6 Nucleic acids and protein synthesis

Nucleic acids have roles in the storage and retrieval of genetic information and in the use of this information to synthesise polypeptides. DNA is an extremely stable molecule that cells replicate with extreme accuracy. The genetic code is used by cells for assembling amino acids in correct sequences to make polypeptides. In eukaryotes this involves the processes of transcription in the nucleus to produce short-lived molecules of messenger RNA followed by translation in the cytoplasm.

6.1 Structure and replication of DNA

Understanding the structure of nucleic acids allows an understanding of their role in the storage of genetic information and how that information is used in the synthesis of proteins.

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

a) describe the structure of nucleotides, including the phosphorylated nucleotide ATP
b) describe the structure of RNA and DNA and explain the importance of base pairing and the different hydrogen bonding between bases
c) describe the semi-conservative replication of DNA during interphase

Nucleic acids – the information molecules

Nucleic acids are the **information molecules** of cells; they are the genetic material of all living organisms and also of viruses. Within the structure of nucleic acid are coded the 'instructions' that govern all cellular activities. This code (known as the **genetic code**) is a universal one – it makes sense in all organisms.

There are two types of nucleic acids found in living cells, **deoxyribonucleic acid (DNA)**, and **ribonucleic acid (RNA)**. DNA is the genetic material and occurs in the chromosomes of the nucleus. But, whilst some RNA also occurs in the nucleus, most is found in the cytoplasm – particularly in the ribosomes. Both DNA and RNA have roles in the day-to-day control of cells and organisms, as we shall see shortly.

Nucleic acids are giant molecules known as **macromolecules**. Chemically they are described as **polymers** because they are made by linking together smaller building blocks called **monomers**. These smaller molecules that make up DNA and RNA are the **nucleotides**, and it is these molecules that are condensed together to form long, thread-like nucleic acids. First you will look at the structure of the nucleotides, themselves.

Nucleotides–the building blocks

A nucleotide consists of three components combined together:

- a **nitrogenous base**, either cytosine (C), guanine (G), adenine (A), thymine (T) or uracil (U)
- a **pentose sugar**, either deoxyribose (in DNA) or ribose (in RNA)
- **phosphoric acid**.

The nitrogenous bases are derived from one of two parent compounds, purine or pyrimidine. The purine bases are **cytosine** or **guanine**; the pyrimidine bases are **adenine**, **thymine** or **uracil**. These molecules are illustrated in Figure 6.1. You do not need to know their structural formulae. However, you should note that purines have a double ring structure and pyrimidines have a single ring structure. We will note the special significance of this size difference, shortly. It has become traditional in biology to abbreviate the names of these five bases to their first letters, **C**, **G**, **A**, **T** and **U**.

The three components, nitrogenous base, pentose sugar and phosphoric acid, are combined by **condensation reactions** to form a nucleotide with the formation of two molecules of water. The structure of these components and how they are combined together are illustrated in Figure 6.1. Here a diagrammatic way of representation has been used to emphasise their spatial arrangement.

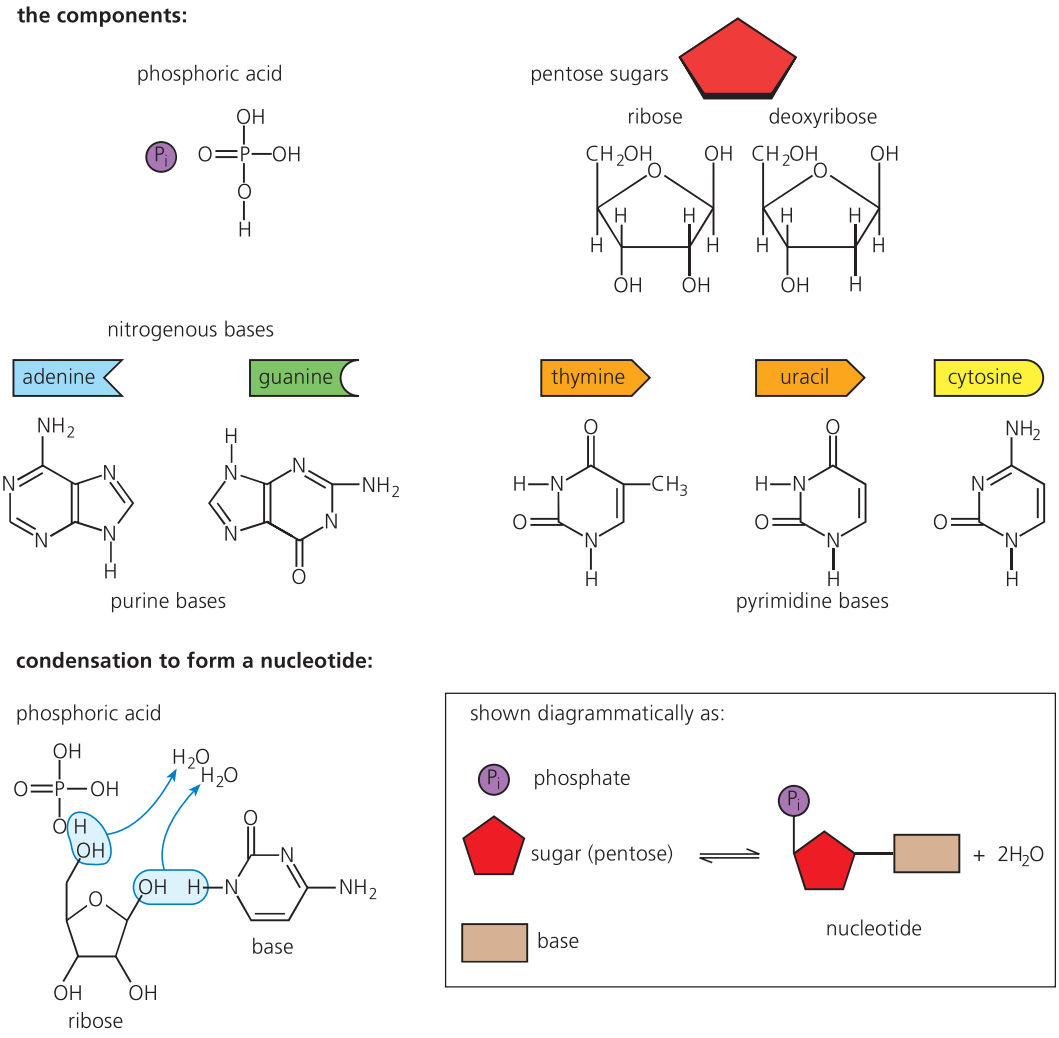


Figure 6.1 The components of nucleotides

Question

1 Distinguish between a 'nitrogenous base' and a 'base' found in inorganic chemistry.

ATP – a nucleotide with unusual features

Energy made available within the cytoplasm may be transferred to a molecule called **adenosine triphosphate** (**ATP**). This substance occurs in all cells at a concentration of 0.5–2.5 mg cm^{-3}. It is a relatively small, soluble organic molecule – a **nucleotide**, which carries three phosphate groups linked together in a linear sequence (Figure 1.22, page 21).

Look at Figure 1.22 now.

ATP is formed from **adenosine diphosphate** (**ADP**) and a **phosphate ion** (**Pi**) by transfer of energy from other reactions. ATP is referred to as '**the universal energy currency**' because, like money, it can be used in different contexts, and it is constantly recycled. ATP contains a good deal

of chemical energy locked up in its structure. What makes ATP special as a reservoir of stored chemical energy is its role as a common intermediate between energy-yielding reactions and energy-requiring reactions and processes.

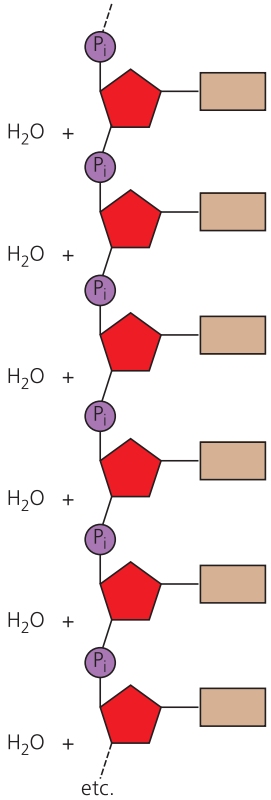
Energy-yielding reactions include the photophosphorylation reactions of photosynthesis (page 267), and many of the reactions of cell respiration (page 243).

Energy-requiring reactions include the synthesis of cellulose from glucose, the synthesis of proteins from amino acids and the contraction of muscle fibres.

The free energy available in ATP is approximately 30–34 kJ mol^{-1}, made available in the presence of a specific enzyme, referred to as ATPase. Some of this energy is lost as heat in a reaction, but much free energy is made available to do useful work, more than sufficient to drive a typical energy-requiring reaction of metabolism.

- Sometimes ATP reacts with water (a hydrolysis reaction) and is converted to ADP and Pi. Direct hydrolysis of the terminal phosphate groups like this happens in muscle contraction, for example.
- Mostly, ATP reacts with other metabolites and forms phosphorylated intermediates, making them more reactive in the process. The phosphate groups are released later, so both ADP and Pi become available for reuse as metabolism continues.

Direction in the nucleic acid molecule:
in a nucleic acid molecule, the phosphate groups are bridges between **carbon-3** of one sugar molecule and **carbon-5** of the next. Consequently, we can identify '***direction***' in the nucleic acid molecule, indicating one end as 5′ and the other end as 3′.



When the polynucleotide chain grows by addition of another nucleotide, the nucleotides are always added to the 3′ end.

Figure 6.2 'Direction' in the nucleic acid molecule

Nucleotides become nucleic acid

Nucleotides themselves then combine together, one nucleotide at a time, to form huge molecules called nucleic acids or **polynucleotides** (Figure 6.2). This occurs by a condensation reaction between the phosphate group of one nucleotide and the sugar group of another nucleotide. Up to 5 million nucleotides, condensed together, form a polynucleotide. So, nucleic acids are very long, thread-like macromolecules with alternating sugar and phosphate molecules forming the 'backbone'. This part of the nucleic acid molecule is uniform and unvarying. However, also attached to each of the sugar molecules along the strand is one of the bases, and these project sideways. Since the bases vary, they represent a unique sequence that carries the coded information held by the nucleic acid.

Introducing DNA

The **DNA molecule** consists of two polynucleotide strands, paired together. These strands are held by hydrogen bonds between their paired bases. The two strands are of the order of several million nucleotides in length and they take the shape of a **double helix** (Figure 6.3). In all DNA molecules the nucleotides contain deoxyribose.

The other chemical feature that distinguishes DNA from RNA is the organic bases present and the way they pair up. In DNA the nitrogenous bases are **cytosine** (**C**), **guanine** (**G**), **adenine** (**A**) and **thymine** (**T**), but never uracil (U). The pairing of bases is between adenine and thymine (**A** and **T**) and between cytosine and guanine (**C** and **G**). This is because there is just enough space between the two sugar–phosphate backbones for one purine and one pyrimidine molecule. So a purine in one strand must be opposite a pyrimidine in the other, and vice versa. These are the only combinations that fit together along the helix. Pairing of bases is accompanied by the formation of hydrogen bonds; two are formed between adenine and thymine and three are formed between cytosine and guanine. This pairing, known as **complementary base pairing**, also makes possible the very precise way that DNA is copied in a process called replication.

We have noted there is direction in a nucleic acid molecule. *Look again at Figure 6.2 to make certain you understand what 'direction' means.*

So, in the DNA molecule, the phosphate groups along each strand are bridges between carbon-3 of one sugar molecule and carbon-5 of the next. Furthermore, within the DNA double helix the two chains are arranged **antiparallel**. In Figure 6.3 you can see that one chain runs from 5′ to 3′ whilst the other runs from 3′ to 5′ – which is what 'antiparallel' means.

The existence of direction in DNA strands becomes important in replication and also when the genetic code is transcribed into messenger RNA.

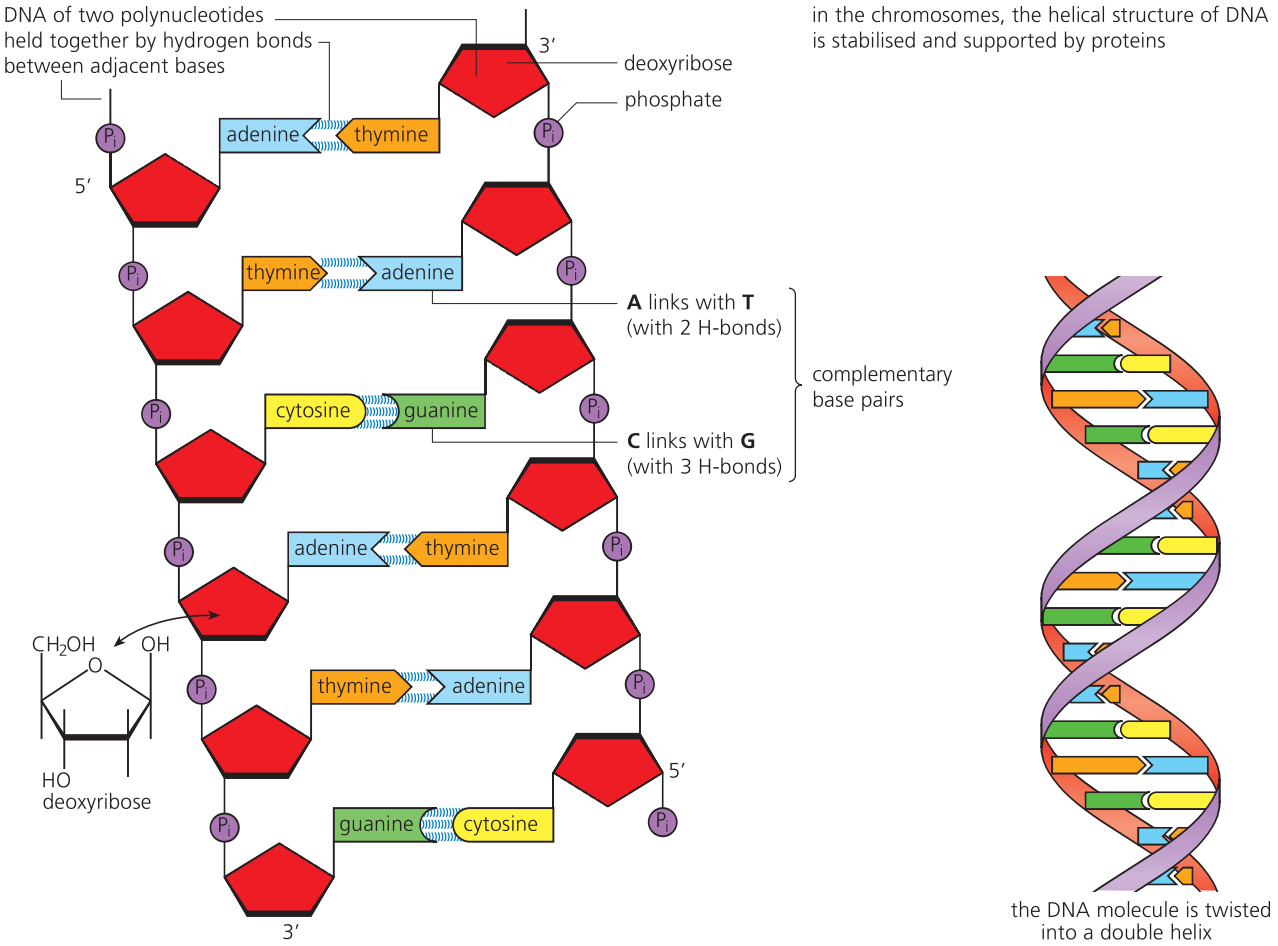


Figure 6.3 The structure of DNA

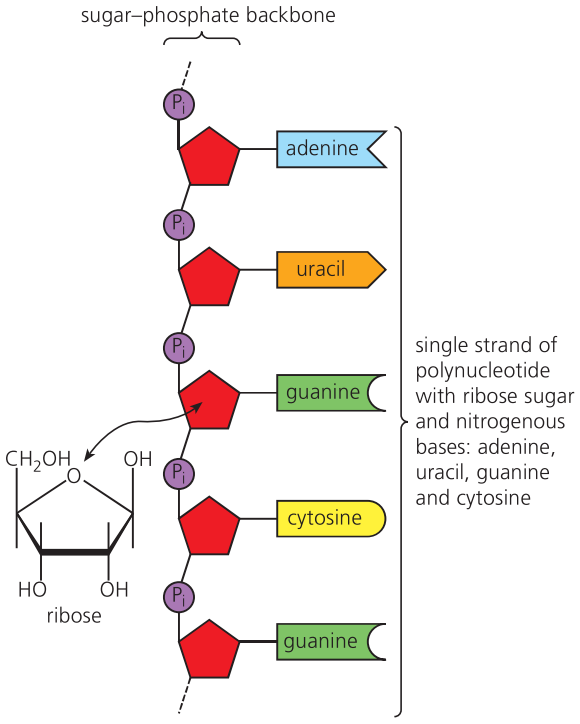


Figure 6.5 The structure of RNA

Introducing RNA

The **RNA molecule** is relatively short in length, compared with DNA. In fact, RNAs tend to be between one hundred and one thousand nucleotides long, depending on the particular role they have. In all RNA molecules the nucleotides contain ribose.

The other chemical feature that distinguishes RNA is that the bases are cytosine (C), guanine (G), adenine (A) and **uracil** (**U**), but never thymine (T). RNA occurs as a single strand.

In the 'information business' of cells there are two functional types of RNA. These are **messenger RNA** (**mRNA**) and **transfer RNA** (**tRNA**). Their roles are in the transfer of information from the nucleus to the cytoplasm and in the construction of proteins at the ribosomes, located in the cytoplasm. We will return to this process, shortly.

Table 6.1 The differences between DNA and RNA

	DNA	RNA
Length	Very long strands, several million nucleotides long	Relatively short strands
Sugar	Deoxyribose	Ribose
Bases	C, G, A and T (not U)	C, G, A and U (not T)
Forms	Consists of two polynucleotide strands in the form of a double helix with complementary base pairs: • C with G and A with T • held by hydrogen bonds.	Consists of single strands and exists in two functional forms: • messenger RNA (mRNA) • transfer RNA (tRNA).

Extension

The existence of the DNA double helix was discovered, and the way DNA holds information was suggested by Francis Crick and James Watson in 1953, for which a Nobel Prize was awarded (Figure 6.4).

Francis Crick (1916–2004) and **James Watson** (1928–) laid the foundations of a new branch of biology – cell biology – and achieved this while still young men. Within two years of their meeting in the Cavendish Laboratory, Cambridge (1951), Crick and Watson had achieved their understanding of the nature of the gene in chemical terms.

Crick and Watson brought together the experimental results of many other workers, and from this evidence they deduced the likely structure of the DNA molecule.

- **Erwin Chergaff** measured the exact amount of the four organic bases in samples of DNA, and found the ratio of A:T and of C:G was always close to 1. Chergaff's results suggest consistent base pairing in DNA from different organisms.

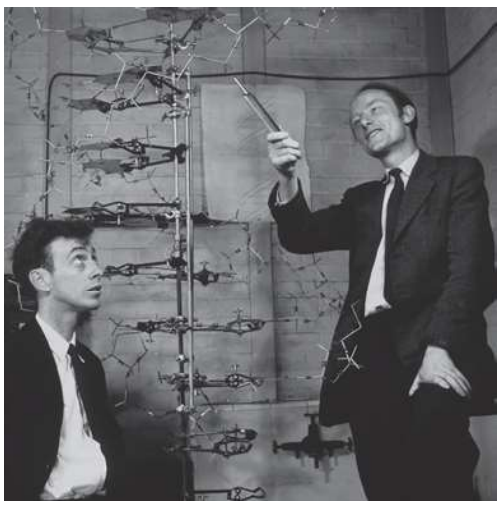
Organism	Ratio of bases in DNA samples	
	Adendine : Thymine	Guanine : Cytosine
Cow	1.04	1.00
Human	1.00	1.00
Salmon	1.02	1.02
Escherichia coli	1.09	0.99

- **Rosalind Franklin** and **Maurice Wilkins** produced X-ray diffraction patterns by bombarding crystalline DNA with X-rays.

Crick and Watson concluded that DNA is a double helix consisting of:

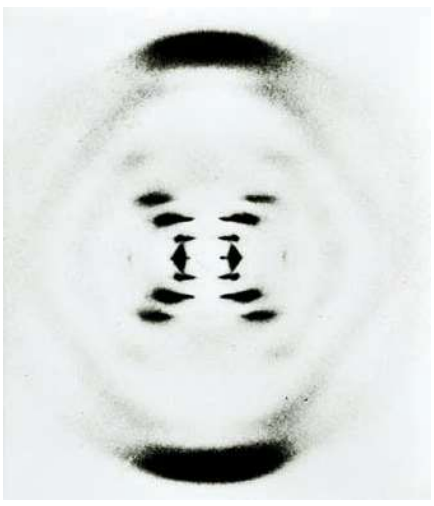
- two polynucleotide strands with nitrogenous bases stacked on the inside of the helix (like rungs on a twisted ladder)
- parallel strands held together by hydrogen bonds between the paired bases (A–T, C–G)
- ten per turn of the helix
- two antiparallel strands of a double helix.

They built a model. See Figure 6.3 for a simplified model of the DNA double helix.



a) Watson and Crick with their demonstration model of DNA

b) Rosalind Franklin produced the key X-ray diffraction pattern of DNA at Kings College, London



c) X-ray diffraction pattern of DNA

Figure 6.4 The discovery of the role of DNA

Replication – how DNA copies itself

Some features of the DNA of the chromosome and how it is copied or **replicated** have already been established.

- Replication must be an extremely accurate process since DNA carries the genetic message.
- Replication is quite separate from cell division, for replication of DNA takes place in the nucleus during **interphase**, well before the events of nuclear division.
- Strands of the DNA double helix are built up individually from free **nucleotides**. The structure of nucleotides is shown in Figure 6.1.

Before nucleotides can be replicated, the DNA double helix has to **unwind** and the hydrogen bonds holding the strands together must be broken, allowing the two strands of the helix to separate. The enzyme **helicase** brings about the unwinding process and holds the strands apart for replication to occur. Each point along the DNA molecule where this occurs is called a **replication fork** (Figure 6.6). At a replication fork, both strands act as **templates**. A molecule of the enzyme DNA polymerase binds to each of the DNA strands and begins to move along them. As each base along a strand is reached, a free nucleotide that is complementary is paired with it (A with T, C with G). These free nucleotides are already 'activated' by the presence of extra phosphates. Each free nucleotide is held in place by the hydrogen bonds that form between complementary bases. Then, from each free nucleotide, the extra phosphates are broken off. This provides the energy by which the remaining phosphate is joined to the sugar molecule of the neighbouring nucleotide. This condensation reaction is brought about by DNA polymerase. The new strands of DNA 'grow' in this way.

DNA polymerase moves along the single strands of DNA in the 5′ → 3′ direction. If you look at the illustration of replication in Figure 6.6, you will see that only one strand can be copied continuously (the strand where the DNA polymerase enzyme is travelling in the same direction as

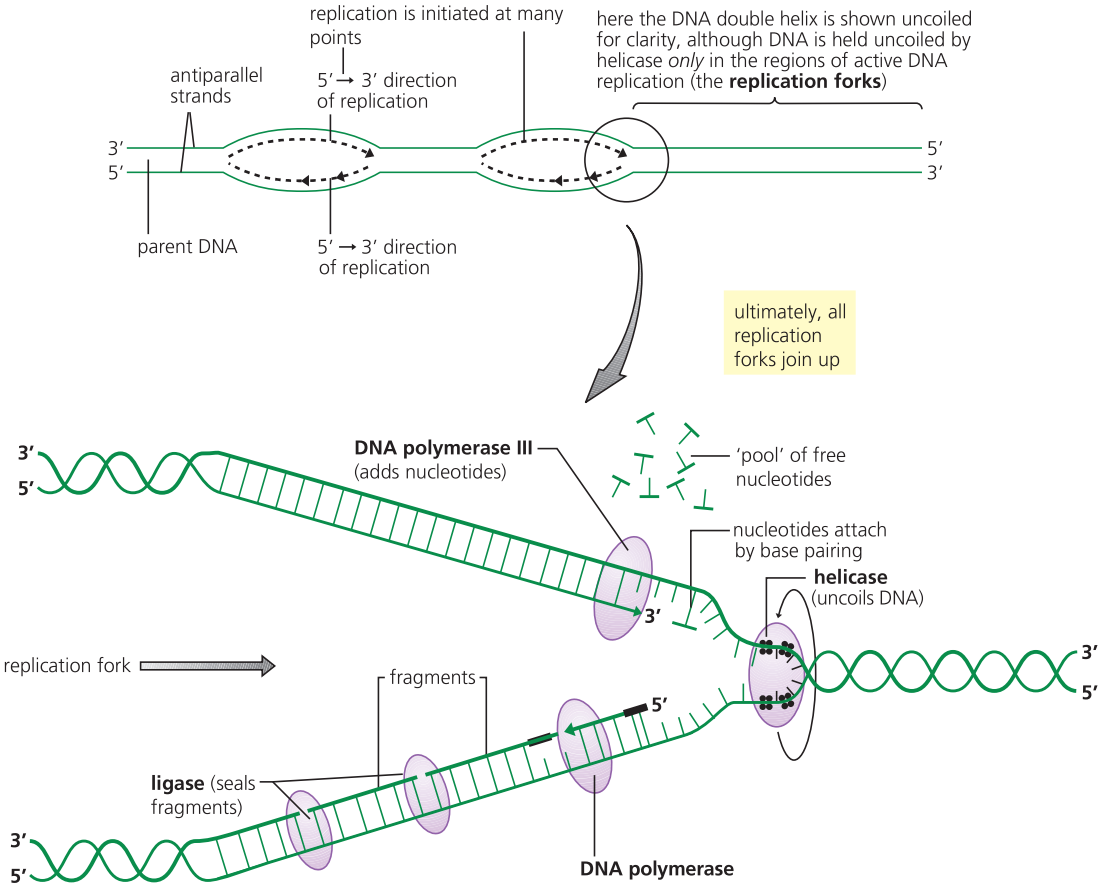


Figure 6.6 DNA replication

the helicase enzyme uncoils the DNA). The copying of the other strand is moving away from the unwinding site. Here, copying has to be frequently restarted. The resulting fragments in this new DNA strand have to be joined up. This is done by another enzyme, known as **ligase**.

DNA polymerase also has a role in **'proof reading'** the new strands. Any 'mistakes' that start to happen (such as the wrong bases attempting to pair up) are immediately corrected so that each new DNA double helix is exactly like the original. Then each pair of double strands winds up into a double helix. The outcome is that one strand of each new double helix came from the parent chromosome and one is a newly synthesised strand. This arrangement known as **semi-conservative replication** because half the original molecule is conserved. We return to this issue shortly.

Question

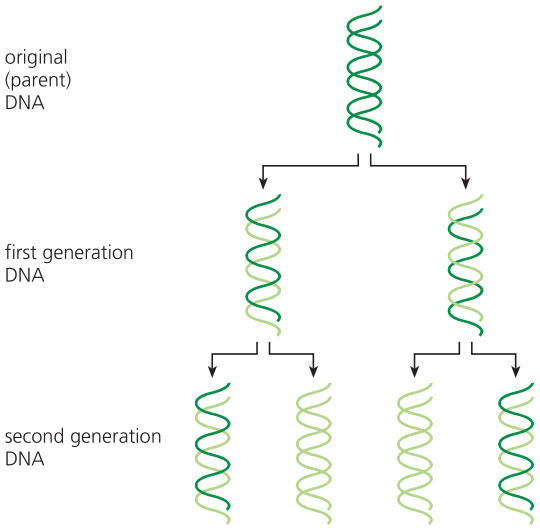
2 Why are many replication forks required in the replication of a single strand of DNA?

The evidence for semi-conservative replication of DNA

By semi-conservative replication we mean that in each DNA molecule formed, one strand is new and one strand is from the original DNA molecule – so is 'old'. On the other hand, if replication was 'conservative', then one DNA molecule formed would be of two new strands and the other molecule would consist of the two originals strands.

Crick and Watson suggested replication of DNA would be 'semi-conservative', and this has since been shown experimentally, using DNA of bacteria 'labelled' with a 'heavy' nitrogen isotope.

In **semi-conservative replication** one strand of each new double helix comes from the parent chromosome and one is a newly synthesised strand (i.e. half the original molecule is conserved).



If an entirely new double helix were formed alongside the original, then one DNA double helix molecule would be conserved without unzipping in the next generation (i.e. **conservative replication**).

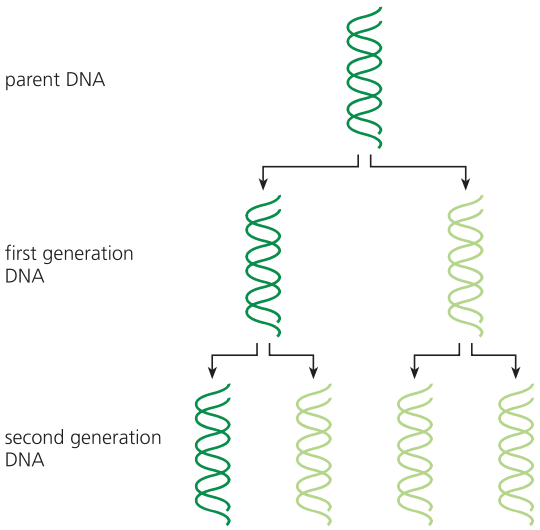


Figure 6.7 Semi-conservative versus conservative replication

The confirmation that DNA is replicated semi-conservatively came from an investigation by Meselson and Stahl. In this experiment, a culture of a bacterium (*Escherichia coli,* page 2) was grown in a medium (food source) where the available nitrogen contained *only* the heavy nitrogen isotope (nitrogen-15, ^{15}N) for many generations. The DNA of the bacterium became entirely 'heavy' as a result. (Isotopes are explained in Appendix I, on the CD).

These bacteria were then transferred to a medium of the normal (light) isotope (nitrogen-14, ^{14}N). The result was that the new DNA manufactured by the cells was now made of nitrogen-14. The experimenters then observed and measured the change in concentration of nitrogen-15 and nitrogen-14 in the DNA of succeeding generations. Note that these observations were possible because the bacterial cell divisions in a culture of *E. coli* are naturally synchronised; every 60 minutes they all divide. The DNA was extracted from samples of the bacteria from each succeeding generation and the DNA in each sample was separated.

Separation of DNA was carried out by placing each sample on top of a salt solution of increasing density in a centrifuge tube. On being centrifuged, the different DNA molecules were carried down to the level where the salt solution was of the same density. Thus, DNA with 'heavy' nitrogen ended up nearer the base of the tubes, whereas DNA with 'light' nitrogen stayed near the top of the tubes. Figure 6.8 shows the results that were obtained.

1 Meselson and Stahl 'labelled' nucleic acid (i.e. DNA) of the bacterium *Escherichia coli* with 'heavy' nitrogen (^{15}**N**), by culturing in a medium where the only nitrogen available was as $^{15}NH_4^+$ ions, for several generations of bacteria.

2 When DNA from labelled cells was extracted and centrifuged in a density gradient (of different salt solutions) all the DNA was found to be 'heavy'.

3 In contrast, the DNA extracted from cells of the original culture (before treatment with ^{15}**N**) was 'light'.

4 Then a labelled culture of *E.coli* was switched back to a medium providing unlabelled nitrogen only, i.e. $^{14}NH_4^+$. Division in the cells was synchronised, and:

- after **one generation** all the DNA was of intermediate density (each of the daughter cells contained (i.e. *conserved*) one of the parental DNA strands containing ^{15}**N** alongside a newly synthesised strand containing DNA made from ^{14}N)
- after **two generations** 50% of the DNA was intermediate and 50% was 'light'. This too agreed with semi-conservative DNA replication, given that labelled DNA was present in only half the cells (one strand per cell).

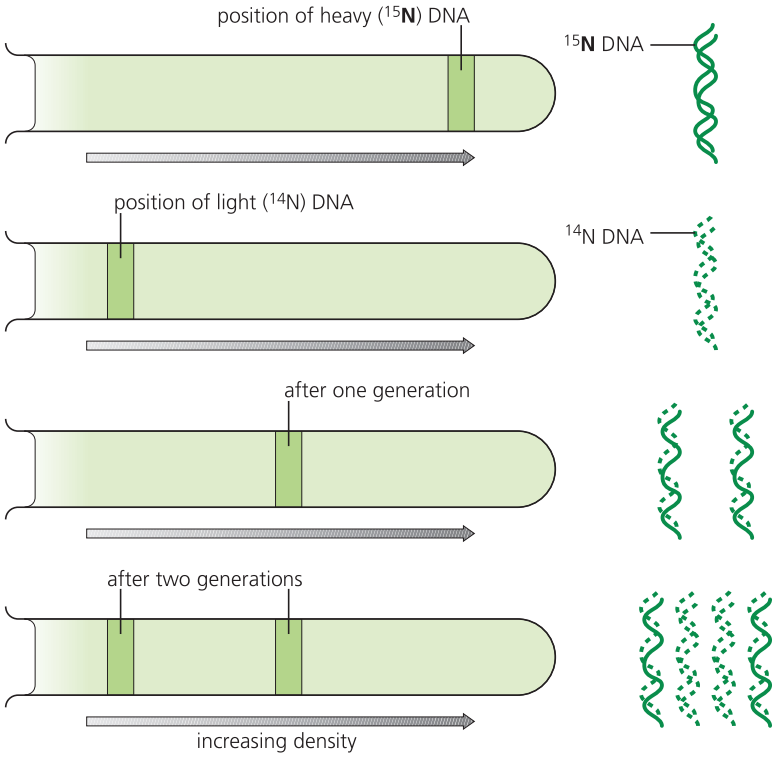


Figure 6.8 Experiment to show that DNA replication is semi-conservative

Question

3 a Predict the experimental results you would expect to see if the Meselson–Stahl experiment (Figure 6.8) was carried on for three generations.

b If replication was by a 'conservative' process, what would be the result of this experiment after one generation in nitrogen-14 (^{14}N)?

6.2 Protein synthesis

The genetic code specifies the amino acids that are assembled to make polypeptides. The way that DNA codes for polypeptides is central to our understanding of how cells and organisms function.

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

a) state that a polypeptide is coded for by a gene and that a gene is a sequence of nucleotides that forms part of a DNA molecule
b) state that a gene mutation is a change in the sequence of nucleotides that may result in an altered polypeptide
c) describe the way in which the nucleotide sequence codes for the amino acid sequence in a polypeptide with reference to the nucleotide sequence for Hb^A (normal) and Hb^S (sickle cell) alleles of the gene for the β-globin polypeptide
d) describe how the information in DNA is used during transcription and translation to construct polypeptides, including the role of messenger RNA (mRNA), transfer RNA (tRNA) and the ribosomes

DNA and the genetic code

Genetic dictionary: a list of the particular base sequences that correspond with particular amino acids. This will vary depending on whether messenger RNA, transfer RNA or either of the two DNA base sequences is given.

In effect, the role of DNA is to instruct the cell to make specific polypeptides and proteins. The huge length of the DNA molecule in a single chromosome carries the codes for a very large number of polypeptides. Within this extremely long molecule, the relatively short length of DNA that codes for a single polypeptide is called a **gene**.

Polypeptides are very variable in size and, therefore, so are genes. A very few genes are as short as 75–100 nucleotides long. Most are at least one thousand nucleotides in length, and some are more.

Polypeptides are built up by condensation of individual amino acids (Figure 2.19, page 46). There are only 20 or so amino acids which are used in protein synthesis; all cell proteins are built from them. The unique properties of each protein lie in:

- which amino acids are involved in its construction
- the sequence in which these amino acids are joined.

The DNA code is in the form of a sequence of the four bases, cytosine (C), guanine (G), adenine (A) and thymine (T). This sequence dictates the order in which specific amino acids are assembled and combined together.

Table 6.2 The case for a triplet code

DNA contains four bases: Adenine, **T**hymine, **G**uanine, **C**ytosine. These are the alphabet of the code.
If one base = one amino acid (singlet code), 4 amino acids can be coded.
If two bases = one amino acid (doublet code), 4^2 (= 16) amino acids can be coded.
If three bases = one amino acid (triplet code), 4^3 (= 64) amino acids can be coded.
If four bases = one amino acid (quadruplet code), 4^4 (= 256) amino acids can be coded.
Conclusion Whilst the doublet code has too few combinations to code for 20 amino acids, the triplet code has too many (44 'spares'!). Each sequence of three bases • in DNA is called a triplet code • in messenger RNA is called a **codon** • in transfer RNA is called an **anticodon**. It has been found that some triplet codes are punctuations (signifying a stop or the end of a gene) and some amino acids are coded by more than one triplet code.

The code lies in the sequence in one of the two strands of the DNA double helix. This is called the **coding strand**. The other strand is complementary to the coding strand. The coding strand is always read in the same direction (from 5′ → 3′).

Despite the fact that there are only four 'letters' to the base code 'alphabet' with which to code 20 amino acids, it was immediately clear that combinations of three bases would be the most likely to code for the amino acids (Figure 6.9). The reason for this is shown in Table 6.2.

So the fact that the code was a three-base one was first deduced. Since then experiments have established that the code is a three-base or **triplet code**. Each sequence of three of the four bases codes for one of the 20 amino acids.

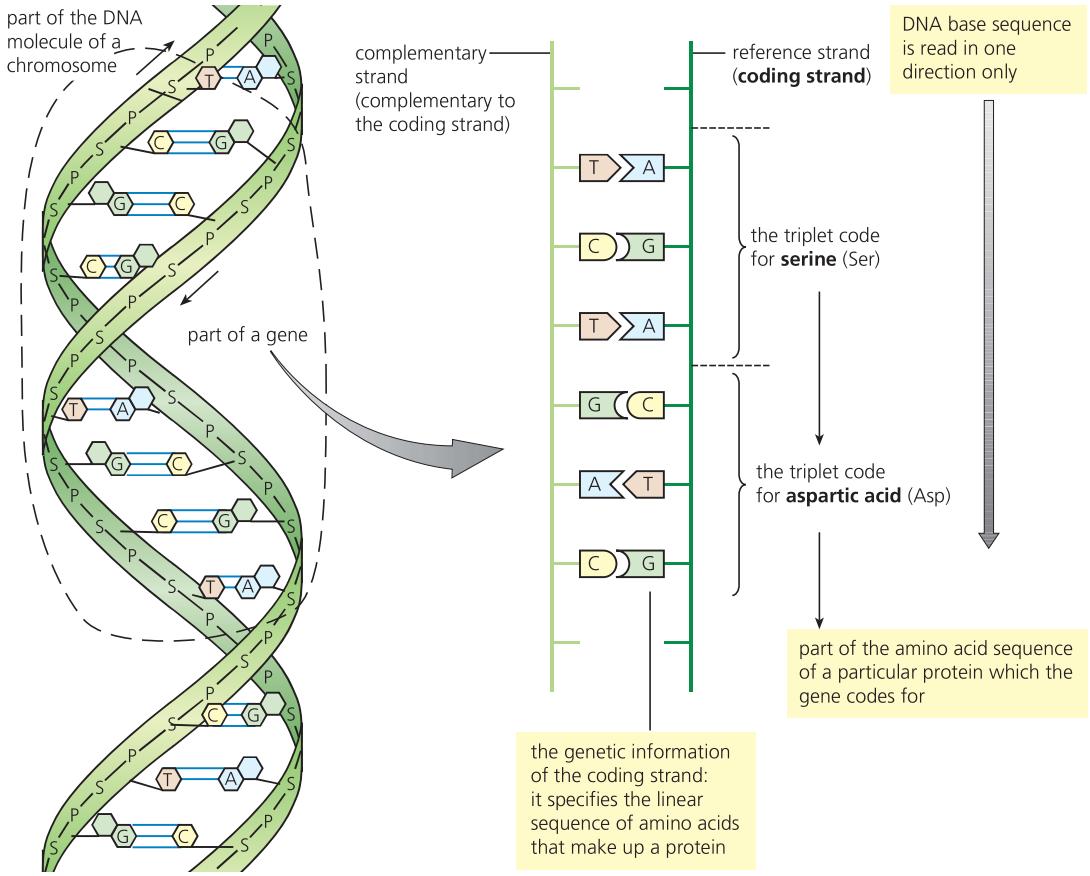


Figure 6.9 Part of a gene and how its DNA codes for amino acids

Amino acid	Abbreviation
alanine	Ala
arginine	Arg
asparagine	Asn
aspartic acid	Asp
cysteine	Cyc
glutamine	Gln
glutamic acid	Glu
glycine	Gly
histidine	His
isoleucine	Ile
leucine	Leu
lysine	Lys
methionine	Met
phenylalanine	Phe
proline	Pro
serine	Ser
threonine	Thr
tryptophan	Trp
tyrosine	Tyr
valine	Val

Most amino acids have two or three similar codons that code for them (Figure 6.10). The DNA triplet code 'TAC' codes for the amino acid methionine. It is always the first triplet code of a gene, so a polypeptide starts with this amino acid. Three of the DNA triplet codes are 'stop' codes, as we shall see.

First base	Second base: A	Second base: G	Second base: T	Second base: C
A	AAA Phe AAG Phe AAT Leu AAC Leu	AGA Ser AGG Ser AGT Ser AGC Ser	ATA Tyr ATG Tyr ATT *Stop* ATC *Stop*	ACA Cys ACG Cys ACT *Stop* ACC Trp
G	GAA Leu GAG Leu GAT Leu GAC Leu	GGA Pro GGG Pro GGT Pro GGC Pro	GTA His GTG His GTT Gln GTC Gln	GCA Arg GCG Arg GCT Arg GCC Arg
T	TAA Ile TAG Ile TAT Ile TAC Met	TGA Thr TGG Thr TGT Thr TGC Thr	TTA Asn TTG Asn TTT Lys TTC Lys	TCA Ser TCG Ser TCT Arg TCC Arg
C	CAA Val CAG Val CAT Val CAC Val	CGA Ala CGG Ala CGT Ala CGC Ala	CTA Asp CTG Asp CTT Glu CTC Glu	CCA Gly CCG Gly CCT Gly CCC Gly

Figure 6.10 The amino acids used in proteins and their DNA triplet codes

Protein synthesis–the stages

So, in the working cell, the information in DNA dictates the structure of polypeptides and proteins, many of which form the enzymes in metabolic process. There are three stages to the process.

Stage one: transcription

This occurs in the nucleus. A complementary copy of the code is made by the building of a molecule of **messenger RNA** (**mRNA**). In the process, the DNA **triplet codes** are transcribed into **codons** in the messenger RNA. This process is called **transcription**. It is catalysed by the enzyme **RNA polymerase** (Figure 6.11).

In transcription, the DNA double helix first unwinds and the hydrogen bonds are broken at the site of the gene being transcribed. There is a pool of free nucleotides nearby. Then, one strand of the DNA molecule, the coding strand, is the template for transcription. This occurs by complementary base pairing. First, the enzyme RNA polymerase recognises and binds to a 'start' signal. The polymerase enzyme then matches free nucleotides by base pairing (A with U, C with G), working in the 5′ → 3′ direction. Note that in RNA synthesis it is uracil which pairs with adenine. Hydrogen bonds then form between complementary bases, holding the nucleotides in place. Each free nucleotide is then joined by a condensation reaction between the sugar and the phosphate groups of adjacent nucleotides of the RNA strand. This is brought about by the action of the enzyme RNA polymerase.

Once the messenger RNA strand is formed it leaves the nucleus through pores in the nuclear envelope and passes to tiny structures in the cytoplasm called **ribosomes** where the information can be 'read' and is used. Then the DNA double stand once more re-forms into a compact helix at the site of transcription.

Question

4 State the sequence of changes catalysed by RNA polymerase.

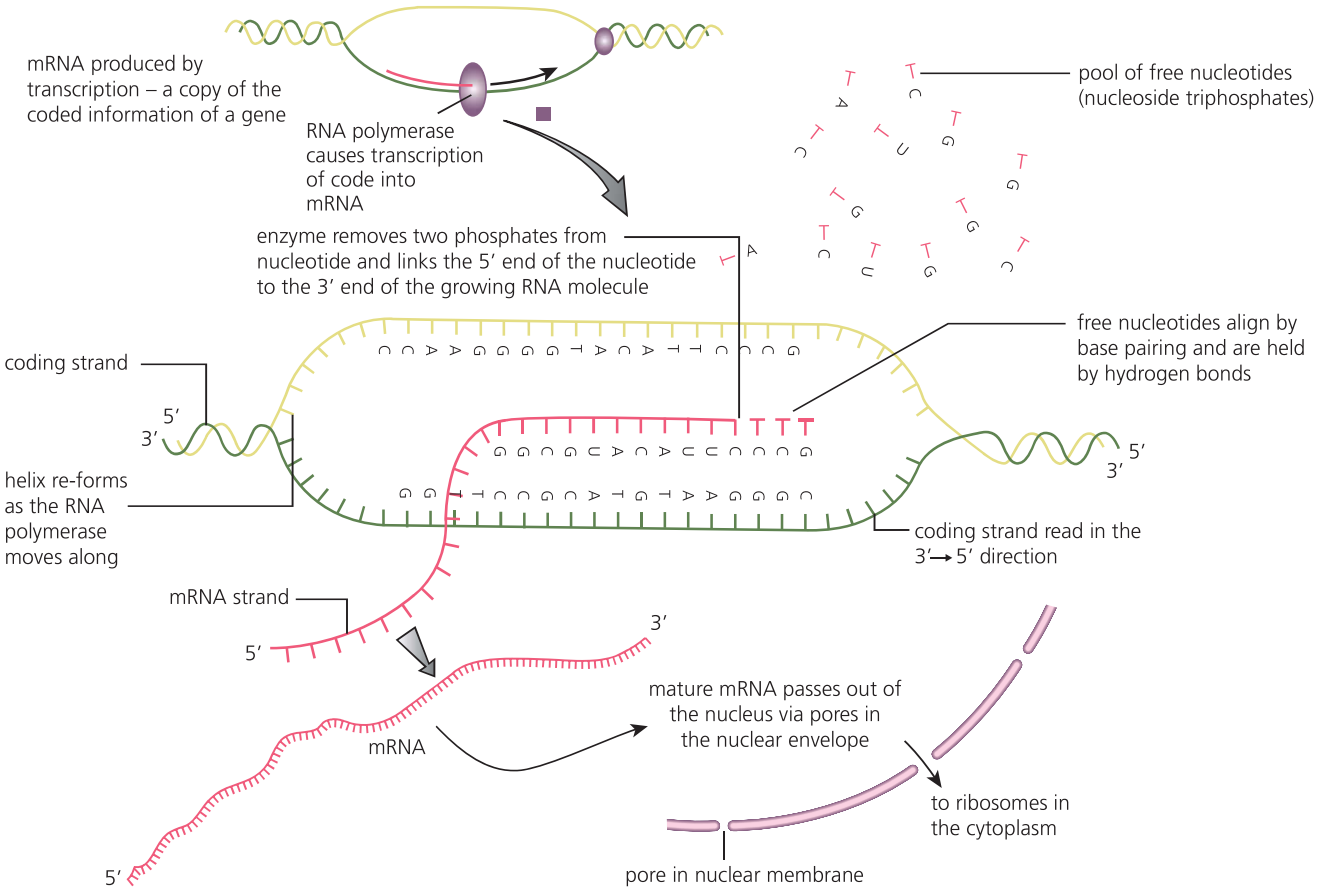


Figure 6.11 Transcription

Stage two: amino acid activation

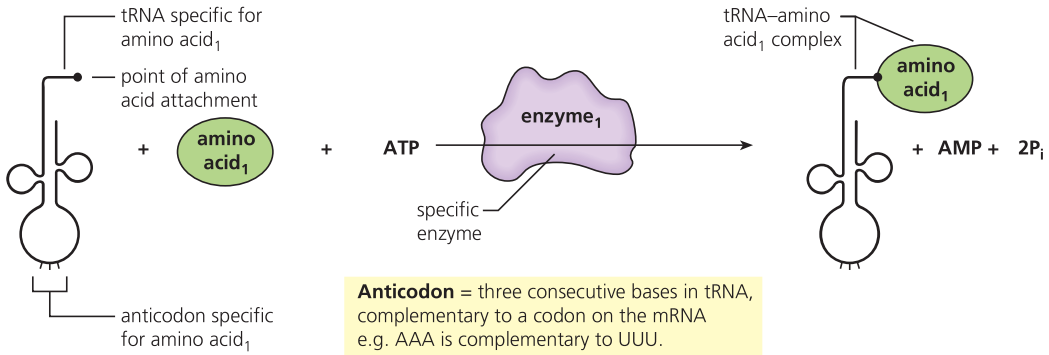
This occurs in the cytoplasm. In stage two, amino acids are activated by combining with short lengths of a different sort of RNA, called **transfer RNA (tRNA)**. Whilst transfer RNAs are relatively short, they are long enough to be folded into the shape of a clover leaf. Their most important feature is that there is a different transfer RNA for each of the 20 amino acids involved in protein synthesis.

So each amino acid has its 'own' type of transfer RNA. At one end of each transfer RNA molecule is a site where the particular amino acid can be joined. The other end of the transfer RNA molecule is folded into a loop. Here, a sequence of three bases forms an **anticodon**. This anticodon is complementary to the codon of messenger RNA that codes for the specific amino acid.

The amino acid is attached to its transfer RNA by an enzyme. These enzymes, too, are specific to the particular amino acids (and types of transfer RNA) to be used in protein synthesis. The specificity of the enzymes is a further way of ensuring the code is implemented correctly and the amino acids are used in the correct sequence.

ATP provides the energy for the attachment of amino acid to its specific tRNA.

Each amino acid is linked to a specific transfer RNA (tRNA) before it can be used in protein synthesis. This is the process of amino acid activation. It takes place in the cytoplasm.



Anticodon = three consecutive bases in tRNA, complementary to a codon on the mRNA e.g. AAA is complementary to UUU.

Figure 6.12 Amino acid activation

Stage three: translation

In stage three a protein chain is assembled by a process called **translation**. This occurs in the **ribosome** (Figure 6.13).

You can see that the ribosome consists of a large and a small subunit. Both parts of the ribosome are intimately involved in the process of translation.

- On arrival at the ribosome, the messenger RNA molecule binds to the small subunit at an attachment site. In this position there are six bases (two codons) of the messenger RNA exposed to the large subunit at any time.
- The first three exposed bases (codon) of messenger RNA are always AUG. A molecule of transfer RNA with the complementary anticodon UAC forms hydrogen bonds with this codon. The amino acid methionine is attached to this transfer RNA molecule.
- A second transfer RNA molecule bonds with the next three bases of the messenger RNA molecule, bringing another amino acid alongside the methionine molecule. Which amino acid this is depends on the second codon, of course.
- Whilst the two amino acids are held close together within the ribosome, a peptide bond is formed between them by condensation reaction. This reaction is catalysed by an enzyme found in the large subunit. A dipeptide has been formed.
- Now the ribosome moves along the messenger RNA molecule in the 5' → 3' direction and the next codon is read. At the same time, the first transfer RNA molecule (with anticodon UAC) leaves the ribosome, minus it amino acid.

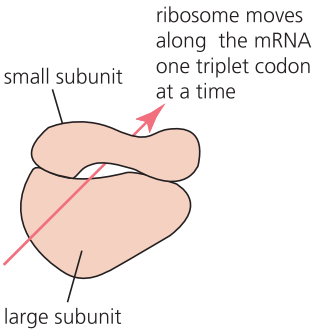


Figure 6.13 Ribosome structure

Question

5 Draw and label the structure of a peptide bond between two amino acids.

- Then a third transfer RNA molecule bonds with the next codon of the messenger RNA molecule, bringing another amino acid alongside the second amino acid residue of the dipeptide. Immediately, a peptide bond is formed between them by a condensation reaction. A tripeptide has been formed and starts to emerge from a hole within the large subunit.
- Again, the ribosome moves along the messenger RNA molecule and the next codon is read. At the same time, the second transfer RNA molecule leaves the ribosome, minus its amino acid.
- A fourth transfer RNA brings another amino acid to lie alongside the third amino acid residue. Whilst these amino acids are held close together another peptide bond is formed.
- By these steps, constantly repeated, a polypeptide is formed and emerges from the large subunit. Eventually a 'stop' codon is reached. This takes the form of one of three codons – UAA, UAG or UGA. At this point the completed polypeptide is released from the ribosome into the cytoplasm.

The steps of translation are shown in Figure 6.15.

Where ribosomes occur in cells

Question

6 What different forms of RNA are involved in 'transcription' and 'translation' and what are the roles of each?

Many ribosomes occur freely in the cytoplasm. These are the sites of synthesis of proteins that are to remain in the cell and fulfil particular roles there. It is common for several ribosomes to move along the messenger RNA at one time; the structure (messenger RNA, ribosomes and their growing protein chains) is called a **polysome** (Figure 6.14).

Other ribosomes are bound to the membranes of the endoplasmic reticulum (known as rough endoplasmic reticulum, RER) and these are the site of synthesis of proteins that are subsequently secreted from cells or packaged in lysosomes there.

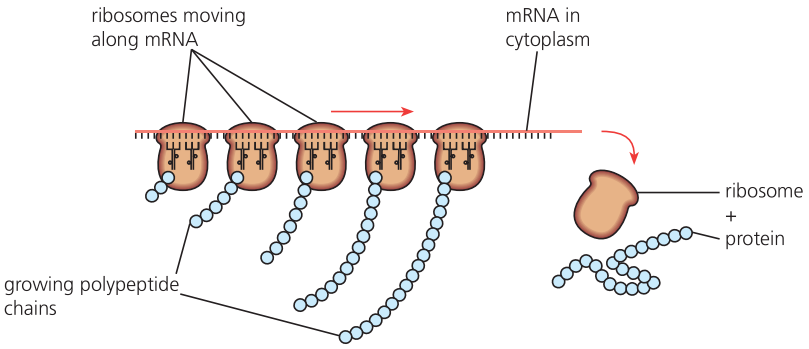


Figure 6.14 A polysome

Triplet codes, codons and anticodons

The sequence of bases along DNA molecules is the **genetic code**. This sequence ultimately determines the sequence of amino acids from which polypeptides are constructed.

- The sequence of bases in the DNA molecule is transcribed to a complementary sequence of bases in messenger RNA. DNA 'codes' are referred to as **triplet codes** and those in messenger RNA are called **codons**.
- Free amino acids in the cytoplasm attach to specific transfer RNA molecules. There is one transfer RNA molecule for each of the twenty different amino acids. The triplet of bases of a transfer RNA molecule is known as an **anticodon**.
- In ribosomes, transfer RNA anticodons align with RNA codons by **complementary base pairing**. This brings amino acids side by side and peptide bonds are formed between them. A polypeptide molecule is formed.

By these steps, the DNA code is translated into an amino acid sequence in a polypeptide. There are different ways of presenting the genetic code and this is the significance of the term '**genetic dictionary**'. The genetic dictionary is a list of the particular base sequences that correspond with particular amino acids. It depends on whether it is the code in DNA, messenger RNA (or transfer RNA) that is quoted.

Can you say why?

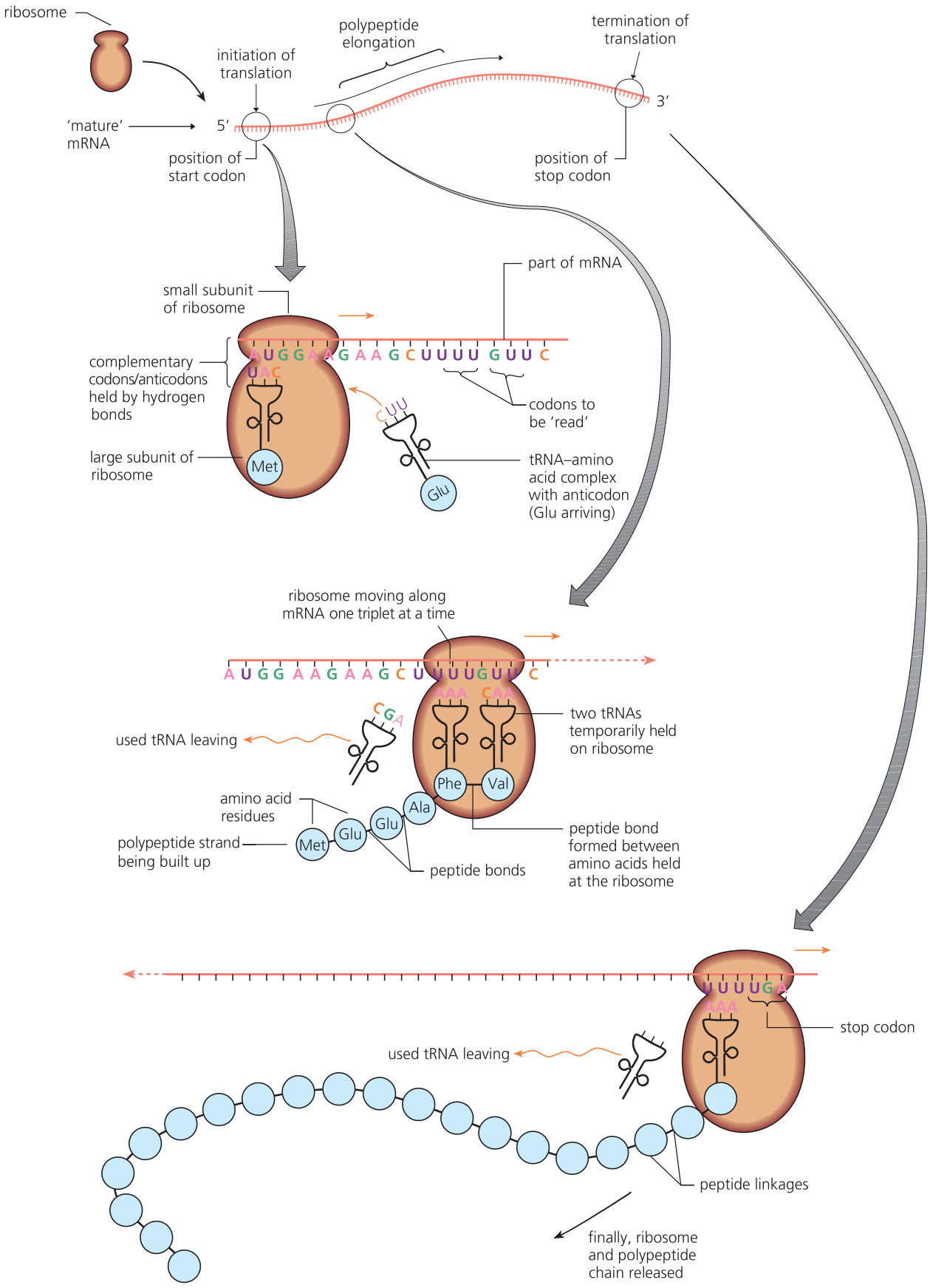


Figure 6.15 Translation

Amino acid	Abbreviation
alanine	Ala
arginine	Arg
asparagine	Asn
aspartic acid	Asp
cysteine	Cyc
glutamine	Gln
glutamic acid	Glu
glycine	Gly
histidine	His
isoleucine	Ile
leucine	Leu
lysine	Lys
methionine	Met
phenylalanine	Phe
proline	Pro
serine	Ser
threonine	Thr
tryptophan	Trp
tyrosine	Tyr
valine	Val

The DNA triplet codes are recorded in Figure 6.10 whilst those of messenger RNA are listed in Figure 6.16. These codes do not need memorising but their relationship is important and you do need to be able to use them.

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

Figure 6.16 The messenger RNA genetic dictionary

Extension

DNA also codes for the RNA of cells

The DNA of the chromosomes codes for all the RNA molecules that a cell contains, as well as the proteins. There is a different enzyme system that 'reads' the DNA that specifically codes for these RNA molecules, additional to the enzymes that catalyse the formation of messenger RNA.

Question

7 A sequence of bases in a sample of messenger RNA was found to be:
GGU, AAU, CCU, UUU, GUU, ACU, CAU, UGU.

a What sequence of amino acids does this code for?

b What was the sequence of bases in the coding strand of DNA from which this messenger RNA was transcribed?

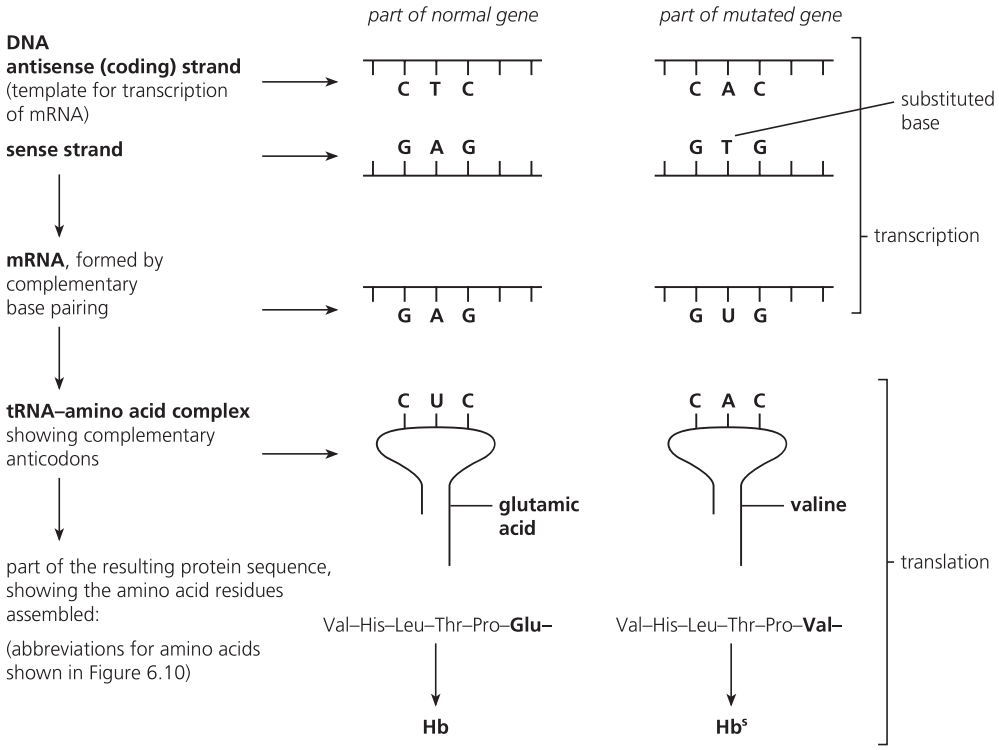
c Within a cell, where are triplet codes, codons and anticodons found?

DNA can change

We have seen that a gene is a sequence of nucleotide pairs which codes for a sequence of amino acids. Normally, the sequence of nucleotides in the DNA is maintained without changing but very occasionally it does. If a change occurs we say a mutation has occurred.

A **gene mutation** involves a change in the sequence of bases of a particular gene. At certain times in the cell cycle mutations are more likely than at other times. One such occasion is when the DNA molecule is replicating. We have noted that the enzyme DNA polymerase that brings about the building of a complementary DNA strand also 'proof reads' and corrects most errors. However, gene mutations can and do occur spontaneously during this step. Also, certain conditions or chemicals may cause change to the DNA sequence of bases (Table 5.2, page 105).

The **mutation** that produces sickle cell haemoglobin (**Hg^S**) is in the gene for β-haemoglobin. It results from the substitution of a single base in the sequence of bases that make up all the codons for β-haemoglobin.



drawing based on a photomicrograph of a blood smear, showing blood of a patient with sickle cells present among healthy red blood cells

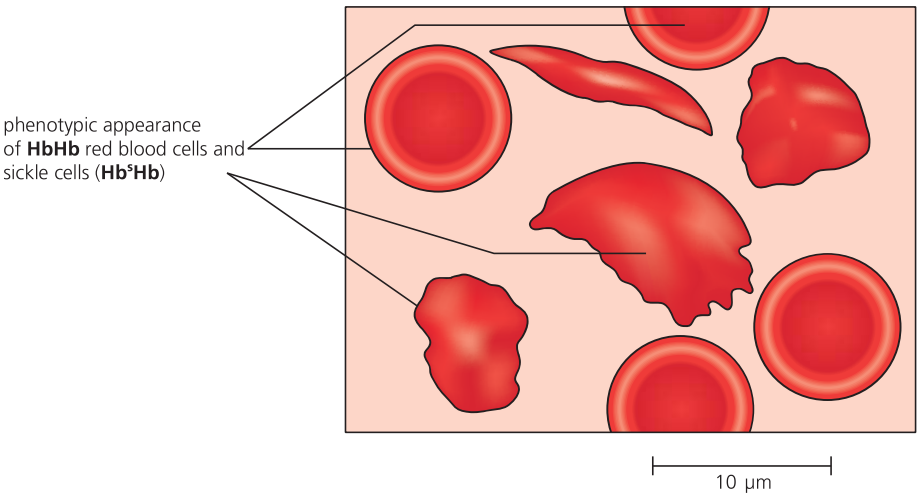


Figure 6.17 Sickle cell anaemia – an example of a gene mutation

An example of a gene mutation occurs in the condition known as **sickle cell anaemia**. Here the gene that codes for the amino acid sequence of a part of the respiratory pigment haemoglobin of red blood cells occasionally mutates at one base pair (Figure 6.17). The result of this change is that the base adenine is replaced by thymine at one position along the coding strand. The messenger RNA produced from this DNA contains the triplet code GUG in place of GAG. This causes the amino acid valine (a non-polar amino acid) to appear at that point, instead of polar glutamic acid. The presence of non-polar valine in the beta chain of haemoglobin gives a hydrophobic spot in the otherwise hydrophilic outer section of the protein. This tends to attract other haemoglobin molecules to bind to it.

When a person who is a sickle cell 'carrier' undertakes sudden physical exercise (or moves to a high altitude) the oxygen content of the blood is lowered. In these conditions the sickle cell haemoglobin molecules readily clump together into long fibres. These fibres distort the red blood cells into sickle shapes. In this condition the red blood cells cannot transport oxygen. Also, sickle cells get stuck together, blocking smaller capillaries and preventing the circulation of normal red blood cells. The result is that people with sickle cell trait suffer from anaemia – a condition of inadequate delivery of oxygen to cells.

Summary

- **Nucleic acids** are **polymers**. They are polynucleotides composed of long chains of nucleotide **monomers** (a pentose sugar, a phosphate molecule and a nitrogenous base) combined together. There are two forms of nucleic acid, **DNA** and **RNA**. These forms differ in the sugar and in the bases they contain, and in their **roles in the cell**. DNA is found in the chromosomes and RNA in both the nucleus and the cytoplasm. DNA also occurs in mitochondria and chloroplasts.
- **ATP** is the universal energy currency molecule by which energy is transferred to do useful work. ATP is a soluble molecule, formed in the mitochondria but able to move into the cytosol by facilitated diffusion. It diffuses freely about cells.
- The **sequence of four bases**, adenine (**A**), thymine (**T**), guanine (**G**) and cytosine (**C**), along a DNA molecule is the **genetic code**. This is a three-letter code – three of the four bases code for a particular amino acid. The sequence of bases in the DNA molecule may be transcribed to a complementary sequence in **messenger RNA (mRNA)**. Once formed, messenger RNA passes out of the nucleus into the cytoplasm and to a **ribosome** where **polypeptide synthesis** occurs. DNA 'codes' are referred to as **triplet codes** and those in messenger RNA are called **codons**.
- **Transfer RNA (tRNA)** molecules occur in the cytoplasm. These molecules have a particular triplet of bases at one end (known as an **anticodon**) and a point of attachment for a **specific amino acid** at the other end. The role of transfer RNA is to pick up free amino acids and bring them to the messenger RNA in the ribosome. Complementary base pairing between the codons of messenger RNA and the anticodons of transfer RNA results in the original DNA code being translated into an amino acid sequence in the process of **polypeptide and protein synthesis**.
- There are **three stages** to polypeptide and protein synthesis.
  - In the nucleus, a copy of the genetic code of a gene is made in the form of a single strand of **messenger RNA** in the stage called **transcription**. The messenger RNA passes out into the cytoplasm.
  - In the cytoplasm **amino acid activation** occurs. Each of the 20 amino acids is attached to a specific **transfer RNA** molecule.
  - Finally a **new polypeptide** is assembled when the information of the messenger RNA is **translated** ('read') in a ribosome and the transfer RNA molecules with their attached amino acids are aligned in the order given in the code. Peptide bonds are formed between the amino acids.

Examination style questions

1 Fig. 1.1 shows the replication of one strand of a DNA double helix.

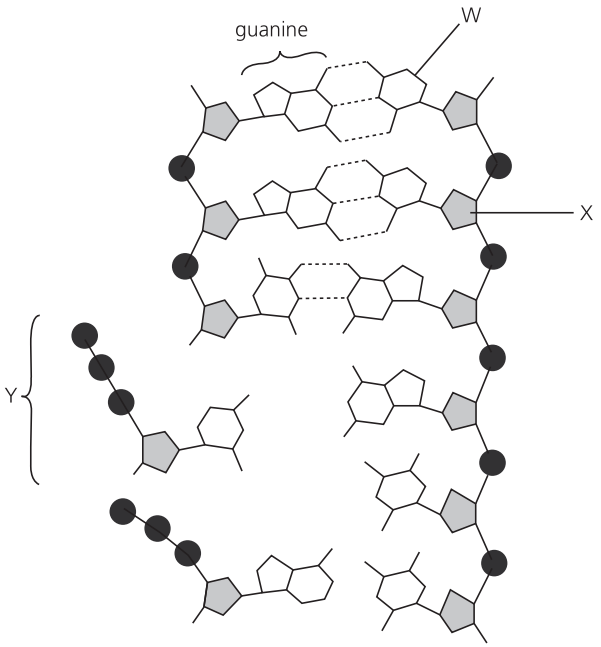


Fig. 1.1

a) Name **W** to **Y**. [3]

b) Explain how the structure of DNA enables it to replicate semi-conservatively. [3]

c) Explain why it is important that an exact copy of DNA is made during replication. [2]

[Total: 8]

(Cambridge International AS and A Level Biology 9700, Paper 02 Q1 June 2005)

2 **a)** A gene in the form of a double-stranded length of DNA contains 12 000 nucleotides. Explain the maximum number of amino acids that could be required to transcribe it. [6]

b) What may the triplet codes of a gene represent other than the amino acids of the polypeptide that it codes for? [2]

c) The following is a short sequence of the triplet codes for part of the primary structure of a polypeptide:

AAA ATA GTA TAA CAC TGC CTC TCG

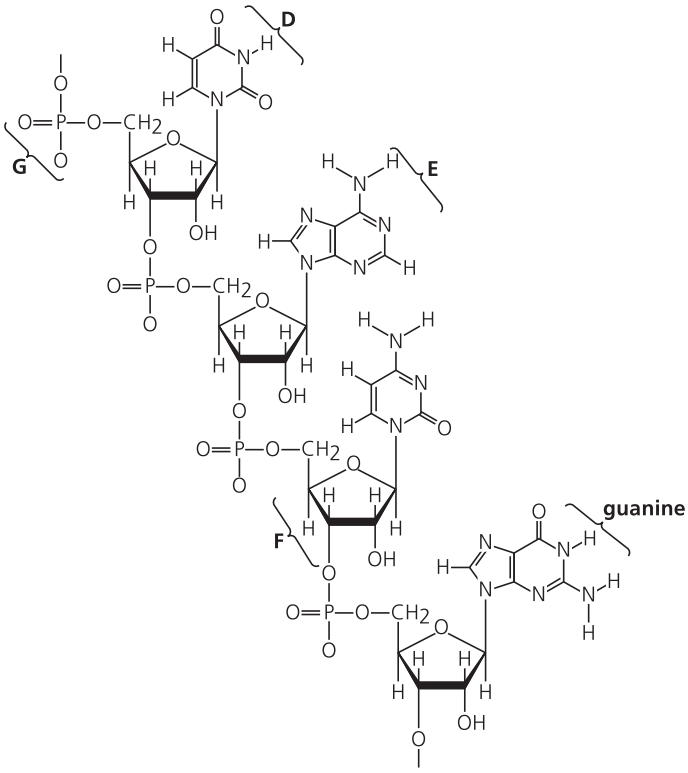
Give the sequence of codons of mRNA transcribed from this part of the gene, and the sequence of anti-codons that will base pair with them. [8]

d) How does the structure of tRNA differ from that of mRNA? [4]

[Total: 20]

3 A molecule of messenger RNA (mRNA) was produced during the transcription of a gene. Part of the template sequence of DNA was ATGC.

The diagram shows the part of the molecule of messenger RNA corresponding to that sequence of four bases.



a) Name the parts of the mRNA molecule shown in the diagram above labelled **D**, **E**, **F** and **G**. [4]

b) Copy and complete the table to show **three** ways in which mRNA differs from DNA. [3]

	mRNA	DNA
1		
2		
3		

c) Describe the role of mRNA after it leaves the nucleus and enters the cytoplasm of a eukaryotic cell. [4]

[Total: 11]

(Cambridge International AS and A Level Biology 9700, Paper 21 Q3 June 2011)