ELEC3223 Intro to BioNanotechnology
ELEC6205 BioNanotechnology

DNA

Size matters

0.1nm  1nm  10nm  100nm  1μm  10μm  100μm

Atoms
Blood Cells
HSV
Phage

DNA

Drugs, Amino acids

Nerve Cells

Proteins

HSV

Bacteria

Drugs, Amino acids etc

Proteins

Virus (Phage)

Blood Cells

Bacteria

Nerve Cells
DNA

- **Deoxyribo Nucleic Acid** – double stranded string of molecules called nucleotides
- Consists of 4 different building blocks - **NUCLEOTIDES** (also called bases) (recall proteins made from 20 amino acids)
  - and a **SUGAR** (deoxyribose)
  - and a **PHOSPHATE**
- EVERY KNOWN LIVING THING CONTAINS DNA
The Book of Life

DNA is found within all cellular life (microbes, plants, animals) is a set of instructions on how to build and reproduce a living organism.

The instructions are written in a code of 4 chemical letters:

- Guanine
- Adenine
- Thymine
- Cytosine

The Book of Us

Our instruction book consists of 3.2 billion letters.

Bases

- A – Adenine
- T – Thymine
- G – Guanine
- C – Cytosine

U – Uracil (sometimes found instead of T and in mRNA)

The sequence of these codes encodes the amino acid sequence of all proteins – i.e. life.

Each amino acid in a protein is coded by three consecutive nucleotides (called a triplet code).

DNA: The Alphabet of Life

- The entire sequence can be described by the sequence of only one of the strands.
- Two nucleotides that bind together are called a base pair.
- Two single strands that bind together are said to be “complementary.”
- Hydrogen bonding keeps the DNA together
  - A forms two hydrogen bonds with T
  - C forms three hydrogen bonds with G
- The more bases that are complementary the stronger the association of the two single strands.

Some facts
DNA is a **right handed** double helix

Each strand has a different end –
Free sugar (3′) or
Free phosphate (5′)
Sequence **starts** at 5′ end

- OH on 3′ end

- Transition from double stranded to single stranded DNA is characterised by a melting curve.
- The 50% point is called the $t_m$.
- (Optical absorbance of ss-DNA is different from ds-DNA).
- The more C-G in the DNA the higher the melting temperature (C-G has three H-bonds, A-T has two).
- $t_m$ is v. important in characterising the DNA

**Denaturation and renaturation**

When dsDNA is heated to 95°C the two strands fall apart, the H-bonds are broken. It denatures like a protein. When it cools it re-forms the original shape – not like a protein.

Denaturation of DNA is called melting

**Two complementary strands will bind to each other**

**Hybridisation - solution**

Different bits of DNA - called oligo-nucleotides (oligos)

Two complementary single strands (oligos) join to form one piece of ds DNA
**Hybridisation – on surfaces**

Different bits of DNA can also be stuck on micron sized beads.

**DNA in Cells - chromosomes**

- Humans have 23 pairs, 46 chromosomes (Yeast has 16)
- Choromosomes are v. long bits of DNA (also contain some proteins)
- Gene- bits of DNA that code for a particular protein. (approx. 20,000 genes in a human, that code for proteins)
- Loci – small section of a gene

**A Gene is a piece of DNA with a particular function**

- Usually encodes a protein
- ~20,000 genes in a human found in large pieces
- ~5,000 genes in *E. coli*

**DNA – storage of information**

**messenger RNA** (mRNA) – “working copy” used to make proteins

**Protein** – molecular machines and structures

**The key fact is that all of these are extended 1-D polymers made of similar subunits.**

We contain $10^{14}$ cells, each with 2m DNA, giving a total length of $2 \times 10^{11}$ km (earth to sun = $1.5 \times 10^8$km).

**Sequence information**

- The sequence of nucleotides along a DNA strand defines the **genetic code**
- The code consists of **three** letter words – called **CODON** (e.g ACT, TTC, CAG)
- Codons are translated via **messenger** (mRNA) and then **transfer** (tRNA) to eventually build protein molecules.
- 64 possible codons (4 bases in 3 places = $4^3$) encoding for 20 amino acids.
- There are three **STOP** (or nonsense) Codons that signify the end of a region – UAA; UGA; UAG
• There are many regions of DNA that are not translated – these are called intervening sequences or **introns**.

• Humans are 98.8% genetically identical to chimpanzees, Why does the tiny 1.2% make such a difference?

• The coding regions are called **exons**.
• Many more introns than exons
• Only 1.5% of the human genome consists of protein coding regions of the DNA – exons.
• Rest is unknown junk DNA – the introns

---

**Central dogma of molecular biology**

• Three major processes in cellular utilisation of genetic information:
  1. **Replication** – copying of parental DNA to form daughter DNA with identical sequence information.
  2. **Transcription** – the genetic message on the DNA is copied to RNA
  3. **Translation** – genetic message in RNA is translated on the ribosomes into a polypeptide with a particular sequence of amino acids and thence a **protein**.
1. Replication of DNA in cells

- When cells reproduce they need to make copies of DNA.
- Double strands are divided into two complementary single strands.
- These single strands are used as templates to produce a complementary strand.
- These are then copied to produce two new (almost) identical copies.
- Each template and its new complementary strand then forms a new DNA double helix, identical to the original.
- Mistakes can occur – mutations.

- Before replication, the DNA double helix must be unwound and the strands separated, like the two sides of a zip.
- Done by breaking the weak hydrogen bonds that link the paired bases, by enzymes – helicases. Can also be broken by heat.
- Once the DNA strands have been unwound, they must be held apart to expose the bases so that new nucleotide partners can hydrogen-bond to them.
- The enzyme **DNA polymerase** then moves along the exposed DNA strand, joining newly arrived nucleotides into a new DNA strand that is complementary to the template.
REPLICATION occurs at many places simultaneously

- Human DNA is very long (80 million base pairs in a single chromosome) it unzips at multiple places along its length
- The replication process proceeds simultaneously at hundreds of places along the length of the chain.
- Eventually these areas run together to form a complete chain.
- In humans, DNA is copied at about 50 base pairs per second. The process would take a month (rather than the hour it actually does) without these multiple places on the chromosome where replication can begin.

DNA replication is extraordinarily accurate.

- DNA polymerase makes very few errors, and those that are made are quickly corrected by DNA polymerase and other enzymes that "proofread" the nucleotides added into the new DNA strand.
- If a newly added nucleotide is not complementary to the one on the template strand, these enzymes remove the nucleotide and replace it with the correct one. With this system, a cell's DNA is copied with less than one mistake in a billion nucleotides.
- This is equal to a person copying 100 large (1000 page) dictionaries word for word, and symbol for symbol, with only one error for the whole process!

2. Transcription

- Enzymes convert the genetic information in the dsDNA into an RNA strand with a base sequence complementary to one of the DNA strands.
- Three kinds of RNA are made:
  - messenger RNA (mRNA) encodes the amino acid sequence for the polypeptides (corresponds to a gene).
  - transfer RNA (tRNA) reads the mRNA information and transfers the appropriate amino acid onto a growing polypeptide strand.
  - ribosomal RNA (rRNA) - constituents of the ribosomes. Ribosomes are the machines that make the proteins from the polypeptides.

3. Translation

Reading the code on the mRNA tells us exactly which proteins the DNA is encoding for. Remember most of the DNA is junk!
mRNA is much more useful for decoding a cell's protein manufacturing capacity than reading the DNA

- Synthesis of proteins from the mRNA code is done using tRNA and rRNA, with ribosomes. Complex process.
(See text books for further details if you are interested)
mRNA codes for one strand of the DNA, replacing the T with U. Note the coding in triplets.

A DNA sequence is called "sense" if the sequence is the same as that of the mRNA copy that is translated into protein.

This one is antisense
Replication of DNA in a test tube

Polymerase Chain Reaction PCR (Kary Mullis 1983)
Revolutionised biotechnology and molecular biology. Allows a **SINGLE** molecule of DNA to be replicated into billions of identical copies, for subsequent detailed chemical analysis (sequence determination)

---

How PCR Works

**Step 1 - Denaturation (optimal temperature is 94°C)**
By heating the DNA, the double strand melts and opens to single stranded DNA.

**Step 2 - Annealing (optimal temperature is 60°C)**
Single-stranded DNA primers bind to their complementary single-stranded bases on the denatured DNA.

Primers are little **fragments** of DNA (20-30bases) that are added to the solution for the next stage to proceed. They match exactly the beginning and the end of the DNA fragment to be amplified. Their $T_m$ is usually 50 to 60°C.

**Step 3 - Extension**
72°C is the ideal temperature for the Taq polymerase to attach and start copying the template. The result is two new identical helices instead of one.
Note: the polymerase cannot start adding nucleotides to nothing. It has to add a nucleotide to an existing piece of DNA—the primer. This is a short strand that is complementary to the DNA to be copied.
Repeating the cycle several times gives large amounts of DNA. Starting with one DNA molecule after just 20 cycles there will be a million copies and after 30 cycles there will be a billion copies.

The bacterium *Thermus aquaticus* was first discovered in several springs in the Great Fountain area of the Lower Geyser Basin at Yellowstone National Park.

**Secret compound – Taq polymerase**

*Thermus aquaticus*

- Enzyme that polymerises the synthesis of the DNA BUT is stable at high temperatures, allowing denaturation of the DNA and repeated cycles
- The taq-polymerase needs ca. 1 min to synthesise 1 kbp. So the synthesis time depends on the length of your product.