The Structure and Function of Nucleic Acids

Revised edition

C.F.A. Bryce* and D. Pacini†
*Department of Biological Sciences, Napier University, Edinburgh, and
†Bolton School Boy's Division, Bolton

The Biochemistry Across the School Curriculum Group (BASC) was set up by the Biochemical Society in 1985. Its membership includes education professionals as well as Society members with an interest in school science education. Its first task has been to produce this series of booklets, designed to help teachers of syllabuses which have a high biochemical content.

Other topics covered by this series include: *Essential Chemistry for Biochemistry*; *Enzymes and their Role in Biotechnology*; *Metabolism*; *Immunology*; *Photosynthesis*; *Recombinant DNA Technology*; *Biological Membranes*; and *The Biochemical Basis of Disease*.

More information on the work of BASC and these booklets is available from the Education Officer at the Biochemical Society, 59 Portland Place, London W1N 3AJ.

Comments on the content of this booklet will be welcomed by the Series Editor Mrs D. Gull at the above address.

ISBN 0 90449 834 4

First edition published 1991

Printed by Holbrooks Printers Ltd, Portsmouth, U.K.

Contents

Foreword

A study of the structure and function of nucleic acids is needed to be able to understand how information controlling the characteristics of an organism is stored in the form of genes in a cell and how these genes are transmitted to future generations of offspring. The rapid developments in the area of genetic engineering and recombinant DNA technology (which are covered in Booklet 7 of the BASC series) have only been possible as a result of detailed understanding of the structure of DNA and RNA. It is therefore not surprising to see nucleic acids included in the compulsory core subject matter of all of the linear and modular A-level Biology syllabuses. The major purpose of these guidance notes is to provide an account of the role of DNA and RNA in these processes in at least sufficient detail for A-level study. Additional information more suited for general interest, Oxbridge preparation or first-year undergraduate study is placed in shaded boxes of text.

In order to maximize the benefit of this booklet for A-level students and their teachers, a set of learning objectives have been included at the start of chapters 1–6. These objectives are based upon the relevant knowledge and understanding statements of current A-level Biology syllabuses. Chapter 7 consists of recent A-level examination questions and answers and this should help students get a feel for the level of detail required. We are grateful to the various Examinations Boards, especially the Northern Examinations and Assessment Board (NEAB), for their permission to include this material.

1

Nucleic acids and genetic information transfer

Learning objectives

Each student should, without reference to his or her notes, be able to:

- state that the genotype describes the genetic composition of an organism in terms of specific genes, i.e. alleles that it contains;
- state that the phenotype is the collection of visible characteristics which identify an individual organism;
- state that the phenotype of an individual depends not only on its genotype but also on the environment in which it lives; and
- state that DNA, which is found in the nucleus of a cell, contains and is responsible for the transfer of genetic information, i.e. the genotype.

Introduction

Normally, when we talk about genes and heredity we perhaps think immediately of how children resemble their parents or, depending on our background knowledge, we may even think about Mendel and the height and colour of his pea plants. This level of genetics is known as phenotypic expression, where the phenotype of an individual is the collection of visible and recognizable characteristics by which that individual is identified.

The phenotype is influenced by the environment as well as the genotype; but this aspect is more appropriate to classical genetics and will not concern us at the present time. In this section we are more concerned with the collection of the inherited factors which determine these traits,

and this is a cell's genotypic expression. The genotype is the genetic composition of an organism in terms of the forms of specific genes, i.e. alleles, that it contains, and its study is referred to as molecular genetics.

The very fact that a child can resemble one or both parents means that there was some way in which the genetic information was passed from the parent to the child. What then is the nature of the genetic material? The answer to this question came from the experimental observations shown in Table 1.

More to heredity than DNA

In similar types of nuclear transplantation experiments, scientists have recently revealed that the environment of the nucleus (i.e. the cytoplasm in which it finds itself) can alter the functioning of key genes, including those which determine body size. These have been referred to as **epigenetic** changes, these being changes which affect the way DNA works rather than sequence changes. These findings could have implications for *in vitro* fertilization in humans.

New Scientist (19 April 1997)

Single-cell organisms take a siesta

In 1997 a short report appeared which indicated that a study on the patterns of DNA synthesis in single-cell organisms revealed that cells synthesized DNA from sunrise to about noon, then shut down synthesis for between three and six hours, then restarted before sunset. The explanation given for this phenomenon is that this represents an evolutionary throwback to the Precambrian times when there was no ozone layer and hence the sun's ultraviolet light at its peak could have damaged the DNA. This may have a future significance if the ozone layer does continue to degrade.

New Scientist (4 January 1997)

Such studies represent the most conclusive evidence that DNA is responsible for the transfer of genetic information, at least in microorganisms and viruses. The evidence that this is also true in higher organisms (plants and animals), although somewhat less conclusive, leaves no doubt that this is the case. Since some virus particles have been found that have no DNA, but which have RNA instead, it is more correct to say that 'nucleic acids are the genetic information carriers'.

The exact roles of DNA and RNA in the complex process of the transfer of genetic information are the subjects of subsequent sections of this booklet.

Table 1. Steps in the discovery of DNA as the genetic material

Experimenter(s) and dates	Details of procedure	Conclusion
Late 19th century biologists	Removal of the nucleus of the cell produced death, but removal of the same volume of cytoplasm did not.	'Something' in the nucleus is responsible for a cell's survival.
Late 19th century biologists	Removal of a cell nucleus followed by transplantation of the nucleus from a different cell type changed the shape and function of the donor cell.	'Something' in the nucleus is responsible for the phenotype of a cell.
Griffiths (1929)	Dead, disease-producing, virulent strains of *Pneumococcus* were able to transform harmless cells of non-virulent *Pneumococcus* into lethal organisms.	'Something' had been transferred from the dead organisms that altered the phenotype of the living cells.
Avery and MacLeod (1944)	These workers extracted the DNA from the virulent *Pneumococcus* bacterium. On adding it to non-virulent strains these were transformed into disease-causing organisms. The DNA contained no detectable protein and was unaffected by proteases, but the transformation was affected by use of DNase.	DNA is responsible for cell transformation i.e. a cell's phenotype.
Hershey and Chase (1953)	These workers infected *Escherichia coli* bacteria using a radioactive strain of T2 bacteriophage virus in which the phosphorus atoms in the DNA and the sulphur atoms in the coat protein were labelled. This revealed that only the DNA entered the bacteria, causing transformation.	DNA is responsible for cell transformation i.e. a cell's phenotype.

2

Isolation and structure of nucleic acids

Learning objectives

Each student should, without reference to his or her notes, be able to:

- state that DNA is composed of phosphate, deoxyribose and the four major bases: adenine, guanine, cytosine and thymine;
- state that RNA is composed of phosphate, ribose, adenine, guanine, cytosine and uracil;
- state that DNA and RNA are polymers of nucleotide subunits;
- state that a nucleotide is composed of a phosphate group, a pentose sugar and one of the four corresponding bases;
- state that the backbone of a DNA molecule is a chain of repeating deoxyribose–phosphate units;
- state that the backbone of an RNA molecule is a chain of repeating ribose–phosphate units;
- state that each molecule of DNA is usually composed of two chains;
- state that in DNA, adenine will only bind with thymine on opposite chains and guanine will only bind with cytosine on opposite chains;
- state that the two linked chains in DNA are arranged in a double helix;
- state that there are three principal types of RNA, these being messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA), which are all single-stranded molecules;
- state the function of the three principal types of RNA;
- draw a diagram of a simple ladder-like representation of a DNA molecule; and
- state that DNA is a stable polynucleotide which contains coded genetic information for inherited characteristics.

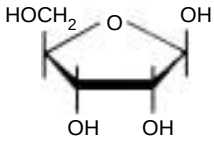
Introduction

Although the nucleic acids were first discovered in 1868, by Friedrich Miescher working with pus cells obtained from discarded surgical bandages, it was not really until the early 1940s that the chemistry and biology of the nucleic acids were set on firm foundations.

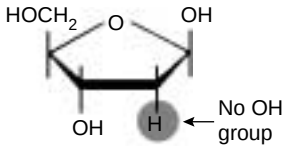
Basically, nucleic acids can be subdivided into two types: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Both DNA and RNA have been shown to consist of three groups of molecules: pentose (5-carbon-atom) sugars; organic bases; and inorganic phosphate.

Sugars

There are only two types of sugar present in nucleic acids, ribose which is present solely in RNA (hence its name) and deoxyribose which is present solely in DNA (again, the sugar gives rise to the name deoxyribonucleic acid). The chemical structures for these compounds are shown here:

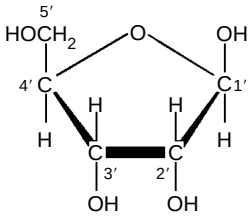


Ribose



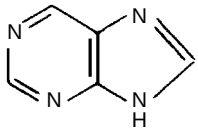
Deoxyribose

The prefix 'deoxy' means 'without oxygen', and we can see from the structures that the only difference between them is the absence of an oxygen in the deoxyribose sugar (see shaded area). Both sugars contain 5 carbon atoms (pentose sugars) and for convenience we number these as shown in the next figure. The 'dash' or 'prime' (′) on, for example, the 5 indicates the carbon in the ribose ring. The purine or pyrimidine rings (see Bases, below) are numbered without primes in order to distinguish them.

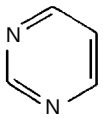


Bases

The nucleotide bases found in nucleic acids are related either to the purine ring system or to the pyrimidine ring system.



Purine

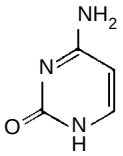


Pyrimidine

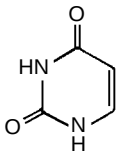
In DNA we find principally four different bases: adenine (A), guanine (G), cytosine (C) and thymine (T). The first two are derived from purine whereas the remaining two are derived from pyrimidine. In RNA we find principally four different bases; adenine, guanine and cytosine as in DNA. The fourth base in RNA, however, is not thymine but instead the pyrimidine-derived base, uracil (U).

The chemical structures for each of the five bases are shown below.

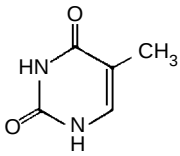
Pyrimidines



Cytosine

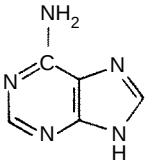


Uracil

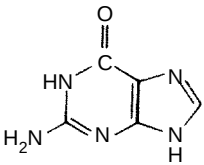


Thymine

Purines



Adenine



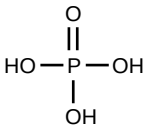
Guanine

Note: A-level students are not expected to recognize these structural formulae. For further help in understanding the chemistry of these compounds, refer to BASC Booklet 1, Essential Chemistry for Biochemistry.

In addition to these major bases there is also a large range of so-called minor bases which occur less frequently than the others, e.g. 5-methyl cytosine.

Inorganic phosphate

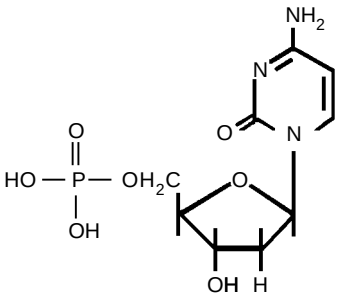
There are phosphate residues in nucleic acids and they are of the type derived from phosphoric acid, the structure of which is shown below.



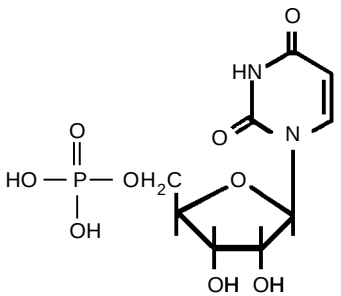
Building nucleic acids from their building blocks

When any one of the bases is joined to either one of the two sugar molecules, we have a compound known as a nucleoside. If the sugar residue is ribose then we have a ribonucleoside, whereas if it is deoxyribose then we have a deoxyribonucleoside. The bond linking these structures is known as a glycoside bond.

Addition of a phosphate group to the sugar residue of a nucleoside molecule produces a different molecule called a nucleotide. Examples of various nucleotides are shown below.



Deoxycytidine 5′-monophosphate

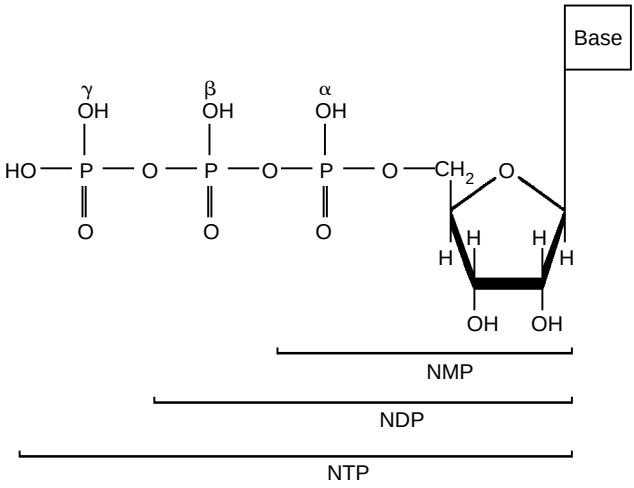


Uridine 5′-monophosphate

Nucleotides can be regarded as the building blocks for the larger nucleic acid molecules, DNA and RNA.

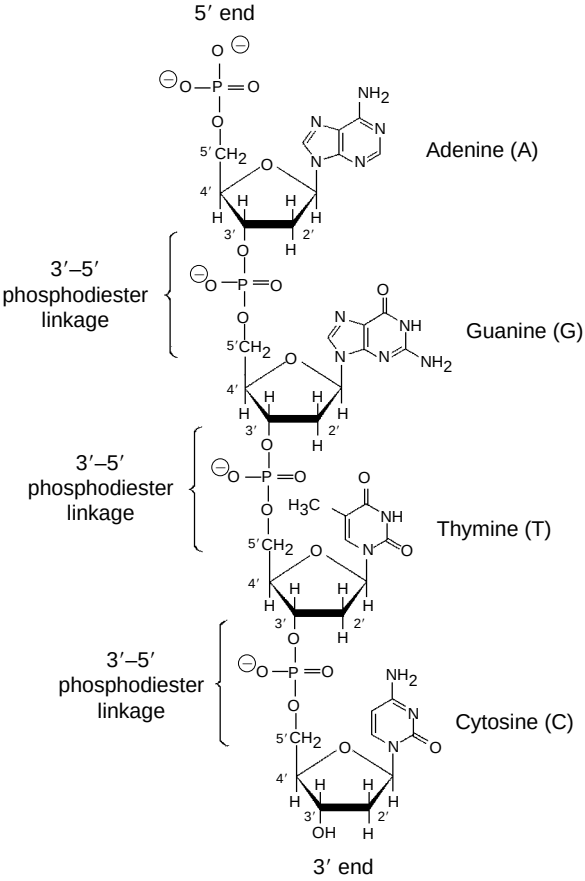
Nucleoside di- and triphosphates

One or two additional phosphates can be added to the first phosphate group of a nucleoside molecule (a nucleoside monophosphate, NMP) by means of a pyrophosphate linkage. The molecules formed in this way are called nucleoside diphosphates (NDPs) and nucleoside triphosphates (NTPs). The most important base involved in these compounds is adenine, forming the adenosine mono-, di- and triphosphate molecules (AMP, ADP and ATP), which fulfil vital roles in many cellular processes as you may know from your study of respiration and photosynthesis [for further details see BASC Booklets 4 (*Metabolism*) and 6 (*Photosynthesis*)].

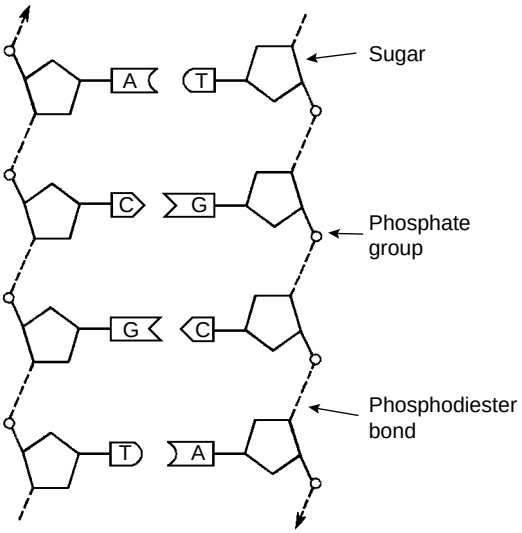


Nucleic acids

Nucleic acids are formed by the combination of nucleotide molecules through sugar–phosphate bonds known as phosphodiester linkages. Because a nucleic acid is a polymer of many nucleotide molecules, DNA and RNA molecules are called polynucleotides.

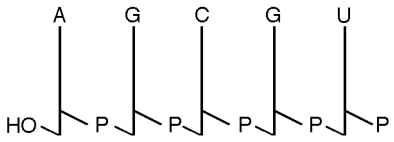


The structure of a polynucleotide is shown diagrammatically above. In common with the formation of other biological polymers, e.g. starch and proteins, water molecules are produced by condensation reactions and energy from ATP molecules, produced in cellular respiration, is used.



From this figure it can be seen that the phosphodiester bond is formed between the 5′carbon of one sugar molecule and the 3′ carbon of the next. The double-stranded nature of DNA is shown schematically above.

Since it would be very time consuming to have to write out such a structure every time it was needed, abbreviated forms have been introduced. These are essentially of three types:



≡ ApGpCpGpUp ≡ $^{5'}$AGCGU$^{3'}$

If the ‘P’ is on the left of a base then it is attached to the 5′-carbon (five); if it is on the right then it is attached to the 3′-carbon (three).

DNA isolation, base composition and structure

Isolation

In order to study the chemistry of DNA it must be isolated and purified from the nucleus of a cell. In the isolation of DNA from cell samples, the major contaminant is histone protein which is closely associated with DNA to form visible chromatin threads in the nucleus. To remove this contaminant the cell is first deproteinized by treatment with phenol and the DNA is precipitated using ice-cold ethanol. A more detailed account of the isolation of DNA is given in one of the practical schedules later (see Chapter 8) and in BASC Booklet 7, *Recombinant DNA Technology*.

Once isolated, DNA can be broken down or degraded by either: (i) treatment with perchloric acid at 100°C, which breaks down the DNA to its constituent bases; or (ii) treatment with enzymes (DNase, phosphodiesterase, etc.) which break down the DNA to its constituent nucleotides.

The mixture of either of these products can be separated by chromatography or electrophoresis and the amount of each of the constituent bases can be measured.

Composition

Using the techniques outlined above it was noted by Chargaff and coworkers in the late 1940s that:
(i) the base composition of an organism's DNA is characteristic of that organism;
(ii) different cells or tissues of the same organism have identical base composition;
(iii) closely related organisms exhibit similar base compositions (this property is used in biology as a basis for chemical taxonomy)
(iv) the majority of DNA molecules exhibit certain chemical regularities that are defined as Chargaff's base pairing rules (base equivalence), namely the amount of A equals the amount of T and the amount of G equals the amount of C, as can be seen (within the bounds of experimental error) from the data shown in Table 2.

In considering the data shown in Table 2, in terms of base equivalence it is necessary to combine the values for cytosine and 5-methyl cytosine where these apply. The only exception to base equivalence is the

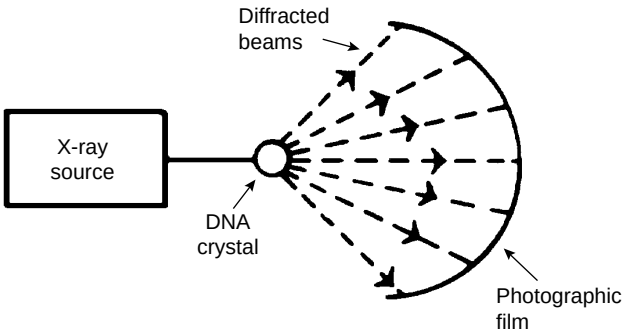
DNA from the virus ϕX174. Can you think of an explanation? (See the Structure section in this chapter for the solution.)

Table 2. The percentage composition of bases in the DNA from a variety of organisms

Source of DNA	Adenine	Guanine	Cytosine	Thymine	5-Methyl cytosine
Cattle sperm	28.7	22.2	20.7	27.2	1.3
Rat bone marrow	28.6	21.4	20.4	28.4	1.1
Sea urchin	32.8	17.7	17.3	32.1	1.1
Wheat germ	27.3	22.7	16.8	27.1	6.0
Yeast	31.3	18.7	18.1	31.9	–
Human sperm	30.9	19.9	19.8	29.4	–
Escherichia coli	26.0	24.9	23.9	25.2	–
ϕX174 Phage virus	24.3	24.5	18.2	32.3	–

Structure

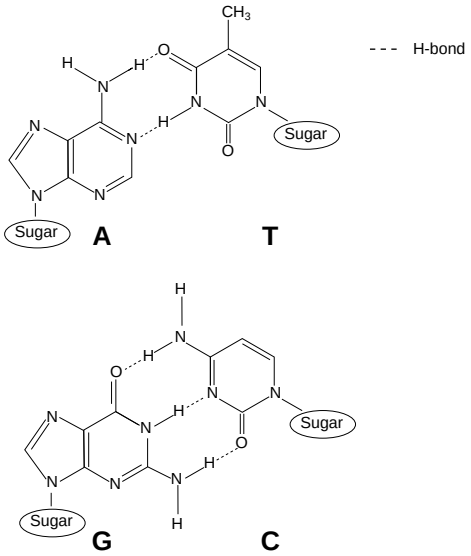
In 1947 Astbury studied DNA using the technique of X-ray diffraction, shown schematically below.



When the beam of X-rays collides with an electron-dense region in the DNA, the beam is diffracted. From the pattern of the diffracted beam it is possible to determine the arrangement of atoms in the DNA. In this way, Astbury was able to deduce that there was a regular feature occurring within the DNA molecule every 3.34 Å. Such a regular feature is referred to as a crystallographic repeat or identity period.

Later work by Franklin, Goslin, Wilkins and co-workers (1953) demonstrated that there were two forms of DNA, a crystalline form known as the A form and a paracrystalline form known as the B form.

Based on the X-ray diffraction work and Chargaff's base equivalence rules, Watson and Crick in 1953 deduced the structure of DNA as a double helix. This comprised two strands of sugar–phosphate backbone with the bases positioned in between, with adenine (A) on one strand being hydrogen (or H-) bonded to a thymine (T) on the other, and a guanine (G) on one strand being H-bonded to a cytosine (C) on the other. The H-bonding arrangement of these base pairs is shown below.



A single-ringed pyrimidine base opposite a double-ringed purine base would ensure equal distance between the two polynucleotide strands but why A–T and G–C specifically? The answer lies in the number of H-bonds that can be formed between the bases as shown (see also BASC Booklet 7, *Recombinant DNA Technology*).

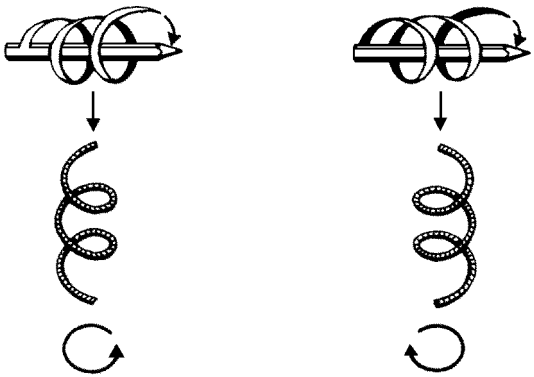
Molecular impostor that cannot form hydrogen bonds

In a recent study a synthetic chemical which resembled thymine but which had a benzene ring instead of the pyrimidine ring and fluorine atoms in place of the oxygen and nitrogen atoms was used in an assay with DNA polymerase. Much to the researchers' surprise, a DNA helix was produced thus raising doubts on the role of hydrogen bonds in the DNA structure. A further study showed that molecular mimics which have the wrong shape distort the helix and that this may represent the key feature of the structural stability.

New Scientist (22 February 1997)

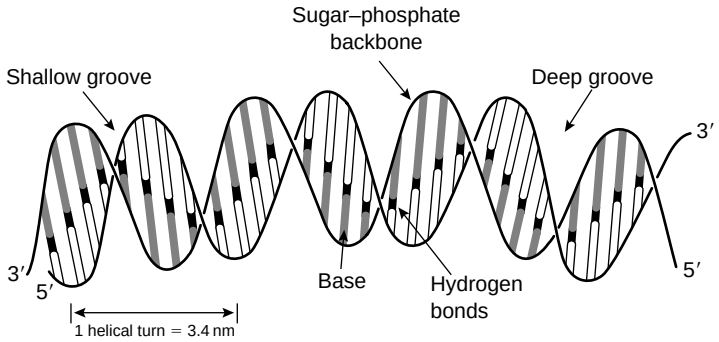
Because the two polynucleotide strands have 'direction', two configurations are possible, one in which both strands run in the same direction, which is known as a parallel arrangement, and one in which the strands run in the opposite direction, which is known as an antiparallel arrangement. It has been shown that the DNA duplexes found in nature have their chains in an antiparallel arrangement. This has a profound significance when we come to consider the biological function of the molecules (see DNA replication section).

In coiling the backbone into a helical configuration, two arrangements in space are possible, one a right-handed helix, the other a left-handed helix. The difference between these two arrangements can be easily demonstrated using two pipe cleaners. It has been shown that the double helix of DNA found in nature is in the right-handed configuration.



Left-handed Right-handed

Subsequent studies using space-filling models showed that there were two grooves in the DNA molecule, one shallow (approx. 12 Å) and one deep (approx. 22 Å). This feature was found to be important in relation to the binding of certain antibiotics, polymerase enzymes, etc. The principal features of the structure of DNA are summarized in the schematic drawing shown below.



Earlier in this chapter it was observed that the DNA isolated from the virus ϕX174 did not exhibit base equivalence. The reason for this is that the DNA of this virus is single-stranded and not the normal duplex form. This has a significance for its mode of replication.

Triple helices are also possible

Triple helical forms of DNA are now known to exist *in vivo* and these would appear to have a key function in certain biological processes, in controlling gene expression and as tools in genome mapping strategies. Such molecules are now being actively investigated.

Triple-helical nucleic acids (V.N. Soyfer, 1996, ISBN 0-387-94495-8)

DNA as the logic for new breed of computers

The notion of organic computers has been with us for some time now, but recently the use of short DNA fragments to simulate the logic AND and OR gates of a computer has been investigated. The claim is that this may now help to produce a new form of computer that would significantly out-perform those currently based on silicon.

New Scientist (8 February 1997)

Breast tissue traps carcinogens

From recent research work it would appear that the fatty tissue that constitutes 80% of women's breasts is capable of absorbing fat-soluble organic molecules, some of which are carcinogenic, causing DNA damage and mutation. It would appear that certain foodstuffs may be the source of these carcinogens and as such it might be impossible or difficult to avoid ingesting these.

New Scientist (7 December 1996)

Table 3. DNA structure and function

Structural feature	Function
Strong, covalently bonded sugar–phosphate backbone	Preserves the base sequence of genetic information during the lifetime of the cell and allows shortening of the DNA during chromosome formation in cell division
Complementary purine–pyrimidine base pairing	Keeps an equal distance between the two polynucleotide strands, increasing the stability of the double helix
Numerous but weak H-bonds between complementary bases	Enhance stability of the helix but are easily broken by enzymes to allow DNA replication and transcription
Large, insoluble molecule (molecular mass approx. 100 000 units)	Restricts DNA to the nucleus, protecting it from biochemical damage, thus preserving the genetic code

RNA isolation, base composition and structure

Isolation

The isolation procedure is very similar to that for DNA isolation although it is essential to include an RNase inhibitor to prevent degradation of the RNA. Unlike DNA, which exists as a more or less homogeneous molecule for any one cell, RNA exists as a family of molecules, each member of which has a distinctive structure and function (see later sections in this chapter). Because of this wide variety of RNA species it is necessary to separate further the initial RNA pool into its constituent molecules. The methods employed include:

- density-gradient centrifugation;
- ion-exchange chromatography;

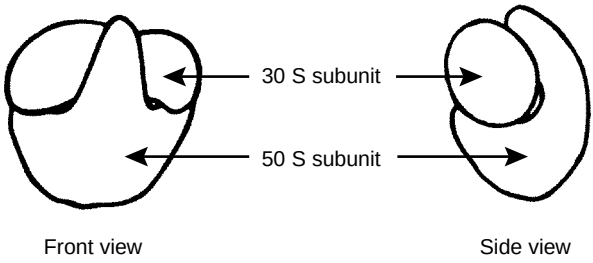
- gel filtration;
- electrophoresis.

Base composition

Purified RNA molecules can be degraded by chemical means (alkali) or by enzymes (RNase) and the base composition determined in a manner similar to that described for DNA (see section on DNA isolation in this chapter). One major difference between DNA and RNA is that in RNA there are a fairly large number of minor bases. Also, in the majority of cases there is no base equivalence, signifying that RNA molecules are usually single-stranded.

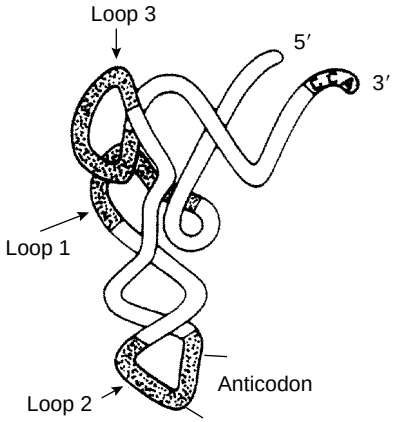
Structure

There are three major forms of RNA: ribosomal RNA (rRNA), transfer RNA (tRNA) and messenger RNA (mRNA) and these exist in all the major life forms, together with other distinct RNA molecules which are not universal. rRNA accounts for about 80% of the total cellular RNA and is associated with protein to form the cytoplasmic particles known as ribosomes. As shown below the ribosome itself can be considered as two subunits, a large and a small, both of which are RNA–protein associations with about 65% RNA: 35% protein. The RNA has high molecular mass and is metabolically stable.



Transfer RNA is the next most abundant species and accounts for about 15% of the total RNA. These molecules are of much lower molecular mass than the rRNA and are also referred to as 4S RNA. As we shall see later, these molecules function as adaptors for amino acids in

the course of protein synthesis and many different tRNAs exist, each being specific for one amino acid. The tRNA molecule is single-stranded but the chain folds back on itself in a very distinctive way (see diagram below) such that about 50–60% of the structure is base-paired.



Most of the remainder of the cell RNA (less than 5%) is accounted for by mRNA which, in eukaryotes, originates in the nucleus and migrates to the cytoplasm during protein synthesis (see Chapters 5 and 6). It is of high molecular mass and is metabolically very labile, i.e. easily broken down. As we shall see in a later section (see Chapters 5 and 6), mRNA is centrally involved in the transfer and expression of the genetic information and is responsible for the sequence of amino acids in each of the different proteins in the cell. Because the size of the messenger is variable, no S-value is associated with it.

Note: the S-values refer to how fast a particular molecule sediments in an ultracentrifuge and the S refers to Svedberg units, the unit of measurement in such studies. Clearly this relates both to the mass and shape of a molecule in that an object of greater mass will move faster than one with lower mass but the same shape, and one with a small volume will move faster than one with a similar mass but larger volume (i.e. less dense).

RNAs which act as enzymes

In the late 1980s it was shown that certain RNA molecules could act as enzymes, capable of splicing out specific sequences of RNA either on itself or other RNA strands. The name given to such molecules was **ribozyme** and this work earned a Nobel Prize for one of the researchers (Tom Cech of the University of Colorado). Two of these ribozymes are now being clinically tested as potential treatment against HIV (see BASC Booklet 3, *Enzymes and their Role in Biotechnology*, for further details).

New Scientist (7 December 1996)

A role for RNA in the treatment of debilitating disease

Specific RNA molecules are being selected from pools of randomly generated and synthesized RNA as a means of treating the debilitating disease myasthenia gravis. Sufferers of this disease produce auto-antibodies which block the normal signals from nerves to muscles by blocking the acetylcholine receptor. It has been shown that specific RNA molecules are able to bind to the antibodies, so interfering with their binding to the receptors and hence restoring the normal nerve-to-muscle signals (see BASC Booklets 5, *Immunology*, and 9, *The Biochemical Basis of Disease* for further information).

New Scientist (18 January 1997)

Table 4. Differences between DNA and RNA structure and properties

DNA	RNA
Deoxyribose sugar	Ribose sugar
Thymine base	Uracil base
Double-stranded helix	Single-stranded molecule
Very large molecular mass	Much smaller molecular mass
Insoluble	Soluble

3

DNA replication

Learning objectives

Each student should, without reference to his or her notes, be able to:

- explain the biological significance of DNA replication;
- explain the suitability of DNA structure for replication;
- outline the main features of semi-conservative replication of DNA;
- outline the Meselson and Stahl experiment; and
- name the major enzymes involved in replication.

Introduction

In the previous chapter it was shown how we knew that the genetic information of a cell is contained in the DNA of that cell. For the cell to divide and produce daughter cells in mitosis and meiosis it is essential that the DNA is copied (replicated) and an identical copy is passed to the daughter cell. DNA is replicated during interphase of both mitosis and meiosis.

The basic results on which our present day assumptions of the mechanism of DNA replication are built are:

(i) the evidence that the genetic information was contained within the nucleic acids;

(ii) the fact that most DNA consists of two strands of polynucleotide chains, each of which consists of deoxyribonucleotide residues joined by 3′–5′ phosphodiester bonds; and

(iii) the fact that H-bonds occur between the bases in the two strands, adenine in one strand is always bonded to thymine in the other and likewise cytosine with guanine.

Each of these we have already considered, and thus, if we are given the sequence of bases in one strand, we could immediately write down the sequence of bases in the complementary strand.

Basically, there are five theoretically possible modes of replication.

Conservative

By this mechanism one daughter cell receives the original DNA molecule whilst the other receives a completely new copy.

Semi-conservative

Here the original DNA molecule is split into two strands and each strand acts as a template on which a new strand is synthesized.

Non-conservative

Here the original DNA is destroyed completely during the course of synthesis of two new identical DNA molecules.

Dispersive

Here the original DNA molecule is dispersed or distributed into all nascent chains.

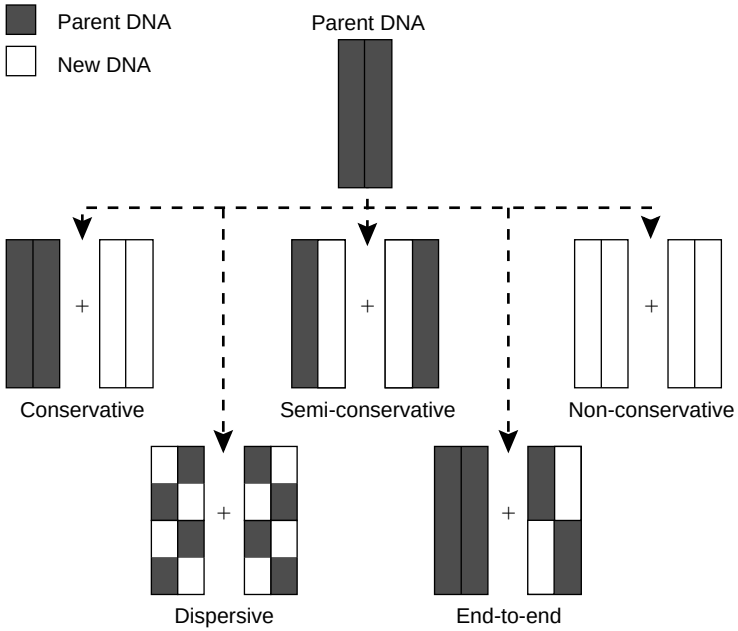
End-to-end

Here the original DNA molecule is present as half of one chain and the other half of the complementary chain for both nascent molecules.

Note: A-level students are only expected to consider evidence for or against the first two possible modes of replication outlined above.

The five possible modes of replication are shown schematically in the figure which follows.

In fact it has been proved that DNA replicates semi-conservatively, first in *E. coli*, and subsequently in all higher organisms as well: it has been shown to be the case for eukaryotic cells, including human cells, in



tissue culture. The original experiment which proved this mechanism was carried out by Meselson and Stahl in 1958.

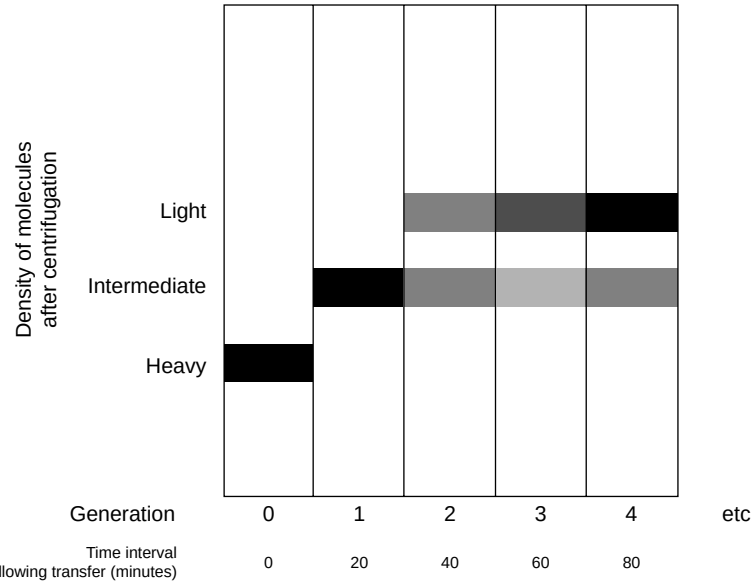
The Meselson and Stahl experiment

(a) Meselson and Stahl grew *E. coli* in a medium containing $^{15}NH_4Cl$ (i.e. ^{15}N is the heavy isotope of nitrogen). The result of this was that ^{15}N was incorporated into newly synthesized DNA (as well as other nitrogen-containing compounds).

(b) They allowed the cells to grow for many generations in the $^{15}NH_4Cl$ so that all the DNA was 'heavy' — that is, had a greater density than normal. This 'heavy' DNA could be distinguished from the normal DNA by centrifugation in a caesium chloride (CsCl) density gradient. Note that ^{15}N is not a radioactive isotope.

(c) They then transferred the cells into a medium with normal $^{14}NH_4Cl$ and took samples at various definite time intervals as the cells multiplied, and extracted the DNA that remained as double-stranded helices.

The various samples were then run independently on CsCl gradients to measure the densities of DNA present. The results of this experiment were as follows:



Thus the DNA that was extracted from the culture one generation after the transfer from ^{15}N to ^{14}N medium (i.e. after 20 minutes) had a hybrid or intermediate density, exactly half-way between that of DNA containing only ^{15}N and that containing only ^{14}N; each molecule must therefore have contained one old and one new strand.

DNA extracted from the culture after another generation (i.e. after 40 minutes) was composed of equal amounts of this hybrid DNA and of 'light' DNA. As time went on, more and more 'light' DNA accumulated.

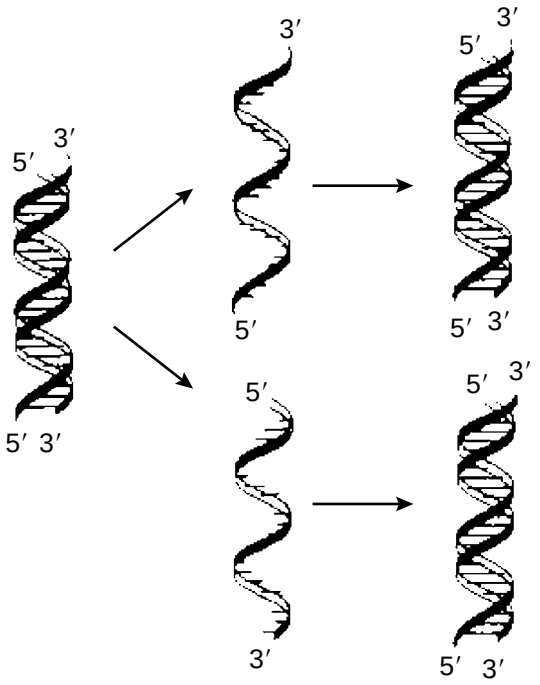
The results of the Meselson and Stahl experiment are not consistent with non-conservative (only ^{14}N-DNA bands would be expected from 20 minutes onwards) or conservative (one band of ^{14}N-DNA and one of ^{15}N-DNA at 20 minutes would be expected) modes of replication.

Thus far, however, the data are consistent with the three remaining modes of replication. To distinguish between these, and hence determine the mode of replication, Meselson and Stahl carried out an additional

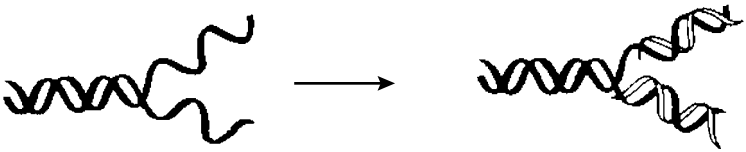
experiment in which they took the DNA isolated at 20 minutes after transfer to ^{14}N-medium, strand separated this and ran the resultant sample on a CsCl gradient. For dispersive or end-to-end models we would predict a single hybrid band, whereas for semi-conservative replication we would expect two bands, one light, one heavy. The latter was found to be the case and hence DNA replication is a semi-conservative process.

Mechanism of replication

Watson and Crick, in their 1953 paper in which they proposed the double-helical structure for DNA, conclude as follows: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material". The implication was that, owing to the strict complementary nature of the strands, the base sequence of one strand would determine the base sequence of the other. It could be imagined therefore that replication could follow a scheme as shown below.

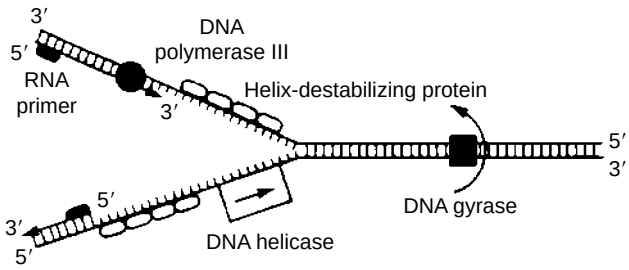


The problem, however, with this scheme is that the energy required for the first step, namely strand separation, would be far too high to be plausible. This led to the suggestion that unfolding took place a little at a time, so forming a replication fork as shown.



Here, as elsewhere in biochemical reactions, an enzyme is used as a catalyst. Enzymes lower the energy needed to activate a reaction, making it more likely to occur. For *E. coli* this is an enzyme called DNA polymerase III. The simplest scheme would be for polymerase molecules to attach themselves to the ends of the single-stranded DNA template and to synthesize a complimentary strand of each.

The first problem encountered with this proposal arose because the polymerase enzyme was found to operate in one direction only, that is it synthesized the nascent DNA strand in the direction 5′→3′. This meant that one of the parent strands (called the leading strand) could be copied in a continuous fashion, whereas the other (called the lagging strand) could only be copied in a discontinuous way as a series of short lengths (approx. 1000 bases long) of DNA, each of which is called an Okazaki fragment (see Figure below). It was also discovered early on that the DNA polymerase enzyme could not actually synthesize a DNA molecule from individual DNA nucleotides but instead required a short stretch of existing DNA helix duplex on which to build. In this sense it is truly a DNA-dependent DNA polymerase III.



The full details of the process of replication are beyond the scope of the A-level syllabus but it may be of interest to know the roles of some of the major enzymes/proteins involved in the process. The details required for A-level are given in bold in Table 5.

Table 5. Enzymes involved in DNA replication

Enzyme	Function
DNA polymerase III	**The 'replicase' enzyme**
DNA polymerase I	A repair enzyme involved in removing errors and filling in gaps in the sequence
DNA ligase	**Seals 'nicks' in the sugar-phosphate backbone**
RNA polymerase or RNA primase	Make RNA primers to get the DNA polymerase III started
RNase H	Removes the RNA primers once they have completed their function
DNA helicase and DNA gyrase	Unwind the DNA duplex
Helix-destabilizing protein	Prevents the unwound DNA from immediately rewinding

Exposure to ionizing radiation and carcinogenic chemicals, e.g. phenols in cigarette tar, probably 'overwhelm' DNA-repair enzymes resulting in gene mutation and cancerous cells.

4

The genetic code

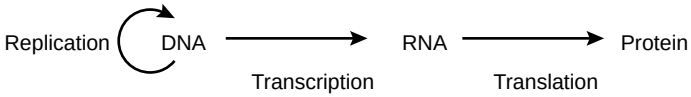
Learning objectives

Each student should, without reference to his or her notes, be able to:

- state that the unit of information is the codon — a sequence of three bases (on DNA and mRNA) representing one amino acid;
- state that since there are four bases there are 64 possible triplet codons;
- state that all codons have been assigned to amino acids or start/stop signals;
- give details of the characteristics of the genetic code; and
- predict the amino acid sequence of a protein from an mRNA or DNA base sequence (with the aid of a copy of the genetic code).

Introduction

The middle to late 1950s was a particularly active and productive phase with respect to the research to aid our understanding of the process of gene expression. At about this time Crick formulated the central dogma which, in the earliest forms, was expressed as:



Thus far, we have seen that the genetic information which dictates a cell's phenotype is stored in the DNA and this can be passed to progeny

cells via the process of replication. We will now consider how this stored information is expressed to yield the specific phenotype.

A key feature in this respect is the concept of a gene which has been defined "operationally as a region of the genome that segregates as a single unit during meiosis and gives rise to a definable phenotypic trait such as a red or white eye in *Drosophila* or a round or wrinkled seed in peas".

The genome is the sum of all DNA nucleotide/base sequences in an individual organism or species.

Later work demonstrated that a gene or gene region corresponds to a length of DNA on the genome that contains the information for the synthesis of a single protein. This led to the useful concept that one gene corresponds to one protein.

However, as a great many biologically active proteins comprise more than one polypeptide chain, it is more correct to consider this as 'one gene — one polypeptide chain'.

Although serving as a useful memory aid, even this statement is not entirely correct since some genes code for single structural RNA molecules, such as mRNA and tRNA, rather than polypeptide chains and, also, more recent studies have revealed additional complexities in terms of the structural organization of the genetic information. For the purposes of the present section, however, it is acceptable to treat the gene in terms of the 'one gene — one polypeptide chain' hypothesis.

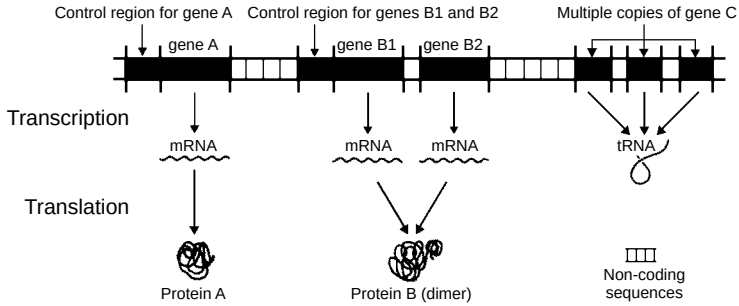
Building a human chromosome

The first synthetic chromosome, that of yeast, was built about a decade ago. To extend this to the synthesis of a human chromosome required detailed consideration of further structural complexities. Thus, in addition to the structural genes there is a need for telomeres, long repeating DNA sequences which protect the ends of the chromosome and prevent them binding together, and centromeres, which provide the scaffolding that enables duplicate chromosomes to split during cell division. Using this approach, totally artificial chromosomes have been produced and it is hoped that these may help in gene therapy in the future.

New Scientist (5 April 1997)

To transfer the genetic information from DNA to a sequence of amino acids that constitutes a polypeptide chain involves two discrete but consecutive processes:
(i) transcription, in which an mRNA strand of nucleotides is synthesized from one of the strands of the DNA in the gene region; and
(ii) translation, in which the genetic information coded as a sequence of bases in mRNA is translated or converted into a sequence of amino acids to form a discrete polypeptide chain.

The gene regions specific to protein synthesis collectively only represent a small part of the genome for many cells. The remaining regions are accounted for as non-coding/'nonsense' sequences, multiple repeat sequences, control regions and regions specifying the synthesis of functional RNA molecules such as tRNA and rRNA. A summary of these features is shown below.



How many bases code for an amino acid?

In the next section we shall consider details of the mechanism by which the information coded in the mRNA is used to synthesize proteins of a specific sequence. An important concept underpinning this topic is the nature of the code used to translate the sequence of bases in the mRNA molecule into a sequence of amino acids in a polypeptide chain. This code is referred to as the genetic code and in the present section we shall consider the general properties of the code.

For mRNA to control the synthesis of specific proteins we must ask how a molecule made up from only four different kinds of base can determine the order of about 20 different amino acids. Obviously there cannot be a 1:1 correspondence between the RNA bases and the amino

acids. Similarly there cannot be a 2:1 correspondence, since this only allows for 16 different arrangements (4×4). If, however, there are three bases coding for each amino acid, the number of possible triplets is 64 ($4 \times 4 \times 4$), which would be enough combinations to code for each amino acid with plenty to spare. Clear-cut evidence has been produced that the triplet theory is correct and we will now look at the evidence for such a view. The name given to the sequence of three nucleotides is a codon.

Evidence for a triplet code

The evidence for a triplet code comes from genetic studies by Crick and co-workers using mutations caused by acridine orange or proflavine in fruit flies. The mode of action of these compounds is not fully understood but the net effect is to cause the insertion of an additional base into a DNA strand or to delete one of the bases in a DNA strand, possibly during DNA replication.

It is easier to explain the results of these studies by considering a simple model and studying its behaviour. If we assume a DNA sequence to be a series of triplet codes of bases, or codons, for five amino acids:

– CAT – CAT – CAT – CAT – CAT –

The insertion of a base, T:

– CAT – **T**CA – TCA – TCA – TCA –

Or deletion of a base, C, from the original sequence:

– CAT – ATC – ATC – ATC – ATC –

The result of a single mutation (either insertion or deletion) is to cause the genetic message/amino acid sequence everywhere to the right of the mutation to be misread.

The result would almost certainly be a defective protein or enzyme, especially if the amino acid sequence is in or near the active site of an enzyme.

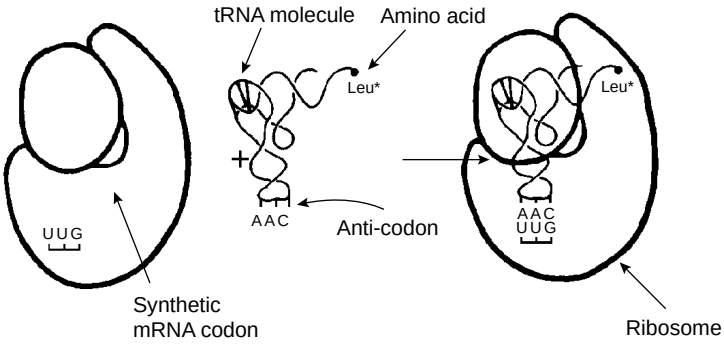
Also in these studies it was shown that if we insert an extra base, then delete another base a little further on in the sequence, the net result is an almost identical product (demonstrate this for yourself). The important point here is that the insertion and deletion are relatively close

together (why?). There are all sorts of other ways of combining mutation, but the net results provide very strong evidence in favour of a triplet code. These results also provide additional data concerning the mode of translating the code.

Deciphering the code

The problem of deciphering the code, that is assigning amino acids to each of the codons, has been approached using a variety of techniques, the most productive being the ribosome-binding technique devised by Leder and Nirenberg. This method, like many of the others, makes use of synthetic mRNA, which in this case is only three bases long, i.e. a single codon. Clearly the advantage of using an mRNA of known sequence is that it makes data interpretation a relatively straightforward process.

The ribosome-binding technique involved mixing a sample of pure synthetic codon with ribosomes and then, in turn, each of the different amino acids (which had been radioactively labelled), each linked to their corresponding tRNA. Each mixture was filtered through a Millipore filter which caused the ribosome and any bound components to be retained on the filter. The filter could then be tested for radioactivity. If we consider as an example the triplet UUG, then it was found that the sample which contained radioactive serine did not bind to the filter, and likewise for the amino acids tryptophan, glycine, methionine, tyrosine, etc. In fact, the only sample which retained radioactivity on the filter occurred when leucine was used (see diagram below) and therefore it can be concluded that the codon UUG codes for leucine (Leu).



This process is then repeated exactly as above for a different triplet code. Using this simple yet highly efficient technique (with some minor modifications) it was possible to assign 61 of the 64 possible triplets unambiguously to specific amino acids, and the accepted code is as shown here.

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

Nonsense codons

As has just been seen, 61 of the 64 triplets have been assigned to particular amino acids — what then is the function of the remaining three, UAA, UAG and UGA? These were initially labelled as the nonsense codons but it has since been shown that they have an important function in protein biosynthesis as chain-terminating signals. It is these codons which inform the protein-synthesizing system that the nascent polypeptide chain has been completely synthesized and to thus allow it to be released from the ribosome.

From the experiments just described, plus other related studies, the general features of the code that are important are as follows:

- the code is sequential (i.e. it is read in strict sequence from one end to the other); and
- the code is degenerate.

A most important feature is that many amino acids are coded by more than one triplet (all except methionine and tryptophan). Indeed it seems that for most amino acids only the first two bases of the codon need be specified. A code of this type is termed degenerate. It follows from this that more than one kind of tRNA may code for the same amino acid. This is in fact known to be the case.

- The code is universal or ubiquitous.

At present, as a result of experimental data, it is generally accepted that the code is universal, e.g. Lipmann and others demonstrated using yeast, *Micrococcal* or *E. coli* tRNA, rabbit reticulocyte ribosomes and mRNA for haemoglobin that a perfectly normal protein, indistinguishable from wild-type, could be formed. More recently, however, some exceptions to this generalization have been identified in mitochondria, leukaemic cells and in some ciliates, and these observations are presently posing a challenge for evolutionary theorists.

5

Transcription

Learning objectives

Each student should, without reference to his or her notes, be able to:

- explain the meaning of the term transcription;
- explain the role of mRNA;
- explain the relationship of mRNA to the DNA template;
- outline the necessity of moving the genetic information required to synthesize proteins from the nucleus to the cytoplasm;
- state that the genetic information is carried to the cytoplasm by an mRNA strand which is complementary to the DNA template strand;
- state that only a relatively short segment of DNA coding for a specific polypeptide is transcribed as a unit from the template strand of DNA;
- state that mRNA is synthesized by RNA polymerase using ribonucleoside triphosphates as substrates and DNA as a template; and
- outline with the aid of simple labelled diagrams the process of transcription.

Introduction

Transcription involves the synthesis of an mRNA molecule complementary to a gene region/specific nucleotide sequence of DNA. This being the case, we require the basic building blocks for RNA, namely free RNA nucleotides and ATP molecules for energy together with a DNA molecule, part of which will be copied, the latter being called a DNA template.

For such a process to be undertaken at an appropriate and sufficiently fast rate for the cell requires the help of an enzyme, which in this case is called DNA-dependent RNA polymerase or just RNA polymerase. This enzyme cannot synthesize an mRNA strand from free RNA nucleotides unless a DNA molecule is present.

It is worth remembering that enzymes have three key advantages.
(i) They act as biological catalysts to speed up biochemical reactions by lowering the energy needed to activate a reaction. This means a reaction is far more likely to proceed under the conditions found in cells, e.g. moderate temperature and low pressures.
(ii) They usually have high specificity in that one very specific reaction (or a few related reactions) is catalysed because of the specific three-dimensional shape of the active site.
(iii) Their activity can be controlled, e.g. by their products and other compounds.

Mechanism of transcription

The enzyme RNA polymerase, like DNA polymerase, operates in the direction 5′→3′, that being the direction in which the mRNA is synthesized. Unlike DNA polymerase it only transcribes one strand of DNA, i.e. 'genes' are located on a single DNA template strand.

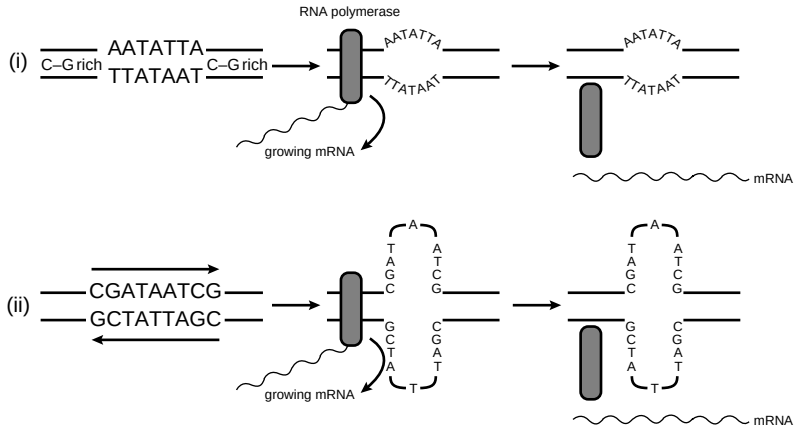
To understand the mechanism of this process it is clearly of interest to know the way in which the start of each gene is correctly recognized. The picture which emerged from a study of the DNA sequences near the start of specific gene regions was that of AT-rich sequences of bases. These sequences were situated a few bases upstream from the starting point of the mRNA transcript, as shown here.

Specific gene	Base sequence near the start of specific gene region
fd	CTGAC**TATAATA**GACAGGGTAAAGACCTG
T7 A2	TGCAG**TAAGATA**CAAATCCGTAGGTAACA
Lac-UV-5	GCTCG**TATAATG**GTTACAAATAAAGCAAT
SV40	CAGCT**TATAATG**GTTACAAATAAAGCAAT
E. coli Tyr tRNA	TTTGA**TATGATG**CGCCCCGCTTCCCGATA

The AT base pair has two H-bonds compared with three for the GC base pair (see Chapter 2); and consequently a region rich in AT pairs tends to melt or strand separate more readily, so forming a bubble. The advantages of forming this bubble are that (i) it helps the RNA polymerase to identify the start of the gene region and (ii) it allows the RNA polymerase to select more easily the template or coding strand and so allow asymmetric synthesis to take place. In fact it is the non-TATA strand (i.e. the 3′→5′ strand) that acts as the template strand in transcription.

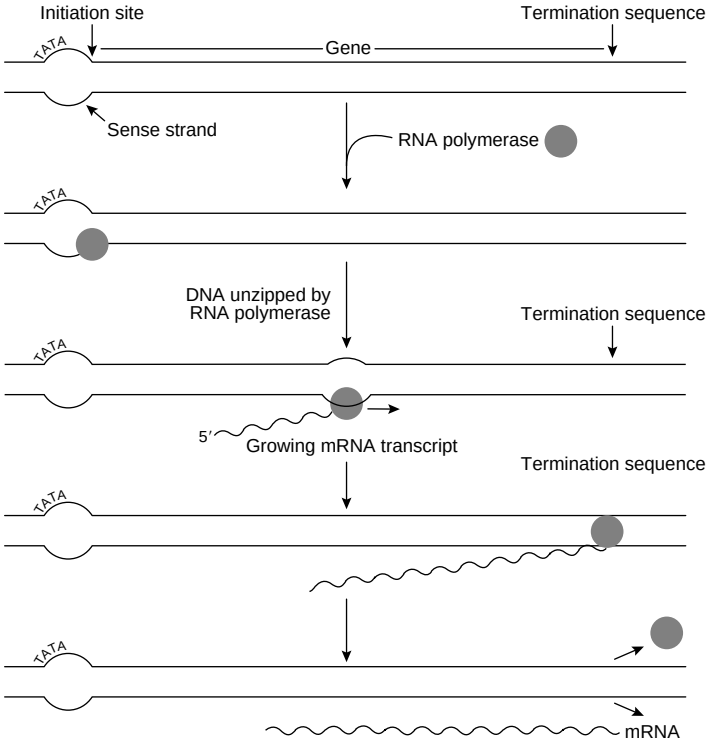
Once correct initiation has been achieved the RNA polymerase moves along the template strand synthesizing an mRNA from RNA nucleotides. The base on each successive RNA nucleotide is complementary to the DNA template strand.

The mechanisms for termination of transcription vary from gene to gene. For some genes an additional protein called rho (ρ) identifies the termination site and causes the RNA polymerase to stop synthesis. Other mechanisms, as shown schematically below, include (i) an AT-rich region flanked on either side by a GC-rich region or (ii) a palindromic region. In both these cases a bubble structure is formed which can be considered to act as a physical barrier, so causing the RNA polymerase to terminate synthesis as shown.



Each mRNA strand then exits the nucleus through pores in the nuclear membrane and attaches to ribosomes in the cell cytoplasm.

The mechanism of transcription is summarized schematically below.



6

Translation

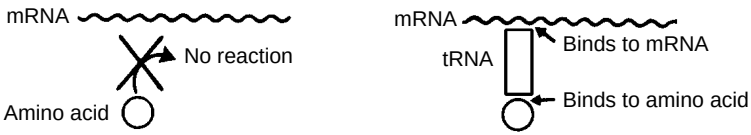
Learning objectives

Each student should, without reference to his or her notes, be able to:

- state that ribosomes are responsible for the translation of the genetic information encoded as a sequence of bases in the mRNA strand into a specific sequence of amino acids in a polypeptide chain;
- outline the functions of ribosomes, mRNA, tRNA and amino acids in translation; and
- outline the processes involved in translation using the terms: codon, anticodon, peptide-bond formation and termination, using simple diagrams.

Introduction

It can be shown that no specific affinity/binding occurs between amino acids and the mRNA, to neither the sugar–phosphate backbone nor the bases (see the Figure below). This being the case, we cannot have a direct 'read-off' in translation as we did in replication and transcription.

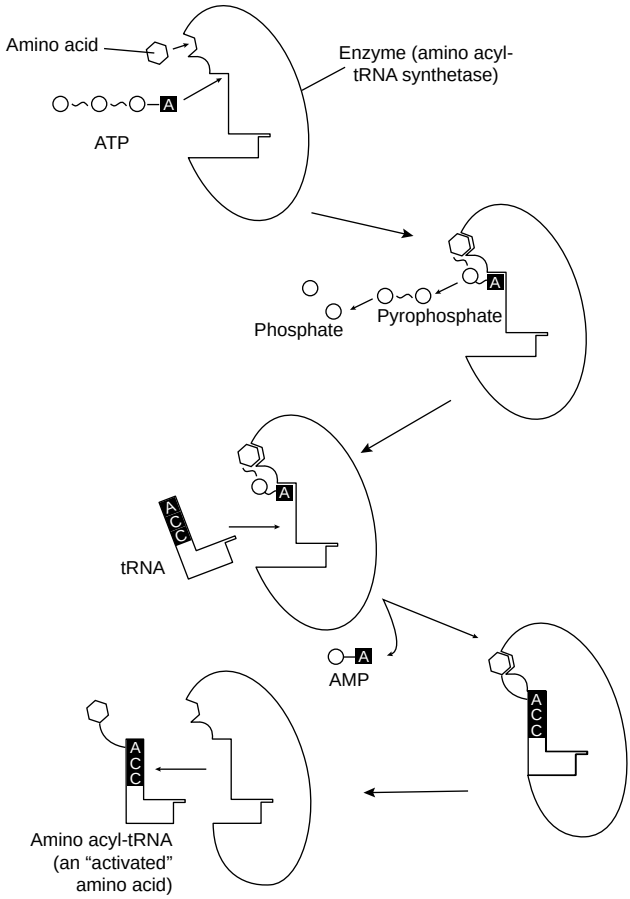


The solution to this problem is to have an indirect association achieved by coupling the specific amino acid to the mRNA via an 'adaptor molecule', which turned out to be tRNA.

The tRNA molecules must be activated before they can attach to a specific amino acid and perform their role in translation of the genetic code.

Activation of tRNA

To bind the amino acid to the tRNA molecule requires ATP energy and a specific enzyme called an amino acyl-tRNA synthetase. The reaction occurs in two stages: in the first the amino acid is activated and in the second the activated amino acid is transferred to its specific tRNA as shown schematically below.

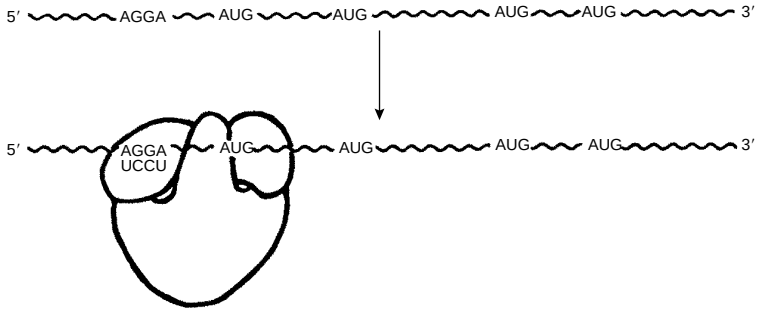


The amino acyl-tRNA synthetase enzyme, apart from catalysing the reaction, also displays high specificity in that it will only bind a particular amino acid to its correct, specific tRNA molecule thus ensuring high accuracy/fidelity in the translation event. The amino acids are linked to the 3′ end of the tRNA (all end with the sequence CCA-OH) via their carboxyl group (COOH).

Initiation

Translation of the mRNA strand starts from the 5′ end at a specific initiation site or triplet codon, AUG, which normally codes for the amino acid methionine. This initiation site AUG codon is distinguished from all other AUG sequences as it is preceded in the mRNA sequence by an AGGA base sequence, known as the Shine–Dalgarno consensus sequence, named after the discoverers.

This sequence is recognized by a UCCU complementary sequence in the tRNA of the small ribosomal subunit and so causes the ribosome to bind to the correct starting point on the mRNA as seen below.

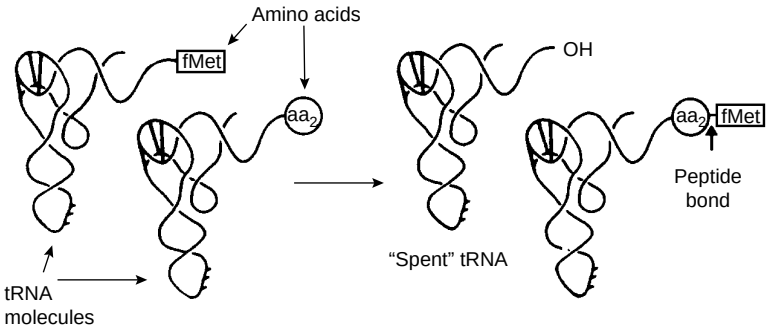


It was also shown that this unique AUG binds a special amino acid (linked to its corresponding tRNA as described previously). This special amino acid is a methionine residue in which the amino group has a formyl group (CHO) bound to it and is called N-formylmethionine (fMet) in prokaryotes. The advantage of this added complexity is that this effectively blocks the amino group and inhibits it reacting, thereby directing the polypeptide chain to be synthesized in the N-terminal→C-terminal direction. Eukaryotes do not use the formyl group.

By such means an initiation complex is formed in which the ribosome is bound to the correct site on the mRNA and the first amino acid (linked to its corresponding tRNA) is in position.

Elongation

After this the next amino acid specified by the codon following the AUG sequence (again an amino acyl-tRNA) is bought into position alongside the first residue in order that these can be linked together via a peptide bond, an example of a strong covalent bond. The energy required for the formation of a peptide bond comes from the high-energy bond between the tRNA molecule and its specific amino acid produced by the breakdown of ATP molecules.

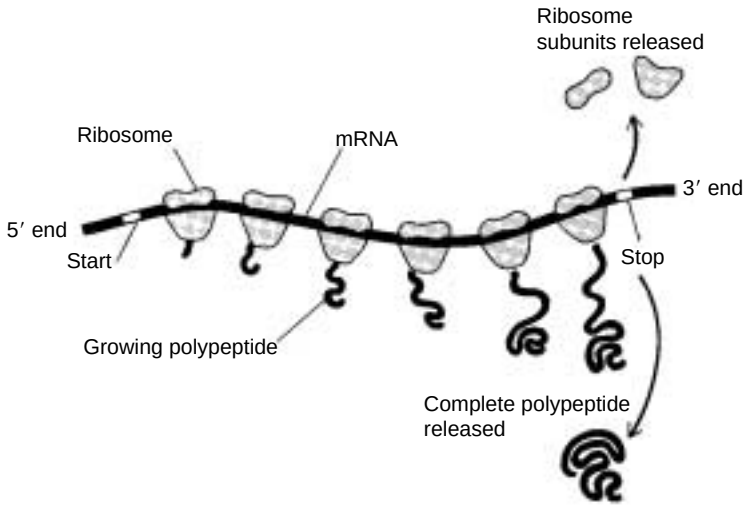


The condensation of (or peptide bond between) the two amino acids is catalysed by an enzyme called peptidyl transferase, one of the constituent proteins of the RNA–protein complex that makes up the large ribosomal subunit. In this way, the fMet is transferred on to the other amino acyl-tRNA and its tRNA (called a 'spent' tRNA) leaves the complex. The ribosome moves along three bases by a process called translocation and this allows the next amino acyl-tRNA to bind to its codon on the mRNA. The individual amino acyl-tRNA molecules bind to the correct codon sequences by virtue of each tRNA having a three base sequence, known as the anticodon, which is complementary to the codon sequence and so H-bonds to it. Because only a specific tRNA molecule attaches to the mRNA codon the correct amino acid is placed in the growing polypeptide sequence. The growing chain is transferred to the incoming amino acyl-tRNA via a second peptide bond and, as

before, the 'spent' tRNA leaves the complex because of the relative weakness of the H-bonds between the codon and anticodon bases.

Termination

The above process continues exactly as described until the ribosome encounters one of the codons referred to earlier as the nonsense codons (see Chapter 4). As those codons do not code for specific amino acids this serves as the signal to terminate the process and consequently the ribosome dissociates from the mRNA and the polypeptide chain is hydrolysed from the tRNA to which it is currently bound.

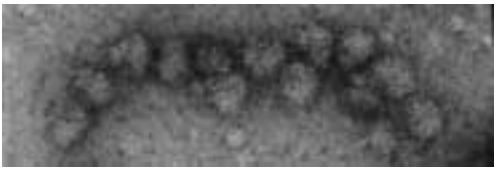


Once protein synthesis is under way, the formyl group is cleaved from the methionine by an enzyme called deformylase and, for many proteins, a number of amino acids are also removed from the N-terminal end by peptidase enzymes. This type of processing is referred to as post-translational modification. The polypeptide chain will now fold up into the secondary and tertiary structure of the functional protein/enzyme and is transported to its required location in the cell through the rough endoplasmic reticulum membrane system.

Therefore, by the processes of transcription and translation the large, permanent, insoluble DNA molecule, which is restricted to the nucleus, can control the expression of its genetic code by producing a protein/enzyme molecule which can now affect a biochemical process throughout the cell.

Polysomes

In the living cell, each strand of mRNA usually carries several ribosomes at any one time, each at a different stage in the formation of the protein which is being synthesized. Such groups of ribosomes strung out on a thread of mRNA are known as polyribosomes or polysomes, as shown here in the electron micrograph.



× 260000

The polysome mechanism protects the mRNA molecule, delaying its breakdown, and greatly increases the efficiency of translation, i.e. many protein molecules can be produced from a single mRNA strand, thus saving the cell energy.

Polysomes are found both free in the cytoplasm and closely associated with the outer face of the rough endoplasmic reticulum membrane system.

Recreating an extinct bird

Groups of researchers in Japan and New Zealand are studying the genetic makeup of the giant moa, a flightless bird which became extinct about 300 years ago. These workers have isolated DNA from the femur of a giant moa and plan to introduce these into chicken embryos. The area of particular interest is the homeobox, which controls embryonic development by switching specific genes on or off. It is hoped in this way to switch on common ancestor genes and to see if they can reproduce some of the features of the extinct moa.

New Scientist (4 January 1997)

7

Assessment: past examination questions and outline solutions

As stated earlier, it is hoped that the inclusion of detailed learning objectives within each section of these guidance notes will not only help students to identify key topics in the syllabus but also help them to gauge the level of expected outcome, and hence be of value in the formal assessment of the material.

Many of the Examination Boards publish detailed analyses of the exam-question performance and it may be worthwhile to read these carefully and modify the guidance notes where appropriate to reflect the varying emphasis placed on the different themes within this broad topic.

The following questions are actual A-level Biology examination questions together with their solutions provided by the authors. Other solutions and approaches may also be valid. Presented in this way they might serve as useful source material in the review of each of the sections at the end of formal lessons. Once again, we are grateful to the various Examination Boards, especially the Northern Examinations and Assessment Board, for allowing past examination questions to be used. Answer all the questions in the spaces provided.

Question 1

The diagram shows the sequence of bases on one strand of a short length of DNA. This sequence should be read from left to right.

CGACCCCAG

(a) Give:

(i) the base sequence that will be produced as a result of transcription of the complete length of DNA shown in the diagram;

. . . . . (2)

(ii) the three bases of the tRNA that will correspond to the sequence of bases shown by the underlined bases on the diagram.

. . . . . (1)

As a result of a mutation, the first base in the length of DNA shown in the diagram is lost (deleted).

(b) (i) Use Table 6 to identify the first two amino acids for which the mutated DNA codes;

. . . . . (1)

Table 6. DNA base sequences and the amino acids for which they code

DNA base sequence	Amino acid
ACC	Tryptophan
CAG	Valine
CCA	Glycine
CCC	Glycine
CGA	Alanine
GAC	Leucine

(ii) explain why mutations involving the deletion of a base may have greater effects than those involving substitution of one base for another.

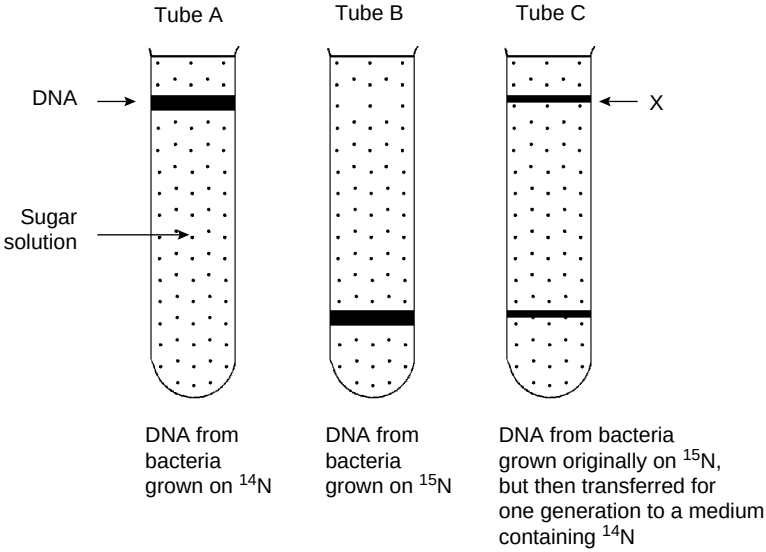
. . . . . .
. . . . . .
. . . . . (2)

Question 2

In an investigation of DNA replication, a species of bacterium was grown for many generations in a medium containing the ^{15}N isotope of nitrogen, rather than its more common form, ^{14}N.

DNA was isolated from the bacterial cells and separated into single strands by mild chemical treatment. The density of this single-stranded DNA was measured using density-gradient centrifugation. In this

technique, samples of different densities settle at different levels in the tube. Some of the results obtained by using this technique are shown in the Figure. All samples were treated in the same manner.



(a) Density-gradient centrifugation enables different types of DNA to be distinguished. Explain how the use of the ^{15}N isotope in this experiment makes this possible.

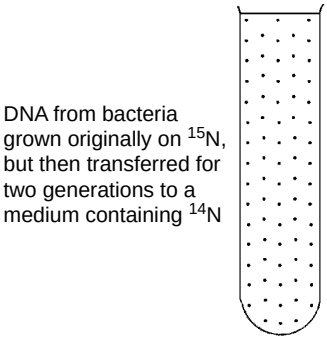
. . . . . .
. . . . . (2)

(b) Account for the results obtained in Tube C, explaining why no DNA of intermediate density was observed at this stage.

. . . . . .
. . . . . .
. . . . . (3)

(c) Bacteria like those that gave the result shown in Tube C of the Figure were grown for a further generation on a medium containing ^{14}N. DNA was isolated from these bacteria and treated as in the first investigation.

Show, on the following Figure, the result expected with density-gradient centrifugation. Also, indicate the relative amount of each type of DNA you would expect to be present.



. . . . . (2)

Studies were carried out to determine the nitrogenous base composition of DNA in this bacterium. This involved finding values for double-stranded DNA as well as for each individual (+ and −) DNA strand.

The Table below gives some of the results obtained.

Table 7. Base composition of DNA

	Percentage of base present in DNA sample				Ratio of (A+G)	Ratio of (A+T)
DNA sample	**A**	**G**	**C**	**T**	**to (C+T)**	**to (C+G)**
Double-stranded	30	20	20	…	1.00	1.5
+ Strand	28	26	…	32	1.17	…
− Strand	32	…	…	…	0.85	…

(d) Enter the seven 'missing' values in the table. (4)

(e) What ratio of (A+G) to (C+T) would you expect in the band of DNA marked as 'X' in Tube C in the first Figure? Explain your answer.

. . . . . .
. . . . . (2)

Question 3

Albino mice are completely white because they cannot produce coat pigment. They cannot produce this pigment because of a gene mutation that affects the formation of an enzyme in the metabolic pathway that

produces the pigment. The enzyme produced by the mutant gene does not function effectively.

(a) (i) Explain what is meant by gene mutation;

. . . . . (1)

(ii) explain how a gene mutation could result in a change in the amino acid sequence in the enzyme produced by the coat-pigment gene.

. . . . . .

. . . . . .

. . . . . (4)

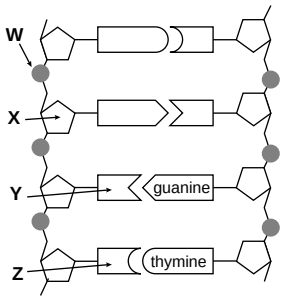
(b) Explain why mice that are heterozygous for this mutant gene are never albinos.

. . . . . .

. . . . . (2)

Question 4

The diagram shows part of a DNA molecule.



(a) How are the two strands of the DNA molecule held together?

. . . . . (1)

(b) What do each of the following letters on the diagram represent?

W . . . . . .

X . . . . . .

Y . . . . . .

Z . . . . . (4)

Question 5

One technique used to produce human insulin by genetic engineering involves inserting a gene for human insulin into the DNA of a bacterium.

(a) Name the enzyme that would be used to:
(i) cut the bacterial DNA;
. . . . . (1)
(ii) insert the DNA for human insulin into the cut bacterial DNA.
. . . . . (1)

(b) There are 51 amino acids in insulin, made up of 16 amino acids out of the 20 that are coded for by DNA. What is the minimum number of different types of tRNA molecule necessary for the synthesis of insulin? Explain your answer.
. . . . . .
. . . . . (2)

(c) The base sequence below is part of the DNA sequence that codes for insulin.

CCATAGCAC

(i) Write down the corresponding mRNA sequence;
. . . . . (1)
(ii) A mutation occurred that replaced guanine in this DNA sequence with a different base. Explain the possible effects of this mutation on the structure of the insulin molecule.
. . . . . .
. . . . . .
. . . . . (3)

Question 6

Scientists have shown that kidney beans are resistant to cowpea weevils and adzuki bean weevils, two of the most serious pests of African and Asian pulses (vegetables related to peas and beans). This is because the beans produce a protein that inhibits one of the weevils' digestive enzymes. Weevils that eat the pulses soon starve to death. The researchers have identified the gene that produces the inhibitor and removed it from the kidney bean DNA. They inserted the gene that produces the inhibitor into the DNA of a bacterium called *Agrobacterium tumefaciens*. Using this bacterium they have been able to add

the inhibitor gene to peas. They hope soon to be able to add the gene to African and Asian pulses.

(a) Describe how scientists could:
(i) remove the gene that produces the inhibitor from kidney beans;
. . . . . .
. . . . . (2)
(ii) insert this gene into the DNA of a bacterium.
. . . . . .
. . . . . (2)

(b) The DNA in the bacterium is able to replicate to produce many copies of itself for insertion into pea cells. Describe the structure of a DNA molecule and explain how this structure enables the molecule to replicate itself.
. . . . . .
. . . . . .
. . . . . .
. . . . . (8)

Question 7

(a) Draw a labelled diagram of a representative length of the DNA (deoxyribonucleic acid) molecule (the chemical structure of the different constituents is not required). (3)

(b) Name the repeating units that are joined to form the macromolecule of DNA.
. . . . . (1)

(c) Suppose that a section of DNA has the following sequence of bases:

GGATTAACACCT

What will be the complementary messenger RNA (ribonucleic acid) sequence?

. . . . . (2)

(d) Using the Table of amino acid triplets below, determine the sequence of the amino acids of the protein synthesized from this messenger RNA.

Table 8. Selected amino acids and their related mRNA triplet codes

Triplet code	Amino acid
AAU	Asparagine
CCU	Proline
GAA	Glutamic acid
AUG	Methionine
CUA	Leucine
GGA	Glycine
ACA	Threonine
GCU	Alanine
UGU	Cysteine

. . . . . .
. . . . . (2)

(e) Name the two other kinds of RNA that participate in protein synthesis.

. . . . . (2)

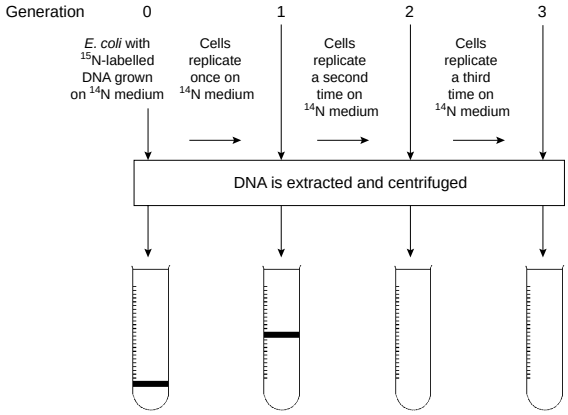
Question 8

In 1958, Meselson and Stahl published the results of an experiment which provided strong evidence that cells produce new DNA by a process of semi-conservative replication.

(a) Why is replication of DNA described as semi-conservative?

. . . . . .
. . . . . (2)

Meselson and Stahl's experiment is outlined in the following Figure (^{15}N is a heavy isotope of nitrogen).



(b) Which component of the DNA was labelled with the ^{15}N?

. . . . . (1)

(c) Explain why centrifugation separates the DNA labelled with different isotopes of nitrogen.

. . . . . (2)

(d) On the diagram above, draw in the results you would expect for generations 2 and 3.

. . . . . (2)

Question 9

Read the following passage.

For more than 30 years, Professor Stanley Miller of the University of California has been trying to show how life began. From a primitive mixture of hydrogen, ammonia, methane and other simple chemicals exposed to lightning and ultraviolet radiation, he has been trying to produce a self-replicating chemical that might have been the first form of life.

His latest experiments support the view that early life was based on RNA, rather than DNA to which it is closely related. RNA still plays a central role in life as the agent for transferring the information stored in

DNA and used to make proteins. In early life it seems to have been more important still. RNA can not only act as a store of genetic information and a means of transferring it, but can also act as a catalyst.

Professor Miller's latest experiments go back to an earlier stage of the process, seeking to show how the RNA itself might have been produced. He has shown that it is likely that all the components on which RNA is based could have been made in the conditions on the Earth 3.5 billion years ago.

Adapted from an article by Nigel Hawkes in *The Times* (3 July 1995).

(a) In what ways are the structures of RNA and DNA alike, and in what ways are they different?

. . . . . .
. . . . . .
. . . . . (4)

(b) Explain how RNA acts 'as the agent for transferring the information stored in DNA that is used to make proteins' (lines 7 and 8).

. . . . . .
. . . . . .
. . . . . .
. . . . . (6)

(c) Suggest why a nucleic acid molecule would have been essential in the first forms of life.

. . . . . .
. . . . . (2)

Answers

Question 1

(a) (i) GCUGGGUC.(2 marks if fully correct; 1 mark if only G and C correct)
(ii) CAG. (1)

(b) (i) Leucine and glycine. (1)
(ii) Deletion affects all (subsequent) triplets in the DNA sequence and so may alter entire protein; substitution may alter only a single amino acid/triplet codon; if the third base in the triplet changed, it may have no effect/degenerate code. (any 2)

Question 2

(a) ^{15}N is incorporated into (nitrogenous) base of DNA. Its presence increases the density of the DNA. (2)

(b) The DNA in Tube C consists of a dense ^{15}N strand, the template used in replication. Also present is a less-dense ^{14}N strand, the newly replicated

strand. No intermediate exists because the DNA was made single stranded before it was centrifuged. (3)

(c) There will be two bands present, one of lower density (top of the tube) and one of higher density. The former contains about three times as much DNA as the latter. (2)

(d)

DNA Sample	Percentage of base present in DNA sample A	G	C	T	Ratio of (A+G) to (C+T)	Ratio of (A+T) to (C+G)
Double-stranded	30	20	20	**30**	1.00	1.5
+ Strand	28	26	**14**	32	1.17	**1.5**
− Strand	32	**14**	**26**	**28**	0.85	**1.5**

(4)

(e) 1.00, because it is composed of equal numbers of + and − strands. (2)

Question 3

(a) (i) Change in sequence of bases/nucleotides in gene/codon. (1)
(ii) Any four of the following: (substitution) leading to different codon/bases at this point (in DNA) chain; leads to 'wrong' amino acid insertion at this point in polypeptide; deletion/addition alters all subsequent amino acids; leads to changed sequence of all subsequent amino acids; copy of changed (codon) in mRNA; reference to ribosomes; reference to anti-codons on tRNA; reference to specific amino acid attached to tRNA. (4)

(b) Heterozygotes always have one dominant *allele*, therefore one DNA chain will always produce enough functional enzyme. (2)

Question 4

(a) Hydrogen bonding/specific base pairing. (1)

(b) W = phosphate or phosphoric acid group; X = deoxyribose/sugar/pentose; Y = cytosine; Z = adenine. (4)

Question 5

(a) (i) Restriction enzyme/endonuclease/named enzyme e.g. *Eco*R1, *Hin*dIII, *Bam*HI. (1)
(ii) DNA ligase (see BASC Booklet 7, *Recombinant DNA Technology* for further details). (1)

(b) 16, because each tRNA is specific to one amino acid. (2)

(c) (i) GGUAUCGUG; (1)
(ii) any four of the following: might code for a stop codon; might code for a different amino acid; this would make a modified protein; protein might not function; amino acid might not alter (because the change occurs on third triplet); therefore no change to protein made. (3)

Question 6

(a) (i) Restriction enzyme/endonuclease. Removes section of DNA. (2)
(ii) Two of the following: cut bacterial DNA/plasmid; ligase; splices in DNA; reference to vectors, e.g. phage. (2)
(See BASC Booklet 7, *Recombinant DNA Technology* for further details.)

(b) Eight of the following (allow points on diagram if appropriately labelled/annotated): reference to nucleotides; four bases named; reference to sugar–phosphate 'backbone'; reference to two (polynucleotide) strands;

reference to specific base-pairing; example, e.g. A–T/C–G; reference to bonding/hydrogen bonds; reference to 'uncoiling'/'unzipping'; reference to new complementary strand formation. (8)

Question 7

(a) Phosphate/sugar/base: in correct sequence; antiparallel strands. (3)
(b) Nucleotides. (1)
(c) CCUAAUUGUGGA (2; 1 mark if single error)
(d) Proline, asparagine, cysteine and glycine (2; 1 mark if single error)
(e) Transfer RNA and ribosomal RNA. (2)

Question 8

(a) Both daughter helices contain one (conserved) strand from the parent molecule and one newly synthesized complementary DNA strand. (2)
(b) The nitrogenous (nitrogen-containing) bases. (1)
(c) The ^{15}N DNA has a higher density than that of ^{14}N and the difference in buoyant density is sufficiently great to be able to be resolved as separate bands in a caesium chloride gradient, ^{15}N moves further down in gradient. (2)
(d) Generation 2 would have two bands of equal intensity, one at the hybrid density and one at the light density. Generation 3 would also have two bands at the same densities but the ratio of amount of DNA would be 3:1, light:hybrid. (2)

Question 9

(a) Alike: contain the same bases adenine, guanine and cytosine; basic building blocks of nucleotides consist of sugar, phosphate and base; nucleotides linked by phosphodiester bonds.
Different: DNA contains the base thymine whilst RNA contains the base uracil; DNA contains the sugar deoxyribose whilst RNA has the sugar ribose; DNA is usually double stranded and large; RNA is usually single stranded and relatively small. (4; maximum of 2 for each section)
(b) Specific portion of the DNA helix coding for the protein 'unzips'; complementary RNA nucleotides line up opposite exposed DNA bases; template DNA strand is used; mRNA strand synthesized by DNA-dependent RNA polymerase; mRNA strand released and exits nucleus through pore in membrane; attaches to ribosomes (contains rRNA); each tRNA molecule carries a specific amino acid; ribosomes move along mRNA strand; codon/anticodon interaction ensures correct sequence of amino acids; adjacent amino acids joined by peptide bond to produce polypeptide/protein. (any 6)
(c) The need to have a stable, compact store of information, a self-replicating molecule which can mutate to allow evolutionary survival of the fittest. (any 2)

8

Laboratory practicals

Extraction and characterization of DNA from *Micrococcus lysodeikticus*

As DNA is a large fragile molecule, isolation methods must be mild in order to avoid excessive breakage of DNA chains. Extreme conditions (either of pH or temperature) must be avoided and the following procedure describes a method of DNA preparation from a bacterial source. Students should record carefully any changes which occur at each stage of the method and try to account for any changes.

Method (day 1, 2 hours)

Suspend 0.5 g of freeze-dried *Micrococcus lysodeikticus* cells in 20 ml of 0.15 M NaCl/0.1 M EDTA, pH 8.0. Add 20 mg of lysozyme and incubate at 37°C for 15 min. Add 1.6 ml of 25% (w/v) sodium dodecyl sulphate (SDS) solution, mix and incubate for 10 min at 65°C. Cool and add 5.5 ml of 5 M sodium perchlorate followed by 25 ml of chloroform-isoamyl alcohol (25:1 v/v). Mix thoroughly for 15 min. Centrifuge for 10 min at high speed. Carefully draw off the top aqueous layer with a Pasteur pipette and place the solution in a narrow beaker. Carefully layer on top an equal volume of ice-cold absolute alcohol. Using a clean glass rod collect the DNA precipitate from the interface and transfer to a clean portion of ethanol. Wash the DNA by unwinding and rewinding onto the rod. Transfer to a tube containing 3 ml of SSC (0.15 M NaCl/0.015M sodium citrate, pH 7.0). Leave until next practical period to dissolve.

Properties of DNA

Hyperchromicity

This feature of DNA arises from the fact that when double-stranded DNA is denatured (i.e. strands are separated) the absorbance at 260 nm increases substantially (up to 40%).

Base composition

The base compositions of DNA and RNA differ, and this can be demonstrated by chromatography of hydrolysates of samples of each. By eluting the separated bases and determining the amount of each spectrophotometrically the base ratios can be calculated and compared.

Method (day 2, 2.5 hours)

Make a 1-in-10 dilution of 1 ml of your DNA sample using SSC. If a spectrophotometer is available record the ultraviolet absorption spectrum using SSC as blank (further dilution may be necessary). Heat a sample of your dilute DNA to 100°C for 15 min and cool quickly. Re-record the spectrum.

Place approx. 20 mg of DNA and RNA provided and your DNA solution [an A_{260} of 1 optical density (OD) unit is approx. 50 μg/ml] in separate clean test tubes and add 0.5 ml of 70% $HClO_4$. Heat in a boiling-water bath for 1 h in a fume cupboard. Cool, add an equal volume of water, and spot 10 μl samples on to a chromatogram together with 10 μl samples of thymine, adenine, guanine, cytosine and uracil solutions. Develop the chromatogram using 65% isopropanol/16% concentrated HCl/19% water as solvent.

Method (day 3, 1.5 hours)

Examine the chromatogram under ultraviolet light and mark in pencil lines around the spots (wear a face mask!). Measure R_F values and cut out the separated bases with scissors (including approx. equal areas of the chromatograms as a blank). Transfer to clean test tubes and elute the bases by adding 10 ml of 0.1 M HCl. Heat to 45°C for 10–15 min and allow to cool. Record the ultraviolet absorption spectrum of each eluate and calculate concentrations using molar extinction coefficients.

Purification of RNA

Before use in biochemistry practical classes, commercially available RNA must first be purified and treated in the following sequence:
(i) 10 g of RNA (BDH-yeast) taken in some distilled water and dissolved by bringing to pH 6.8–7.0 with NaOH;
(ii) dilute to 200 ml with distilled water;
(iii) spin down insoluble residue in a centrifuge (5000 revolutions/min for 15 min);
(iv) dialyse for 48 h in three changes of 2 litres of ice-cold distilled water;
(v) precipitate with 2 vols. of 95% ethanol and solid sodium acetate to make 2% solution;
(vi) spin down RNA, redissolved to 200 ml with distilled water;
(vii) dialyse for 24 h against distilled water;
(viii) precipitate with 2 vols. of 95% ethanol and bring to 2% with solid sodium acetate;
(ix) RNA spun down and wash the precipitate twice with ethanol:ether (1:1);
(x) wash precipitated RNA with ether and air dry (stirring well).
Store cold and use within 8 weeks.

Further experiments relating to the physical and chemical properties of DNA and RNA can be performed using the DNA isolated in this way or using material obtained from commercial sources. The choice of experiments is best made by the individual teacher based on such factors as availability of specialist equipment, cost of consumables, safety and practical time available. Titles for some practicals in this area are given below together with the appropriate reference. Practicals relating to functional aspects of the nucleic acids that are suitable for school are sparse, largely because many depend on the use of very specialist equipment or expensive and potentially hazardous radiochemicals. In these cases use may be made of computer simulations and/or videotape material and possible titles are included below and in the list that follows. School teachers may also wish to take the opportunity of linking with a neighbouring institute of higher education in order to establish access to specialist equipment of this type. At the present time many such institutions would welcome such an approach.

Further practicals

Falconer, A.C. and Hayes, L.J. (1986) The Extraction and Partial Purification of Bacterial DNA as a Practical Exercise for GCE Advanced Level Students. *Journal of Biological Education* **20**, 25–26

Wood, P. (1984) Use of Computer Simulations in Microbial and Molecular Genetics. *Journal of Biological Education* **20**, 309–312

Goodridge, F. (1983) The Teaching of Protein Synthesis — a Microcomputer Based Method. *Journal of Biological Education* **17**, 222–224

New, J. and Jolley, R. (1986) Ultraviolet Radiation Induction of Mutation in *Penicillium claviforme*. *Journal of Biological Education* **20**, 155–157

Franco, R. and Canela, E.I. (1985) SIMCODE; A Program for Simulating Point Mutations in Genomic DNA. *Biochemical Education* **13**, 66–67

Spinder, S., Siebert, P., Coffman, F. and Jurnak, F. (1984) Isolation of Biologically Active mRNA. *Biochemical Education* **12**, 22–25

Rosenfield, I.Z.H.E. and Van Loon, B. (1983) *DNA for Beginners*, Writers and Readers Publishing Co., location ISBN 0 86316 023 9

9

Source materials

Textbooks

Boyer, R.F. (1986) *Modern Experimental Biochemistry*, Addison–Wesley Publishing Co. Inc., Wokingham [experiments 15–20]

Calladine, C.R. and Drew, H.R. (1997) *Understanding DNA* (2nd edition), Academic Press, London

Campbell, N.A. (1996) *Biology* (4th edition), Benjamin/Cummings Publishing Company, California

Drlica, K. (1984) *Understanding DNA and Gene Cloning: A Guide for the Curious*, John Wiley and Sons, Chichester

Garrett, R.H. and Grisham, C.M. (1997) *Biochemistry*, Saunders College Publishing, California

Matthews, C.K. and van Holde, K.E. (1996) *Biochemistry* (2nd edition), Benjamin/Cummings Publishing Company, California

Matthews, H.R., Freedland, R. and Miesfeld, R.L. (1997) *Biochemistry: A Short Course*, John Wiley and Sons, Chichester

Rosenfield, I., Riff, E. and Van Loon, B. (1983) *DNA for Beginners*, Writers and Readers Publishing Co., ISBN 0 86316 023 9

Sinden, R.R. (1994) *DNA Structure and Function*, Academic Press, London

Solomon, E.P., Berg, L.R., Martin, D.W. and Villee, C. (1996) *Biology* (4th edition), Saunders College Publishing, California

Voet, D. and Voet, J.G. (1995) *Biochemistry* (2nd edition), John Wiley and Sons, Chichester

Articles

Darnell Jr., J.E. (1985) RNA. *Scientific American*, **253**, 54–72

Felsenfeld, G. (1985) DNA. *Scientific American* **253**, 54–72

The Molecular Basis of Evolution. (1985) *Scientific American* **253**, 148–157

Weinberg, R.A. (1985) The Molecules of Life. *Scientific American* **253**, 34–43

Teaching aids

Fletterick, R.J., Schroer, T. and Matela, R.J. (1985) *Blackwell Molecular Models* (Patent No. 437218) together with *Molecular Structure: Macromolecules in Three Dimensions*, Blackwell Scientific Publications, Oxford

Parish J.H. (1986) *Nucleic Acid Structure and Synthesis*, IRL Press Ltd., Oxford

Bleasby, A.J. and Parish J.H. (1986) *The Genetic Code and Protein Synthesis*, IRL Press Ltd., Oxford

Lowrie, R.S. *Minit Molecular Modeling System.* DNA and RNA model kits. Cochranes of Oxford Ltd., Oxford

Nicholson Molecular Models. DNA and RNA model kits. Labquip, Wokingham

The Bio-Bombshell (1984) *New Scientist* Video, IPC Magazines Ltd., London

The DNA Story, John Wiley and Sons Ltd., Chichester

DNA - Blueprint of Life, John Wiley and Sons Ltd., Chichester

Molecular Models Kit (DNA and RNA), Molecular Design Inc., Academic Press, London

Pipe cleaners: useful for distinguishing left and right-handed helices

The Biochemical Basis of Biology (video programme), volumes 2 and 3, The Biochemical Society, London

DNA Kit, NCBE, The University of Reading, Reading

Plant DNA Investigation Kit, NCBE, The University of Reading, Reading

DNA Model, NCBE, The University of Reading, Reading

Subject index