# 1 Synthesis of Proteins

**Definition 1.1. Condensation reaction** involves the removal of a molecule of water from the reagents.

If $n$ aminoacids are stitched together, $n-1$ peptide bonds form in a **polypeptide chain** with two ends, one called *C-terminus* close to a carbonyl carbon and the other *N-terminus* close to an amide nitrogen, with the alkyls attached to a main chain/peptide bond. The N-terminus is to the left by convention.

In the carbonyl bond, a carbonyl carbon is bonded to an amide nitrogen.

The order of aminoacids is very important.

**e.g.** Leu-Enkephalin, a pentapeptide of Tyr-Gly-Gly-Phe-Leu, is a natural opioid peptide which modulates the perception of pain. Leu-Phe-Gly-Gly-Tyr does not have any effects.

Hydrogen bonds consists of a polypeptide backbone, with hydrogen bonds as stabilisors forming between the carbonyl-oxygen and amide hydrogen. The bonds are 4 aminoacids apart and within the same strand.. Alkyl groups do not increase the stability of the helix.

**Question.** Why does an alpha helix has a secondary structure?

## 1.1 $\beta$ sheet

**Remark 1.2.** Oxygen is coded as red. Carbon is coded as black. Hydrogen is coded as white. Nitrogen is coded as black. Alkyl groups are coded as purple.

The $\beta$ sheet also has a backbone with hydrogen bonds as stabilisors forming between the amide hydrogen and carbonyl oxygen.

Typically there is a hydrogen bonding between aminoacids in different strands (usually in the same polypeptide chain).

## 1.2 Specialised Alpha Helix: Coiled Coil

Although the alkyl groups do not stabilise the helix, if two stripes of side chains in the helix are such that one of them is hydrophobic and the other is hydrophilic, then a coil will be coiled, giving it different biochemical and biophysical properties in different basis. This coiled coil is called *amphipathic*, has a supersecondary structure.

## 1.3 Tertiary Structure: Rhodopsin

Phodopsin has a three-dimensional structure held together by hydrophobic interactions, non-covalent bonds and covalent disulfide bonds. In this case, a single polypeptide chain makes up the proteins.

## 1.4 Protein Domains

The domains are often specialized for different functions:

- Structural and functional unit
- Independent folding
- Independent/semi-independent function
- Important in the evolution of proteins

Note that one polypeptide chain may have multiple domains.

## 1.5 Quaternary Structure: Hemoglobin

- Hemoglobin protein formed from different subunits: $2\alpha$, $2\beta$
- Each subunit is a separate polypeptide
- Sickle cell anaemia is caused by a mutation in the $\beta$ subunit

Haemoglobin:

- Transports $O_2$ from lungs to tissues
- Heterozygotes fro the sickle cell anaemia mutation in the $\beta$ globin gene are partially protected against malaria.
- Frequency of the sickle cell allele has reached highest levels in Africa and India
- A related molecule, myoglobin, is

## 1.6 Multiprotein Complexes

- 1 polypeptide in 1 functional unit
- Many proteins in one structure/molecular machine

Specific data is needed to differentiate between quaternary structures and multiprotein complexes.

## 1.7 Proteomics

**Definition 1.3. Proteomics** is a study of methods to separate, isolate and analyze proteins.

One of the methods involves two-dimensional gel electrophoresis followed by mass spectrometry.

Gel electrophoresis separates the proteins and allows to choose a method for their isolation.

An unknown protein can be treated with trypsin, which breaks the protein into peptide fragments and then analyze them by mass spectrometry.

# 2 Genes and Chromosomes

Genome is the entirety of an organism's hereditary information.

Almost all genetic information is stored in DNA, but some viruses have the genetic information stored in RNA.

## 2.1 Human genome

23 maternal chromosomes are maternal, 23 chromosomes are paternal. There are 3 billion base pairs per genome, with 25 000 genes spread across 23 chromosomes.

Genomes can come in all sizes. Bacteriophages, for example, have 48000 base pairs.

*E.coli*, a standard prokaryote, has 4.6 billion base pairs, while mitochondria have 16000 base pairs. Over the evolutionary periods, some of the mitochondrial DNA has moved to the nuclear DNA.

Eucalyptus globulus have chloroplast DNA with 160000 base pairs.

## 2.2 Comparing genome sizes

There are about 21000 genes encoding proteins, while having about three billions of base pairs.

Taegleria fowleri, a brain-eating amoeba, has 670 billion base pairs.

About 501.5

Non-repetitive DNA that is neither in introns nor codons are not transcribed, but they can help to determine how much and when to transcribe.

Thousands to millions of sequences are repeated. Transposons are regions that can cut themselves out of the DNA, make a copy and then get inserted back. There are also remnants of virus infections (SINEs), as well as long interspersed nuclear elements (LINEs, typically greater than 500 base pairs).

## 2.3 Packaging of DNA in the Cell

In a non-packaged state, even the small prokaryotic genome would occupy a considerable portion of the cell volume, which is even a greater probelem in eukaryotes.

In prokaryotes, the DNA forms the prokaryotic nucleoid.

Topoisomerases are enzymes that wind and unwind DNA.

There are 6 billion base pairs per cell, 2 meters of DNA per cell, located in the nucleus 6 $\mu$m wide, which is a geometric equivalent of packing 40 km of fine thread into a tennis ball.

## 2.4 Structure of the Nucleus

All DNA is in the nucleus.

There are two envelopes of phospholib

Nuclear pore is used for movement in and out of the nucleus.

Chromatin structures hold DNA together in the nucleus.

Nucleoplasm is a fluid in the nucleus.

Nucleolus is the place where a ribosome begins to be built.

The eukaryotic genome is packed into cells using chromosomes.

In this way, a karyotype can be constructed which pairs chromosomes in numerical order, which is an excellent diagnostic aid.

Chromosome painting hybridization can be perfomed in order to distinguish a variety of chromosomes (get the probes, mark them with fluorescent labels and heat the solution with the chromosomes up).

FISH (fluorescence in situ hybridization) can also be administered to test for the presence of a particular DNA sequence (heat the probe, heat the sample and cool it down).

Each chromosome contains a single, long, linear DNA molecule and associated proteins (clled chromatin). Chromatin is tightly packaged and dynamic.

**Question.** What is a chromatid?

There are 8 different proteins forming core histones. The DNA is wrapped around $1.6 - 1.8$ times in a nucleosome core particle, with about 146 base pairs in the loop. Linker DNA connects histones (proteins rich in lysine and arginine) and thus constitutes a nucleosome if taken together with a nucleosome core particle (consider an analogy of beads on a string).

A linker histone clips the DNA to a histone.

Positive charge of histones neutralizes a negative charge of DNA.

DNA is not always completely loose or tightly packages and changes its topology depending on the conditions.

Cohesin may be involved in forming chromatin loops, which can also be attached to a nuclear scaffold to enable increased compactification.

To sum up, each DNA molecule is packaged into a mitotic chromosome that is 10000 shorter than its extended length. Note that packaging and unpackaging requires ATP.