Rewarding Learning

eGUIDE//Biology

Biochemistry, Genetics and Evolutionary Trends

Unit A2 2 5.3 DNA as the Genetic Code

This e-book is designed to complement other support materials and enhance the understanding of this unit for students at GCE level. The topics covered are in accordance with those topics present in the current specification.

Contents

DNA as the Genetic Code

5.3

Learning Outcomes from A2 2 5.3

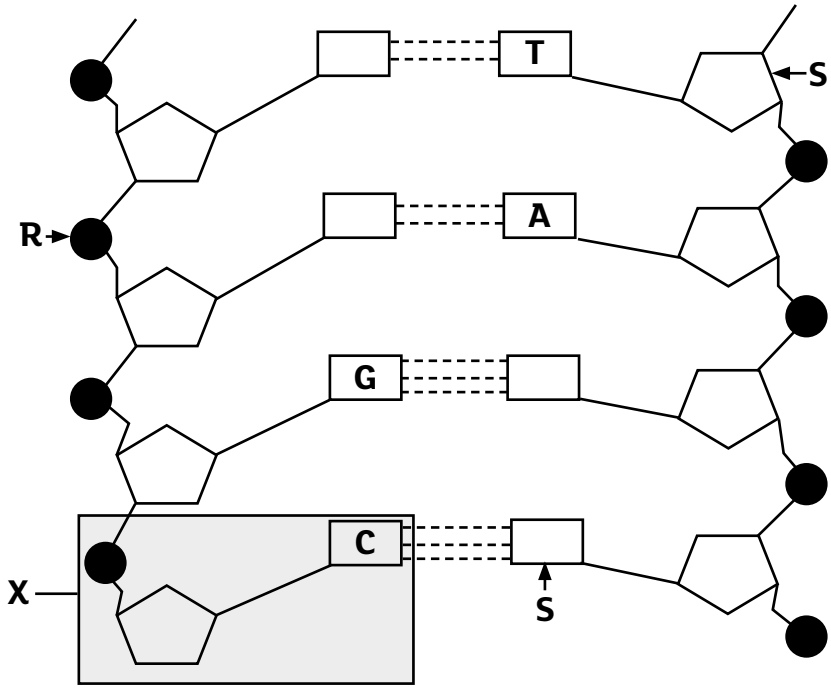
Students should be able to:

- Demonstrate knowledge and understanding of the nature of the genetic code
- Demonstrate knowledge and understanding of the process of transcription in synthesising proteins
- Demonstrate knowledge and understanding of the process of translation in synthesising proteins
- Demonstrate knowledge and understanding of the one gene / one polypeptide theory
- Demonstrate knowledge and understanding of the concept of epigenetics.

The structure of DNA

As quick reminder watch a short video 'DNA Structure' by Teachers Pet (4.21mins)
https://www.youtube.com/watch?v=C1CRrtkWwu0

The diagram below represents a section of a DNA molecule.



Task
Identify the parts indicated by the letters Q, R, S, X and complete the missing letters corresponding to the bases on the diagram.

The two antiparallel strands of the DNA molecule are twisted into a DOUBLE HELIX.

Functions of DNA

DNA has two unique properties;
- it can make accurate copies of itself by the process of REPLICATION (covered in AS Unit 1)
- it possesses the information or GENETIC CODE to make proteins.

DNA is the material of heredity and it is important that the information it carries can be passed on to daughter cells during mitosis or meiosis. The fact that it can replicate itself accurately means that the information it contains is passed on with integrity during cell division. The information codes for amino acids, which are the building blocks of proteins. Each protein has its own specific sequence of amino acids required to function effectively. Each section of DNA that codes for the sequence of amino acids in a particular protein or polypeptide is known as a GENE. A GENE is a sequence of bases on a DNA molecule that codes for a sequence of amino acids in a polypeptide chain.

In humans, genes may vary in size from a few hundred to more than 2 million bases.

DNA as a protein

The information in DNA codes for polypeptide chains. A single polypeptide may be a protein, or a protein may consist of more than one polypeptide chain.

Q. Think about the different types of proteins you know about and write down some examples. Put the examples of proteins that you have listed into the following categories;
- Structural proteins
- Chemical messengers
- Enzymes
- Carrier proteins

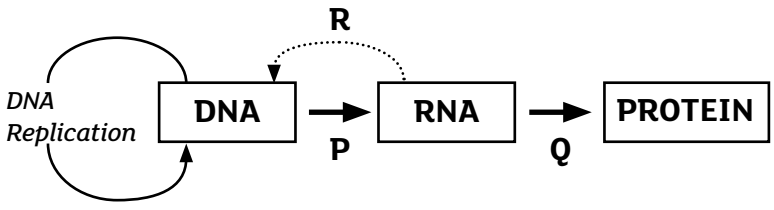
From the examples it is clear that proteins have a wide range of functions, including acting as biological catalysts determining which reactions can and cannot happen within an organism. By determining what chemical reactions occur, enzymes determine the characteristics of cells and organisms.

The webpage below gives information on how genes work.
http://genetics.thetech.org/about-genetics/how-do-genes-work

Protein Synthesis

Protein synthesis takes place in two stages:
1. **Transcription** – the information encoded in the DNA molecule is transcribed into an RNA molecule, which acts as a messenger. It carries the information from the nucleus to ribosomes in the cytoplasm. This molecule is called messenger RNA or mRNA.
2. **Translation** – the information carried in the mRNA is translated into the correct amino acid sequence for the specific polypeptide required.

The pathway of information flow from DNA via RNA to protein is shown below. This is called the Central Dogma and was first proposed by Francis Crick in 1956. The process of DNA replication is also represented on the diagram.



A short video from the DNA Learning Center on the Central Dogma shows how:
1. The DNA genetic code is transcribed into a complementary mRNA molecule.
2. The mRNA molecule formed is spliced to remove non-coding sections called introns (*think of intervening regions*), leaving only the coding regions or exons (*think of regions expressed*) which then leave the nucleus.
3. The genetic information on the modified mRNA is translated into a specific sequence of amino acids to form a polypeptide at ribosomes in the cytoplasm.

A useful transcript for the video is also provided.
https://www.dnalc.org/resources/3d/central-dogma.html

Transcription

Transcription is the process in which DNA serves as a template for the formation of a mRNA molecule. It has the following properties:

- single–stranded
- complementary to a section of a DNA molecule
- synthesised alongside its complementary DNA sequence by the enzyme RNA polymerase
- contains the base uracil (U) instead of thymine (T)
- contains ribose sugar (not deoxyribose)
- smaller than DNA
- can travel out of the nucleus into the cytoplasm via the nuclear pores.

Transcription takes place in 3 stages:

1. **Initiation** – a molecule of the enzyme RNA polymerase binds to the DNA at an initiation site and unwinds and unzips the DNA molecule in that region.
2. **Elongation** – the RNA polymerase copies one of the two strands – the template strand, bringing into position the complementary RNA nucleotides and joining them together to form a molecule of RNA.
3. **Termination** – when the RNA polymerase reaches a stop signal, the transcription stops and the mRNA produced leaves the site of transcription.

The sequence on mRNA matches that of the non-template or coding strand of DNA, except that U (uracil) replaces T (thymine). However the mRNA formed (sometimes called pre-mRNA) must undergo some modification before it leaves the nucleus.

DNA; Transcription and translation by Prof Amann. Transcription is covered up until 29 minutes; 25 seconds.
https://www.youtube.com/watch?v=sOwUZ8fF1Fk

Modification of mRNA – Splicing

Only about 2% of human DNA codes for protein. Most human genes and their mRNA 'transcripts' have long non-coding sections of nucleotides that lie between the coding sections. The non-coding regions are said to intervene between the coding sections and are called introns. The coding regions; those regions that will be expressed as a sequence of amino acids, are called exons. The immature mRNA (pre-RNA) undergoes a modification process called splicing to cut out the introns and join together a continuous stretch of coding exons. The process of splicing is carried out by spliceosomes which recognise the splicing sites. The introns removed are then degraded.

Watch an animation of the splicing process here.
https://www.youtube.com/watch?v=aVgwr0QpYNE

Complete the table below showing a comparison of features of DNA and mRNA.

Feature \ Nucleic acid	DNA	mRNA
Length	50–250 million base pairs	75–3000 nucleotides
Nitrogenous bases		adenine, guanine, cytosine, uracil
Where made in cell	nucleus	
Location in cell	nucleus	

Q. Explain why the length of the DNA is measured in base pairs and the mRNA in nucleotides.
Q. Explain the very large difference in length between DNA and mRNA.

Characteristics of the Genetic Code

4 bases on mRNA code for the 20 amino acids that make up proteins
For example, if one base coded for one amino acid there would only be 4 possible codons. Having a combination of 2 bases coding for 1 amino acid also would not give enough codes.

Table showing the possibilities for combinations of bases

Combination of number of bases in code	Number of amino acids possible
1	$4^1 = 4$
2	$4^2 = 16$
3	$4^3 = 64$

However a combination of 3 bases coding for 1 amino acid provides more than enough possibilities. It is called a *'triplet code'*. The table below shows that most amino acids have more than one codon. They are described as being a *'degenerate code'*.

The genetic code is:
- A triplet code – 3 bases code for 1 amino acid.
- Degenerate – a single amino acid may be coded for by more than one codon.
- Non-overlapping – Each base is only part of one codon. The codons are read sequentially, they do not overlap. For example, the sequence UACGCCUAUCGC is read as UAC GCC UAU CGC.

The table below shows the 'genetic dictionary' indicating the amino acids coded for by mRNA codons.

first base in codon	second base in codon: U	C	A	G	third base in codon
U	phenylalanine	serine	tyrosine	cysteine	U
	phenylalanine	serine	tyrosine	cysteine	C
	leucine	serine	stop	stop	A
	leucine	serine	stop	tryptophan	G
C	leucine	proline	histidine	arginine	U
	leucine	proline	histidine	arginine	C
	leucine	proline	glutamine	arginine	A
	leucine	proline	glutamine	arginine	G
A	isoleucine	threonine	asparagine	serine	U
	isoleucine	threonine	asparagine	serine	C
	isoleucine	threonine	lysine	arginine	A
	methionine and start	threonine	lysine	arginine	G
G	valine	alanine	aspartate	glycine	U
	valine	alanine	aspartate	glycine	C
	valine	alanine	glutamate	glycine	A
	valine	alanine	glutamate	glycine	G

AUG is the codon corresponding to the amino acid methionine (met) and signifies the starting position for protein synthesis. There are 3 STOP codons. These signify the termination of the polypeptide chain.
These are: UAA, UAG and UGA.

This resource shows how a specific sequence of nucleotides codes for insulin. It also shows the degenerate nature of the genetic code. Show the first 13 frames only.
https://www.dnalc.org/view/16515-Animation-23-A-gene-is-a-discrete-sequence-of-DNA-nucleotides-.html

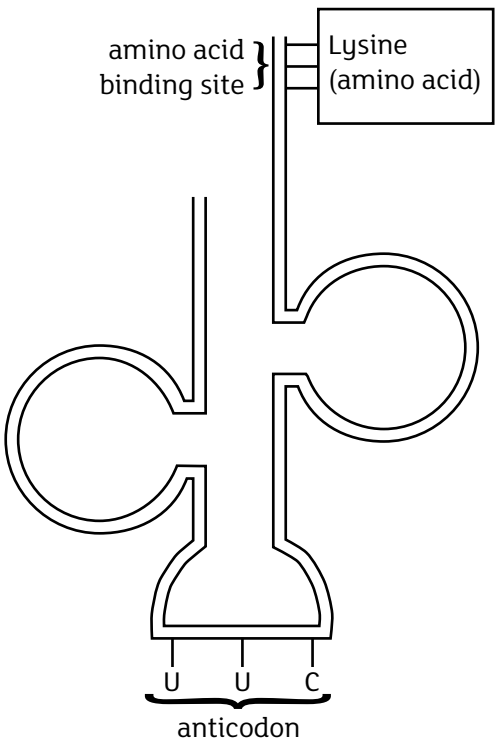
Translation

Two other types of RNA are needed to translate (or decode) the information encoded in the mRNA molecule into a sequence of amino acids.
These are:
1. Transfer RNA (tRNA)
2. Ribosomal RNA (rRNA)

Transfer RNA (tRNA)

The role of tRNA is to 'pick up' specific amino acids from the amino acid pool in the cytoplasm and carry them to the ribosome. Here they are attached to a growing polypeptide chain in a specific sequence according to the order of the codons on mRNA. The structure of tRNA is very important in this function.

Look at the diagrammatic representation of tRNA and the computer artwork image of its structure below.

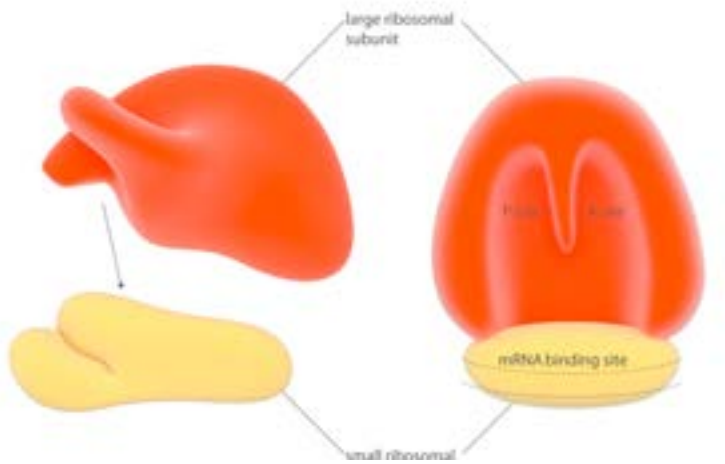


Note the distinctive three-leafed clover shape of the molecule.
It is generally between 76 to 90 nucleotides in length.
Two important sites are labelled in the diagram;
– the amino acid binding site
– the anticodon.

The amino acid that attaches to each tRNA is dependent on, and corresponds to, the anticodon on the molecule. This anticodon is complementary to a triplet codon on mRNA and so will determine the order in which amino acids will be added to the polypeptide being synthesised. Each type of tRNA can only attach to one type of amino acid and so there are many different types of tRNA molecules (remember that the genetic code is degenerate, with a number of different codons coding for one amino acid).

Ribosomal RNA (rRNA) and the ribosomes

Ribosomes are made up of ribosomal RNA (rRNA) and protein.



rRNA is the predominant type of molecule in the ribosome, making up approximately 60% of its weight. There are over 50 different types of protein also associated with the ribosome. The ribosome is made up of two sub-units, the large and small ribosomal units, which come together during the process of translation, when mRNA sits in a groove between the small and large subunits.

A ribosome has two binding sites which are also shown in the illustration. These are:
1. The P or peptidyl site which binds the tRNA with the growing peptide chain.
2. The A or aminoacyl site which accepts incoming tRNA's with their associated amino acids.

These two sites are where protein synthesis takes place.

The process of translation

Translation takes place in four stages. All take place in the cytoplasm of the cell, either on free ribosomes or ribosomes located on the rough endoplasmic reticulum (RER).

1. **Initiation** – the large and small subunits of the ribosome assemble and bind to a mRNA molecule at the start of the sequence it is going to "translate' into a chain of amino acids. The starting sequence is usually AUG and codes for the amino acid methionine (met). The tRNA with an attached met (amino acid) moves into the P site, and a second tRNA with an attached amino acid moves into the A site. The anticodons on the tRNA molecules are complementary to the codons on mRNA at that point.
2. **Elongation** – the amino acid of the tRNA in the P site forms a peptide bond (by a condensation reaction) with the amino acid in the A site, and is released from the first tRNA, which then exits the ribosome.
3. **Translocation** – the ribosome moves to the next mRNA codon and as it does so the tRNA with two amino acids attached (the growing polypeptide chain) is transferred into the P site. The process is then repeated with a third tRNA complementary to the next codon moving into the A site. The two amino acids then join to the third amino acid with a peptide bond and the second tRNA is released. The polypeptide chain grows as the process is repeated. This is shown in the diagram below.

4. **Termination** – the process is terminated when a STOP codon is reached (UAA, UAG or UGA). No tRNA can bind to a STOP codon. The rRNA subunits dissociate and the mRNA and polypeptide are released.

The polypeptide then folds into an active protein (secondary/tertiary structure) to perform its functions.

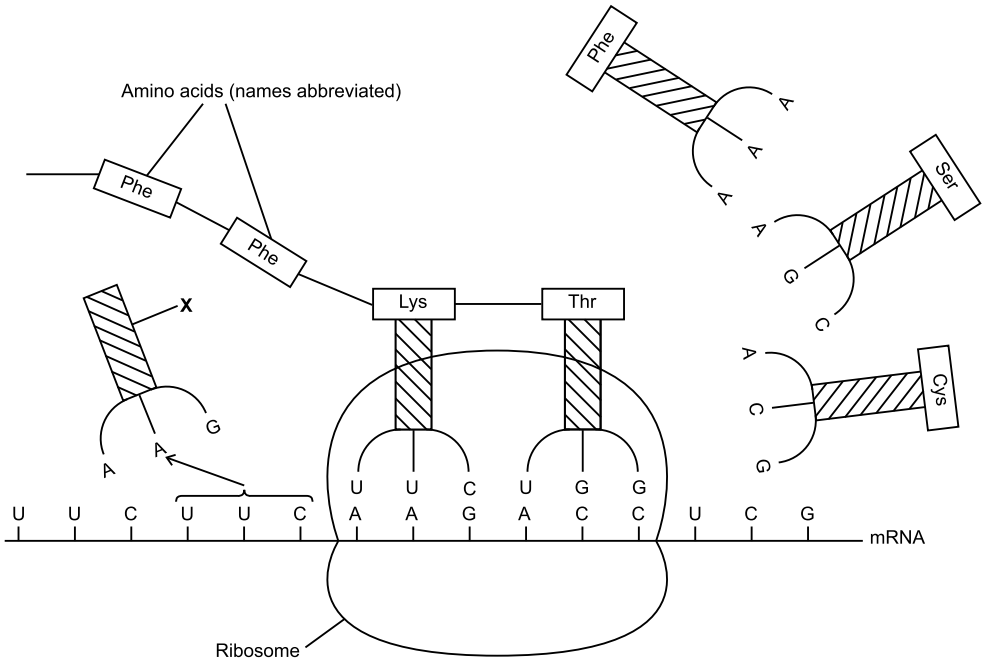


Diagram showing the process of translation

Q. What is the molecule labelled X in the diagram above?
Q. Draw and label the P and A sites on the diagram.

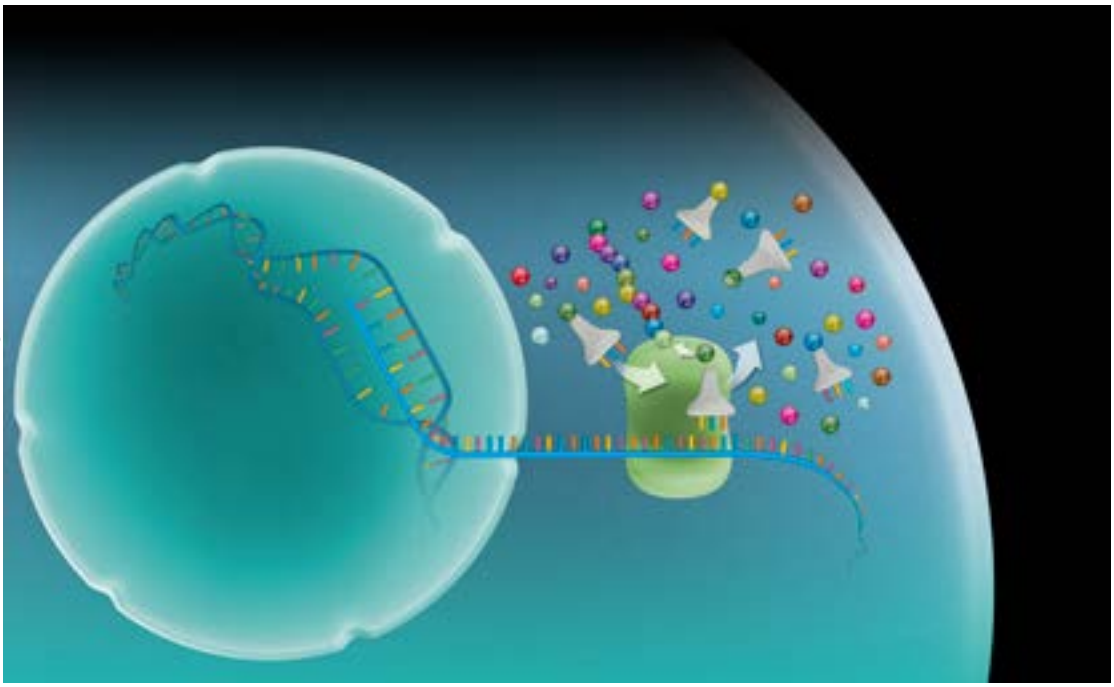
Watch this animation of the process of translation by ndsuvirtualcell.
https://www.youtube.com/watch?v=5bLEDd-PSTQ

If the ribosome is attached to the surface of the rough endoplasmic reticulum (RER), the newly synthesised polypeptide chain enters the cisternae of the RER for modification, vesicle transport and secretion out of the cell. A number of ribosomes can translate an mRNA molecule at the same time, forming a polyribosome or polysome. In this way many polypeptides can be produced simultaneously from a single mRNA strand.

A diagram of a polysome is in action here.
http://www.phschool.com/science/biology_place/biocoach/translation/poly.html

An overview of protein synthesis.

A computer illustration showing how proteins are synthesised from amino acids (coloured spheres) within a cell. Proteins are synthesised according to the information carried in segments of a DNA (deoxyribonucleic acid) strand (upper left) known as genes. A strand of messenger ribonucleic acid (mRNA) is copied from the DNA strand in the nucleus (left) and moves to a ribosome (beige, centre-right). Short strands of transfer RNA (tRNA, bell-shaped) then transport the corresponding amino acids to the ribosome, where they are assembled into a protein chain (upper centre) according to the sequence encoded by the mRNA strand.



Task

Label the following structures in the illustration of protein synthesis above.

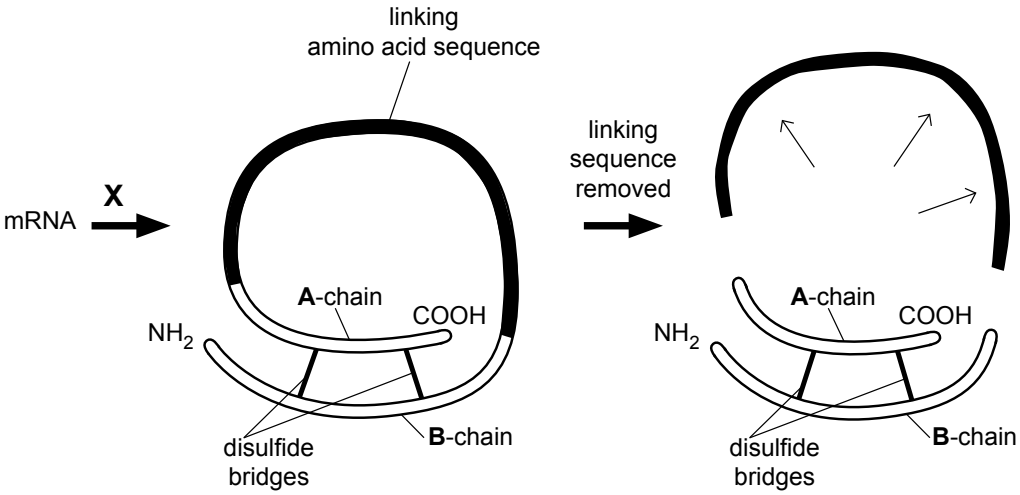
1. Nucleus
2. Cytoplasm
3. DNA double helix
4. mRNA
5. DNA template strand
6. Non-template strand
7. tRNA
8. Amino acid
9. Growing polypeptide chain
10. Ribosome.

You can make your own polypeptide chain by transcribing a DNA sequence and translating it into a sequence of amino acids at the link below
http://learn.genetics.utah.edu/content/molecules/transcribe/

One gene, one polypeptide hypothesis

This hypothesis states that each structural gene carries the information that encodes a single polypeptide chain, rather than a protein. Some proteins are made up of more than one polypeptide chain (quaternary structure) and so require more than one gene to encode their amino acid structure. Some proteins are produced as an inactive precursor, which are subsequently 'activated' by the removal of a specific sequence of amino acids. An example is the protein insulin, which regulates blood glucose levels. It is produced in specialised cells in the pancreas and consists of two polypeptide chains (A and B) made up of 51 amino acids in total. Following the formation of an initial precursor molecule, proinsulin, a linking sequence of amino acids is removed to leave two separate chains which form insulin.

This is summarised in the diagram below.



Proinsulin (precursor molecule) **Insulin (active molecule)**

Epigenetics

It is already known that genes are functional units, which provide instruction for cells through the production of specific proteins. Epigenetics is the study of markers, small *chemical tags*, which act as a set of instructions sitting on top of DNA and the histone proteins they are packed around (see the illustration below). Epi – means on top of, Genetic – is to do with genes.

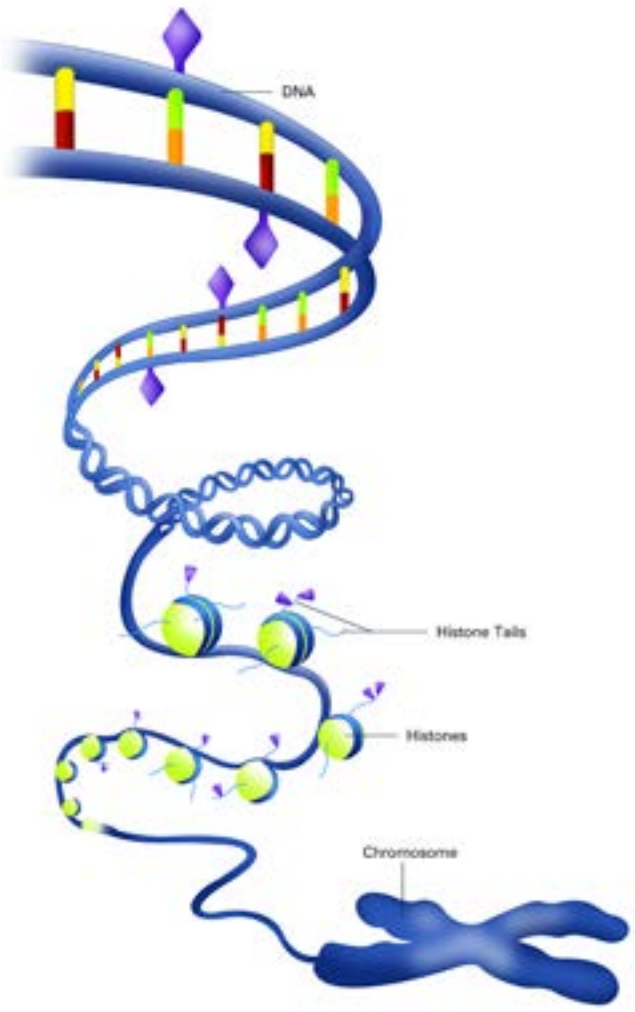
Genes can only be expressed or 'switched on' if the DNA can unwind and separate so that one of the strands can act as a template for the production of a complementary molecule of mRNA. If DNA is highly compacted, this process cannot happen and so the genes in this region of the DNA will be effectively 'switched off', and therefore not expressed. Epigenetics discriminates between different types of cells in an organism, determining which genes are 'switched on' or 'switched off' in a given cell, effectively changing the phenotype of the individual without changing the genotype (DNA sequence). Epigenetic changes are preserved when a cell divides, so are inheritable.

Types of epigenetic modification
Two types of epigenetic modifications are:

1. DNA methylation (addition of a CH_3 group) to cytosine bases of DNA. Methylated genes are not transcribed.
2. Histone modifications – histones are basic proteins that act as spools around which DNA winds in chromatin. The histones enable *compacting* of the DNA molecule to take place, which allows the long strands of DNA to fit into a much smaller space than would otherwise be possible. Methyl groups and other chemical tags can attach to different locations on the histones, each one having a different effect. Some markers make the DNA more accessible for transcription, whereas others have the opposite effect and prevent transcription.

DNA modification is usually long term. Once methylated, the DNA remains methylated through generations. Histone modification is usually associated with short-term epigenetic memory and can be reversed after a few cycles of cell division.

Much epigenetic modification happens during embryonic development. As cells begin to divide, epigenetic markers begin to accumulate. Epigenetic markers can change throughout your lifetime and are influenced by environmental factors such as nutrients, diet, exercise and stress, and passed on to future generations. For example, mothers who don't eat well during pregnancy can develop epigenetic markers that may have longer-term health consequences on their children. Epigenetic markers are also reversible, but it is not yet known exactly which exposures affect which epigenetic markers, or the mechanisms by which this happens.



Epigenetic code. Illustration of the different packaging levels of the structure of DNA (deoxyribonucleic acid), showing the chemical tags that control the epigenetic code. This is a mechanism that controls gene activity. One sort of chemical tag (purple diamonds, upper left) marks the DNA helix (DNA methylation). Another chemical tag (purple triangles, lower left) marks the tails (light blue strands) of histone proteins (histone modification). These markings determine whether genes will be transcribed by RNA polymerase. Genes hidden from RNA polymerase are not expressed. One form of DNA packaging, a chromosome, is at lower right.

Useful Resources

This article on ribosomes, transcription and translation has useful diagrams and images.
http://www.nature.com/scitable/topicpage/ribosomes-transcription-and-translation-14120660

From DNA to protein 3D by your genome shows how the information in DNA is converted into protein structure.
https://www.youtube.com/watch?v=gG7uCskUOrA

This short animation shows the process of translation in real time
https://www.dnalc.org/resources/3d/15-translation-basic.html

.....or recap the whole picture of protein synthesis here
https://www.youtube.com/watch?v=D3fOXt4MrOM

Watch the Crash Course video 'DNA, Hot Pockets, & The Longest Word Ever: Crash Course Biology #11'
https://www.youtube.com/watch?v=itsb2SqR-R0

A quick revision aid showing transcription and translation 'Protein Synthesis (English) by Victors.
https://www.youtube.com/watch?v=lpb5s2F1pyM

The following 2 suggested resources require the teacher to Register with TES

The powerpoint – 'Protein Synthesis – Revision' by D Greenwood on the TES resources website is an excellent revision tool. It contains a simple animation, which aids understanding of the processes of transcription and translation and highlights the important points that students need to know.
https://www.tes.com/teaching-resource/protein-synthesis-6158175

This card sort on the TES website covers some of the main terminology in transcription and translation.
https://www.tes.com/teaching-resource/protein-synthesis-triangle-card-sort-6314533

Quizlet flash cards on transcription.
https://quizlet.com/98954886/transcription-flash-cards/

Quizlet flash cards on translation.
https://quizlet.com/59458486/translation-flash-cards/

Quizlets flash cards on DNA and protein synthesis.
https://quizlet.com/3669302/dna-replication-protein-synthesis-flash-cards/

CCEA past examination questions on 'DNA as the Genetic Code'
(all from Assessment Unit 2 2)

May/June 2015	Question 2
May/June 2014	Question 2
May/June 2013	Question 4
May/June 2012	Question 3
May/June 2011	Question 4
May/June 2010	Question 9