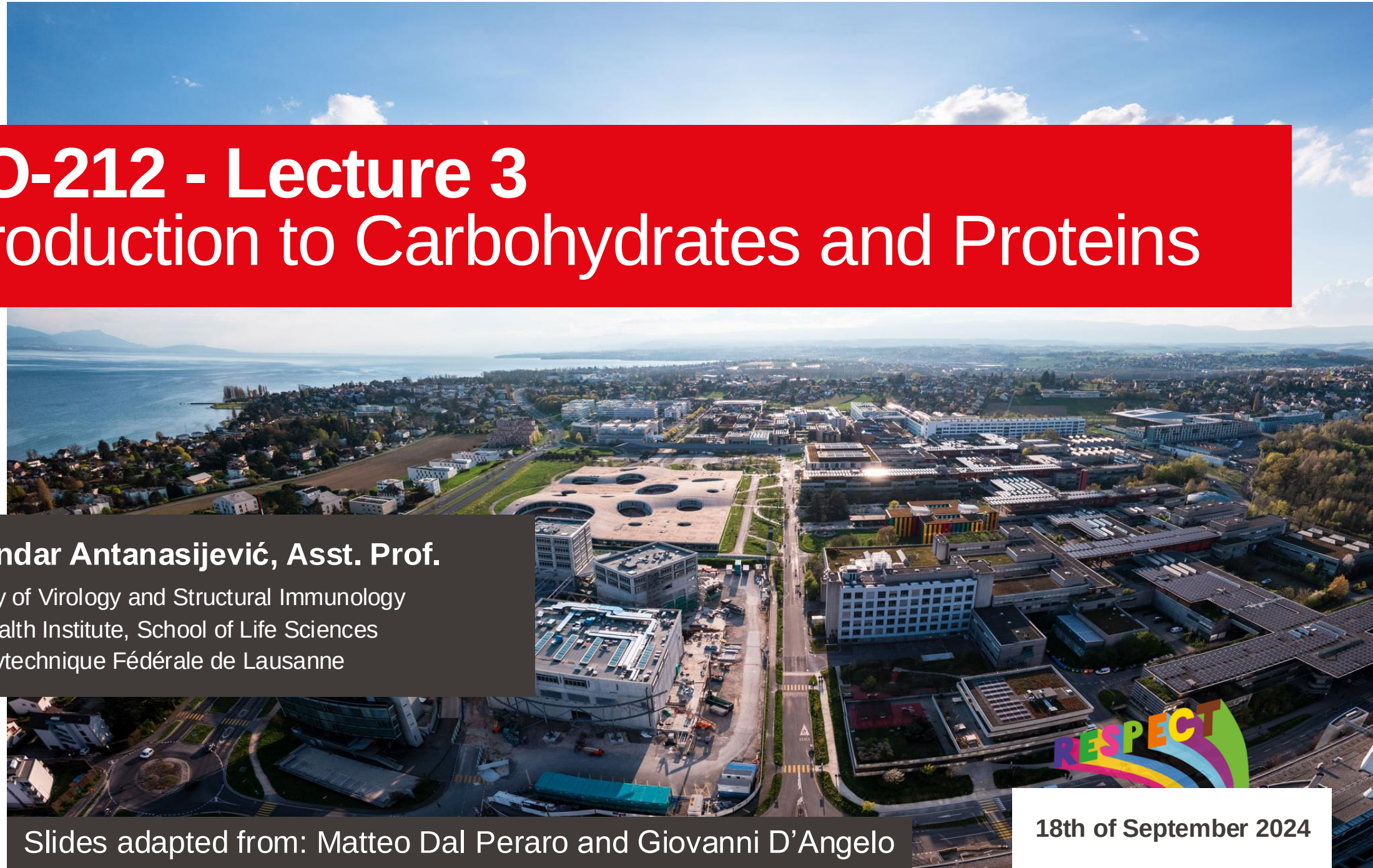
BIO-212 - Lecture 3
Introduction to Carbohydrates and Proteins



Aleksandar Antanasijević, Asst. Prof.

Laboratory of Virology and Structural Immunology
Global Health Institute, School of Life Sciences
École Polytechnique Fédérale de Lausanne

École polytechnique fédérale de Lausanne

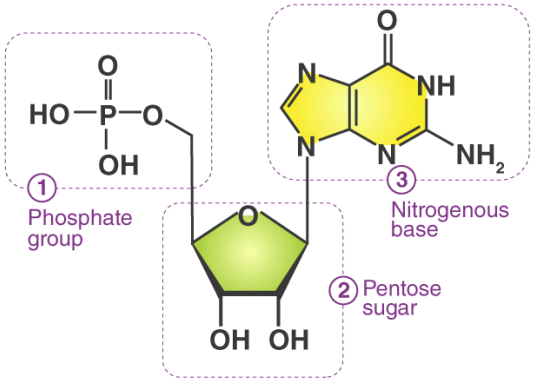
Slides adapted from: Matteo Dal Peraro and Giovanni D'Angelo

18th of September 2024

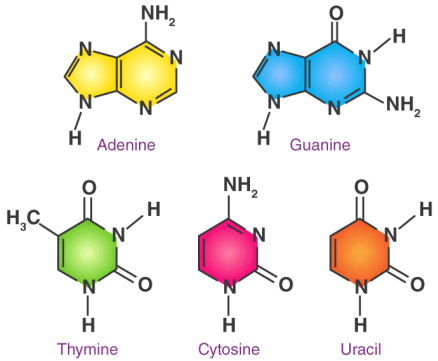
Lecture 2 – Quick Summary

• Nucleic Acid building blocks

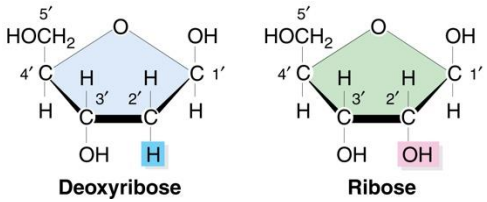
- 3 main components: **base, pentose, and phosphate**



- 5 base options



- 2 pentose options



- 2 x 4 sets of nuceotides to produce DNA and RNA

• Nucleic Acids (DNA and RNA)

- Linear **polymers of nucleotides**

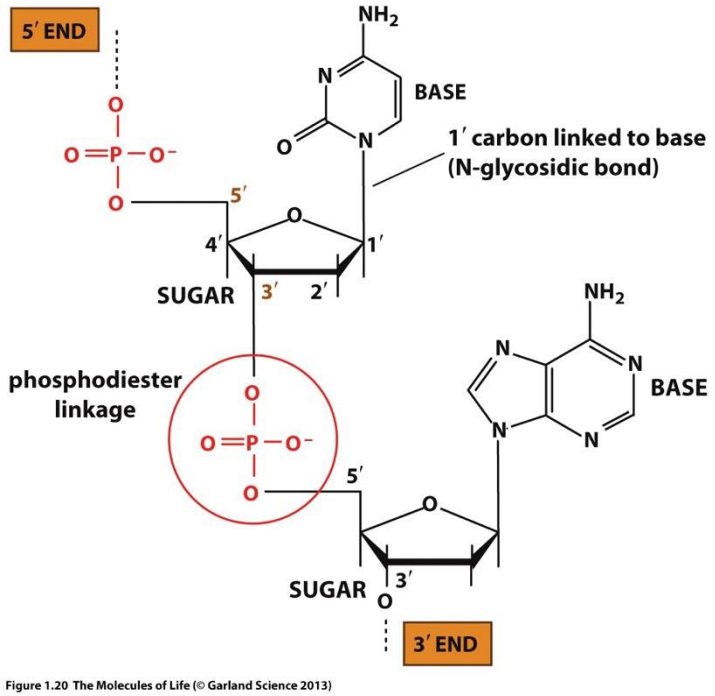


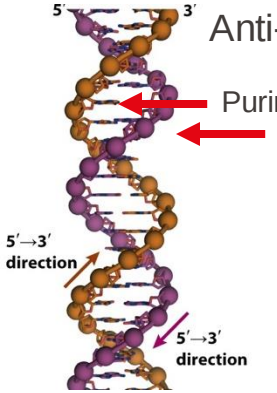
Figure 1.20 The Molecules of Life (© Garland Science 2013)

- 5'->3' directionality in addition of nucleotides
- Attachment via phosphodiester linkages

Lecture 2 – Quick Summary

• Nucleic Acid structure and assembly

- Double-stranded helical assembly of DNA (B form)

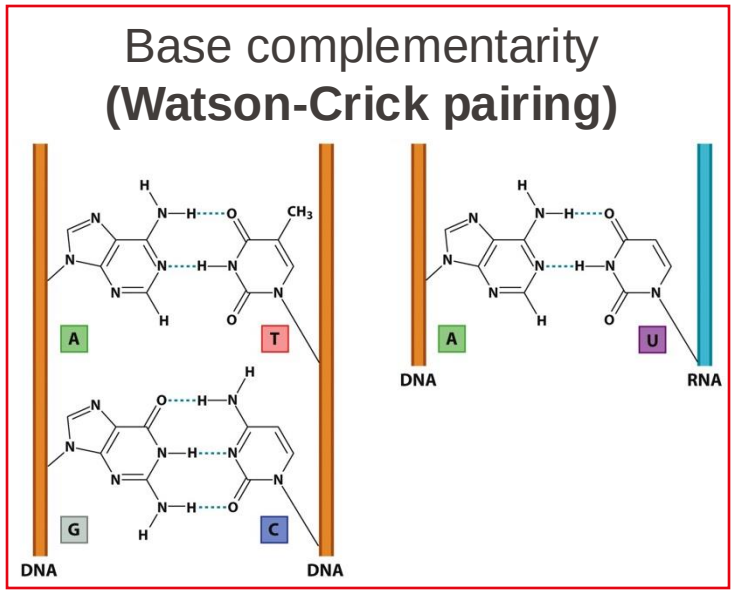


Anti-parallel complementary strands

Purine-Pyrimidine Base Pairs (inwards)

Phosphate-pentose backbone (outwards)

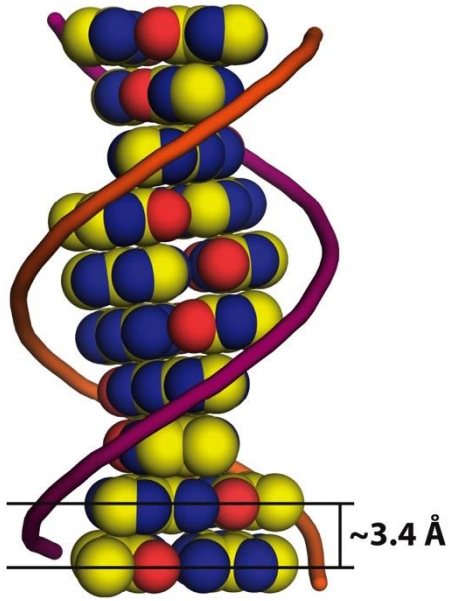
10 base pairs per helical turn



- RNA can be double- and single-stranded and features greater conformational and functional diversity

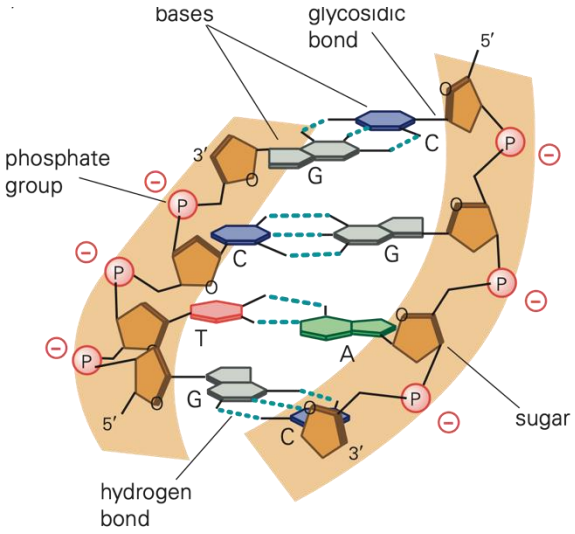
• Important non-covalent interactions

Van der Waals interactions (layered base stacking)



- Stacking of arromatic π-orbitals (π-π stacking)

Hydrogen-bond network (base pairing)

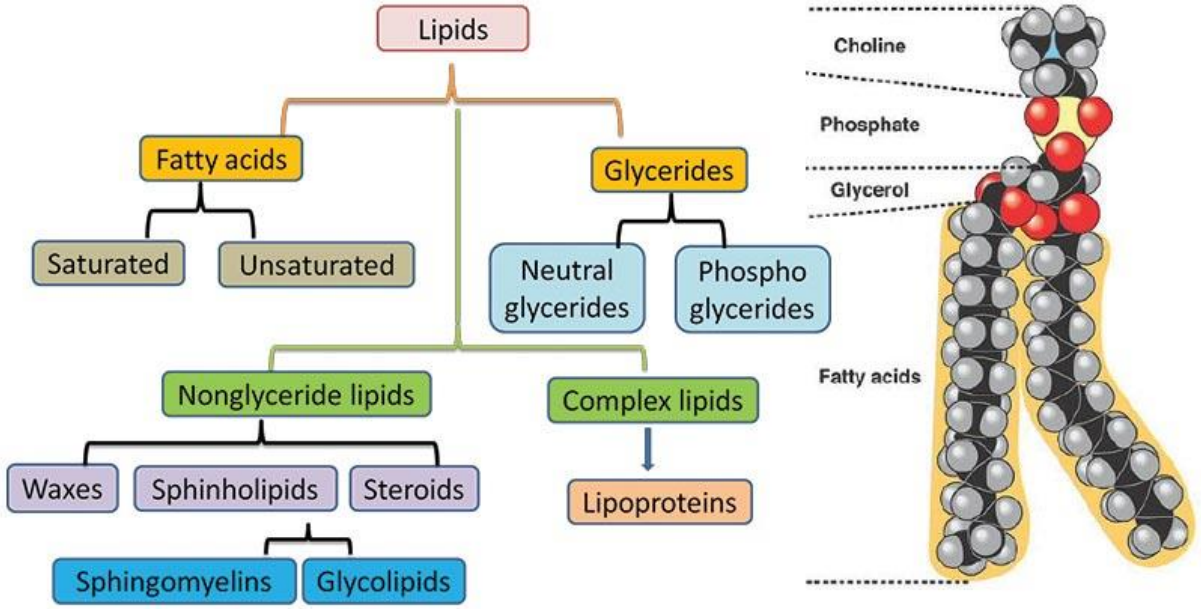


- Charged, polar backbone attracts water and positively charged ions

Lecture 2 – Quick Summary

• Main lipid types and roles

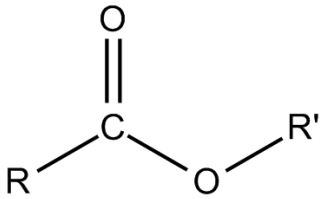
- Main lipid types based on chemical properties



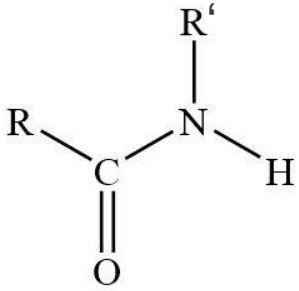
- There are >1000 building blocks which can assemble in different ways (very diverse)
- Their main roles include energy storage, assembly of biological membranes, cell and hormone signaling

• Important bonds and interactions

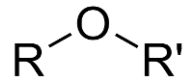
Ester bonds (triglycerides)

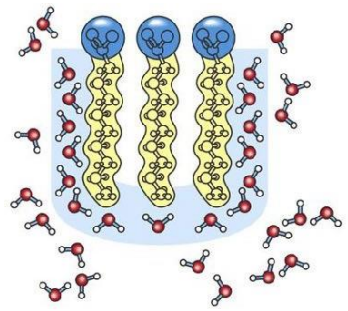


Amide bonds (sphingolipids)

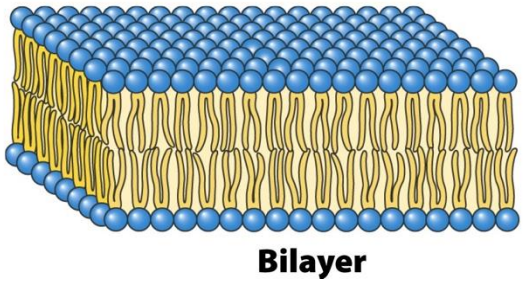


Ether bonds (head domains)





- Amphipatic molecules (polar and hydrophobic) that form bilayers in aqueous solutions

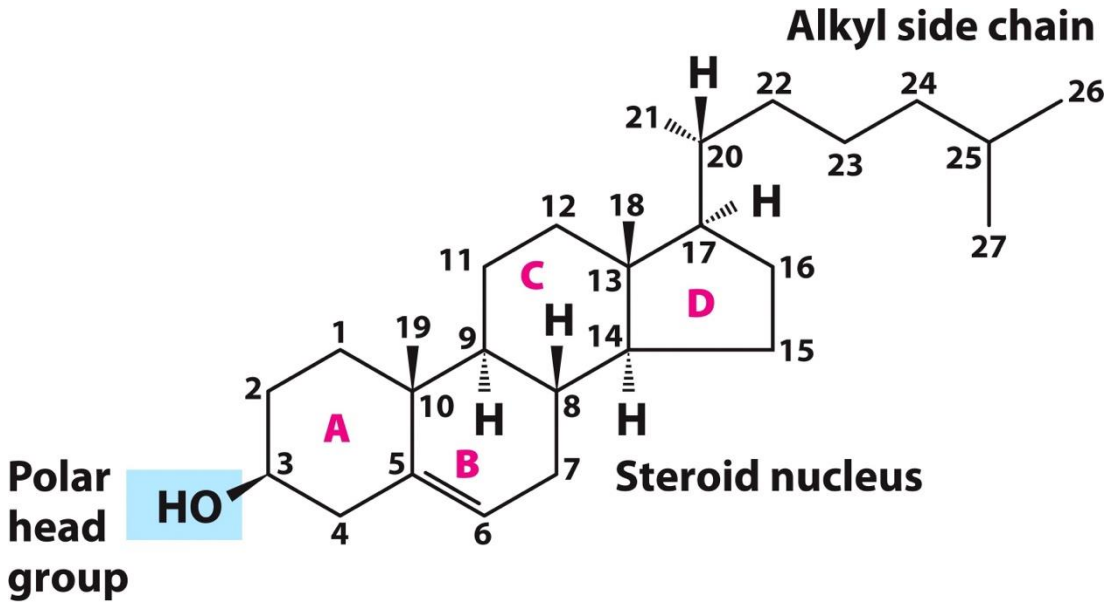


- Van der Waals interactions in the tail
- Hydrogen bonding and charged interactions using the head domain

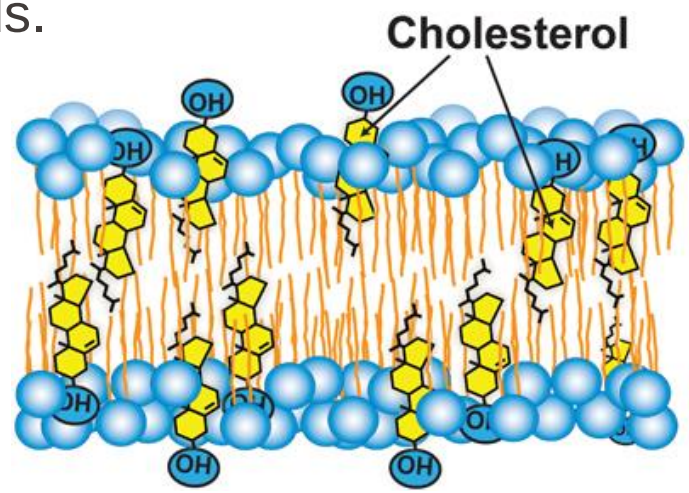
Sterols

• Sterols consist of a rigid steroid nucleus containing four fused rings, an alkyl side chain of 8 carbons, and a single hydrophilic hydroxyl group attached to C-3 of ring A.

• They are synthesized in humans but also acquired through diet, and they are present in the membranes of most eukaryotic cells.



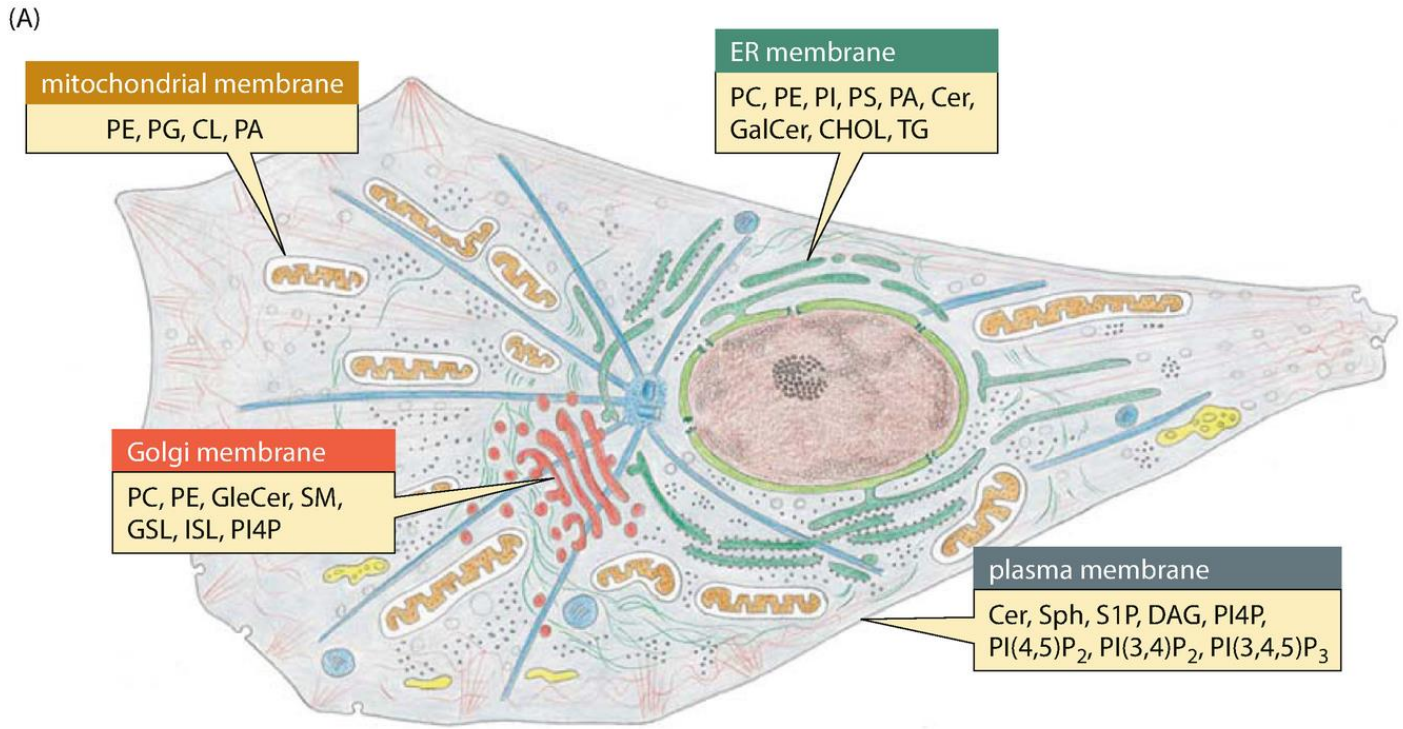
• The steroid nucleus is nearly planar, and the molecule efficiently packs with the acyl chains of membrane glycerophospholipids and sphingolipids.



• Cholesterol impacts membrane rigidity/mobility

Lipid composition in different cell compartments

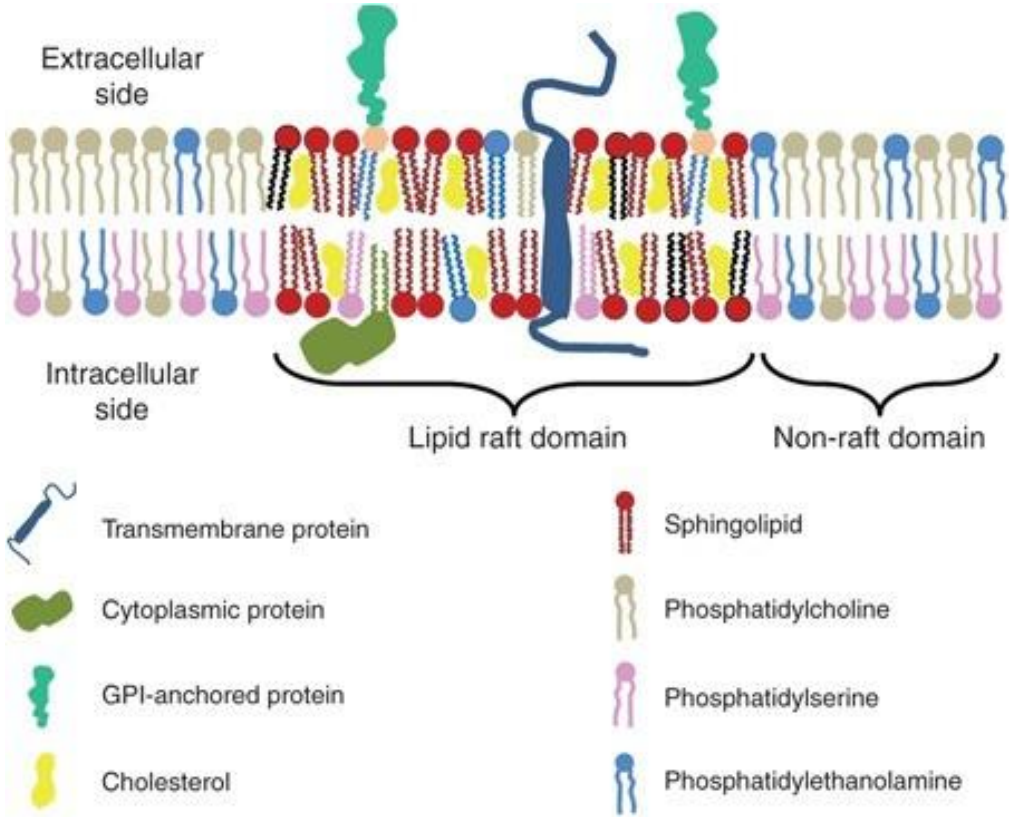
- Phospholipids (PE, PC, PI, PS) are relatively uniform across most membranes
- Sphingolipids (CER, GalCER, Sph) largely present on the plasma membrane but also Golgi and ER
- Cholesterol comprises all membranes but is the highest at the cell membrane where it represents 40-50% of all lipids.



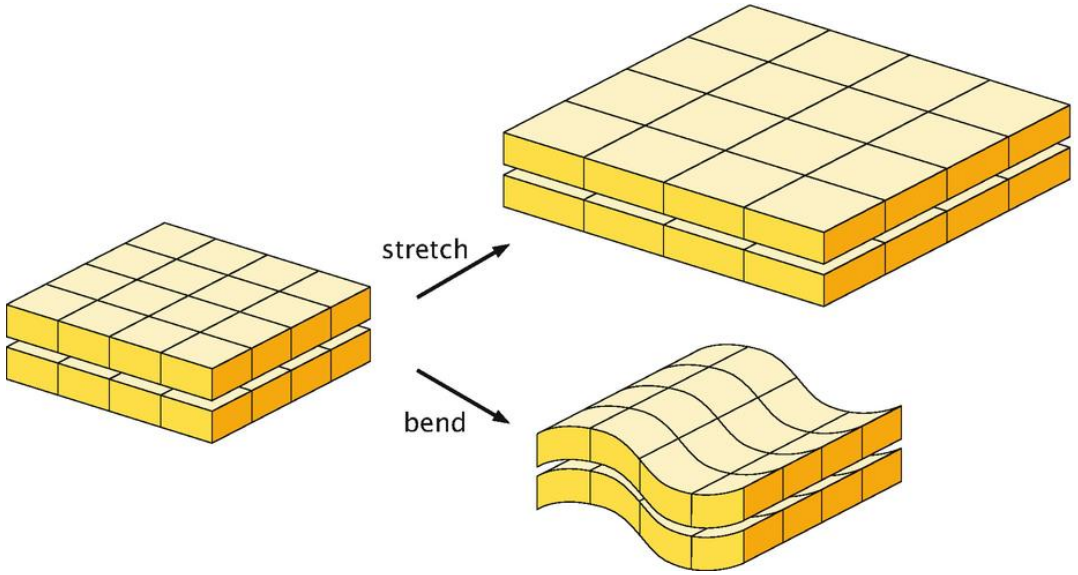
Cholesterol/Phospholipid ratio:

Membranes are asymmetric dynamic structures

- Inner and outer leaflet have different compositions



- Membranes are dynamic and pleomorphic

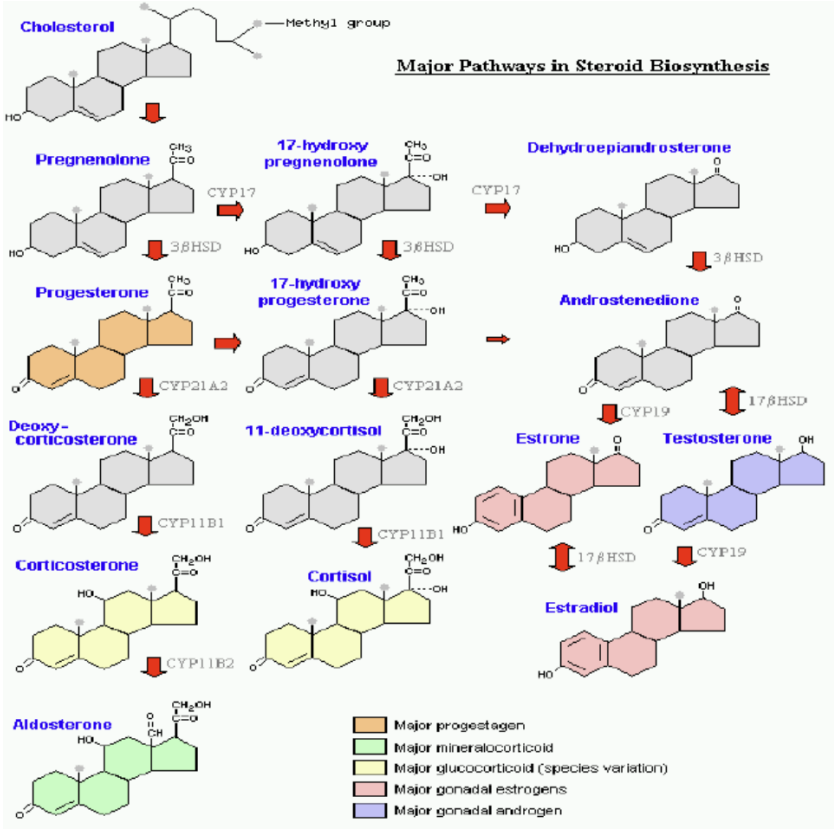


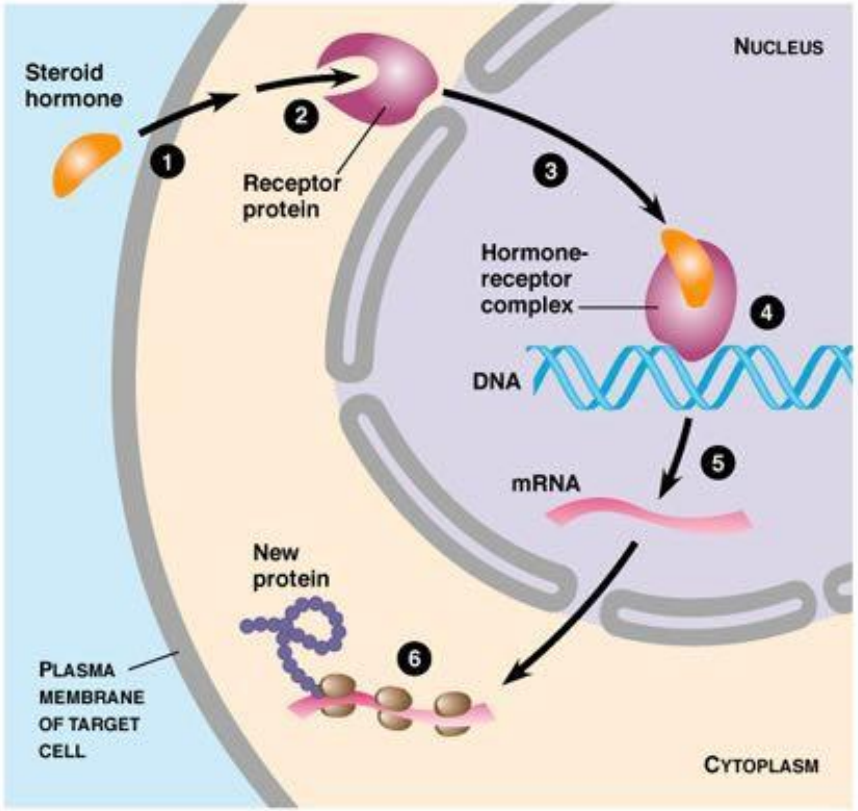
- **Flippases, Floppases** and **Scramblases** are transmembrane proteins that regulate transport of lipids between leaflets

- Membrane flexibility is determined by lipid composition and environmental factors (e.g., mechanical forces)

Bioactive lipids – Steroid hormones

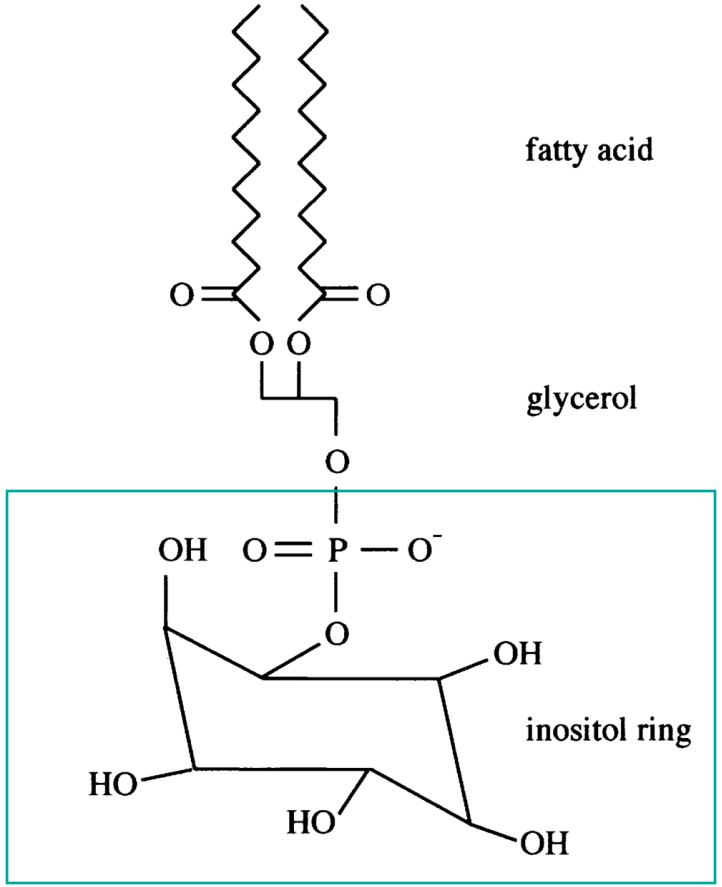
- Bioactive lipids affect cell function in a concentration dependent manner.
- Most commonly they serve as hormones and secondary messengers
- Sterols are precursors of steroid hormones and are therefore also considered bioactive lipids

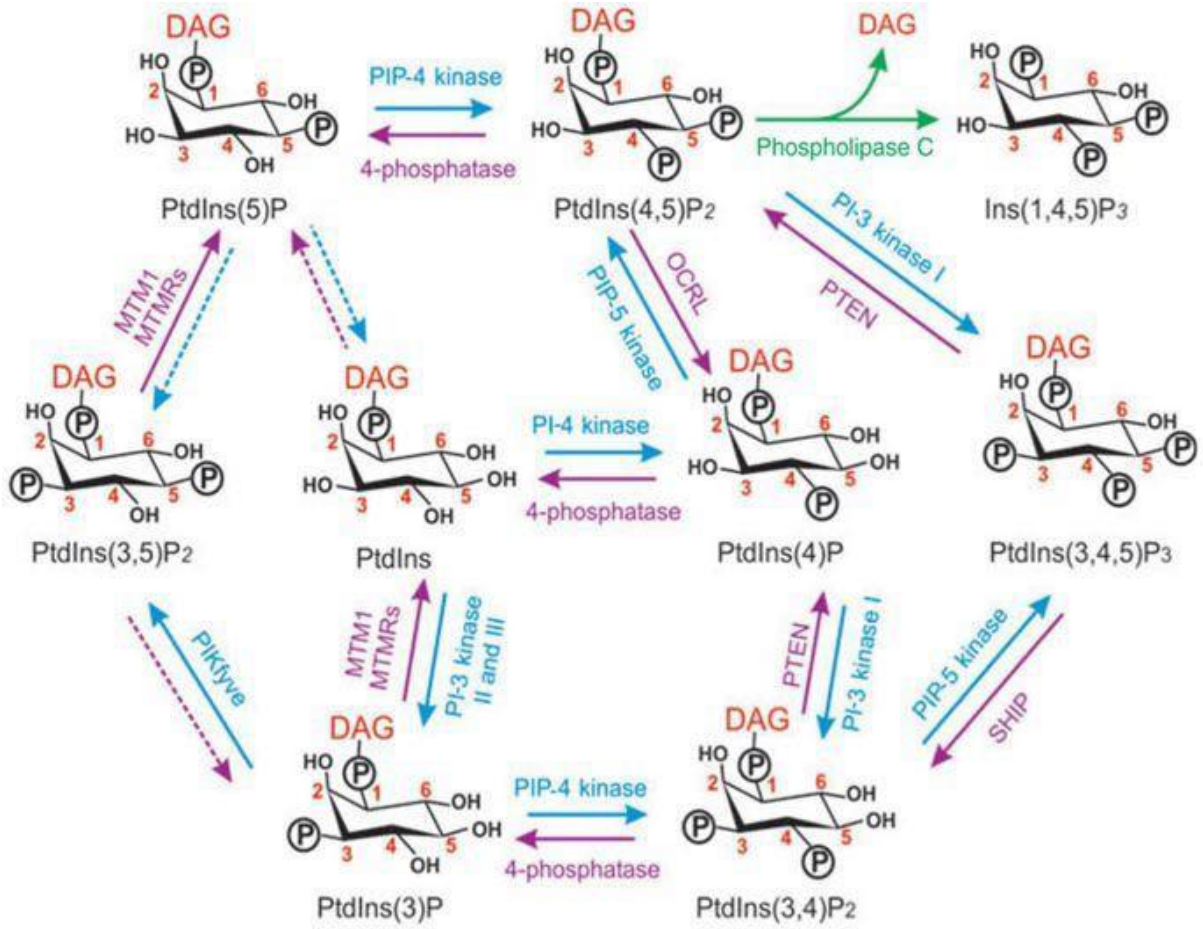




Bioactive lipids – Phosphoinositides (PI)

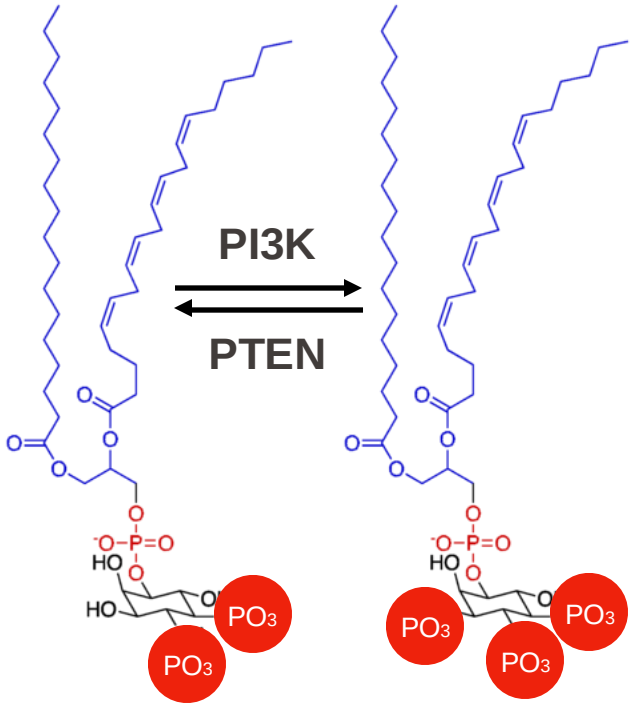
- Phosphoinositides are minority phospholipids on cellular membranes that have a phosphoglycerol backbone esterified with 2 fatty acids and inositol head group.
- Inositol group can be phosphorylated (P) at one or more sites in a reversible fashion (kinase – phosphatase)

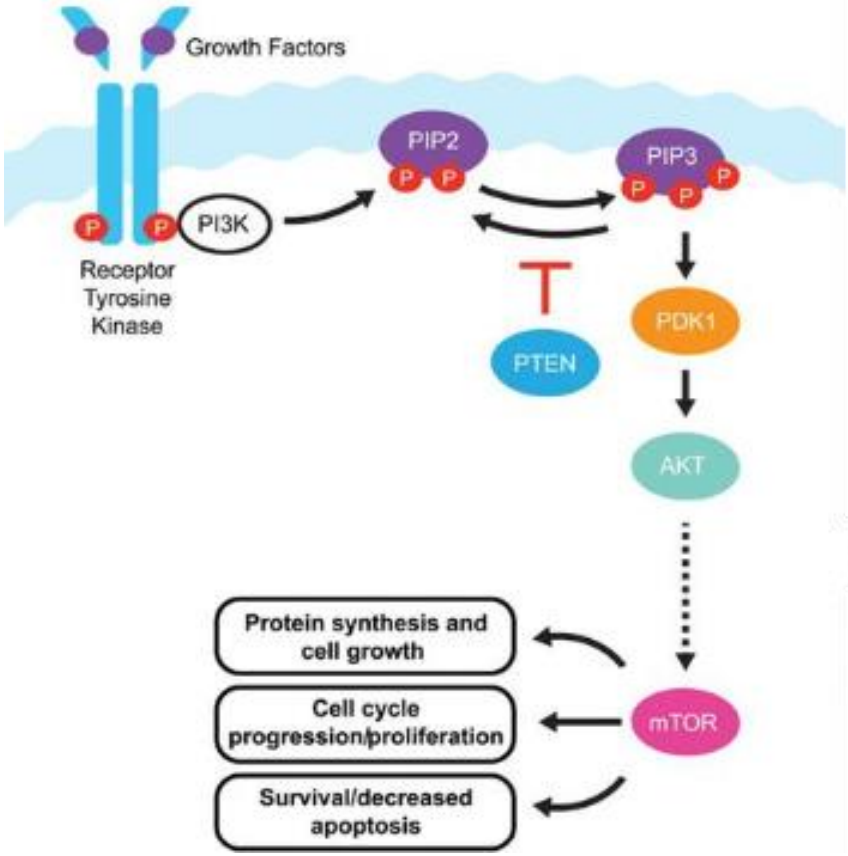




Bioactive lipids – PtdIns(3,4,5)P3

- Phosphorilation of PtdIns(3,4)P2 to PtdIns(3,4,5)P3 catalyzed by receptor typrosine kinase is recognized by PDK1 resulting in activation of the downstream signaling cascade and leading to different outcomes to the cell
- In this case, phosphoinositides serve as **secondary messengers**

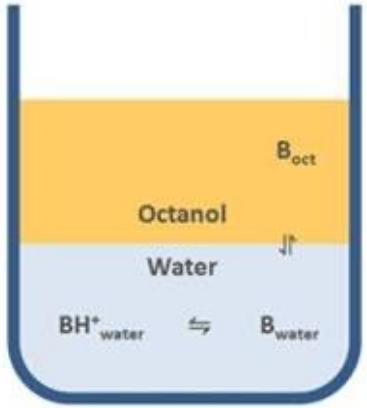




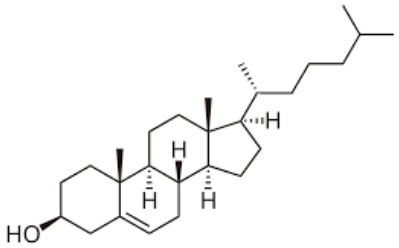
Measuring the relative hydrophobicity of a lipid

• The partition coefficient, P, is a measure of the differential solubility of a lipid in two immiscible solvents. The most commonly used solvent system is **octan-1-ol and water**.
• Higher LogP value -> More hydrophobic

$$\text{Log}P = \log_{10}(\text{Partition Coefficient}) \qquad \text{Partition Coefficient, } P = \frac{[\text{Compound}]_{\text{octanol}}}{[\text{Compound}]_{\text{water}}}$$

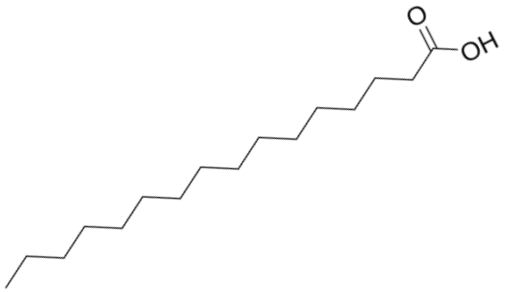


Cholesterol



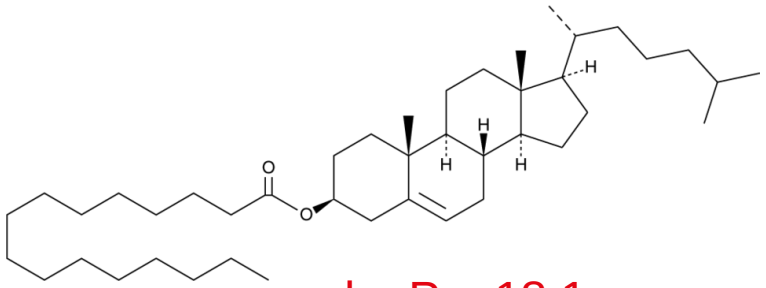
logP = 8.7

Palmitate



logP = 7.1

Cholesteryl Palmitate



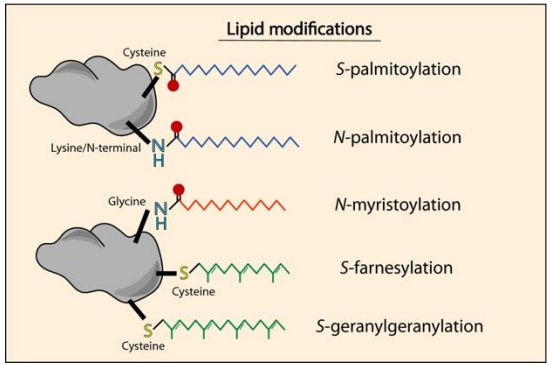
logP = 18.1

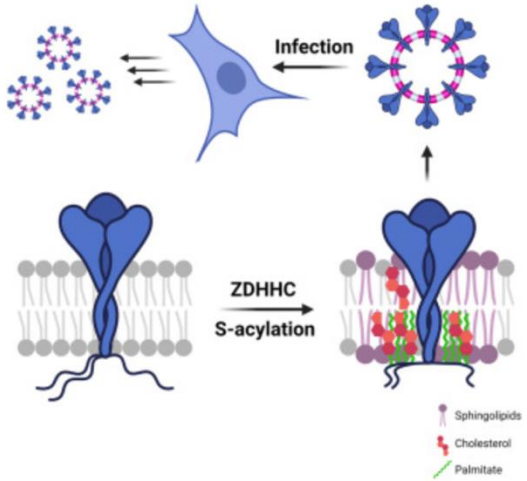
• The differing hydrophobicity of lipids can be used for their separation by chromatography methods (e.g., reverse phase chromatography)
• This parameter is also calculated for drugs and therapeutics as a measure of their bioavailability (more water soluble -> easier to traffic to different tissues)

Lipids in bioengineering

- Conventional applications include food, cosmetics, soap, biofuel, chemical feedstock etc.

Lipid-protein interactions



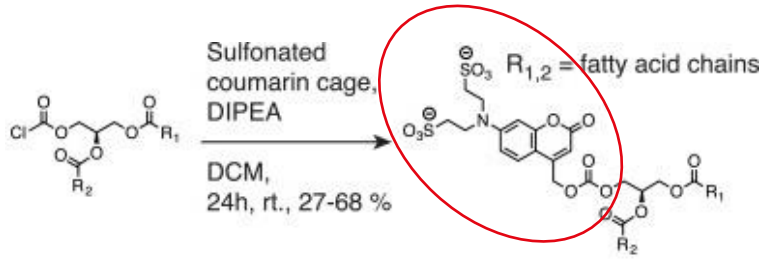


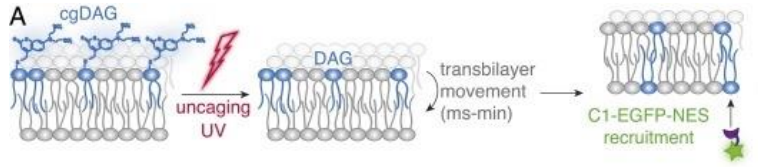
https://www.epfl.ch/labs/vdg/

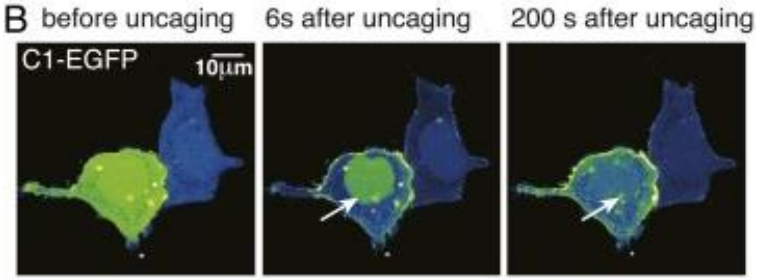
Membrane development

https://www.epfl.ch/labs/dangelo-lab/

Chemical toolkit for biology







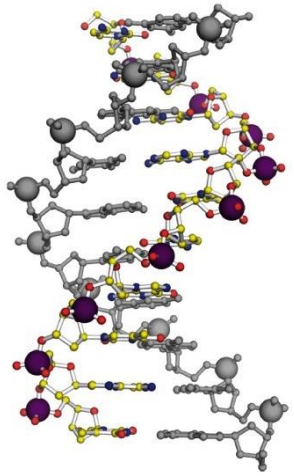
https://www.epfl.ch/labs/gr-schuhmacher/

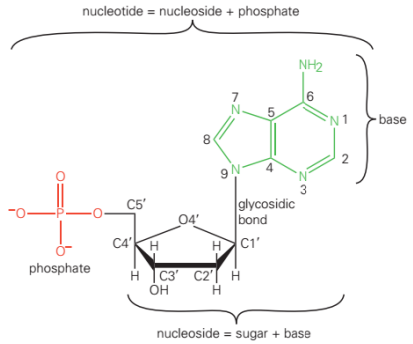
The molecules of Life

Macromolecular Structure

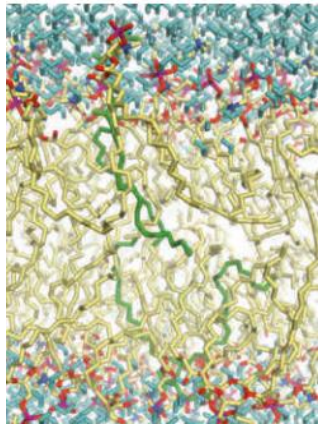
Building Block

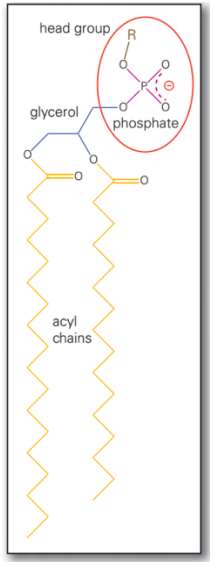
Nucleic Acids



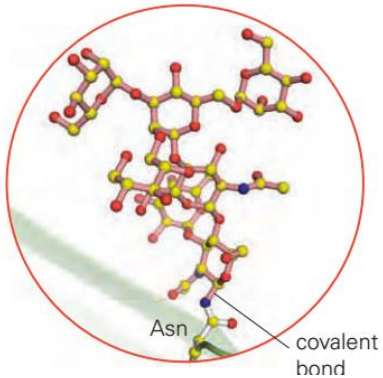


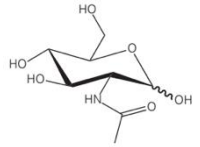
Lipids



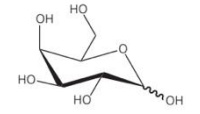


Carbohydrates

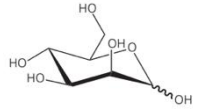




N–acetylglucosamine (GlcNAc)

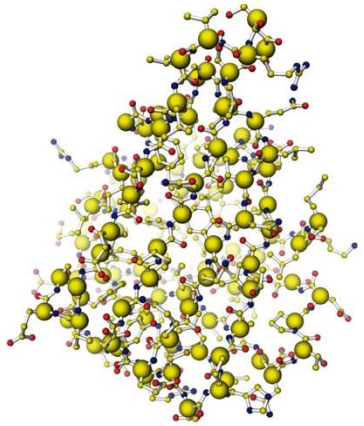


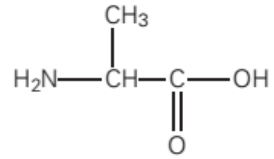
galactose (Gal)

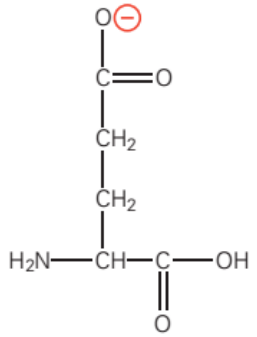


mannose (Man)

Proteins







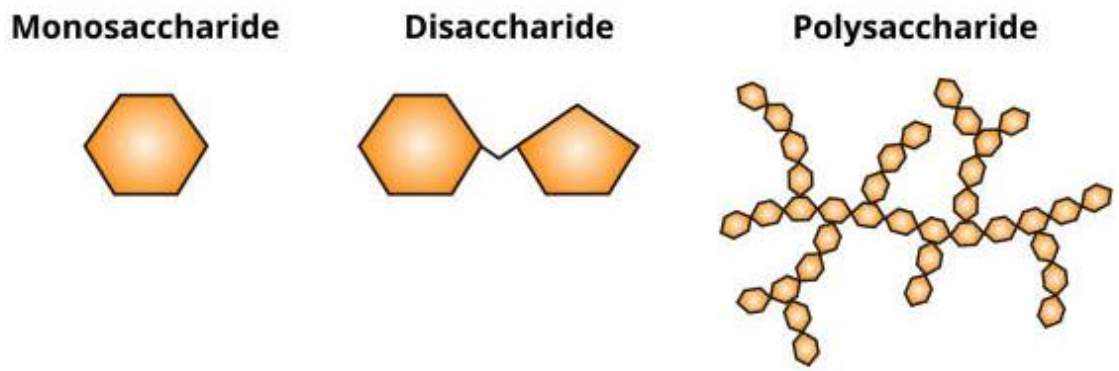
Carbohydrates - Intro

• **Carbohydrates, glycans, sugars or saccharides** are a diverse group of biomolecules that represent 2-10% of dry matter in humans and as high as 60-90% in plants (depending on the source).

• They serve many roles in cells:
  - energy metabolism and storage (e.g., glucose, glycogen, and amylose)
  - markers of cellular identity (e.g., glycolipids and glycoproteins),
  - structural components (e.g., cellulose in plants),
  - constituents of nucleotides (e.g., ribose in RNA, deoxyribose in DNA)

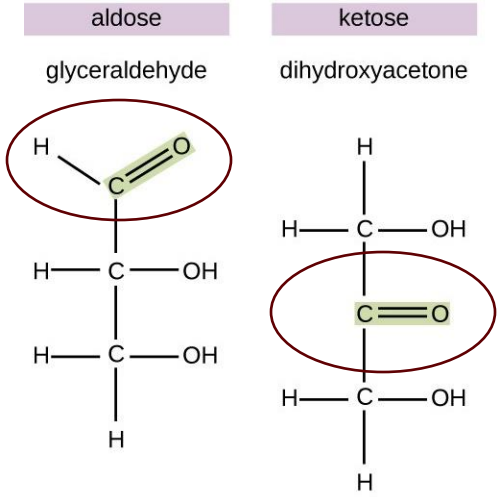
• General Formula: $\mathbf{C_nH_{2n}O_n}$
(due to the 1:1 ratio of C to H_2O, the name **carbohydrates** was proposed)

• Like nucleic acids they also form polymers from smaller building blocks (**monosaccharides**)



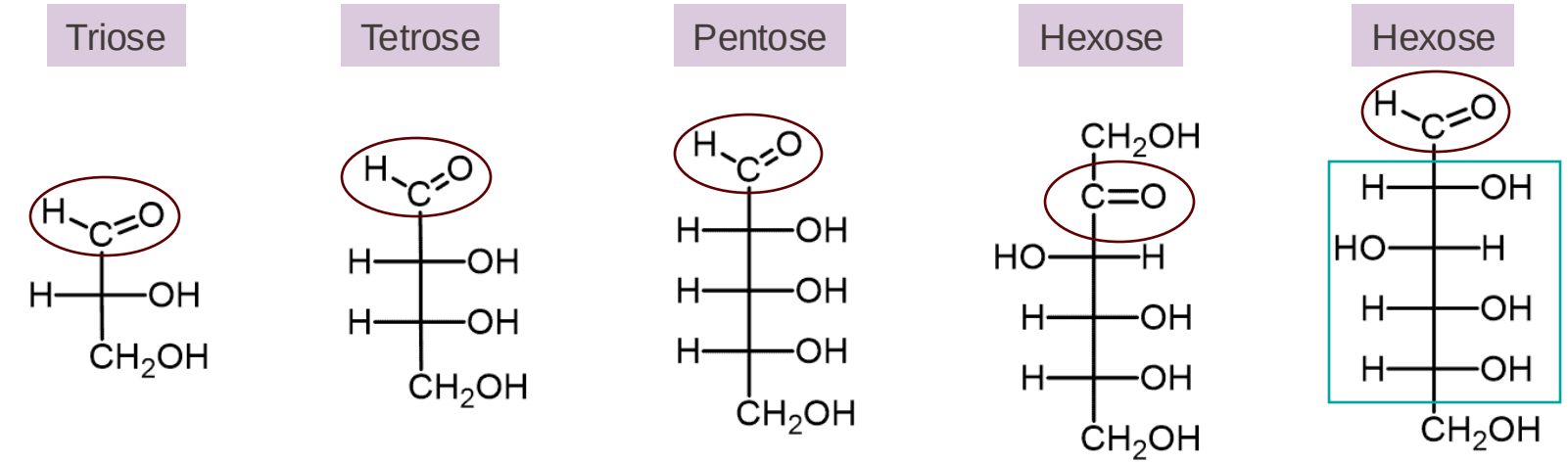
Monosaccharides

- Smallest groups – building blocks for larger carbohydrates
- Chain of 3 or more carbons, one carrying a carbonyl (C=O), and others with hydroxyl groups (OH)
- Their names carry the extension "**ose**"
- There are 2 primary chemical groups: aldehyde and ketone



- Aldoses have an aldehyde group: HC=O
- Ketoses have a keto group: C=O

- Categorized based on the length of sugar chain and the primary chemical group:

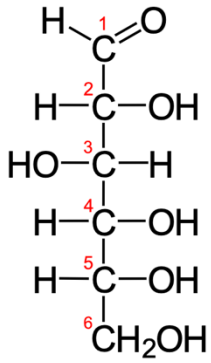


The diversity comes from the locations of H and OH groups around each carbon

Monosaccharides – Some important examples

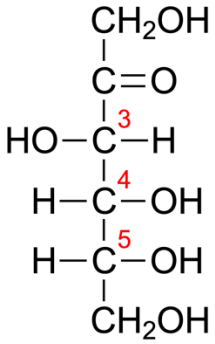
- For most of them historic names have been preferentially used, rather than the exact chemical names

• Glucose



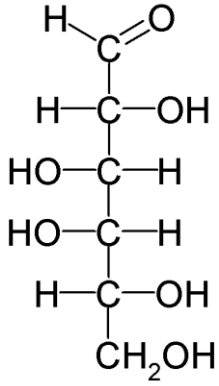
- The most abundant monosaccharide
- Important in metabolism, energy storage, structural assembly etc.

• Fructose



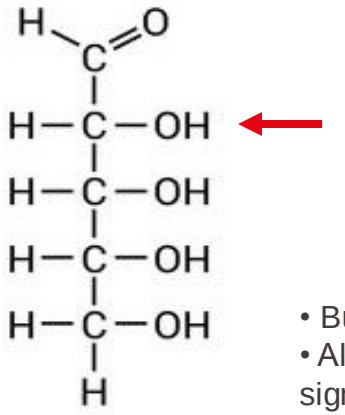
- Fruit sugar
- Component of sucrose

• Galactose



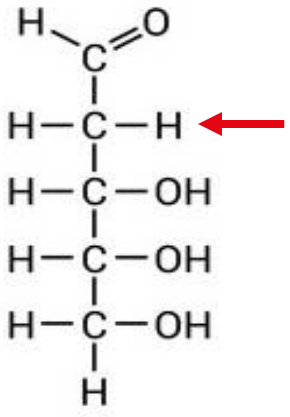
- Component of lactose (milk sugar)
- Present in glycoproteins and glycolipids (e.g., blood group antigens)

• Ribose

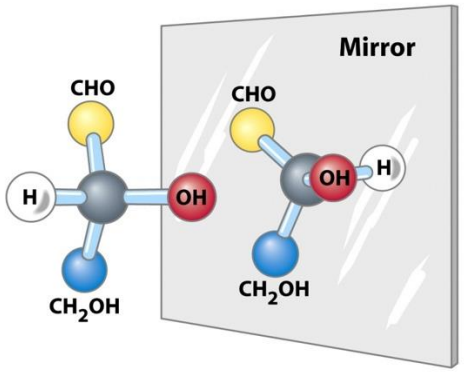


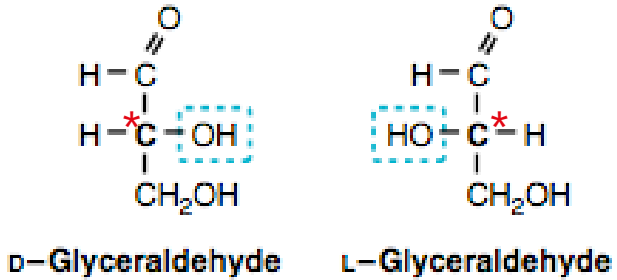
- Building blocks of nucleic acids
- Also important for intracellular signaling and metabolism

• Deoxyribose



Chirality of monosaccharides





- Chiral Center appears when there are **4 different substituents**

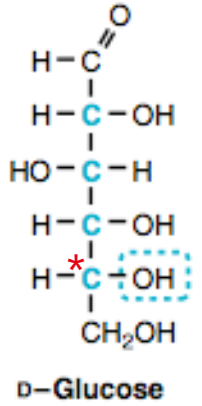
- D and L mirror images (nomenclature based on rotation of polarized light, D=dexter, L=laevus)

- Glyceraldehyde has a single chiral center, while longer sugars have multiple.

- Sugars defined with respect to glyceraldehyde (defined by asymmetric carbon)

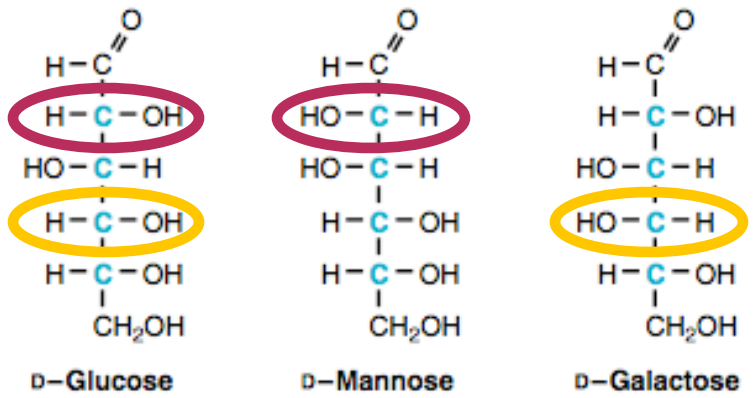
- **Most human sugars are D**

The main chiral center is always the first asymmetric carbon from the distal end (*).



D-glucose also called **dextrose** in the clinic

Stereoisomers and epimers

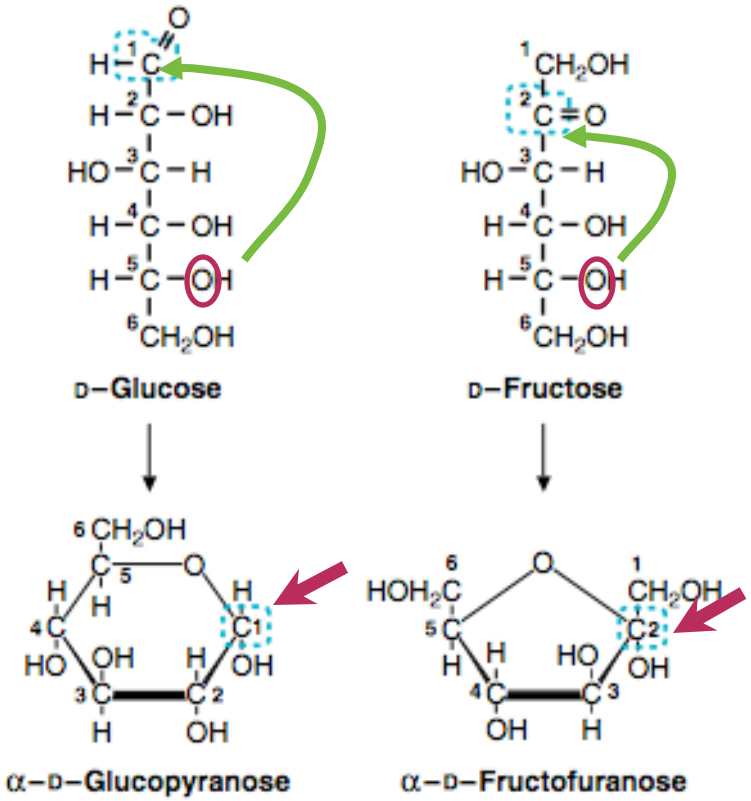


• Stereoisomers: same formula ($C_6H_{12}O_6$ in this example) but differ in position of OH at one or more chiral carbons

• Epimers: stereoisomers that differ at only one chiral carbon (compare glucose-mannose with glucose-galactose)

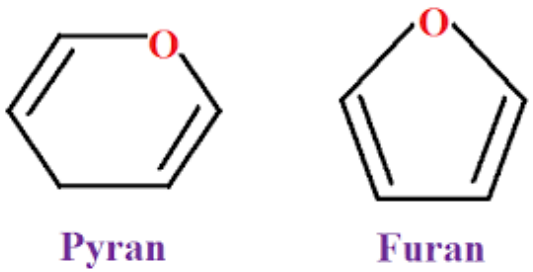
• Epimers interconverted by enzymes called epimerases (i.e. the OH changed to other chiral position)

Formation of cyclic structures in solution



- In solution monosaccharides form ring structures (C=O group attacked by the OH group of the chiral carbon)
- C=O becomes anomeric carbon (i.e. a stereo center)
- This is only possible for sugars with 5 or more carbons
- At anomeric carbon, OH below ring=alpha, above ring=beta

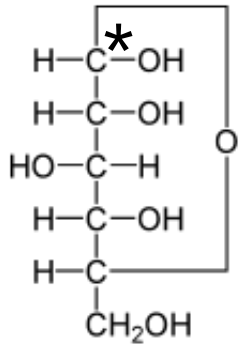
Naming scheme based on:



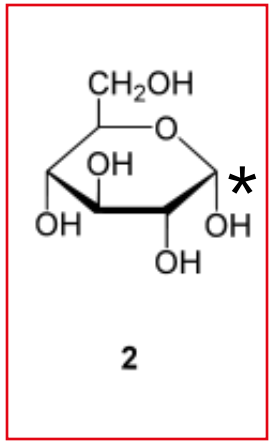
Although note that there are no double bonds in actual sugars

Different representations of glucose

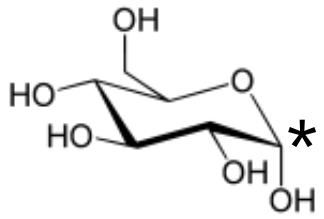
- Please learn how to convert from linear to simple cyclic representation



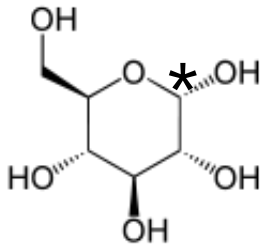
1



2



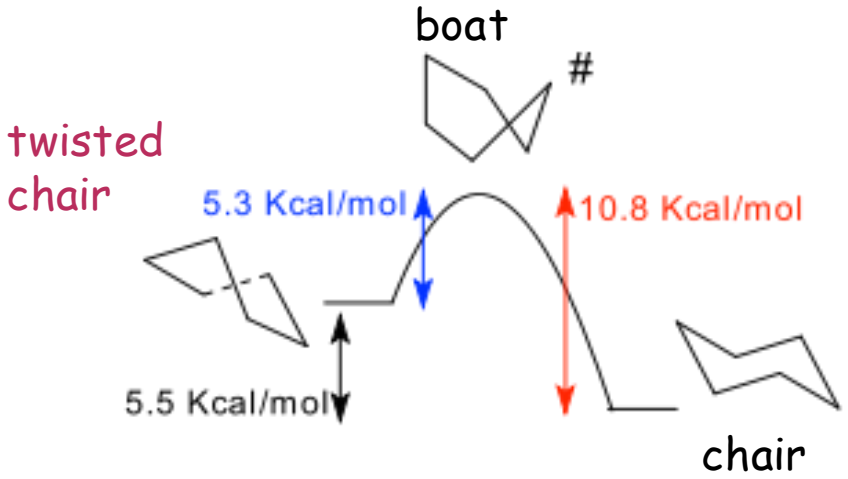
3



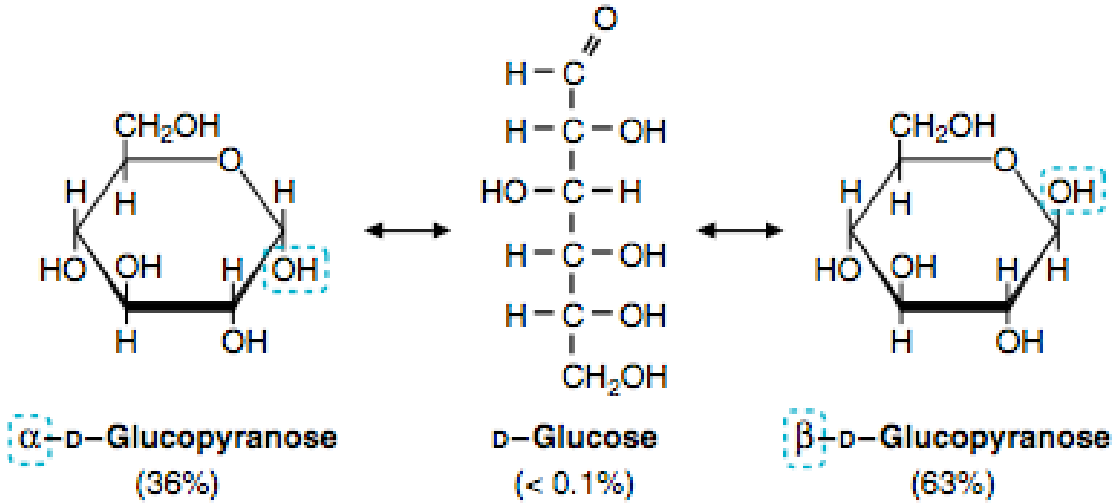
4

- Anomeric carbon denoted by asterisk
- H's not always shown
- Wedges represent view (toward and away)

- Due to the lack of double bonds the structure is not planar
- Predominant structure is the "chair" (lowest energy)



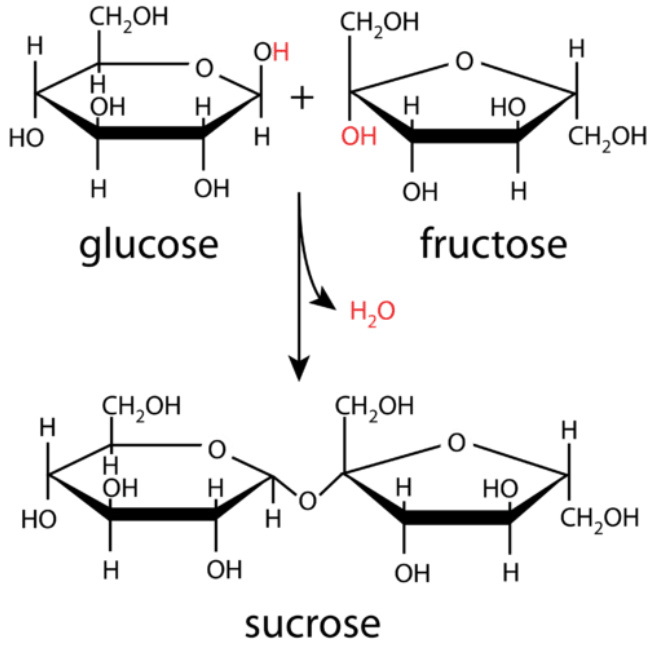
Mutarotation around the anomeric carbon



- In solution there is an equilibrium between 3 forms (alpha, beta and linear)
- Note how the H and OH groups around the anomeric carbon swap places
- Mutarotases: enzymes that catalyze mutarotation (i.e. speed up reaction)
- Enzymes often react with only one stereoisomer (alpha or beta)

Formation of larger carbohydrate chains

- Monosaccharides may be covalently linked to form longer chains
- This reaction is catalyzed by an enzyme group called synthases
- Water molecule is generated from the leaving H and OH groups
- The linkage between glycans is called “glycosidic”
- The inverse reaction (sugar breakdown) is called hydrolysis

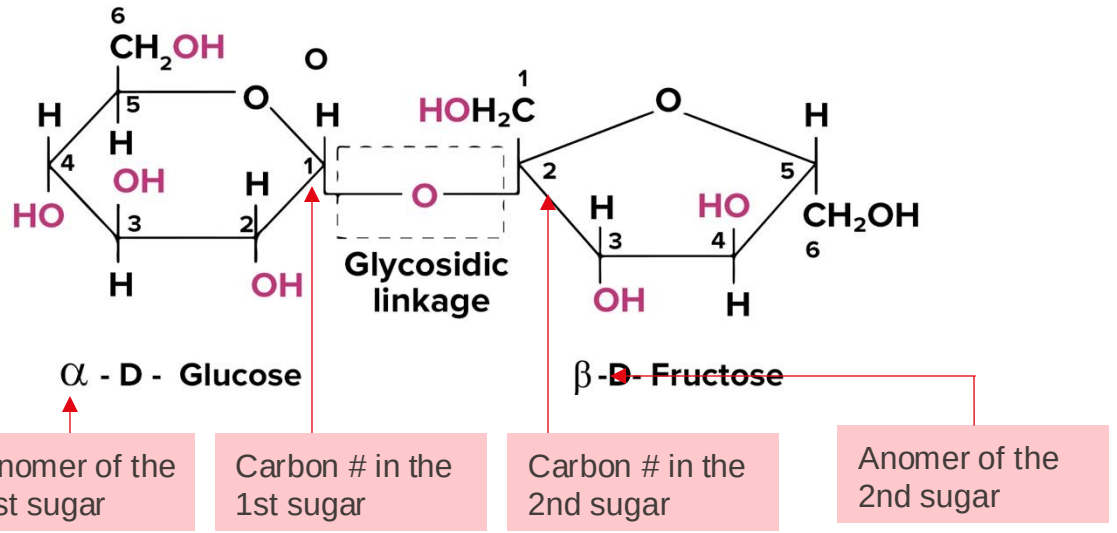


Glycosidic bond naming:

α1 -> β2 O-glycosidic bond

or

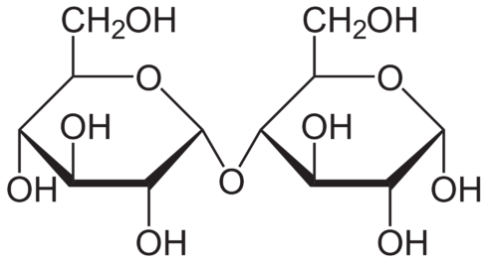
α(1-2) O-glycosidic bond



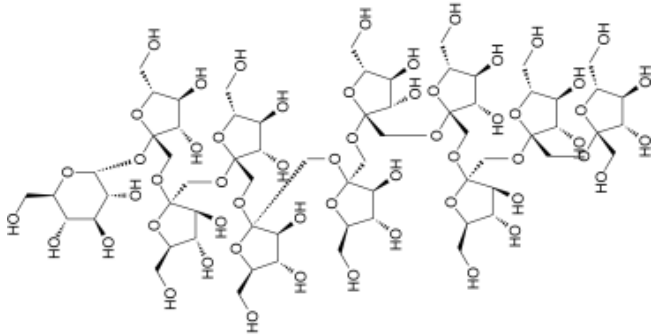
Categories of larger carbohydrate chains

- Disaccharide: 2 monosaccharides linked by O-glycosidic bond (e.g. sucrose lactose)
- Oligosaccharides: 3-12 monosaccharides linked by O- or N-glycosidic bonds (e.g. glycoprotein or glycolipid moieties)
- Polysaccharides: very large number of linked monosaccharides (e.g. starch, cellulose)

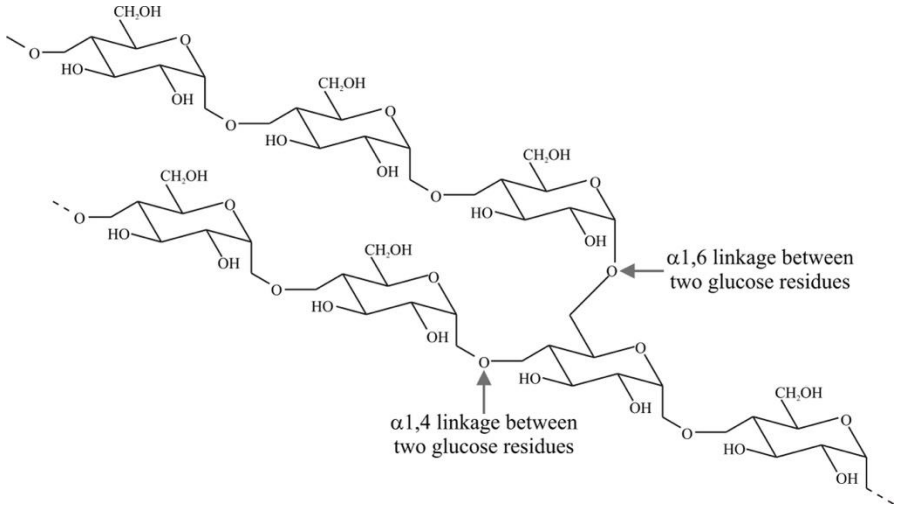
Disaccharide
(Maltose)



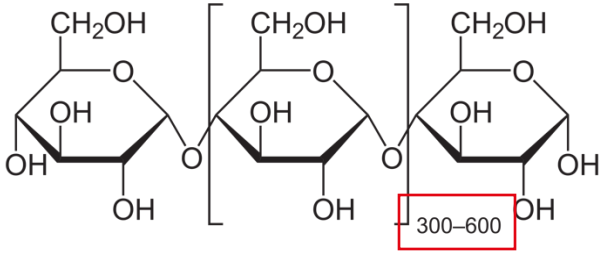
Oligosaccharide
(Oligofructose)



Branched polysaccharide
(Amylopectin)

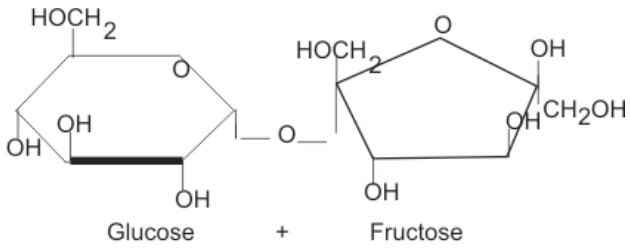


Linear polysaccharide
(Amylose)



Components of starch

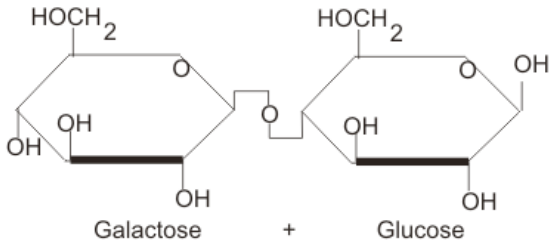
Common disaccharides



Sucrose

Glucose + Fructose
α(1-2) linkage

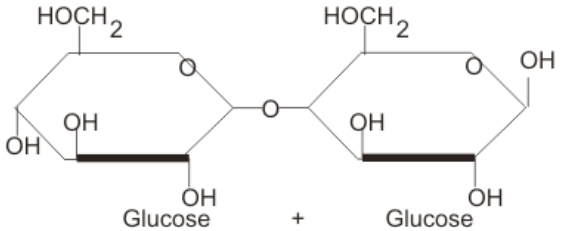
Table sugar



Lactose

Galactose + Glucose
β(1-4) linkage

Milk sugar



Maltose

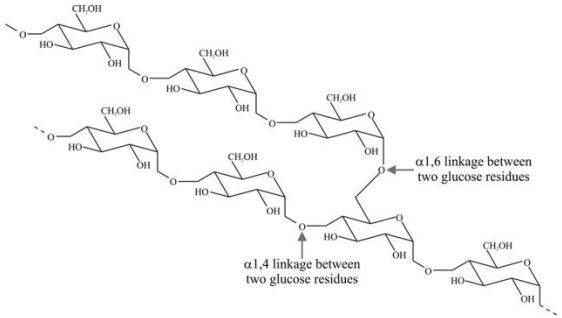
Glucose + Glucose
α(1-4) linkage

Malt sugar
(breakdown product of starch)

Common polysaccharides

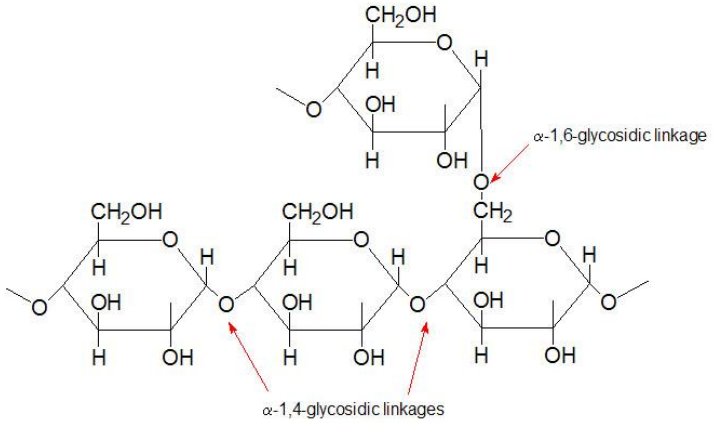
Starch

- Poly-glucose found in plants
- Storage of energy, source of carbon for photosynthesis
- 2 components:
  - Amylose: **alpha(1-4)** glucose linkage
  - Amylopectin: **alpha(1-4) and alpha(1-6)** glucose linkage



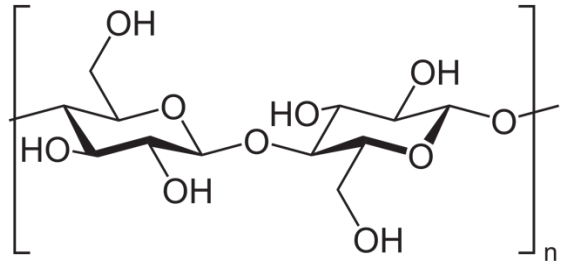
Glycogen

- Poly-glucose found in animals
- Storage form of glucose created in the liver under **insulin** control
- Glycogenolysis is controlled by the **glucagon** hormone
- Similar to amylopectin but with more **alpha(1-6)** linkages



Cellulose

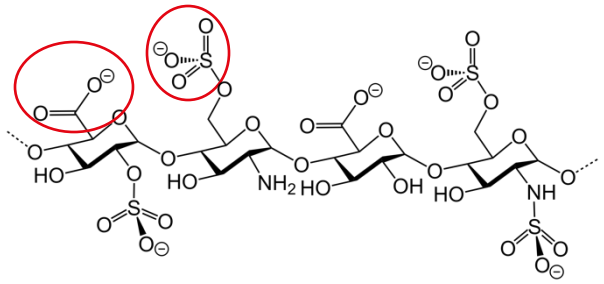
- Poly-glucose found in plants where it assembles the cell wall
- The most abundant biomolecule on the planet
- Glucose linked by **beta(1-4) bonds**



Other important sugar molecules

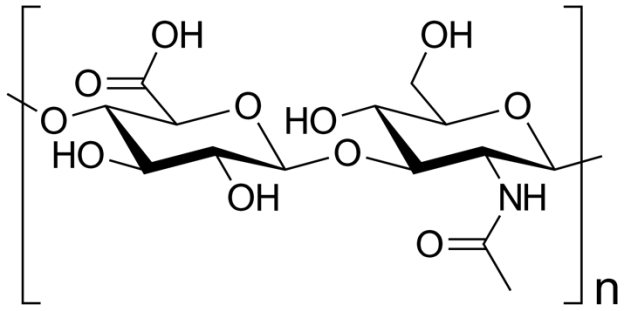
Glucosaminoglycans

- Heparin



Used as anticoagulant in clinic

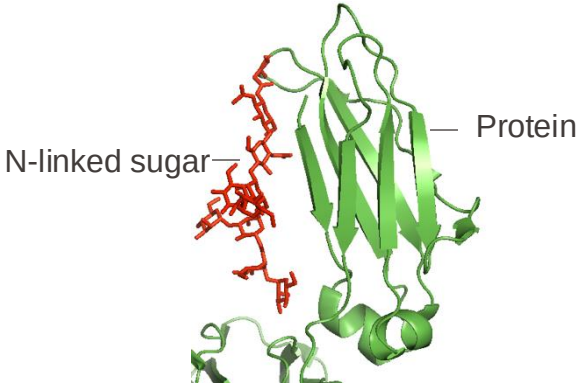
- Hyaluronic acid

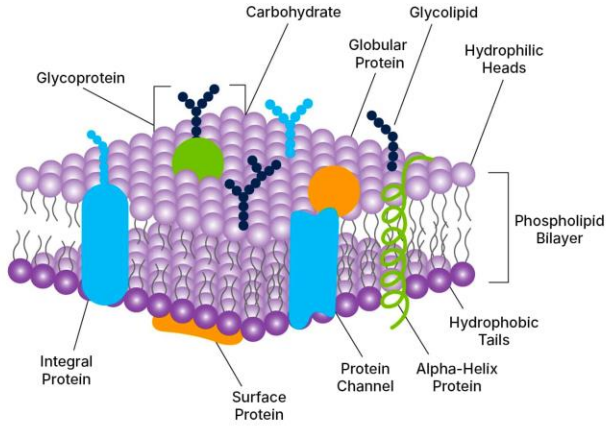


Connective tissue, extracellular matrix

Note that monosaccharides can incorporate amino, sulfate and carboxylic acid groups

Glycoproteins

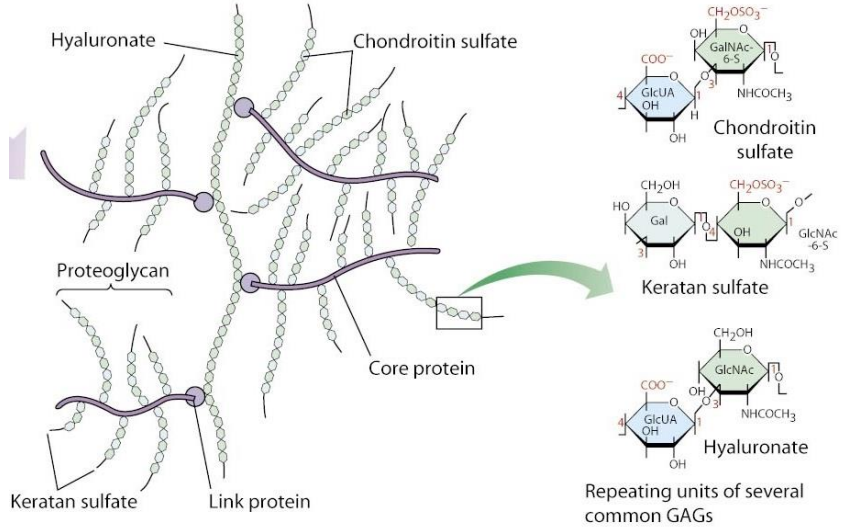




Glycoproteins and glycolipids serve as cell biomarkers (but can have other roles)

Proteoglycans

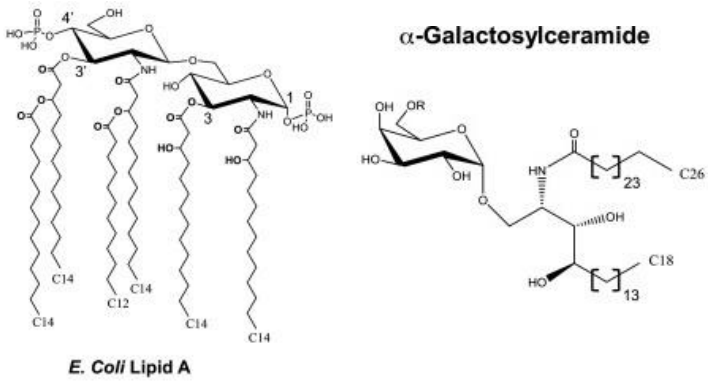
- Glycoproteins with very large amounts of sugar attached (negatively charged)
- Components of extracellular matrix



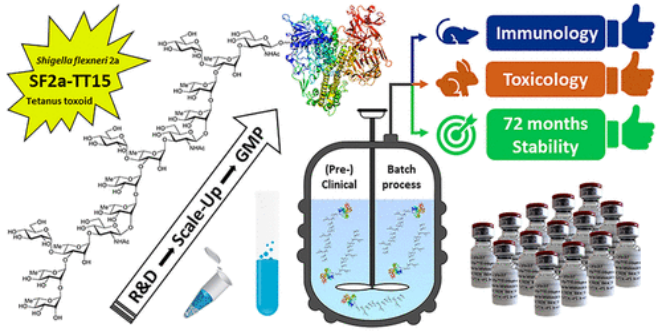
Carbohydrates in bioengineering

Vaccines and adjuvants

Bacterial glycolipids (LPS)



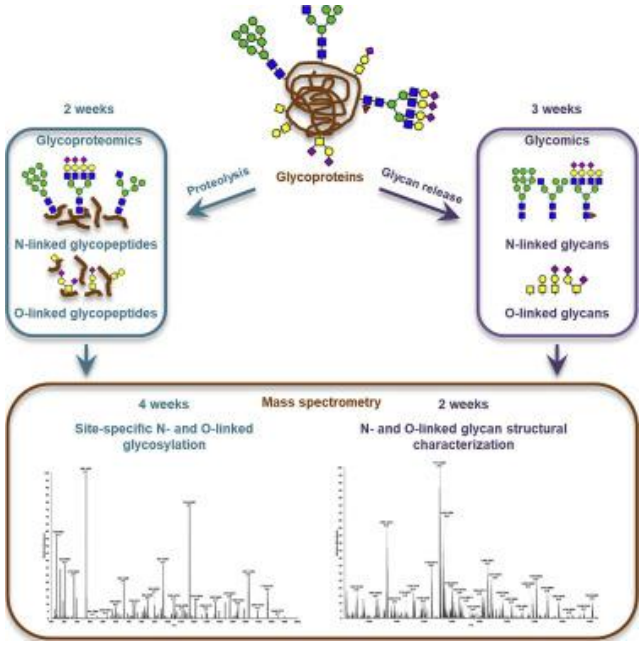
Fully synthetic glycan vaccine for Shigella



https://research.pasteur.fr/fr/team/chemistry-of-biomolecules/

Glycoproteomics

Identification of biomarkers



Bioorthogonal chemistry of sugars

https://bertozzigroup.stanford.edu/

Nobel Prize in Chemistry, 2022

Nutrition

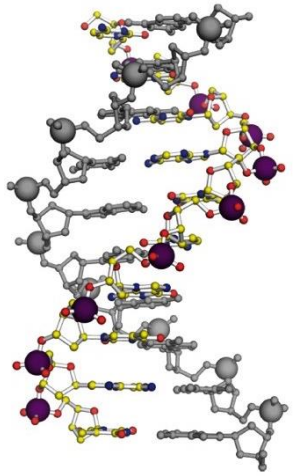
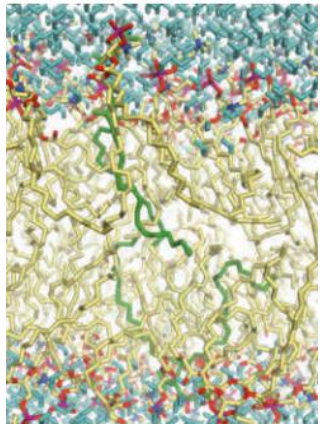
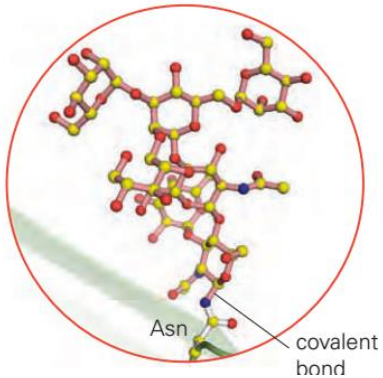
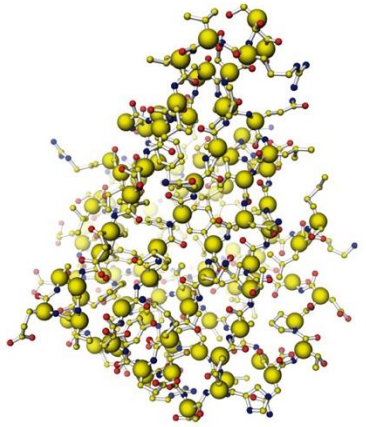
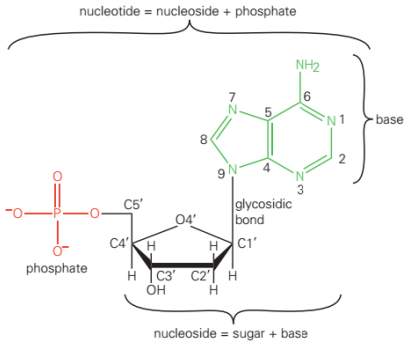
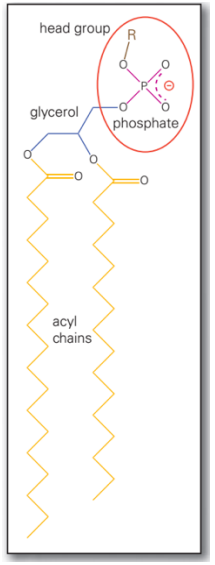
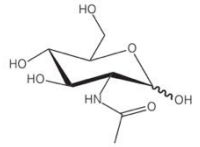
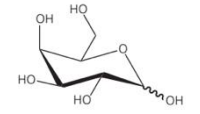
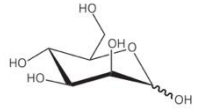
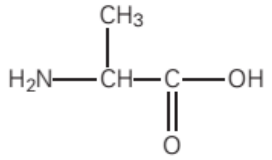
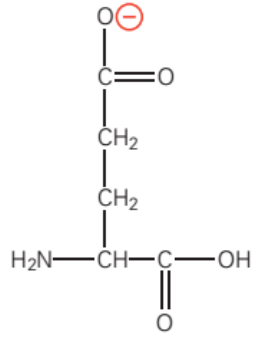
AI in food and nutrition

https://www.epfl.ch/labs/salathelab/

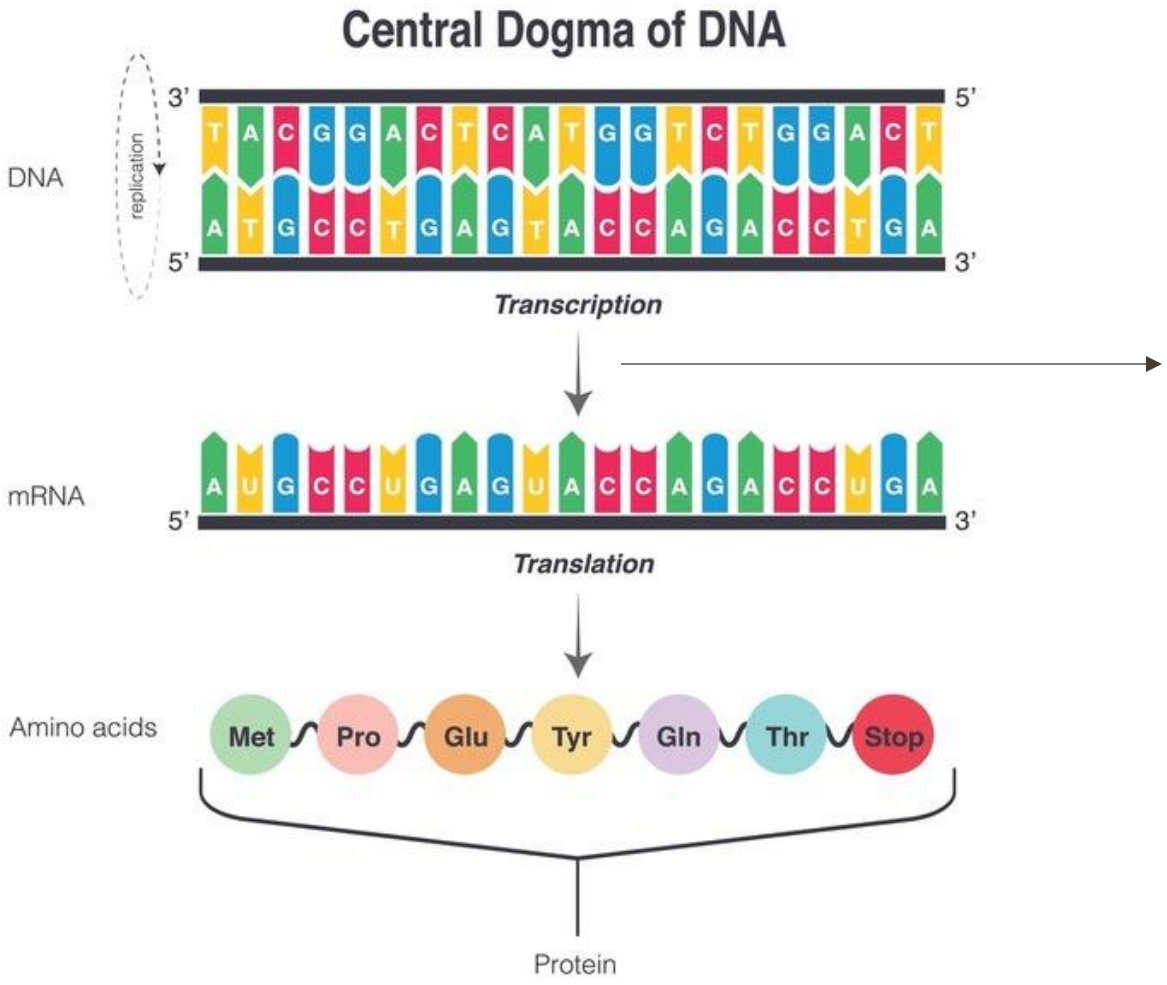
Carbohydrates - Take home messages

- Carbohydrates are a diverse family of biomolecules with primary function in energy storage and metabolism.
- Monosaccharide represent the smallest building blocks of carbohydrates with the general formula of $\mathbf{C_nH_{2n}O_n}$ but can include other groups (e.g., sulfate, amino, and carboxylic acid)
- Monosaccharides with 5 or more carbons exist as cyclic molecules in solution
- Longer carbohydrate chains are built from monosaccharide units connected via glycosidic bonds
- Polymers are very dynamic and lack a structured 3D assembly
- They can also be conjugated to proteins and lipids and presented on cell surface (signaling role) or as part of extracellular matrix (structural role)

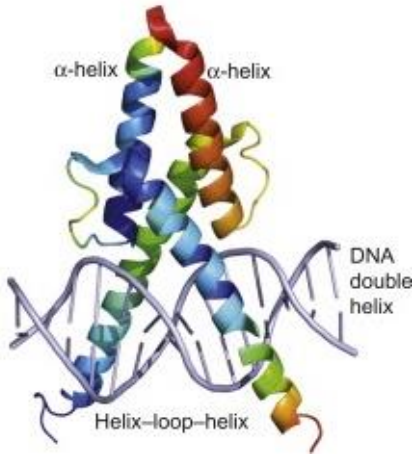
The molecules of Life

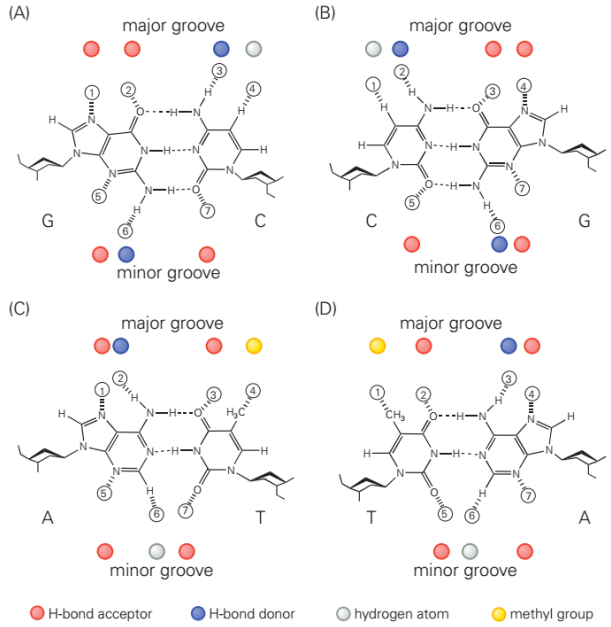
	Nucleic Acids	Lipids	Carbohydrates	Proteins
Macromolecular Structure				
Building Block			  	 

Genetic code is translated to protein sequence



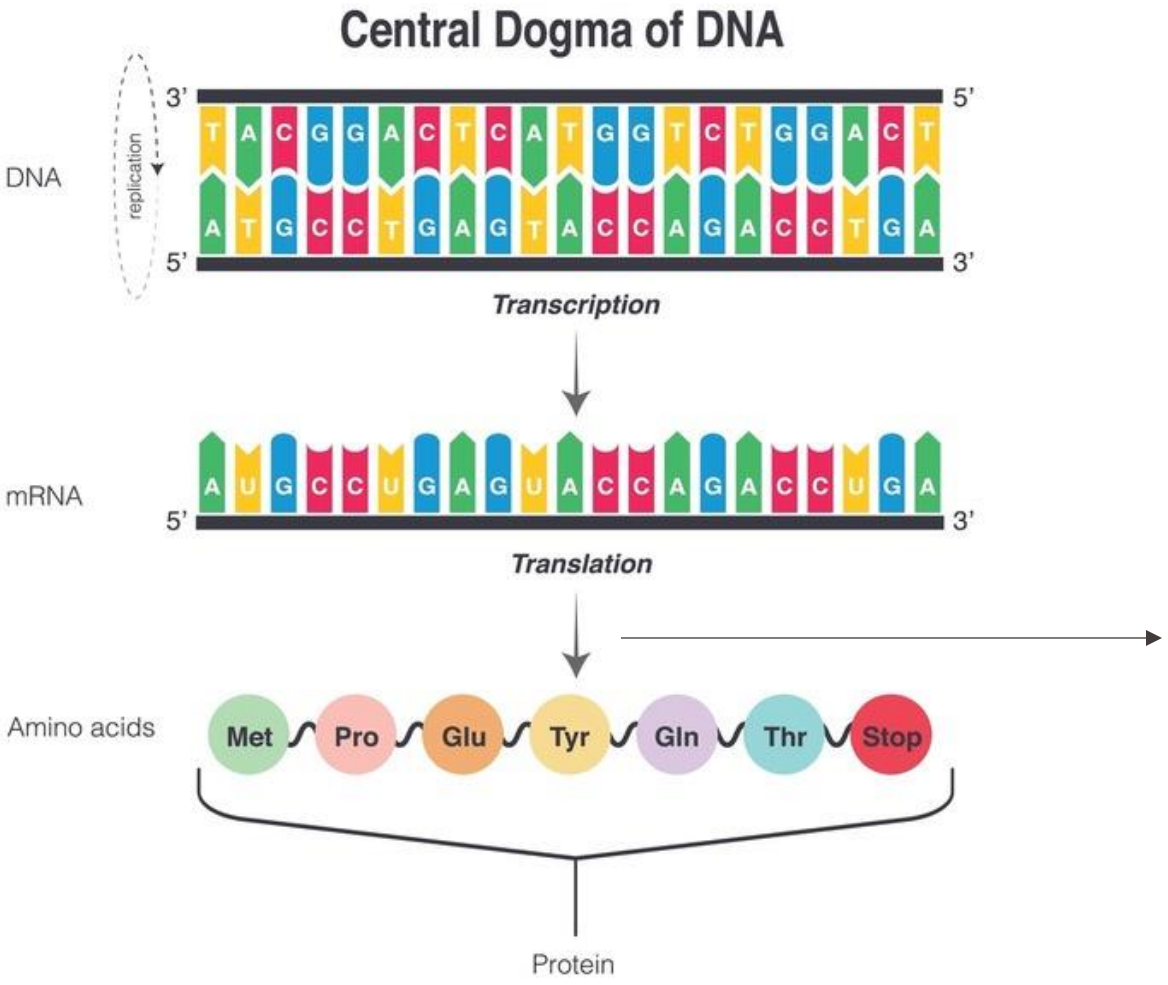
• Transcription factors recognize the gene locations through the major and minor groove of the DNA chain



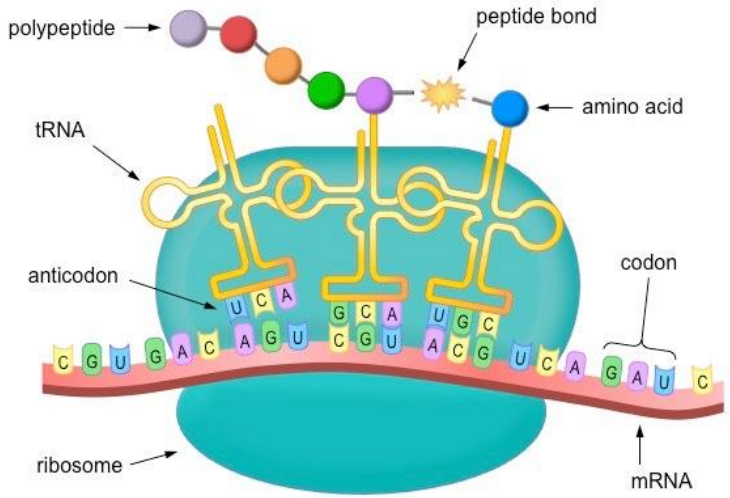


• The sensing is mediated by interactions with available hydrogen bond donor and acceptor group (=barcode)
• These interactions are very specific and lead to transcription of downstream gene sequence into RNA

Genetic code is translated to protein sequence

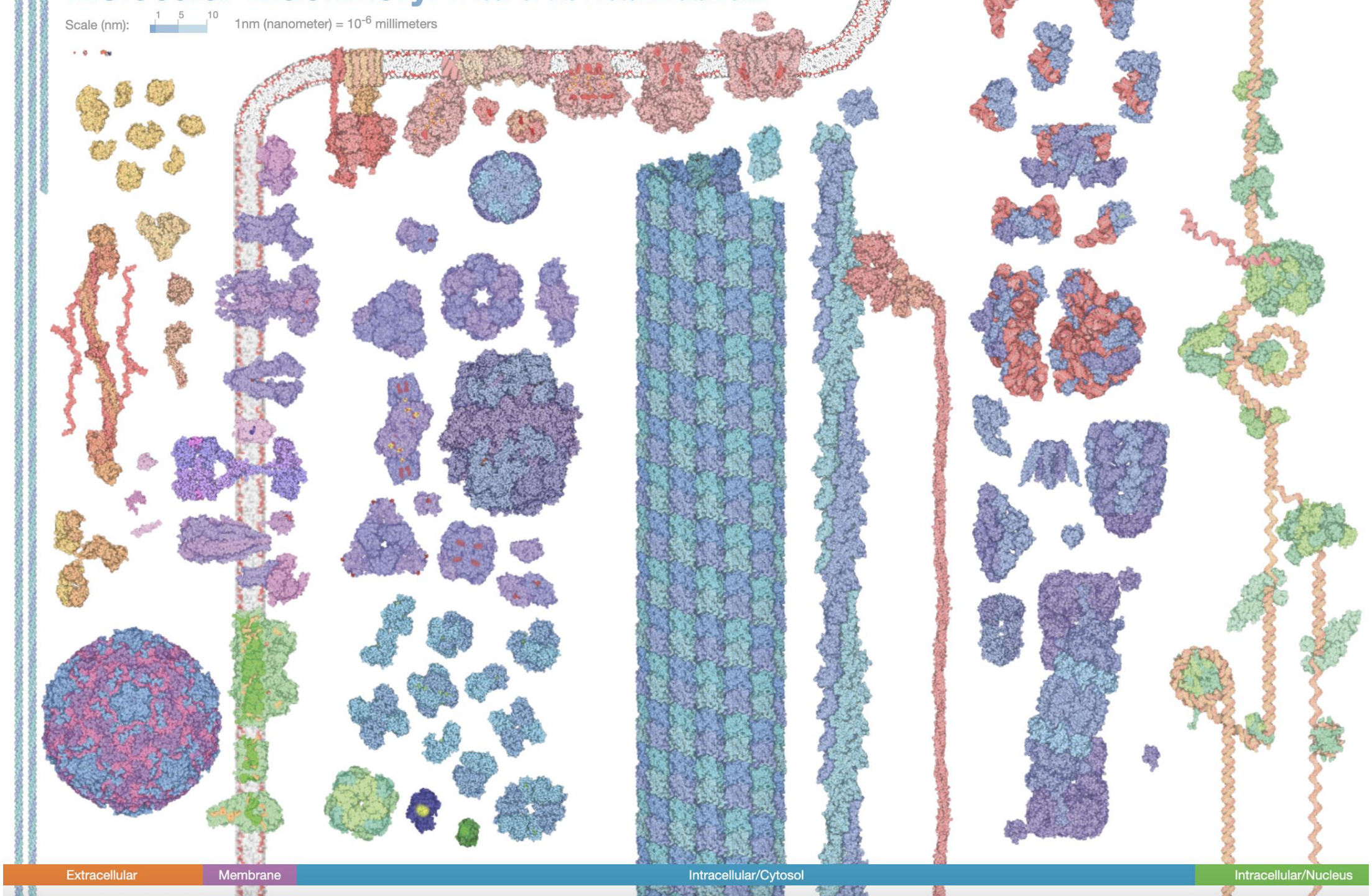


- mRNA moves to the ribosome where it is translated into protein sequence
- 3 nucleotides = 1 codon -> 1 amino acid

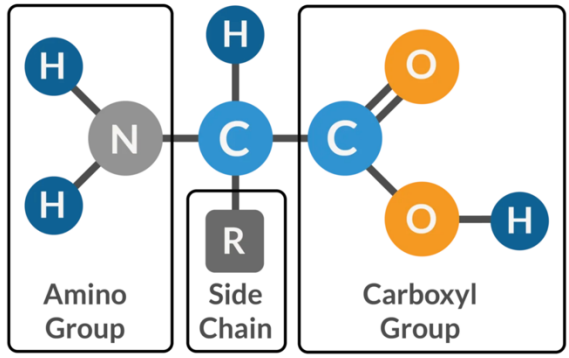


- Transfer RNA (tRNA) molecules carry a different amino-acid for each different codon in mRNA
- Recognition is driven by base complementarity between codons in mRNA and anticodons in tRNA

Protein diversity

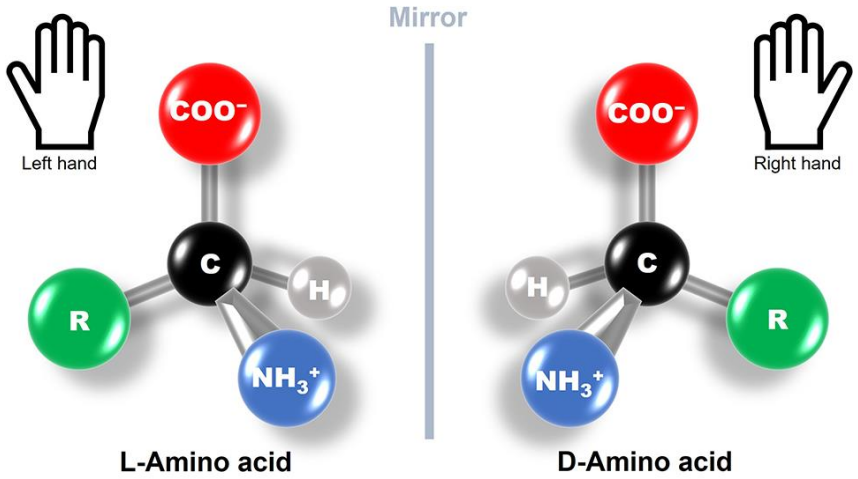


Amino acids are the building blocks of proteins

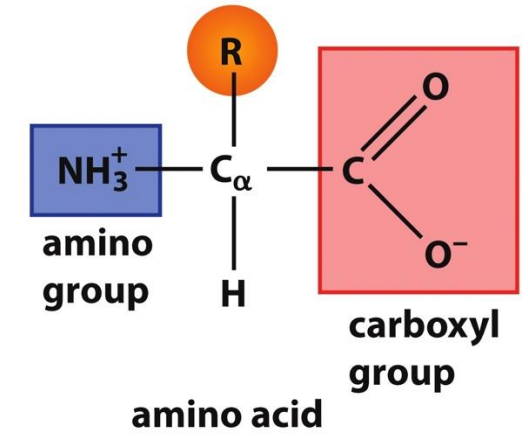


- Protein building blocks are the amino acids encoded by the DNA.
- The general structure of an amino acid is :
- Note that the amino acids are zwitterionic possessing positive and negative charge
(net charge is a function of the pKa value of the acid/base group and the pH of the solution)

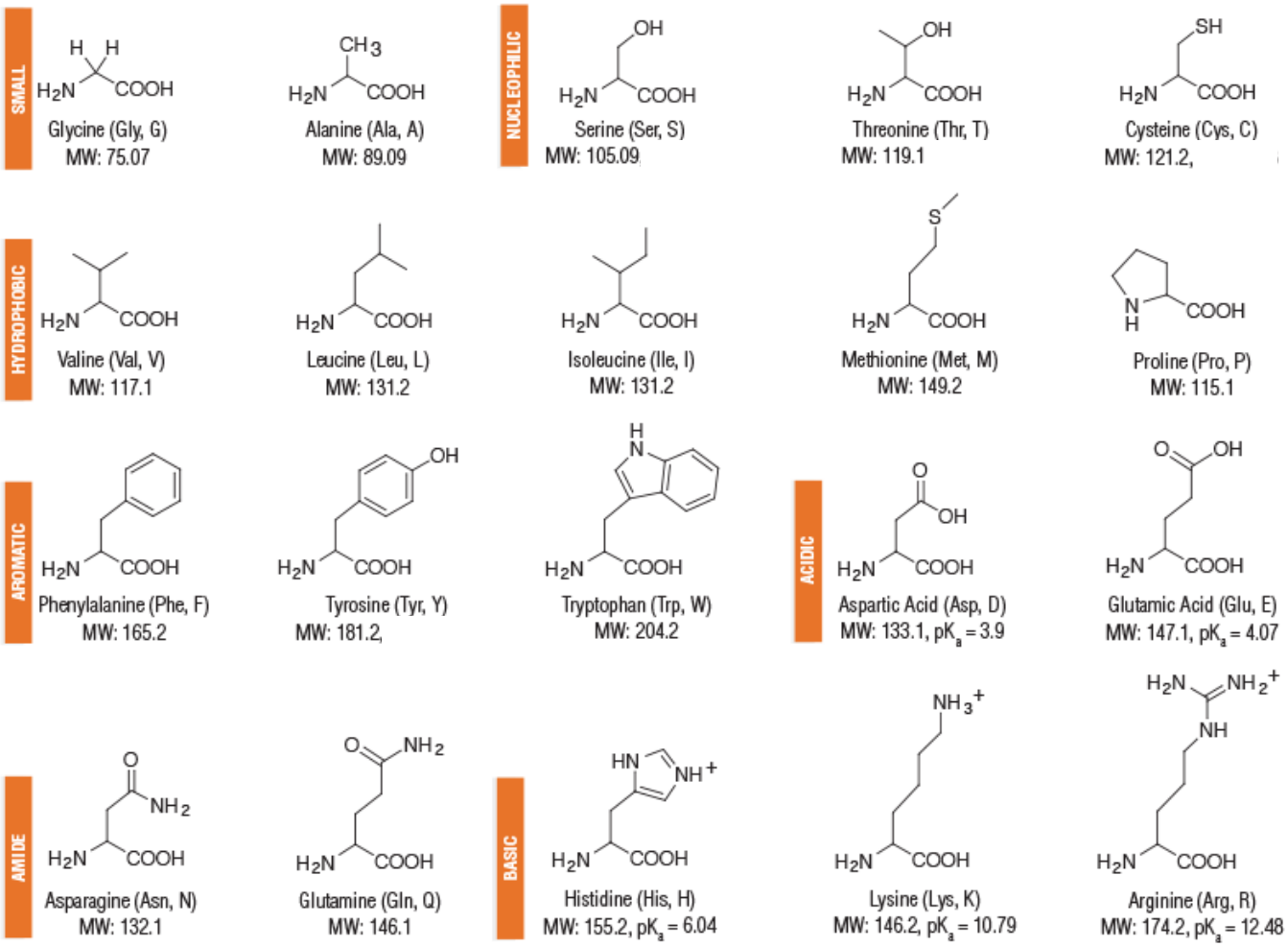
- Amino acids are chiral and genetically encoded amino acids have the **L form**
(Note that L amino acids can't be converted in D just by rotation)

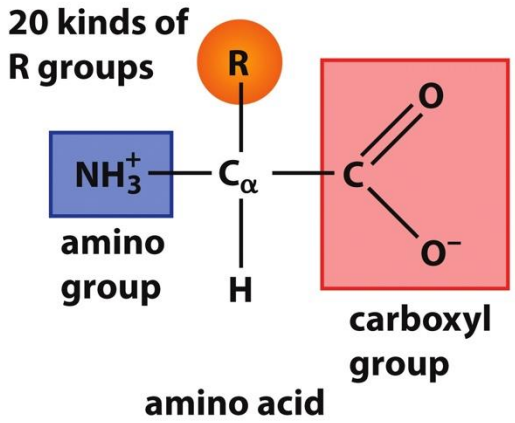


Zwitterionic form in solution



Amino acids are the building blocks of proteins



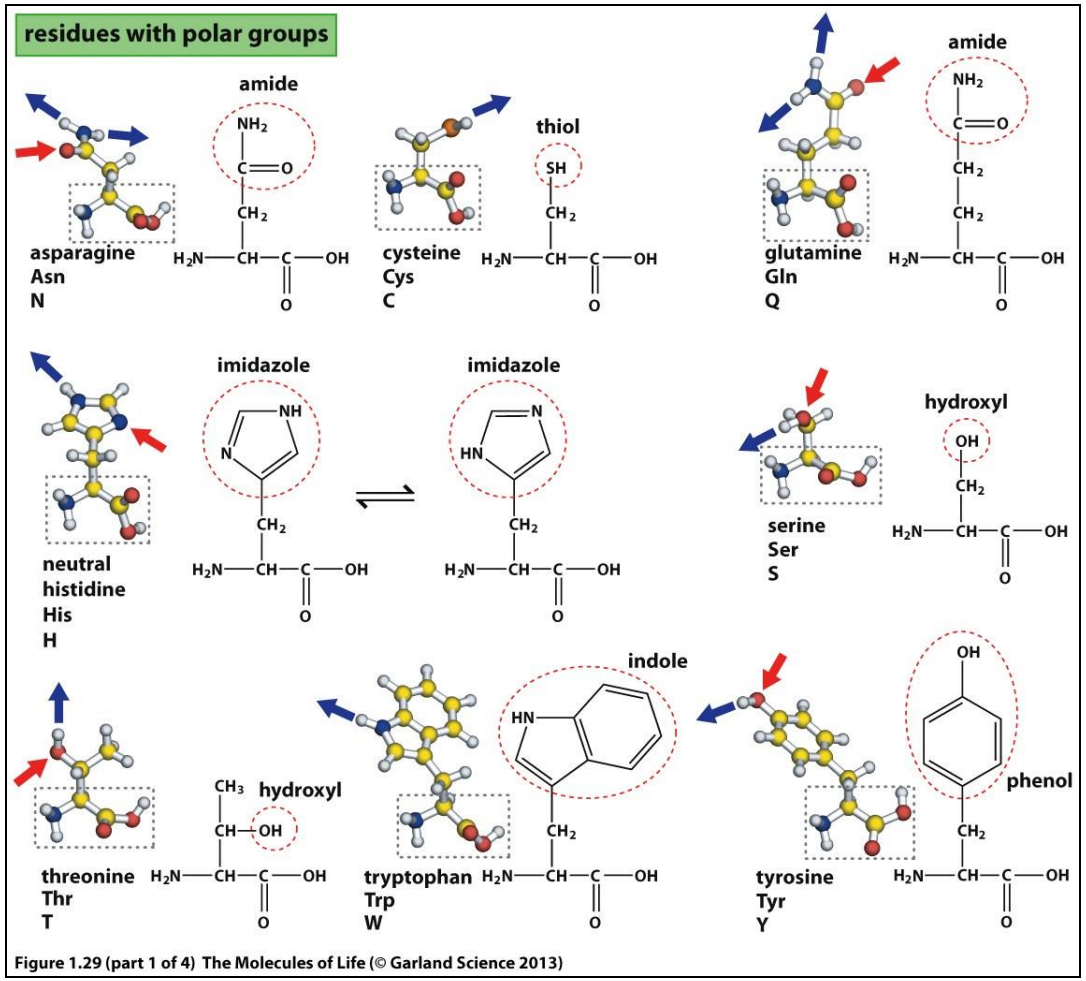


- There are 20 primary amino acids that assemble all proteins
- Amino acids have the same scaffold but differ only in the side chain group (R)
- Note 3- and 1- letter abbreviations for each amino acid

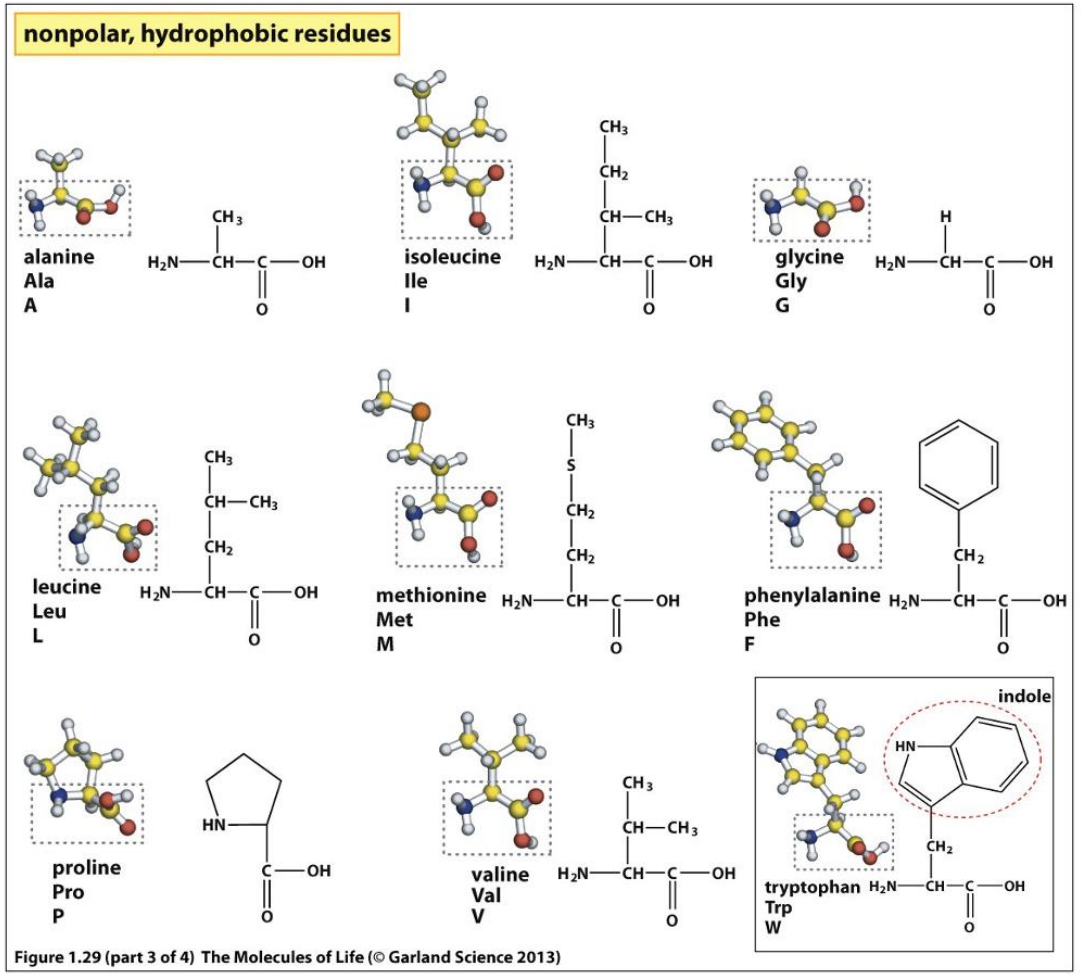
* There are also non-canonical amino acids: Selenocysteine (Sec) and Pyrrolysine (Pyl) are known as the 21st and 22nd amino acids in the genetic code. Sec and Pyl are encoded by UGA and UAG codons, respectively, which normally serve as stop signals. But these are not very common among proteins.

Different groups of amino acids based on side-chain

- Imperfect categorization system but very useful to understand different properties of amino-acids



Note the H-bond donors (blue) and acceptors (red)

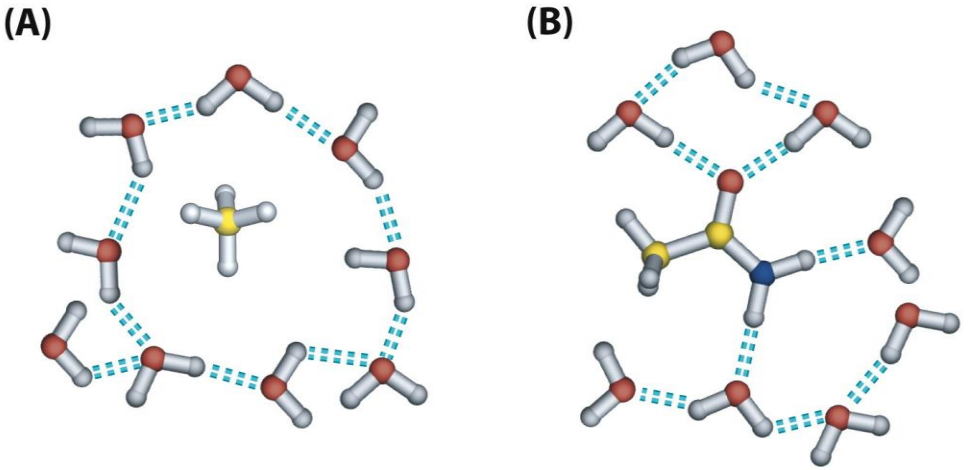


Similar to lipid tails, largely hydrophobic

Hydrophobicity of amino acids

• **Hydrophobicity** of the side chains of amino acids is one of the most important features for protein folding (e.g., exposed on surface versus buried inside the protein core) and localization (e.g., transmembrane domains are very hydrophobic)

Hydrophobic



Hydrophilic

• Relative scale of hydrophobicity:

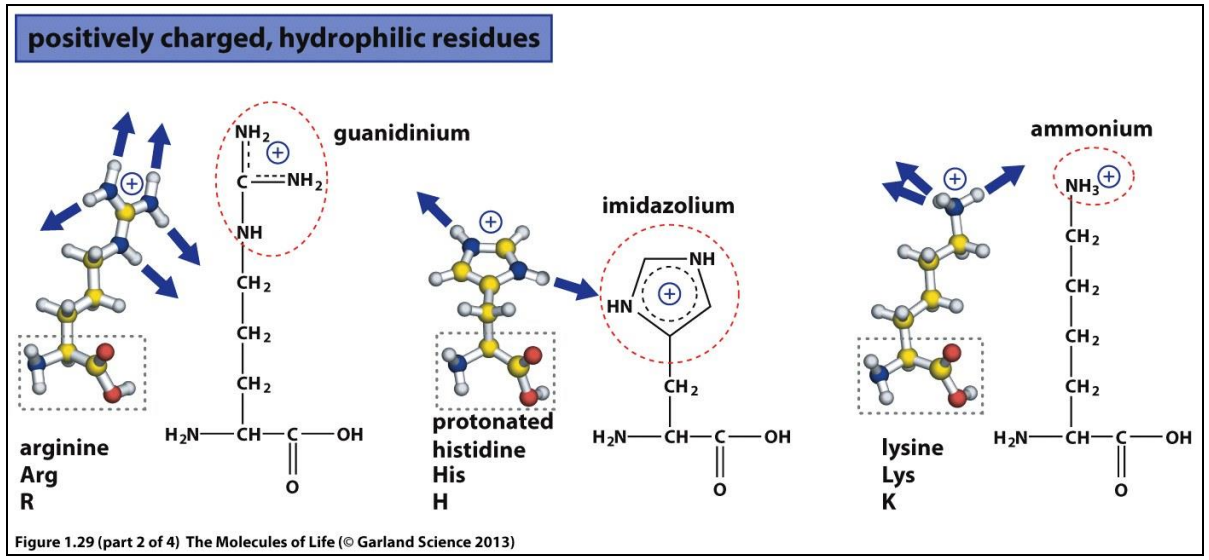
Water/Octanol transfer free energy kJ/mol

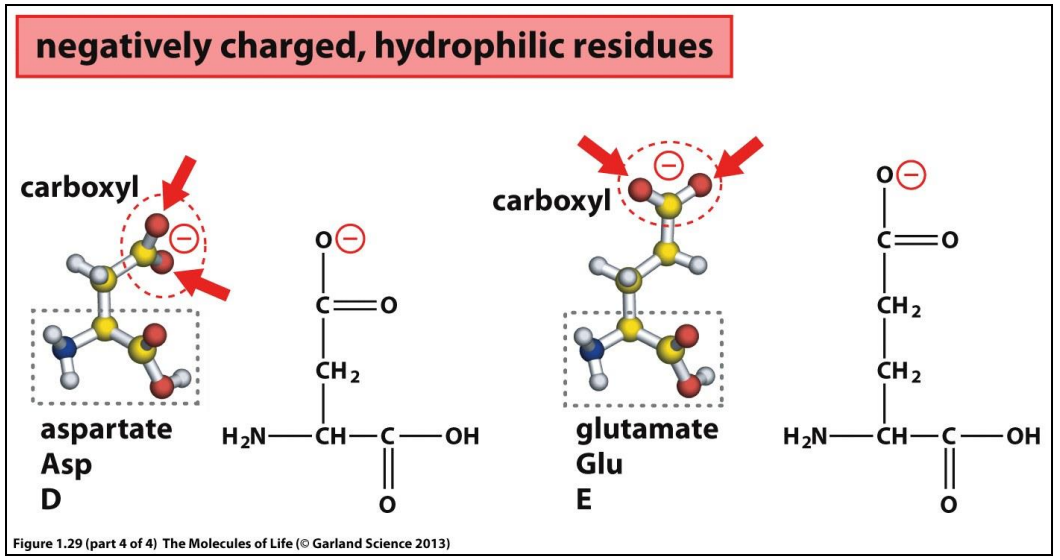
Trp	Phe	Leu	Ile	Met	Tyr	Val	Cys	Pro	His^0	Thr	Ser	Ala	Gln	Asn	Gly	Arg	His^+	Lys	Glu	Asp
−8.8	−7.1	−5.0	−4.5	−2.9	−2.9	−2.1	0.0	+0.4	+0.4	+0.8	+2.1	+2.1	+3.3	+3.8	+4.5	+7.5	+9.5	+11.7	+15.0	+15.0

Hydrophobic **Hydrophilic**

Different groups of amino acids based on side-chain

- Asp and Glu are amino-acids with negatively charged carboxyl groups in their side chains
- Arg, Lys and His have positively charged amine-based groups in their side chains
- The net charge depends on the nature of amino-acid side-chain (pKa) and the pH of the environment





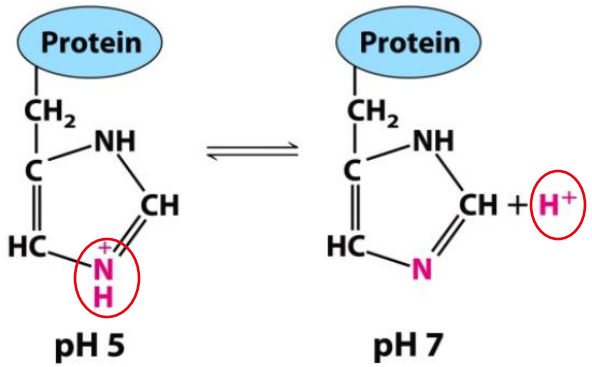
pKa values: 12.48 | **6.04** | 10.79 | 3.90 | 4.07

What about His?

pH > pKa -> Deprotonated group (e.g., COO-)
pH < pKa -> Protonated group (e.g., COOH)

Interesting amino acids

• **Histidine** can get protonated if the pH drops below ~6, giving it a positive charge
• Sometimes serves as pH-sensitive trigger in proteins



• **Cysteine** can form covalent disulfide bonds with other cysteines under oxidizing conditions (e.g., outside of a cell).
• Reversed under reducing conditions

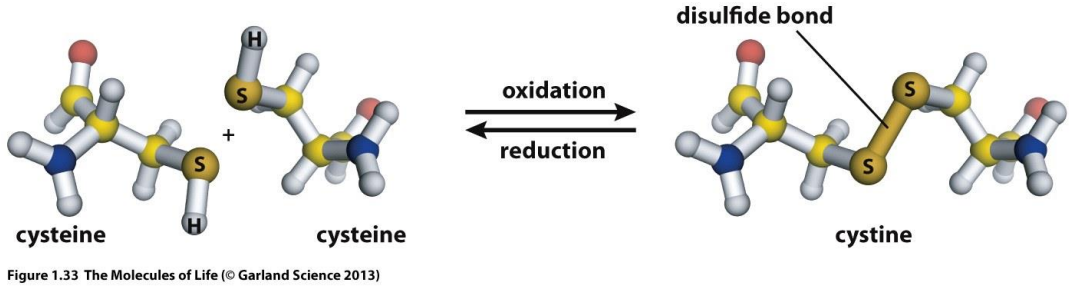
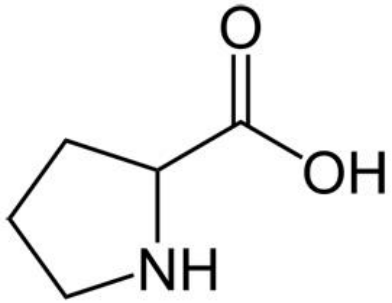
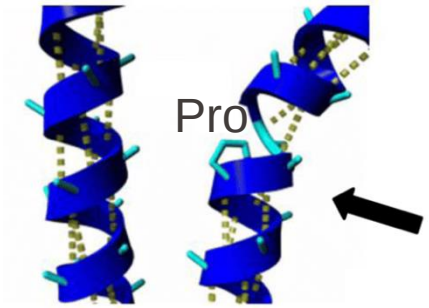


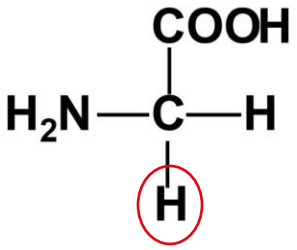
Figure 1.33 The Molecules of Life (© Garland Science 2013)

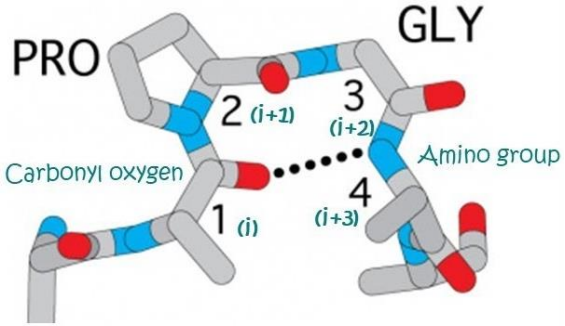
• **Proline** is not a real amino acid but rather an **imino** acid.
• Disruptive to **α**-helices and **β**-strands in proteins, but commonly found in flexible loops and turns





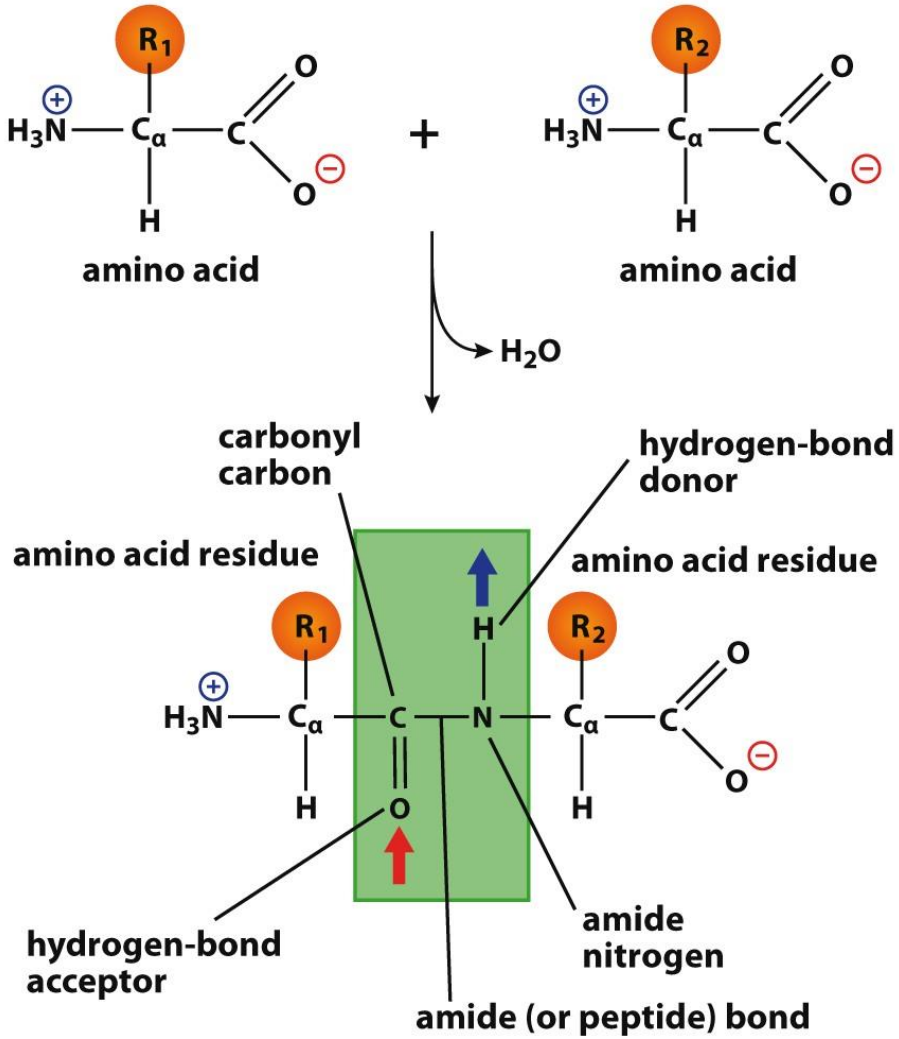
• **Glycine** only has a single hydrogen as a side-chain, which allows for more movement in proteins.
• Commonly found in flexible loops or peptide turns





Proteins are polymers of amino acids

- Proteins are assembled from amino acids through translation of genetic material



- The synthesis of proteins involves a **condensation reaction** in which the amino group of one amino acid combines with the carboxyl of another.

- This reaction forms a **peptide bond** and the elimination of a water molecule

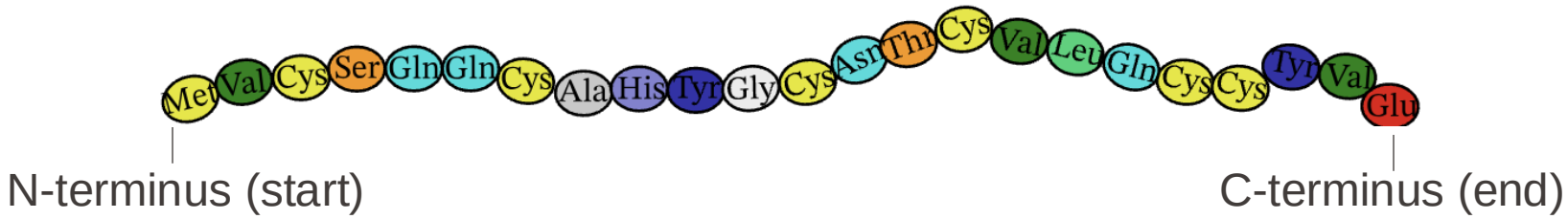
- This reaction is catalyzed by a large assembly of proteins and RNA called the **ribosome**

- The end product is a **polypeptide chain** where amino acids are added sequentially (no branching)

Figure 1.30 The Molecules of Life (© Garland Science 2013)

Proteins are polymers of amino acids

• The sequence of a protein is written from the **N-terminus** (the one with free NH_3^+) to the **C-terminus** (the one with the COO^-)



• Chemical and three-dimensional structure of a short peptide sequence

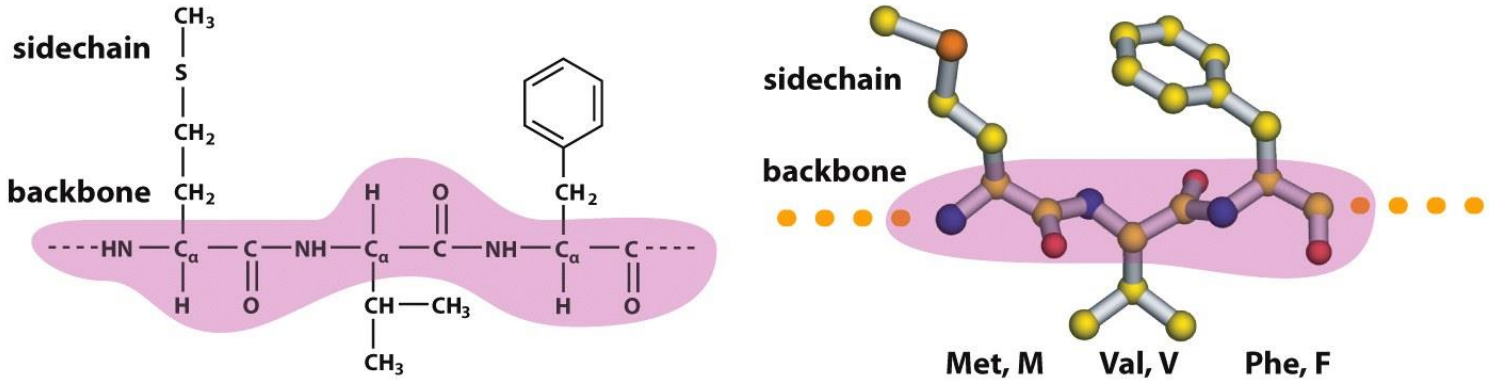
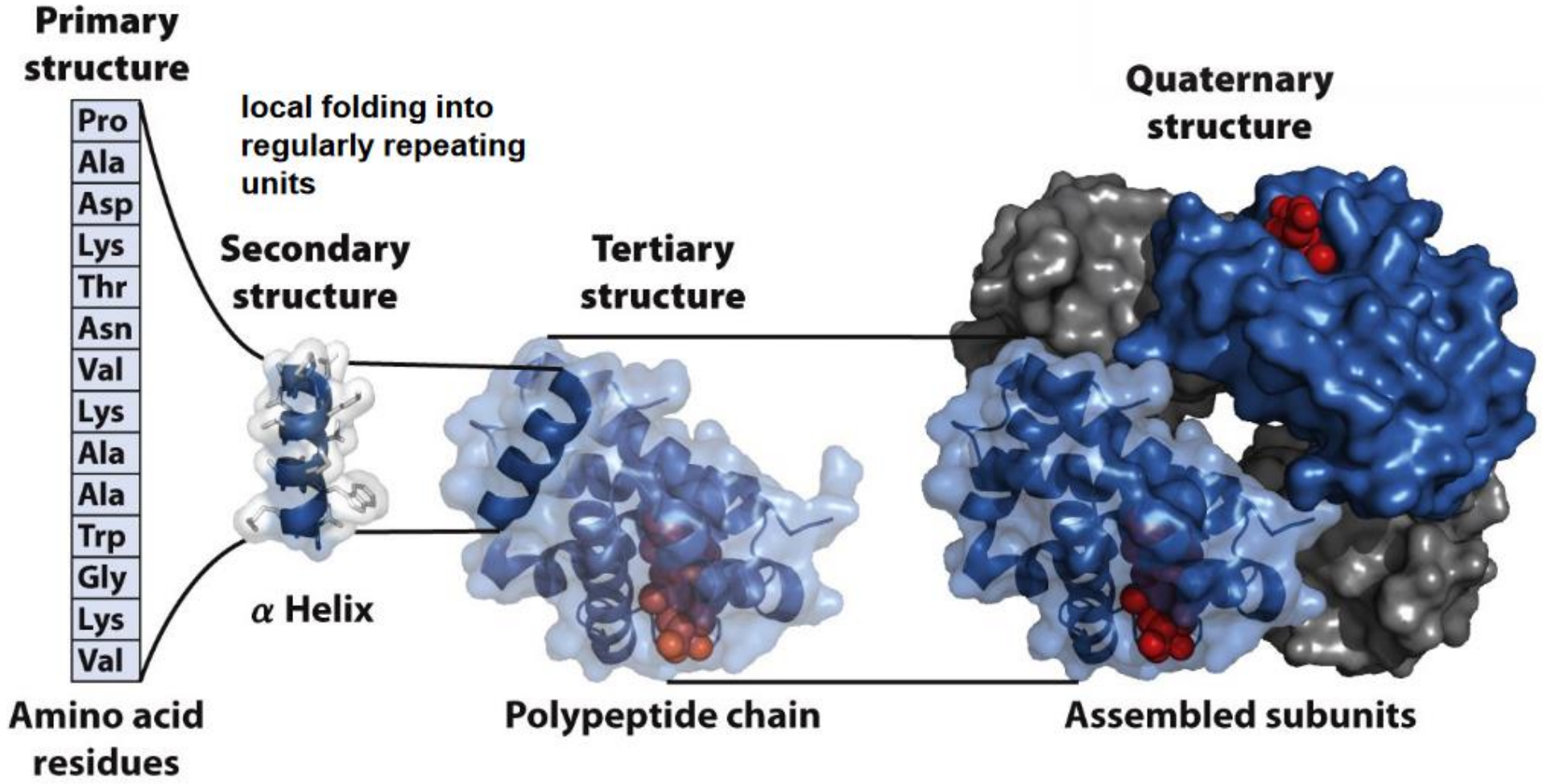


Figure 1.31 The Molecules of Life (© Garland Science 2013)

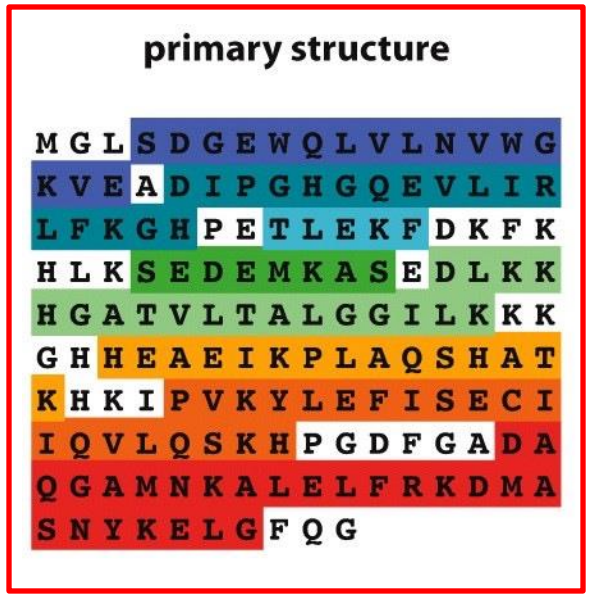
Hierarchical organization of protein structure

- From polypeptide chain sequence to full protein assembly into functional (multi)molecular complexes



- Biological role of a protein is determined by its structure which is encoded through the sequence

Primary structure



- **Sequence**: the pattern of amino acids along a linear polypeptide chain
- Always starts with **Methionine** encoded by the **AUG** (ATG in DNA) start codon
- Can be used for bioinformatic analyses, calculations of biophysical properties, molecular weight etc.

Secondary structure

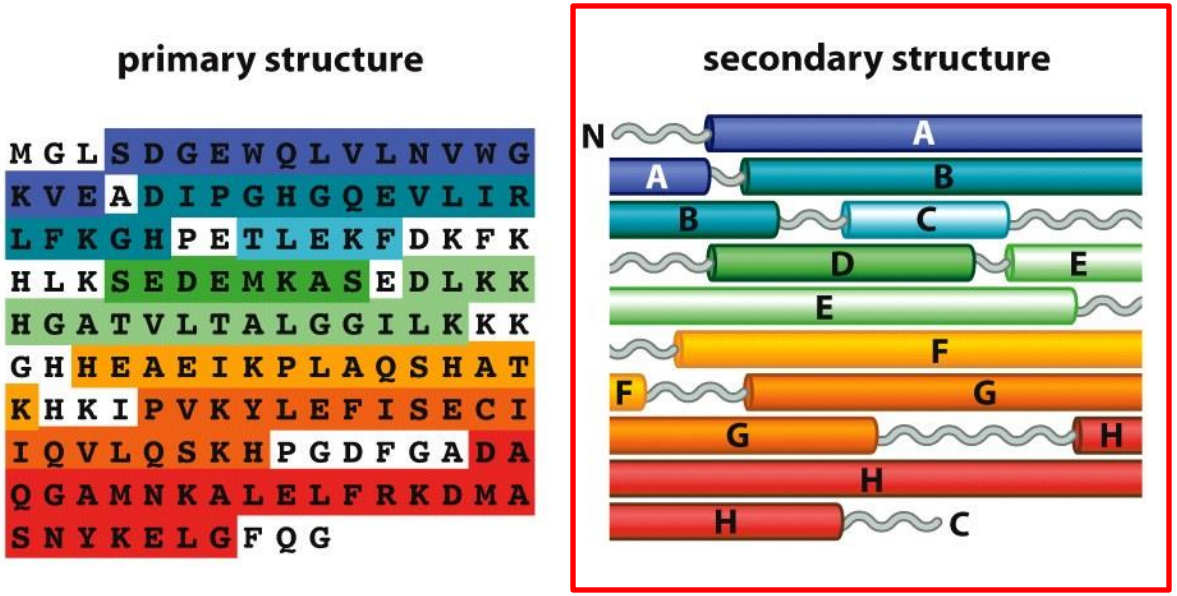


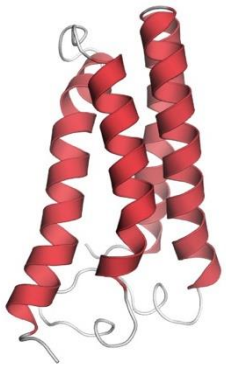
Figure 4.1 The Molecules of Life (© Garland Science 2013)

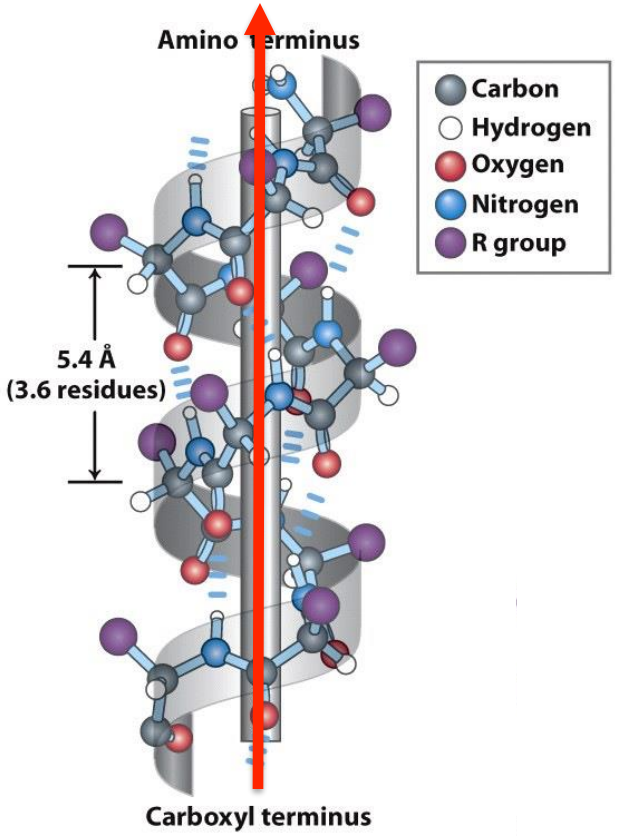
• The local conformation of the polypeptide chain (**α-helices, β-strands and loops**) is the **secondary structure**.

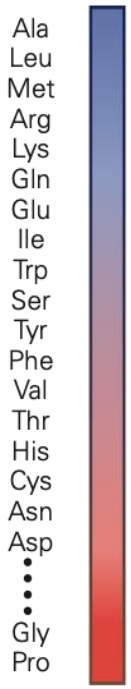
Secondary structure elements: α-helix

• Segments of amino acids adopt **α-helices** these structural motifs are the architectural elements of protein structure

α-helices







→ Gly and Pro are not good at assembling into α-helices

• In the α-helix, the carbonyl oxygen of residue “i” forms a hydrogen bond with the amide of residue “i+4”.

• Although each hydrogen bond is relatively weak in isolation, the sum of the hydrogen bonds in a helix makes it quite stable.

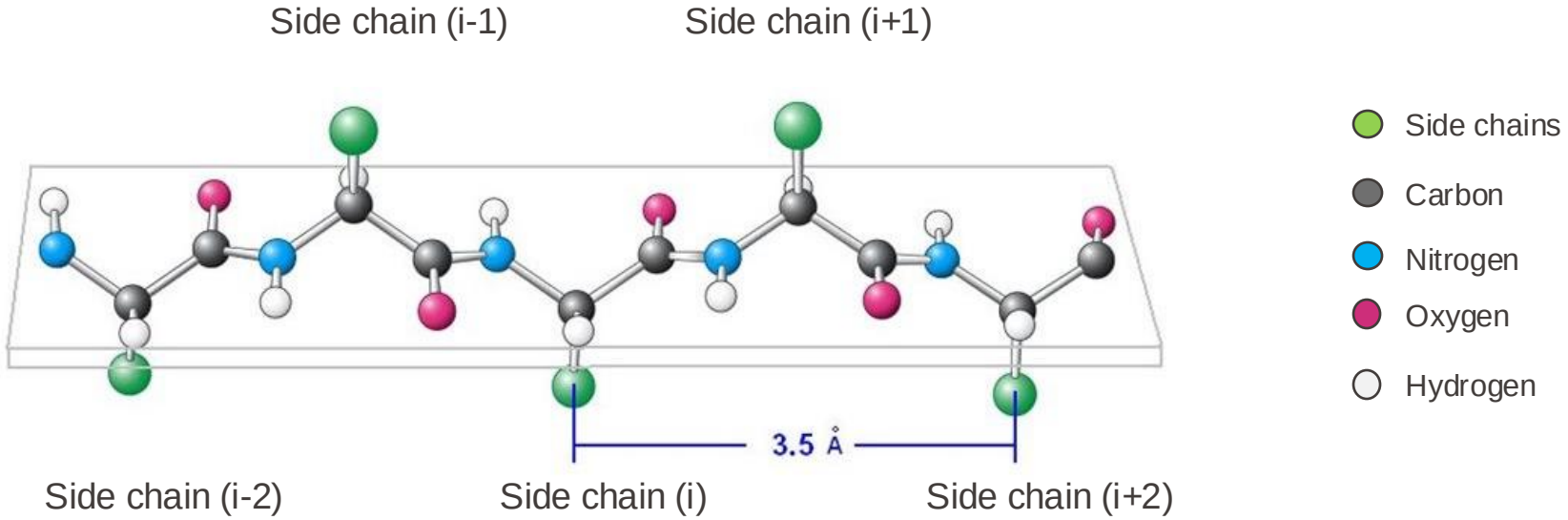
• A helical dipole is thus developed

• Amino-acid side chains face towards the outside

• The propensity of a peptide for forming an α-helix also depends on its sequence (see scale on the left).

Secondary structure elements: β-strands and sheets

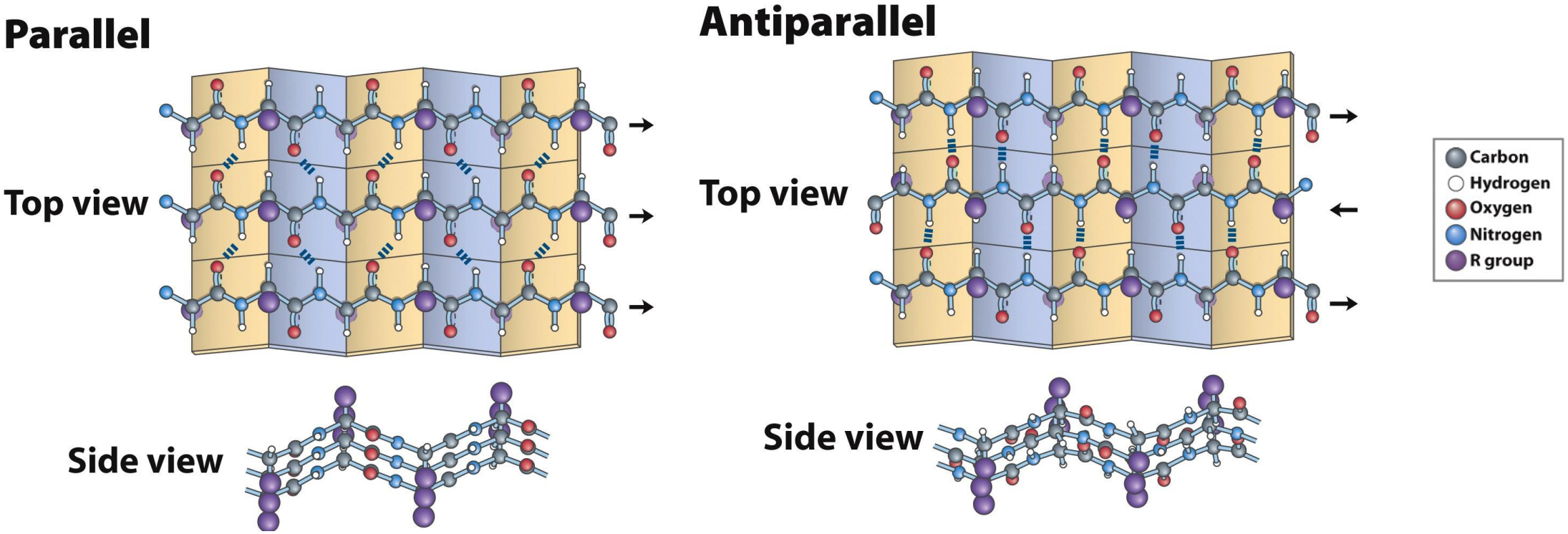
β-strands



- In a β-strand amino acids are arranged so that the side chains of neighboring residues face in the opposite directions
- Typically, 3-10 amino acids long and can be linear or partially twisted
- The propensity of a peptide for forming β-strands also depends on its sequence.

Secondary structure elements: β-strands and sheets

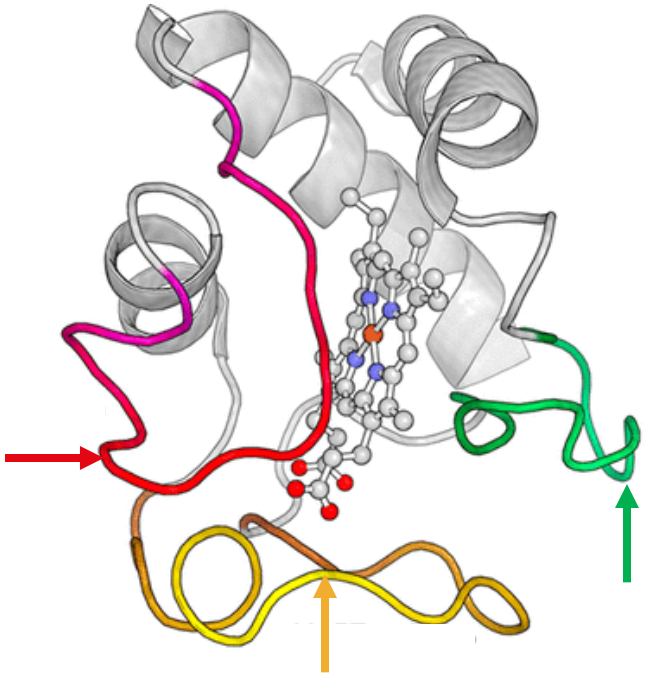
β-sheets



- β-sheets consist of multiple β-strands oriented in a **parallel** or **antiparallel** structure
- In a β-sheet, carbonyl oxygens and amides form hydrogen bonds **between** the strands, i.e. between aa far away from each other in the primary sequence.
- Individual strands do not have to be directly connected in a sequence (i.e., they can be distal or even come from different proteins altogether)

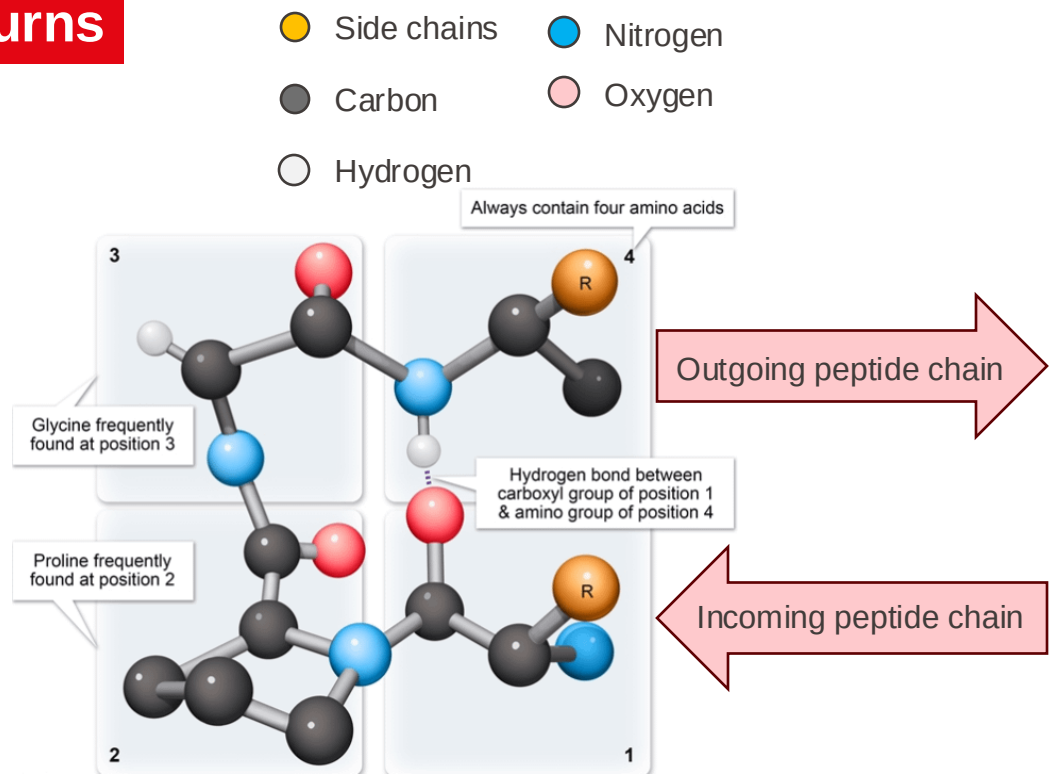
Secondary structure elements: Loops and Turns

Loops



- Loops follow the backbone geometry rules but do not have any regular assembly
- Often used to connect different helix- and strand-based domains
- Sometimes called “random coils”

β-turns

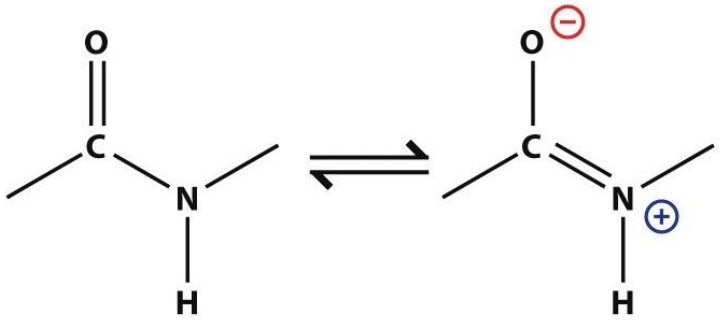


- Turns are there to change direction to connect two other elements of secondary structure.
- Key residues are Pro and Gly
- Hydrogen bond between the C=O of residue 1 and the N-H of residue 4

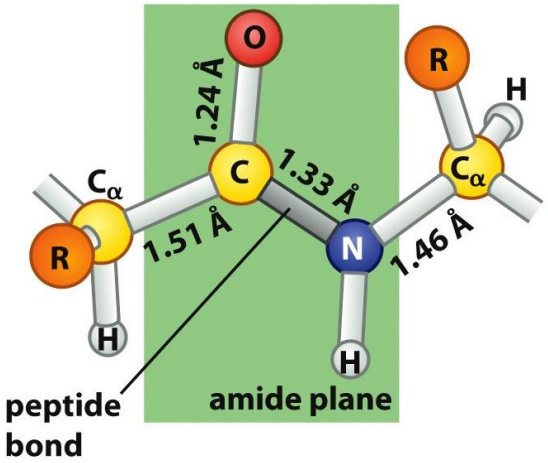
Backbone torsions and secondary structure

- Protein backbone torsions are not random. Instead, they are driven by chemical properties of different groups

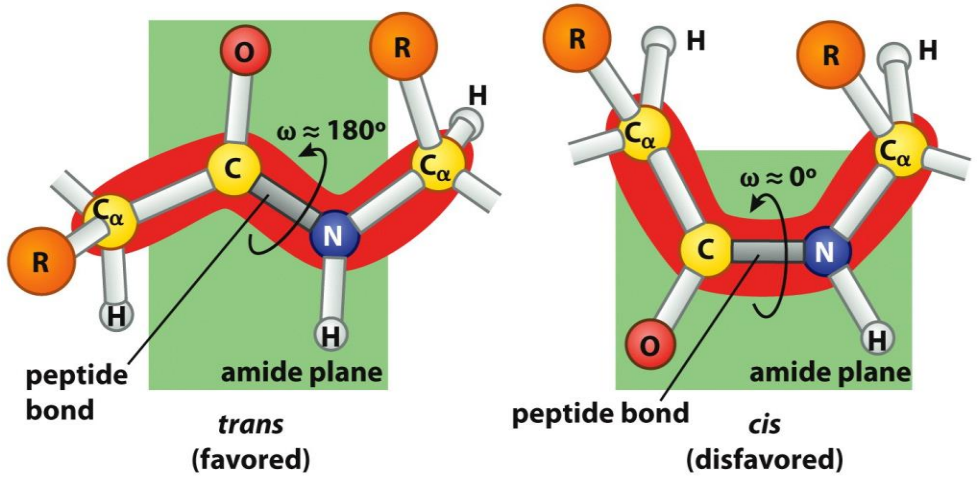
Peptide bond has a partial double bond character



This makes it planar (i.e., HN-CO atoms in one plane)



Peptide bond can exist as *trans* or *cis*



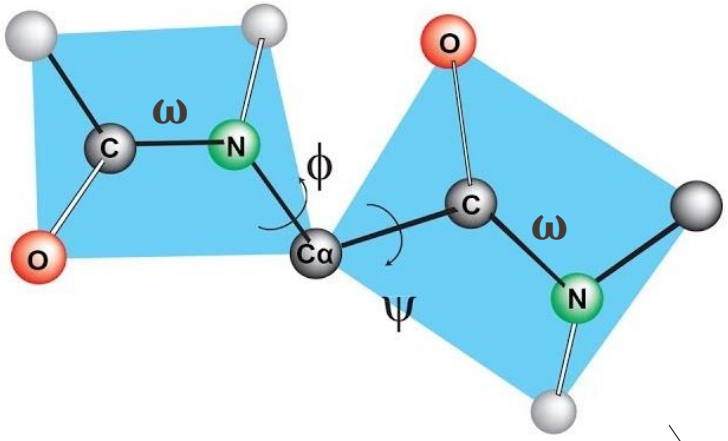
Omega (ω) is the bond angle of the peptide bond

ω=180˚ for *trans*

ω=0˚ for *cis*

Backbone torsions and secondary structure

- While there is not much allowed rotation around the peptide bond due to partial double-bond character, the surrounding bonds can accommodate rotation
- They are called dihedral angles (ϕ and ψ) and **specific angular rotations around the corresponding bonds (i.e., N-C_a, C_a-C) within the polypeptide chain defines the secondary structure**



(ϕ1, ψ1)

α-helix

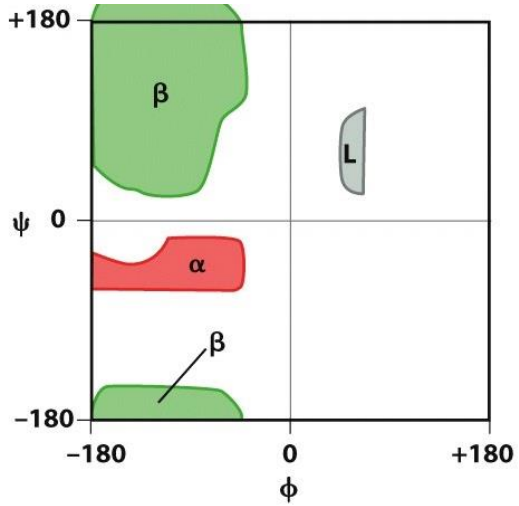
(ϕ2, ψ2)

β-strands

(ϕ3, ψ3)

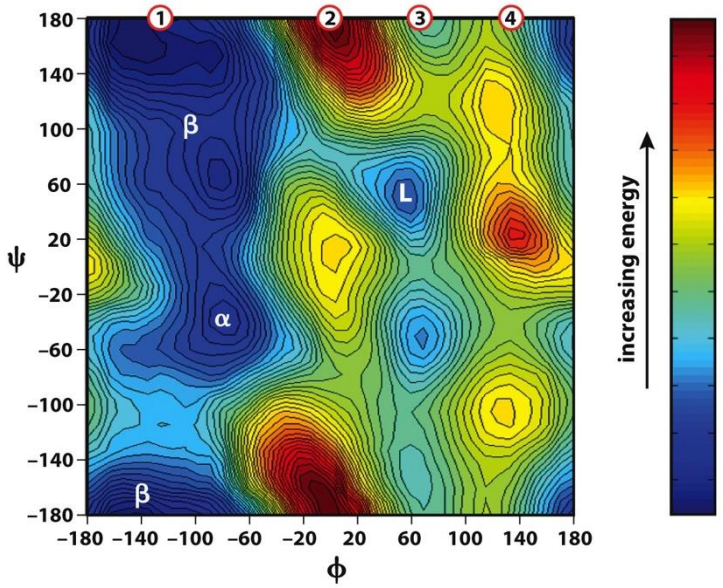
Loops

Ramachandran Plot



Backbone torsions and secondary structure

- The backbone torsion angles **ϕ** (phi) and **ψ**(psi) determine that conformation of the protein chain
- The **Ramachandran diagram** defines the restrictions on backbone conformation. It is a map of energetically favored and unfavored regions



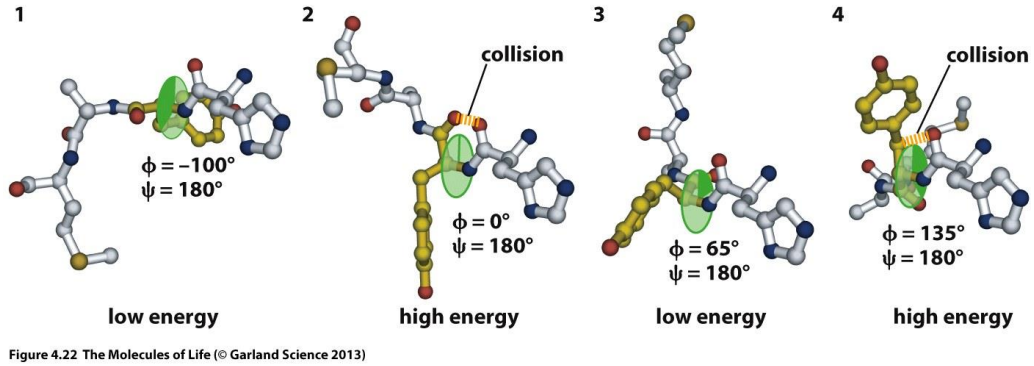


Figure 4.22 The Molecules of Life (© Garland Science 2013)

- The values will be different for every amino acid due to unique steric restraints

G.N. Ramachandran

(1922-2001)

Backbone torsions and secondary structure

Why is it that the Ramachandran plot seems so different for glycine?

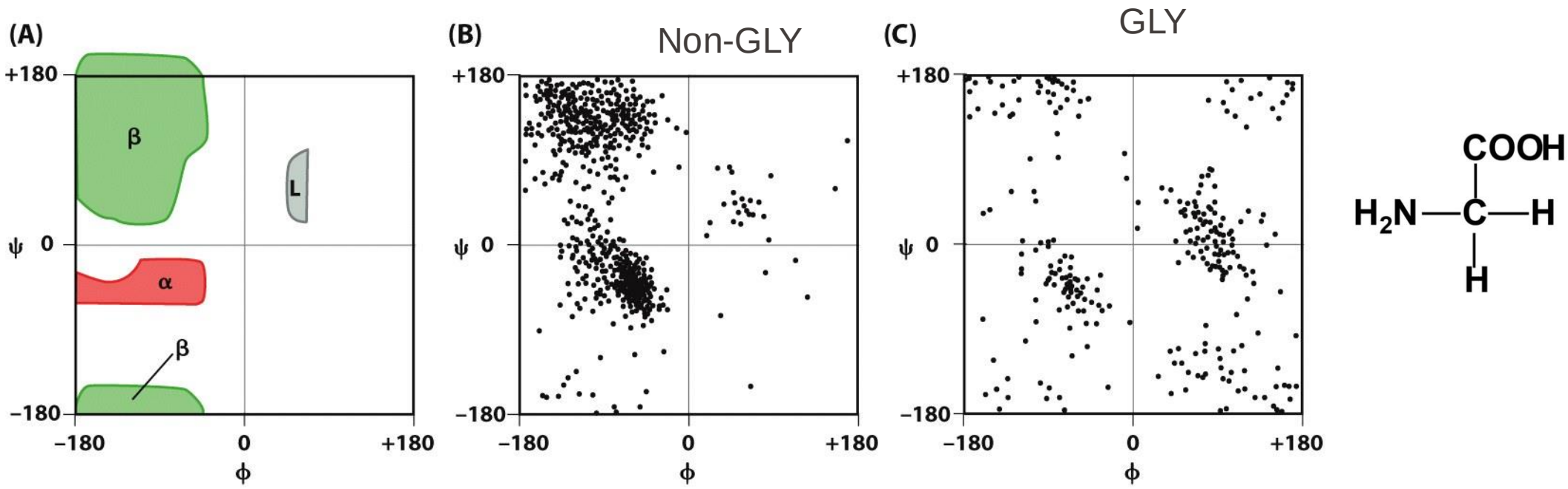


Figure 4.20 The Molecules of Life (© Garland Science 2013)

Secondary structure elements combine to create motifs

- Motifs are assemblies of two or more secondary structure elements. They usually have a functional role

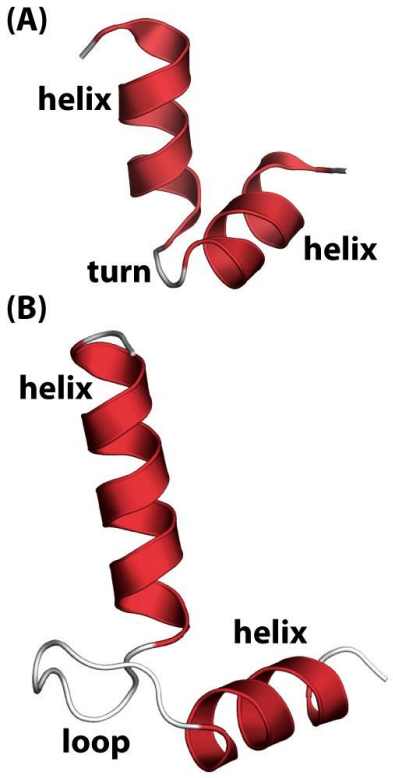


Figure 4.33 The Molecules of Life (© Garland Science 2013)

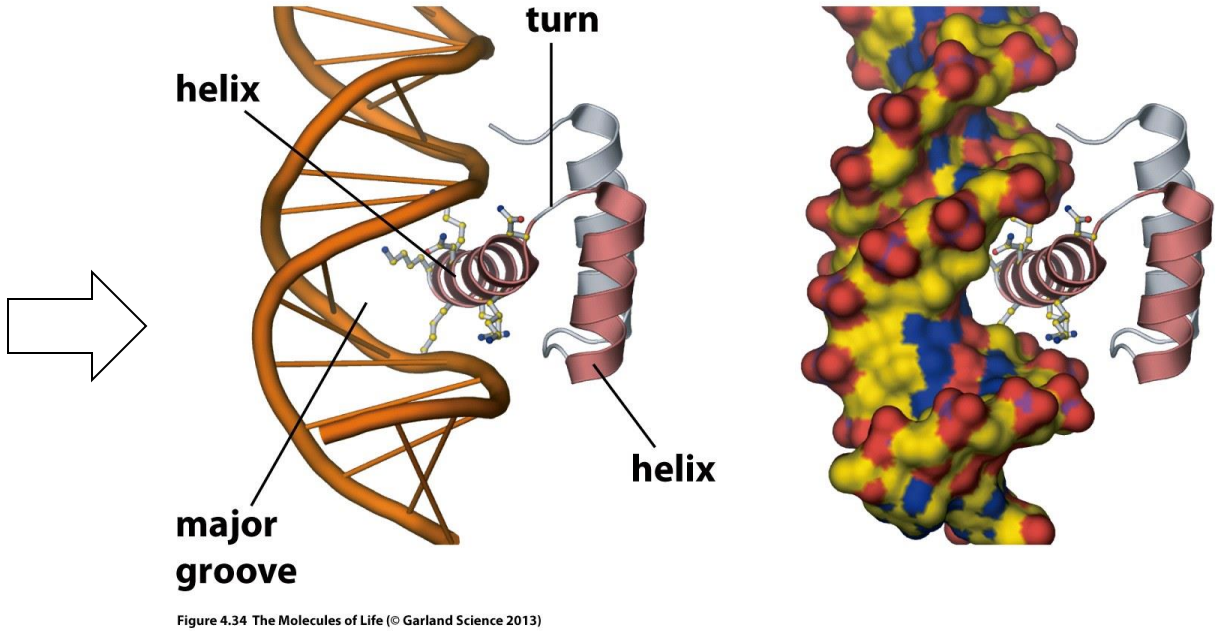


Figure 4.34 The Molecules of Life (© Garland Science 2013)

Helix-turn-helix motif

DNA-binding role

Secondary structure elements combine to create motifs

- Many diverse motifs (note how different secondary structural elements are depicted)

Greek Key

Figure 4.36 The Molecules of Life (© Garland Science 2013)

β-α-β motifs

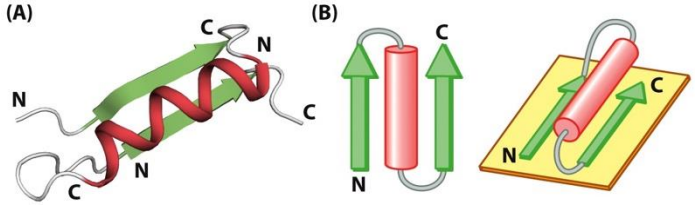


Figure 4.37 The Molecules of Life (© Garland Science 2013)

Helix-helix

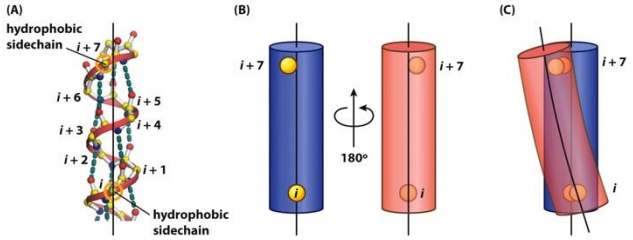


Figure 4.39 The Molecules of Life (© Garland Science 2013)

4 helices

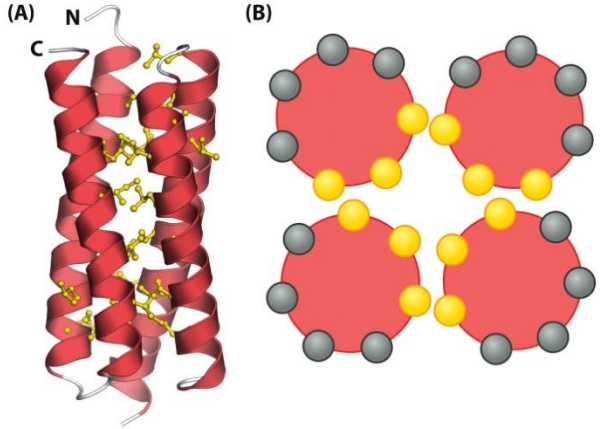


Figure 4.44 The Molecules of Life (© Garland Science 2013)

α/β domains

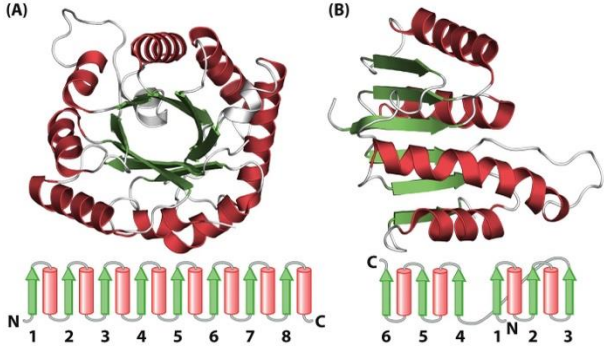


Figure 4.49 The Molecules of Life (© Garland Science 2013)

Coiled-coil

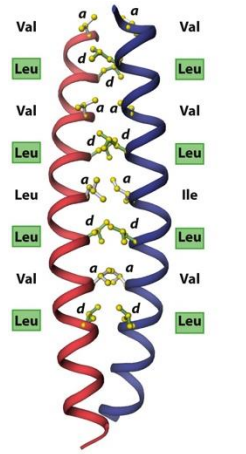


Figure 4.41 The Molecules of Life (© Garland Science 2013)

"Leucine zipper"

Tertiary structure

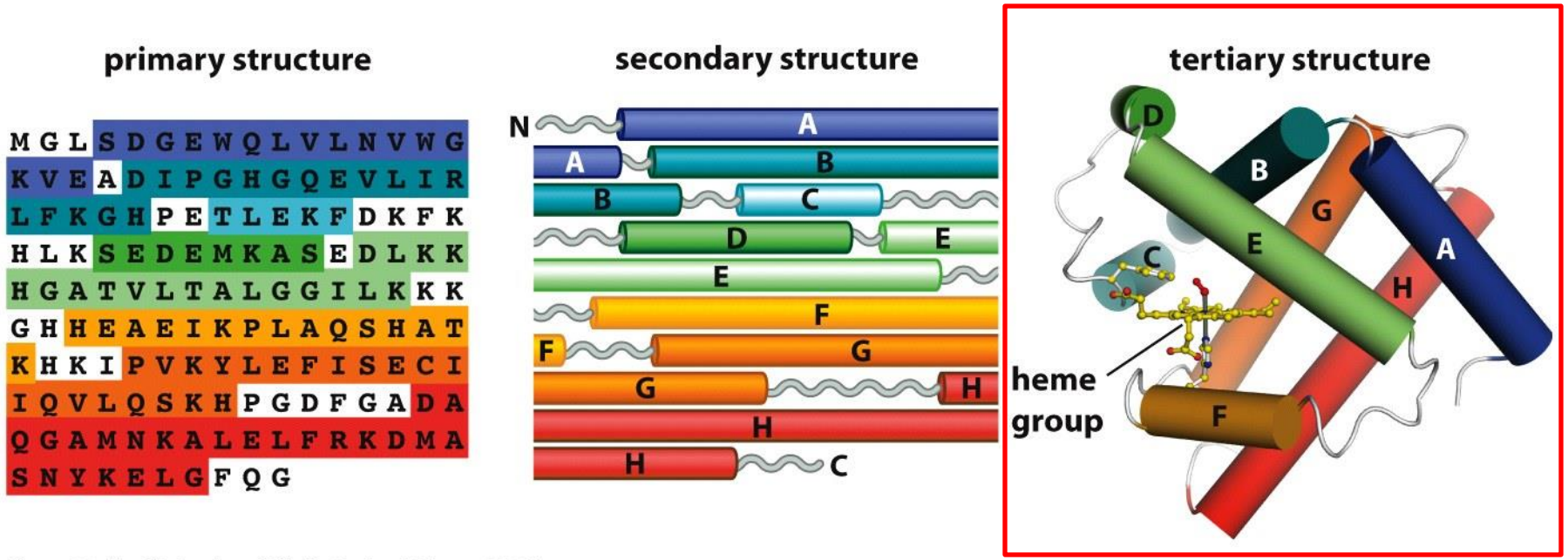


Figure 4.1 The Molecules of Life (© Garland Science 2013)

• Tertiary structure refers to the three-dimensional organization of the secondary structure elements from a single polypeptide chain into domains.

• Sometimes referred to as **protein fold**.

Using the secondary structure building blocks nature has built many folds

- They serve as structural support, biochemical catalysts, hormones, enzymes, building blocks etc.
- Some are more abundant than others

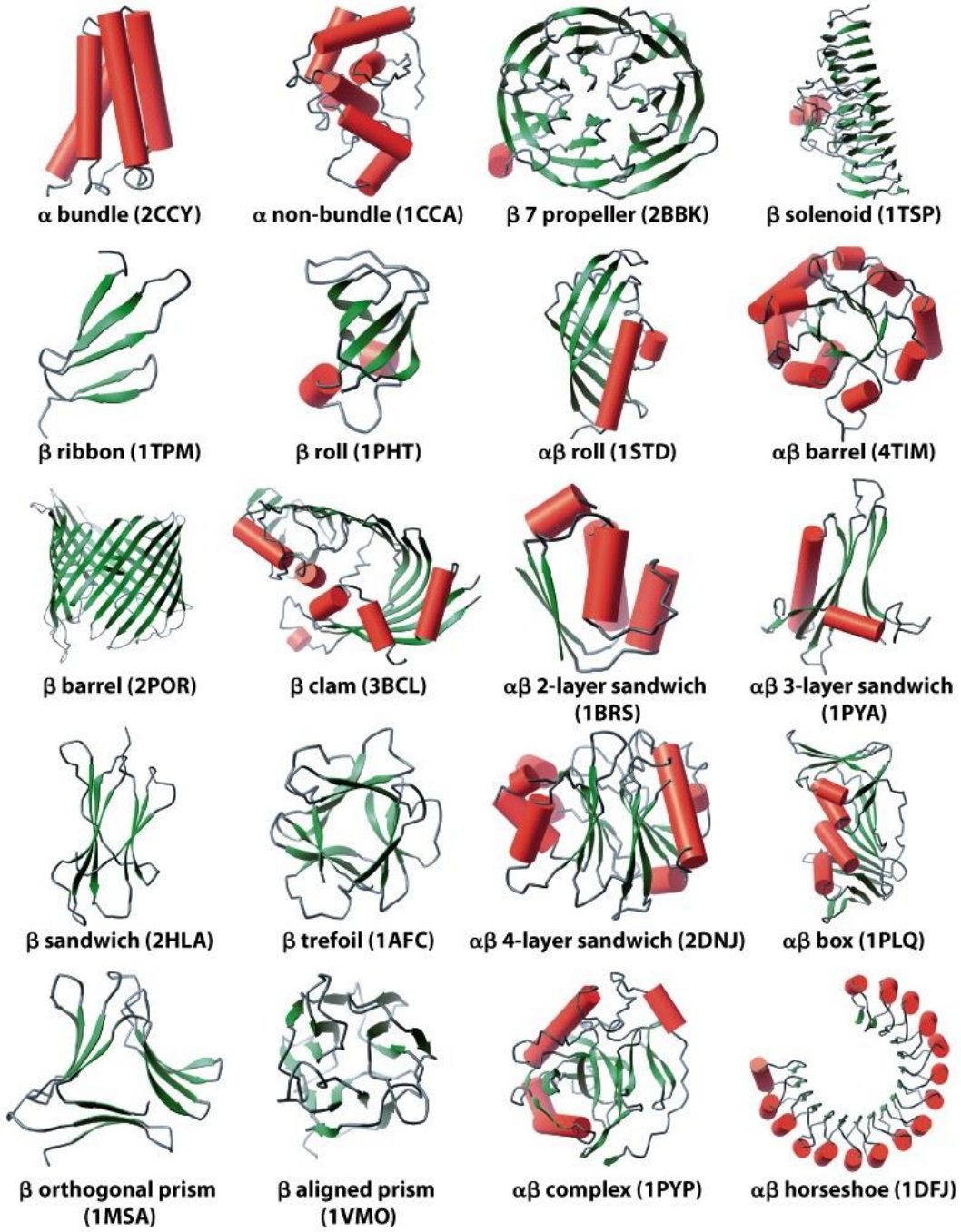


Figure 5.32 The Molecules of Life (© Garland Science 2013)

Quaternary structure

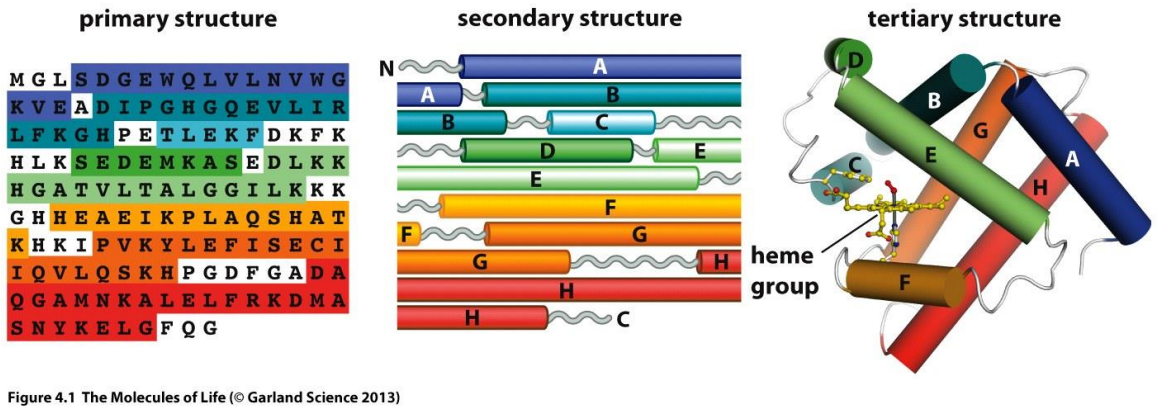
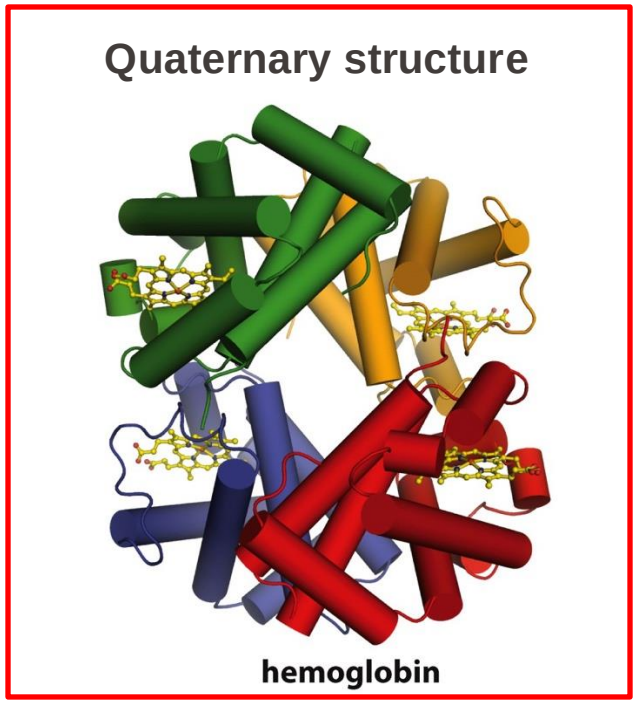


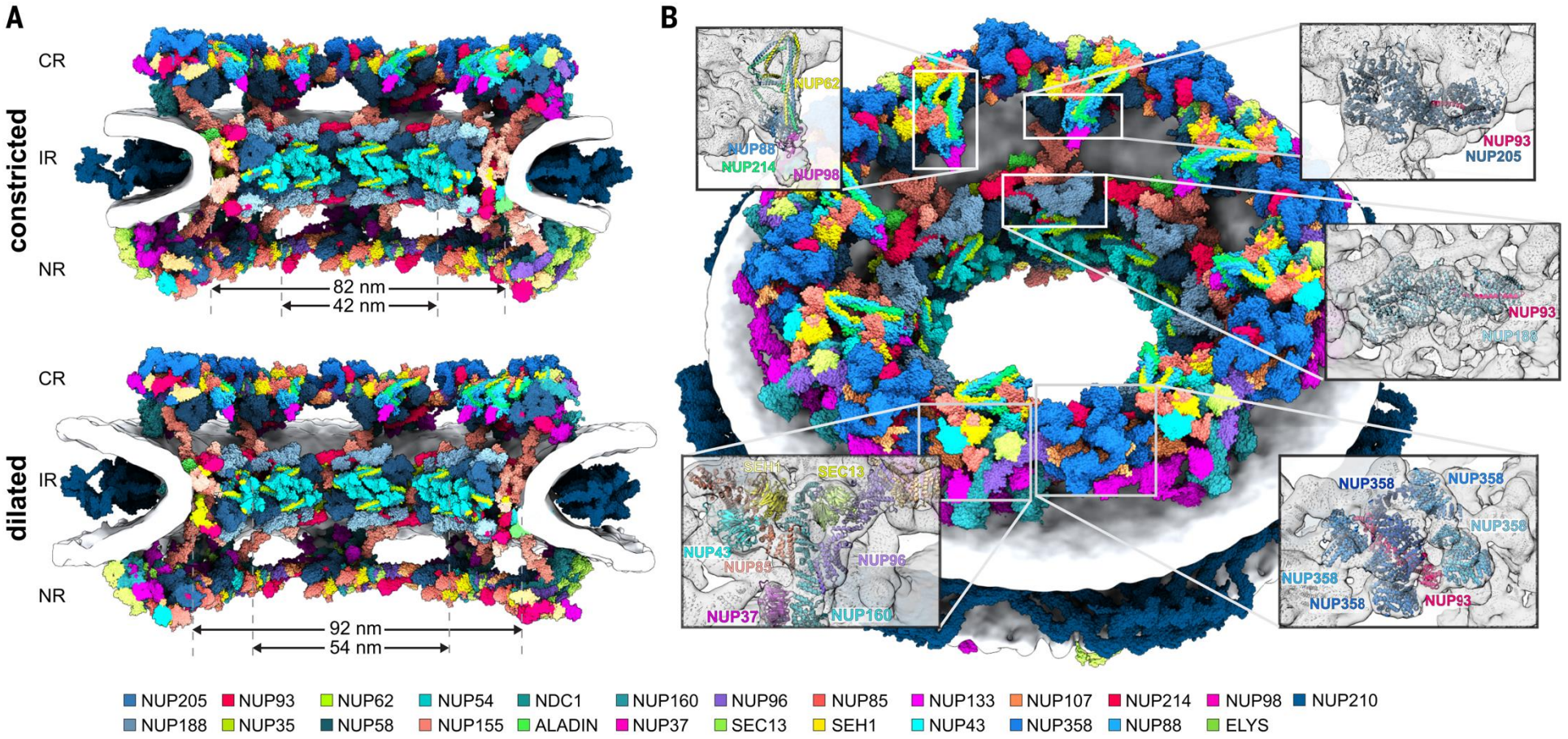
Figure 4.1 The Molecules of Life (© Garland Science 2013)



• Quaternary structure refers to the association of different polypeptide chains (subunits) into a multimeric complex with some molecular function.

• For example, hemoglobin is composed of 4 subunits of the same protein (+ heme group). Quaternary structure can also refer to complexes with DNA, RNA, lipids or any other biomolecule.

The nuclear pore puzzle



~456 different proteins

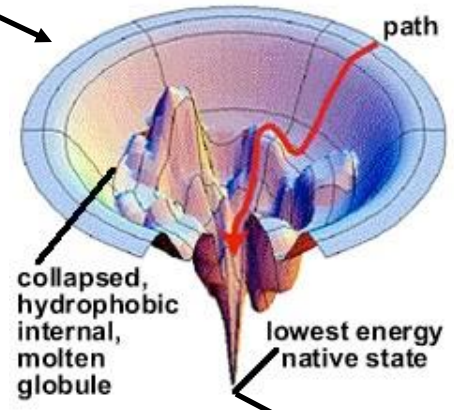
How do proteins fold into functional conformation

```
MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS
TPDDHLVARWCATPVAEVAEKSARF
PPRPEELLLPREETLPEMYSLSFTA
MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS
```

Primary sequence

Anfinsen dogma (1954):

Protein structure is determined by its sequence



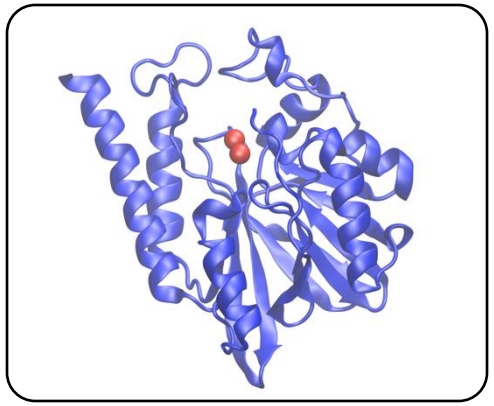
Dill@UCSF

Folding pathway

Levinthal paradox (1969)

100 residue-long peptide

~10^{94} torsional degrees of freedom



Folded native structure

? Why is this a paradox ?

Because all these conformations are not sampled, and the native state is quickly found within milliseconds thanks to thermodynamics

Proteins - Take home message

• Proteins are formed from amino acids connected by peptide bonds

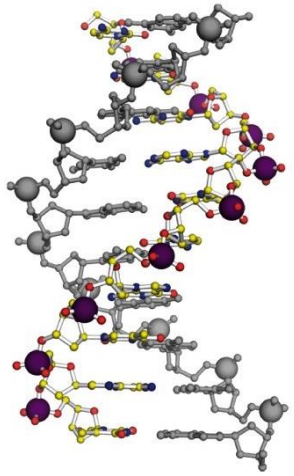
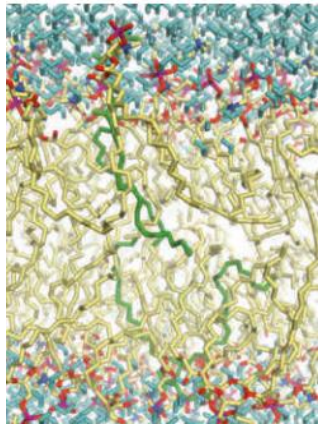
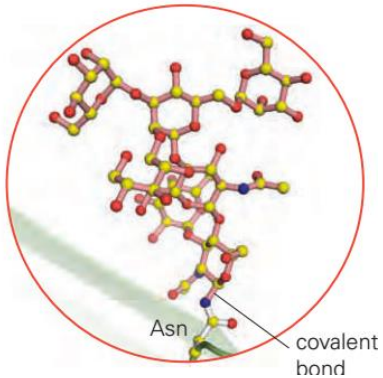
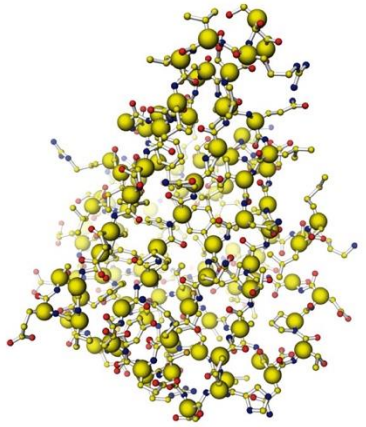
• The 20 genetically encoded amino acids are classed as either hydrophilic (charged, or polar but neutral) or hydrophobic (nonpolar) depending on the properties of their sidechains

• Proteins fold into specific three-dimensional structures that are highly diverse

• **α-helices** and **β-sheets** are the regular structural elements on which the structures of proteins are based

• Protein folds are assembled from a single polypeptide chain, and they can co-assemble with other biomolecules to form many diverse complexes with diverse molecular and biological functions

The molecules of Life

	Nucleic Acids	Lipids	Carbohydrates	Proteins
Macromolecular Structure				
Building Block