Lesson Seven

DNA

Aims

By the end of this lesson you should be able to:

- know the structure of DNA, including
  - the nucleotides (purines and pyrimidines)
  - base pairing
  - the two sugar-phosphate backbones
  - phosphodiester bonds
  - hydrogen bonds
- understand how DNA is replicated
  - semi-conservatively
  - using DNA helicase, polymerase and ligase
- know that a gene
  - is a sequence of bases on a DNA molecule
  - coding for a sequence of amino acids in a polypeptide chain

Context

This lesson covers sections 1.4 i - iii of the Edexcel AS and A-level Biology specifications.

Edexcel A Level Biology 1, pages 48 – 57.

Oxford Open Learning

Prior knowledge

The blue "Prior knowledge" box on page 48 of the textbook lists the relevant facts you should already know from your study of (I)GCSE Biology or Double Award Science. Use this, and the following "Test yourself on prior knowledge" box and its answers on the website, to revise this knowledge and assess whether you understand it.

If you encounter trouble with this, refer back to the "Assumed background" section on page 4 of the Introduction.

Introduction

To construct a working organism requires **information** as well as materials: information about its structures and about the enzymes which control its chemical reactions. This information is stored in the **DNA** inside cells.

DNA stands for **deoxyribonucleic acid**. Along with the similar **RNA** (ribonucleic acid) it is one type of **nucleic acid**. Its function is information storage. It contains, in coded form, the genetic information which determines the characteristics of the organism in which it is found.

We shall first study the **structure** of DNA and see how that structure encodes the genetic information. Then we shall see how DNA is **replicated** – how it is copied before being passed on when a cell divides. We shall see how this information is used in the construction of proteins in Lesson 8.

DNA structure

The two sugar-phosphate backbones

The structure of one half of a DNA molecule is illustrated in Figure 3.3 on page 50 of the textbook, while the structure of the complete molecule is shown in Figure 3.5 on page 52. DNA is a very long molecule containing three sorts of **subunit**, each coloured differently in the diagrams. In Figure 3.3:

- **deoxyribose** is coloured red. This is a 5-carbon sugar, a **pentose** (see Lesson 2). It has one oxygen atom less than the similar molecule **ribose** found in RNA, hence the name deoxyribose. The structures of both molecules are shown top right of Figure 3.1 on page 49.
- **phosphate** is coloured grey/mauve. This is derived from phosphoric acid, shown top left of Figure 3.1. The abbreviation $\mathbf{P_i}$ stands for "inorganic phosphate".

- **nitrogenous base** (or simply "base") is coloured brown. More about this below.

These three subunits are joined together by **condensation reactions**, eliminating a water molecule in each case, as shown at the bottom left of Figure 3.1. This leaves deoxyribose and phosphate **residues** rather than molecules. The combined structure, with three subunits, is called a **nucleotide**.

Figure 3.3 shows alternating deoxyribose and phosphate residues making up a **sugar-phosphate backbone** within the DNA molecule. We can also think of each phosphate residue as forming one big **phosphodiester bond** joining two deoxyribose residues together.

A base is attached to each deoxyribose residue.

Activity 1	Use the diagram of the deoxyribose molecule at the top right of Page 49 of the textbook:
	(a) Count the number of carbon atoms In the molecule. Remember that the ones at the junctions of the bonds are not shown. (b) The carbon atoms are numbered from the right of the diagram, moving via the front to the left. Now look at the bottom left of Figure 3.1. To which carbon atom is the phosphate group being attached?

Purine and pyrimidine bases

There are four different types of nitrogenous base found in DNA: **adenine** (A), **guanine** (G), **thymine** (T) and **cytosine** (C). Their structures are drawn in the middle of Figure 3.1 on page 49 of the textbook, but you do not need to remember them in detail. Notice that they all contain nitrogen atoms, hence the term nitrogenous base. A fifth sort of base, **uracil**, is also drawn. This replaces thymine in RNA molecules, as we shall see later.

The bases come in two sizes:

- those with two rings in the molecule, adenine and guanine, are called **purines**
- those with a single ring, cytosine and thymine (and uracil) are called **pyrimidines**.

The fact that there are different sorts of base is crucial. It is the sequence of bases on the backbone (the **base sequence**) that encodes the genetic information of the cell.

Base pairing and the double helix

A sugar-phosphate backbone with its attached bases is called a **polynucleotide**. But a complete DNA molecule possesses two polynucleotides, as shown in Figure 3.5 on page 52 of the textbook. The two polynucleotides are held together by **hydrogen bonds** between the bases.

The bases are hydrogen-bonded in a particular pattern to form **complementary base pairs**:

- adenine (A) always bonds with thymine (T)
- guanine (G) always bonds with cytosine (C)

This is **specific** base pairing – in other words, no other pairs can be formed. Notice that a purine pairs with a pyrimidine in each case. Because of their different sizes, this keeps the two sugar-phosphate backbones a regular distance apart.

The whole, double-stranded, molecule is coiled up into a regular spiral shape, as shown on the right of Figure 3.5. This shape is called a **double helix**: "double" because of the two strands, and "helix" meaning spiral.

Activity 2	A polynucleotide has the base sequence A T T G C A C T. Write down the base sequence of the polynucleotide with which it is paired in a DNA molecule.
	

Extension: the discovery of DNA structure

You will not be examined directly on the content of this section.

The structure of DNA was discovered in England in the early 1950s, the work being done in London and Cambridge.

Because the double helix is a regular spiral, it produces a "diffraction pattern" when exposed to a beam of X-rays (see the picture below). Rosalind Franklin, working at Kings College London, worked out the exact dimensions of the spiral using such patterns in 1952.

The X ray diffraction pattern produced by crystallized DNA. The dimensions of the double helix can be calculated from the positions of the black regions.

http://upload.wikimedia.org/wikipedia/commons/4/49/Xray_DNA.gif?uselang=en-gb
Accessed 10.5.2015

James Watson and Francis Crick at Cambridge, using scale models of the DNA subunits, constructed a model of DNA to fit Franklin's data. This was the double helix described above, with the bases joined in the middle.

Watson and Crick's model, on display at the London Science Museum

http://upload.wikimedia.org/wikipedia/commons/5/50/DNA_Model_Crick-Watson.jpg?uselang=en-gb
Accessed 10.5.2015

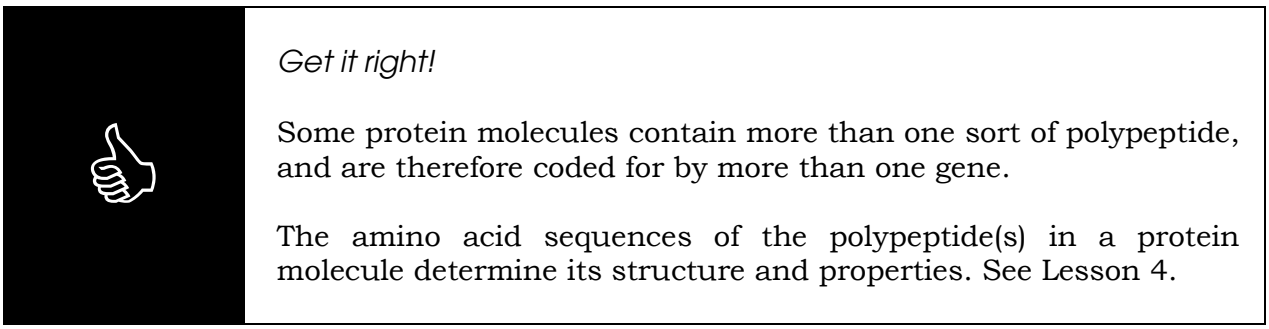
Watson and Crick's proposed structure of DNA was published in a very brief article in the periodical *Nature* in 1953. It is rare example of a ground-breaking scientific article that can be understood by someone studying AS or A Level. You can download the complete text, entitled "A Structure for Deoxyribose Nucleic Acid", from www.ool.co.uk/0705ba.

James Watson wrote up the events in a racy autobiography entitled *The Double Helix* (1968) which is an exciting read. Francis Crick is immortalised in the name of the cutting-edge Francis Crick Institute for Medical Research, which opened in London in 2015.

The Genetic Code

Before the discovery of DNA structure in 1953, the term **gene** was already in use for a single unit of **inherited** information: information passed on from parent to offspring. A single gene might determine what your blood group is, for example, or whether a pea plant is tall or short, or the eye colour of a fruit fly.

We now know that *a gene is a section of DNA whose base sequence determines the amino acid sequence of a single polypeptide.* This is **the one gene one polypeptide hypothesis**.



Get it right!

Some protein molecules contain more than one sort of polypeptide, and are therefore coded for by more than one gene.

The amino acid sequences of the polypeptide(s) in a protein molecule determine its structure and properties. See Lesson 4.

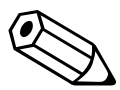
Polypeptides are constructed from 20+ different sorts of amino acid (see Lesson 4) but there are only 4 different sorts of base in DNA. So it cannot be that each type of amino acid is coded for by a different type of base. Instead, the DNA code is a **triplet code**. Every sequence of *three* bases, called a **codon**, codes for a single amino acid in a polypeptide. As there are 64 possible combinations of three bases, this provides enough triplets to code for all of the 20+ amino acids, with several to spare.

The list of amino acids found in proteins is shown in Figure 3.8 on page 57 of the textbook, along with the codons that code for them. You do not need to remember any of the codons, but notice that:

- the code is expressed here in its messenger RNA form (see Lesson 8), with uracil (U) replacing thymine (T).
- the code is **degenerate**, meaning that some amino acids are produced by more than one codon. For example, tyrosine (Tyr) is produced by both UAU and UAC.
- some codons do not code for an amino acid at all. UAA, UAG and UGA are "stop" codons. These signifies the end of a gene, and therefore of the polypeptide chain that it codes for.

Figure 3.9 on page 58 of the textbook shows how this works in practice. The strand on the left of the double-stranded DNA (the **complementary** or **sense strand**) contains the codons. Reading from the top of this strand, the first three bases (TCT) code for the amino acid serine, the second three (GAC) code for the amino acid aspartic acid, and so on. The other DNA strand, called the **coding** or **antisense strand**, does not contain the codons.

Because many of the proteins produced in this way are enzymes, and because enzymes control the chemistry of the cell, the genes control the cell and the nature of the organism that is formed. That is why a gene is also "a unit of inherited information", as in the old definition of the term.

Activity 3	Use Figure 3.8 on page 57 of the textbook to find (a) the two codons which code for the amino acid phenylalanine (b) the amino acid coded for by the codon AGG.
	

Non-coding DNA

The entire base sequence of all the DNA of some organisms has now been determined. In 2003 this was also achieved for a human being by the **human genome project**.

It has become clear that much of the DNA in a cell (more than 98% of it in human cells) does not actually code for the amino acid sequences of proteins as described above. The DNA includes, among other things:

- sections within genes called **introns**, whose information is ignored during protein synthesis (the making of proteins using the information on the DNA)
- control sections next to genes, which decide whether the information in a gene is used or not at a particular time

We shall examine these issues in greater depth in the second year of the A Level course.

DNA replication

Every cell in a multicellular organism contains a complete set of the organism's DNA. Before the cell divides, this DNA is copied or **replicated** so that a copy can be passed to each of the daughter cells formed by division.

Get it right!

Eukaryotic organisms usually have two copies of the DNA per cell, as we shall see in Lesson 12. But this is still all replicated before the cell divides.

Because of DNA replication, every cell in your body possesses the same DNA (and therefore the same genes) that was in the single-celled zygote formed by the fusion of your father's sperm and your mother's egg at your conception.

There are some cells which have no DNA at all, and which cannot divide to form other cells. Human red blood cells are one example.

The replication process

DNA replication happens in four steps, as illustrated in Figure 3.6 on page 53 of the textbook.

- An enzyme called **DNA helicase** binds at a place on the DNA molecule called a **replication site**, breaks the hydrogen bonds between the bases on the two strands to separate ("unzip") them, and holds the two strands apart. Each strand then acts as a **template** for the formation of a new strand.

- New, free nucleotides attach to the now-exposed bases on both strands by specific base pairing. Adenine attaches to an exposed thymine, guanine to an exposed cytosine, and so on. The new nucleotides are held in place by hydrogen bonds.

- Another enzyme, **DNA polymerase**, joins the phosphate and deoxyribose residues of these newly-bound nucleotides together, by condensation reactions, to form two new polynucleotides.

- Because there are several replication sites on each DNA molecule, the new polynucleotides are formed in sections. Another enzyme, **DNA ligase**, joins these together.

DNA polymerase also "proof reads" the new polynucleotides, and corrects any mistakes that have been made during the copying process. The result is two double-stranded DNA molecules, each identical to the original.

Note the following points:

- Each new double-stranded DNA molecule contains *one* of the two strands from the original DNA molecule. This pattern is called **semi-conservative replication**.
- The enzymes work in one particular direction along the DNA molecule, from the 3′ end ("three prime end") to the 5′ end ("five prime end") of the template. The numbers refer to the position of the phosphate group on the deoxyribose residue in the backbone (see Activity 1 above).
- **DNA ligase**, which can join together sections of DNA, is important in genetic engineering, as we shall see in the second year of the A Level course.

Activity 4	The antisense strand does not code for amino acids in a polypeptide chain. Explain how its sequence of bases is nevertheless essential.
	

Evidence for semi-conservative replication

That DNA replicates semi-conservatively was first demonstrated by Meselson and Stahl in 1958. Their experiment made use of two **isotopes** of nitrogen: two sorts of nitrogen atom containing different numbers of neutrons and therefore having different masses:

- normal **^{14}N** atoms with 7 neutrons
- heavy **^{15}N** atoms with 8 neutrons

They first grew some of the bacteria ***E. coli*** in a growth medium containing only heavy **^{15}N** nitrogen for several generations. The DNA in the bacteria now contained only heavy nitrogen atoms in its nitrogenous bases. They then transferred these bacteria to a growth medium containing only normal **^{14}N**.

They harvested DNA from the bacteria before it had replicated, and then after it had replicated once and twice (i.e. after one and two generations). The extracted DNA was whirled in a **centrifuge**, and it moved to a position in its centrifuge tube corresponding to its density and therefore to the mass

of its nitrogen atoms. The results are shown in Figure 3.7 on page 55 of the textbook.

After one generation, all of the extracted DNA had an intermediate density. This suggests that it all contained one heavy and one light polynucleotide strand, as shown on the right hand side of Figure 3.7. This confirms that the DNA had replicated semi-conservatively, the heavy strand being the old DNA, and the light strand the newly-constructed DNA.

After two generations, 50% of the DNA was light and 50% was of intermediate density. This confirms that replication was semi-conservative, as again shown on the right of the diagram.

Activity 5	Interpretation of Meselson and Stahl's experiment is a favourite with A Level examiners. Work carefully through the Activity on pages 54 – 55 of the textbook concerning it, and check your answers to the questions on the Hodder site as usual.

Extension: density gradient centrifugation

You will not be examined directly on the content of this section.

A centrifuge is rather like a spin drier, but holds glass centrifuge tubes filled with water rather than damp washing. As the centrifuge spins, material denser than water is forced rapidly down to the bottom of the centrifuge tubes.

A centrifuge. The centrifuge tubes are placed in the holes around the rim of the central part, which spins fast when the lid is closed.

http://commons.wikimedia.org/wiki/Centrifuge#mediaviewer/File:Tabletop_centrifuge.jpg
Accessed 10.5.2015

Meselson and Stahl used an adaptation of this method called density gradient centrifugation. Their tubes were filled with solution that got more concentrated, and therefore more dense, going from the top to the bottom of the tube. The DNA, suspended in the solution, moved down the tube to the point at which it had the same density as the solution, and then stopped. The position of the DNA was then found by using a stain (as in staining nuclei for viewing under a microscope).

Activity 6	Use the diagram of Meselson and Stahl's experiment shown in Figure 3.7 on page 55 of the textbook. Predict the bands that would be obtained after a third generation.
	

Activity 7	Read examination questions 1 - 6 on pages 64 - 65 of the textbook. Guidance on answering them is given at the end of the lesson.

Read *Edexcel A Level Biology 1,* pages 48 – 57.

Answer the "Test yourself" questions on pages 52, 56 and 57, and check your responses.

Keywords

adenine, guanine, thymine, cytosine, uracil
base sequence
centrifuge
coding / antisense strand
codon
complementary base pairs

complementary / sense strand
degenerate
deoxyribose, ribose, pentose
DNA, RNA, nucleic acid
DNA helicase, DNA polymerase, DNA ligase
double helix
E. coli
gene
human genome project
introns
isotope
nitrogenous base
nucleotide, polynucleotide
one gene one polypeptide hypothesis
P_i / inorganic phosphate
phosphodiester bond
polynucleotide
purine, pyrimidine
replication
semi-conservative replication
specific base pairing
subunit
sugar-phosphate backbone
triplet code

Summary

Lesson Seven: DNA

DNA structure
- sugar-phosphate backbone
- purine and pyrimidine bases
- base pairs and the double helix

The genetic code

DNA replication
- the replication process
- evidence for semi-conservative replication

What you need to know

- the meanings of the keywords listed above
- the detailed structure of the DNA molecule

What you might be asked to do

- explain why DNA has a triplet code

- explain the process of DNA replication
- explain why DNA replication is semi-conservative
- describe and interpret the Meselson Stahl experiment
- deduce the base sequence of a complementary polynucleotide
- deduce the codons corresponding to an amino acid sequence, and vice versa

Suggested Answers to Activities

Activity 1

(a) 5
(b) The fifth carbon atom – C^5

Activity 2

TAACGTGA

Activity 3

(a) UUU and UUC
(b) arginine

Activity 4

It is the template for the construction of a new sense strand by complementary base pairing.

Activity 6

Light and medium bands only (with the light band stronger than the medium).

Activity 7

Q1: C

Q2: C

Q3: (a) The strong covalent bonds holding each polynucleotide together, and the hydrogen bonds binding the two polynucleotides to each other.

(b) In the sequence of nitrogenous bases. Each triplet of bases codes for one amino acid in a polypeptide.

Q4: (a) L = guanine, so M = cytosine; N = adenine

(b) This happens during replication. The enzyme DNA helicase binds at a replication site. The hydrogen bonds between the pairs of nitrogenous bases are broken.

Q5: (a) It is in circular form, rather than long strands; it is not contained within a membrane-bound nucleus. (There are other differences as well.)

(b) In all DNA, the percentage of G = the percentage of C; the percentage of A = the percentage of T.

Q6: (a) It contains deoxyribose rather than ribose; it contains thymine (T) rather than uracil (U).

(b) The bottom two of the three nucleotides are purines – purines have bases with a double-ring structure.

(c) To the top. The enzymes work from the 3′ to the 5′ direction.