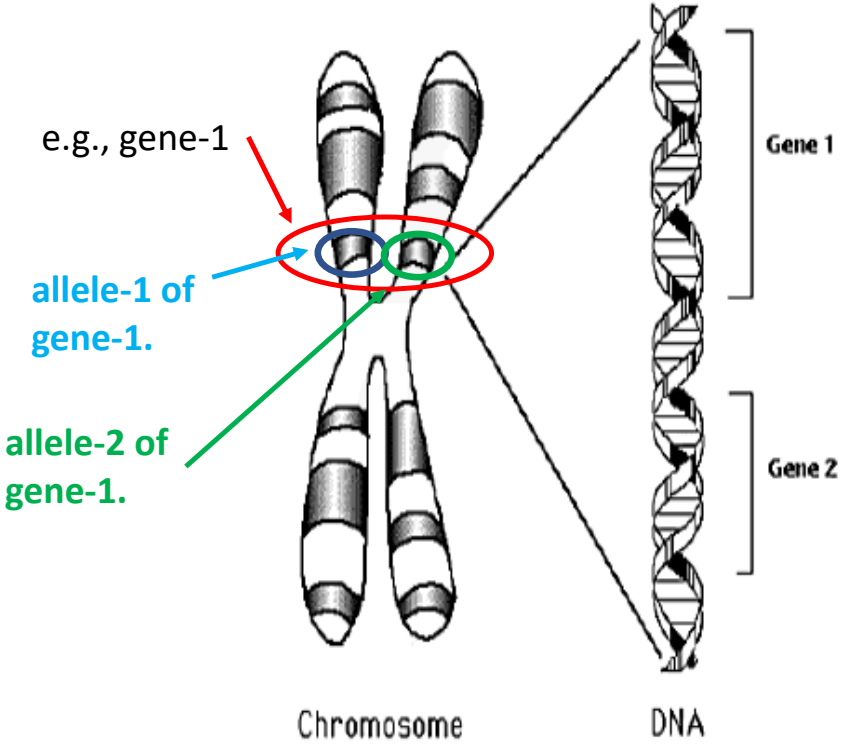
Chromosomes and Genome





Human Chromosomes

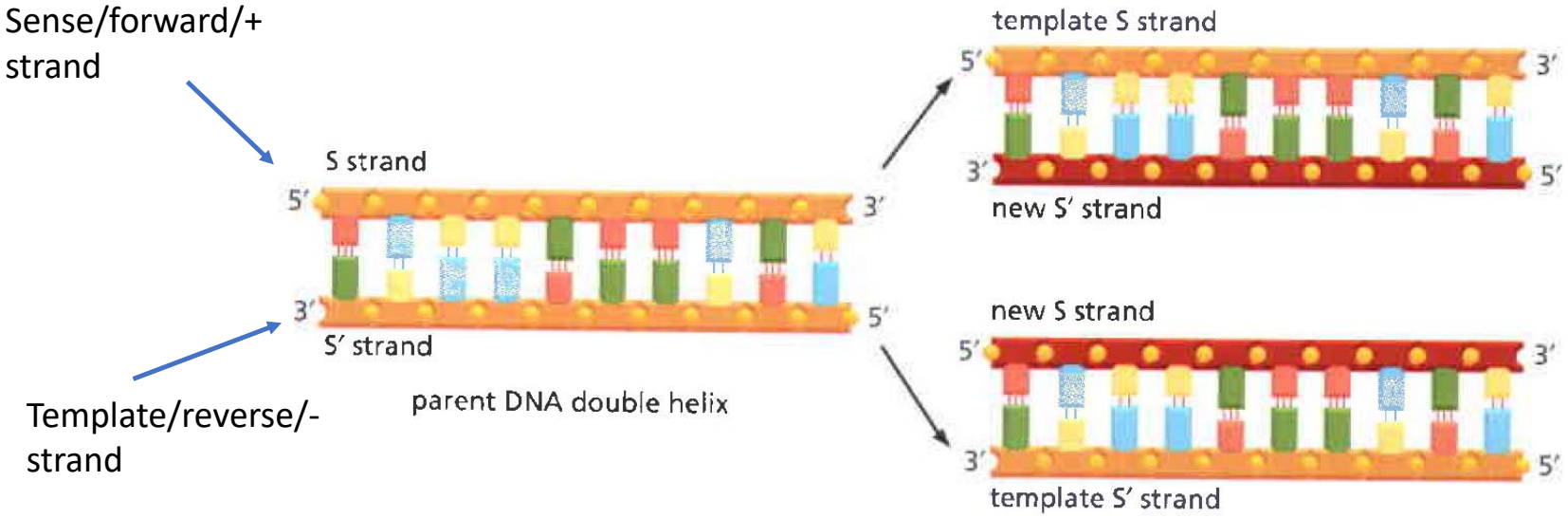
A gene consists of two ***alleles,*** one from each parent.



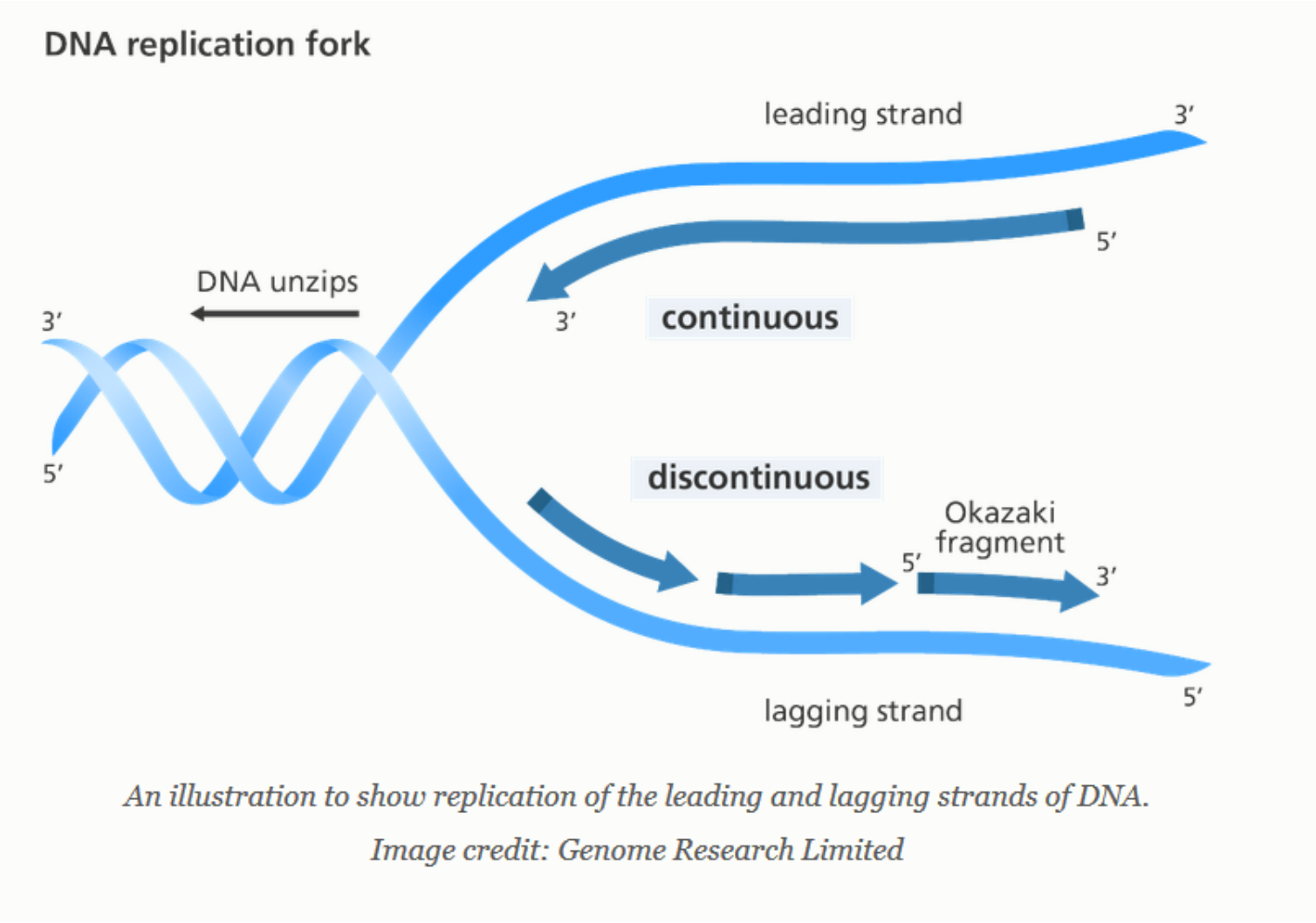
Genes

- A **gene** is a specific sequence of nucleotide bases along a chromosome carrying information for constructing a protein. A gene encodes a protein (or an RNA).
- The distance between **genes** is often much larger than the genes themselves.
- The human genome has around 23,500 genes.

DNA Replication: The Process



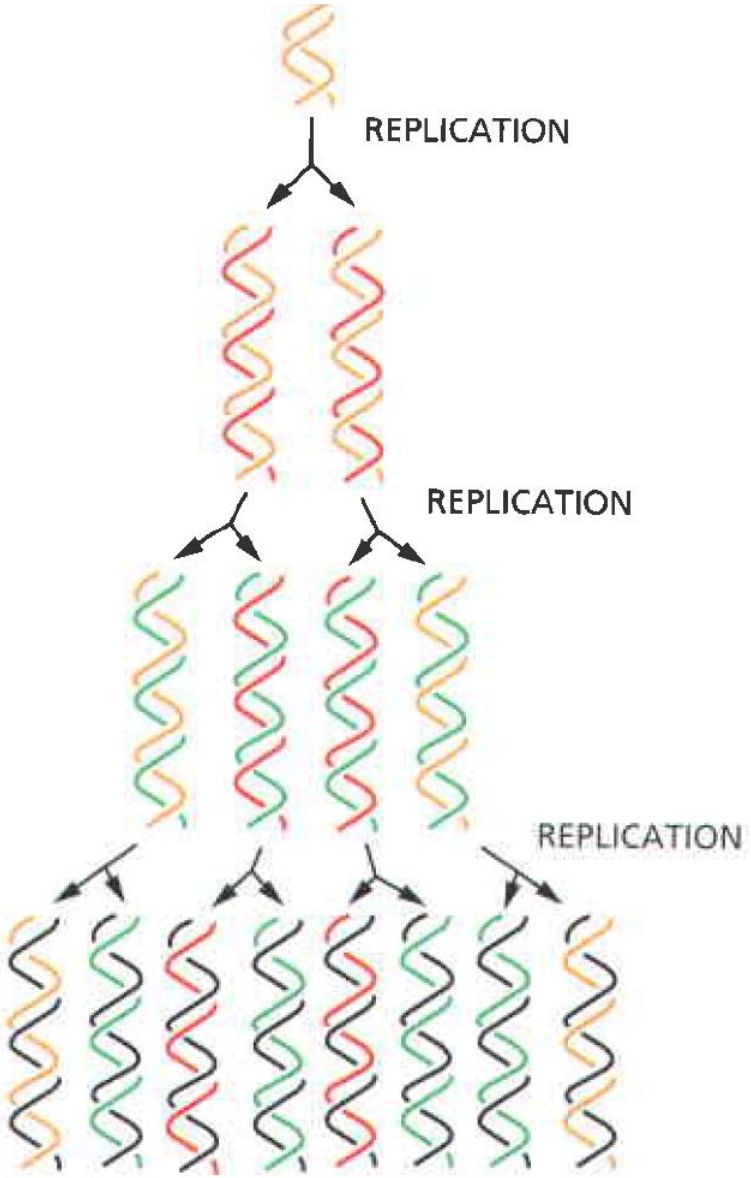
The DNA double helix acts as a template for its own duplication. Because the nucleotide A will pair successful only with T and G only with C, each strand of DNA can serve as a template to specify the sequence of nucleotides in its complementary strand by DNA base pairing. In this way, a double-helical DNA molecule can be copied precisely.



An illustration to show replication of the leading and lagging strands of DNA.

Image credit: Genome Research Limited

[Video DNA Replication]



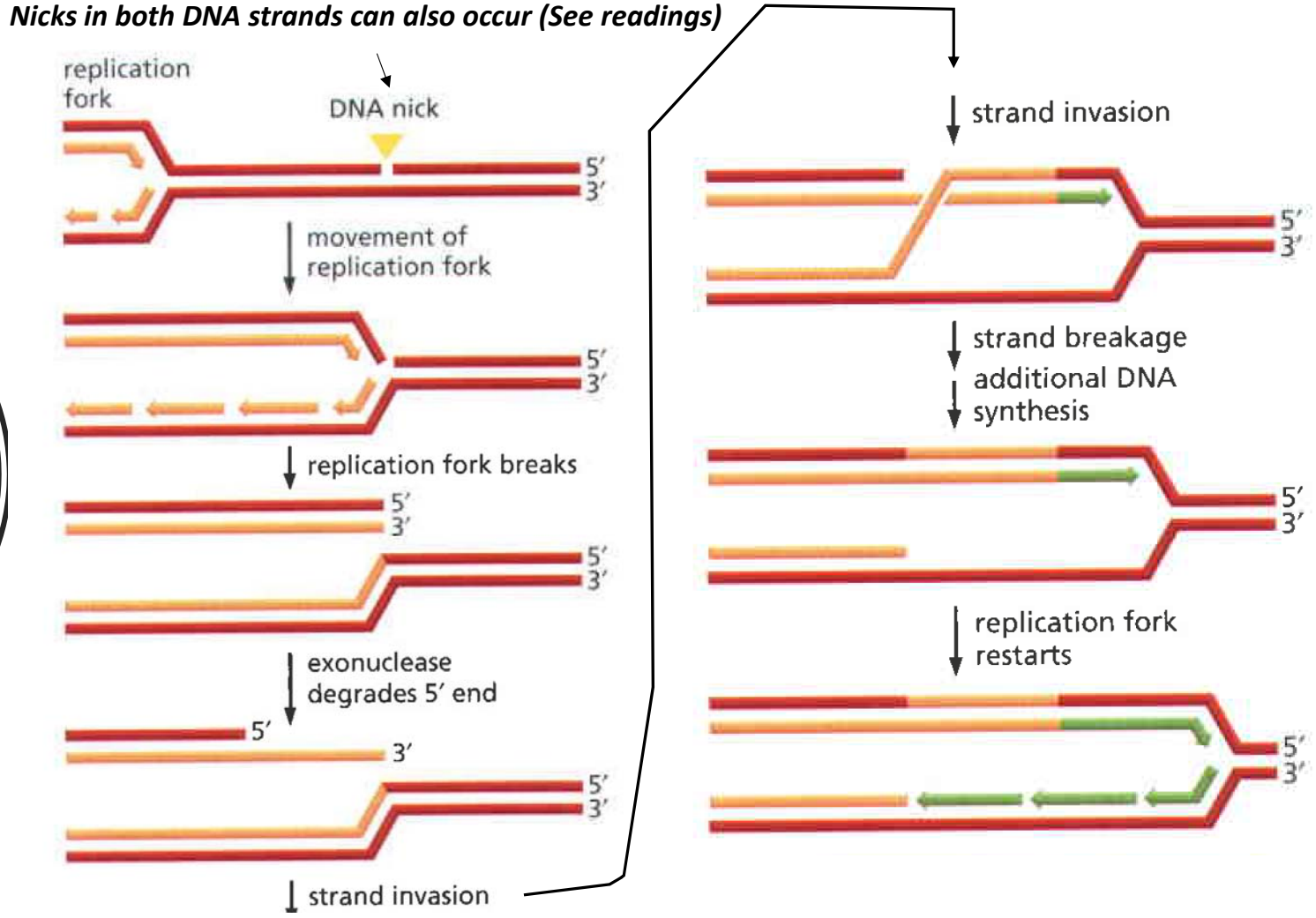
DNA Replication Happens With Fidelity

- **Mistakes occur every ~10^9 nicleotides copied.**
- **Is a semiconserved process.**
- **There are mechanisms to help repair nicks/breaks in the DNA.**
- **There are mechanisms to repair nucleotide mismatches.**

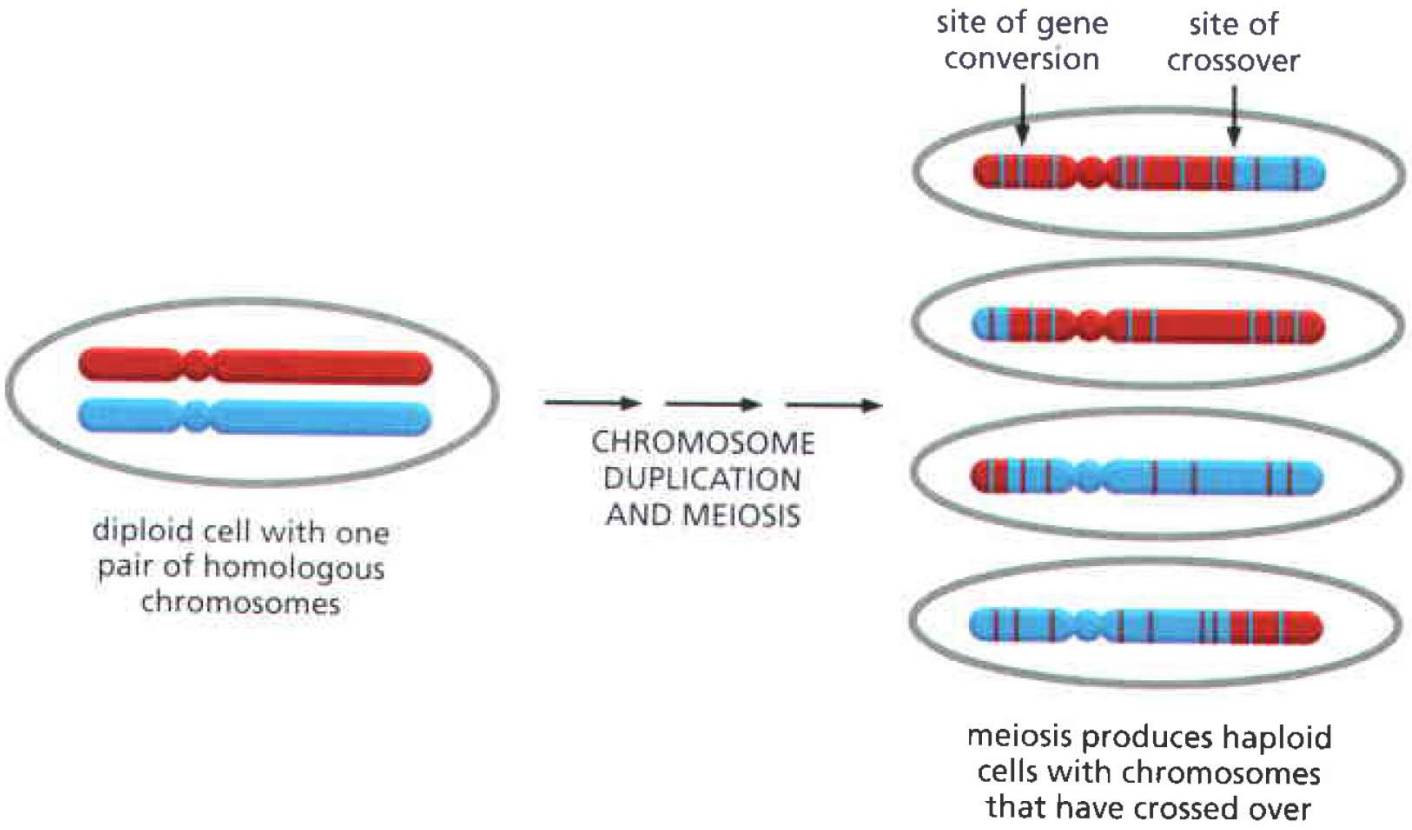
Things That Can Go Wrong With Replication

Homologous Recombinat-ion

Nicks in both DNA strands can also occur (See readings)



Chromosome crossing over occurs in meiosis.

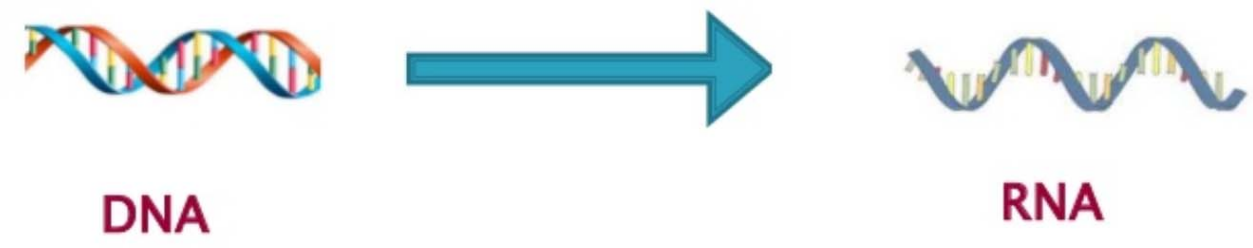


Transcription

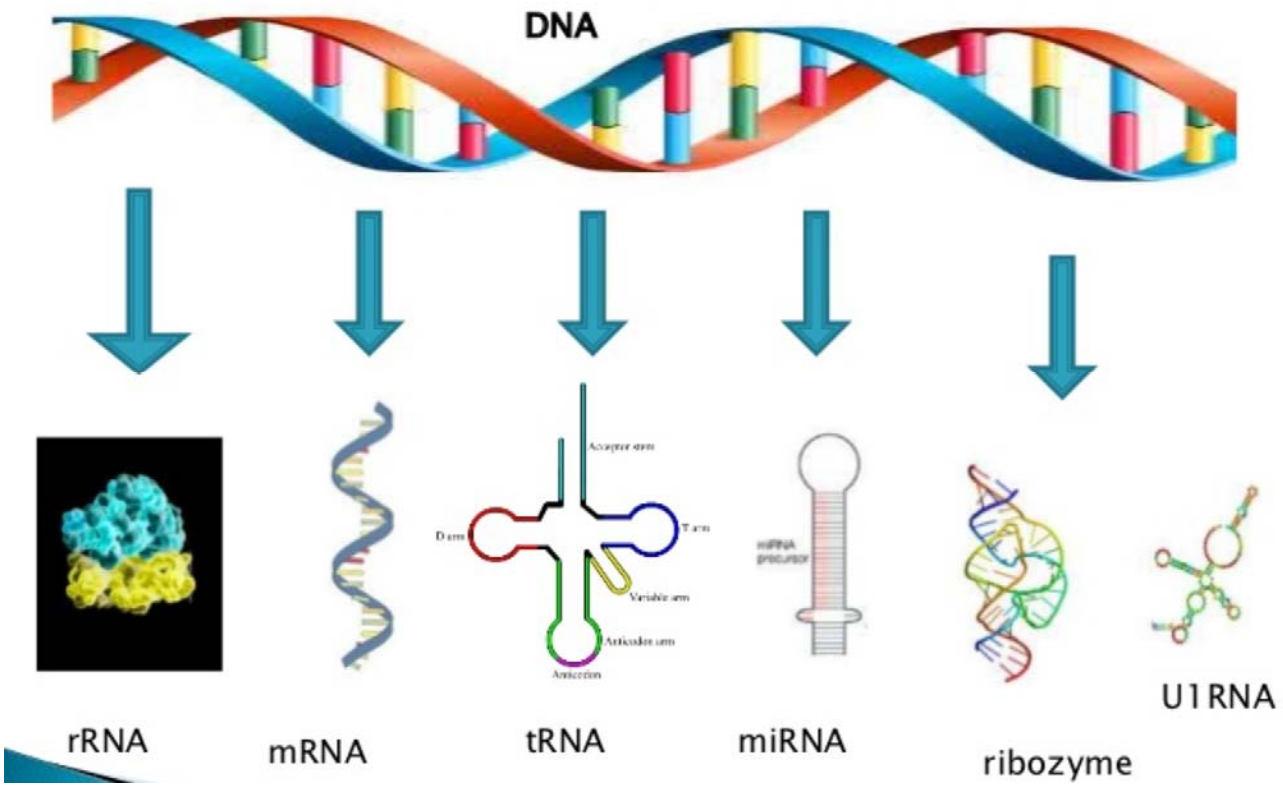
Transcription

Definition

Cellular process in which RNA is synthesized using DNA as a template known as **TRANSCRIPTION**.



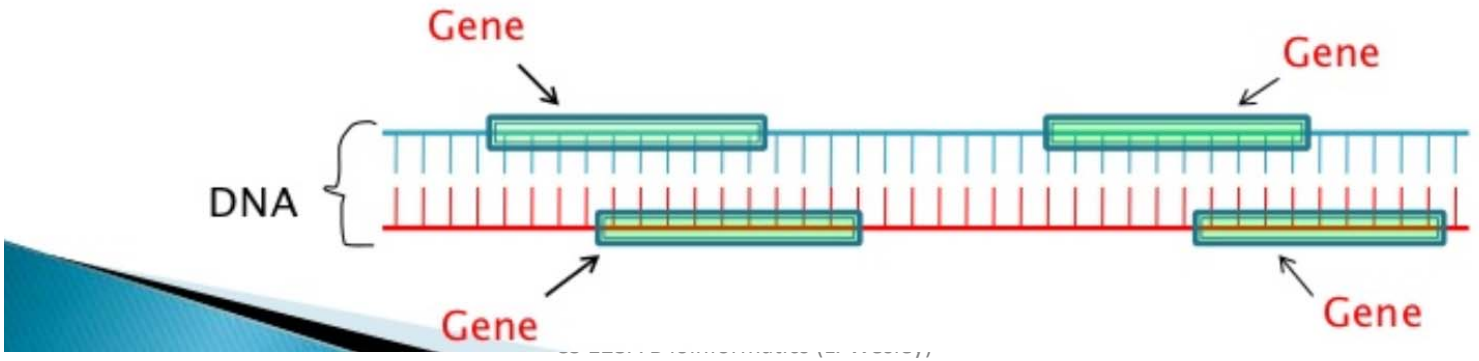
DNA Is The Blueprint For Many Types of RNA



Transcription

- In transcription a particular gene or group of genes are copied at any time, & some portions of DNA are never transcribed.

- Gene is a segment of DNA that codes for a type of **Protein** or for **RNA** & may present on any strand of DNA (contain many genes.)

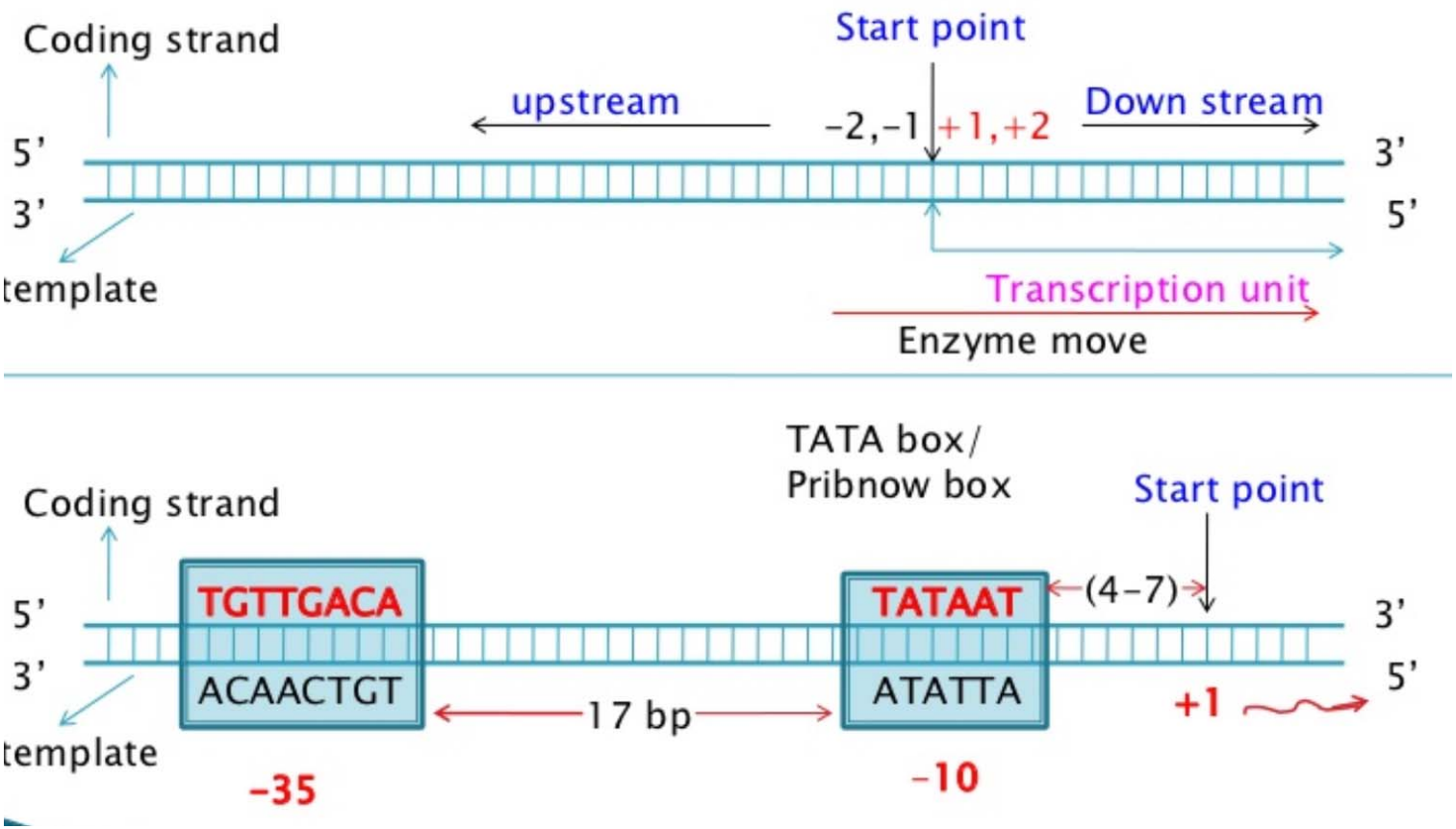


Transcription is Highly Selective

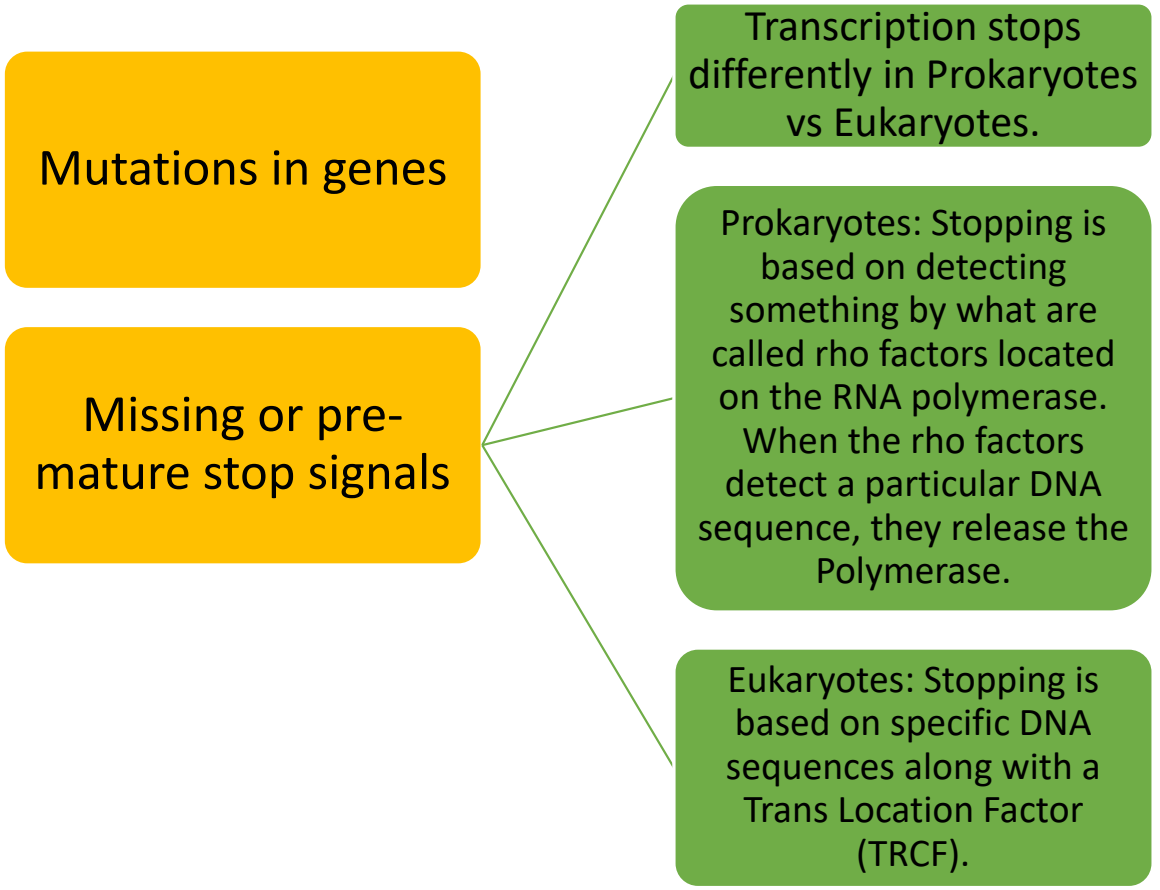
- This selectivity is due to signals embedded in the nucleotide sequence of DNA.
- Specific sequences mark the beginning and end of the DNA segment which is to be transcribed.
- This signals instruct the enzyme

where to start & stop the transcription
when to start,
how often to start .

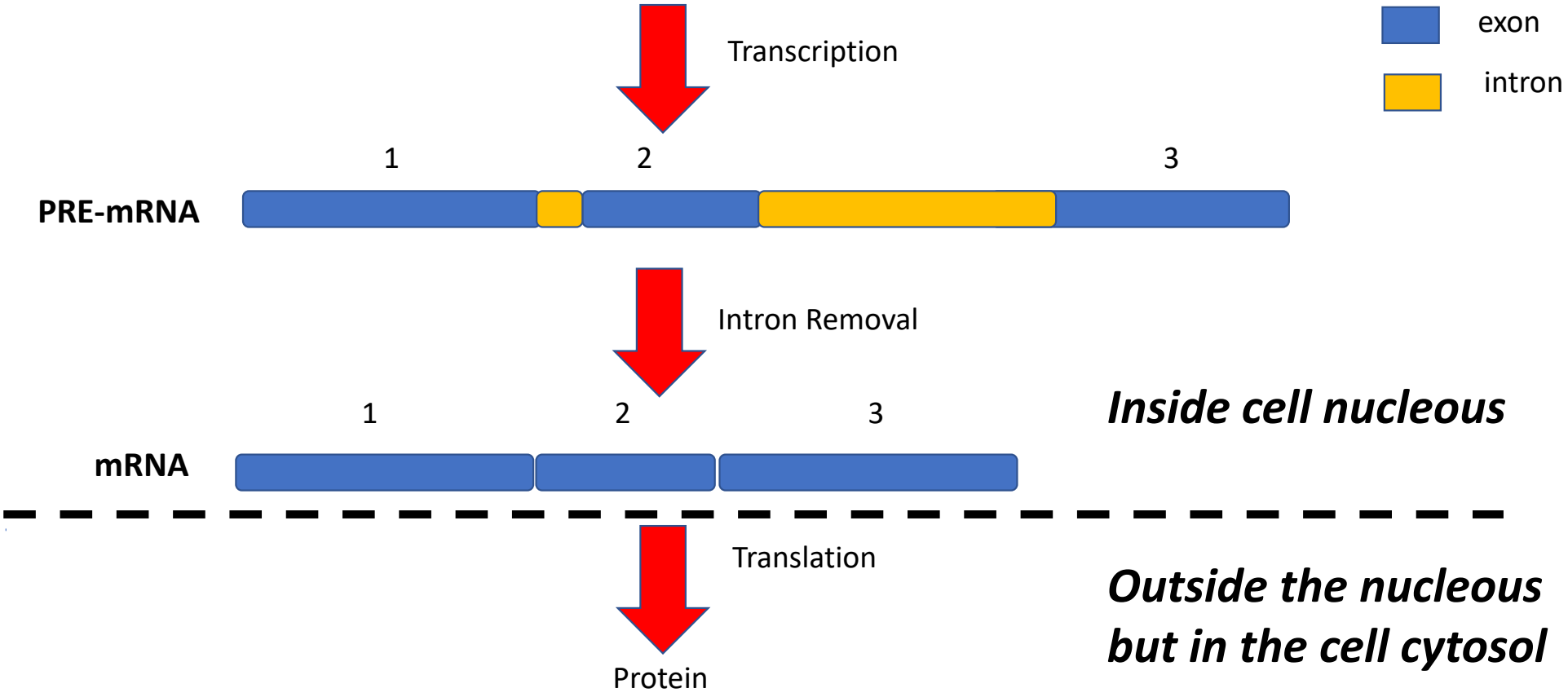
The Starting Position For Transcription



Potential Problems With Transcription



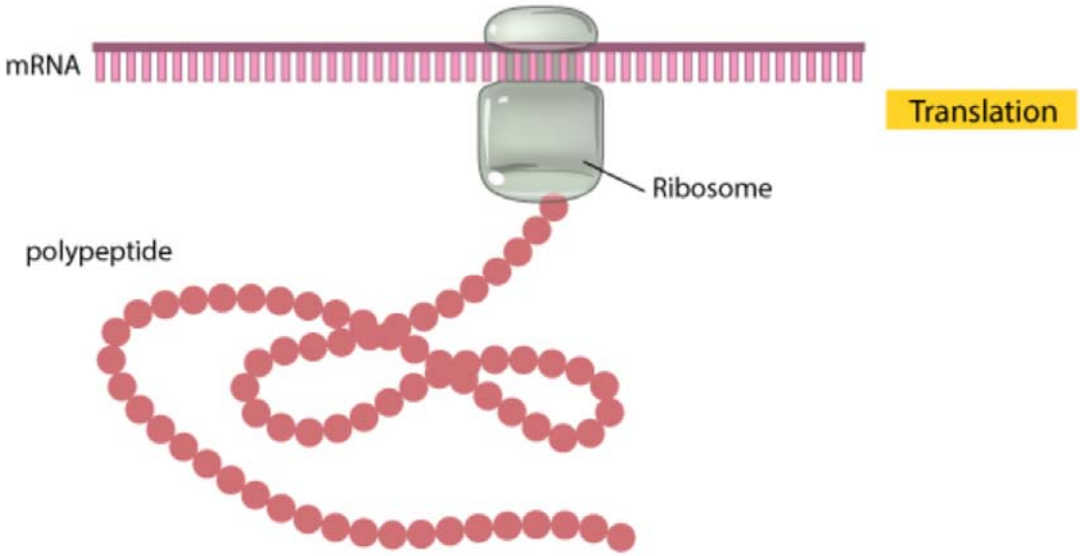
Pre-mRNA → mRNA (messenger RNA)



Inside cell nucleous

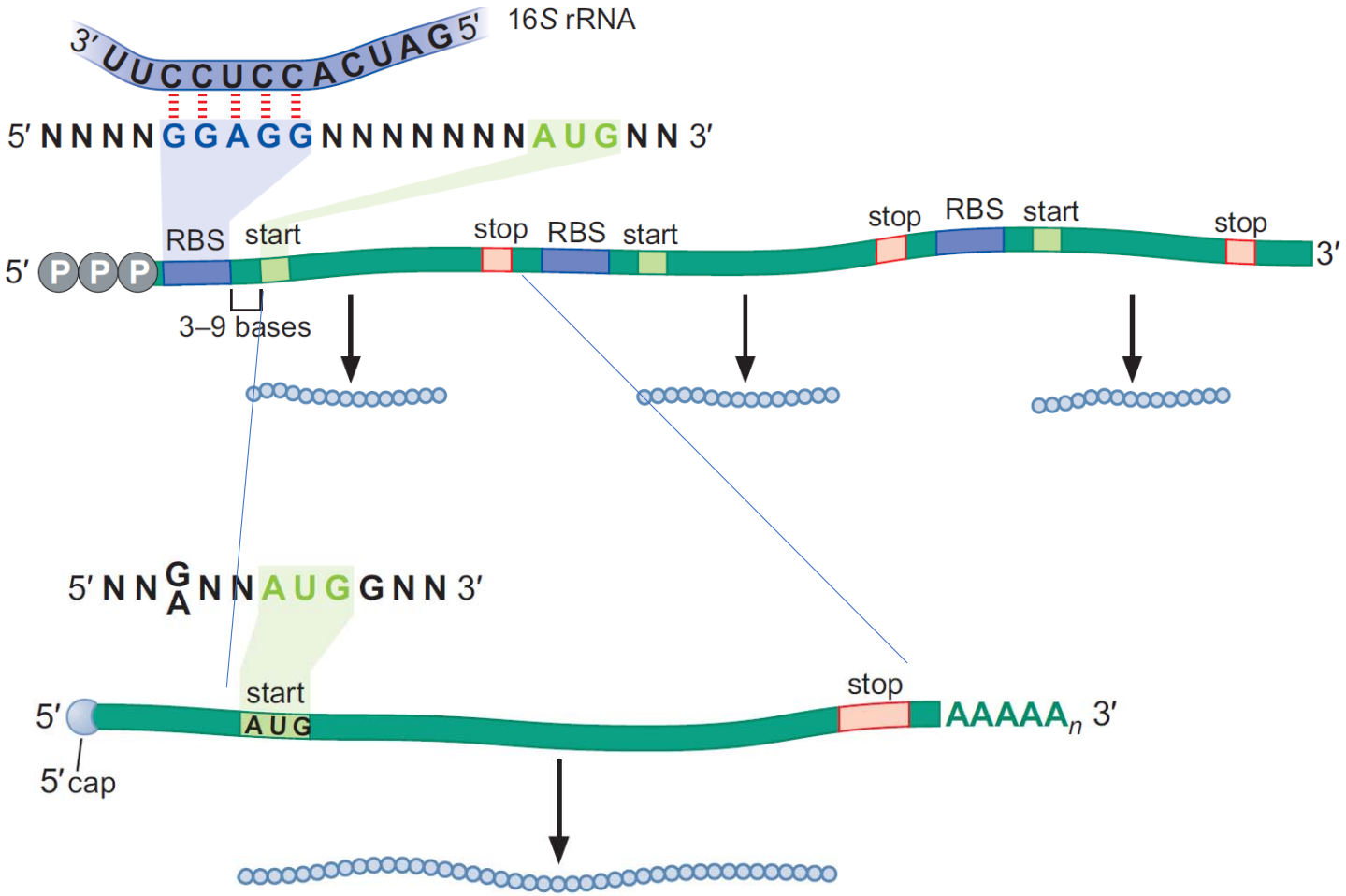
Outside the nucleous but in the cell cytosol

Ribosomes Translate mRNA Into Proteins

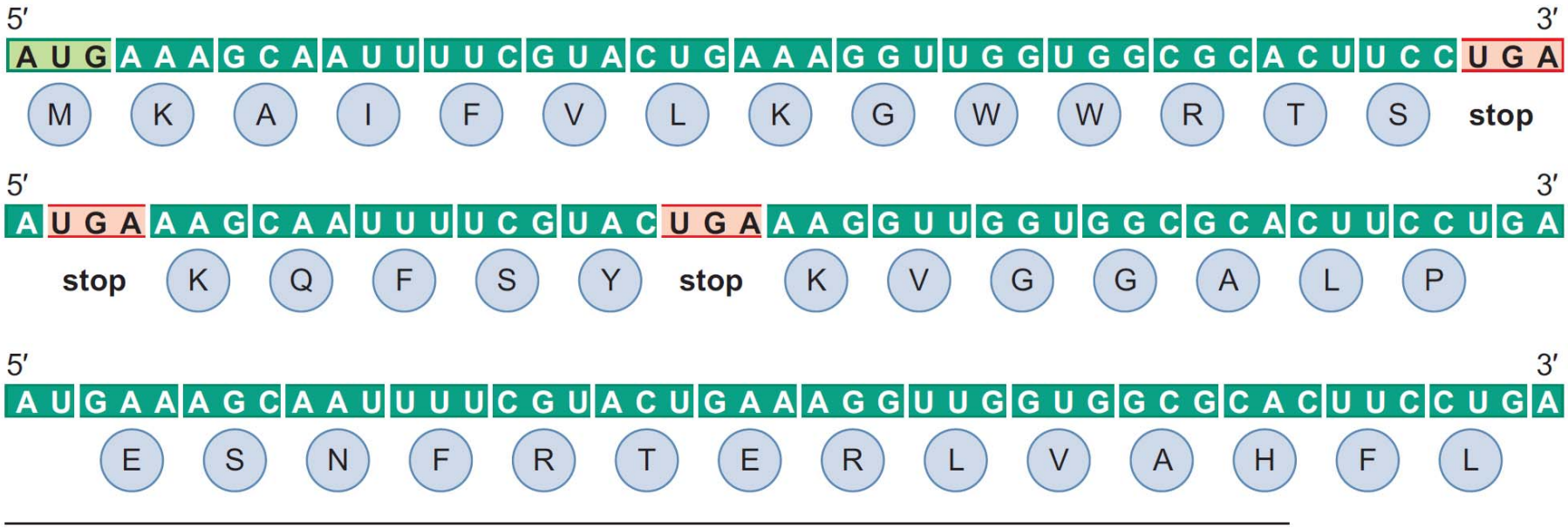


This is accomplished using a "code" …

Structure of mRNA



Triples Of RNA nucleoties Specify An Amino Acid



RNA Codon Table

		2nd base			
		U	**C**	**A**	**G**
1st base	**U**	UUU (Phe/F)Phenylalanine UUC (Phe/F)Phenylalanine UUA (Leu/L)Leucine UUG (Leu/L)Leucine	UCU (Ser/S)Serine UCC (Ser/S)Serine UCA (Ser/S)Serine UCG (Ser/S)Serine	UAU (Tyr/Y)Tyrosine UAC (Tyr/Y)Tyrosine UAA Ochre (*Stop*) UAG Amber (*Stop*)	UGU (Cys/C)Cysteine UGC (Cys/C)Cysteine UGA Opal (*Stop*) UGG (Trp/W)Tryptophan
	C	CUU (Leu/L)Leucine CUC (Leu/L)Leucine CUA (Leu/L)Leucine CUG (Leu/L)Leucine	CCU (Pro/P)Proline CCC (Pro/P)Proline CCA (Pro/P)Proline CCG (Pro/P)Proline	CAU (His/H)Histidine CAC (His/H)Histidine CAA (Gln/Q)Glutamine CAG (Gln/Q)Glutamine	CGU (Arg/R)Arginine CGC (Arg/R)Arginine CGA (Arg/R)Arginine CGG (Arg/R)Arginine
	A	AUU (Ile/I)Isoleucine AUC (Ile/I)Isoleucine AUA (Ile/I)Isoleucine AUG (Met/M)Methionine, *(Start)*	ACU (Thr/T)Threonine ACC (Thr/T)Threonine ACA (Thr/T)Threonine ACG (Thr/T)Threonine	AAU (Asn/N) Asparagine AAC (Asn/N)Asparagine AAA (Lys/K)Lysine AAG (Lys/K)Lysine	AGU (Ser/S) Serine AGC (Ser/S)Serine AGA (Arg/R) Arginine AGG (Arg/R)Arginine
	G	GUU (Val/V)Valine GUC (Val/V)Valine GUA (Val/V)Valine GUG (Val/V)Valine	GCU (Ala/A)Alanine GCC (Ala/A)Alanine GCA (Ala/A)Alanine GCG (Ala/A)Alanine	GAU (Asp/D)Aspartic acid GAC (Asp/D)Aspartic acid GAA (Glu/E)Glutamic acid GAG (Glu/E)Glutamic acid	GGU (Gly/G)Glycine GGC (Gly/G)Glycine GGA (Gly/G)Glycine GGG (Gly/G)Glycine

https://biopeptek.com/peptide-resources/rna-codon-table/

Exercise

- What is the amino acid (AA) sequence that would be formed from the following DNA sequence? Show the mRNA sequence.

3'-T A C G C C T A T T T C A A C A G G C G C G G G A C T - 5'

- Is this the only DNA sequence that can produce the AA sequence? If not, provide an alternative.

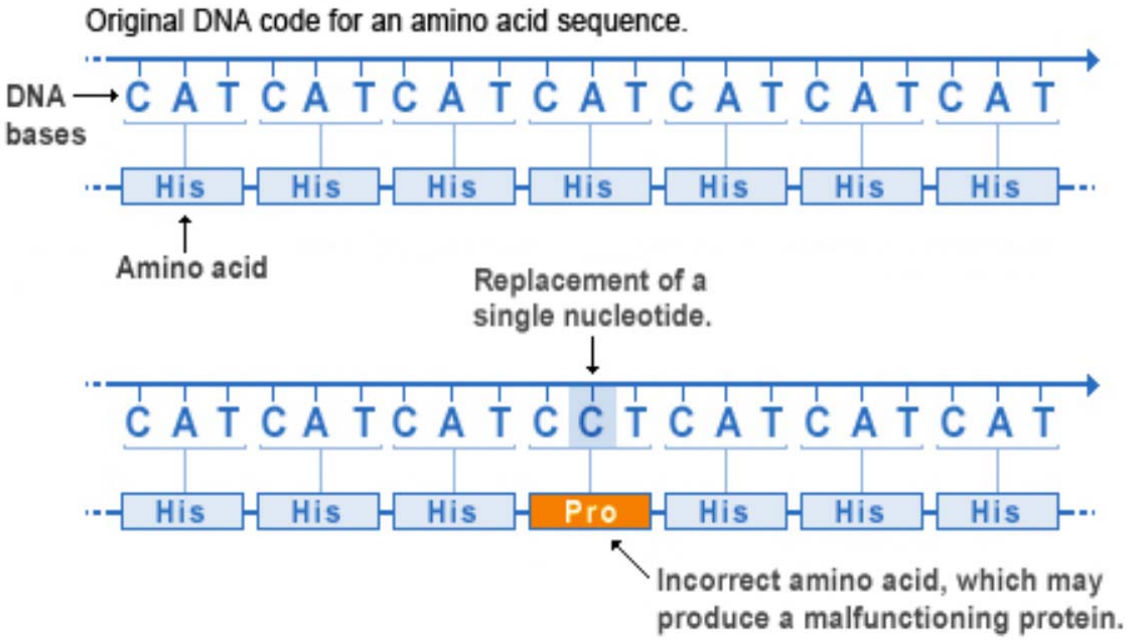
- What is the sense strand sequence that might have resulted in the following AA sequence? Are there other possible sense strands?

M P P K I L L Q E T T V F L D G

Types of Mutations

Missense Mutation

This type of mutation is a change in one DNA base pair that results in the substitution of one amino acid for another in the protein made by a gene.

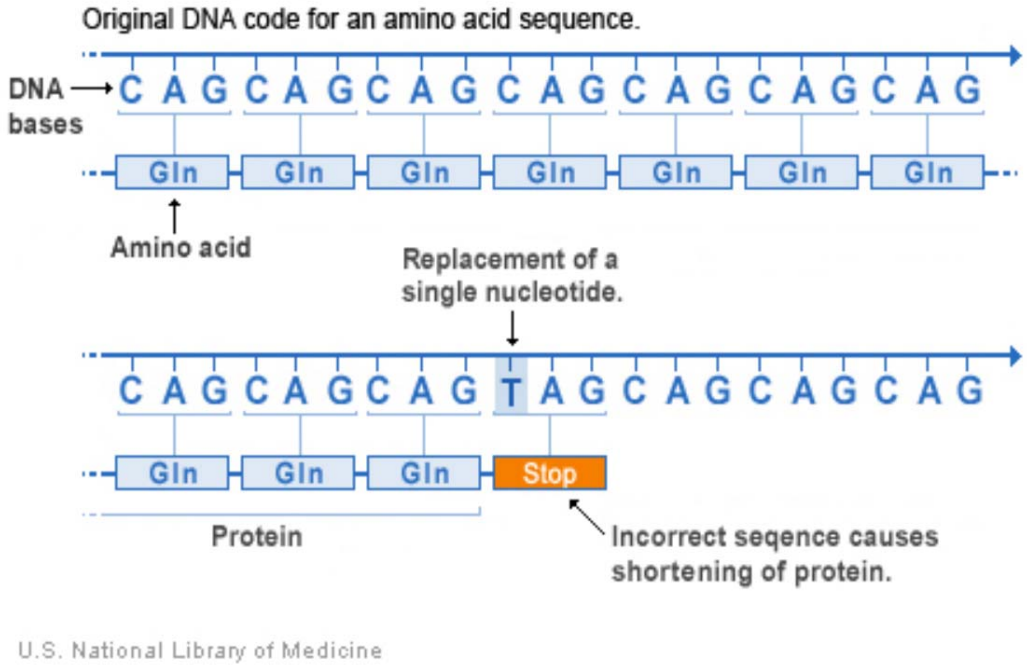


U.S. National Library of Medicine

Nonsense Mutation

A nonsense mutation is also a change in one DNA base pair. Instead of substituting one amino acid for another, however, the altered DNA sequence prematurely signals the cell to stop building a protein.

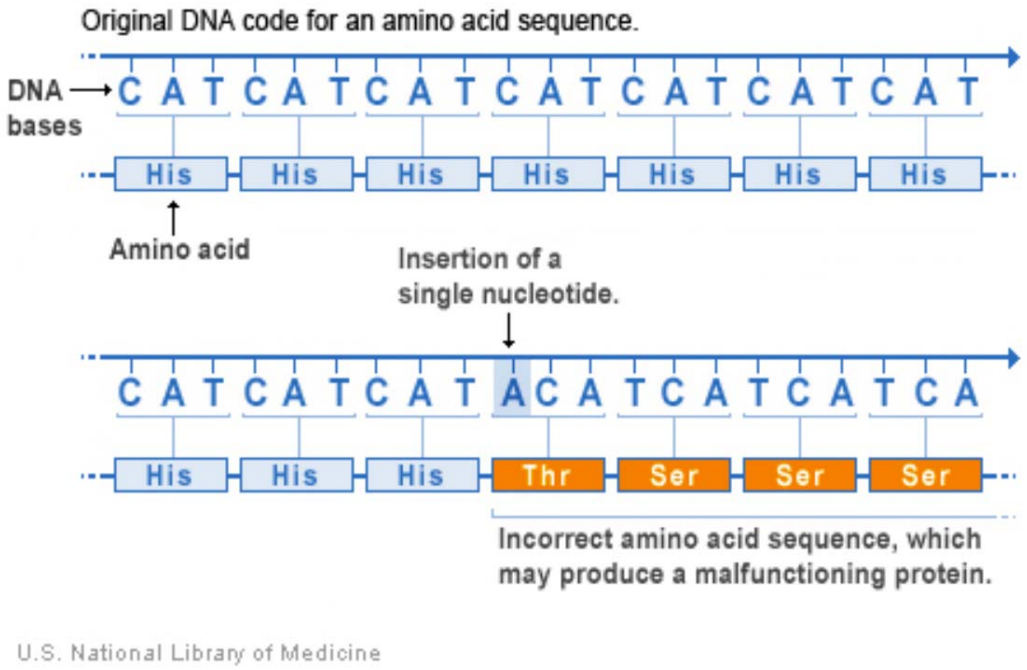
This type of mutation results in a shortened protein that may function improperly or not at all.



U.S. National Library of Medicine

Insertion Mutation

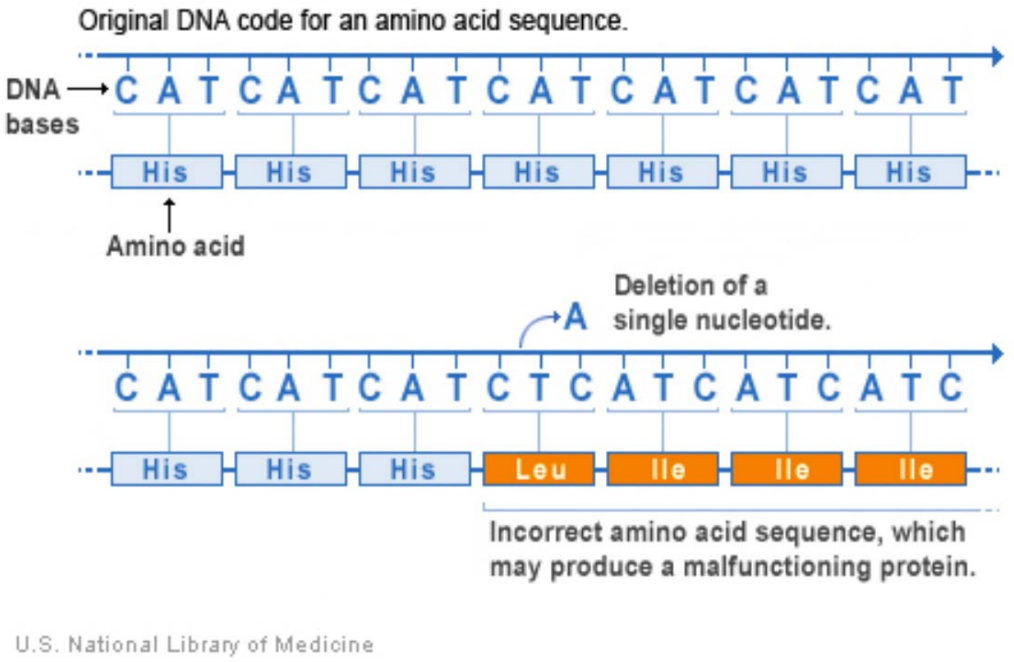
An insertion changes the number of DNA bases in a gene by adding a piece of DNA. As a result, the protein made by the gene may not function properly.



Deletion Mutation

A deletion changes the number of DNA bases by removing a piece of DNA.

Small deletions may remove one or a few base pairs within a gene, while larger deletions can remove an entire gene or several neighboring genes. The deleted DNA may alter the function of the resulting protein(s).

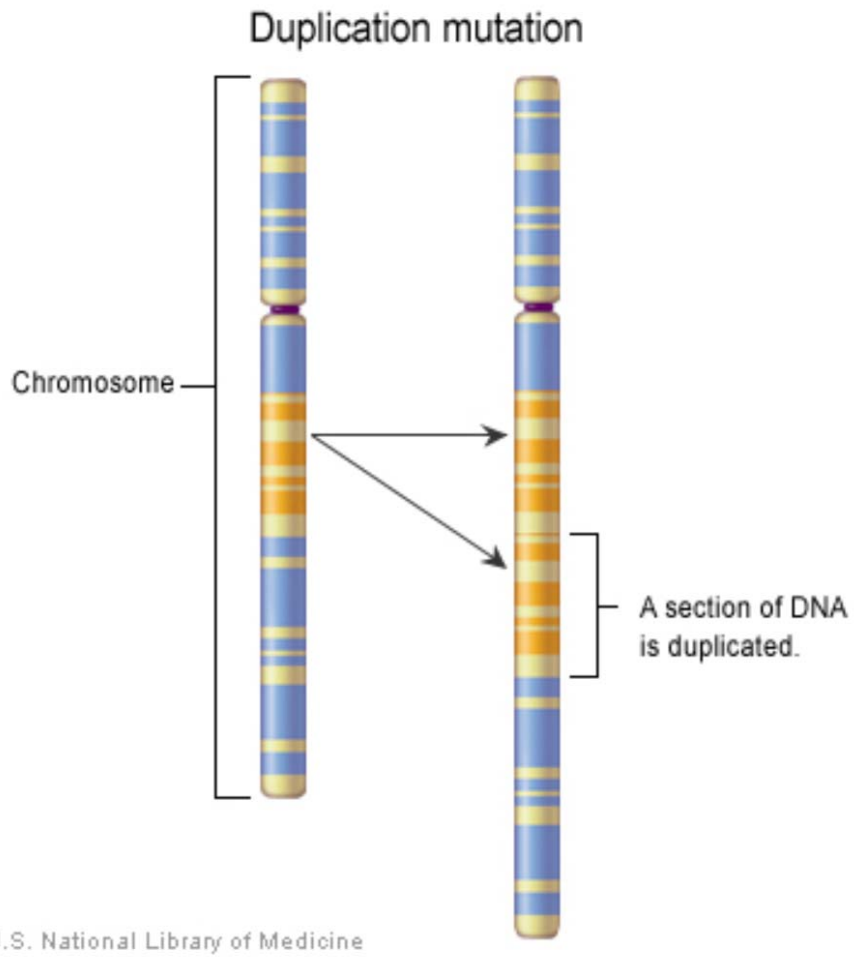


U.S. National Library of Medicine

Duplication Mutation

A duplication consists of a piece of DNA that is abnormally copied one or more times. This type

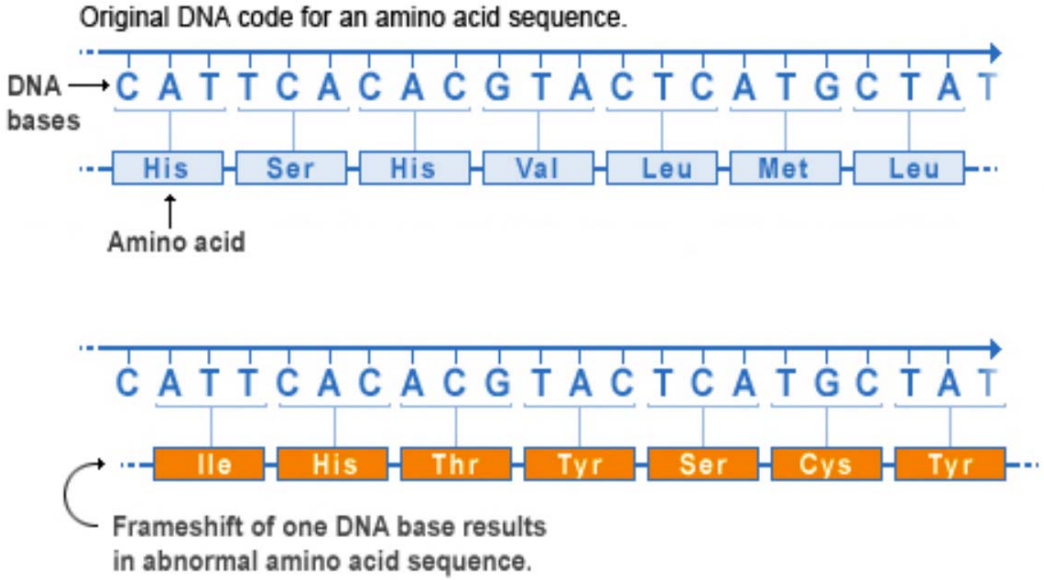
of mutation may alter the function of the resulting protein.



Frameshift Mutations

This type of mutation occurs when the addition or loss of DNA bases changes a gene's reading frame. A reading frame consists of groups of 3 bases that each code for one amino acid. A frameshift mutation shifts the grouping of these bases and changes the code for amino acids.

The resulting protein is usually nonfunctional. Insertions, deletions, and duplications can all be frameshift mutations.

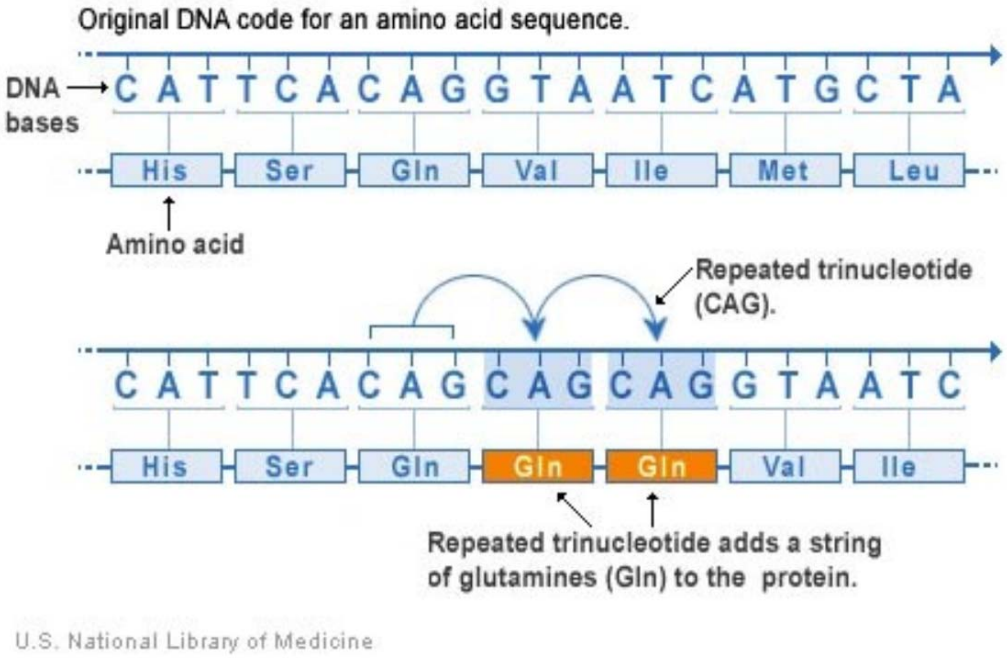


U.S. National Library of Medicine

Repeat Expansion Mutations

Nucleotide repeats are short DNA sequences that are repeated a number of times in a row. For example, a trinucleotide repeat is made up of 3-base-pair sequences, and a tetranucleotide repeat is made up of 4-base-pair sequences.

A repeat expansion is a mutation that increases the number of times that the short DNA sequence is repeated. This type of mutation can cause the resulting protein to function improperly.



U.S. National Library of Medicine

Gene Mutations: The Good, Bad, and The Ugly

- Evolution is the process by which populations of organisms change over generations. Genetic variations underlie these changes.

- Genetic variations can arise from gene mutations or from genetic recombination (a normal process in which genetic material is rearranged as a cell is getting ready to divide). These variations often alter gene activity or protein function, which can introduce different traits in an organism.

- If a trait is advantageous and helps the individual survive and reproduce, the genetic variation is more likely to be passed to the next generation (a process known as natural selection).

- Over time, as generations of individuals with the trait continue to reproduce, the advantageous trait becomes increasingly common in a population, making the population different than an ancestral one. Sometimes the population becomes so different that it is considered a new species.

Gene Mutations: The Good, Bad, and The Ugly

- Not all mutations lead to evolution. Only hereditary mutations, which occur in egg or sperm cells, can be passed to future generations and potentially contribute to evolution.

- Some mutations occur during a person's lifetime in only some of the body's cells and are not hereditary, so natural selection cannot play a role. Also, many genetic changes have no impact on the function of a gene or protein and are not helpful or harmful.

- The environment in which a population of organisms lives is integral to the selection of traits. Some differences introduced by mutations may help an organism survive in one setting but not in another—for example, resistance to a certain bacteria is only advantageous if that bacteria is found in a particular location and harms those who live there.

Gene Mutations: The Good, Bad, and The Ugly

- So why do some harmful traits, like genetic diseases, persist in populations instead of being removed by natural selection? There are several possible explanations, but in many cases, the answer is not clear.

- For some conditions, such as the neurological condition Huntington disease, signs and symptoms do not occur until after a person has children, so the gene mutation can be passed on despite being harmful.

- For other harmful traits, a phenomenon called reduced penetrance, in which some individuals with a disease-associated mutation do not show signs and symptoms of the condition, can also allow harmful genetic variations to be passed to future generations.

- For some conditions, having one mutated copy of a gene in each cell is advantageous, while having two mutated copies causes disease. The best-studied example of this phenomenon is sickle cell disease: Having two mutated copies of the HBB gene in each cell results in the disease, but having only one copy provides some resistance to malaria.

Proteins Will Be Covered At The Time Protein DBs Are Introduced