



1785

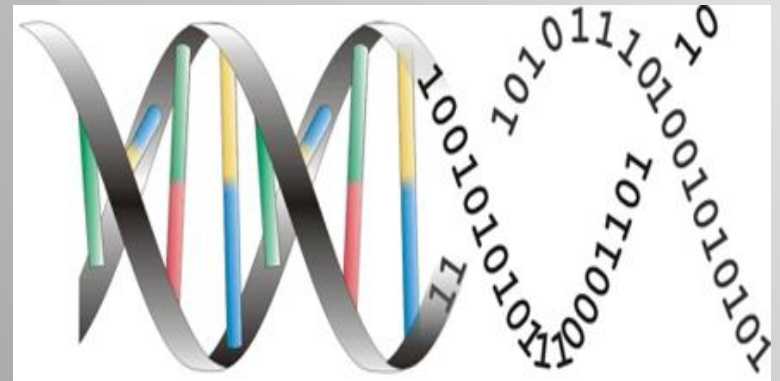
The University of Georgia



Programming for Computational & Systems Biology

Instructor: Paul Xie

Tue. & Thr. 9:35~10:50



Overview of Course

- Introduction to programming skills in computational and systems biology. Topics include **real world examples**, such as **processing** genome or proteome data, and **analyzing** large-scale data. The idea of “big data” will be emphasized to help students with their coding skills to **discovering** new knowledge in biomedical sciences and **solving** biomedical problems.

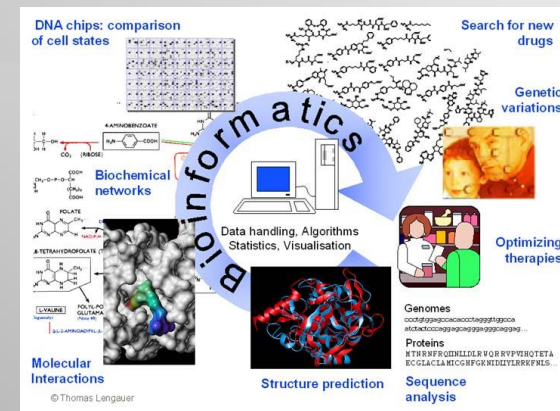
Course Format

- Computer labs + lectures (3 hours/week)
- Coding concepts (some knowledge about biological data)
- 6-8 assignments, published on eLC, (30-40%)
 - Please upload them by the due days
- Paper review (10-20%)
- 1 term paper (50-60%)

Topics



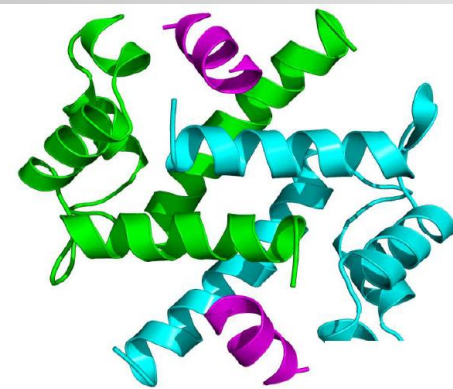
- -Omics data, e.g. Genomics, Proteomics...
- Basic Programming, e.g. I/O, variables, string, loop, regular express, array...etc.
- Data retrieval & processing
- Data analysis
- Model & building model
- Clustering and Classification
- Prediction





Regular Expression Amino Acid & Protein Structure

Instructor: Paul Xie (7)



Last Week

- Sequence alignment
 - Global (N-W) and Local (S-W)
- Statistics model & Blast
- List
- Random number

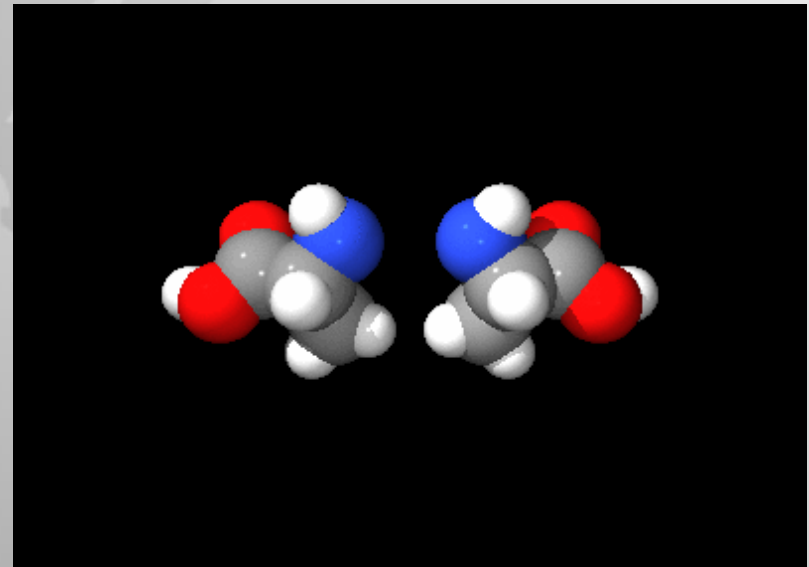
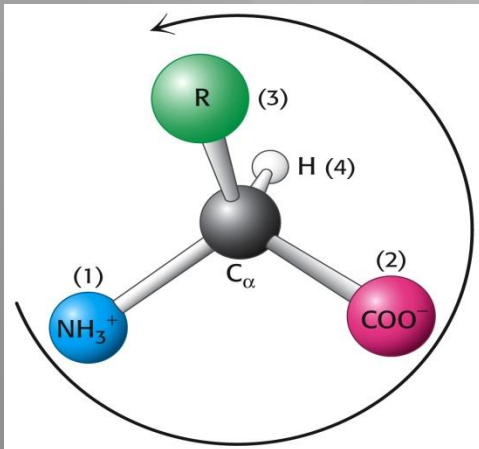
Histone H1 (residues 120-180)

HUMAN	KKASKPKKAASKAPT	KKPKATPVKKAKKKLAATP	KKAKPKTVKAKPVKASKPKKAKPVK
MOUSE	KKAAKPKKAASKAPS	KKPKATPVKKAKKKPAATP	KKAKPKVVKVPVKASKPKKAKTVK
RAT	KKAAKPKKAASKAPS	KKPKATPVKKAKKKPAATP	KKAKPKIVKVKPVKASKPKKAKPVK
COW	KKAAKPKKAASKAPS	KKPKATPVKKAKKKPAATP	KKTKKPKTVKAKPVKASKPKKTKPVK
CHIMP	KKASKPKKAASKAPT	KKPKATPVKKAKKKLAATP	KKAKPKTVKAKPVKASKPKKAKPVK
	:**:	*****:*****	**.******:*
NON-CONSERVED AMINO ACIDS	Conservative	Conservative	Non-conservative Conservative Non-conservative Semi-conservative Non-conservative

This Week



- Amino Acids & Protein
- Regular expression



Python Time



Syntax

- Regular expression

- `import re`

- `re.search`

- Example:

- `print(x)`

- `print(x.string)`

- `Print(x.group())`

- `re.match`

- `re.findall`

- `re.split`

- `re.sub`

Example

- S = 'it rains in Spain'
 - re.search(pattern, string)

- Regular expression
 - metacharacters
 - . ^ \$ * + ? { } [] \ | ()
 - [] → any character in the bracket, e.g. [A-Z]
 - () → the specific pattern should be stored
 - | → or e.g. A|T|C|G == [ATCG]
 - \ → escape the original meaning e.g. \d == [0-9],
\w == [a-zA-Z0-9], \s == any spaces, \n == new
line

Example

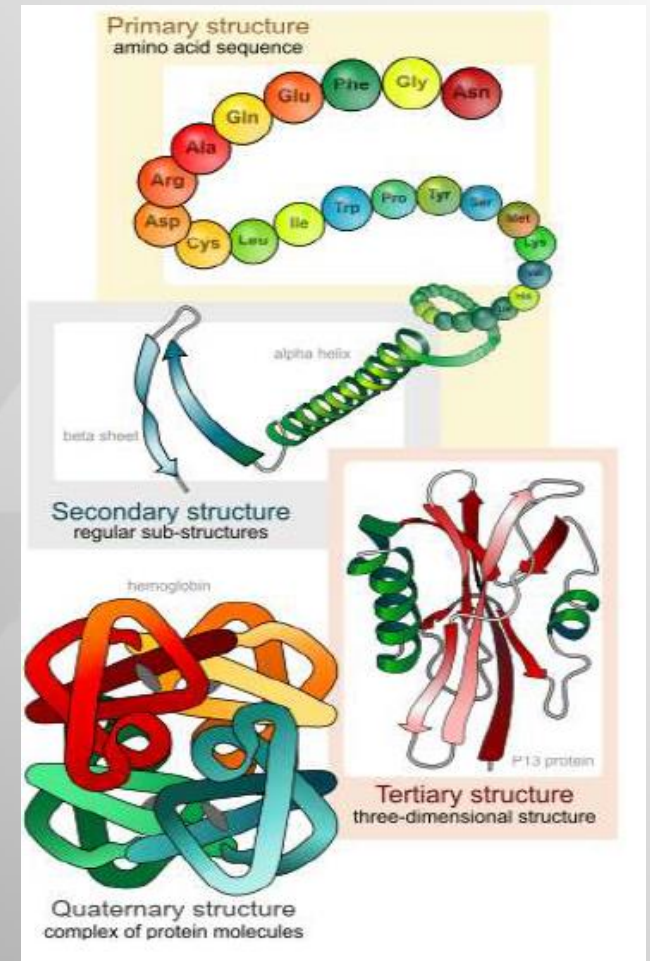
- fruit = ['1 @pple#\$&', '3 oranges', '5 grapes', '7 peaches', '9 blue berries', '11 pineapples', '12 lemons']



Structures



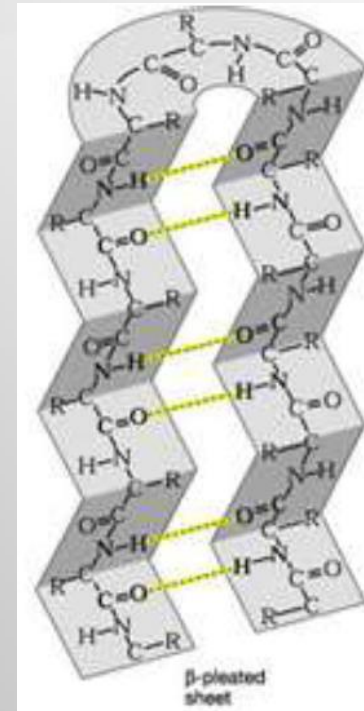
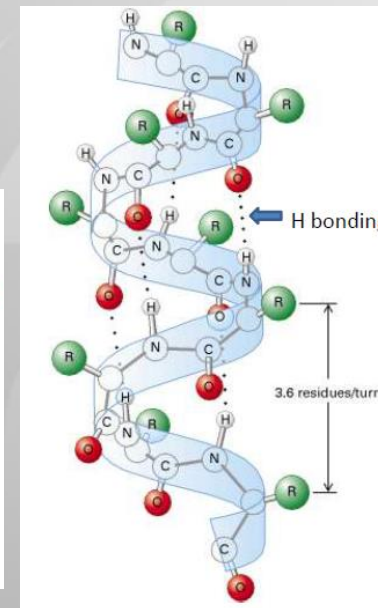
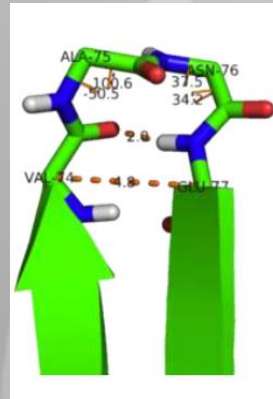
- Protein structures
 - Primary (sequence)
 - Secondary (helix, sheet, turn)
 - Motifs (Super-secondary)
 - Tertiary
 - Quaternary



Secondary Structures



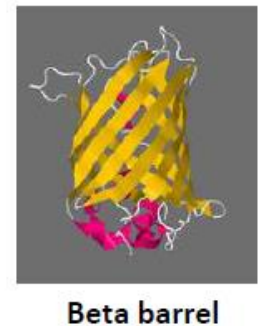
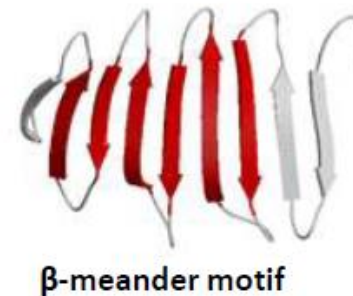
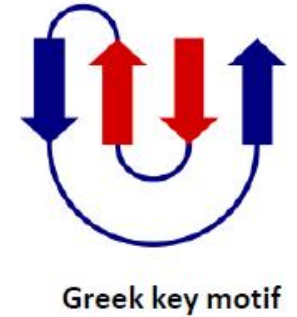
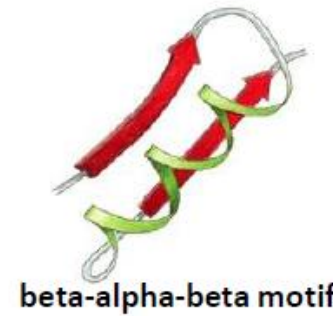
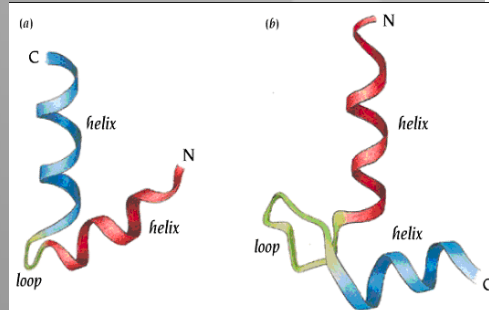
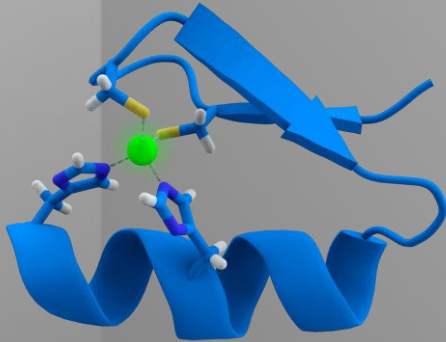
- Secondary structures
 - Alpha helix
 - Beta sheet
 - Tight turn
 - Loop



Motifs

- Beta-alpha-beta
- HLH
- Greek key
- Zinc finger

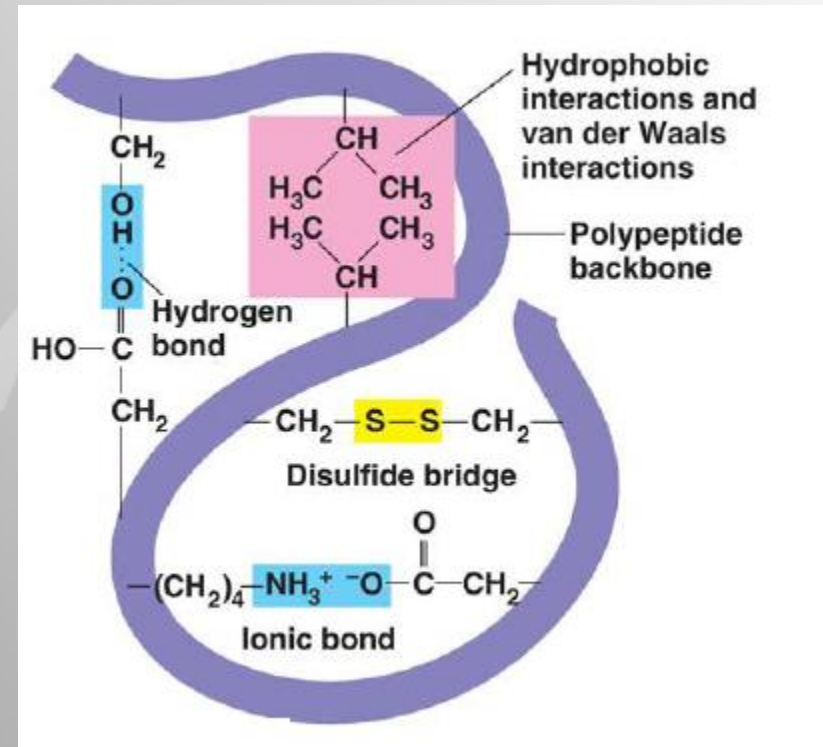
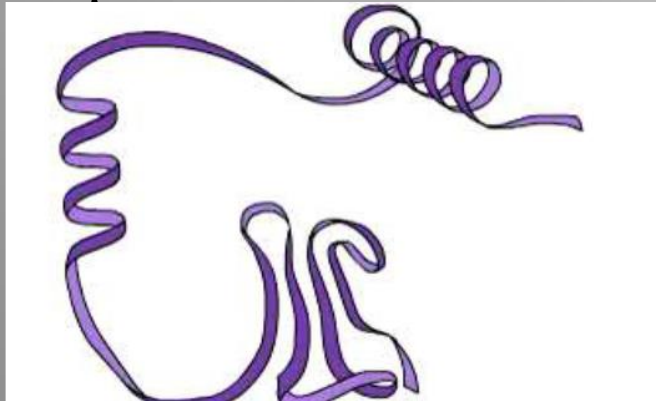
– 2 His + 2 Cys + Zn



Interactions

- Interactions stabilize 3D structures

- Hydrogen bond
- Ionic bonds
- Disulfide bonds
- Hydrophobic interactions



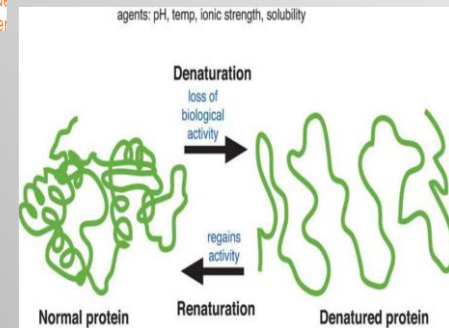
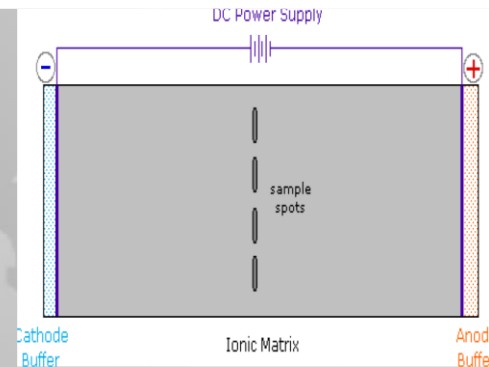
Classification

- Enzymes (Ligase)
- Transport proteins (Hemoglobin)
- Storage Proteins (Ferritin)
- Contractile/Motile Proteins (Actin)
- Regulatory Proteins (Hormones, insulin)
- Receptors (Estrogen receptor)

- Henderson-Hasselbalch Equation

Henderson-Hasselbalch Equation:
$$\text{pK}_a = \text{pH} + \log \frac{[\text{HA}]}{[\text{A}^-]}$$

- Isoelectric point (pI)
- Electrophoresis
- Folding & denature
 - Alcohol, salt, acid & base



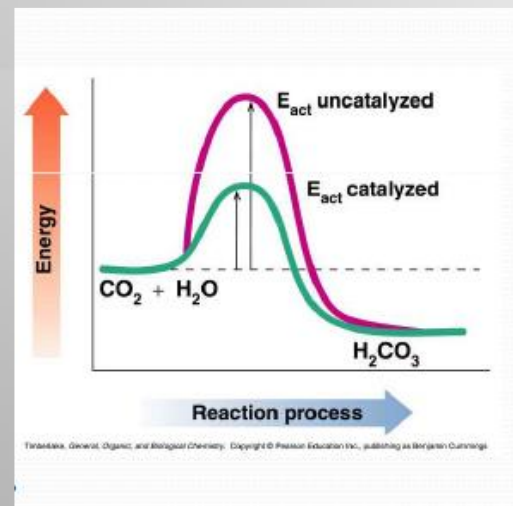
Properties and Conventions Associated with the Standard Amino Acids

Amino acid	Abbreviated names		M_r	pK_a values			pI	Hydropathy index ^a	Occurrence in proteins (%) [†]
				pK_1 (—COOH)	pK_2 (—NH ₃ ⁺)	pK_R (R group)			
Nonpolar, aliphatic R groups									
Glycine	Gly	G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala	A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val	V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu	L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile	I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met	M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups									
Phenylalanine	Phe	F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr	Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp	W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups									
Serine	Ser	S	105	2.21	9.15		5.68	−0.8	6.8
Proline	Pro	P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr	T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys	C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn	N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln	Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups									
Lysine	Lys	K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His	H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg	R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups									
Aspartate	Asp	D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu	E	147	2.19	9.67	4.25	3.22	−3.5	6.3

Enzyme



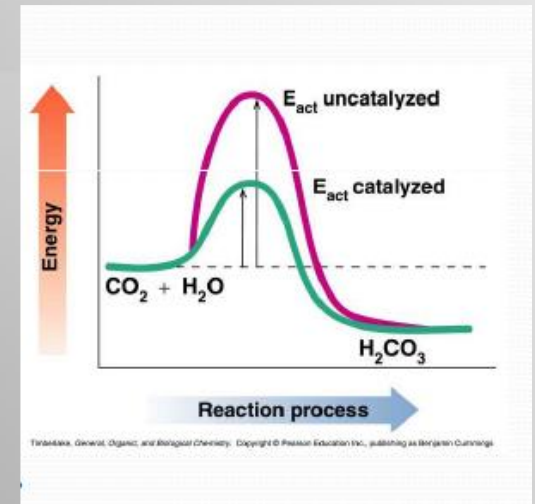
- In Greek means in living (en=in, zyme=living)
- Biocatalysts or Organic catalysts
- Increase the reaction rate
 - By lowering the energy of activation
- Catalyze the chemical reactions in the cell



Enzyme



- In Greek means in living (en=in, zyme=living)
- Biocatalysts or Organic catalysts
- Increase the reaction rate
 - By lowering the energy of activation
- Catalyze the chemical reactions in the cell



Nomination & Classification

- Usually ends in –ase (DNA Polymearse)
- Substrate & reaction catalyzed
 - Lactate dehydrogenase
- EC (Enzyme commission) no.
 - e.g. 1.1.1.1 alcohol dehydrogenase

EC Number

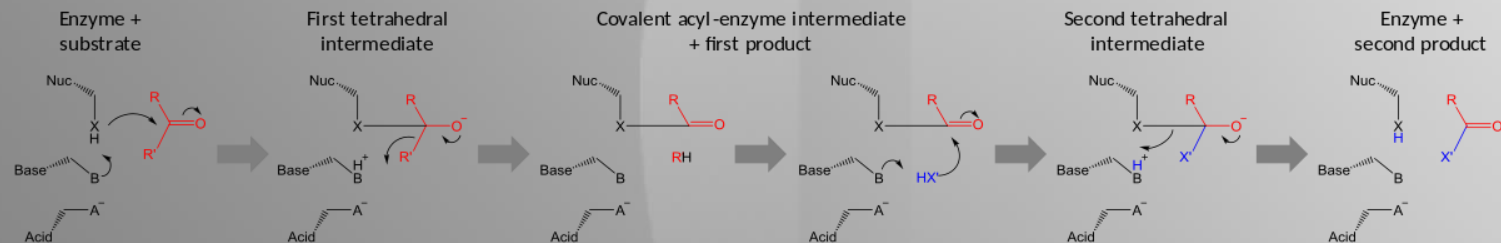
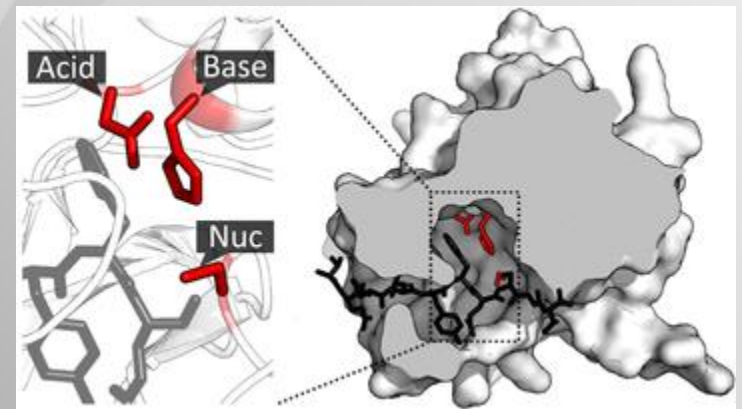
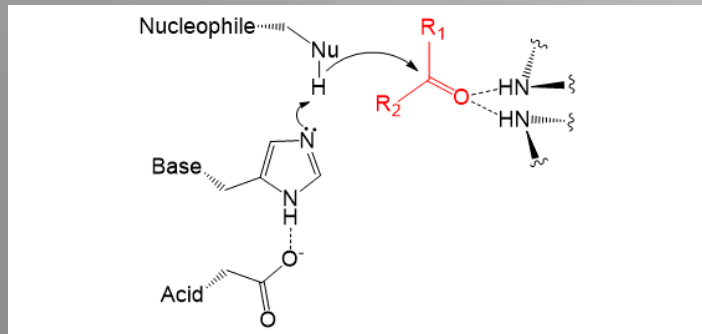
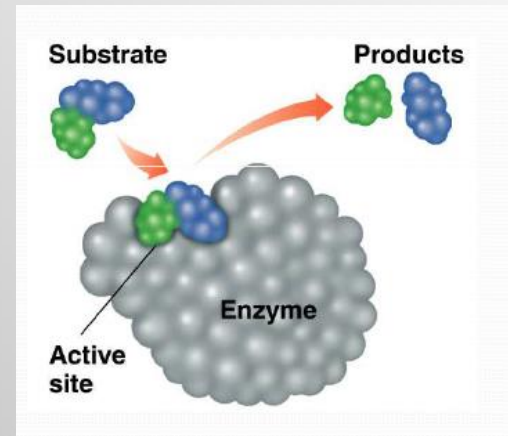
- 6 groups

Top-level EC numbers^[5]

Group	Reaction catalyzed	Typical reaction	Enzyme example(s) with trivial name
EC 1 <i>Oxidoreductases</i>	To catalyze oxidation /reduction reactions; transfer of H and O atoms or electrons from one substance to another	$AH + B \rightarrow A + BH$ (reduced) $A + O \rightarrow AO$ (oxidized)	Dehydrogenase, oxidase
EC 2 <i>Transferases</i>	Transfer of a functional group from one substance to another. The group may be methyl-, acyl-, amino- or phosphate group	$AB + C \rightarrow A + BC$	Transaminase, kinase
EC 3 <i>Hydrolases</i>	Formation of two products from a substrate by hydrolysis	$AB + H_2O \rightarrow AOH + BH$	Lipase, amylase, peptidase
EC 4 <i>Lyases</i>	Non-hydrolytic addition or removal of groups from substrates. C-C, C-N, C-O or C-S bonds may be cleaved	$RCO_2COOH \rightarrow RCOH + CO_2$ or $[X-A+B-Y] \rightarrow [A=B + X-Y]$	Decarboxylase
EC 5 <i>Isomerases</i>	Intramolecule rearrangement, i.e. isomerization changes within a single molecule	$ABC \rightarrow BCA$	Isomerase, mutase
EC 6 <i>Ligases</i>	Join together two molecules by synthesis of new C-O, C-S, C-N or C-C bonds with simultaneous breakdown of ATP	$X + Y + ATP \rightarrow XY + ADP + P_i$	Synthetase

Catalysis

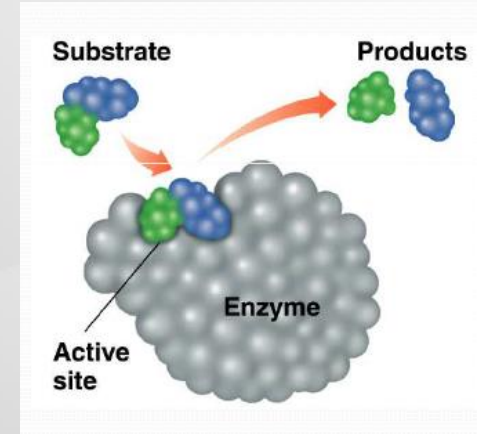
- Active sites
- Catalytic triad
 - Hydrolase or transferase
 - Acid, base, & nucleophile



Inhibition



- Different ways to inhibit enzymes
 - Drug binding
 - Competitive inhibition
 - Conformational change
 - Uncompetitive inhibition
 - Covalent bonding
 - Denature
 - Change pH

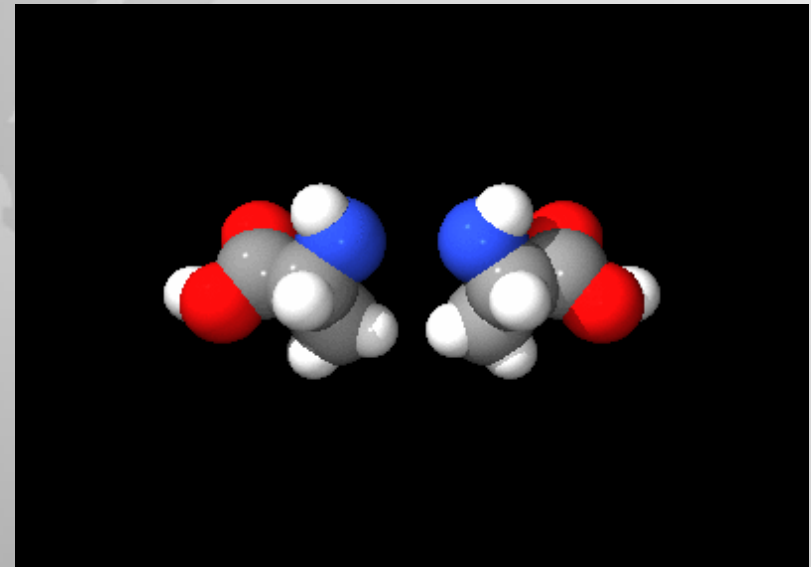
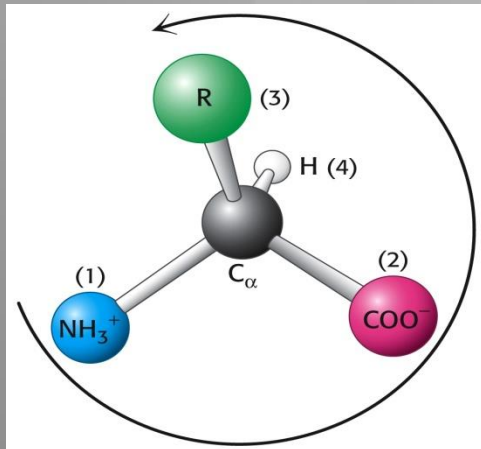




This Week



- Amino Acids
- Regular expression



Functions of amino acids



- A variety of roles in metabolism
 - the building blocks of proteins
 - forming parts of coenzymes
 - as precursors for the biosynthesis of molecules such as heme

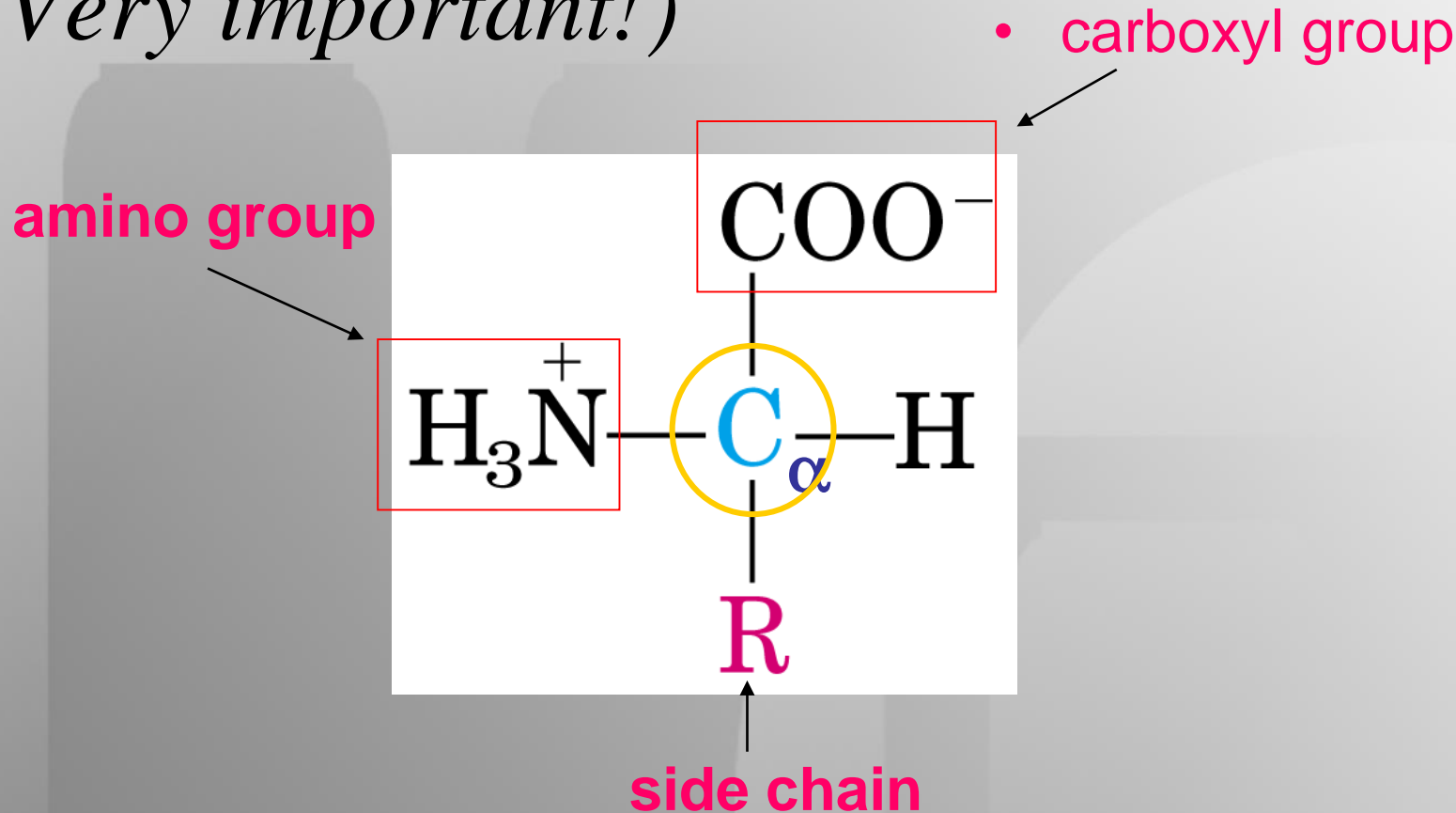
Standard and nonstandard amino acids

- More than 300 different amino acids have been described in nature.
 - Standard α -amino acids:
 - Only 20 are commonly found as constituents of proteins
 - Nonstandard amino acids:
 - Amino acid derivatives found in protein
 - Non-protein amino acid

- Why are amino acids uniquely suited to their role as the building blocks of proteins?

General structure of α -amino acids

(Very important!)



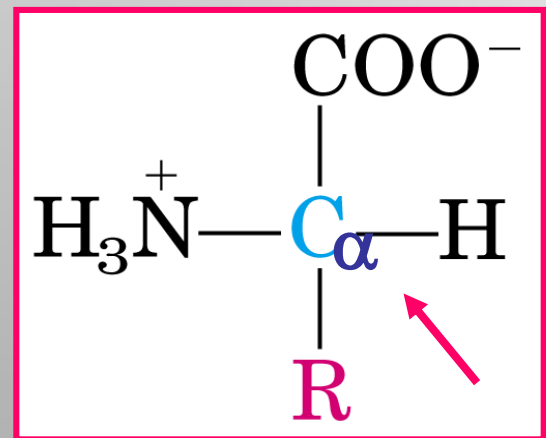
(or R group, R="Remainder of the molecule")



Stereochemistry of amino acids

- Configuration

- For all the common amino acids except Glycine, the α -carbon is bonded to four different groups.
- The α -carbon atom is thus a chiral center.



Chiral



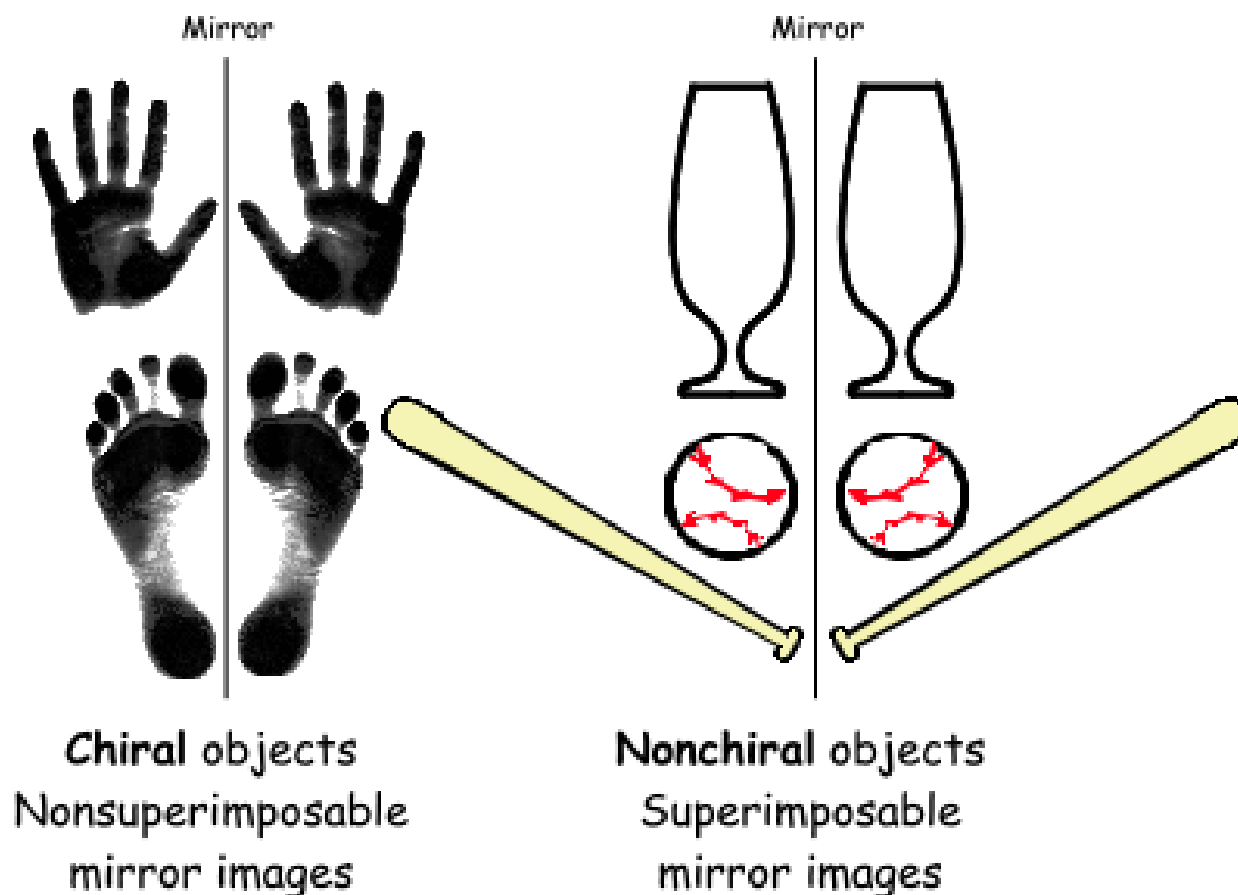
- Chiral (from Greek *cheir*, meaning “hand”):
 - An object or a system cannot be superimposed on its mirror image.
 - One hand does not match the other when superimposed.



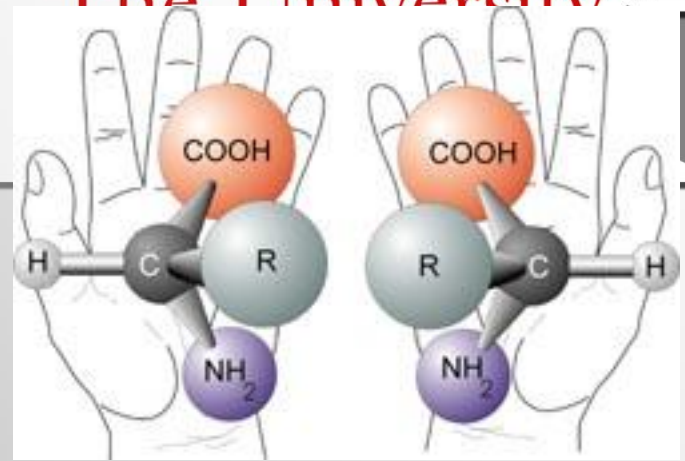
Copyright © 2007 Pearson Prentice Hall, Inc.

CHIRALITY

An object that cannot be superimposed on its mirror image is called chiral

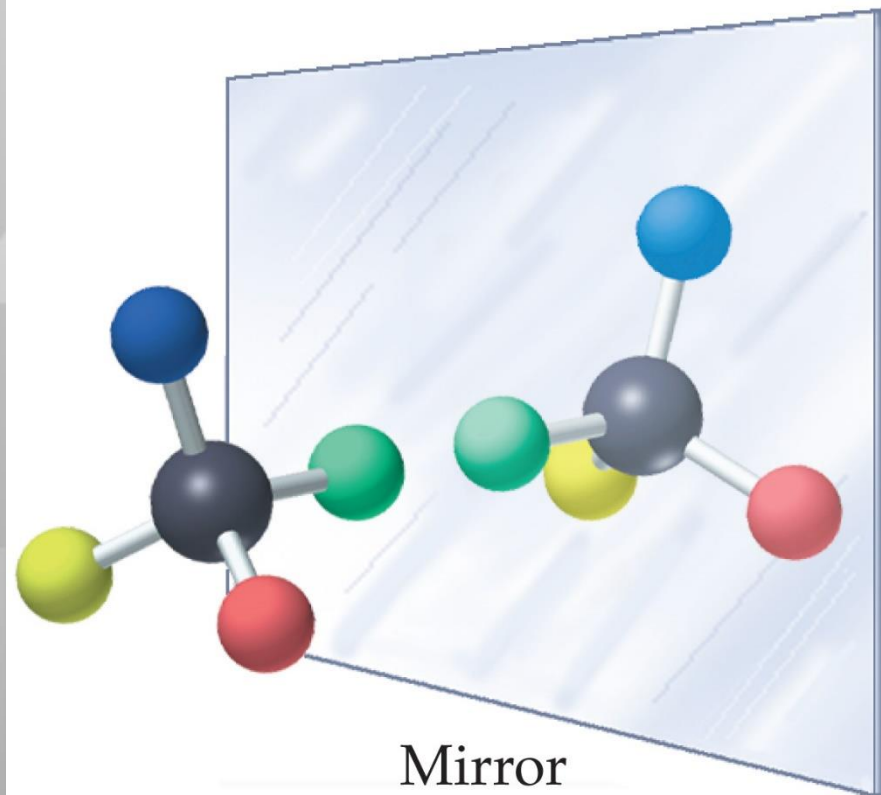


Chiral molecule

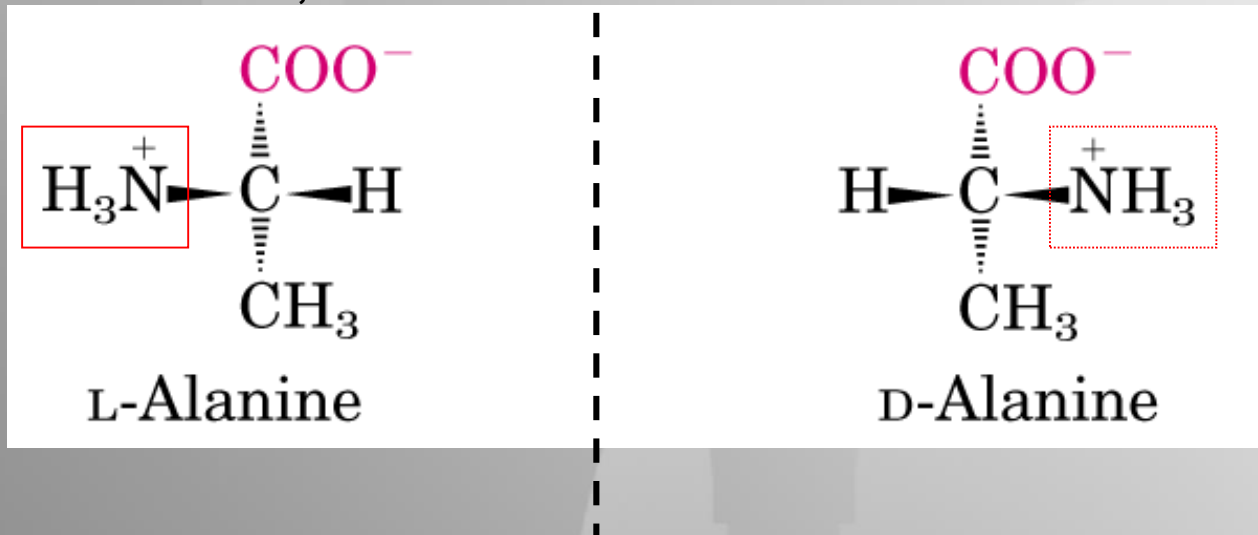


- Chiral molecule:

- if a molecule has an atom bonded to **four different groups**, it can be chiral.



- ▶ The two enantiomers of each amino acid are designated by D,L system according to the D- and L-glyceraldehyde.
- ▶ D: Dextrorotation; L: Levorotation



- Only the *L-amino acids* have been found in proteins.
- (D-isomers have been found only in small peptides of bacteria cell walls or in some peptide antibiotics).

The Classification of 20 standard Amino Acids

(*very important!!!*)

- The name and abbreviation of amino acids
 - All the AAs were given a trivial (common) name.
 - Glutamate from wheat gluten.
 - Tyrosine from cheese (“tyros” in Greek).
 - Each AA is given a 3 letter abbreviation and 1 letter symbol.
 - They often the first three letter and the first letter. When there is confusion, an alternative is used.
- They should be remembered !

Classification of Amino Acids

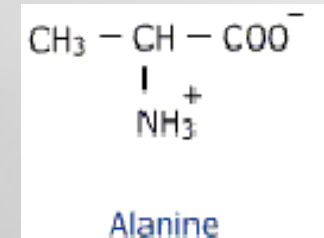
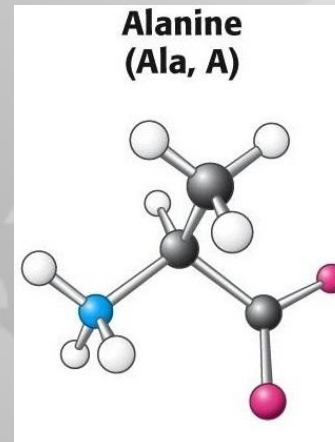
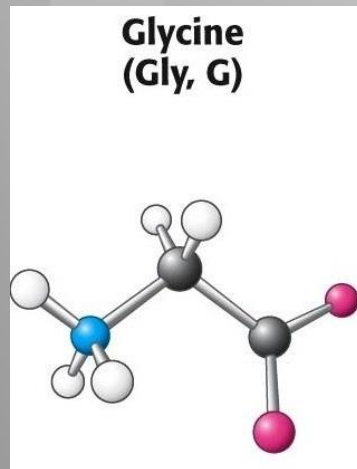
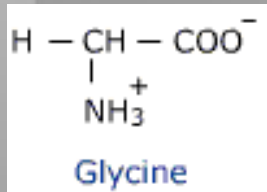
- Based on **polarity**
 - Nonpolar, aliphatic amino acids
 - Polar, uncharged amino acids
 - Acidic amino acids
 - Basic amino acids
 - Aromatic amino acids

Aliphatic Amino Acids:



a) Mono-amino mono-carboxylic acids:

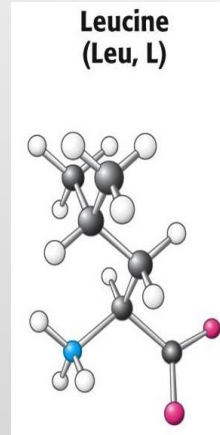
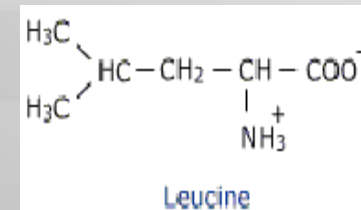
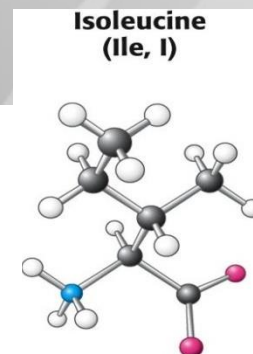
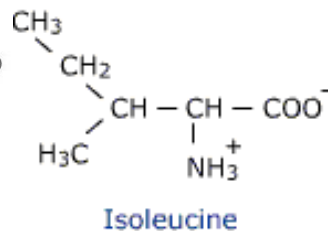
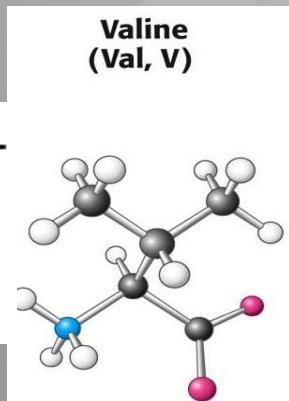
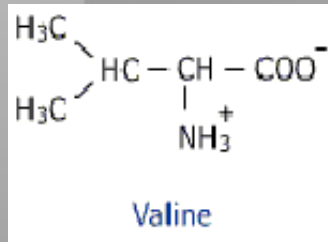
❑ Simple amino acids: Glycine , Alanine



Aliphatic Amino Acids:

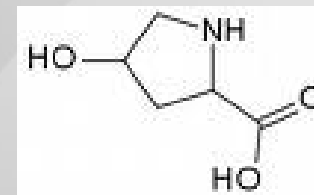
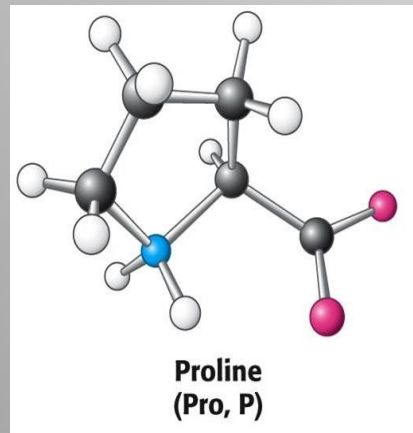
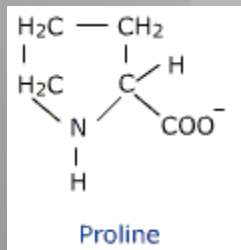
a) Mono-amino mono-carboxylic acids:

❑ **Branched chain amino acids:** Valine, Leucine and Isoleucine



Aliphatic Amino Acids:

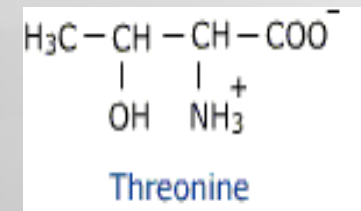
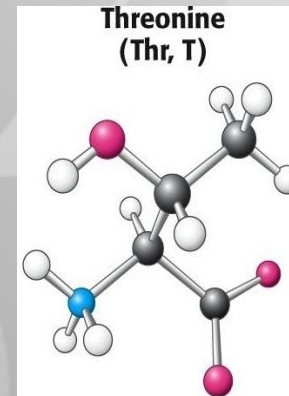
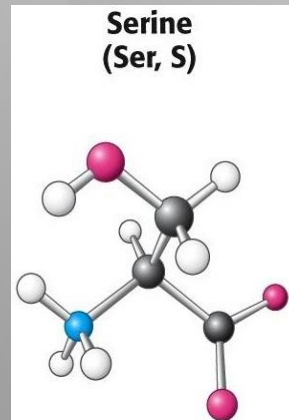
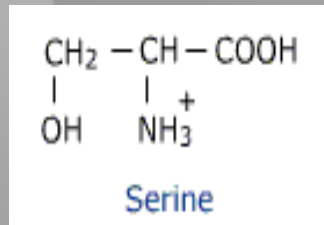
iv) Imino acid- **Proline**



Polar Amino Acids:

a) Mono-amino mono-carboxylic acids:

☐ -OH group-containing amino acids: Serine and Threonine

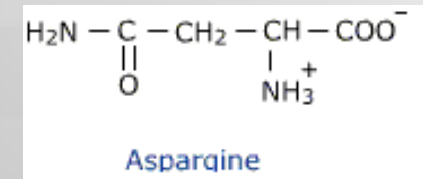
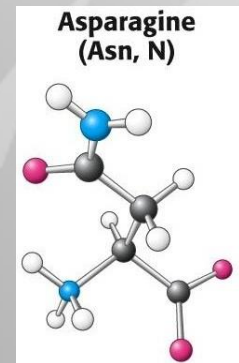
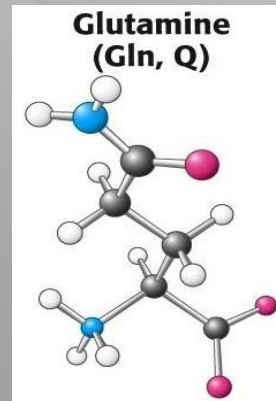
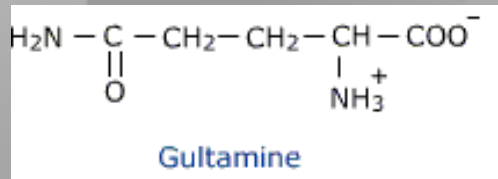


Polar Amino Acids:

a) Mono-amino mono-carboxylic acids:

❑ Amide group-containing amino acids:

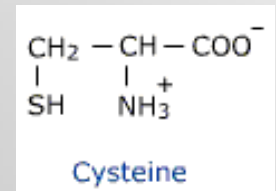
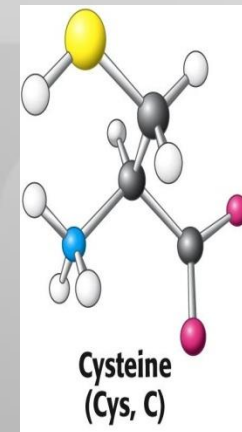
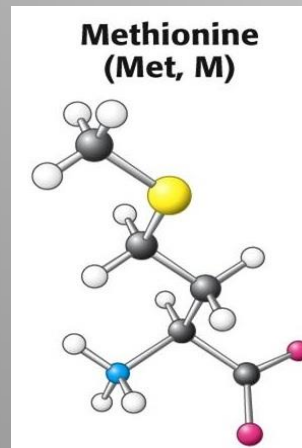
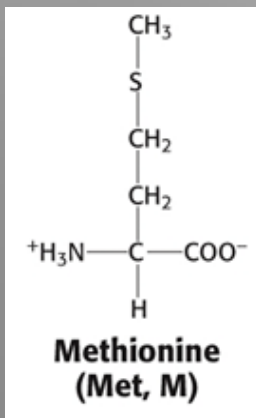
Glutamine and Asparagine



Non-Polar Amino Acids:

a) Mono-amino mono-carboxylic acids:

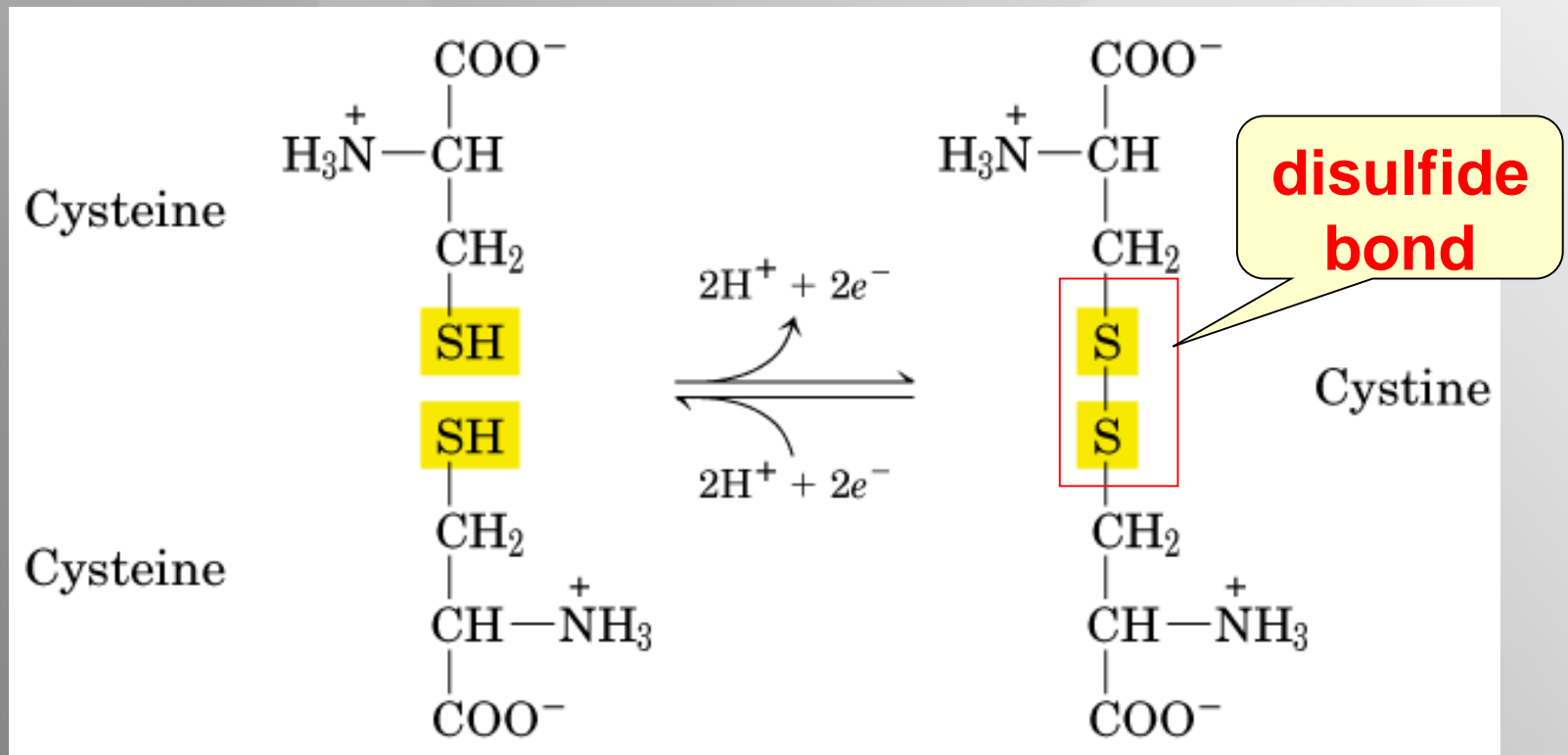
❑ **Sulfur-containing amino acids:** Cysteine, Cystine (Formed by linking of two cysteine residues) and Methionine.



➤ Disulfide bonds

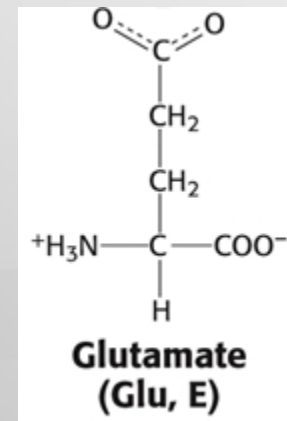
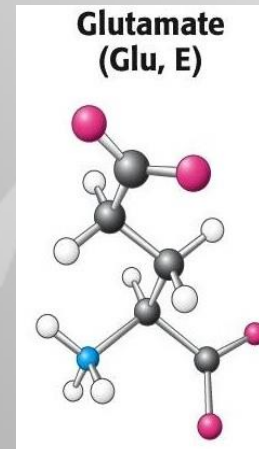
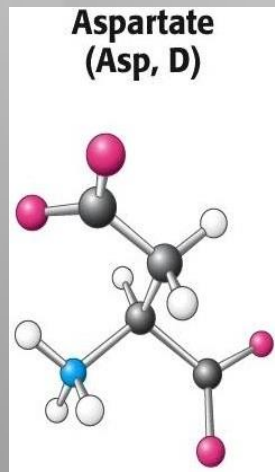
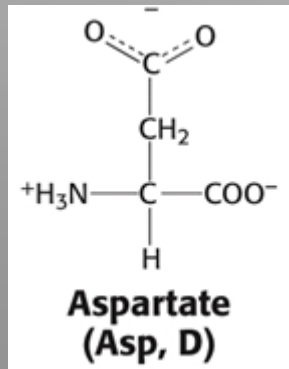
➤ -SH group of two Cys in proteins can be *oxidized to form a covalent disulfide bond*.

➤ Disulfide bonds: play a special role in the structures of many proteins by forming covalent links.



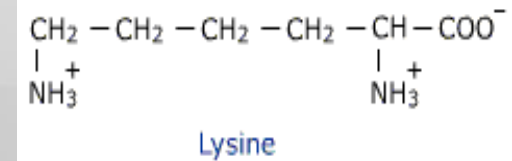
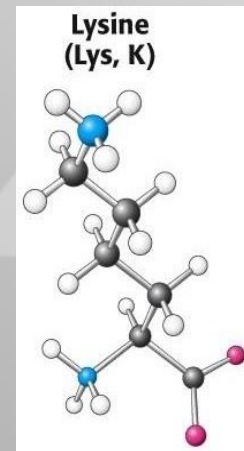
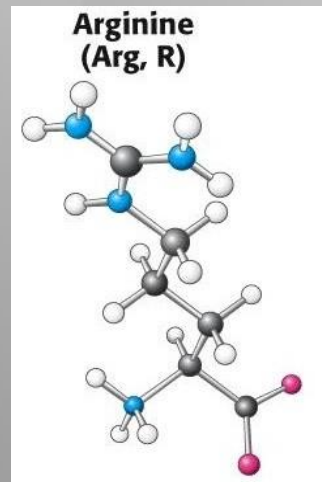
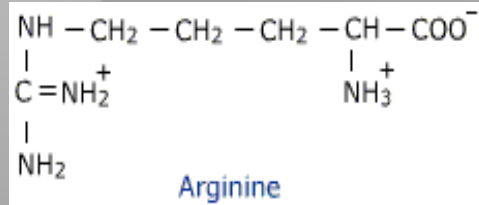
Polar Amino Acids:

b) Mono-amino di-carboxylic acids: Aspartic acid and Glutamic acid



Polar Amino Acids:

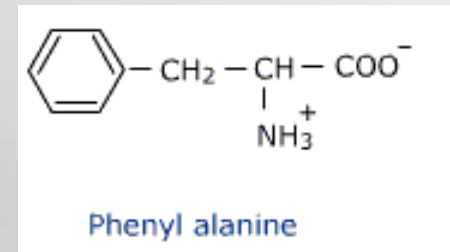
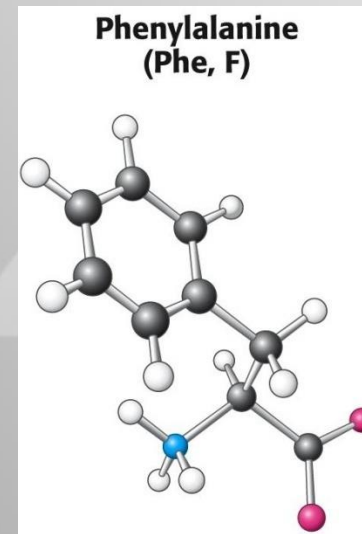
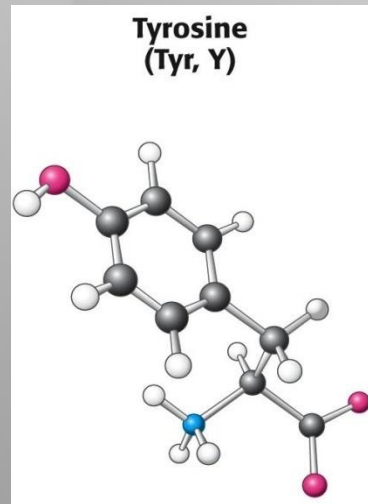
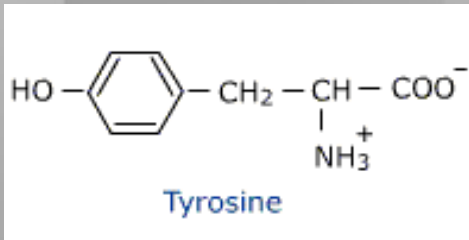
c) Di- basic mono-carboxylic acids: Arginine and Lysine



Polar Amino Acids:

a-3) Aromatic amino acids-

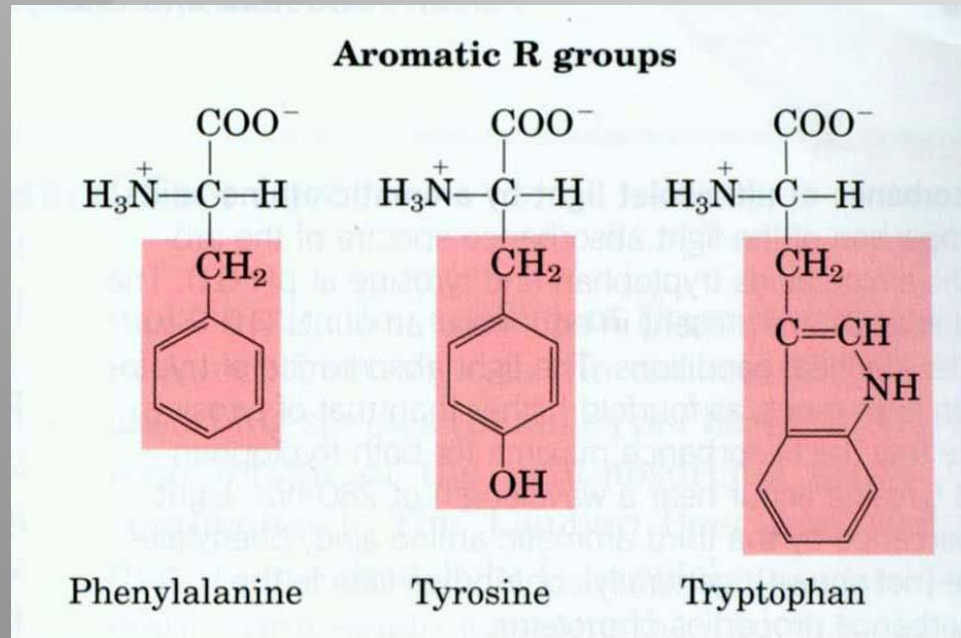
❑ Phenyl alanine and tyrosine



Aromatic amino acids

- **Phe, Tyr and Trp.**

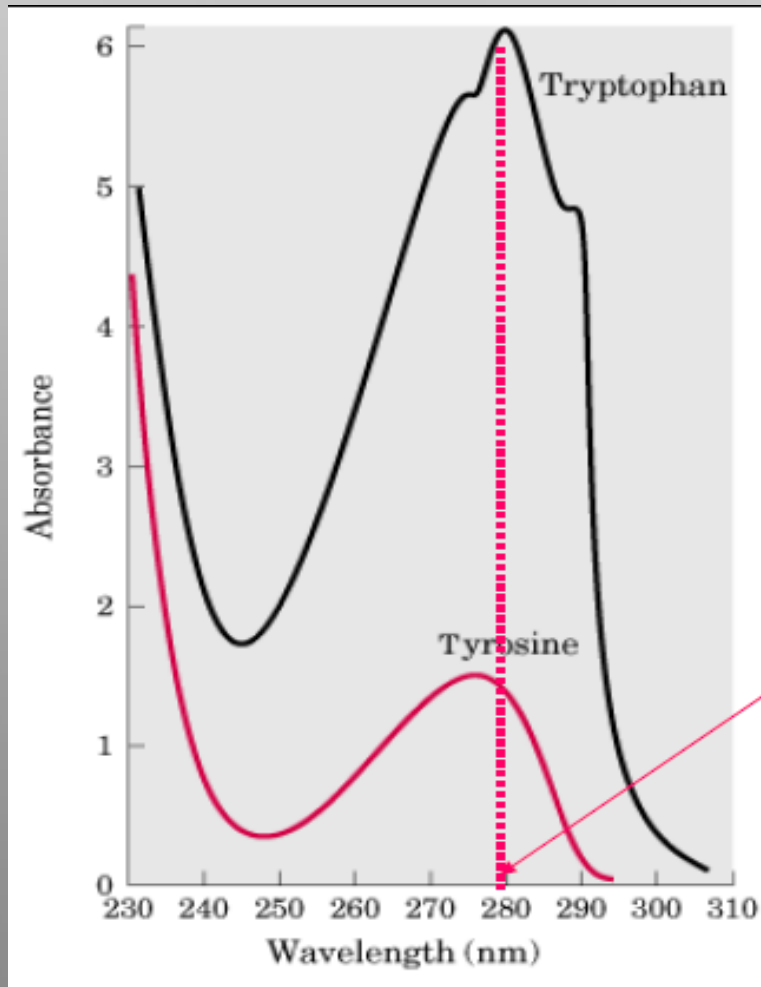
- Phe and Tyr: benzene rings.
- Tryptophan: indole ring.



- The -OH group in Tyr is an important functional group in proteins. (phosphorylation, hydrogen bond, etc), polar



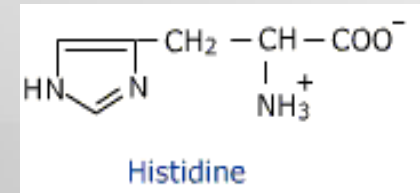
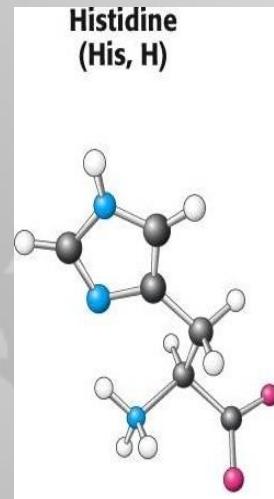
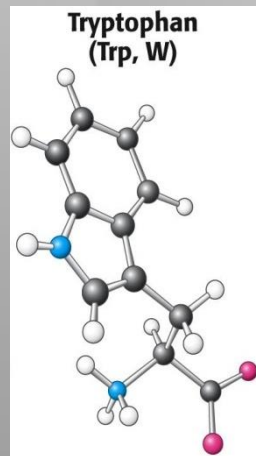
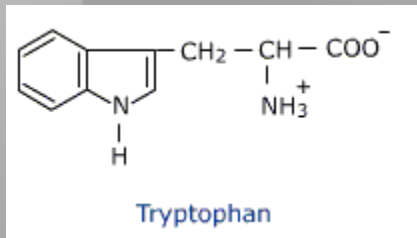
- They are jointly responsible for the light absorption of proteins at 280 nm.



- Proteins in solution absorb UV light with absorbance maximum at 280nm.
- Measuring protein content by photo spectrometry.

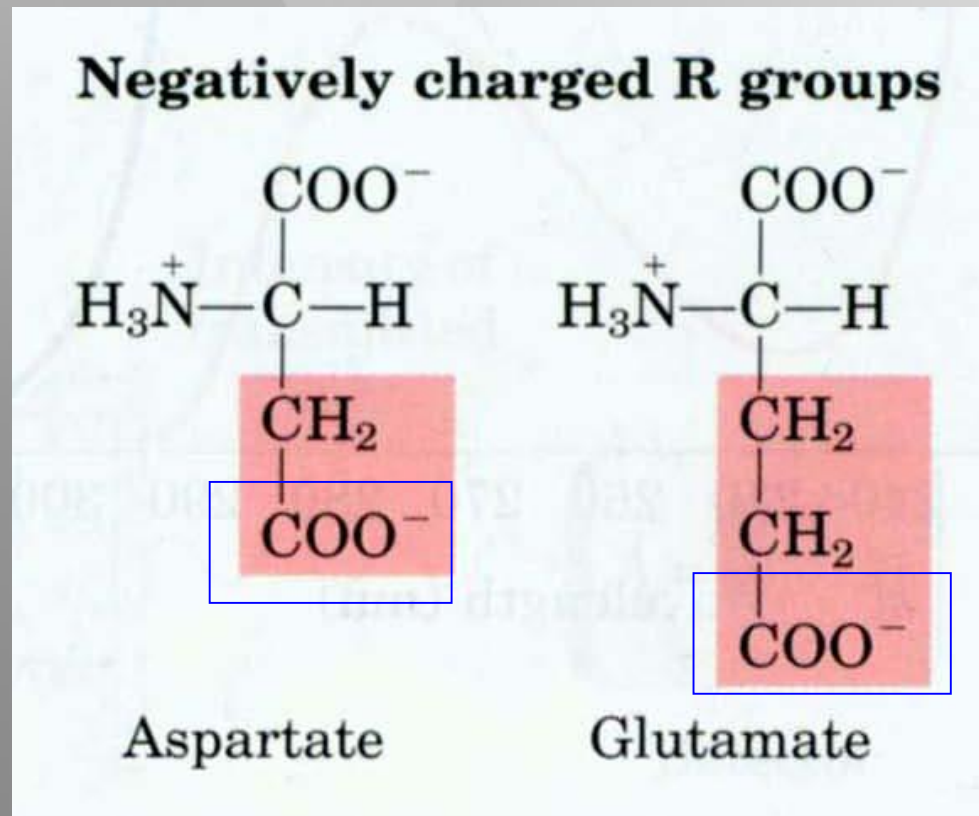
Polar Amino Acids:

a-4) Heterocyclic Amino Acids: Tryptophan and Histidine



Acidic amino acids

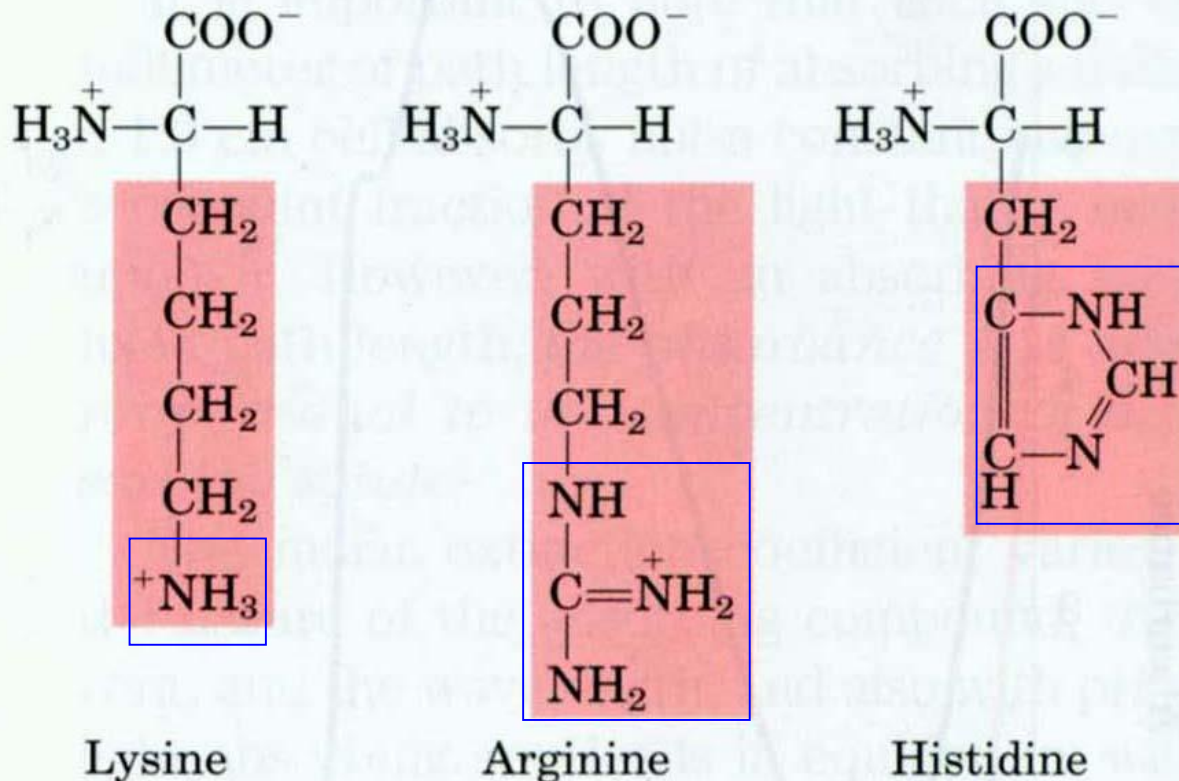
- Asp and Glu
 - Have **carboxyl** in their R groups.



Basic amino acids

➤ Lys, Arg, and His.

Positively charged R groups



➤ R groups

➤ Amino

➤ Guanidino

➤ Imidazole

➤ Positive
charged R
groups at pH 7.0

- **Note these structural features**

1. All 20 are α -amino acids
2. For 19 of the 20, the α -amino group is primary; for **proline**, it is **secondary** (imino acid)
3. Except **glycine**, the α -carbons for 19 of them are asymmetric (or **chiral**).

- **Nonpolar Amino Acids**

Ala Val Leu Ile Gly Pro Phe Trp

- **Polar, Uncharged Amino Acids**

Ser Thr Cys Asn Gln Met Tyr

Aromatic



- **Acidic Amino Acids**

Asp Glu

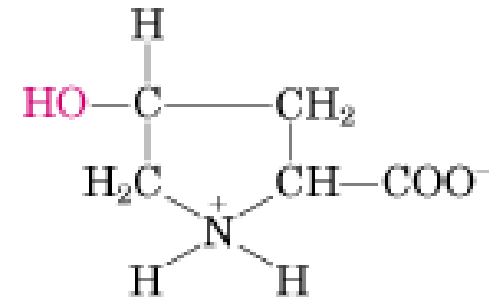
- **Basic Amino Acids**

Lys Arg His

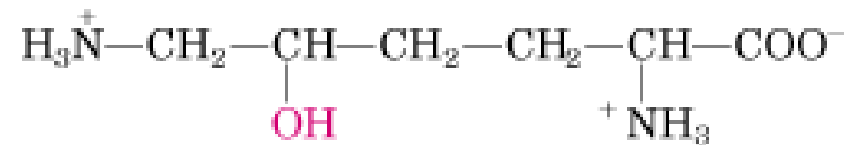
Nonstandard amino acids

- **Amino acid derivatives found in proteins**

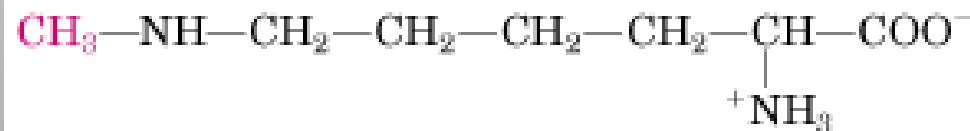
- **4-Hydroxyproline and 5-hydroxylysine in collagen.**
- **6-N-Methyllysine in myosin.**



4-Hydroxyproline

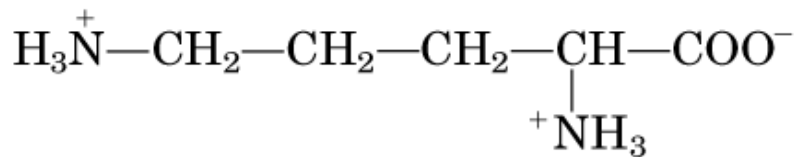


5-Hydroxylysine

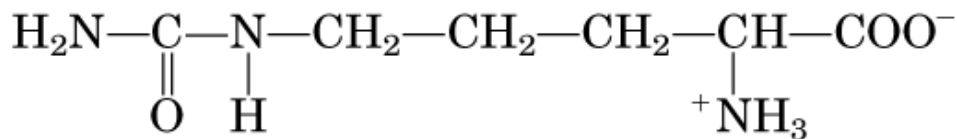


6-N-Methyllysine

Many additional nonstandard amino acids are found in cells, but not in proteins



Ornithine



Citrulline

- Ornithine and citrulline
- Intermediates in amino acid metabolism.

Essential and non-essential amino acids

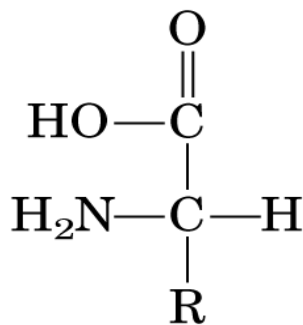


- Essential amino acids (or indispensable amino acids):
 - Cannot be synthesized by the humans, must be supplied in the diet
 - 8: Phe, Val, Thr, Trp, Ile, Met, Leu, Lys
- Semi-essential amino acids:
 - 2: His and Arg
 - Required by infants and growing children

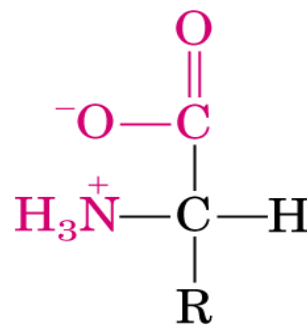
Acid/base properties of AAs (*Very important!*)



- Amino acid has both a basic amine group and an acidic carboxylic acid group.
- In neutral solution (pH 7.0), the amino acid contains a negative charge and a positive charge. It is called a *zwitterion* (German for “hybrid ion”).

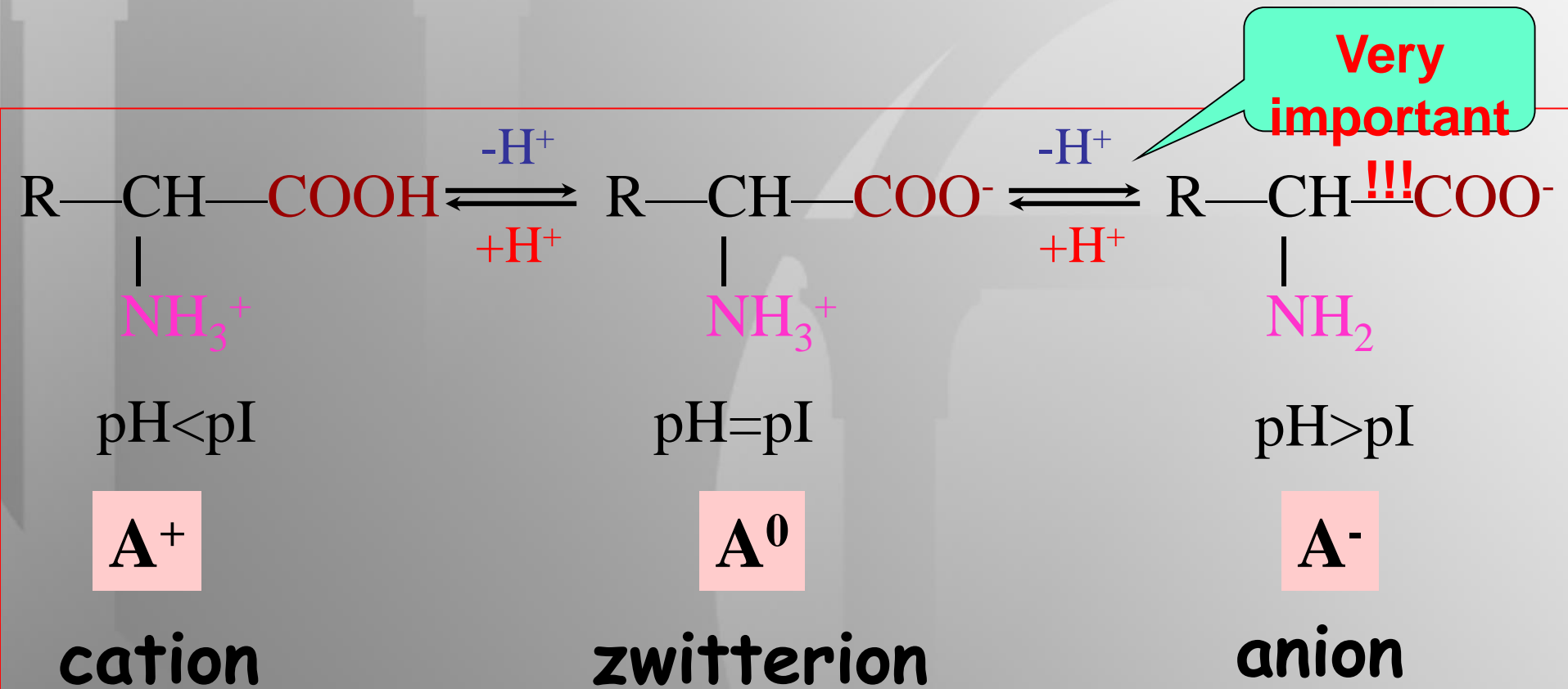


Nonionic
form



Zwitterionic
form

- AAs ionize to various states depending on pH values.
- pI: there is a specific pH (designated isoelectric point, pI) at which an AA has equal positive and negative charge (no net electric charge) .

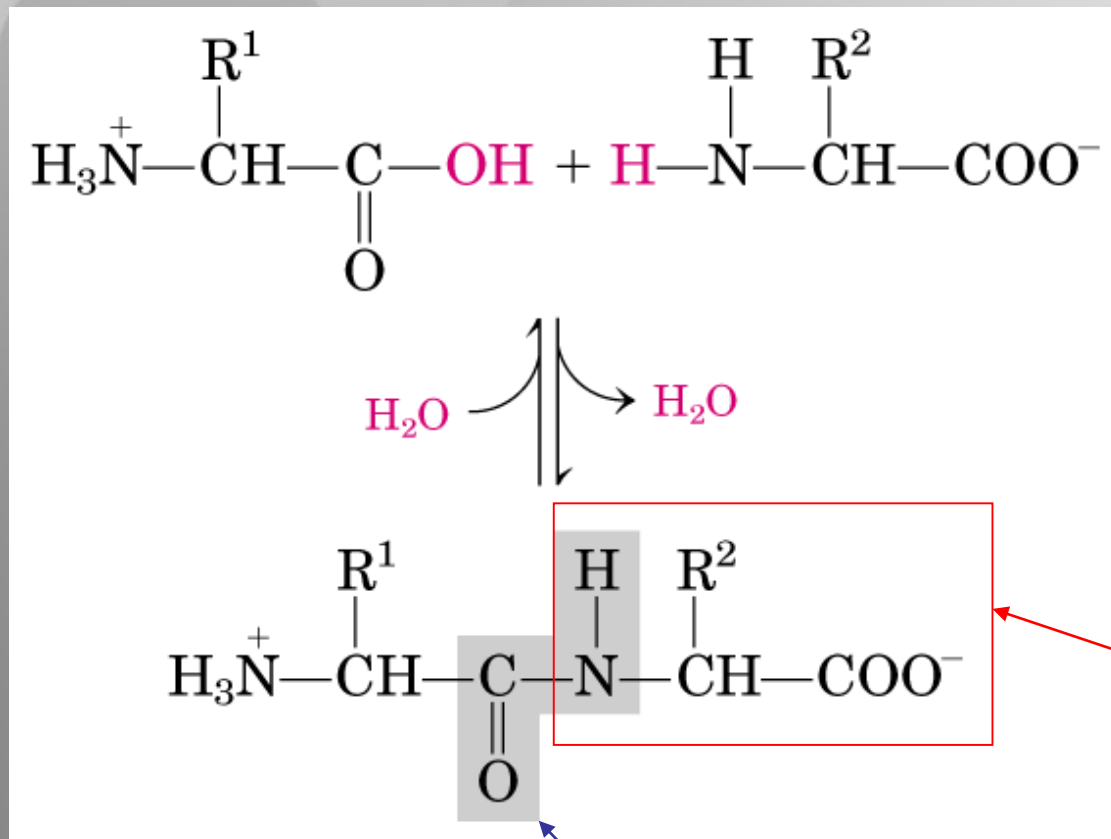


Properties and Conventions Associated with the Standard Amino Acids

Amino acid	Abbreviated names		M_r	pK_a values			pI	Hydropathy index ^a	Occurrence in proteins (%) [†]
				pK_1 (—COOH)	pK_2 (—NH ₃ ⁺)	pK_R (R group)			
Nonpolar, aliphatic R groups									
Glycine	Gly	G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala	A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val	V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu	L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile	I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met	M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups									
Phenylalanine	Phe	F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr	Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp	W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups									
Serine	Ser	S	105	2.21	9.15		5.68	−0.8	6.8
Proline	Pro	P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr	T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys	C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn	N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln	Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups									
Lysine	Lys	K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His	H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg	R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups									
Aspartate	Asp	D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu	E	147	2.19	9.67	4.25	3.22	−3.5	6.3

Polypeptides

Peptide bond: the special name given to the amide bond between the carboxyl group of one amino acid and the -amino group of another.



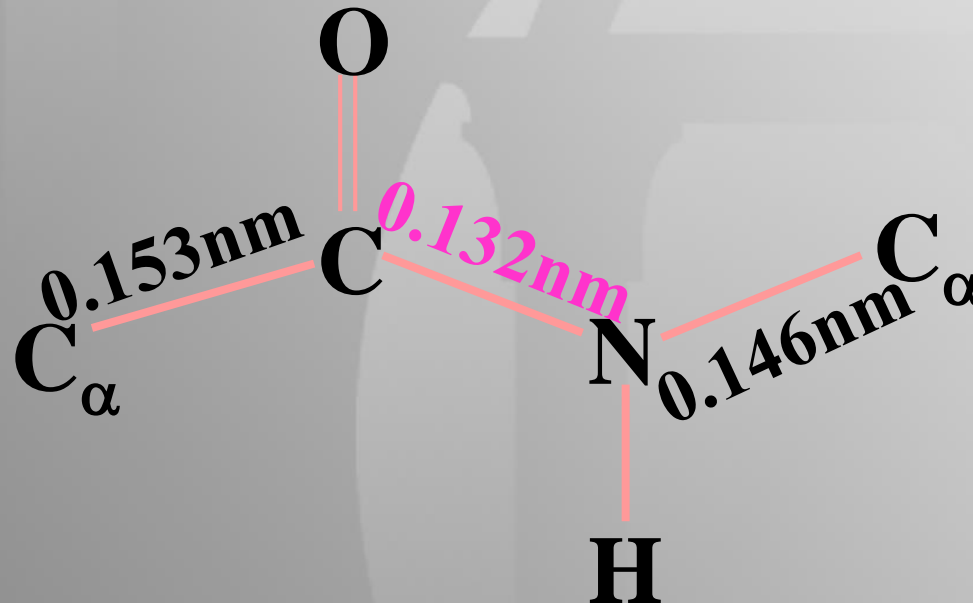
**Amino acid
residue**

peptide bond

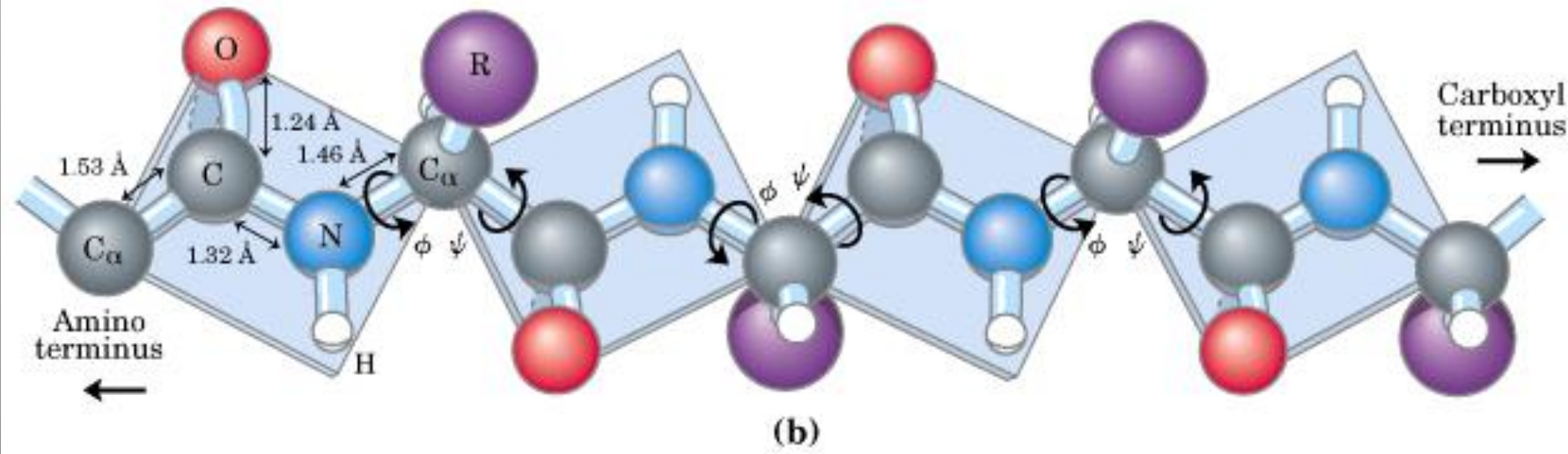
Characteristics of the peptide bond



- The peptide bond **have partial double bond feature**
 - about **0.132nm** (C-N single bond, 0.149nm; C=N double bond, 0.127nm),
 - *rigid and unable to rotate freely.*



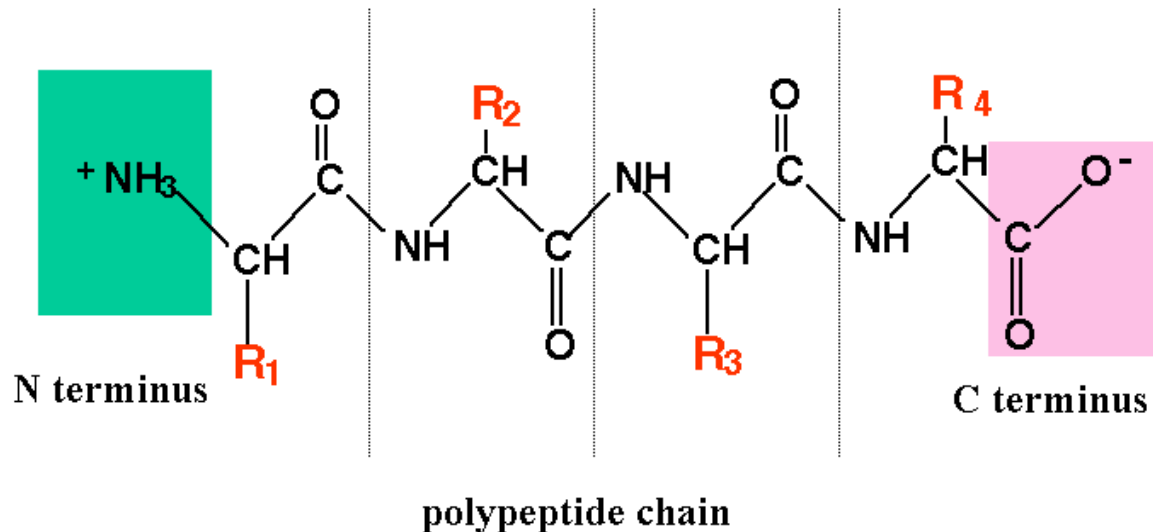
- The peptide bond is planar, trans-configuration and uncharged.
- Peptide plane: the six atoms attached to the peptide bond are coplanar.
- the carbonyl oxygen and the amide hydrogen are in trans positions.



The peptide chain is directional.

- Amino-terminal or N-terminal: the end having a free α -amino group.
- Carboxyl-terminal or C-terminal: the end having a free α -carboxyl group.
- By convention, the N-terminal is taken as the beginning of the peptide chain, and put at the left (C-terminal at the right).

Peptide = chain of amino acids



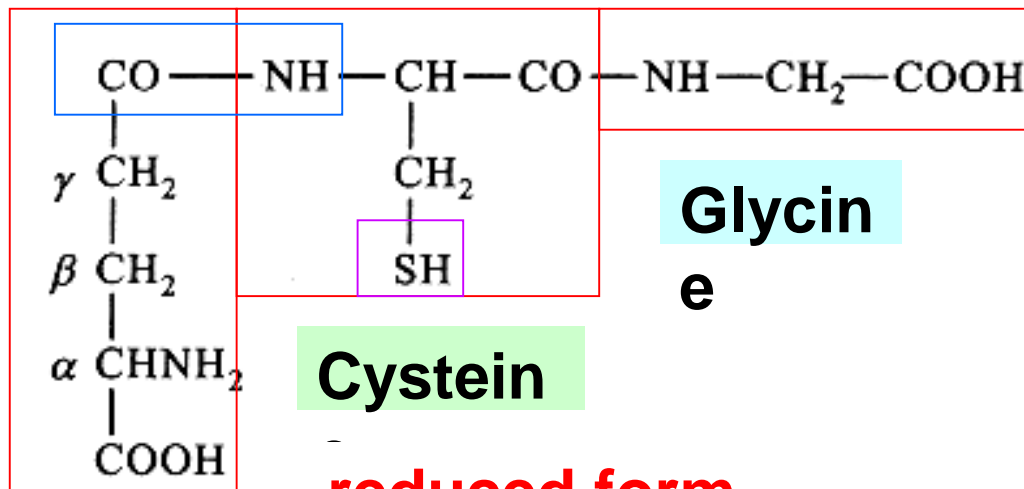
- Peptides can be classified according to how many amino acids they contain
 - **Dipeptide:** 2 amino acid residues, **tripeptide:** 3 residues, and so on
 - **Oligopeptide:** 12~20 residues
 - **Polypeptide:** many residues

Biologically important peptide

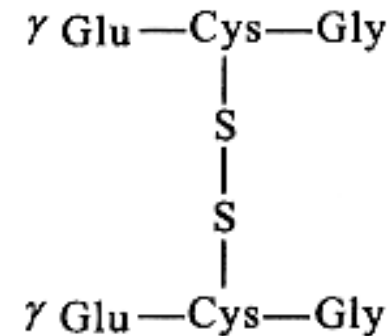


- Glutathione (GSH)

- Tripeptide: glutamic acid, cysteine and glycine;
- Function: important in biological oxidation-reduction reactions, has reduced and oxidized form.
- It's the most important molecule you need to stay healthy and prevent disease.



reduced form
GSH



oxidized form
glutathione



PDB



- Protein databank
- 162,269 entries
(4/6/2020)

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

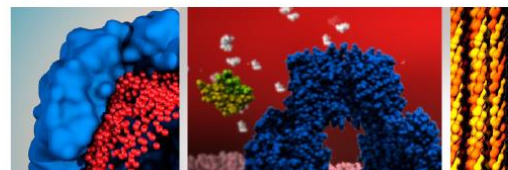
A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

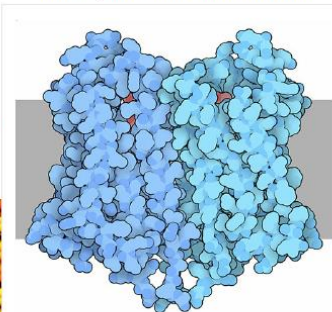
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

New Video: What is a Protein?



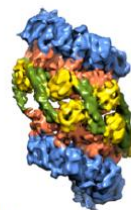
January Molecule of the Month



Opioid Receptors

Latest Entries

As of Tuesday Jan 09



EMD-3894

EMDB Entry

Regulatory coiled-coil domains promote head-to-head assemblies of AAA+ chaperones essential for tunable activi...

Features & Highlights



wwPDB News
A new FTP re



Improved Te
Simple text se
now powered

News

Publications ▼

wwPDB News: Time-stamped Copies of the PDB Archive Available
A snapshot of the PDB archive (<ftp://ftp.wwpdb.org>) as of January 1, 2018 has been added to <ftp://snapshots.wwpdb.org/> and <ftp://snapshots.pdbj.org/>. » 01/09

A Year of Antimicrobial Resistance » 01/02

Season's Greetings » 12/26

Access Irving Geis' Early Molecular Images in 3D » 12/19

Browse the Biology of Plants and More at PDB-101 » 12/12

Overview of PDB Validation Reports published in Structure » 12/05

Structural Databases as Teaching Tools » 11/28

New Video Answers "What is a Protein?" » 11/21

PDB Format

The University
of Georgia



- The PDB format consists of a collection of fixed format records that describe :

- Atomic coordinates,
- Chemical and biochemical features
- Experimental details of the structure determination
- Some structural features such as

- Secondary structure assignments,

- Hydrogen bonding

- Biological assemblies

- Active sites



PDB Format



```

SHEET 3 L 3 THR H 212 VAL H 214 -1 0 THR H 212 N HIS H 207
SHEET 1 M 3 THR H 158 TRP H 161 0
SHEET 2 M 3 TYR H 201 HIS H 207 -1 0 ASN H 206 N THR H 158
SHEET 3 M 3 LYS H 217 VAL H 218 -1 0 VAL H 218 N TYR H 201
SHEET 1 N 3 TRP C 28 LEU C 29 0
SHEET 2 N 3 VAL C 13 ALA C 18 -1 N VAL C 17 0 LEU C 29
SHEET 3 N 3 LEU C 36 ALA C 38 -1 0 LEU C 36 N HIS C 15
SHEET 1 O 5 TRP C 28 LEU C 29 0
SHEET 2 O 5 VAL C 13 ALA C 18 -1 N VAL C 17 0 LEU C 29
SHEET 3 O 5 TYR C 151 ALA C 156 -1 0 ILE C 154 N ALA C 14
SHEET 4 O 5 GLY C 54 GLN C 67 -1 N TYR C 59 0 GLY C 153
SHEET 5 O 5 PRO C 113 LEU C 126 -1 0 GLY C 122 N ILE C 58
SHEET 1 P 5 GLU C 42 ARG C 44 0
SHEET 2 P 5 GLN C 47 VAL C 49 -1 0 VAL C 49 N GLU C 42
SHEET 3 P 5 ARG C 131 ILE C 136 -1 0 LEU C 132 N LEU C 48
SHEET 4 P 5 LEU C 76 ILE C 83 -1 N THR C 79 0 GLU C 135
SHEET 5 P 5 LYS C 98 LYS C 98 -1 0 LEU C 94 N ILE C 80

SSBOND 1 CYS L 23 CYS L 88 1555 1555 2.05
SSBOND 2 CYS L 134 CYS L 194 1555 1555 2.03
SSBOND 3 CYS H 22 CYS H 98 1555 1555 2.05
SSBOND 4 CYS H 147 CYS H 203 1555 1555 2.03
SSBOND 5 CYS C 69 CYS C 101 1555 1555 2.06
CISPSEP 1 SER L 7 PRO L 8 0 2.35
CISPSEP 2 TRP L 94 PRO L 95 0 6.74
CISPSEP 3 TYR L 140 PRO L 141 0 1.27
CISPSEP 4 GLU H 1 VAL H 2 0 3.36
CISPSEP 5 GLY H 140 GLY H 141 0 9.68
CISPSEP 6 PHE H 153 PRO H 154 0 -3.35
CISPSEP 7 GLU H 155 PRO H 156 0 0.91
CISPSEP 8 CYS C 101 GLN C 102 0 -3.69
CRYST1 153.663 153.663 99.279 90.00 90.00 120.00 H 3 9
ORIGX1 1.000000 0.000000 0.000000 0.000000
ORIGX2 0.000000 1.000000 0.000000 0.000000
ORIGX3 0.000000 0.000000 1.000000 0.000000
SCALE1 0.006508 0.003757 0.000000 0.000000
SCALE2 0.000000 0.007514 0.000000 0.000000
SCALE3 0.000000 0.000000 0.010073 0.000000
ATOM 1 N ASP L 1 24.141 -2.203 -3.932 1.00 34.47 N
ATOM 2 CA ASP L 1 25.160 -2.458 -2.905 1.00 44.54 C
ATOM 3 C ASP L 1 25.722 -1.169 -2.301 1.00 43.71 C
    
```

how file was produced

the 17 numbered atoms

HEADER	REMARK	Spartan	ST	exported	M001	x	y	z
HETATM	1	Cu	UNK	0001		0.000	0.000	0.000
HETATM	2	N	UNK	0001		-0.870	0.000	1.728
HETATM	3	H	UNK	0001		-1.884	0.000	1.654
HETATM	4	H	UNK	0001		-0.607	0.825	2.263
HETATM	5	H	UNK	0001		-0.607	-0.825	2.263
HETATM	6	N	UNK	0001		-1.728	0.000	-0.870
HETATM	7	H	UNK	0001		-1.654	0.000	-1.884
HETATM	8	H	UNK	0001		-2.263	0.825	-0.607
HETATM	9	H	UNK	0001		-2.263	-0.825	-0.607
HETATM	10	N	UNK	0001		0.870	0.000	-1.728
HETATM	11	H	UNK	0001		1.884	0.000	-1.654
HETATM	12	H	UNK	0001		0.607	0.825	-2.263
HETATM	13	H	UNK	0001		0.607	-0.825	-2.263
HETATM	14	N	UNK	0001		1.728	0.000	0.870
HETATM	15	H	UNK	0001		1.654	0.000	1.884
HETATM	16	H	UNK	0001		2.263	0.825	0.607
HETATM	17	H	UNK	0001		2.263	-0.825	0.607
CONECT	1	2	6	10	14			
CONECT	2	3	4	5	1			
CONECT	3	2						
CONECT	4	2						
CONECT	5	2						
CONECT	6	7	8	9	1			
CONECT	7	6						
CONECT	8	6						
CONECT	9	6						
CONECT	10	11	12	13	1			
CONECT	11	10						
CONECT	12	10						
CONECT	13	10						
CONECT	14	15	16	17	1			
CONECT	15	14						
CONECT	16	14						
CONECT	17	14						
END								

How the 17 atoms are connected

PDB Format

The University
of Georgia



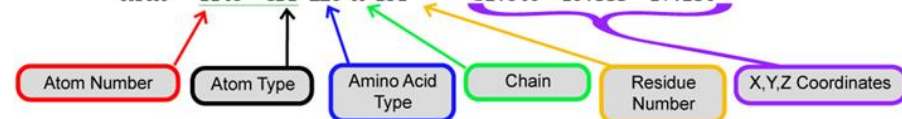
Atomic Coordinates: PDB Format

	Amino Acid		Chain name		Sequence Number		Coordinates		
	Element						X	Y	Z (etc.)
ATOM	1	N	ASP	L	1		4.060	7.307	5.186 ...
ATOM	2	CA	ASP	L	1		4.042	7.776	6.553 ...
ATOM	3	C	ASP	L	1		2.668	8.426	6.644 ...
ATOM	4	O	ASP	L	1		1.987	8.438	5.606 ...
ATOM	5	CB	ASP	L	1		5.090	8.827	6.797 ...
ATOM	6	CG	ASP	L	1		6.338	8.761	5.929 ...
ATOM	7	OD1	ASP	L	1		6.576	9.758	5.241 ...
ATOM	8	OD2	ASP	L	1		7.065	7.759	5.948 ...

Element position within amino

```

ATOM 1132 NH1 ARG A 149 31.814 -31.597 16.995
ATOM 1133 NH2 ARG A 149 32.203 -32.934 18.816
ATOM 1134 N ASN A 150 29.346 -24.359 18.812
ATOM 1135 CA ASN A 150 28.480 -23.190 18.933
ATOM 1136 C ASN A 150 28.606 -22.168 17.808
ATOM 1137 O ASN A 150 27.803 -21.276 17.678
ATOM 1138 CB ASN A 150 28.732 -22.524 20.282
ATOM 1139 CG ASN A 150 28.284 -23.389 21.447
ATOM 1140 OD1 ASN A 150 27.205 -23.981 21.430
ATOM 1141 ND2 ASN A 150 29.110 -23.463 22.466
ATOM 1142 N LEU A 151 29.629 -22.313 16.996
ATOM 1143 CA LEU A 151 29.868 -21.415 15.894
ATOM 1144 C LEU A 151 29.953 -22.205 14.597
ATOM 1145 O LEU A 151 30.149 -23.422 14.614
ATOM 1146 CB LEU A 151 31.208 -20.735 16.100
ATOM 1147 CG LEU A 151 31.436 -19.884 17.337
ATOM 1148 CD1 LEU A 151 32.846 -19.333 17.256
    
```



PDB Format

The University
of Georgia



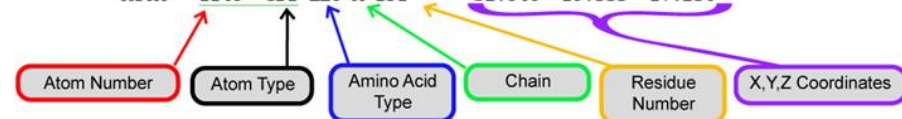
Atomic Coordinates: PDB Format

					-----Coordinates-----			
					X	Y	Z	(etc.)
ATOM	1	N	ASP L	1	4.060	7.307	5.186	...
ATOM	2	CA	ASP L	1	4.042	7.776	6.553	...
ATOM	3	C	ASP L	1	2.668	8.426	6.644	...
ATOM	4	O	ASP L	1	1.987	8.438	5.606	...
ATOM	5	CB	ASP L	1	5.090	8.827	6.797	...
ATOM	6	CG	ASP L	1	6.338	8.761	5.929	...
ATOM	7	OD1	ASP L	1	6.576	9.758	5.241	...
ATOM	8	OD2	ASP L	1	7.065	7.759	5.948	...

Element position within amino

```

ATOM 1132 NH1 ARG A 149 31.814 -31.597 16.995
ATOM 1133 NH2 ARG A 149 32.203 -32.934 18.816
ATOM 1134 N ASN A 150 29.346 -24.359 18.812
ATOM 1135 CA ASN A 150 28.480 -23.190 18.933
ATOM 1136 C ASN A 150 28.606 -22.168 17.808
ATOM 1137 O ASN A 150 27.803 -21.276 17.678
ATOM 1138 CB ASN A 150 28.732 -22.524 20.282
ATOM 1139 CG ASN A 150 28.284 -23.389 21.447
ATOM 1140 OD1 ASN A 150 27.205 -23.981 21.430
ATOM 1141 ND2 ASN A 150 29.110 -23.463 22.466
ATOM 1142 N LEU A 151 29.629 -22.313 16.996
ATOM 1143 CA LEU A 151 29.868 -21.415 15.894
ATOM 1144 C LEU A 151 29.953 -22.205 14.597
ATOM 1145 O LEU A 151 30.149 -23.422 14.614
ATOM 1146 CB LEU A 151 31.208 -20.735 16.100
ATOM 1147 CG LEU A 151 31.436 -19.884 17.337
ATOM 1148 CD1 LEU A 151 32.846 -19.333 17.256
    
```



PDBsum

The University
of Georgia



EMBL-EBI

Enter Text Here **Find** Help | Feedback

Databases Tools Research Training Industry About Us Help Site Index

Highlights
List of PDB codes
Het Groups
Ligands
Drugs
Enzymes
ProSite
Pfam
Species
Generate
Gallery
Figure stats
Documentation
Downloads
Contact us

EBI > Databases > Structure Databases > PDBsum

PDBsum

PDB code (4 chars) **Find** Example: "1kfy"

PDBsum provides an at-a-glance overview of every macromolecular structure deposited in the Protein Data Bank (PDB), giving schematic diagrams of the molecules in each structure and of the interactions between them.

Text and sequence searches

Submit your PDB file for PDBsum analysis **Search**

Scans all TITLE, HEADER, COMPND, SOURCE and AUTHOR records in the PDB (eg to find a given protein by name).

Search by sequence

Search

Perform FASTA search vs all sequences in the PDB to get a list of the closest matches.

Notes

You can use the **Generate** option on the left to submit your own structure and get a password protected PDBsum analysis generated for it.

The development of PDBsum has been partly funded by the [Wellcome Trust](#)

Contents

PDBsum contains 79,259 entries, including 1,620 superseded
Last update: 8 October, 2011

Related databases

BC-PDB
Enzyme 3D structures organized by the E.C. numbering hierarchy.

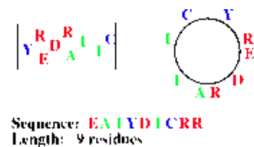
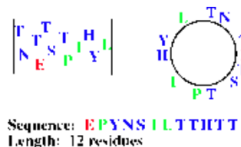
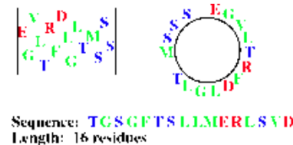
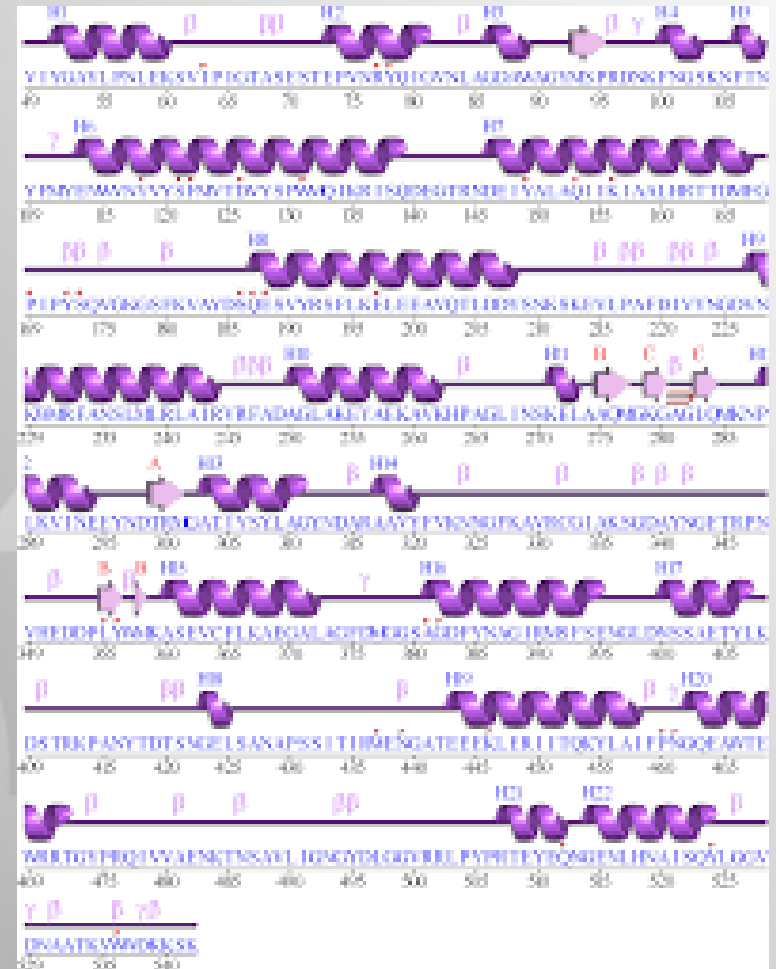
DrugPort
Structures of drugs and their target proteins in the PDB.

SAS
Searches sequence against all PDB sequences and structurally annotates alignment.

ProFunc
Prediction of protein function from 3D structure.

Arch Schema
NEW Graphs of protein sequences having related Pfam domain architectures.

Terms of Use | EBI Funding



PDB



- Protein databank
- 126,978 entries

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

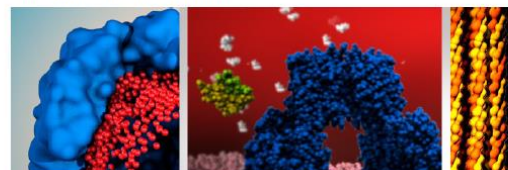
A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

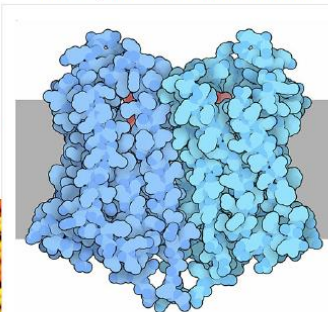
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

New Video: What is a Protein?



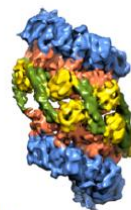
January Molecule of the Month



Opioid Receptors

Latest Entries

As of Tuesday Jan 09



EMD-3894

EMDB Entry

Regulatory coiled-coil domains promote head-to-head assemblies of AAA+ chaperones essential for tunable activi...

Features & Highlights



wwPDB News
A new FTP re



Improved Te
Simple text se
now powered

News

Publications

wwPDB News: Time-stamped Copies of the PDB Archive Available
A snapshot of the PDB archive (<ftp://ftp.wwpdb.org>) as of January 1, 2018 has been added to <ftp://snapshots.wwpdb.org/> and <ftp://snapshots.pdbj.org/>. » 01/09

A Year of Antimicrobial Resistance » 01/02

Season's Greetings » 12/26

Access Irving Geis' Early Molecular Images in 3D » 12/19

Browse the Biology of Plants and More at PDB-101 » 12/12

Overview of PDB Validation Reports published in Structure » 12/05

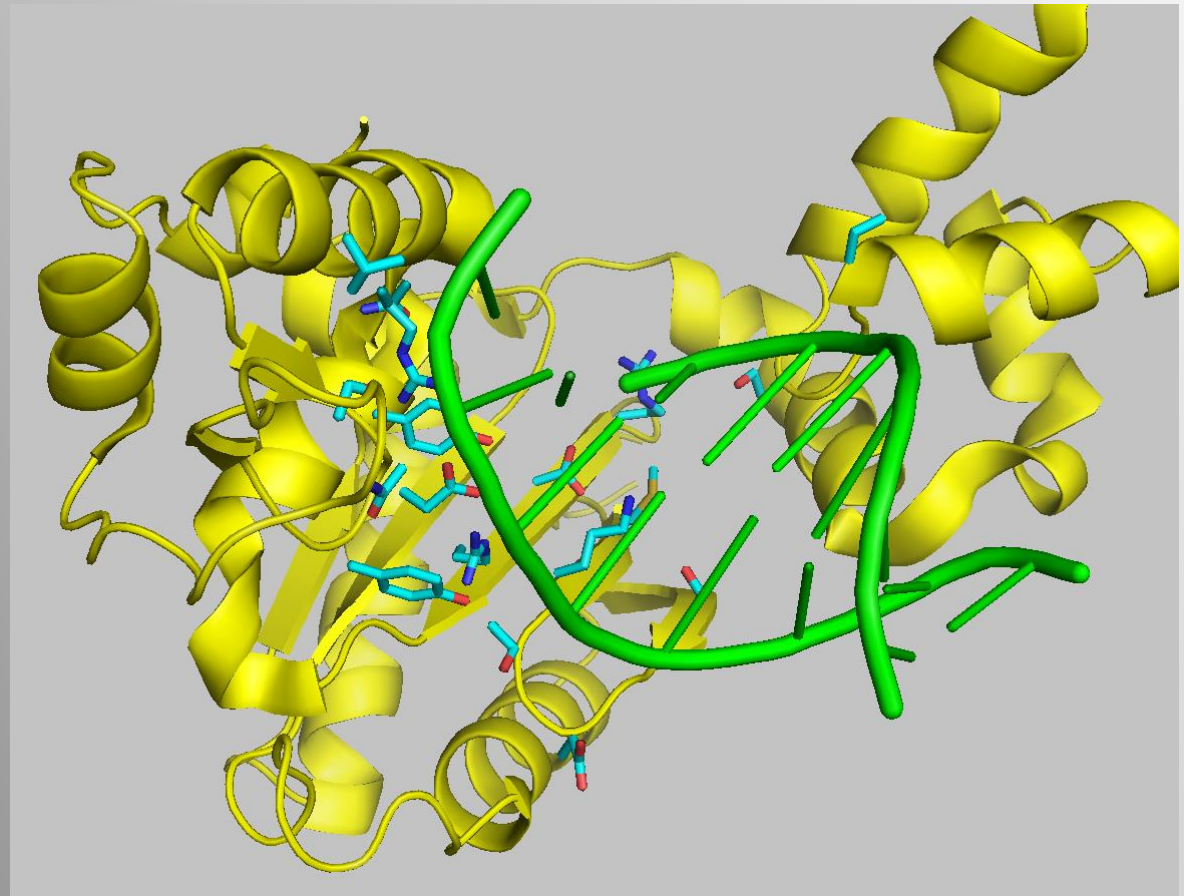
Structural Databases as Teaching Tools » 11/28

New Video Answers "What is a Protein?" » 11/21

Software



- VMD
- Pymol



Python Time

