TOPIC: GENE EXPRESSION

Key Knowledge:

- the structure of genes: exons, introns and promoter and operator regions
- the genetic code as a universal triplet code that is degenerate and the steps in gene expression, including transcription, RNA processing in eukaryotic cells and translation by ribosomes
- the role of rough endoplasmic reticulum, Golgi apparatus and associated vesicles in the export of proteins from a cell via the protein secretory pathway
- the basic elements of gene regulation: prokaryotic *trp* operon as an example of a regulatory process

GENES

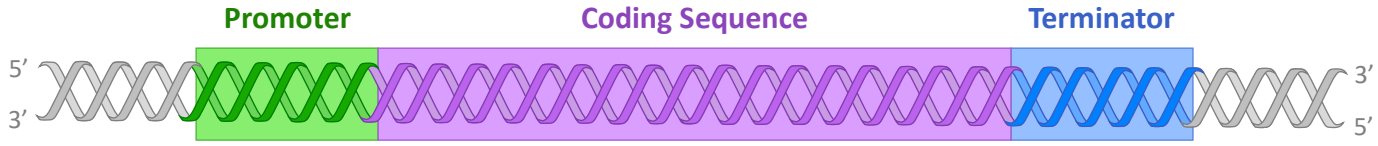
Genes are sequences of DNA that encode a specific characteristic (via the production of RNA or protein). They function as the basic units of inheritance. Genes occupy specific positions on chromosomes *(loci)*.

Genes can be categorised as being either structural or regulatory:

- **Structural genes:** Code for proteins that contribute to either the form or function of the cell
- **Regulatory genes:** Code for transcription factors that control gene expression *(turns genes on or off)*

STRUCTURE OF A GENE

A gene is composed of three key sections. The **promoter** is the site to which the enzyme RNA polymerase will bind – it is responsible for initiating transcription. The **coding sequence** is the region of DNA that is transcribed into RNA, while the **terminator** sequence functions to stop transcription by RNA polymerase. Basically, genes consist of start regions (promoter), copying regions (coding) and stop regions (terminator).



TRANSCRIPTION

Transcription is the process by which a DNA sequence (gene) is copied into complementary RNA sequences by **RNA polymerase**. This enzyme binds to the promoter and then separates the double-stranded DNA of the coding sequence (by breaking the hydrogen bonds between base pairs). **Free RNA nucleotides** then align opposite their complementary base partner and RNA polymerase joins them together with covalent bonds (between the sugar-phosphate backbone). When the enzyme reaches the terminator sequence the synthesised RNA transcript is released and the double helix reforms. Transcription occurs in the nucleus.

POST-TRANSCRIPTIONAL MODIFICATIONS

In eukaryotic cells, there are three post-transcriptional modifications that must occur in order to convert the RNA transcript into **messenger RNA** (mRNA) that is then capable of being translated by the ribosome:

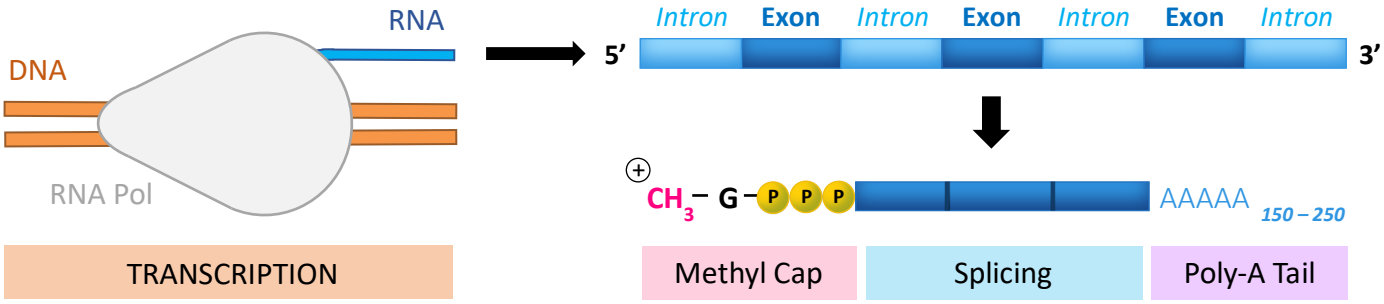
- **Capping:** A methyl group is added to the 5'-end of a transcript *(prevents degradation by exonucleases)*
- **Polyadenylation:** A poly-A tail is added to the 3'-end of a transcript *(improves transcript stability)*
- **Splicing:** Non-coding sequences *(called introns)* must be removed from within the eukaryotic gene

SPLICING

Introns are ***intr****uding sequences* within eukaryotic genes that do not code for protein, whereas **exons** are the ***ex****pressing sequences* within the gene that encode the relevant amino acid sequence. Prokaryotic cells do not typically possess introns and hence do not require splicing of the RNA transcript prior to translation.

It is possible to remove specific exon sequences from a gene to create different forms of the same protein (e.g. removing an exon could determine whether a protein is membrane-bound or found in the cytosol). The differential removal of exon segments is called **alternative splicing**.

Eukaryotic genes *without introns* can be artificially synthesised by converting mRNA into a DNA sequence via the enzyme **reverse transcriptase** (from retroviruses). The intron-free gene sequence is called **cDNA**.



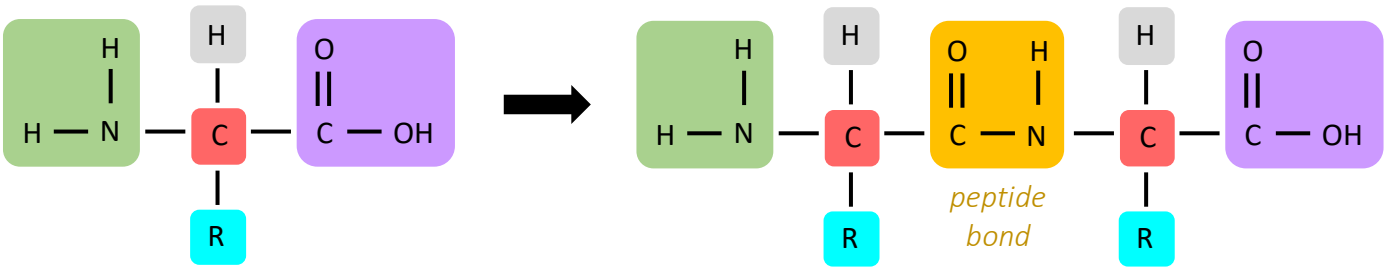
TRANSLATION

Translation is the process of **protein synthesis**, in which genetic information encoded in mRNA is translated into a sequence of amino acids (polypeptide).

Hint: MR CAT APP

- **Messenger RNA** (mRNA) is transported to the ribosomes (in the cytosol)
- **Ribosomes** read the mRNA sequence in triplets of bases called codons
- **Codons** code for specific amino acids according to a genetic code
- **Amino acids** are brought to the ribosome by transfer RNA (tRNA)
- **Transfer RNA** binds to specific codons via complementary anticodons
- **Anticodons** cause tRNA molecules to line up according to codon order
- **Peptide bonds** form between amino acids (catalysed by the ribosome)
- **Polypeptides** are produced as the ribosome moves along the mRNA

Once the amino acid has been added to a growing polypeptide chain by the ribosome (see diagram below), the tRNA molecule is released to collect a new amino acid. Translation is initiated by a ribosome when it reaches a **start codon** (AUG) and is terminated when the ribosome reaches a **stop codon**. The start codon functions to establish the appropriate **reading frame** in which triplets of bases are read.



GENETIC CODE

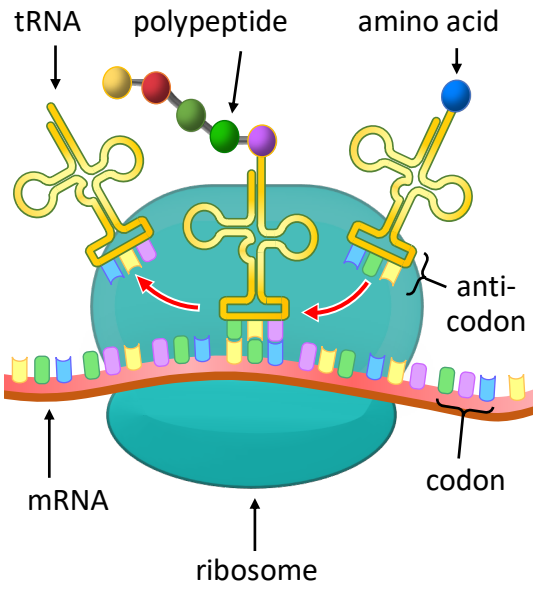
The genetic code is the set of rules that identifies which amino acid corresponds to each mRNA codon. It is typically represented as a table. The genetic code possesses two key characteristics:

- **Universality:** Almost every living organism uses the same code (there are a few viral exceptions)
- **Degeneracy:** Some codons code for the same amino acid (there are 64 codons and only 20 amino acids)

The Genetic Code

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

Translation



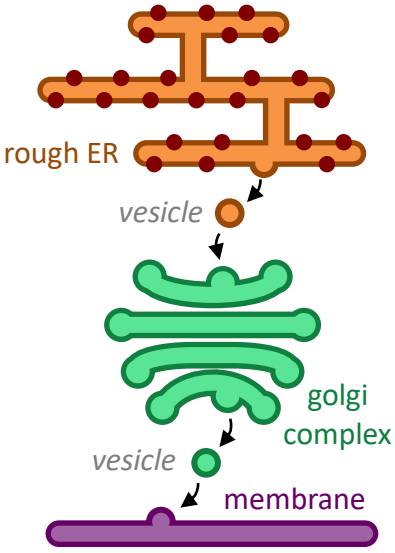
PROTEIN DESTINATIONS

Proteins destined to remain within the cytosol (for intracellular use) are synthesised by free ribosomes. Proteins destined for secretion (extracellular use) are synthesised by ribosomes embedded on rough ER. This secretory pathway will involve the transport of proteins within membranous containers (vesicles) that will fuse with the plasma membrane (via **exocytosis**) to expel the proteins from a cell for extracellular use. The destination of a protein is determined by a short sequence at the start of the polypeptide chain (called the signal recognition peptide). If present, the SRP directs a ribosome to the rough ER (secretory pathway).

VESICULAR TRANSPORT

Proteins destined for secretion are moved between organelles via a membrane network called the **endoplasmic reticulum (ER)**. The rough endoplasmic reticulum is embedded with **ribosomes** which synthesise proteins into the lumen of the rough ER. The rough ER then shuttles the proteins to other organelles via **vesicles**. Proteins destined for secretion are transported to the **Golgi apparatus**, where they are sorted, stored or modified prior to export. The Golgi complex transports the proteins to the **cell membrane** via a vesicle, which fuses with the phospholipid bilayer to release the protein externally. The Golgi body can regulate the secretion of proteins from the cell in one of two distinctive ways:

- **Constitutive:** Products are released immediately as they are made
- **Regulatory:** Products are stored for release in response to a signal



CELL DIFFERENTIATION

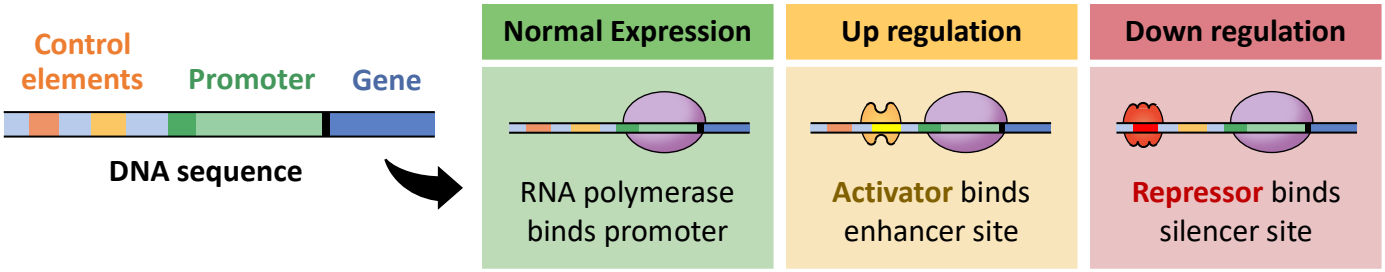
All cells in a multicellular organism contain the same genetic instructions. The totality of DNA sequences (both genes and non-coding DNA) within a cell or organism is called the **genome**. However, different genes may be activated in certain tissues, leading to the production of different proteins. The totality of proteins expressed within a cell or organism at a particular time is called the **proteome**. Because different cell types express different genes and produce different proteins, they may differ in both morphology and function. Cells that are undifferentiated and retain the capacity to form any cell type are known as **stem cells**.

GENE EXPRESSION

The expression of genes is coordinated by **transcription factors**, which are produced by regulatory genes. Transcription factors either mediate or impede the binding of RNA polymerase to the promoter, and hence function to help switch genes on and off. There are two main types of transcription factors:

- **Activator proteins** bind to enhancer sites and essentially function to increase transcription rates
- **Repressor proteins** bind to silencer sites and essentially function to decrease transcription rates

The presence of certain transcription factors may be tissue-specific, leading to the differentiation of cells and tissues. Additionally, **chemical signals** can moderate the activity of transcription factors and hence change gene expression (e.g. hormones may activate target tissues by altering gene expression patterns). The study of changes in organisms as a result of variations in gene expression levels is called **epigenetics**.



TRP OPERON

An **operon** is a cluster of genes under the control of a single promoter. The prokaryotic *trp* operon controls the production of enzymes that are responsible for the synthesis of tryptophan (non-essential amino acid). The expression of the operon is controlled by the ***trp* repressor**, which is produced by a regulatory gene. The *trp* repressor binds to an **operator** site and inhibits the expression of the operon. Tryptophan acts as a positive **inducer** molecule by binding to the repressor and activating it. As a consequence of this:

- If tryptophan is present, the *trp* repressor is activated and the operon is off (tryptophan is not made)
- If tryptophane is absent, the *trp* repressor remains inactive and the operon is on (tryptophan is made)

The inducer-repressor regulation of the *trp* operon is an example of a **negative feedback** loop, whereby the response is the opposite of the stimulus (when levels are low, production of tryptophan increases).