Protein Biochemistry

Chem 660
Spring, 2016

The term "*protein*" first appeared in literature in 1838.

mason.gmu.edu/~bbishop1

"Proteins hold the key to the whole subject of the molecular basis of biological reactions."

Linus Pauling. "Signs of Life." *Electronic Medical Digest*, 35-36. 1949.

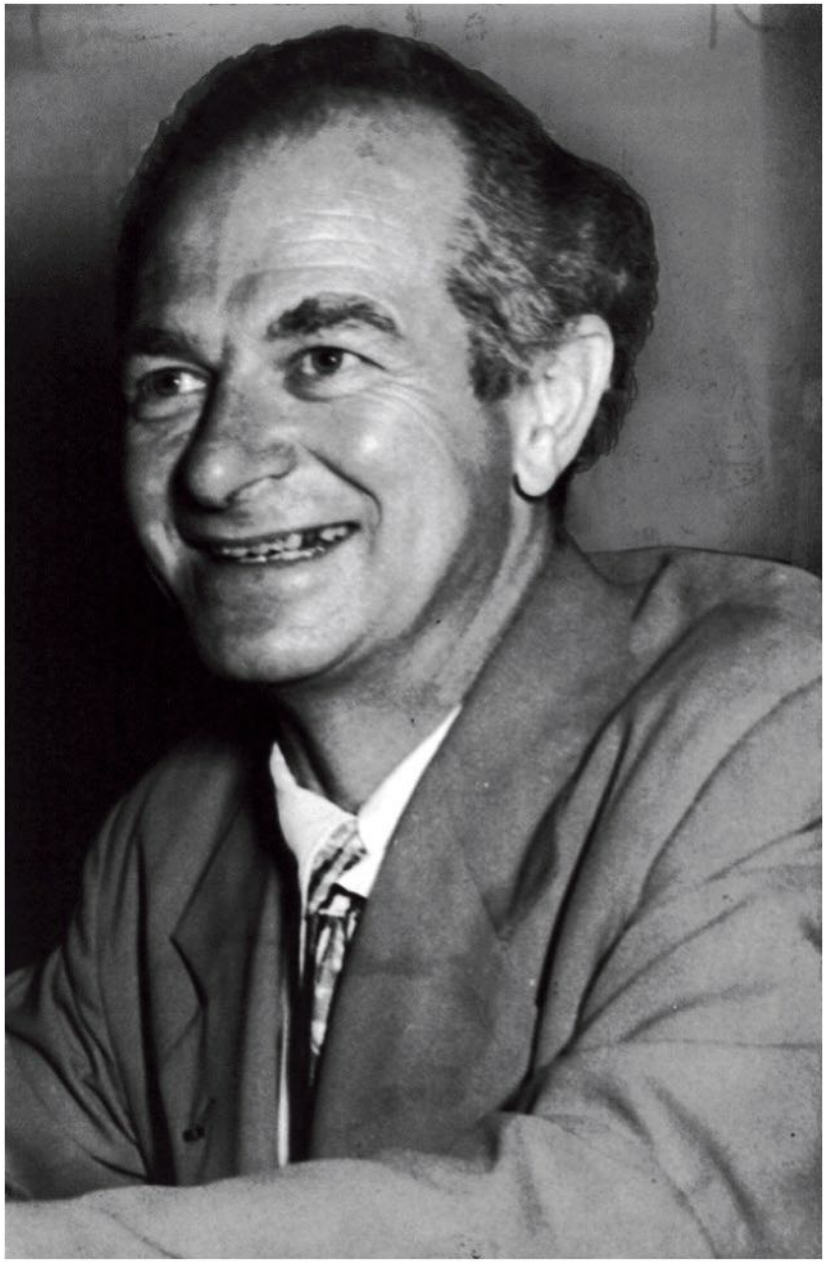
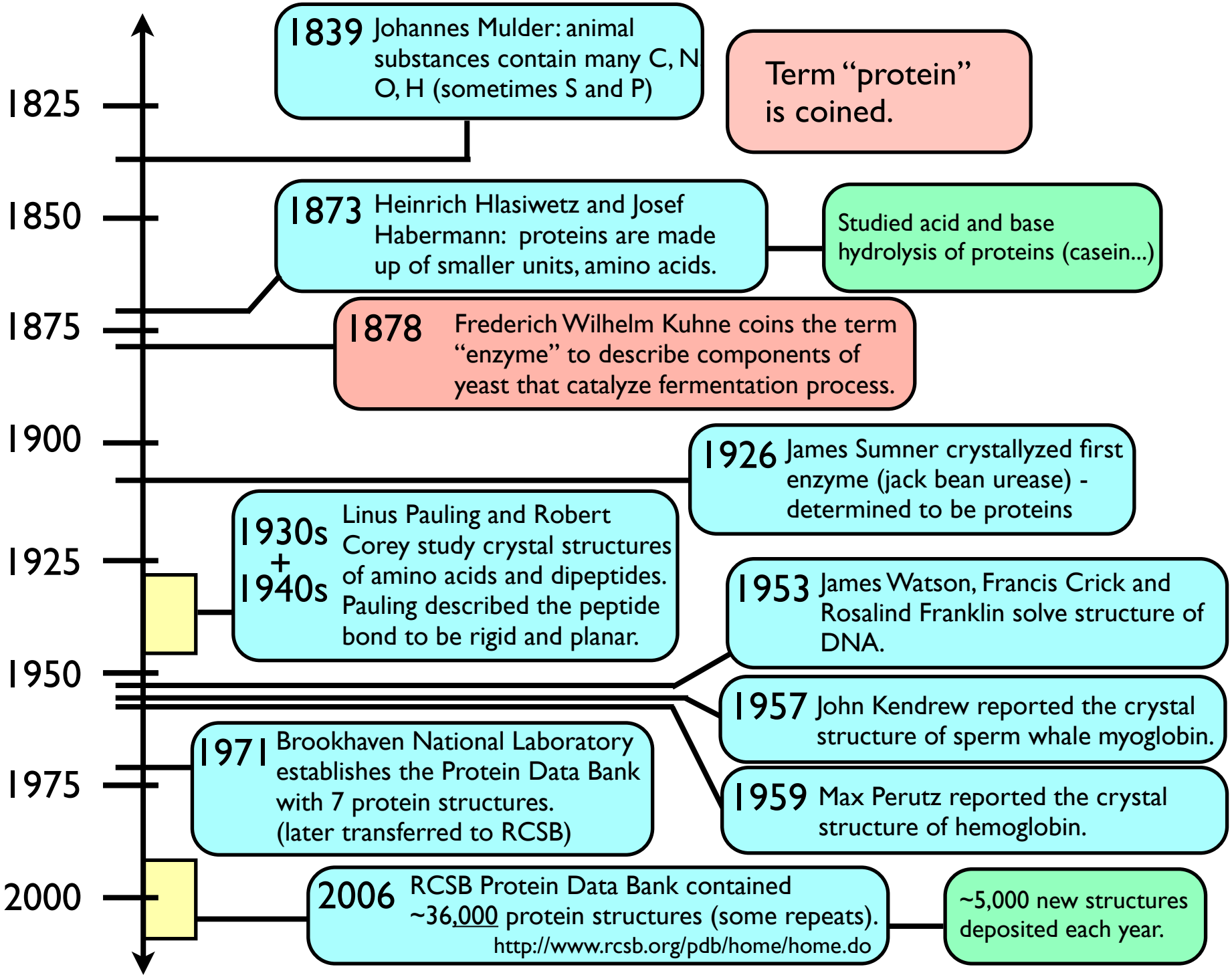


Figure 1.8.1 How Proteins Work (©2012 Garland Science)


1825
1850
1875
1900
1925
1950
1975
2000
1839 Johannes Mulder: animal substances contain many C, N O, H (sometimes S and P)
Term "protein" is coined.
1873 Heinrich Hlasiwetz and Josef Habermann: proteins are made up of smaller units, amino acids.
Studied acid and base hydrolysis of proteins (casein...)
1878 Frederich Wilhelm Kuhne coins the term "enzyme" to describe components of yeast that catalyze fermentation process.
1926 James Sumner crystallyzed first enzyme (jack bean urease) - determined to be proteins
1930s + 1940s Linus Pauling and Robert Corey study crystal structures of amino acids and dipeptides. Pauling described the peptide bond to be rigid and planar.
1953 James Watson, Francis Crick and Rosalind Franklin solve structure of DNA.
1957 John Kendrew reported the crystal structure of sperm whale myoglobin.
1959 Max Perutz reported the crystal structure of hemoglobin.
1971 Brookhaven National Laboratory establishes the Protein Data Bank with 7 protein structures. (later transferred to RCSB)
2006 RCSB Protein Data Bank contained ~36,000 protein structures (some repeats). http://www.rcsb.org/pdb/home/home.do
~5,000 new structures deposited each year.

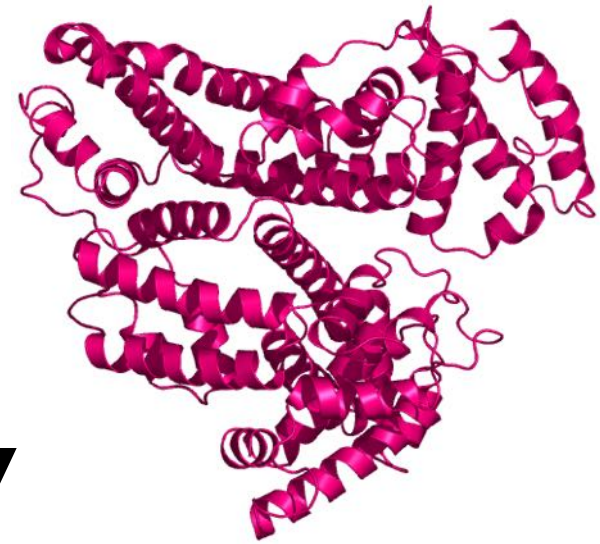
Proteins

- Proteins are involved in almost every process in living organisms.
- The diversity of cellular processes reflects the complexity and versatility of proteins.
- Each protein is usually tailored to a specific function or group of functions.
- Constitute more than 50% of the dry weight of cells.
- More abundant than any other biomolecule.

Lecture	Topic	Reading
Week 1 (Jan. 18)	No Class	
Week 2 (Jan. 25)	Snow Day	
Week 3 (Feb. 1)	Introduction and Amino Acids	
Week 4 (Feb. 8)	Protein Structure	
Week 5 (Feb. 15)	Protein Structure	
Week 6 (Feb. 22)	Protein Structure / Oligomers	
Week 7 (Feb. 29)	Test 1	Lectures Weeks 3-6
Week 8 (Mar. 7)	No Class (Spring Break)	
Week 9 (Mar. 14)	Protein Interactions	Writing assignment
Week 10 (Mar. 21)	Enzymes	
Week 11 (Mar. 28)	Protein Flexibility and Dynamics	
Week 12 (Apr. 4)	Protein Complexes	Writing assignment due
Week 13 (Apr. 11)	Protein Biosynthesis, Posttranslational Modification	Handouts/Papers Papers TBA
Week 14 (Apr. 18)	Test 2	Lectures Weeks 9-13
Week 15 (Apr. 25)	Student Presentations	NA
Week 16 (May 2)	Student Presentations	NA
May 9, 4:30-6:30 pm	Final Exam	

Platypus Venom
42 amino acids
5,100 Da

GM-CSF
127 amino acids,
14,500 Da

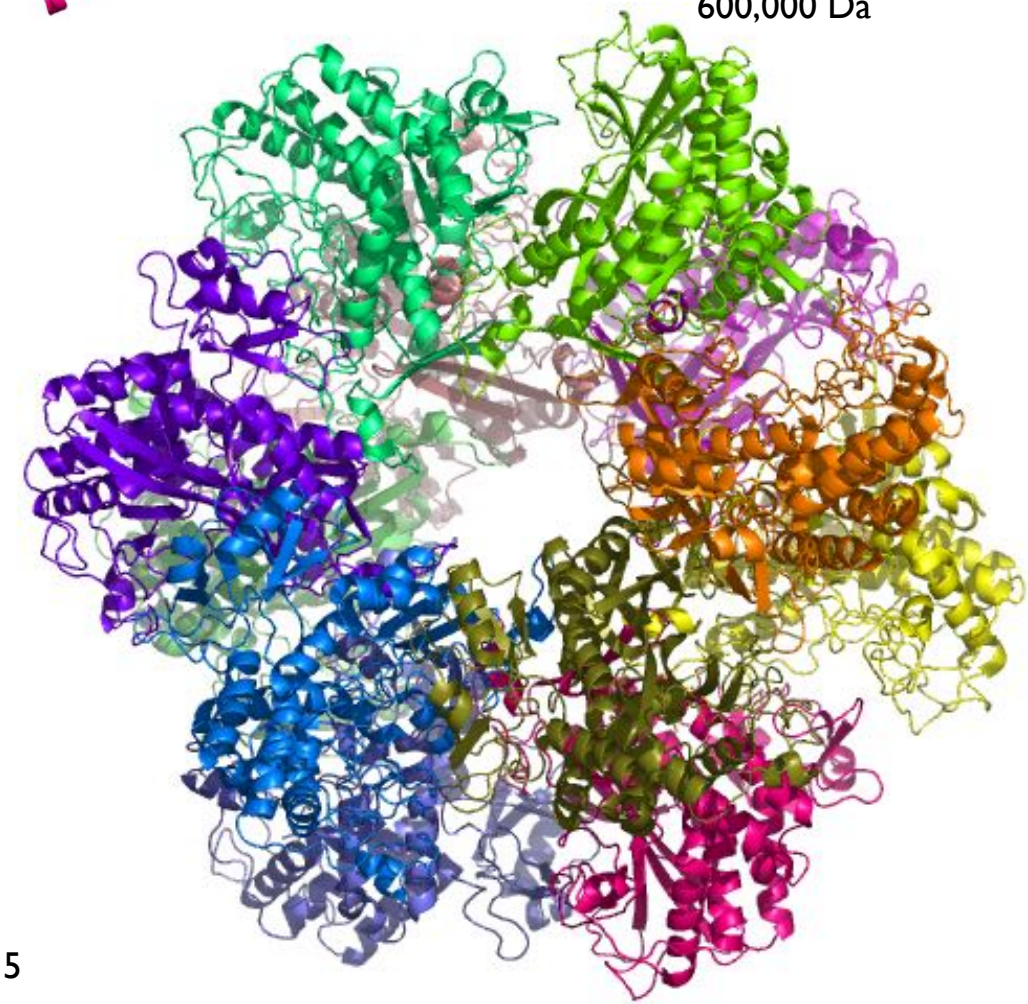


Serum Albumin
550 amino acids,
68,500 Da

Complexity

- Proteins are a complex family of molecules.
- Proteins can range in size from under 100 to around 2,000 amino acid residues.
- Some proteins are monomeric others can form complex multimeric assemblies.

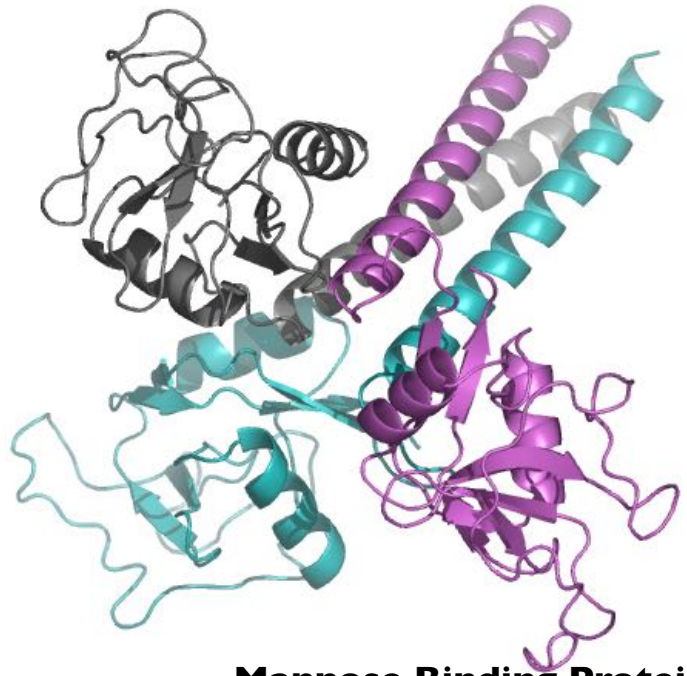
Glutamine Synthetase
12 polypeptide chains, 468 amino acids per chain, total molecular weight 600,000 Da



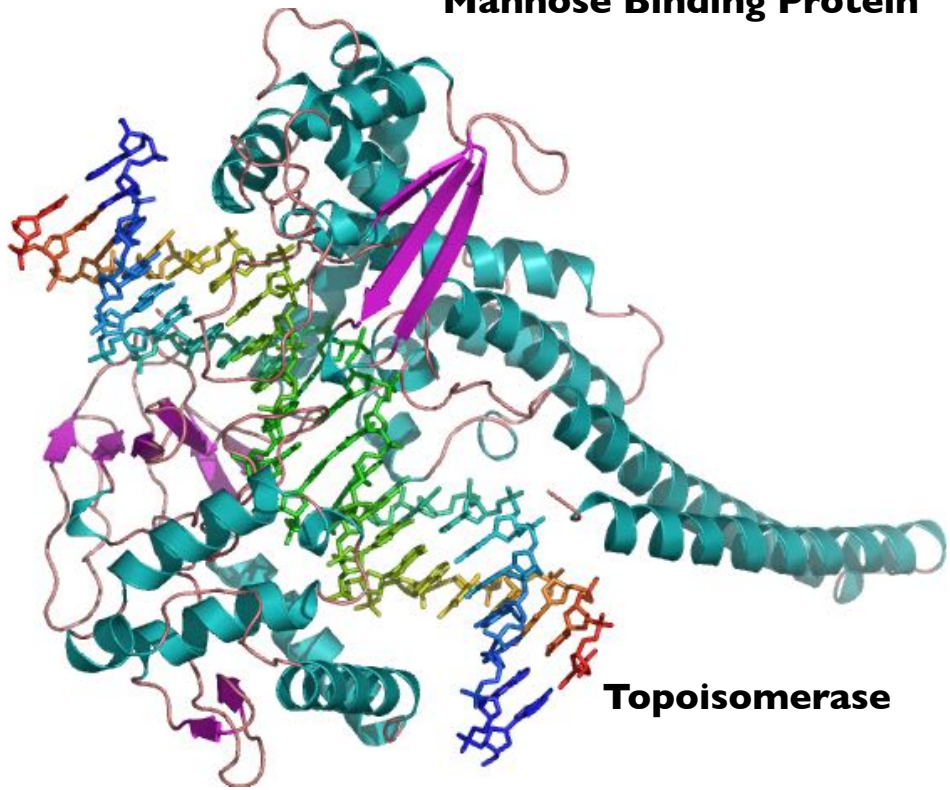
Interactions

Proteins can bind and interact with a broad spectrum of small molecules and macromolecules.

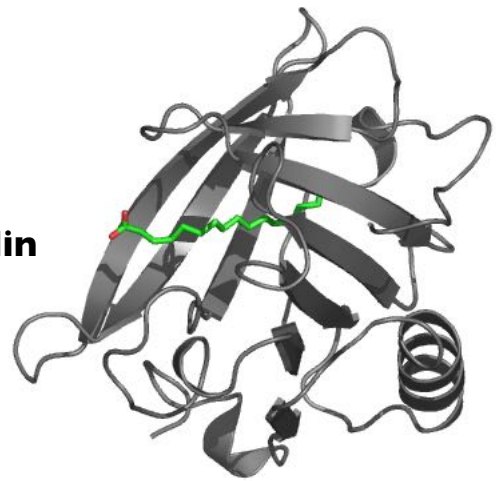
- Polypeptides and amino acids.
- Nucleic acids and nucleotides.
- Membranes and lipids.
- Metal ions.
- Other small molecules and ions.



Mannose Binding Protein



Topoisomerase



Lipocalin

Trace Elements and Protein Structure and Function

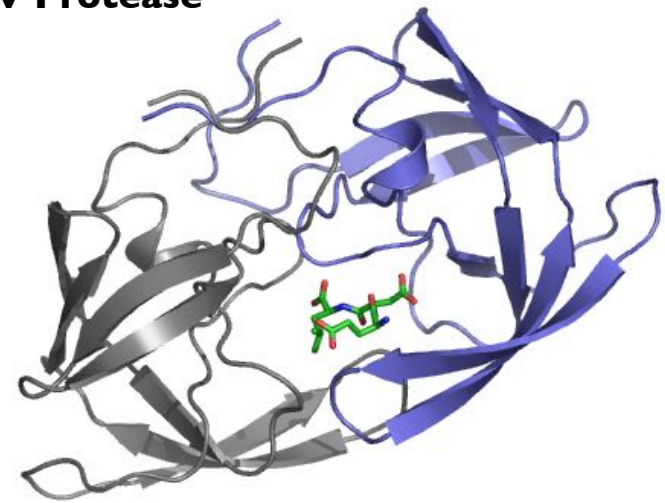
Element	Functional Role
Sodium (Na^+)	Principal intracellular ion, osmotic balance.
Potassium	Principal intracellular ion, osmotic balance.
Magnesium	Bound to ATP/GTP in nucleotide binding proteins, found as a structural component of hydrolases and isomerases.
Calcium	Activator of calcium binding proteins such as calmodulin.
Vanadium	Bound to enzymes such as chloroperoxidase.
Manganese	Bound to pterin co-factor in enzymes such as xanthine oxidase or sulphite oxidase. Also found in nitrogenase as component of water splitting enzyme.
Iron	Important catalytic component of heme enzymes inolved in oxygen transport and electron transfer (i.e. hemoglobin, cytochrom oxidase and catalase).
Cobalt	Metal component of vitamin B_{12} found in many enzymes.
Nickel	Co-factor found in hydrogenase enzymes.
Copper	Involved as co-factor in oxygen transport system and electron transport proteins (i.e. hemocyanin and plastocyanin).
Zinc	Catalytic component of enzymes such as carbonic anhydrase and superoxide dismutase
Chlorine	Principal intracellular anion, osmotic balance.
Iodine	Iodination of tyrosine residues form part of hormones thyroxine and liothyronine
Selenium	Found in active site of glutathione peroxidase

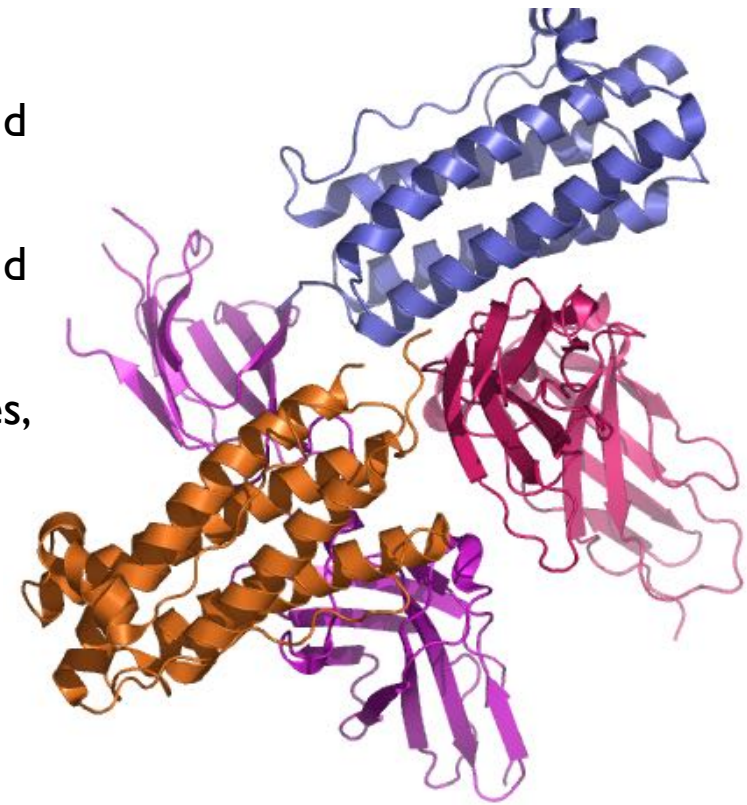
Diverse Function

Proteins:

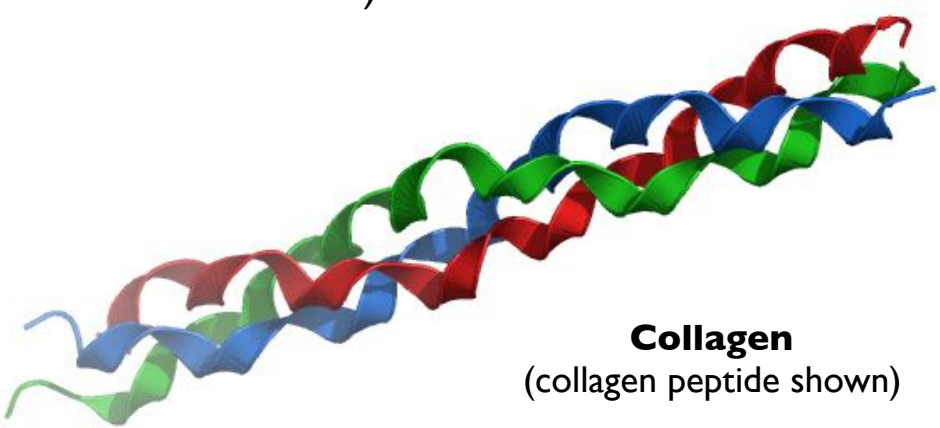
- Enzymes or catalytic proteins (i.e. trypsin, DNA polymerases and ligases).
- Contractile proteins (i.e. actin, myosin tubulin and dynein).
- Structural or cytoskeletal proteins (i.e. collagen and keratin).
- Transport proteins (i.e. hemoglobin, myoglobin, serum albumin and transthyretin).
- Effector proteins (i.e. cytokines, chemokynes, receptors and other hormones).
- Receptors (CD4, acetylcholine receptor and cytokine and chemokyne receptors).
- Control gene expression (histones, repressors, polymerases, ribosomes... etc.).

HIV Protease





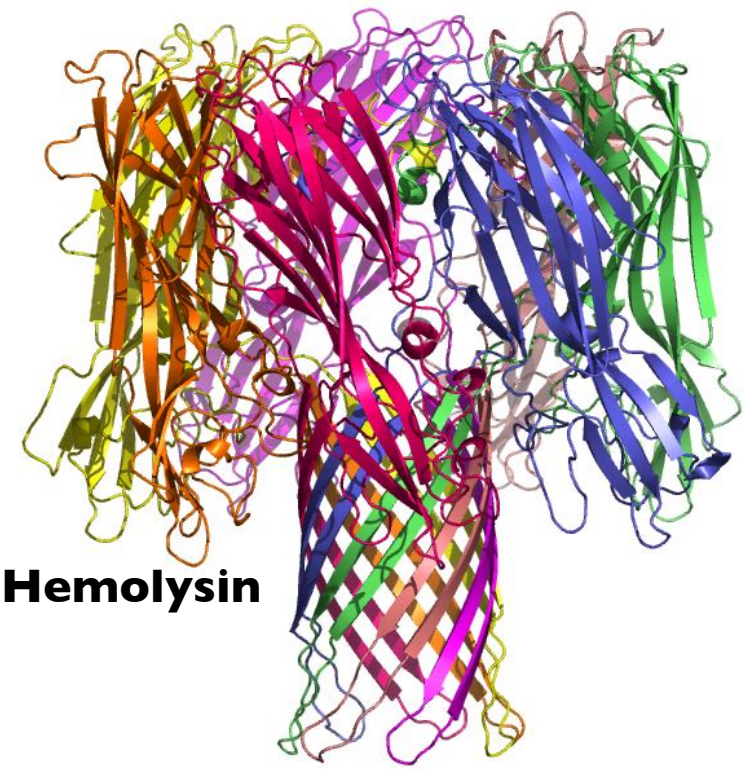
G-CSF and cytokine-binding domains of G-CSF receptor



Collagen
(collagen peptide shown)

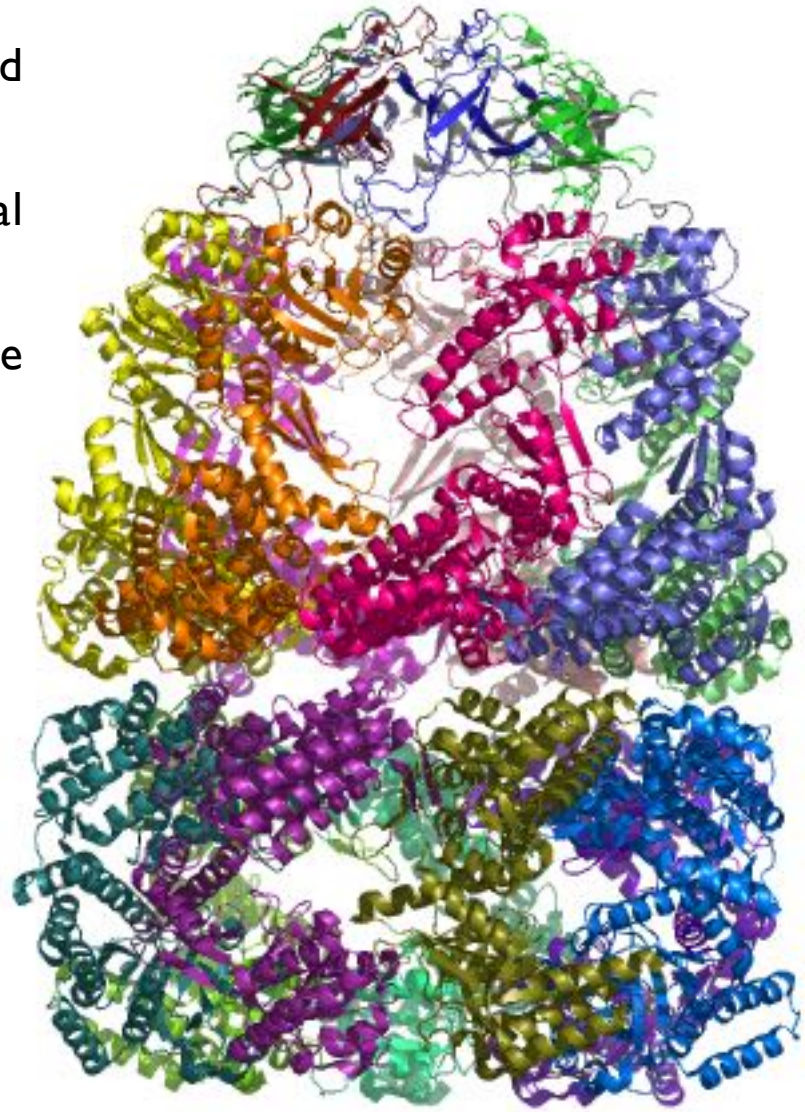
Diverse Function

- Chaperones - folding accessory proteins (i.e. GroEL, and DnaK).
- Electron transfer (i.e. Cytochrome oxidase, bacterial photosynthetic reaction center and ferredoxin)
- Active components of immunity (antibodies, cell-surface receptors, and defensins.)
- Toxins and venoms.
- Storage Proteins (i.e. ferritin and gliadin)



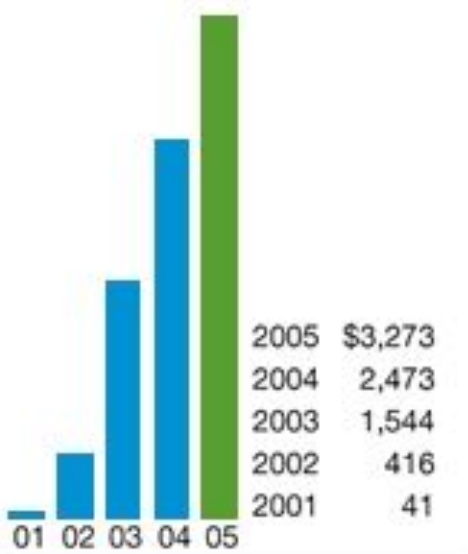
α-Hemolysin

Scorpion Toxin

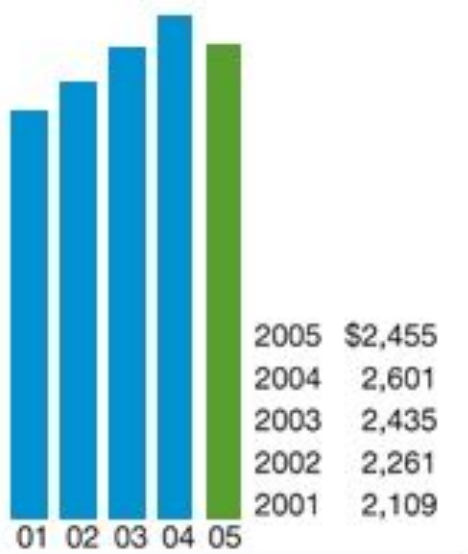


GroEL-GroES

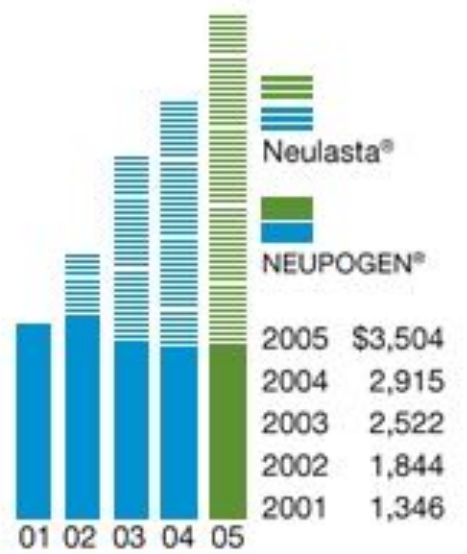
Protein Therapeutics



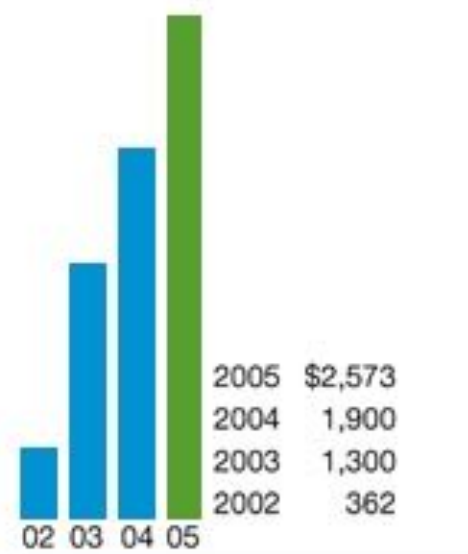
Aranesp® product sales
($ in millions)



EPOGEN® product sales
($ in millions)



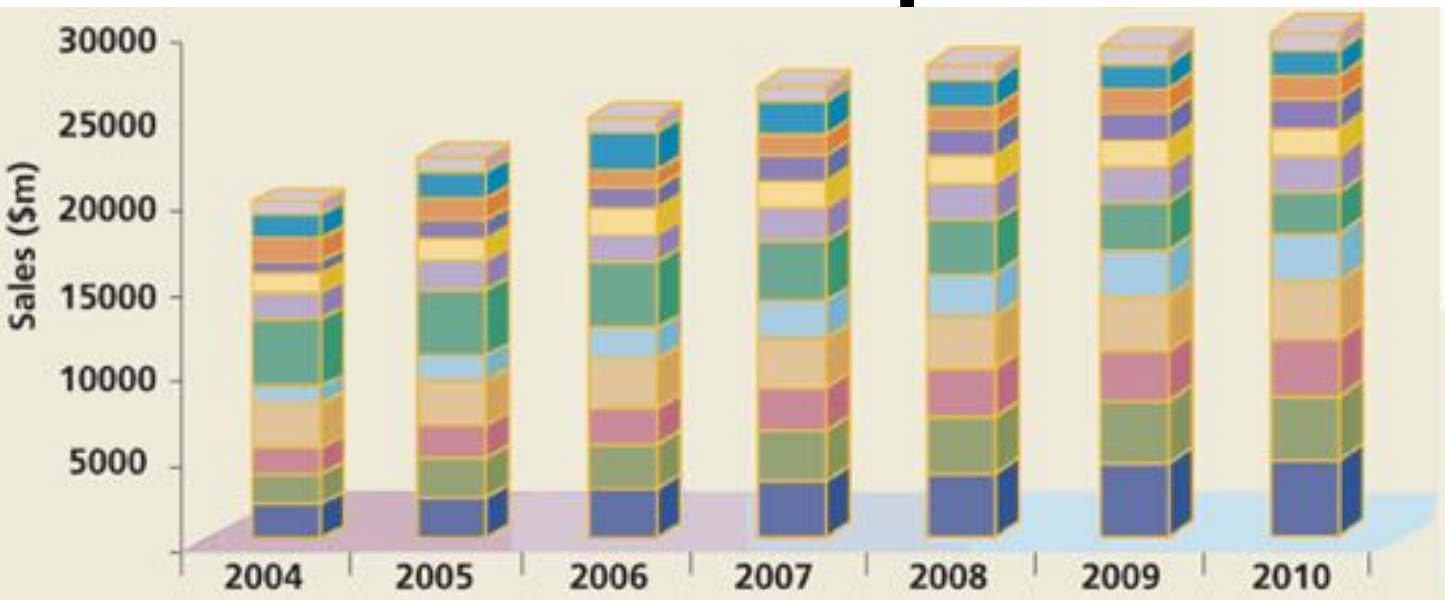
Neulasta®/NEUPOGEN® product sales
($ in millions)

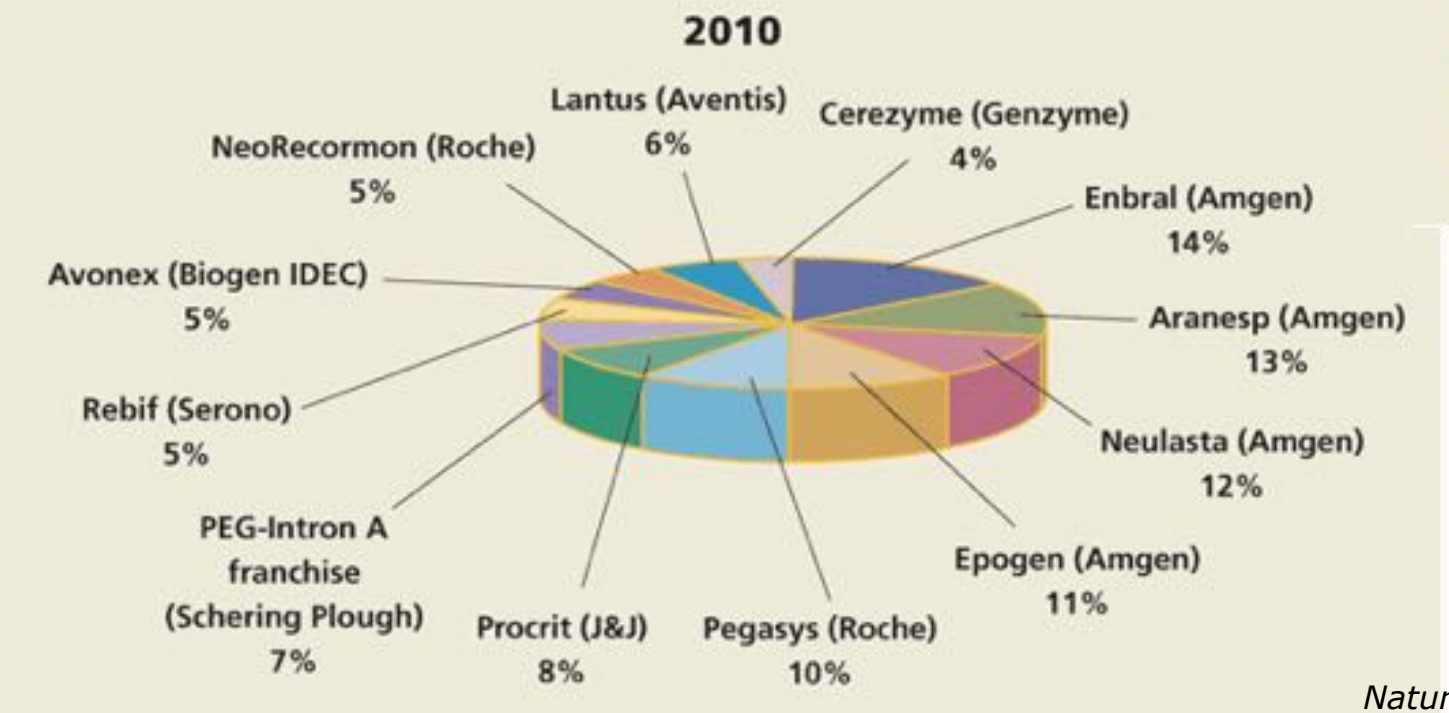


ENBREL product sales*
($ in millions)

Amgen 2005 Annual Report:
http://www.amgen.com/investors/AnnualReport2005/financials_review.html

Protein Therapeutics: Projections





Nature Biotechnology **22**, 1513 - 1519 (2004)

Protein Therapeutics: 2012

Table 1. Top 20 best-selling drugs in 2012 (modified from [6])

Rank	Drug	Company	Small Molecule/Biologic	Sales 2011 [MUSD]	Sales 2012 [MUSD]	Change* [%]
1	Humira	AbbVie	Biologic	7932	9265	19.3
2	Remicade	Johnson and Johnson	Biologic	8159	8215	0.7
3	Enbrel	Amgen and Pfizer	Biologic	7367	7963	8
4	Advair/seretide	GSK	Small Molecule	7928	7904	-1
5	Rituxan/MabThera	Roche	Biologic	6523	7285	9
6	Lantus	Sanofi	Biologic	5249	6648	19.3
7	Herceptin	Roche	Biologic	5706	6397	11
8	Crestor	AstraZeneca	Small Molecule	6622	6253	-4
9	Avastin	Roche	Biologic	5747	6260	6
10	Cymbalta	Eli Lilly	Small Molecule	4161	4994	20
11	Plavix	Sanofi & BMS	Small Molecule	9823	5318	-45.9
12	Neulasta	Amgen	Biologic	3952	4092	3.5
13	Lycria	Pfizer	Small Molecule	3693	4158	12.6
14	Januvia	Merck&Co.	Small Molecule	3324	4086	22.9
15	Lipitor	Pfizer	Small Molecule	9577	3948	-58.8
16	Nexium	Astra Zeneca	Small Molecule	4429	3944	-10
17	Singulair	Merck&Co.	Small Molecule	5479	3853	-29.7
18	Atripla	Gilead Sciences	Small Molecule	3225	3574	10.8
19	Symbicort	AstraZeneca	Small Molecule	3148	3194	5
20	Truvada	Gilead Sciences	Small Molecule	2875	3181	10.6

Pohlscheidt, M. and Kiss, R. "Recent Advances and Trends in the Biotechnology Industry - Development and Manufacturing of Recombinant Proteins and Antibodies", Amer. Pharm. Rev., October, 2013.

Amino Acids and Chemical Properties of Polypeptides

Polymeric Nature of Proteins

- Despite the diversity of biological functions proteins (and peptides) perform, they are a relatively homogeneous class of molecules.
- Proteins are linear polymers assembled from varied combinations of 20 different amino acids.
- Unlike most synthetic polymers, proteins are assembled with absolute control of the amino acid sequence. Therefore, a specific protein will have a unique amino acid sequence.
- The linear polymeric chain of almost all natural proteins are able to assume a specific three-dimensional folded conformation.
- *The chemical and biophysical properties and biological activities of a protein arise from its amino acid sequence and the three-dimensional structure of the protein (which is determined by the amino acid sequence).*

Polymerized Amino Acids: Terms

- **Peptide:** a short chain of amino acid residues with a defined sequence. The chemical properties of the peptide generally reflect the sum of the properties of the amino acids. Usually lack defined three-dimensional structures.
- **Polypeptide:** a longer chain of amino acid residues... usually have defined sequence and length.
- **Polyamino acids:** random sequences of amino acids of varied lengths... usually result of nonspecific polymerization.
- **Protein:** term used to describe polypeptides that have a defined three-dimensional structure under physiological conditions. The folded conformation is a major factor in defining the properties of the protein.

Levels of Organization

- **Primary structure (1° structure):** the amino acid sequence of polypeptide chain.

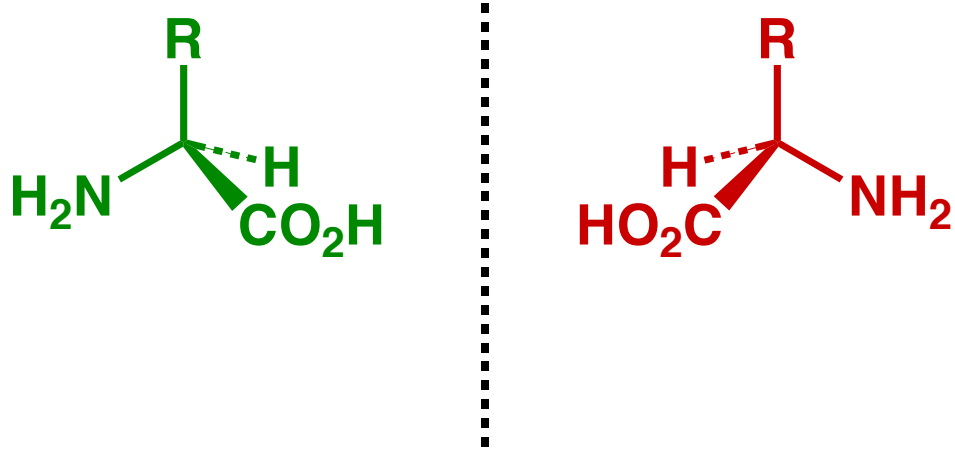
- **Secondary structure (2° structure):** local spatial organization and arrangement of the peptide backbone. Generally refers to easily localized structural elements (i.e. helices and sheets).

- **Tertiary structure (3° structure):** the comprehensive three-dimensional structure of a protein (single polypeptide chain).

- **Quaternary structure (4° structure):** assembly through noncovalent interactions) of a larger protein structure from 2 or more polypeptide chains (subunits), and the organization of these subunits.

Amino Acids: Optical Activity

- All amino acids other than glycine are optically active (chiral).
- They demonstrate an asymmetry such that their mirror images are not superimposable.
- Asymmetric centers ⇔ chiral centers.
- Enantiomers are molecules that are nonsuperimposable mirror images of each other.
- Diastereomers are stereoisomers that differ by at least one but not all asymmetric centers.



Residue	Mass (daltons)	Van der Waals Volume (Å^3)
Ala (A)	71	67
Arg (R)	156.19	148
Asn (N)	114.11	96
Asp (D)	115.09	91
Cys (C)	103.15	86
Gln (Q)	128.14	114
Glu (E)	129.12	109
Gly (G)	57.05	48
His (H)	137.14	118
Ile (I)	113.16	124
Leu (L)	113.16	124
Lys (K)	128.17	135
Met (M)	131.19	124
Phe (F)	147.18	135
Pro (P)	97.12	90
Ser (S)	87.08	73
Thr (T)	101.11	93
Trp (W)	186.21	163
Tyr (Y)	163.18	141
Val (V)	99.14	105
Weighted Avg.	119.4	161

Not All Amino Acids are Created Equal

- Each amino acid is unique and the amino acid sequence and composition of a protein contribute to its biophysical and biochemical properties.
- The chemical and physical properties of a protein are more complex than just the sum of the properties of the amino acid residues that comprise the protein.
- Not all amino acids are used with equal frequency.

TABLE 1.2 Frequency of occurrence of amino acids in proteins

Amino acid	Frequency in intracellular proteins (%)	Frequency in membrane proteins (%)	Number of codons
Ala	7.9	8.1	4
Arg	4.9	4.6	6
Asp	5.5	3.8	2
Asn	4.0	3.7	2
Cys	1.9	2.0	2
Glu	7.1	4.6	2
Gln	4.4	3.1	2
Gly	7.1	7.0	4
His	2.1	2.0	2
Ile	5.2	6.7	3
Leu	8.6	11.0	6
Lys	6.7	4.4	2
Met	2.4	2.8	1
Phe	3.9	5.6	2
Pro	5.3	4.7	4
Ser	6.6	7.3	6
Thr	5.3	5.6	4
Trp	1.2	1.8	1
Tyr	3.1	3.3	2
Val	6.8	7.7	4

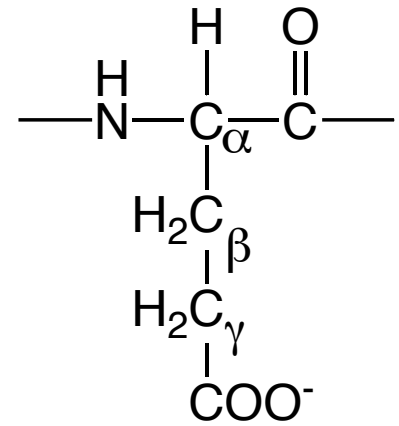
(Data taken from J. Cedano et al., *J. Mol. Biol.* 266:594–600, 1997.)

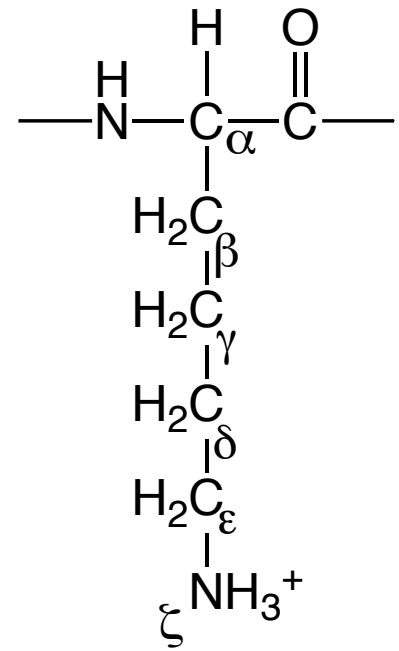
Not All Amino Acids are Created Equal

- Each amino acid is unique and the amino acid sequence and composition of a protein contribute to its biophysical and biochemical properties.
- The chemical and physical properties of a protein are more complex than just the sum of the properties of the amino acid residues that comprise the protein.
- Not all amino acids are used with equal frequency.

Amino Acids: Nomenclature

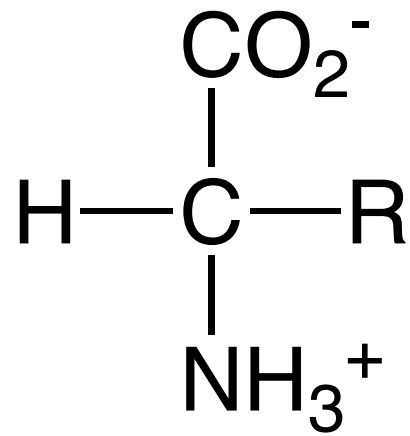
- In a peptide or protein sequence, amino acid residues are named by replacing -ine with -yl.
- Peptide sequences are written (left to right) from the amine terminus (*N*-terminus) to the carboxyl terminus (*C*-terminus).
- Glx reflects uncertainty between Glu and Gln.
- Asx reflects uncertainty between Asp and Asn.
- Position of nonhydron side chain atoms indicated by Greek alphabet (α, β, γ, δ, ε, ζ...)





Classification of Amino Acids

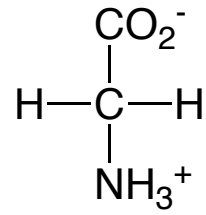
- Classification based on the properties and characteristics of the amino acid side chains.



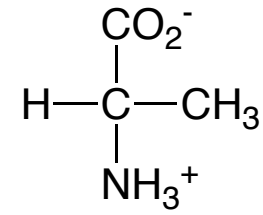
- Nonpolar.
- Uncharged Polar.
- Charged Polar.

Amino Acids: Nonpolar Side Chains

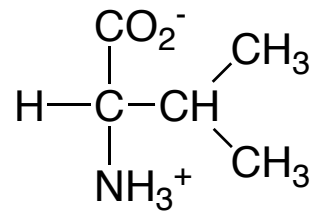
- Nine amino acids with hydrophobic side chains.
- Aliphatic side chains
- Aromatic side chains.



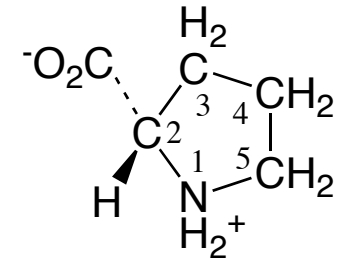
Glycine (Gly,G)



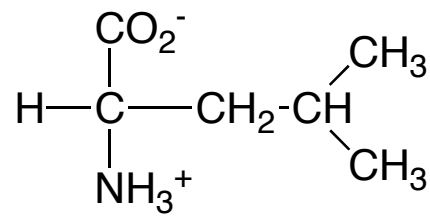
Alanine (Ala, A)



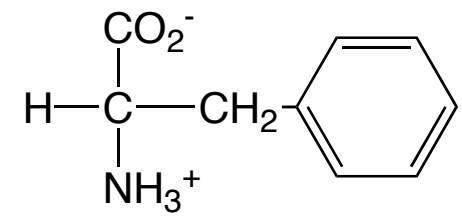
Valine (Val, V)



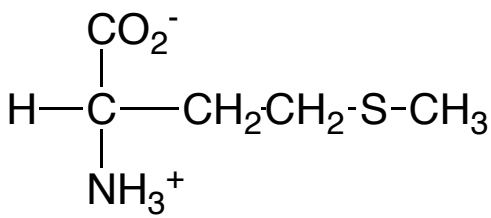
Proline (Pro, P)



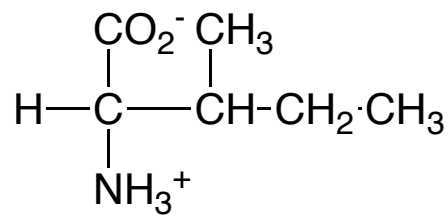
Leucine (Leu, L)



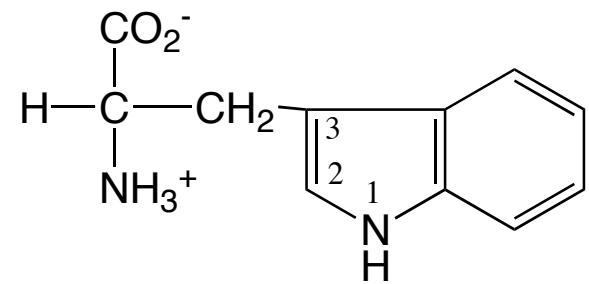
Phenylalanine (Phe, F)



Metihonine (Met, M)



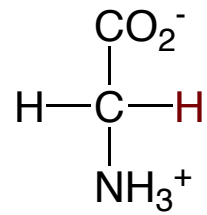
Isoleucine (Ile, I)



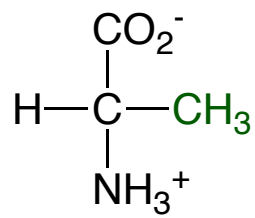
Tryptophan (Trp, W)

Gly and the Aliphatic Residues (Ala, Val, Leu and Ile)

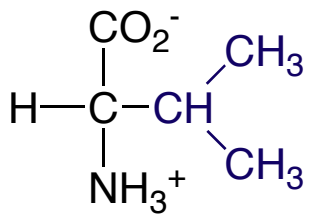
- The side chains of these amino acids are chemically inert (having no reactive functional groups.)
- Gly is the simplest amino acid, having only an H atom for a side chain. This also makes Gly the only achiral amino acid.
- Methyl group on Ala makes it the second smallest.
- The large aliphatic side chains of Val, Leu and Ile do not interact favorably with water (hydrophobic). Interact more favorably with other nonpolar groups (i.e. each other).
- Side chains of Val, Leu and Ile demonstrate varying degrees of conformational flexibility.



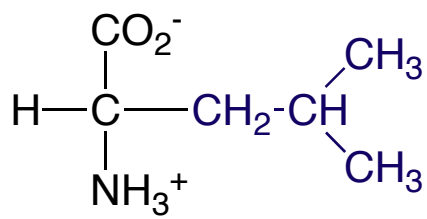
Glycine (Gly,G)



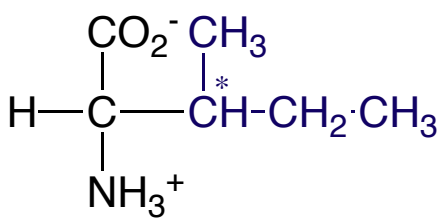
Alanine (Ala, A)



Valine (Val, V)



Leucine (Leu, L)

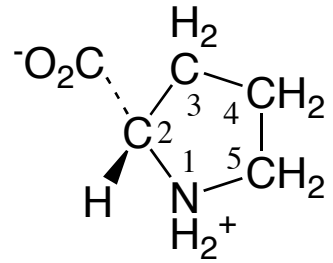


Isoleucine (Ile, I)

Proline: the Cyclic Amino Acid

- The Pro side chain is aliphatic and is covalently bonded to the backbone nitrogen, making Pro the only cyclic amino acid (five-membered ring).

- Unique arrangement results in the absence of an amide hydrogen atom present in other amino acids. Therefore, Pro residues lack an amide hydrogen for backbone hydrogen bonding.



Proline (Pro, P)

- Cyclic structure imposes conformational constraints on the peptide backbone:

  - Very limited degrees of rotation around the N-C_α bond.

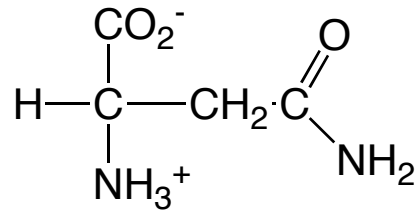
  - Preceding amide bond more likely to assume *cis* configuration.

  - The five membered pyrrolidine ring is puckered.

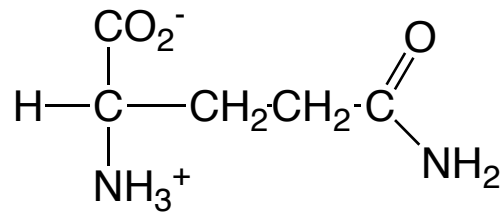
- Presence of a proline residue can disrupt/break peptide structure.

Amino Acids: Uncharged polar side chains

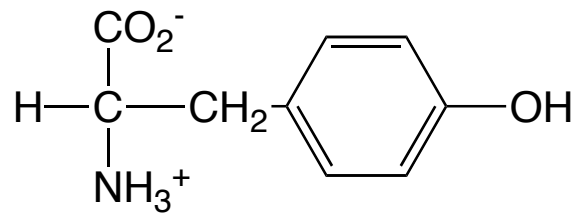
- Six amino acids with uncharged polar side chains.
- Side chain hydroxyl.
- Side chain amide.
- Side chain phenol.
- Side chain thiol.



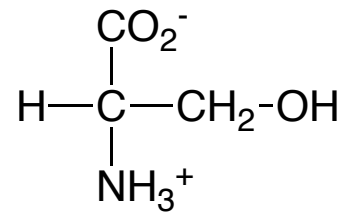
Asparagine (Asn, N)



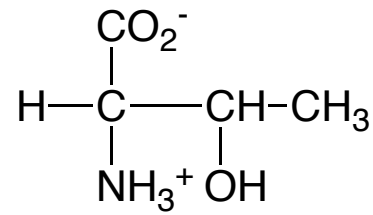
Glutamine (Gln, Q)



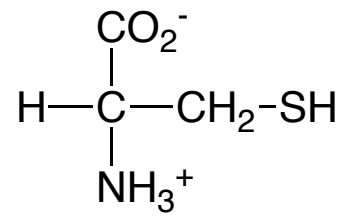
Tyrosine (Tyr, Y)



Serine (Ser, S)



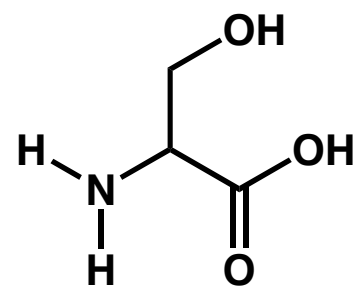
Threonine (Thr, T)



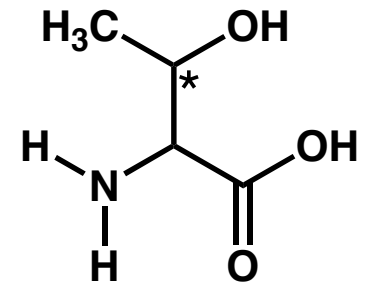
Cysteine (Cys, C)

Serine and Threonine

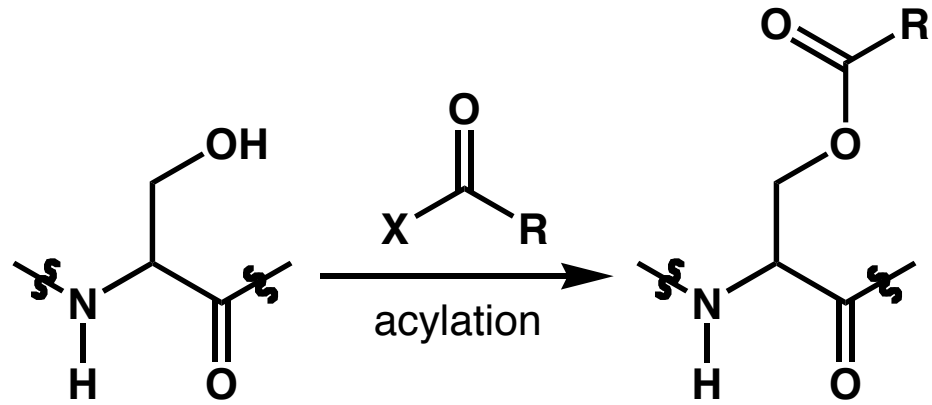
- The hydroxyl groups of serine and threonine are relatively unreactive (chemical reactivity similar to hydroxyl group of ethanol).
- Like ethanol can be acylated to form esters.
- Threonine (like Ile) has a second stereocenter at the C_β position.



Serine (Ser, S)

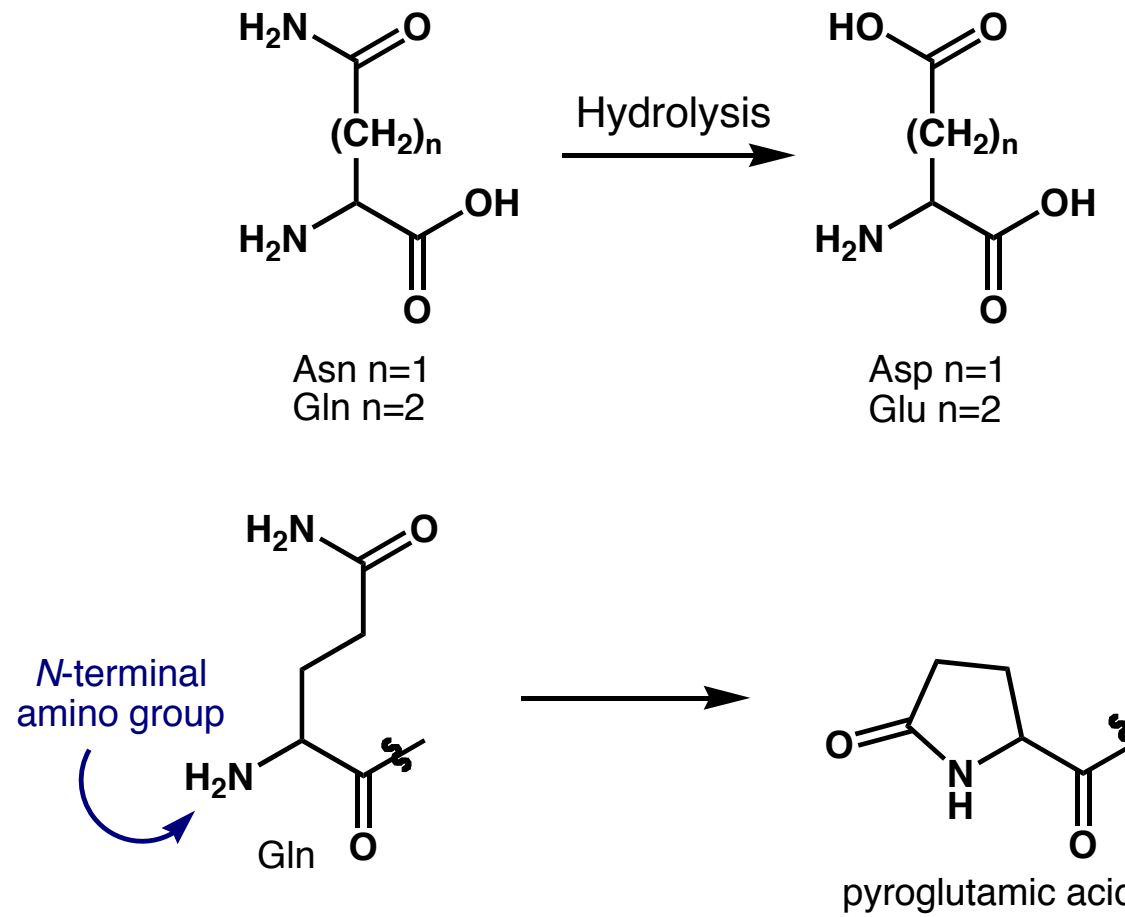


Threonine (Thr, T)



Asparagine and Glutamine

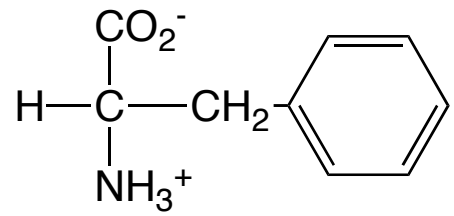
- The Gln and Asn side chain amide groups provide hydrogen bond donor and acceptor.
- Gln and Asn side chain amide groups generally unreactive.
- However the side chain amide bonds of Asn and Gln are readily hydrolyzed under extremes of pH and at high temperatures (forming Asp and Glu respectively).
- *N*-terminal Gln residues will spontaneously cyclize. The resulting pyrrolidone carboxylic acid blocks the *N*-terminus. Can be removed using pyroglutamyl amino peptidase.



Aromatic Residues (Phe, Tyr and Trp)

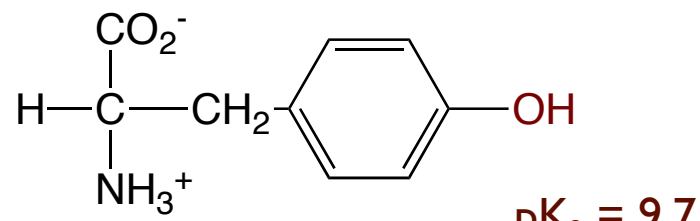
- Responsible for most of the UV absorbance and fluorescence properties of proteins. Phe, Tyr and Trp spectral properties are greatly influenced by environment.

- The hydrophobic side chain of Phe is similar to benzene or toluene and is relatively chemically inert.



Phenylalanine (Phe, F)

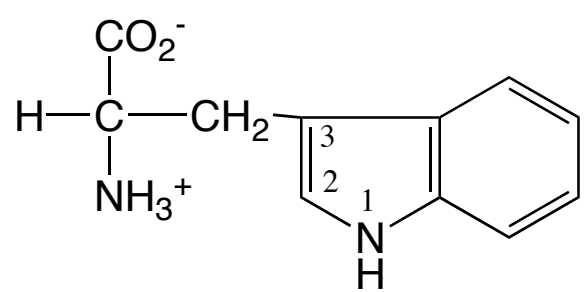
- The side chain of Tyr bears a phenolic group.
  - The hydroxyl present in the ring makes the ring relatively reactive towards electrophilic substitution reactions.
  - The hydroxyl group can be deprotonated under alkaline conditions, and can also participate in hydrogen bonding.



$pK_a = 9.7$

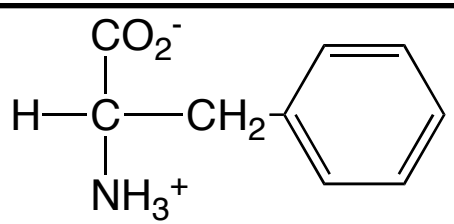
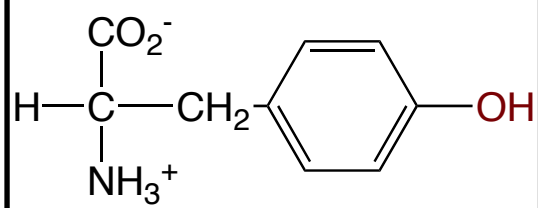
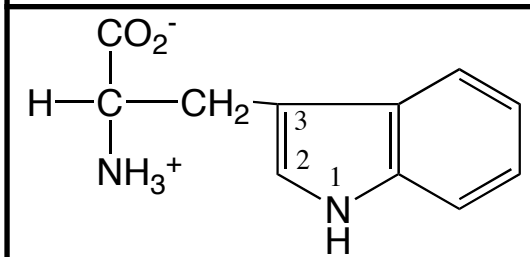
Tyrosine (Tyr, Y)

- The indole side chain of Trp is the largest and the most fluorescent. It also occurs least frequently. Trp fluorescence is very sensitive to environmental conditions.
  - The indole ring is susceptible to irreversible oxidation
  - The nitrogen in the indole group can be reversibly formylated
  - The nitrogen can also participate in hydrogen bonding as a hydrogen donor.



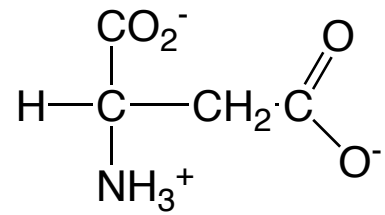
Tryptophan (Trp, W)

Spectroscopic Properties of Aromatic Amino Acids

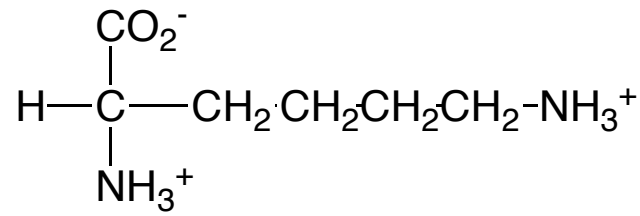
	Absorbance		Fluorescence	
Amino Acid	λ_{max} (nm)	ε ($M^{-1}cm^{-1}$)	λ_{max} (nm)	Quantum Yield
 Phenylalanine (Phe, F)	257.4	197	282	0.04
 Tyrosine (Tyr, Y)	274.6	1420	303	0.21
 Tryptophan (Trp, W)	279.8	5600	348	0.2

Amino Acids: Charged Polar Side Chains

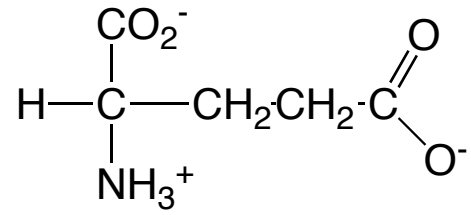
- Five amino acids have charged polar side chains.
- Side chain carboxylic acid groups.
- Side chain amine and guanidino groups.
- Side chain imidazole group.



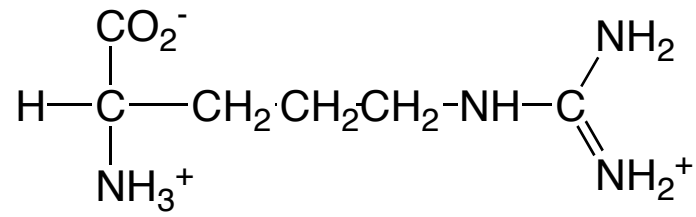
Aspartic Acid (Asp, D)



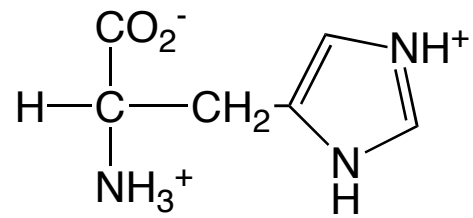
Lysine (Lys, K)



Glutamic Acid (glu, E)



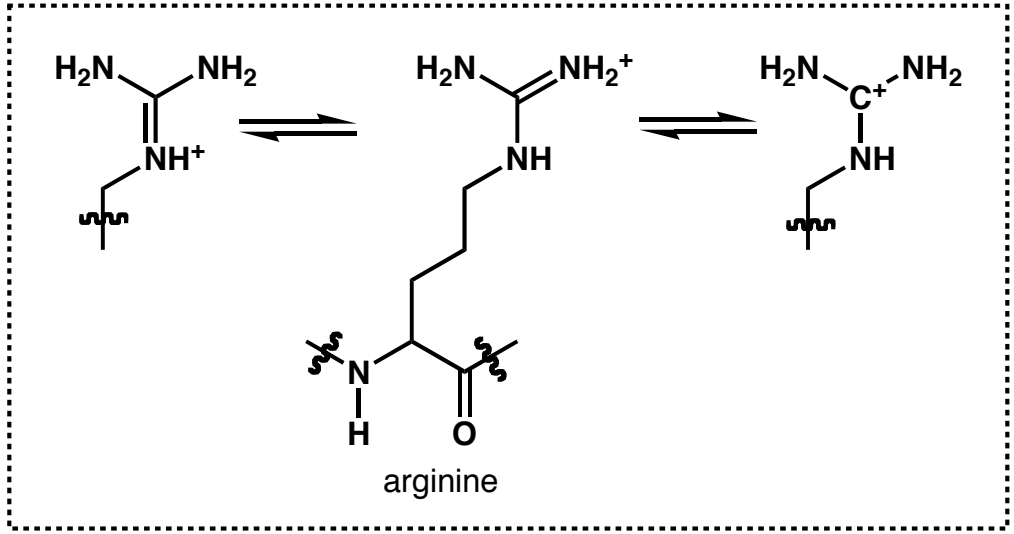
Arginine (Arg, R)

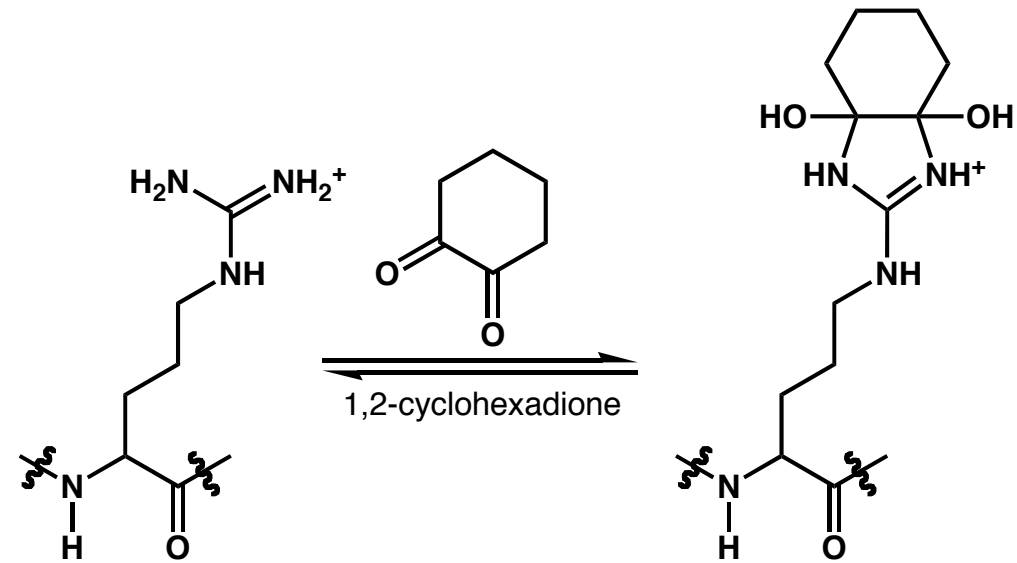


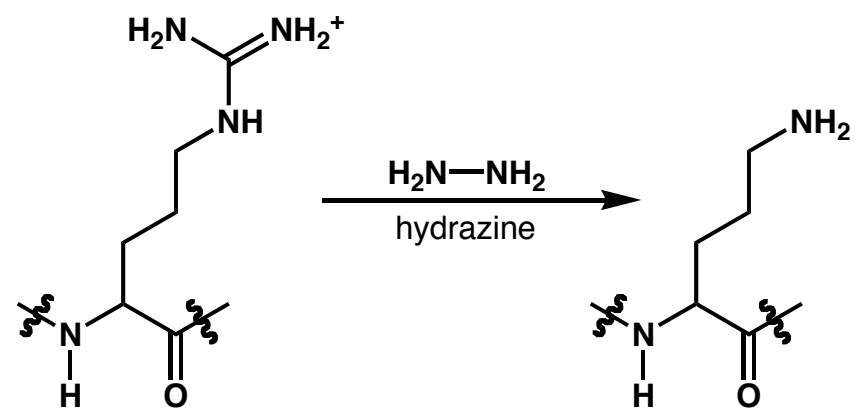
Histidine (His, H)

Arginine

- Side chain of Arg bears a strongly basic guanidino group (base pK_a = ~12.5).
- Positively charged over entire pH range proteins usually encounter.
- Positive charge of guanidino group is resonance stabilized.
- δ-Guanidino group reacts with 1,2- and 1,3-dicarbonyl compounds forming heterocyclic products.
- δ-Guanidino group also reacts with hydrazine leaving a primary amine on the δ-carbon of the side chain.

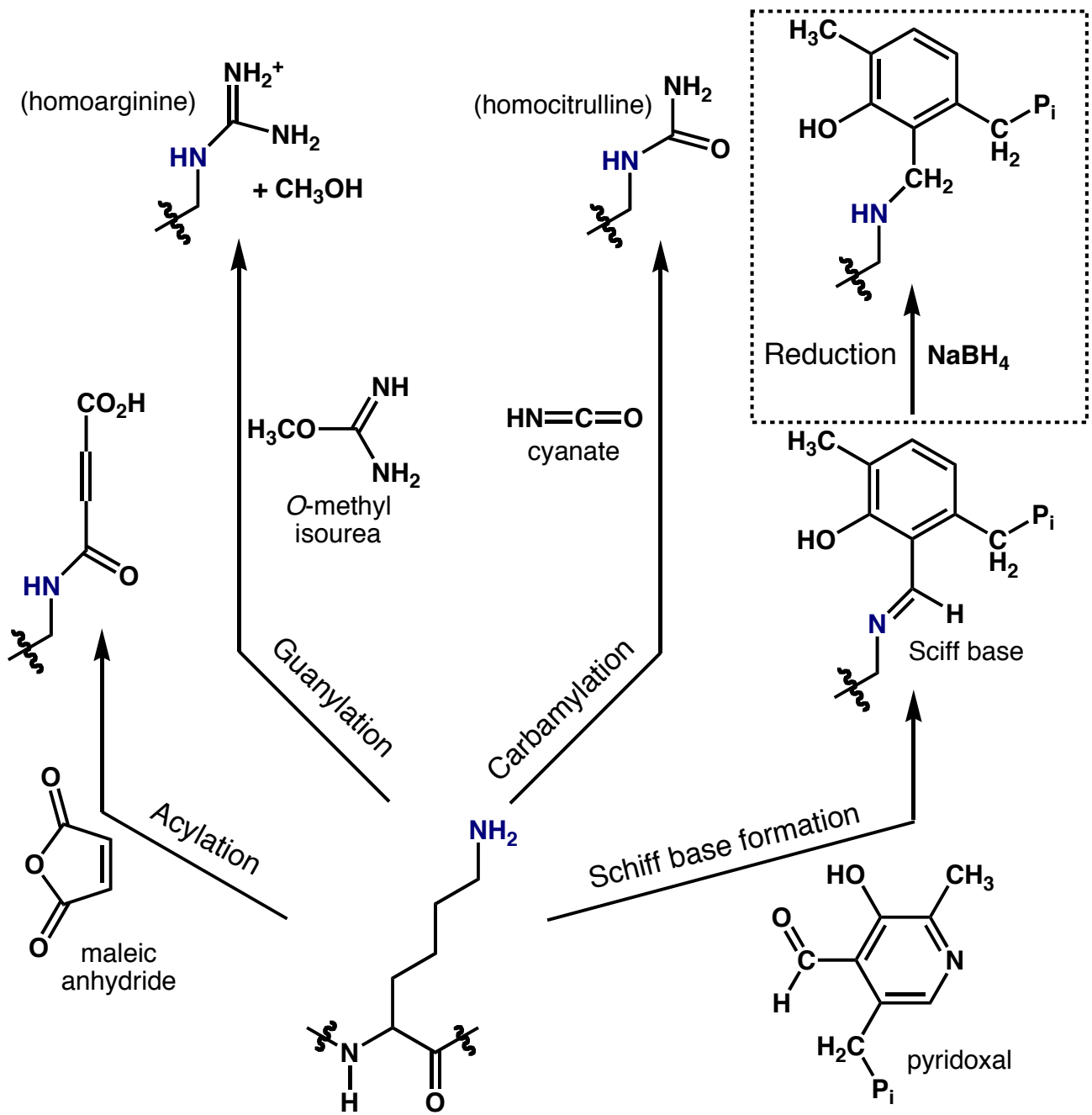






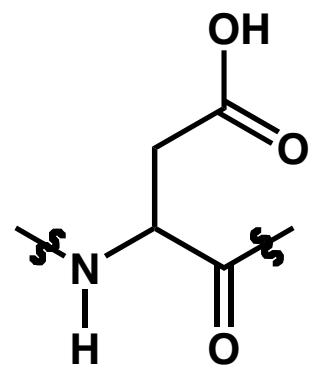
Lysine

- Hydrophobic chain capped with a terminal primary amino group (base pK_a = 10.5).
- While the majority of Lys ε-amino groups are protonated under physiological conditions, a small fraction of them are not.
- These amino groups are good nucleophiles and may participate in acylation, alkylation, arylation, carbamylation and amidination reactions.
- Rates of these reactions are greatly influenced by pH.
- The ε-amino group of Lys can also form a Schiff base (imine) with aldehyde groups.
- Can use 2,4,6-trinitrobenzene sulfonate (TNBS) to quantitate the number of free amino groups.

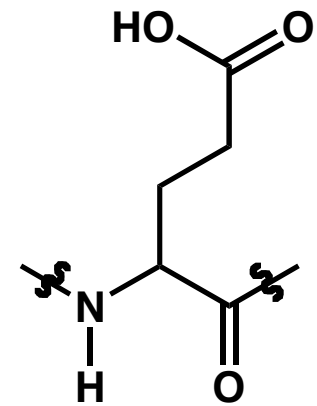


Glutamic and Aspartic Acids

- Structurally their side chains differ only by one methylene group.
- Significant differences in the way that they interact with the peptide backbone and effect on backbone conformation and chemical properties.
- Asp and Glu carboxyl groups typically have base pK_a's of 3.9 and 3.2 respectively.
- Side-chain pKa's in proteins can range from 2.0-6.7
- Chemical reactivity of the carboxyl groups on Glu and Asp side chains similar to corresponding organic molecules such as acetic acid.



Asp

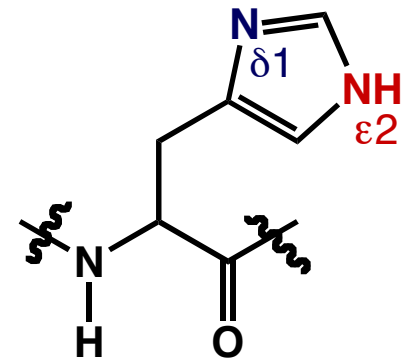


Glu

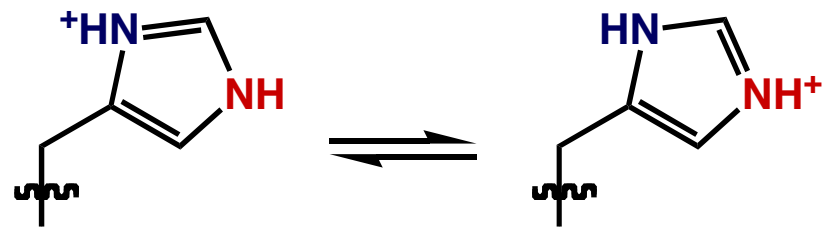
Histidine

- Unique properties of imidazole side chain of histidine make it ideally suited as a nucleophilic catalyst and as a ligand for coordinating metal ions.
- Imidazole has a pK_a of ~6, making it one of the strongest bases that can exist at neutral pH. The nitrogen is readily protonated, which kills its nucleophilicity.
- In its nonionized form, the nitrogen atom with the H atom is H-bond donor and the other nitrogen atom is a nucleophile and H-bond acceptor.
- The nonionized imidazole ring has two tautomers, differing in which ring nitrogen bears the H atom.
- The ring hydrogen of His can be removed with an apparent pK_a of 14.4.
- While imidazole can participate in several types of reactions, it is less reactive than amino and thiol groups therefore difficult to selectively modify His.

Two nitrogen atoms in ring designated δ1 and ε2 (or π and τ respectively).



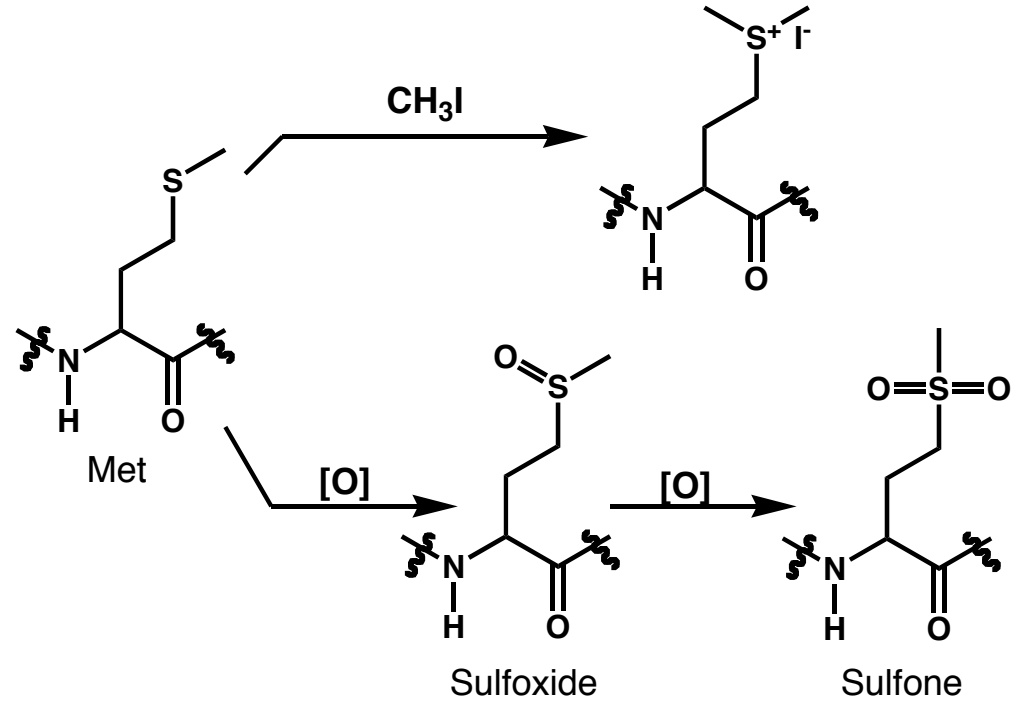
In the neutral form, the hydrogen usually resides on ε2 nitrogen.

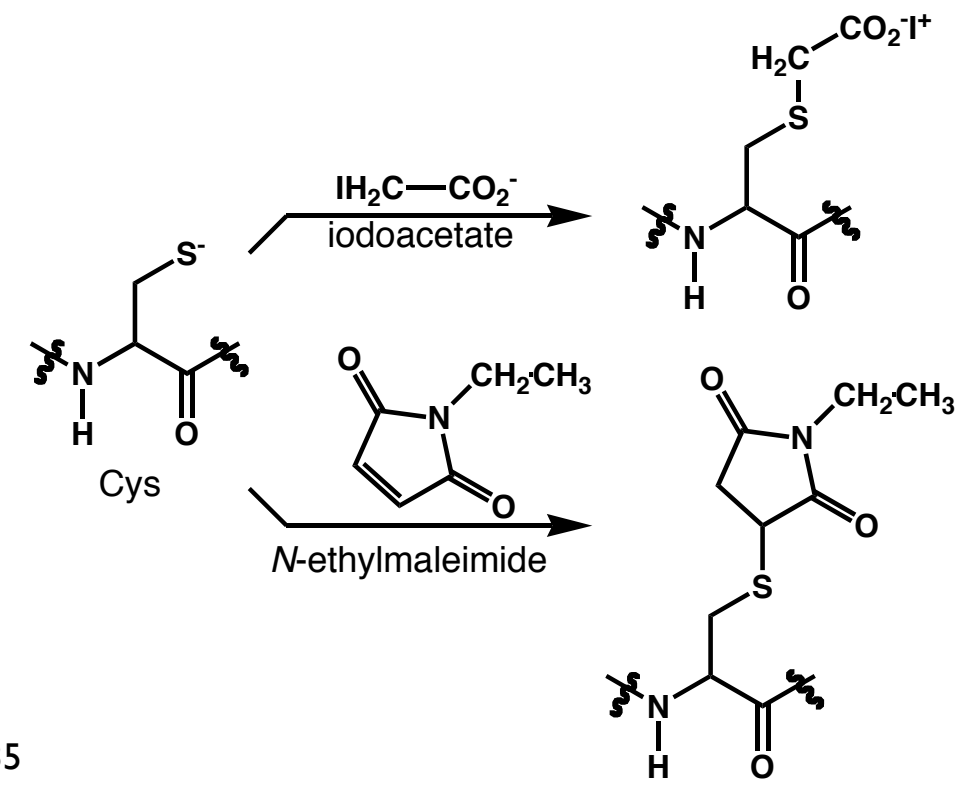


When protonated, charge distributed on both nitrogen atoms.

Cysteine and Methionine

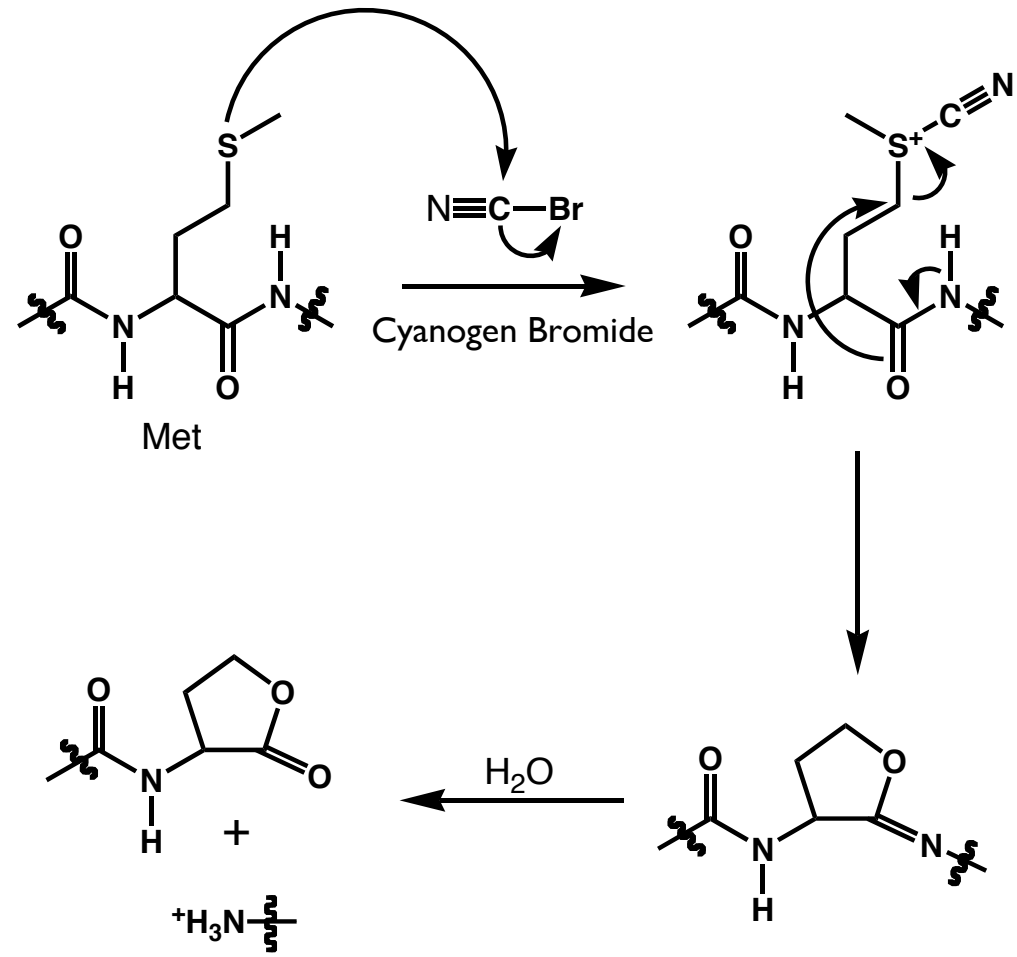
- **Methionine:** nonpolar relatively unreactive.
- Sulfur atom is somewhat nucleophilic, but cannot be protonated.
- Sulfur atom is susceptible to oxidation.
- **Cysteine:** thiol group is very reactive. Thiol group has a pK_a of 8.4-9.5 readily deprotonated under slightly basic conditions. Good nucleophiles.
- Thiolate ion is very reactive with alkyl halides.
- Thiol group can add across double bonds (*N*-ethylmaleimide).
- From complexes with various metal ions. Most stable complexes are formed with divalent Hg^{2+}. Also forms complexes with Cu, Fe, Zn, Mo, Mn and Cd ions.





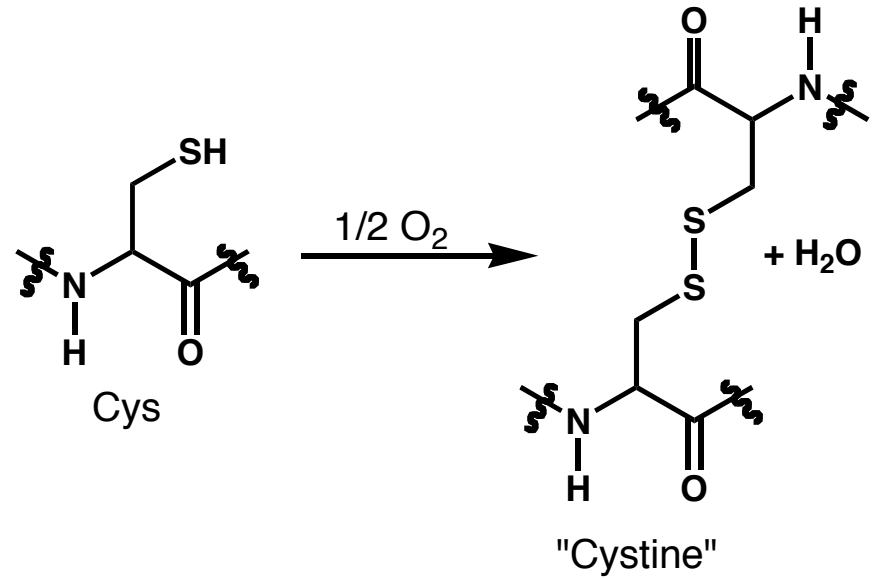
Reaction of Methionine with Cyanogen Bromide

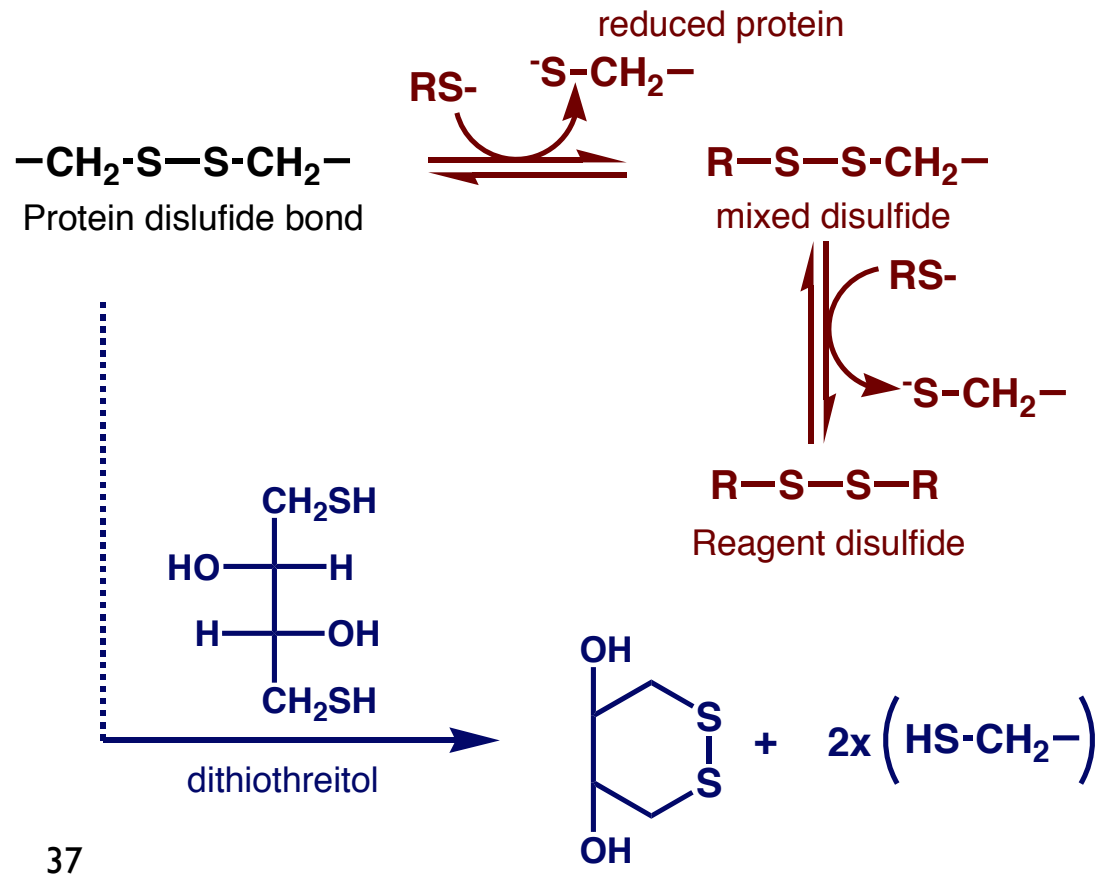
- The reaction between methionine residues and cyanogen bromide allows for the controlled fragmentation of peptides and proteins.
- Results in breaking of the peptide bond on the C-terminal side of a methionine residue.
- The methionine is converted to a homoserine lactone, and the C-terminal fragment is released with a free N-terminal amino group.



Cysteine Oxidation

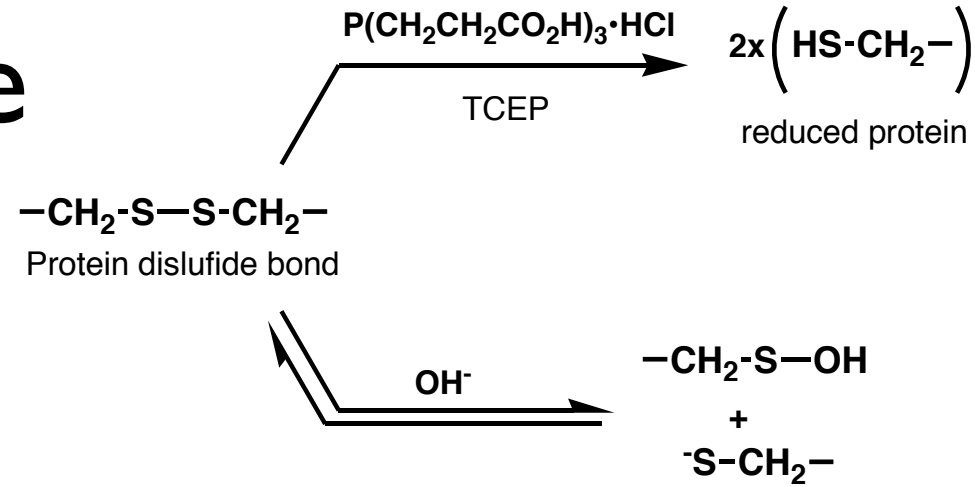
- Sulfur of Cys side chain can exist in several oxidation states. Besides thiol, only two oxidation states (disulfide and sulfonic acid) are usually encountered.
- Two Cys residues bound by disulfide bonds (sulfur-sulfur bond) often referred to as cystine (older nomenclature)
- Disulfide bonds are covalent bonds, and are relatively stable depending on conditions. With preferred dihedral angles of approx. + or - 90°.
- Disulfide bonds exchange rapidly under neutral or alkaline pH. Stable to acidic conditions.
- Disulfide bonds can be reduced by thiol-disulfide exchange with thiol reagent (RSH), such as mercaptoethanol and dithiothreitol or dithioerythreitol.

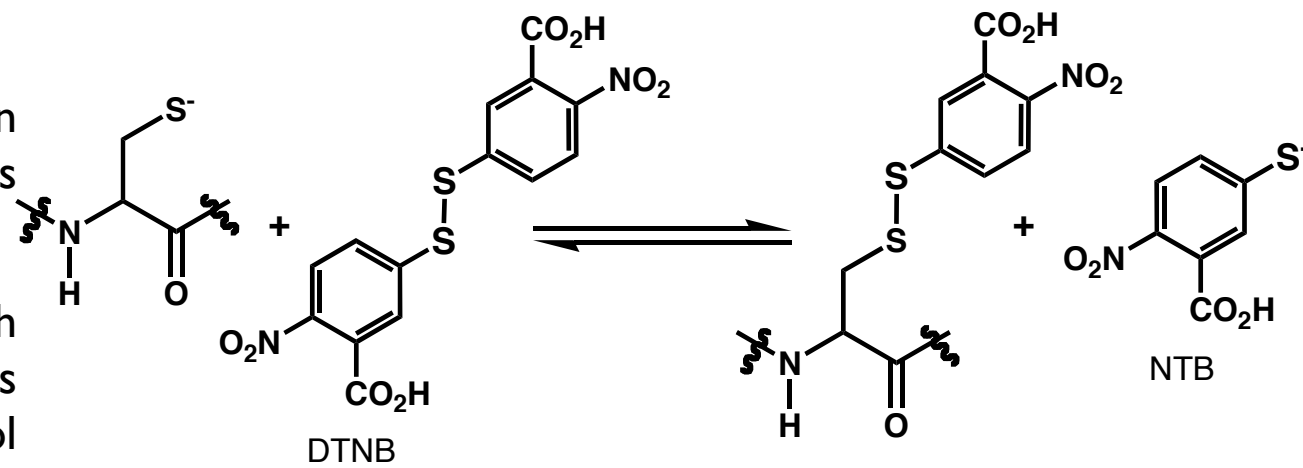


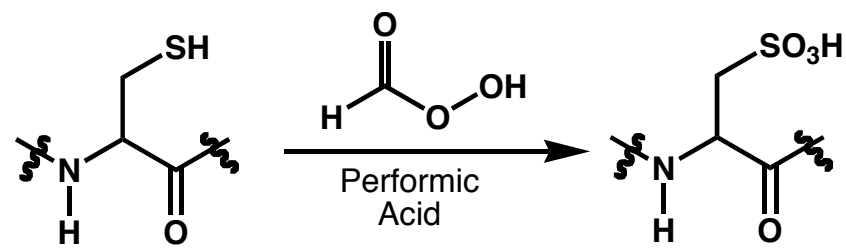


More Cysteine Chemistry

- Disulfide bonds can be reduced by phosphine reagents [i.e. tris(2-carboxyethyl) phosphine].
- Disulfide bonds can be broken by nucleophiles such as cyanide, sulfide or hydroxide.
- Thiol-disulfide exchange with aromatic disulfides provides means of assaying for free thiol groups.
- The thiol of cysteine can be oxidized to sulfonic acid using strong oxidizing agents (such as performic acid).







3-Cysteine

Non-standard Amino Acids

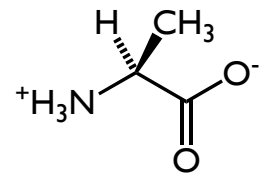
(Amino acids not directly coded for by genes)

- Considerable diversity in both structure and function.
- **Stereoisomers:**

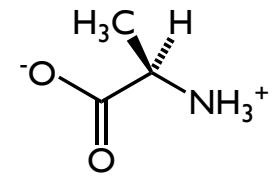
  D amino acids are relatively common in microorganisms.

  D-alanine and D-isoglutamate are incorporated in the cell walls of Gram-positive bacteria.

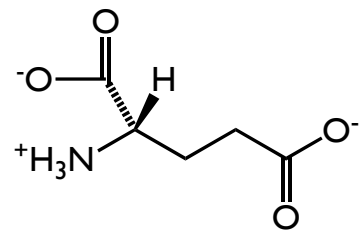
  Some micorobes are known to produce small peptides (ionophores such as gramicidin A) that form channels in membranes. These peptides consist of alternating L and D amino acid residues.



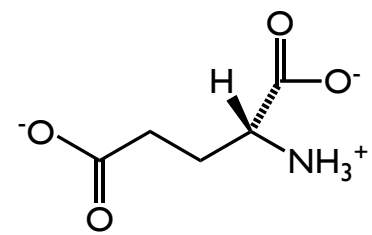
L-Ala



D-Ala



L-Glu*



D-Glu*

* drawn consistent with incorporation into peptide backbone in "iso" orientation.

Gramicidin A:

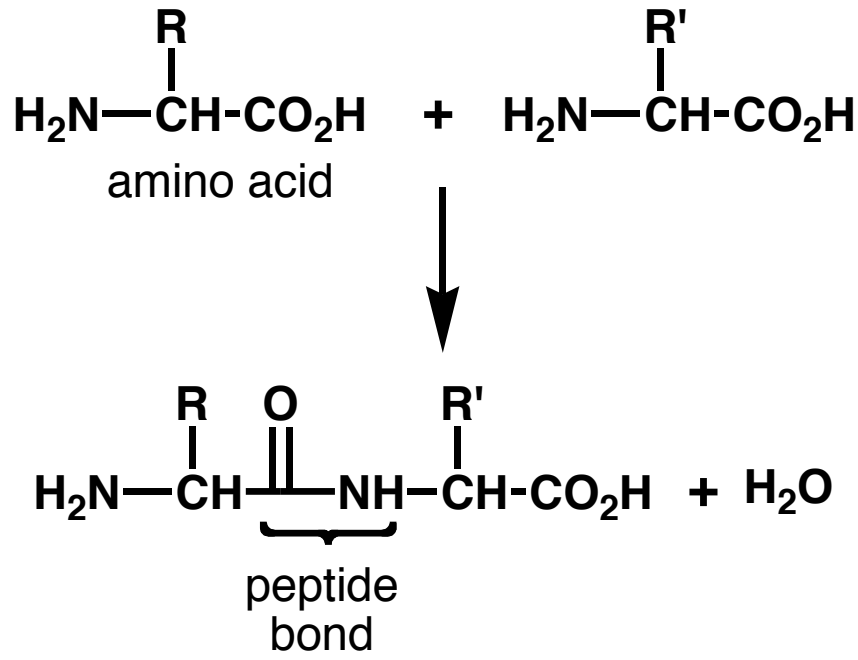
HCO-NH-Val-Gly-Ala-*Leu*-Ala-*Val*-Val-*Val*-Trp-*Leu*-Trp-*Leu*-Trp-*Leu*-Trp-*Leu*-Trp-CO-NH-CH_2CH_2OH

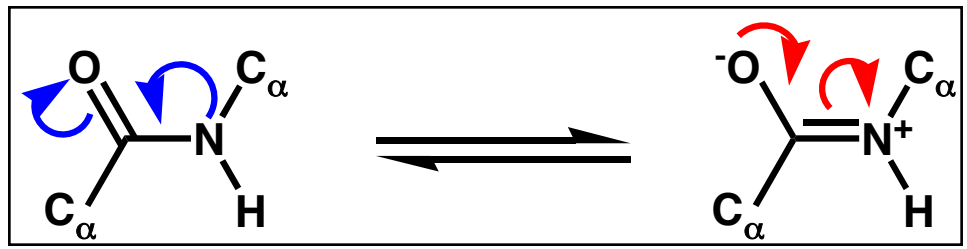
(D amino acid residues indicated in *italics*)

Polypeptide Backbone

Polypeptide Backbone

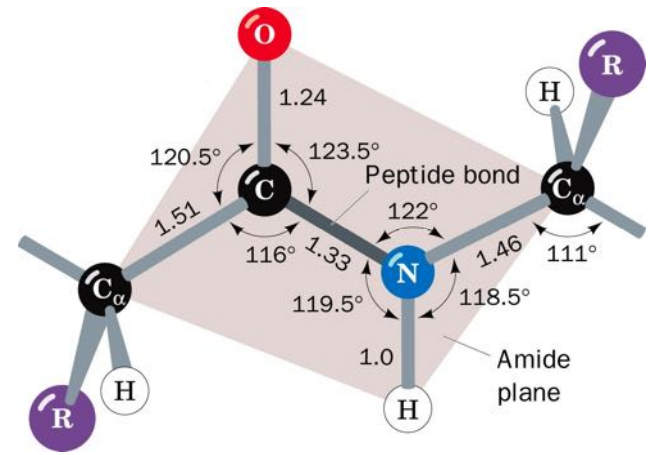
- Amino acid residues of a protein are linked by amide bonds ("peptide bonds").
- Formation of a peptide bond also produces a molecule of water. (referred to as a condensation reaction)
- Peptide backbone consists of repeated pattern of amide N, C_α and carbonyl carbon atoms.
- The peptide bond has partial (~40%) double bond character, which restricts rotation around the bond.



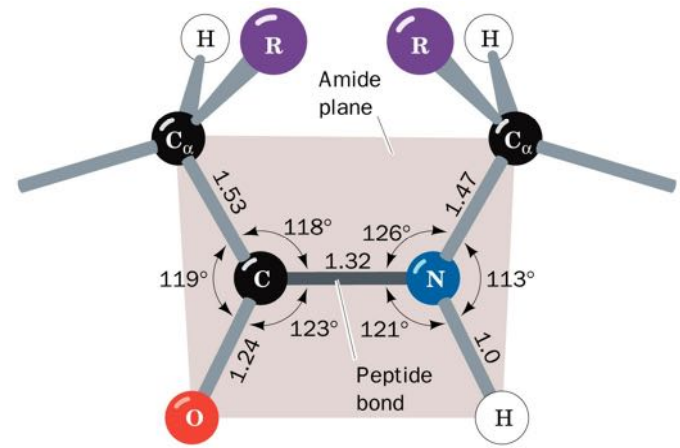


Polypeptide Backbone

- The planar peptide bond can assume a configuration where the C_α atoms are *trans* and one where they are *cis* (usually in *trans* configuration).
- Planar "Peptide group" defined as peptide bond and flanking C_α atoms.
- The presence of an asymmetric center at the C_α carbon atom and only L-amino acids results in the polypeptide backbone having an inherent asymmetry.
- This combination of inherent asymmetry and restricted rotation around peptide bond are important in the conformational properties of polypeptides and proteins.



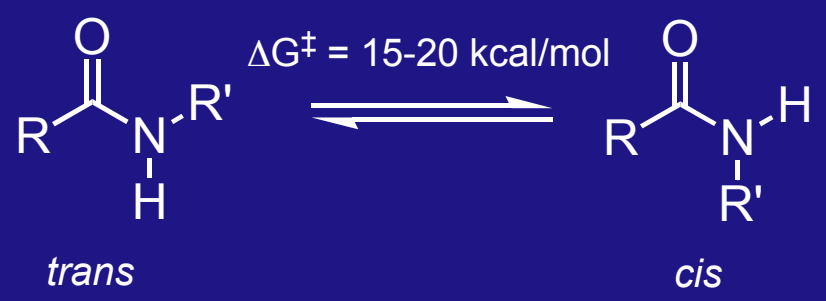
***trans* peptide bond**



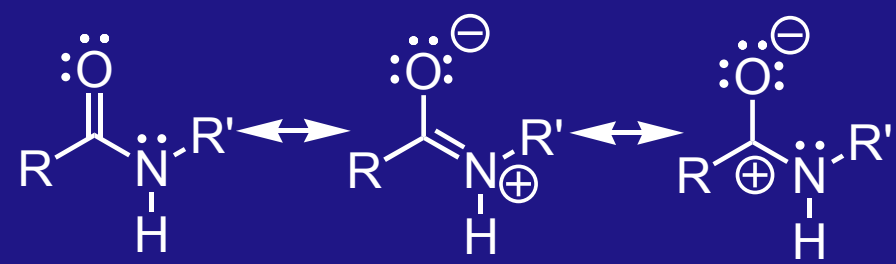
***cis* peptide bond**

Resonance, Dipoles and Peptide Bonds

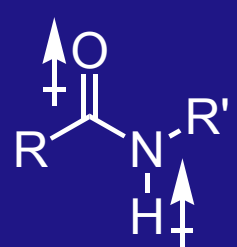
- Historically barrier to rotation and preference for trans configuration has been attributed to resonance and double-bond character in the C-N bond.
- Recent data suggests that the properties of the peptide bond more reflects dipole interactions associated with the C=O and N-H bonds.
- When two atoms of differing electronegativities are bonded, the electrons in the bond are not distributed equally, resulting in a dipole with one end of the bond will be δ^+ and the other will be δ^-.
- Dipole moment has both magnitude and directionality.
- Dipole moment provides a means of comparing bond polarities and evaluating the relative force that the dipole exerts on neighboring charges or dipoles.



Resonance



Dipoles



Peptide Conformation and Torsion Angles

- Peptide backbone is a linked sequence of nearly planar peptide groups
- ϕ = N-C_α
- ψ = C_α-$C_{carbonyl}$
- Steric constraints associated with ϕ and ψ angles. Influenced by substituents on the amino acid side chains.
- Some conformations can become sterically forbidden.

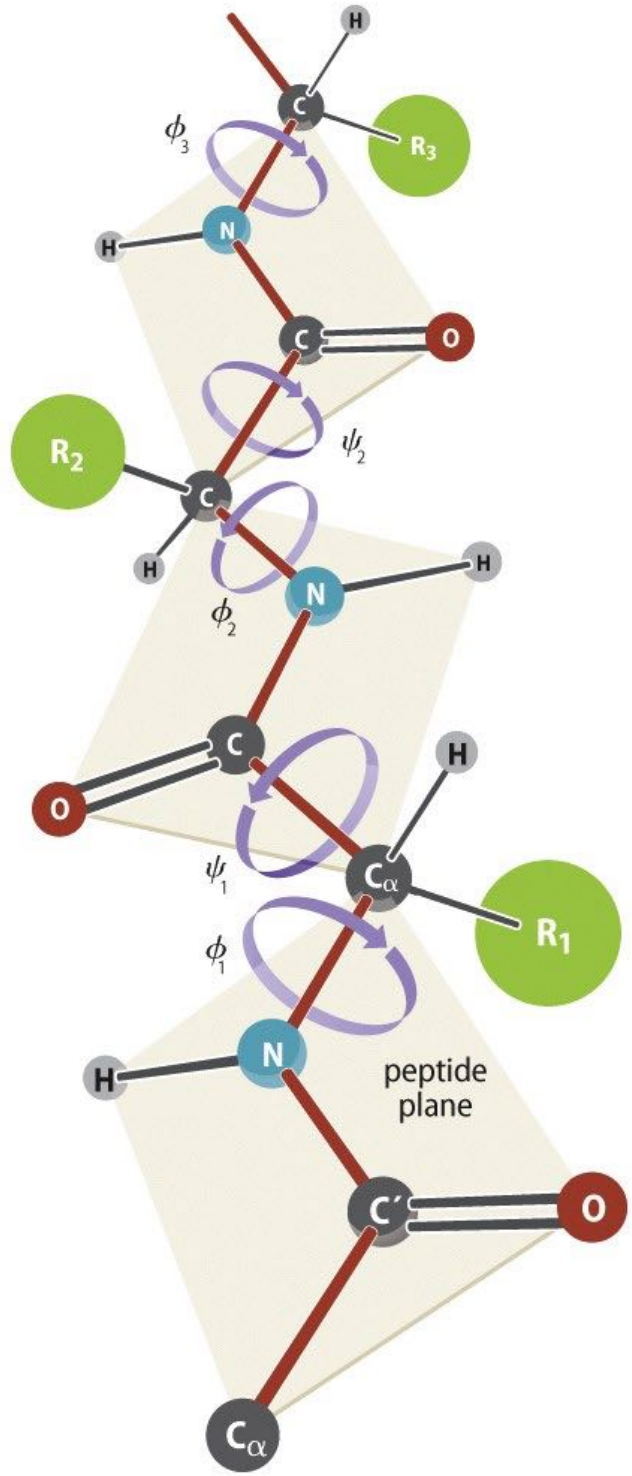
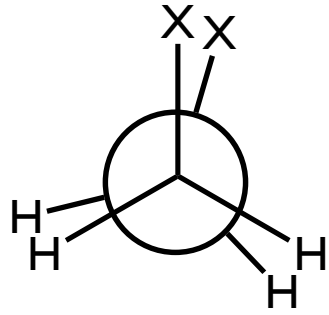


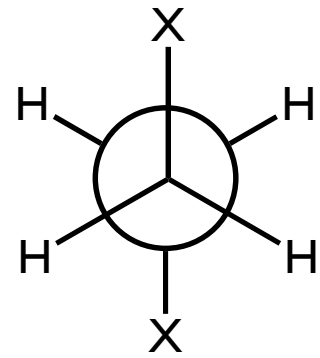
Figure 1.7 How Proteins Work (©2012 Garland Science)

Three-Dimensional Conformations

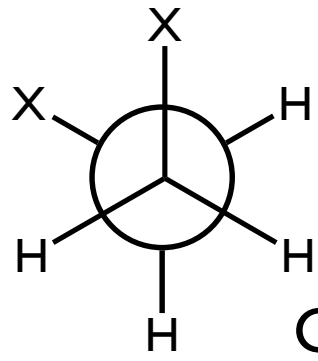
- The three-dimensional structure is important for biomacromolecules, which contain many bonds and can assume many conformations.
- **Conformations**: nonsuperimposable three-dimensional arrangements of atoms in a molecule that are interconvertible without breaking covalent bonds.
- Even a simple molecule might be considered to exist in an infinite number of conformations.
- Only energetically stable arrangements are usually classified as distinct conformations.
- Each amino acid in a polypeptide contains three bonds in the peptide backbone plus the side chain and can exist in a number of conformations.
- The peptide bond has double bond character and is limited to planar conformations (*cis*/*trans*). The other backbone and side chain bonds are primarily single bonds.

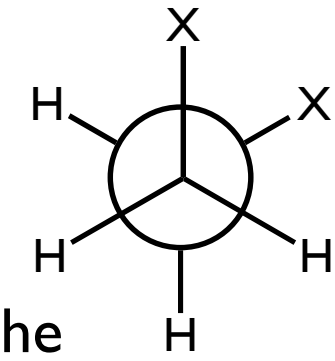


Eclipsed



Staggered





Gauche

Three-Dimensional Conformations

- Each amino acid in a polypeptide contains three bonds in the peptide backbone plus the side chain and can exist in a number of conformations.
- Not all of the theoretical amino acid conformations are possible because they would result in steric conflicts (overlapping atoms and excluded volume effect).
- Calculating the number of suitable conformations presents a significant challenge, only rough estimates are possible.
- Conformational diversity makes adoption of one conformation entropically unfavorable (conformational entropy: ΔS_{conf}).

$$\Delta S_{conf} = R \ln N$$

- For a conformation to be stable, it requires stabilizing interactions that overcome the loss in conformational freedom.
- Proteins and some peptides assume particular conformations that are stabilized by weak interactions.

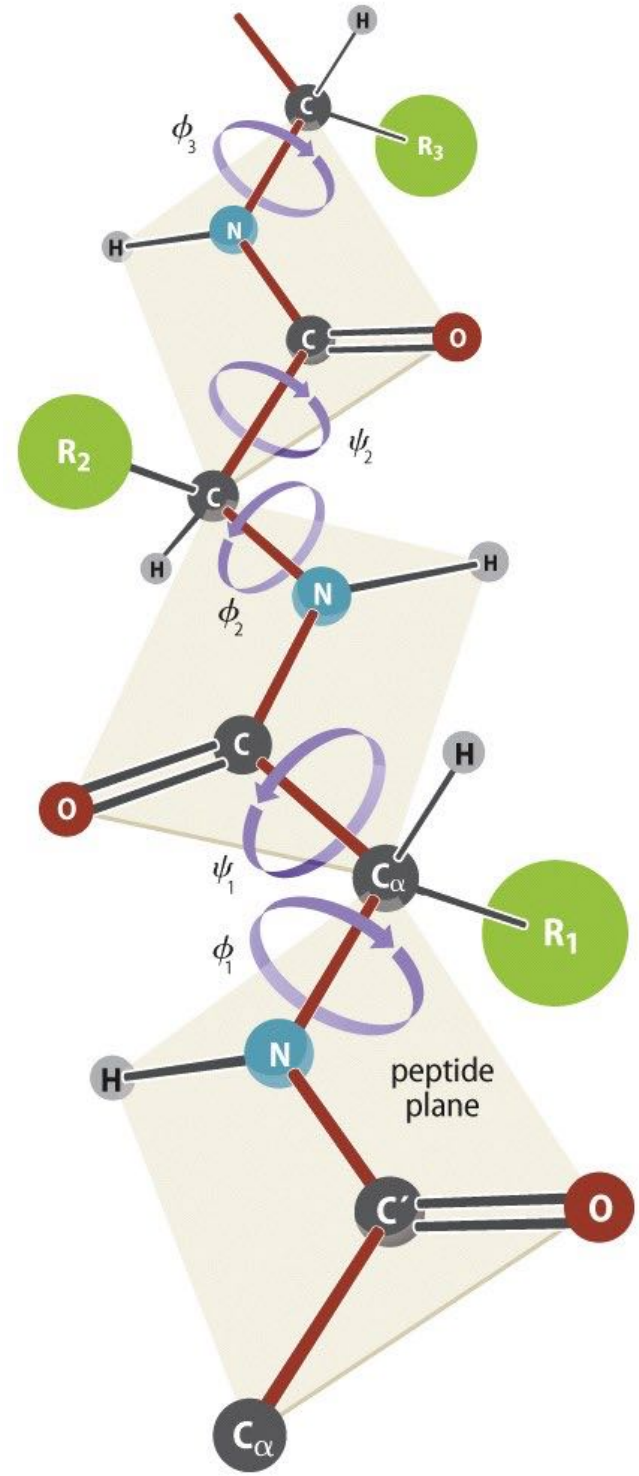


Figure 1.7 How Proteins Work (©2012 Garland Science)

Polypeptides as Random Polymers

- The conformational properties of random polypeptides are best calculated statistically using methods developed for synthetic polymers.
- The peptide bond is usually planar and the group of atoms usually functions as a rigid unit (peptide unit).
- Rotation about bonds described as torsion or dihedral angles (-180° to +180°).
  - ω = C'-N
  - ϕ = N-C_α
  - ψ = C_α-C'
  - x_j = side chain torsion angles
- For most amino acids, the peptide bond (ω) prefers the *trans* conformation 1000:1 over the *cis* form.
- When the residue *i*+1 is Pro, there is very little difference between the *cis* and *trans* forms of the peptide bond (*trans* form favored only 4:1).
- The values of ϕ and ψ that are possible are constrained geometrically due to steric clashes with neighboring atoms.

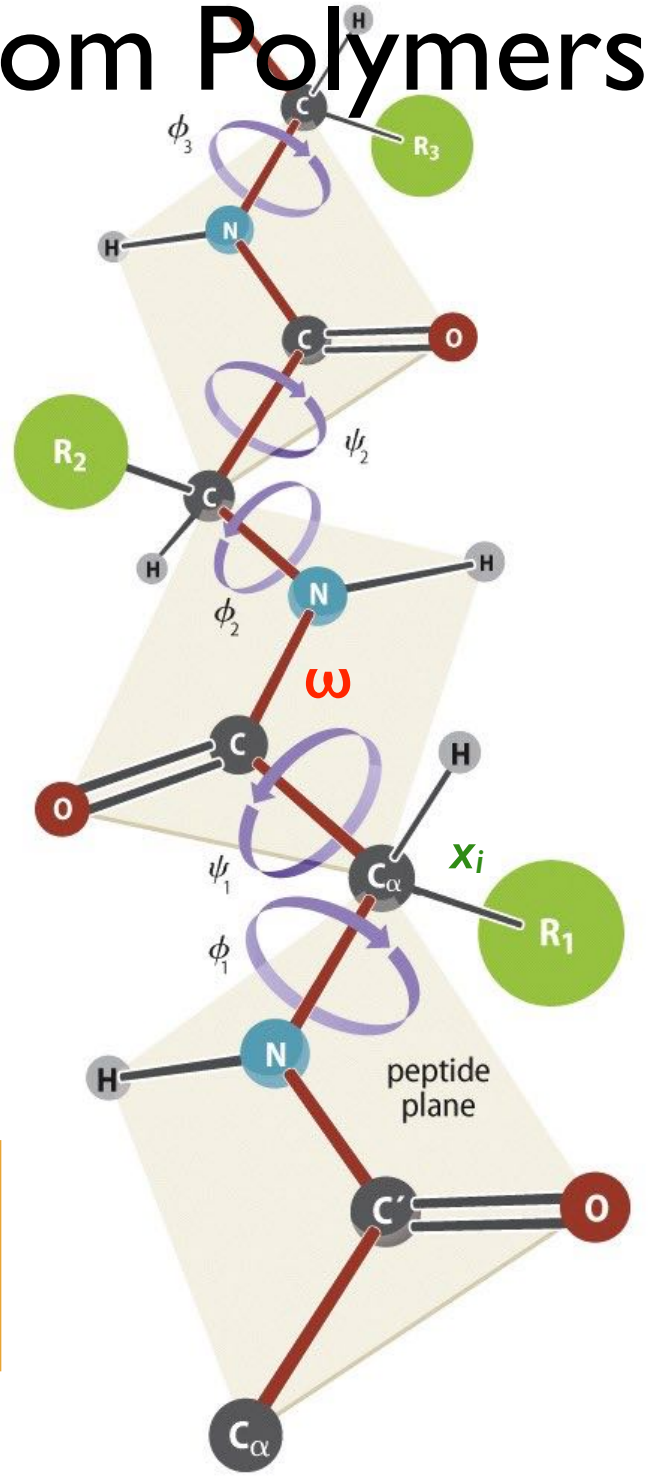
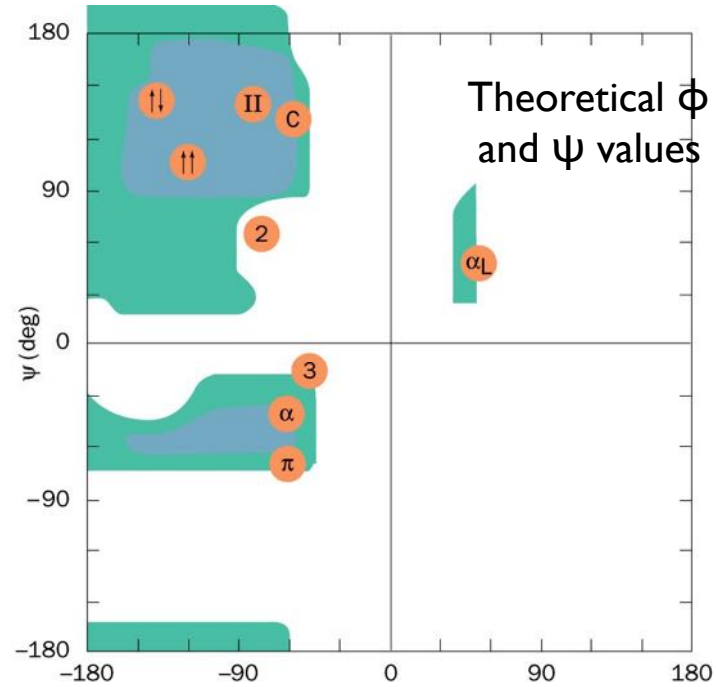
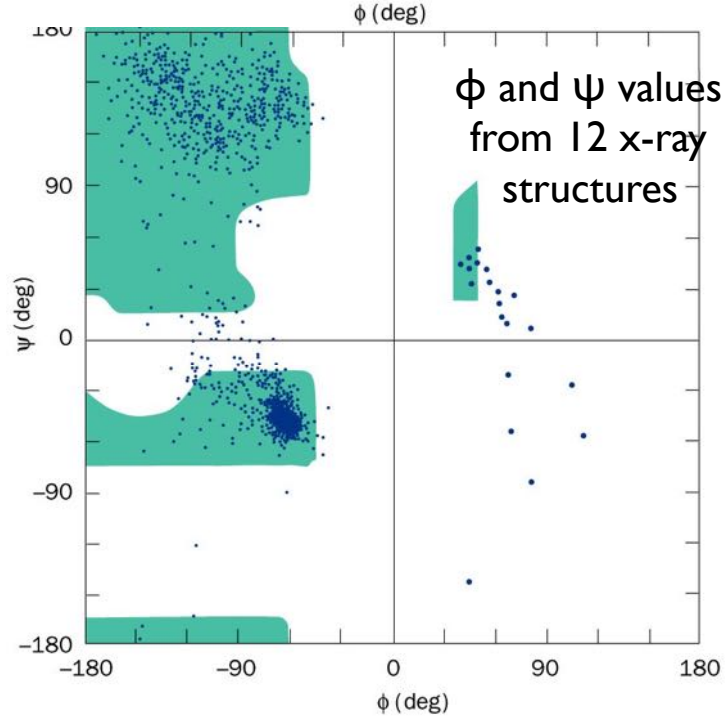


Figure 1.7 How Proteins Work (©2012 Garland Science)

Polypeptides as Random Polymers

- The permitted values of ϕ and ψ can be illustrated using a two dimensional map known as a Ramachandran plot.
- Only three small regions, accounting for ~30% of the Ramachandra diagram, represent combined fully and partially allowed ϕ and ψ combinations.
- Distribution of allowed ϕ and ψ in part reflects the inherent chirality of most amino acids.
- Gly is the most conformationally flexible.
- Other amino acids with longer and larger side chains have additional restrictions on ϕ and ψ.
- Amino acids with β-branched side chains are more constrained than those without.
- Pro is the most constrained.
- Energy differences between allowed and disallowed conformations are smaller than expected.
- Torsion angles (ϕ and ψ) associated with the common secondary structures fall within the allowed regions.





Side Chain Conformational Freedom

- Side chain conformational restrictions arise from potential overlap with neighboring residues and with the peptide backbone.
- Side chain branching and steric bulk are major factors in limiting conformational freedom. (particularly true for β-branched side chains)
- Steric restriction around the C_α-C_β bond results in discrete rotamer populations.
- x_1 = C_α-C_β dihedral angle.
  - Gly, Ala and Pro lack a x_1.
- x_2 = C_β-C_γ dihedral angle.
  - x_2 for Ser, Thr and Cys is difficult to assign.
  - Preferred x_2 for Arg, Glu, Gln, Ile, Leu, Lys and Met are well known.
- Rotamer libraries based on preferred amino acid side chain conformations: take into account inherent preferences for specific amino acids as well as constraints associated with secondary structure.

Rotamer	x_1 angle
g^+ (gauche $^+$)	-120°-0°
trans	120°-240°
g^- (gauche $^-$)	0°-120°

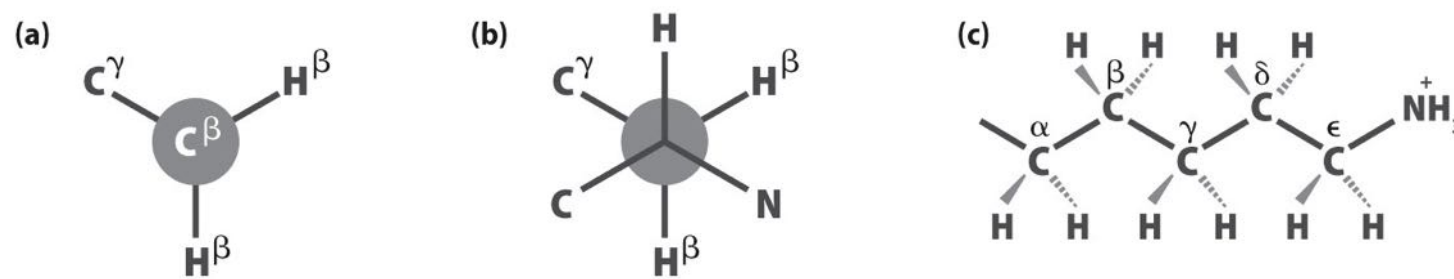
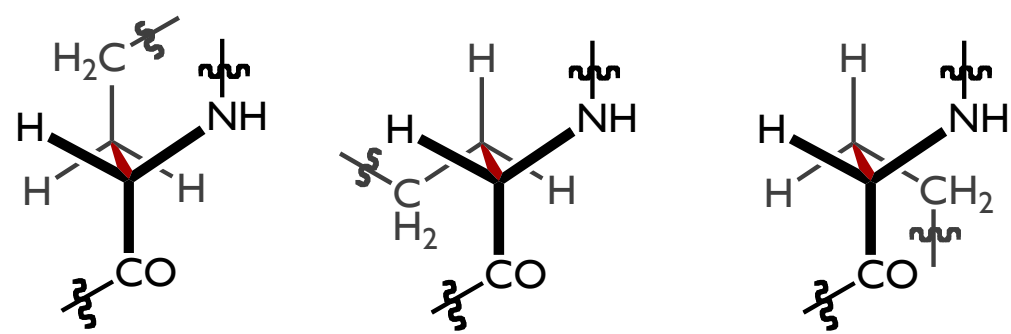


Figure 1.9 How Proteins Work (©2012 Garland Science)



Backbone Conformations and Secondary Structure

Levels of Organization

- **Primary structure (1° structure):** the amino acid sequence of polypeptide chain.

- **Secondary structure (2° structure):** local spatial organization and arrangement of the peptide backbone. Generally refers to easily localized structural elements (i.e. helices and sheets).

- **Tertiary structure (3° structure):** the comprehensive three-dimensional structure of a protein (single polypeptide chain).

- **Quaternary structure (4° structure):** assembly through noncovalent interactions) of a larger protein structure from 2 or more polypeptide chains (subunits), and the organization of these subunits.

β-Sheet

- The most populated region of backbone conformational space is the **β-sheet** region.
- The β-sheet is characterized by peptide chains in extended conformations with a repeating pattern of ϕ and ψ angles (approx. -130° and +125° respectively).
- Extended conformation of an isolated chain is not stable. β-strand is only stable when incorporated within a β-sheet.
- In a β-sheet, hydrogen bonds formed between backbone amide C=O of one strand with amide -NH of an adjacent strand - with near ideal geometry for hydrogen bonds.

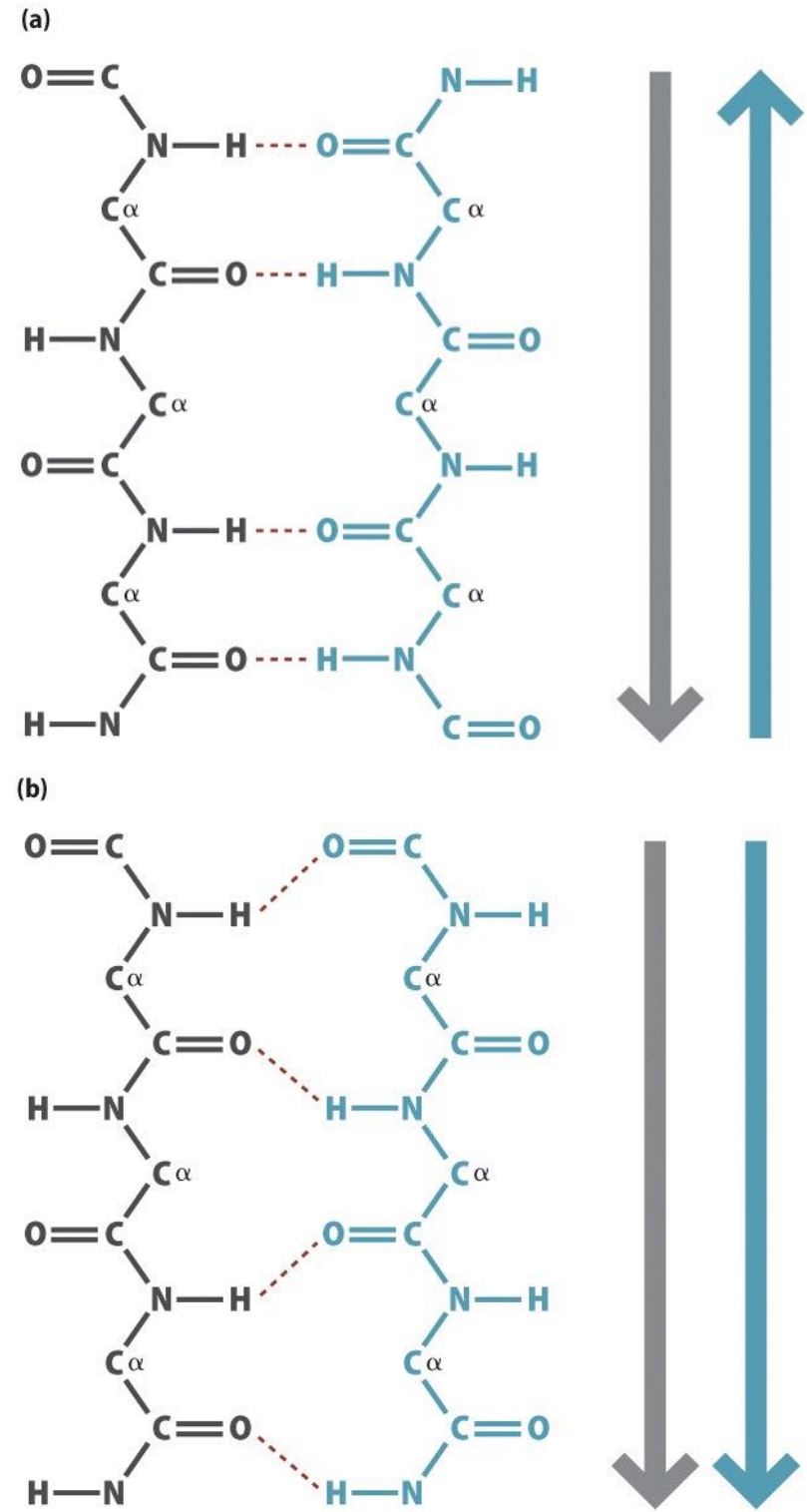


Figure 1.10 How Proteins Work (©2012 Garland Science)

β-Sheet

- **Two flavors of β-sheet:**

  - ✦ Antiparallel β-sheet ➔ H-bonded β-strands run in opposite directions.

  - ✦ Parallel β-sheet ➔ H-bonded β-strands run in the same direction.

  - ✦ β-strands can also combine in mixed β-sheets - strong bias against mixed β-sheets.

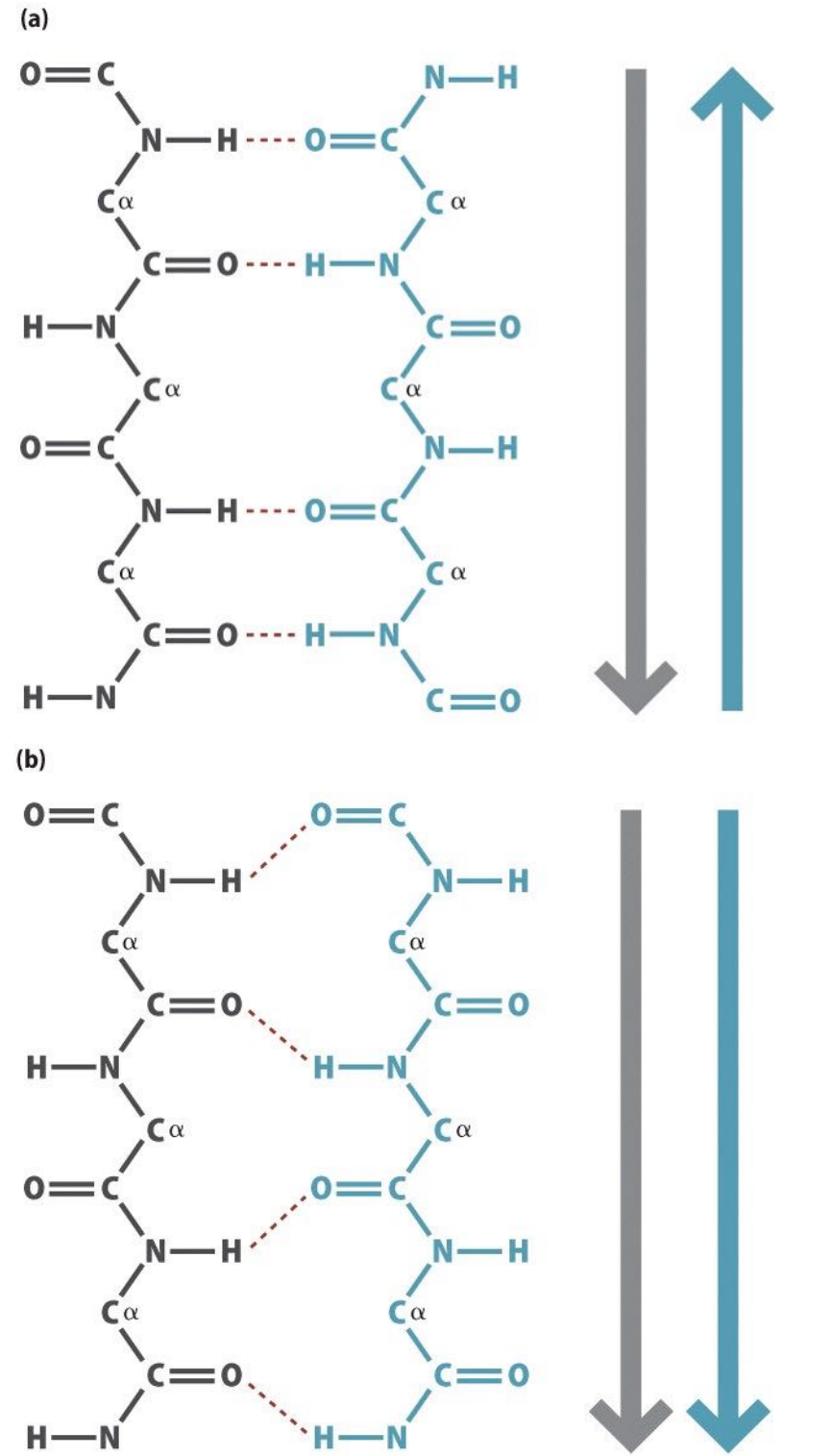


Figure 1.10 How Proteins Work (©2012 Garland Science)

β-Sheet

- In a β-sheet, side chains from adjacent residues lie on opposite sides of the sheet and do not interact.
- Adjacent strands in β-sheets tend to be adjacent in the sequence as well.
- β-sheets may involve the interaction of different strands that can be far apart in the amino acid sequence.
- An intramolecular β-sheet is not a completely regular structure because it requires turns and loops in order for strands to align.
- β-sheets can be involved with protein-protein interactions and interfaces.
- Poly(Tyr), poly(Lys) and poly(S-carboxymethyl-Cys) form soluble β-sheets under certain conditions.

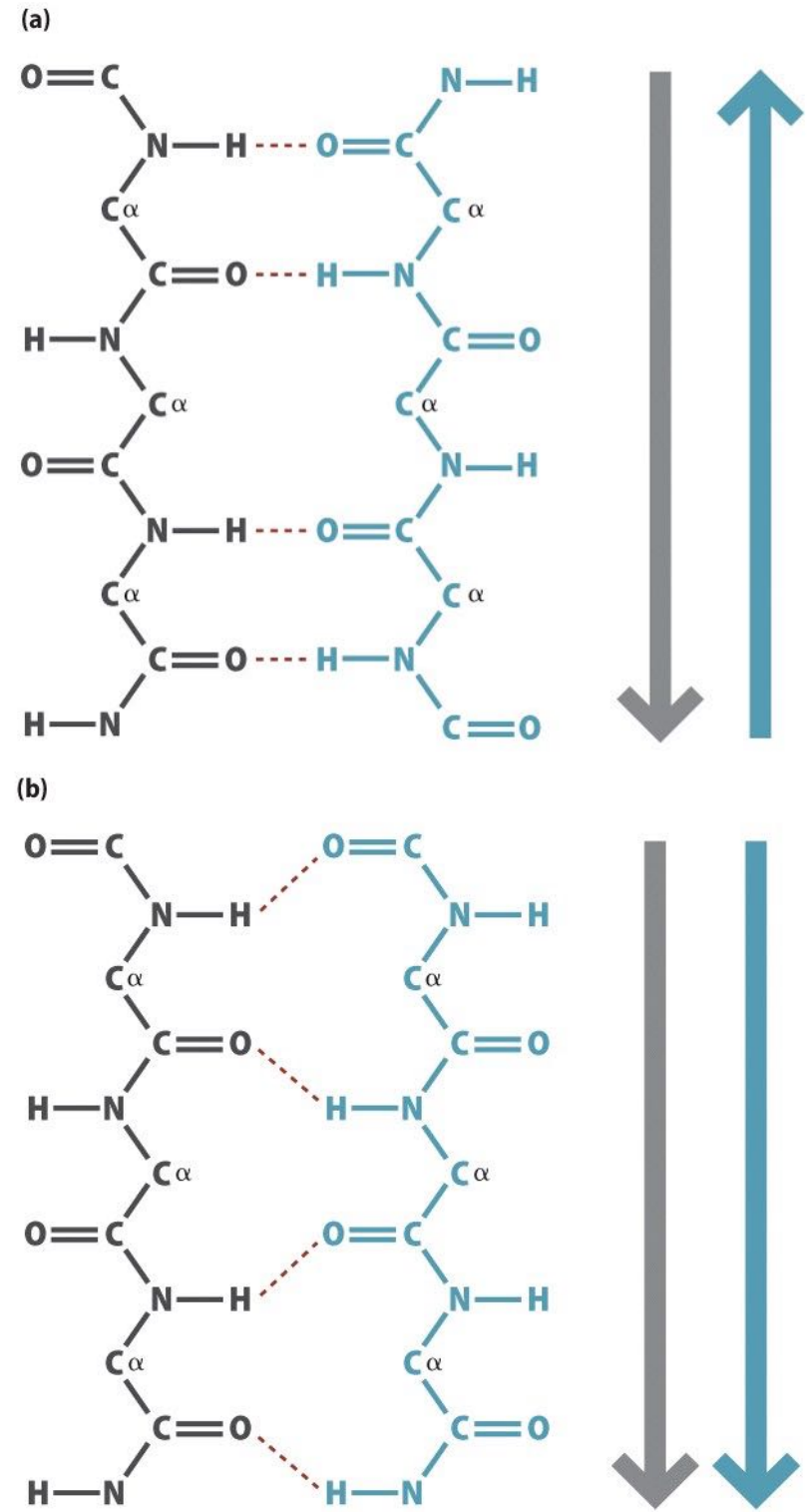


Figure 1.10 How Proteins Work (©2012 Garland Science)

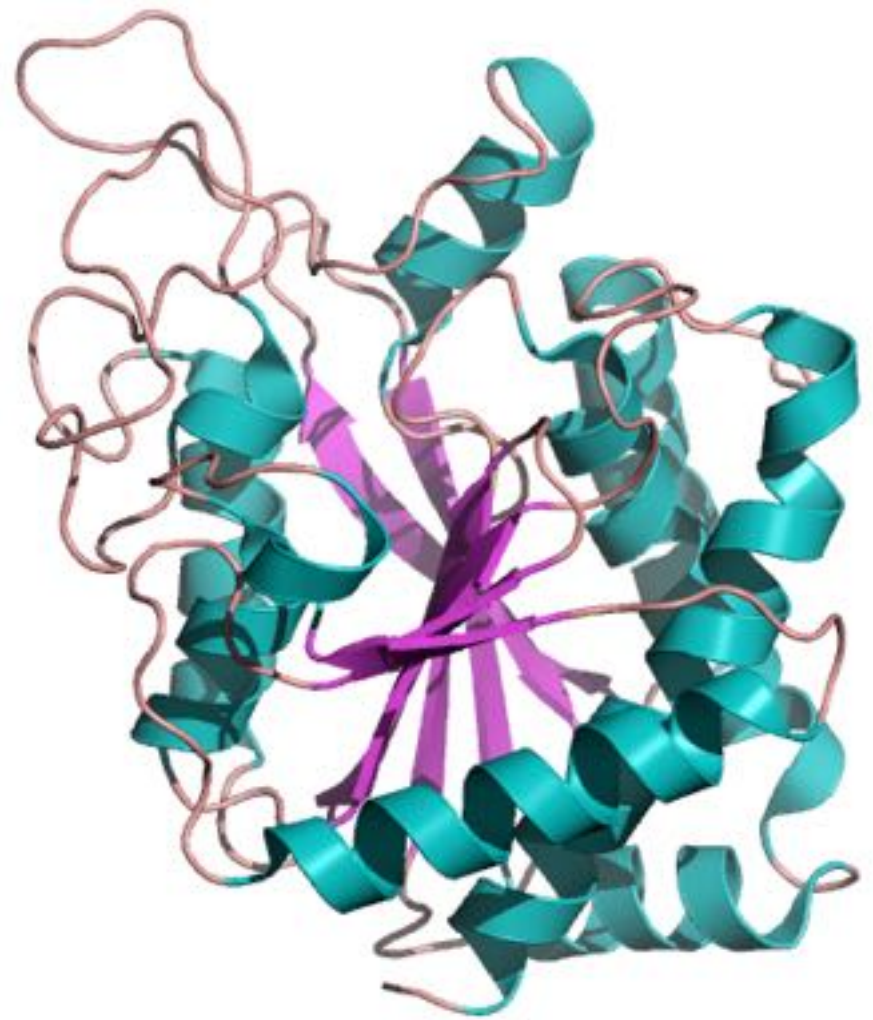
Secondary Structure

- Most β-sheets in globular proteins are twisted rather than planar – with a right-handed twist of 0°-30° between strands. Likely due inherent chirality of the amino acids and non bonding interactions.

- The conformational parameters of the peptide backbone can also deviate from ideality. More positive ϕ and ψ values are generally observed in twisted sheets.

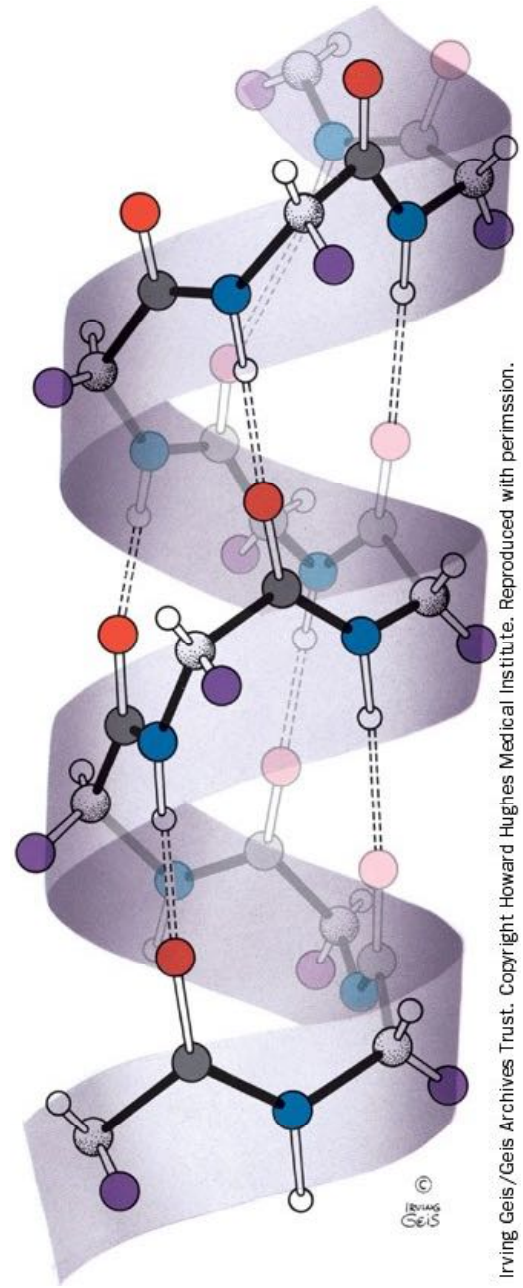
- Further distortions are also observed in mixed β-sheets because of differences in the backbone conformations of parallel and antiparallel β-sheets.

- Isolated β-sheets have a propensity to aggregate and grow indefinitely from the edges. Therefore, there is no ideal model for isolated β-sheets.



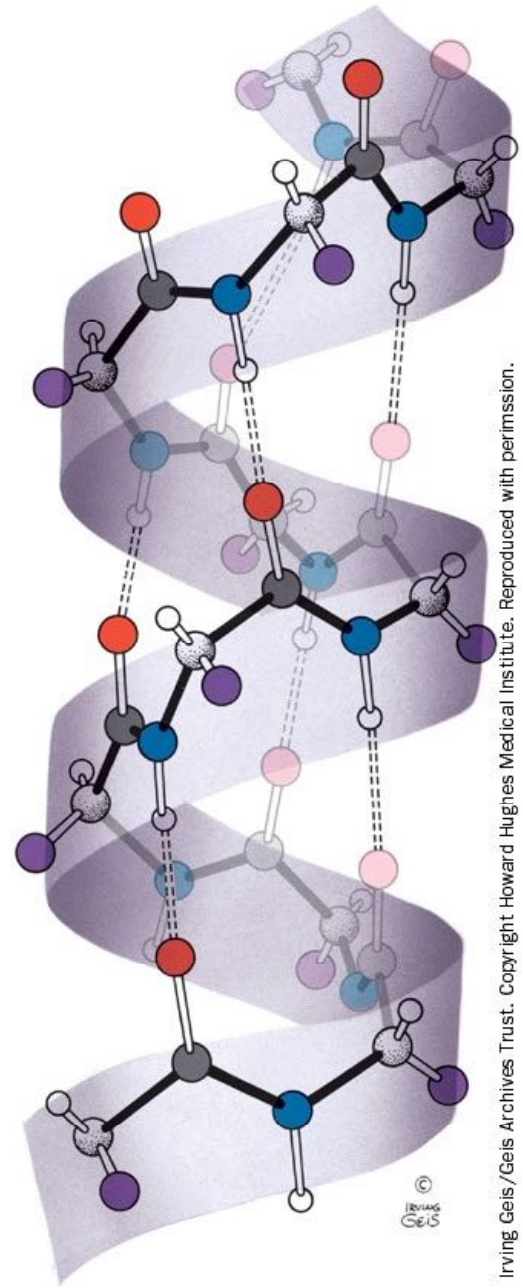
The α-Helix

- The other major structural region is the **α-helical** region.
- The right handed α-helix is the best known and most recognizable of the polypeptide regular structures.
- The α-helix combines favorable conformational angles, van der Waals interactions and backbone hydrogen bonding.
  - ✦ ϕ = -57° (-62°) and ψ = -47° (-41°).
  - ✦ 3.6 residues/turn with pitch of 5.4Å.
- H-bonds between N-H group (donor) of n^{th} residue and the C=O group (acceptor) of the $n\text{-}4^{th}$ residue .



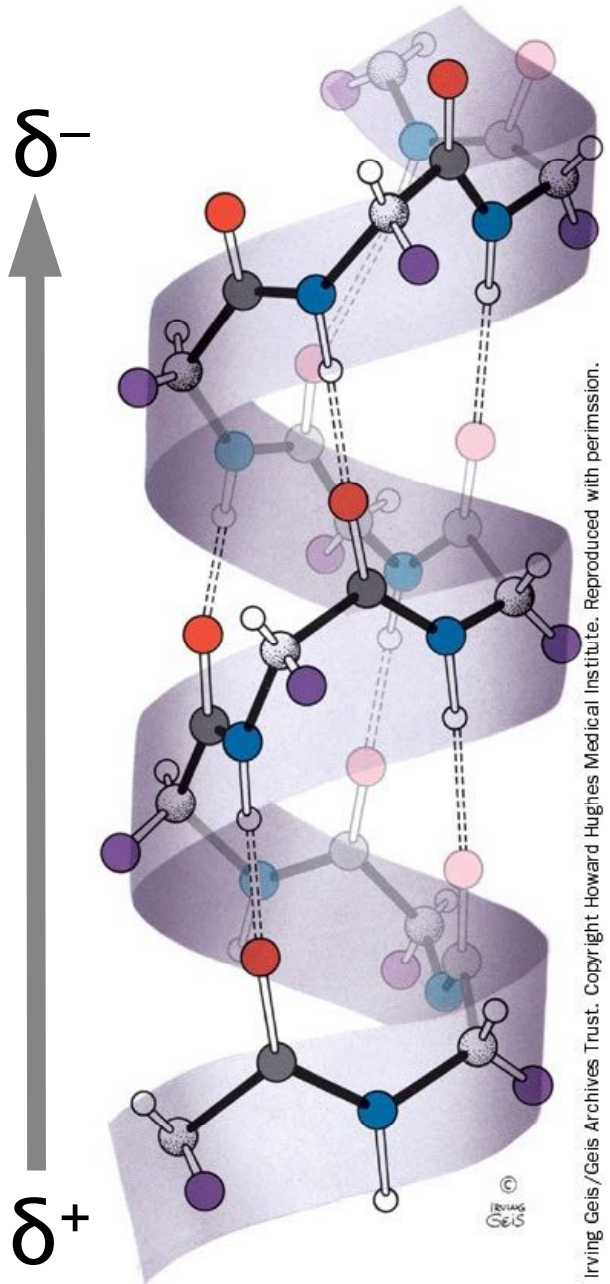
The α-Helix

- Side chains are directed outward and slightly backwards (towards N-terminus). [restrictions on side chain conformations]
- The detailed geometry of the α-helix is found to vary somewhat in folded proteins.
- Slightly different geometry is adopted by natural proteins with ϕ = -62°, ψ = -41° and H-bonds directed slightly out-away from helix (believed more favorable than classic conformation).



The α-Helix

- All backbone hydrogen bonds and peptide groups point in the same direction in the α-helix.
- Alignment of hydrogen bonds results in helices having a net dipole with the N- and C-termini having partial positive and negative charges (respectively) ~0.5-0.7 unit charge at each end.
- Frequently negatively-charged groups/species bind at N-terminus of helix, but positively-charged groups only rarely bind at C-terminus of helix.
- Polarization of hydrogen bonding may increase the dipole moment of each peptide bond as much as 50%.



End Lecture 1