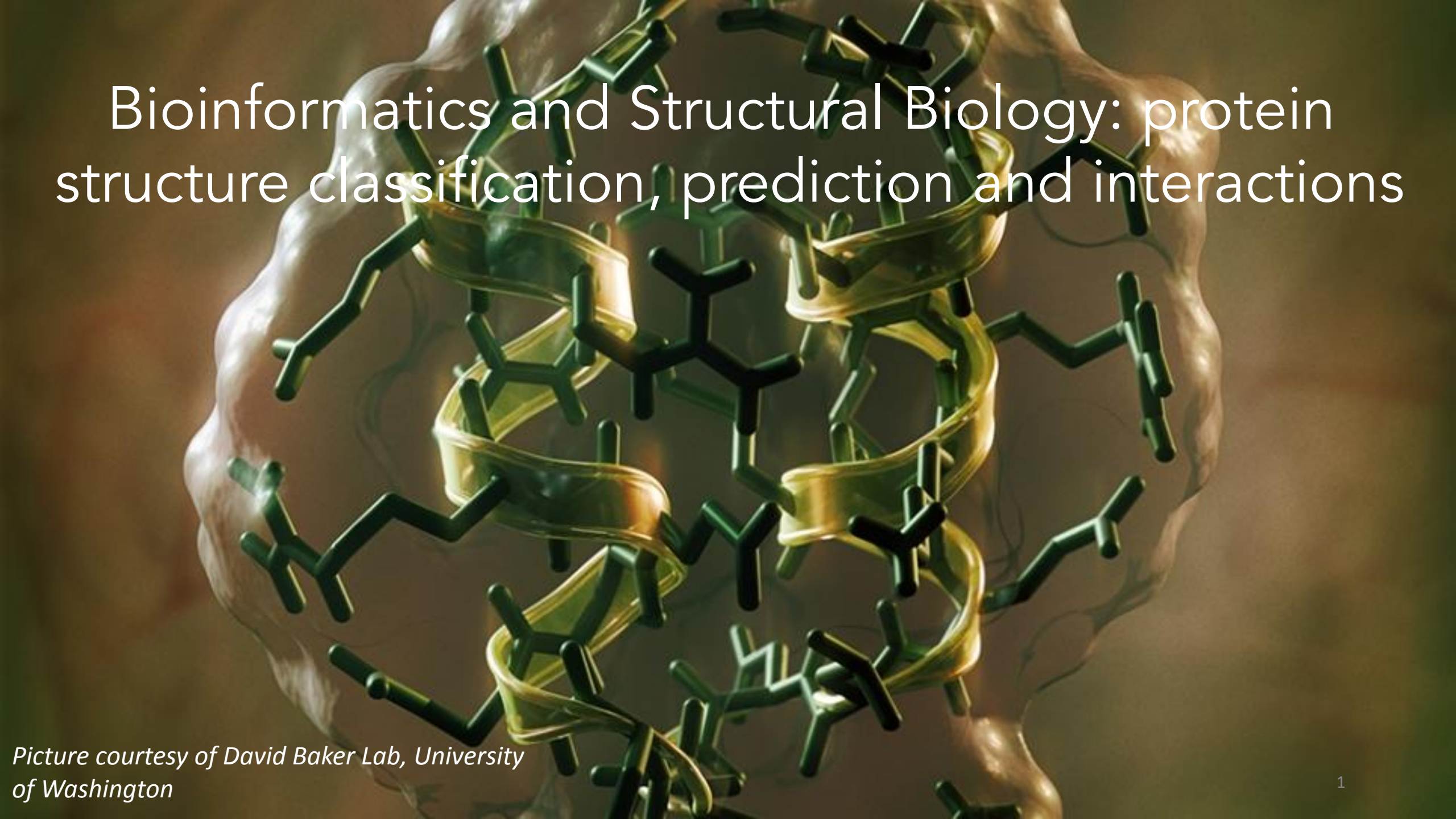
Bioinformatics and Structural Biology: protein structure classification, prediction and interactions

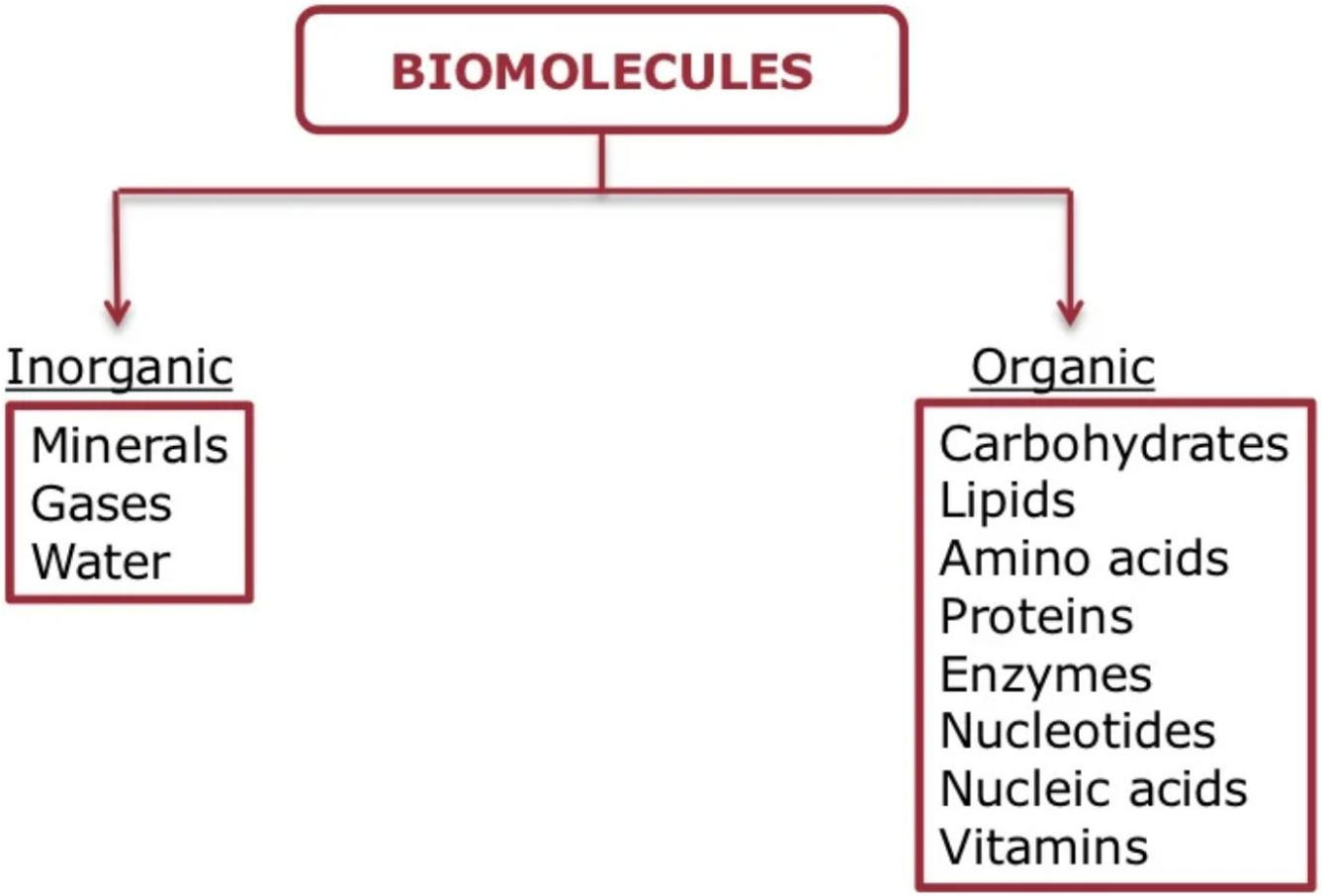


Picture courtesy of David Baker Lab, University of Washington

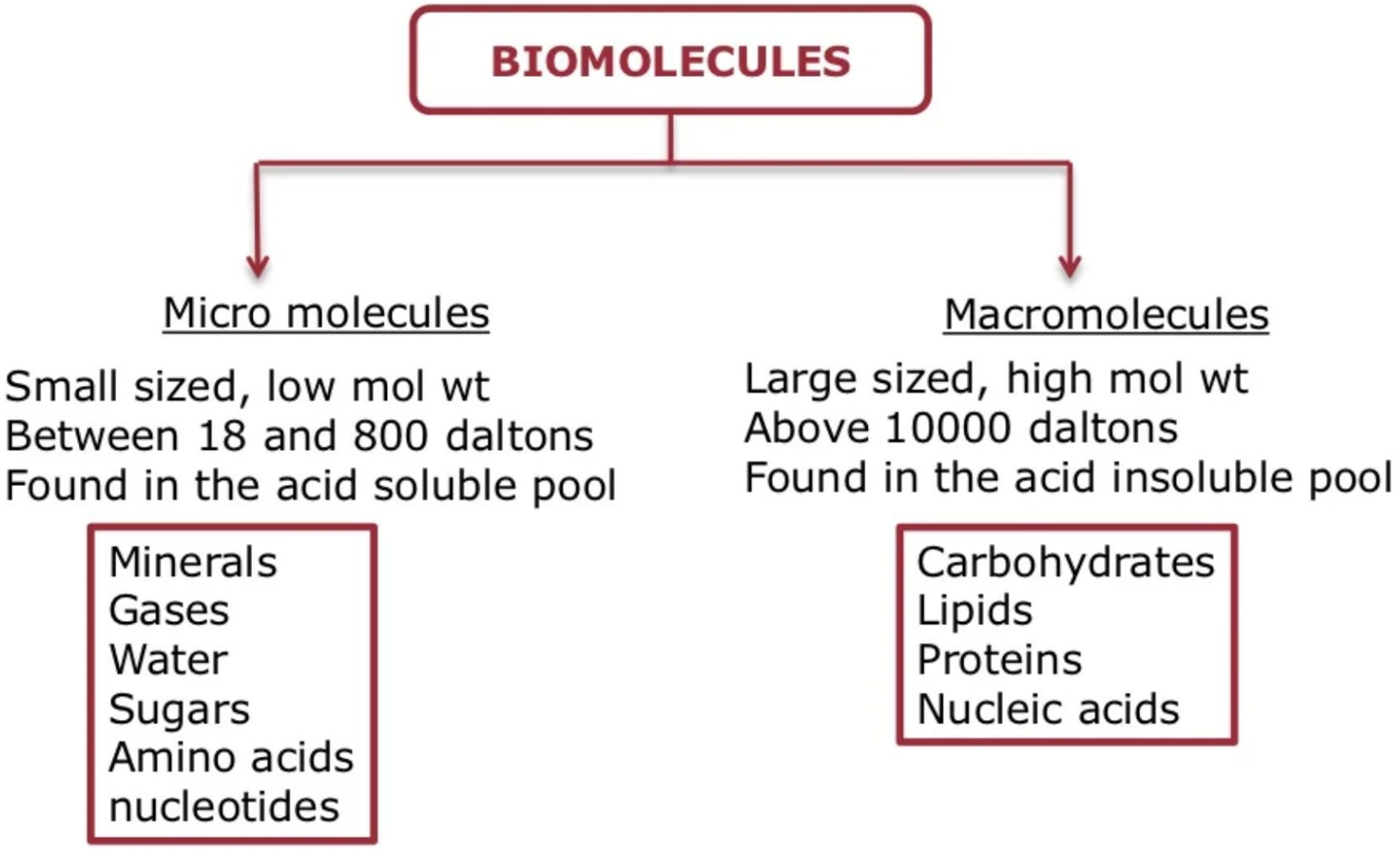
Outline

1. Introduction
2. Structure of Biomolecules
3. Different Levels of protein structure

Biomolecules



Biomolecules

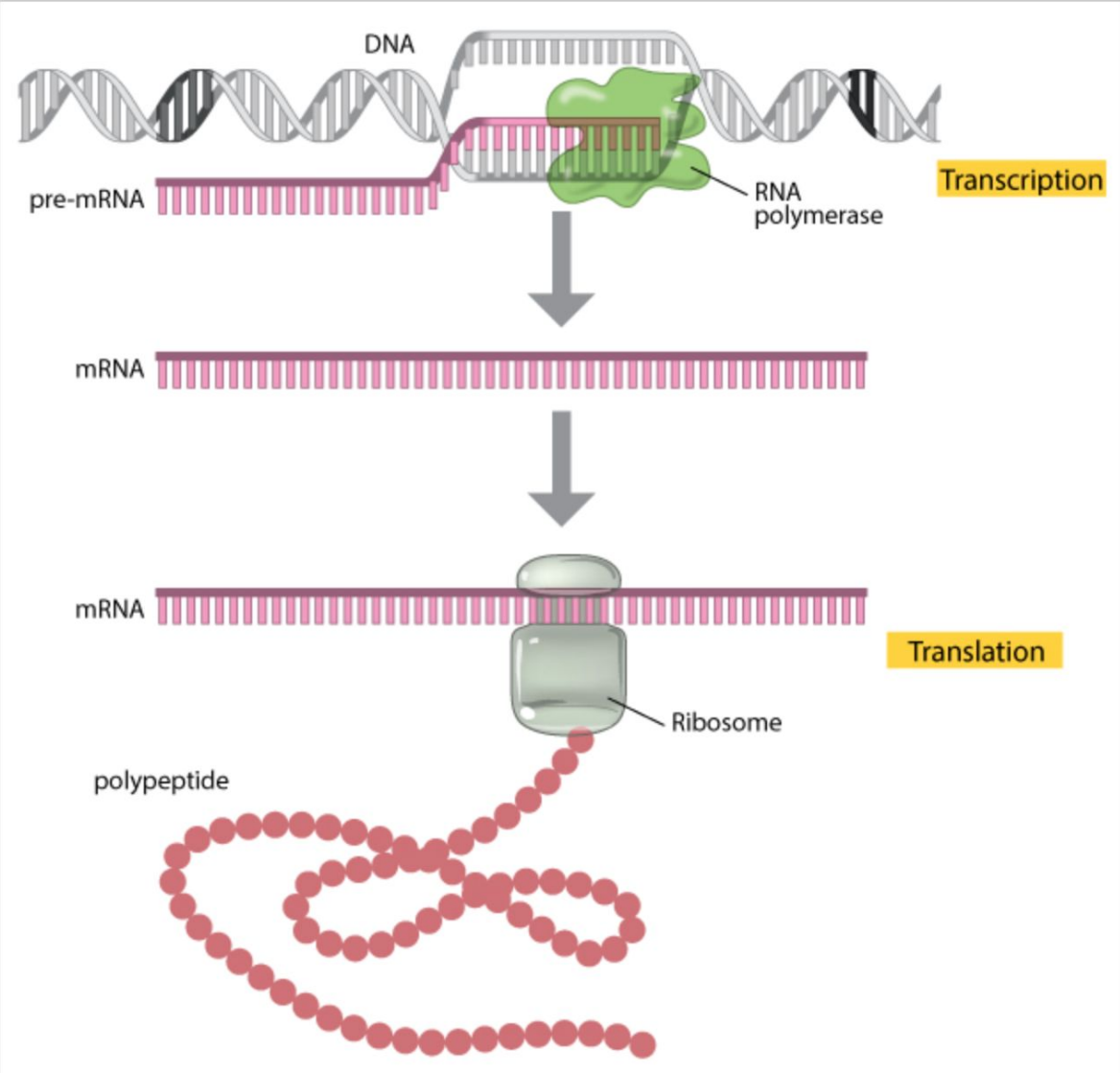


Biomolecules

The major complex biomolecules of cells

Biomolecule	Building block	Major functions
Protein	Amino acid	Basic structure and function of cell
DNA	Deoxyribonucleotide	Hereditary information
RNA	Ribonucleotide	Protein synthesis
Polysaccharide	Monosaccharide	Storage form of energy
Lipids	Fatty acids & glycerol	Storage form of energy to meet long term demands

From DNA to Protein



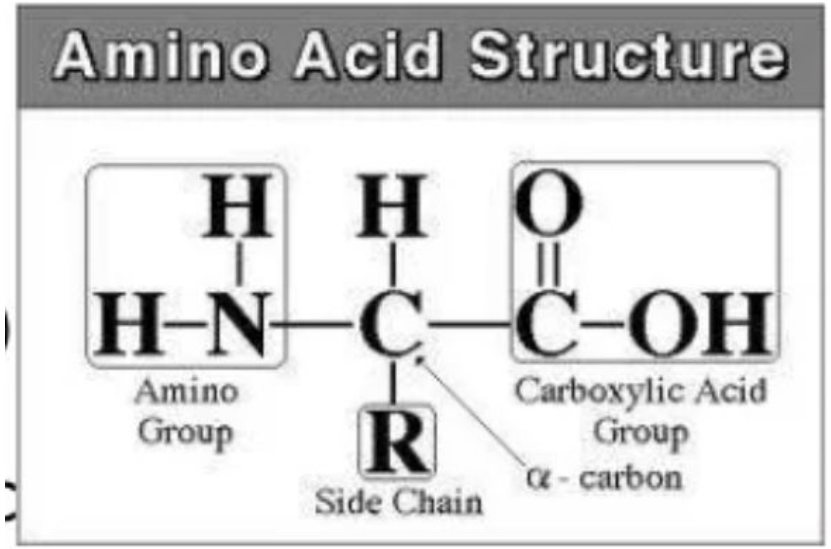
"Central dogma is the process in which the genetic information flows from DNA to RNA, to make a functional product protein."

Proteins

- Most abundant organic molecules of the living system.
- They form about 50% of the dry weight of the cell. They are most important for the architecture and functioning of the cell.
- Proteins are polymers of amino acids and on complete hydrolysis yields Amino Acids
- There are 20 standard amino acids which are repeatedly found in the structure of proteins – animal, plant or microbial.
- Collagen is the most abundant animal protein and Rubisco is the most abundant plant protein

α-Amino Acids and Proteins

- Proteins consist of α-amino acids linked by *peptide* bonds
- Each amino acid consists of:
  - a central carbon atom
  - an amino group
  - a carboxyl group and
  - a side chain
- Differences in side chains distinguish the various amino acids (20)



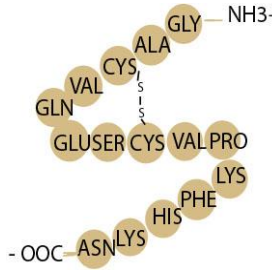
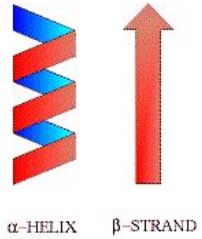
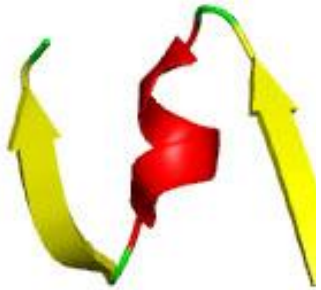
β-amino acid not suitable – has another free rotation

Amino Acids and side chains

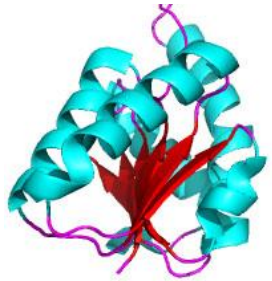
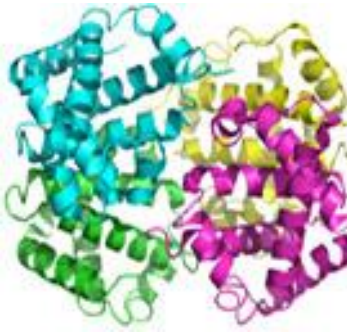
- The common amino acids are grouped according to whether their side chains are:
  - **acidic** D, E
  - **basic** K, R, H
  - **uncharged polar** N, Q, S, T, Y
  - **nonpolar** G, A, V, L, I, P, F, M, W, C

- Hydrophilic amino acids (uncharged polar) are usually on the outside of a protein whereas nonpolar residues cluster on the inside of protein
- Basic or acidic amino acids are very polar and are generally found on the outside of protein molecules

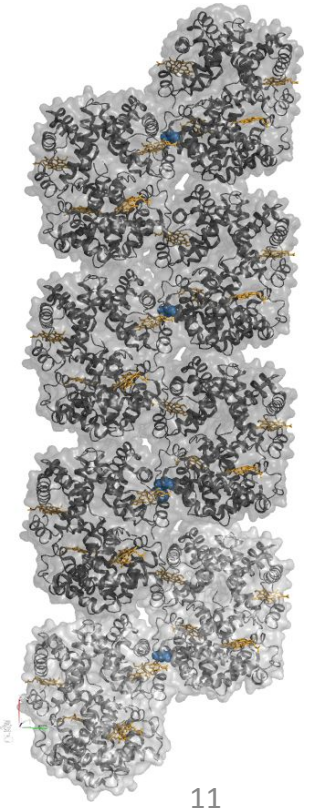
Structure of protein-macromolecule

Level	Description	Representation
I **Primary Structure**	Amino acid sequence of the polypeptide chain(s), without regard to spatial arrangement.	
II **Secondary Structure**	Local spatial arrangement of its main-chain atoms without regard to the conformation of its side chains.	
Motifs / Superseconda ry structures	Elements of secondary structure called motifs, or super-secondary structure. E.g β α β motif, which creates a fold.	

Structure of protein-macromolecule

Level	Description	Representation
III **Tertiary Structure**	Arrangement of all atoms in space, without regard to relationship with neighboring molecules / subunits.	
Domain	Section of a protein that folds into a structurally independent functional unit called a domain	
IV **Quaternary Structure**	Arrangement of subunits in space and the ensemble of inter subunit contacts and interactions.	

Quinary structures – (Hb)



Why is protein structure important?

Aspects

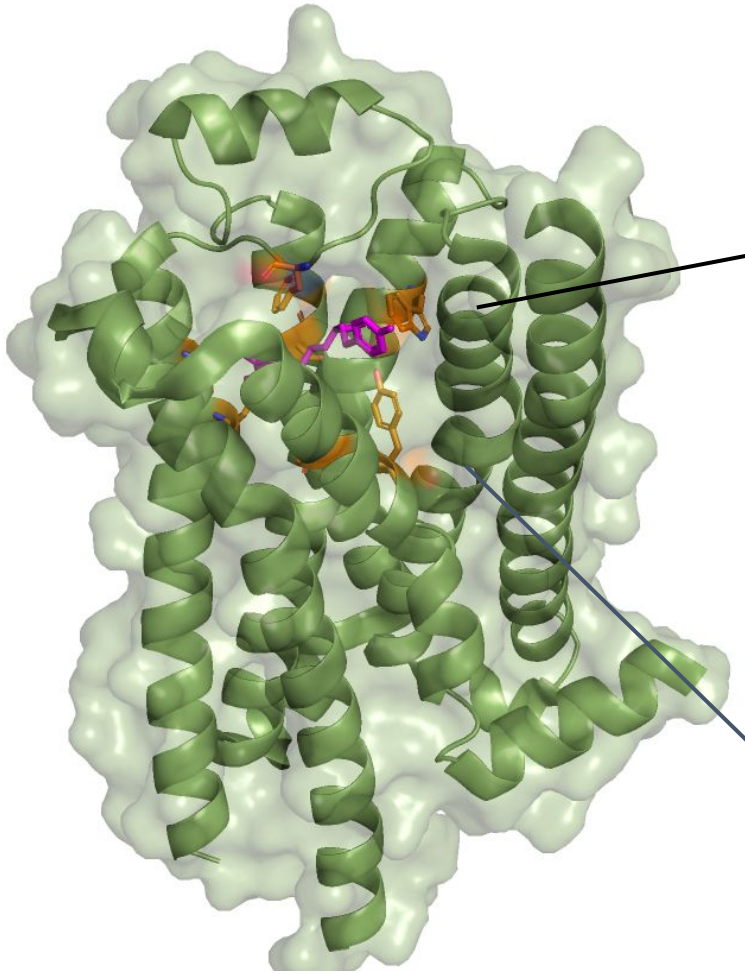
Quaternary Structure
- Biological form

Evolutionary information

Location of SNPs, mutations, conserved residues

Residue solvent accessibility

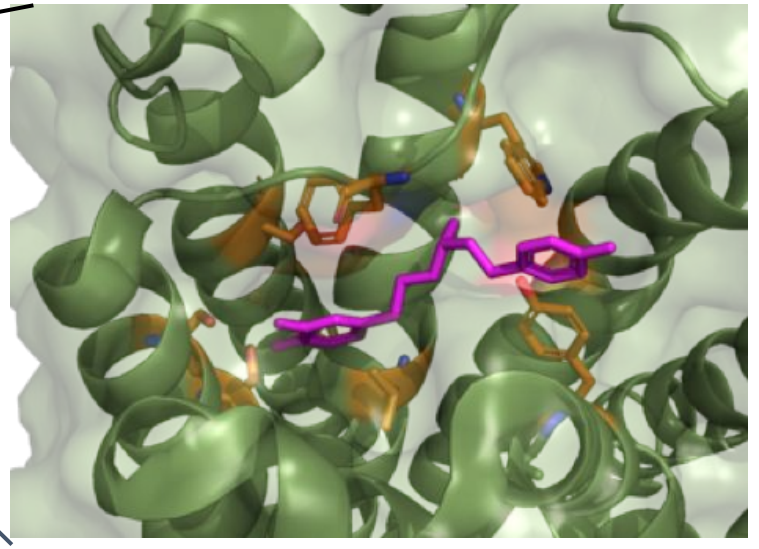
Functional sites, ligand binding



ADRB1 - Dobutamine

Areas

Structure-Function Relationship (Basic Science)



Designing of therapeutics (Pharma Industry)

The RCSB PDB PROTEIN DATA BANK is a tremendous source of information on proteins

Currently >180K structures

Protein Data Bank