Summary of Weeks 5 - 7

DNA replication.
From DNA to mRNA.
mRNA to Protein.

Components of DNA

DNA is built a bit like a ladder with sides made of deoxyribose and phosphate groups linked together.
Notice the anti-parallel nature of the sides.
The rungs are made of pairs of nitrogen bases either –
Adenine with Thymine or Guanine with Cytosine.
To make the model, label the ribose shapes and number the carbon atoms 1 – 5.
Roughly lay out the ribose shapes with carbon 1 of each molecule adjacent to carbon 1 of the ribose on the other side. Then join them using the phosphate groups through carbons 3 and 5.
Label the nitrogen bases. Adenine A and guanine G are the large molecules (shapes), and thymine T and cytosine C are the small shapes. They link to ribose through carbon 1 and to one another by hydrogen bonds.

Discovery of DNA structure

The sentence "This structure has novel features which are of considerable biological interest" may be one of science's most famous understatements. It appeared in April 1953 in the scientific paper where James Watson and Francis Crick presented the structure of the DNA-helix, the molecule that carries genetic information from one generation to the other.
Nine years later, in 1962, they shared the Nobel Prize in Physiology or Medicine with Maurice Wilkins, for solving one of the most important of all biological riddles. Half a century later, important new implications of this contribution to science are still coming to light.
Way back in 1868, a young Swiss physician named Friedrich Miescher, isolated something no one had ever seen before from the nuclei of cells. He called the compound "nuclein." This is today called nucleic acid, the "NA" in DNA (deoxyribo-nucleic-acid) and RNA (ribo-nucleic-acid).

Two years earlier, the Czech monk **Gregor Mendel**, had finished a series of experiments with peas. Mendel was able to show that certain traits in the peas, such as their shape or colour, were inherited in different packages. These packages are what we now call genes.

In 1944 the American scientist Oswald Avery managed to transfer the ability to cause disease from one strain of bacteria to another. But not only that: the previously harmless bacteria could also pass the trait along to the next generation. What Avery had moved was nucleic acid. This proved that genes were made up of nucleic acid.

The genetic code

ATGC are the four letter alphabet used to code for amino acids in the sequence that make proteins.
We've seen proteins control the processes of the cell. So if DNA controls which proteins are made and when, then DNA controls the cell and the organisms that cells form.
There are only 20 amino acids to code for so only 3 letters of the four possible are needed to code for any one amino acid.
You remember from last week that sickle cell anaemia occurs when one amino acid is wrong in the haemoglobin molecule and that in turn is caused by one letter being wrong in the code.
We'll look in detail at how DNA produces proteins (Protein synthesis) next time.
For now lets consider what happens when a cell divides.

DNA replication

DNA replication starts at replication origins.
Each strand acts as a template to make a new half – **semi conservative.**
DNA replication involves the enzymes DNA polymerase.
DNA polymerase can add subunits only to the 3' end of each strand so "growth" occurs from the 5' to the 3' end.
The subunits are **nucleotides** (sugar, phosphate and nitrogen base) which exist in the cell already.
One strand, the leading strand is easy to replicate but the other, the lagging strand has to be done in sections (Okazaki fragments) which are later joined together.
DNA replication is quite well understood though some details have only very recently been worked out.

Mutations are changes to the genetic code.

Mutations in germ cells (eggs, sperm and pollen) are passed on to offspring – all the cells in the offspring have the mutation.
Mutations in somatic cells must also be avoided to prevent cell functions being compromised in particular cells.
DNA replication rarely makes mistakes and has a system the **mismatch repair system** to correct things. Recently colon cancer has been linked to a mutation in a gene responsible for producing a mismatch repair protein. Since most cancers result from cells that have accumulated multiple mutations, a cell deficient in mismatch repair has an increased chance of being cancerous.
DNA is also damaged by processes naturally going on in the cell.

DNA replication some more details.

The units that are added to the template strand of DNA are nucleoside triphosphates. In this form energy is available for the polymerisation reaction.
Replication forks – where the DNA double helix has been opened up– replication machines move away from one another and the replication origin eventually meeting up with another further along the chain.

Replication origins often occur where there is a lot A-T which are held together with fewer H bonds and are therefore easier to open.
Because DNA polymerase can only add units to the C3 of dioxyribose, replication only occurs in the 5' to 3' direction. For the other strand – lagging strand – replication takes place in section that are later joined. The leading strand is done in a continuous sequence.

The replication Machine.

The replication machine is a group of proteins that can move along the DNA strand giving rise to replication in both strands.
The machine includes –
Helicase that uses the energy of ATP to open the double helix,
Single-strand binding protein, this stops the bases rebinding again,
Sliding clamp protein holds the polymerase molecules in place.
These proteins are held together in a multienzyme complex the replication machine.

Repair mechanism

Mismatched pairs –– distorts the double helix. Mismatch repair proteins detect this distortion and cut out a part of the new strand contain the mismatch. DNA polymerase and DNA ligase – joins lengths of DNA – repairs the gap.
Thousands of random chemical changes occur every day to human DNA but because there are two complimentary strands one can be used as a template for repair as it is for replication.

Sequence similarity between Species

Humans and Chimpanzees diverged about 5 million years ago but their DNA is still 98% similar because of the faithful replication of DNA that is more detailed and longer than any book.
A human chromosome may be 250 million base pairs long. Here is the simplest genome, that of a virus

From DNA to Protein

Those parts of the DNA which code for proteins (we might call them genes) do not directly take part in protein synthesis. DNA is much too important. So the relevant section of the code is copied in a complimentary fashion - except that RNA has Uracil (U) instead of Thymine - to Messenger RNA or mRNA. This process is called **transcription** and is catalysis by **RNA polymerase**
The mRNA then leaves the nucleus and enters the cytoplasm where protein synthesis takes place.
Consider this sequence of DNA –
ATTCGTGCGTATTATCGSTA
What would the sequence be in mRNA?

mRNA production

DNA ATTCGTGCGTATTATCGTTA

RNA UAAGCACGCAUAAUAGCAAU

mRNA synthesis is fast 1500 bases per minute and many can be synthesised in rapid succession.

RNA polymerase has no proofreading mechanism as RNA is not a permanent store of genetic information. It makes about 1 error in 10^4.There are 3 other types of RNA two of which are involved in protein synthesis.

Starting transcription

RNA needs to be able to recognise the beginning of a gene so it “knows” when to start transcription. This process differs in prokaryotes and eukaryotes and is the way cells regulate which protein are produced. We’ll look at this **gene regulation** later.

In prokaryotes – bacteria – polymerase binds weakly with DNA when it randomly collides with it. It then slides along the DNA until it encounters a promoter (start signal region) where it binds more tightly.

The double helix of DNA is then opened and nucleoside triphosphates then form a complimentary strand of mRNA.

This is a bit like DNA replication though mRNA is only single stranded.

Since DNA is a double strand two RNA molecules could in theory be made however the promoter is asymmetrical and polymerase binds in only one orientation and proceeds in the 5’ to 3’ direction.

Different genes will be transcribed from different strands and therefore in different direction.

Why is mRNA in the cytoplasm shorter than that in the nucleus?

Early in the 1970s it was observed in eukaryotes that only 5% of the mRNA made in the nucleus got into the cytoplasm.

In 1977 it was discovered that while bacteria genes were in an uninterrupted sequence, in eukaryotes the coding sequence was interrupted by noncoding sections called **introns**. Introns range in length from 80 to 10,000 bases.

So after the mRNA is initially coded the introns are cut out by **RNA splicing** carried out a complex of enzymes and RNA called **small nuclear ribonuleoprotein particles** (snRNAs, pronounces snurps)

How long does mRNA last in the cell?

It varies. In bacteria each molecule last for about 3 minutes. In eukaryotes it is between 30 minutes and 10 hours. In general proteins made in large quantities will have mRNA that lasts a longer time

mRNA to protein

Transcription simply produces a complimentary version of the genetic code of one gene that can be safely allowed into the cytoplasm.
Translation is the process where the code is used to produce a particular sequence of amino acids i.e. a protein.
The genetic code, as we saw in DNA, is made of three out of the four letters. That gives 4^3 or 4x4x4=64 possible combinations. Since only 20 amino acids are commonly found in proteins more than one **codon** (triplet) can be used for the same amino acid.
It is vital that mRNA is read from the correct starting point to get the right sequence, this is known as the **reading frame.**

Transfer RNA (tRNA) comes in as many kinds as there are amino acids

The mRNA message is decoded on Ribosomes

DNA	ATTCGTGCGTATTATCGTTA
mRNA	UAAGCACGCAUAAUAGCAAU
tRNA	?

Protein breakdown in the cells helps regulate the amount of protein in the cell

Proteins vary enormously in their lifespan.
Structural proteins in bone and muscle may last for months or even years. Is this a surprise?
Some enzymes concerned with controlling **metabolic** (remember metabolism = all chemical reactions in the cell) **processes** last much shorter times days or hours.
Proteins are destroyed by the enzymes **proteases.**
Ubiquitin marks proteins for destruction.
Faults in the protein can probably be detected that lead to there being tagged for destruction.
The amino acids can be used again to make more protein

Life may have begun with RNA

The processes that led to the formation of the first living cells may have begun with RNA as it is able to carry information – like DNA – and catalyse reactions.
RNA molecules with catalytic properties are called ribozymes
DNA and proteins may therefore have been later additions.