

Synopsis: Molecules of interest to biochemists may be classified as **small molecules** and **macromolecules**. Small molecules are similar to those encountered in conventional organic chemistry, and are important in metabolism, which we deal with in the second half of the course. Macromolecules are huge by comparison - molar masses from 10^4 to over 10^9 g.mol⁻¹. What makes it possible to comprehend structures of this magnitude is their **modular construction** from much simpler smaller molecular units. The basis of macromolecule assembly is the **reversible formation** of certain kinds of bonds, e.g. **ester or amide bonds** to link up smaller subunits into long chains. Proteins are chains of linked **amino acids**. Each amino acid has a unique side chain. Since the α -amino/ α -carboxylate core is constant, the **side chain R** determines the specific properties of a particular amino acid and the role it plays in a protein.

REVIEW: CHEM*1040 notes regarding electronegativity and Lewis structures.

Classes of molecules found in biochemistry:

Small molecules

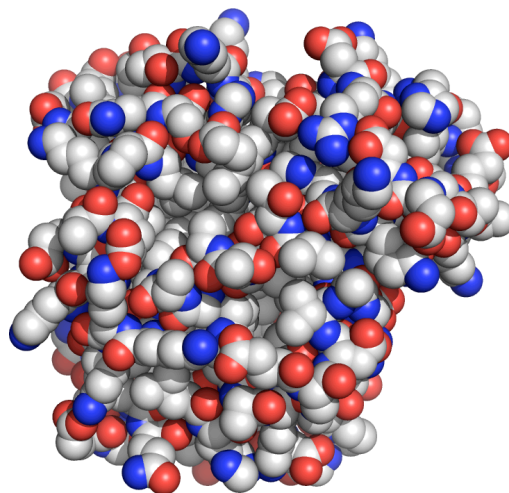
- **Sugars, amino acids, nucleotides, fatty acids, simple carboxylic acid derivatives**
- Interconversions of small molecules may be used to store or release energy, which is the basis of **metabolism**
- Particular kinds of small molecules may serve as building blocks for **macromolecules**

Macromolecules

- **Proteins**, made as chains of amino acids
- **Nucleic acids**, made as chains of nucleotides
- **Polysaccharides**, made as chains of simple sugars

Proteins form complex structures capable of many functions, including structural components of cells, catalysis of reactions and communication processes. For this reason, the first half of the semester will focus on proteins and their role.

For example, **myoglobin** (right) is a protein that stores O₂ in muscle tissue.



Typical protein molecules have molecular masses between 10 000 and 100 000 g.mol⁻¹, so they contain literally thousands of atoms. Myoglobin has a molecular mass of 16 500 g.mol⁻¹.

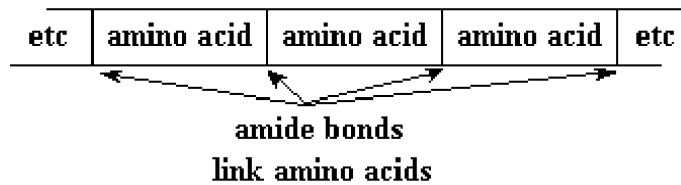
Because proteins and other macromolecules are so large, biochemists use a unit called the **kiloDalton (kDa)**. One Dalton is simply 1 g.mol⁻¹, so 11 000 g.mol⁻¹ becomes 11 kDa. Typical proteins are therefore between 10 and 100 kDa, while myoglobin is 16.5 kDa. The largest known single protein molecule is titin at 10 000 kDa.

The building block principle of macromolecule structure

Proteins are chains of linked **amino acids**:

Each protein has a **unique sequence of different amino acids**, and a well-defined **size and structure**. The arrangement of amino acids in the chain determines the properties and function of the protein. A

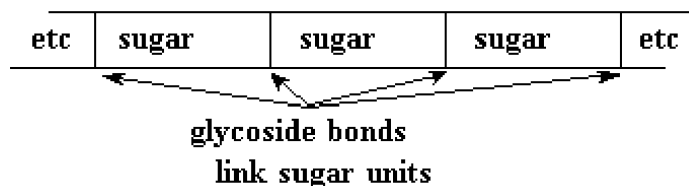
protein of 100 amino acids has a mass of about 11 000 g.mol⁻¹, about 110 g.mol⁻¹ per amino acid. Proteins are between 10 and 10 000 kDa (10⁴ to 10⁷ g.mol⁻¹).



Two other kinds of macromolecule will be dealt with later in the semester:

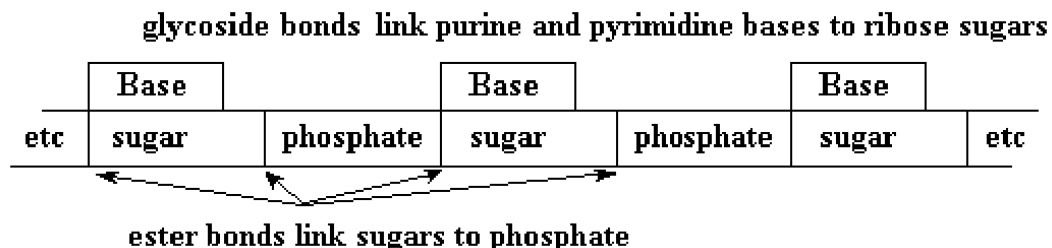
1. A **polysaccharide** is a chain of sugars:

Most polysaccharides, e.g. starch, are **simple repetitive structures** of one or two sugars, with no definite size. Some polysaccharides are used for storage of sugars; others act in simple structural roles.



2. **Nucleic acids** DNA and RNA are a bit more complex:

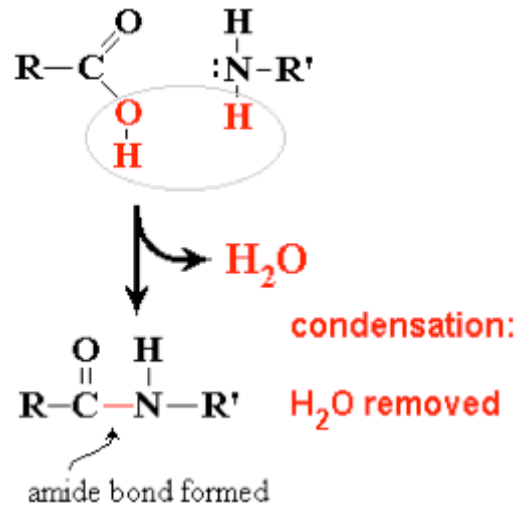
The backbone is simple and repetitive; but different bases are attached giving nucleic acids unique and characteristic sequences. The repeating unit, base + sugar + phosphate is called a **nucleotide**.



Bonding between subunits

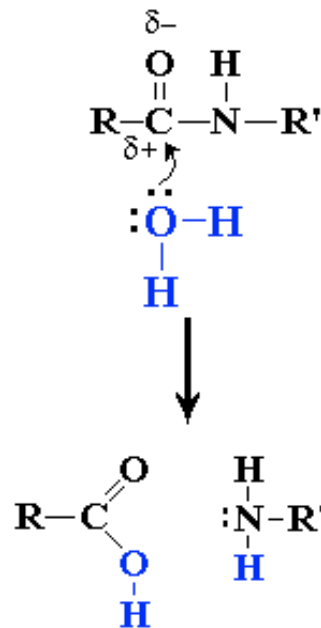
The types of bond that link subunits in macromolecules are formed by a process called **condensation**, since the process involves **elimination of the elements of H₂O**.

e.g **amino acids** contain both carboxylic acid and amino groups, and these allow the formation of an **amide bond** by condensation:



The converse of condensation is the **attack of H₂O on the amide bond**, which restores the original carboxylic acid and amino group and thus **unlinks** the two units. This is called **hydrolysis**.

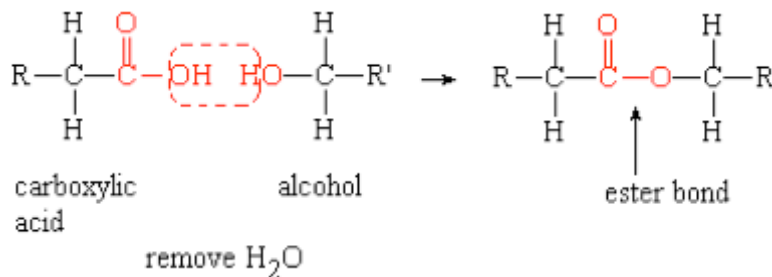
The **carbonyl group C=O** of the amide is the **point of weakness** that allows H₂O to attack.



Bonds formed by condensation and broken by hydrolysis:

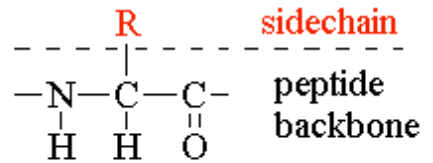
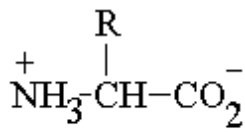
carboxylic acid + amino group ⇌ amide

carboxylic acid + alcohol ⇌ ester



Proteins are chains of amino acids

Twenty different amino acids are found in proteins. They have in common the following structure, distinguished by different side chain groups, shown as R.



Each amino acid has an **amino group** and a **carboxylate group**, and can thus form either end of an amide bond.

as single amino acid

as part of a polypeptide chain

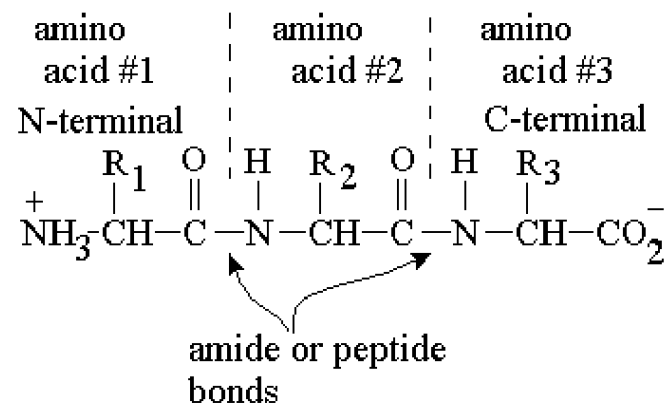
(What's different about the amino acid **proline**? - look it up and find out!)

After removal of the H₂O in condensation, the portion of the amino acid found in the chain is called an amino acid **residue**.

Large numbers of amino acids can be strung together in a chain:

Each amino acid in the chain can have a different side chain **R**; **these give the protein molecule as a whole its unique properties.**

On average, each amino acid contributes about **110 Da** to the molar mass of a protein. 153 amino acids make up the 16.5 kDa myoglobin molecule.



Some terminology to learn:

One end of the chain has an amino acid with an uncombined amino group, usually protonated; this is called the **N-terminal amino acid**. The other end has an amino acid with an uncombined carboxylate group, and is called the **C-terminal amino acid**.

Amino acid chains linked in this way are called peptide **chains**; the **amide bonds linking them are called peptide bonds**.

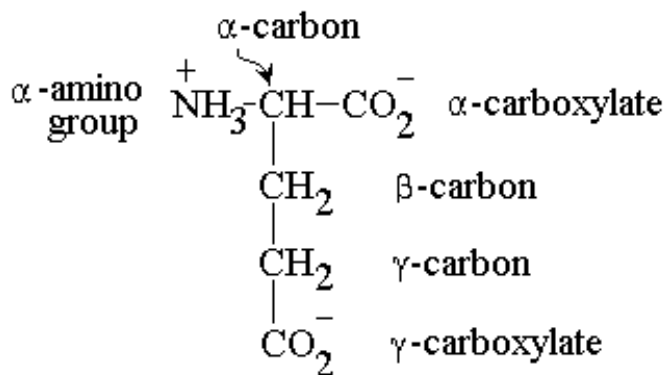
A **polypeptide** is a large peptide chain, usually the complete amino acid chain of a protein. Typical proteins are several hundred amino acids long.

Poly = Greek for "many"

An **oligopeptide** is a smaller peptide chain, often applied to a fragment of a larger protein
Oligo = Greek for "a few"

Structural layout of amino acids

Amino acids in a peptide chain have identical backbone but **unique side chains R**. Specific properties of a protein are determined by the particular functional groups present in the side chains. The amino acid **glutamate (Glu for short)** is an example.



In biochemical nomenclature, Greek letters identify the carbon atoms of the structural core of an amino acid, as in the example at left:

Functional groups are labeled according to the core atom to which they are attached, e.g. **α -amino**, **α -carboxylate**, **γ -carboxylate**. The α -amino and α -carboxylate become linked up in the peptide bonds making up the backbone.

The side chain properties that have most influence on the behaviour of the protein include:

- polarity
- hydrogen bonding ability
- charge

There are twenty different amino acids commonly found in protein chains, described by a full name, a three-letter abbreviation, or a single letter symbol. Full structures of amino acids and other properties can be found in Lehninger p.73-75 and on the amino acids Euchre deck (available online).

You should be prepared to **reproduce** the structures of the amino acid constituents of proteins and the complete covalent structures of proteins, and to know the single- and three-letter abbreviations for the amino acids.

Classification of amino acids by properties

The "inverted pyramid" below is a convenient memory aid which groups amino acids according to **common properties and structures**. There is also a Euchre deck of the amino acids and nucleotides in DNA. If you play euchre, try these cards out!

6 amino acids with very non-polar side chains:

<p>Alanine Ala, A</p>	<p>Valine Val, V</p>	<p>Leucine Leu, L</p>	<p>Isoleucine Ile, I</p>	<p>Phenylalanine Phe, F</p>	<p>Methionine Met, M</p>
----------------------------------	---------------------------------	----------------------------------	-------------------------------------	--	-------------------------------------

5 amino acids with medium to moderately non-polar side chains:

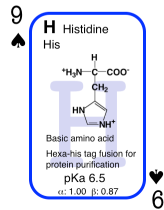
<p>Glycine Gly, G</p>	<p>Cysteine Cys, C</p>	<p>Proline Pro, P</p>	<p>Tryptophan Trp, W</p>	<p>Tyrosine Tyr, Y</p>
----------------------------------	-----------------------------------	----------------------------------	-------------------------------------	-----------------------------------

4 amino acids with polar uncharged side chains:

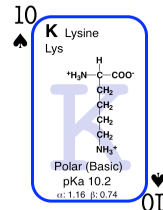
<p>Serine Ser, S</p>	<p>Threonine Thr, T</p>	<p>Asparagine Asn, N</p>	<p>Glutamine Gln, Q</p>
---------------------------------	------------------------------------	-------------------------------------	------------------------------------

3 amino acids with positively charged side chains:

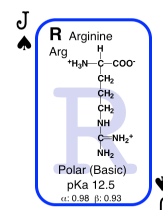
Histidine
His, H



Lysine
Lys, K

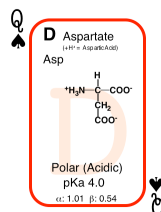


Arginine
Arg, R

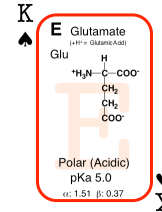


2 amino acids with negatively charged side chains:

Aspartate
Asp, D



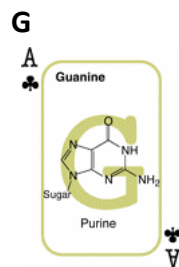
Glutamate
Glu, E



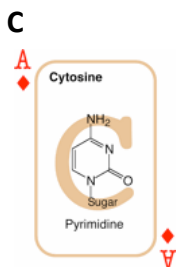
And, just for the record, here are the 4 aces in the euchre deck:

4 nucleotides in DNA

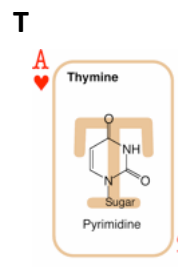
Guanine



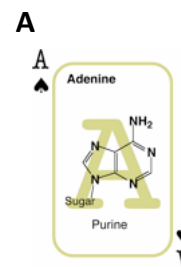
Cytosine



Thymine



Adenine



(You'll get into the nucleotides in the 2nd half of the course)