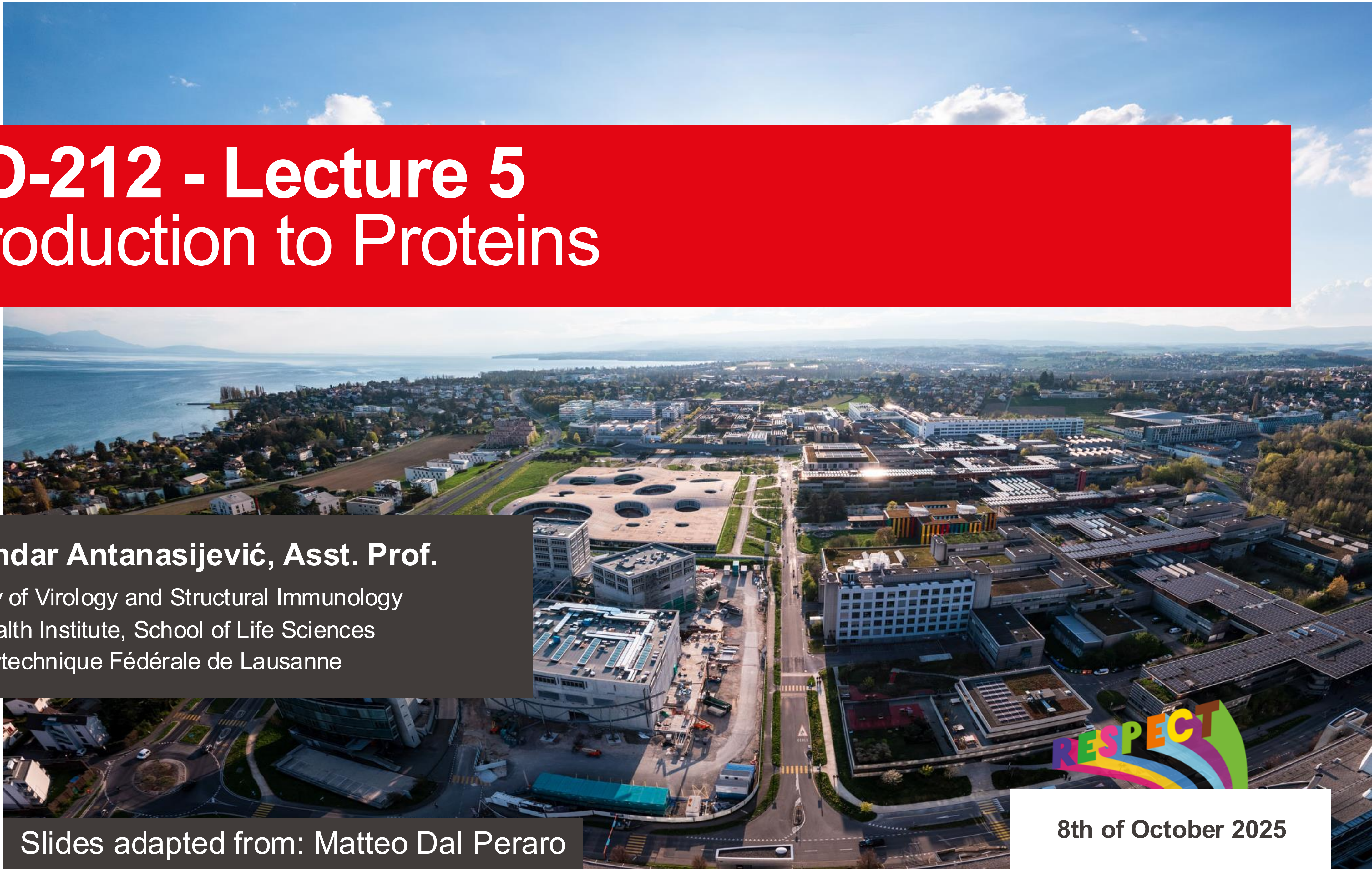


BIO-212 - Lecture 5
Introduction to Proteins

Aleksandar Antanasijević, Asst. Prof.
Laboratory of Virology and Structural Immunology
Global Health Institute, School of Life Sciences
École Polytechnique Fédérale de Lausanne

Slides adapted from: Matteo Dal Peraro

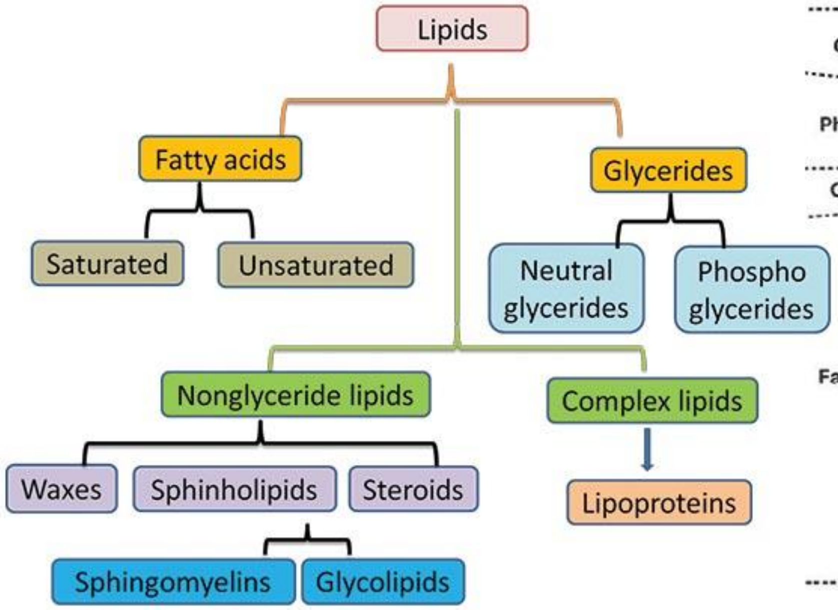
8th of October 2025

École
polytechnique
fédérale
de Lausanne

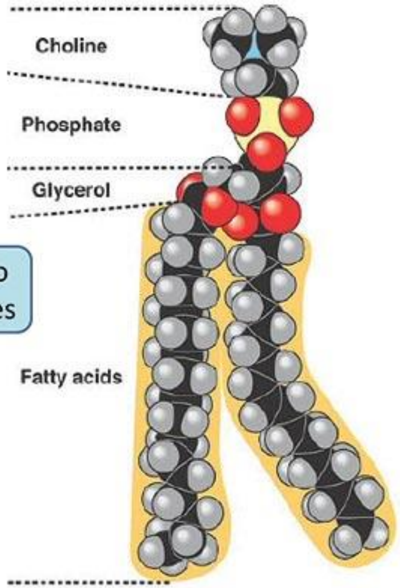
Lecture 4 – Quick Summary

• Main lipid types and roles

- Main lipid types based on chemical properties



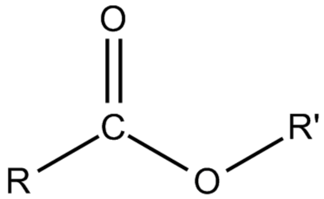
- Lipids
  - Fatty acids
    - Saturated
    - Unsaturated
  - Glycerides
    - Neutral glycerides
    - Phospho glycerides
  - Nonglyceride lipids
    - Waxes
    - Sphinholipids
      - Sphingomyelins
      - Glycolipids
    - Steroids
  - Complex lipids
    - Lipoproteins



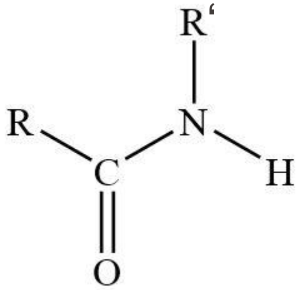
- There are >1000 building blocks which can assemble in different ways (very diverse)
- Their main roles include energy storage, assembly of biological membranes, cell and hormone signaling

• Important bonds and interactions

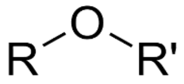
Ester bonds (triglycerides)

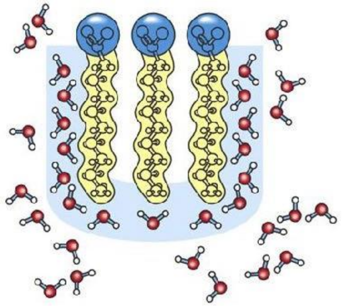


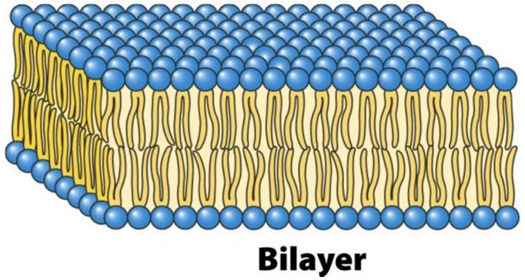
Amide bonds (sphingolipids)



Ether bonds (head domains)







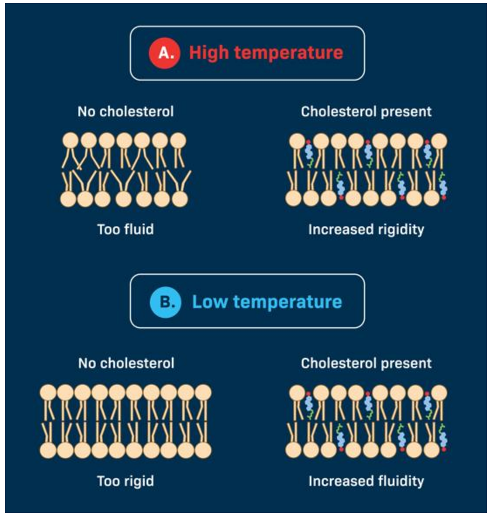
- Amphipatic molecules (polar and hydrophobic) that form bilayers in aqueous solutions

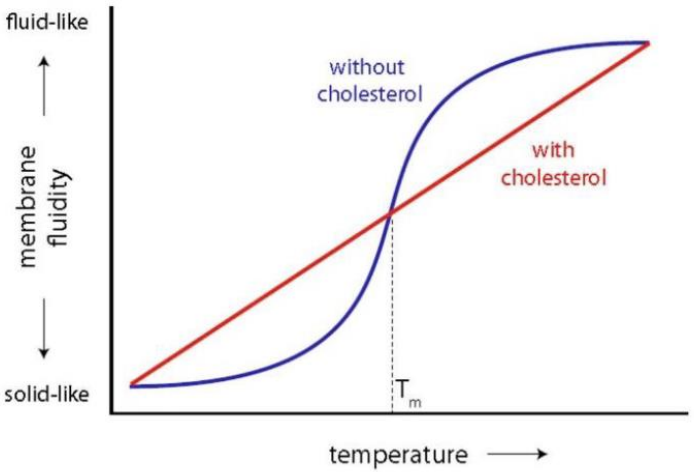
- Van der Waals interactions in the tail
- Hydrogen bonding and charged interactions using the head domain

Lecture 4 – Extra Clarifications

• Cholesterol in lipid membranes

- Cholesterol acts as a fluidity buffer in lipid membranes, increasing fluidity at low temperatures and decreasing it at high temperatures by interfering with phospholipid interactions

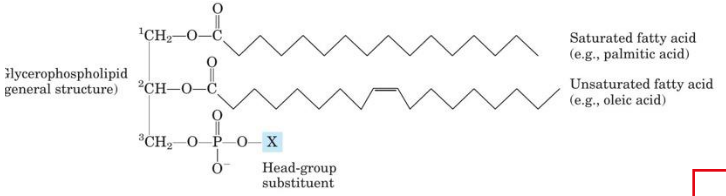


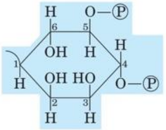


- This effect is concentration dependent, and very high cholesterol content (above ~40-50%) may make membranes too rigid and impermeable.
- Normal cholesterol content in plasma membranes is ~30-50 molar %

• Net charge of different lipids and membranes

- Glycerophospholipids are the major constituents of membranes and due to the presence of a phosphate group they will be negatively charged (-1) unless the head group is also charged.



Name of glycerophospholipid	Name of X	Formula of X	Net charge (at pH 7)
Phosphatidic acid	—	$-H$	−1
Phosphatidylethanolamine	Ethanolamine	$-CH_2-CH_2-\overset{+}{N}H_3$	0
Phosphatidylcholine	Choline	$-CH_2-CH_2-\overset{+}{N}(CH_3)_3$	0
Phosphatidylserine	Serine	$-CH_2-\underset{COO^-}{CH}-\overset{+}{N}H_3$	−1
Phosphatidylglycerol	Glycerol	$-CH_2-\underset{OH}{CH}-CH_2-OH$	−1
Phosphatidylinositol 4,5-bisphosphate	*myo*-Inositol 4,5-bisphosphate		−4

- Depending on the composition the membrane charge will range from neutral to negative, which has biological consequences (e.g., interactions with proteins, ions etc.)

The molecules of Life

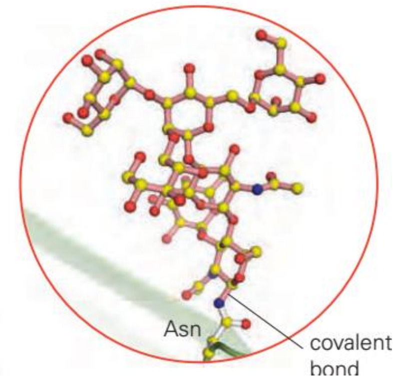
Carbohydrates

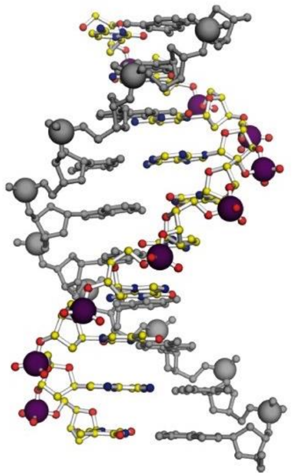
Nucleic Acids

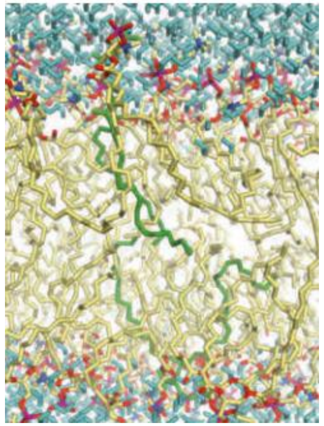
Lipids

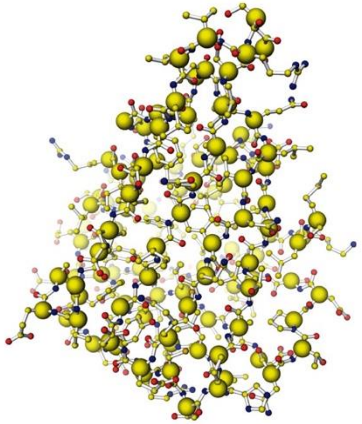
Proteins

Macromolecular Structure

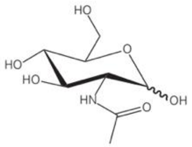




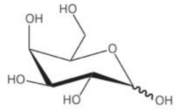




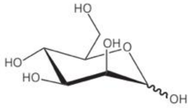
Building Block



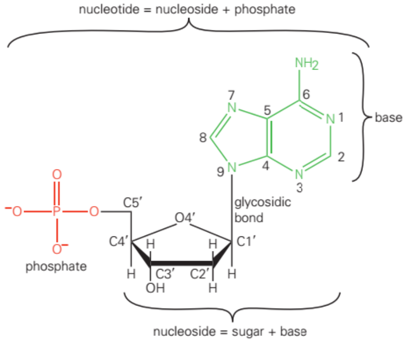
N-acetylglucosamine
(GlcNAc)

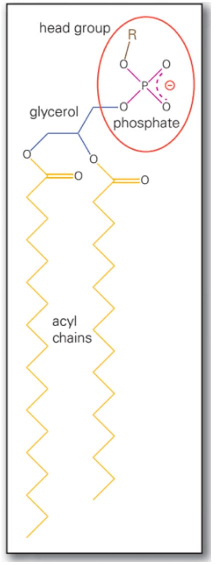


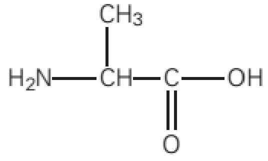
galactose
(Gal)

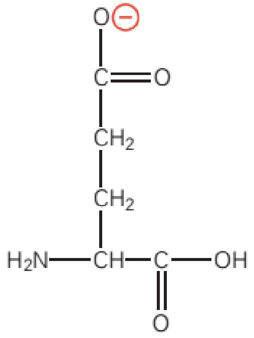


mannose
(Man)









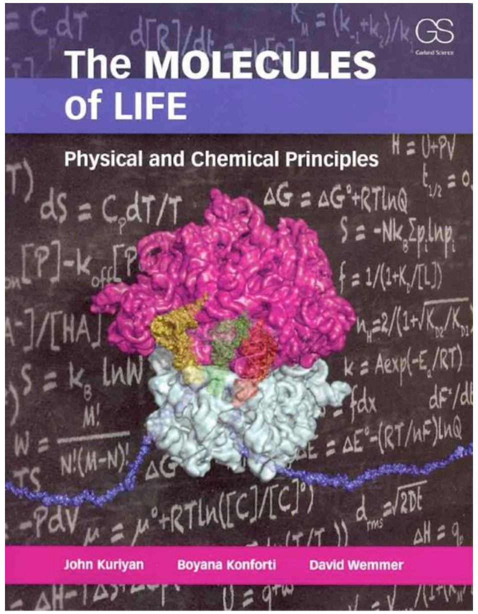
Lecture 5 - Outline

Today:

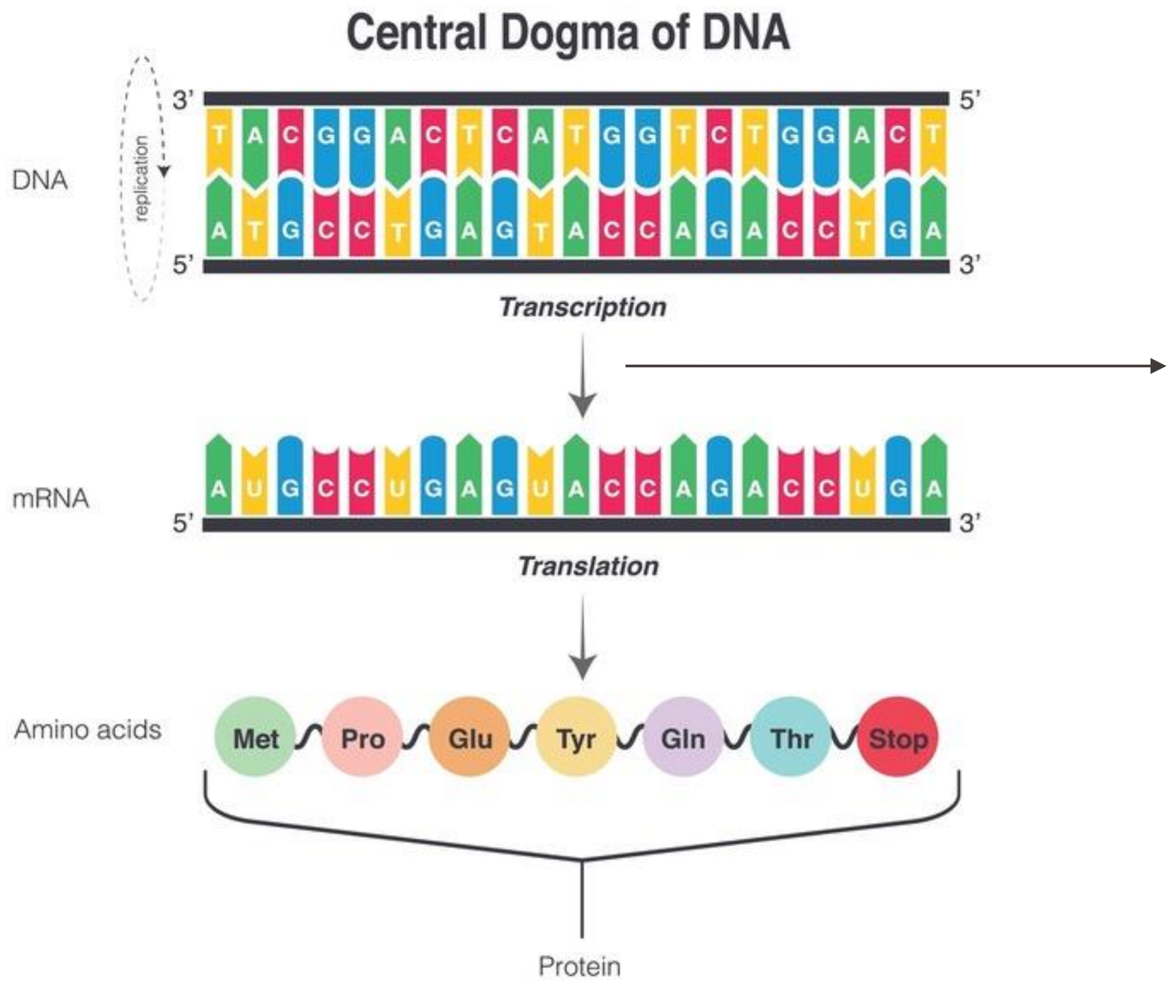
- Amino acid structures and categorization
- Peptide bond and polypeptide chains
- Protein structure and assembly

Reading suggestions:

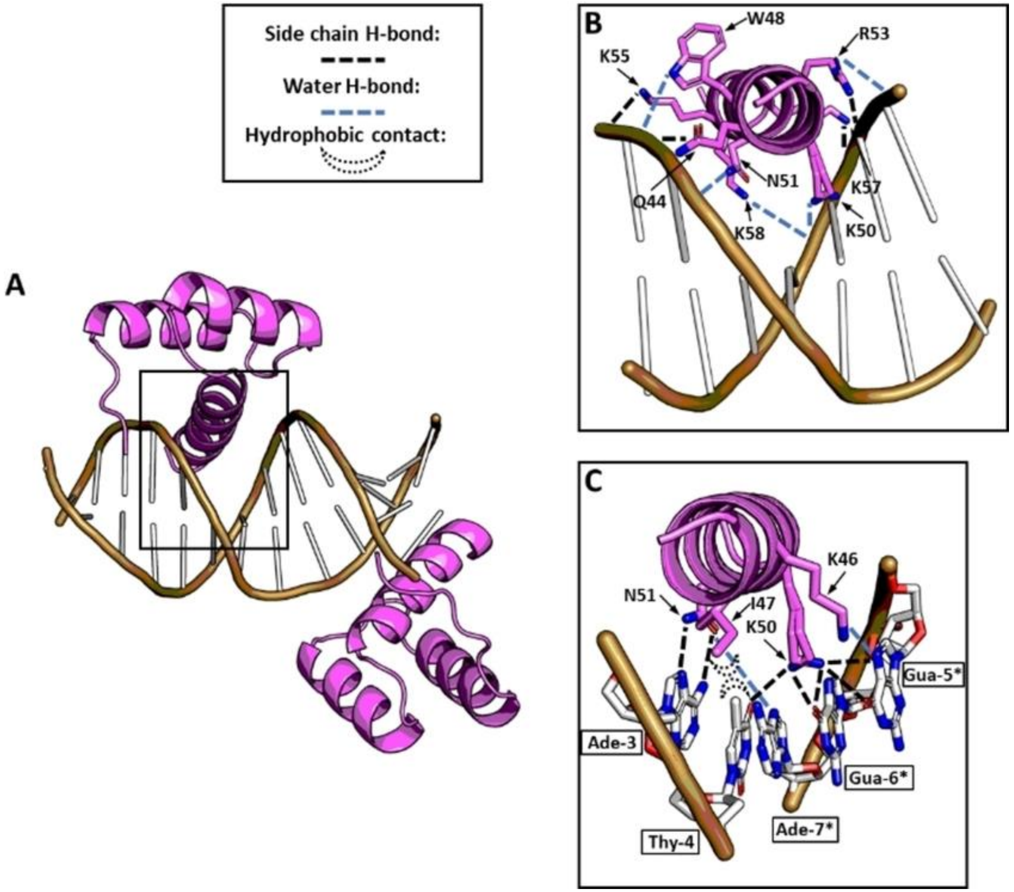
- The Molecules of Life (Chapters 1 and 4)



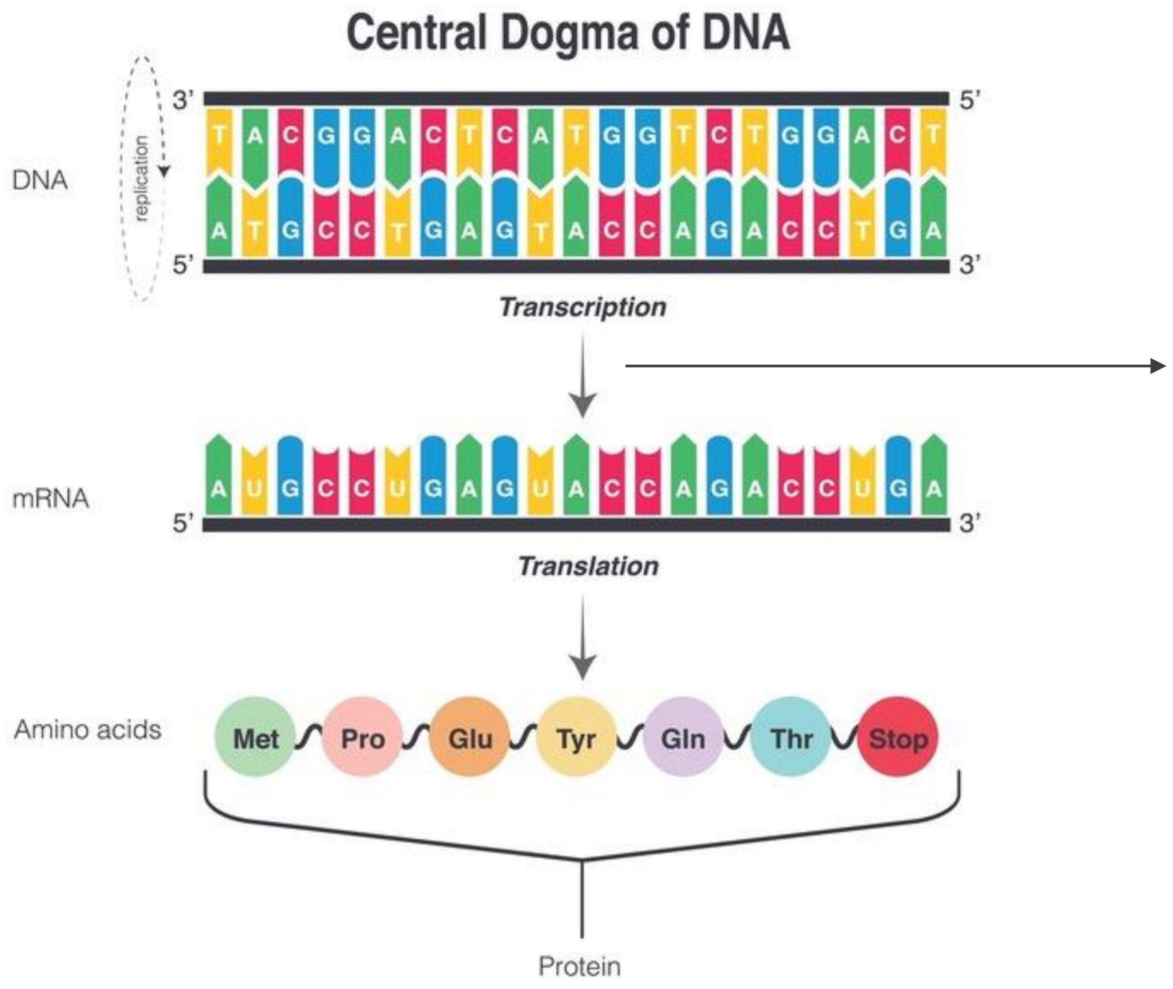
Genetic code is translated to protein sequence



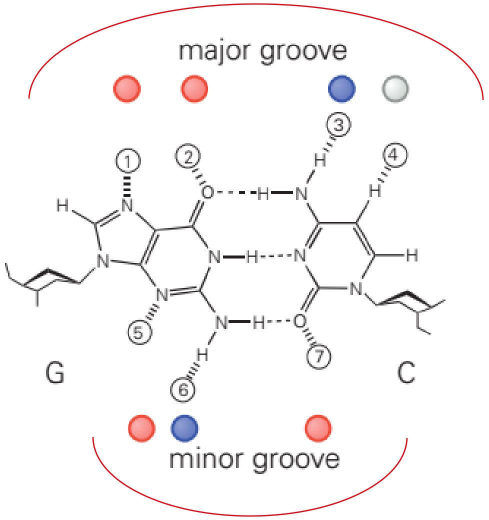
- **Transcription factors** recognize the gene locations through the major and minor groove of the DNA chain
- In the process it minimally disturbs the double helix structure

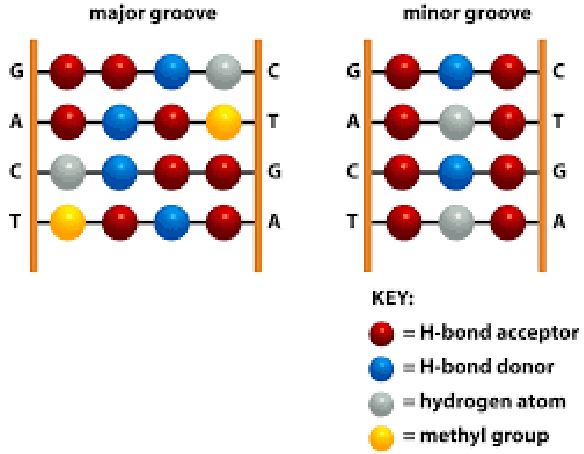


Genetic code is translated to protein sequence



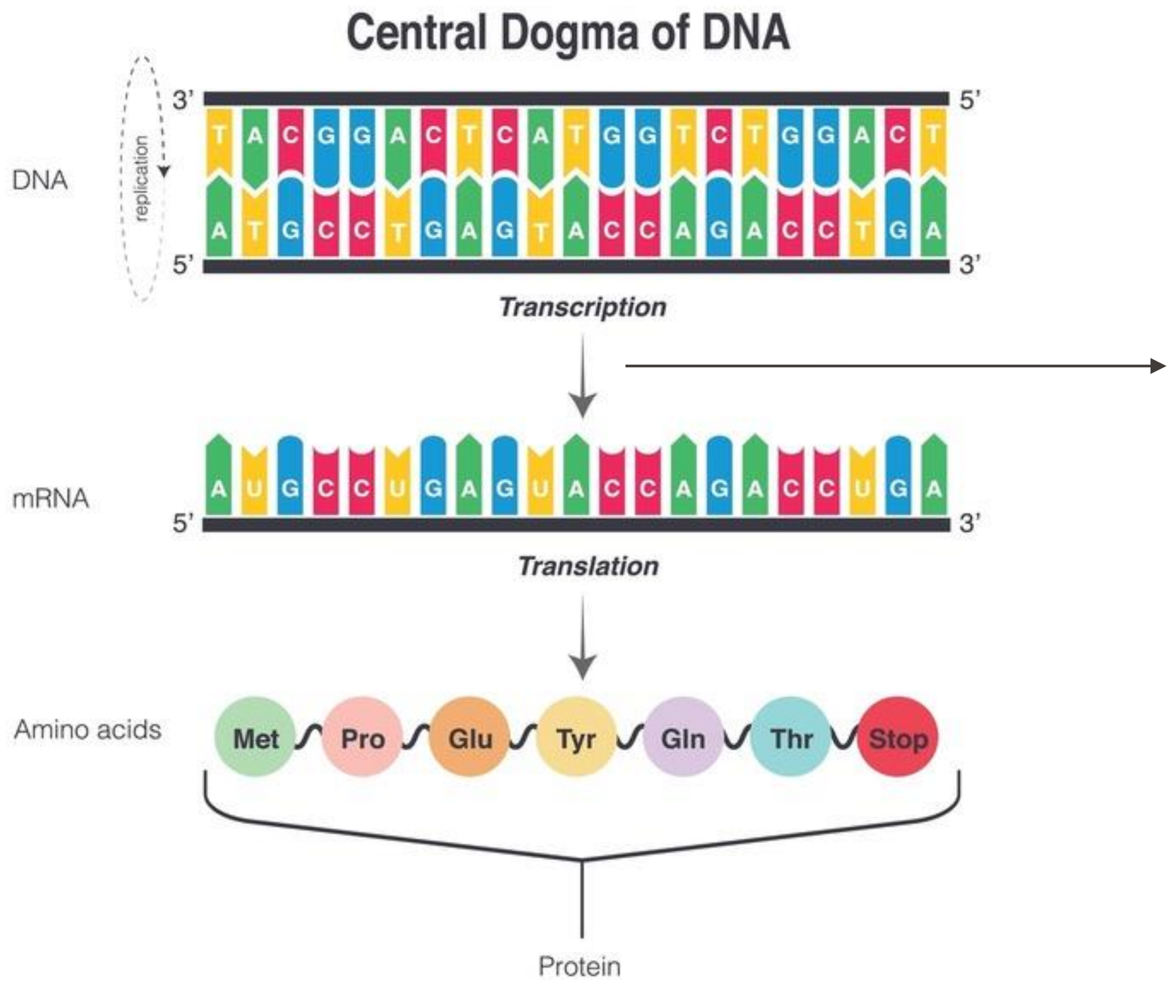
- **Transcription factors** recognize the gene locations through the major and minor groove of the DNA chain
- In the process it minimally disturbs the double helix structure



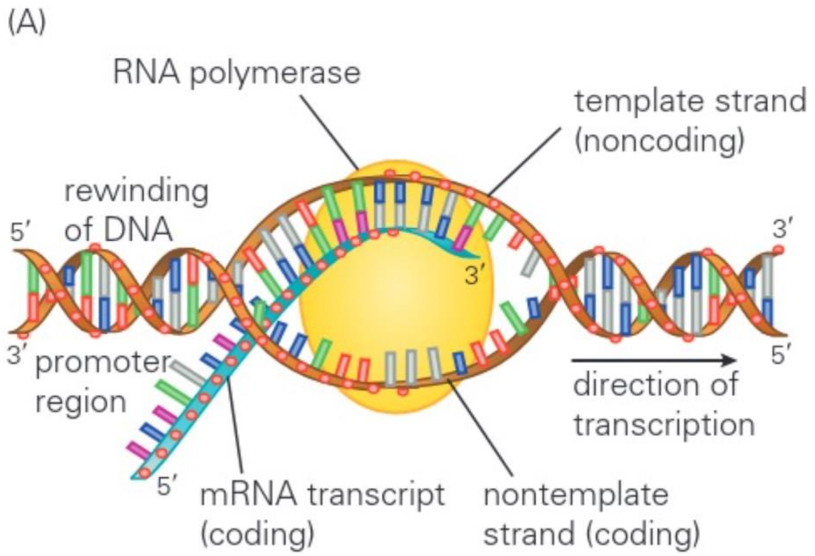


- The sensing is mediated by interactions with available hydrogen bond donor and acceptor group (=barcode)
- These interactions are very specific and lead to transcription of downstream gene sequence into RNA

Genetic code is translated to protein sequence

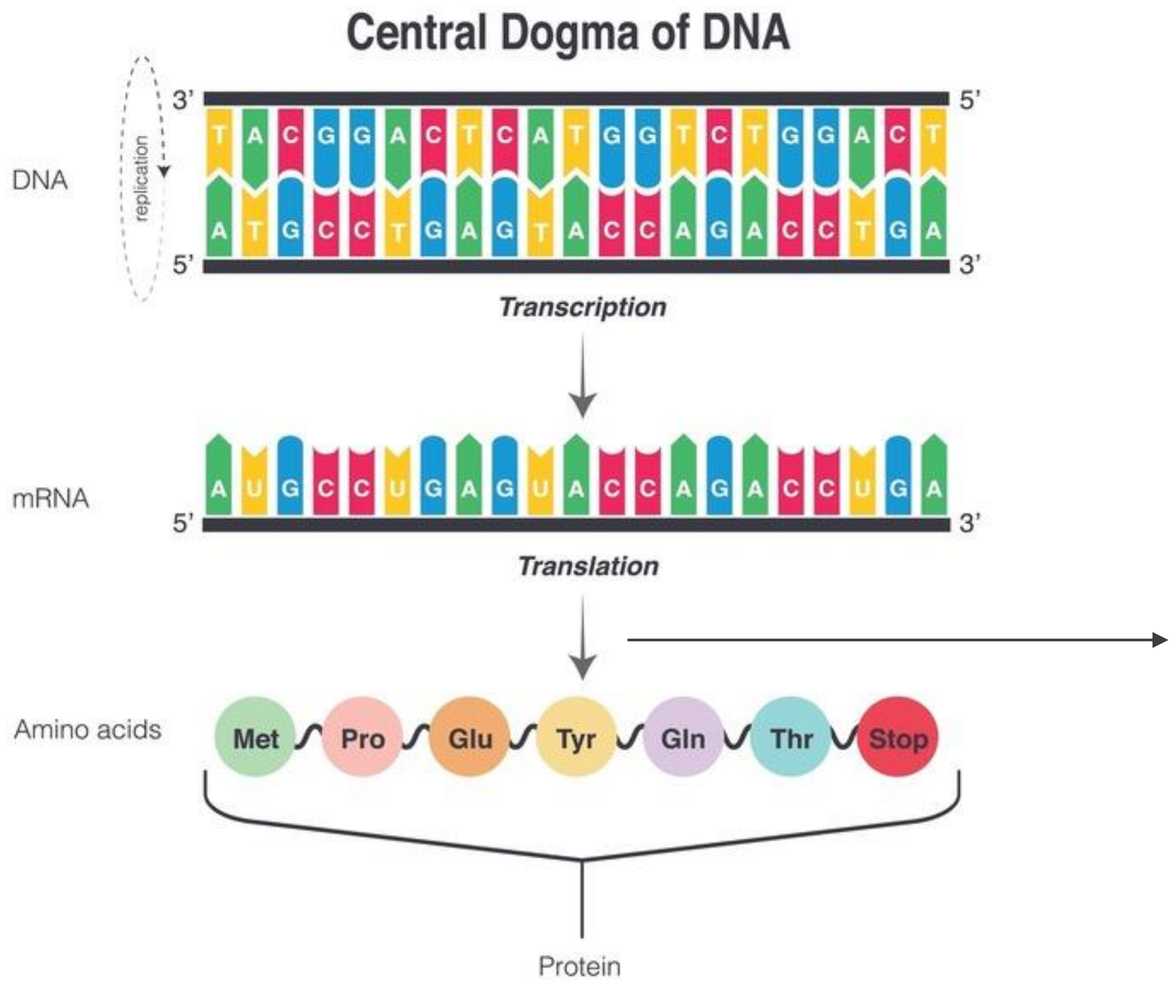


- RNA polymerase is recruited onto the site, binding to the **promoter** region of the DNA
- RNA polymerase creates a small (~17bp) unwound region on the DNA and initiates **transcription**.

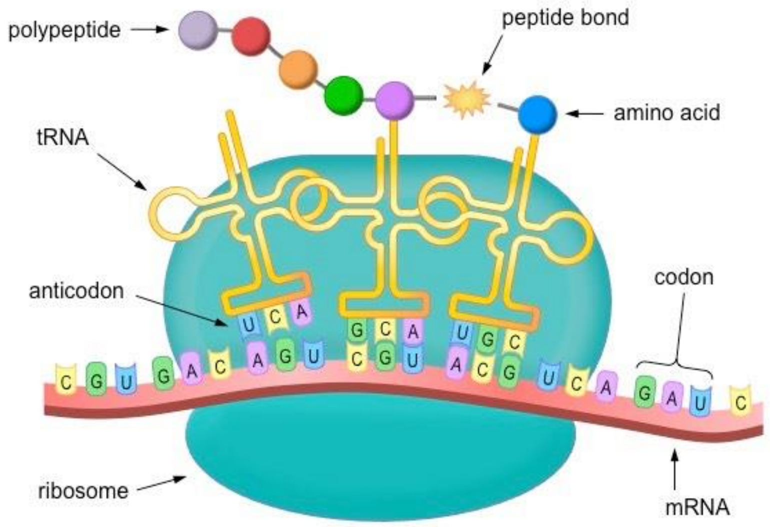


- RNA elongates in the 5' -> 3' direction following Watson-Crick base pairing until it reaches a polyadenylation signal
- The pre-mRNA generated through this process is then spliced to generate **messenger RNA (mRNA)**

Genetic code is translated to protein sequence

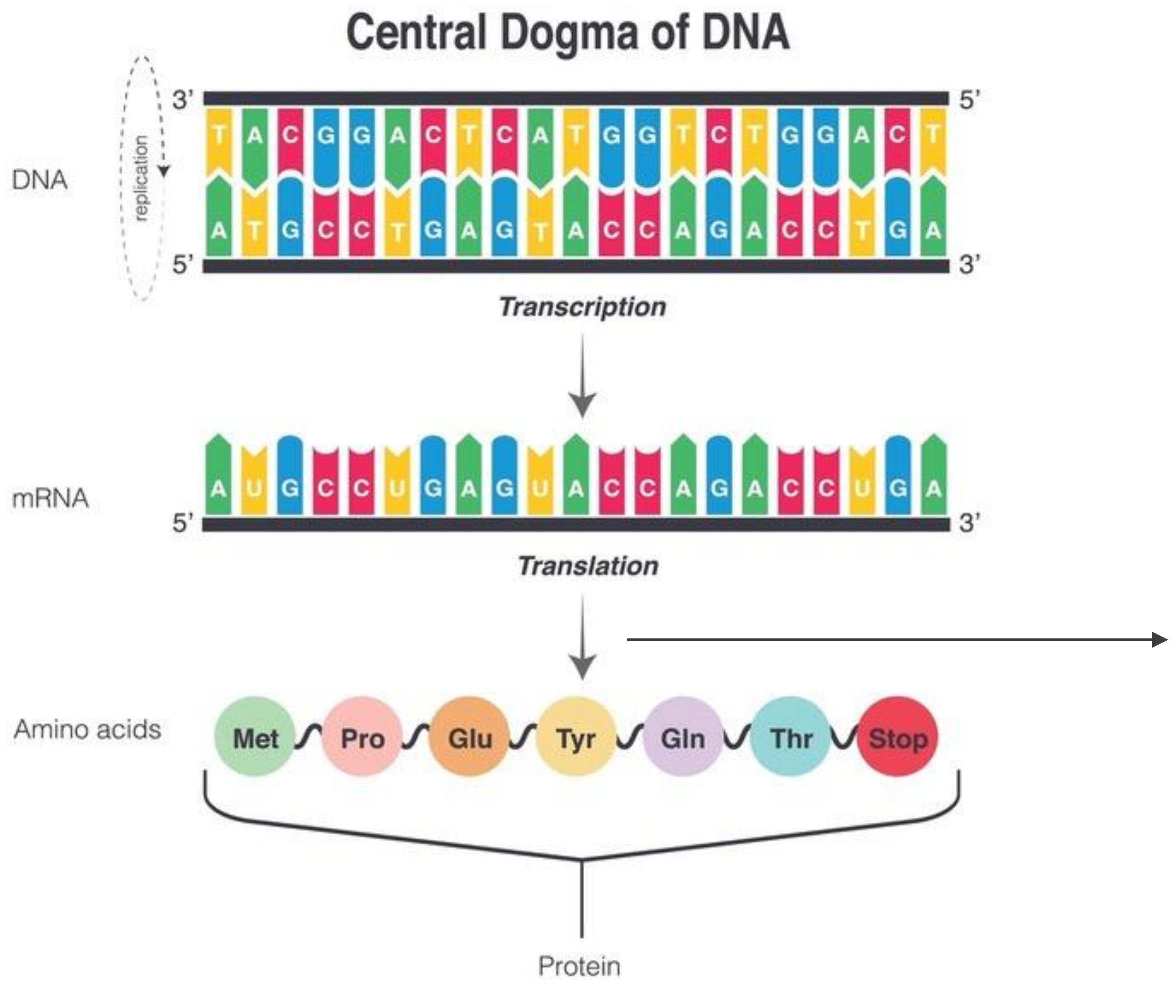


- mRNA moves to the **ribosome** where it is translated into protein sequence
- **3 nucleotides = 1 codon -> 1 amino acid**

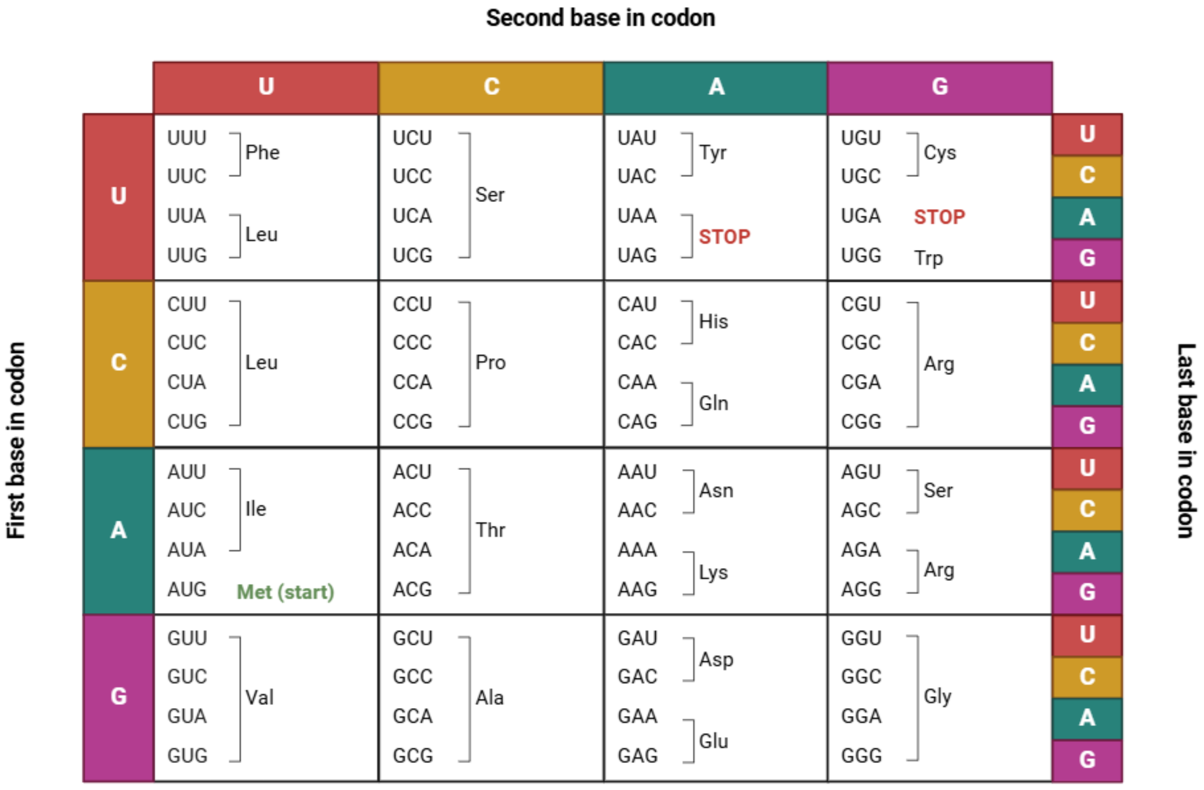


- **Transfer RNA (tRNA)** molecules carry a different amino-acid for each different codon in mRNA
- Recognition is driven by base complementarity (Watson-Crick) between **codons in mRNA** and **anticodons in tRNA**

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**

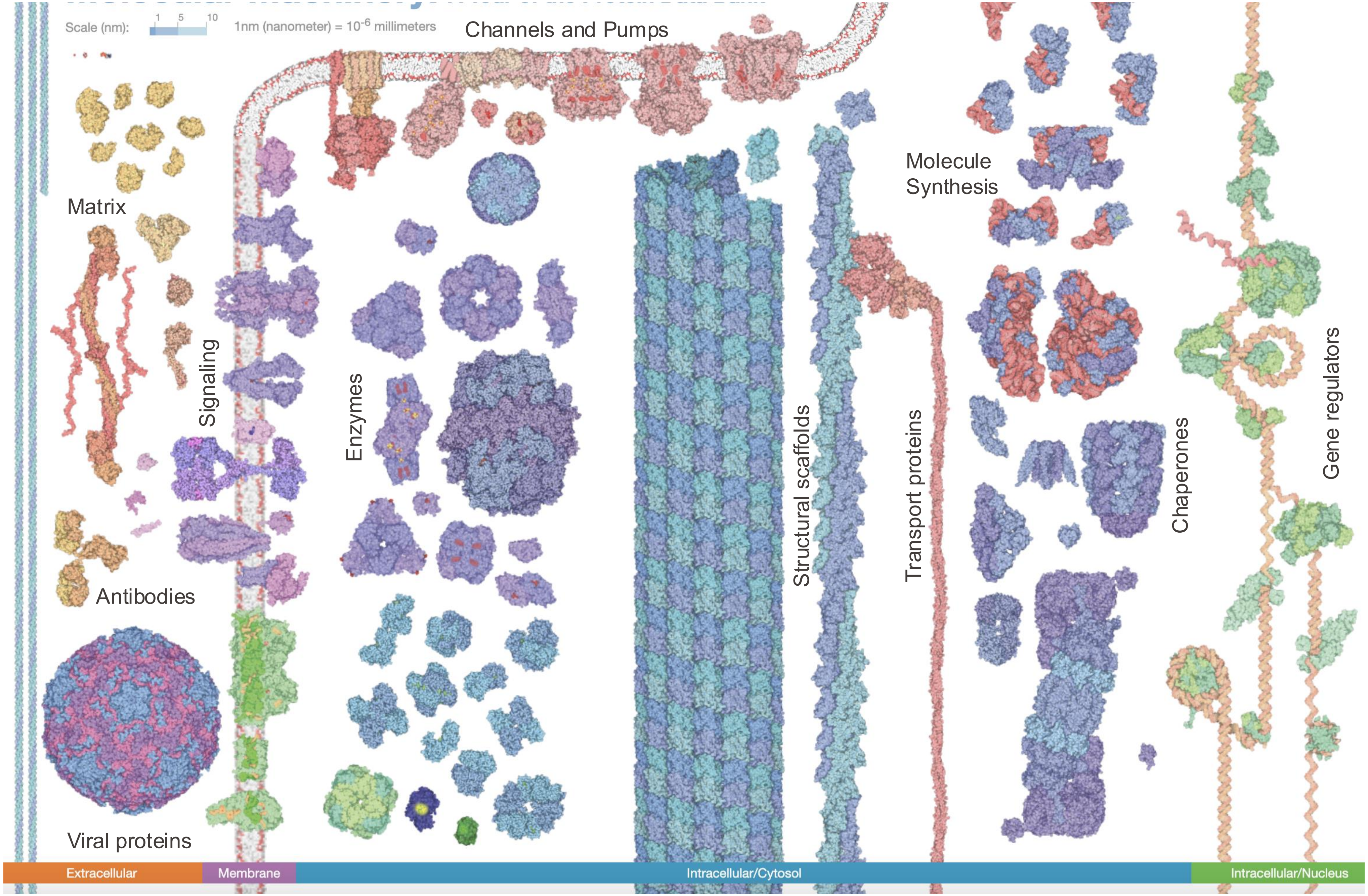


Second base in codon

First base in codon	U		C		A		G		Last base in codon
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG		UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met (start)	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Codons and corresponding amino acids

Protein diversity



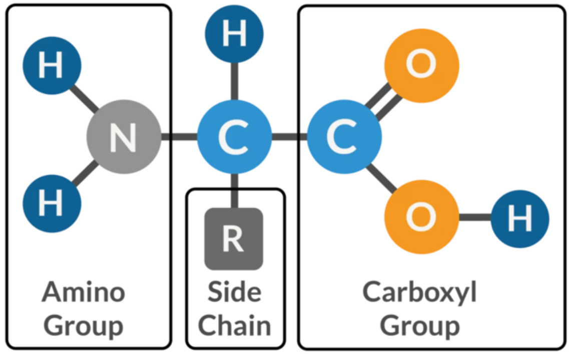
Bacterial cells ~ 4’000 proteins

Human cells ~ 80’000-200’000 proteins

Amino acids are the building blocks of proteins

- An **amino acid** is an **organic molecule** that serves as the **basic building block of proteins**.

- Each amino acid has the same general structure:
  - A **central carbon atom (α-carbon)**
  - Attached to a **hydrogen atom (H)**
  - An **amino group (–NH_2)**
  - A **carboxyl group (–COOH)**
  - A **side chain (R group)** that determines its chemical properties.



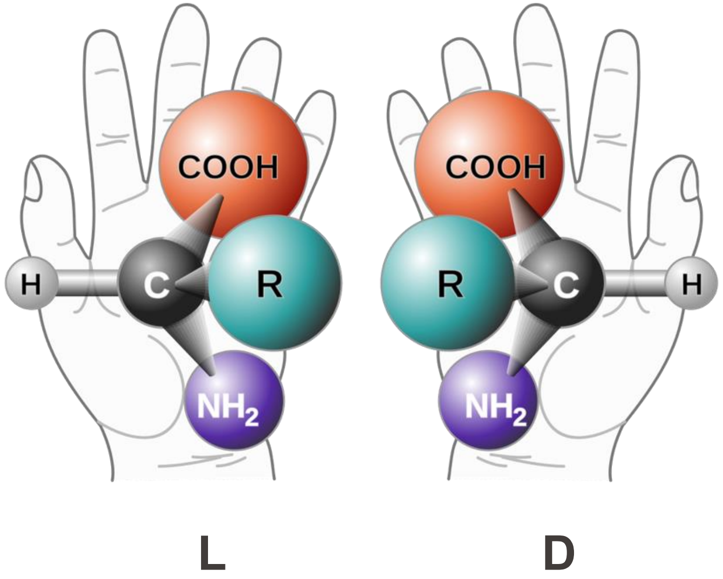
- Amino acids are chiral and can exist in L or D forms

(Note that L amino acids can't be converted in D just by rotation)

- **All genetically encoded amino acids are in the L form**

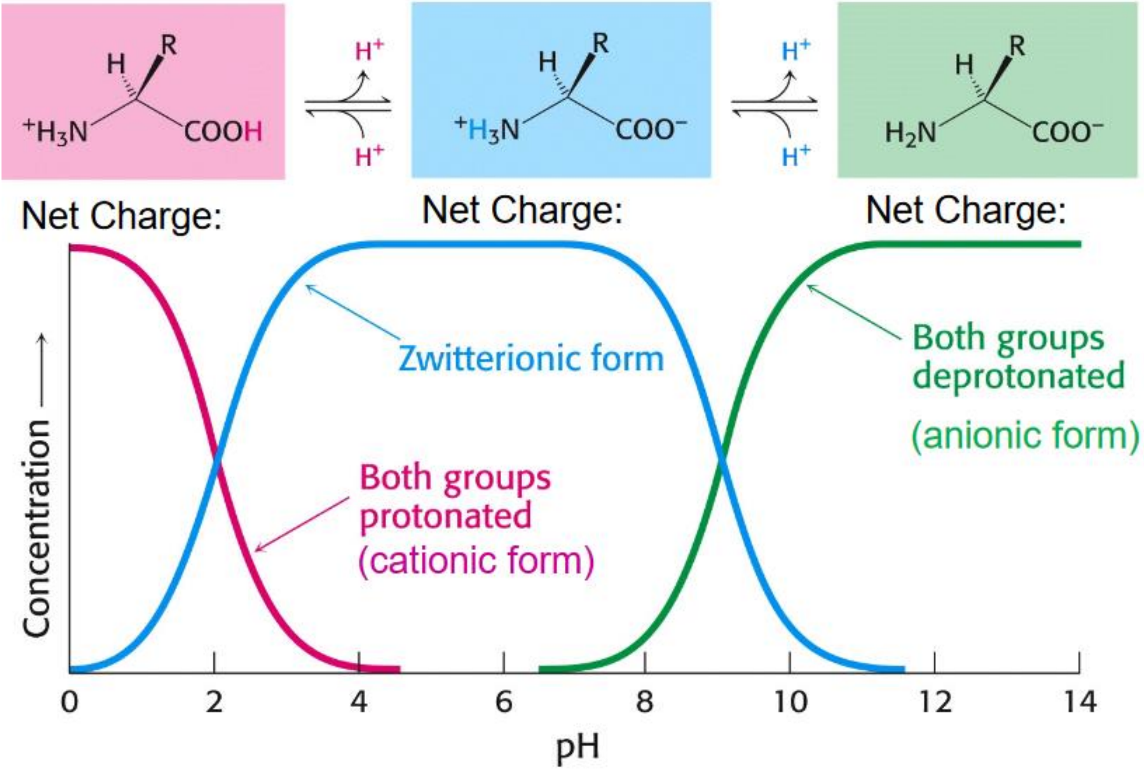
- D-amino acids exist but they usually have different roles

(e.g., neurotransmitters, cell wall structure assembly, secondary metabolites)

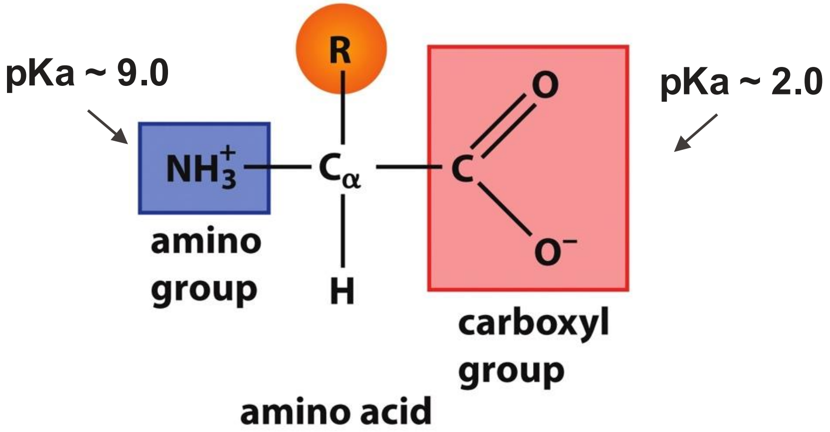


Amino acids exist as zwitterions in solution

- In aqueous solution, the amino and carboxyl groups of amino acids can gain or lose protons, causing them to switch between **positively** and **negatively** charged forms.
- The extent of protonation or deprotonation depends on **solution pH** and the **pKa** value of the chemical group

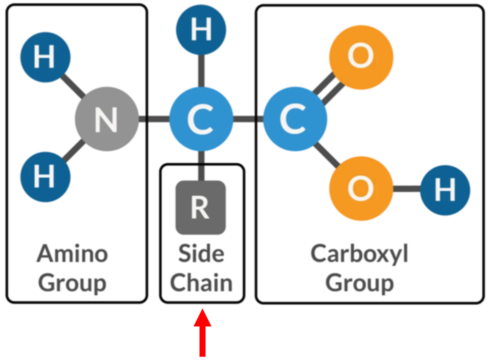


In aqueous solutions at neutral pH (=7) amino-acids are **zwitterionic**

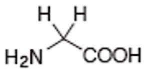


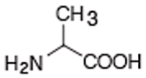
- pKa is the measure of acid's strength (= pH at which 50% of the acid group is dissociated).

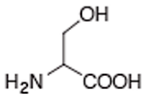
20 proteinogenic amino acids

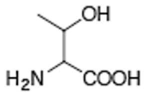


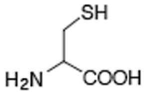
- There are **20 primary amino acids** that assemble all proteins
- Amino acids have the same scaffold but **differ only in the side chain group (R)**

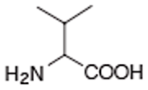

Glycine (Gly, G)
MW: 75.07

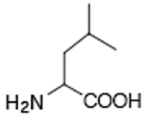

Alanine (Ala, A)
MW: 89.09

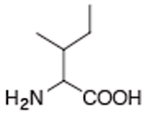

Serine (Ser, S)
MW: 105.09

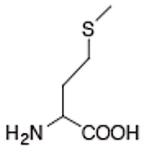

Threonine (Thr, T)
MW: 119.1,

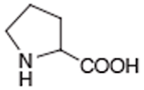

Cysteine (Cys, C)
MW: 121.2

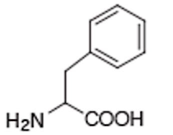

Valine (Val, V)
MW: 117.1

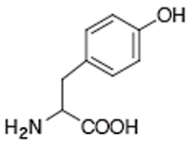

Leucine (Leu, L)
MW: 131.2

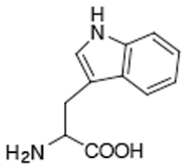

Isoleucine (Ile, I)
MW: 131.2

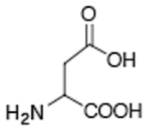

Methionine (Met, M)
MW: 149.2

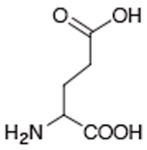

Proline (Pro, P)
MW: 115.1

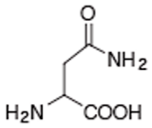

Phenylalanine (Phe, F)
MW: 165.2

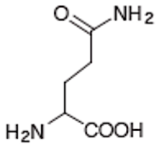

Tyrosine (Tyr, Y)
MW: 181.2

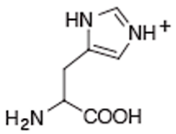

Tryptophan (Trp, W)
MW: 204.2

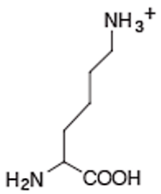

Aspartic Acid (Asp, D)
MW: 133.1, $pK_a = 3.9$

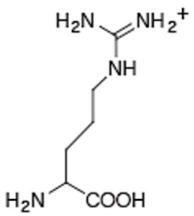

Glutamic Acid (Glu, E)
MW: 147.1, $pK_a = 4.07$


Asparagine (Asn, N)
MW: 132.1


Glutamine (Gln, Q)
MW: 146.1


Histidine (His, H)
MW: 155.2, $pK_a = 6.04$


Lysine (Lys, K)
MW: 146.2, $pK_a = 10.79$


Arginine (Arg, R)
MW: 174.2, $pK_a = 12.48$

- Their names are derived based on the source, chemical properties or structure
  - Asparagine - Isolated from asparagus
  - Glycine - Greek for sweet
  - Argininine - Silver-white
  - Tyrosine - Greek for cheese
  - Leucine - Greek for white

- Note the 1- and 3-letter abbreviations

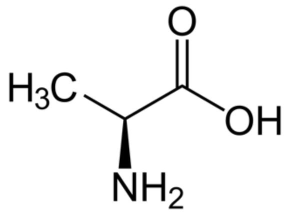
All amino acid names, abbreviations and structures need to be learned

* 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.

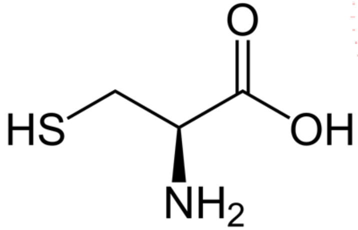
Essential and Non-essential amino acids

- Amino acids can be categorized based on whether the body can **synthesize** them (**non-essential**) or must **obtain them through diet (essential)**

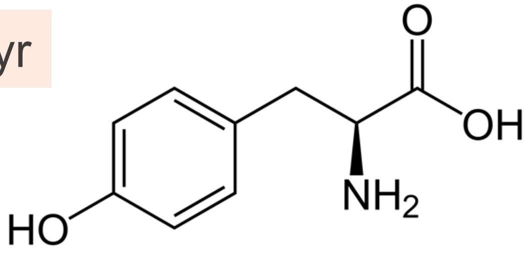
Ala



Cys



Tyr



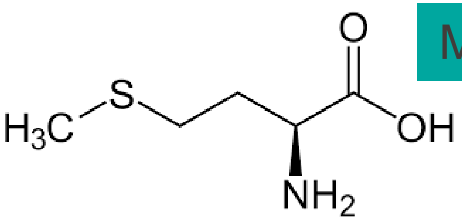
Non-Essential

- Alanine
- Arginine
- Asparagine
- Aspartate
- Cysteine
- Glutamate
- Glutamine
- Glycine
- Proline
- Serine
- Tyrosine

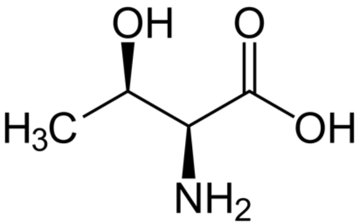
Essential

- Histidine
- Isoleucine
- Leucine
- Lysine
- Methionine
- Phenylalanine
- Threonine
- Tryptophan
- Valine

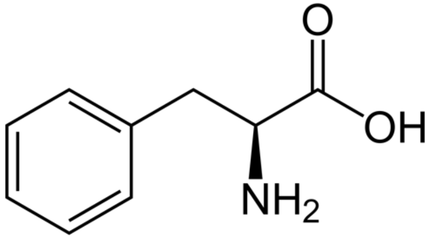
Met



Thr



Phe



Why is Tyr non-essential while Phe is?

- Plants and microbes can produce all 20 amino acids, but most mammals have evolved to produce ~10-11 amino acids and acquire other through diet.

Different groups of amino acids based on side-chain

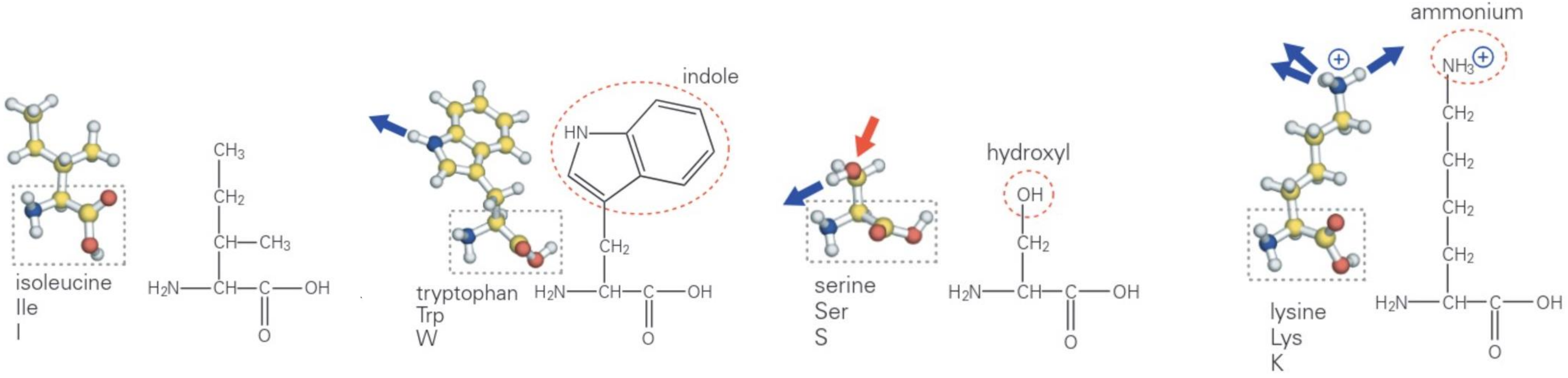
- Amino acids are often categorized based on the chemical properties of their side chains

Non-polar
(Hydrophobic)

Aromatic
(Hydrophobic*)

Polar but uncharged
(hydrophilic)

Charged polar
(hydrophilic)

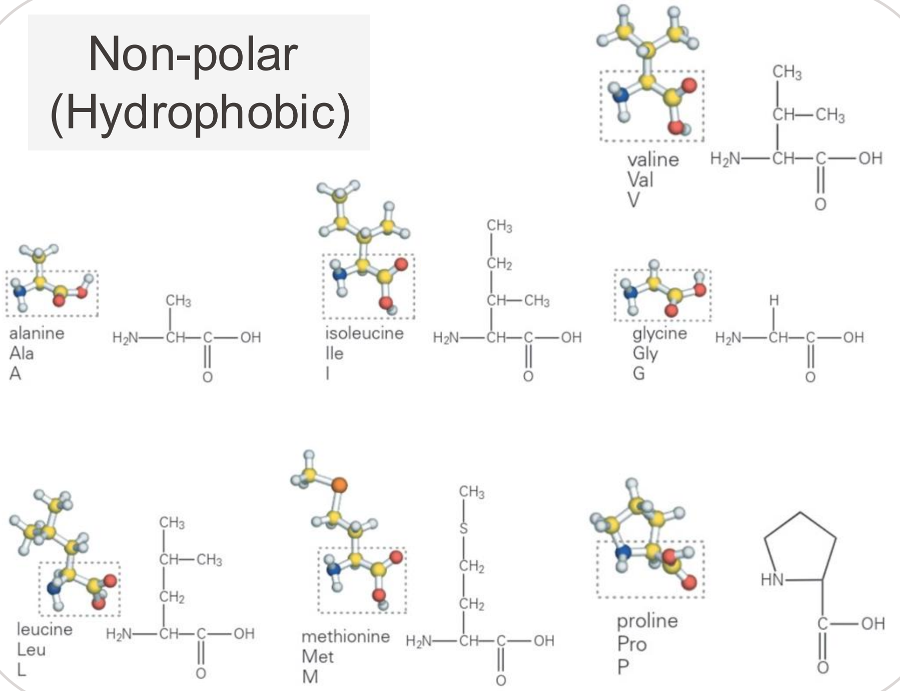


🡒 Hydrogen bond acceptors 🡒 Hydrogen bond donors

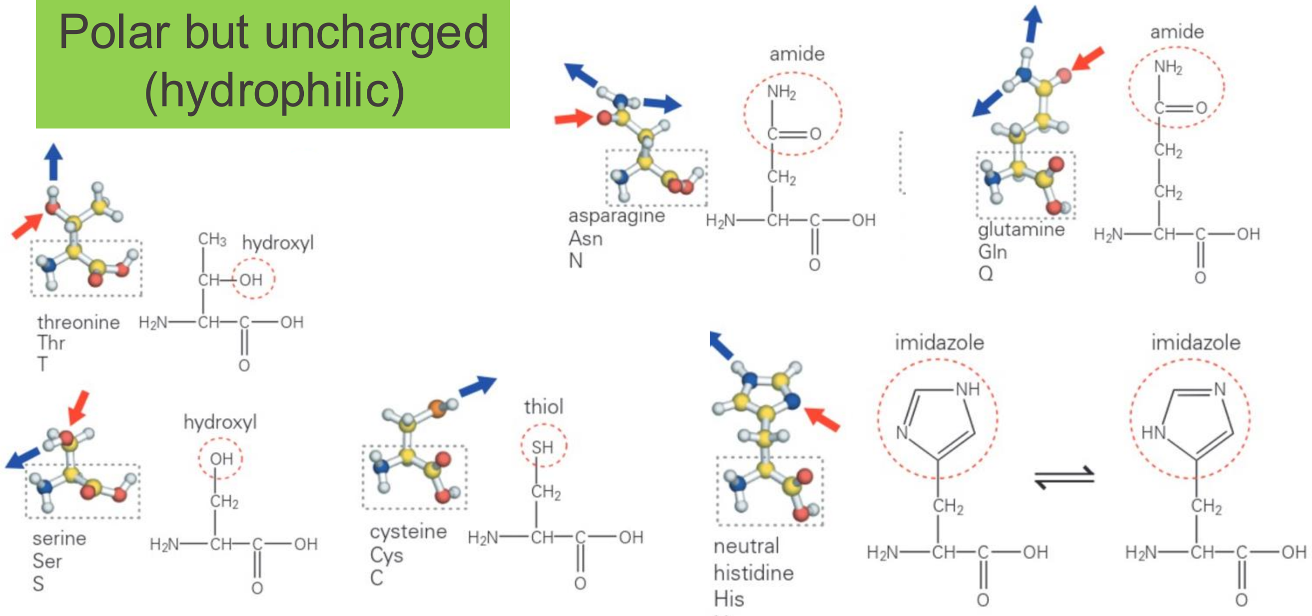
- Note that Tryptophan also has a polar NH group but the rest of its side chain is very hydrophobic
- Imperfect categorization system but very useful to understand different properties of amino-acids

Different groups of amino acids based on side-chain

Non-polar (Hydrophobic)

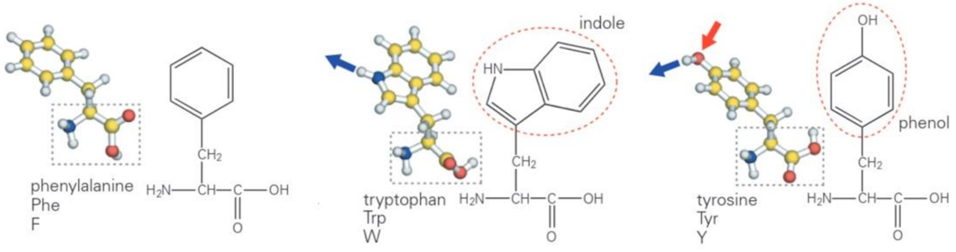


Polar but uncharged (hydrophilic)



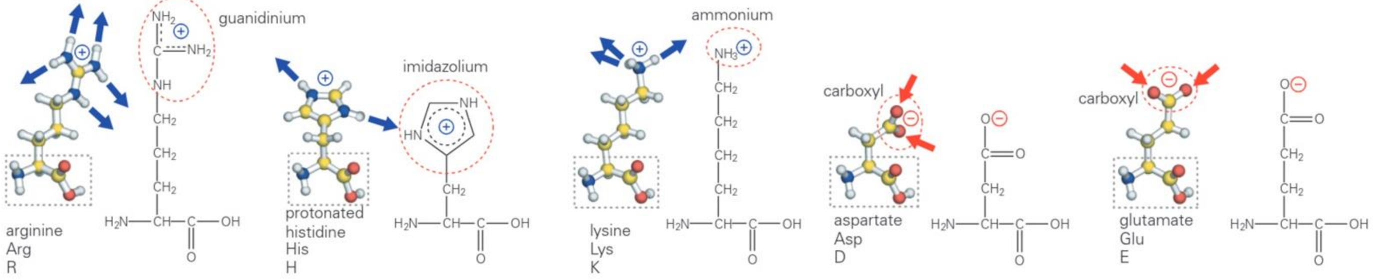
* Imidazole can be charged

Aromatic (Hydrophobic*)



* Trp and Tyr also have polar groups

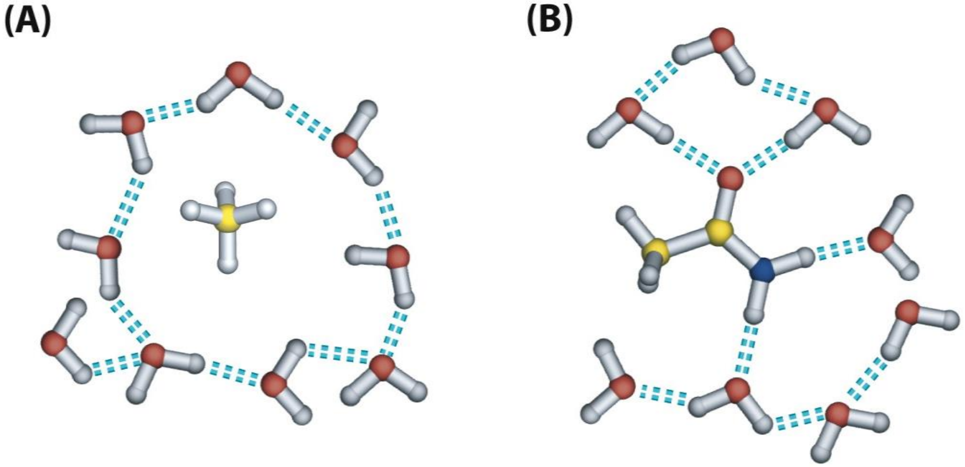
Charged polar (hydrophilic)



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:

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**

Water/Octanol transfer free energy

Amino acids with charged side chains

- 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

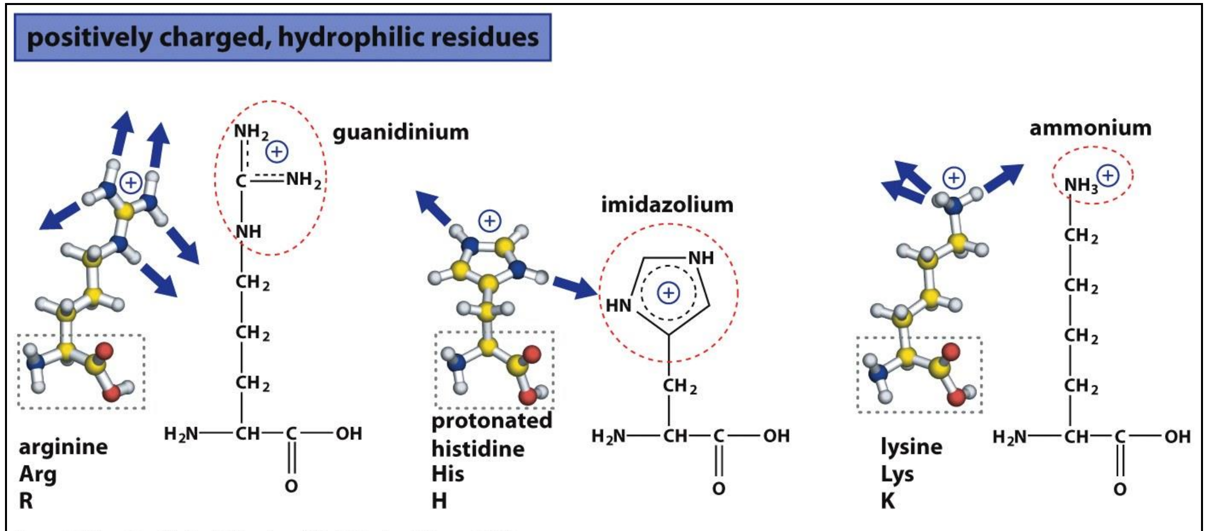


Figure 1.29 (part 2 of 4) The Molecules of Life (© Garland Science 2013)

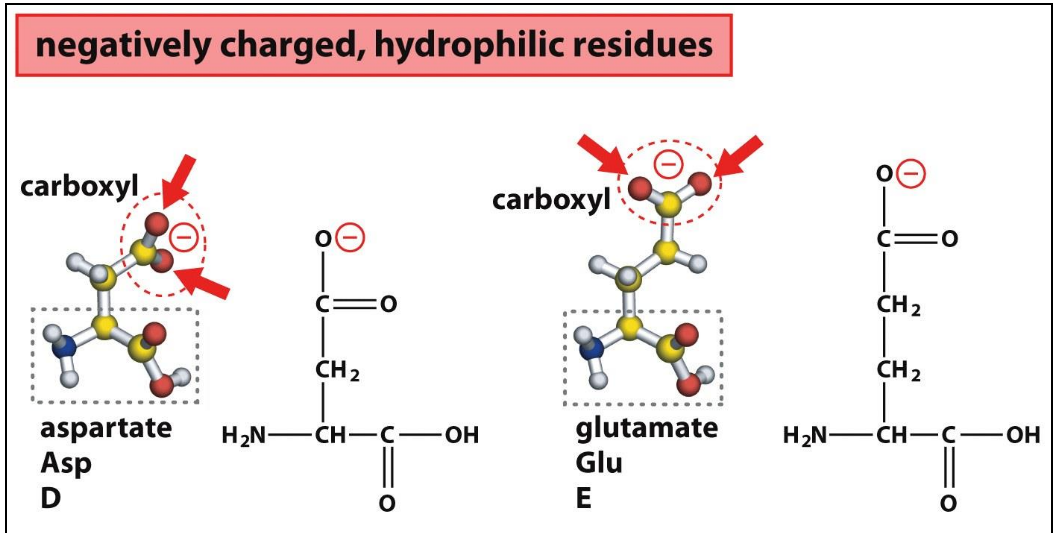


Figure 1.29 (part 4 of 4) The Molecules of Life (© Garland Science 2013)

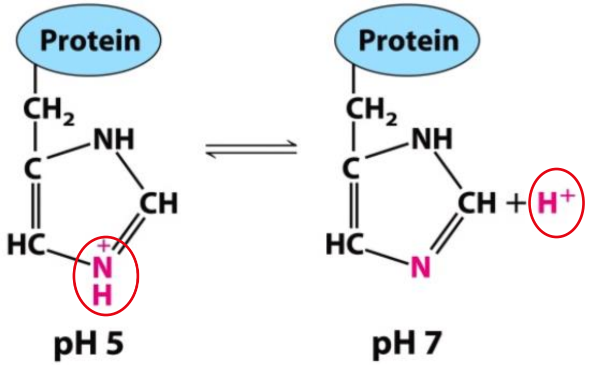
pKa values: 12.48 6.04 10.79 3.90 4.07

What about His?

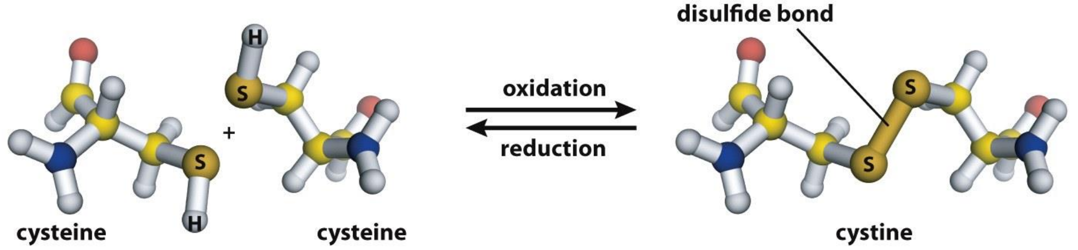
pH > pKa -> Deprotonated group (e.g., COO-)
pH < pKa -> Protonated group (e.g., COOH)

Versatile amino acids

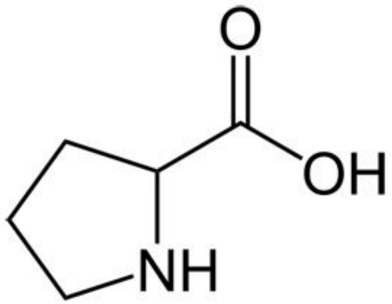
- **Histidine** can get protonated if the pH drops below ~6, giving it a positive charge
- Sometimes serves as pH-sensitive trigger in proteins

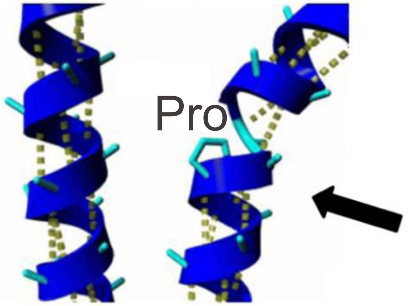


- **Cysteine** can form covalent disulfide bonds with other cysteines under oxidizing conditions (e.g., outside of a cell).
- Reversed under reducing conditions

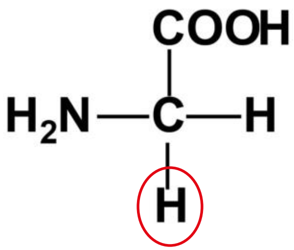


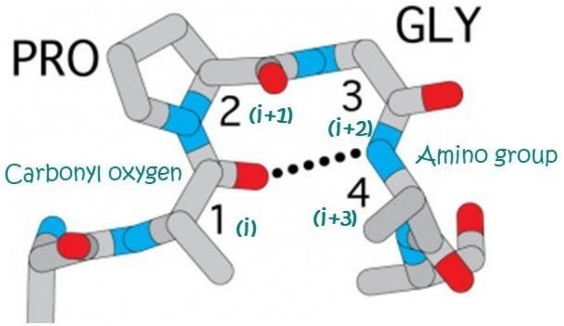
- **Proline** is not a real amino acid but rather an **imino** acid.
- Disruptive to α-helices and β-strands in proteins, but commonly found in flexible loops and turns





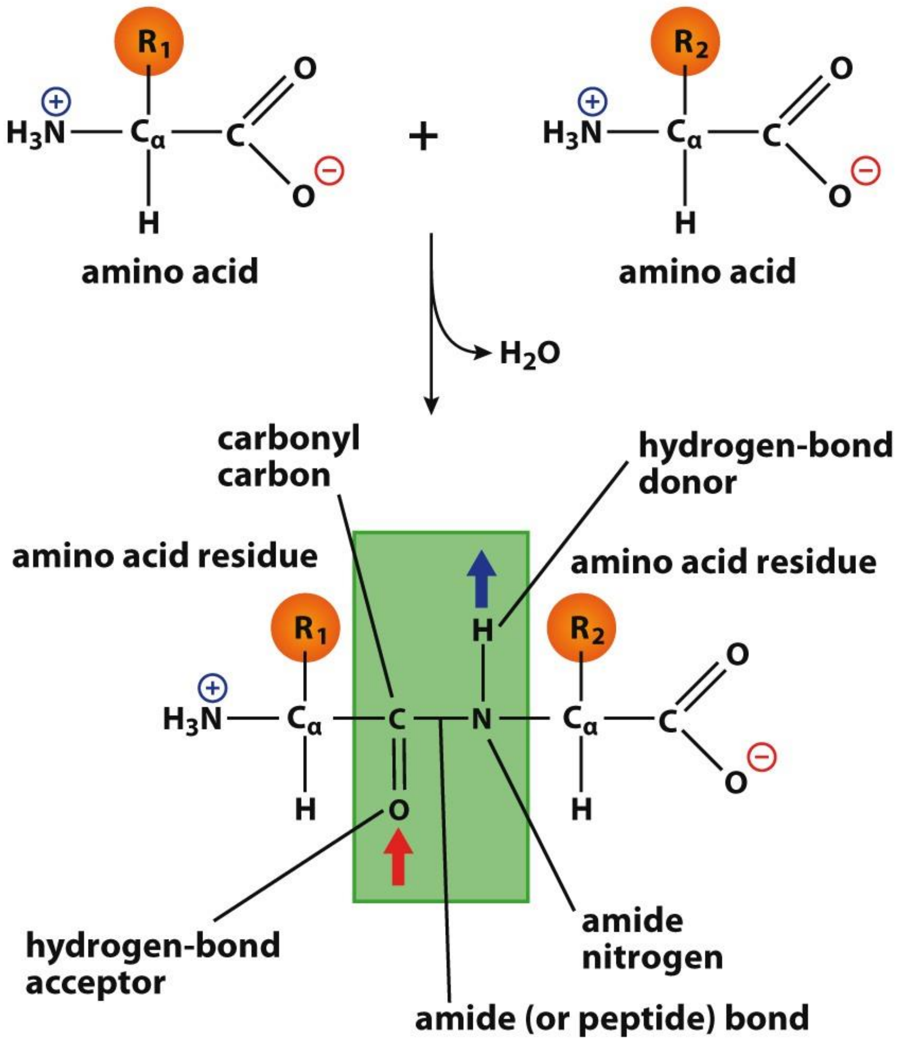
- **Glycine** only has a single hydrogen as a side-chain, which allows for more movement in proteins.
- Commonly found in flexible loops or peptide turns





Proteins are linear 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

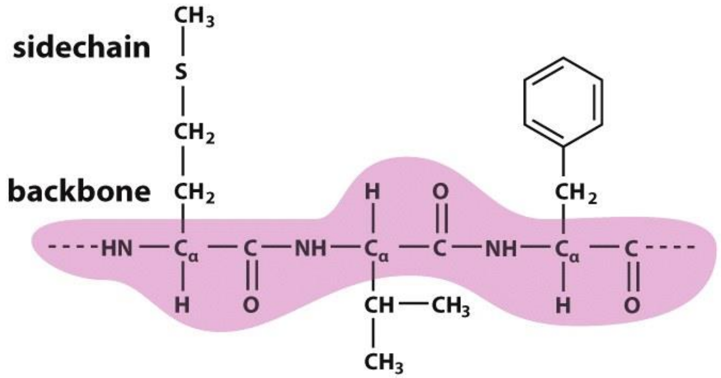
- This reaction is catalyzed by a large assembly of proteins and RNA called the **ribosome**

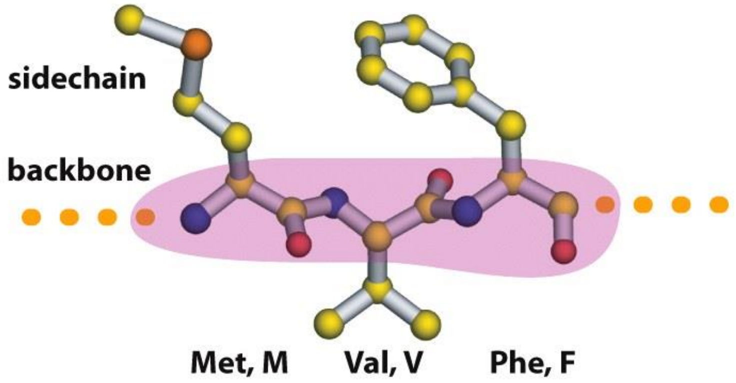
- The end product is a **polypeptide chain** where amino acids are added sequentially (no branching)

Proteins are linear 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^-)

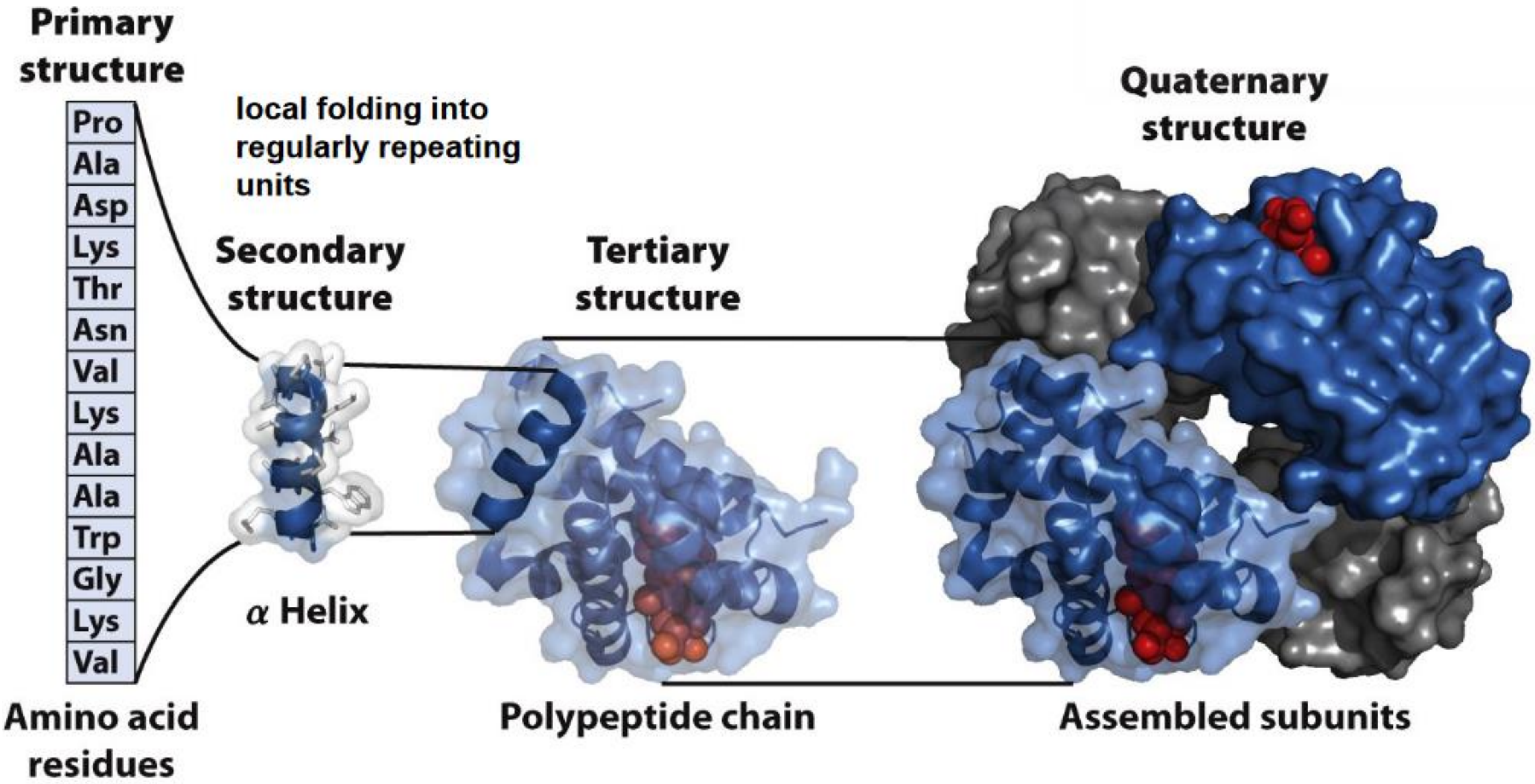
- Chemical and three-dimensional structure of a short peptide sequence





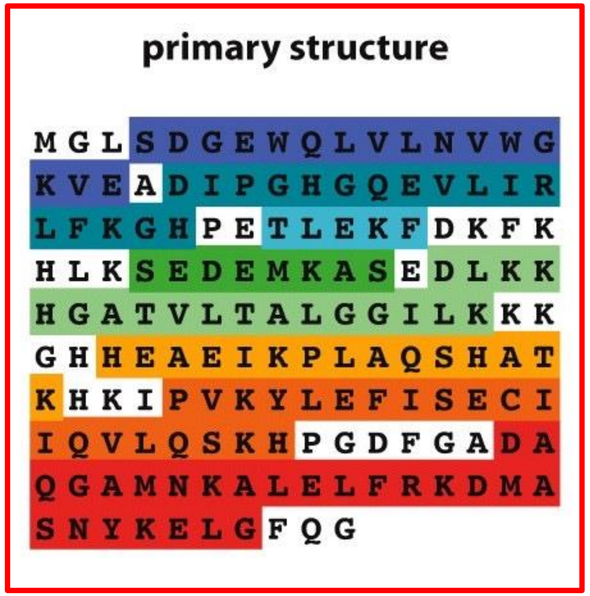
Hierarchical organization of protein structure

- 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

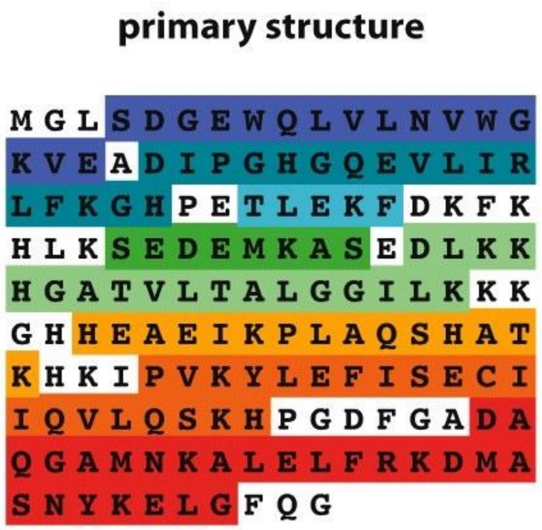


primary structure

M G L S D G E W Q L V L N V W G
K V E A D I P G H G Q E V L I R
L F K G H P E T L E K F D K F K
H L K S E D E M K A S E D L K K
H G A T V L T A L G G I L K K K
G H H E A E I K P L A Q S H A T
K H K I P V K Y L E F I S E C I
I Q V L Q S K H P G D F G A D A
Q G A M N K A L E L F R K D M A
S N Y K E L G F Q G

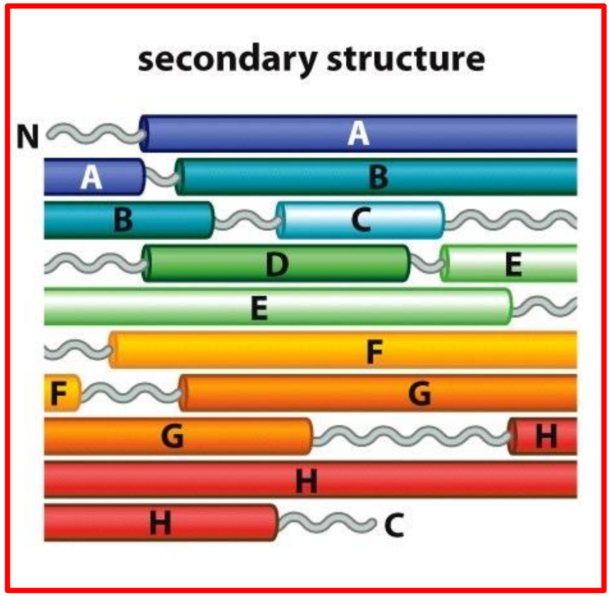
- **Sequence**: the pattern of amino acids along a linear polypeptide chain
- Always starts with **Methionine** encoded by the **AUG** (ATG in DNA) start codon
- Can be used for bioinformatic analyses, calculations of biophysical properties, molecular weight etc.

Secondary structure



primary structure

M G L S D G E W Q L V L N V W G
K V E A D I P G H G Q E V L I R
L F K G H P E T L E K F D K F K
H L K S E D E M K A S E D L K K
H G A T V L T A L G G I L K K K
G H H E A E I K P L A Q S H A T
K H K I P V K Y L E F I S E C I
I Q V L Q S K H P G D F G A D A
Q G A M N K A L E L F R K D M A
S N Y K E L G F Q G



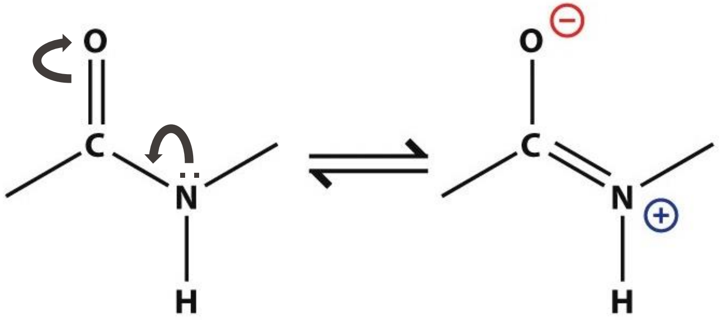
secondary structure

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

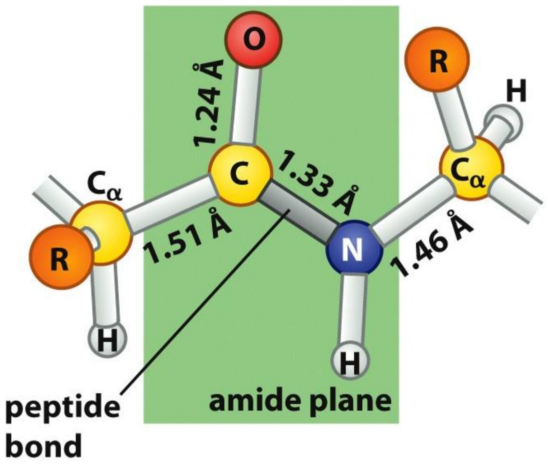
Backbone torsions and secondary structure

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

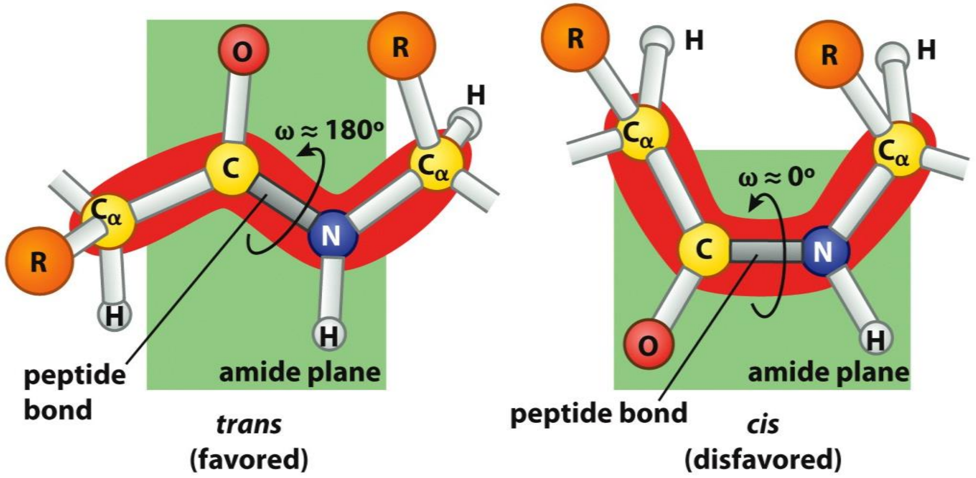
Peptide bond has a partial double bond character



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



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



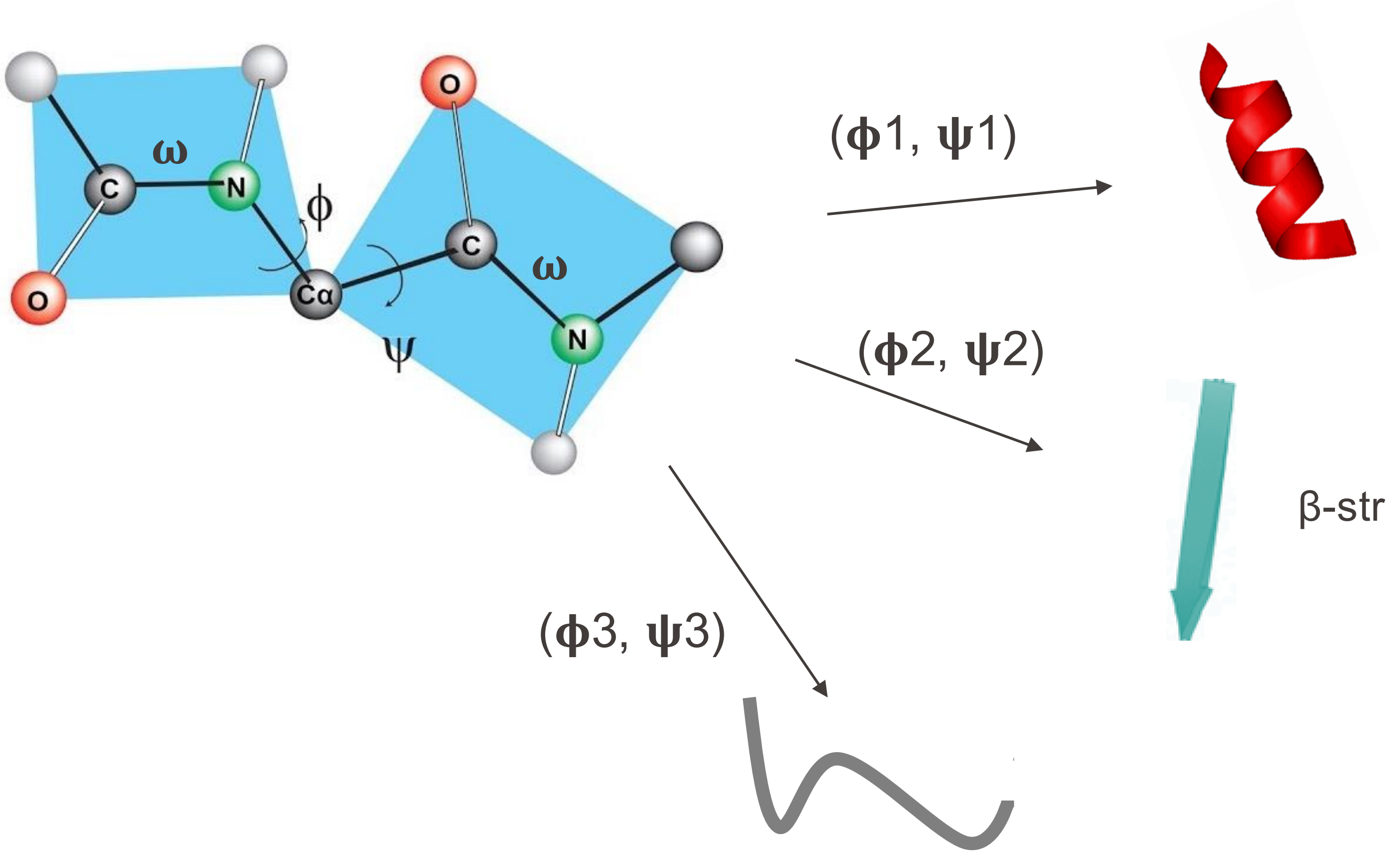
Omega (**ω**) is the bond angle of the peptide bond

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

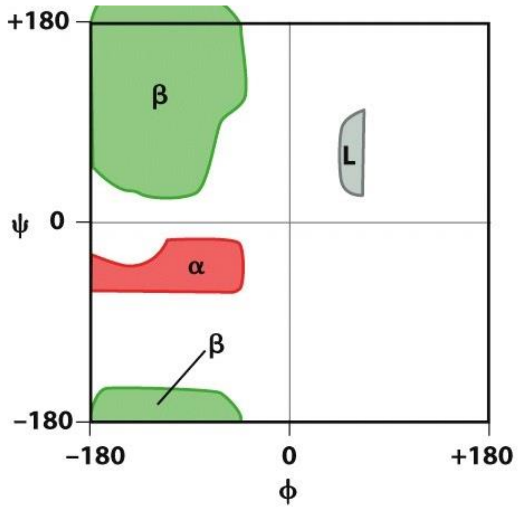
(ϕ2, ψ2)

β-strands

(ϕ3, ψ3)

Loops

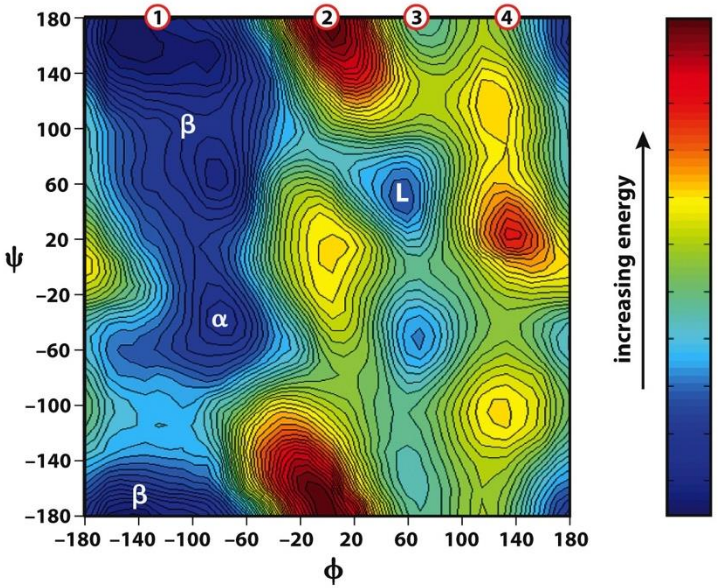
Ramachandran Plot

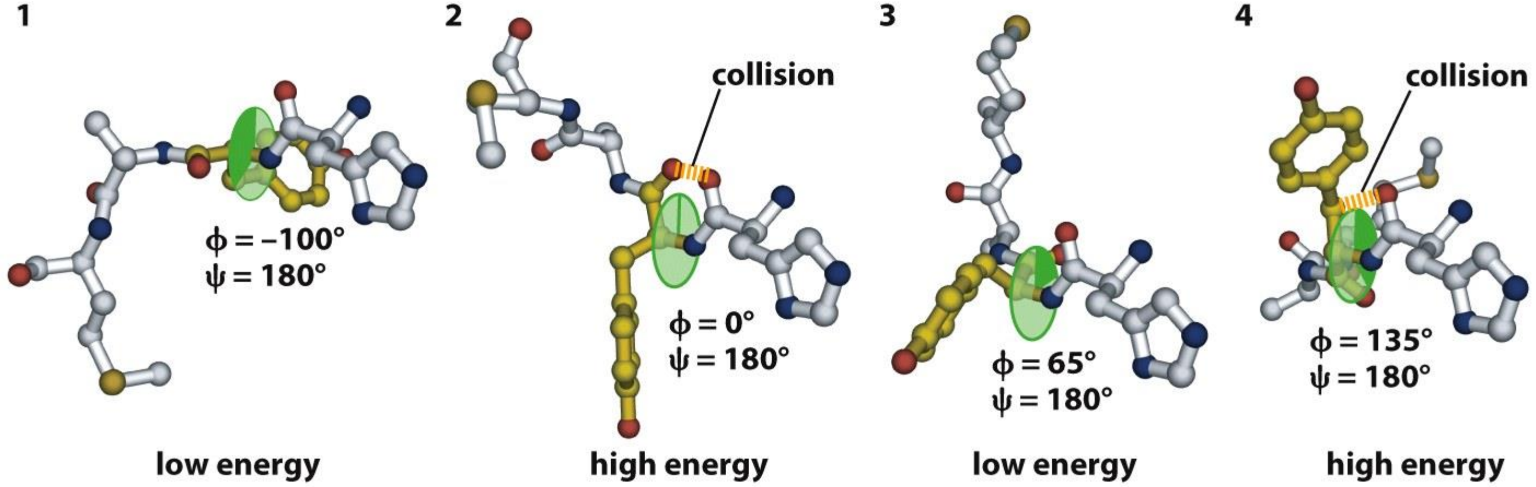


Backbone torsions and secondary structure

G.N. Ramachandran
(1922-2001)

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





- The values will be different for every amino acid combination due to unique steric restraints due to the size and chemical nature of the side chains

Backbone torsions and secondary structure

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

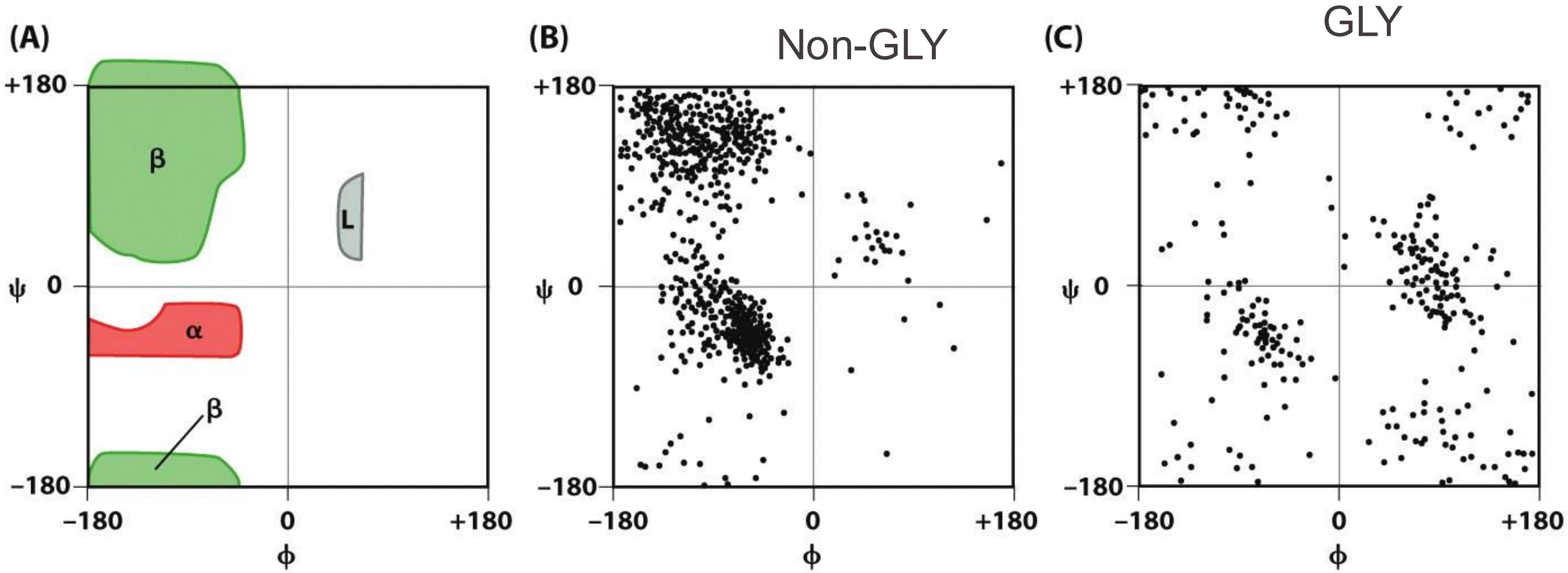
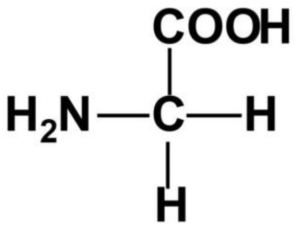


Figure 4.20 The Molecules of Life (© Garland Science 2013)

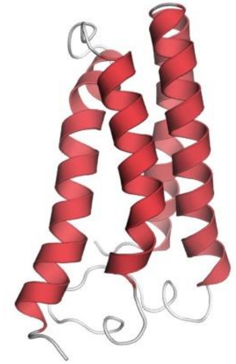


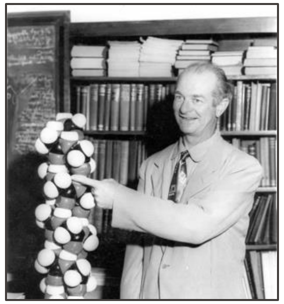
$$H_2N-\overset{\displaystyle COOH}{\underset{\displaystyle H}{C}}-H$$

Secondary structure elements: α-helix

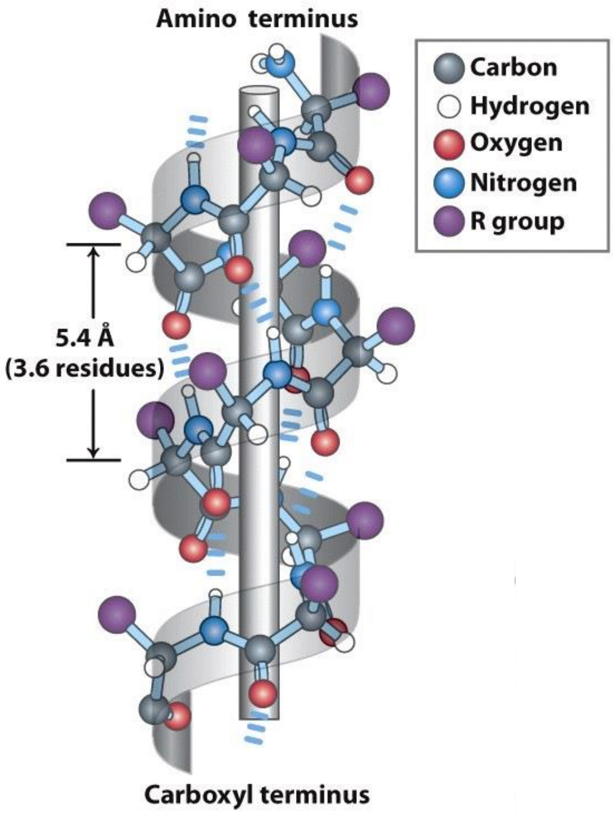
- Segments of amino acids adopt **α-helices** these structural motifs are the architectural elements of protein structure

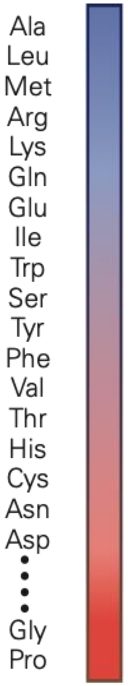
α-helices





Linus Pauling
(1901-1994)



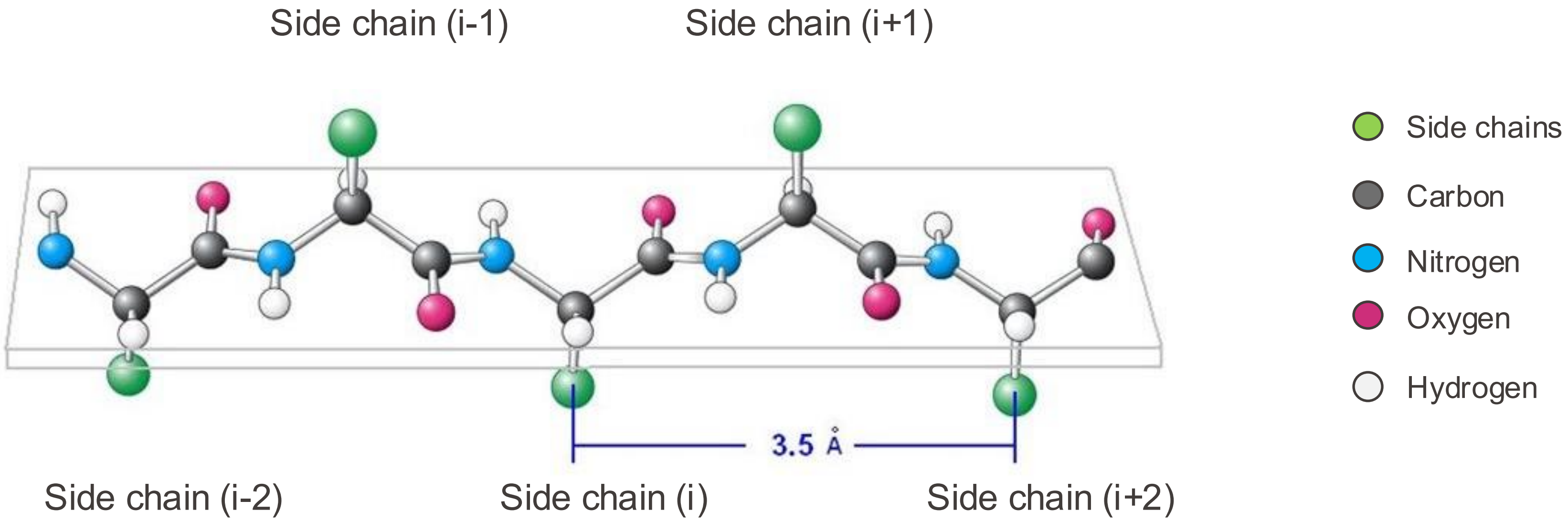


- Right-handed helix with **a rise of 5.4Å**. Each turn consists of **3.6 amino acids**
- In the α-helix, the **carbonyl oxygen** of residue “i” forms a hydrogen bond with the **amide hydrogen** of residue “i+4”.
- Although each hydrogen bond is relatively weak in isolation, the sum of the hydrogen bonds in a helix makes it quite stable.
- 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).

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

Secondary structure elements: β-strands and sheets

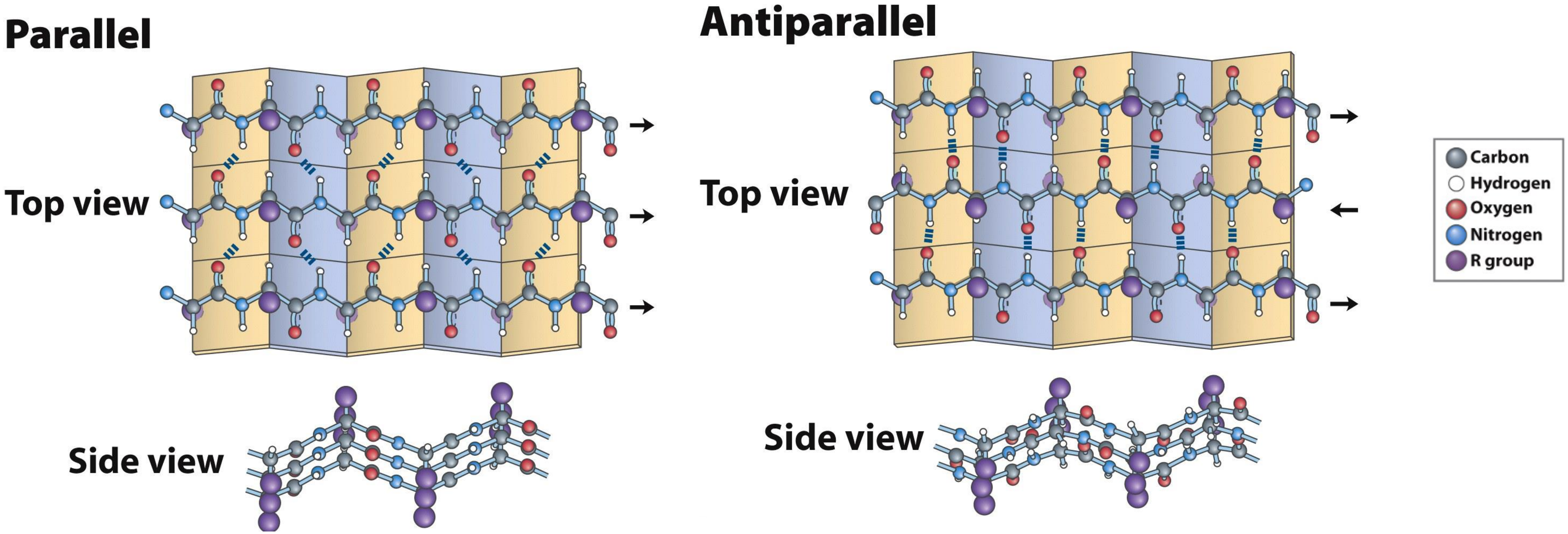
β-strands



- In a β-strand amino acids are arranged so that the side chains of neighboring residues face in the opposite directions (“zig-zag” conformation)
- The same is also true for alternating main-chain groups
- 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

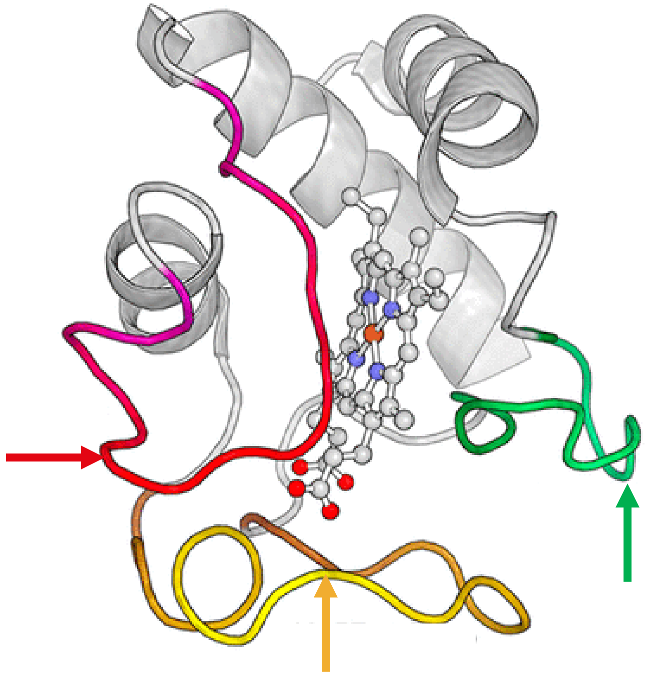
β-sheets



- 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)

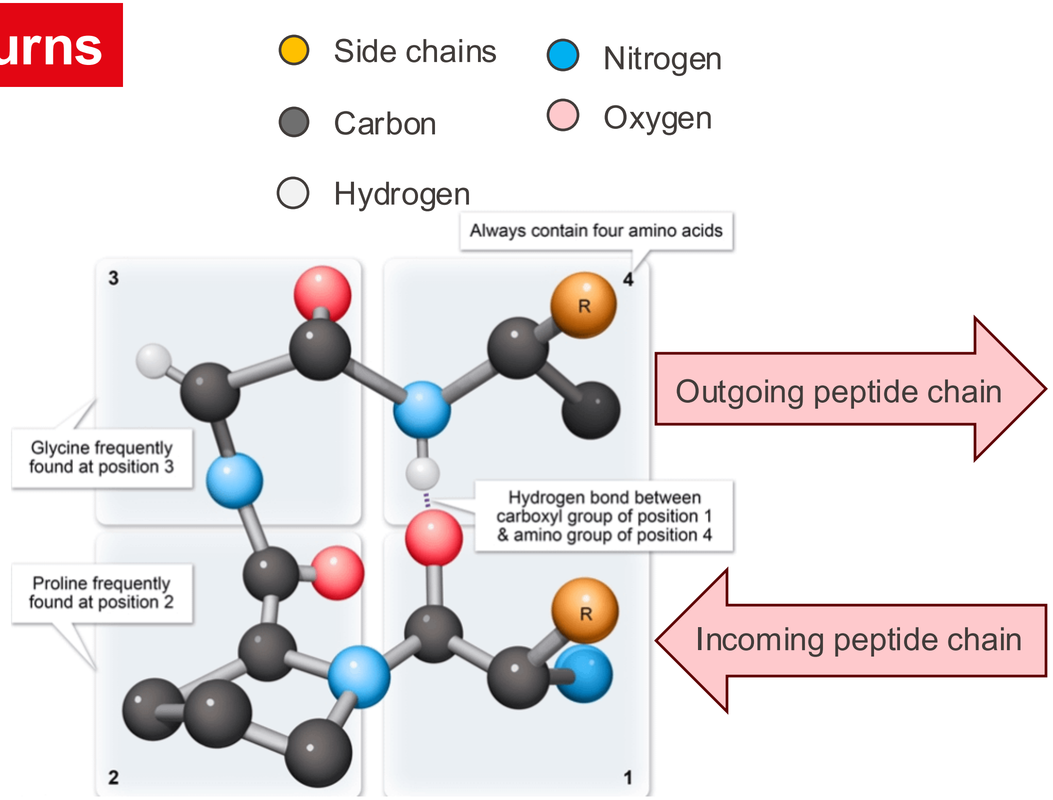
Secondary structure elements: Loops and Turns

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

Secondary structure elements combine to create motifs

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

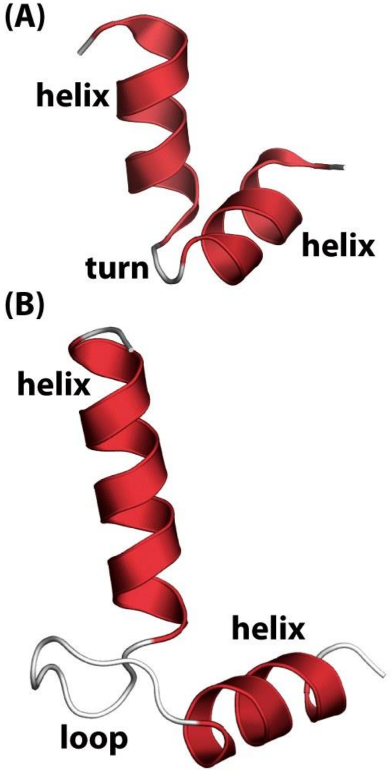


Figure 4.33 The Molecules of Life (© Garland Science 2013)

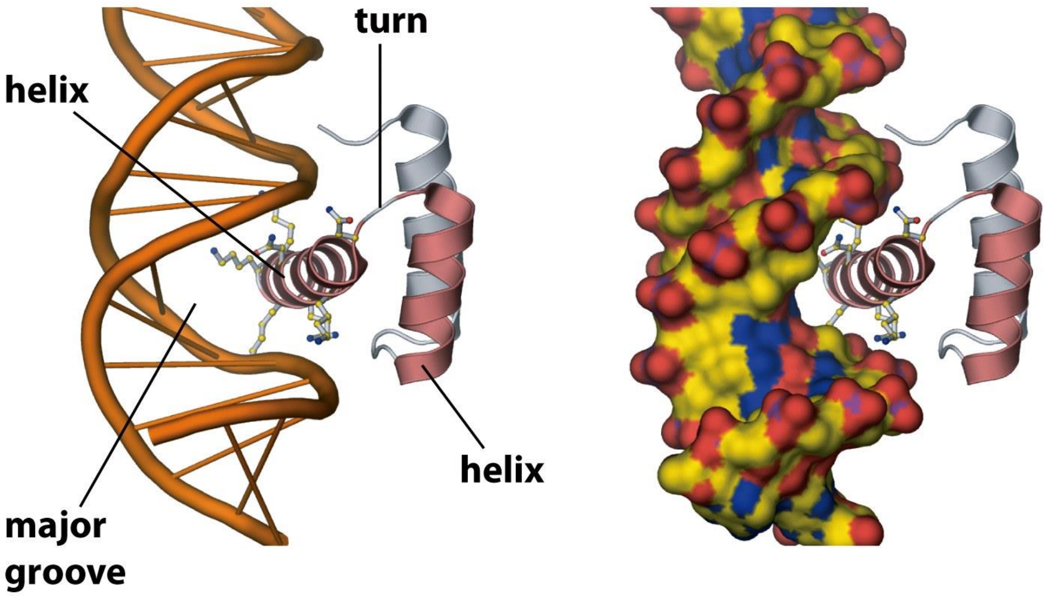


Figure 4.34 The Molecules of Life (© Garland Science 2013)

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

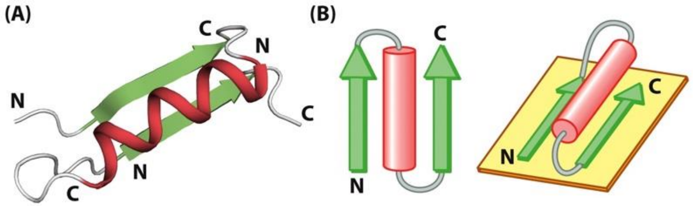


Figure 4.37 The Molecules of Life (© Garland Science 2013)

Helix-helix

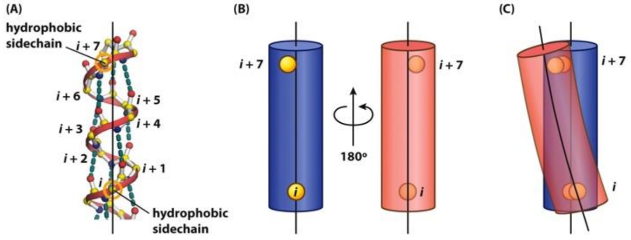


Figure 4.39 The Molecules of Life (© Garland Science 2013)

α/β domains

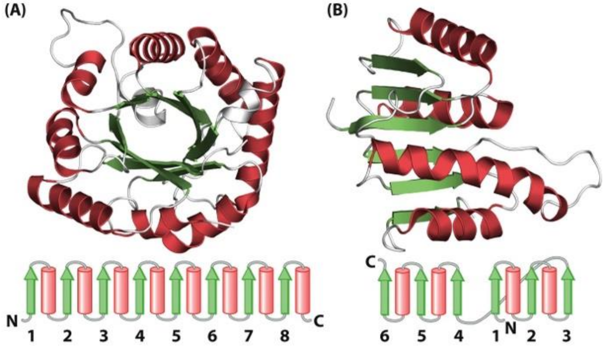


Figure 4.49 The Molecules of Life (© Garland Science 2013)

4 helices

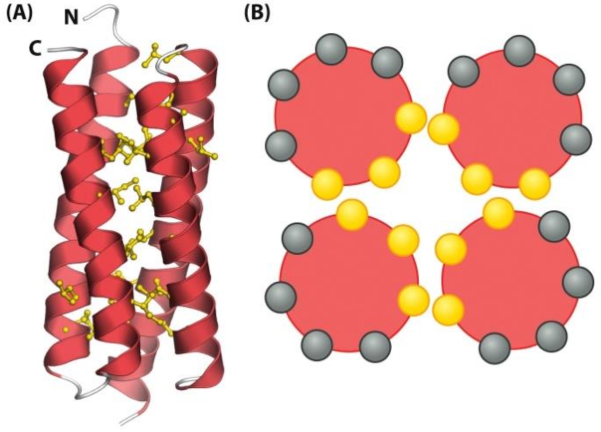


Figure 4.44 The Molecules of Life (© Garland Science 2013)

Coiled-coil

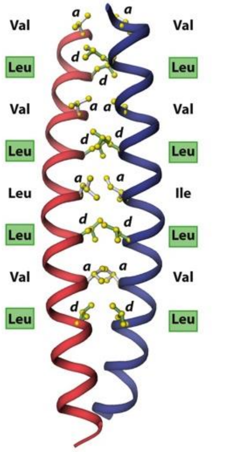
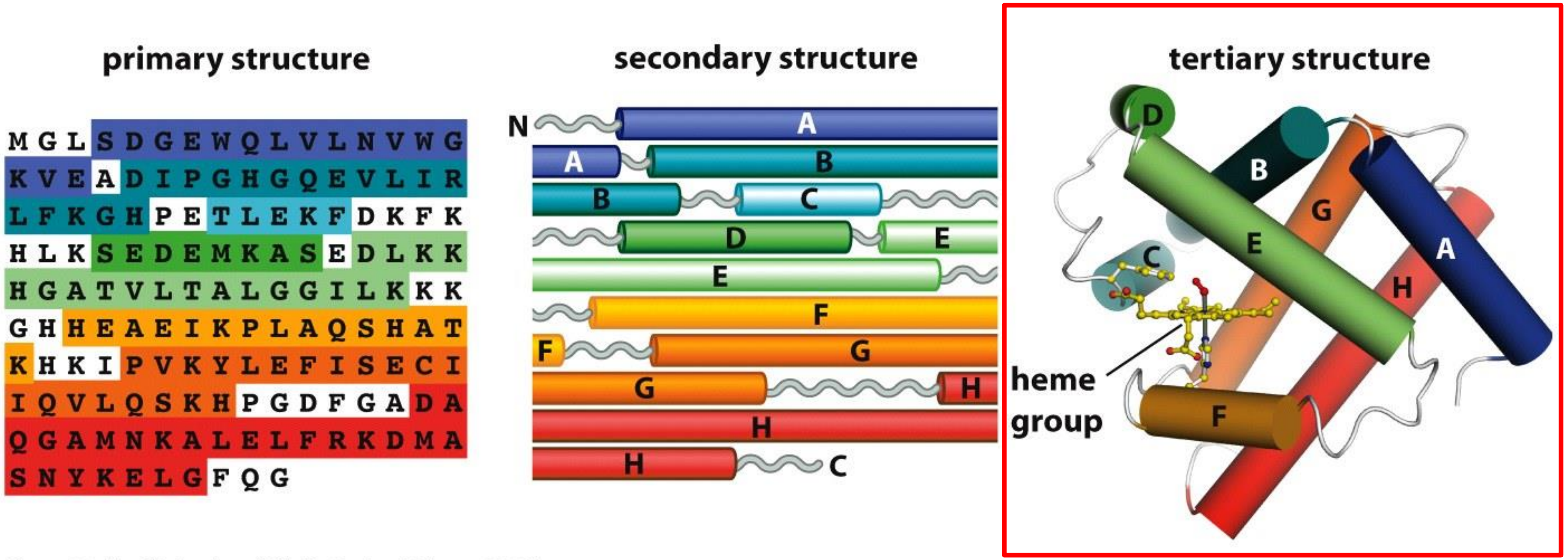


Figure 4.41 The Molecules of Life (© Garland Science 2013)

“Leucine zipper”

Tertiary structure



M G L S D G E W Q L V L N V W G
K V E A D I P G H G Q E V L I R
L F K G H P E T L E K F D K F K
H L K S E D E M K A S E D L K K
H G A T V L T A L G G I L K K K
G H H E A E I K P L A Q S H A T
K H K I P V K Y L E F I S E C I
I Q V L Q S K H P G D F G A D A
Q G A M N K A L E L F R K D M A
S N Y K E L G F Q G

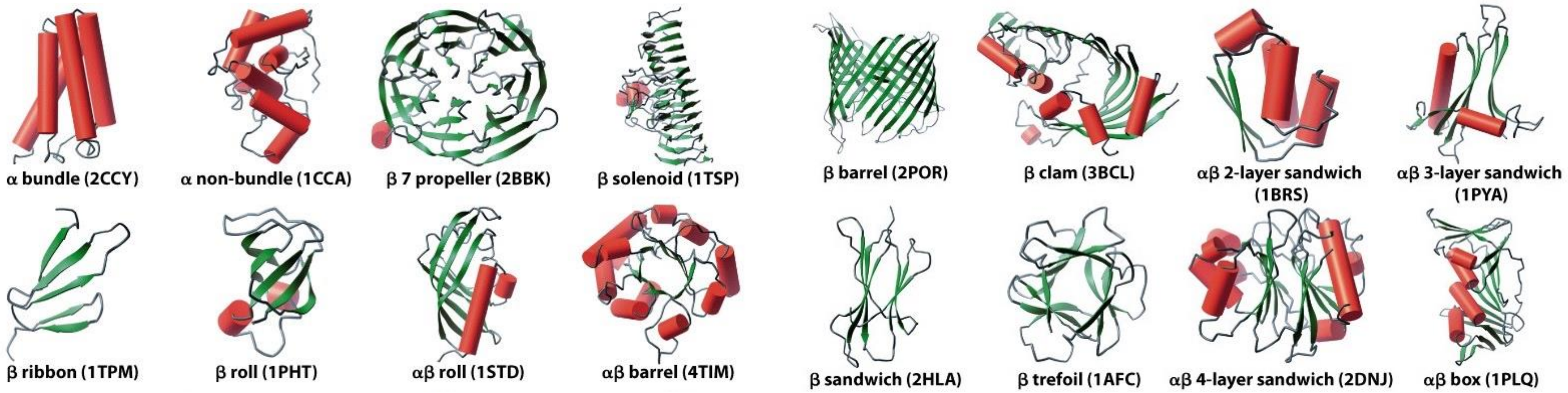
Figure 4.1 The Molecules of Life (© Garland Science 2013)

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

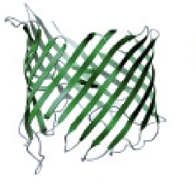
- Sometimes referred to as **protein fold**.

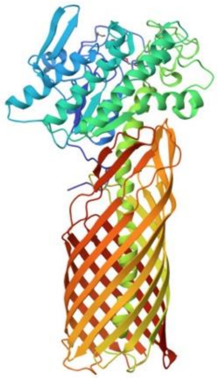
Folds are created from secondary structure blocks

- They serve as structural support, biochemical catalysts, hormones, enzymes, building blocks etc.
- Some are more abundant than others (regulated by transcription, translation or degradation)

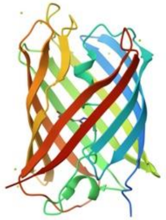


- Very similar folds can have very different functions and locations (depending on their sequence)


β barrel (2POR)


Autrotransporter
(PDB ID: 3KVN)


Pore-forming toxin
(PDB ID: 7AHL)


Green fluorescent protein
(PDB ID: 2GX2)

Protein Domains are the units of Tertiary structure

- A structural domain is a polypeptide chain or a part of it chain that forms an independent structural unit.

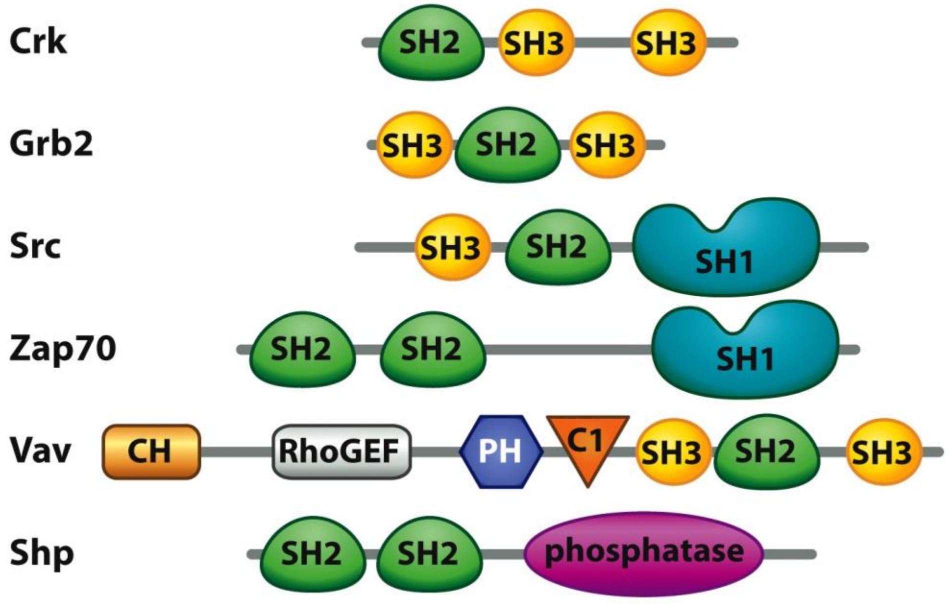


Figure 4.5 The Molecules of Life (© Garland Science 2013)

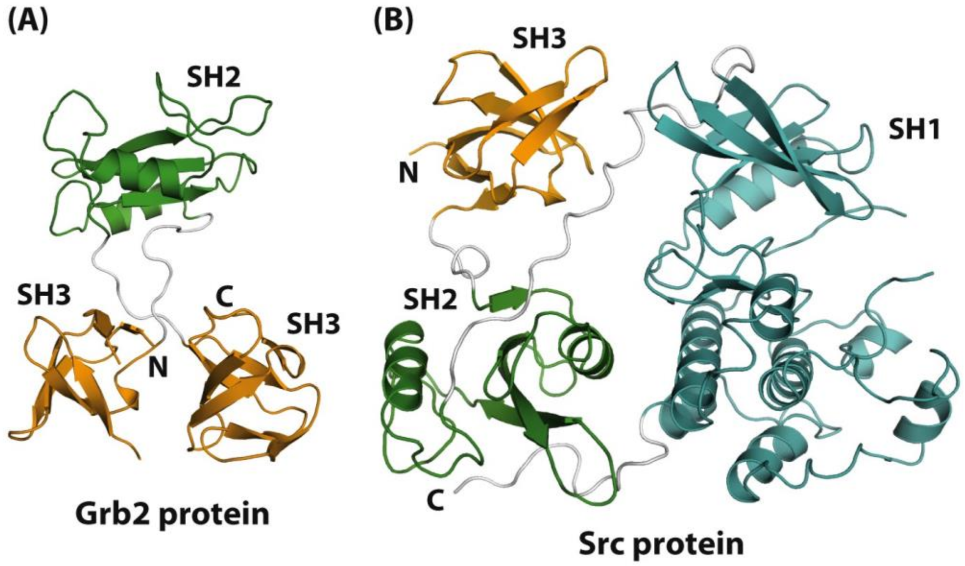


Figure 4.6 The Molecules of Life (© Garland Science 2013)

- Domains are often units of function and proteins may comprise a single domain or several
- They are typically connected by flexible loops (random coil)

Quaternary structure

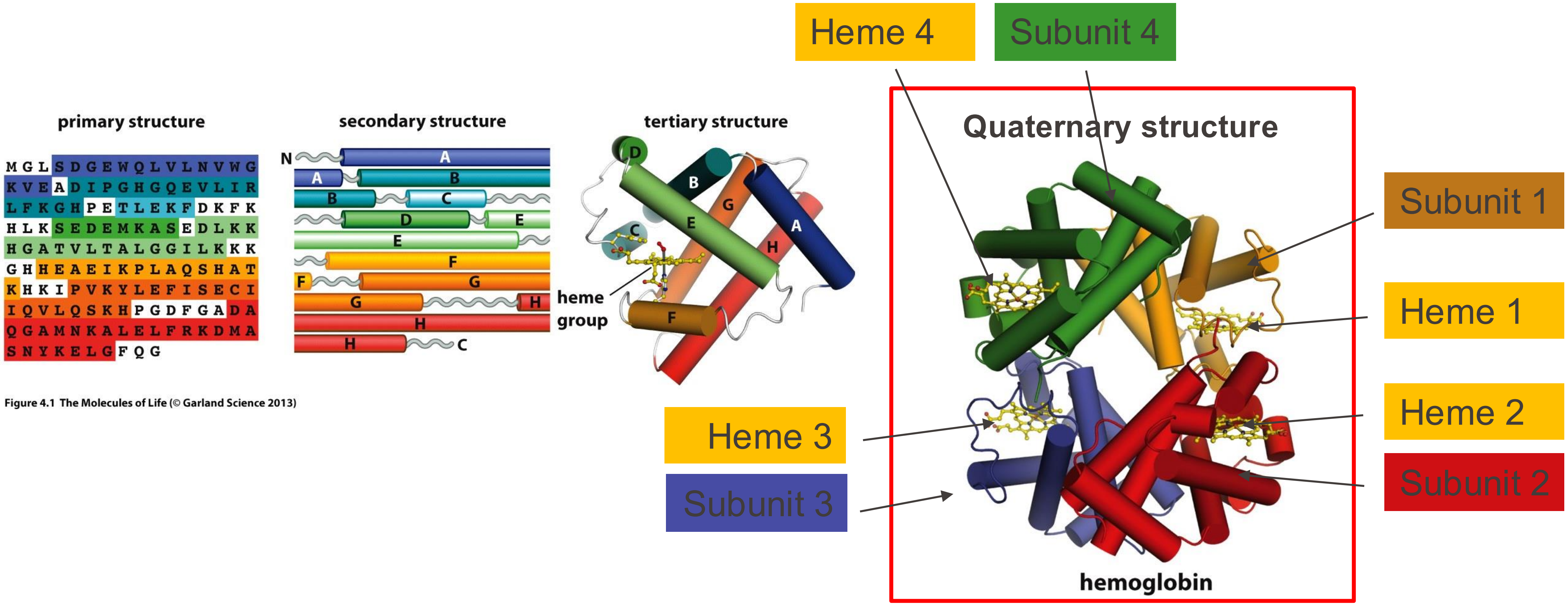


Figure 4.1 The Molecules of Life (© Garland Science 2013)

- 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.

How do proteins fold into functional conformation?

MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS
TPDDHLVARWCATPVAEVAEKSARF
PPRPEELLLPREETLPEMYSLSFTA
MNKITTRSPLEPEYQPLGKPHHDLQ
GQKGDGLRAHAPLAATFQPGREVGL
DRVESIINALMPLAPFLEGVTCETG
VQSLNPAADGAEVMIWSVGRDTLAS

Primary sequence

↓

Levinthal paradox (1969)

100 residue-long peptide

~10^{94} torsional degrees of freedom

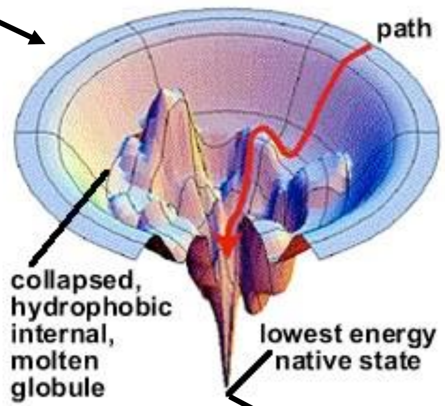
? Why is this a paradox ?

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

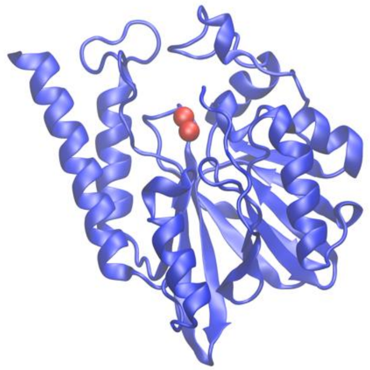
Anfinsen dogma (1954):

Protein structure is determined by its sequence

Folding pathway



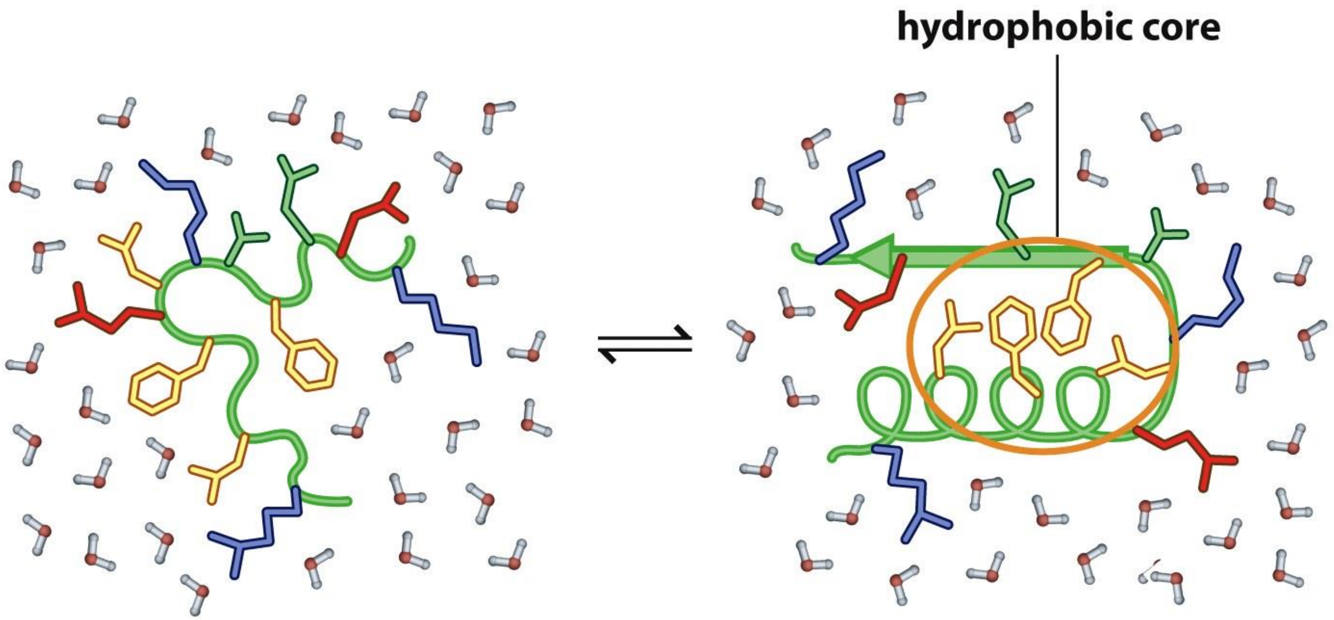
Dill@UCSF

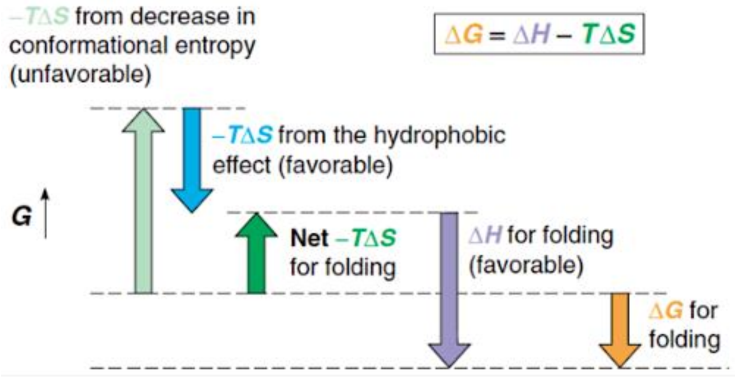


Folded native structure

Protein folding is driven by energy minimization

- The **thermodynamic hypothesis** (also known as *Anfinsen's hypothesis*): Protein folding is governed by the principle that, given the proper environment, a polypeptide will spontaneously adopt the conformation that minimizes its Gibbs free energy (**G**).
- Protein folding is a subtle free energy optimization exercise, $\Delta G = \Delta H - T\Delta S$



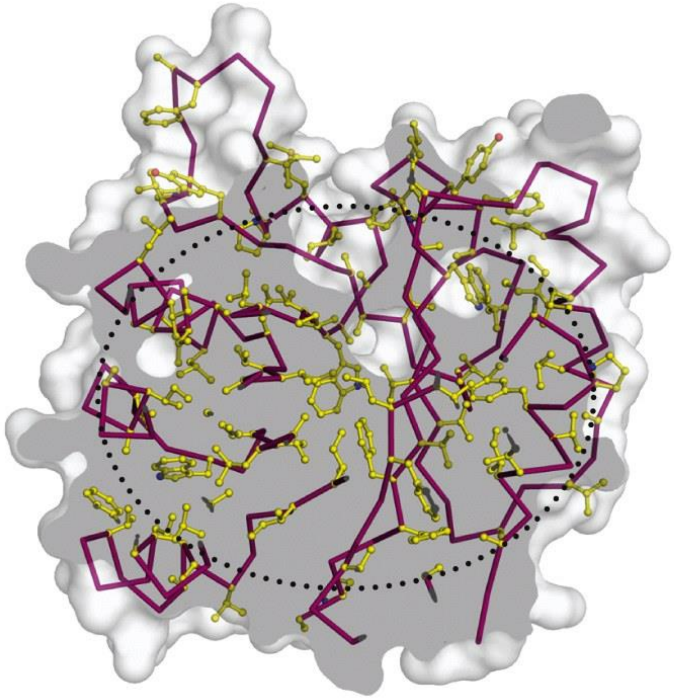


ΔG - Change in Gibbs free energy
ΔH - Change in enthalpy
ΔS - Change in entropy

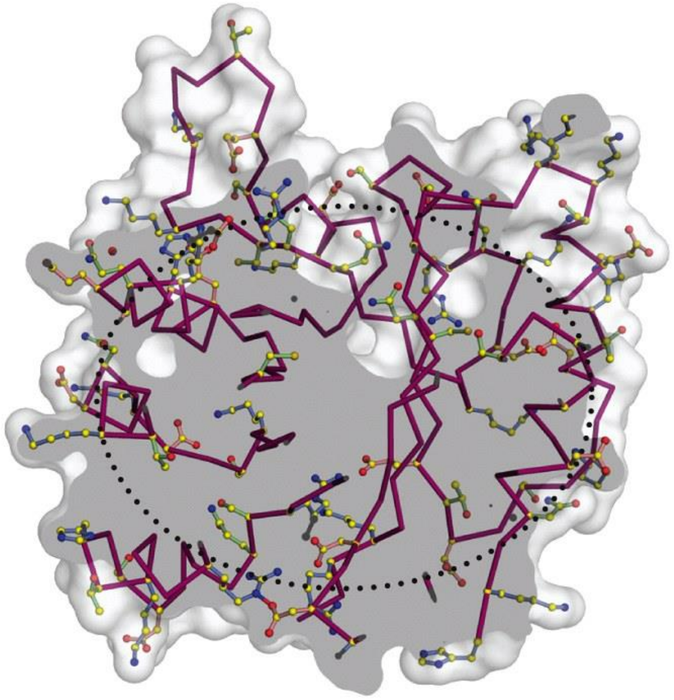
- The stability of the folded structure results primarily from the hydrophobic side chains clustering together away from the water (**the hydrophobic effect**)
- Main-chain bond restraints and non-covalent interactions further drive the process to (i) allow for optimal packing without clashes and (ii) maximize attractive interactions

Proteins have hydrophobic cores

- Proteins orient the hydrophobic residues towards the inside of the molecule and hydrophilic residues towards the solvent
- Note the differences in orientations of different amino acid residues in the central slice of a globular protein shown below



Hydrophobic residues
(yellow)



Hydrophilic residues
(green)

Transmembrane proteins

- Transmembrane proteins have their membrane-spanning domains coated in hydrophobic amino acids

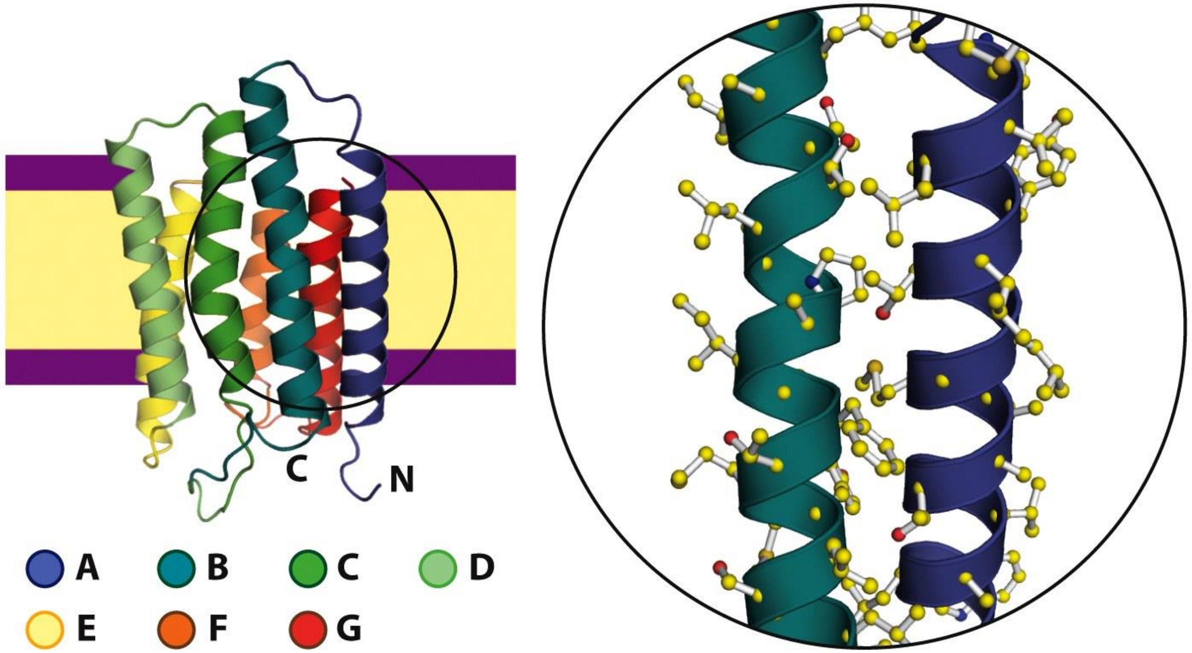


Figure 4.70 The Molecules of Life (© Garland Science 2013)

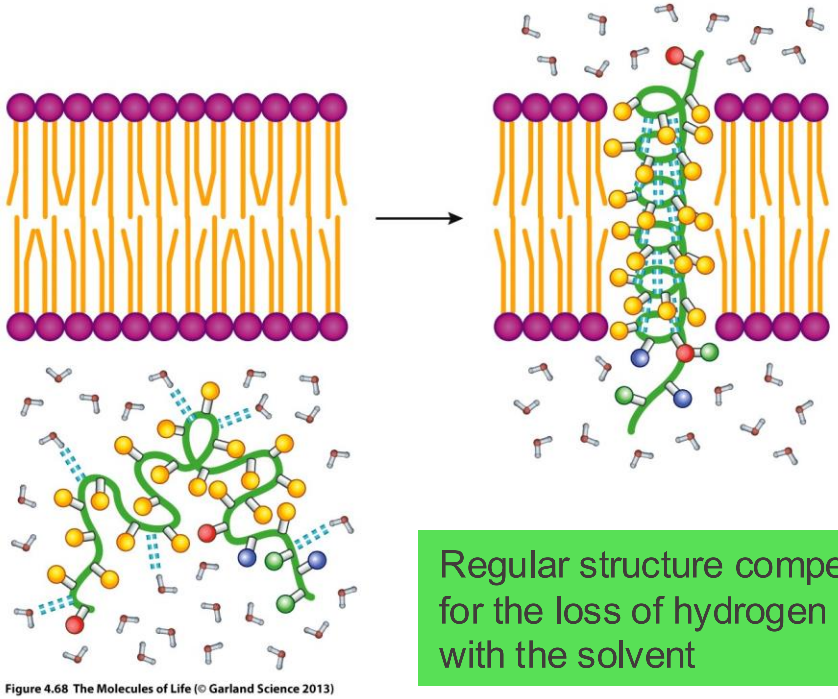


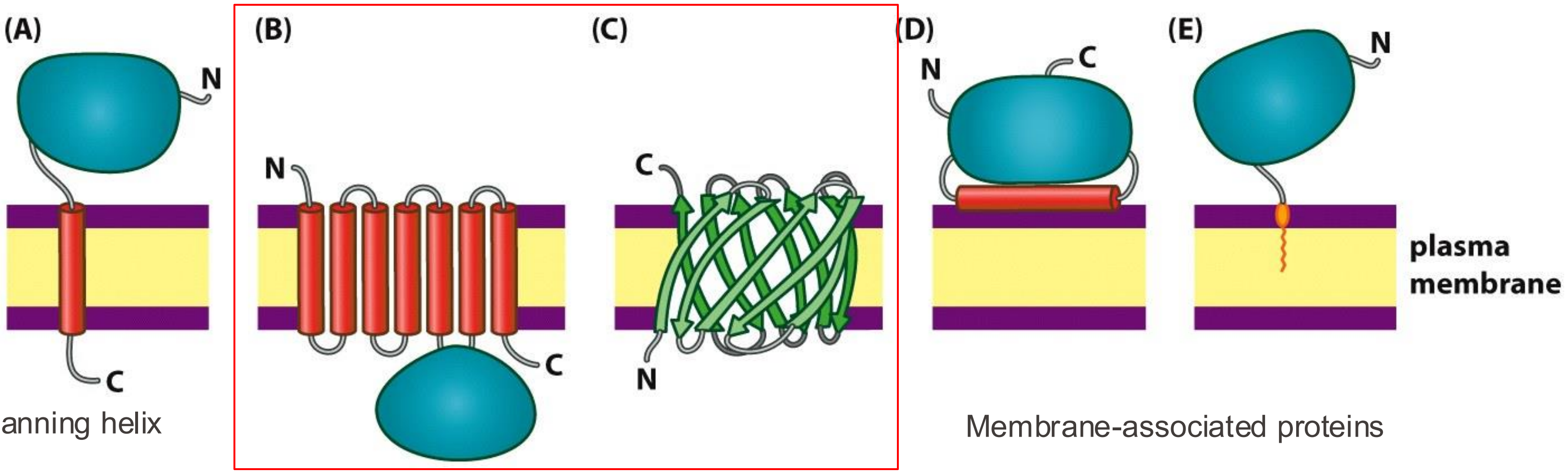
Figure 4.68 The Molecules of Life (© Garland Science 2013)

Regular structure compensates for the loss of hydrogen bonds with the solvent

- Usually, the TM domains have regular secondary structure (i.e., α-helices, β-strands) with hydrogen bonds formed between peptide elements.
- This is because of the hydrophobic environment within the lipid bilayer and the lack of hydrogen bond

Transmembrane proteins

- Proteins can be bound to the lipid membrane in different ways

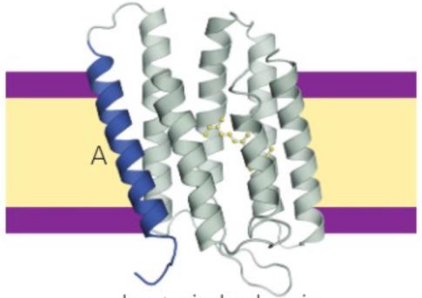


Single spanning helix

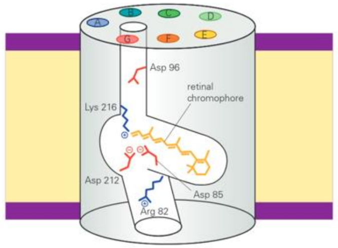
Multiple-pass transmembrane proteins

Membrane-associated proteins

- Multiple pass membrane proteins can form pores and channels for detection and transport of nutrients and/or signaling molecules from one side of the membrane to the other



bacteriorhodopsin



Bacteriorhodopsin (bR) is a **light-driven proton pump** found in the membrane of *Halobacteria*. It converts **light energy into a proton gradient**, which the cell uses to generate ATP.

Proteins - Take home message

- Proteins are formed from amino acids connected by peptide bonds which assemble during translation

- The 20 genetically encoded amino acids are classified as **hydrophilic** (charged, polar) or **hydrophobic** (non-polar) depending on the properties of their sidechains

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

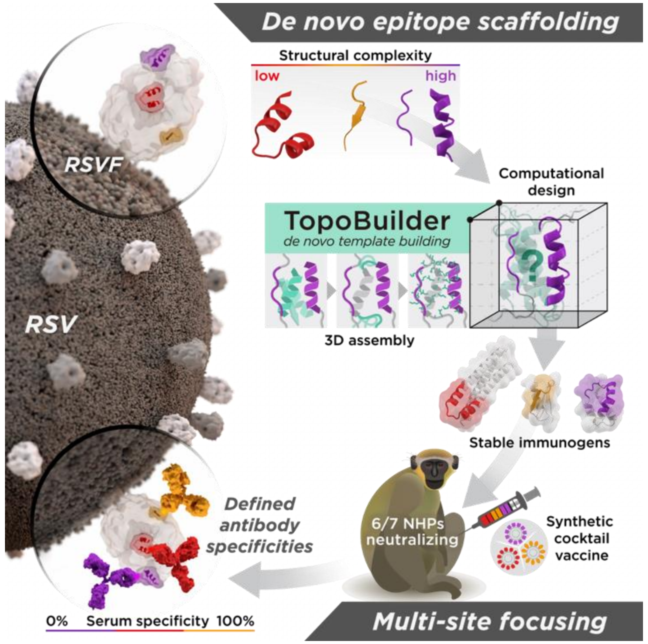
- **α-helices** and **β-sheets** are the regular structural elements on which the structures of proteins are based

- 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

Proteins in bioengineering

- Most labs at SV study proteins, their function, or their regulation on the level of genes

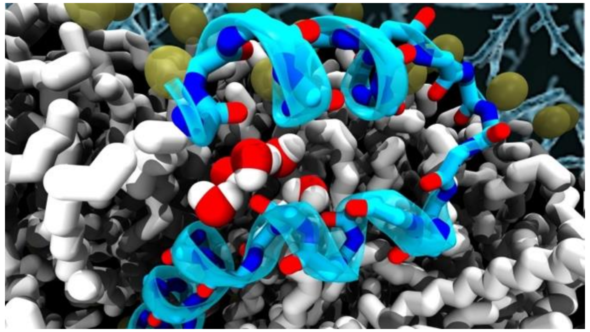
- **Protein engineering**

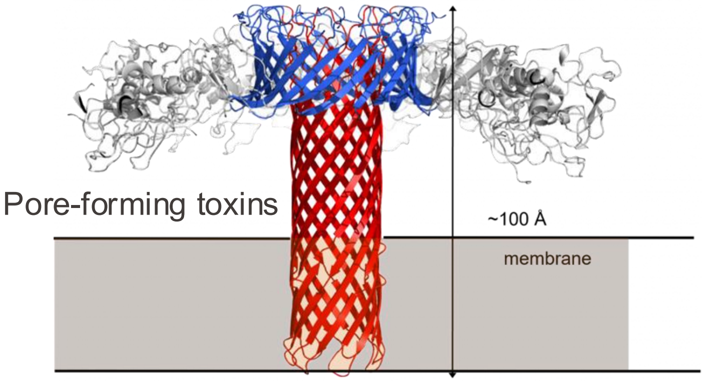


Vaccines, biologics etc.

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

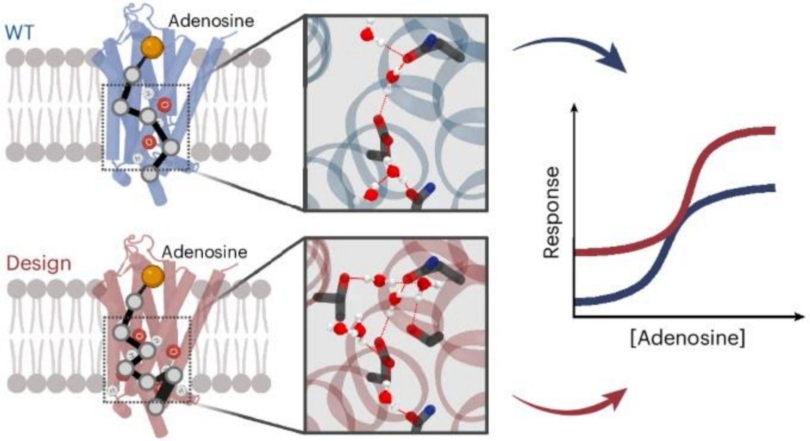
- **Protein-Protein Interactions**





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

- **Engineering cell signaling**



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

The molecules of Life

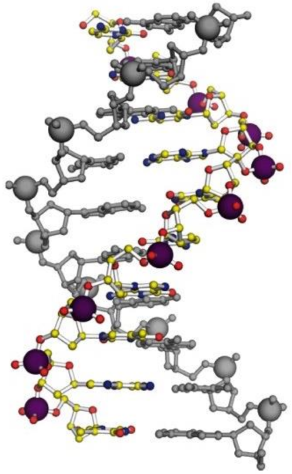
Nucleic Acids

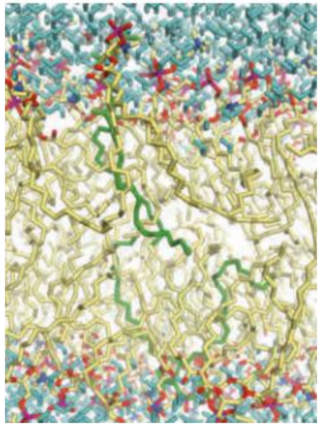
Lipids

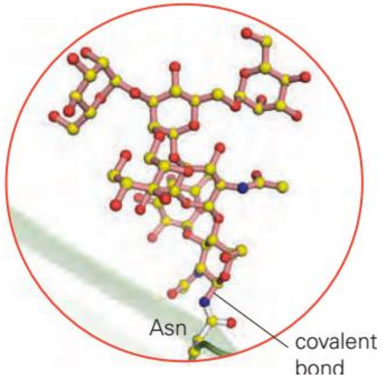
Carbohydrates

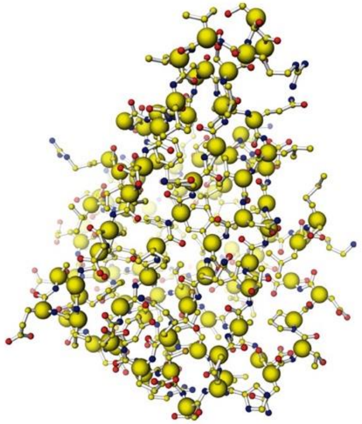
Proteins

Macromolecular Structure

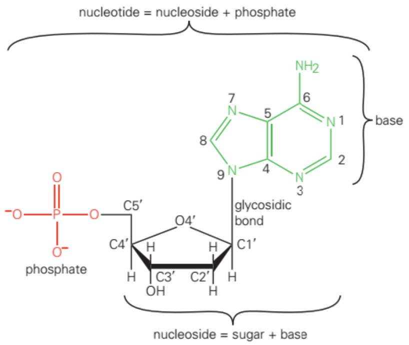


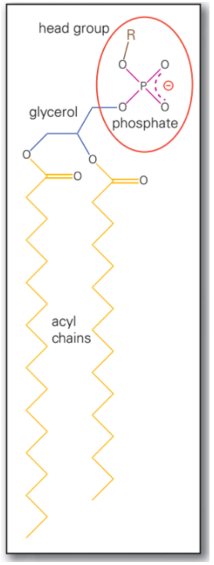


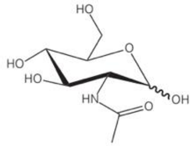




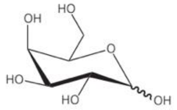
Building Block



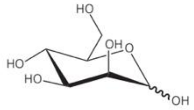




N-acetylglucosamine
(GlcNAc)



galactose
(Gal)



mannose
(Man)