CORE IDEA 2: GENETICS AND INHERITANCE
TOPIC 2.4: GENE EXPRESSION AND REGULATION (PART 1)

Learning Outcomes:

(c) Describe how the information on DNA is used to synthesize polypeptides in prokaryotes and eukaryotes (description of the processes of transcription, formation of mRNA from pre-mRNA and translation is required)

(d) **[modified]** Describe the structure and organisation of **eukaryotic genomes** (including DNA / RNA, single- / double-stranded, number of nucleotides, packing of DNA, linearity/circularity and presence of introns).

(i) Describe the structure and function of non-coding DNA in eukaryotes (i.e. portions that do not encode protein or RNA, including introns, centromeres, telomeres, promoters, enhancers and silencers) *(Knowledge of transposons, satellite DNA, pseudo-genes and duplication of segments is not required)*

(j) explain how differential (i.e. spatial and temporal) gene expression in eukaryotes can be regulated at different levels:

(i) chromatin level (histone modification and DNA methylation)
(ii) transcriptional level (control elements, such as promoters, silencers and enhancers, and proteins, such as transcription factors, including activators and repressors)
(iii) post-transcriptional level (processing of pre-mRNA in terms of splicing, polyadenylation and 5' capping)

Use the knowledge gained in this section in new situations or to solve related problems.

References:

Reece, J. B. et al. (2018). Campbell Biology (11th Ed). Chapter 18: Regulation of Gene Expression, pp. 368 – 385.

Note: This textbook is available in our library. You may wish to borrow the book to supplement your reading when necessary.

Contents

1. Glossary of Terms

Retrieve knowledge from Topic 2.1 Structure and Function of Nucleic Acids:
DNA is a long, double-stranded molecule that forms a double helix. Each strand is made up of deoxyribonucleotides, which are composed of a pentose sugar, a phosphate group, and one of four nitrogenous bases (adenine, thymine, guanine, and cytosine). The two strands are held together by hydrogen bonds between the complementary base pairs (adenine pairs with thymine, and guanine pairs with cytosine).

Definition list

1. **Genome:** The complete set of genetic instructions or DNA present in an organism. It contains all the information needed for the development, growth, and functioning of that organism.

2. **DNA (Deoxyribonucleic Acid)**: A long, double-stranded molecule that carries the genetic instructions for the development, functioning, and reproduction of living organisms. It is composed of four nucleotide bases: adenine (A), thymine (T), guanine (G), and cytosine (C).

3. **Chromosome:** A thread-like structure composed of DNA and proteins, found in the nucleus of cells. Chromosomes carry the genetic information in the form of genes

4. **Coding Region:** The portion of a gene's DNA or RNA sequence that provides instructions for the synthesis of a specific protein. It is the region that is translated into the amino acid sequence of a protein.

5. **Non-coding Region:** The portion of a gene's DNA or RNA sequence that does not code for a protein. These regions may have regulatory functions or may not have any known function.

6. **Coding Sequence:** The specific sequence of nucleotides in a gene's DNA or RNA that encodes the amino acid sequence of a protein. It is the same as the coding region.

7. **Non-coding Sequence:** The sequence of nucleotides in a gene's DNA or RNA that does not encode a protein. These sequences may have regulatory or structural roles.

8. **Centromere:** A constricted region of a chromosome that holds the two sister chromatids together during cell division. It is the attachment point for the spindle fibers that separate the chromosomes during cell division.

9. **Telomere:** A repetitive nucleotide sequence at the end of a chromosome that protects the chromosome from deterioration during cell division.

10. **Promoter:** A specific DNA sequence located near the beginning of a gene that serves as a binding site for RNA polymerase and other transcription factors, initiating the process of transcription.

11. **Enhancer:** A short DNA sequence that can increase the transcription of a gene by binding to specific proteins called transcription factors, even if the enhancer is located far away from the gene it regulates.

12. **Silencer:** A DNA sequence that binds to specific proteins and inhibits the transcription of a gene, effectively silencing or reducing its expression.

13. **Intergenic Introns:** Non-coding regions of DNA that are located between genes.

14. **Intragenic Introns:** Non-coding regions of DNA that are located within a gene, interrupting the coding sequence.

15. **Exons:** The coding regions of a gene that are ultimately transcribed into messenger RNA (mRNA) and translated into a protein.

16. **Regulatory Sequence:** A DNA sequence that controls the expression of a gene by binding to specific proteins or other regulatory molecules.

17. **Regulatory Gene:** A gene that encodes a protein involved in the regulation of other genes, such as transcription factors or signaling molecules.

18. **Gene:** A specific sequence of nucleotides in DNA provides the instructions for the synthesis of a functional product, such as a protein or a non-coding RNA molecule.

1.1 Organisation of DNA

(i) The DNA double helix is coiled 1¾ turns (146 bp) by forming ionic bonds around an **histones octamer** forming the **<u>nucleosome</u>**.

(ii) The nucleosome is further stabilised by a **H1 histone**.

Note:

- *Histones are positively charged due to abundance of the amino acids, lysine and arginine.*
- *Histone tails containing amino acids available for modification, protrudes from the nucleosome and interacts with DNA.*

(iii) **<u>Linker DNA</u>** links nucleosomes to form a **<u>10nm 'beads-on-string' structure</u>**.

(iv) Further coiling occurs to result in a **<u>30 nm chromatin fibre</u>**.

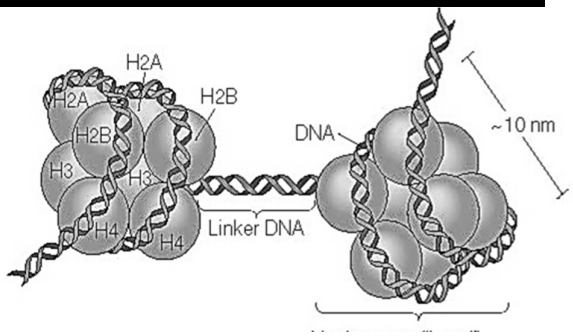


Fig. 1.1.2: Nucleosome

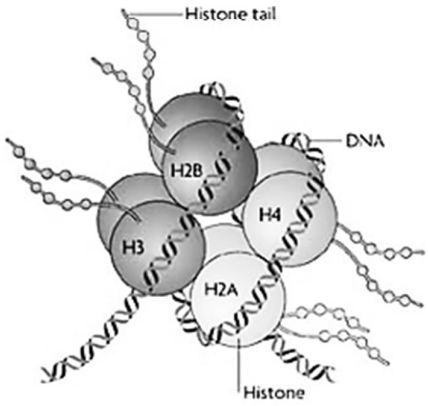


Fig. 1.1.1: Formation of nucleosome

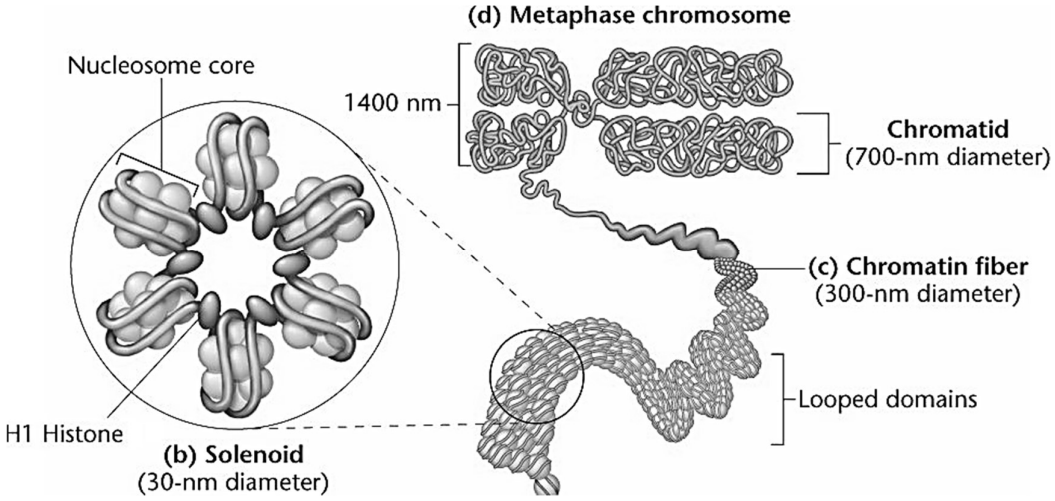


Fig. 1.1.3: Formation of solenoid & looped domains

(v) The 30 nm chromatin fibre forms **looped domains** by attaching to a scaffold of non-histone proteins, thus forming the **300 nm chromatin fibre**.

(vi) The 300 nm chromatin fibre further coil and fold upon themselves to produce the **700 nm chromatid**.

(vii) At prophase, each chromosome consists of 2 sister chromatids held at the centromere.

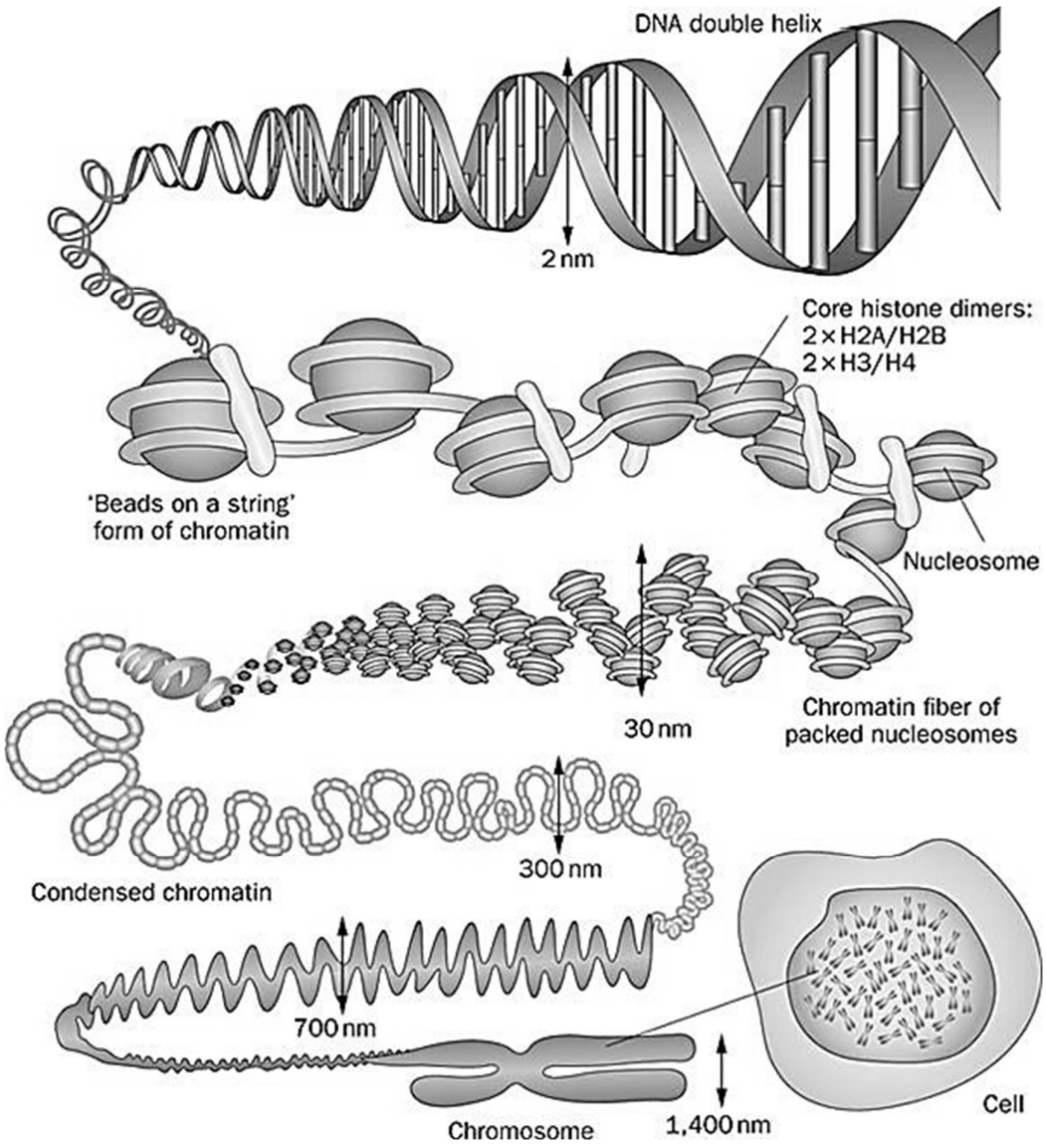


Fig. 1.1.4: Organisation of DNA into chromosomes

1.2 Definition of a Gene

- A gene is a sequence of deoxynucleotides that code for a functional protein or RNA. Although most genes encode proteins, some (e.g. tRNA genes, rRNA genes) only encode RNAs such as tRNAs and rRNAs.

- Gene contains the coding regions (exons) and the noncoding regions such as transcription control regions (e.g. promoter, enhancer and silencer) and intragenic introns.

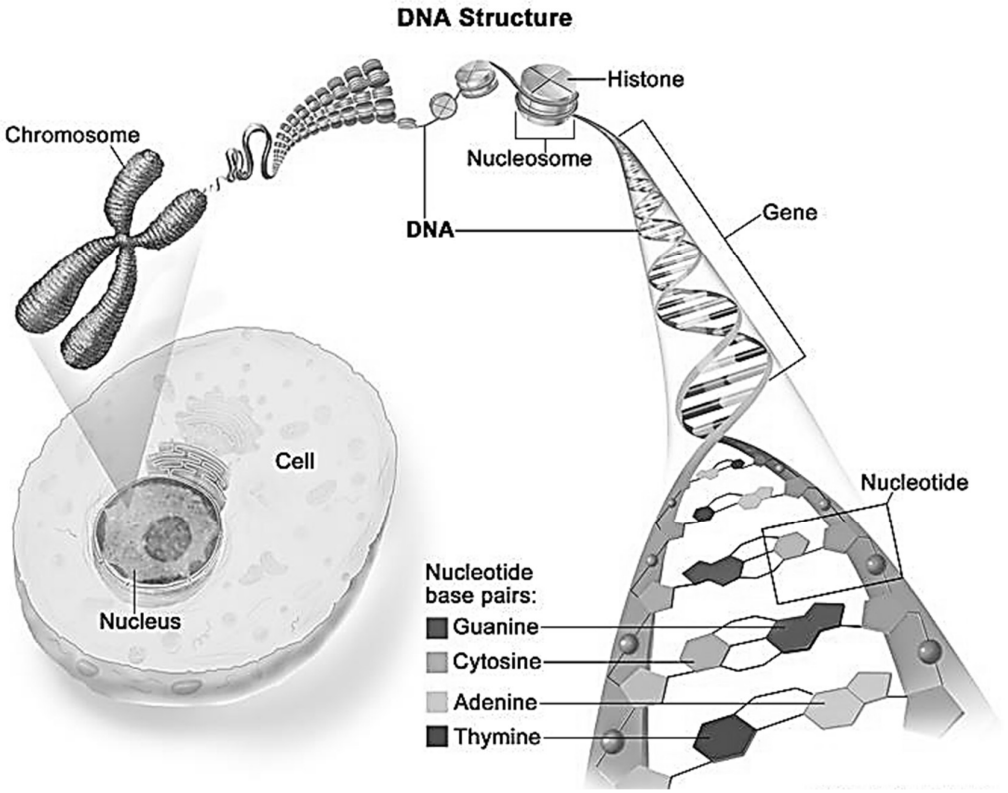


Fig. 1.2.1: Organisation of genes in a chromosome

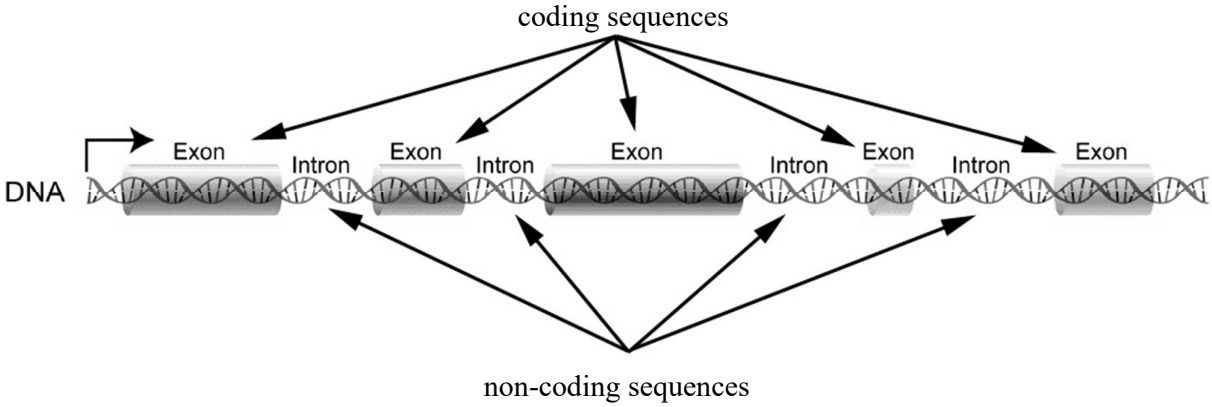


Fig. 1.2.2: Organisation of coding (exons) and non-coding (intragenic introns) sequences within a gene

Learning Outcome (d) [modified]:

Describe the structure and organisation of **eukaryotic genomes** (including DNA / RNA, single-/double-stranded, number of nucleotides, packing of DNA, linearity/circularity and absence/presence of introns).

1.3 Structure and Organisation of Eukaryotic Genomes

- Eukaryotic genomes are complex and highly organized, consisting of both DNA and RNA molecules including genetic materials in mitochondrion and/or chloroplasts.

From Topic 2.1 Structure and Function of Nucleic Acids

DNA and RNA (Fig. 1.3.1)

- DNA (Deoxyribonucleic Acid) is the primary genetic material in eukaryotic cells. It is **double-stranded**, forming a double helix structure.

- RNA (Ribonucleic Acid) exists in several forms (mRNA, tRNA, rRNA) and is usually **single-stranded**. It plays crucial roles in protein synthesis and gene regulation.

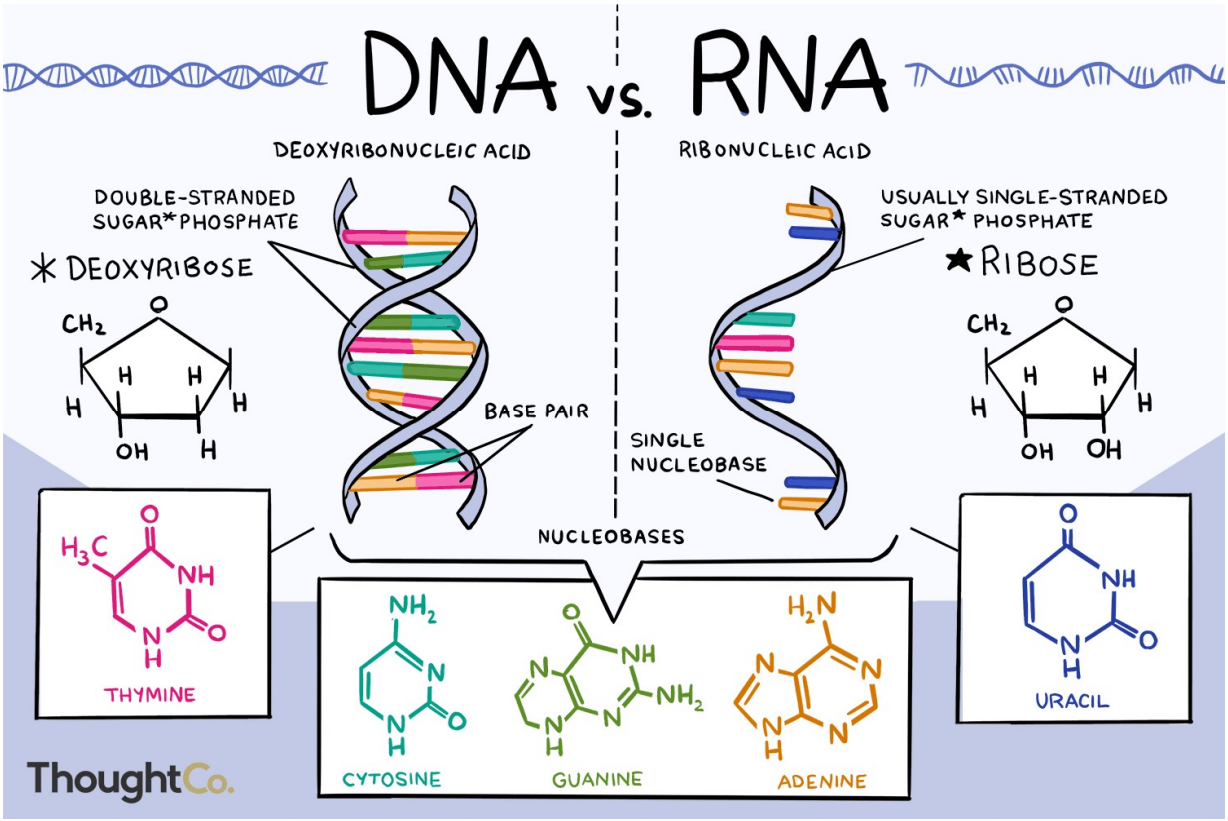


Fig 1.3.1: Difference between DNA and RNA

For a full detailed description of structure of DNA and RNA, refer to Topic 2.1 Structure and Function of Nucleic Acids.

Number of Nucleotides

- Eukaryotic genomes contain **millions to billions of nucleotides**. For example, the human genome has about 3 billion base pairs.
    - Eukaryotes have **complex structures and functions**, requires a larger amount of genetic information, resulting in more nucleotides.
    - Eukaryotic cells perform a **wide variety of specialized functions**. For example, different cell types in multicellular organisms have unique roles, necessitating a vast and varied set of genes.
    - Eukaryotic genomes contain **extensive regulatory sequences** that control gene expression. These include promoters, enhancers, and silencers, which add to the overall number of nucleotides.
    - Eukaryotic genes often undergo **alternative splicing**, where different combinations of exons are joined together to produce multiple proteins from a single gene. This process requires additional nucleotide sequences.
    - Many eukaryotic genes (Fig. 1.3.2) contain **introns**, non-coding sequences that are removed during RNA splicing. Introns can be quite long and contribute significantly to the total nucleotide count.
    - Eukaryotic genomes often contain large amounts of **repetitive non-coding DNA**, including tandem repeats (e.g., microsatellites) and interspersed repeats (e.g., transposable elements). These sequences do not code for proteins but play roles in genome structure and regulation.
    - Over evolutionary time, eukaryotic genomes have undergone **duplications**, leading to **additional copies of genes and other sequences**. Whole-genome duplications and segmental duplications contribute to the increased number of nucleotides.

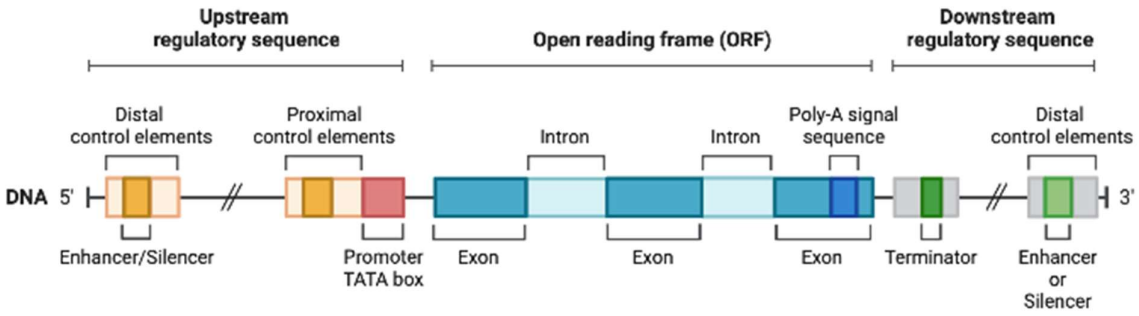


Fig. 1.3.2: Structure of eukaryotic gene

From page 5 – 6: Organisation of DNA

Packing of DNA (Fig. 1.3.3)

- **Chromatin**: DNA is wrapped around histone proteins, forming nucleosomes, which look like beads on a string.

- **Chromosomes**: During cell division, chromatin further condenses into visible chromosomes. Eukaryotic cells typically have **multiple linear chromosomes** (e.g., humans have 46 chromosomes).

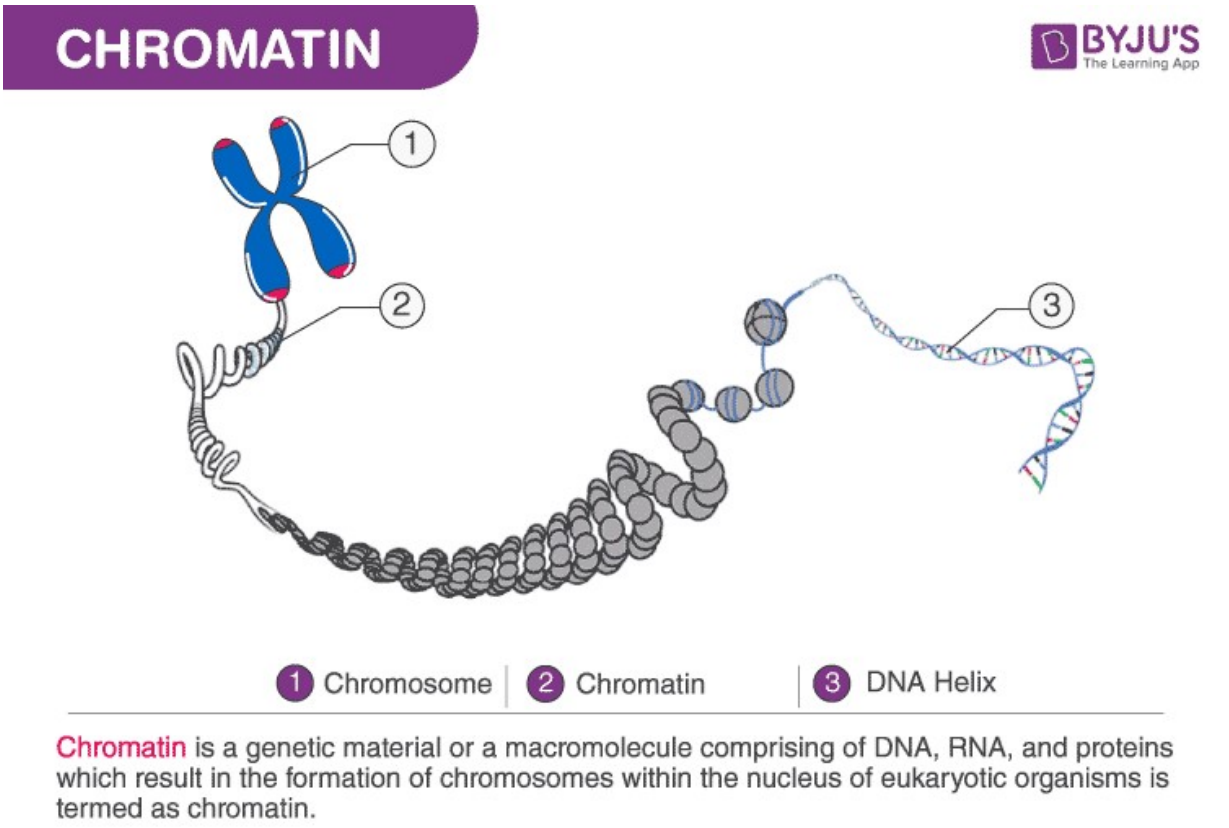


Fig. 1.3.3: Packing of DNA – from double helix to chromosome

For a full detailed description of packing of DNA, refer to Topic 2.7 Gene Expression and Regulation Part (I) page 5 to page 6.

Shape

- Linear DNA: In eukaryotic cells, the primary genetic material is organized into **multiple linear chromosomes**. Each chromosome is a **long, linear** DNA molecule **associated with proteins** to form a complex called chromatin.
    - Telomeres: The ends of linear chromosomes are protected by **telomeres,** repetitive nucleotide sequences that prevent the loss of genetic information during DNA replication.
    - Centromeres: Each chromosome has a **centromere**, a specialized region that plays a crucial role during cell division by ensuring proper segregation of chromosomes to daughter cells.
- Circular Mitochondrial DNA: Mitochondrial DNA is **typically circular** and **much smaller** than nuclear DNA.
    - Human mitochondrial DNA is about 16,569 base pairs long and encodes 37 genes essential for mitochondrial function.
    - Mitochondrial DNA is usually **inherited maternally**, meaning it is passed down from mother to offspring.
- Circular Chloroplast DNA: Chloroplasts, the photosynthetic organelles in plants and some algae, also have their own **circular DNA**.
    - Chloroplast genomes **vary in size but are generally larger than mitochondrial genomes**. For example, the chloroplast genome of Arabidopsis thaliana is about 154,478 base pairs long.
    - Chloroplast DNA encodes genes involved in photosynthesis and other functions essential to chloroplast operation.

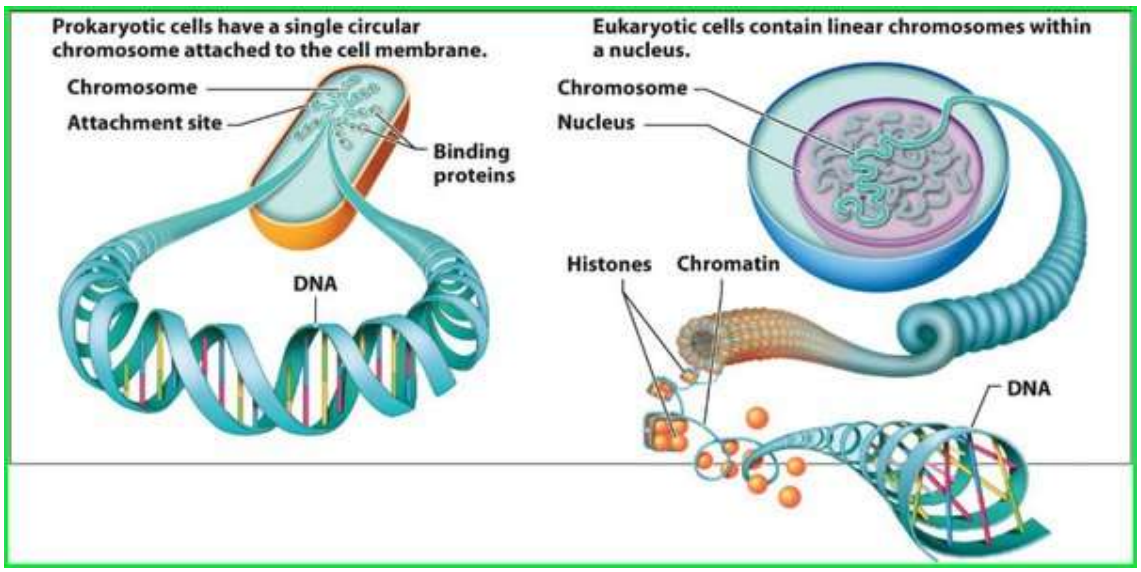


Fig. 1.3.4: Circular chromosome versus linear chromosome

Introns

- Introns: Eukaryotic genes often contain introns, **non-coding sequences** that are removed during RNA splicing.
- Exons: The **coding sequences** that remain after introns are removed and are joined together to form the final mRNA.

Checkpoint 1
Identify whether the statements below are true or false by circling either T / F and provide your reasons for your answers.

Q	Statement	Choice	Justification
1	In eukaryotic cells, both nuclear DNA and mitochondrial DNA are organized into linear chromosomes.	T / F	
2	Eukaryotic genomes contain a higher number of nucleotides compared to prokaryotic genomes because eukaryotic cells are more complex and have more regulatory sequences.	T / F	
3	All genes in eukaryotic genomes contain introns, which are removed during RNA splicing.	T / F	

Learning Outcome (h):

Describe the structure and function of non-coding DNA in eukaryotes (i.e. portions that do not encode protein or RNA, including introns, centromeres, telomeres, promoters, enhancers and silencers)

1.4 Non-coding DNA

- Non-coding DNA refers to the portions of the eukaryotic genome that do not encode for proteins or functional RNA molecules.
- Despite not coding for gene products, non-coding DNA plays crucial roles in regulating gene expression, chromosome structure, and genome organization.

1.4.1 Introns

(i) Structure

- Intragenic introns are non-coding sequences found within genes. They are located between exons (coding sequences).
- Intergenic introns are non-coding sequences found between genes.

(ii) Function

- Intragenic introns are transcribed into pre-mRNA and will be removed from the pre-mRNA during RNA splicing to produce mature mRNA.
- They allow for alternative splicing, which enables a single gene to produce multiple protein variants, increasing protein diversity.

1.4.2 Centromere

(i) Structure

- Centromeres are **specialized DNA sequences**. They consist of **highly repetitive** DNA sequences and **associated proteins (e.g. kinetochore)**.

- The centromere is a **constricted region** of a chromosome. Each chromosome / chromatid has its own centromere.

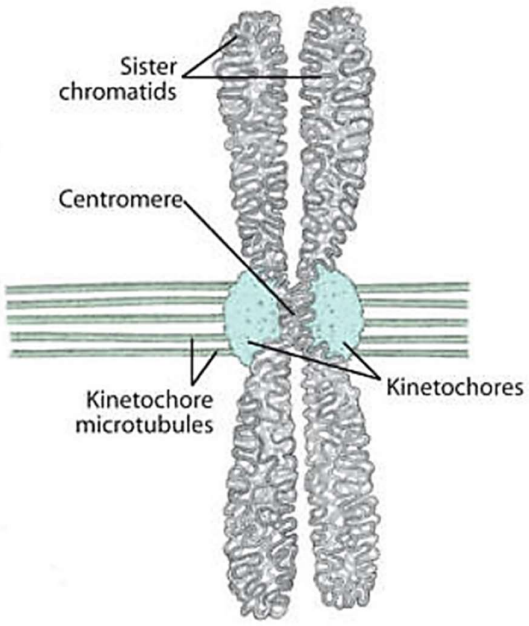


Fig 1.4.2.1: Each sister chromatid has its own centromere with kinetochore.

(ii) Function

- Centromeres serve as **attachment points for spindle fibers** during cell division, ensuring proper chromosome segregation.

- Spindle fibers bind to the kinetochore proteins. When the kinetochore microtubules contract, the **centromeres divide** and the sister chromatids will be **pulled to opposite poles** of the cell during anaphase of mitosis or anaphase II of meiosis.

- Without centromeres, improper chromosomal alignment and segregation will result in **aneuploidy** e.g. Trisomy 21 or Down syndrome.

(iii) Distribution

- Centromeres are found **only in eukaryotes** as they are found in **<u>linear chromosomes only</u>**. There are no centromere on circular prokaryotic DNA.

1.4.3 Telomere

(i) Structure

- Telomeres are repetitive DNA sequences at the ends of linear chromosomes. A prophase chromosome has four telomeres.

- Telomeres consist of a series of short tandem repeat sequences where each repeat unit is about 5 – 10 nucleotides long. E.g. **<u>TTAGGG</u>** is repeated 2500 times in human telomeres at birth.

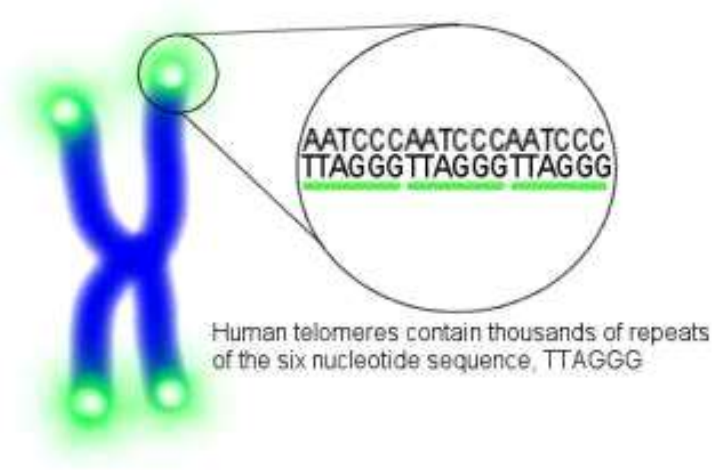


Fig 1.4.3.1: Structure of the human telomere

- Telomeres have a **single-stranded region** of DNA at their 3' ends known as **<u>3' overhang</u>**. This region of DNA does not have a complementary strand (Fig. 1.4.3.2)

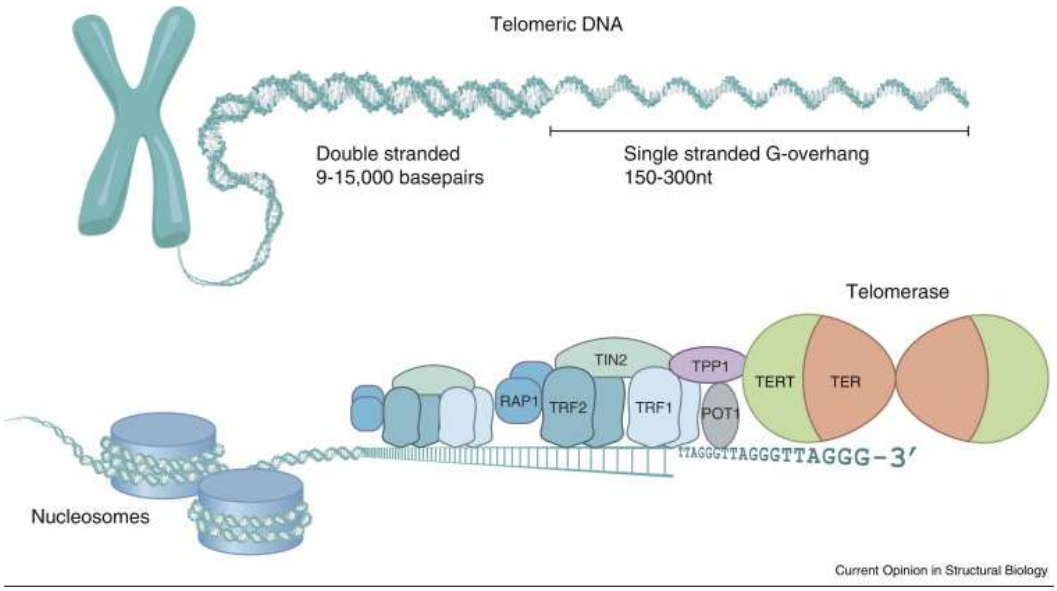


Fig. 1.4.3.2: Structure of telomere

- The **3' single-stranded overhang folds back and forms a T-loop (Fig. 1.4.3.3),** hiding the single-stranded DNA overhang. The overhang **bonds with an earlier repeat of the same complementary sequence in the opposite strand**.

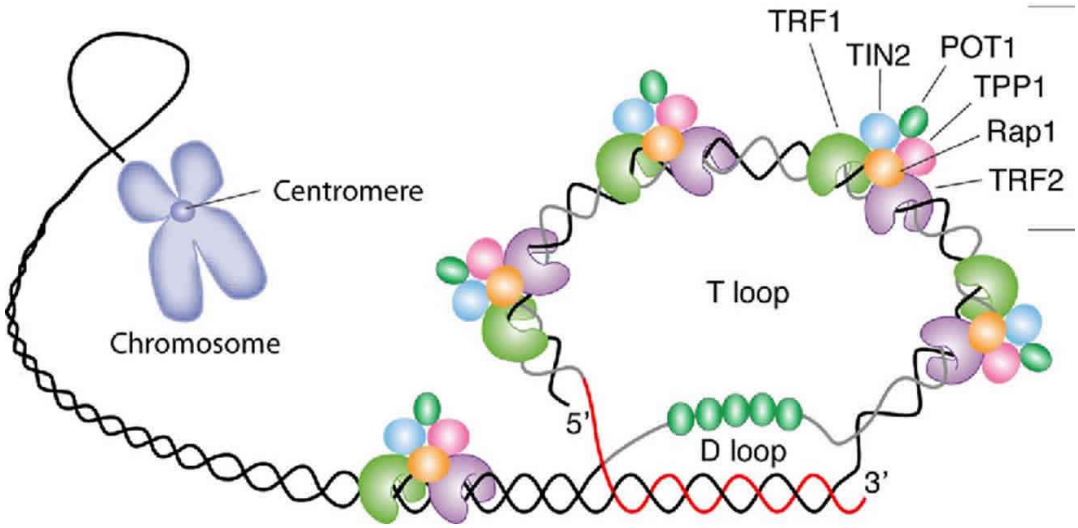


Fig 1.4.3.3: Model of Telomere Structure

- **<u>Telomere capping proteins</u>** bind at the T-loop juncture to maintain the stability of this structure hence preventing degradation by exonucleases.

(ii) Function

- Telomeres **play a role in aging and cellular senescence** by shortening with each cell division.
    - Due to the end-replication problem, **telomeres shorten with every round of DNA replication**. Since telomeres are non-coding, shortening of the telomeres has no deleterious effects.
    - Telomeres **protect the genes within the chromosome from gene erosion** with each round to DNA replication**, preventing loss of vital genetic information**.
    - Telomeres provide a counting mechanism for the number of cell division a cell has undergone (Hayflick limit). Hayflick limit states that a normal human cell can only replicate and divide forty to sixty times before it cannot divide anymore and will break down by programmed cell death or apoptosis.
        - ♦ As a cell replicates, its telomeres will eventually **shortened to a critical length** and trigger signal transduction pathways that lead to cell cycle arrest or apoptosis.
        - ♦ The shortening of telomeres thus **indirectly prevents the development of cancer by preventing unlimited proliferation** of cells in adult tissue.
- Telomeres **protect and stabilize the terminal ends of chromosomes**
    - Telomere's T-loop **prevents the ends of chromosomes from being degraded by exonucleases**. A cell can detect the open end of a DNA molecule as DNA damage and may trigger signal transduction pathways leading to cell cycle arrest or cell death (apoptosis).
        - ✦ If there is no formation of T-loop, the **single stranded 3' overhang** may **anneal to a complementary single-stranded region of the terminal end of another chromosome**. This causes joining of different chromosomes leading to chromosomal aberration. ***(You will learn more about chromosomal aberrations in Topic 2.7 Chromosomal Aberrations.)***
        - ✦ Thus by forming the T loop (Fig. 1.4.3.3), the telomeres **stabilise the ends** of the chromosomes by
            - ▪ **preventing them from fusing with other chromosomes**, and
            - ▪ preventing DNA repair machinery from recognising the ends of chromosomes as DNA breaks, **protecting the chromosome**, hence **preventing apoptosis**.
- Telomeres **allow their own extension,** by providing an attachment point for the correct positioning of the enzyme telomerase**.**
    - In cells such as **stem cells and germ cells** (i.e. cells which give rise to gametes) that need to **divide repeatedly for prolonged period,** an enzyme **telomerase** is expressed to **maintain telomere length.**

- Telomerase catalyses the **lengthening of telomeres**. Hence, in cells that express telomerase, the end-replication problem is **overcomed as telomeres can be lengthened** with each round of division. (*Note: End replication is **NOT prevented**.)*

  - Telomerase is a **ribonucleoprotein** containing both RNA and protein subunits.

  - Telomeric RNA provides the **template sequence 3'-AAUCCC-5'** to guide the synthesis and insertion of the telomeric DNA sequence -5'-TTAGGG-3'.

  - Protein subunits form the **hTERT** (**human telomere reverse transcriptase**) provides the **catalytic action** of telomerase. Reverse transcriptase synthesises a double-stranded DNA from a single-stranded RNA template.

  - The following mechanism (Fig. 1.4.3.4) is the currently accepted model as to how telomeres replicate:

    1. Telomerase binds to telomere via **complementary base pairing** between the 3' overhang of the telomere and the RNA component of the enzyme.

    2. The **3' overhang is extended using the telomerase RNA as a template**. DNA nucleotides complementary to the telomerase RNA are added.

    3. The telomerase **translocates towards the 3' end** of the newly added sequence for further extension.

    4. The **extended 3' end is used as a template for addition of primers and DNA polymerase synthesises** a strand complementary to the 3' overhang.

  - In most somatic cells, telomerase is usually absent or inactive. This allows the cells to age and undergo cellular senescence. Adult stem cells are then triggered to divide asymmetrically to replace these aged cells.

  - However, if telomerase becomes active in somatic cells, a signal is conveyed for **cells to keep dividing**. Hence, such cells achieve limitless replicative potential, a hallmark of cancer. ***(You will learn more about this in Topic 2.8 Molecular Biology of Cancer.)***

(iii) Distribution

- Telomeres are found **only in eukaryotes, at both ends of linear chromosomes**.

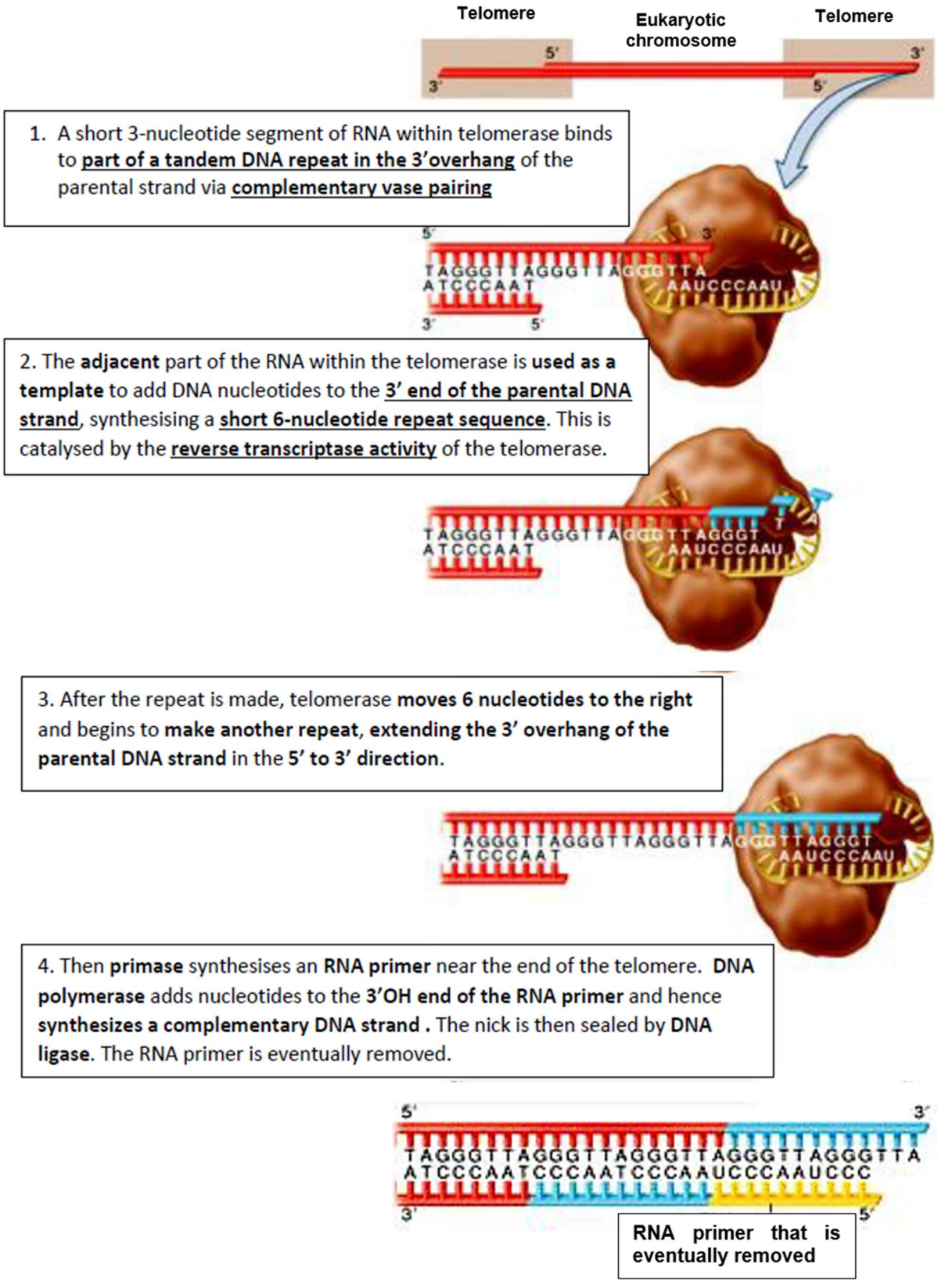


Fig 1.4.3.4: Action of Telomerase

1.4.4 Promoter

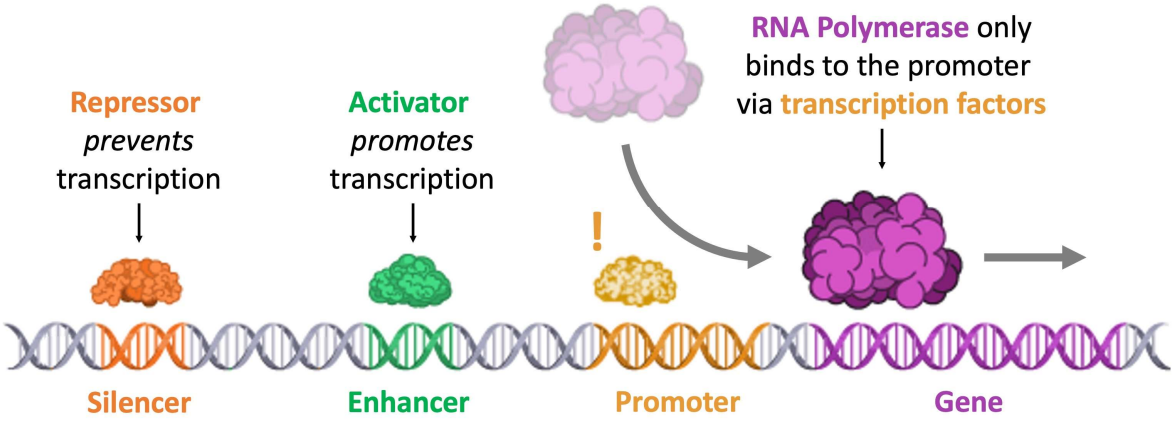


Fig. 1.4.4.1: Regulatory sequences in a gene

(i) **Structure**

- Core Promoter: The core promoter is located **immediately upstream of the gene** and includes the **TATA box** (a conserved DNA sequence found in many promoters) and the transcription initiation site. The TATA box is typically located about 25-30 base pairs upstream of the transcription initiation site.

- Proximal Promoter Elements: These are regulatory sequences located **close to the core promoter**. Examples include the CAAT box and **GC-rich regions**.

(ii) **Function**

- **Binding Site for RNA Polymerase**: The core promoter is the binding site for RNA polymerase II and **general transcription factors**, which are necessary for the initiation of transcription.

- **Regulation of Gene Expression**: Proximal promoter elements help regulate the **efficiency and rate of transcription initiation** by interacting with specific transcription factors.

1.4.5 Enhancer

(i) Structure

- Location: Enhancers can be **located far from the gene they regulate**, sometimes thousands of base pairs away. They **can be upstream, downstream, or within introns of the gene**.

- DNA Sequences: Enhancers are composed of **multiple binding sites for specific transcription factors, i.e. activators**.

(ii) Function

- **Increase Transcription Rates:** Enhancers enhance the transcription rate of associated genes by binding specific transcription factors, i.e. activators and co-activators.

- **DNA Looping:** Enhancers can interact with promoters through DNA looping, bringing the enhancer-bound transcription factors into close proximity with the promoter region. This **facilitates the assembly of the transcription initiation complex at the promoter**.

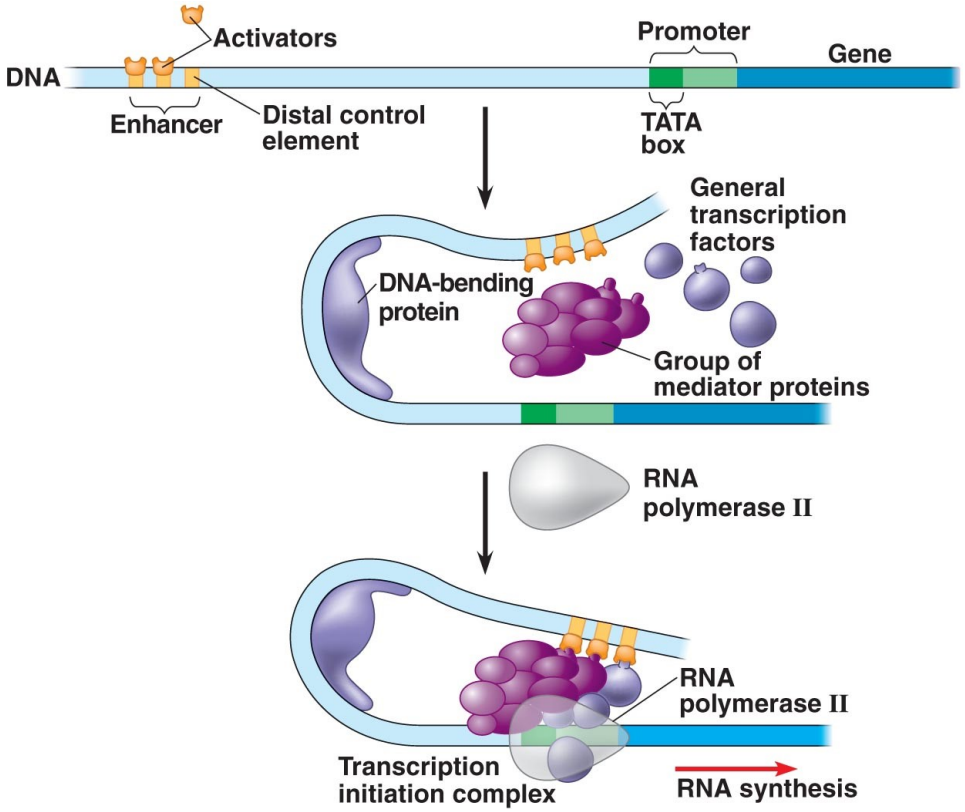


Fig 1.4.4.2: DNA-bending proteins cause DNA to loop such that multiple factors may interact at promoter region to regulate transcription initiation.

- **Tissue-Specific Expression:** Enhancers often contribute to tissue-specific and temporal regulation of gene expression, **ensuring that genes are expressed in the right cells at the right time**.

1.4.6 Silencer

(i) Structure

- Location: Silencers can be **located at various distances from the gene they regulate**. They can be **upstream, downstream, or within introns**.

- DNA Sequences: Silencers contain **binding sites for transcriptional repressors**.

(ii) Function:

- **Decrease Transcription Rates**: Silencers reduce the transcription rate of associated genes by binding transcriptional repressors and co-repressors.

- **Inhibit Transcription Machinery**: When bound by repressors, silencers can interfere with the assembly or activity of the transcription initiation complex at the promoter. This can happen through direct interaction or by inducing changes in chromatin structure that make the promoter region less accessible to RNA polymerase and general transcription factors.

Checkpoint 2

Identify whether the statements below are true or false by circling either T / F and provide your reasons for your answers.

Q	Statement	Choice	Justification
1	Introns are sequences within eukaryotic genes that are removed during RNA splicing and do not contribute to the coding sequence of proteins.	T / F	
2	Telomeres are regions at the ends of chromosomes that contain coding DNA sequences essential for protein synthesis.	T / F	
3	Enhancers are non-coding DNA sequences that decrease the rate of transcription of associated genes by binding transcriptional repressors.	T / F	

2. Central Dogma of Molecular Biology

Learning Outcome 2(c):

Describe how the information on DNA is used to synthesize polypeptides in prokaryotes and eukaryotes (description of the processes of transcription, formation of mRNA from pre-mRNA and translation is required)

In the discovery of the DNA structure, James Watson and Francis Crick built on the work of many scientists. This includes the works of Swiss physiological chemist, Friedrich Miescher on his identification of nuclein, DNA associated with proteins in cell nuclei in 1869, Russian biochemist Phoebus Levene on the discovery of the 3 major components of nucleotide in the 1900s, Austrian biochemist Erwin Chargaff's discovery that the total number of pyrimidines equals the total number of purines and X-ray crystallography work by English researchers Rosalind Franklin and Maurice Wilkins.

Further exchange of ideas between Crick and other scientists generated ideas on the possible relationship between DNA, RNA and protein. Crick took these ideas and the experimental data that increasingly suggested that RNA was intermediate between DNA and protein and developed a scheme to explain the relations between these three classes of biological molecules.

In 1956, Crick wrote a set of notes to detail his ideas on protein synthesis and introduced the hypothesis for the Central Dogma, which is related to the flow of information between genes and proteins. In this principle known as the central dogma of the chemistry of life, the genetic information flows from the DNA in our genetic material via RNA to proteins, which in turn construct cells with different functions.

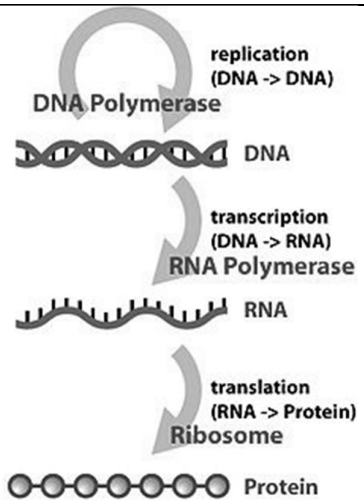


Fig. 2.3: The Central Dogma of molecular biology (Francis Crick, 1958)

- The central dogma of molecular biology describes the **flow of genetic information within a biological system**. It states that the **information stored in DNA is transcribed into RNA, which is then translated into proteins**. This process can be summarized as:DNA → RNA → Protein

- Here's how it works (Fig. 2.1):

    - **DNA** (Deoxyribonucleic acid) is the genetic material that **stores all the instructions for making proteins**. It is a double-stranded molecule found in the nucleus of eukaryotic cells.

    - **Transcription**: During transcription, a segment of **DNA is used as a template to synthesize a single-stranded RNA molecule called messenger RNA (mRNA)**. This process is carried out by an enzyme called RNA polymerase.

    - **Translation**: The mRNA molecule travels from the nucleus to the cytoplasm, where it is read by ribosomes. Ribosomes are cellular machines that **translate the genetic code from the mRNA into a sequence of amino acids**, forming a polypeptide which folds into protein.

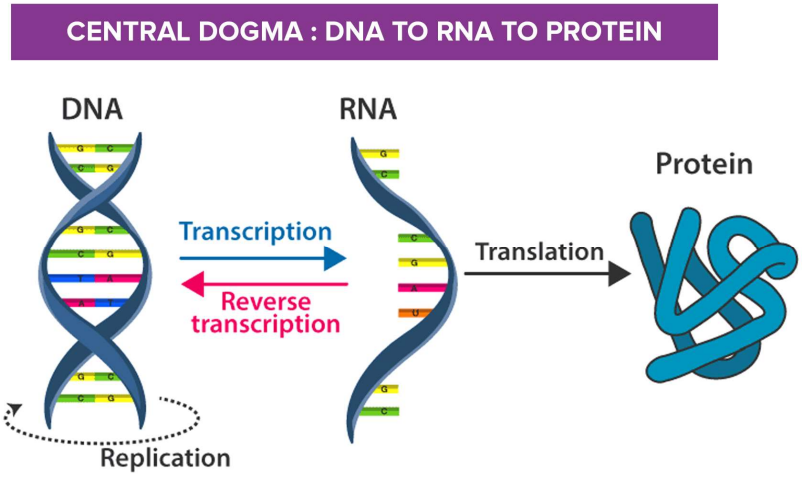


Fig. 2.1: Central Dogma of Molecular Biology

- The central dogma emphasizes that the **flow of genetic information is unidirectional**, meaning it goes from DNA to RNA to protein, but not in reverse. This is because DNA and RNA are nucleic acids that can store and transfer genetic information, while proteins are made up of amino acids and cannot store or transfer genetic information back to nucleic acids.
  Note that DNA replication in Fig 2.1 is <u>NOT</u> part of the central dogma of molecular biology.

EXCEPTION:
While the central dogma describes the general flow of genetic information, there are some exceptions and additional processes involved in gene expression and regulation, such as reverse transcription in retroviruses, e.g. HIV. ***(You will learn more in the topic of Genetics of Viruses.)***

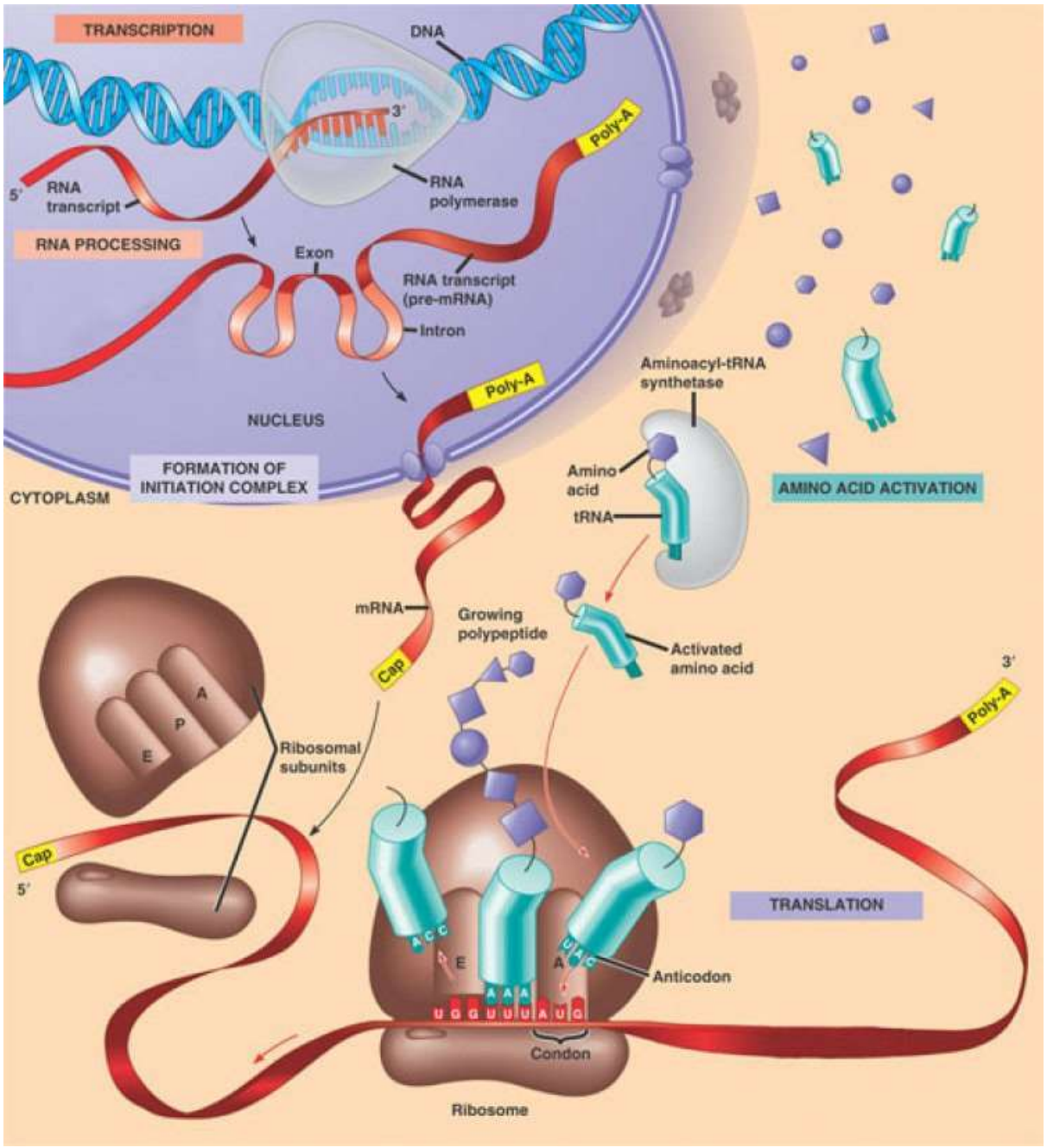


Fig. 2.2: Central Dogma of Molecular Biology

3. Protein Synthesis (Part I)

3.1 Transcription

Learning Outcome 2(c) modified:

Describe how the information on DNA is used to synthesise polypeptides in prokaryotes and eukaryotes. (Description of the processes of **transcription**)

- In both prokaryotes and eukaryotes, the information stored in DNA is used to synthesize polypeptides (proteins) through a two-step process: transcription and translation.

- Transcription is the first step, where the genetic information from DNA is copied into a messenger RNA (mRNA) molecule.

- Transcription is divided into three main stages: initiation, elongation and termination.

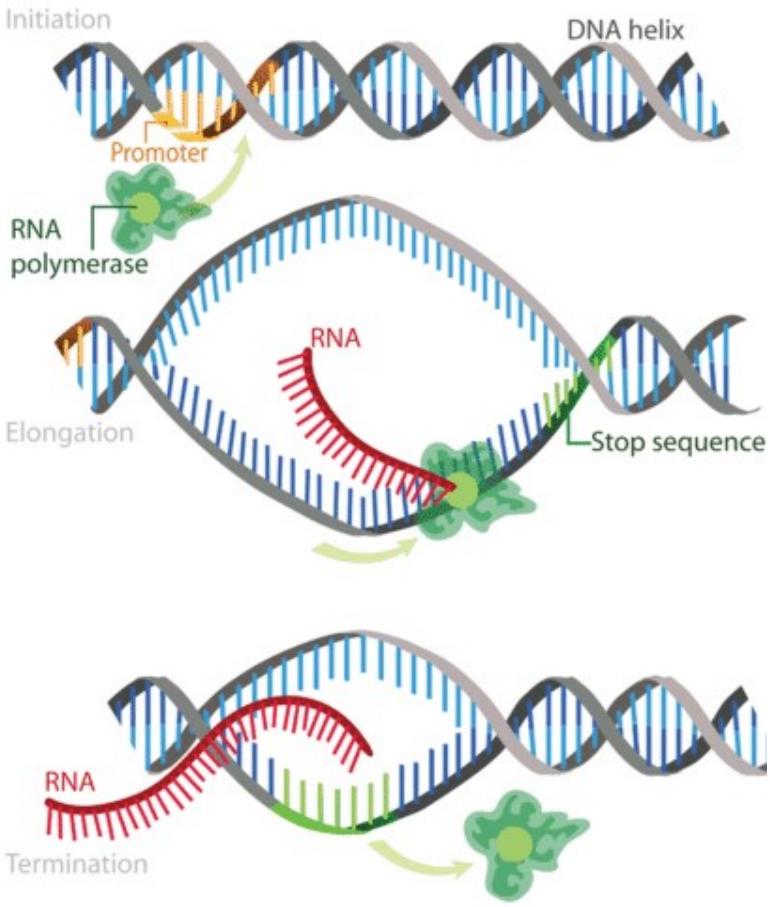


Fig. 3.1.1: Overview of transcription.

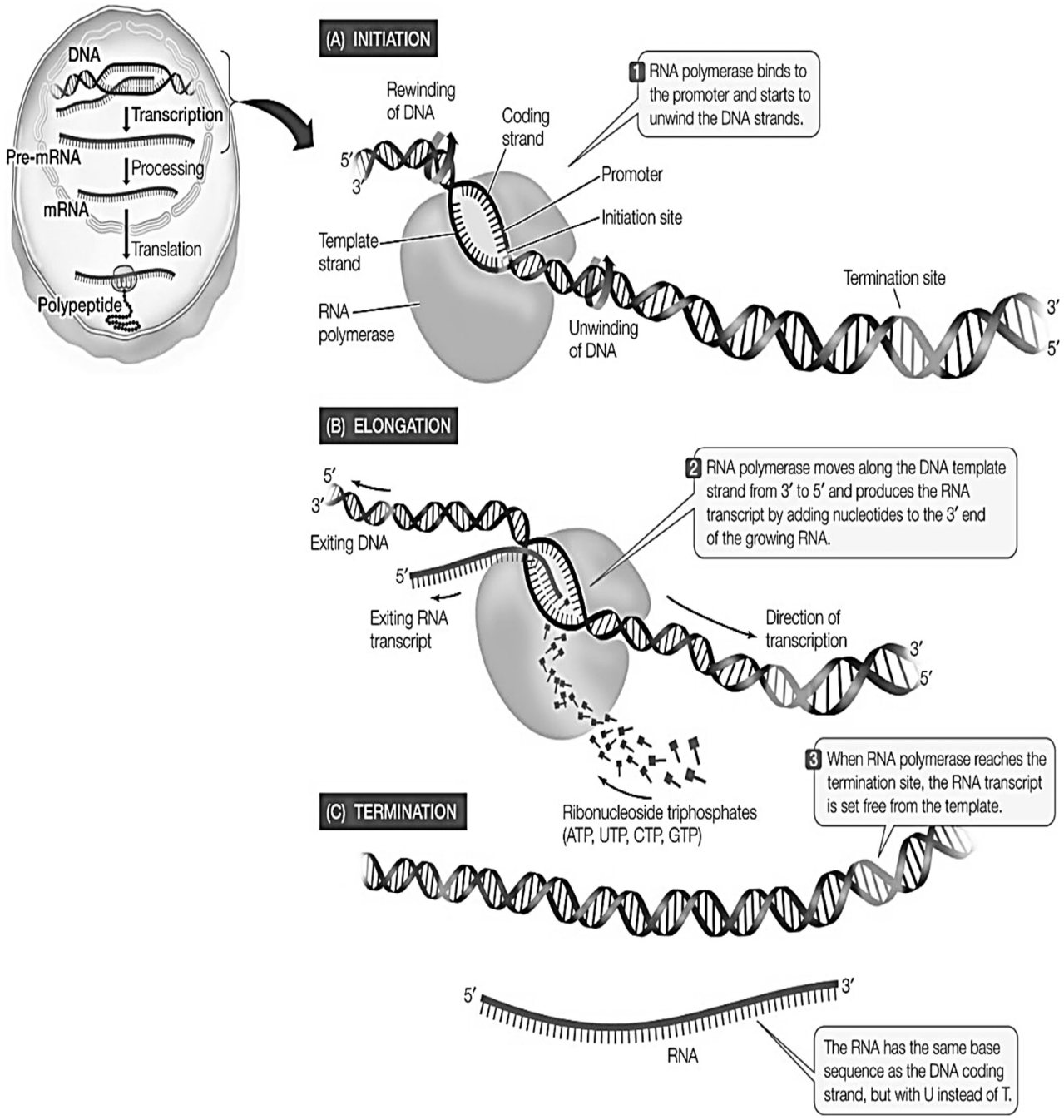


Fig. 3.1.2: Mechanisms of Transcription, e.g. synthesis of messenger RNA (mRNA)

Eukaryote	Prokaryote
Initiation	

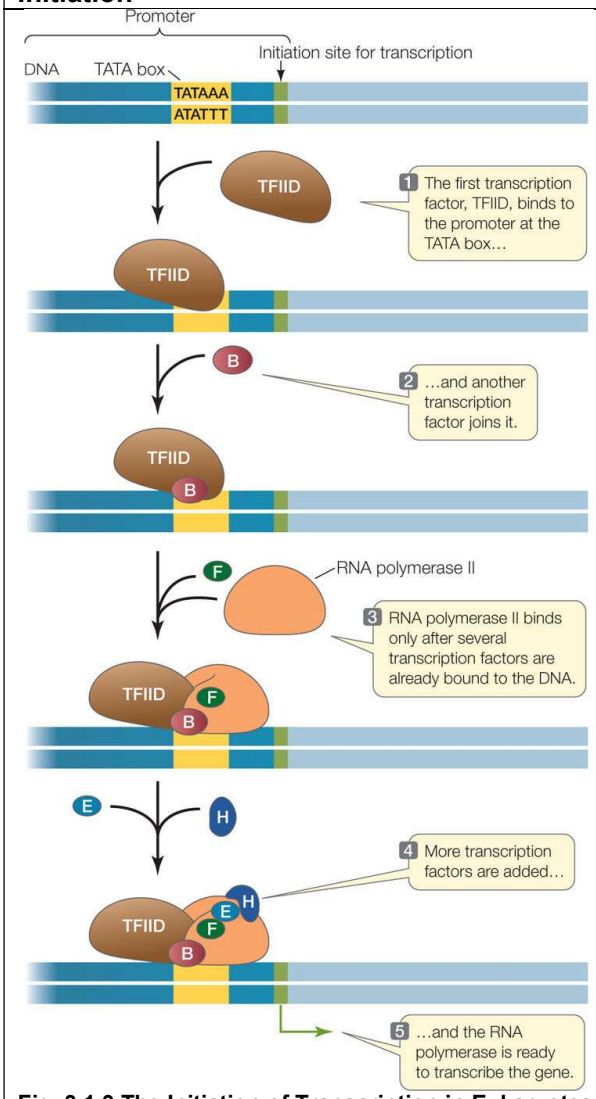


Fig. 3.1.3 The Initiation of Transcription in Eukaryotes.

- General transcription factor, TFIID, recognises and **binds to TATA box** and recruits other **general transcription factors**.
- The assembly of general transcription factors at the TATA box **recruits RNA polymerase** to bind and **form the transcription initiation complex**.

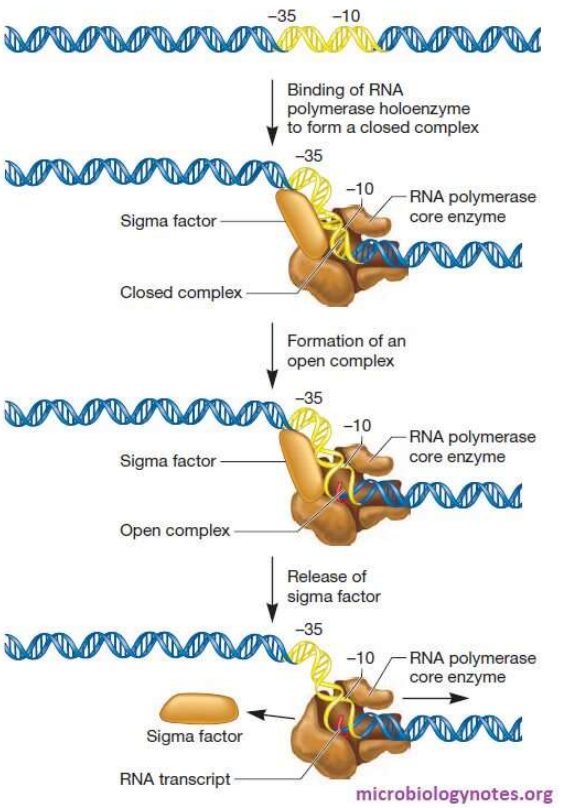


Fig. 3.1.4 The Initiation of Transcription in Prokaryote.

- **Sigma factor** associates with **core RNA polymerase** forming the RNA polymerase holoenzyme.
- As the holoenzyme scans along the DNA, its sigma factor recognizes and binds to the promoter.
- ***NOTE: No transcription factors are involved in transcription initiation in prokaryotes.***

- RNA polymerase **unwinds and unzips** the DNA, by breaking the hydrogen bonds between the complementary base pairs, to **separate the strands of DNA**.
- Only the **template / non coding strand (3' → 5' strand)** is **transcribed**. The other strand not used for transcription is known as the **non-template / coding strand**.

Elongation

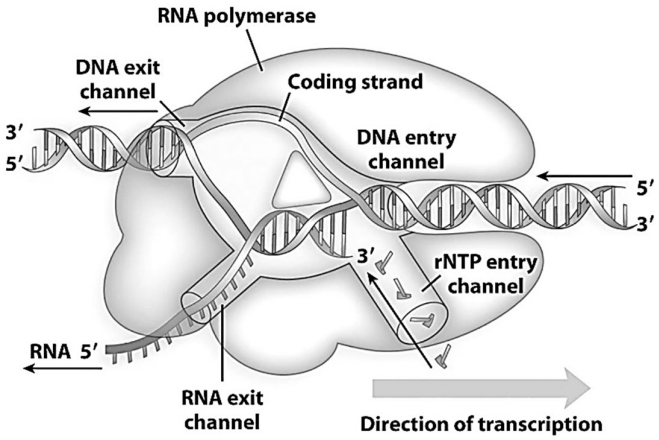


Fig. 3.1.5: Elongation of RNA in the 5' to 3' direction.

- RNA polymerase **reads** the template strand in the **3' → 5' direction**.
- RNA polymerase then selects free **ribonucleoside triphosphates** (rNTPs) from the nucleoplasm with bases that are **complementary** to the DNA template strand.
- The **pre-mRNA strand** is **synthesized** in the **5' → 3' direction**.

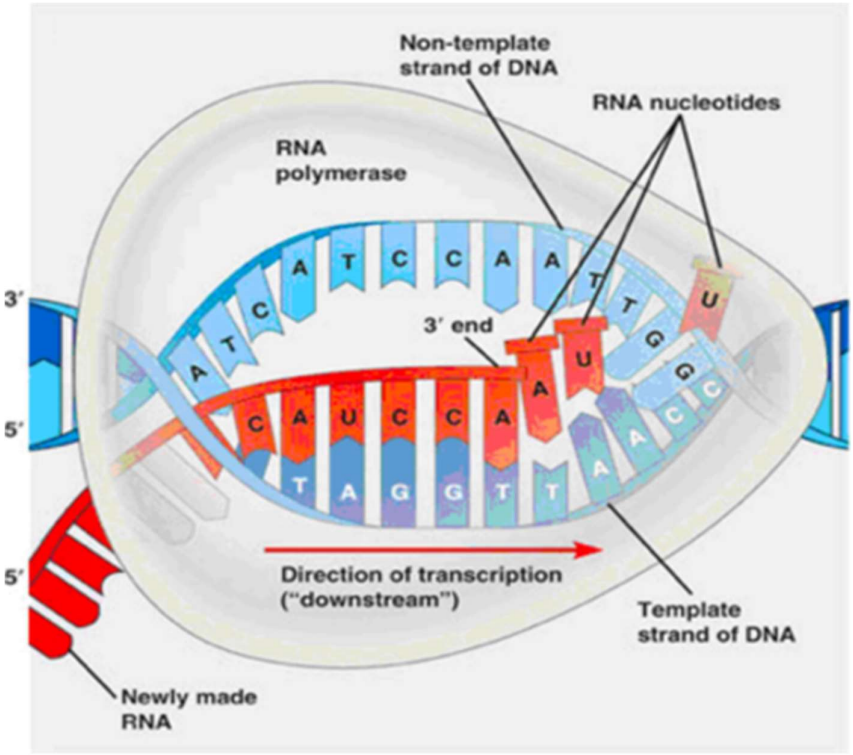


Fig. 3.1.6: Elongation of RNA using RNA nucleotides.

- As each ribonucleoside triphosphate is brought in, **2 of its terminal phosphate groups are removed** to **release energy** for formation of a **phosphodiester bond** between **adjacent ribonucleotides** (3' OH of the growing RNA strand and the 5' phosphate group of the incoming ribonucleotide) via **condensation**.

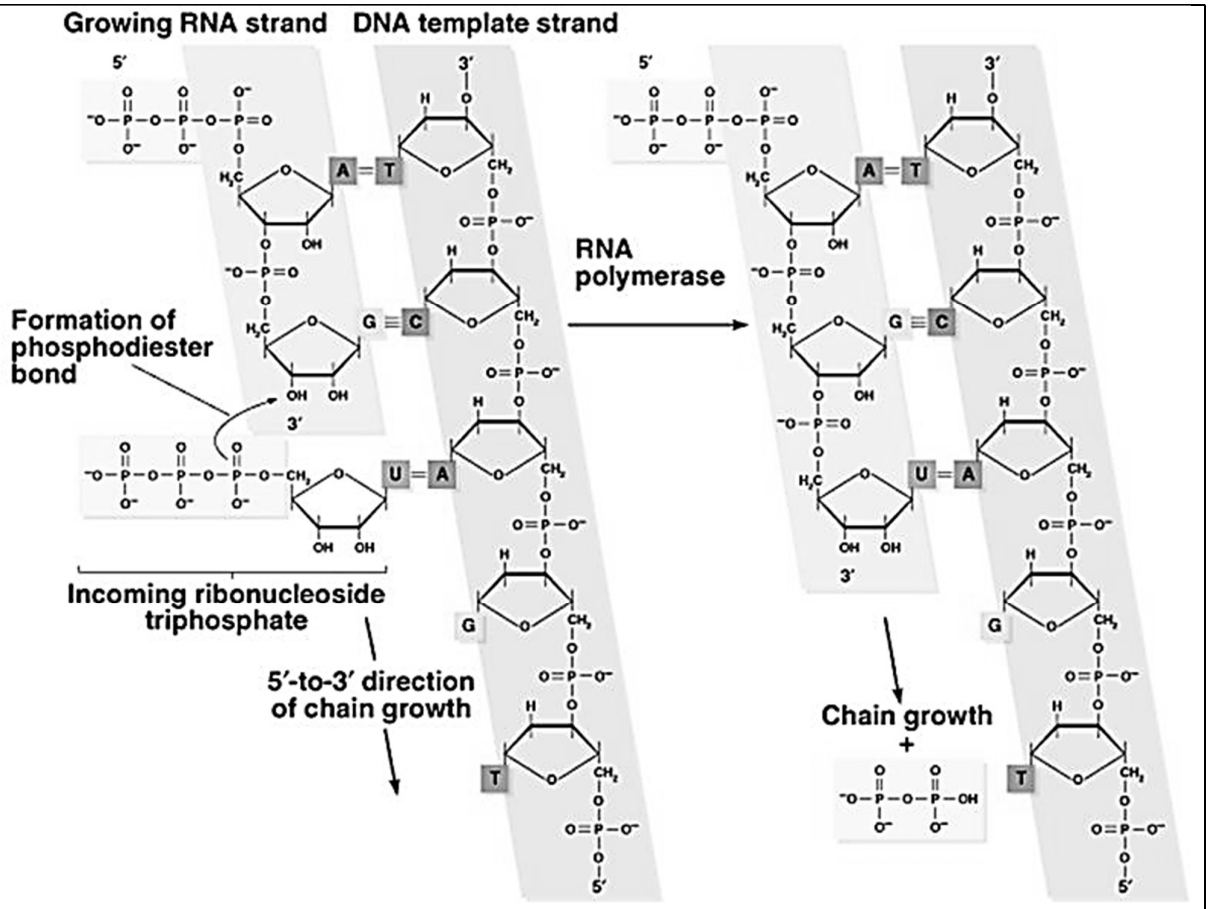


Fig. 3.1.7: Ribonucleoside triphosphate hydrolysing to form ribonucleotides for elongation.

- This reaction is catalysed by RNA polymerase.

Termination

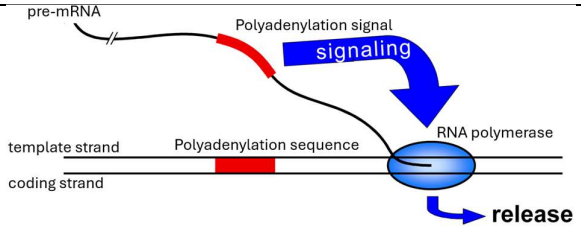


Fig. 3.1.8: Termination via polyadenylation seq.

- In **eukaryotes**, termination occurs **after** the **polyadenylation sequence** is transcribed. The polyadenylation signal sequence found on the DNA template strand. It codes for **AAUAAA** in the pre-mRNA, known as **polyadenylation signal**.

- There are two major termination strategies found in prokaryotes: Rho-dependent and Rho-independent.

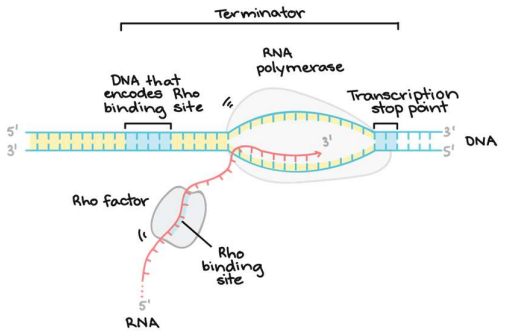


Fig. 3.1.9: Rho-dependent termination.

- **Rho factor binds to Rho binding site on mRNA** and translocates towards RNA polymerase.

• **RNA polymerase continues to transcribe past the polyadenylation sequence** before the **pre-mRNA will then be cleaved** in a region at about **10-35 nucleotides downstream from the polyadenylation sequence**. • The pre-mRNA will be **released**, **RNA polymerase dissociates** from the DNA template and the **DNA renneals**.	• **RNA polymerase reads the termination sequence** and **stop transcribing**. This will allows the **Rho factor to catch up with the RNA polymerase**. • **Rho factor pulls the RNA transcript and the template DNA strand apart**, releasing the RNA molecule and **ending transcription**.

In a nutshell, these are the biomolecules that are involved in the transcription in eukaryotic and prokaryotic cells.

Eukaryote	**Prokaryote**
• **General transcription factors:** also known as basal transcriptional factors, are a class of protein transcription factors that bind to specific sites (promoter) on DNA to activate transcription. • **RNA polymerase:** is the enzyme that catalyzes the formation of phosphodiester bonds, polymerizing a new RNA molecule exactly complementary to the strand of DNA used as template. • **Ribonucleoside triphosphate:** nucleoside containing a nitrogenous base bound to a 5-carbon ribose sugar with three phosphate groups bound to the ribose. • **Ribonucleotide**: A nucleotide in which a purine or pyrimidine base is linked to a ribose molecule.	• **Sigma factor:** prokaryotic proteins that enable specific binding of RNA polymerase to promoter regions within the DNA. • **RNA polymerase:** is the enzyme that catalyzes the formation of phosphodiester bonds, polymerizing a new RNA molecule exactly complementary to the strand of DNA used as template. • **Ribonucleoside triphosphate:** nucleoside containing a nitrogenous base bound to a 5-carbon ribose sugar with three phosphate groups bound to the ribose. • **Ribonucleotide**: A nucleotide in which a purine or pyrimidine base is linked to a ribose molecule. • **Rho factor:** a protein that acts in bacterial cells to mediate termination of transcription in prokaryote.

Checkpoint 3

Identify whether the statements below are true or false by circling either T / F and provide your reasons for your answers.

S/N	Statement	Choice	Reasons
1.	The enzyme involved in transcription is a DNA polymerase.	T / F	
2.	During initiation, the enzyme binds to the origin of replication and unwinds the DNA helix.	T / F	
3.	The direction of elongation of the RNA is in the direction of 3' → 5'.	T / F	
4.	Transcription synthesizes tRNA.	T / F	
5.	The polyadenylation sequence AAUAAA is found on the RNA synthesized.	T / F	

3.2 Post-Transcriptional Modification

Learning Outcome 2(c) modified:

Describe how the information on DNA is used to synthesise polypeptides in prokaryotes and eukaryotes. (Description of the processes of **formation of mRNA from pre-mRNA**).

- In **prokaryotes**, **transcription and translation is simultaneous** (i.e. translation starts as transcription continues) and takes place in the cytoplasm. This is because there is no nucleus to separate both processes in prokaryotes.

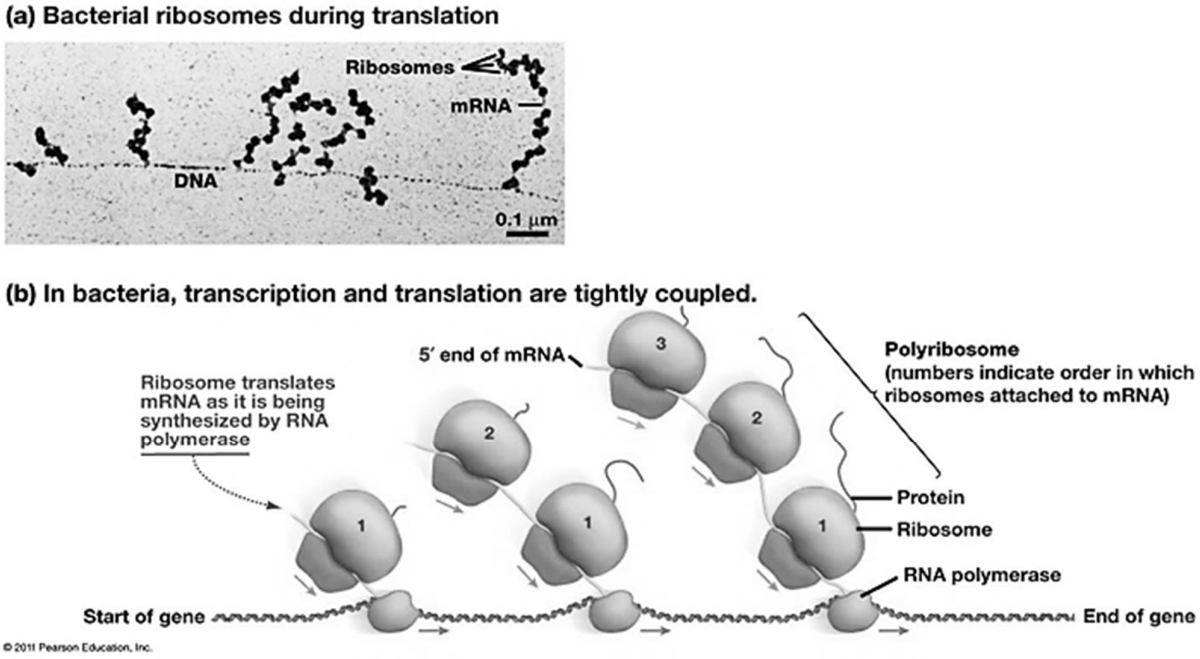


Fig. 3.2.1: Post-trancriptional RNA processing in prokaryotes

- In **eukaryotes**, **transcription and translation is spatially and temporally separated** due to the nucleus**.** This means that transcription has to complete in the nucleus before the **mature mRNA** is exported out of the nucleus to the cytoplasm for translation to initiate the synthesis of a polypeptide.

- After transcription, the newly synthesized mRNA is known as **pre-mRNA**. pre-mRNA must undergo **post-transciptional modification** processes to form **mature mRNA** which then leaves the nucleus via the nuclear pores for translation into proteins in the cytoplasm by the ribosomes.

The **post-transcriptional modification** includes:

i. **5' capping** is a process, catalysed by guanylyl transferase, where a 7-methylguanosine nucleotide is added to the 5' end of the pre-mRNA forming a 5' cap.

ii. **3' polyadenylation** is a process, catalysed by Poly (A) polymerase, where 50 to 250 ribonucleotides with adenine are added to the 3' end of the pre-mRNA forming a 3' poly(A) tail.

Functions of the 5' cap and 3' poly (A) tail:

- Confer stability to the mature mRNA and protects it from enzymatic degradation.
- Facilitate export of the mature mRNA from the nucleus into cytoplasm.
- Recruit and promote attachment of ribosomes to the 5' end of the mature mRNA.

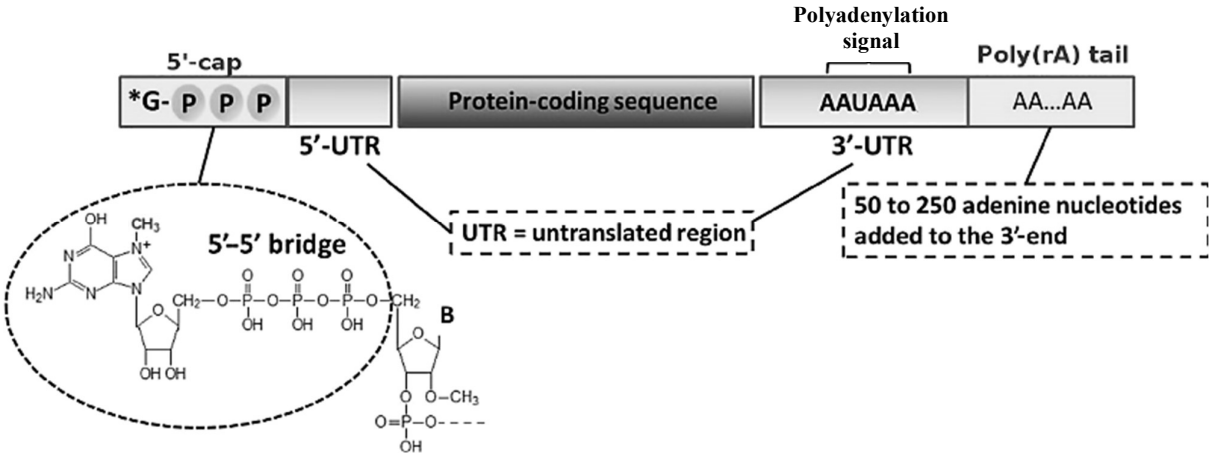


Fig. 3.2.2: RNA processing at the ends of the transcript

iii. Splicing of pre-mRNA to join exons and remove introns to produce the mature mRNA

Different types of RNA employ different modes of splicing. The ribonucleotide sequences at the 5' and 3' ends of introns act as splice sites where splicing recognition take place.

RNA splicing can be catalysed by a **spliceosome** formed from the assembly of several different small nuclear ribonucleoproteins (snRNPs, *pronounced as "snurps"*) together with proteins in the nucleus.

The spliceosome will break the phosphodiester bonds between the ribonucleotides, release the introns for degradation and joins the exons together by forming phosphodiester bonds.

The spliceosome then dissociates to release the **mature mRNA** containing **only the exons**.

In some organisms, RNA splicing can occur without proteins. The intron RNA functions as a ribozyme and catalyses its own excision of introns.

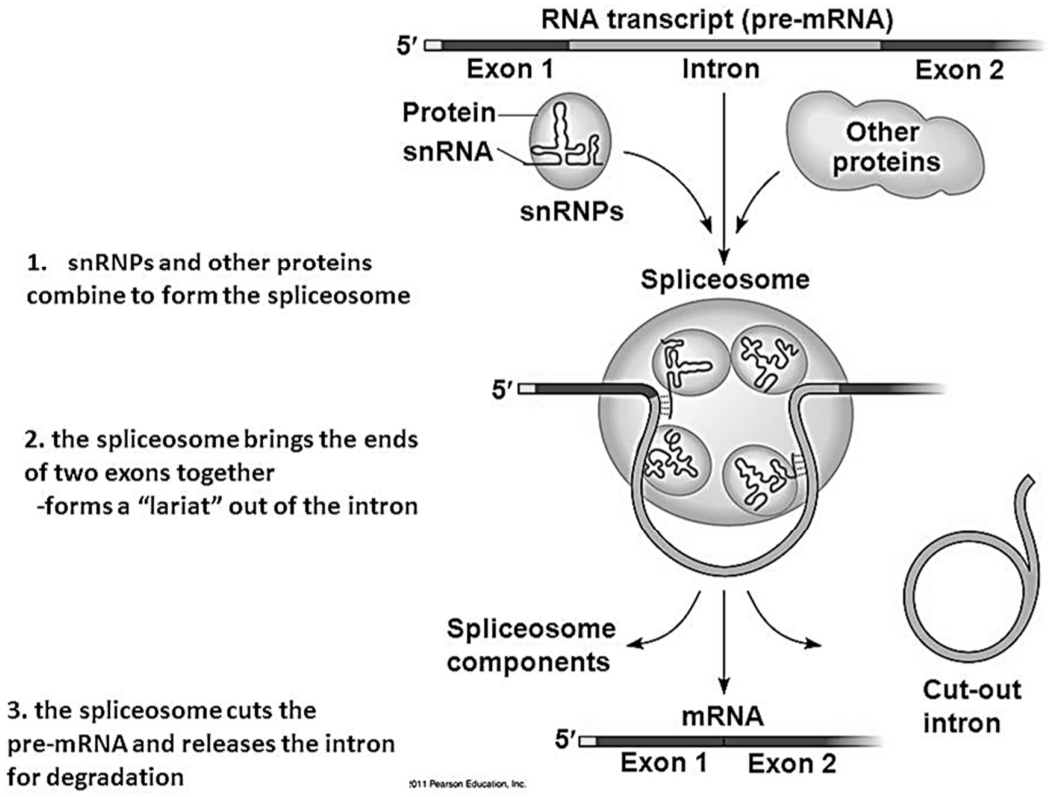


Fig. 3.2.3: RNA Splicing mediated by spliceosome

4. Importance of regulatiing gene expression

Learning Outcome (j) (modified):

Explain why eukaryotes carry out differential (i.e. spatial and temporal) gene expression.

- Gene regulation is the process used **to control the timing, location and amount in which genes are expressed**.

- Differential gene expression, which refers to the spatial (tissue-specific) and temporal (stage-specific) regulation of gene activity, is **crucial for the proper development and functioning of eukaryotic organisms**.

- **Spatial Gene Expression**:
    - **Cell differentiation**: Different cell types within an organism, such as muscle cells, nerve cells, and skin cells, have distinct functions and structures. This is achieved by **expressing different sets of genes in each cell type**, a process known as cell differentiation.

    - **Tissue formation**: During embryonic development, spatial gene expression patterns **guide the formation of various tissues and organs**. Specific genes are expressed in particular regions, leading to the **development of specialized structures** like the heart, brain, and limbs.

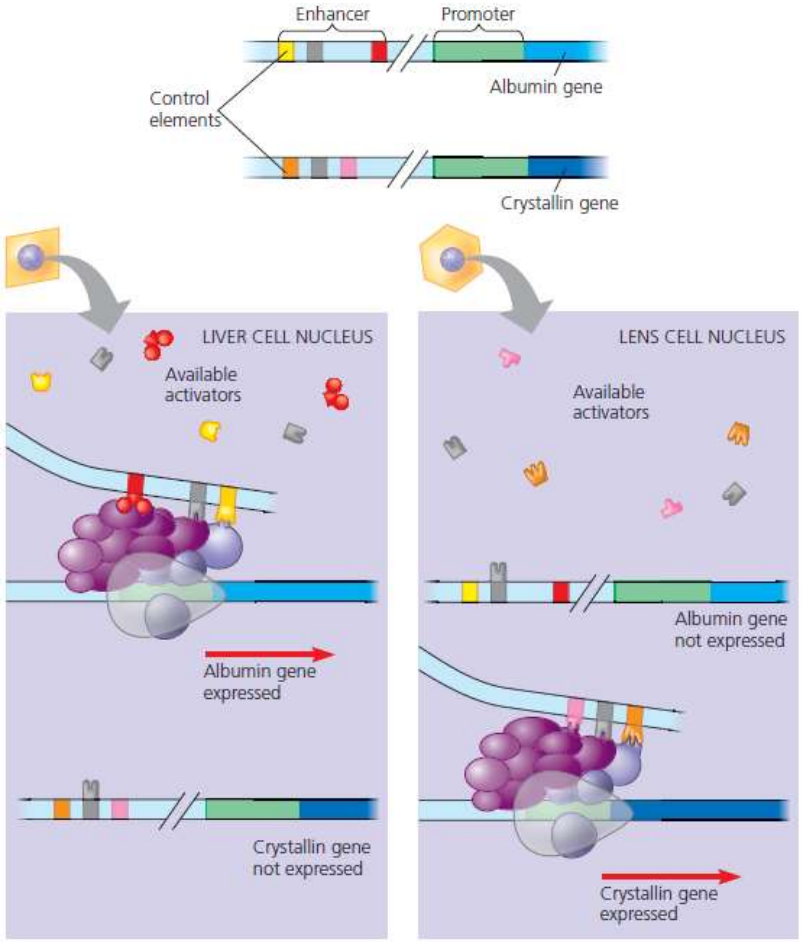


Fig. 4.1: Cell type-specific transcription.

- **Organ function**: Within organs, different cell types **express specific genes to carry out specialized functions**. For example, both liver cells and lens cells have the genes for making the proteins albumin and crystalline, but only liver cells make albumin (a blood protein) and only lens cells make crystalline (the main protein of lenses)(Fig. 4.1).

- **Temporal Gene Expression**:
  - **Developmental stages**: Gene expression patterns change dynamically during different stages of development, such as embryogenesis, larval stages, and adulthood. This temporal regulation **ensures that the right genes are expressed at the right time**, enabling proper growth and maturation.

  - **Cellular processes**: Within cells, genes are expressed in a temporal manner to **coordinate various cellular processes** like cell division, differentiation, and apoptosis (programmed cell death).

  - **Environmental responses:** Temporal gene expression allows organisms to **adapt to changing environmental conditions**, such as temperature, light, or nutrient availability, by activating or repressing specific genes at the appropriate times.

- Gene expression can be **<u>regulated at various levels</u>**:

  (i) Chromatin level;
  (ii) Transcriptional level;
  (iii) Post-transcriptional level;
  (iv) Translational level;
  (v) Post-translational level.

  This set of notes will only cover regulation at (i) chromatin level, (ii) transcription level and (iii) post-transcription level. (iv) Translation level and (v) post-translation level will be covered in the Gene Expression Part II notes.

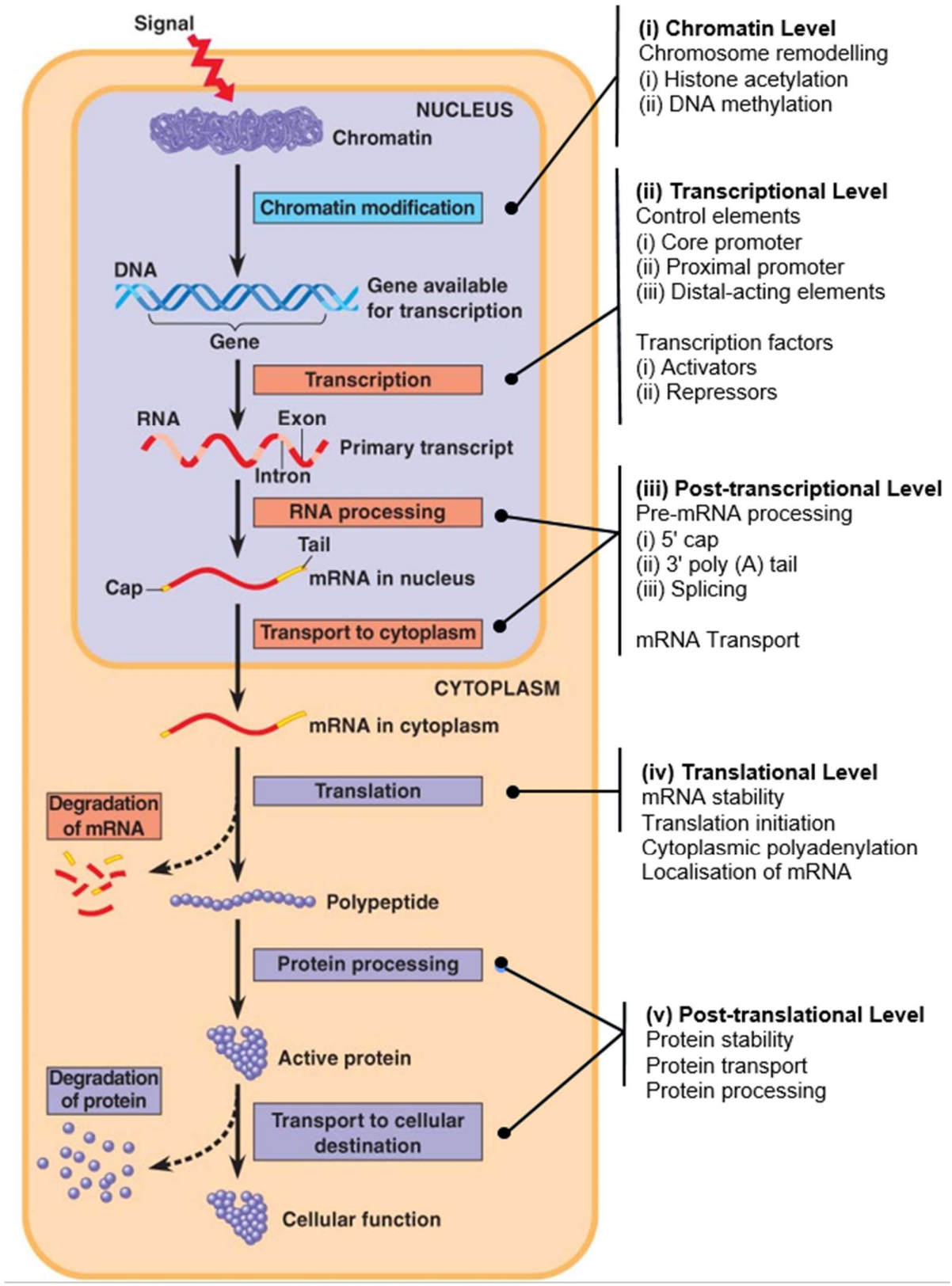


Fig 4.1: Regulation of eukaryotic gene expression at various levels

Learning Outcome (j):

Explain how differential (i.e. spatial and temporal) gene expression in eukaryotes can be regulated at different levels:

i. Chromatin level (histone modification and DNA methylation)

4.1 Regulation at Chromatin Level

- Chromatin level regulation alters the structure of chromatin via **histone modification (acetylation)** and **DNA methylation**.

- Organization of DNA into chromatin helps to:
  1. pack DNA into a compact form that fits inside the nucleus of a cell.
  2. regulate gene expression - physical state of DNA in / near a gene determines if the gene is **accessible** for transcription.

4.1.1 Histone modification

- DNA is packed into chromatin by associating with **histone proteins**. Histone modification (histone acetylation) is a modification to the histone protein which can impact gene expression by **altering chromatin structure**.

- The compactness of chromatin structure influences the access of transcription machinery (RNA polymerase and transcription factors) to the region of DNA, thus **regulating the rate of transcription**.

- DNA in **heterochromatic** regions are **highly** compacted (more condensed) and usually **transcriptionally inactive** as RNA polymerase and transcription factors cannot access and bind to the promoter. Hence, this **prevents** the formation of the **transcription initiation complex**.

- DNA in **euchromatic** regions are more **loosely** packed (less condensed) and usually **transcriptionally active** as RNA polymerase and transcription factors is able to access the promoter site on the DNA. Hence, this **allows** the formation of the **transcription initiation complex**.

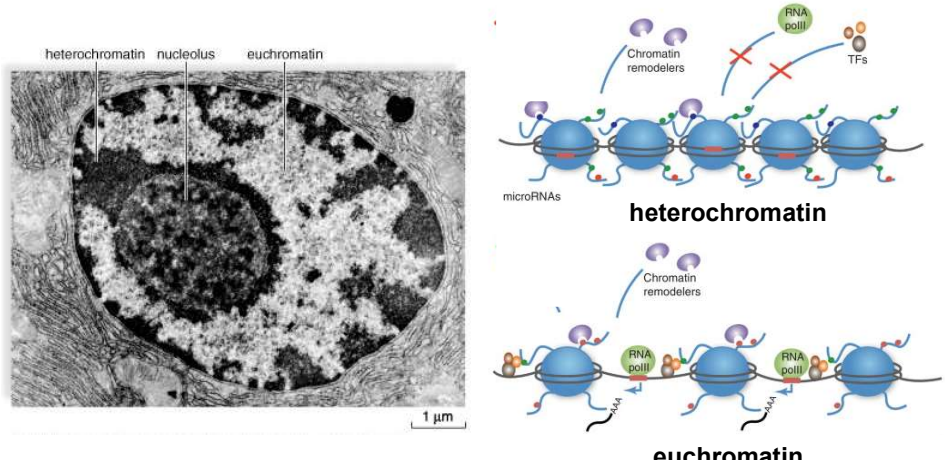


Fig 4.1.1.1: Structure of heterochromatin and euchromatin

- **Histone acetylation** involves adding an acetyl group (–$COCH_3$) to lysine residues (positively-charged amino acid) at the N-terminal of **histone tails**.

- This reaction is catalysed by **histone acetyl transferase** (HATs) which reduces the positive charge of histone.

- **Reduction of positive charge of histone** reduces the affinity of histones for negatively-charged phosphate group in DNA.

- This results in a **more relaxed chromosome conformation** allowing access of RNA polymerase and transcription factors to promoter site thus **increasing transcription rate**.

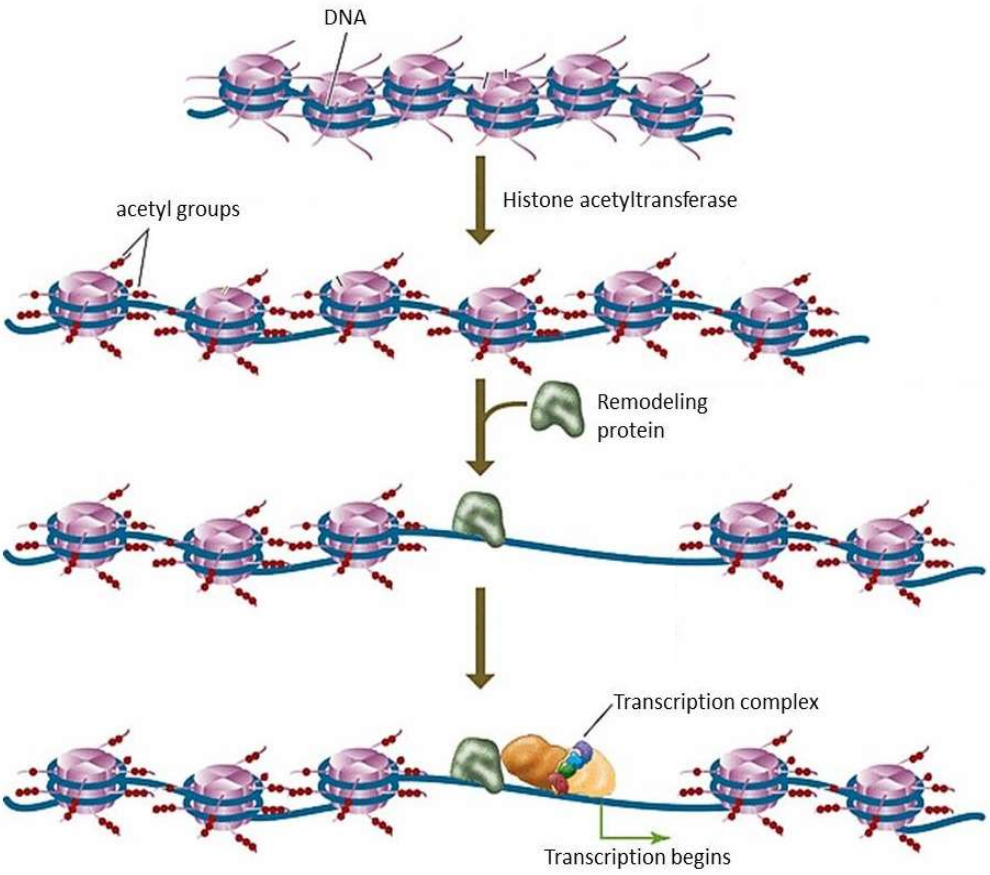


Fig 4.1.1.2: Histone acetylation activating transcription

- **Deacetylation** of histone involves removal of an acetyl group (–$COCH_3$) from lysine residues at the N-terminal of **histone tails**.

- This reaction which is catalysed by **histone deacetylases** (HDACs) makes the DNA to coil more tightly around the histone, making it harder for the RNA polymerase and transcription factors to access the promoter site thus **decreasing transcription rate**.

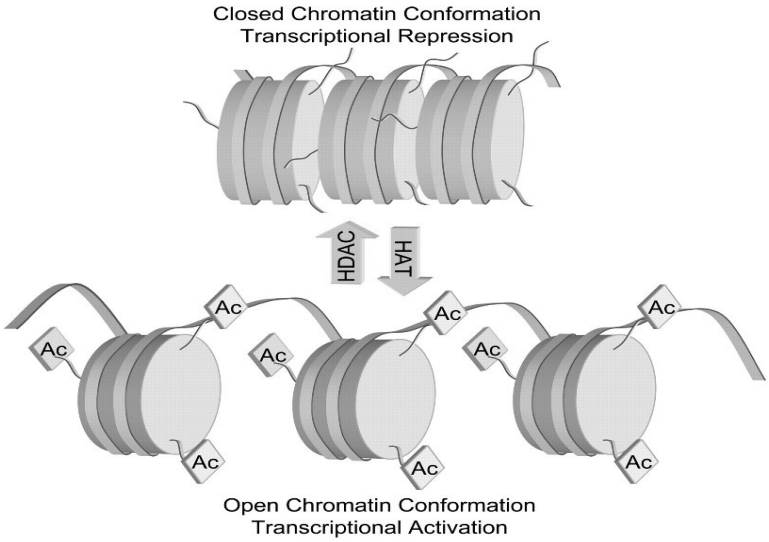


Fig 4.1.1.3: Histone Acetylation and deacetylation regulating gene expression

4.1.2 DNA Methylation

- DNA methylation involves the **covalent addition** of a methyl group (-CH_3) to the DNA. This reaction is catalysed by **DNA methyltransferase** which results in conversion of cytosine bases (nucleotide) in the CpG islands to 5-methyl cytosine.

  Note: CpG Island is a non-coding region in the DNA rich with C and G sequences which are usually near the transcription start sites.

- Methylation of CpG islands **decreases transcription rates of genes** in their proximity. Non-methylated CpGs are found in high density **near promoters of actively transcribed genes** like housekeeping genes.

- DNA methylation can impact the transcription of genes in two ways:

  1. The addition of the methyl group may **physically block the binding of transcriptional proteins to the promoter**.

     - Methylation changes the shape of the cytosine residues.
     - Methylated cytosine residues are not complementary in shape and charge to the binding sites of transcription proteins.

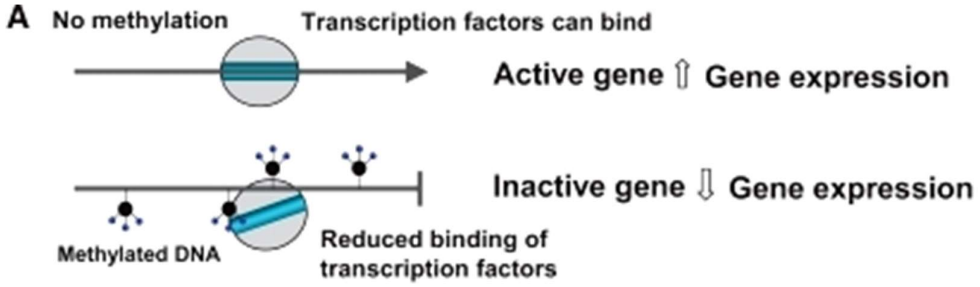


Fig 4.1.2.1: Effects of DNA methylation in regulating gene expression

2. Methylated DNA may be bound by **methyl-CpG-binding proteins**. These proteins then **recruit additional proteins** that **suppress transcription.**

   - For example, a methyl-CpG-binding protein may recruit a **histone deacetylases** and other chromatin remodelling proteins that can modify histones, resulting in the formation of **closed conformation of chromatin (heterochromatin)** that is **transcriptionally inactive**.

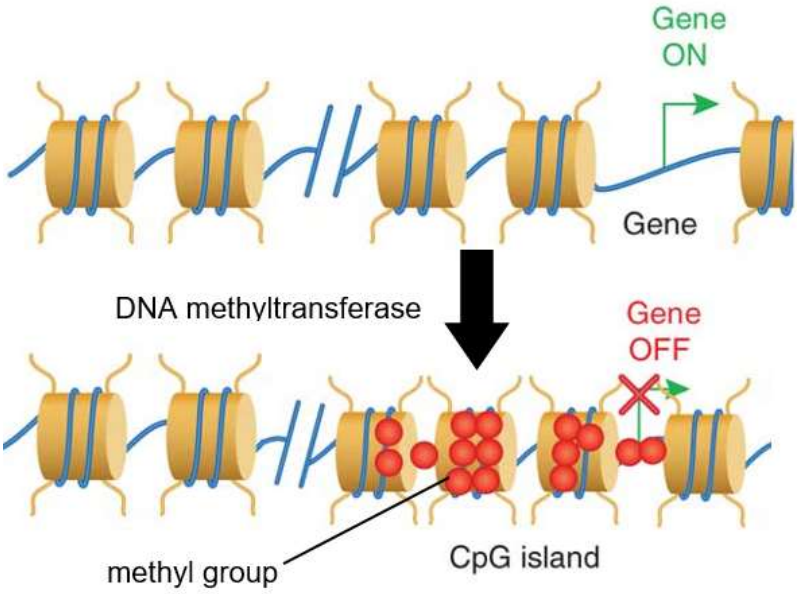
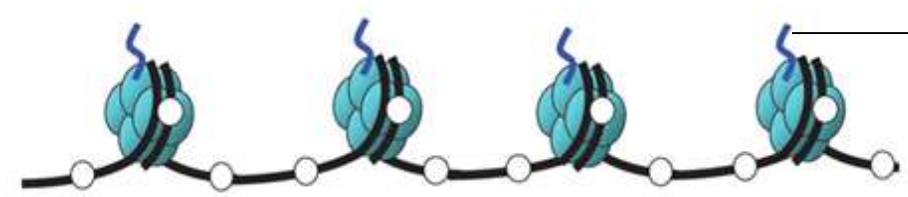


Fig 4.1.2.2: DNA methylation regulating gene expression

CHECKPOINT 4

Fill in the blanks to summarise how chromosome remodelling influences transcription.

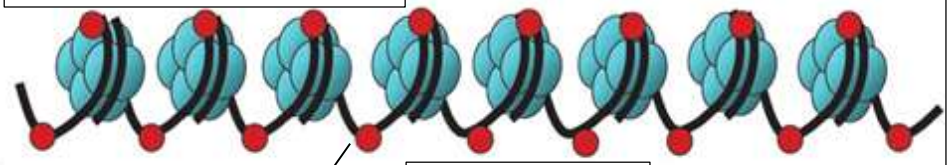


_____ lysine on N-terminal of histone tails

enzymes involved:

enzyme involved:

results in: _____ chromosome conformation leading to _____ transcription rate



_____ cytosine of CpG islands in DNA

results in: _____ coiled around histone leading to _____ transcription rate

Learning Outcome (j):

Explain how differential (i.e. spatial and temporal) gene expression in eukaryotes can be regulated at different levels:

ii. Transcriptional level (control elements, such as promoters, silencers and enhancers, and proteins, such as transcription factors, including activators and repressors)

4.2 Regulation at Transcription Level

- Gene expression can be regulated by influencing the rate of transcription initiation. Transcription initiation is greatly dependent on the assembly of the **transcription initiation complex** which consists of RNA polymerase and various transcription factors.

- Specific proteins binding to **control elements** (regulatory DNA sequences such as promoters, enhancers and silencers) can influence the formation and stability of the **transcription initiation complex.**

- Control elements are **regulatory DNA sequences** located on the **chromosome** which includes **promoters** and **distal acting elements such as silencers and enhancers**.

4.2.1 Control Elements

Refer to section 1.4.4 to section 1.4.6 for the structure and function of promoters, enhancers and silencers.

(i) Promoter

- Core promoters are **non-coding DNA sequences** found **upstream** of the gene that serve as recognition sites for **transcription factors** and **RNA polymerase** binding and directs the **initiation of transcription**.

- Core promoter contains the **TATA box**. TATA binding protein (TBP) recognizes and binds to the TATA box and helps in the subsequent binding of other general transcription factors and RNA polymerase to initiate transcription.

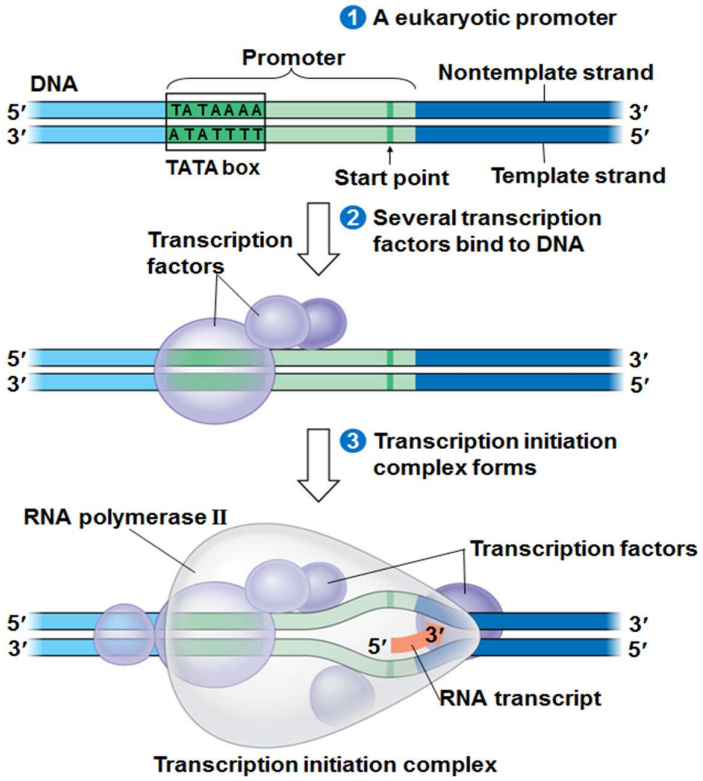


Fig 4.2.1.1: Function of the core promoter in regulating transcription

(ii) Proximal Promoters

- Proximal promoters are **non-coding DNA sequences** found **further upstream** of **core promoters**. They are bound by various **general transcription factors** to promote transcription of downstream genes.

- Proximal promoters contain **CAAT box** consisting of GGCCAATCT sequence usually located at -70 to -80. It is a **strong promoter** often associated with genes that encode proteins essential to normal cellular functioning i.e. **housekeeping genes**.

  Housekeeping genes are defined as being stably expressed in all cells and conditions, essential, belong to cellular maintenance pathways and be conserved.

- Proximal promoters also contains **GC box** consisting of variants of the consensus sequence GGGCGG, often found within 100 base pairs **upstream of the transcriptional initiation site**.

- The CAAT box and GC box help **improve efficiency** of the promoter, by helping to recruit **general transcription factors and RNA polymerase to the core promoter.**

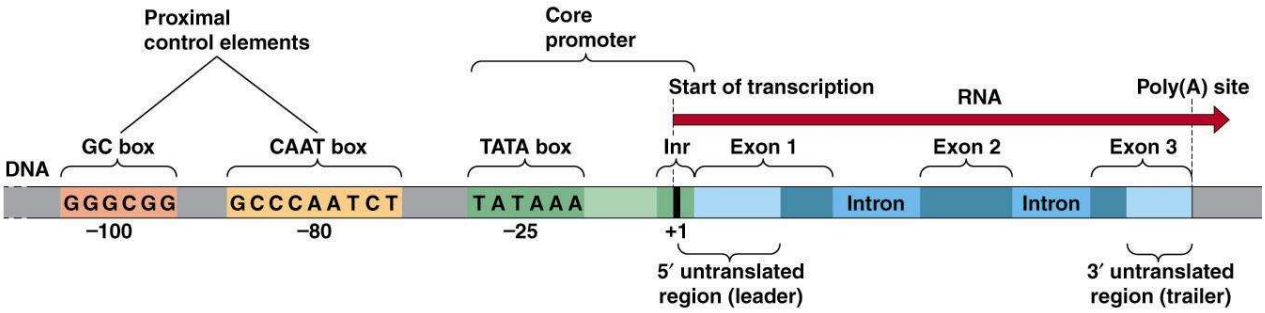


Fig 4.2.1.2: Proximal promoter and core promoter upstream of the gene

(iii) Distal-acting Elements

- Distal-acting elements are **non-coding DNA sequences** that can be **located upstream or downstream of the gene.**

- Distal acting elements are **regulatory DNA sequences** which consist of **enhancers** and **silencers** which are bound by **specific transcription factors**, the **activators** and **repressors** respectively.

Enhancers

- **Enhancers** are regulatory DNA sequences which are required for **maximal rate of transcription of a gene**.

- Enhancers are **positive regulatory elements** involved in the **upregulation** of transcription as they **promote the assembly of the transcription initiation complex** via their interaction with **specific transcription factors** called **activators**.

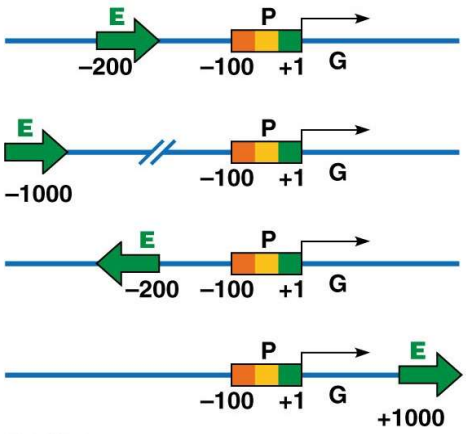


Fig 4.2.1.3: Various positions and orientations of the enhancer

Silencers

- Silencers are **negative regulatory elements** involved in the **downregulation** of transcription as they **prevent the assembly of the transcription initiation complex** via their interaction with **specific transcription factors** called **repressors.**

- **Enhancers and silencers** exhibit **three general properties**:

  I. They **act over relatively large distances** – up to several thousand base pairs from their regulated gene(s);

  II. Their influence on gene expression is **independent on orientation** – they function equally well in either the normal or inverted orientation within the DNA; and

III. Their effects are **independent of position** – they can be **located upstream, downstream** or **within an intron of a gene** and **still have profound effects on the gene's expression.**

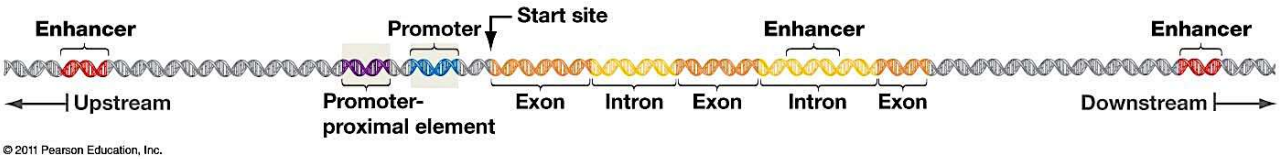


Fig 4.2.1.4: Various positions and orientations of the enhancers

4.2.2 Transcription Factors

- Transcription factors are **gene regulatory proteins** required for transcription. They **bind to control elements** as well as **other transcription factors/ proteins**.

(i) General Transcription Factors

- **General / Basal transcription factors** are essential for the initiation of transcription of all protein-coding genes. The first step in formation of a transcription complex is the **binding of a general transcription factor (TFIID)** to the **TATA box**. They are generally very weak in increasing the efficiency of transcription and usually results in a **low rate of transcription initiation** and few mRNA transcripts are produced.

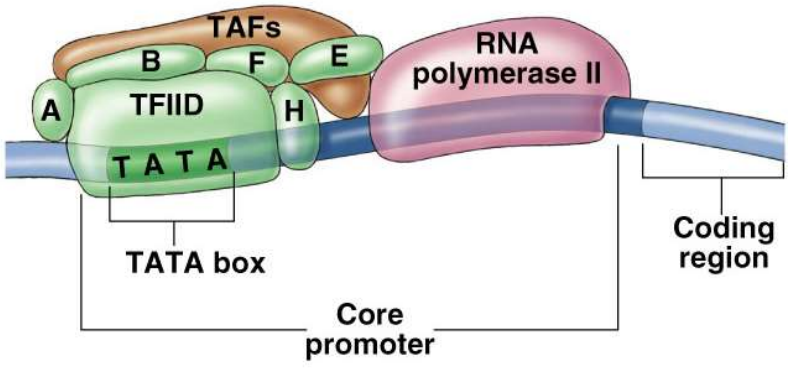


Fig 4.2.2.1: Binding of general transcription factors and RNA polymerase at the core promoter stimulates basal transcription rate.

(ii) Specific Transcription Factors

- Specific transcription factors are regulatory protein molecules consist of **activators** and **repressors** which binds to **enhancers** and **silencers** respectively.

Activators	Repressors
✓ Activators are proteins that bind to enhancers and interacts with basal transcription factors and RNA polymerase to **increase** rate of transcription initiation.	✓ Repressors are proteins that bind to **silencers** to **decrease** the rate of transcription initiation.
✓ Activators have **DNA-binding domain** and **activation domain.** o DNA binding domain is a three-dimensional structure that binds to enhancers in the DNA sequence. o Activation domain interacts with other regulatory proteins or components of transcription machinery (transcription factors and RNA-polymerase) facilitating protein-protein interaction to stimulate transcription.	✓ Repressors have **DNA-binding domain** and **repression domain**. o DNA-binding domain binds to **silencers** in the DNA sequence. o Repression domain interacts with **general transcription factors and activators.**
• Activators bind to enhancers to increase frequency of transcription by 2 ways: o **promoting assembly of transcription initiation complex (TIC)** through **direct interaction** with transcription factors and RNA polymerase (Fig. 4.2.2.2) ✓ Activators can interact with **transcription factor-RNA polymerase complex**, making it more energetically favourable to assemble on a promoter. o **increasing accessibility to the promoter region through chromatin remodeling (Fig. 4.2.2.3)** ✓ Activators can recruit **histone acetyl transferases (HATs)** to change **chromatin structure** at the regulatory sequences and promoters to increase access of transcription factors and RNA polymerases to promoter sites.	✓ Repressors suppress transcription by (Fig. 4.2.2.4): o **Competitive DNA binding** ✓ Repressors compete for binding with activators to the same regulatory DNA sequence (overlap of enhancer and silencer gene regulatory sequences) blocking activator binding to the enhancer. o **Masking activation surface** ✓ Repressors bind to the activation domain of the activator preventing it from binding to other proteins carrying out its activation of transcription. o **Direct interaction with basal transcription factors** ✓ Repressors can interact with transcription factors to block the assembly of the transcription initiation complex.

	○ **Chromatin remodelling** ✓ Repressors can recruit histone deacetylase (HDAC) to change chromatin structure at regulatory sequences and promoters to decrease access of transcription factors and RNA pol to promoter sites. This reduces the transcription rate of the gene.

Mechanism 1: Promoting assembly of transcription initiation complex (TIC) through direct interaction with transcription factors and RNA polymerase

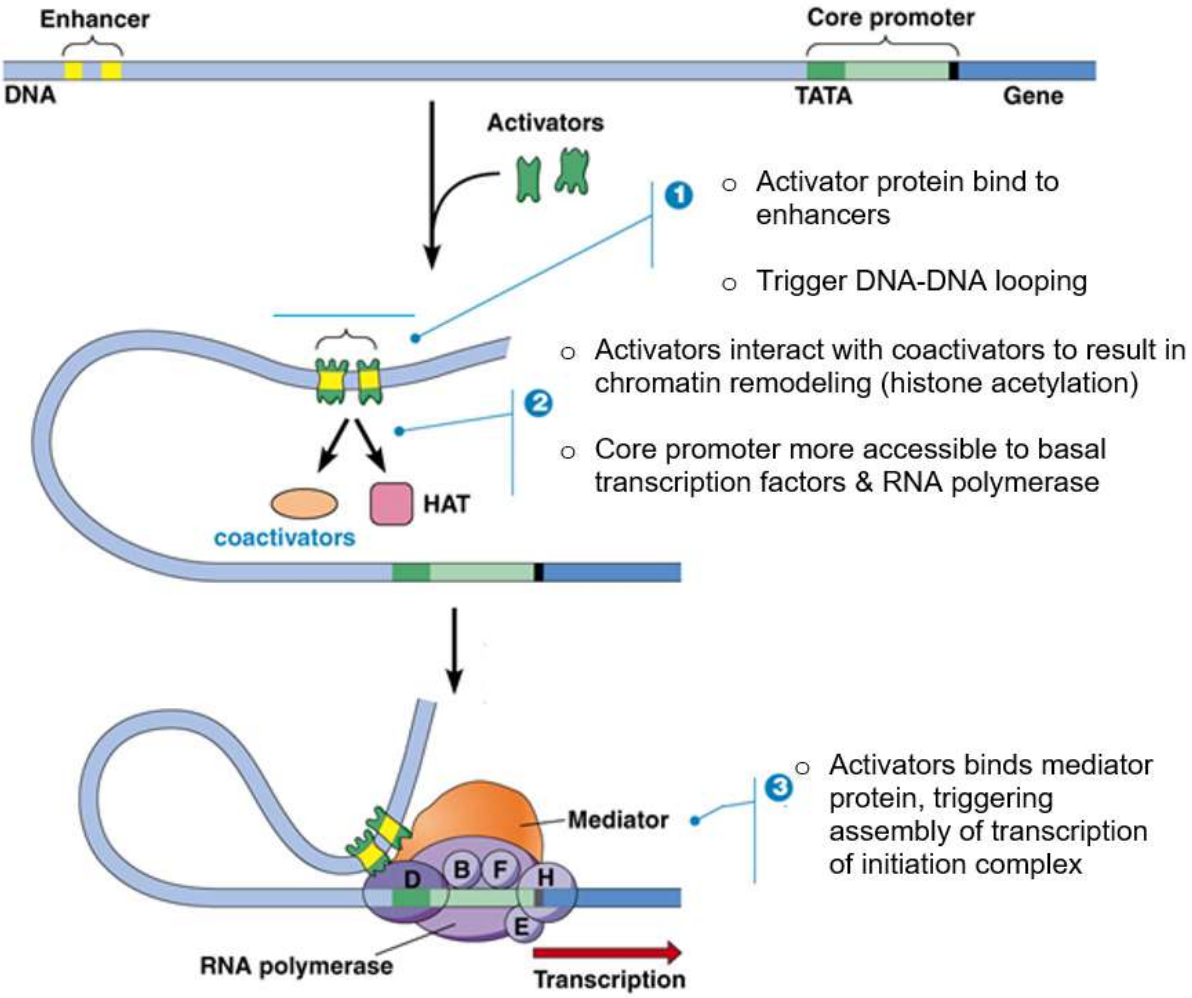


Fig. 4.2.2.2: Activator proteins bind to enhancer region, causing DNA bending to facilitate TIC formation

Mechanism 2: Increasing accessibility to the promoter region through chromosome remodelling

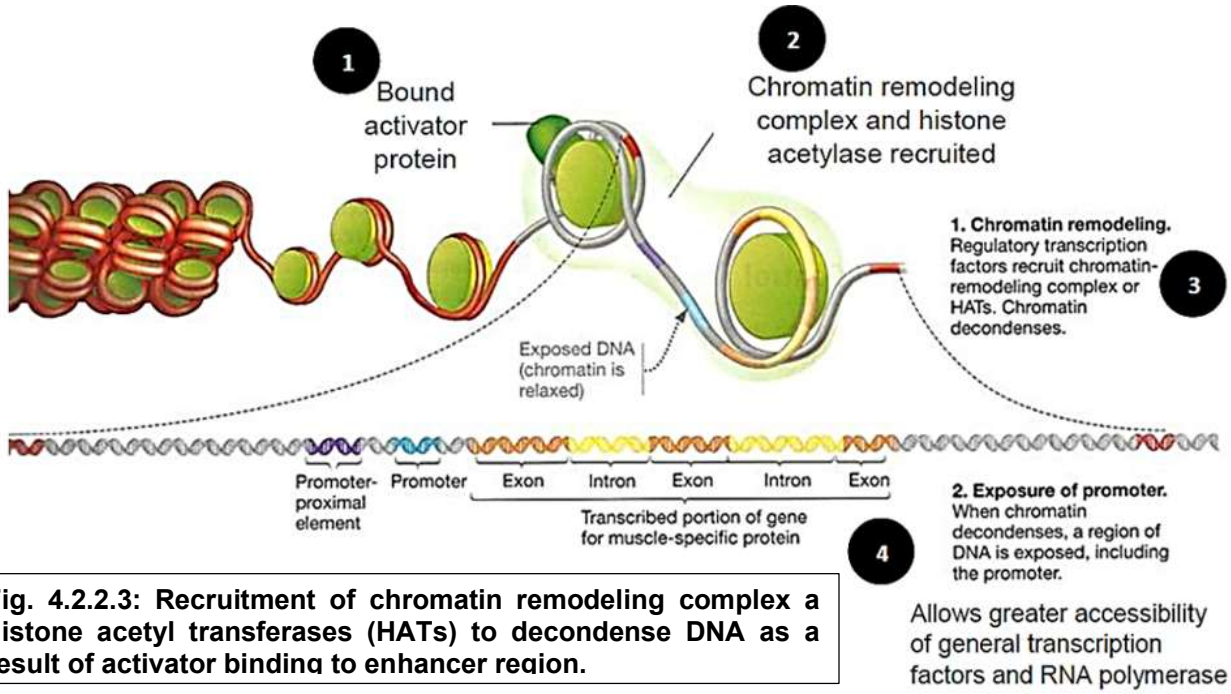


Fig. 4.2.2.3: Recruitment of chromatin remodeling complex a histone acetyl transferases (HATs) to decondense DNA as a result of activator binding to enhancer region.

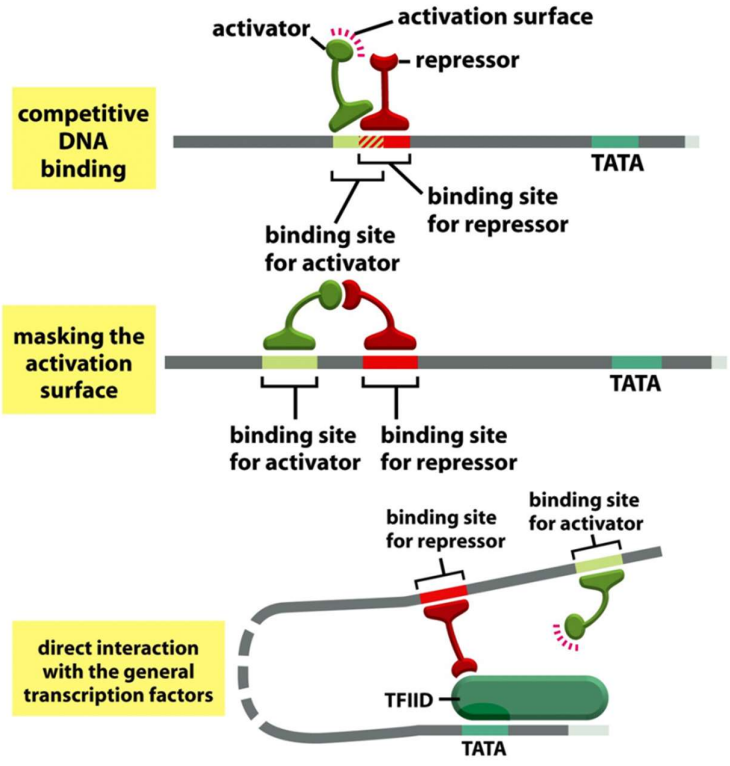
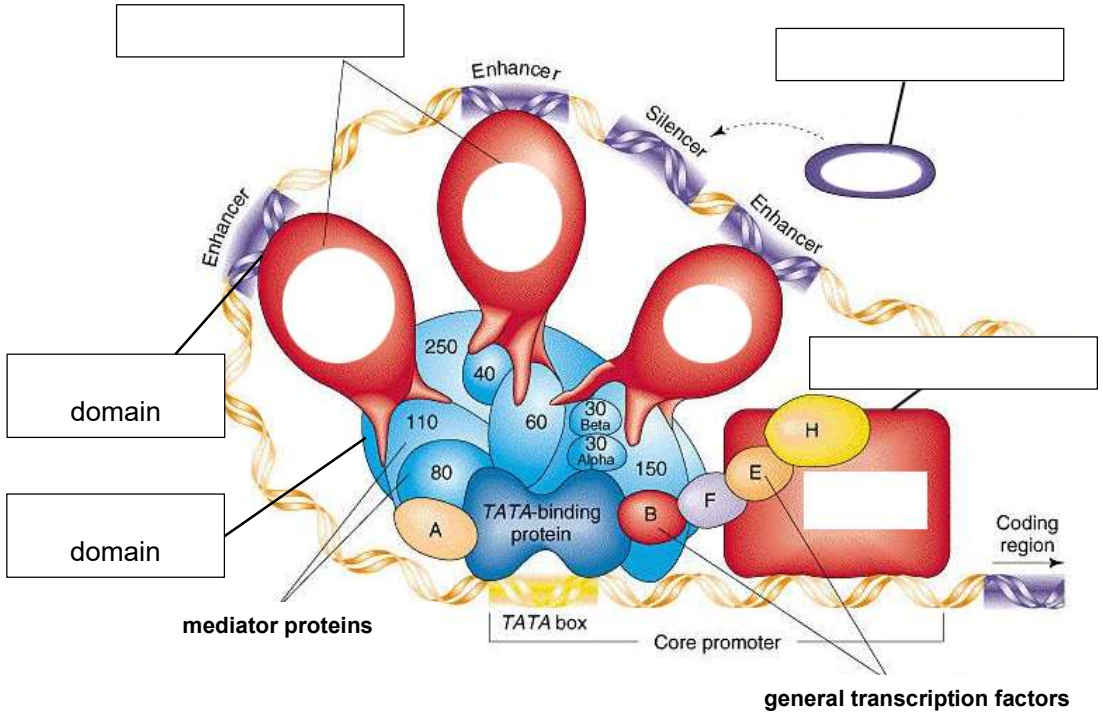


Fig 4.2.2.4: Mode of action of repressor proteins.

CHECKPOINT 5

Label the following structures.



Learning Outcome (j):

Explain how differential (i.e. spatial and temporal) gene expression in eukaryotes can be regulated at different levels:

iii. Post-transcriptional level (processing of pre-mRNA in terms of splicing, polyadenylation and 5' capping)

4.3 Regulation at Post-transcriptional Level

- Regulation at post transcriptional level refers to regulation of gene expression **once a gene has been transcribed to form the pre-mRNA**.

4.3.1 Pre-mRNA Processing (Refer to Section 3.2 for details)

- Transcription and translation in eukaryotes are two processes that are **spatially** and **temporally** separated. mRNA transcripts must move from the nucleus to the cytosol to be translated. The spatial separation of transcription and translation allows for the primary transcripts (pre-mRNA) to undergo **post-transcriptional processing** before becoming **mature mRNAs**.

Processes	Function
(a) 5' Capping	**1.** Confer stability to the mature mRNA and protects it from enzymatic degradation. **2.** Facilitate export of the mature mRNA from the nucleus into cytoplasm. **3.** Recruit and promote attachment of ribosomes to the 5' end of the mature mRNA.
(a) 3' Polyadenylation	
(b) Splicing	**4.** to join exons and remove introns to produce the mature mRNA

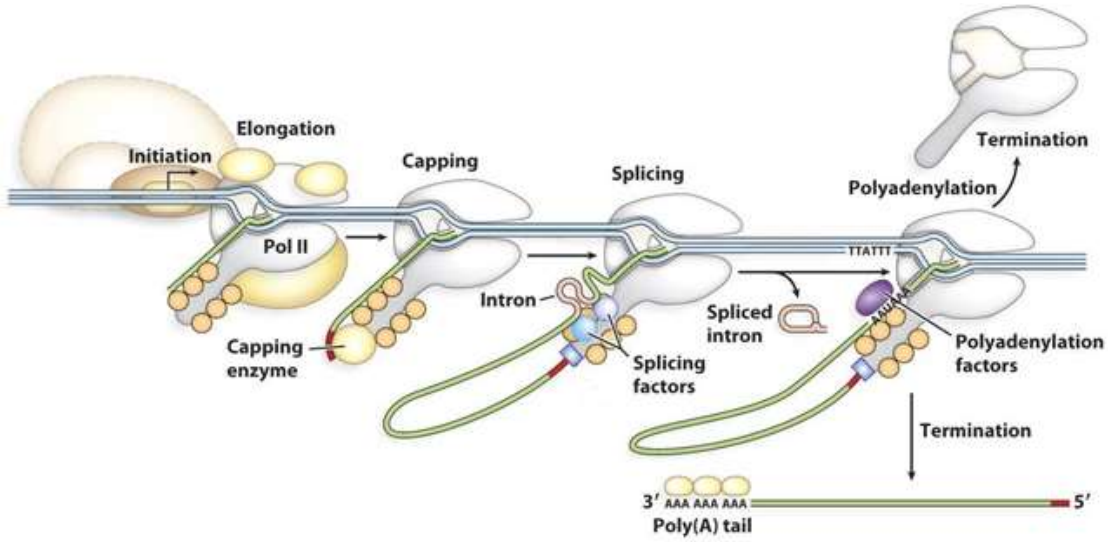


Fig 4.3.1.1: pre-mRNA processing

- mRNA splicing can be classified into 2 types; **<u>constitutive splicing</u>** and **<u>alternative splicing</u>**.

  - ✓ **<u>Constitutive</u>** splicing involves removing of **all introns** and including **<u>every exon.</u>**

  - ✓ **<u>Alternative</u>** splicing removes **all introns** but includes **<u>different exons</u>**.

    - **a single pre-mRNA can produce different mature mRNA**, depending on which **combinations of exons** are spliced together. Regulatory proteins specific to a cell type control intron-exon choices within the primary transcript.

    - The presence of exons within a gene **allows the gene to potentially encode several different polypeptides**. This means that **one gene can code for more than one type of polypeptide**.

    - **Different snRNPs (with different snRNA sequences) recognise different splice sites** and therefore **cut out all introns and splice together different combinations of exons** respectively.

    - This allows different proteins to be produced from the same pre-mRNA, increasing the **<u>diversity of proteins</u>** produced thus **<u>increasing the coding capacity</u>** of the genome. It also offers another level at which regulation of gene expression can occur.

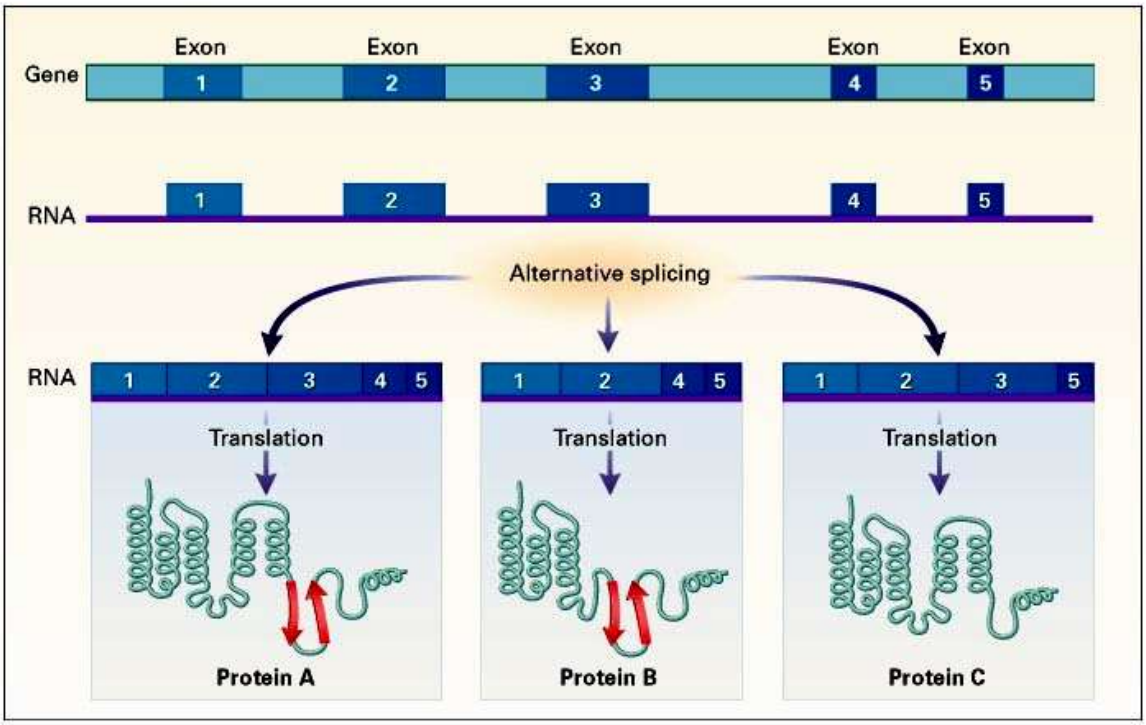


Fig 4.3.1.2: Production of different proteins from the same gene/pre-mRNA

4.3.2 mRNA Transport

- All mRNAs synthesized in the nucleus must be **exported to the cytosol** before they can be translated.

- Mature mRNAs are associated with proteins forming a **nuclear-export signal** that stimulates their transport through **nuclear pores complexes**.

- Thus, gene expression may be repressed by **removing the nuclear-export signal** to prevent mRNAs from exiting the nucleus.

CHECKPOINT 6

How many different proteins can the following pre-mRNA result in?