# Bioinformatics I introductions to genetic martials

**Dr Manaf A Guma**
**University of Anbar- college of applied sciences-Hit**
**Department of chemistry**

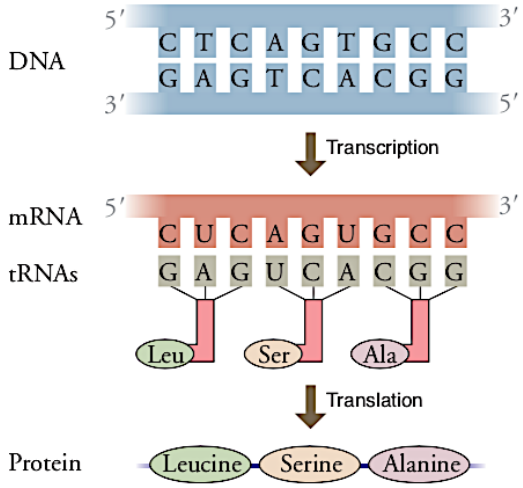
1

## Protein synthesis: Genes Encode Proteins !

- DNA is essential for inheriting the genetic information.
- The complete set of genetic information of an organism is called its **genome**
- *A part of the DNA, a **gene,** is **transcribed** to produce a complementary strand of RNA; then the RNA is translated into protein.*


2

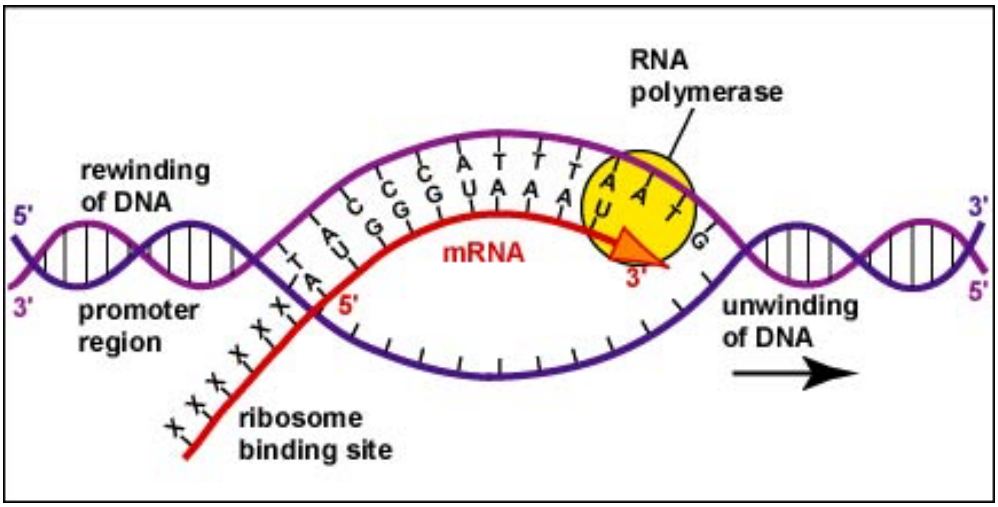
**The protein's amino acid sequence therefore depends on the nucleotide sequence of the DNA.**




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

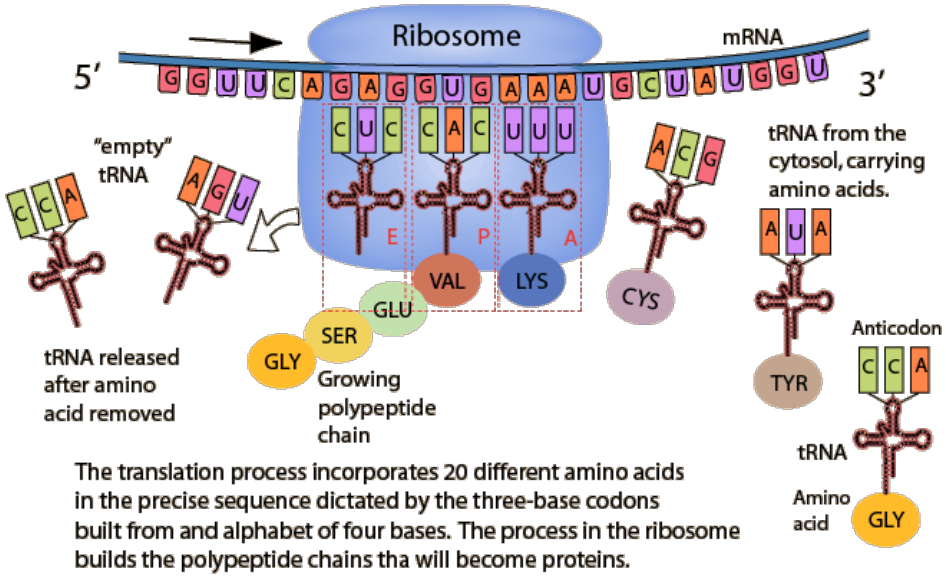
3

# From DNA to RNA: Transcription (making a copy)




4

# The translation to a protein



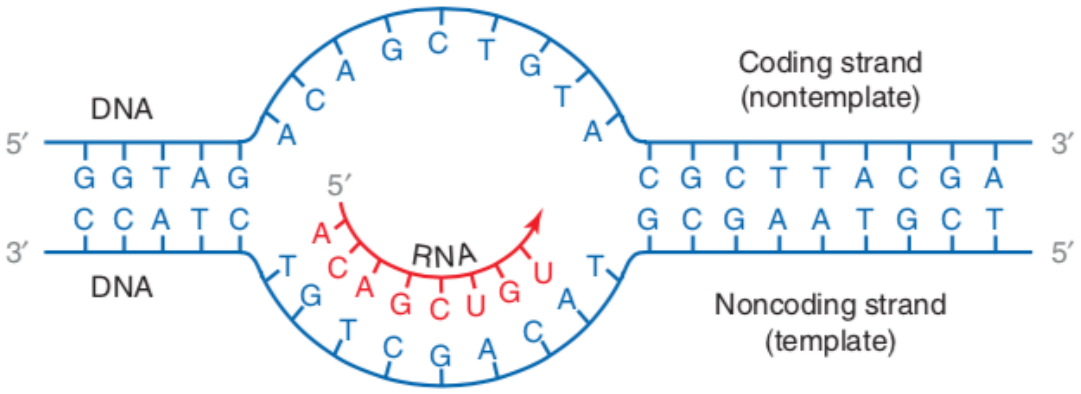

5

# So, How is protein synthesized:

- Gene expression begins with the process called **transcription**, which is the synthesis of a strand of mRNA that is complementary to the gene of interest.
- A region of DNA un-winds and then the two strands separate.
- However, only that small part of the DNA will be split apart.
- The triplets within the gene on this section of the DNA molecule are used as the template to transcribe the complementary strand of RNA.

6

# Transcription of RNA from the DNA



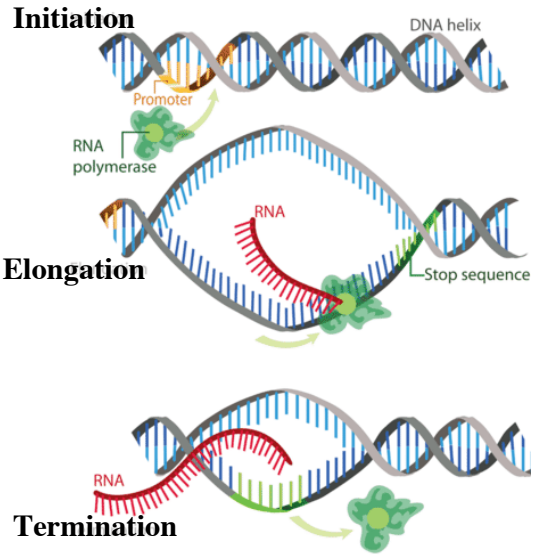

7

# Three stages of transcription

a. *Stage 1:* ***Initiation:*** **by** a **promoter.**

b. *Stage 2 :* ***Elongation:*** **RNA polymerase** is an enzyme that adds new nucleotides to a growing strand of RNA.

c. *Stage 3:* ***Termination****:* (UAA, UAG, or UGA) codes a "stop codons".




8

- **There are a total of 64 codons: 3 of these are "stop" signals that terminate translation, and the remaining 61 represent, with some redundancy, the 20 standard amino acids found in proteins**

| First Position (5′ end) | Second Position | | | | Third Position (3′end) |
|---|---|---|---|---|---|
| | U | C | A | G | |
| U | UUU Phe | UCU Ser | UAU Tyr | UGU Cys | U |
| | UUC Phe | UCC Ser | UAC Tyr | UGC Cys | C |
| | UUA Leu | UCA Ser | UAA Stop | UGA Stop | A |
| | UUG Leu | UCG Ser | UAG Stop | UGG Trp | G |
| C | CUU Leu | CCU Pro | CAU His | CGU Arg | U |
| | CUC Leu | CCC Pro | CAC His | CGC Arg | C |
| | CUA Leu | CCA Pro | CAA Gln | CGA Arg | A |
| | CUG Leu | CCG Pro | CAG Gln | CGG Arg | G |
| A | AUU Ile | ACU Thr | AAU Asn | AGU Ser | U |
| | AUC Ile | ACC Thr | AAC Asn | AGC Ser | C |
| | AUA Ile | ACA Thr | AAA Lys | AGA Arg | A |
| | AUG Met | ACG Thr | AAG Lys | AGG Arg | G |
| G | GUU Val | GCU Ala | GAU Asp | GGU Gly | U |
| | GUC Val | GCC Ala | GAC Asp | GGC Gly | C |
| | GUA Val | GCA Ala | GAA Glu | GGA Gly | A |
| | GUG Val | GCG Ala | GAG Glu | GGG Gly | G |

- **U=Uracil which is a nucleotide found in the RNA only.**
- **We read it in the DNA seq. as T=Thymine.**

A **codon** is a three-base sequence of mRNA, so-called because they directly encode amino acids.

9

# Translation DNA to a protein ?! Is it possible?

- You can also find the DNA sequence for a specific protein using the following web site:
- https://www.ncbi.nlm.nih.gov/gene
- This is the opposite to the translation of DNA to proteins ?why?
- What are the possibilities for converting a protein to DNA based on the codon chart.
- Is it possible? What do you know about codon optimization.

10

# Translation DNA to protein using Expasy Translate website

- We can use the website: https://web.expasy.org/translate/
- You will have 6 translation for your query DNA sequence which are based on the frameshift.
- Be aware of the translation:
- You could have a problem while you copy and paste the sequence.
- You could the wrong frameshift. ACTGCAGTGCAA ….

11

# Tutorial

- Translate the following DNA sequence to protein: how many possibilities you can get?
- `CGAGCATGGACGCGATCAAGAAGAAGATGCAAATGCTGAAACTGGAC AAAGAAAATGCGCTGGACCGTGCGGAACAGGCGGAGGCGGACAAGAA AGCGGCGGAGGATCGTAGCAAGCAGCTGGAAGACGAGCTGGTGAGCC`
- Can you do the opposite? Protein to DNA ?

12

# Expasy tool param database to find the chemical properties for a peptide of a protein

- An engine can calculate the following parameters for the protein.
- https://web.expasy.org/cgi-bin/protparam/protparam.
- **You can calculate the** the chemical properties for the following seq:
  MDDIYKAAVEQLTEEQKNEFKAAFDIFVLGAEDGCISTKELGKVMRMLGQNPTPEELQEM
  IDEVDEDGSGTVDFDEFLVMMVRCMKDDSKGKSEEELSDLFRMFDKNADGYIDLDELKIM
  LQATGETITEDDIEELMKDGDKNNDGRIDYDEFLEFMKGVE

1. **The umber of amino acids: Amino acid composition and the negative and the positives amino acids.**
2. **Molecular weight:**
3. **Theoretical pI ?**
4. **Formula. Like** $C_{46}H_{61}N_{11}O_{18}$
5. **Total number of atoms.**
6. **Extinction coefficient ? What is it ? Next page …..**

13

# How to calculate Molecular weight for a peptide or a protein?

- The **extinction coefficient** is the absorbance divided by the concentration and the pathlength, according to Beer's Law
- = (epsilon = absorbance/concentration/pathlength).
- The units of **extinction coefficients** are usually $M^{-1}cm^{-1}$, but for **proteins** it is often more convenient to use $(mg/ml)^{-1}cm^{-1}$.
- Molar Extinction Coefficient = (Number of Tryptophan residues X 5500) + (Number of Tyrosine residues X 1490)= gm/l = A0.1%mg/ml
- And then divide the check?

14