

Contents

- 0 Learning unit 0: Introduction
- 1 Learning unit 1: Introduction to cells and biological molecules
- 2 Learning unit 2: The internal organisation of the cell
- 3 Learning unit 3: DNA, chromosomes and genes
- 4 Learning unit 4: DNA replication and repair
- 5 Learning unit 5: Transcription and translation
- 6 Learning unit 6: Regulation of gene expression
- 7 Learning unit 7: The cell cycle
- 8 Learning unit 8: Laboratory techniques in molecular biology

Announcements

Discussions

LEARNING UNIT 0:

Welcome and introduction to module

0.1 Welcome message

Welcome to this module and the study of molecular biology (BMI2604). Molecular biology focuses on the interaction between large molecules such as proteins, lipids, DNA and RNA that act together in a synchronised way in biological systems.

On completion of the module you should have an in-depth understanding of some of the important processes that occur within cells, specifically with regard to DNA replication and how the information within the DNA is transformed into RNA and protein molecules. The control of these processes is also fundamentally important for the cell to function correctly and we will discuss this. These processes are essential to the life of a cell and form the foundation of molecular biology.

Molecular biology presents fascinating challenges to the researcher. To assist you in getting the most from this module, we strongly recommend that you access molecular biology, biotechnology, biochemistry and microbiology internet websites so that you can add to the notes that you will prepare from your textbook.

This module is a 12-credit course, meaning you will need at least 120 study hours for it. You will need to prepare a portfolio of your studies as you work through the material in the course. The myUnisa website will be the main teaching medium for this module. You need to visit this site frequently to interact with your fellow students and to participate in discussions about certain topics within the course. Try to go online at least once a week. We will explain the structure of the course in more detail in the next sections.

Molecular biology, although a complex subject, can be fascinating. We hope that you will find this course interesting and we wish you success on your journey.

0.2 Course information

0.2.1 How this course is organised

Before starting on the first learning unit of this study guide, we are going to present some details of the course. There are a number of central concepts that are important for understanding how cellular systems function and how this knowledge can be applied in a molecular biology laboratory. These central concepts form the basis for the learning units in this module.

- Unit 1: Introduction to cells and biological molecules
- Unit 2: The internal organisation of the cell
- Unit 3: DNA, chromosomes and genes
- Unit 4: DNA replication and repair
- Unit 5: Transcription and translation
- Unit 6: Regulation of gene expression
- Unit 7: The cell cycle
- Unit 8: Laboratory techniques in molecular biology

0.2.2 Textbook

The prescribed textbook for Molecular Biology (BMI2604) that you need to use in conjunction with the online material is:

Alberts, B, Johnson, A, Lewis, J, Raff, M, Martin, K & Walter, P. 2012. *Molecular biology of the cell*. New York: Garland Science. (ISBN: 978-0-8153-44643)

The textbook is a comprehensive guide to the molecular biology of the cell. You will not be required to learn the whole textbook. Use the online study material to guide you in what you need to learn. If you find a topic particularly interesting, then feel free to do further reading on that topic. Note that if you purchase the latest edition of the textbook, you may find that the pages in the study guide do not directly correlate to the pages in this latest edition of the textbook. However, I am sure that you will find it easy to locate the relevant section, rather than the specific pages, in the newer textbook.

In the study guide we will refer to the textbook as Alberts.

0.2.3 Purpose of this module

This introductory course deals with the fundamentals of molecular biology. The disciplines of biochemistry, microbiology, molecular biology and genetics are applied and adapted as foundation studies to further investigate your understanding of the interactions of macromolecules such as proteins, DNA, lipids and RNA within cells that result in life. You will also use this knowledge to practically manipulate these molecules to understand life and also to benefit humankind. Bear in mind that changes to these molecules may contribute to changes in human cells that affect various aspects of health and disease in humans. In summary, the purpose of this module is, therefore, to enable you to identify and apply practices, processes and principles of molecular biology to solve problems at molecular and cellular level. We trust that you will find the module interesting and useful, and that you will be able to apply the knowledge not only in your career, but also towards a better understanding of your own health.

0.2.4 Outcomes of the module

This module is designed to introduce you to some of the important processes that are essential to life and that form the foundation of molecular biology.

After completing this module, you should be able to

- explain the following processes and why they are important in biological organisms:
 - DNA replication
 - DNA repair
 - transcription
 - translation
 - regulation of gene expression
- discuss some modern molecular laboratory techniques and how they can be used to isolate and characterise DNA for further research
- describe some applications of molecular biology in biotechnology and medical diagnosis

0.3 Distance learning

Distance studies are unique, with particular requirements for success that should not be underestimated. Once you have received your study material, plan how you will approach and complete this module. Draw up a reasonable study schedule that can guide you through the whole module. Take into consideration the due dates of the assignments as given in Tutorial Letter 101.

0.3.1 Independent study

A crucial phase in the process of understanding and learning the basics of molecular biology is to articulate your ideas about these principles, both orally and in writing. Only when you have tried this process for yourself will you understand the full value of this exercise.

The assessment following your studies measures an aspect of success. This module will have formative (ongoing) and summative (final) assignments and examinations mainly in the form of written work. Your reflections on your process of learning are therefore also an important part of your work. Since the focus in this module is to understand and apply the concepts of molecular biology, your assessment will focus on the

imbedded abilities to be competent at doing this.

We will work through the study guide, using the guidelines in the next section (“Improved study skills”). This involves drawing up mind maps and making your own summaries of the objectives and contents of chapters. Restrict summaries to one page. Be conscientious about this: it is vital to keep consulting the textbook if you are to master its contents. Additional textbooks and articles give alternative views or provide more insight into issues under discussion and are optional additional reading.

Be focused in your studies. Build up your own study and exam preparation portfolio (with your assignments, activities, reflections, summaries, self-assessments and notes) throughout your academic and/or experiential learning. This portfolio will not be assessed by the lecturer, but you will need to prepare it so that you can complete the assignments and ultimately pass the final examination. It is also very important to use this portfolio, in combination with your assignments and subsequent feedback (tutorial letters), for your exam preparation. The advantage is that you engage personally in your learning, you set goals, you evaluate your own progress through reflective actions and you evaluate your ability to realise the learning outcomes, thus becoming a more independent and self-directed learner.

What is a portfolio? A portfolio is a folder/file in which you gather and compile additional and/or summarised information during the year as you work through the study guide. This portfolio will help you to prepare for the examination by focusing on the most important facts and issues.

Your portfolio should comprise

- answers to each activity in each learning unit
- a mind map/summary of each learning unit
- your marked assignments (or a copy you made prior to submitting your assignment)
- your reflections on each learning unit
- extra reading material taken from the internet, additional books, medical and/or scientific journals
- a new vocabulary of words or glossary of new terms in your own words

To ensure that you master this module, use the following study skills guidelines.

0.3.2 Improved study skills

It is critical that you think independently and learn to look beyond the prescribed book alone. Many additional references are included in this study guide and we encourage you to consult them. In addition, as a more advanced distance education student, it is essential that you become familiar with searching for research/scientific articles on the internet.

0.3.3 How to search for research/scientific articles

As a general source of information, the Google website has now created an additional search engine under “advanced search”, called Google Scholar. This has its own advanced search with a number of linked fields. This allows a brief description of your subject query, say in four to six words, to be linked to specific phrases. On pressing “search enter”, a variety of websites relating to the query will appear. The advantage of using this portal is that you can access most of the journal references from any internet site, other than myUnisa. Certain journals, such as *ScienceDirect* on the Elsevier site, can only be accessed through a tertiary academic institution such as Unisa. To access this journal, you need to do this:

1. Go to Unisa online at <http://www.unisa.ac.za/>
2. Click on Library at the top of the page.
3. In the maroon area at the top of the page, click on “Find e-resources”.
4. Follow the guidelines if you are a first-time user.
5. Click on the option “A-Z list of all electronic resources”.
6. Various links for databases will now be on your screen. Click on any database to do a search. For

Molecular Biology we recommend clicking on ScienceDirect, Nature or Springer Link. (Remember, to find ScienceDirect, click on s on top and a list of all the databases starting with s will appear; if you want to go to Nature, click on the n on top, etc.).

7. Once you enter one of these databases, you can search for scientific articles by typing in the relevant keywords in the “search” box. Be very specific with the keywords. One word will usually result in too much information and much of it won't be on the specific topic you are looking for.
8. You will need to do some independent searches yourself, as part of your portfolio, assignments and exam preparation. Especially since this is a distance education course, you need to supplement it with information from internet sources.

Contact Unisa Library if you have any difficulties or for assistance: +27 12 429 3206 or see the library website for the local branch library's telephone number.

0.3.4 Skimming, scanning, study-reading strategy (SSS strategy)

There are a number of strategies that can help you study, one of them being the SSS strategy. The three techniques in the SSS strategy are **skimming**, **scanning** and **study-reading**. The strategy comprises six steps altogether. What do these steps involve?

Skimming

1. **Page through and explore.** First, read the section quickly, forming a rough idea of the contents. Concentrate on headings and subheadings, bold and italic type, boxes, tables and illustrations, and - in the case of a chapter – introductions and summaries. The objectives of a chapter are important. (Think of how you would page through a magazine. When starting a new learning unit, scan it and concentrate on the concepts that catch your eye.)
2. **Make a cursory survey.** Ask yourself while you read: What key terms occur in this section or chapter? Stop when you identify a key term and read carefully what is said about it. Mark it in the book. What you are trying to ascertain is: **Where** is it? In other words, where is the information that you will need to discuss later?

Scanning and reflecting

3. **Scan** the section or chapter.
4. **Start a mind map** (for the whole section or chapter or for parts of it, as in starting a summary). You are looking for items and concepts while reading the information in the section or chapter in a more evaluative way. Reflect on interrelationships between concepts. The question now is: **What** is it? What is the meaning and the purpose? Visualisation is important and you will certainly start writing down key concepts. You can omit parts of the text.
5. **Deeper reflection.** Start by giving your mind map a structure. As you work through the prescribed activities of the section or chapter, keep returning to the mind map to fill in the detail. Reflect on the value and meaning of categories, concepts, motivations, variables and key terms.

Study-reading

6. **Study-read.** This follows directly from stages 2, 4 and 5 and is done carefully, thoroughly and thoughtfully. The key terms and concepts you have pinpointed have to be linked up, and here the mind map and summaries are important. (Remember to put your detailed mind map in your portfolio.) Pause while reading, consolidate what you remember and consider how new information fits in with what you already have. This will give you a good idea of the whole picture.
7. **Activity-based approach**

Whenever you get to an activity in your study guide, complete it in full on loose pages which you then insert in your portfolio, grouped together per section and learning unit. Supplement this with your own

notes from your portfolio. (You don't need to submit activities or the portfolio to the lecturer, but these are essential for exam preparation.)

8. Understanding what you read

Take time to note new vocabulary words. Use a dictionary to understand the meaning of new words, or use Google to define a word for you. You could compile a page for each learning unit and add it to your portfolio.

0.3.5 Managing your self-paced study time

You require at least 120 study hours for this module if you are an average student (this time may vary substantially). You should, therefore, plan to use eight study hours per week per module, meaning you should finish one module in 15 weeks. Remember, if you have registered for more than one module, you should plan time for each module accordingly. We advise you to keep a study schedule or diary so that you have a clear idea of the time you have available for study. This will help you to manage your studies within your available time and balance study with work and family life. In Tutorial Letter 101 and on myUnisa you will find a list of due dates for various assignments, which you should keep in your normal diary. Break the large assignments into a series of smaller tasks to complete one step at a time.

To manage your workload, your best working method is to study frequently and regularly in this subject. Establish a routine in an environment with low noise and good lighting. Reward yourself after a productive session.

0.3.6 Academic specialist guidance

If you get into difficulties, please contact the staff in the Department of Life and Consumer Sciences who are responsible for this module. A positive, encouraging attitude towards the course as a whole will keep you motivated to persevere in your studies.

0.4 Plagiarism

Never try to pass off other people's work (or our lecture notes and tutorial matter) as your own. If you wish to use other people's words and ideas or our notes in your own answers, you must use quotation marks and acknowledge your source (use the Harvard method). If you are unsure about the correct way of acknowledging sources, contact Unisa's Library Information Desk. Students who fail to acknowledge quotations or who borrow from lecture notes and outside sources or who copy someone else's answers may be refused permission to write the examination or may be penalised in the assignment.

0.5 Assessment in the module

This course will have formative (ongoing) assessment in the form of assignments and summative (final) assessment in the form of a written examination. In addition to learning a new subject (essentially a new language), it is important for you to reflect on the subject and also on your process of learning.

0.6 In conclusion

After reading this general introduction, you should now have a better understanding of what the module involves and what the aims are for completing your studies in Molecular Biology. Learning unit 1 will introduce molecular biology and will cover the chemistry of some important biological molecules. Learning unit 2 will focus on the structure of the cell. Learning unit 3 deals with the structure of chromosomes. In learning unit 4 you will learn about DNA replication and repair. In learning unit 5 we will discuss the transcription and translation processes, and in learning unit 6 their regulation. Learning unit 7 will be about the cell cycle and finally learning unit 8 will focus on some common molecular biology techniques.

0.7 Getting started

To get to know your online environment and fellow students, we would like you to work through an activity

called an ice breaker.

What is an ice breaker?

An ice breaker helps you to

- understand the technologies that will be used in the course
- get to know and connect with your fellow students

There is one ice breaker activity that you need to do: a blog entry.

Ice breaker: Personal blog entry

Create your own blog entry and share what molecular biology means to you and what made you choose to study this subject.

Go to the Blog tool by clicking on Blogs in the left-hand tool list of this website. You can find the instructions on how to use the blog in the FAQ section in the left-hand tool list of this website.

If you like, you can add links, bullets, lists and colour by using the editing buttons. You can also go back and edit your blogs. The next time we ask you to use the blog, just click on "Add blog entry" again and create a new blog, which will appear under your name.

Learning unit 1:

Introduction to cells and biological molecules

1.1 Introduction to learning unit 1

Molecular biology is the study of biological organisms at the molecular level. It deals primarily with the formation, structure, function and interactions between the molecules essential for life. The molecules within a cell interact to provide energy, create new molecules, replicate DNA and interpret signals from outside the cell. The control of these processes results in growth and life as we know it.

The field of molecular biology overlaps with genetics, biochemistry and other disciplines. It focuses mainly on understanding the interactions between the different macromolecules of the cell, specifically the interactions between DNA, RNA, the manufacture of protein molecules and how these processes are regulated. Molecular biologists essentially isolate, characterise and manipulate these macromolecules to gain further understanding of cellular processes.

To work through the learning unit, refer to chapter 1 and chapter 2 in Alberts. Internet sources that may help you are indicated in the text.

The following website has a glossary with useful information that can be used for the entire course:

<http://seqcore.brcf.med.umich.edu/doc/educ/dnapr/mbglossary/mbgloss.html>

1.2 Learning outcomes

After completing this learning unit, you should be able to

- describe the difference between eukaryotic and prokaryotic cells
- explain what is meant by the central dogma
- discuss the origins of mitochondria and chloroplasts
- describe the three domains of life
- describe the basic structure and composition of common biological macromolecules

1.3 The central dogma

Recommended reading: chapter 1, pages 1–10 in Alberts

All living cells on earth store their hereditary information in DNA, a long, unbranched double-stranded molecule made up of four different monomers or bases. The linear sequence of these four bases encodes all the information a cell needs to reproduce and function. For cells to use the information in the DNA, the DNA needs to be expressed, resulting in the production of other macromolecules within the cell. In the first step of expression the DNA code is “copied” into an intermediary form called RNA during the process of transcription. Many identical RNA transcripts can be produced from a portion of the DNA.

A portion of RNA molecules then act as templates for the production of proteins in a process called translation. The proteins produced have a number of functions within the cell and, as enzymes, are involved in the catalysis of many reactions or they may act as structural elements. Some RNA molecules are not translated and have different roles within the cell. As the DNA determines the RNA sequences and therefore specifies what proteins are produced, the DNA determines the chemistry of the cell.

The sequence of information transfer, from DNA to RNA through the process of transcription and then to protein during translation, is central to molecular biology and is referred to as the central dogma.

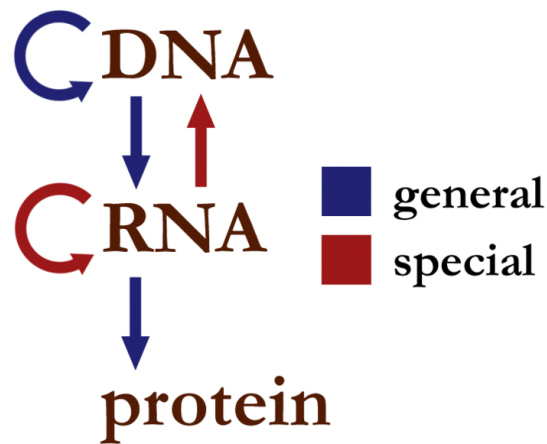


Figure 1.1: The central dogma

To understand these processes, you need to have a basic understanding of the cell structure and the chemistry of the molecules found within the cell. In the rest of this learning unit we will discuss the tree of life and the basic chemistry of a number of important molecules.

1.4 Prokaryotic and eukaryotic cells

Recommended reading: chapter 1, pages 10–22 (entitled “The diversity of genomes and the tree of life”) in Alberts

If you study basic cell structure, you will see that all living organisms can be classified into two groups: prokaryotes or eukaryotes. Note that both the words “prokaryote” and “eukaryote” can be spelt with a “c” or a “k”, for example procaryote or prokaryote – either spelling is correct. Eukaryotic cells are generally bigger and more complex than prokaryotic cells. Many eukaryotic cells form complex multicellular organisms, although many also live as unicellular organisms. Plants, animals and fungi are all eukaryotes.

The genetic material of eukaryotes is found within a membrane-bound structure called a nucleus (refer to figure 1-32 in Alberts, or figure 1.2 in this study guide). Prokaryotes do not have a membrane-bound section to house their DNA. The DNA region is referred to as the nucleoid (refer to figure 1-18 in Alberts or figure 1.3 in this study guide). Most prokaryotes are small and live as independent organisms rather than as multicellular organisms. Prokaryotes can be further divided into two groups based on genetic analysis: the bacteria and archaea resulting in three domains of living organisms (refer to figure 1-21 in Alberts). Bacteria and archaea look similar but at a molecular level the archaea are more closely related to the eukaryotes.

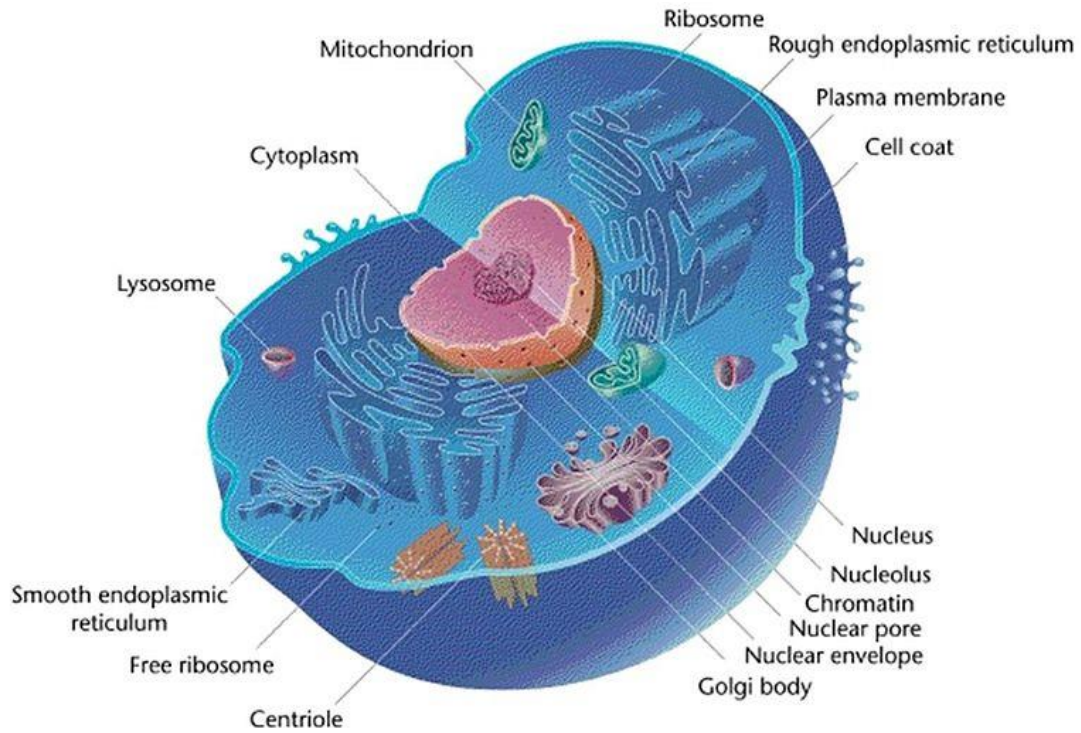


Figure 1.2: Simplified structure of a eukaryotic cell
 (http://commons.wikimedia.org/wiki/File:Eukaryotic_Cell_%28animal%29.jpg)

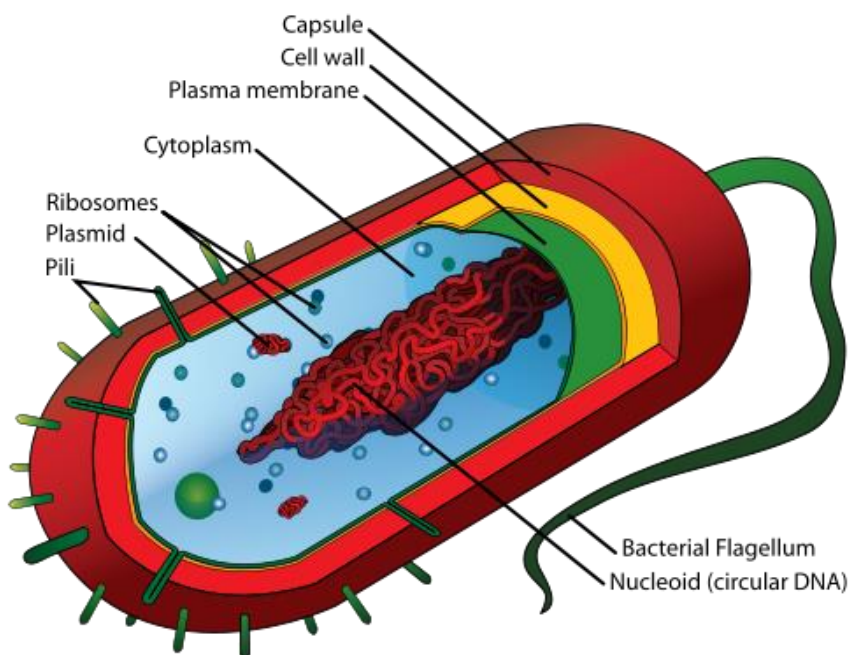


Figure 1.3: General structure of a bacterial cell
 (http://commons.wikimedia.org/wiki/File:Average_prokaryote_cell-_en.svg)

1.4.1 Activity 1.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

The following article contains additional information that may help you when answering the questions:

Koonin, EV. 2010. The two empires and three domains of life in the postgenomic age. *Nature Education*, 3(9):27, <http://www.nature.com/scitable/topicpage/the-two-empires-and-three-domains-of-14432998>.

Using your textbook, answer the following questions:

- a) Describe some characteristics and differences between prokaryotic and eukaryotic cells.
- b) How do the three domains of life differ from one another?
- c) Two of the organelles in eukaryotes, the mitochondrion and the chloroplast are believed to have originated through a process of symbiosis. Discuss some of the theories of the origins of mitochondria and chloroplasts (refer to pages 27-30 in Alberts).

1.5 Model organisms

Prokaryotes are genetically simpler than most eukaryotes and have, therefore, been used extensively as a model organism. *Escherichia coli* (*E. coli*) is the most well-studied prokaryote. It is a small, rod-shaped bacterium that can be found in the gut of many animals and humans. It is easily cultured in the laboratory, reproduces rapidly and is easily modified genetically. Although some strains of *E. coli* can cause serious food poisoning and other diseases in humans, most *E. coli* strains are harmless. The harmless strains are part of the normal gut flora, and benefit their hosts by preventing colonisation of the intestine by other pathogenic bacteria.

Although there are differences between eukaryotic cells and *E. coli*, the basic genetic mechanisms are highly conserved and, therefore, studies on *E. coli* yield information about the processes occurring in all cells. Yeasts are unicellular eukaryotes and are often used as model organisms to study cellular functions specific to eukaryotes. *Saccharomyces cerevisiae* is a yeast that has been used extensively in genetics and cell biology research, and is easily cultured in the laboratory. For thousands of years yeasts have been used by humans to make bread rise and in the production of alcoholic beverages. Some yeast species are pathogenic and can cause disease in humans. The best-studied pathogenic yeast is *Candida albicans*, which causes thrush.

1.6 Biological molecules

When studying the structure and function of macromolecules, it is important to keep in mind the chemistry and chemical composition of these macromolecules. The chemical interactions between molecules and the catalytic facilitation of chemical reactions by enzymes are fundamental to life.

To understand the complex reactions that occur between the molecules of the cell, let's review the basic chemistry of organic molecules.

To help you understand which chemical components are important in biological systems, read pages 45-55 in Alberts on the chemical components of a cell. Make sure you understand what chemical elements are typically found in biological molecules, how atoms and molecules interact to form bonds and the difference between the different types of bonds and their relative strength.

1.6.1 Activity 1.2

Do the following activity and add it to your portfolio:

- a) Describe the difference between a covalent and non-covalent bond.

- b) Write a brief description of the three different types of non-covalent bond:
- electrostatic attraction
 - hydrogen
 - Van der Waals attraction
- c) What are hydrophobic forces?
- d) Why is water a polar molecule?

The panels at the end of the chapter on pages 106-123 are also great as a review of basic biological chemistry concepts.

1.7 Macromolecules of cells

Cells contain four main families of small organic molecules that associate to form large macromolecules:

- monosaccharides - carbohydrates
- fatty acids - lipids
- amino acids - proteins
- nucleotides – nucleic acids

We will now discuss these common molecules of the cell briefly.

1.7.1 Carbohydrates

Recommended reading: chapter 2, pages 55–58 (entitled “Sugars provide an energy source for cells and are the subunits of polysaccharides”) in Alberts

Monosaccharides are the basic units of carbohydrates and glucose is an example. The structures of sugars are highly diverse: they can be linear or ring structures, they can have hydroxyl, aldehyde or ketone groups, they may be branched or linear. The monomeric subunits can interact to form disaccharides or polysaccharides with two or more subunits, respectively. The monosaccharide glucose, when broken down, releases energy and is the main energy source within an animal cell. Polysaccharides of glucose do not only serve as a source of energy – they also form important structural components of cells, for example the cellulose of plant cell walls.

1.7.2 Activity 1.3

Do the following activity and add it to your portfolio:

- a) Describe the basic structure of monosaccharides and how they join to form polysaccharides.
- b) Explain what hydrolysis and condensation reactions are. Describe the role of water in these reactions.
- c) What is a disaccharide?

1.7.3 Lipids

Recommended reading: chapter 2, pages 58–59 (entitled “Fatty acids are components of cell membranes, as well as a source of energy”) in Alberts

Lipids are biological molecules that are insoluble in water but are soluble in fat and organic solvents such as benzene. They encompass molecules such as the fatty acids which contain long hydrocarbon chains or the steroids with multiple linked rings. They may be hydrophobic or amphipathic.

Fatty acids are used to store energy in cells in the form of triacylglycerol molecules that can be broken down when needed. Fatty acids are also the main structural component of cell membranes, the most common fatty acids being phospholipids. Glycolipids are lipids with a carbohydrate moiety attached and often serve as

markers on cell surfaces.

1.7.4 Activity 1.4

Do the following activity and add it to your portfolio:

- Look at the example of the fatty acid, palmitic acid in the textbook, figure 2-21, and describe the structure.
- In your own words, describe the terms “lipophilic”, “hydrophilic” and “hydrophobic”.
- Describe what makes a molecule hydrophilic, hydrophobic or amphipathic.
- Discuss how the amphipathic structure of a phospholipid allows it to spontaneously form lipid bilayers.

1.7.5 Nucleotides

Recommended reading: chapter 2, pages 61-62 (entitled “Nucleotides are the subunits of DNA and RNA”) in Alberts

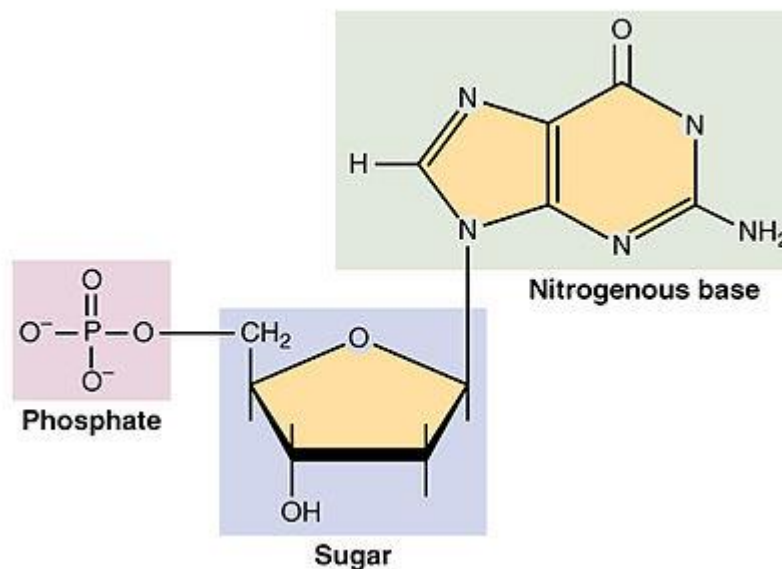


Figure 1.4: Structure of a nucleotide; image from OpenStax College, *Anatomy & physiology* (<http://cnx.org/content/col11496/1.6/>)

Nucleotides are the subunits of DNA and RNA. They are molecules composed of a five-carbon sugar (either a ribose or deoxyribose sugar) linked to a phosphate group and a nitrogenous base (nitrogen containing ring structure). There are a number of different nitrogenous bases and the type of base determines what nucleotide it is. There are five nucleotides found within DNA and RNA: cytosine (C), thymine (T) and uracil (U) which are pyrimidines, and adenine (A) and guanine (G) which are purines. The nucleotides are covalently joined together by phosphodiester bonds to form long chains called nucleic acids (refer to figure 2-28 in Alberts).

The nucleic acid DNA contains nucleotides with a deoxyribose sugar and the bases A, G, C and T. It exists as a double-stranded helix with the bases of the two antiparallel strands connected by hydrogen bonds. The ability of the bases to interact specifically with one another via hydrogen bonding is called base pairing. The sequence of the bases in DNA encodes the genetic information of the cell. RNA contains ribose sugars and the bases A, G, C and U. It typically occurs as a single chain that can fold upon itself to form three-dimensional structures. You will learn more about the structure of DNA in learning unit 3.

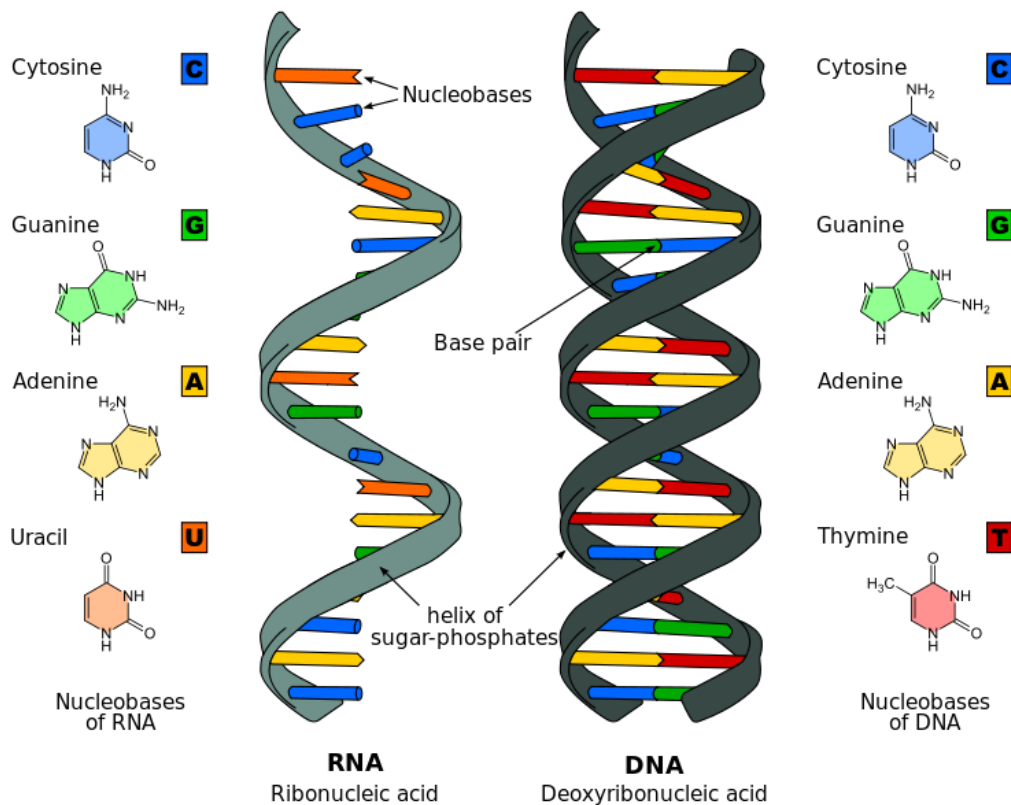


Figure 1.5: Simplified structures of RNA and DNA and the structures of the nitrogenous bases they contain (http://commons.wikimedia.org/wiki/File:Difference_DNA_RNA-EN.svg)

The nucleotide adenosine triphosphate (ATP) acts as a short-term chemical energy storage molecule in the cell. It is formed during the oxidative breakdown of foodstuffs or during photosynthesis. It is composed of three phosphates linked by phosphoanhydride bonds. When these bonds are broken by hydrolysis, energy is released. This energy can be used to drive other energy-requiring reactions.

1.7.6 Activity 1.5

Do the following activity and add it to your portfolio:

- Draw and describe the basic structure of a nucleotide. Make sure you know the difference between a ribose and deoxyribose sugar.
- Describe the general structure of a purine and pyrimidine nitrogenous base.
- Describe how a phosphodiester bond is formed. Does the formation of a phosphodiester bond release or require energy? Take note of the 3' hydroxyl (OH) group and the 5' phosphate group.
- Read the section in Alberts entitled "ATP is the most widely used activated carrier", pages 80-81. Describe how the conversion of ATP to ADP yields energy and how this energy can be used to drive other processes.

1.7.7 Amino acids

Recommended reading: chapter 2, pages 59-60 (entitled "Amino acids are the subunits of proteins") in Alberts

Amino acids are the building blocks of proteins. They all have a carboxylic acid group (COOH), an amino group (NH₂) and a side chain linked to an α-carbon atom.

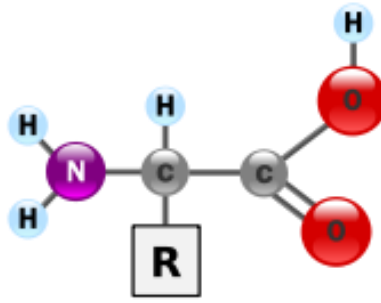


Figure 1.6: Structure of an amino acid (R represents the side chain)
 (<http://en.wikipedia.org/wiki/File:AminoAcidball.svg>)

The bond formed between two amino acids is called a peptide bond and a long chain of amino acids joined together is referred to as a polypeptide or protein. There are 20 common amino acids (see panel 3-1, pages 128-129, in Alberts) each specified by the side chain or R group attached to the α -carbon. Some of the side chains can become charged in neutral aqueous solutions, other are not charged at all. Some are polar and therefore hydrophilic, and others are non-polar and hydrophobic. These differences in the properties of the side chains allow the proteins to fold into the unique structures. Each protein molecule folds into a specific three-dimensional form with certain amino acids positioned at the active site, thereby allowing the protein to catalyse a number of reactions that can control the formation or disruption of covalent bonds.

1.7.8 Activity 1.6

Do the following activity and add it to your portfolio.

Remember, answering the activity questions will help you in preparing for the exam!

- What is a peptide bond?
- Why is a protein said to have an N-terminal and a C-terminal end?
- The amino acids can be grouped according to the properties of their side chains into four groups: acidic, basic, uncharged polar and non-polar. Draw up a table grouping the 20 different amino acids according to their properties. In your table include the full name and the one- and three-letter abbreviation for each amino acid. Refer to panel 3-1, pages 128-129, to help you.

1.7.9 Feedback on activity 1.6

When amino acids are joined together to form a peptide by peptide bonds, one end of the peptide will have a free amine group ($-\text{NH}_2$) and this is referred to as the N-terminus or N-terminal end. The other end of the peptide will have a free carboxyl group ($-\text{COOH}$) and is called the C-terminal end or C-terminus. The convention for writing peptide sequences is to write the sequence from the N-terminus to the C-terminus.

Learning unit 2

The internal organisation of the cell

2.1 Introduction to learning unit 2

The cell is the smallest structural and functional unit of all known living organisms and is typically only visible under a microscope. Cells can replicate autonomously, and are often called the "building blocks of life". The most recent estimate of the number of cells in the human body is about 37 trillion (3.7×10^{13}) cells. Cells grow and divide as a result of cellular metabolism. Metabolism is usually divided into two types: catabolism, in which the cell breaks down organic matter to produce energy, and anabolism in which the cell uses energy to construct complex molecules such as proteins and nucleic acids. These processes take part in the different compartments or organelles of the cell.

This section is to remind you of the structure of the cell and its components, as you should have studied the composition of the cell in previous courses. You learnt about the overall structure of prokaryote and eukaryote cells in learning unit 1. In this learning unit we will focus on the individual components of a eukaryotic cell. The sections in the textbook dealing with cell structure are very detailed, but for this course you need only have a basic knowledge of cell structure. Use the study guide to guide you in what you need to know.

2.2 Learning outcomes

After completing this learning unit, you should be able to

- describe the structure and function of the plasma membrane, including how molecules are transported in and out of the cell
- describe the basic structure and function of the
 - nucleus
 - endoplasmic reticulum
 - mitochondrion
 - Golgi apparatus
 - lysosome
 - peroxisome
 - chloroplast
- describe the role of the cytoskeleton and the main protein filaments that make up the cytoskeleton

2.3 Plasma membranes

Recommended reading: in Alberts, chapter 10, pages 617-622, the section on the lipid bilayer and chapter 11, pages 651-655, the section on principles of membrane transport

The following website also has information on the cell membrane:

<http://cnx.org/content/m46021/latest/?collection=col11496/1.6>

All living cells are enclosed by a plasma membrane that physically separates the inside components of the cell from the outside environment. It is selectively permeable and regulates the movement of substances that enter or leave the cell. It has a fluid nature and is composed mainly of phospholipids. Phospholipids are amphipathic molecules containing both hydrophilic and hydrophobic regions. In aqueous solutions phospholipids will spontaneously come together forming liposomes, micelles or bilayer sheets with the hydrophilic regions on the outside in contact with the aqueous environment and the hydrophobic regions together on the inside (refer to figure 2.1 in the study guide). In this way the membrane is held together via the non-covalent interaction of the hydrophobic tails.

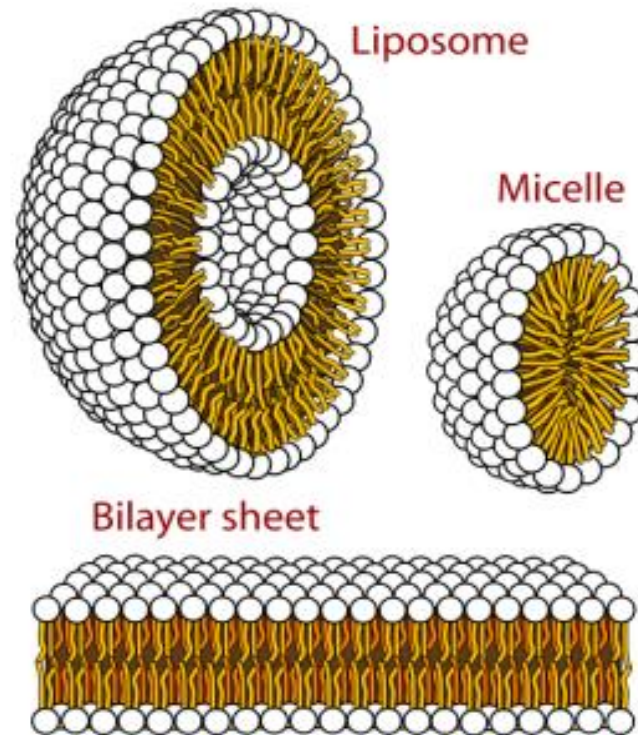


Figure 2.1: The structure's amphipathic molecules spontaneously form in aqueous solutions. (http://commons.wikimedia.org/wiki/File:Phospholipids_aqueous_solution_structures.svg)

Lipid bilayers are said to be fluid as they are not rigidly held together. Sterols are an important component of the membrane and have an impact on the membrane dynamics. Various proteins and glycolipids are attached to or embedded within the phospholipid membrane (refer to figure 2.2 in the study guide). Proteins that are found extending outside of the cell often have attached carbohydrates and are called glycoproteins. They act as recognition molecules. Integral proteins extend through the membrane and act as ion channels or receptors.

A glycolipid you may have heard of is CD₄. This is the receptor to which human immunodeficiency virus (HIV) binds before it enters the cell.

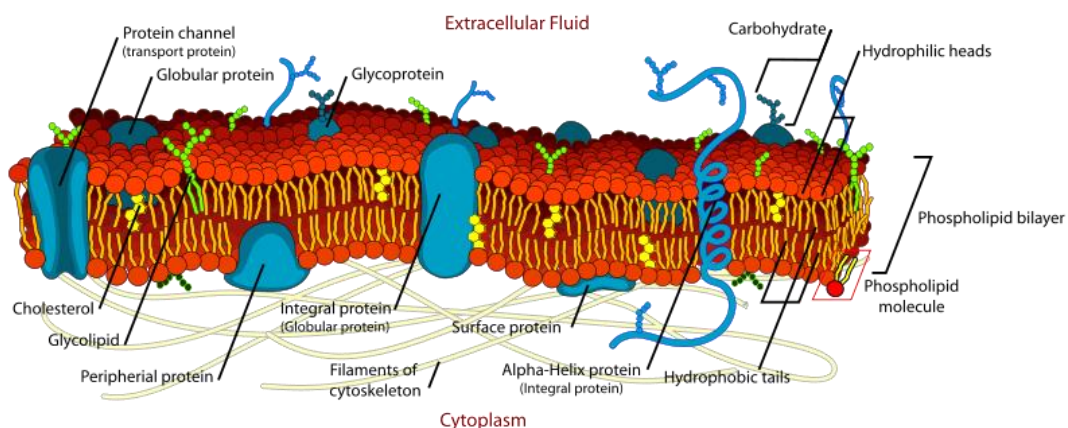


Figure 2.2: The plasma membrane of a cell contains a variety of proteins, phospholipids and other molecules. It also serves as the attachment point for both the intracellular cytoskeleton and, if present, the cell wall. (http://commons.wikimedia.org/wiki/File:Cell_membrane_detailed_diagram_en.svg)

2.3.1 Movement of substances across membranes

The main function of the membrane is to control the movement of substances in and out of the cell and thus influence the cell's internal environment. Ions, sugars, fatty acids, amino acids and CO₂ are some of the substances that are transported across the membrane. The movement of small molecules across the membrane may be by passive transport, where the concentration gradient determines the direction of movement of the molecule, or active transport which is against the concentration gradient and requires energy. There are two main classes of transport proteins: transporters and channels. Movement across the membrane mediated by channel proteins is always passive. Transporter proteins, on the other hand, can passively or actively transport molecules.

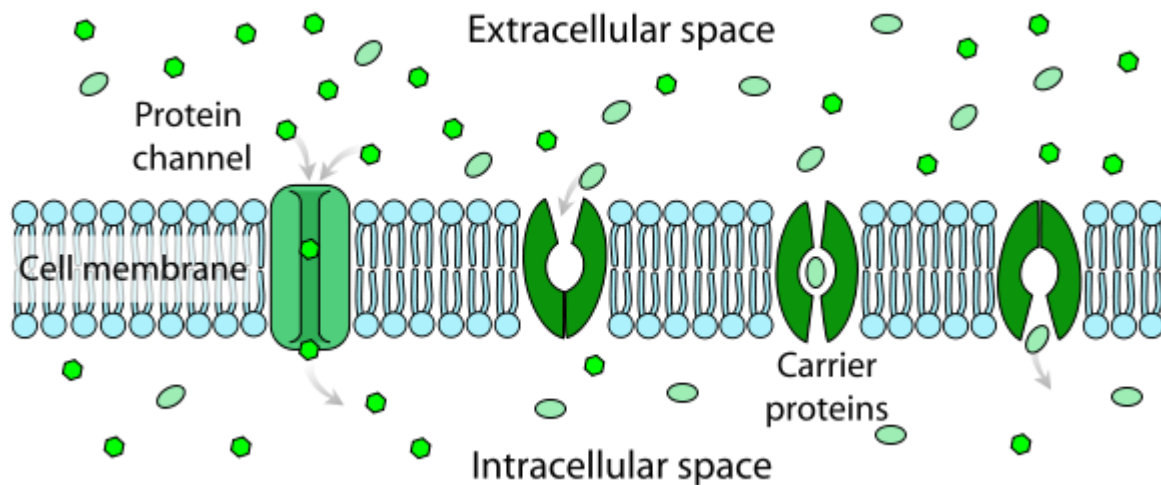


Figure 2.3: Possible methods of transporting molecules across the plasma membrane (http://commons.wikimedia.org/wiki/File:Scheme_facilitated_diffusion_in_cell_membrane-en.svg)

Messages can be transmitted across the membrane by lipid-soluble molecules that can enter or leave the cell, or by molecules that bind the outside of the cell to a specific receptor which then triggers an internal cascade.

2.3.2 Activity 2.1

Do the following activity and add it to your portfolio:

Refer to Alberts, pages 651-655, and answer the following questions:

- Discuss the permeability of the cell membrane to different sizes and classes of molecules.
- Describe the two main classes of transport proteins.
- What is the difference between active transport and passive transport?
- The inherited disorder cystinuria is as a result of a problem with transport proteins. Briefly describe the disease.

2.3.3 Feedback on activity 2.1

When describing the permeability of the cell membrane to different sizes and classes of molecule, you should have mentioned that smaller molecules that are soluble in oil (hydrophobic) will diffuse across a lipid membrane faster than larger hydrophilic molecules and that many molecules, for example ions (charged), will not diffuse across the membrane at all. Small uncharged polar molecules can also diffuse across the lipid membrane but at a lower rate than small hydrophobic molecules.

Transporters and channels are the two main classes of transport proteins and are described on page 652 in the textbook.

2.4 Intracellular organelles

In eukaryotes the cell is divided into membrane-enclosed compartments or organelles that are functionally distinct with their own complement of enzymes and other molecules. The organelles are found within the cytoplasm and provide additional membrane surface area for biochemical reactions to take place on, in addition to forming an enclosed compartment with a different chemical composition to the cytosol. Each organelle membrane has different transport proteins that import and export the metabolites and proteins that are required by that organelle.

The main organelles of the cell include the

- nucleus
- endoplasmic reticulum
- mitochondria
- Golgi apparatus
- lysosomes
- peroxisomes
- endosomes
- chloroplasts (only in photosynthetic organisms)

Let's examine some of these in more detail.

2.4.1 Nucleus

Recommended reading: in Alberts, chapter 4, pages 200-201 (the section entitled "In eukaryotes, DNA is enclosed in a cell nucleus")

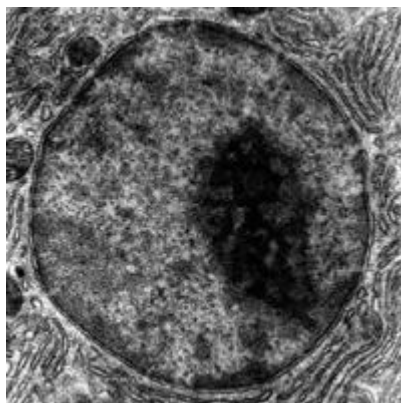


Figure 2.4: Electron micrograph of a nucleus
(<http://commons.wikimedia.org/wiki/File:Nucleus%26Nucleolus.gif>)

The nucleus is a large organelle found in most eukaryotic cells which contains the genome (aside from the DNA in the mitochondria and chloroplasts) and is the main site of DNA and RNA synthesis. It is enclosed by a double membrane (nuclear envelope) that keeps the nuclear and cytosolic enzymes separate (see figure 2.5 in this study guide). The nuclear envelope is directly connected to the endoplasmic reticulum and is supported by a network of protein fibres called the nuclear lamina. The membrane is covered in nuclear pores which control the movement of molecules in and out of the nucleus. Within the nucleus is a dense region called the nucleolus where ribosomal RNA and ribosomes are synthesised.

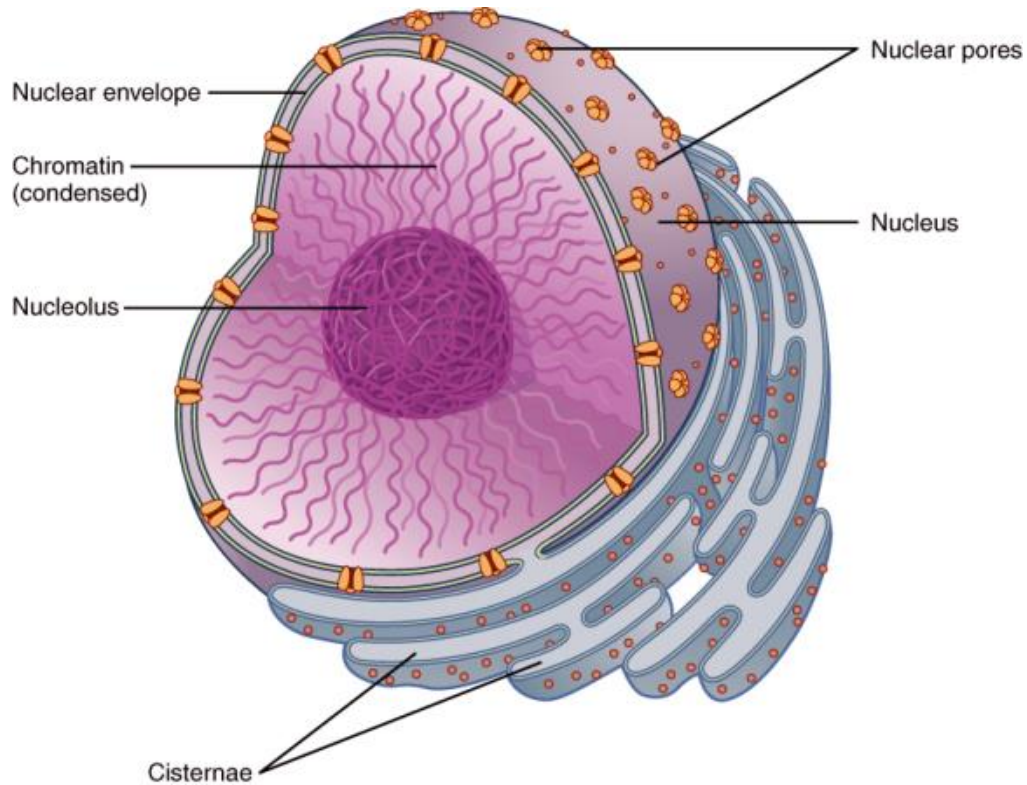


Figure 2.5: Schematic diagram of a nucleus; image from OpenStax College, *Anatomy & physiology* (<http://cnx.org/content/m46073/latest/?collection=col11496/1.6>)

2.4.2 Mitochondria

Recommended reading: chapter 14, pages 815-822 in Alberts, the section on the mitochondrion

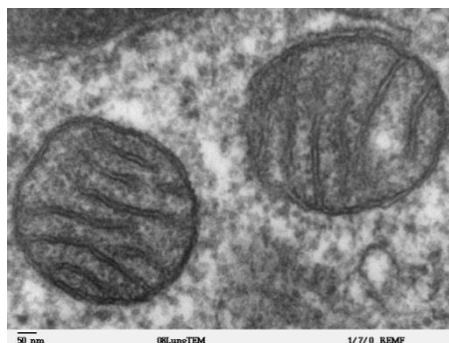


Figure 2.6: Transmission electron micrograph of two mitochondria (http://en.wikipedia.org/wiki/File:Mitochondria,_mammalian_lung_-_TEM.jpg)

Mitochondria are sausage-shaped organelles with diameters of 0.5 – 1 μm found in all eukaryotic cells and are large enough to be seen in a light microscope. They generate most of the STP that is used by the rest of

the cell to drive reactions requiring free energy.

They have an outer membrane and an inner membrane that is folded into many contiguous layers called cristae. The cristae separate the mitochondrion into two compartments: the internal matrix and an intermembrane space, each containing unique constituents (refer to figure 2.7 in this guide of the mitochondrion). Figure 14-8 in Alberts also illustrates these different areas.

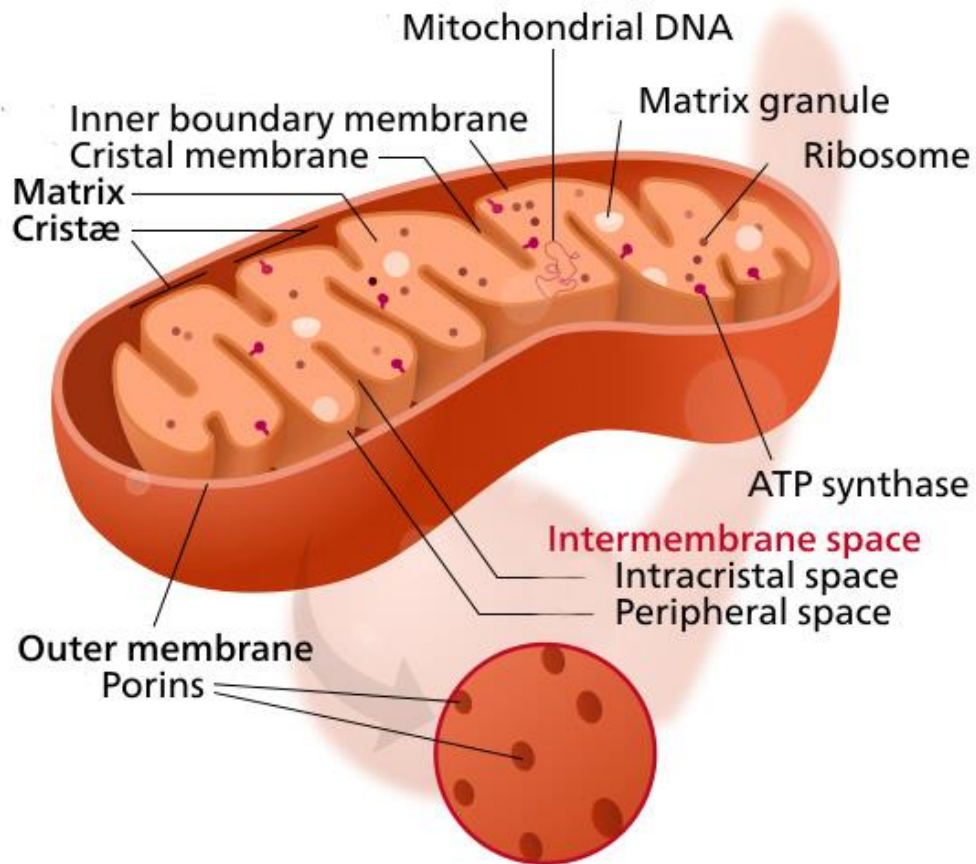


Figure 2.7: A mitochondrion (http://en.wikipedia.org/wiki/File:Mitochondrion_mini.svg)

Mitochondria have their own DNA, encoding for a number of essential proteins, although the bulk of the proteins required are encoded for by the nuclear DNA and are imported into the mitochondrion through pores in the outer membrane.

2.4.3 Activity 2.2

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- Draw a diagram of a mitochondrion showing the cristae, outer membrane, inner membrane and the location of the internal matrix and intermembrane space.
- The mitochondria utilise pyruvate and fatty acids to generate energy. Refer to figure 4-10 in Alberts and the associated text and describe how the mitochondria generate energy using these two molecules. Make sure you understand how the proton gradient is used to generate ATP.

2.4.4 Chloroplast and plastids

Recommended reading: chapter 14, pages 840-844 in Alberts, the section on chloroplasts and photosynthesis

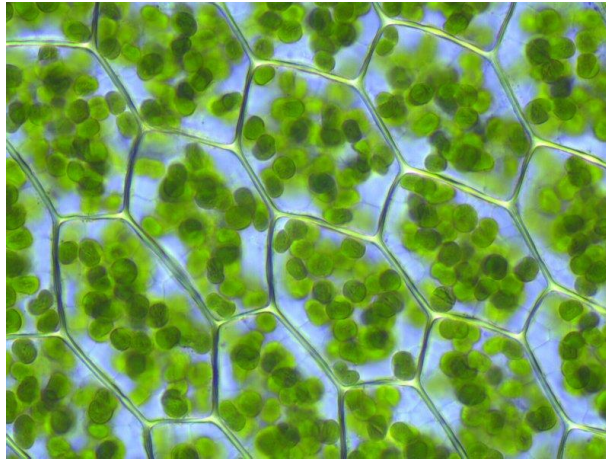


Figure 2.8: Chloroplasts in plant cells (http://en.wikipedia.org/wiki/File:Plagiomnium_affine_laminazellen.jpeg)

Photosynthetic organisms use CO_2 and the energy from sunlight to produce organic materials in a process called carbon fixation. Plants and algae have specialised structures called chloroplasts for photosynthesis. Chloroplasts are larger than mitochondria and have a more complex internal organisation. They are surrounded by two closely associated membranes (refer to figure 2.9 in this guide). The inner membrane encloses the space called the stroma. Within the stroma are stacked membranes called grana. The grana are made up of stacks of disc-shaped structures called thylakoids which contain the photosynthetic, light-capturing pigments, electron transport chains and ATP synthase. The lumens of the thylakoids are all interconnected to form the thylakoid space.

Chloroplasts also contain DNA which encodes for a small portion of the chloroplast components. The rest are encoded for by genes within the nucleus.

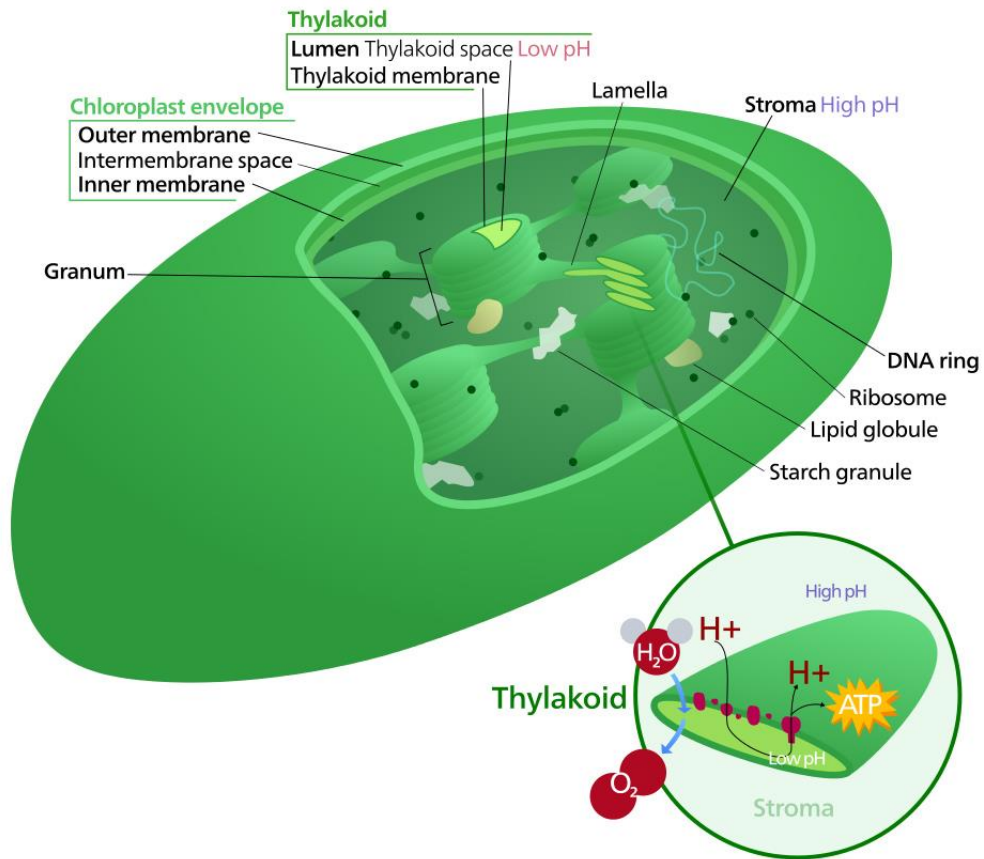


Figure 2.9: A chloroplast (http://commons.wikimedia.org/wiki/File:Chloroplast_%28standalone_version%29-en.svg)

Chloroplasts belong to the plastid family of organelles. Plastids are also involved in the storage of molecules. An amyloplast is a common form of leucoplast that is found within plant cells and stores starch. Figure 2.10 shows amyloplasts in a potato cell.

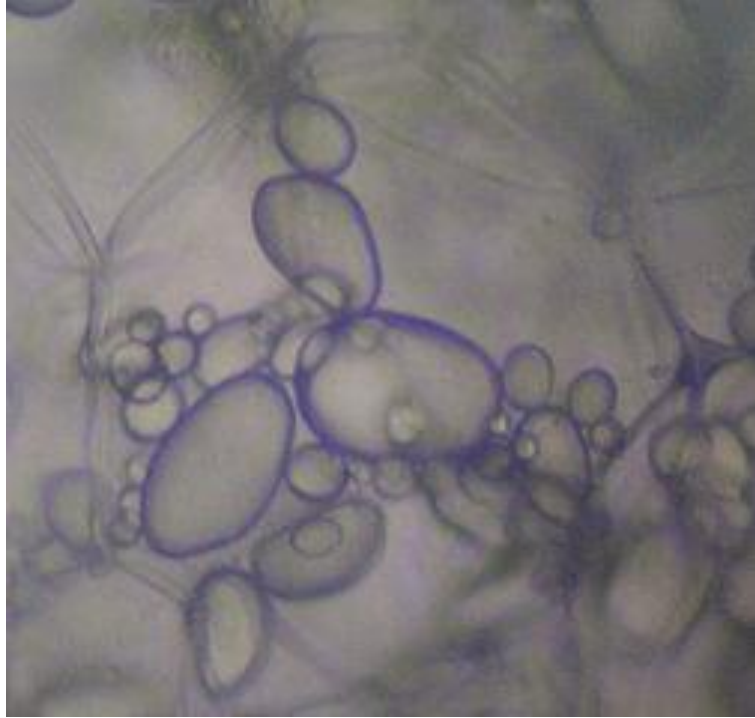


Figure 2.10: Amyloplasts in potato cells ([http://commons.wikimedia.org/wiki/File:Potato - Amyloplasts.jpg](http://commons.wikimedia.org/wiki/File:Potato_-_Amyloplasts.jpg))

2.4.5 Activity 2.3

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- a) Chloroplasts capture energy from sunlight and use it to fix carbon. What is meant by this statement? Briefly discuss photosynthetic electron transfer and the carbon fixation reaction.
- b) Draw up a table stating some of the similarities and differences between mitochondria and chloroplasts.

2.4.6 Endoplasmic reticulum and Golgi apparatus

Recommended reading: chapter 12, pages 723-726 in Alberts, the section on the endoplasmic reticulum

The following article also contains further information on the endomembrane system:

<http://www.nature.com/scitable/topicpage/endoplasmic-reticulum-golgi-apparatus-and-lysosomes-14053361>

The following web page has more information on the endoplasmic reticulum:

<http://micro.magnet.fsu.edu/cells/endoplasmicreticulum/endoplasmicreticulum.html>

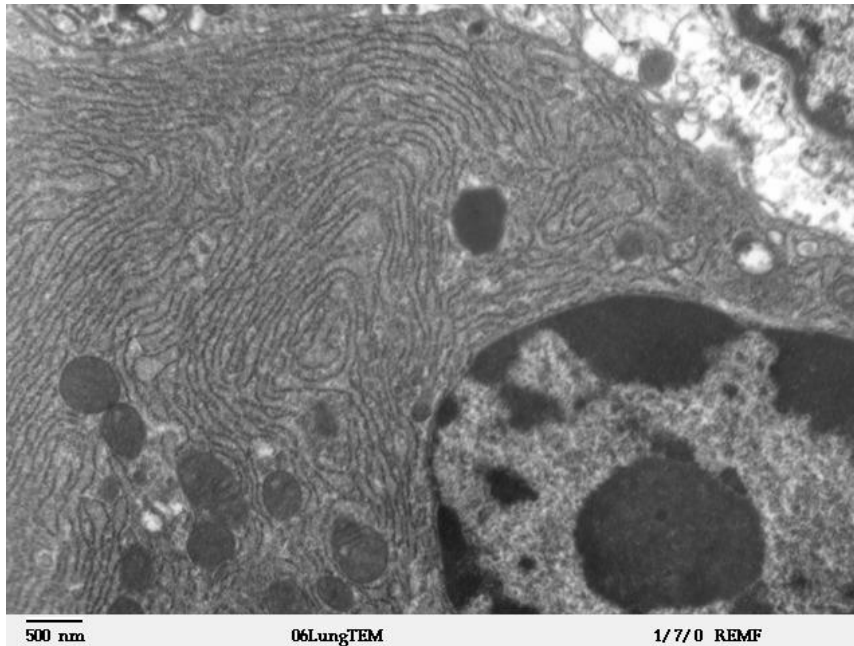


Figure 2.11: Micrograph of rough endoplasmic reticulum network around the nucleus (shown in lower right-hand side of the picture). The dark small circles in the lower left-hand side are mitochondria. (http://en.wikipedia.org/wiki/File:Clara_cell_lung_-_TEM.jpg)

The endoplasmic reticulum (ER) is a winding network of membrane that is connected to and continuous with the nuclear membrane. The space enclosed by the ER is called the ER lumen. The ER is involved in the synthesis, transport and storage of protein and lipid molecules. The part of the ER involved in protein production is called the rough ER as it is studded with ribosomes (visible as dots). As proteins are synthesised by the ribosomes, they pass through the ER membrane into the lumen where they fold and are glycosylated. Most of the protein produced in the ER are secreted out the cell or are transported to other organelles. The smooth ER does not have ribosomes attached to the surface but has enzymes involved in lipid biosynthesis.

The Golgi is composed of stacks of membrane-bound structures known as cisternae. The ER, Golgi apparatus and vesicles form part of the endomembrane system of the cell. The ER transports many of its synthesised proteins to the Golgi apparatus where they are sorted, packaged, processed and modified. Once finished, it sends the proteins to different locations around the cell.

2.4.7 Activity 2.4

Do the following activity and add it to your portfolio:

Proteins can be translocated co-translationally or post-translationally. What does this mean and where do these different types of translocation happen? Refer to figure 12-35 and the associated text in Alberts to help you answer this question. If you are having trouble answering this question, come back to it after you have completed learning unit 5.

2.4.8 Lysosomes and peroxisomes

Lysosomes contain hydrolytic enzymes that are involved in intracellular digestion. The enzymes are acid hydrolases which need an acidic environment to function optimally and therefore the lysozymes maintain a pH of 5 within the organelle. Lysosomes carry out intracellular digestion of phagocytised particles which are contained within a vacuole. The lysosome fuses with and releases its enzymes into a vacuole and uses its enzymes to break down the ingested material. The resulting molecules are released into the cytosol where they can be reused. Worn out cell parts are also recycled by the lysozymes in a process called autophagy. Peroxisomes contain oxidative enzymes.

2.4.9 Activity 2.5

Do the following activity and add it to your portfolio:

- a) The cytosol has a neutral pH. If some of the acid hydrolases of the lysosome leaked into the cytosol, would they digest the cellular material? Refer to Alberts, pages 779-780, the section entitled "Lysosomes are the principal sites of intracellular digestion".
- b) What is the function of the peroxisome?

2.4.10 Feedback on activity 2.5

To find the function of peroxisomes, look in the index of the textbook under peroxisomes. There is a reference to the pages dealing with functions. Go to these pages and read up on peroxisome function.

2.5 The cytoskeleton

Recommended reading: chapter 16, pages 965-969 in Alberts, the section on the cytoskeleton

The cytoskeleton gives the cell structure and shape and plays a vital role in intracellular transport (the movement of organelles) and cell division. Cells need to be able to reorganise their components as they grow and divide. Some cells need to be able to move and the cytoskeleton is responsible for enabling this movement.

There are three main types of protein filaments that form part of the eukaryotic cytoskeleton:

- microfilaments (actin filaments)
- microtubules
- intermediate filaments

2.5.1 Activity 2.6

Do the following activity and add it to your portfolio:

- a) Describe some of the functions of the cytoskeleton. (Refer to the recommended reading at the beginning of the section on the cytoskeleton.)
- b) Refer to panel 16-1, page 968 in Alberts, and write a brief description of each of the main types of filaments that can be found in the cytoskeleton.

Learning unit 3

DNA, chromosomes and genes

3.1 Introduction to learning unit 3

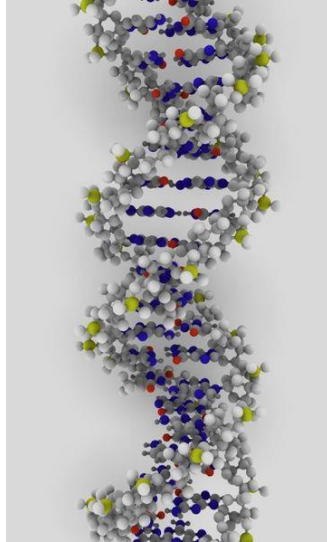


Figure 3.1: DNA molecule (<http://commons.wikimedia.org/wiki/File:DNA-molecule3.jpg>)

All living organisms grow and develop according to the information in their genetic material. The genetic material is DNA and is organised into pieces called chromosomes. The chromosomes can be divided into short segments of DNA called genes which encode for all proteins and functional RNA chains. Genes contain the information needed to construct and sustain an organism's cells and determine the characteristics of the organism. All organisms have a mixture of biological traits; some are clearly visible, such as number of fingers or hair colour, and some are not, such as blood type or increased risk for specific diseases. These traits are encoded for by the genes and are passed on from one generation to the next through the hereditary molecule, DNA.

To work through the learning unit, refer to chapter 4 in Alberts. Additional internet sources indicated in the text may also help you.

3.2 Learning outcomes

After completing this learning unit, you should be able to

- describe the structure of DNA and how the structure provides a mechanism for heredity
- describe how the structure of DNA was determined
- describe the global structure of chromosomes in eukaryotes
- explain how genomes evolve

3.3 Structure of DNA

Recommended reading: chapter 4, pages 195–202 in Alberts, on the structure and function of DNA

The following website also has further information on the structure of DNA:
http://www.blc.arizona.edu/molecular_graphics/dna_structure/dna_tutorial.html

The structure of the DNA molecule was first described in the 1950s by Watson and Crick. A DNA molecule consists of two complementary chains of nucleotides that are held together by hydrogen bonds and wound into a helical form (refer to figure 4-3 in Alberts). The sugar phosphate backbone is found on the outside of

the helix with the bases pointing inwards. There are four different types of nucleotide units in DNA differing only in their nitrogenous bases: adenine (A), guanine (G), cytosine (C) and thymine (T).

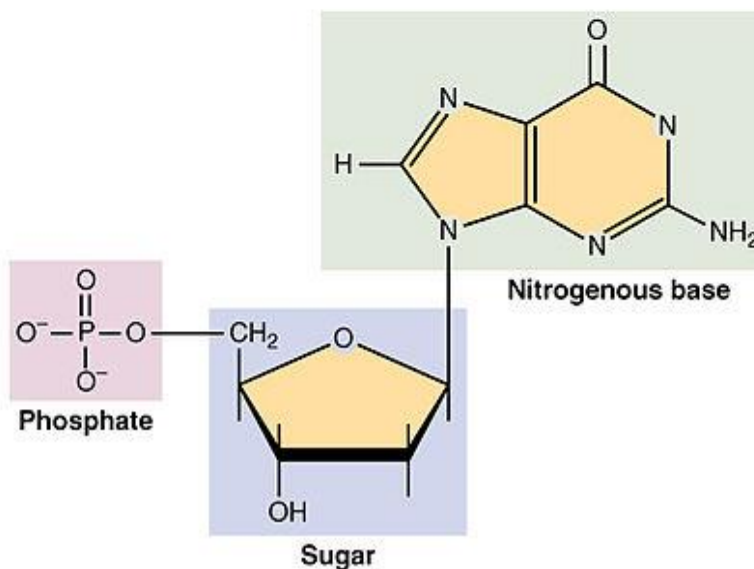
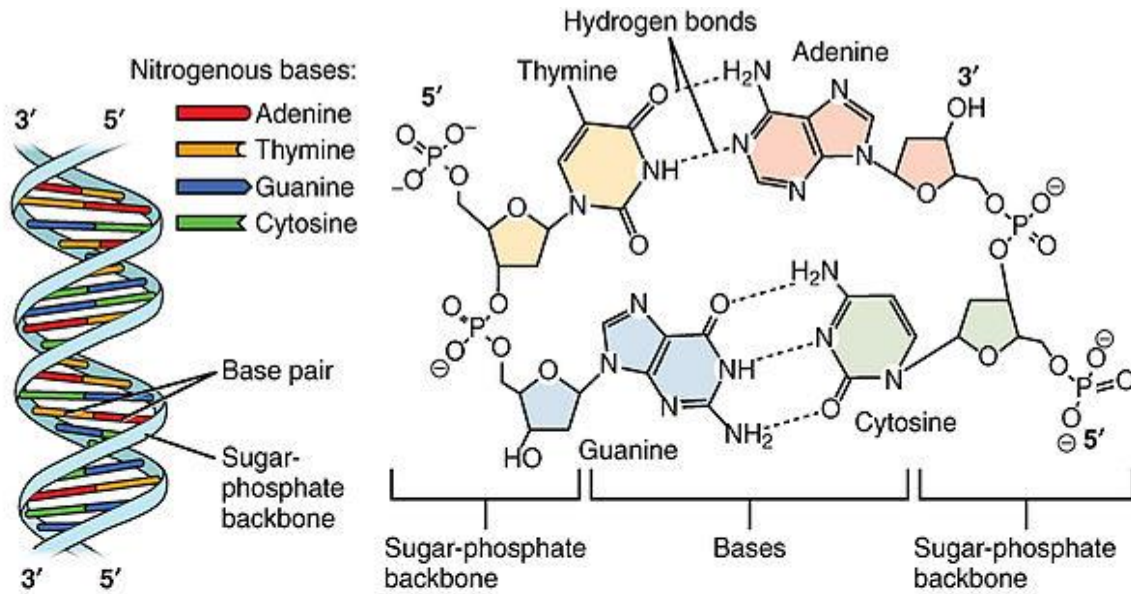


Figure 3.2: The structure of DNA; image from OpenStax College, *Anatomy & physiology* (<http://cnx.org/content/col11496/1.6/>)

Guanine forms three hydrogen bonds with cytosine and adenine forms two hydrogen bonds with thymine on the opposite strand; this is referred to as complementary base pairing. This means that each strand of DNA is joined to another strand that is exactly complementary to itself. If the sequence of one strand is known, then the sequence of the other strand can be determined. Therefore, a double strand of DNA that contains 20% G residues will have 20% C residues and 30% A and 30% T residues, giving you a total of 100%.



Figure 3.3: Section of double-stranded DNA showing the two strands with complementary bases (http://commons.wikimedia.org/wiki/File:Image_DNA_sequence_-_png.png)

Each base pair is a joining of a purine with a pyrimidine and is the same size, giving the DNA strand a uniform diameter. The way in which the nucleotide subunits are joined together gives the DNA strand a chemical polarity, with a 5' phosphate end and a 3' hydroxyl end. The two polynucleotide chains run in opposite directions and the DNA strands are said to be antiparallel. When writing a single strand of DNA sequences, we always write it in the 5' to 3' direction.

3.3.1 Activity 3.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- Read up on the major structural elements of DNA. Make sure you can identify the deoxyribose sugar, the four different types of nitrogenous base, the phosphate groups and the 3' and 5' ends.
- What are the four nucleotide bases found in DNA? Which are purines and which are pyrimidines?
- If a section of DNA contains 22% adenine residues, what is the percentage of guanine, cytosine and thymine residues?
- Read the article: Pray, L. 2008. Discovery of DNA structure and function: Watson and Crick. *Nature Education*, 1(1):100 (<http://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397>) or other additional material from the library and discuss how the structure of DNA was determined.

3.3.2 Feedback on activity 3.1

Remember when calculating the percentage of each nucleotide base in a DNA molecule that the percentage of adenine residues will be the same as the percentage of thymine residues, and the percentage of guanine residues will be the same as the percentage of cytosine residues. Obviously the total percentage of all four residues should be 100%.

3.4 The genome

The complete DNA sequence of an organism is referred to as its genome. The genome can be subdivided into regions called genes, each specifying the information for a particular protein or functional RNA. Genomes contain large quantities of information – the amount depends on the organism. The human genome is approximately 3.2×10^9 nucleotides in length and encodes for approximately 24 000 distinct proteins. *E. coli* is about 4.6×10^6 nucleotides in length and encodes for roughly 4 500 different proteins, although the number of genes can vary greatly between subtypes. During cell division the cell needs to make a copy of the DNA, because each strand is complementary to the other, and each can act as a template for a new strand. This form of duplication is referred to as semiconservative replication.



Figure 3.4: Semiconservative replication of DNA

3.4.1 Activity 3.2

Do the following activity and add it to your portfolio:

Describe how the complementarity of DNA provides the mechanism for heredity and how DNA acts as a template for its own duplication.

3.4.2 Feedback on activity 3.2

DNA contains two complementary chains of nucleotides that are held together by hydrogen bonds. During replication the two complementary DNA strands are separated and new complementary DNA strands are then synthesised using the original strands as templates. Because the strands are complementary, the two new strands that are produced are the same as the original DNA strand.

3.5 Structure of chromosomes

Recommended reading: chapter 4, pages 202–219 in Alberts, the section on chromosomal DNA and its packaging in the chromatin fibre

Eukaryotic DNA is tightly coiled around various proteins called histones and is packaged into structures called chromosomes. The complex of DNA and proteins is referred to as chromatin. The chromosomes of a cell are not generally visible under the microscope unless the cell is dividing, as the DNA becomes more tightly packed during division, allowing it to be seen. When cells are not dividing, the chromatin exists as long thin entangled threads. Each chromosome has a constriction area called a centromere that divides the chromosome into two arms. The short arm is called the p arm and the long arm the q arm. The centromere ensures that one copy of each duplicated chromosome will be moved into each daughter cell during cell division. The telomeres at the ends of the chromosome are responsible for ensuring that the ends of the chromosome are replicated.

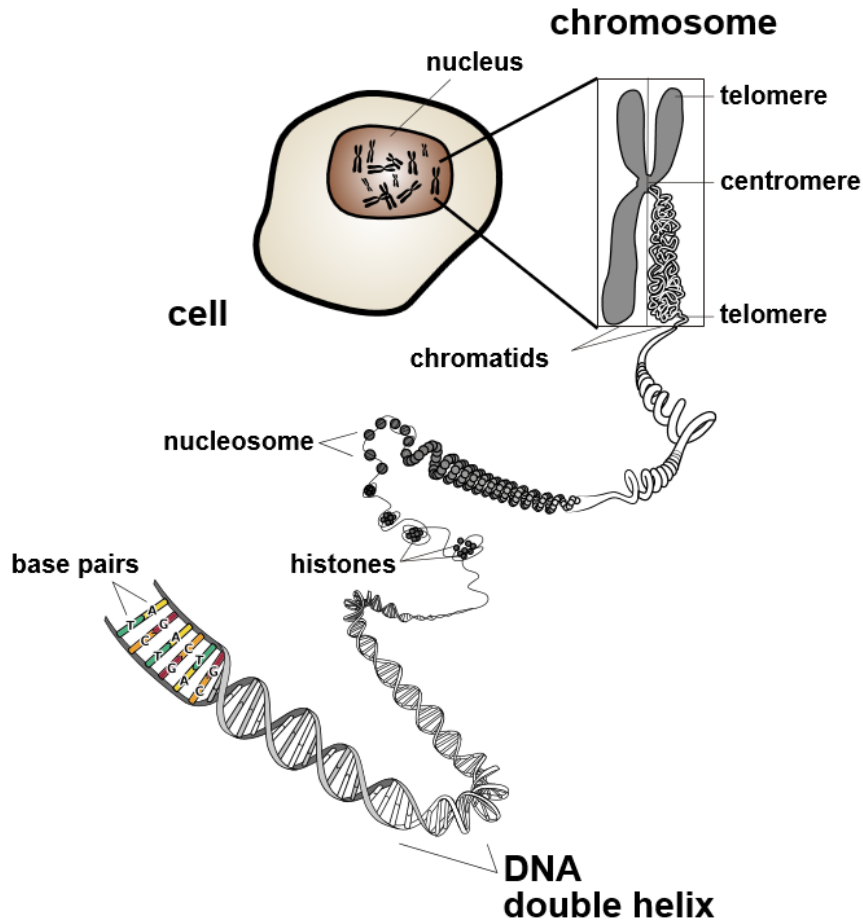


Figure 3.5: Structure of a chromosome (http://commons.wikimedia.org/wiki/File:Chromosome_fr.svg)

How the DNA is divided into chromosomes varies from one species to the other and there is no straightforward relationship between the number of chromosomes, the total genome size and the species complexity. Humans have 46 chromosomes, 22 pairs of autosomal chromosomes and a pair of sex chromosomes. For the pairs of chromosomes, one is of paternal origin and the other of maternal origin. Each pair of chromosomes is referred to as homologs, as they carry the same genes in the same locations or loci as one another. The sex chromosomes in males are non-homologous as the X and Y chromosomes are very different from each other. The appearance of the chromosomes during mitosis (cell division) is called the karyotype. Figure 3.6 here is a picture of a normal human male karyotype. Can you identify the X and Y chromosomes? They are the pair containing different-sized chromosomes.



Figure 3.6: Karyotype of a human male
(http://en.wikipedia.org/wiki/File:NHGRI_human_male_karyotype.png)

Abnormalities in chromosome structure that result in missing, extra, or irregular portions of chromosomal DNA can cause disease. There are numerous different types of chromosome anomalies. One type is called aneuploidy. This is when a person has an abnormal number of chromosomes, and will either be missing a chromosome from a pair or have more than two chromosomes of a pair. Down syndrome is an example of a disease that is caused by having an extra copy of chromosome 21, also called Trisomy 21. An example of a disease that occurs when a portion of a chromosome is missing is Wolf-Hirschhorn syndrome. This is caused by partial deletion of the short arm of chromosome 4.

3.5.1 Packaging of DNA

The chromosome structure is dynamic and changes, depending on the needs of the cell. During mitosis the DNA is tightly packaged; at other times it is less so, with regions decondensing during DNA repair, replication and gene expression. Nucleosomes are the basic unit of DNA packing and consist of a section of DNA wound around eight histone subunits (refer to figure 3.7 in this guide). The formation of nucleosomes reduces the length of the DNA by about a third. The string of nucleosomes is then wound more, condensing the DNA further, forming a fibre with a diameter of about 30 nm. When cells are not undergoing division, most of the chromatin is in this form.

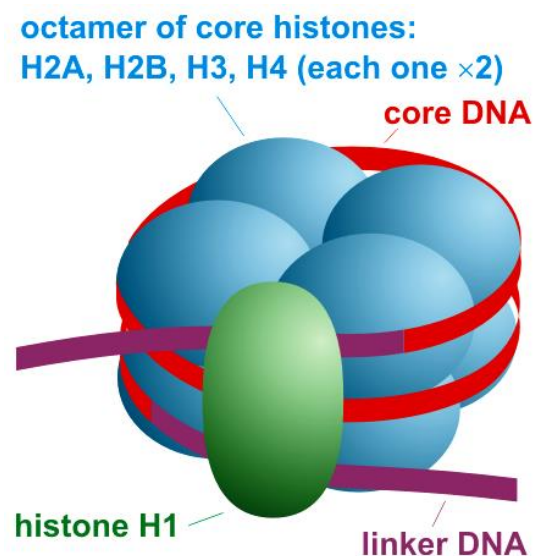


Figure 3.7: Nucleosome structure (http://en.wikipedia.org/wiki/File:Nucleosome_organization.png)

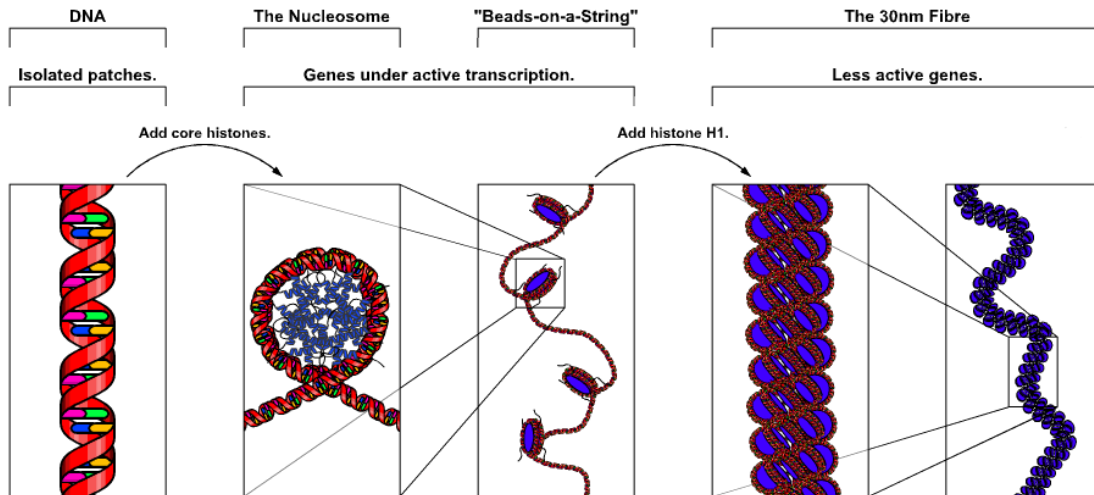


Figure 3.8: DNA packing (http://en.wikipedia.org/wiki/File:Chromatin_Structures.png)

3.5.2 Activity 3.3

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- Draw a labelled diagram of a human chromosome as it appears during cell division. Label the p and q arms, the centromere and telomeres.
- What is a nucleosome?
- Refer to Alberts, pages 211-219, and describe how DNA is packaged to form nucleosomes and chromatin fibres.

3.5.3 Feedback on activity 3.3

When discussing how DNA is packaged to form nucleosomes and chromatin fibres, did you mention the following terms?

- histones
- non-histone chromosomal proteins
- chromatin
- nucleosome core particle
- histone octamer
- linker DNA
- hydrogen bonds
- chromatin fibre

3.6 Genes

Genes are sections of DNA that encode for proteins or functional RNA molecules and have regulatory regions in addition to transcribed regions. The regulatory regions are generally upstream of a gene and are responsible for controlling when the gene is switched on or off and the level of expression. The regulatory regions control which genes are expressed in the different cell types, as not all cells produce the same complement of molecules. For example, insulin is produced in pancreatic cells and not brain cells.

Only a small percentage of the human genome encodes for proteins or functional RNAs. The rest is made up

of “junk DNA” (regions of DNA that serve no obvious function), although some of these regions are important for the correct expression of certain genes. A large portion of the “junk DNA” is made up of short mobile pieces of DNA called transposable elements.

Bacterial chromosomes are generally a single DNA molecule that may be circular or linear and is associated with proteins that help condense the DNA. These proteins are very different from those associated with eukaryotic DNA. The chromosomes of prokaryotes are relatively gene dense and do not contain much “junk DNA”.

3.7 Genome mapping

The complete genome sequence of many different organisms, including humans, has been determined. This gives us the location of every sequence that could potentially be a gene (looks like a gene) within the genome. The problem is that the function of many of these genes/potential genes is unknown. Linkage between genes can help researchers identify the location of the gene responsible for a certain phenotype or disease. Therefore, chromosome and linkage mapping can be used to assign function to these genes.

The genes responsible for certain genetic disorders like cystic fibrosis and muscular dystrophy have been mapped to specific locations on specific chromosomes. These diseases are monogenic, meaning a single defective gene is responsible for causing the disease. They also follow classical inheritance patterns. Many diseases are polygenic and involve many different genes. There is a lot of research on polygenic diseases where inheritance patterns are complex and it is difficult to map which genes are responsible for the disease. Once the diseases have been mapped to specific genes, this information can be used to improve the treatment and diagnosis of these diseases.

The following web page contains more information on gene mapping:

<http://www.nature.com/scitable/ebooks/gene-mapping-then-and-now-16553637/contents>

3.8 How genomes evolve

Recommended reading: chapter 4, pages 245–257 in Alberts, the section on how genomes evolve

Genome evolution is the change in the sequence of a genome over time, resulting in increased genetic diversity. Generally evolution of the genome is as a result of mistakes that occur during genome replication repair or recombination. The changes in the genome may be small as in single base substitutions, or may be large owing to genome rearrangements, deletions, duplications, inversions and translocations of DNA between chromosomes. The movement of transposable elements within genomes also plays an important role in genetic changes. Transposable elements of DNA (transposons) are regions of DNA that can move or spread within the genome. They often disrupt the function or alter the regulation of existing genes.

The genomes of many different organisms have been sequenced and a comparison between the nucleotide sequences of these genomes has influenced how we understand genome evolution. By comparing the genomes of closely related organisms and distant ancestors, the differences and similarities can be used to understand the mechanisms by which genomes evolve over time.

Sequences that serve an important function are relatively similar in diverse life forms. These regions are referred to as conserved regions. Sequences that are not critical often undergo many changes, resulting in little similarity between the sequences in distantly related organisms.

A phylogenetic tree is used to depict the relationship between different organisms and can be generated using different criteria. In figure 3.9, the genome sequences of a number of different organisms were compared and the relatedness between the organisms is represented as a circular tree.

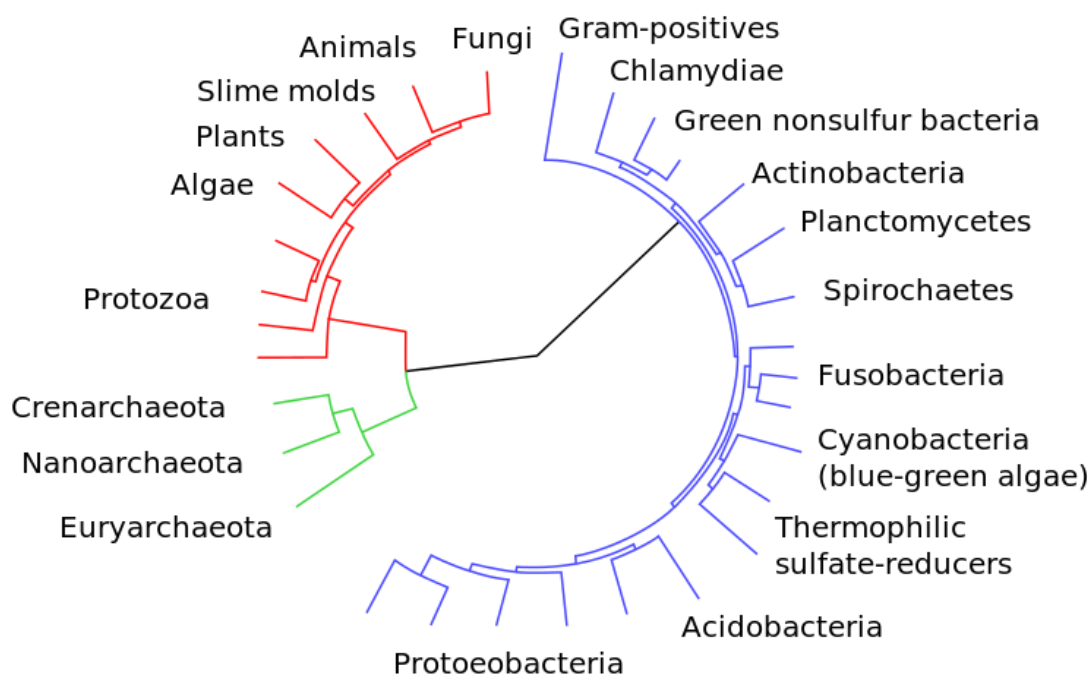


Figure 3.9: A tree of life, based on completely sequenced genomes. Eukaryotes are coloured red, archaea green and bacteria blue. (<http://en.wikipedia.org/wiki/File:CollapsedtreeLabels-simplified.svg>)

Human and chimpanzee genomes are more alike than human and mouse genomes. DNA sequence comparisons show that there has been extensive sequence divergence between the genomes of mice and humans in non-essential regions. Protein-coding sequences and regulatory sequences in the DNA that are essential, and, therefore, cannot be changed easily, are amazingly conserved.

3.8.1 Gene duplication and divergence

Generally there is an increase in the number of genes with increasing biological complexity. These additional genes are thought to have originated as a result of the expansion of families of closely related genes, owing to gene duplication and divergence. It is probable that all present-day genes are descendants of a few ancestral genes that existed in early life forms.

Many organisms lose genes when they are no longer required. This is often seen in organisms that have adapted to a parasitic lifestyle where they can obtain nutrients from the host. Genome reduction and gene loss can be seen by the presence of pseudogenes. These are defective genes that are no longer able to produce functional proteins derived from previously functional gene relatives. An example of pseudogenes found within the human genome includes some of the olfactory gene families, which explains why humans have such a poor sense of smell in comparison to other mammals.

3.8.2 Activity 3.4

Do the following activity and add it to your portfolio:

- In a few sentences describe what a phylogenetic tree is.
- A comparison of human and mouse chromosomes shows how the structures of genomes diverge. Read the section "A comparison of human and mouse chromosomes" in Alberts, pages 249-250, and comment on the divergence of mouse and human genomes.
- Discuss how gene duplication and divergence result in the generation of new genes. (Refer to

Alberts, pages 253-255, the sections “Gene duplication provides an important source of genetic novelty during evolution” and “Duplicated genes diverge”.)

- d) In your own words, explain what a pseudogene is (remember there is a glossary at the back of your textbook).

3.8.3 Feedback on activity 3.4

A phylogenetic tree is a drawing that shows branching depending on the supposed evolutionary relationships between several biological species. The degree of relatedness is based on similarities and differences in their physical or genetic characteristics. The groups that are joined together in the tree are assumed to have descended from a common ancestor.

Learning unit 4

DNA replication and repair

4.1 Introduction to learning unit 4

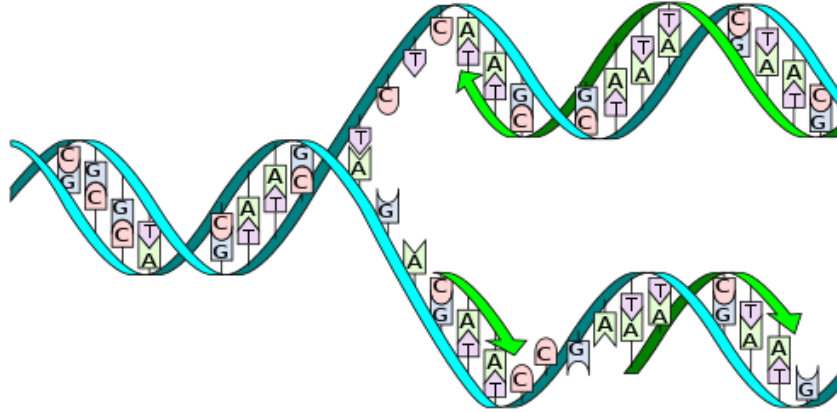


Figure 4.1: DNA replication (http://commons.wikimedia.org/wiki/File:DNA_replication_split_horizontal.svg)

For organisms to grow and reproduce they need to be able to replicate (duplicate) their DNA accurately, and correct any errors that may occur during the replication process. The cells also need to be able to fix any damage that may be done to the DNA. Changes in the DNA are called mutations, and they may have no effect, be harmful, lethal or beneficial. If the mutations occur within the germ cells, the changes will be heritable.

To work through the learning unit, refer to chapter 5 in Alberts and to the internet sources as indicated in the text. The following web page explains the basic process of DNA replication:
<http://cnx.org/content/m46073/latest/?collection=col11496/1.6>

4.2 Learning outcomes

After completing this learning unit, you should be able to

- describe how DNA is replicated and what enzymes are involved in the process
- discuss a number of methods used by the cell to repair damaged DNA
- describe the role of methylation in DNA repair

4.3 DNA replication

Recommended reading: chapter 5, pages 263–281 in Alberts, in the section “DNA replication mechanisms”

The link http://www.youtube.com/watch?v=yqESR7E4b_8 is to a very good video showing the replication process; if possible watch it before **and** after studying this section.

During DNA replication the two strands of the DNA need to be separated so that each can serve as a single-stranded template for a new complementary DNA strand. The region where the strands separate and are copied is called the replication fork. A number of enzymes are involved in this replication process, each with its own role. Some of the enzymes involved are

- helicases - separate and unwind the DNA
- topoisomerases – relieve the supercoiled tension (torsional strain) ahead of the replication fork
- DNA polymerases – catalyse the synthesis of DNA in the 5’-3’ direction
- single-stranded binding (SSB) proteins – bind the DNA and keep the strands apart once they have

been separated

- DNA primase – synthesises short RNA strands that act as primers
- DNA ligase – joins DNA strands by catalysing the formation of a phosphodiester bond

DNA replication is initiated at an AT-rich site called the origin of replication. The DNA is unwound at this point and separated, resulting in a single-stranded region of DNA. The unwinding of the helix is driven by helicases, a class of proteins that catalyse the unwinding of DNA double helices. SSB proteins then bind to the single-stranded DNA to keep the strands apart.



Figure 4.2: The replication bubble, with two replication forks each moving in opposite directions; adapted from http://commons.wikimedia.org/wiki/File:DNA_bubbles2.png

DNA polymerase is the enzyme that catalyses the synthesis of the new DNA strand. It can only add nucleotides in a 5' to 3' direction to an existing 3' OH group; therefore another enzyme is required to add the first new nucleotides. This enzyme is called DNA primase and generates short complementary RNA sequences at the replication fork. Primase is used as it has the ability to initiate RNA synthesis without a pre-existing 3' OH group, something DNA polymerase cannot do.

As we discussed in learning unit 3, the two DNA strands are antiparallel. So when the strands separate, one of the strands runs in a 5'-3' direction and the other in the 3'-5' direction. As DNA polymerase can only add nucleotides in a 5' to 3' direction, only one strand can be synthesised continuously; the other needs to be synthesised in parts, called Okazaki fragments. The strand synthesised continuously is the leading strand and the other is called the lagging strand. Refer to figure 4.3 here and figure 5-7 in Alberts. Make sure you understand why only one of the strands can be synthesised continuously.

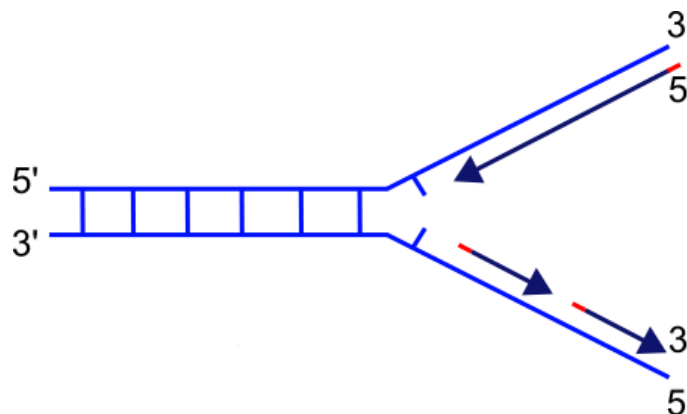


Figure 4.3: The direction of the DNA strands at the replication fork (http://commons.wikimedia.org/wiki/File:Replication_fork.svg)

In the leading strand only one primer is required and the DNA polymerase will continue to add nucleotides to the 3' end. In the lagging strand many primers will be required to initiate the synthesis of each Okazaki fragment. In bacteria each Okazaki fragment is approximately 1 000 nucleotides, whereas in eukaryotes it is approximately 100 nucleotides. The RNA primers are eventually removed and replaced with DNA, by DNA polymerase, and the separate Okazaki fragments are joined by DNA ligase.

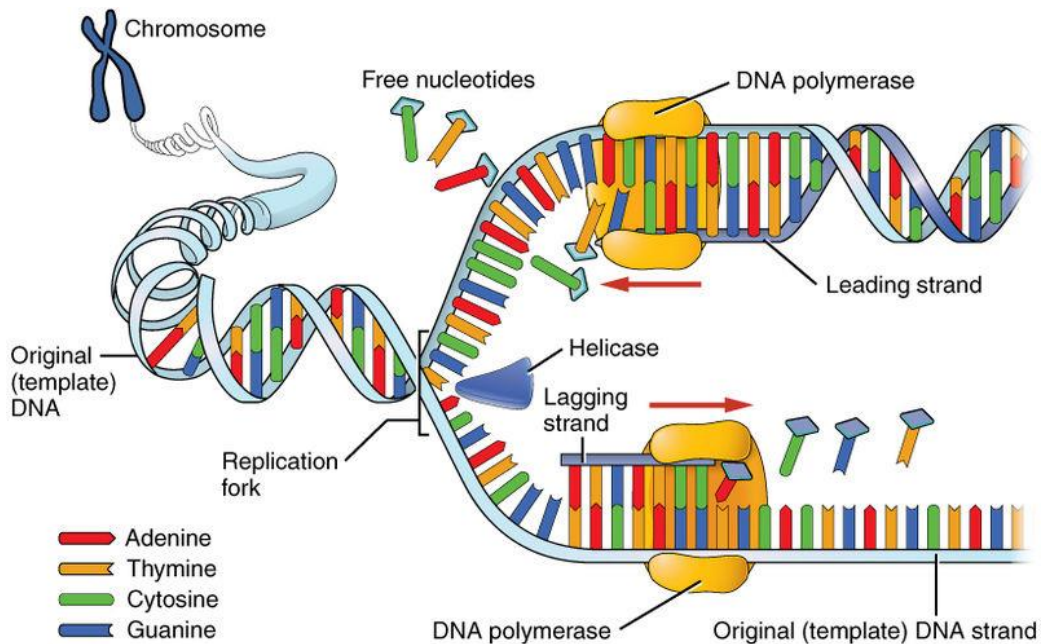


Figure 4.4: DNA replication process; image from OpenStax College, *Anatomy & physiology* (<http://cnx.org/content/m46073/latest/?collection=col11496/1.6>)

Most DNA polymerases have a tendency to fall off the DNA during synthesis. To prevent this, the DNA polymerase is associated with another protein called a sliding clamp. The sliding clamp protein keeps the DNA polymerase connected to the single-stranded DNA as it moves along the DNA.

The proteins at the replication fork do not all work separately but rather interact to form a replication machine. This can be seen clearly in the YouTube video we mentioned at the beginning of this learning unit (http://www.youtube.com/watch?v=yqESR7E4b_8).

4.3.1 Activity 4.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- In your own words, describe the DNA replication process.
- Explain what the sliding clamp and clamp loader are and how they are involved in keeping the DNA polymerase attached to the DNA.
- Refer to the section “DNA topoisomerases prevent DNA tangling during replication” on page 278 in Alberts and describe the role of topoisomerases and how they function.
- Describe the difference between topoisomerase I and topoisomerase II.

4.3.2 Feedback on activity 4.1

The sliding clamp and clamp loader are discussed on page 274 in Alberts.

When describing the DNA topoisomerases did you mention that they can be called reversible nucleases? Do you know what a nuclease is? And that topoisomerase I uses single-stranded breaks to relieve the tension ahead of the replication fork because there is free rotation around the single phosphodiester backbone? Topoisomerase II makes transient double-stranded breaks and can be used to separate two interlocked circles of DNA.

4.4 Proofreading by DNA polymerase

Recommended reading: chapter 5, pages 268–272 in Alberts, in the section “DNA replication mechanisms”

The replication process is highly accurate with only approximately one mistake for every 10^9 nucleotides added. This high accuracy is in part due to the proofreading abilities of the DNA polymerase. The first proofreading check is done before the nucleotide is covalently added to the growing chain. The DNA polymerase undergoes a conformational change that allows it to detect if it has the correct base before it catalyses the addition of the nucleotide. However, if it does make a mistake, the DNA polymerase also has 3'-5' exonucleolytic proofreading activity. This means if the incorrect base is covalently attached, it has the ability to recognise that a mistake has been made and to remove the incorrect base before it adds the correct base. Refer to figures 5-8 and 5-9 on page 270 in Alberts.

There is an additional strand-directed mismatch repair system that checks the DNA immediately after replication. This system detects distortions in the helix owing to incorrect base pairing. The process involves three steps:

- recognition of a mismatch
- excision of the segment of incorrect DNA from the new strand
- resynthesis of DNA using the old strand as the template

The mismatch repair system can differentiate between the original strand and the daughter strand owing to methylation. After replication the A residues in certain sequences are methylated (if you don't know what methylation is, look it up). The methylation process is not immediate; therefore for a short time the old strand can be distinguished from the new strand and the repair machinery will correct the error in the unmethylated strand, using the methylated strand as a template.

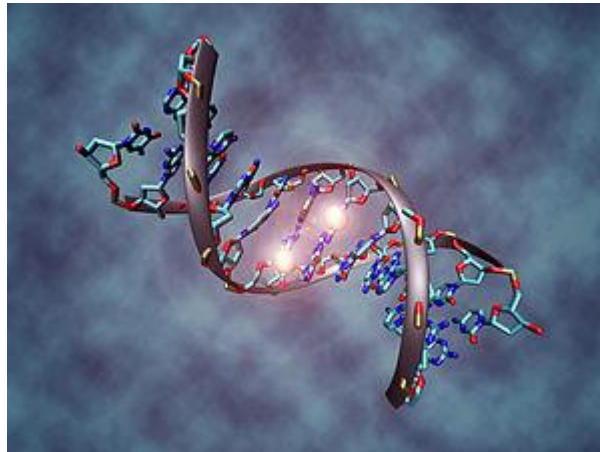


Figure 4.5: Model showing a DNA molecule with two methylated residues (http://en.wikipedia.org/wiki/File:DNA_methylation.jpg)

4.4.1 Activity 4.2

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- Explain why DNA replication in the 5'-3' direction allows for efficient error correction, and why replication in the opposite direction would make error correction harder. Refer to the section “Only DNA replication in the 5' -to-3' direction allows efficient error correction” pages 271-272 in Alberts.
- An enzyme that starts new nucleotide chains cannot be efficient in error correction. Why then is a

RNA primer better than a DNA primer?

- c) What is methylation of DNA?
- d) In your own words, describe the mismatch repair system.

4.4.2 Feedback on activity 4.2

Could you explain why DNA replication in the 5'-3' direction allows for efficient error correction? If synthesis of DNA were in the 3' to 5' direction, then the 5' end of the chain rather than the incoming nucleotide would provide the activating triphosphate. This situation is fine if there are no errors during synthesis. However, if an error was detected and the last nucleotide was removed, then the resulting 5' end would not have a triphosphate and DNA synthesis would have to halt. Refer to figure 5-10 in Alberts which illustrates the reason for the 5'-3' directionality of DNA synthesis.

An RNA primer is better than a DNA primer because if there were more errors in the primer sequence than in the rest of the synthesised DNA, then the cell would want to remove and replace these sequences. If they were DNA sequences, it would be difficult to distinguish them from the rest of the DNA. But because they are RNA, they can be easily detected, removed and replaced with DNA that has been synthesised with a more accurate enzyme.

The mismatch repair system is described in detail on pages 276-277 in the section "A strand-directed mismatch repair system removes replication errors that escape from the replication machine". Remember to look at the figures as well.

4.5 DNA repair

Recommended reading: chapter 5, pages 295–304 in Alberts, the section on DNA repair

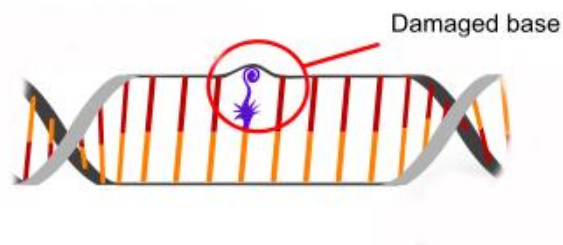


Figure 4.6: Damaged DNA (http://commons.wikimedia.org/wiki/File:Dna_repair_base_excision_en.svg)

DNA within a cell can be damaged by a number of mechanisms including

- chemicals (external or internally generated)
- radiation
- heat

Random changes in DNA that are not corrected will result in genetic mutations, which may have negative consequences for the cell. This can be seen in cases where an organism has defective DNA repair genes and large numbers of mutations accumulate. Fortunately, the DNA helix is easily repaired. There are two major DNA repair pathways:

- base excision repair
- nucleotide excision repair

A different repair mechanism fixes double-stranded breaks.

4.5.1 Activity 4.3

Do the following activity and add it to your portfolio:

- Describe how the chemical modification of nucleotides can produce mutations. (Refer to Alberts, figure 5-47.)
- Referring to Alberts, explain base excision repair and nucleotide excision repair and when each type of repair mechanism is used.

4.5.2 Feedback on activity 4.3

Base excision repair and nucleotide excision repair are described on pages 297 and 298 in the textbook. Can you draw a series of images and describe both forms of repair mechanisms, similar to figure 5-48? Did you mention DNA glycosylases and AP endonuclease with regard to base excision repair, and bulky lesions, multienzyme complex, DNA polymerase and DNA ligase when discussing nucleotide excision repair?

4.6 Chromosome replication

Recommended reading: chapter 5, pages 281–284 and 292 in Alberts, the section on the initiation and completion of DNA replication in chromosomes

The chromosomes of bacteria and eukaryotes are different, so although the general replication machinery is similar, the initiation of replication and number of replication forks is different. Many bacteria have a single circular chromosome with one origin of replication. Replication then begins at a single point and moves around the chromosome bidirectionally outward from the origin of replication. This results in two replication forks that move around the chromosome until the entire chromosome is copied. Refer to figure 5-27, page 283, in Alberts.

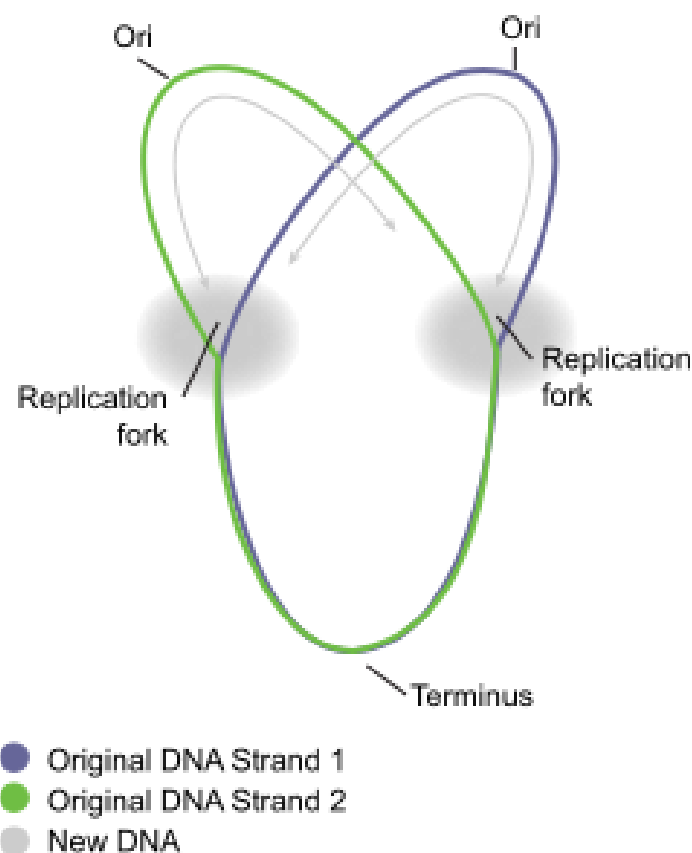


Figure 4.7: Replication of a circular bacterial chromosome, showing DNA replication proceeding

bidirectionally (http://commons.wikimedia.org/wiki/File:Circular_DNA_Replication.svg)

Eukaryotes have long linear chromosomes. Because of their size, there are a number of replication forks that copy the DNA simultaneously. These multiple replication forks move along the DNA and stop when they collide with another replication fork moving in the opposite direction. The ends of the linear chromosomes are replicated by telomerase.

4.6.1 Activity 4.4

Do the following activity and add it to your portfolio:

Refer to your textbook and answer the following questions:

- a) What is an origin of replication? Is this region AT or GC rich? Explain why this is the case.
- b) What is bidirectional replication?
- c) Describe how eukaryotic chromosomes are replicated using multiple origins of replication.
- d) Explain how telomerase replicates the ends of chromosomes. (Refer to figure 5-41 and page 292 in Alberts.)

Learning unit 5

Transcription and translation

5.1 Introduction to learning unit 5

For an organism to survive, the information in the DNA needs to be converted into a functional form, i.e. the genes need to be expressed. All cellular organisms use the same basic mechanism of gene expression. When a particular gene is expressed, the relevant region of DNA is unwound and copied into an RNA molecule. This process is called transcription. If the gene encodes for a protein, mRNA is produced. The mRNA is then read by a ribosome to form a protein in a process called translation. Therefore protein-coding regions of DNA are not converted directly into proteins, but are first converted into an intermediary mRNA molecule. The process of converting DNA into RNA and then into protein is referred to as the central dogma of molecular biology. Remember we discussed it in learning unit 1?

To work through the learning unit, refer to chapter 6 in Alberts and to the internet sources as indicated in the text.

5.2 Learning outcomes

After completing this learning unit, you should be able to

- explain how DNA is transcribed
- explain how RNA is translated into protein
- describe a number of different types of RNA that are produced in cells and their functions

5.3 Transcription

5.3.1 RNA synthesis

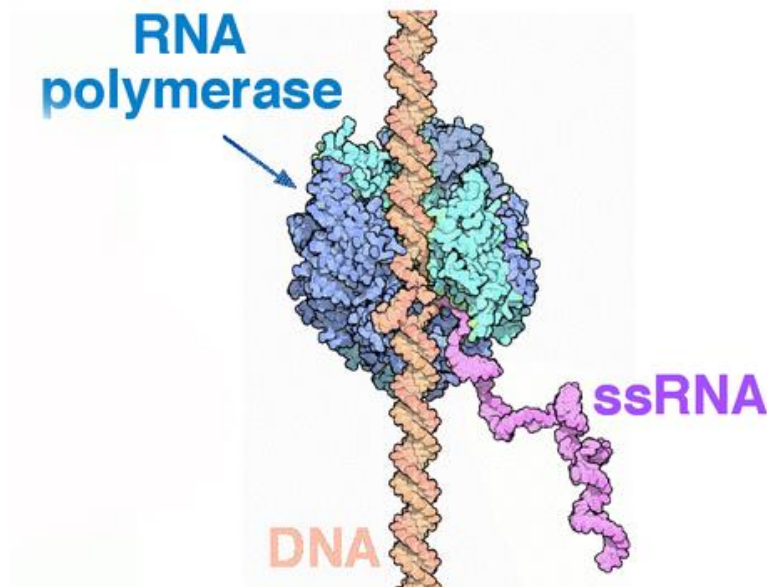


Figure 5.1: The synthesis of RNA (http://commons.wikimedia.org/wiki/File:Transcription_both_strands.png)

Recommended reading: chapter 6, pages 329–348 in Alberts, the section “From DNA to RNA”

Transcription is the process of converting a portion of DNA into an RNA copy. It involves the unwinding of a section of the DNA helix while RNA polymerase catalyses the formation of an RNA chain complementary to the DNA. The reaction that is catalysed by the RNA polymerase is similar to that of DNA polymerase and also occurs in the 5'-3' direction. If mRNA is produced, it is read by a ribosome to form a protein in a process

called translation. Many genes are not translated into proteins and RNA is the final product. These RNA molecules fold into three-dimensional shapes and may have catalytic or regulatory functions.

There are a number of different forms of RNA that are produced by cells:

- mRNA - messenger RNA, which is the message for protein synthesis
- tRNA - transfer RNA carries the amino acids during protein synthesis
- rRNA - ribosomal RNA forms important parts of the ribosome
- snRNAs - small nuclear RNAs are involved in rRNA and mRNA processing
- miRNAs - micro-RNAs cause inhibition of gene expression
- siRNAs - small interfering RNA cause inhibition of gene expression

5.3.2 The bacterial RNA polymerase

RNA polymerase (RNAPol), also known as DNA-dependent RNA polymerase, is an enzyme that produces RNA from a DNA template (refer to figure 6-8 in Alberts). The bacterial RNA polymerase enzyme is made up of the core enzyme and the σ (sigma) factor. The σ factor is not involved in catalysis but it helps the enzyme recognise the promoter region, which is the region where transcription starts. When the σ factor and core enzyme are joined together, we refer to it as the RNA polymerase holoenzyme. The holoenzyme is required to start transcription, but only the core enzyme is required to continue synthesis of the RNA.

5.3.3 Activity 5.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- In what respects does RNA differ from DNA?
- In your own words, describe the different types of RNA produced in cells and the roles they play in the cell.
- Discuss some of the differences between DNA polymerase and RNA polymerase. (Refer to pages 333-334 in Alberts, the section "Transcription produces RNA complementary to one strand of DNA".)
- Why does RNA polymerase not need to be as accurate as DNA polymerase?

5.3.4 Transcription process in bacteria

The promoter region is a sequence of nucleotides that indicates where RNA synthesis must begin. Transcription in bacteria is simpler than in eukaryotes and we will talk about that first (refer to figure 6-11 in Alberts).

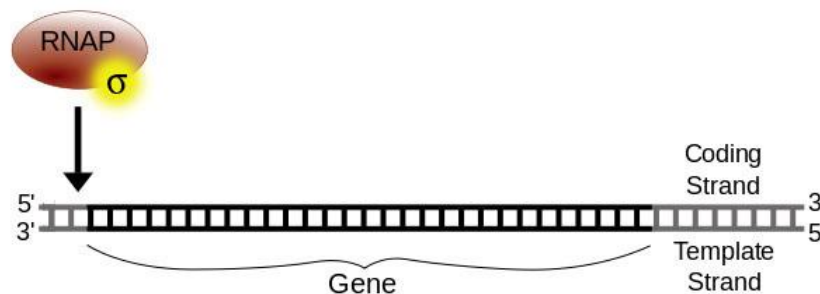


Figure 5.2: Binding of RNA polymerase (RNAP) to the DNA; adapted from http://commons.wikimedia.org/wiki/File:Simple_transcription_initiation1.svg

When the RNA polymerase holoenzyme comes into contact with a promoter, it binds tightly to the DNA. The σ factor help the RNA polymerase recognise the promoter region. Different σ factors will recognise different

promoters. This differential recognition helps to control which genes are transcribed, and we will discuss this further in learning unit 6. The RNA polymerase holoenzyme then unwinds and opens a region of the DNA. One of the exposed DNA strands act as the template for the start of transcription. After the synthesis of approximately ten nucleotides of RNA, the σ factor dissociates. These initial steps are referred to as the initial stage.

The transcription process can be divided into three stages

- initiation (binding of holoenzyme, opening of helix, addition of first nucleotides)
- elongation (the RNA chain is extended in length)
- termination (signals indicate the end of transcription and the RNA is released)

The initiation step is important as a regulatory step; if transcription is not initiated, then the gene will not be expressed. After initiation, the RNA polymerase shifts to elongation mode, and chain elongation continues at a rate of approximately 50 nt/sec in bacteria.

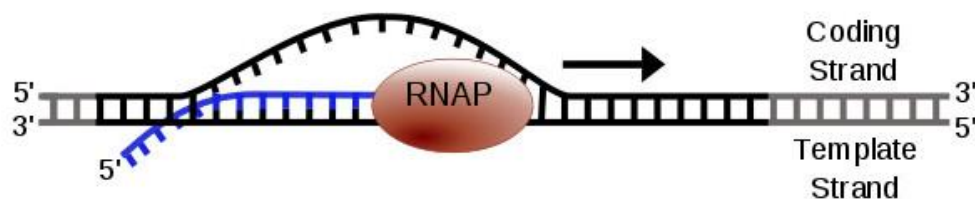


Figure 5.3: Elongation of RNA transcript; adapted from http://commons.wikimedia.org/wiki/File:Simple_transcription_elongation1.svg

The final step is termination, which is directed by a termination sequence in the DNA. During termination, the RNA polymerase stalls and the DNA and RNA chains are released. The RNA polymerase can then bind a new σ factor and the process is repeated with the same or a different gene.

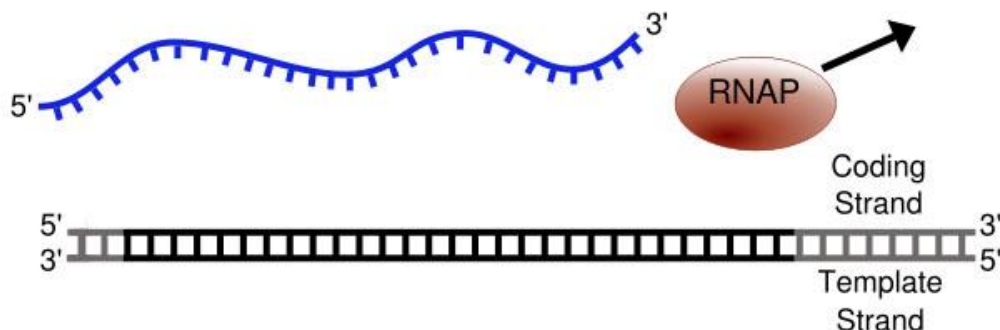


Figure 5.4: Termination of transcription; adapted from http://commons.wikimedia.org/wiki/File:Simple_transcription_termination1.svg

5.3.5 Activity 5.2

Do the following activity and add it to your portfolio:

Refer to your textbook and answer the following questions:

- In your own words, describe transcription in bacteria. Make sure you discuss the promoter, coding strand, template strand, σ factor, RNA polymerase and 5' to 3' direction
- Refer to pages 338-339 in Alberts, the section "Transcription start and stop signals are heterogeneous in nucleotide sequence". Describe what a consensus nucleotide sequence is.

5.3.6 Feedback on activity 5.2

You **must** be able to describe transcription, as this is a vital process that occurs within cells and is universal to all life on earth.

After reading the section in the textbook, you should have realised that a consensus nucleotide sequence is the average or most representative form of a sequence that occurs with minor variations in specific regions of DNA or RNA. For a σ factor to be able to recognise and bind to different promoters, there will have to be some common sequence elements that the σ factor recognises. These elements are the consensus sequence.

5.3.7 Transcription in eukaryotes

Transcription in eukaryotes is more complex than in bacteria and is performed by three different RNA polymerases.

There are three major RNA polymerases in eukaryotes:

- RNA polymerase I which transcribes genes encoding tRNAs, rRNAs and other small RNAs
- RNA polymerase II which transcribes most genes, including protein-encoding genes
- RNA polymerase III which transcribes genes encoding tRNAs, rRNAs and other small RNAs

In this learning unit we will only look at RNA polymerase II and the polymerase that transcribes protein-encoding genes. RNA polymerase II requires a number of general transcription factors to initiate transcription, unlike the bacterial RNA polymerase which requires just the σ factor. Once the RNA polymerase has started transcription, most of the general transcription factors are released.

RNA polymerase II also requires activator, mediator and chromatin-modifying proteins when transcribing DNA in eukaryotic cells. The transcription activator proteins bind sequences in the DNA to help sequester the RNA polymerase to the start of transcription. The mediator protein helps the RNA polymerase II to interact with the activator proteins and general transcription factors. The chromatin-modifying proteins help the RNA polymerase II to manage the complex packaging of the DNA. They alter the histones and therefore change the chromatin structure which influences gene transcription. These proteins are all important in the regulation of gene expression, but we won't go into them in detail in this module.

Coupled transcription

In bacteria, transcription and translation are often coupled with translation of the mRNA beginning before transcription is complete. In eukaryotes the mRNA is first modified and transported from the nucleus to the cytoplasm; therefore, translation only starts after transcription is finished.

5.3.8 Activity 5.3

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

Read up on the general transcription factors of RNA polymerase II (pages 339-342 in Alberts, the section entitled "RNA polymerase II requires general transcription factors"). Discuss how they are involved in the

initiation of transcription of eukaryotic genes. Be sure to discuss TFIID, the TATA sequence and the transcription initiation complex.

5.3.9 mRNA processing

mRNA produced by eukaryotes undergoes a number of processing steps within the nucleus. The mRNA is modified by the addition of a 5' cap and a 3' poly-A tail and the introns are also removed in a process called RNA splicing.

Before you read further, refer to the following diagram, which gives an overview of the process.

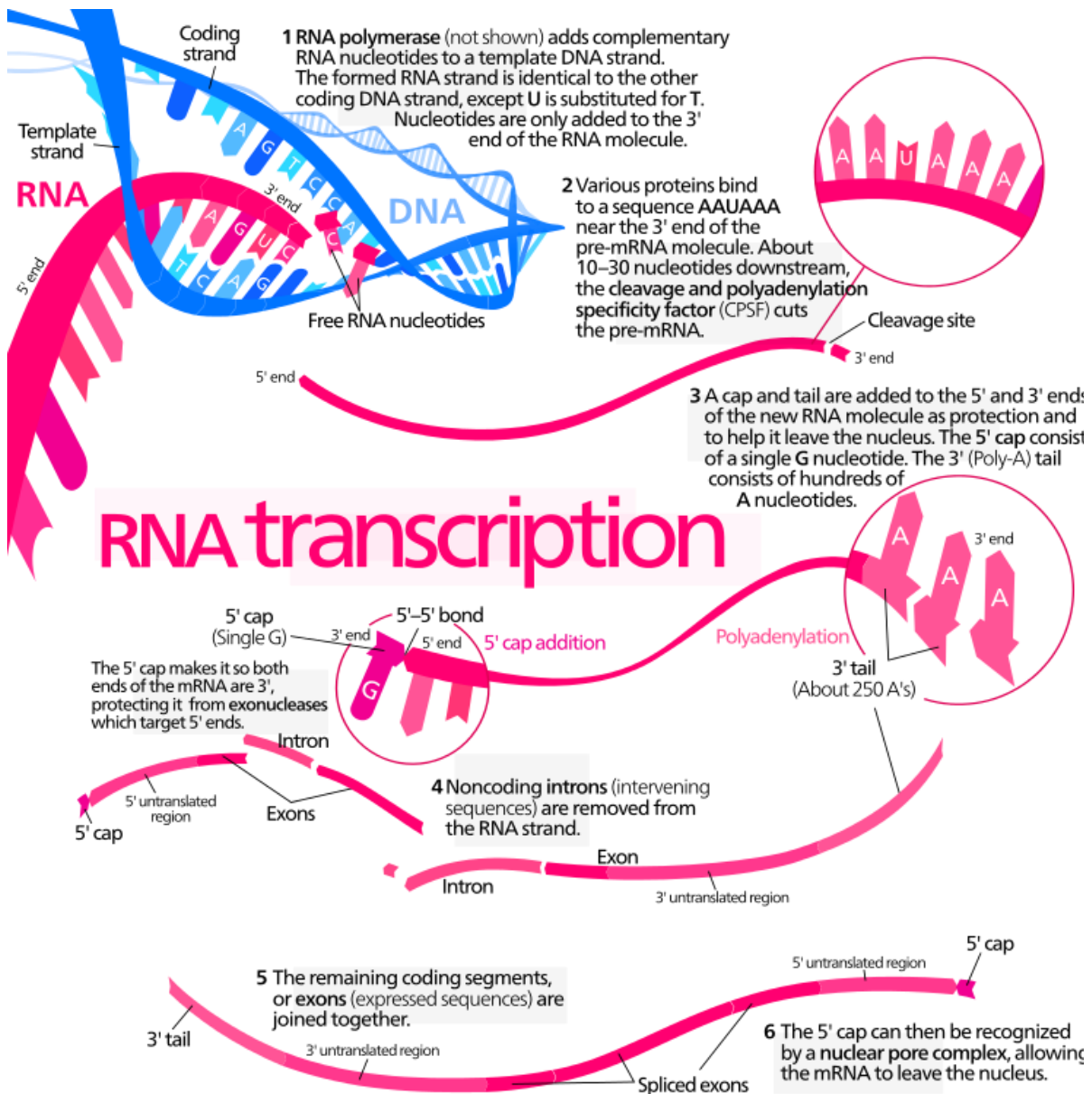


Figure 5.5: Transcription of a protein-coding gene in eukaryotes (<http://en.wikipedia.org/wiki/File:MRNA.svg>); image credit: Kelvinsong

The 5' end is modified by the addition of a methylated guanine cap. The 5' cap helps the cell to distinguish the mRNAs from the other RNAs that are produced in the cell, as only the mRNAs are capped. The 3' end is polyadenylated, which results in the addition of a number of adenine residues, known as the poly-A tail. The modification of the 5' and 3' ends of the mRNA also protects the mRNA from nuclease attack.

In eukaryotes the RNA molecules transcribed contain both coding (exon) and non-coding (intron) sequences. Therefore, before the RNA can be translated into a protein, the introns need to be removed in a process termed splicing.

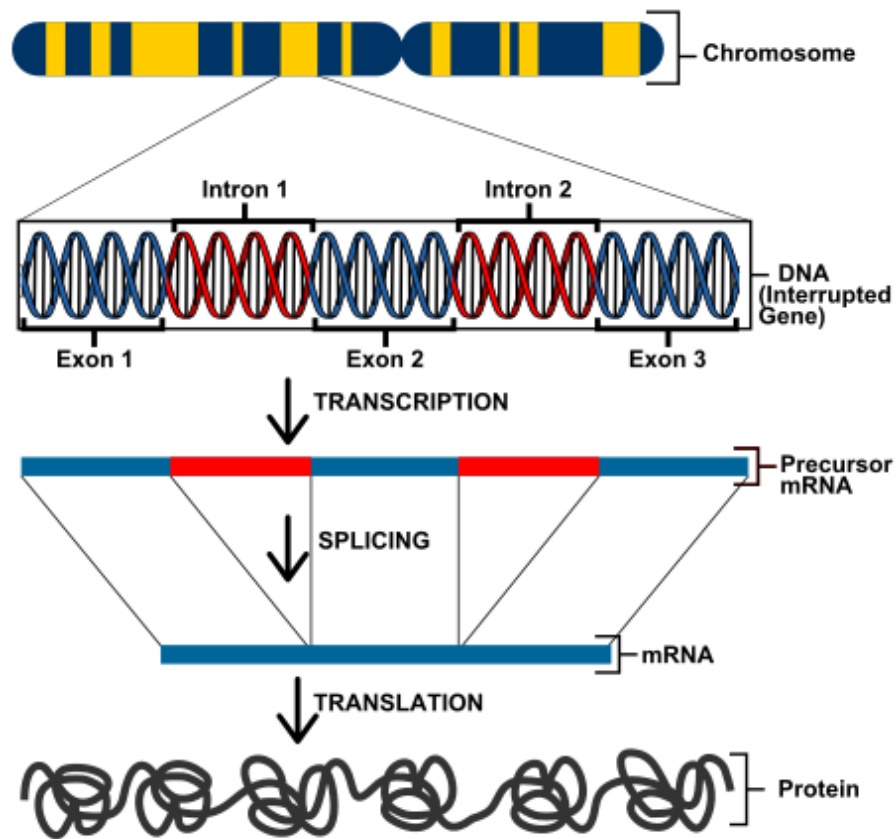


Figure 5.6: RNA splicing (http://commons.wikimedia.org/wiki/File:Eukaryotic_Transcription_.svg)

5.3.10 Activity 5.4

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- Briefly describe how eukaryotic mRNAs are modified after they have been synthesised.
- Describe what RNA splicing is and how it is performed.
- Compare transcription in eukaryotes and bacteria. Remember to mention transcription factors, sigma factor, activator, mediator, consensus sequence, capping, poly-A tail and splicing.

5.3.11 Feedback on activity 5.4

The following YouTube video shows the RNA splicing process. Although it is not essential that you watch it, it is very interesting and informative: <https://www.youtube.com/watch?v=aVgwr0QpYNE>

When describing the splicing process, did you mention exons, introns and the lariat structure? (Refer to pages 347 and 348 in Alberts.)

After you have compared transcription in eukaryotes and bacteria, refer to figure 6-21 in Alberts to see if you discussed the main differences.

Remember to watch the video showing the transcription process in eukaryotes again now that you

understand the process a bit more: <http://www.youtube.com/watch?v=SMtWvDbfHLo>

5.4 Translation

Recommended reading: chapter 6, pages 366–382 in Alberts, the section entitled “From RNA to protein”, which describes the basics of translation

5.4.1 What is translation?

Most genes in a cell encode for mRNAs that serve as templates for the production of proteins. In molecular biology, translation is the process where the information in the mRNA sequence, which contains four different subunits, is converted into an amino acid sequence containing 20 different subunits. The resulting polypeptide chain then folds into a three-dimensional structure forming the active protein.

Before you read further, examine figure 5.7 in this study guide to get an overview of the basic process of translation.

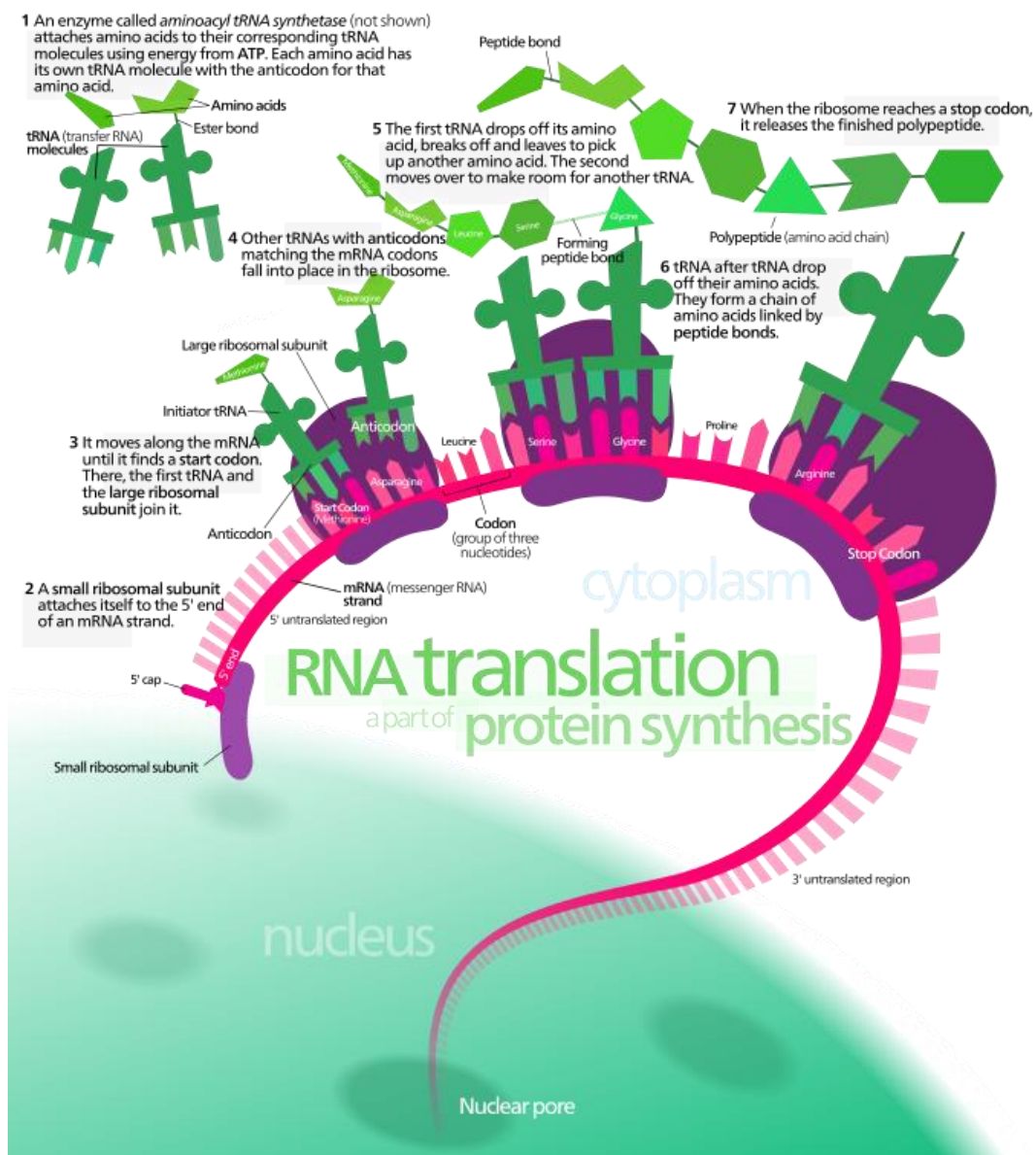


Figure 5.7: RNA translation (http://en.wikipedia.org/wiki/File:Protein_synthesis.svg); image credit: Kelvinsong

5.4.2 How is an mRNA sequence decoded?

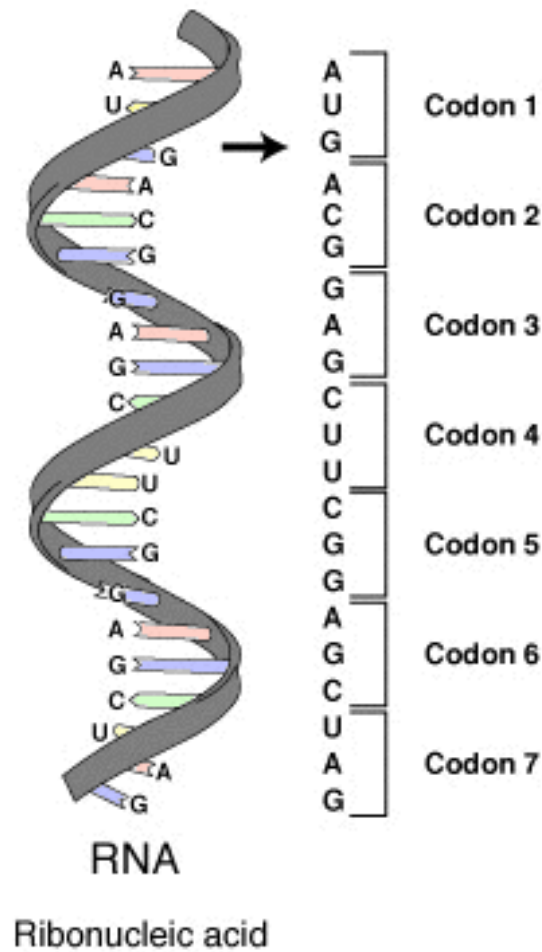


Figure 5.8: mRNA molecule with the codons shown (<http://en.wikipedia.org/wiki/File:RNA-codons.png>); image credit: TransControl

As DNA and RNA are similar, the DNA can act as a direct template for the synthesis of RNA by complementary base pairing. This is not possible for the synthesis of proteins. During translation, the mRNA sequence is decoded using sets of three nucleotides called codons. Each codon specifies a specific amino acid or a stop to translation. The codons do not bind directly to the amino acids; rather, tRNAs act as adaptor molecules that recognise both the codon, in the RNA, and the amino acid. tRNAs are about 80 nucleotides in length and have an acceptor stem which binds to a specific amino acid and an anticodon loop which binds to a specific codon on the mRNA. Enzymes called aminoacyl-tRNA synthetases couple each amino acid to the correct tRNA molecule. Refer to figures 5.9 and 5.10 in the study guide and figure 6-52 in Alberts, which depict the general structure of a tRNA molecule.

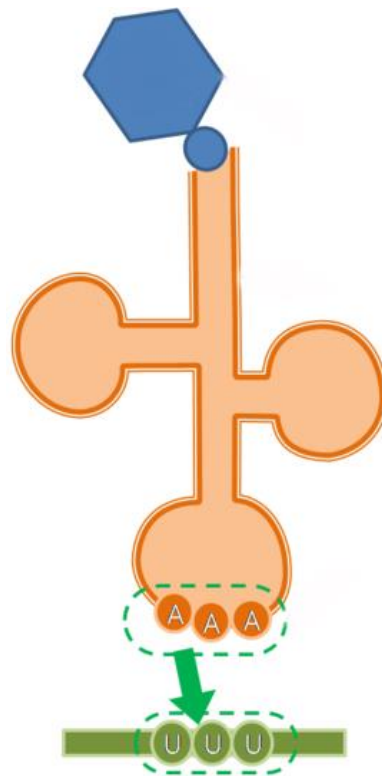


Figure 5.9: A tRNA molecule bound to an amino acid (blue) and the codon of an mRNA chain (green); adapted from http://commons.wikimedia.org/wiki/File:TRNA_illustration_Korean.png

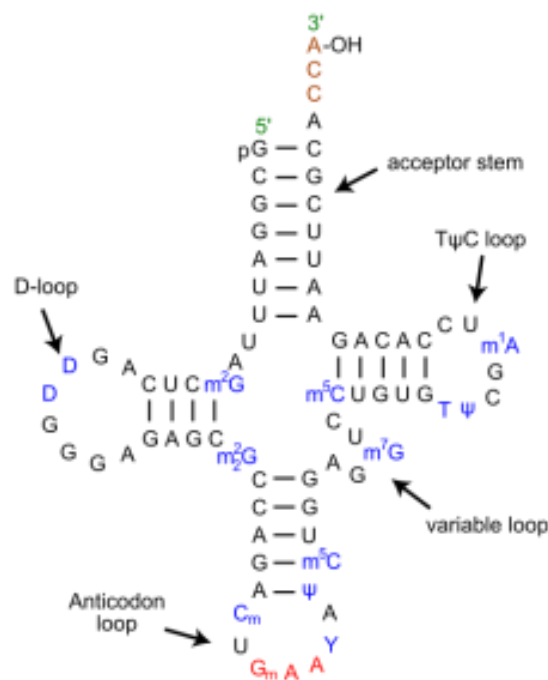


Figure 5.10: A tRNA molecule (http://commons.wikimedia.org/wiki/File:TRNA-Phe_yeast_en.svg)

The anticodon of the tRNA is the region that binds to the codon of the mRNA. Which codon corresponds to which amino acid is referred to as the genetic code. A codon usage table (table 5.1) can be used to predict the amino acid sequence of a protein from the DNA sequence of a gene.

Table 5.1: Codon usage table (<http://openwetware.org/wiki/CH391L/S12/TranslationRBSandCodons>)

		Second Letter					
		U	C	A	G		
1st letter	U	UUU Phe UUC UUA Leu UUG	UCU Ser UCC UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G	
	C	CUU Leu CUC CUA CUG	CCU Pro CCC CCA CCG	CAU His CAC CAA Gln CAG	CGU Arg CGC CGA CGG	U C A G	
	A	AUU Ile AUC AUA AUG Met	ACU Thr ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G	
	G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU Gly GGC GGA GGG	U C A G	

Using the table you can determine the amino acid sequence if you have the DNA sequence. If you have the RNA sequence AUGCUUCAUGAUGA, this will encode for the peptide with the sequence Met, Leu, His, Arg. Note the last codon is a stop codon and does not encode for an amino acid. Some tRNAs require accurate base pairing of only the first two positions of the codon and can handle a mismatch at the third position. This is referred to as the wobble base. This means it is possible to match 20 amino acids to 61 different codons (there are 64 possible codons, but 3 are stop codons) with only 31 different tRNAs. Therefore some tRNAs can base pair with more than 1 codon. The actual number of tRNAs present is different in each organism. There are 3 possible reading frames in an mRNA sequence.

5.4.3 Activity 5.5

Do the following activity and add it to your portfolio:

- What is the sequence of the polypeptide translated from the mRNA with the following sequence: 5'-AUGCGAGCUAUUUCUGCUUAA? (Refer to the codon usage table.)
- What is meant by reading frame? Draw a diagram showing the three reading frames for the sequence in question (a).
- In your own words, describe the structure of a tRNA molecule.
- What enzyme couples the amino acid to the appropriate tRNA molecule?

5.4.4 Feedback on activity 5.5

If you are having a problem understanding reading frames, look at figure 6-51 on page 368 in Alberts.

5.4.5 Ribosome structure

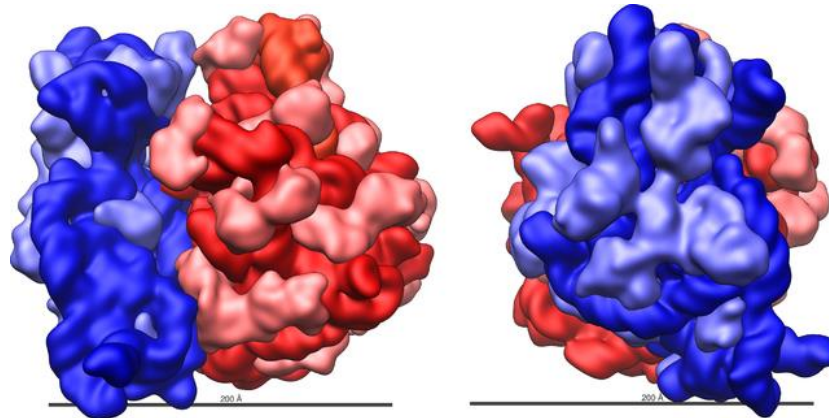


Figure 5.11: The large (red) and small (blue) subunit of a ribosome
(http://en.wikipedia.org/wiki/File:Ribosome_shape.png)

We know that amino acids are joined to tRNA molecules that can also bind mRNA, but how are the amino acids joined together to form proteins? During protein synthesis each amino acid is joined by a peptide bond. This reaction is catalysed by the ribosome, a complex macromolecular machine made from ribosomal (rRNAs) and more than 50 different proteins. Eukaryotic ribosomes are assembled in the nucleolus where the rRNA is transcribed. The ribosomal proteins are made in the cytoplasm and are imported into the nucleus. The assembled ribosomal subunits are then transported out the nucleus and into the cytoplasm.

5.4.6 Activity 5.6

Do the following activity and add it to your portfolio:

- Ribosomes are made up of two subunits, one small and one large. Discuss the differences between the ribosomal subunits found in bacteria and eukaryotes (refer to figure 6-63 in Alberts to help you).
- The ribosome is a ribozyme. What does this mean? (Refer to page 378 in Alberts.)

5.4.7 The translation process

Similar to transcription, translation can be divided into three stages:

- chain initiation
- chain elongation
- chain termination

The link http://www.youtube.com/watch?v=TfYf_rPWUdY is to a video showing the translation process. If possible, watch it before **and** after studying this section.

Initiation

Initiation at the correct nucleotide is important because if initiation is shifted over by a nucleotide, the translation will no longer be in frame. Translation always begins at the codon AUG. A specific tRNA, the initiator tRNA, is required to start translation. The initiator tRNA carries the amino acid methionine (in bacteria it is modified to formylmethionine). Although methionine (Met) is the first amino acid in any new protein, the methionine residue may be removed after translation, and may therefore not be present in the mature protein.

During initiation, initiation factor proteins bind to the small subunit of the ribosome forming the pre-initiation complex. This complex and the methionine-carrying tRNA then bind to the mRNA, forming the initiation complex.

After the initiation complex is formed, the large ribosomal subunit binds to the complex and the initiation factors are released. The large ribosomal subunit has three sites where tRNAs can bind. They are the

aminoacyl site (A site), the peptidyl site (P site) and the exit site (E site). Refer to figure 6-64 in Alberts. At this stage the met-tRNA will occupy the P site.

Elongation

During elongation, a charged tRNA enters the A site of the ribosome. Refer to figure 6-66 in Alberts. If the tRNA is the correct one, such that the anticodon of the tRNA is complementary to the codon in the A site, peptidyl transferase will catalyse the formation of a peptide bond between the amino acid in the A site and the amino acid in the P site. The ribosome then moves one codon down, in a process called translocation, so the tRNA that was in the A site moves to the P site, and the tRNA that was in the P site moves to the E site. The E site is the site where the tRNA that has given up its amino acid sits before it is released back into the cytoplasm. The A site is now empty and is ready to receive the next charged tRNA. This process is repeated and new amino acids are added to the polypeptide chain until all the codons in the mRNA have been read. The elongation process is aided by a number of elongation factors.

Termination

Once the polypeptide has been fully translated it needs to be released from the ribosome. There are three termination codons: UAG, UAA and UGA. They do not have matching tRNAs and signal the ribosome to stop translating. Release factors, types of protein, bind to the ribosome with a stop codon in the A site, causing the polypeptide and mRNA to be released from the ribosome. The inactive ribosome complex then dissociates into its two subunits, which are ready for another round of translation.

The translation process in both prokaryotes and eukaryotes is similar, although the initiation, elongation and termination factors are different.

Post-translational processing

In eukaryotes, proteins destined for export to specific parts of the cell or from the cell have leader sequences at their N-terminal ends. These leader sequences are recognised and direct the proteins to their proper destination. They are eventually removed by specific proteases.

5.4.8 Activity 5.7

Do the following activity and add it to your portfolio:

Read the article: Clancy, S & Brown, W. 2008. Translation: DNA to mRNA to protein. *Nature Education*, 1(1):101, <http://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393>. Using your textbook and the article, answer the following questions:

- In your own words, describe the translation process. Discuss chain initiation, elongation and termination. Refer to your textbook for additional information.
- What is the untranslated region (UTR) or leader sequence of an mRNA molecule?
- What is the function of initiation, elongation and termination factors in the translation process?
- Compare the initiation of translation in prokaryotes and eukaryotes. Include the terms "Shine-Dalgarno", "AUG", "polycistronic", "5' cap" and "initiation factor". (Refer to pages 379-380 in Alberts, the section "Nucleotide sequences in mRNA signal where to start protein synthesis".)

Learning unit 6

Regulation of gene expression

6.1 Introduction to learning unit 6

Look at the following images. They are two different cells from the human body – one is an epithelial cell from the cheek, the other is a cell from the central nervous system. Both of these cells have a different structure and biochemical makeup and perform different functions within the human body. Yet both contain the same genetic code. What allows cells to become so different?

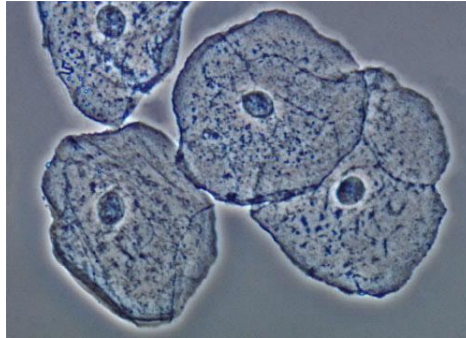


Figure 6.1: Epithelial cells from cheek (<http://cellfunctioning.wikispaces.com/Epithelial+cells>)

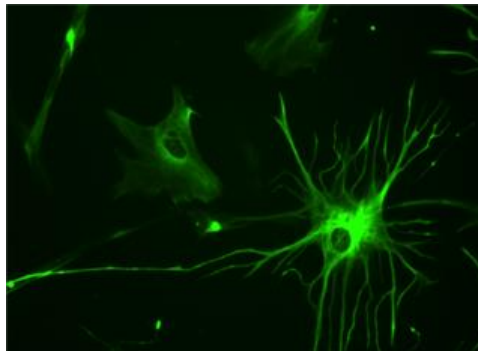


Figure 6.2: Astrocyte http://en.wikipedia.org/wiki/File:Human_astrocyte.png

During the development of a multicellular organism, the cells undergo differentiation owing to the expression of different genes. The variations in gene expression at different times result in the production of different proteins and other molecules within the various cell types. This in turn affects the structure and function of the cell. In this learning unit we will focus on some of the mechanisms used by cells for controlling which genes are switched on and which are switched off. To work through this learning unit, refer to chapter 7 in Alberts and the internet sources, as indicated in the text of the learning unit.

6.2 Learning outcomes

After completing this learning unit, you should be able to

- describe why gene regulation is important
- describe the structure of some DNA-binding motifs found in proteins
- explain how bacteria regulate their trp and lac operons
- describe the various methods eukaryotes use to regulate gene expression

6.3 Gene regulation

Recommended reading: chapter 7, pages 411–415 in Alberts, an overview of gene control

All the cells in a multicellular organism (bar a few exceptions) contain the same genome, yet may be very different from one another (e.g. astrocyte vs. epithelial cell). These differences come about through variations in gene expression resulting in the presence of different RNAs and proteins and are not due to changes in their DNA sequence.

Distinctive cell types produce different proteins, because signals control which genes are expressed at which time. This results in cells with different biochemical makeups that perform diverse functions. External signals can cause a cell to change the expression of its genes and different cell types will also often respond differently to the same extracellular signals, as a result of having different biochemical pathways active within the cell. For example, glucocorticoids cause liver cells to increase their production of glucose and cause fat cells to reduce their production of tyrosine aminotransferase.

There are six steps at which eukaryotic gene expression can be controlled (refer to figure 7-5, page 415 in Alberts):

- transcriptional control
- RNA processing control
- RNA transport and localisation control
- rate of mRNA degradation
- translational control
- regulation of protein activity

We will focus on the first four points in this learning unit.

6.3.1 Activity 6.1

Do the following activity and add it to your portfolio.

Remember, these activities serve as part of your preparation for the exam!

- a) Discuss why the regulation of genes is important.
- b) Discuss what evidence there is that the different cell types of a multicellular organism contain all the genetic information required to generate a complete organism.

6.3.2 Feedback on activity 6.1

If you are having trouble answering question (b), see figure 7-2 in Alberts, where it is clearly illustrated.

6.4 Transcriptional control

Recommended reading: chapter 7, pages 416–425, 432-441 in Alberts

For most genes the initiation of transcription is the first and most important level of regulation of gene expression. In bacteria the level of transcription is often controlled by a regulatory region of DNA close to the transcription start site. A regulatory protein then interacts with this region of DNA, either stopping or allowing transcription to occur, and thus the gene is either switched on or off.

Gene regulatory proteins (transcription factors) were initially discovered in bacteria during the 1950s and were found to control the expression of sets of genes. These regulatory proteins were found to bind specific DNA sequences close to the genes they regulated. The lac repressor is a well-studied regulator protein that turns off the genes encoding for the proteins that metabolise lactose when lactose is not present as a food source.

6.4.1 Regulatory protein structure

Regulatory proteins work by binding the DNA, then moving along it and scanning for the sequence they are looking for without having to unwind and open the helix. They generally bind to DNA as dimers.

There are a number of structural motifs that are present in DNA-binding proteins that allow them to read DNA sequences. Refer to the textbook or the internet for pictures of the motifs described below. The helix-turn-helix (HTH) is a common DNA-binding motif found in proteins that are capable of binding DNA. It comprises two α helices joined by a short strand of amino acids and is found in many proteins that regulate gene expression.

The homeodomain proteins are a special class of HTH proteins that were initially discovered in *Drosophila*. In the homeodomain proteins the HTH motif is always surrounded by the same structural elements. The zinc finger motif and leucine zipper motif are two other important DNA-binding motifs. Zinc fingers can be classified into a number of distinct structural groups, but all of them share the common feature of binding to the major groove of the DNA and containing zinc as a structural element that is used to stabilise the fold of the protein.

Other types of DNA regulatory proteins have two stranded β sheets that can recognise specific sequences or may have protruding peptide loops to recognise nucleotide sequences.

The helix-loop-helix (HLH) motif contains two α helices connected by a loop, one longer than the other. One helix folds and packs against another helix because of the flexibility of the loop. This structure then binds to an additional HLH motif of another protein. If the two HLH proteins are the same, a homodimer is formed; if they are different, a heterodimer is formed.

6.4.2 Activity 6.2

Do the following activity and add it to your portfolio:

Refer to pages 416-426 in Alberts, the section on DNA-binding motifs in gene regulatory proteins, and answer the following questions:

- How do regulatory proteins read the DNA sequence without opening the helix? (How do the proteins distinguish the different nucleotides from one another?)
- Watch the YouTube video (http://www.youtube.com/watch?v=GRL_rdB30GY) or refer to your textbook (there should be a video on the media DVD that came with the textbook) and describe the zinc finger domain.
- Watch the YouTube video (<http://www.youtube.com/watch?v=uUD9Epy9NOI>) or refer to your textbook (there should be a video on the media DVD that came with the textbook) and describe what a leucine zipper motif looks like and what part of the protein binds to the DNA.

6.4.3 Genetic switches

Recommended reading: chapter 7, pages 432-441 in Alberts

The following web page contains additional information on the topic:
<http://www.nature.com/scitable/topicpage/operons-and-prokaryotic-gene-regulation-992>

In bacteria clusters of genes are often under the control of a single promoter. This grouped gene setup is called an operon.

The tryptophan (*trp*) operon in bacteria is regulated primarily by a protein called the tryptophan repressor. There are five genes involved in tryptophan biosynthesis that are arranged as a single operon (refer to figure 7-34 in Alberts). When tryptophan is freely available in the environment, it would be a waste of energy to manufacture tryptophan and the enzymes required for its synthesis. Therefore the cell switches off the genes of the *trp* operon.

Upstream of the protein-coding genes is a region called the promoter. This region contains the sequence of

DNA that the RNA polymerase binds to in order to start transcription. Within the trp promoter there is another region called the operator that is recognised by the trp repressor protein. When the trp repressor binds the operator, the RNA polymerase is blocked from binding the promoter and transcription does not occur (refer to figure 7-35 in Alberts). For the trp repressor to bind the operator, it needs to be bound to two free tryptophan residues. The binding of the tryptophan to the repressor results in a conformational change in the protein, allowing it to bind the operator (refer to figure 7-36 in Alberts). If there is a shortage of tryptophan in the cell, the trp repressor will not be bound to tryptophan and will not be able to bind the operator and the genes will be transcribed.

In the case of the trp operon a repressor protein binds the promoter region and stops transcription of the genes. On the other hand, transcriptional activators are molecules that bind and switch genes on. They generally act by facilitating the binding of the RNA polymerase to the DNA and therefore increasing transcription.

6.4.4 Activity 6.3

Do the following activity and add it to your portfolio:

- What is an operon?
- In your own words, explain how the trp operon is regulated in *E. coli*.
- Discuss the regulation of the lac operon in *E. coli* by both negative and positive transcriptional controls. (Refer to pages 435-438 in Alberts.)
- What role does DNA looping play in the control of the lac operon?

6.4.5 Feedback on activity 6.3

The regulation of the trp operon is discussed in Alberts on pages 433-434 in the section "The tryptophan repressor is a simple switch that turns genes on and off in bacteria". When discussing the regulation of the trp operon, did you mention the promoter, operator, repressor protein (tryptophan repressor), negative control and the fact that binding of tryptophan to the tryptophan repressor alters its conformation and binding ability?

The regulation of the lac operon and the role of DNA looping are discussed in Alberts on pages 435-438. Make sure you have described negative and positive regulation and that you understand the difference between the two.

6.4.6 Regulation using transcription factors

As we discussed in learning unit 5 the bacterial RNA polymerase requires a sigma (σ) factor to facilitate its binding to the DNA promoter. Bacteria have a number of different σ factors they use to control the expression of groups of genes. Each σ factor recognises different groups of promoters; therefore if the cell produces lots of one type of σ subunit, that group of genes will be recognised and expressed. Under different conditions the same σ factor might not be produced and those genes will not be transcribed.

The regulation of eukaryotic genes is much more complex than the regulation of bacterial genes. Eukaryotic genes are generally controlled by many gene regulatory proteins that bind to sequences that may or may not be close to the promoter where transcription starts. The term "gene control region" is used to describe all the DNA involved in regulating and initiating transcription of a gene, and may be as large as 50 000 nucleotides in length. Eukaryotic activators and repressors act by a number of different mechanisms. Unlike for bacteria where genes occur in operons, each gene in eukaryotes is independently regulated.

Eukaryotes regulate the transcription of genes by

- controlling the assembly of general transcription factors
- altering chromatin structure, making certain parts of the DNA accessible or not
- using a mediator, which serves as an intermediary between gene regulatory proteins and RNA polymerase

Gene regulation in eukaryotes involves many more proteins than the bacterial examples. Eukaryotic RNA polymerase II needs a number of general transcription factors to function, compared to the bacterial RNA

polymerase which uses a single σ factor. Refer to figure 7-44 in Alberts showing the gene control region for a typical eukaryotic gene.

Eukaryotic gene activator proteins attract, position and modify the transcription factors, RNA polymerase II and mediator, controlling transcription. They can do this by acting directly with the components or by changing the chromatin structure.

Simple activator proteins have two distinct domains, a DNA-binding domain and an activation domain. Eukaryotes also have gene repressor proteins that are involved in regulating the transcription of genes. They act using a variety of mechanisms. We won't be discussing these in this module. If you wish to know more, you will find further information in the textbook.

6.4.7 Activity 6.4

Do the following activity and add it to your portfolio:

- a) What are the functions of the DNA-binding domain and an activation domain of an activator protein? (Refer to page 441 in Alberts, the section on eukaryotic gene activator proteins.)
- b) What happens if you take the DNA-binding domain of one protein and fuse it with the activation domain of another activator protein? (Refer to figure 7-45, page 442, in Alberts.)

6.4.8 Riboswitches

Riboswitches are short sequences of RNA that alter their folding pattern in response to the binding of an effector molecule. This conformational change can control gene expression by blocking the binding of the RNA polymerase or causing it to terminate transcription before the entire gene is transcribed. Riboswitches are very common in bacteria (refer to figure 7-93 in Alberts).

6.5 Post-transcriptional controls

Recommended reading: chapter 7, pages 477–486 in Alberts, the section on post-transcriptional controls

Eukaryotic post-transcriptional controls include

- RNA processing control (alternative splicing, change in site of RNA transcript cleavage and RNA editing)
- RNA transport and localisation
- rate of mRNA degradation
- translational control
- regulation of protein activity

RNA processing

Cells can splice RNA differently and thereby produce different mRNAs resulting in different polypeptide chains. It is estimated that 75% of the genes expressed in humans undergo alternative splicing. Therefore the regulation of RNA splicing can generate distinct (but related) proteins in different cell types according to the requirements of the cell.

The 3' end of a eukaryotic mRNA is generated by a cleavage reaction catalysed by different factors. RNAs cleaved at different sites will generate transcripts of different lengths, resulting in proteins with different C-terminal ends.

RNA editing is another regulatory method in which the nucleotide sequence of an RNA transcript is changed. One or more nucleotides may be added or removed or changed to different nucleotides, resulting in changes in the amino acid sequence of the translated protein.

RNA transport and localisation control

RNA molecules are only exported from the nucleus once processing is complete. Anything that changes the

regulation of nuclear transport of mRNA can be used to regulate gene expression. After export from the nucleus, some RNAs are directed to specific intracellular locations before translation starts. This is called RNA localisation and is also a potential point of regulation.

Rate of mRNA degradation

mRNAs can have different half-lives ranging from minutes to hours. The longer the RNA is able to persist in the cell before being degraded, the more protein will be produced from that RNA. Therefore changes in RNA stability can regulate the amount of protein synthesised.

Regulation by non-coding RNAs

MicroRNAs (miRNAs) are short RNA sequences that can base pair with complementary regions on mRNAs, regulating their stability and translation. They function in regulation of gene expression because they can block translation of the mRNA and the complexes are often actively degraded by the cell, thereby removing the mRNA.

Translational control

A number of steps in translation can be controlled. Obviously if the translation process is stopped, no protein will be produced. We won't discuss this in this module. If you wish to know more, read pages 488-491 in Alberts.

6.5.1 Activity 6.5

Do the following activity and add it to your portfolio:

- a) Explain and draw a diagram to show how alternative mRNA splicing can result in multiple unique polypeptides. (Refer to Alberts, page 479, the section "Alternative RNA splicing can produce different forms of a protein from the same gene".)
- b) Discuss how alternative RNA splicing can result in increased protein complexity of an organism.
- c) Discuss how changes in the site of RNA transcript cleavage can change the C-terminus of a protein. (Refer to pages 482-483, the section on "A change in the site of RNA transcript cleavage" in Alberts.)

Discussion forum question

This is a discussion activity that you should answer in the Discussions tool on the module website. If you don't have internet access, do the activity in writing and add it to your portfolio.

Do an internet search, or consult other relevant resources, on how problems with gene regulation may cause disease in humans. Use the article: Importance of gene regulation for common human disease (<http://www.sciencedaily.com/releases/2007/09/070916143515.htm>) as a starting point for your research.

Then go to the discussion topic entitled: Gene regulation and disease, and post some points that you find particularly interesting.

You may discuss with your peers different diseases that occur as a result of problems with gene regulation and how you think changes in gene regulation may have caused the illness.

Also read the other students' postings and respond to at least one other posting. Mention anything you found particularly interesting about the other posting, or ask for clarification or more details of any aspect if you wish.

Learning unit 7

The cell cycle

7.1 Introduction to learning unit 7

Cell division and death are important in the development and maintenance of multicellular organisms. Many of our tissues are constantly being replaced with new cells through cell division (e.g. skin, blood and intestinal lining). The process of cell division is controlled by the cell cycle and is balanced with cell death, ensuring that the total number of cells within the body does not change significantly. This cell death is a highly controlled process and is referred to as programmed cell death. Apoptosis is the most common form of programmed cell death.

To work through the learning unit, refer to sections of chapters 17 and 18 in Alberts. The following website has additional information on the cell cycle and mitosis:

<http://cnx.org/content/m46034/latest/?collection=col11496/1.6>

7.2 Learning outcomes

After completing this learning unit, you should be able to

- discuss the phases of the cell cycle
- describe the steps in the mitotic process
- discuss the morphological changes that occur in an apoptotic cell
- describe the role of caspases in apoptosis

7.3 Cell cycle overview

Recommended reading: chapter 17, pages 1053-1060 in Alberts

For organisms to grow and reproduce, cells need to be able to duplicate themselves. The cell cycle is the series of events that lead to a cell duplicating its contents and then dividing in two. In unicellular species (bacteria and yeasts) cell division produces a complete, new organism. In multicellular organisms complex division and differentiation are required for growth and reproduction.

In fully grown humans cell division replaces dead cells and millions of cells are duplicated on a daily basis. A number of steps are required to replicate a cell:

- The DNA of the cell needs to be duplicated.
- The macromolecules and organelles need to be duplicated.
- The DNA and other structures then need to be accurately separated into two daughter cells.

This is all brought about by the cell cycle control system, in which an ordered series of biochemical changes initiate the various steps of the cell cycle to coordinate the duplication, growth and division stages. When this system fails, excessive division can result in cancer.

The eukaryotic cell cycle is divided into four phases:

- G₁ phase (gap phase 1)
- S phase (DNA synthesis)
- G₂ phase (gap phase 2)
- M phase (mitosis occurs)

G₁, S and G₂ are referred to as interphase and the M phase incorporates mitosis and cytokinesis.

7.3.1 Activity 7.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

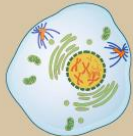
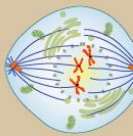
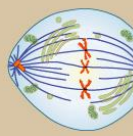
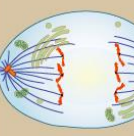
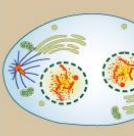
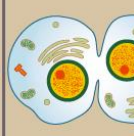
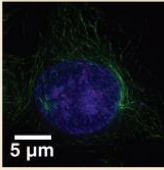
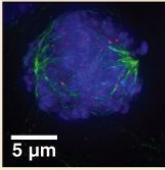
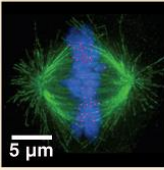
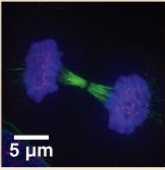
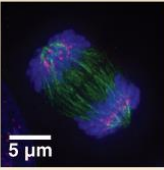

Read the overview of the cell cycle in the textbook, pages 1054-1055, and describe what happens in the four different phases of the cell cycle and how long they typically last. Make sure you understand what happens in the S and M phases and the purpose of the G₁, G₂ and the special G₀ phase.

7.4 Mitosis

Before mitosis, the cell replicates its DNA during the S phase. We discussed the molecular mechanisms involved in DNA duplication extensively in learning unit 4. At the end of S and G₂ phases, the entire DNA molecule in each chromosome has been duplicated along with the chromatin proteins that associate with the DNA. The next stage involves the actual division process and is called mitosis.

Mitosis is divided into five stages: prophase, prometaphase, metaphase, anaphase and telophase. Each stage can be easily distinguished under a microscope (refer to table 7.1 in this guide). After mitosis is complete, cytokinesis divides the cells in half, each containing an identical nucleus. Cytokinesis typically begins in anaphase and ends after telophase.

Table 7.1: Cell division by mitosis (OpenStax College, *Anatomy & physiology* (<http://cnx.org/content/m46034/latest/?collection=col11496/1.6>))

Prophase	Prometaphase	Metaphase	Anaphase	Telophase	Cytokinesis
					
<ul style="list-style-type: none"> Chromosomes condense and become visible Spindle fibers emerge from the centrosomes Nuclear envelope breaks down Centrosomes move toward opposite poles 	<ul style="list-style-type: none"> Chromosomes continue to condense Kinetochores appear at the centromeres Mitotic spindle microtubules attach to kinetochores 	<ul style="list-style-type: none"> Chromosomes are lined up at the metaphase plate Each sister chromatid is attached to a spindle fiber originating from opposite poles 	<ul style="list-style-type: none"> Centromeres split in two Sister chromatids (now called chromosomes) are pulled toward opposite poles Certain spindle fibers begin to elongate the cell 	<ul style="list-style-type: none"> Chromosomes arrive at opposite poles and begin to decondense Nuclear envelope material surrounds each set of chromosomes The mitotic spindle breaks down Spindle fibers continue to push poles apart 	<ul style="list-style-type: none"> Animal cells: a cleavage furrow separates the daughter cells Plant cells: a cell plate, the precursor to a new cell wall, separates the daughter cells
					
MITOSIS					

Meiosis is another form of division involved in reproduction and results in the formation of haploid gametes. We won't cover meiosis in this module. If you want to know more about it, you can find further information in your textbook.

7.4.1 Activity 7.2

Do the following activity and add it to your portfolio:

- a) In your own words, describe the different stages of mitosis (refer to table 7.1 in the study guide and panel 17-1 in Alberts).
- b) Define the following terms and describe how they are involved in mitosis:
 - spindle fibres
 - kinetochore
 - chromatid
 - contractile ring

7.5 Apoptosis

Recommended reading: chapter 18, pages 1115-1119 in Alberts

Apoptosis is essentially cell suicide by a sequence of biochemical events. It is especially important in embryogenesis where it helps to shape the developing organism by removing unwanted cells. For example, the separation of the fingers and toes in a developing human embryo is as a result of the cells between the digits undergoing apoptosis. Billions of cells die each day in the average human adult through apoptosis. This is especially evident in the bone marrow and intestines of humans. During apoptosis, apoptotic bodies are formed and are quickly phagocytised by macrophages. This process is quick and results in the contents of the cell being removed before they can spill onto the surrounding cells and cause an inflammatory response. Cells that die accidentally generally die by necrosis. Unlike apoptosis, necrosis is a disorganised disintegration of a damaged cell and triggers an inflammatory response.

Apoptosis is characterised by a number of morphological and biochemical changes. The morphological changes include

- cell shrinkage
- cytoskeleton collapse
- nuclear envelope disassembly
- chromatin condensation and fragmentation
- cell membrane blebbing and the formation of apoptotic bodies

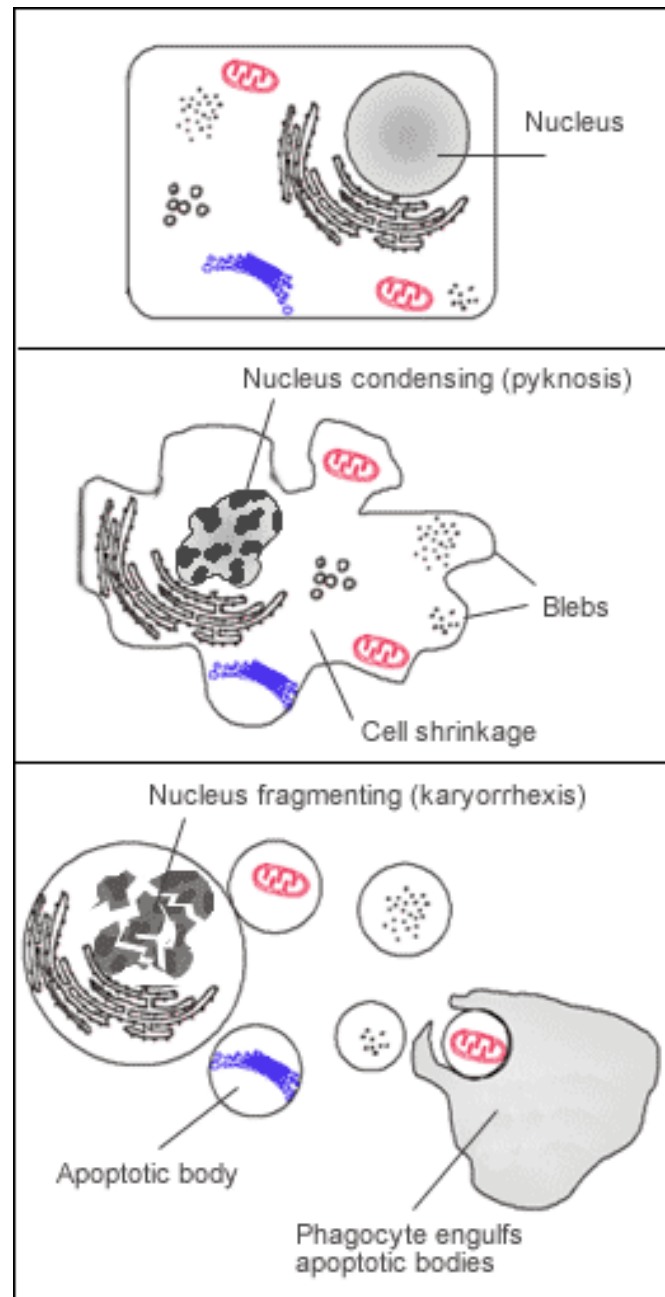


Figure7.1: A cell undergoing apoptosis (<http://en.wikipedia.org/wiki/File:Apoptosis.png>)

These morphological changes are brought about by several biochemical changes that occur within the cell. Some of these changes include

- cleavage of chromosomal DNA by endonucleases
- loss of the electric potential across the inner mitochondrial membrane
- movement of phosphatidylserine to the outer portion of the phospholipid membrane, which signals the macrophages to phagocytise the apoptotic bodies without initiating an inflammatory response

7.5.1 Caspases

The apoptotic pathways in cells are very complex but caspases play a central role in mediating the proteolytic cascade that results in apoptosis. Caspases are a family of proteases that have a cysteine at the active site and cleave target proteins at specific aspartic acid residues. They themselves require cleavage to become active and in their inactive forms are called procaspases. A cascade of about nine different caspases is

involved in the apoptotic process. Some are involved in initiating the apoptotic cascade and are called initiator procaspases. These then cleave downstream executioner procaspases, which act on further caspases or cell constituents such as DNA repair enzymes, cytoskeletal components and nuclear proteins. A DNAase is also activated, which fragments the genomic DNA. Once the cascade is started by the initiator procaspases, the process is irreversible.

There are two main signalling pathways that trigger the apoptotic caspases cascade in mammalian cells. They are the intrinsic pathway and extrinsic pathway (we won't cover these in this module). Inhibitors of apoptosis (IAPs) can bind and inhibit activated caspases and can therefore stop apoptosis. They are used by some viruses to prevent the host cell from killing itself. Excessive or inadequate apoptosis may cause disease. Many cancer cells do not regulate the apoptotic pathway in the normal manner and result in excessive cell growth.

7.5.2 Activity 7.3

Do the following activity and add it to your portfolio:

- a) Briefly discuss the role of apoptosis during animal development (pages 1116 and 1117 in Alberts).
- b) List a few of the proteins that are targeted by the caspases and relate this to the observed biochemical and morphological changes of apoptosis (pages 1118 and 1119 in Alberts).

7.5.3 Feedback on activity 7.3

After reading about apoptosis during animal development, you should understand how important this process is in the development of functional organisms.

The caspases system targets a number of proteins including

- nuclear laminins – result in the breakdown of the nuclear membrane
- exonuclease inhibitors – this allows the specific exonucleases to become active and cleave the DNA in the nucleus
- cytoskeletal components – cells round up
- cell-cell adhesion proteins – cells detach from their neighbours

You can clearly see that the proteins targeted by the caspases result in the observed biochemical and morphological changes associated with apoptosis.

Learning unit 8

Laboratory techniques in molecular biology

8.1 Introduction to learning unit 8

In the previous learning units we covered some of the theoretical aspects of molecular biology. Most of the research performed in molecular biology involves applying theoretical knowledge to ask a question about a biological system. This is then tested in a laboratory using different instruments and techniques. It is therefore important to understand what techniques are available and how these can be applied to study biological macromolecules.

In this learning unit we will show you some of the practical methods that are used in the laboratory to isolate, characterise and manipulate biological macromolecules, specifically what techniques can be used to characterise DNA, RNA and proteins. It is critical that you not only learn the various techniques, but also understand when they are used, and how they can be used in combination with one another. The terms *in vitro* and *in vivo* are used to describe whether an experiment has taken place in an organism (*in vivo*) or in a test tube (*in vitro*).

To work through the learning unit, refer to chapter 8 in Alberts which deals with the techniques used to manipulate DNA, RNA and proteins.

8.2 Learning outcomes

After completing this learning unit, you should be able to

- discuss the isolation and growth of cells in culture
- discuss how proteins can be purified using chromatography
- describe various techniques that can be used to analyse proteins
- describe what restriction endonucleases are
- describe how DNA can be amplified by polymerase chain reaction (PCR)
- explain how to clone DNA and create a genomic or cDNA library
- describe various hybridisation techniques such as Northern and Southern blotting

8.3 Isolating and growing cells in culture

Recommended reading: chapter 8, pages 501-510 in Alberts



Figure 8.1: Tissue culture flasks (http://en.wikipedia.org/wiki/File:Tissue_culture_vials_nci-vol-2142-300.jpg)

Most biological research requires sizeable amounts of biological material isolated from large numbers of cells. Cells extracted directly from tissues provide the most realistic source of material but tissues are heterogeneous and are made up of different cells with different internal processes. To allow biologists to study a specific cell type, the different cells need to be separated according to their cell type or grown up as a uniform population.

It is not always feasible to harvest cells directly from tissues; a more common method is to grow cells up in culture. A cell culture is derived from a single cell and provides a homogeneous population of cells from which cellular material can be extracted. Most plant, animal and bacterial cells can be manipulated to grow in the laboratory in culture media in the correct conditions. They can then be observed microscopically or analysed biochemically. Cultures prepared directly from the tissues of an organism are called primary cultures. Once the cells have been subcultured, they are referred to as secondary cultures. Most vertebrate cells cannot divide indefinitely and will die in culture owing to replicative cell senescence. Various techniques can be used to modify cells to make them immortalised. Once immortalised, they will divide indefinitely in culture. Cell lines derived from cancer cells are referred to as transformed cell lines.

8.3.1 Activity 8.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- Refer to the section “Cells can be isolated from intact tissues”, on page 502 in Alberts, and write a paragraph in your own words describing how cells may be isolated from tissues.
- In regard to the culturing of cells, discuss transformed cell lines, the action of telomerase and inactivation of cell cycle checkpoints (refer to Alberts, page 505).
- Read up on how embryonic stem cells may be used in medicine to replace and repair damaged tissues that are causing disease (see Alberts, pages 505-506).

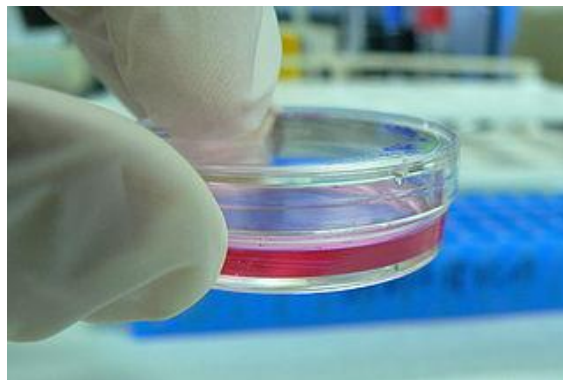


Figure 8.2: Culture in a petri dish (http://en.wikipedia.org/wiki/File:Cell_Culture_in_a_tiny_Petri_dish.jpg)

8.3.2 Culturing bacterial cells

Many bacteria are easily grown in a laboratory either on a solid media, for example an agar plate, or in liquid culture in culture flasks. Have a look at these images.



Figure 8.3: The bacterium *Klebsiella pneumoniae* grown on an agar plate (http://commons.wikimedia.org/wiki/File:K_pneumoniae_M_morganii_providencia_styphimuriuma.JPG)

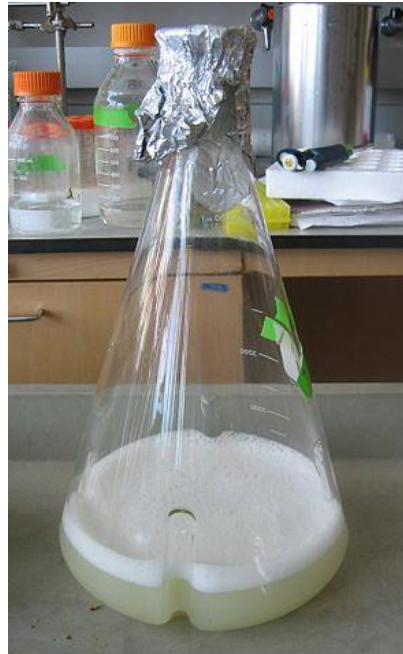


Figure 8.4: Bacteria grown in liquid culture (<http://commons.wikimedia.org/wiki/File:Shakeflask.JPG>)

The media contains all the nutrients the bacteria require to grow and divide. Bacteria can be used to produce large quantities of specific DNA or a protein of interest. The expression of proteins is accomplished by inserting a recombinant gene into what is called an expression vector which can be introduced into a bacterium. This gene will then be transcribed into RNA and translated into protein. The advantage of this system is that large amounts of protein can easily be produced and purified from the rest of the bacteria cell contents.

8.4 Manipulation of DNA

Recommended reading: chapter 8, pages 532-552 in Alberts, the section entitled “Analysing and manipulating DNA”

DNA can easily be isolated, manipulated and sequenced. It is also relatively simple to make multiple copies

of any region of a genome by amplifying the DNA. Using various techniques, genes can be altered (engineered) and reintroduced into an organism, changing the heritable part of the genome. Genes may also be introduced into bacteria or other organisms where the gene products can be expressed and studied. There are a number of techniques that are very important in genetic engineering:

- cleavage of DNA at specific sites by enzymes called restriction endonucleases
- joining of DNA fragments in a process called ligation
- DNA separation by gel electrophoresis
- replication of DNA using either cloning vectors or polymerase chain reaction (PCR) to generate billions of copies of a DNA molecule
- nucleic acid hybridisation which can be used to find specific sequences of DNA or RNA
- sequencing of DNA fragments or entire genomes
- simultaneous monitoring of mRNA levels of different genes within the cell using nucleic acid microarrays

We will describe these techniques in more detail in the following section. Refer to the textbook where many of the processes are well described.

8.4.1 Gel electrophoresis

Gel electrophoresis can be used to separate DNA or RNA molecules of different sizes. Agarose is generally used for the gel, except for DNA less than 500 nucleotides in length. In that case, polyacrylamide gels are usually used. This technique is used extensively in DNA work and is easy to perform. During electrophoresis DNA is loaded into precast wells in an agarose gel (refer to figures 8.5 and 8.6 in this guide) which have been placed in an electrophoresis tank with buffer.

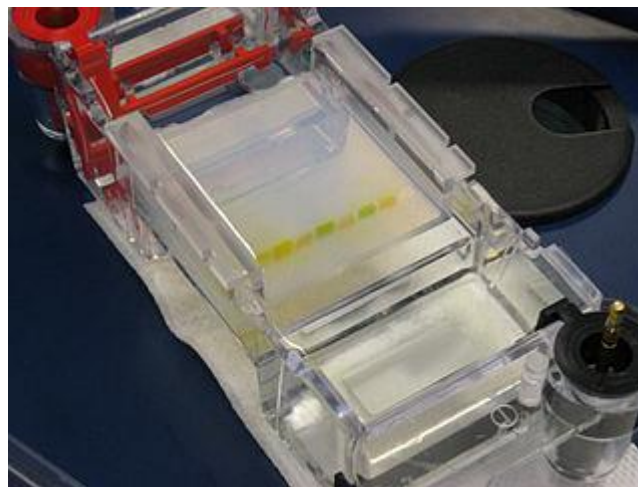


Figure 8.5: Agarose gel in an electrophoresis tank showing filled precast wells (http://commons.wikimedia.org/wiki/File:Agarose_gel_electrophoresis.jpg)

A current is then applied, as the DNA is negatively charged at a neutral pH (the buffer will have a neutral pH) and it will migrate towards the positive electrode through the gel. DNA has a uniform charge to mass ratio and the fragments will separate according to size. The rate of migration is determined by the

- size of the DNA molecule
- concentration of the agarose
- DNA conformation
- strength of the electric field applied (voltage used)
- composition of the electrophoresis buffer

DNA cannot be seen directly, so a stain needs to be added to see it. Ethidium bromide intercalates between

the bases of DNA and RNA and is the most commonly used stain for DNA and RNA gels. Ethidium bromide fluoresces under ultraviolet light when it is bound to DNA. The DNA can then be seen when the gel is placed on an ultraviolet transilluminator. Figure 8.6 shows a typical DNA gel where each band represents a DNA fragment. The ladders on the sides are the molecular weight markers with known sizes and can be used to estimate the sizes of the other bands.

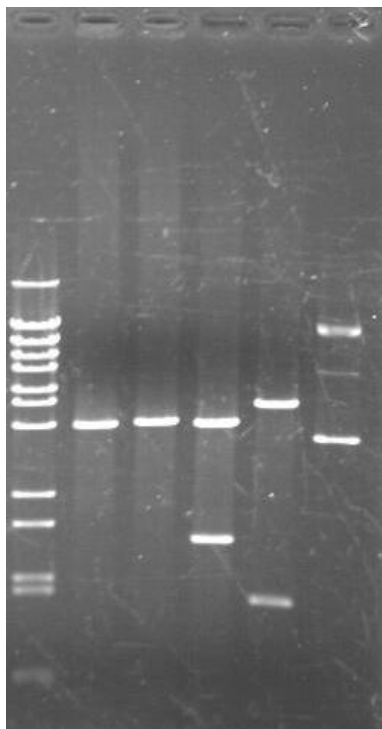


Figure 8.6: Typical DNA gel after electrophoresis and staining with ethidium bromide. The DNA ladder can be seen on the left-hand side. (<http://commons.wikimedia.org/wiki/File:Miniprep-restricci%C3%B3n.JPG>)

8.4.2 Activity 8.2

Do the following activity and add it to your portfolio:

- In your own words, explain the principle of how gel electrophoresis of DNA separates the molecules according to size.
- Will the DNA move towards the anode or cathode? Why?
- What is pulse field gel electrophoresis? When is it used?

8.4.3 Feedback on activity 8.2

Pulse field gel electrophoresis is discussed in the textbook on page 534.

8.4.4 DNA digestion by restriction endonucleases

A number of enzymes are used to manipulate DNA. Some of the most useful are the restriction endonucleases. These enzymes recognise specific short nucleotide sequences usually between four and eight nucleotides in length, and cleave the DNA at a specific site. Different species of bacteria produce different restriction endonucleases that recognise different DNA sequences. They are named after the bacterium from which they are isolated, for example *Bam*HI is from *Bacillus amyloliquefaciens*. The first three letters in the restriction enzyme name consist of the first letter of the genus (B), the first letters of the species (am), possibly follow by a strain designation if there is one (H) and roman numeral (I, II or higher) to indicate

the order of discovery. *Bam*HI will always recognise the same sequence (GGATCC). The enzymes cut such that blunt ends or overlapping (sticky) ends are produced (refer to figure 8-31 in Alberts). DNA fragments produced from enzyme digestion can be analysed or isolated by gel electrophoresis. The fragments of DNA can be joined using another enzyme called DNA ligase.



Figure 8.7: *Eco*RI restriction enzyme recognition site

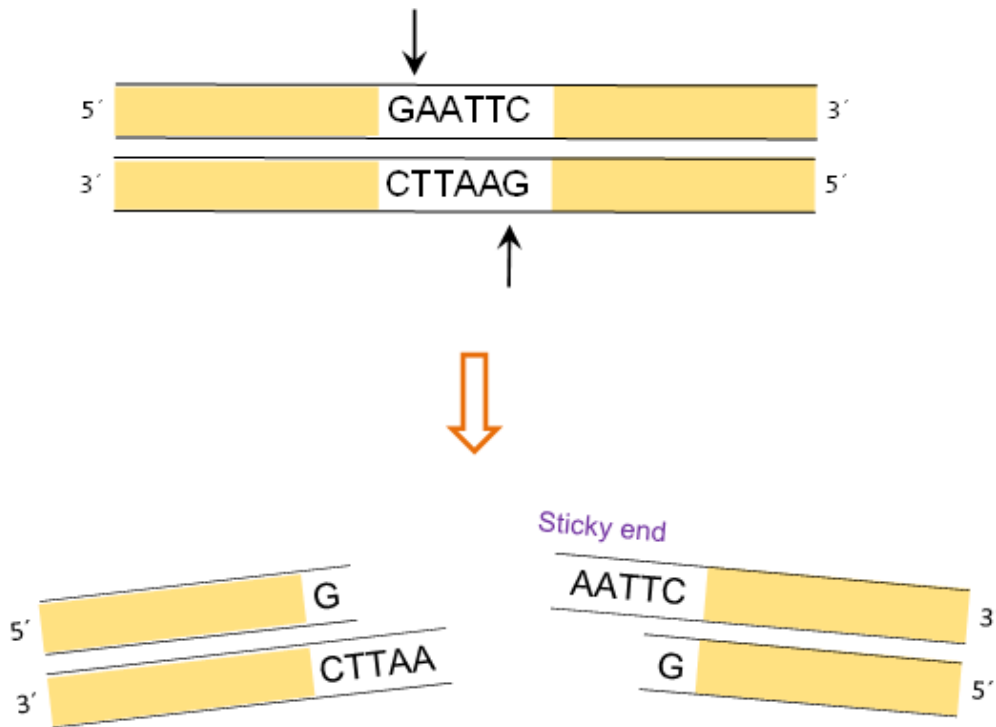


Figure 8.8: Digestion of DNA with *Eco*RI. The *Eco*RI cuts between the G and the A on each strand of DNA, within the recognition sequence GAATTC, leaving single-stranded DNA overhangs called sticky ends.

8.4.5 Activity 8.3

Do the following activity and add it to your portfolio:

Given the DNA sequence AGCATCGGATTCATTCTAGTCGGTAAG, do the following:

- a) Draw the double-stranded DNA sequence using your knowledge of base complementarity.
- b) Label the 5' and 3' ends of the strands (they should be antiparallel).
- c) On your diagram indicate where the restriction enzyme *Bam*HI will cleave the DNA.

*Bam*HI recognition sequence and the position of the cuts:



- d) Draw the two separate fragments of DNA, showing the sticky ends.

8.4.6 Ligation

DNA ligases catalyse formation of a phosphodiester bond between the 5' phosphate of a nucleotide on one fragment of DNA and the 3' hydroxyl of another. The joining of DNA fragments together is called ligation. Because DNA ligase can join two pieces of DNA, it is used extensively in genetic engineering. The DNA ligase most widely used in the lab is derived from the bacteriophage T4.

If restriction-digested fragments of DNA are placed together under appropriate conditions, the DNA fragments from two sources with the same sticky ends will anneal. However, the two strands are not covalently bonded by phosphodiester bonds and DNA ligase is required to covalently bond the two strands. T4 DNA ligase will also ligate fragments with blunt ends, but the reaction is less efficient and higher concentrations of the enzyme are usually required.

8.4.7 DNA amplification

Multiple copies of specific regions of DNA can be produced in two ways:

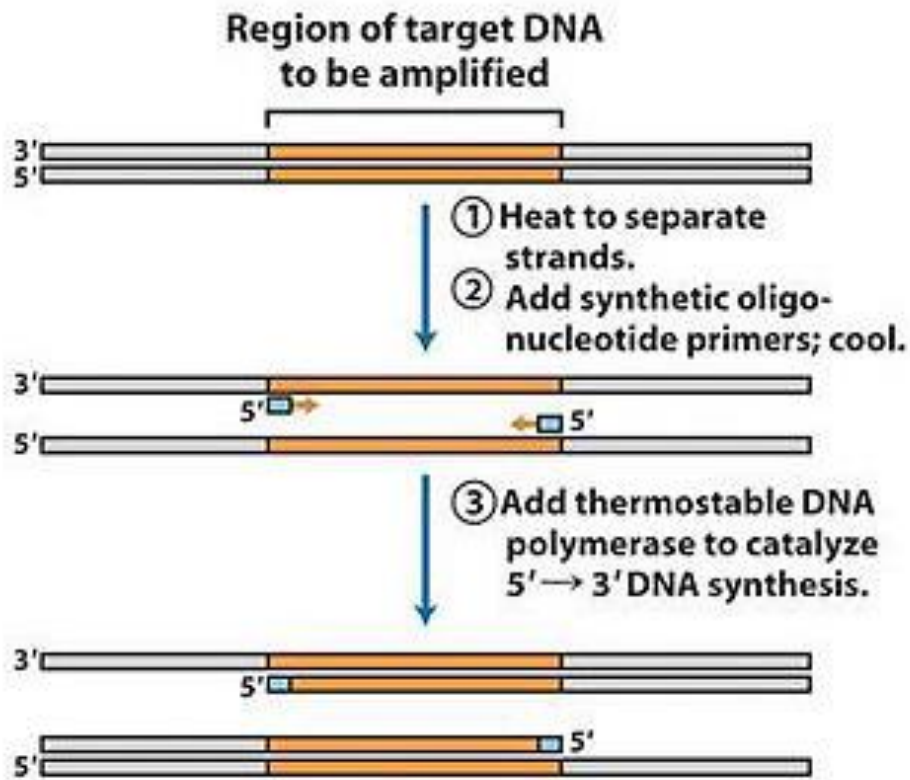
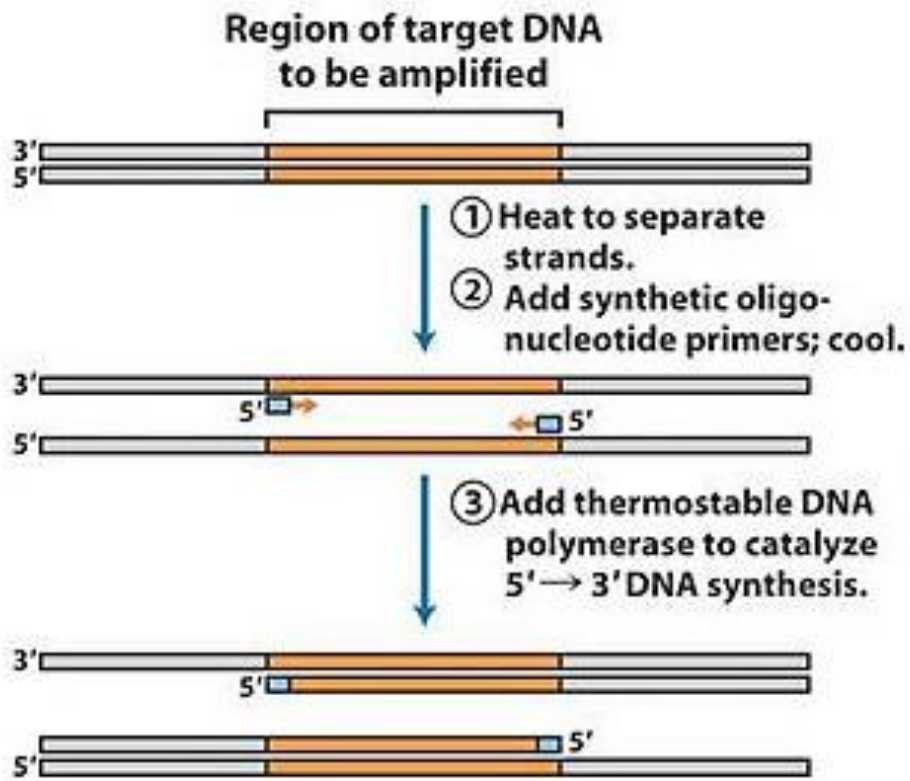
- through cloning into a DNA vector, which is then replicated within a host cell
- through the use of polymerase chain reaction (PCR) where multiple copies of a selected region of DNA sequence are generated using a DNA polymerase

In the PCR process the temperature is raised to 95 °C and therefore the DNA polymerase used is *Taq* DNA polymerase as it is able to withstand high temperatures that would normally denature and render inactive other DNA polymerases. It is named after the thermophilic bacterium *Thermus aquaticus* from which it was originally isolated. This bacterium lives in hot springs and thermal vents and needs to be able to replicate its DNA at high temperatures.

8.4.7.1 Polymerase chain reaction (PCR)

Refer to figure 8.9 in the study guide and figure 8-45 in Alberts illustrating the PCR process. In the first stage two synthetic DNA primers (short DNA oligonucleotides) complementary to sequences flanking the gene of interest are synthesised. One primer is complementary to a sequence on one strand of the double helix and the other is complementary to a sequence on the other strand, at opposite ends of the region of interest.

In the first step of the PCR process the DNA is heated to about 95 °C. This separates the strands of DNA. The process of separating the strands is called denaturation. The temperature is then dropped, allowing the primers to anneal (bind) to the complementary sequences within the DNA. The temperature is then raised slightly in the presence of DNA polymerase and dNTPs and the primers are extended from their 3' ends. This step is called elongation. The cycle is repeated 20-30 times, and each time more DNA is synthesised. The newly synthesised strands also act as templates and within a few cycles there are thousands of copies of the region of interest. Only the sequence flanked by the two primers is amplified because DNA polymerase requires a free 3' OH to synthesise DNA, and the primers will not bind anywhere else, resulting in all the products being the same length. The resulting fragments can be seen by running them on an agarose gel.



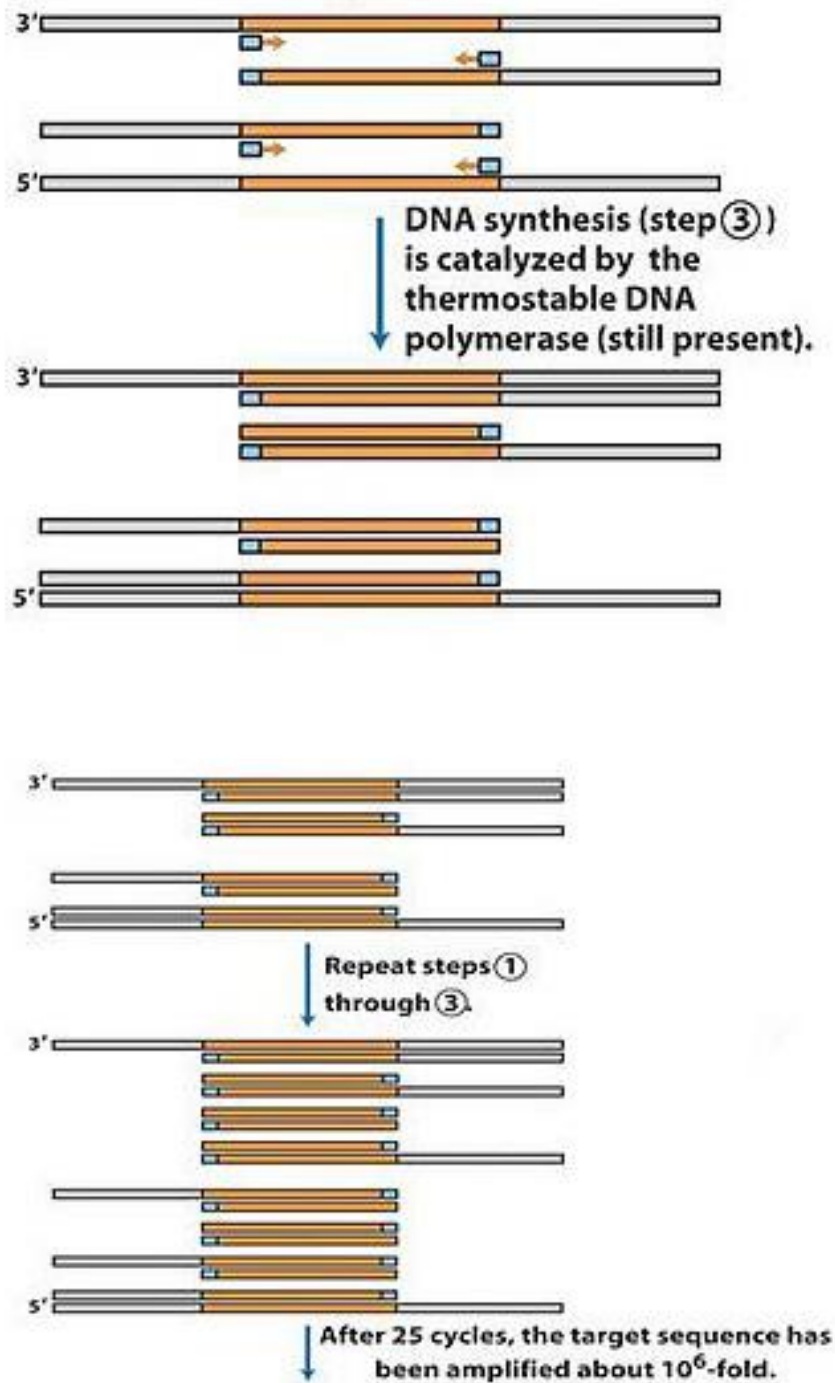


Figure 8.9: Amplification of DNA during PCR (<http://commons.wikimedia.org/wiki/File:PCR.jpg>)

A standard PCR reaction requires the following:

- DNA template
- primers
- DNA polymerase
- dNTPs
- buffer

PCR can also be used in forensic analysis to determine if a specific sequence of DNA is present in a sample. For example, if you have a sample of blood of unknown origin, it can be added to a PCR reaction with primers that flank a specific gene. If the DNA is amplified, then you can say that the DNA contains that gene. If you do this for a number of genes, you can generate what is called a genetic "fingerprint" that can be used to identify exactly whose blood it is. Refer to figure 8-47 in Alberts.

8.4.8 Activity 8.4

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- a) What is DNA denaturation?
- b) In your own words, describe how specific regions of DNA can be amplified by PCR.
- c) How can PCR be used in forensic analysis?

8.4.9 Feedback on activity 8.4

PCR is one of the most widely used techniques in a molecular biology laboratory. It is described in the textbook on pages 544-546 – read the section carefully. It is essential that you understand and can explain what PCR is and how it is performed. Read back over your answer to question (b) and make sure you have answered the question properly.

The use of PCR in forensic science is illustrated in figure 8-47.

8.4.9.1 Vector cloning

Cloning of DNA refers to the synthesis of many copies of DNA. One means of preparing multiple copies of DNA is to insert a DNA fragment into a cloning vector. This is then replicated by the host cell. There are a number of different cloning vectors, bacterial plasmids, phages, or cosmids, which then replicate in a host cell; thus the DNA is amplified. In this study guide we will discuss bacterial plasmid cloning vectors. The principles are similar for other types of cloning vectors.

Plasmids are small circular DNA molecules that occur as single or multiple copies in bacterial cells. A number of naturally occurring plasmids have been modified to be used in the laboratory as cloning vectors. All plasmid cloning vectors have certain features in common: an origin of replication, a multiple cloning site (MCS) and one or more selective markers. The origin of replication allows the plasmids to replicate independently within a specific bacterium. The MCS is a region of the plasmid that contains a number of restriction sites that can be used to insert the DNA of interest. The selective marker is used to select for bacteria that contain the plasmid and are typically genes that confer antibiotic resistance. The DNA vector is introduced into the bacterial cell in a process called transformation.

DNA is a hydrophilic molecule and as a result will not normally pass through a bacterial cell's membrane. Therefore, it is essential to first make the bacteria "competent" to take up DNA. Competent bacterial cells can then be transformed by heat shock. Electroporation is another method that can be used to transform bacteria.

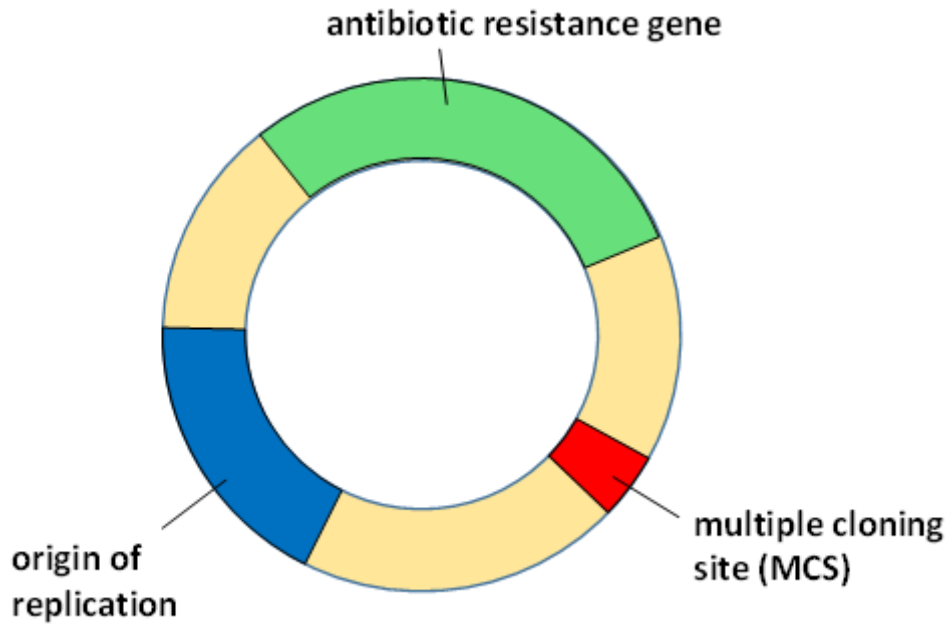


Figure 8.10: Schematic diagram of a plasmid vector

Culturing of a transformed bacterial cell will give rise to millions of identical bacterial cells, from which millions of copies of the recombinant DNA molecule can be isolated. Refer to figure 8-39 in Alberts which shows the insertion of a DNA fragment into a bacterial plasmid.

8.4.10 DNA libraries

A DNA library is basically a collection of cloned DNA fragments from a cell or tissue of an organism. The goal is to have each gene of the organism represented at least once in the library. Libraries can be created using different vectors which are generally kept in bacterial cells. When a particular DNA fragment is needed, the specific bacterial clone is cultured and the DNA can be isolated.

During the preparation of a DNA library, the DNA of the organism being used in the construction of the library is cut with a restriction enzyme. At the same time the library plasmids are cut by the same restriction endonuclease. The two cut samples are then added together and the DNA is allowed to anneal. DNA ligase is added to join the fragments and the recombinant plasmids can then be introduced into bacterial cells. Refer to figure 8.9 here showing the construction of a DNA library. The transformed bacteria are spread on agar plates containing antibiotic. This is the antibiotic that the plasmids confer resistance to and, therefore, only the bacteria that have taken up plasmids will grow. Each separate bacterial clone should contain a plasmid with a different fragment of DNA from the organism being used in the construction of the library. The different clones grow as colonies on the plate and the bacterial clones can be stored for later use.

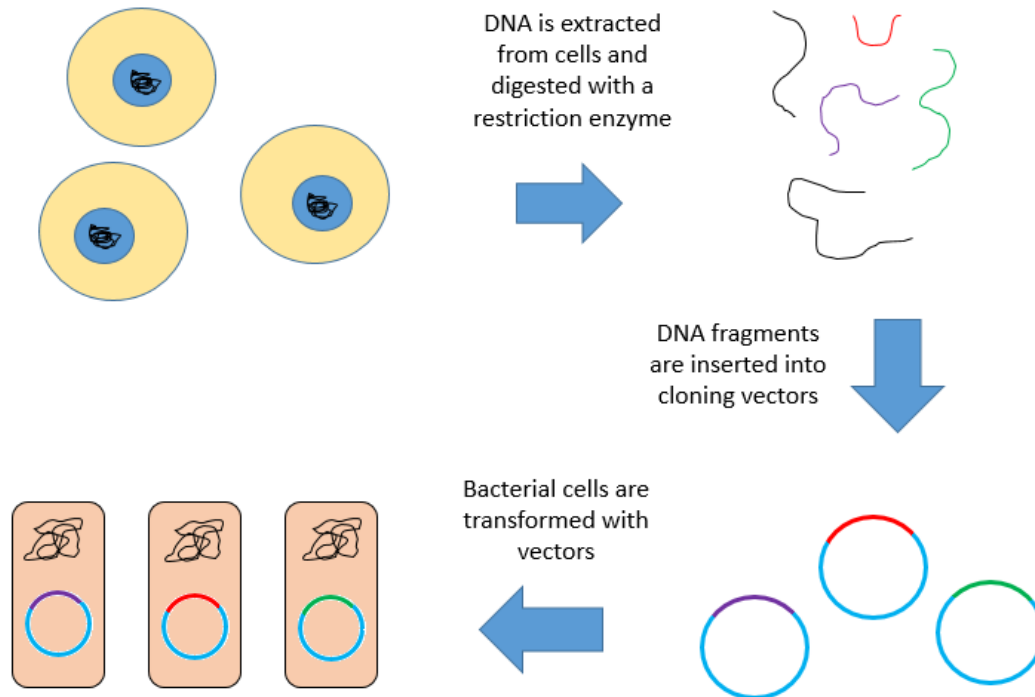


Figure 8.11: Construction of a DNA library
 (http://en.wikipedia.org/wiki/File:Genomic_Library_Construction.png), image credit: Aluquette

There are two main types of libraries:

- A genomic library is created from the genomic DNA of a cell line or tissue.
- A cDNA library is constructed from mRNA and therefore only contains the protein-coding genes of the organism.

During the generation of the cDNA library the mRNA is first extracted and reverse transcribed into complementary DNA or cDNA (refer to figure 8-43 in Alberts). The cDNA is then inserted into cloning vectors.

cDNA libraries contain only the regions of DNA that have been transcribed into mRNA at the time in which the RNA was isolated. Therefore, different cDNA libraries will be generated for each different tissue type of an organism, as each tissue type will express different genes. During the production of mRNA the introns are spliced out, and therefore the cDNA clones, which are produced from mRNA, contain the uninterrupted coding sequences of a gene. The cDNA library will also include alternatively spliced mRNAs that are produced within a tissue.

Most general cloning plasmids can only accept inserted DNA fragments that are less than 15 kbs in size. Bacterial artificial chromosomes (BACs) can accept fragments of up to a million nucleotides in length and are the preferred vector when making DNA libraries of organisms with large genomes.

8.4.11 Activity 8.5

Do the following activity and add it to your portfolio:

- In your own words, describe the steps involved in the generation of a DNA library.
- Discuss the difference between cDNA clones and genomic DNA clones that have been derived from the same region of DNA (refer to figure 8-44 in Alberts).
- Discuss how BACs differ from plasmids in terms of copy number, stability and the size of the insert they can accommodate.

8.4.12 Nucleic acid hybridisation and DNA sequencing

Nucleic acid hybridisation can be used to detect specific nucleotide sequences in both DNA and RNA. Hybridisation reactions rely on the fact that two single-stranded nucleic acid chains that have complementary nucleotide sequences can base pair. The probe is the single-stranded DNA sequence that is complementary to the specific DNA being investigated.

DNA can be sequenced rapidly by automated processes. This has resulted in many different genomes being completely sequenced. These DNA sequences can be used to predict the amino acid sequences of proteins. Comparisons of the genome sequences of different organisms can also be used to trace the evolutionary relationships between organisms. A common sequencing process is the dideoxy method.

8.4.13 Activity 8.6

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

- a) Discuss the process of northern and Southern blotting of nucleic acids and how a specific DNA or RNA molecule can be detected using gel-transfer hybridisation (refer to the section “Northern and Southern blotting facilitate hybridization with electrophoretically separated nucleic acid molecules”, pages 538-539 in Alberts).
- b) Briefly describe the dideoxy method of DNA sequencing (refer to figure 8-50 and the associated text, page 549 in Alberts).

8.4.14 Microarrays

A DNA microarray is essentially an assortment of microscopic DNA spots attached to a solid surface. Each DNA spot contains an oligonucleotide with a specific DNA sequence and is known as the probe. In a gene expression study, mRNA from cells is converted to cDNA and labelled with a fluorescent dye. The microarray is then incubated with the cDNA to allow hybridisation to take place. Unbound cDNA is washed off and the spots where hybridisation has occurred are detected, identifying the genes that were being expressed.

DNA microarrays can be used to monitor the expression of thousands of genes simultaneously by detecting the RNA products that are produced. This can allow us to see which genes are switched on and off when cells are undergoing the normal processes of growing, dividing and differentiating. They can also be used to analyse the gene expression profile of cells that have been exposed to different chemicals (hormones and toxins) or biological agents. Figure 8.12 in this study guide shows a large microarray with over 37 500 probes on it.

Some specific examples of the use of microarrays include

- changes in the gene expression profiles of uninfected as opposed to infected tissues in response to exposure by a specific pathogen
- gene expression profiling of the effects of specific drugs used in treatment on the expression profile of a tissue or cell line
- detection of single nucleotide polymorphisms (SNPs); this information can be used in genotyping, prediction of predisposition to diseases etc.
- identification of pathogens by hybridising the DNA from an infected tissue to an array containing DNA probes from a number of different species of pathogens

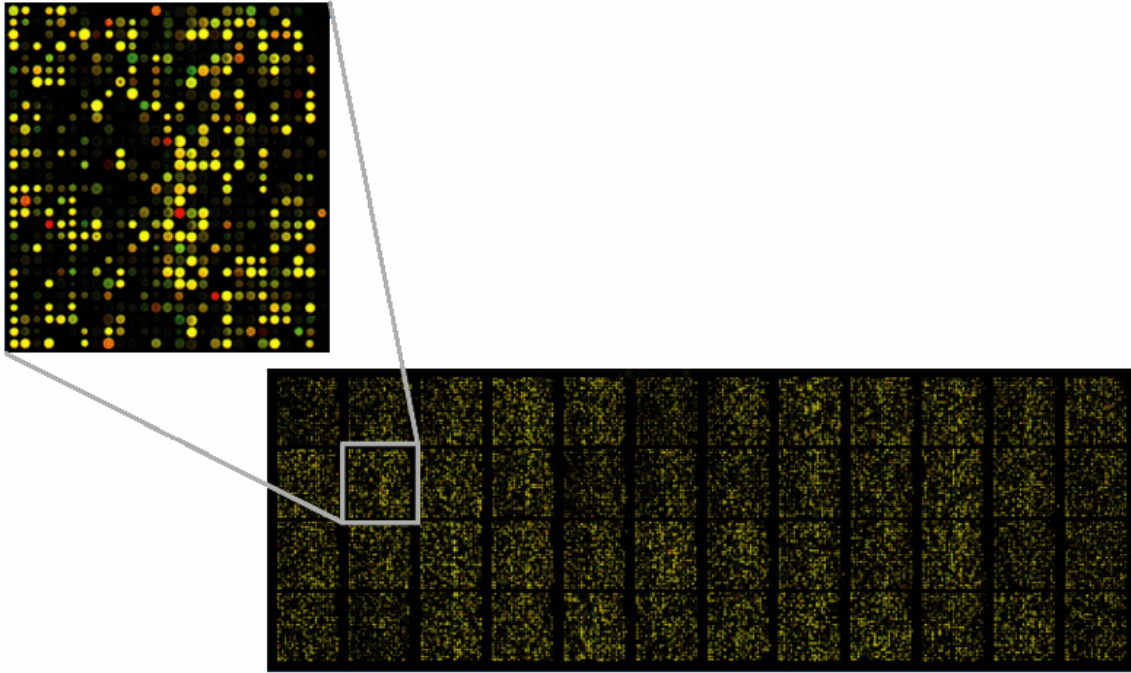


Figure 8.12: Example of an oligo microarray spotted with approximately 37 500 probes with enlarged inset to show detail (<http://en.wikipedia.org/wiki/File:Microarray2.gif>)

8.4.15 Activity 8.7

Do the following activity and add it to your portfolio:

- Describe what a microarray is and how it is made. (Refer to pages 574-575, the section on microarrays and figure 8-73 in Alberts.)
- Discuss how a microarray can be used to monitor levels of gene expression.
- In your own words, describe what genotyping and single nucleotide polymorphisms are.

8.5 Genetic markers in the identification of disease

Genetic engineering has been essential in the development of a number of medical applications, including the production of vaccines, human proteins (e.g. insulin), diagnosis of existing diseases, gene therapy and the prediction of the risk of developing a specific disease.

Sickle cell anaemia or sickle cell disease is a hereditary blood disorder, where the red blood cells assume an abnormal, sickle shape. Individuals with two copies of the defective gene (homozygotes, SS) have the disease. Individuals with one copy of the defective gene (heterozygotes, AS) have both normal and abnormal haemoglobin and therefore normal and abnormal red blood cells, and the condition is referred to as sickle cell trait.

Sickle cell disease is an example of a genetic disease that is due to a single point mutation (single nucleotide polymorphism). There is a single base change, a T-to-A DNA substitution, that results in a change in the sixth codon of β -globin gene, and a glutamic acid is changed to a valine, resulting in a structural change in the β -globin protein.

The point mutation in the DNA results in the loss of the recognition site for the restriction enzyme *Mst*II (refer to figure 8.13 in this guide). Therefore the DNA of normal, heterozygous and homozygous individuals cut by *Mst*II will result in different size fragments and different banding patterns on a gel. The position of the fragments can be detected by Southern blot using labelled β -globin cDNA as a probe (remember there will be lots of fragments from the rest of the genome). Because digestion with a restriction enzyme will result in different size fragments for people with the normal and sickle cell genes, it is known as a restriction fragment

length polymorphism (RFLP). This illustrates how a disease can be screened for using the techniques we discussed in this study guide. This type of diagnosis may be carried out prenatally to determine if a foetus has the disease.

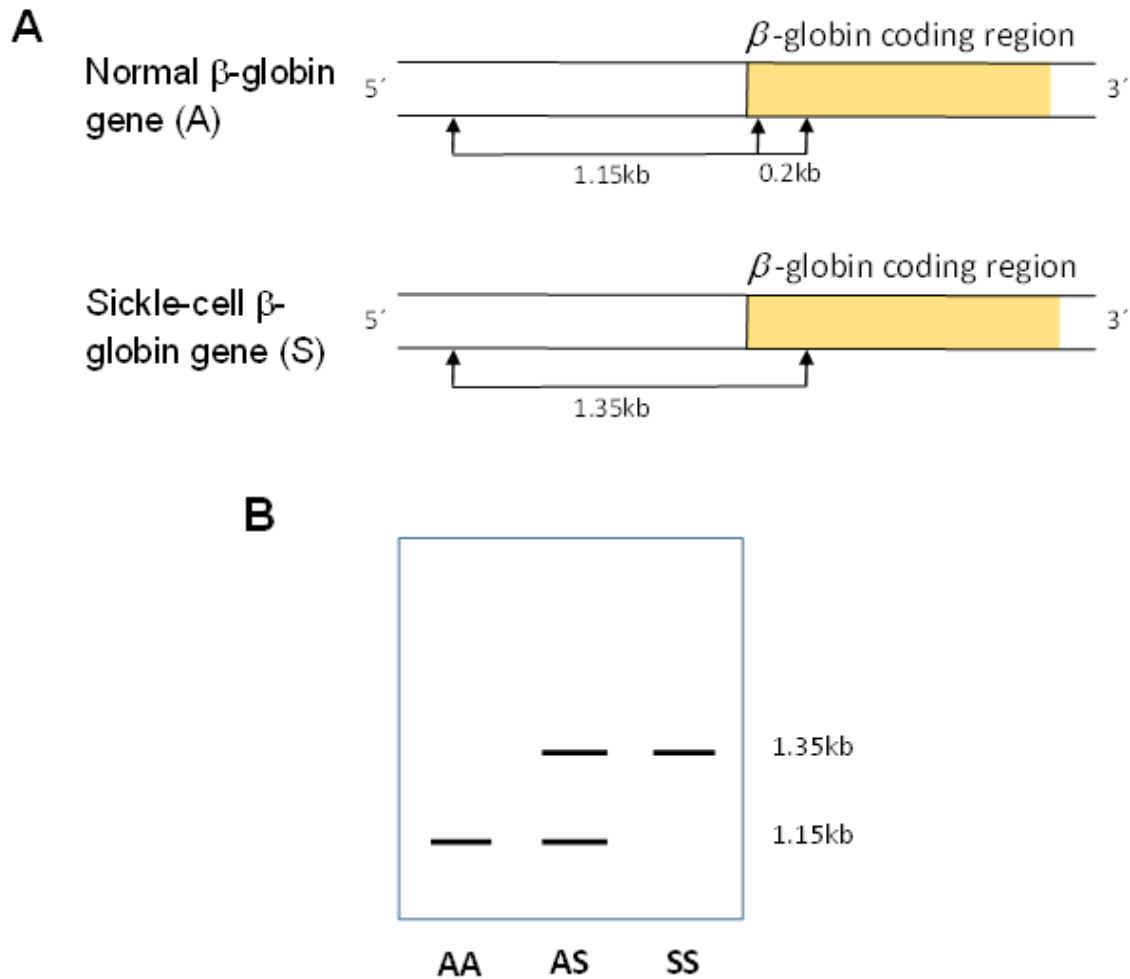


Figure 8.13: The top part of figure (A) shows the start of the β -globin gene and *MstII* restriction sites in the normal (A) and sickle cell (S) β -globin genes. Digestion with *MstII* results in fragments of 0.2 kb and 1.15 kb length in people with the normal genes and a single restriction fragment of 1.35 kb length in people with sickle cell disease. The bottom part of the figure (B) shows the expected sizes of fragments after digestion with *MstII* for individuals who are normal (AA), heterozygous (AS) and homozygous (SS) for the sickle cell gene. The difference in size can be detected by Southern blot using labelled β -globin cDNA as a probe. The 0.2 kb fragment would have run off the gel.

8.5.1 Activity 8.8

Do the following activity and add it to your portfolio:

- What do homozygous and heterozygous mean?
- In your own words, describe what a restriction fragment length polymorphism is.

Forum 1: Student Lounge

Use this forum to discuss general matters among yourselves

Discussion 1: Introduce yourself

Use this space to get to know your fellow classmates.

Tell one another about your current work situation, professional background and something personal (\pm 250 words).

Discussion 2: Fellow student contact details

Use this space to share your contact details with your classmates and to form study groups.

Forum 2: Learning units

Use this forum to discuss specific topics you are trying to understand in the course

Discussion 1: Learning unit 1

Discussion 2: Learning unit 2

Discussion 3: Learning unit 3

Discussion 4: Learning unit 4

Discussion 5: Learning unit 5

Discussion 6: Learning unit 6

Discussion 7: Learning unit 7

Discussion 8: Learning unit 8

Announcement 1:

Welcome and getting started

Dear Student

Welcome to Molecular Biology. You should have received a Getting Started letter in the mail explaining what we expect of you as an online student. If you have not received this letter, click on this link to read it.

Getting Started Letter

But for now, we would like you to first go to the **Discussion Forums** link on the left side of your screen and access Forum 1: Student Café. In Discussion 1 we want you to participate in your first online activity where you need to introduce yourself to your fellow students. Please participate in this discussion during February/July.

We are looking forward to meeting you online!

Your lecturers