1 Tree of Life

- Prokaryotes: Eubacteria

  e.g. Cyanobacteria are capable of photosynthesis.

- Prokaryotes: Archaea

  DNA replication, transcription, translation are similar to these of eucarya, and hence archaea are closer to eucarya than to eubacteria.

  e.g. Methanococcus: predominantly live in hydrothermal vents on the bottom of the ocean

  e.g. Methanobacterium: anaerobic, low pH conditions

- Eucarya

  e.g. animals

2 Study of Diversity

1. XIX and early XX: use large collections and study each species independently
2. Late XX and XXI: use model organisms

Definition 2.1. Model organisms are species that have been studied intensively over a long period of time and thus serving as models for deriving fundamental biological principles.

General attributes:

- Rapid development with short life cycles
- Small size
- Readily available
- Tractable – easy to manipulate and modify
- Have understandable genomics

e.g. elephant is a horrible model organism!

e.g. eukaryotes: zebra fish

Question. How does the diversity can be explained?

Answer. Central dogma:

$$\text{DNA} \xrightarrow{\text{transcription}} \text{RNA} \xrightarrow{\text{translation}} \text{protein}$$

2.1 Central Dogma

There are different kinds of RNA:

- mRNA – produces protein by translation
- tRNA – transport AAs and for protein synthesis
- rRNA – part of the ribosome

There are other kinds of RNAs, some have nothing to do with making proteins.

2.2 Elaborated Central Dogma

genome: complete set of DNA sequences in a cell or organism

transcriptome: complete set of RNA sequences in a cell or organism

proteome: complete set of protein sequences in a cell or organism

DNA $\xrightarrow{\text{transcription}}$ RNA $\xrightarrow{\text{translation}}$ Protein

organisation
replication

interactome
a complete set of interactions in a cell or organism

metabolome
a complete set of small molecule metabolites in a cell or organism

phenome: a complete set of phenotypes

3 Review

DNA, RNA and proteins are linear chains of information. The order of aminoacids contains the information.

Information in nucleic acid sequence is translated into an aminoacid sequence via a genetic code which is essentially universal among all species.

3.1 Eukaryotes

Nuclear agenda:

1. Transcription
2. RNA processing
3. Nuclear export
4. Translation
5. Protein folding and modification

3.2 Nucleic Acids

1. The genetic material in a cell is a blueprint for an organsim.
2. DNA is a deoxyribonucleic acid
3. Ribonucleic acid

Three parts of a nucleic acid:

1. Pentose sugar – scaffold for a base
2. Nitrogenous base – varies
3. Phosphate group – backbone, can be 1P, 2P's or 3P's

3.2.1 Bases

- Purines – guanine and adenine (Al Gore stings - Pu!)
- Pyrimidines – cytosine, thymine and uracil (U C The Pyramides)

3.2.2 DNA vs RNA

1. DNA: G, C, A, T, RNA: G, C, A, U
2. DNA: Deoxyribose, RNA: ribose

3.2.3 Nucleic Acid Nomenclature

1. Nucleoside monophosphate: sugar + base + 1P
2. Nucleoside diphosphate: sugar + base + 2P
3. Nucleoside triphosphate: sugar + base + 3P

   Nucleoside: base + sugar Nucleotide: base + sugar + at least one P

Remark 3.1.
Adenoside: nucleoside
AMP: nucleotide, nucleoside monophosphate
ADP: nucleotide, nucleoside diphosphate
ATP: nucleotide, nucleoside triphosphate

4 Molecular Interactions and Introduction to Proteins

In nature, structure and function are closely intertwined.

Structure of DNA makes it stable and allows it to function as a hereditary material.

4.1 Molecular Interactions

Interactions between individual molecules usually mediateed by noncovalent attactions (within a molecule, covalent nature of interactions is prevalent, cf Karp's p.33-42)

1. Electrostatic attractions
2. Hydrogen bonds
3. van der Waals attractions
4. Hydrophobic interactions the tendency of nonpolar molecules to aggregate to minimize interaction with surrounding polar molecules

4.2 Properties of Nucleic Acids

1. DNA is synthesised from deoxyribonucleoside triphosphates, otherwise known as dNTPs
2. RNA is synthesised from ribonucleoside triphosphates, also known as NTPs

3. Nucleotides are linked by phosphodiester bonds (links 1' carbon of one nucleotide to 3' carbon of the other nucleotide)

Base pairing holds the DNA double helix together with the H-bonding:

1. A - T: 2 H-bonds
2. G - C: 3 H-bonds

However, H-bonds are not the only interactions holding DNA together – van der Waals attractions (in general, between the bases) and hydrophobic interactions (in general, between the bases) also play a role.

Between two strands the interactions are noncovalent – strictly speaking, DNA consists of two molecules. However, some scientists consider DNA as a single molecule composed of two strands. Both are fine.

DNA of one complete turn have on average 10.5 base pairs.

Gaps on either side are not the same – there is a major groove on one side, and a minor groove on the other.

5' end is close to a phosphate (-PO_4), 3' is close to OH.

The strands in a double helix are antiparallel, with DN backbone having a negative charge. Base stacking contributes significantly to the stability of the helix.

At high enough temperatures (called melting temperature) or at high pH, the noncovalent bonds are broken, which denaturates DNA. If cooled or if pH is lowered, the bases will recombine.

Covalent bonds can also be broken above 100 °C, but this situation is not a focus of biology.

The sequence of the two strands is complementary, and the strands can be unzipped. This is important for DNA replication (including PCR) and RN synthesis.

G-C rich DNA is more stable at high temperatures.

4.3 Introduction to Proteins

1. Primary (sequence)

   e.g. AA sequence

2. Secondary (local folding)

   e.g. a helix, β sheet

3. tertiary (long-range folding)

   e.g. 3D structure

4. Quaternary (multimeric organization)

   e.g. multiple polypeptide chains

5. Multiprotein complexes

   e.g. molecular machines

4.4 Protein Structure

Proteins are composed of amino acids.

The amino acid side-chain, or R group, is variable and determines the type of amino acid.

4 main categories

- polar charged
- polar uncharged
- nonpolar
- those with unique properties

α carbon is attached to the side chain, amino and carboxyl group.

Cysteine looks different in reduced conditions (found in cytosol) and oxidized conditions (lumen of organellesd). A disulphide bond often acts as a base to stabilise the protein structure.

Definition 4.1. The **genetic code** is the set of rules specifying the correspondence between nucleotide triplets (codons) in DNA or RNA and amino acids in proteins.

It is degenerate (multiple options for some amino acids) and almost universal.

Question. What is the minimum number of mutational steps between amino acids?

Answer. The number of mutations between codons for different amino acids.

Example 4.2

How many mutational steps required to get from a codon for proline to one for cysteine, at a minimum?

Answer. Two, CCG $\rightarrow$ UGC

Groups of aminoacids with similar properties tend to be clustered in the genetic code table. Codons of aminoacids with similar properties tend to have fewer mutational steps between them. One random mutation in a codon is less likely to result in a dramatic change in amino acid properties than two random mutations.