

6

Nucleic acids: structure and function

In this chapter we discuss the structure and function of nucleic acids. Drug action at nucleic acids is discussed in chapter 9 and in other chapters throughout the text. Although most drugs act on protein structures, there are several examples of important drugs which act directly on nucleic acids. There are two types of nucleic acid – DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). We first consider the structure of DNA.

6.1 Structure of DNA

Like proteins, DNA has a primary, secondary, and tertiary structure.

6.1.1 Primary structure of DNA

The primary structure of DNA is the way in which the DNA building blocks are linked together. Whereas proteins have over 20 building blocks to choose from, DNA has only four—the nucleosides **deoxyadenosine**, **deoxyguanosine**, **deoxycytidine**, and **deoxythymidine** (Fig. 6.1). Each nucleoside is constructed from two components—a **deoxyribose** sugar and a base. The sugar is the same in all four nucleosides and only the base is

different. The four possible bases are two bicyclic purines (**adenine** and **guanine**) and two smaller pyrimidine structures (**cytosine** and **thymine**) (Fig. 6.2).

The nucleoside building blocks are joined together through phosphate groups which link the 5'-hydroxyl group of one nucleoside unit to the 3'-hydroxyl group of the next (Fig. 6.3). With only four types of building block available, the primary structure of DNA is far less varied than the primary structure of proteins. As a result, it was long thought that DNA had only a minor role to play in cell biochemistry, since it was hard to see how such an apparently simple molecule could have anything to do with the mysteries of the genetic code. The solution to this mystery lies in the secondary structure of DNA.

6.1.2 Secondary structure of DNA

Watson and Crick solved the secondary structure of DNA by building a model that fitted all the known experimental results. The structure consists of two DNA chains arranged together in a double helix of constant diameter (Fig. 6.4). The double helix can be seen to have a major groove and a minor groove, which are of some importance to the action of several anticancer agents acting as intercalators (section 9.1).

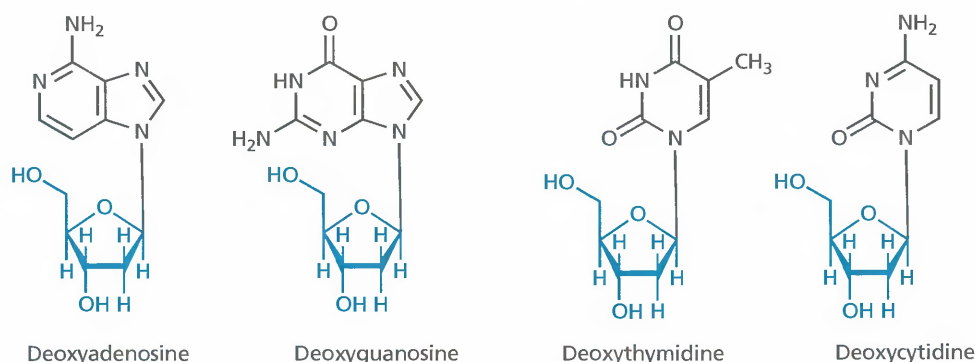


FIGURE 6.1 Nucleosides—the building blocks of DNA.

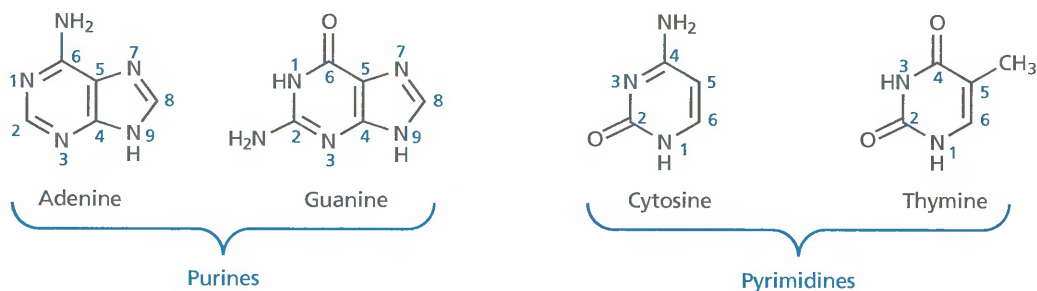


FIGURE 6.2 The nucleic acid bases for DNA.

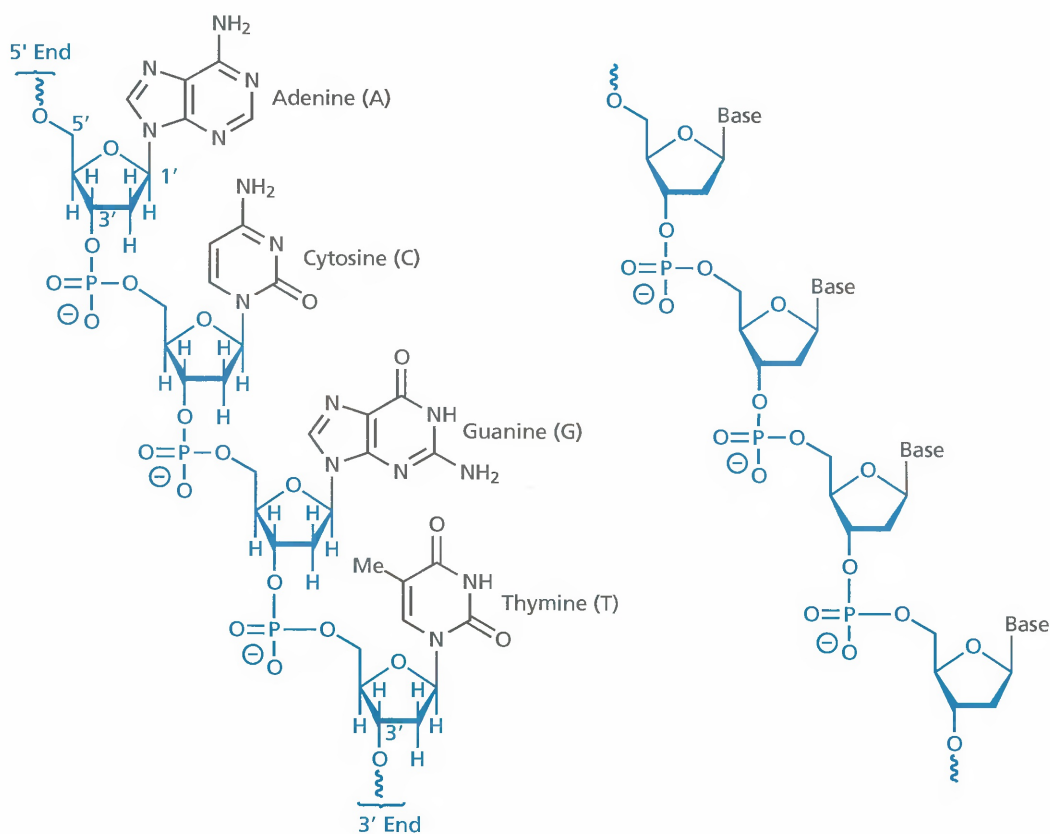


FIGURE 6.3 Linkage of nucleosides through phosphate groups.

The structure relies crucially on the pairing up of nucleic acid bases between the two chains. Adenine pairs with thymine via two hydrogen bonds, whereas guanine pairs with cytosine via three hydrogen bonds. Thus, a bicyclic purine base is always linked with a smaller monocyclic pyrimidine base to allow the constant diameter of the double helix. The double helix is further stabilized by the fact that the base pairs are stacked one on top of each other, allowing hydrophobic interactions between the faces of the heterocyclic rings. The polar sugar–phosphate backbone is placed to the outside of the structure and can form favourable polar interactions with water.

The fact that adenine always binds to thymine, and cytosine always binds to guanine means that the chains are complementary to each other. It is now possible to see how **replication** (the copying of genetic information) is feasible. If the double helix unravels, a new chain can be constructed on each of the original chains (Fig. 6.5). In other words, each of the original chains acts as a template for the construction of a new and identical double helix. The mechanism by which this takes place is shown in Figs. 6.6 and 6.7. The template chain has exposed bases which can base pair by hydrogen bonding with individual nucleotides in the form of triphosphates. Once a

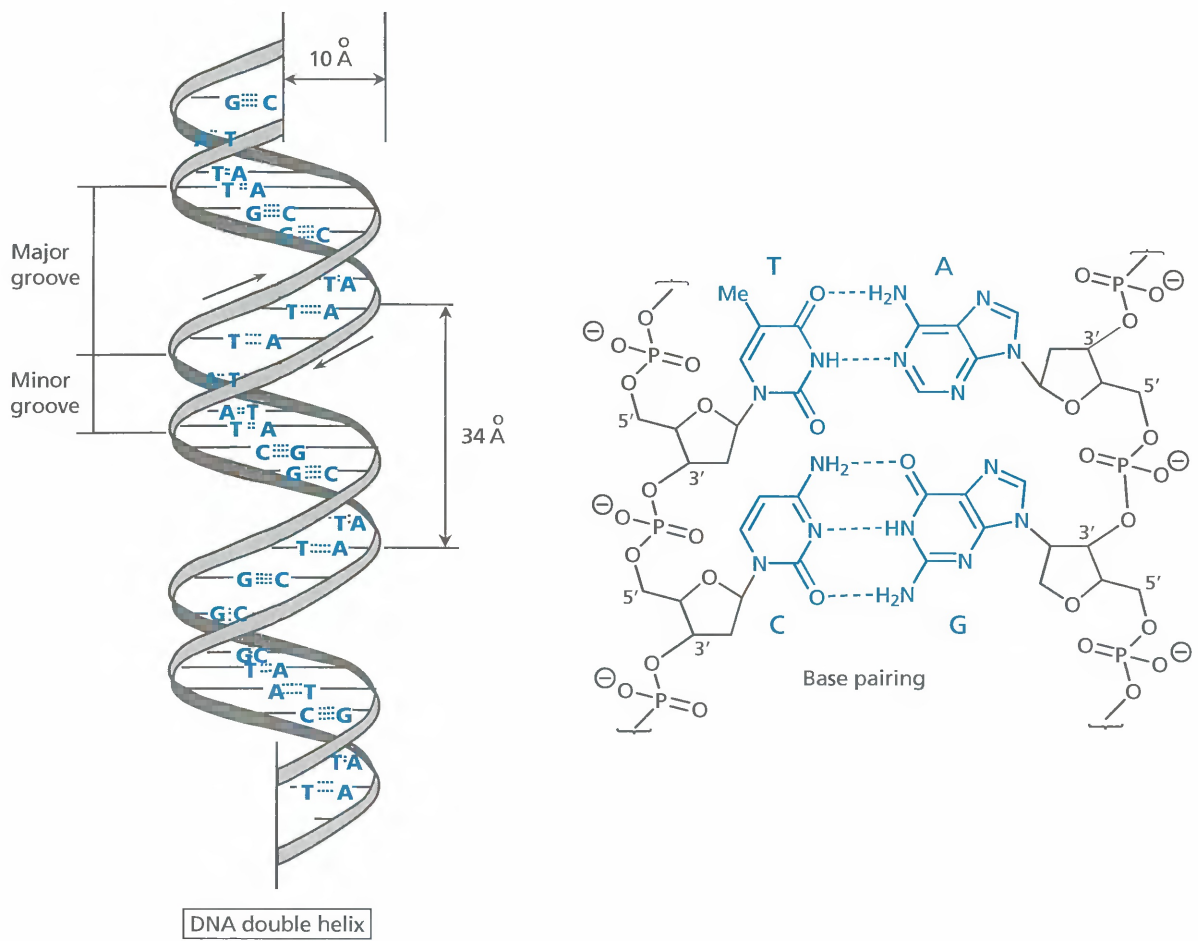


FIGURE 6.4 The secondary structure of DNA.

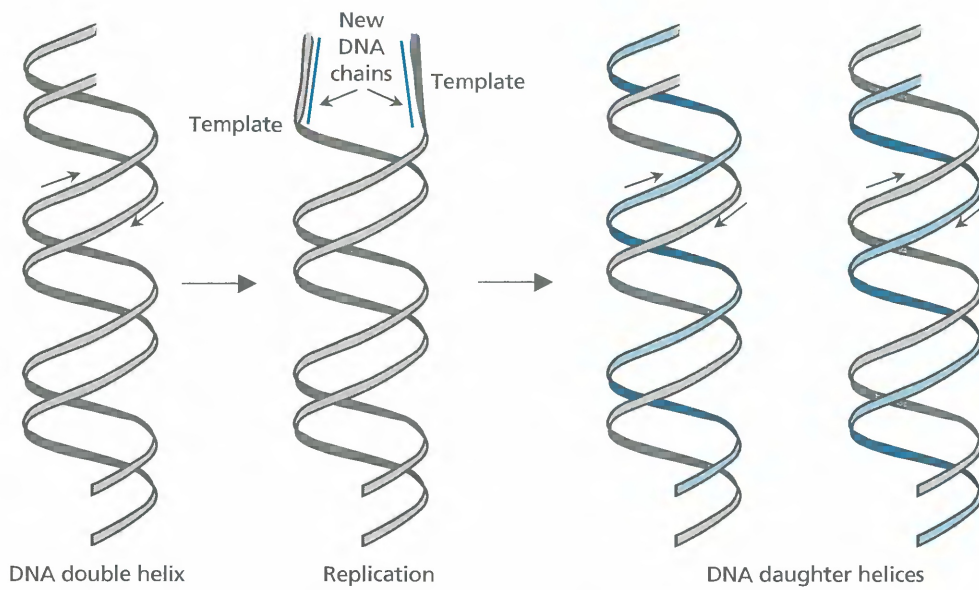


FIGURE 6.5 Replication of DNA chains.

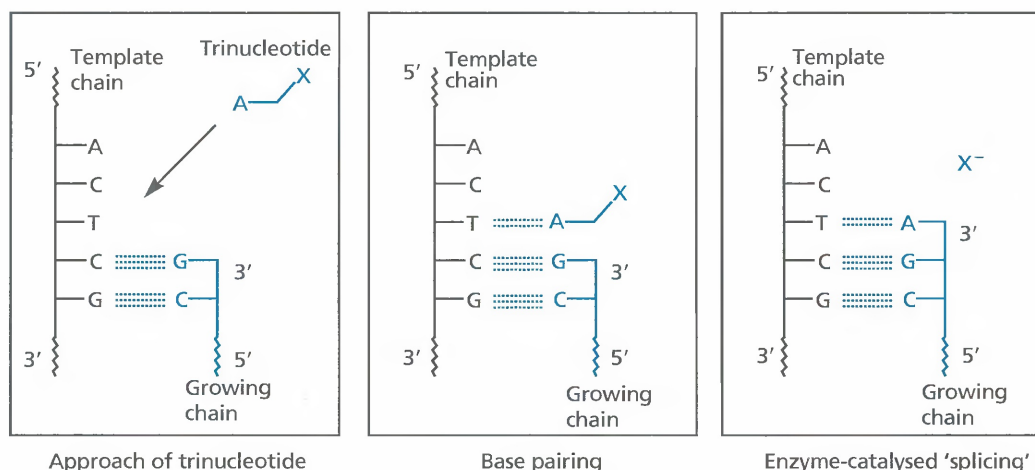


FIGURE 6.6 Base pairing of a trinucleotide and extension of the growing DNA chain.

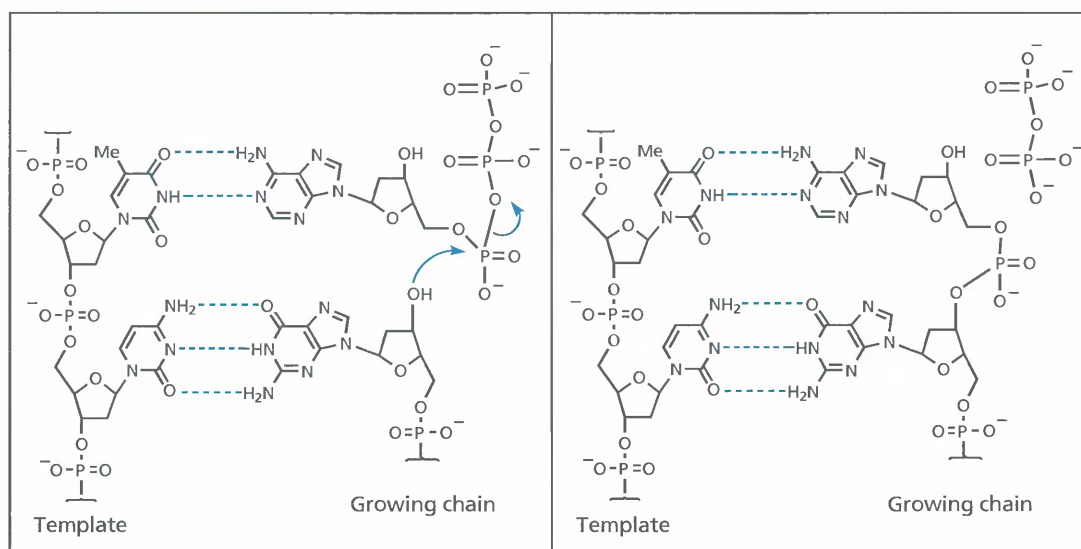


FIGURE 6.7 Mechanism by which a nucleotide is linked to the growing DNA chain.

nucleotide has base paired, an enzyme-catalysed reaction takes place where the new nucleotide is spliced on to the growing complementary chain with the loss of a diphosphate group—the latter acting as a good leaving group. Note that the process involves each new nucleotide reacting with the 3' end of the growing chain.

We can now see how genetic information is passed on from generation to generation, but it is less obvious how DNA codes for proteins. How can only four nucleotides code for over 20 amino acids? The answer lies in the **triplet code**. In other words, an amino acid is coded not by one nucleotide, but by a set of three. There are 64 (4^3) ways in which four nucleotides can be arranged in sets of three – more than enough for the task required. Appendix 2 shows the standard genetic code for the various

triplets. We shall look at how this code is interpreted to produce a protein in section 6.2.

6.1.3 Tertiary structure of DNA

The tertiary structure of DNA is often neglected or ignored, but it is important to the action of the quinolone group of antibacterial agents (section 9.2) and to several anticancer agents (sections 9.1–9.2). DNA is an extremely long molecule, so long in fact that it would not fit into the nucleus of the cell if it existed as a linear molecule. It has to be coiled into a more compact three-dimensional shape which *can* fit into the nucleus—a process known as **supercoiling**. This process requires the action of a family of enzymes called **topoisomerases**, which can catalyse the

seemingly impossible act of passing one stretch of DNA helix across another stretch. They do this by temporarily cleaving one or both strands of the DNA helix to create a temporary gap, then resealing the strand(s) once the crossover has taken place. Supercoiling allows the efficient storage of DNA, but the DNA has to be uncoiled again if replication and transcription (section 6.2.2) are to take place. If uncoiling did not take place, the unwinding process (catalysed by **helicase** enzymes) that takes place during replication and transcription would lead to increased tension due to increased supercoiling of the remaining DNA double helix. You can demonstrate the principle of this by pulling apart the strands of rope or sisal. The same topoisomerase enzymes are responsible for catalysing the

uncoiling process, so inhibition of these enzymes would effectively block transcription and replication.

Topoisomerase II is a mammalian enzyme that is crucial to the effective replication of DNA. The enzyme binds to parts of DNA where two regions of the double helix are in near proximity (Fig. 6.8). The enzyme binds to one of these DNA double helices and a tyrosine residue is used to nick both strands of the DNA (Fig. 6.9). This results in a temporary covalent bond between the enzyme and the resulting 5' end of each strand, thus stabilizing the DNA. The strands are now pulled in opposite directions to form a gap through which the intact DNA region can be passed. The enzyme then reseals the strands and departs.

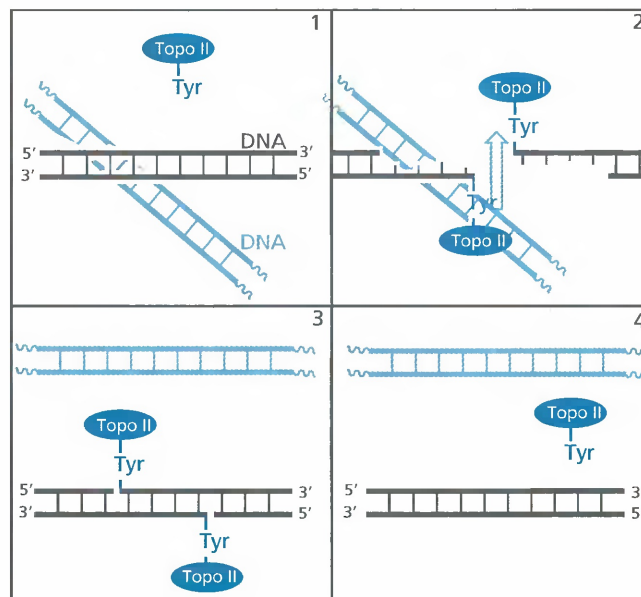


FIGURE 6.8 Method by which topoisomerase II catalyses the crossover of DNA strands.

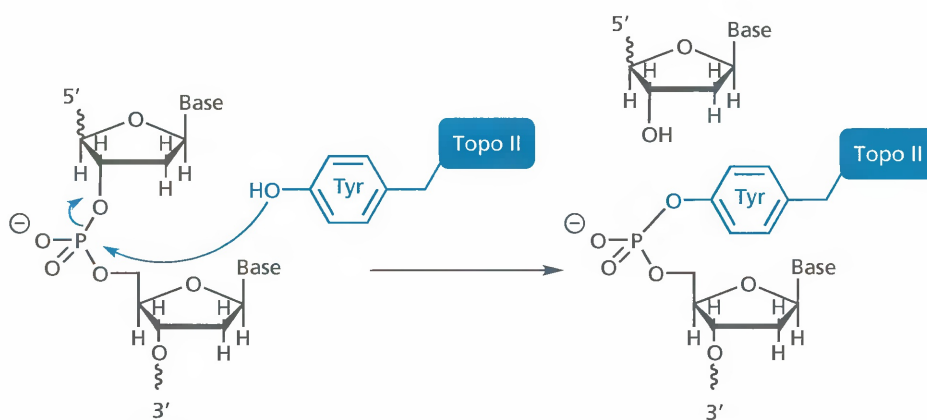


FIGURE 6.9 Mechanism by which topoisomerase II splits a DNA chain.

Topoisomerase I is similar to topoisomerase II in that it relieves the torsional stress of supercoiled DNA during replication, transcription, and the repair of DNA. The difference is that it cleaves only one strand of DNA, whereas topoisomerase II cleaves both strands. The enzyme catalyses a reversible transesterification reaction similar to that shown in Fig. 6.9, but where the tyrosine residue of the enzyme is linked to the 3' phosphate end of the DNA strand rather than the 5' end. This creates a 'cleavable complex' with a single-strand break. Relaxation of torsional strain takes place either by allowing the intact strand to pass through the nick or by free rotation of the DNA about the uncleaved strand. Once the torsional strain has been relieved, the enzyme rejoins the cleaved strand of DNA and departs.

Topoisomerase IV is a bacterial enzyme that carries out the same process as the mammalian enzyme topoisomerase II, and is an important target for the fluoroquinolone antibacterial agents (section 9.2).

6.1.4 Chromatins

So far we have focused on the structure of DNA. However, DNA is not an isolated macromolecule within the nucleus of the cell. It is associated with a variety of proteins such as histones in a structure called a chromatin (Fig. 21.5). The histones and associated DNA form a structure called a nucleosome which occurs regularly along the length of the chromatin and plays a crucial role in the regulation of DNA transcription (section 21.7.4).

6.1.5 Genetic polymorphism and personalized medicine

The process of replication is not 100% perfect and occasionally a mutation can occur. If the mutation does not prove fatal, it will be carried on from generation to generation. This leads to different individuals having subtly different gene sequences. On average, there is a difference of one base pair in every thousand base pairs between individuals, and this is known as genetic polymorphism. Since the nucleic acid bases act as the code for amino acids in proteins, a difference at this level results in a different amino acid being introduced into a protein, which may or may not have an effect on that protein's activity or function (sections 3.5.6 and 4.11). Genetic polymorphism has important consequences with respect to the susceptibility of individuals to disease, and also to the kinds of drug therapies that are best suited for individuals. A detailed knowledge of a patient's genome opens up the possibility of predicting and preventing disease, as well as choosing the ideal drug therapy for that patient should a disease occur. This is known as **personalized medicine** (see also section 21.1.11).

KEY POINTS

- The primary structure of DNA consists of a sugar–phosphate backbone with nucleic acid bases attached to each sugar moiety. The sugar is deoxyribose and the bases are adenine, thymine, cytosine, and guanine.
- The secondary structure of DNA is a double helix where the nucleic acid bases are stacked in the centre and paired up such that adenine pairs with thymine, and cytosine pairs with guanine. Hydrogen bonding is responsible for the base pairing and there are van der Waals interactions between the stacks of bases. Polar interactions occur between the sugar phosphate backbone and surrounding water.
- The DNA double helix is coiled up into a tertiary structure. The coiling and uncoiling of the double helix requires topoisomerase enzymes.
- The copying of DNA from one generation to the next is known as replication. Each strand of a parent DNA molecule acts as the template for a new daughter DNA molecule.
- The genetic code consists of nucleic acid bases, which are read in sets of three during the synthesis of a protein. Each triplet of bases codes for a specific amino acid.
- Knowing a patient's genome opens up the possibility of predicting disease and identifying the best therapies for that individual. This is known as personalized medicine.

6.2 Ribonucleic acid and protein synthesis

6.2.1 Structure of RNA

The primary structure of RNA is the same as that of DNA, with two exceptions: **ribose** (Fig. 6.10) is the sugar component rather than **deoxyribose**, and **uracil** (Fig. 6.10) replaces thymine as one of the bases.

Base pairing between nucleic acid bases can occur in RNA, with adenine pairing to uracil, and cytosine pairing to guanine. However, the pairing is between bases within the same chain, and it does not occur for the whole length of the molecule (e.g. Fig. 6.11). Therefore, RNA is not a double helix, but it does have regions of helical secondary structure.

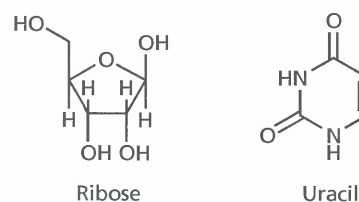


FIGURE 6.10 Ribose and uracil.

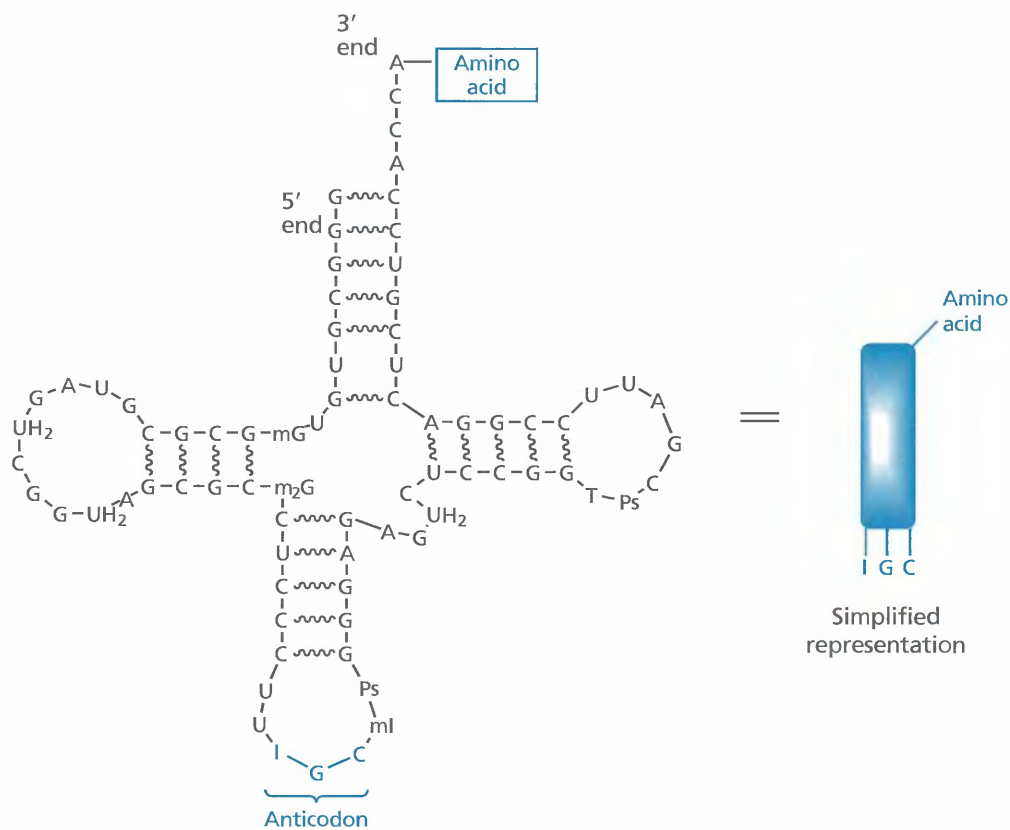


FIGURE 6.11 Yeast alanine transfer RNA. The wiggly lines indicate base pairing. mI, methylinosine; UH₂, dihydrouridine; T, ribothymidine; Ps, pseudouridine; mG, methylguanosine; m₂G, dimethylguanosine.

Because the secondary structure is not uniform along the length of the RNA chain, more variety is allowed in RNA tertiary structure. There are three main types of RNA molecules with different cellular functions. The three are **messenger RNA (mRNA)**, **transfer RNA (tRNA)**, and **ribosomal RNA (rRNA)**. These three molecules are crucial to the process by which protein synthesis takes place. Although DNA contains the genetic code for proteins, it cannot produce these proteins directly. Instead, RNA takes on that role, acting as the crucial ‘middle man’ between DNA and proteins. This has been termed the ‘central dogma’ of molecular biology.

The bases adenine, cytosine, guanine, and uracil are found in mRNA and are predominant in rRNA and tRNA. However, tRNA also contains a number of less common nucleic acids—see for example Fig. 6.11.

6.2.2 Transcription and translation

A molecule of mRNA represents a copy of the genetic information required to synthesize a single protein. Its role is to carry the required code out of the nucleus to a cellular organelle called the **endoplasmic reticulum**. This is where protein production takes place on bodies called

ribosomes. The segment of DNA which is copied is called a gene, and the process involved is called **transcription**. The DNA double helix unravels and the stretch that is exposed acts as a template on which the mRNA can be built (Fig. 6.12). Once complete, the mRNA departs the nucleus to seek out a ribosome, while the DNA re-forms its double helix.

Ribosomal RNA is the most abundant of the three types of RNA and is the major component of ribosomes. These can be looked upon as the production sites for protein synthesis—a process known as **translation**. The ribosome binds to one end of the mRNA molecule, then travels along it to the other end, allowing the triplet code to be read, and catalysing the construction of the protein molecule one amino acid at a time (Fig. 6.13). There are two segments to the mammalian ribosome, known as the 60S and 40S subunits. These combine to form an 80S ribosome. (In bacterial cells, the ribosomes are smaller and consist of 50S and 30S subunits combining to form a 70S ribosome. The terms 50S, etc. refer to the sedimentation properties of the various structures. These are related qualitatively to size and mass, but not quantitatively—that is why a 60S and a 40S subunit can combine to form an 80S ribosome.)

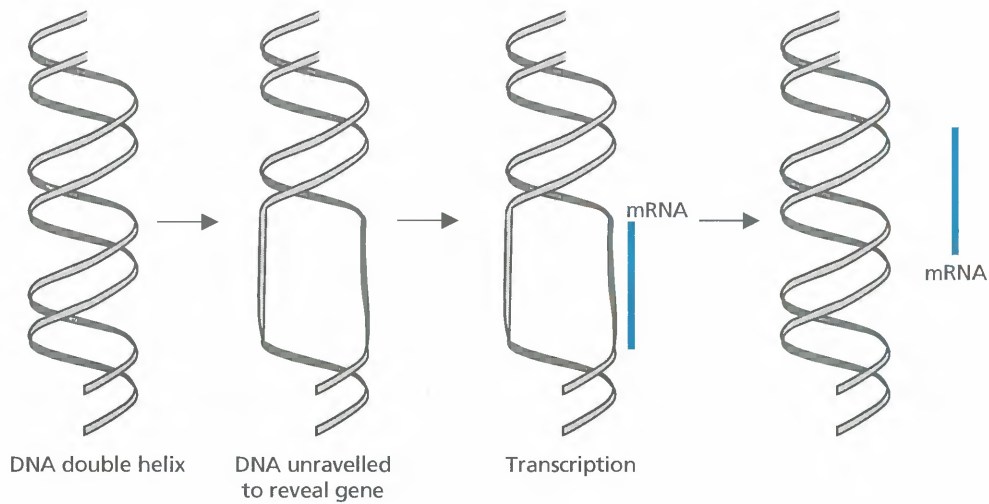


FIGURE 6.12 Formation of messenger RNA (mRNA).

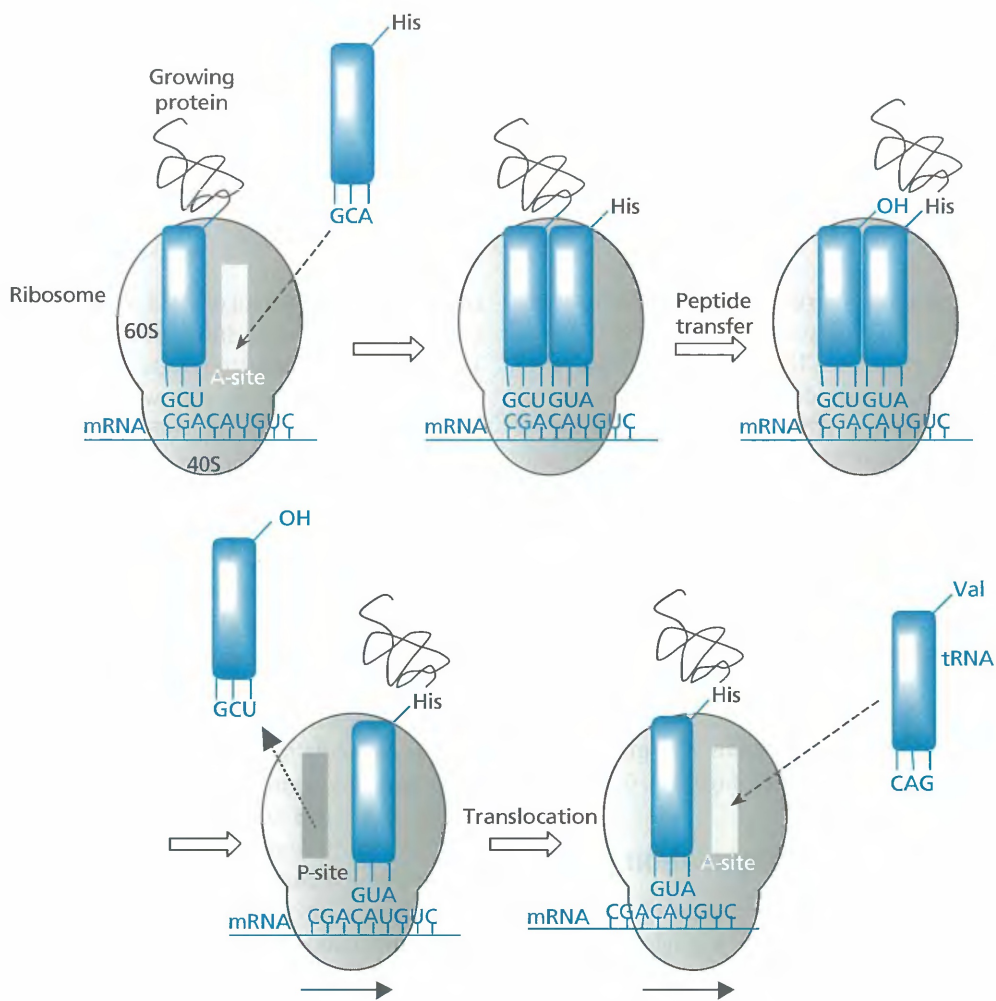


FIGURE 6.13 Protein synthesis—translation.

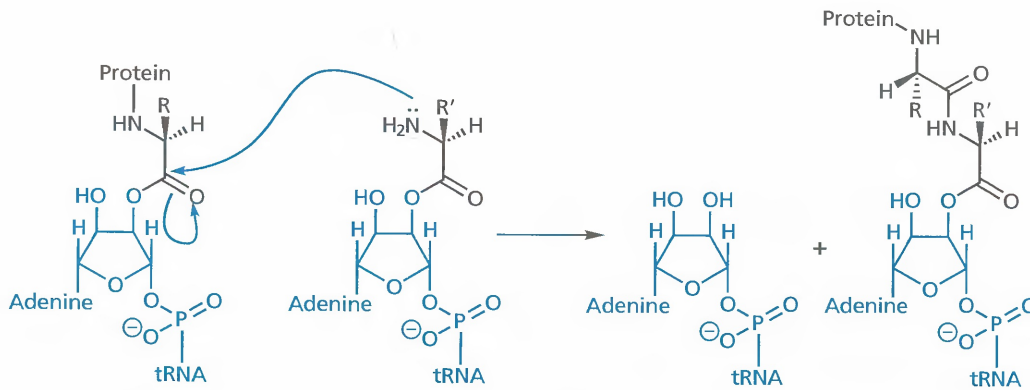


FIGURE 6.14 Mechanism by which a growing protein is transferred to the next amino acid.

rRNA is the major component of each subunit, making up two thirds of the ribosome's mass. The 40S subunit contains one large rRNA molecule along with several proteins, whereas the 60S subunit contains three different-sized rRNAs, again with accompanying proteins. The secondary structure of rRNA includes extensive stretches of base pairing (duplex regions), resulting in a well-defined tertiary structure. At one time, it was thought that rRNA only played a structural role, and that the proteins were acting as enzymes to catalyse translation. The rRNA molecules certainly do have a crucial structural role, but it is now known that they, rather than the ribosomal proteins, have the major catalytic role. Indeed, the key sites in the ribosome where translation takes place are made up almost entirely of rRNA. The proteins are elongated structures which meander through the ribosome structure and are thought to have a fine-tuning effect on the translation process.

tRNA is the crucial adaptor unit which links the triplet code on mRNA to a specific amino acid. This means there has to be a different tRNA for each amino acid. All the tRNAs are clover-leaf in shape, with two different binding regions at opposite ends of the molecule (see Fig. 6.11). One binding region is for the amino acid, where a specific amino acid is covalently linked to a terminal adenosyl residue. The other is a set of three nucleic acid bases (**anticodon**) which will base pair with a complementary triplet on the mRNA molecule. A tRNA having a particular anticodon will always have the same amino acid attached to it.

Let us now look at how translation takes place in more detail. As rRNA travels along mRNA, it reveals the triplet codes on mRNA one by one. For example, in Fig. 6.13 the triplet code CAU is revealed along with an associated binding site called the A site. The A stands for aminoacyl and refers to the attached amino acid on the incoming tRNA. tRNA molecules can enter this site but they are accepted only if they have the necessary

anticodon capable of base pairing with the exposed triplet on mRNA. In this case, tRNA having the anticodon GUA is accepted and brings with it the amino acid histidine. The peptide chain that has been created so far is attached to a tRNA molecule which is bound to the P binding site (standing for peptidyl). A grafting process then takes place, catalysed by rRNA, where the peptide chain is transferred to histidine (Fig. 6.14). The tRNA occupying the P binding site now departs and the ribosome shifts along mRNA to reveal the next triplet (a process called translocation), and so the process continues until the whole strand is read. The new protein is then released from the ribosome, which is now available to start the process again. The overall process of transcription and translation is summarized in Fig. 6.15.

6.2.3 Small nuclear RNA

After transcription, mRNA molecules are frequently modified before translation takes place. This involves a splicing operation where the middle section of mRNA (the intron) is excised and the ends of the mRNA molecule (the exons) are spliced together (Fig. 6.16).

Splicing requires the aid of an RNA-protein complex called a **spliceosome**. The RNA molecules involved in this complex are called small nuclear RNAs (snRNAs). As the name indicates, these are small RNA molecules with fewer than 300 nucleotides that occur in the nucleus of the cell. The role of the snRNAs in the spliceosome is to base pair with particular segments of mRNA such that the mRNA can be manipulated and aligned properly for the splicing process. Splice sites are recognized by their nucleotide sequences, but on occasions a mutation in DNA may introduce a new splice site somewhere else on mRNA. This results in faulty splicing, an altered mRNA, and a defective protein. About 15% of genetic diseases are thought to be due to mutations that result in defective splicing.

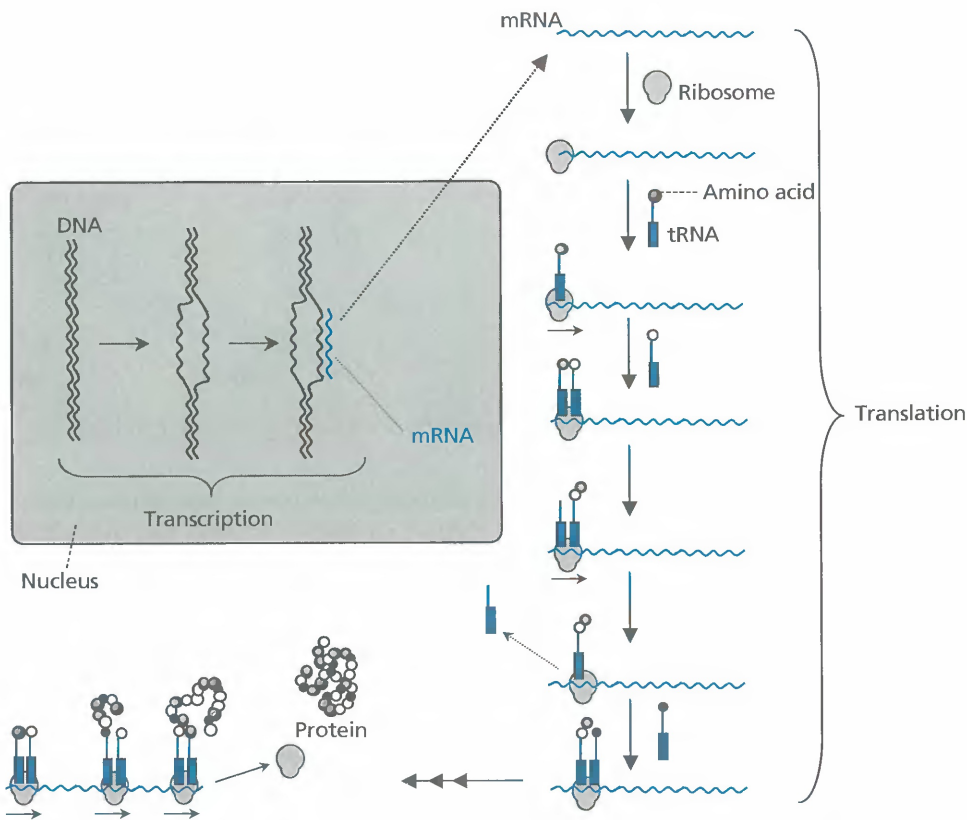


FIGURE 6.15 Transcription and translation.



FIGURE 6.16 Splicing messenger RNA (mRNA).

6.3 Genetic illnesses

A number of genetic illnesses are due to genetic abnormalities that result in the non-expression of particular proteins, or the expression of defective proteins. For example, **albinism** is a condition where the skin, hair, and eyes lack pigment; it is associated with a deficiency of an enzyme called **tyrosinase**. This is a copper-containing enzyme that catalyses the first two stages in the synthesis of the pigment **melanin**. Over 90 mutations of the tyrosinase gene have been identified which lead to the expression of inactive enzyme. Mutations in the triplet code result in one or more amino acids being altered in the resulting protein, and if these amino acids are important to the activity of the enzyme, activity is lost. Mutations which alter amino acids in the active site are the ones most likely to result in loss of activity.

Phenylketonuria is a genetic disease caused by the absence or deficiency of an enzyme called **phenylalanine hydroxylase**. This enzyme normally converts phenylalanine to tyrosine, and in its absence the blood levels of phenylalanine rise substantially, along with alternative metabolic products such as phenylpyruvate. If left untreated, this disease results in severe mental retardation.

Haemophilias are inherited genetic diseases in which one of the blood coagulation factors is deficient. This results in uncontrolled bleeding after an injury. In the past, people with this disease were likely to die in their youth. Nowadays, with the proper treatment, affected individuals should have a normal life expectancy. Treatment in severe cases involves regular intravenous infusion with the relevant coagulation factor. In less severe cases, transfusions can be used when an injury has taken place. The coagulation factors used to be typically derived from

blood plasma, but this meant that people with haemophilia were susceptible to infection from infected blood samples. For example, during the period 1979–1985 more than 1200 people in the UK were infected with HIV as a result of taking infected blood products. For the same reason, they were also prone to viral infections caused by hepatitis B and C. During the 1990s, recombinant DNA technology (section 6.4) successfully produced blood coagulation factors, and these are now the agents of choice because they eliminate the risk of infection. Unfortunately, some patients produce an immune response to the infused factor, which can preclude their use. At present, clinical trials are under way to test whether gene therapy can be used as a treatment. This involves the introduction of a gene which will code for the normal coagulation factor so that it can be produced naturally in the body (section 6.4).

Muscular dystrophy is another genetic disease that affects 1 in every 3500 males and is characterized by the absence of a protein called **dystrophin**. This has an important structural role in cells, and its absence results in muscle deterioration. Gene therapy is also being considered for this disease.

Many cancers are associated with genetic defects which result in molecular signalling defects in the cell. This is covered more fully in Chapter 21.

6.4 Molecular biology and genetic engineering

Over the past few years, rapid advances in molecular biology and genetic engineering have had important repercussions for medicinal chemistry. It is now possible to clone specific genes and to include these genes into the DNA of fast-growing cells such that the proteins encoded by these genes are expressed in the modified cell. As the cells are fast growing, this leads to a significant quantity

of the desired protein which permits its isolation, purification, and structural determination. Before these techniques became available, it was extremely difficult to isolate and purify many proteins from their parent cells due to the small quantities present. Even if one was successful, the low yields inherent in the process made an analysis of the protein's structure and mechanism of action very difficult. Advances in molecular biology and recombinant DNA techniques have changed all that.

Recombinant DNA technology allows scientists to manipulate DNA sequences to produce modified DNA or completely novel DNA. The technology makes use of natural enzymes called **restriction enzymes** and **ligases** (Fig. 6.17). The restriction enzymes recognize a particular sequence of bases in each DNA molecule and split a specific sugar–phosphate bond in each strand of the double helix. The break is not a clean one; there is an overlap between the two chains, resulting in a tail of unpaired bases on each side of the break. The bases on each tail are complementary and can still recognize each other, so they are described as 'sticky' ends. The same process is carried out on a different molecule of DNA and the molecules from both processes are mixed together. As these different molecules have the same sticky ends, they recognize each other such that base pairing takes place in a process called **annealing**. Treatment with the ligase enzyme then repairs the sugar phosphate backbone and a new DNA molecule is formed.

If the DNA molecule of interest does not have the required sequence recognized by the restriction enzyme, a synthetic DNA linker that *does* contain the sequence can be added to either end of the molecule using a ligase enzyme. This is then treated with the restriction enzyme as before (Fig. 6.18).

There are many applications for this technology, one of which is the ability to amplify and express the gene for a particular human protein in bacterial cells. In order to do this it is necessary to introduce the gene to the bacterial cell. This is done by using a suitable **vector** which will carry the

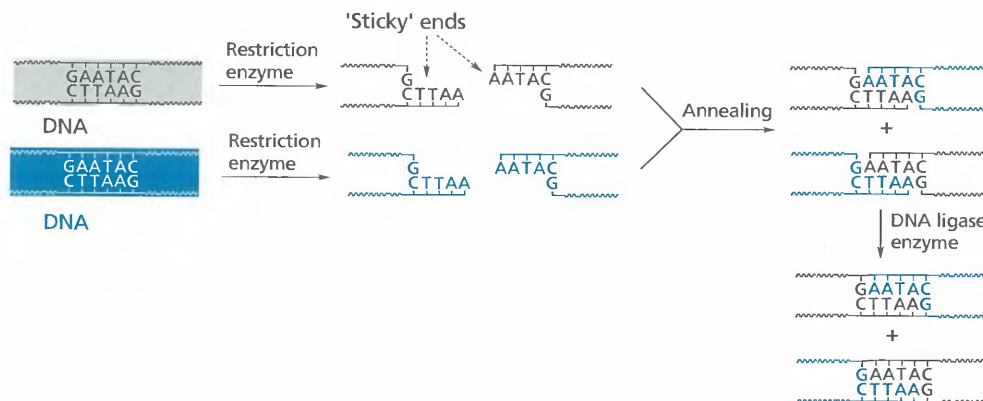


FIGURE 6.17 Recombinant DNA technology.

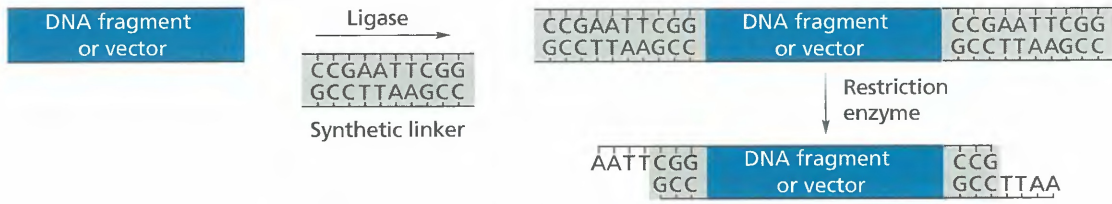


FIGURE 6.18 Attaching sequences recognized by restriction enzymes.

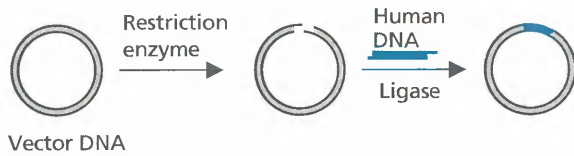


FIGURE 6.19 Inserting a human gene into a plasmid by recombinant DNA technology.

gene into the cell. There are two suitable vectors—**plasmids** and **bacteriophages**. Plasmids are segments of circular DNA which are transferred naturally between bacterial cells and allow the sharing of genetic information. Because the DNA is circular, the DNA representing a human gene can be inserted into the vector’s DNA by the same methods described above (Fig. 6.19). Bacteriophages (phages for short) are viruses which infect bacterial cells. There are a variety of these, but the same recombinant DNA techniques can be used to insert human DNA into viral DNA.

Whichever vector is used, the modified DNA is introduced into the bacterial cell where it is cloned and amplified (Fig. 6.20). For example, once a phage containing modified nucleic acid infects a bacterial cell, the phage takes over the cell’s biochemical machinery to produce multiple copies of itself and its nucleic acid.

Human genes can be introduced to bacterial cells such that the gene is incorporated into bacterial DNA and expressed as if it was the bacteria’s own DNA. This allows the production of human proteins in much greater quantity than would be possible by any other means. Such proteins could then be used for medicinal purposes as described below. Modified genes can also be introduced and expressed to produce modified proteins to see what effect a mutation would have on the structure and function of a protein.

The following are some of the applications of genetic engineering to the medical field.

Harvesting important proteins The genes for important hormones or growth factors such as **insulin** and **human growth factor** have been included in fast-growing unicellular organisms. This allows the harvesting of these proteins in sufficient quantity that they can be marketed and administered to patients who are deficient in these important hormones. Genetic engineering has also been crucial in the production of monoclonal antibodies (section 14.8.3).

Genomics and the identification of new protein drug targets Nowadays, it is relatively easy to isolate and identify a range of signalling proteins, enzymes and receptors by cloning techniques. This has led to the identification of a growing number of isozymes and receptor subtypes which offer potential drug targets for the future. The **Human Genome Project** involved the mapping of human DNA (completed in 2000) and has led to the discovery of new proteins previously unsuspected. These too may offer potential drug targets. The study of the structure and function of new proteins discovered from genomics is called **proteomics** (section 2.6).

Study of the molecular mechanism of target proteins Genetic engineering allows the controlled mutation of proteins such that specific amino acids are altered. This allows researchers to identify which amino acids are important to enzyme activity or to receptor binding. This in turn leads to a greater understanding of how enzymes and receptors operate at the molecular level.

Somatic gene therapy Somatic gene therapy involves the use of a carrier virus to smuggle a healthy gene into cells in the body where the corresponding gene is defective. Once the virus has infected the cell, the healthy gene

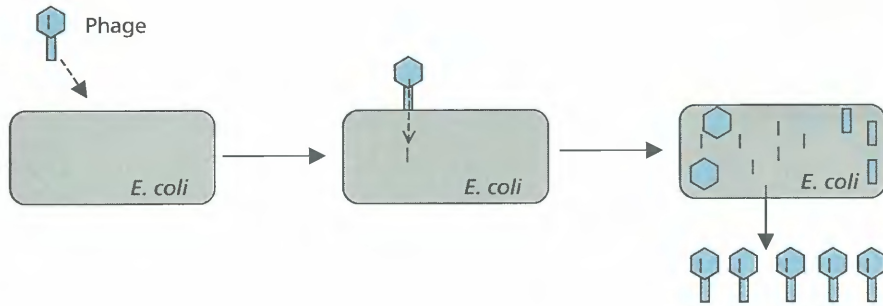


FIGURE 6.20 Infecting *Escherichia coli* with a phage.

is inserted into the host DNA where it undergoes transcription and translation. This approach has great therapeutic potential for cancers, AIDS, and genetic abnormalities such as cystic fibrosis. However, the approach is still confined to research labs and there is still a long way to go before it is used clinically. There are several problems still to be tackled, such as how to target the viruses specifically to the defective cells, how to insert the gene into DNA in a controlled manner, how to regulate gene expression once it is in DNA, and how to avoid immune responses to the carrier virus. Progress in this field was set back significantly in 1999 as a result of a fatality to a teenage volunteer during a clinical trial in America—thought to be due to an over-reactive immune response to the carrier virus. Consequently, there are now studies looking into the use of artificial viruses which would be less likely to cause an immune response. Non-viral delivery systems are also being studied, involving caged molecules called cyclodextrins. In addition, lipids, polyaminoesters, and glycine polymers are being investigated as carriers.

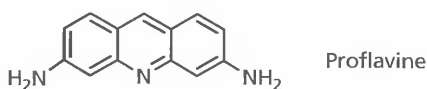
KEY POINTS

- The primary structure of RNA is similar to that of DNA, but it contains ribose instead of deoxyribose. Uracil is used as a base in place of thymine and other bases may be present in smaller quantities.

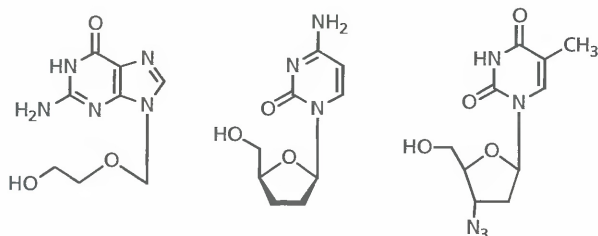
- Base pairing and sections of helical secondary structure are possible within the structure of RNA.
- There are three main types of RNA—messenger RNA, transfer RNA, and ribosomal RNA.
- Transcription is the process by which a segment of DNA is copied as mRNA. mRNA carries the genetic information required for the synthesis of a protein from the nucleus to the endoplasmic reticulum.
- rRNA is the main constituent of ribosomes where protein synthesis takes place. A ribosome moves along mRNA revealing each triplet of the genetic code in turn.
- tRNA interprets the coded message in mRNA. It contains an anticodon of three nucleic acid bases, which binds to a complementary triplet on mRNA. Each tRNA carries a specific amino acid, the nature of which is determined by the anticodon.
- The process of protein synthesis is called translation. The growing protein chain is transferred from one tRNA to the amino acid on the next tRNA and is only released once the complete protein molecule has been synthesized.
- Genetic engineering has been used in the production of important hormones for medicinal purposes, the identification of novel drug targets, the study of protein structure and function, and gene therapy.

QUESTIONS

- Proflavine is a topical antibacterial agent that intercalates bacterial DNA and was used to treat wounded soldiers in the Far East during the Second World War. What role (if any) is played by the tricyclic ring and the primary amino groups? The drug cannot be used systemically. Suggest why this is the case.



- The following compounds are antiviral drugs that mimic natural nucleosides. What nucleosides do they mimic?



- Adenine is an important component of several important biochemicals. It has been proposed that

adenine was synthesized early on in the evolution of life when the Earth's atmosphere consisted of gases such as hydrogen cyanide and methane. It has also been possible to synthesize adenine from hydrogen cyanide. Consider the structure of adenine and identify how cyanide molecules might act as the building blocks for this molecule.

- The genetic code involves three nucleic acid bases coding for a single amino acid (the triplet code). Therefore, a mutation to a particular triplet should result in a different amino acid. However, this is not always the case. For any triplet represented by XYZ, which mutation is least likely to result in a change in amino acid—X, Y, or Z?
- The amino acids serine, glutamate, and phenylalanine were found to be important binding groups in a receptor binding site (see Appendix 1 for structures). The triplet codes for these amino acids in the mRNA for this receptor were AGU, GAA, and UUU respectively. Explain what effect the following mutations might have, if any:

AGU to ACU; AGU to GGU; AGU to AGC
 GAA to GAU; GAA to AAA; GAA to GUA
 UUU to UUC; UUU to UAU; UUU to AUU

FURTHER READING

- Aldridge, S. (2003) The DNA story. *Chemistry in Britain*, April, 28–30.
- Breaker, R. R. (2004) Natural and engineered nucleic acids as tools to explore biology. *Nature*, **432**, 838–845.
- Burke, M. (2003) On delivery. *Chemistry in Britain*, February, 36–38.
- Dorsett, Y., and Tuschl, T. (2004). siRNAs: Applications in functional genomics and potential as therapeutics. *Nature Reviews Drug Discovery*, **3**, 318–329.
- Johnson, I. S. (2003) The trials and tribulations of producing the first genetically engineered drug. *Nature Reviews Drug Discovery*, **2**, 747–751.
- Judson, H. F. (1979) *The Eighth Day of Creation*. Simon and Schuster, New York.
- Langer, R. (2003) Where a pill won't reach. *Scientific American*, April, 32–39.
- Lindpaintner, K. (2002) The impact of pharmacogenetics and pharmacogenomics on drug discovery. *Nature Reviews Drug Discovery*, **1**, 463–469.
- Nicholl, D. S. T. (2008) *An Introduction to Genetic Engineering*, 3rd Edition. Cambridge University Press.
- Opalinska, J. B., and Gewirtz, A. M. (2002) Nucleic acid therapeutics: basic principles and recent applications. *Nature Reviews Drug Discovery*, **1**, 503–514.
- Petricoin, E. F., *et al.* (2002) Clinical proteomics. *Nature Reviews Drug Discovery*, **1**, 683–695.
- Stark, H., *et al.* (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, **409**, 539–542.
- Winter, P. C., Hickey, G. I., and Fletcher, H. L. (1998) *Instant Notes Genetics*. Bios Scientific Publishers, Oxford.