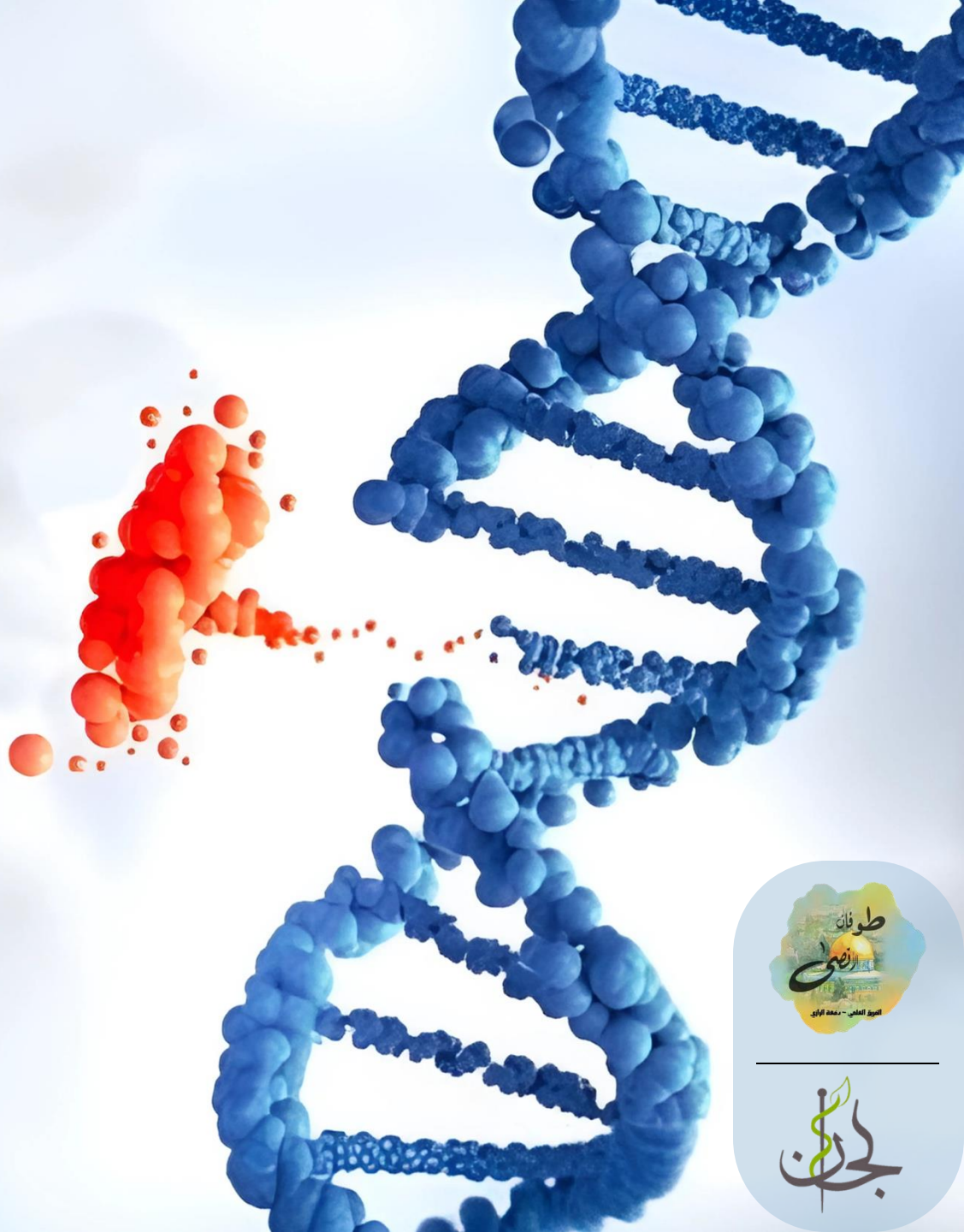
Genetics

Modified no.

Writer: Mais Salman & Lama Abu-Ismail

Editor: Ismail Alarda & Khadijah Naser

Doctor: Zaid Aburubaiha



- This lecture has many repeated info from Tawjihi & basic science + during the lecture many info will be repeated
- Don't get panic, its an easy one 😊, Say بسم الله And let's start

Color code

- Slides
- Doctor
- Additional info
- Important

DNA replication

DNA replication consists basically of the breaking of the weak hydrogen bonds between the bases, leaving a single DNA strand with each base unpaired.

The consistent pairing of adenine with thymine and guanine with cytosine, known as **complementary base pairing**, is the key to accurate replication

- As we've taken before, this process starts with double-stranded DNA that will separate into two single-stranded DNA to begin the process of replication, which starts by a polymerase enzyme that will achieve a complementary base pairing.

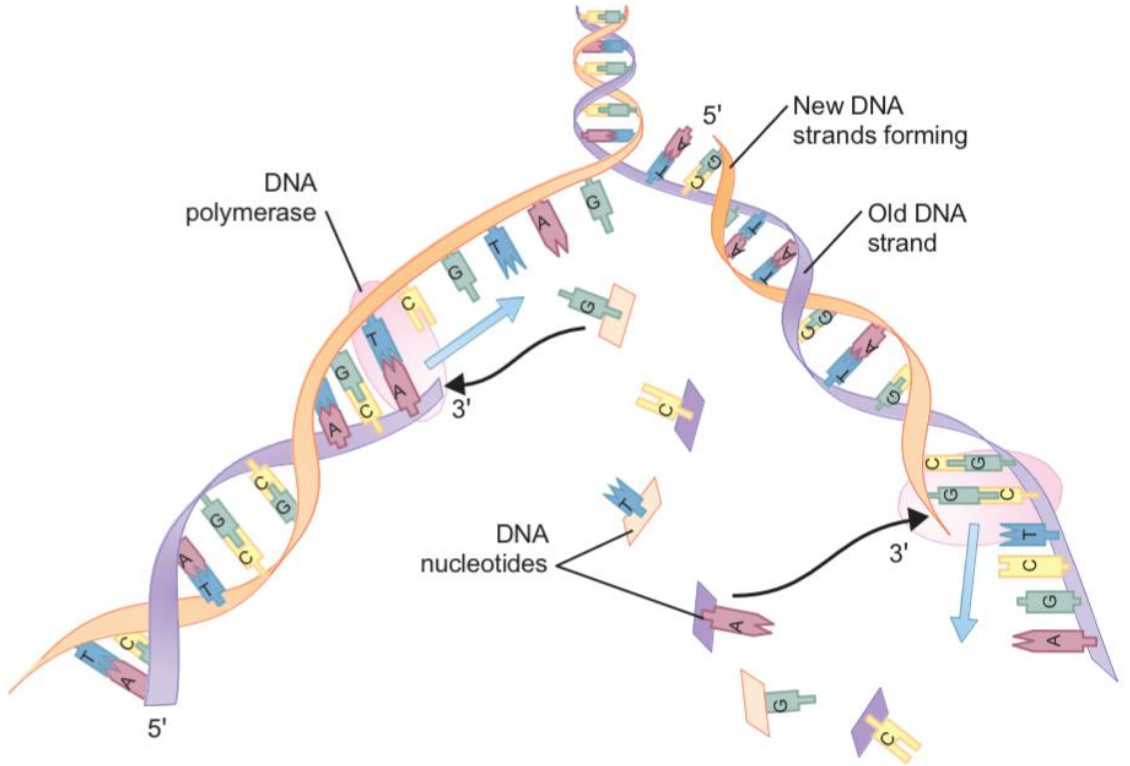


FIG 2-5 DNA replication. The hydrogen bonds between the two original strands are broken, allowing the bases in each strand to undergo complementary base pairing with free bases. This process, which proceeds in the 5′ to 3′ direction on each strand, forms two new double strands of DNA.

- DNA replication is a very important process where Adenine always binds to Thymine and Cytosine to Guanine. When there's a defect in this binding, mutations happen.

- DNA replication is a very important process in mitosis to increase the number of cells either to repair damaged cells or for the human growth (from a baby to a child to an adult).
- Since we have billions of cells, the replication process should be very efficient, otherwise human will build up many mutations and problems.

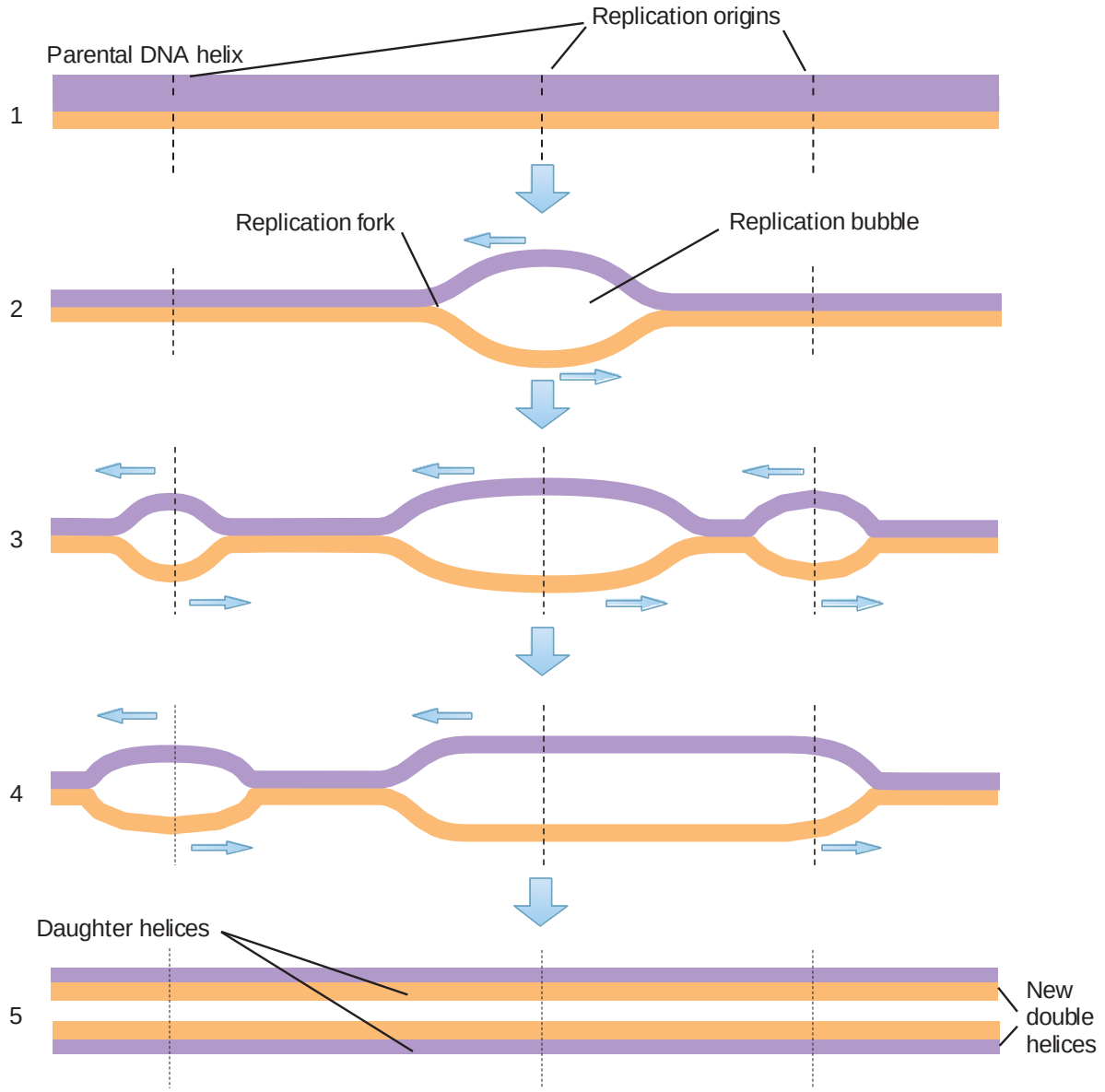


FIG 2-6 Replication bubbles form at multiple points along the DNA strand, allowing replication to proceed more rapidly.

- Since our genome is very big (2.3 billion bases) and polymerase enzyme can add 40 to 50 bases/sec **in humans** (this is considered slow), replication process should be fast, so it starts at multiple different sites of the DNA molecule at the same time " this facilitates an efficient and fast replication process".

- While **in bacteria**, the replication process is considered fast, for example, bacterial polymerase enzyme can add 500 to 1000 bases/sec, so bacterial growth is much faster than our cells growth.

- Remember!! polymerase enzyme has two functions: **1-adding nucleotides** 2-after finishing, it do **proofreading** to ensure there is no replication mistakes.

- To sum up, DNA replication process must happen during mitosis (cell devision, cell growth) and any mistake in the polymerase enzyme function will lead to mutations.

From Genes to Proteins

- The genetic material have another function which is being transcribed and translated into proteins.

- Involves two processes, transcription and translation.

- The DNA code is transcribed into messenger RNA, which then leaves the nucleus to be translated into proteins.

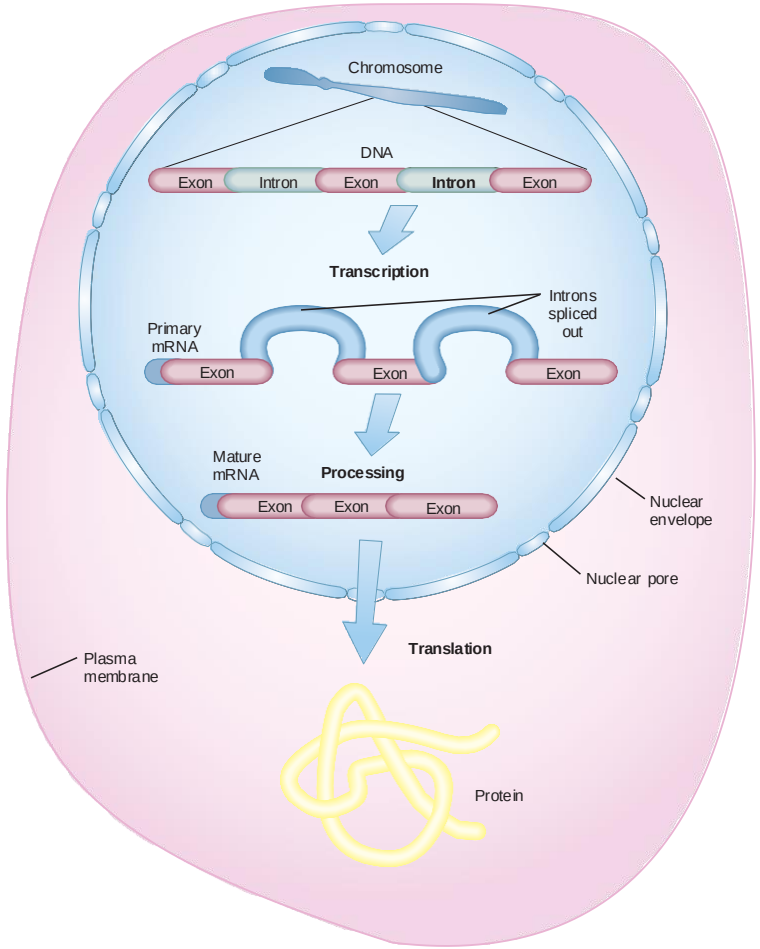
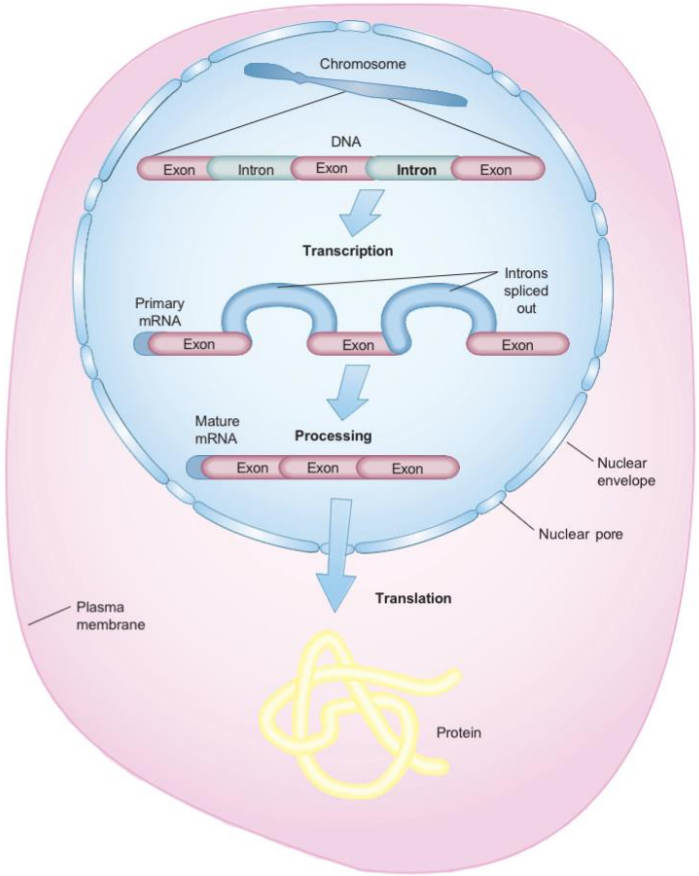


FIG 2-7 A summary of the steps leading from DNA to proteins. Replication and transcription occur in the cell nucleus. The mRNA is then transported to the cytoplasm, where translation of the mRNA into amino acid sequences composing a protein occurs.

- mRNA at the beginning is called primary transcript which contains both **introns** and **exons** (introns form the **majority** of the DNA).
- Then introns are removed since they are not coding sequences.
- Then mature transcript is formed (mature mRNA), which exists the nucleus to the cytoplasm where it will be translated into a protein by the ribosomes.
- So for transcription, I need only one single stranded DNA, where the gene promoter will bind to, and the RNA polymerase to form mRNA.
- Keep in mind that the two strands of the DNA are anti-parallel 5' to 3' // 3' to 5'.
- Also keep in mind that this process (production of proteins) is determined based on the needs of the cell and by other regulatory proteins inside it, which guide the polymerase enzyme to bind and start the transcription.



Transcription

Transcription is the process by which an RNA sequence is formed from a DNA template.

To initiate mRNA transcription, one of the **RNA polymerase** enzymes (RNA polymerase II) binds to a promoter site on the DNA (it detects this sequence specifically).

The RNA polymerase then pulls a portion of the DNA strands apart from one another, exposing unattached DNA bases.

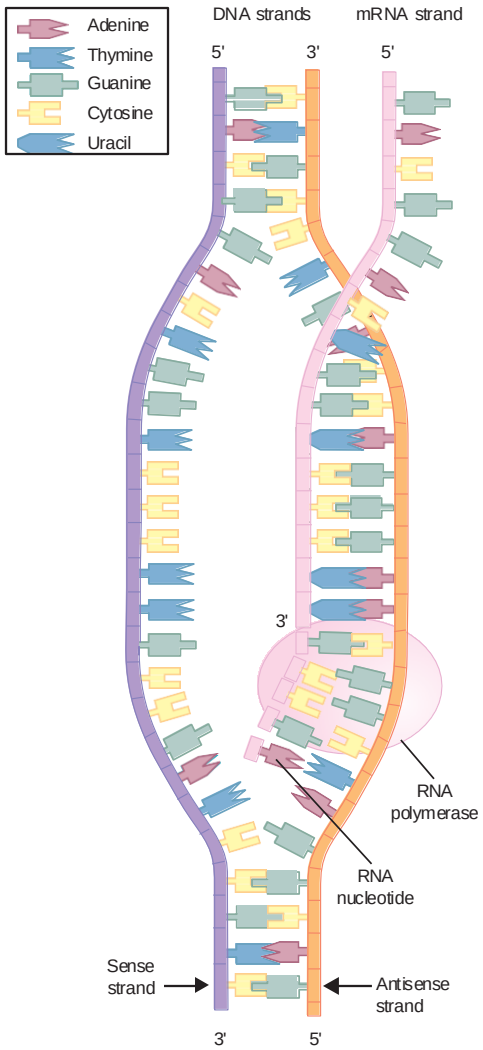


FIG 2-8 Transcription of DNA to mRNA. RNA polymerase II proceeds along the DNA strand in the 3′ to 5′ direction, assembling a strand of mRNA nucleotides that is complementary to the DNA template strand.

Only **one strand** is chosen to serve as a template for mRNA synthesis.

This choice is **determined by the promoter sequence**, which orients the RNA polymerase in a specific direction along the DNA sequence.

mRNA is synthesized only in the **5' to 3' direction**. RNA polymerase moves in the 3' to 5' direction along the DNA template strand.

Because of **complementary base pairing**, the mRNA nucleotide sequence is identical to that of the DNA strand that does not serve as the template.

- In mRNA, thymine base is replaced by uracil
- So, if Adenine is in the DNA template strand, the mRNA will have uracil as a complementary base to it instead of thymine →

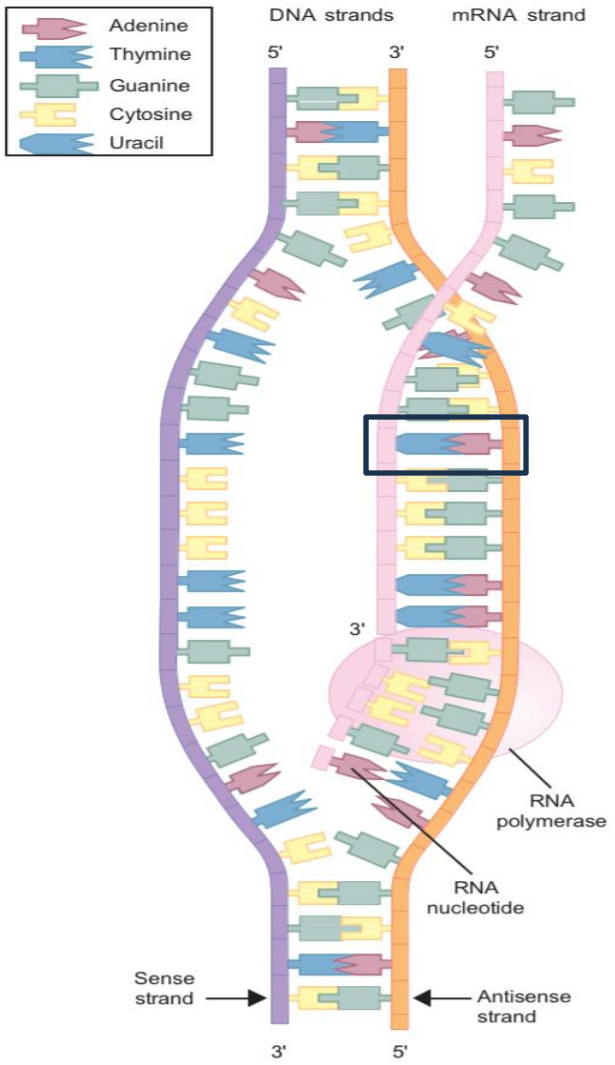


FIG 2-8 Transcription of DNA to mRNA. RNA polymerase II proceeds along the DNA strand in the 3′ to 5′ direction, assembling a strand of mRNA nucleotides that is complementary to the DNA template strand.

Soon after RNA synthesis begins, the 5' end of the growing RNA molecule is "capped" by the addition of a chemically modified guanine nucleotide at the 5’ end of the RNA. (Capping is imp. to protect RNA from degradation once it reaches the cytoplasm, and it is the starting point of translation).

Transcription continues until a group of bases called a termination sequence is reached. Near this point, a series of 100 to 200 adenine bases are added to the 3' end of the RNA molecule (poly- A tail). This tale is also important for stabilizing the RNA in the cytoplasm so it doesn’t get degraded.

Finally, the DNA strands and RNA polymerase separate from the RNA strand, leaving a transcribed single mRNA strand. This mRNA molecule is termed the primary transcript.(as we said before it will be processed to become a mature transcript).

Gene Splicing

In **eukaryotes**, Sections of the RNA are removed by nuclear enzymes, and the remaining sections are spliced together to form the functional mRNA that will migrate to the cytoplasm. (Side note: prokaryotes don't have introns)

The excised sequences are called introns, and the sequences that are left to code for proteins are called exons. Only when gene splicing is completed does the mature transcript move out of the nucleus into the cytoplasm.In gene splicing process, introns are removed.

Some genes contain alternative splice sites, which allow the same primary transcript to be spliced in different ways, ultimately producing different protein products from the same gene.

- **Alternative splicing sites**:
- Within the same gene complex, the expression of different exons leads to different products.
- As in beta globin complex, you know that we have HbA (2alpha,2 beta) & HbF(2alpha, 2gamma), so what happens during fetal life is the expression of gamma chains, but after birth it will be switched into beta chains by changing the splicing sites.

- In between introns & exons, there are **consensus regions** (they're detected by splicing enzymes which will splice them and join exons together. This will give us the mature transcript, which will leave the nucleus to the cytoplasm).
- Remember, capping and poly-A addition contribute to the stability of the mRNA.
- Any mutation in these consensus sequences (splice site mutation or "processing mutation") → leads to splicing at wrong site or absence of splicing → this will lead to expression of exons that shouldn't be expressed = mistake in mRNA processing & forming an abnormal protein.

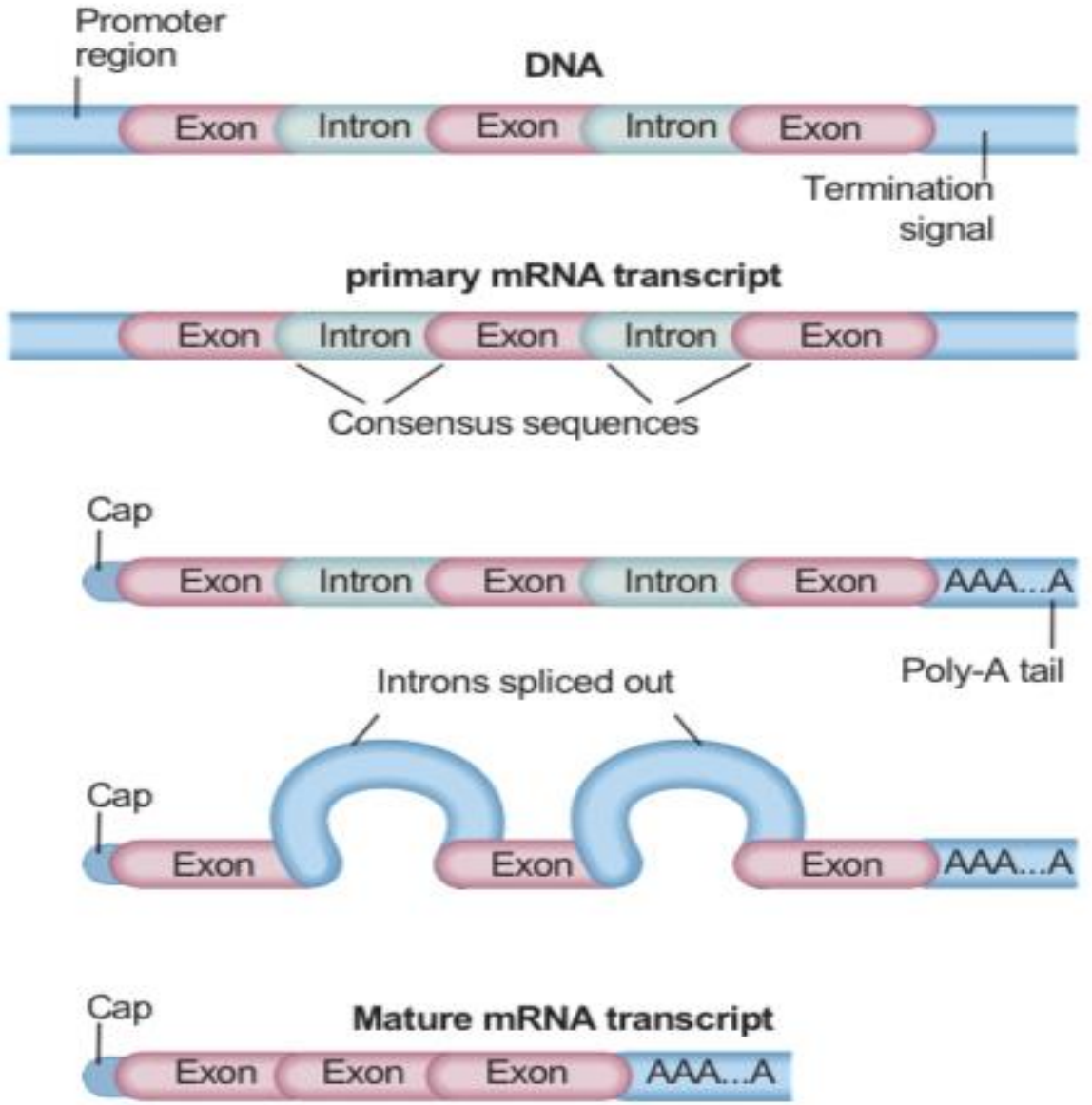


FIG 2-11 Gene splicing. Introns are precisely removed from the primary mRNA transcript to produce a mature mRNA transcript. Consensus sequences mark the sites at which splicing occurs.

Check this extra video, its useful & funny !

https://youtu.be/aVgwr0QpYNE?feature=shared

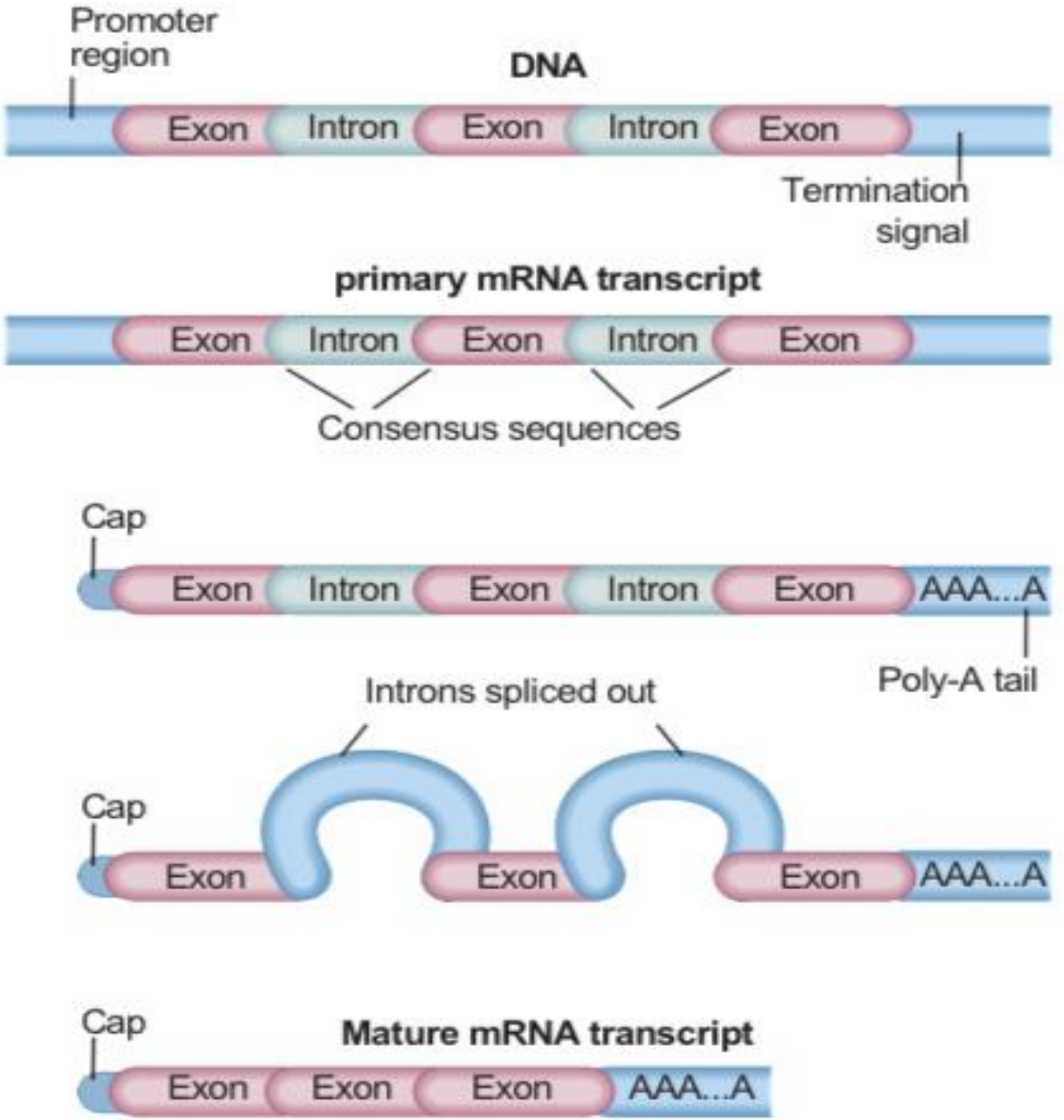


FIG 2-11 Gene splicing. Introns are precisely removed from the primary mRNA transcript to produce a mature mRNA transcript. Consensus sequences mark the sites at which splicing occurs.

The Genetic Code

The body contains **20** different types of amino acids that rise from 4 bases, and the amino acid sequences must in some way be designated by the DNA after transcription into mRNA. (Recently they have discovered a new amino acid so they become 21, its name selenocysteine)

If triplet sets of bases are translated into amino acids, however, 64 (4 X 4 X 4) combinations can be achieved- more than enough to specify each amino acid.This means many codons give the same amino acid

The correspondence between specific codons and amino acids is known as the **genetic code**.

- As you can see in the first line, the codon (UUU) gives Phe amino acid.
- You will see that there are many codons that will give Val, Ala ,etc. And this is normal since as we said before that we have 64 combinations that will give us 20 AA.

But if you noticed, there are only three codons that signal a STOP in translation: UAA, UAG, and UGA

- There is an amino acid that have only one codon for it, can you find it?🤓
- It is Methionine (its codon is AUG, which is the starting codon) ➡️ All of our proteins start with this AA and translation process starts at this codon (AUG). This codon guides the mRNA where to set in the ribosome for translation.
- **Note: Trp also has only one codon, but we are interested here in the one and only starting codon**

TABLE 2-2 The Genetic Code

FIRST POSITION (5′ END) ↓	SECOND POSITION U	SECOND POSITION C	SECOND POSITION A	SECOND POSITION G	THIRD POSITION (3′ END) ↓
U	Phe	Ser	Tyr	Cys	U
U	Phe	Ser	Tyr	Cys	C
U	Leu	Ser	STOP	STOP	A
U	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
C	Leu	Pro	His	Arg	C
C	Leu	Pro	Gln	Arg	A
C	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
A	Ile	Thr	Asn	Ser	C
A	Ile	Thr	Lys	Arg	A
A	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
G	Val	Ala	Asp	Gly	C
G	Val	Ala	Glu	Gly	A
G	Val	Ala	Glu	Gly	G

Ala, Alanine; *Arg*, arginine; *Asn*, asparagine; *Asp*, aspartic acid; *Cys*, cysteine; *Gln*, glutamine; *Glu*, glutamic acid; *Gly*, glycine; *His*, histidine; *Ile*, isoleucine; *Leu*, leucine; *Lys*, lysine; *Met*, methionine; *Phe*, phenylalanine; *Pro*, proline; *Ser*, serine; *Thr*, threonine; *Trp*, tryptophan; *Tyr*, tyrosine; *Val*, valine.

*Examples: UUG is translated into leucine; UAA is a stop codon; GGG is translated into glycine. Under some circumstances the UGA codon can specify an amino acid called selenocysteine, which is often called the 21st amino acid.

Of the 64 possible codons, 3 signal the end of a gene and are known as **stop codons**. These are UAA, UGA, and UAG.

The remaining 61 all specify amino acids. This means that most amino acids can be specified by more than one codon.

The genetic code is thus said to be "degenerate." While a given amino acid may be specified by more than one codon, each codon can designate only one amino acid.

So, one amino acid can be generated from more than one codon , BUT one codon is able only to generate one amino acid.

Translation

Translation is the process in which mRNA provides a template for the synthesis of a polypeptide.

mRNA cannot bind directly to amino acids. It interacts with transfer RNA (tRNA), a cloverleaf-shaped (single stranded) RNA strand of about 80 nucleotides.

The tRNA molecule has a site at its 3' end for the attachment of an amino acid by a covalent bond. At the opposite end of the cloverleaf is a sequence of three nucleotides called the anticodon (which recognize AA & its shape changes based on the AA shape) once this anticodon binds to mRNA, if there is complimentary base pairing it will give the AA and the polypeptide chain will be built.

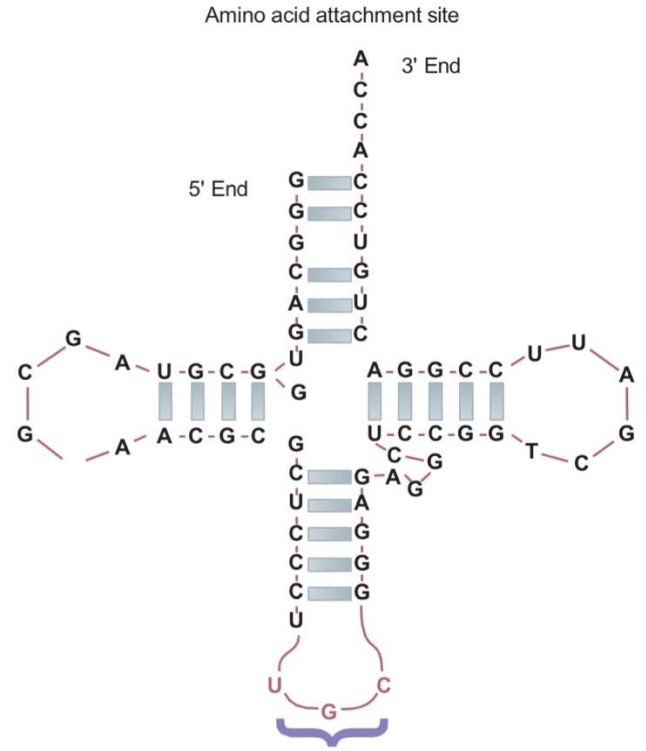


FIG 2-12 The structure of a tRNA molecule. In two dimensions, the tRNA has a cloverleaf shape. Note the 3′ site of attachment for an amino acid. The anticodon pairs with a complementary mRNA codon.

The cytoplasmic site of protein synthesis is the ribosome, which consists of roughly equal parts of enzymatic proteins and ribosomal RNA (rRNA). rRNA is synthesized in the nucleolus & there're small and large fragments of it (2 sections, and in between them, the mRNA will bind at the 5' end , this region is a docking place for tRNA). Each tRNA has anticodon & if there's complementary base pairing, it will give the AA it carries.

During translation, the ribosome first binds to an initiation site on the mRNA sequence, AUG, which specifies the amino acid methionine.

The ribosome then binds the tRNA to its surface so that base pairing can occur between tRNA and mRNA. The ribosome moves along the mRNA sequence, codon by codon, in the usual 5' to 3' direction.

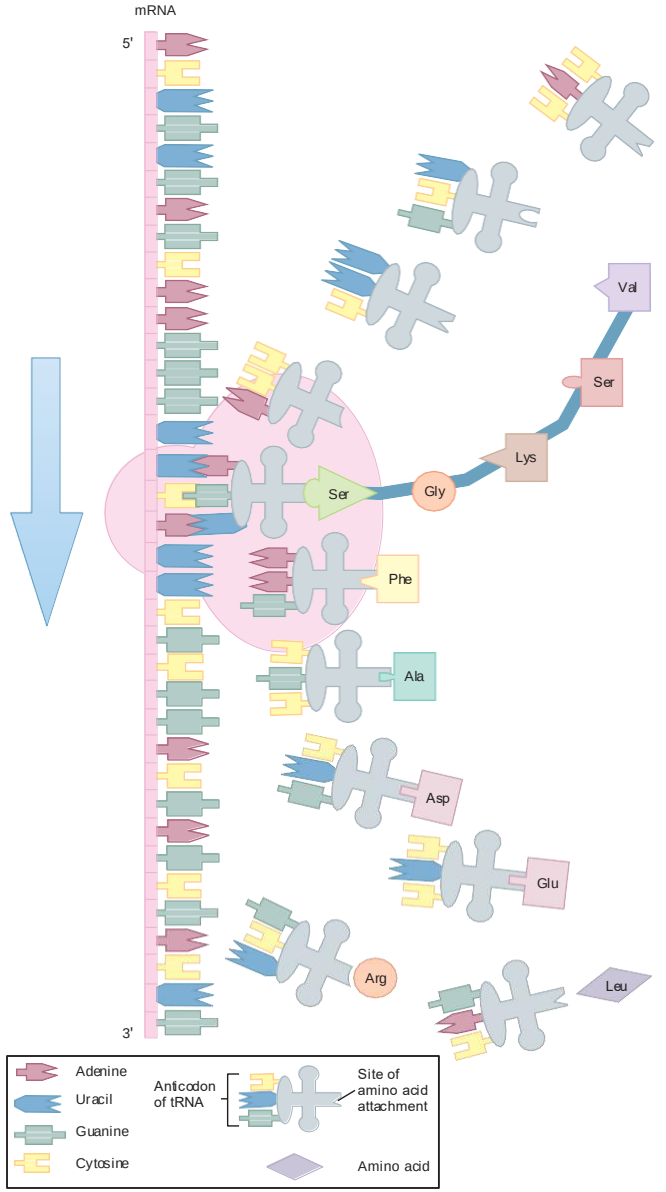


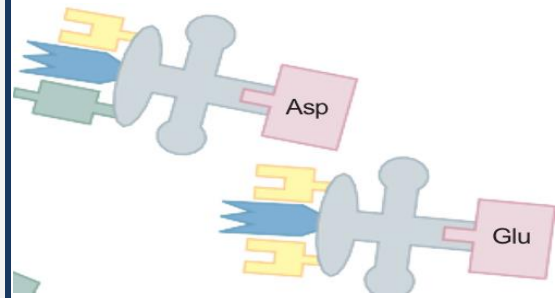
FIG 2-13 Translation of mRNA to amino acids. The ribosome moves along the mRNA strand in the 5′ to 3′ direction, assembling a growing polypeptide chain. In this example, the mRNA sequence GUG AGC AAG GGU UCA has assembled five amino acids (Val, Ser, Lys, Gly, and Ser, respectively) into a polypeptide.

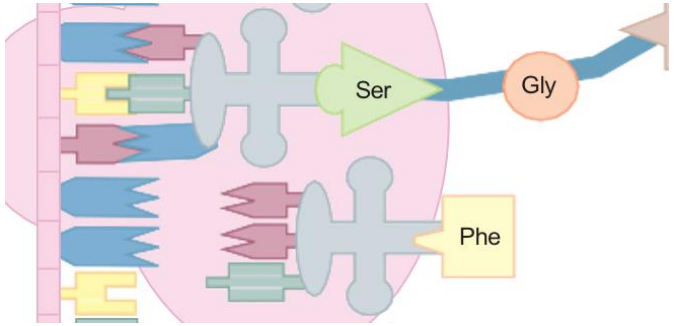
- Since it is a little confusing, I will tidy up things here.
- + you can check this nice video : https://youtu.be/bKIpDtJdK8Q?feature=shared

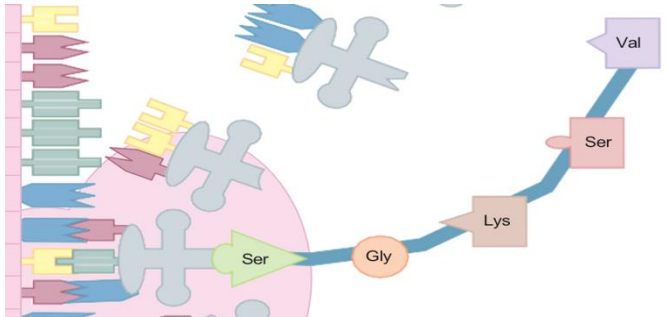
- 1-Here as you see the tRNA with its anti-codon & AA at the other side.

- 2-Here where it binds to the mRNA, if complementary base pairing is achieved (red matches blue ,green matches yellow) then tRNA will give the AA it carries (which is Ser here).

- 3- Here you can see the polypeptide chain (which results from repeating the previous steps with different AAs), it always starts with Met and ends up when there's a stop codon.







the ribosome provides an enzyme that catalyzes the formation of covalent peptide bonds between the adjacent amino acids, resulting in a growing polypeptide.(also tRNA & mRNA codons help in this process with the enzymes as we discussed earlier)

When the ribosome arrives at a stop codon on the mRNA sequence, translation and polypeptide formation cease.

The amino (NH_2) terminus of the polypeptide corresponds to the 5' end of the mRNA strand, and the carboxyl ($COOH$) terminus corresponds to the 3' end.

Before a newly synthesized polypeptide can begin *its* existence as a functional protein, it often undergoes further processing, termed posttranslational modification.like adding sugars or disulfide bonds, and this gives the protein its functionality.

If you remember from biochem, we have discussed the primary structure of the protein (linear AA sequence) and the secondary (coils, sheaths & helices of the AAs sequence) and the tertiary (folding of the helices & sheaths) & quaternary structure (more than 1 chain come together as in Hb).

These modifications can take a variety of forms, including cleavage into smaller polypeptide units or combination with other polypeptides to form a larger protein. Other possible modifications include the addition of carbohydrate side chains to the polypeptide.

These modifications are needed, for example, to produce proper folding of the mature protein or to stabilize its structure.

- As we've mentioned earlier, if there is any abnormality in any process, this would lead to mutations.
- So if **modification enzymes** aren't working (meaning that their genes have some sort of mutations), there will be no proper modification & folding of proteins.

THE STRUCTURE OF GENES AND THE GENOME

- **Introns and Exons**

The intron-exon structure of genes, discovered in 1977, is one attribute that distinguishes eukaryotes from prokaryotes.

Introns are present in eukaryotic genes, While prokaryotic genes lack introns.

Introns form the major portion of most eukaryotic genes. As noted previously, introns are spliced out of the mRNA before it leaves the nucleus.

Splicing is controlled by DNA sequences known as consensus sequences that are located adjacent to each exon.

The regions between exon and intron are conserved and well-known, and any mutation in this region can result in abnormal proteins.

- Additional:
- Consensus sequence: are very specific and commonly found nucleotide or animo acid sequences in a given conserved region
- -10 box, -35 box of *E. coli* promoter, protein binding sites, splice sites and restriction enzymes recognizing sites are several examples of consensus sequences.

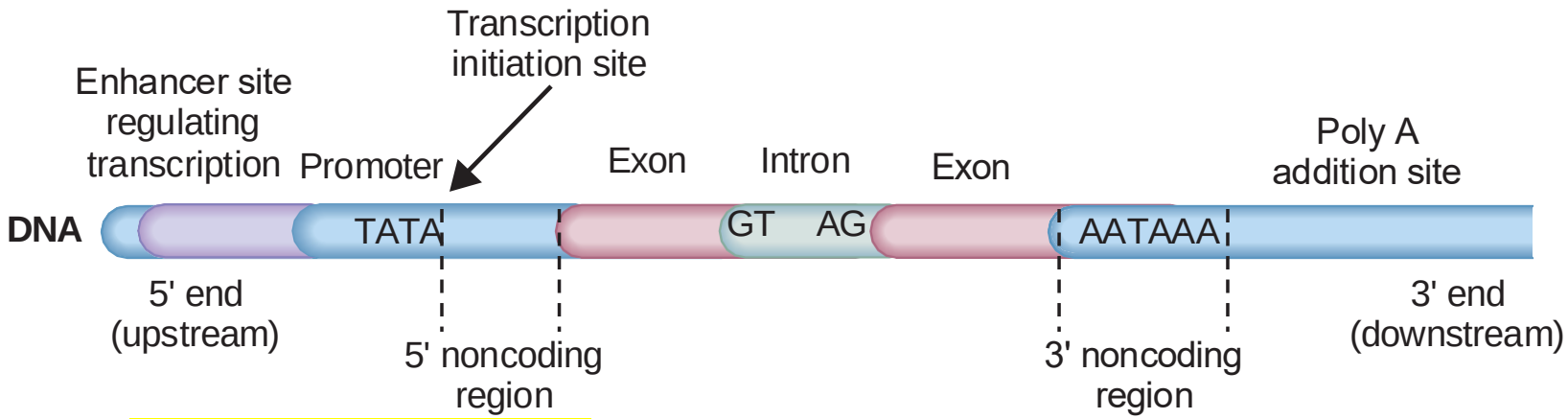


FIG 2-16 Details of gene structure, showing promoter and upstream regulation (enhancer) sequences and a poly-A addition site.

- The promoter is the region where transcription starts. RNA polymerase binds to it to transcribe the gene into mRNA.
- The enhancer site is a regulatory DNA sequence that, when bound by specific proteins called transcription factors, enhance the transcription of an associated gene.
- Between exons and introns, we have the sequences (GT) and (AG) which are constant in all genes. These regions are recognized by enzymes to remove introns, and any mutation in these regions can lead to abnormal proteins.
- The poly-A addition site directs the addition of a poly-A tail at 3' end stabilizes the mRNA, allowing it to move from the nucleus to the cytoplasm without destruction.

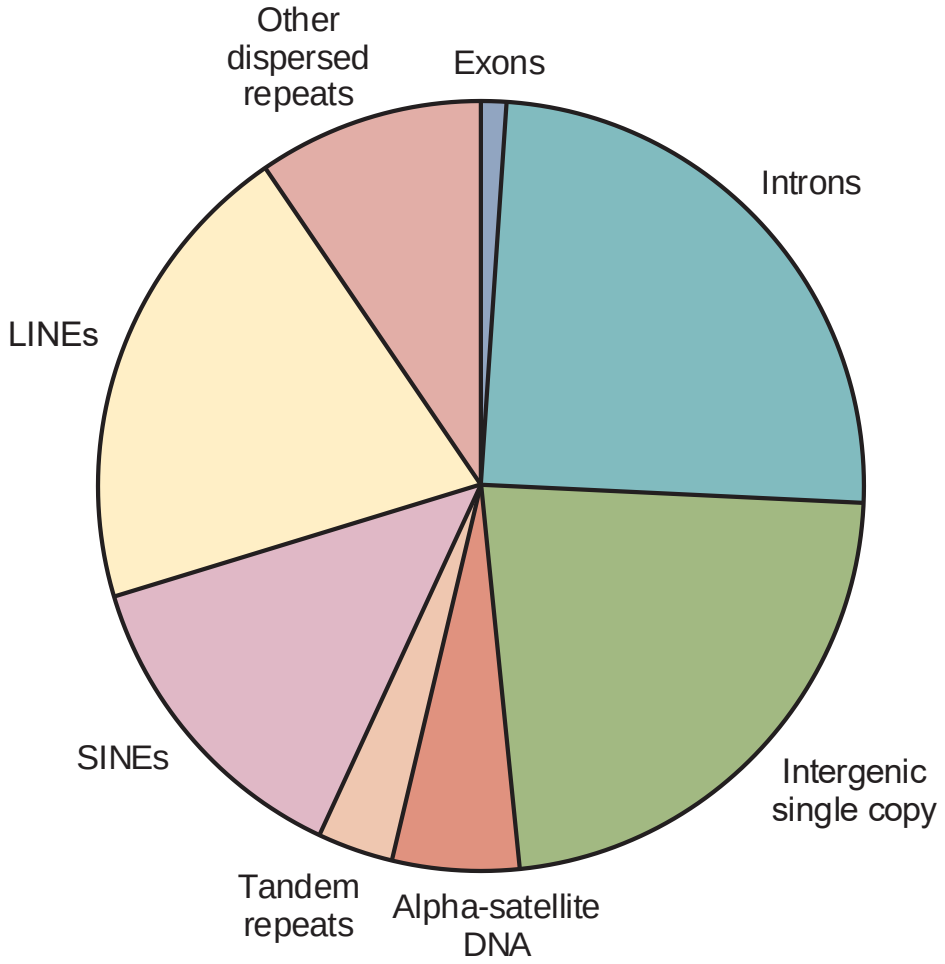


FIG 2-17 Structural composition of the human genome. Because of limitations in mapping repetitive sequences, these figures are approximate. In addition, there is some overlap among categories (e.g., repeat sequences are sometimes found in introns). The category "other dispersed repeats" includes DNA transposons, LTR (long terminal repeat) retrotransposons, and low-copy number duplications.

- **Exons** and **introns** make up 25% of the human genome.
- Exons are very small, comprising less than 1% of the genome.
- **Intergenic single-copy DNA** accounts for 25%, but its function is unknown.

- Repetitive DNA makes up 55% of the genome and includes:
  - **Long interspersed elements (LINEs).**
  - **Short interspersed elements (SINEs).**
  - **Alpha-satellite DNA.**
  - **Tandem repeats.**

- Repetitive DNA is thought to be important for:
  - Genome stability.
  - Maintaining proper spacing between genes to prevent coding sequence disruption during DNA rearrangement.
  - Structural stability of DNA.

However scientists have not yet identified an exact function for repetitive DNA.

Introns: by lengthening genes, encourage the shuffling of genes when homologous chromosomes exchange material during meiosis.

It has also been suggested that introns have evolved to modify the amount of time required for DNA replication and transcription.

Types of DNA

The most common, class of DNA is termed single-copy DNA-- are seen only once (or possibly a few times) in the genome.

Single-copy DNA composes about 45% of the genome and includes the protein-coding genes.

- Single-copy DNA includes:

1. Exons
2. Introns
3. intergenic single copy

Notice the difference: each type is detailed in the next slides

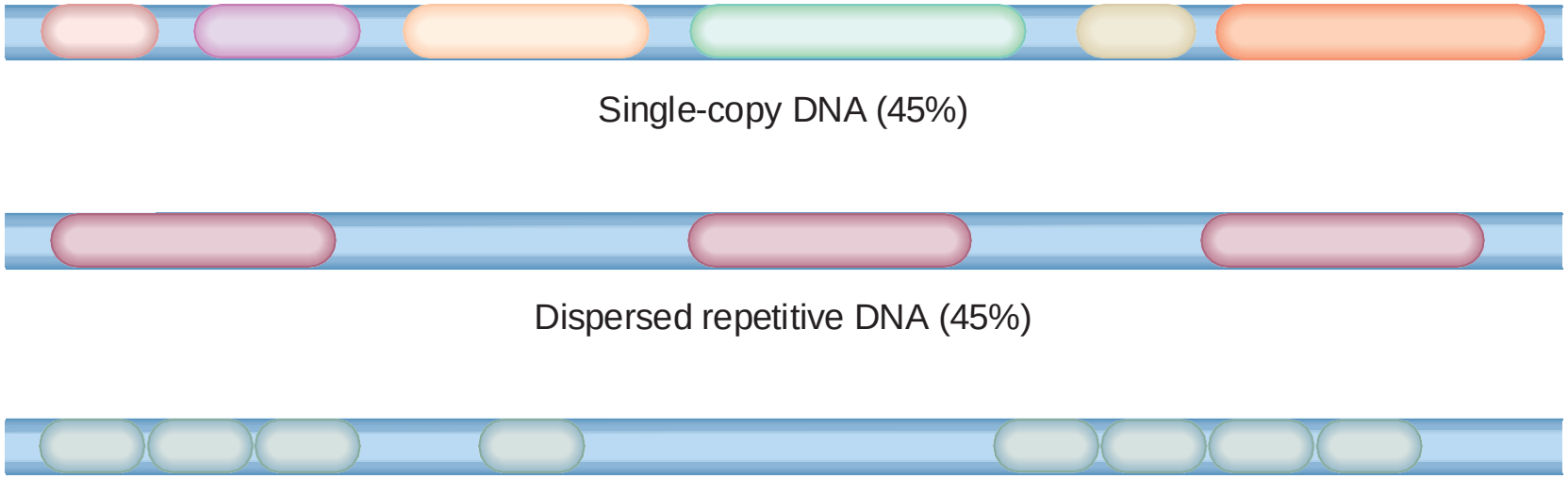


FIG 2-18 Single-copy DNA sequences are unique and are dispersed throughout the genome. Satellite DNA sequences are repetitive elements that occur together in clusters. Dispersed repeats are similar to one another but do not cluster together.

The remaining **55%** of the genome consists of **repetitive DNA**, sequences that are repeated over and over again in the genome, often thousands of times.

There are two major classes of repetitive DNA, **dispersed repetitive DNA** and **satellite DNA**. Satellite repeats are clustered together in certain chromosome locations, where they occur in tandem.

Dispersed repeats, as the name implies, tend to be scattered singly throughout the genome (they do not occur in tandem).

An example on dispersed repeats : a sequence of 20,000 base pairs is followed by a repeated sequence after 100,000 base pairs, and then another repetition occurs after 5,000,000 base pairs. This is the pattern of dispersed repeats.

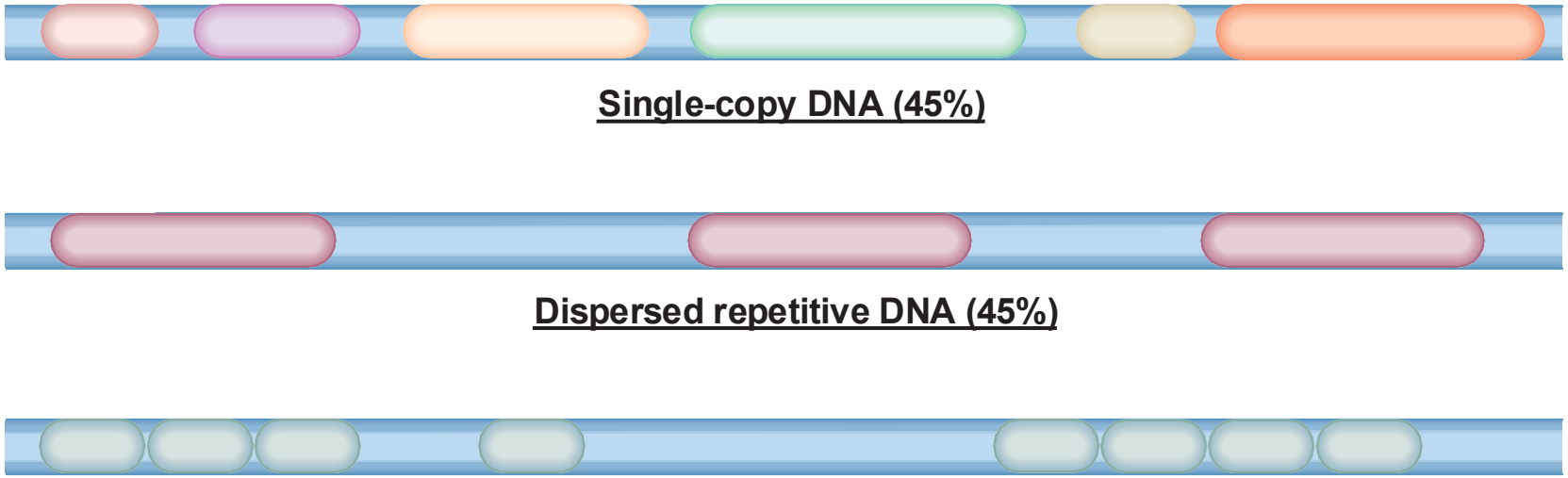


FIG 2-18 Single-copy DNA sequences are unique and are dispersed throughout the genome. Satellite DNA sequences are repetitive elements that occur together in clusters. Dispersed repeats are similar to one another but do not cluster together.

Satellite DNA composes approximately **10%** of the genome and can be further subdivided into several categories.

Alpha-satellite DNA occurs as tandem repeats of a 171-bp (base pairs) sequence that can extend several million bp or more in length. This type of satellite DNA is found near the centromeres of chromosomes.

Minisatellites are blocks of tandem repeats whose total length is much smaller. These repeats, which are 20 to 70 bp in length, usually have a total length of a few thousand base pairs or so.

Microsatellites, are smaller: the repeat units are usually only 2, 3, or 4 bp in length, and the total length of the array is usually less than a few hundred base pairs.

These repeats differentiate between individuals. For example, in one person, a specific repeat may occur 100 times, while in another, it may occur 102 times, and in another, 15 times, and so on. These sequences have high discriminatory power due to variations in the number of repeats among individuals.

Dispersed repetitive DNA makes up about 45% of the genome, and these repeats fall into two major categories, **SINEs** (short interspersed elements) and **LINEs** (long interspersed elements).

Individual SINEs range in size from 90 to 500 bp, while individual LINEs can be as large as 7,000 kb.

One of the most important types of SINEs is termed the "***Alu* repeats.**" The term *"Alu"* derives from the fact that these repeat units, which are about 300 bp in size, contain a DNA sequence that can be cut by the *Alu* restriction enzyme. Using this enzyme will separate *Alu* repeats.

The *Alu* repeats are a family of genes, meaning that all of them have highly similar DNA sequences. About 300,000 to 500,000 *Alu* repeats are scattered throughout the genome; these repeats thus constitute about 2% to 3% of all human DNA.

THE CELL CYCLE

During the course of development, each human progresses from a single-cell zygote (an egg cell fertilized by a sperm cell) to a marvelously complex organism containing approximately 100 trillion (10^{14}) individual cells.

The cell division processes responsible for the creation of new diploid cells from existing ones are termed mitosis (nuclear division) and cytokinesis (cytoplasmic division).

Before dividing, a cell must duplicate its contents, including its DNA; this occurs during interphase. The alternation of mitosis and interphase is referred to as the cell cycle.

Chromosomal abnormalities occur due to problems in spermatogenesis, oogenesis, and meiosis.

- Both the egg and sperm contain 23 chromosomes, which is half the total number of chromosomes in a human cell.
- The egg is produced during oogenesis, and the sperm is produced during spermatogenesis—both through the process of meiosis.
- When fertilization occurs, the zygote is formed, containing a complete set of 46 chromosomes.
- Mitosis plays a key role in growth by increasing the number of cells.
- If there is a reduction in red blood cells (RBCs) or iron (Fe), mitosis is induced in the bone marrow to produce more RBCs.
- Mitosis also occurs when tissues are damaged or in cells that continuously divide, such as skin cells, which are constantly being replaced.
- Mitosis involves nuclear and cytoplasmic division, producing new cells with the same number of chromosomes as the original cell.
- In contrast, meiosis is the process that forms gametes (sperm and egg), reducing the chromosome number by half to ensure proper genetic balance in offspring.

I think we all know these simple informations, so just read them

A typical cell spends most of its life in interphase. This portion of the cell cycle is divided into three phases, G1, S, and G2.

During G1 ("gap 1," the interval between mitosis and the onset of DNA replication), synthesis of RNA and proteins takes place. DNA replication occurs during the S (synthesis) phase.

During G2 (the interval between the S phase and the next mitosis), some DNA repair takes place, and the cell prepares for mitosis. By the time G2 has been reached, the cell contains 2 identical copies of each of the 46 chromosomes.

These identical chromosomes are referred to as sister chromatids. Sister chromatids often exchange material during or after the S phase, a process known as sister chromatid exchange.

- Watch this video to know difference between Homologous chromosomes vs sister chromatides.

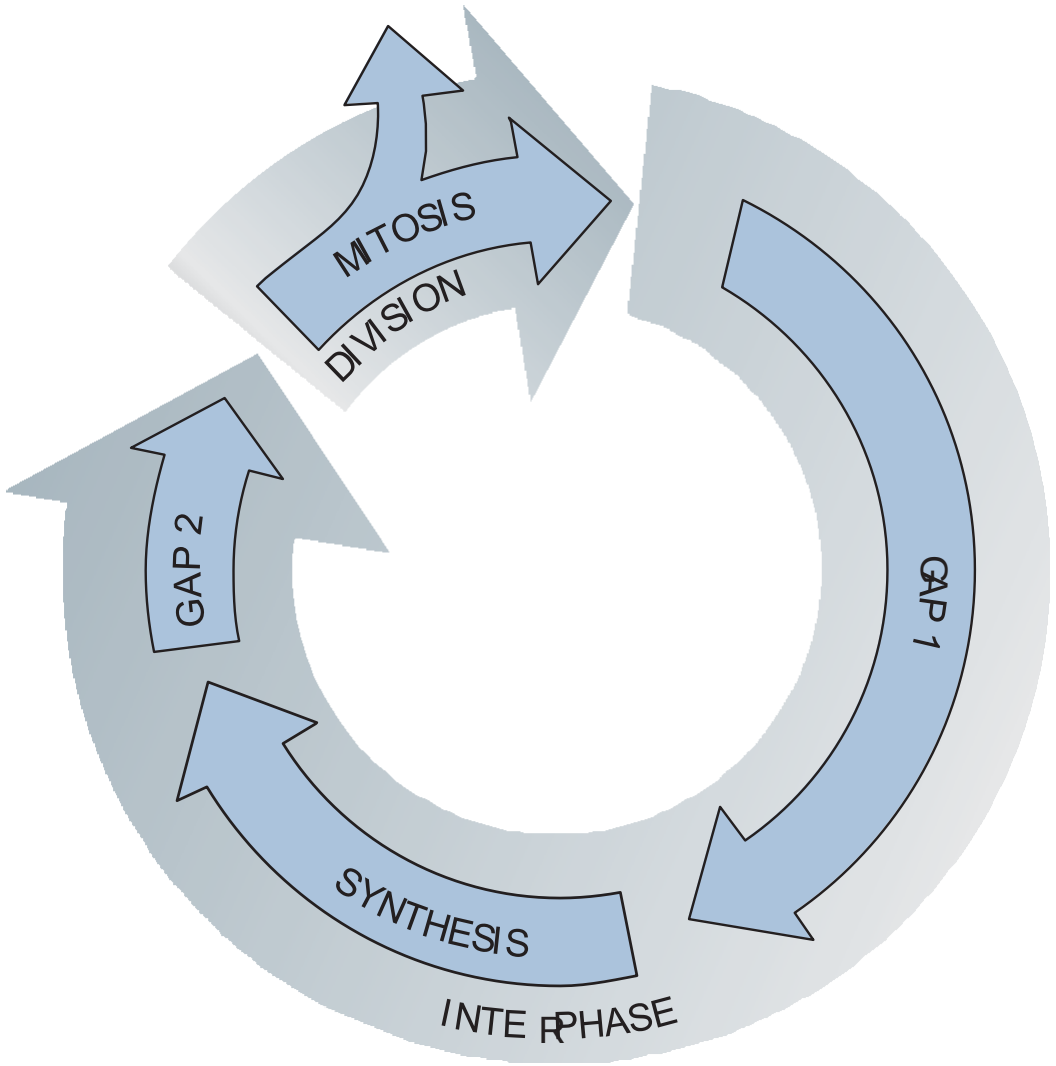


FIG 2-19 Major phases of the mitotic cell cycle, showing the alternation of interphase and mitosis (division).

The length of the cell cycle varies considerably from one cell type to another.

- Rapidly dividing cells:
  - Found in epithelial tissues (skin, intestine, lungs).
  - Complete the cell cycle in a short time (e.g., ~10 hours).
- Poorly dividing or non-dividing cells:
  - Includes liver cells, skeletal muscle cells, and neurons.
  - Divide very slowly or lose the ability to divide entirely.

- Remember (additional):
- in a healthy adult liver, the cells are dormant and rarely undergo cell division. However, if the liver is damaged, the liver cells re-enter the cell cycle to divide and produce more of themselves.

Some cell types, such as skeletal muscle cells and neurons, completely lose their ability to divide and replicate in adults.

The great majority of this variation is due to differences in the length of the G1 phase. When cells stop dividing for a long period, they are often said to be in the G0 stage.
(like skeletal muscle cells and neural cells)

Before entering mitosis, DNA replication has to be accurate and complete and the cell has to have achieved an appropriate size.

The cell must respond to extracellular stimuli that require increased or decreased rates of division.

- Extracellular stimuli:

- Can include growth factors, transcription factors, and hormones.

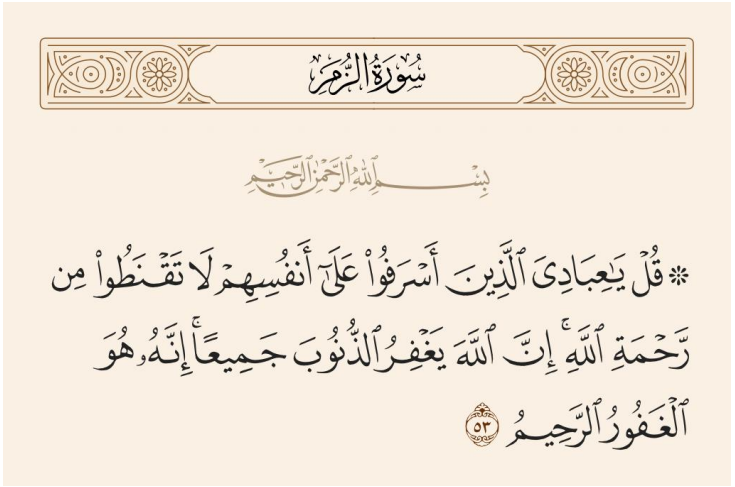
Example: Iron deficiency leads to a lack of hemoglobin and red blood cells (RBCs). In response, the kidneys release erythropoietin, which stimulates the bone marrow to produce more RBCs.

Complex molecular interactions mediate this regulation. Among the most important of these molecules are cyclin-dependent kinases (CDKs), a family of kinases that phosphorylate other regulatory proteins at key stages of the cell cycle.

Cyclin-dependent kinases regulate the cell cycle, so mutations in their genes or associated proteins can lead to cancer.

Additional sources

1. Book pages
2. [Ninja nerd translation : protein synthesis]
3. Webpages…etc



VERSIONS	SLIDE #	BEFORE CORRECTION	AFTER CORRECTION
V1→ V2			
V2→V3			

امسح الرمز و شاركنا بأفكارك لتحسين أدائنا !!