BIOL 451
GENE EXPRESSION, MODIFICATION AND GENETIC ENGINEERING

J. A. LARBI

1

COURSE OUTLINE

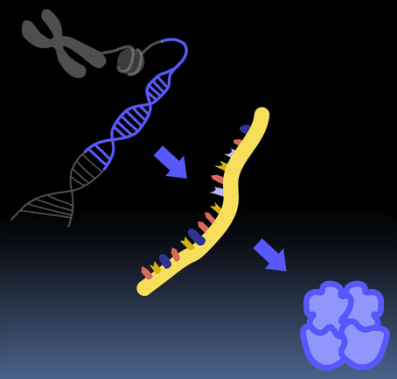
- Bacteria Gene Expression – Transcription-promoter, and structure and consensus sequences.
- Region of the lac and trp operons (SELT STUDY)
- Eukaryotic Gene Expression – Eukaryotic RNA polymerases;
  - Differences between eukaryotic and prokaryotic gene expression;
  - Promoters and enhancers;
  - Transcription factors

2

- Post Transcriptional Events
- Translation and its control
- Restriction Endonucleases and Recombinant DNA
- Complex Vectors for Cloning and Expressions – YACS, Cosmids, λ-phagemids
- Cloning Strategies – Genomic cDNA and Cloning by PCR
- Methods of Site Directed Mutagenesis
- Gene Transfer and Expression in Eukaryotes – Transfer methods in yeasts, animal and plant cell;
  - Electroporation and microprojector bombardment
- Gene Therapy in Man

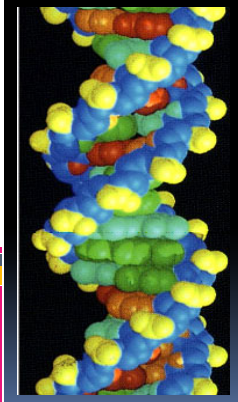
3

GENE EXPRESSION



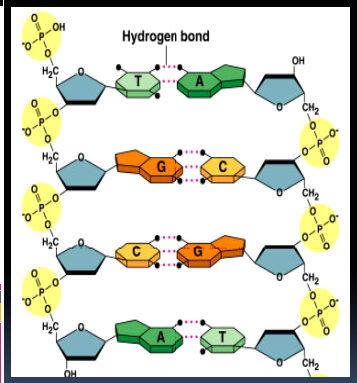
4

DNA



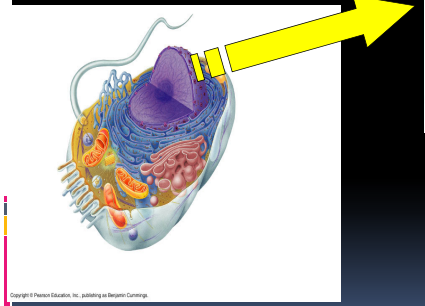
- Deoxyribonucleic Acid
- Double helix
- Carries genetic information
- Located in the nucleus
- The monomer is a nucleotide
  - A phosphate
  - A ribose sugar
  - A nitrogenous base

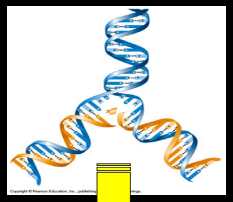
What are the bases in DNA

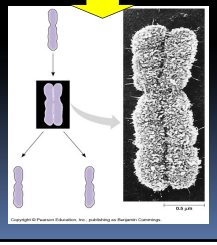


- A – adenine
- T – thymine
- C – cytosine
- G – guanine
- Base pair rules

Where is DNA located in the Cell?

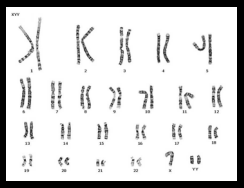






Where are the genes located?

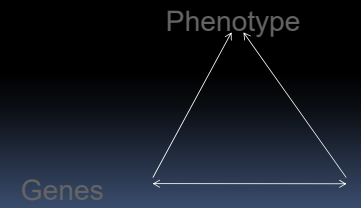
- Genes are located on the chromosomes.
- Every species has a different number of chromosomes.
- There are two types of chromosomes: autosomes and sex chromosomes



- Genes are located on the chromosomes which are found in the nucleus of a cell.
- When a cell is undergoing cell reproduction, the chromosomes are visible.
- Chromosomes appear when the chromatin condenses and become visible.
- Most of the time (90%) the genetic material in the form of chromatin.
- A genome is the complete genetic information contained in an individual.
  - (gene + chromosome)

Modern
Molecular Biology combines:

Biochemistry
Genetics
Cell Biology



10

"Central Dogma"

DNA → RNA → (processing) → protein

Most aspects of Molecular Biology will fit on this simple diagram

11

Current focus:

- **Genome** consist of the entire hereditary information of an organism.
  - either in DNA or RNA.
  - includes both the genes and the non-coding DNA
- **Genomics** (i.e. the study of the structure and function of whole genomes) and **Structural Biology**

12

Main goal: to determine the functions of all genes/proteins.
Approaches:

1. **Structural Genomics (the study of the sequences of genomes)**: sequence and compare as many genomes as possible (DNA sequence data allows one to predict sequence and probable function of protein products).

2. **Functional Genomics -** the study of the functions of genetic information contained within genomes.

3. Proteomics (characterization of all proteins encoded by a genome) is a key element of structural biology: includes determination of their structure, functional domains, interactions and expression levels;

13

GENE EXPRESSION, MODIFICATION & GENETIC ENGINEERING

- **GENE**: A sequence of DNA nucleotides on a chromosome that encodes a protein, tRNA or rRNA molecule or regulates the transcription of such a sequence.
- In summary, it is the study that recognizes how the genetic material (stored in the form of DNA) is transferred to its final product.(Protein or Polypeptide)
- It also involves how these genes can be regulated and manipulated for a desired product.

14

OVERVIEW OF THE GENE CONCEPT

- Genes known through Mendel's work as hereditary factors were associated with specific characters.
- Genes are known to control hereditary traits.
- Its physical nature was puzzling until known to control hereditary.
- Experiments with labeled bacteriophage showed that DNA is the hereditary material.

15

KNOWLEDGE OF DNA

- ***One-Gene-One Enzyme Theory*** postulated that genes control the structure of protein.
- Gene were known to be carried on chromosomes, and the chromosomes made up DNA.
- Grifiths and co. postulate to DNA as the genetic material.
- Genes are connected with specific traits/characters.

16

KNOWLEDGE OF DNA...*contd.*

- The net outcome is that a protein is made up with one of the two basic functions:
  a) Proteins may be structurally contributing to the *physical properties* of the cell organizations (muscle and hair protein)
  b) May be an enzyme-functional protein that catalyzes one of the chemical reactions of the cells by coding two important factors-*Biological structures* and *Biological functions*.
- Genes don't work in isolation because their action is in many ways affected by the environment.

17

DNA AS PART OF GENE EXPRESSION

- The genetic information carried by a gene lies in ***just one*** of the two helixes (strands).
- Thus polynucleotide acts as template for the synthesis of a complementary molecule called ***RNA***
- This is the first stage of gene expression and is called ***TRANSCRIPTION***.
- The amount of information on a gene is unlimited. E.g. 150bp length have anyone of 4^{150} different nucleotide sequences.

18

- Biologically not all these genes would be used meaningfully due to the rules that limit the number of sequence that make sense.
- The number of genes in different organism vary. (e.g. a bacteria cell may have ~4milllion DNA bp or 4000Kbp).
- Thus if a single gene is about 1.1Kbp in length, then a bacteria may have ~3500 genes if all DNA code for protein.
- But not all DNA code for proteins, hence there may be 3000 or less coding genes, of which **ONLY** about 1000-2000 may be expressed at any particular time.

19

- This implies that genes are discrete segments of DNA molecules and may be separated from each other by ***INTERGENIC DNA***.
- They are arranged in different ways in different types of animals.
- E.g.: Viruses have closely packed genes very little intergenic DNA between them.
- However, in other organisms, they are spread out and have long intergenic spaces.
- In higher organisms, they are spread out and separated by very long intergenic regions.

20

- In humans, genes make up 30% of the total DNA in the cell.
- Majority are spaced out more or less randomly, but in some cases they are grouped into distinct clusters.
- Sometimes, the individual genes in a cluster are unrelated but more frequently, the clusters are made up of genes that contain related units of biological information.
- E.g. are the OPERON and MULTIGENE FAMILY.
- Investigators in 1977, found that the biological information carried by genes are split into distinct units separated by intervening segments of DNA.
- Segments containing biological information were called ***EXONS*** whilst the intervening segments were called ***INTRONS***.
- These types of genes are referred to as ***DISCONTINUOUS*** genes, or ***SPLIT*** genes or ***MOZAIC*** genes. introns are common in higher organisms, many viruses and bacteria.

21

- ***ASSIGNMENT:- What is the use of introns?***
- Introns are common in higher organisms.
- In higher organisms, a gene may contain no intron or may have as many as a hundred (100).
- Often introns are much longer than exons, e.g. the human gene for cystic fibrosis transmembrane regulator genes with 24 exons and 23 introns.

22

- Causes cystic fibrosis when it does not function correctly.
- It is 250Kbp in length and split into 24exons and 23introns.
- The exons are scattered throughout the length of the gene and separated by introns ranging in size from 2-35bp.
- Average length of an exon is 277bp. Exons form only 2.4% of the entire gene.

23

TRANSFER OF BIOLOGICAL INFORMATION:

- A number of experiments show that in the expression of a gene, RNA as an intermediate between DNA and protein.

  (Read On Pulse-Chase Experiment or Technique).
- Other experiments have also shown that for each gene, RNA is transcribed for only one of the DNA strands (the template).
- Transcription and Translation are the two main stages in gene expression.

24

- Nevertheless, genetic information in the DNA can either be copied into more DNA during replication or be translated into protein.

25

TRANSFER OF BIOLOGICAL INFORMATION: *contd.*

- There are three (3) stages of information transfer.
  i. Replication
  ii. Transcription
  iii. Translation

- ***GENE EXPRESSION*** is the process through which biological information in a gene is made available to the cell.

26

Two steps are required

1. **Transcription**
   The synthesis of mRNA uses the gene on the DNA molecule as a template.
   It involves the reading of the DNA and changing the code to mRNA.

This happens in the nucleus of eukaryotes

- **Translation**
  The synthesis of a polypeptide chain using the genetic code on the mRNA molecule as its guide. It is changing the mRNA into a trait by using tRNA to interpret the mRNA.

27

TRANSLATION AND TRANSCRIPTION.

- DNA
  ↓ (Replication)
- DNA
  ↓ (Transcription)
- RNA
  ↓ (Translation)
- PROTEIN

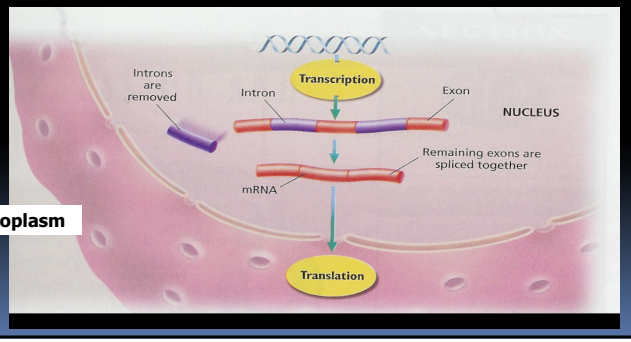
28

TRANSLATION AND TRANSCRIPTION...*contd.*

- F. Crick proposed that the biological information in DNA of a gene is first transferred to RNA and then to protein.
- Furthermore, the information flow was unidirectional, and that, proteins cannot themselves direct the synthesis of DNA and RNA.
- These ideas constituted the Central Dogma of Genetics.
- Howard Temin and David Baltimore independently disapproved the second part of Crick's proposition.
- This is because certain viruses undergo ***REVERSE TRANSCRIPTION (Reverse transcriptase)***.

29

- DNA in eukaryotes has regions of coding and noncoding DNA. The regions of DNA that code for proteins or traits are called **EXONS**, while the regions that do not code for proteins are called **INTRONS.**

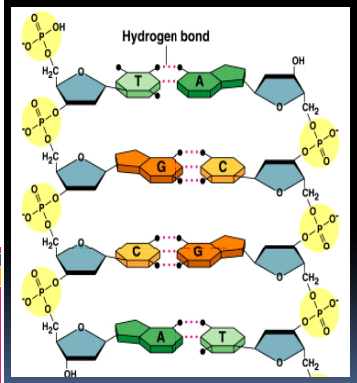


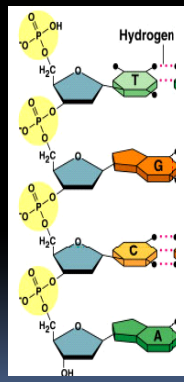
30

TRANSLATION AND TRANSCRIPTION...*contd.*

- **DNA**
  - (Replication)
- **DNA**
  - (Transcription/Reverse Transcription)
- **RNA**
  - (Translation)
- **PROTEIN**

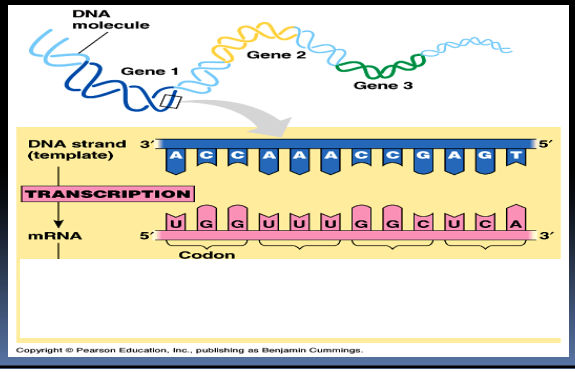
31





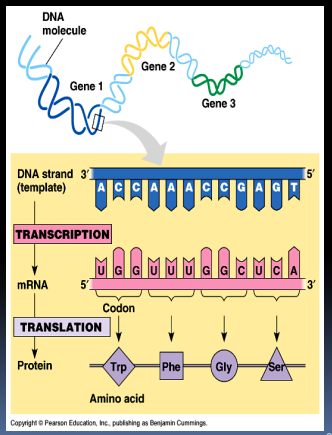
32

- In Eukaryotes, following mitosis or meiosis, DNA recoils but certain regions remain relaxed for transcription. The areas of relaxed DNA are called **euchromatin**.



33

- **RNA**
  - **Single stranded**
  - **Does not contain thymine but has uracil instead.**
- **tRNA carries 3 base pair code for specific amino acid.**
- **Amino acids compose polypeptid chains.**
- **One or more polypeptide chains compose a protein**
- **proteins provide the "blueprints" for our characteristics and functions.**



34

Gene expression takes place differently in prokaryotes and eukaryotes.
What is a prokaryote/Eukaryote?

- Prokaryotes
  - No membrane bound organelles (nucleus)
  - More primitive organisms
  - Only one circular chromosome
  - Bacteria.

- Eukaryotes
  - Membrane bound organelles (specialize in function –nucleus, mitochondria, chloroplast)
  - Chromosomes are in pairs and not circular
  - Protist, fungi, plants and animals

- In prokayotes, transcription and translation occur in the cytoplasm.
- In eukaryotes, transcription occurs inside the nucleus in a two step sequence of events.
  - Pre-mRNA includes both introns and exons for the gene.
  - mRNA is only the coding portion (exons).
- Translation occurs in the cytoplasm at the ribosomes.

36

RIBONUCLEIC ACID (RNA)

Found all over the cell

(nucleus, mitochondria, chloroplasts, ribosomes and the soluble part of the cytoplasm).

37

Types

- Messenger RNA (mRNA) <5%
- Ribosomal RNA (rRNA) Up to 80%
- Transfer RNA (tRNA) About 15%
- In eukaryotes small nuclear ribonucleoproteins (snRNP).

38

Structural characteristics of RNA molecules

- **Single polynucleotide strand** which may be looped or coiled (not a double helix)
- Sugar **Ribose** (not deoxyribose)
- Bases used: Adenine, Guanine, Cytosine and **Uracil** (not Thymine).

39

mRNA

- A long molecule 1 million Daltons
- Ephemeral
- Difficult to isolate
- **mRNA provides the plan for the polypeptide chain**

40

rRNA

- Coiled
- Two subunits:
  a long molecule 1 million Daltons
  a short molecule 42 000 Daltons
- Fairly stable
- Found in ribosomes
- Made as subunits in the nucleolus
- **rRNA provides the platform for protein synthesis**

41

tRNA

- Short molecule about 25 000 Daltons
- Soluble
- At least 61 different forms each has a specific anticodon as part of its structure.
- **tRNA "translates" the message on the mRNA into a polypeptide chain**

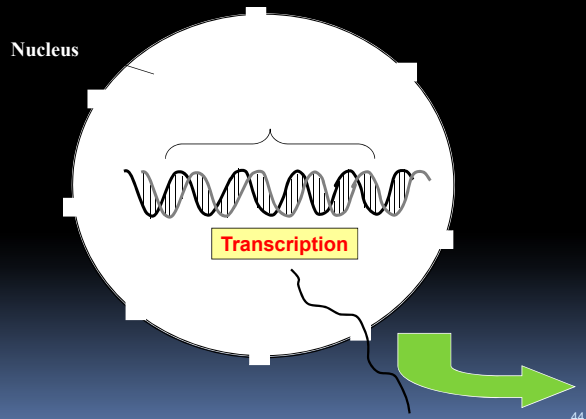
42

The Problem

- Information must be transcribed from DNA for further functions
- REMEMBER:

DNA→RNA→Protein

Transcription plan



44

Transcription: The synthesis of a strand of mRNA (and other RNAs)

- Uses an enzyme **RNA polymerase**
- Proceeds in the same direction as replication (**5′ to 3′**)
- Forms a **complementary** strand of mRNA
- It begins at a **promotor site** which signals the beginning of gene is not much further down the molecule (about 20 to 30 nucleotides)
- After the end of the gene is reached there is a **terminator sequence** that tells RNA polymerase to stop transcribing

NB Terminator sequence ≠ terminator codon.

45

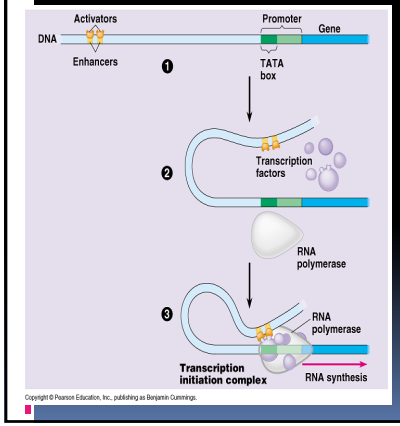
- In Prokaryotes there are three (3) regulatory elements that control gene expression.
  1. Structural genes – genes that code for a specific polypeptide (protein).
  2. Promoter – DNA segment that recognizes RNA polymerase.
  3. Operator – element that serves as a binding site for an inhibitor protein that blocks transcription.

Enhancers are short regulatory elements of accessible DNA that help establish the transcriptional program of cells by increasing transcription of target genes. They are bound by transcription factors, co-regulators

46

THURS. 9th Feb 2023

Enhancer Control



- **Eukaryote genes on a DNA strand also have noncoding control sequences that facilitate transcription.**
- **These are called enhancers.**
- **Transcription factors are additional proteins that bind to RNA polymerase and enhancers to help with transcription.**

48

QUIZ I

1. What is the difference between chromosomes, chromatin, chromatid, DNA and genes?
2. Non-coding DNAs can be put into two groups, name them.
3. What is gene manipulation?
4. What do you understand by the term transformation?
5. State three importance of introns.

49

TRANSCRIPTION

- **TRANSCRIPTION:**
- Synthesis of complementary strand of RNA from a specific gene (DNA template)
- It requires RNA polymerase and the process can be viewed as taking place in three (3) phase.
  - -Initiation (recognition and binding)
  - -Elongation
  - -Termination (and release)
- Only one strand of the DNA serve as a template.

50

Tanscription in Prokaryotes

Polymerization catalyzed by RNA polymerase

- Can initiate synthesis
- Uses rNTPs
- Requires a template
- Unwinds and rewinds DNA

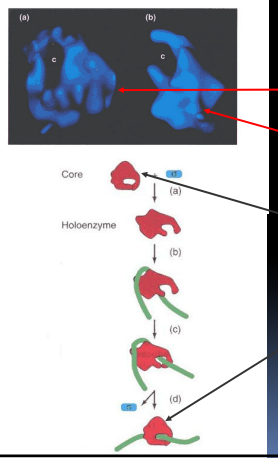
RNA Polymerase

- 5 subunits, 449 kd (~1/2 size of DNA pol III)
- Core enzyme
  - 2 α subunits---hold enzyme together
  - β--- links nucleotides together
  - β'---binds templates
- σ---recognition
- Holoenzyme= Core + sigma

Features of RNA Polymerase

- Starts at a **promoter sequence**, ends at **termination signal**
- Proceeds in 5′ to 3′ direction
- Forms a temporary DNA:RNA hybrid
- Has complete processivity

RNA Polymerase



- X-ray studies reveal a "hand"
- Core enzyme closed
- Holoenzyme open
- Suggested mechanism
- NOTE: when sigma unattached, hand is closed
- RNA polymerase stays on DNA until termination.

- <u>INITIATION</u>:
- Regions along the DNA molecule that signal initiation are called ***Promoter Regions*** /***Promoters***.
- There are ***two (2) types***:
  - the ***-35 promoter region*** and
  - **the *-10 promoter region*** (Pribnow box)

  The numbers refer to distance from the start point
- The pol surrounds the first base pair that is transcribed into RNA, ***START POINT***.

55

TRANSCRIPTION *contd.*

- Sequences prior to the start point are described as ***UPSTREAM*** while sequences after are called ***DOWNSTREAM***.
- Conventionally, sequences are written so that transcription proceeds from left (upstream) to right (downstream) corresponding to the usual 5^1 to 3^1 direction.
- Numbers of nucleotides increase going downstream.
- The base before start point is numbered -1 and negative numbers increase going upstream.

56

TRANSCRIPTION *contd.*

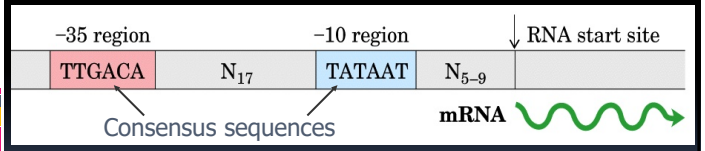
- A Prokaryotic Promoter.
- The two important sequences are the sites to which the RNA polymerase initially binds i.e. -35 or binding site and the sequences which align RNA pol against the antisense strand of DNA i.e. -10 region (also called PRIBNOW or TATABOX).
- In *E. coli* RNA pol recognizes -35 and -10, especially -35 hence called ***RECOGNITION SEQUENCE.***

5^1 — TTGTCAG — TATAAT — 3^1

-35 (~6pb) -10 (6-7bp) +1(Transcription Start Site)

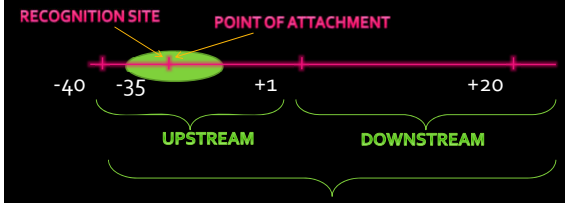
57

- A crucial feature of the promoter region is the spacing between the pribnow and the recognition site.
- For optimal transcriptional rate, it should be exactly ***17nucleotide long***, but actual sequence seems to be of little importance.



- NB: Transcriptions start on the antisense strand in the 3^1 to 5^1whilst RNA is formed in the opposite (5^1 to 3^1) direction.

58



Enzyme covers the whole surface

When the enzyme attaches to the range it moves to cover ~60bp (i.e. from ~40ntd upstream to 20ntd downstream from the actual start point).

59

TRANSCRIPTION *contd.*

- Evidence shows that the -10 region is where breakage of the DNA double occurs with the unwinding of the double helix.
- Melting is essential to expose the base of the template strand in order to direct transcript synthesis.
- The sigma factor dissociates from the rest of the enzyme when the open promoter complex is formed.
- RNA pol binds initiation site and rewind the double helix.
- When it binds initially, it is called ***PROMOTER COMPLEX***.

60

TRANSCRIPTION *contd.*

- This leads the core enzyme to continue the process of transcription.
- The first two ribonucleotides are then base paired to the template strand positions +1 and +2.
- This leads to the first phosphodiester bonds of the RNA molecule synthesized.

61

TRANSCRIPTION *contd.*

- ELONGATION OF TRANSCRIPT
- RNA is synthesized in 5^1-3^1 direction with nucleotide triphosphate (NTP) acting as substrate for the enzyme.
- Each ribonucleotide is added to the following reaction.

$$(NMP)_n + NTP \xrightarrow[ENZYME]{DNA, Mg^{2+}} (NMP)_{n+1} + PP_i$$

62

TRANSCRIPTION *contd.*

- Each step of transcription involves the addition of one ribonucleotide (NMP) to the growing polypeptide chain NMP_{n+1} using a nucleotide triphosphate (NTP) as the precursor or substrate.
- The energy for the reaction is derived from splitting the high energy nucleotide triphosphate (NTP) into monophosphate and releasing the inorganic phosphates (Pp_i ,Pyrophosphate).
- The enzyme traverse the entire gene until it encounters a specific nucleotide sequence that acts as a terminator sequence.

63

TRANSCRIPTION *contd.*

- The enzyme traverse the entire gene until it encounters a specific nucleotide sequence that acts as a terminator sequence.
- The enzyme attaches sigma (σ) factor falls off – core enzyme coils till a sequence signaling termination is reached indicating that that gene is finished.
- Here strands unwind unlike replication where the two strands are separated forever.

64

TRANSCRIPTION *contd.*

- NB:
- 1. The RNA transcript that is made is longer than the gene (template).
- 2. Only a small region of the double helix is unwound at anytime
- This region is called the ***TRANSCRIPTION BUBBLE***
- Once the elongation begins, the polymerized protein of the double helix of the RNA molecule dissociates from the DNA template and allow the double helix of the DNA to return to its original state.

65

TRANSCRIPTION *contd.*

- 3. The rate of elongation is not constant (In E. coli, the rate of elongation is about 50nt/s at 37°C; organism has its rate.
- TERMINATION
- Like initiation, termination occurs at certain positions along the DNA when RNA pol reaches a side of the DNA called the ***TERMINATOR SITE/TERMINATOR SEQUENCE.***
- Such sequences (about 40bp in length) are very important to prokaryote due to close proximity of the end of one gene and upstream sequence of adjacent gene.

66

- Bacteria may use one of two different strategies for transcription termination –
  - Rho-independent termination and
  - Rho-dependent termination.
- In Rho-independent transcription termination, which is also known as intrinsic termination,
  - RNA transcription stops when the newly synthesized RNA molecule forms a G-C-rich hairpin loop followed by a run of Us.

67

TRANSCRIPTION *contd.*

- In some cases, termination is dependent on a termination factor called ***Rho (ρ)***.
- This is called ***Rho-Dependent Termination***, and only occurs when the protein '*ρ*' is present.
- *ρ* attaches to the growing RNA molecule and when the polymerase enzyme pauses.
- It actively disrupts the base pairing between the DNA template and RNA.

68

TRANSCRIPTION *contd.*

- ***Mechanism*:**
- -Binding of **ρ** protein to a specific site on the RNA is termed the RUT (Rho Utilization)
- - ρ then 'pulls' the RNA off the enzyme (RNA pol after transcription)

69

EUKARYOTIC TRANSCRIPTION

- Very similar to prokaryotic transcription but initiation is more complicated.
- ***Major Difference*:**

a) Occurs within the nucleus under three (3) separate forms of RNA pol.

- Unlike prokaryotes, the eukaryotic transcript is not free to associate with ribosome prior to completion of transcription.

70