7. Nucleic acids

7.1 DNA structure and replication

The structure of DNA is ideally suited to its function

In the first half of the 20th century, there was a debate about which molecule would carry genetic information. Two candidates were in the running: proteins and DNA. Proteins are built from amino acids and it was known there are 20 different amino acids. DNA is built from nucleotides and there are 4 different ones involved in DNA. Since it was believed that genetic information would need very complex molecules, it was thought that protein was the most likely molecule to carry genetic information.

However in 1944 *Avery*, *MacLeod* and *McCarty* showed that non-lethal *Pneumococcus* bacteria, when mixed with dead bacteria of the lethal strain, becomes lethal. They investigated the "transforming factor" and found it to be DNA, suggesting that this molecule carries genetic information.

Around that time, the structure of the T2 virus had been viewed with an electron microscope. See *Figure 701*.
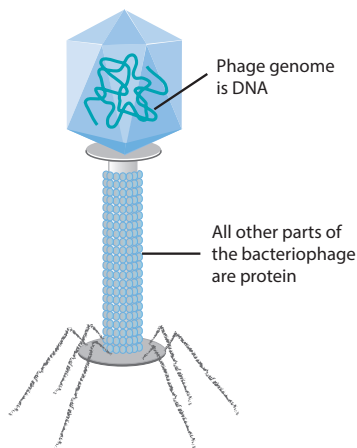

Phage genome is DNA

All other parts of the bacteriophage are protein

Figure 701 The structure of a T2 virus

The Hershey and Chase experiment

It was known that a T2 virus is a bacteriophage (or phage): a virus which has a bacterium as its host. This virus consists of DNA and a protein coat. It was also known that DNA contains phosphorus but never sulfur while proteins contain sulfur but never phosphorus.

Hershey and *Chase* (in 1952) used this information and labelled two batches of bacteriophages. Batch 1 was grown in the presence of radioactive sulfur so the phages' protein coats were radioactive. Batch 2 was grown in the presence of radioactive phosphorus so they contained radioactive DNA.

Hershey and Chase (cont.)

The phages of each batch were allowed to infect their bacterial hosts and go through their life cycle where the host cell is used to produce thousands of copies of the phage (both the DNA and protein component). See *Figure 702*.
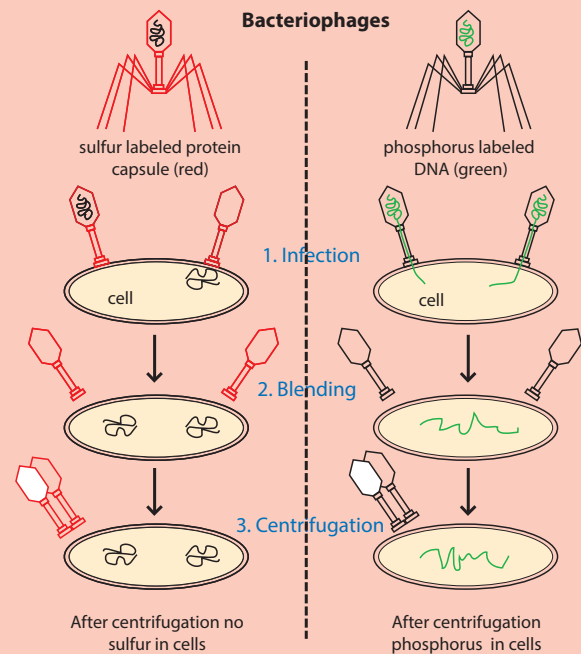

Bacteriophages

sulfur labeled protein capsule (red)

phosphorus labeled DNA (green)

1. Infection

cell

cell

2. Blending

3. Centrifugation

After centrifugation no sulfur in cells

After centrifugation phosphorus in cells

Figure 702 Hershey and Chase's experiment

The phages were given sufficient time to infect the bacteria and have the bacterial cell produce copies of the phage. Before the bacterial cell would burst, *Hershey* and *Chase* would vigorously mix the cells which separated the (empty) protein coats from the bacterial cell which contained their new phages. They then used a centrifuge to separate the lighter (empty) protein coats from the heavier bacterial cells which were full of phages.

They found that for the first batch the fluid near the top of the centrifuge tube was radioactive because this is where the (empty) protein coats were found. For the second batch, they found that the bacteria, containing the new phages, were radioactive because the radioactive DNA had been inserted into the cell.

As the bacterial cells were the place where the new phages were made, the "instructions" on how to make them had to have come from the phages' DNA which was inserted into the bacteria after infection with the phages. This evidence strongly supported the hypothesis that the genetic information was carried in the DNA of the virus and not the protein coat

AHL

The structure of DNA

DNA structure was first investigated by X-ray diffraction

The discovery of the three dimensional structure of DNA involved a large number of scientists who all contributed pieces to this puzzle. *Watson* and *Crick* first published the information and received the Nobel prize for their work together with *Wilkins*. The story includes the appreciation of initially bewildering information, errors in thinking which took time to be corrected, the race to be the first to publish, poor interpersonal relationships and even the hardship of female scientists at the time.

A vital piece of the puzzle came from the measurements of *Rosalind Franklin* (**Figure 703**) and *Maurice Wilkins* from their X-ray diffraction data.

Figure 703 Rosalind Franklin

When an X-ray is taken, for example the X-ray of a knee shown in the **Figure 704**, the x-rays darken the photographic plate. The bones stop or bend some of the rays, making those sections of the leg lighter.

Figure 704 X-ray of a human knee.

The structure of DNA (cont.)

In X-ray diffraction, a narrow beam of X-rays is focussed at a crystalline solid and projected on a photographic plate. The atoms in the DNA, diffract (scatter) the x-rays. This means that the pattern of light and dark on the plate is related to the position of the atoms in the DNA. Refer to **Figure 705**.
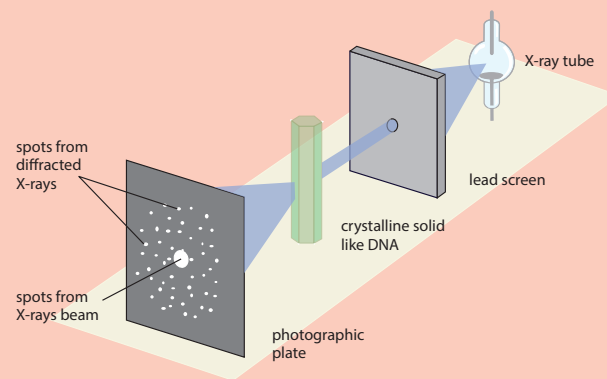
Figure 705 The principle of X-ray diffraction

Franklin and *Wilkins* made many X-ray diffraction photographs but "Photo 51" (below) was extraordinary clear and contributed significantly to determining the final structure. Refer to **Figure 706**.
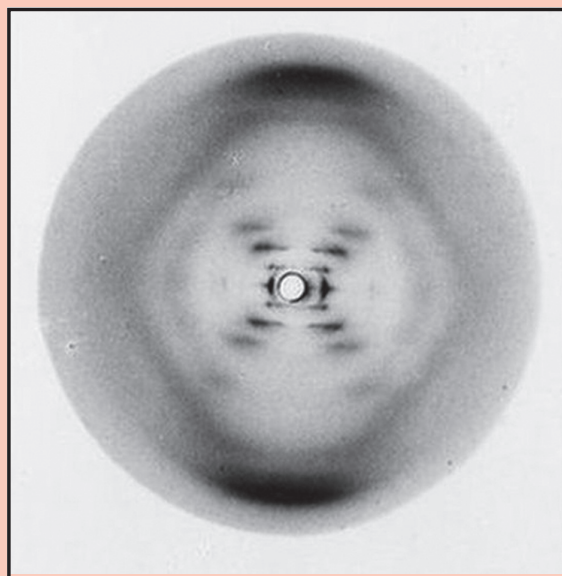
Figure 706 X-ray pattern of the DNA helix

They determined that one helical turn measured 340 nm (1 nm = 10^{-3} µm) and the distance between the base pairs is 34 nm which means that there are 10 base pairs in one helical turn. This was a critical discovery in the elucidation of the double helix as the basis for the structure of DNA.

AHL

7.1.1 Nucleosomes help to supercoil the DNA

© IBO 2014

The DNA double helix is composed of two antiparallel strands, kept together by hydrogen bonds between the organic bases (see also Topic 2.7).

Each DNA strand is made up of a chain of nucleotides. Each nucleotide is composed of a sugar (deoxyribose), a phosphate group and a nitrogenous base.

It is possible to assign numbers to the carbon atoms in deoxyribose. By convention, the numbering of the carbon atoms is done as shown in *Figure 707*.
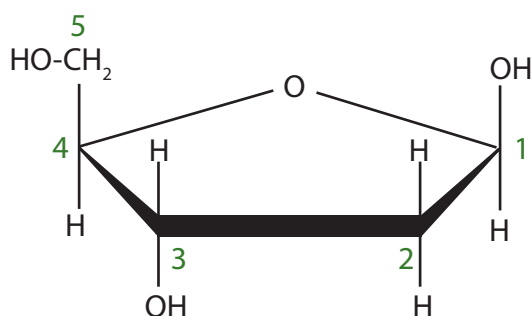

*Figure 707 Deoxyribose sugar
(with carbon atoms numbered)*

If we then draw in the phosphate (circle) and the nitrogenous base (rectangle), the result is a nucleotide, as shown in *Figure 708*.
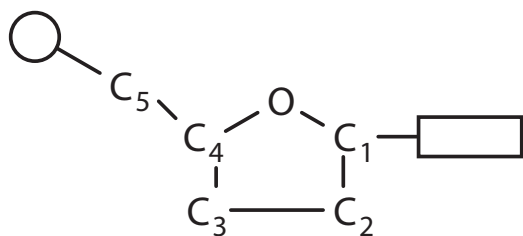

Figure 708 A nucleotide

The phosphate is covalently attached to C5 of the deoxyribose. In a single strand of DNA, the phosphate of the next nucleotide will attach to C3. This is called a 3' - 5' (phosphodiester) bond or linkage.

A DNA strand therefore, ends on one side with a phosphate on the 5 prime (5') end of the deoxyribose and on the other with a deoxyribose at the 3 prime (3') end. Since there are two strands which are antiparallel, the result is shown in *Figure 709*. The sequence of the nucleotides is customarily given in the 5' to 3' direction.
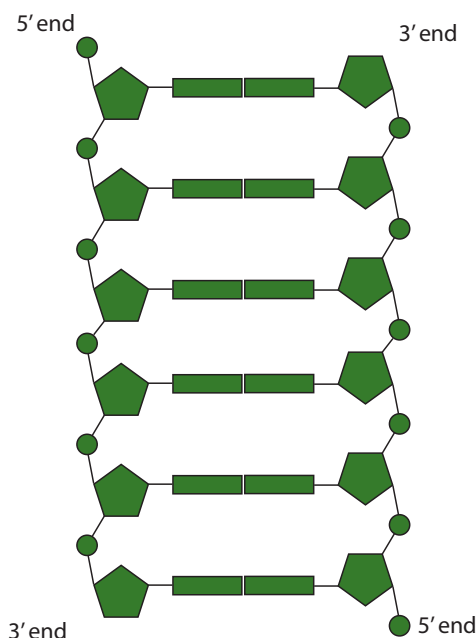

Figure 709 The DNA double strand

Analysis of chromosomes has shown that they are made of DNA and protein and a small amount of chromosomal RNA. (The 'chromosomes' of prokaryotes, like bacteria, are made of only DNA.)

DNA has negative charges along the strand and positively charged proteins are bonded to this by electrostatic forces. These basic proteins are called histones. The complex of DNA and protein is known as chromatin.

The total length of DNA in a human cell is approximately 2 metres (6 billion base pairs). This means that an average chromosome is around 4 cm long. A typical eukaryotic nucleus is between 11 and 22 μm = 22 ×10⁻⁶ m in size. (1 millimetre=1000 micrometres (μm). This means that the length of the DNA needs to be reduced 2000 to 6000 times. You can imagine that this requires good biological organisation to prevent the DNA from getting tangled. The answer to how this packing ratio is accomplished is found in the existence of nucleosomes. The structure of a nucleosome is shown in *Figure 710*.
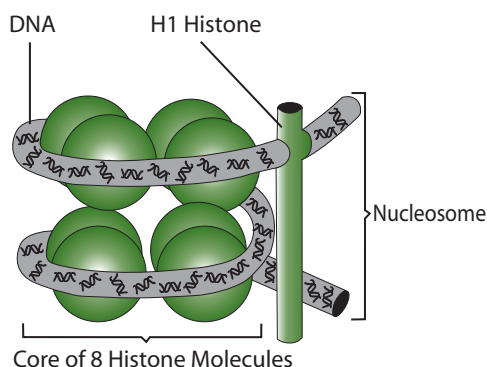

Figure 710 The structure of a nucleosome

AHL

High resolution EM images of DNA have shown that the genetic material looks like beads on a string. This is caused by the DNA helix combining with eight small histone molecules with an additional histone (H1) keeping it all together. These structures are called nucleosomes.

Figure 711 illustrates how the nucleosomes interact with each other and achieve a high degree of organisation in the process of shortening the DNA to a more manageable length.
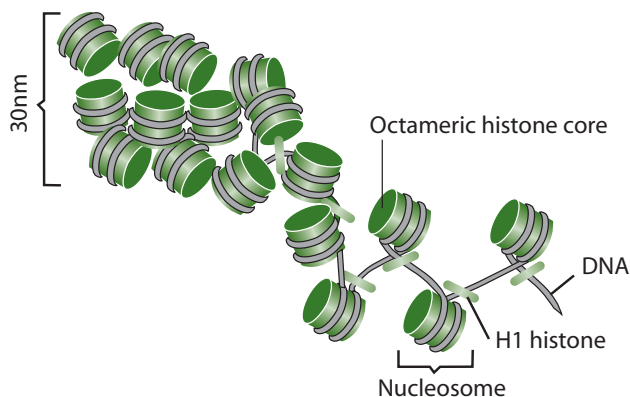

Figure 711 The arrangement of nucleosomes

Nucleosomes are important in two ways. The first was illustrated in *Figure 711*, that is nucleosomes help to organise the DNA so that it can fit into a nucleus.

When they are in this compact arrangement, the chromosomes are described as supercoiled. The second role of nucleosomes is to prevent transcription. (see section 7.2)

When the cell requires transcription to occur, enzymes will alter the shape of the nucleosomes to allow RNA polymerase to attach to the promotor region of the DNA strand and to start the process of transcription.

Visualisation of molecules

APPLICATIONS AND SKILLS

3D molecular visualization software can be used to analyse the association between protein and DNA within a nucleosome.

Visit the websites below, you may need to install 'Java' animation software on your device and possibly adjust security settings.

<http://www.umass.edu/molvis/bme3d/materials/explore.html#nucleosome>

and

<http://www.biochem.umd.edu/biochem/kahn/teach_res/nucleosome/jmol-nsome.html>

7.1.2 DNA structure suggested a mechanism for DNA replication © IBO 2014

As described in Topic 2.7, DNA replication is a semi-conservative mechanism. This was demonstrated by *Meselson* and *Stahl* in 1958. The result of semi-conservative replication is shown below with the parent DNA shown in blue and the newly formed strands shown in green See *Figure 713*.
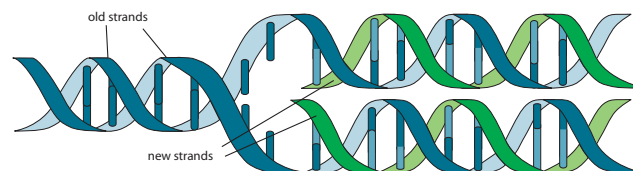

Figure 713 DNA replication

Semi-conservative essentially means that in the process of DNA replication, the DNA double helix 'unzips' and new hydrogen bonds are formed with the organic bases of DNA nucleotides 'floating around' in the cell. DNA polymerase creates the 3'-5' linkages between the nucleotides thus creating a new DNA strand, complementary to the original one and identical to the one that was 'unzipped'. DNA polymerase creates the covalent bonds between the nucleotides of the growing strand. Refer to *Figure 714*.
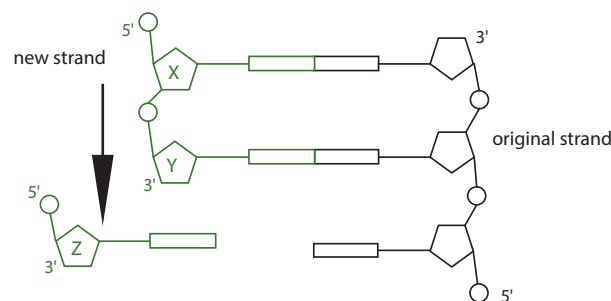

Figure 714 DNA replication

The 'black' nucleotides in *Figure 714* are part of the old, existing chain. The 'green' nucleotides are part of the new, growing chain. Nucleotide X is in place. Nucleotide Y has just been added and nucleotide Z is in the process of forming hydrogen bonds between its organic base and the complementary base on the existing chain.

The DNA polymerase will catalyse the formation of a covalent bond between the 3' end of nucleotide Y (the nucleotide that has just been added) and the 5' end of nucleotide Z (the nucleotide that is being added). So if you draw in an arrow pointing in the direction of the most recently added nucleotide, the arrow will point from the 5' end to the 3' end. This means that DNA replication occurs in a 5' to 3' direction.

AHL

7.1.3 DNA polymerases can only add nucleotides to the 3′ end of a primer

The other new strand, the lagging strand, would have to grow in a 3'-5' direction to keep up. This means that new nucleotides would have to be added to the 5' end of the growing chain. However, DNA polymerase III cannot work in this direction. Refer to *Figure 717*.

7.1.4 DNA replication is continuous on the leading strand and discontinuous on the lagging strand

As a result, the lagging strand is formed in short segments of 100 – 200 nucleotides (called *Okazaki* fragments) in a 5' to 3' direction, adding new nucleotides to the 3' end of the growing strand.

The process is the same:

- RNA primer formed from RNA nucleotides by DNA primase.

- Deoxyribonucleoside triphosphates form hydrogen bonds with the exposed organic bases of the old DNA strand.

- DNA polymerase III, working in a 5' to 3' direction (i.e. away from the direction in which the DNA unwinds and unzips, adding new nucleotides to the 3' end of the growing strand), bonds the DNA nucleotides to form an Okazaki fragment.

- DNA polymerase I will then remove the RNA primer of the Okazaki fragment and replace the RNA nucleotides with DNA nucleotides. The leading strand has no Okazaki fragments and does not need DNA polymerase I except to remove the initial RNA primer.

- DNA ligase then catalyses the formation of the covalent bonds attaching the DNA segments to create one strand. Again, this enzyme is not used much in the leading strand.

The most stable configuration of DNA is in a double helix. So as soon as the strands have been separated by helicase, these complementary strands would tend to reform the double helix. To prevent this from happening and to stabilise the single DNA strand, single strand binding proteins (SBBs) attach to the single stranded DNA. The presence of SBBs also prevents the strand from being digested by nucleases and allows the necessary enzymes to function effectively.

The rate of replication in fruit flies (*Drosophila*) is 2600 nucleotides/minute. The largest chromosome of *Drosophila* is 6.5×10^7 nucleotides.

If DNA replication started at both ends of the chromosome, it would take 8.5 days to replicate the chromosome. In fact, it only takes 3-4 minutes. In order to explain this discrepancy, scientists have determined that replication must start at many points along the same DNA helix at the same time. Refer to *Figure 715*.
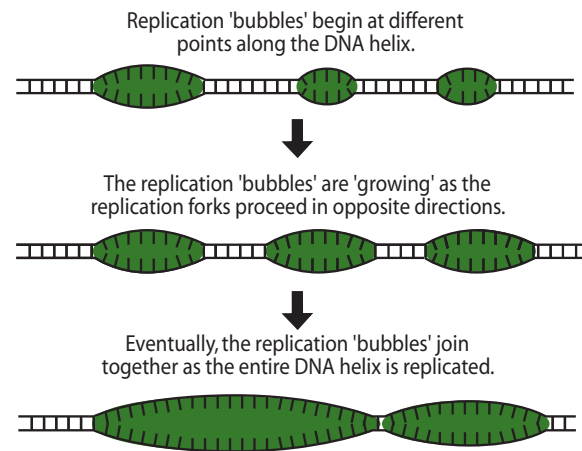

Figure 715 Replication occurs in stages

Stopping DNA replication

Use of nucleotides containing dideoxyribonucleic (*think and say di-deoxy-ribo-nucleic*) acid to stop DNA replication can be used in preparation of samples for base sequencing.

Base sequencing, also known as DNA sequencing or gene sequencing, is the process of determining the sequence of nucleotides in a section of DNA. This can be used to determine if someone has a genetic condition e.g. carries genes which are very likely to cause breast cancer.

The method below uses dideoxyribonucleic acid nucleotides rather than deoxyribonucleic acid nucleotides. The crucial difference is that the dideoxyribonucleic acid nucleotides have a hydrogen (H) attached to the carbon at position 3, rather than a hydroxyl (-OH) as is the case is deoxyribonucleic acids. See *Figure 716*.

When a nucleotide is attached to a growing DNA strand, the next nucleotide will attach its phosphate group to the oxygen of the hydroxyl group attached to C-3. When a dideoxynucleotide attaches to a growing DNA strand, there is no hydroxyl group attached to C-3 so DNA replication will terminate with the last nucleotide added being the dideoxynucleotide. Refer again to *Figure 716*.

This knowledge can be used in gene sequencing. In order to sequence the DNA, single stranded DNA is used. It is allowed to replicate in four different experiments. In each of the experiments, everything that is needed for replication is supplied except one of the deoxyribose nucleotides is replaced with dideoxyribose nucleotides.
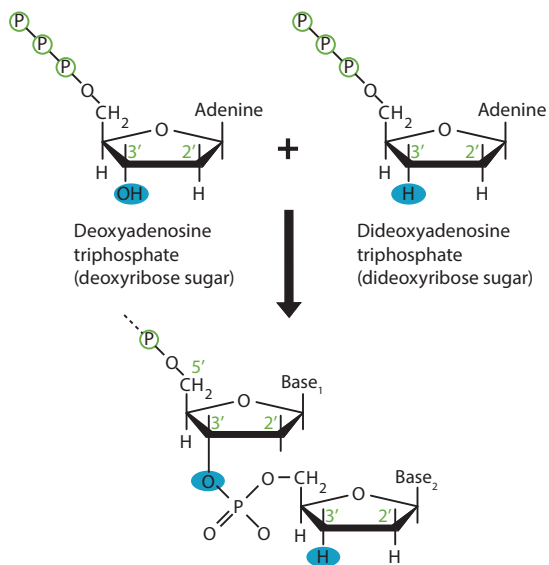
Figure 716 Deoxy- and Dideoxy-ribose nucleotides

As a result, short sections of new DNA strands are produced, with the last nucleotide in each section being a dideoxyribose nucleotide. The sections overlap. By combining the information, it is possible to obtain the sequence of bases in the original DNA single strand (which might be a gene).

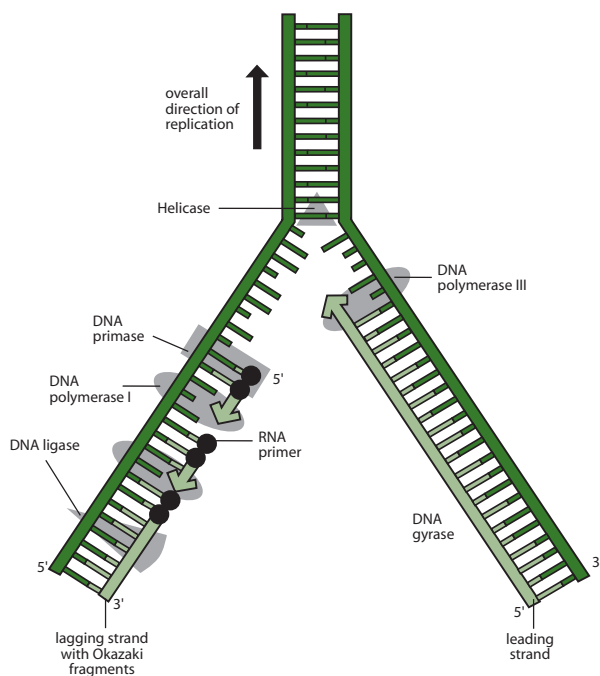### 7.1.5   DNA replication is carried out by a complex system of enzymes

© IBO 2014

Figure 717 DNA replication

Prokaryotic DNA replication is illustrated in *Figure 717*, it involves the following steps. This process involves a number of different protein enzymes. Their names end with *–ase*.

The DNA is a double helix (like a twisted ladder) which needs to be unwound by the enzyme helicase. In order to replicate, the two DNA strands need to be separated as shown in the *Figure 718*.
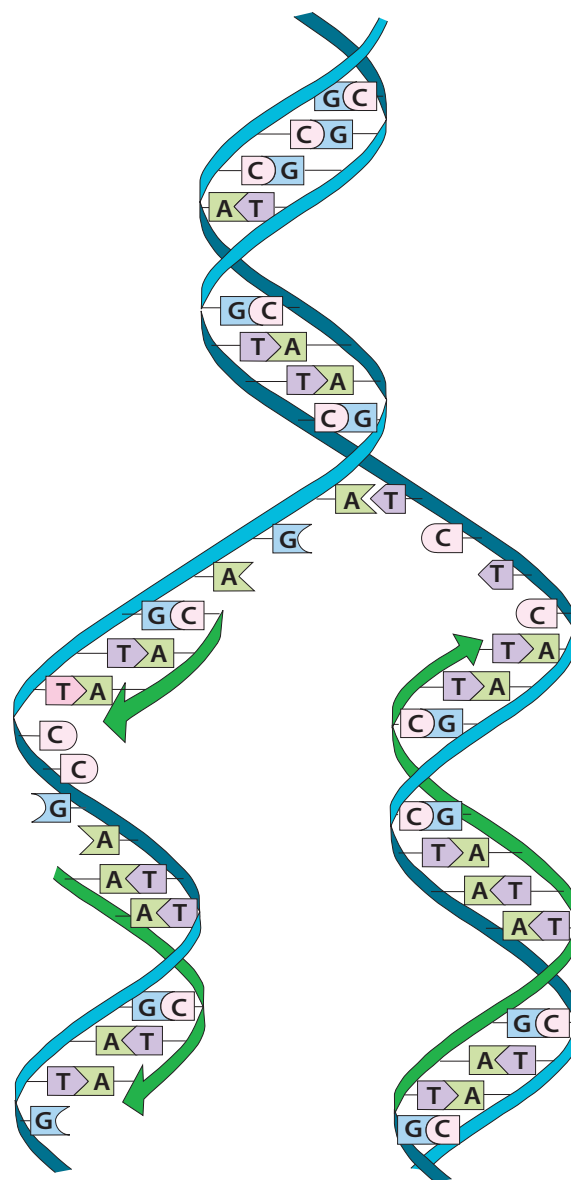
Figure 718 Semi-conservative replication

If you imagine the DNA as two twisted strings and then try to separate them by pulling them apart at the end, you will end up making the twists in the remaining part of the strand tighter. If you keep going, you will end up with some very tangled string.

To avoid this, an enzyme DNA gyrase (also known as topoisomerase II) relieves the strain on the "twisted ladder", by supercoiling the DNA. See *Figure 719*
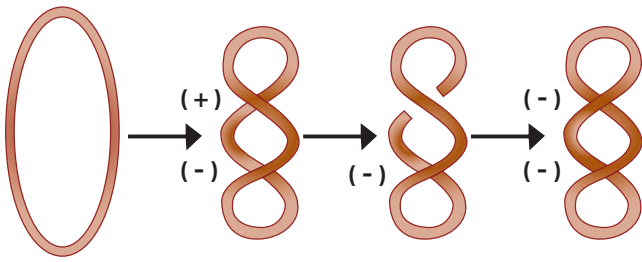
AHL

215

Figure 719 The role of DNA gyrase

When strain occurs due to the separation of the strands and the DNA twists as shown above, DNA gyrase will break the DNA at a point where two strands overlap, allow the intact DNA to pass to the other side and re-attach the broken sections. This relieves strain. Refer to *Figure 719*.

- Helicase also breaks the hydrogen bonds between the two strands (between the base pairs) separating the strands. One of the old strands will be 3' to 5' while the other strand will be 5' to 3'.

- Present in the cell are deoxyribonucleoside triphosphates which are composed of an organic base, a deoxyribose sugar and three phosphate groups. They are sometimes referred to as dATP, dCTP, dGTP and dTTP.

- Before a new strand of DNA can be formed, it is necessary to start with an RNA primer. The RNA primer is a few RNA nucleotides which bind to the old DNA strand (hydrogen bonding between the organic bases). The enzyme DNA primase will bind these RNA nucleotides together (covalent bonds).

- Now that the RNA primer is in place, the deoxyribonucleoside triphosphates will form hydrogen bonds between their organic bases and the complementary base on the exposed strand of the old DNA, i.e. A with T and C with G.

- In the new strand that is forming in the 5' to 3' direction, the 'leading strand', DNA polymerase III will bind the new nucleotides to the growing strand by covalent bonds formed via condensation reactions. DNA polymerase III only works in a 5' to 3' direction, i.e. it can only attach new nucleotides to the 3' end of a growing strand.

- As it attaches to the growing DNA strand, the second and third phosphate groups are removed from the deoxyribonucleoside triphosphate, changing it into deoxyribonucleotide (organic base, deoxyribose sugar, phosphate).

### 7.1.6 Some regions of DNA do not code for proteins but have other important functions © IBO 2014

DNA is transcribed into RNA and then translated into proteins. These proteins may be enzymes which catalyse a specific reaction or they may be any other protein. Some sections of DNA do not code for proteins. Non-coding DNA may be transcribed or not transcribed.

Transcribed non-coding DNA includes the sections of DNA which are transcribed into tRNA which is needed for translation (see section 7.2).

Introns are also transcribed non-coding DNA. They are more common in eukaryotic DNA and rare in prokaryotic DNA. While in prokaryotes, genes are uninterrupted sections of DNA, in eukaryotic cells, coding sections of DNA are interrupted with long non-coding intervening sequences. In other words, many genes are discontinuous. The intervening sequences are called introns, while the coding sequences are called exons (since they are expressed). Refer to *Figure 720*.
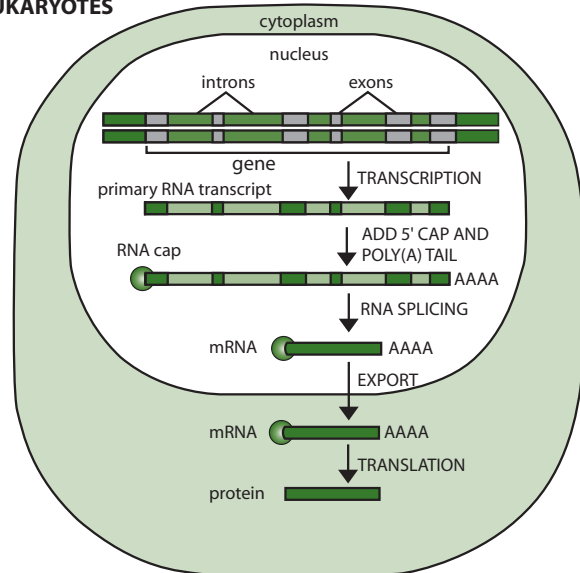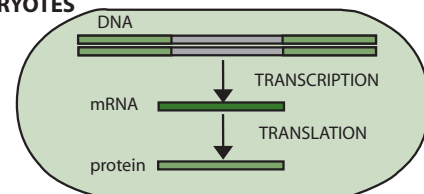

Figure 720 The role of introns and exons

Telomeres are non-coding sections at the end of a chromatid. They are not transcribed. When chromosomes are replicated, it is not possible to continue the process of replication all the way to the end of the chromosome. The non-coding telemores are therefore not completely replicated but at least all other coding DNA is maintained. Telomeres protect the chromosome from loss of essential coding DNA. Over time, with each cell division, the telomeres become shorter.

Regulator genes are genes which determine if a specific other gene is expressed or not. The regulator gene may be an activator or repressor of the other gene.

Tandem repeats

APPLICATIONS AND SKILLS

'Tandem repeats' are used in DNA profiling.

DNA profiling is used to identify individuals by their DNA, to match a sample of DNA to a person or to determine if individuals are genetically related. Almost all DNA is identical for all humans so DNA profiling focusses on the sections that are different.

Some of these sections are called short tandem repeats (STR) and they are found on the same place on the chromosome but their sequence is so variable that they are unlikely to be the same for different individuals (except for identical twins). The short tandem repeats are also more likely to remain intact when DNA is degraded in non-ideal conditions, which might be the case with DNA samples at a crime scene. In addition, STRs are easy to amplify using PCR (polymerase chain reaction (PCR), see section 2.7)

Language and Natural Science

THEORY OF KNOWLEDGE

Interpretation of information is influenced by the labels associated with information. Labels allow us to quickly categorise the importance or relevance of information.

Most of a eukaryote's genome is made up of repetitive sequences of DNA that appears to be non-coding. Initially some Scientists labelled these sequences as 'junk' DNA. This emotive label suggested that repetitive DNA is useless. Such an idea runs contrary to evolutionary theory that would suggest useless code would be unlikely to remain in the genome. The sheer size of the genome that is repetitive and the fact that it is found in all eukaryotes has led scientists to change the label. What is becoming increasingly obvious is that repetitive DNA sequences are likely to have a significant role in organisms that is beyond current human knowledge and understanding.

Discuss other examples in which labels used in the pursuit of knowledge may affect the nature of the knowledge we obtain?

AHL

7.2 Transcription and gene expression

Information stored as a code in DNA is copied onto mRNA

7.2.1 Transcription occurs in a 5′ to 3′ direction
© IBO 2014

Transcription is the enzyme-controlled process of synthesising RNA from a DNA template. It is carried out in a 5' to 3' direction (of the new RNA strand). This means that RNA polymerase adds new RNA nucleotides to the 3' end of the growing RNA strand. mRNA, tRNA and rRNA all need to be present in the cell (and hence have been transcribed) for protein synthesis to take place.

7.2.2 Nucleosomes help to regulate transcription in eukaryotes
© IBO 2014

Nucleosomes are important in two ways. The first was illustrated in *Figure 711*, that is nucleosomes help to organise the DNA so that it can fit into a nucleus. *See also section 7.1*

The second role of nucleosomes is to prevent transcription. Transcription is the process in which the antisense strand of the DNA is used as a template for producing a strand of RNA. In order for transcription to occur, RNA polymerase needs to be able to attach to the promotor region of the 3' end of the structural gene on the DNA. If the DNA is organised into nucleosomes, the promotor region is usually not accessible to RNA polymerase and therefore, transcription will not occur.

7.2.3 Eukaryotic cells modify mRNA after transcription
© IBO 2014

After the DNA has been transcribed, the produced mRNA is also referred to as pre-mRNA. It needs to be modified before translation.

Modifications occur at the 3' end or the 5' end of the pre-mRNA or elsewhere in the pre-mRNA. During modification pre-mRNA becomes mature mRNA.

The modification of the 5' end of mRNA is capping. Capping is the addition of 7-methylguanosine to the 5' end of the mRNA after the 5' phosphate has been removed. The cap protects the mRNA from breakdown by ribonucleases and increases the mRNA's stability during translation. Capping takes place in the nucleus. Next to the 5' cap, there is an untranslated region called 5' UTR. It has a regulator function.

The modification at the 3' end of mRNA is called poly-adenylation. Polyadenylation is the addition of multiple adenosine monophosphates to the 3' end of the pre-mRNA so that a sequence of A (adenine) is created. This is known as a poly(A) tail and can be 250 bases long. The poly(A) tail binds to proteins which enhance stability. Next to the poly(A) tail, there is an untranslated region called 3' UTR. It has a regulator function and affects the formation of the poly(A) tail.

The third modification needed to turn pre-mRNA into mature mRNA is to remove the introns which are not translated and splice together the remaining exons which will be expressed in translation. See *Figure 730*.
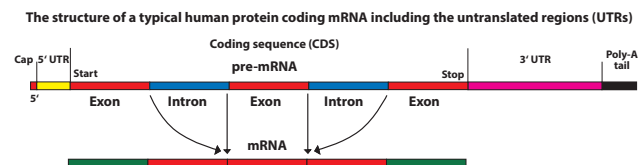

Figure 730 The role of introns and exons

7.2.4 Splicing of mRNA increases the number of different proteins an organism can produce
© IBO 2014

It is possible that a single gene can code for several different proteins, using a process sometimes known as alternative splicing.

In alternative splicing, the DNA is transcribed into mRNA in the usual manner. The pre-mRNA is modified through capping and tailing and introns are removed. However, it is possible that not all exons are found in the mature mRNA. For example, if exons A, B, C and D exist, then several proteins including 3 exons can be made: A, B, C; A, B, D; A, C, D and B, C, D.

In *Drosophila*, sex determination depends on the inclusion of a section of a particular exon. Including this section creates a mature mRNA with an early stop codon so a truncated protein which is not functional. This will lead to the development of a male individual. If the exon is not included, the full protein will be produced in translation and a female will develop.

Another example is alternative splicing of tropomyosin pre-mRNA. The 11 exons in this pre-mRNA can be spliced to produce 5 different proteins, found in different tissues as shown in the *Figure 731*.

Regulation of alternative splicing often involves (*cis*-acting) regulators which are found in the mRNA sequence which are activated or repressed by (trans-acting) protein factors from elsewhere in the genome.
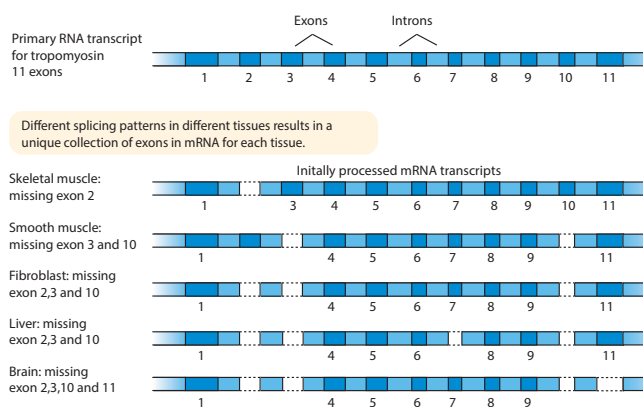
Figure 731 Several proteins from one gene

7.2.5 Gene expression is regulated by proteins that bind to specific base sequences in DNA
© IBO 2014

Promoters as non-coding DNA

A section of DNA which activates transcription of a gene is called a promoter. This is a specific base sequence in the DNA which is non-coding, i.e. the section itself has a regulatory function and its base sequence will not be transcribed nor translated into a protein.

Transcription is carried out in a 5' to 3' direction when looking at the newly formed RNA strand. Since this new RNA strand is anti-parallel to the non-sense strand that is being transcribed, transcription starts at the 3' end of the DNA gene. *See Figure 711.* Promoters are found "upstream" (towards the 3' end) from the genes which they cause to be transcribed. They are found on the same strand of the double helix: the anti-sense strand which is the strand that is transcribed.

The promoter is the site for binding RNA polymerase which will attach the individual nucleotides together to form a single strand of RNA. At the (5') end of the gene, a terminator site is found which will stop the transcription process. The terminator, another section of non-coding DNA, is found at the 5' end of the gene on the non-sense strand of the DNA that is being transcribed.

RNA nucleotides, in the form of ribonucleoside triphosphates, form hydrogen bonds with the complementary nucleotide of the DNA strand and are then linked by RNA polymerase. RNA polymerase is the enzyme which binds the individual RNA nucleotides together to form an RNA strand. It is similar to DNA polymerase and works in almost the same way. however, it can also function in a 3' to 5' direction.

The only difference between ribonucleoside triphosphate and deoxyribonucleoside triphosphate is one hydroxyl group (-OH group) on C2 in the pentose sugar. This changes the deoxyribose into ribose. RNA does not bind with the organic base to produce Thymine but uses Uracil in its place.

As in DNA replication, the ribonucleoside triphosphate will attach (by a covalent bond) to the 3' hydroxyl group of the growing strand. The second and third phosphate groups will be removed, providing the energy required to drive this reaction.

7.2.6 The environment of a cell and of an organism has an impact on gene expression
© IBO 2014

The Lac Operon model (Jacob, Monod. 1961) is an example of a simple system of gene regulation involving environmental factors.

When a bacterium has lactose available, it will need three enzymes to effectively use this lactose. These three (protein) enzymes are produced by one gene. However, it would be rather wasteful to transcribe the gene, splice the pre-mRNA and produce the enzymes if lactose were not present in the environment of the bacteria. See *Figure 732*.

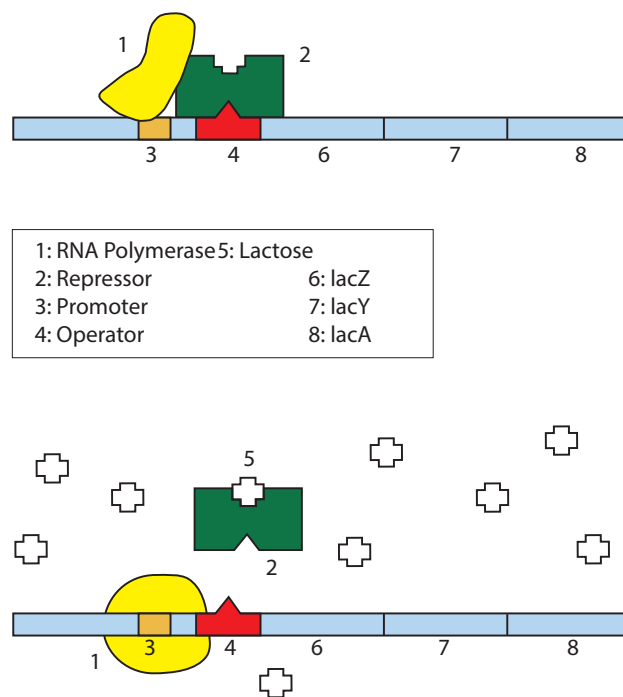|	
1: RNA Polymerase	5: Lactose
2: Repressor	6: lacZ
3: Promoter	7: lacY
4: Operator	8: lacA

Figure 732 The Lac Operon model

At any time, RNA polymerase will attempt to attach to the promoter region. In the absence of lactose (top diagram), it will not be able to proceed because a repressor molecule is attached to the operator.

AHL

When lactose is present, lactose will bind to the repressor which will remove it from the operator region. RNA polymerase can now proceed to bind and transcribe the gene. Splicing of the gene will produce one of the three possible enzymes.

When the enzymes have digested all lactose, the repressor will bind once again to the operator, blocking further transcription of the gene.

DNA methylation

Changes in the DNA methylation patterns can be analysed. DNA methylation adds a methyl group (-CH_3) to cytosine which is followed by guanine in DNA (5' CG 3'). It does not change the genetic code and is said to be an 'epigenetic' modification.

This process alters gene expression and occurs during cell differentiation. It is needed for normal development. Methylation is permanent but removed during zygote formation and subsequently built up again. Methylation tends to occur in coding regions; non-coding regions (particularly the promoter region) is generally unmethylated.

Changes in DNA methylation are related to cancer development and the subject of much ongoing research.

Methylation of coding regions makes transcription more difficult, meaning that the gene will only be transcribed when the polymerase attaches at the beginning and works it way through. It prevents transcription starting just anywhere, in the middle of a gene. Hypomethylation of coding regions may therefore lead to transcription of sections of genes. Also, overall hypomethylation can lead to a less stable chromosome which is more likely to undergo mutations (which could cause cancer).

Hypermethylation can be an indicator of cancer. Hypermethylation of the promotor region may inactivate the gene which normally silences the onco-gene (the cancer causing gene).

Nature verses Nurture

The debate about the relative influence of nature (hereditable material) verses nurture (environment and experience) on the personality of an individual is ongoing. It also has extremes of arguments from those who believe in the concept of '*tabular rasa' (i.e.* we are born as a blank slate) to those who believe that our personality and behaviour is almost entirely inherited.

It is almost impossible to control testing in this area, so studies of identical twins separated at birth have been most helpful in research into this question. Current thought is that a complex interaction between nature, nurture and chance determines individual human behavior. *Steven Pinker* and *Matt Ridley* are two scientists who have written on this topic.

Is it important for Natural Sciences to answer the question of the origin of human behaviour?

AHL

7.3   Translation

Information transferred from DNA to mRNA is translated into an amino acid sequence

7.3.1   Initiation of translation involves assembly of the components that carry out the process
© IBO 2014

Translation is the process of decoding the mRNA to produce a polypeptide chain (section 2.7). The cell organelles responsible for translation are ribosomes.

Ribosomes are found in both prokaryotic and eukaryotic cells, but have a slightly different structure in each and are larger in eukaryotic cells than in prokaryotic cells. Ribosomes are measured by their size and density in units called 'Svedberg' or 'S'. Prokaryotic ribosomes and ribosomes of chloroplasts and mitochondria are 70S, whereas eukaryotic ribosomes are 80S. These units do not necessarily add up mathematically because the 80S eukaryotic ribosome consists of a large 60S and a small 40S subunit.

The nucleolus in the nucleus contains many copies of the information on how to make rRNA. Some cells which are very active in protein synthesis may have many nucleoli.

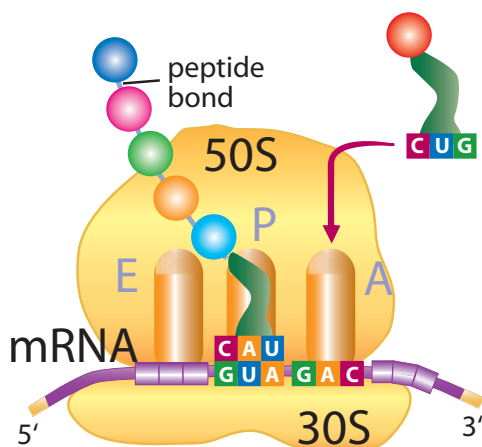For the basic structure of a ribosome see section 2.7. A more detailed diagram shown in *Figure 740*.

Figure 740 The structure of a eukaryotic ribosome

Ribosomes consist of two subunits, a small and a large unit. The smaller subunit is made of one molecule of rRNA and some proteins, the larger subunit is made of two molecules of rRNA and some proteins, including the enzyme peptidyl transferase which bonds together the amino acids brought in by tRNA. The smaller subunit has the binding site for mRNA, the larger subunit has the binding sites (known as the A, P and E sites) for tRNA.

Transfer RNA (tRNA) is a single strand of RNA which is folded back on itself to form the structure shown in *Figure 741*. The primary structure of tRNA would be the sequence of nucleotides in a single strand. Some of the nucleotides will form hydrogen bonds, creating the secondary structure which is the "clover leaf" that you also saw in section 2.7. This structure will fold into the L-shaped tertiary structure that tRNA needs to have to fit into the A, P and E sites of the ribosome. *Figure 741* shows an image of the tertiary structure in both 2D and 3D.
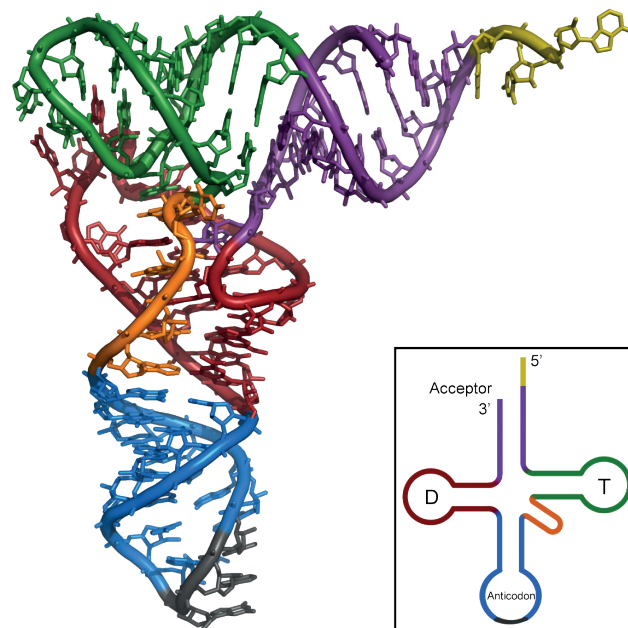

Figure 741 The 3D structure of tRNA

The clover leaf shape can also be drawn as a mirror image so it is useful to remember that the small additional loop is found between the anti codon and the T arm, which is also the side of the 3' end of the tRNA.

The anticodon contains 3 nucleotides which bind to the codon on the mRNA in the ribosome during translation. *Figure 741* also shows the D arm (or Loop 2) and the T arm (or Loop 4). The other section labelled is the acceptor stem which has 7 base pairs (although they are not always C-G or A-U pairs) with an addition 3 nucleotides (CCA) at the 3' end.

The amino acid attachment site which always has CCA as its final sequence on the 3' end. This is again illustrated in *Figure 741*. The tRNA shown has the tRNA anticodon of CGG which matches the mRNA codon of GCC (read from the 5' end of the mRNA). This codon codes for the amino acid Alanine which is covalently bonded to the amino acid binding site on the acceptor arm.

Each amino acid has its own tRNA with a unique amino acid binding site, to bind the correct amino acid, and anticodon, to bind to the correct codon on the mRNA.

AHL

Molecular visualization

'Molecular visualization' software can be used to analyse the structure of eukaryotic ribosomes and a tRNA molecule.

For example, this site

<http://higheredbcs.wiley.com/legacy/college/boyer/0471661791/structure/tRNA/trna.htm>

An animation of the process of translation, and others is shown in

<http://www.youtube.com/watch?v=5bLEDd-PSTQ>

Again you may need to download 'Java' software and adjust security settings on your computer.

The process of translation involves ribosomes, mRNA, tRNA, GTP, initiation factors. It is divided into initiation, elongation and termination.

Inititation

The appropriate tRNA covalently binds to the first amino acid of the polypeptide about to be synthesized, using energy from GTP. A tRNA which is covalently bonded with its amino acid is known as an aminoacyl-tRNA. The enzyme which catalyses the formation of this bond is aminoacyl tRNA synthetase but there is a different aminoacyl tRNA synthetase for each amino acid.

The start codon in eukaryotic mRNA is always AUG, coding for methionine. The first tRNA, also known as the initiator tRNA, will have the anticodon UAC and carry the amino acid methionine.

Initiation factor eIF-2 recognises initiator tRNA and forms a complex.

The ribosome dissociates into its subunits. The small 40S subunit binds with initiation factor eIF-3 to prevent both subunits from immediately combining again. The tRNA and eIF-2 complex bind to the small 40S subunit of the ribosome.

The 5' cap of mRNA (see section 7.2) binds to initiation factor eIF-4F which guides the small ribosomal subunit with its initiator tRNA to the 5' end of the mRNA.

The small ribosomal subunits moves from the 5' cap of the mRNA towards its 3' end until it comes across the start codon AUG. The initiation factors will be released and the large 60S subunit of the ribosome will join the small 40S subunit, forming the eukaryotic 80S ribosome with the initiator tRNA in the P-site.

Elongation

Depending on the next codon on the mRNA, the appropriate aminoacyl tRNA will bind to the vacant A-site. The two amino acids, the methionine attached to the initiator tRNA in the P-site and the next amino acid attached to the tRNA in the A-site will form a peptide bond. The bond between methionine and its tRNA will break and the ribosome will move 3 nucleotides (1 codon) towards the 3' end of the mRNA. The first tRNA, without its amino acid, will be found in the E-site and will leave the ribosome, the second tRNA with a short polypeptide chain of two amino acids, will be found in the P-site and the A-site is ready to receive the next aminoacyl tRNA.

Termination

When the ribosome reaches a stop codon (UAA, UAG or UGA) on the mRNA, no tRNA exists with a matching anticodon. Instead, release factors (RF1 or RF2) will break the bond between the polypeptide chain and last tRNA (found in the P-site) and release the polypeptide from the ribosome. The mRNA and tRNA will be released from the ribose which itself will dissociate and the cycle can start again.
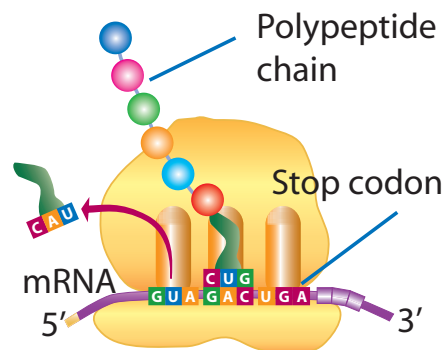

Figure 743 Termination

Often several ribosomes follow each other along the mRNA, showing longer polypeptide chains as they are closer to the 3' end of the mRNA. *See Figure 745*.

Enzyme-substrate specificity

tRNA-activating enzymes illustrate enzyme-substrate specificity and the role of phosphorylation. Phosphorylation of eIF-2 inhibits its action. Without eIF-2, the initiator tRNA cannot bind to the small subunit of the ribosome so initiation of translation cannot occur.

You could possibly do an internet search for animations to illustrate this process.

7.3.2 Synthesis of the polypeptide involves a repeated cycle of events © IBO 2014

As was shown above, the components which come together for protein synthesis are released at the end of the translation. The small ribosomal unit will attach to eIF-3 again, all of the tRNAs used can find another of the correct amino acid and become activated by aminoacyl tRNA synthetase. The aminoacyl tRNA can bind with eIF-2 which will allow it to bind to the small ribosomal subunit and both can then attach to the mRNA which was only read, not altered, during translation

7.3.3 Disassembly of the components follows termination of translation

© IBO 2014

As stated above, when the stop codon is reached, the components required for translation will separate in order to be available to build the next polypeptide. It can be the same sequence synthesized before, using the same mRNA or a different mRNA can be the basis for a different polypeptide.

7.3.4 Free ribosomes synthesise proteins for use primarily within the cell © IBO 2014

The distribution of ribosomes depends on the function of the protein they make; if the proteins are to be used inside the cell, the ribosomes tend to be found throughout the cytoplasm.

7.3.5 Bound ribosomes synthesise proteins primarily for secretion or for use in lysosomes © IBO 2014

If the protein is to be exported (secretion) or used by lysosomes, the ribosomes are generally associated with the endoplasmic reticulum. These proteins will then enter the lumen of the RER as they are produced and from there move to the Golgi apparatus where they are packaged into vesicles (see *Figure 744*).
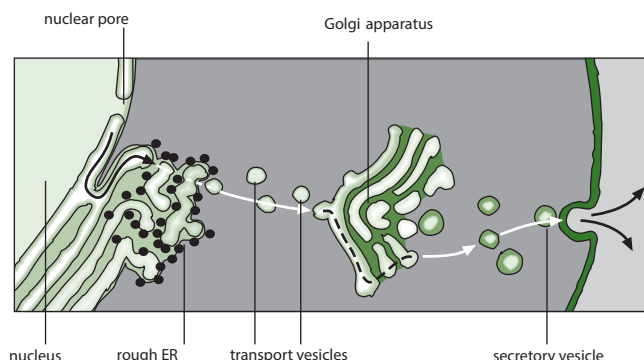

Figure 744 Golgi apparatus and vesicles

Polysomes

Polysomes can be identified in electron micrographs of prokaryotes and eukaryotes.

In the EM photograph (*Figure 745*) a number of individual ribosomes can be seen to form a polysome as they are all translating the same mRNA at the same time but at different stages.
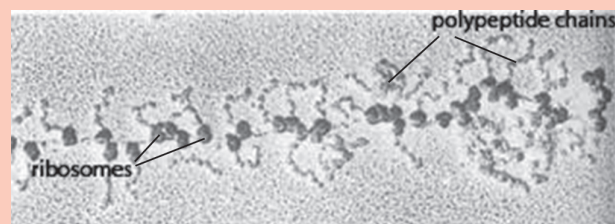

Figure 745 A polysome

7.3.6 Translation can occur immediately after transcription in prokaryotes due to the absence of a nuclear membrane © IBO 2014

In eukaryotic cells, the mRNA is transcribed in the nucleus and then modified (see section 7.2) before leaving the nucleus to go to the cytoplasm where translation takes place.

As there is no distinct nucleus due to the absence of a nuclear membrane in prokaryotes, and there is no mRNA modification, translation can commence immediately after transcription.

7.3.7 The sequence and number of amino acids in the polypeptide is the primary structure © IBO 2014

The primary structure of a protein is the sequence of the amino acids in the chain. Refer to *Figure 746*. The linear sequence of amino acids with peptide linkages affects all the subsequent levels of structure since these are the consequence of interactions between the R groups of the amino acids. Each amino acid can be characterised by its R group. Polar R groups will interact with other polar R groups further down the chain and the same is true for non-polar R groups.
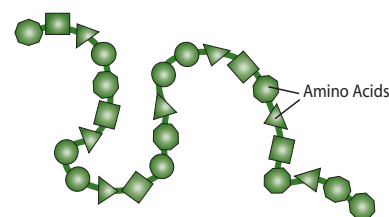

Figure 746 Primary structure

AHL

223

7.3.8 The secondary structure is the formation of alpha helices and beta pleated sheets stabilised by hydrogen bonding
© IBO 2014

The secondary structure of a protein consists of the coils of the chain, for example α helix and β pleated sheet. Refer to *Figure 747*. α (alpha) helix structures are found in the proteins of hair, wool, horns and feathers; β (beta) pleated sheet structures are found in fibroin (spider silk). Hydrogen bonds are responsible for the secondary structure. Fibrous proteins like collagen and keratin are in helix or pleated sheet form which is caused by a regular repeated sequence of amino acids. They are structural proteins.
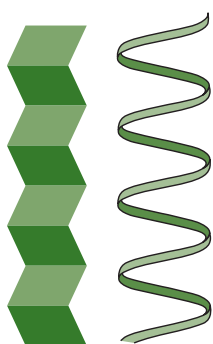
Beta pleated sheet   Alpha helix

Fig 747 Secondary structure

7.3.9 The tertiary structure is the further folding of the polypeptide stabilised by interactions between R groups
© IBO 2014

The tertiary structure of a protein refers to the way the chain is folded. Refer to *Figure 748*. This is caused by the interactions between the R groups of the amino acids in the polypeptide chain. Non-polar groups are hydrophobic and cluster together on the inside of the protein (away from the water) while polar groups are hydrophilic and cluster together on the outside of the protein (near the water). Several amino acids have a sulfur atom in their R group. Two adjacent sulfur atoms may come together, under enzyme control and form a covalent bond called a disulfide bridge (i.e. -SH + HS- → -S-S-)
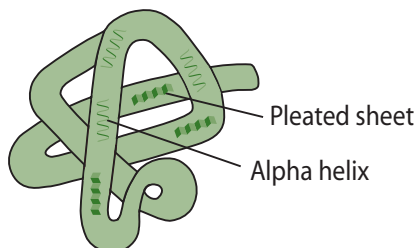

Pleated sheet
Alpha helix

Figure 748 Tertiary structure

Hydrogen bonds are also involved in the formation of the tertiary structure. Bonds between an ion serving as a cofactor and the R group of a certain amino acid may also be responsible for the folds in the polypeptide. Proteins which have a globular (folded) shape, such as hemoglobin, are called globular proteins. Microtubules consist of globular proteins (tubulin) which have a structural function. Enzymes are globular proteins. The precise and unique folding of the polypeptide creates the 'active site', i.e. the location where the substrate binds to the enzyme so that the reaction can take place.

7.3.10 The quaternary structure exists in proteins with more than one polypeptide chain
© IBO 2014

The quaternary structure of a protein involves the combination of different polypeptide chains. Refer to *Figure 749*. The different polypeptide chains are kept together by hydrogen bonds, attraction between positive and negative charges, hydrophobic forces and disulfide bridges or any combination of the above.

Many proteins (especially large globular proteins) are made of more than one polypeptide chain. Together, with the greater variety in amino acids, this causes a greater range of biological activity. Another example of the quaternary structure is the binding of prosthetic groups such as the heme group of hemoglobin. *Figure 749* shows a hemoglobin molecule. It can seen that it is made of 4 polypeptide chains:

- two α chains

- two β chains

$β_1$ $β_2$

$α_2$

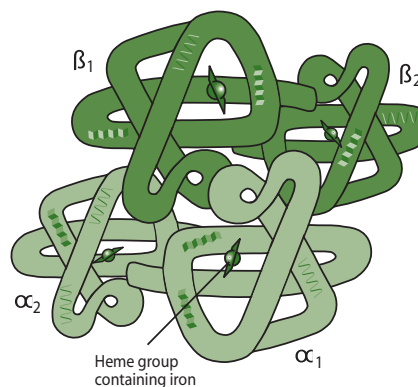$α_1$

Heme group containing iron

Figure 749 Quaternary structure

Each of these four chains is also combined with a prosthetic group, the heme group. This is not a polypeptide but an iron-containing molecule. In *Figure 749*, the heme groups are shown as rectangles with (iron) spheres in the middle.

Proteins with a prosthetic group are called conjugated proteins. Chlorophyll and many of the enzymes in the electron transport chain are conjugated proteins.

Chapter 7 Revision Questions

Multiple Choice Questions

1   The complex of DNA and histone protein found in eukaryotic nuclei is known as:

   A   a histone

   B   a nucleosome

   C   genes

   D   chromatin
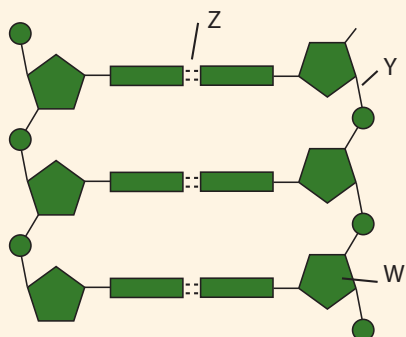
2   What is the function of nucleosomes?

   A   to help supercoil DNA

   B   to facilitate transcription

   C   to facilitate translation

   D   to help DNA replication

3   The units of genetic information are called:

   A   histones

   B   nucleosomes

   C   genes

   D   chromatin

4   The diagram shows a schematic representation of a section of DNA. Identify each of the labelled components.

	W	X	Y	Z
A	deoxyribose	3' end	covalent bond	hydrogen bond
B	deoxyribose	5' end	hydrogen bond	covalent bond
C	ribose	3' end	covalent bond	hydrogen bond
D	ribose	5' end	hydrogen bond	covalent bond

5   DNA replication occurs in a

   A   random direction

   B   helical direction

   C $3' \rightarrow 5'$ direction

   D $5' \rightarrow 3'$ direction

6   The diagram illustrates DNA replication. X marks the:
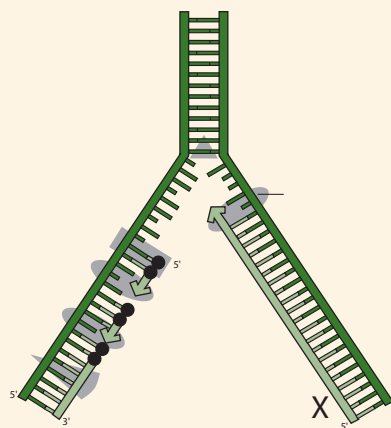
   A   lagging strand

   B   leading strand

   C   DNA polymerase molecule

   D   Okazaki fragment

7   Which of the following is not an enzyme involved in DNA replication?

   A   helicase

   B   DNA ligase

   C   DNA polymerase

   D   amylase

AHL

8   Transcription is carried out in the following way:

   A   3′ → 5′ direction of the new RNA strand, using the DNA sense strand as template

   B   3′ → 5′ direction of the new RNA strand, using the DNA anti-sense strand as template

   C   5′ → 3′ direction of the new RNA strand, using the DNA sense strand as template

   D   5′ → 3′ direction of the new RNA strand, using the DNA anti-sense strand as template

9   Initiation, elongation and termination are the main stages in:

   A   transcription

   B   translation

   C   replication

   D   decomposition

10   Which of the four levels of organisation in proteins may involve a non-protein molecule?

   A   primary structure

   B   secondary structure

   C   tertiary structure

   D   quaternary structure

Short Answer Questions

11   Explain the structure of double stranded DNA. Use a drawing of a segment of 6 linked nucleotides to enhance your explanation.

12   What is the function of each of the following enzymes in DNA replication?

   (a)   DNA gyrase

   (b)   helicase

   (c)   single strand binding proteins

   (d)   RNA primase

   (e)   DNA polymerase III

   (f)   DNA polymerase I

   (g)   DNA ligase

13   Outline the differences and similarities between transcription and translation.

14   What is the role of the following in the process of translation?

   (a)   mRNA

   (b)   ribosomes

   (c)   codons and anticodons

   (d)   tRNA

15   State the types of bonds that are involved in each of the four levels of protein structure.

16   State four functions of proteins. Give an example of each.

17   Explain why both polar and non-polar amino acids are present in the protein molecules that are part of the cell membrane.

18   Describe the secondary structures as shown in the computer-generated model of a protein shown below.