Basic Principles of Protein Structures

- ECS129
- *Patrice Koehl*

Proteins

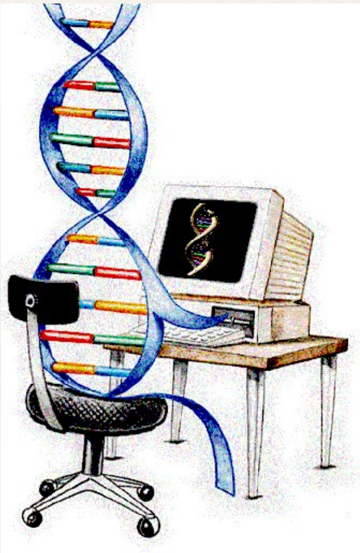
Proteins: The Molecule of Life

Proteins: Building Blocks

Proteins: Secondary Structures

Proteins: Tertiary and Quartenary Structure

Proteins: Geometry



Proteins

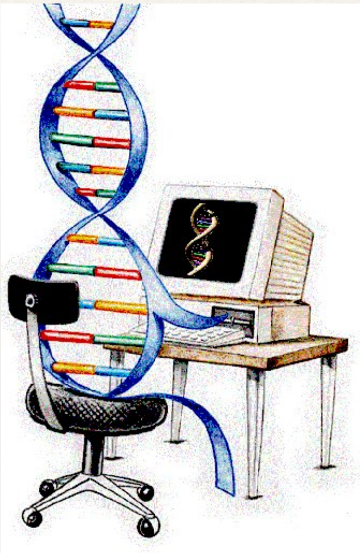
Proteins: The Molecule of Life

Proteins: Building Blocks

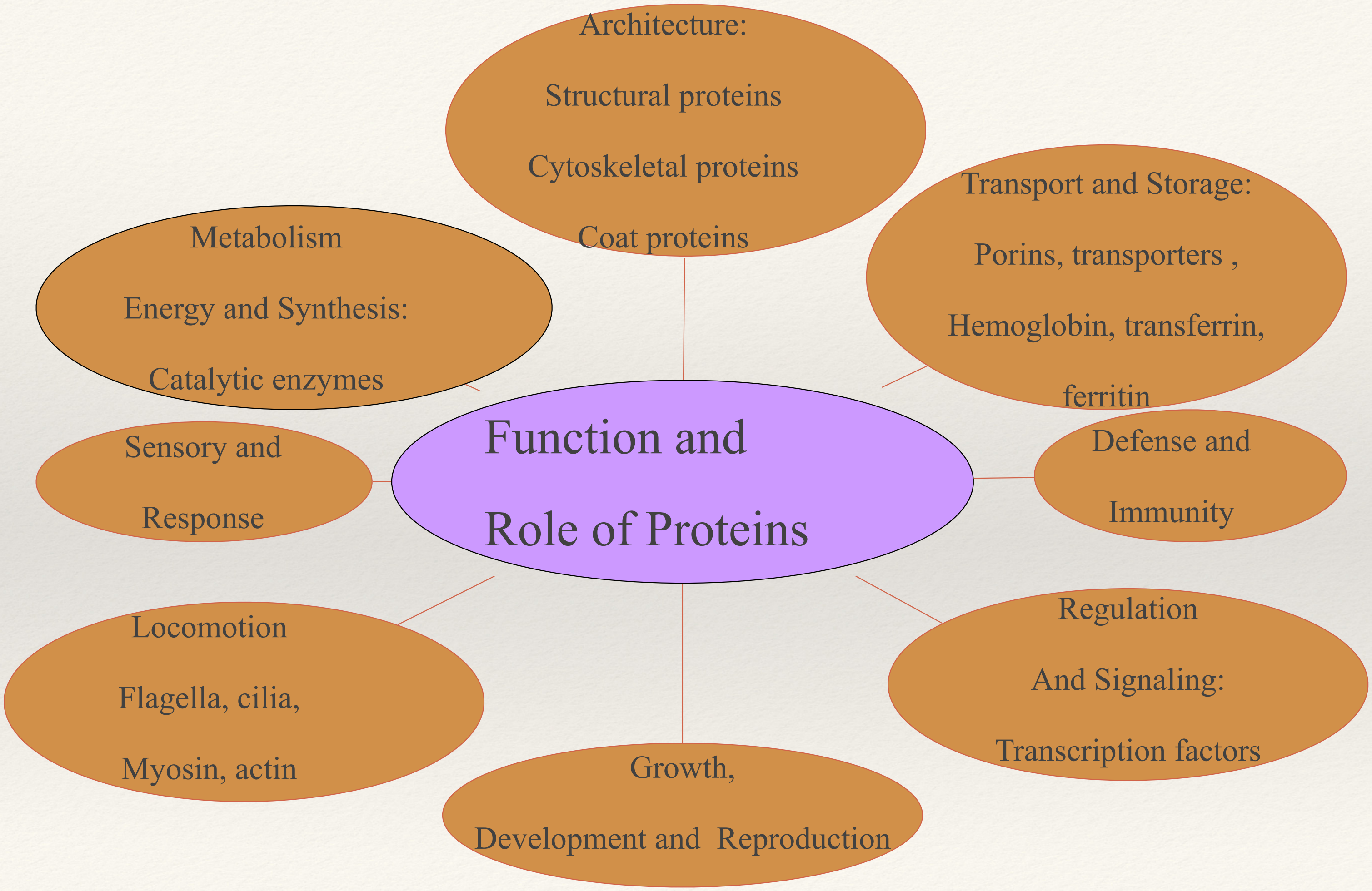
Proteins: Secondary Structures

Proteins: Tertiary and Quartenary Structure

Proteins: Geometry



Why Proteins?



Function and Role of Proteins

Architecture:
Structural proteins
Cytoskeletal proteins
Coat proteins

Transport and Storage:
Porins, transporters ,
Hemoglobin, transferrin,
ferritin

Defense and Immunity

Regulation
And Signaling:
Transcription factors

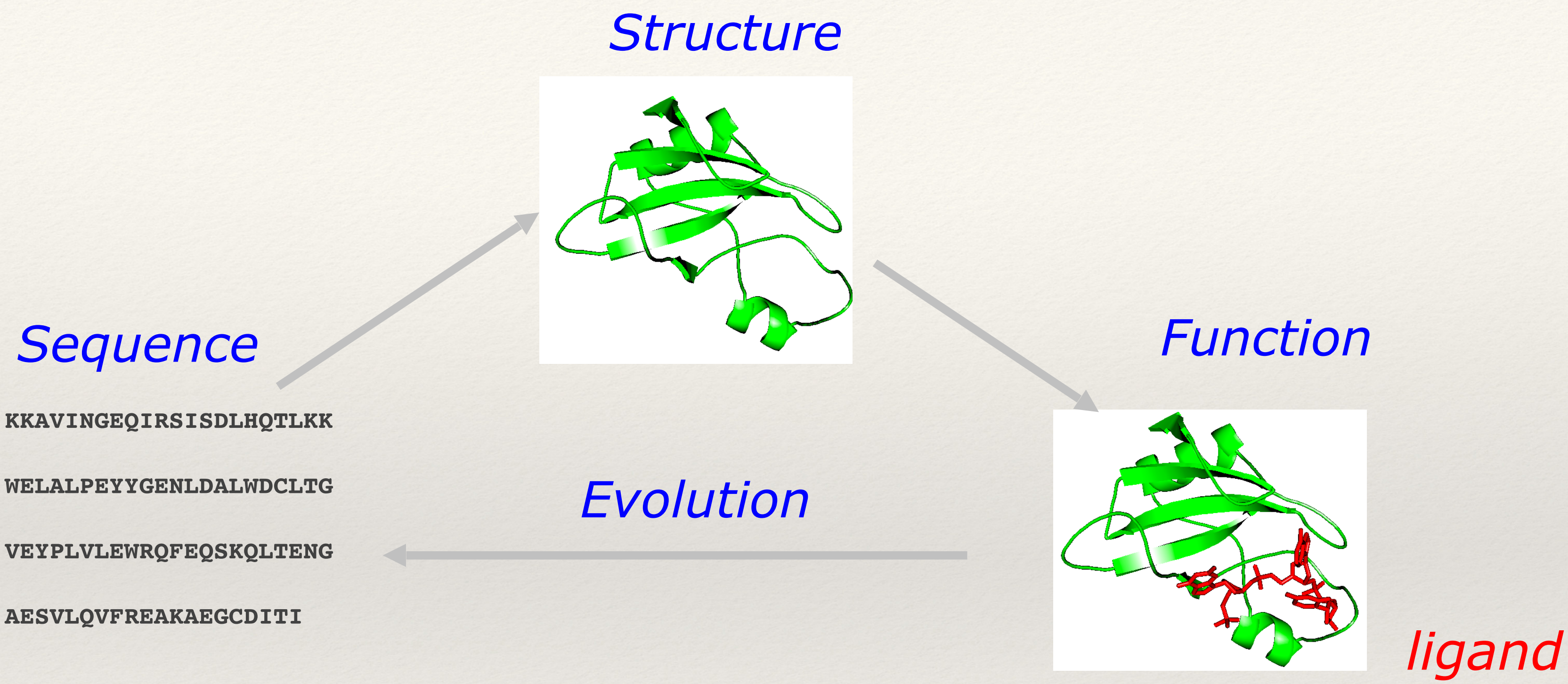
Growth,
Development and Reproduction

Locomotion
Flagella, cilia,
Myosin, actin

Sensory and Response

Metabolism
Energy and Synthesis:
Catalytic enzymes

The Protein Cycle



Structure

Sequence

KKAVINGEQIRSISDLHQTLKK

WELALPEYYGENLDALWDCLTG

VEYPLVLEWRQFEQSKQLTENG

AESVLQVFREAKAEGCDITI

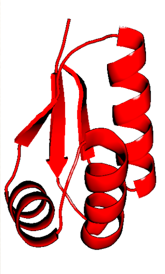
Function

Evolution

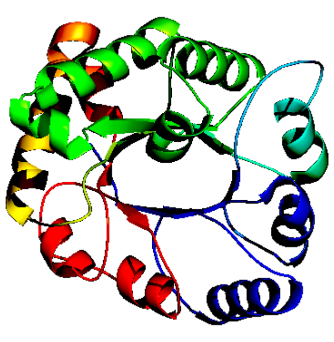
ligand

Protein Structure Diversity

1CTF

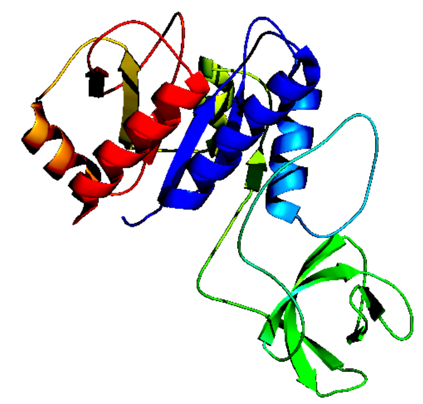


1TIM



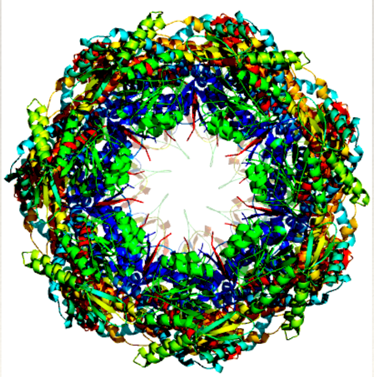
1A1O

1K3R



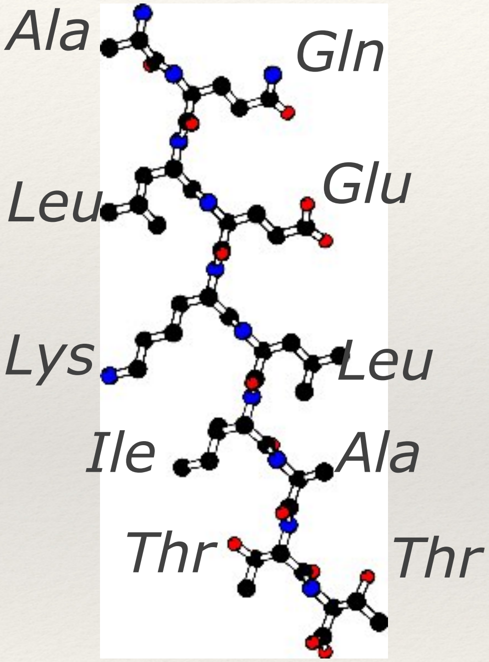
1NIK

1AON



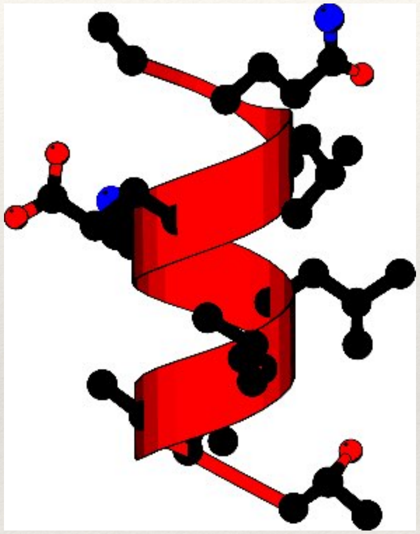
Protein Structure

Primary structure



Sequence of Amino acids

Secondary Structure



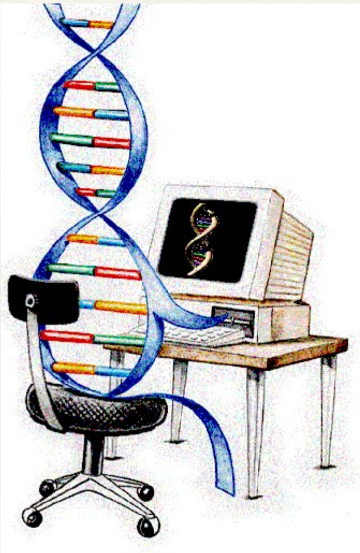
Local interactions

Tertiary Structure

Native protein

Proteins

- Proteins: The Molecule of Life
- Proteins: Building Blocks
- Proteins: Secondary Structures
- Proteins: Tertiary and Quartenary Structure
- Proteins: Geometry



Review of Acid-Base Chemistry

What is an acid or a base?

An acid is a material that can release a proton (or hydrogen ion, H^+), and a base is a material that can donate a hydroxide ion (OH^-) (Arhennius definition), or accept a proton (Lowry Bronsted definition).

Note: It is important to notice that just because a compound has a hydrogen or an OH group does not mean that it can be an acid or a base!!

- The hydrogen of methane (CH4) and usually of methyl groups (-CH3) are all strongly attached to the carbon atom
- Glycerol has three OH groups (CH2OH – CHOH – CH2OH) and all 3 are alcoholic groups.

Review of Acid-Base Chemistry

Acid plus base makes water plus a salt:

$\mathrm{AH + BOH \longrightarrow AB + H2O}$

$\mathrm{(HCL + NaOH \longrightarrow NaCl + H2O)}$

The chemical dissociation of nitric acid is:

$\mathrm{HNO3 \longrightarrow (NO3)\text{-} + H+}$

Which can be rewritten as:

$\mathrm{HNO3 + H2O \longrightarrow (NO3)\text{-} + H3O+}$

acid base conjugate conjugate

Review of Acid-Base Chemistry

pH is a measure of how acidic or alkaline (basic) a solution is. The pH of a solution is the negative log of the hydrogen ion concentration.

$$pH = -\log\left(\left[H^+\right]\right)$$
$$pOH = -\log\left(\left[OH^-\right]\right)$$
$$pH + pOH = 14$$

	$[H^+]$	pH	pOH	$[OH^-]$
Strong base	10^{-14}	14	0	1
Base	10^{-12}	12	2	10^{-2}
Weak base	10^{-9}	9	5	10^{-5}
Neutral	10^{-7}	7	7	10^{-7}
Weak acid	10^{-4}	4	10	10^{-10}
Acid	10^{-2}	2	12	10^{-12}
Strong acid	1	0	14	10^{-14}

Review of Acid-Base Chemistry

Equilibrium constant:

Dissociation of a weak acid:

$\mathrm{HA} \longrightarrow \mathrm{A}^- + \mathrm{H}^+$

$$\boxed{\begin{gathered} K_A = \frac{[H^+][A^-]}{[HA]} \\ pK_A = -\log(K_A) \end{gathered}}$$

Dissociation of a weak base:

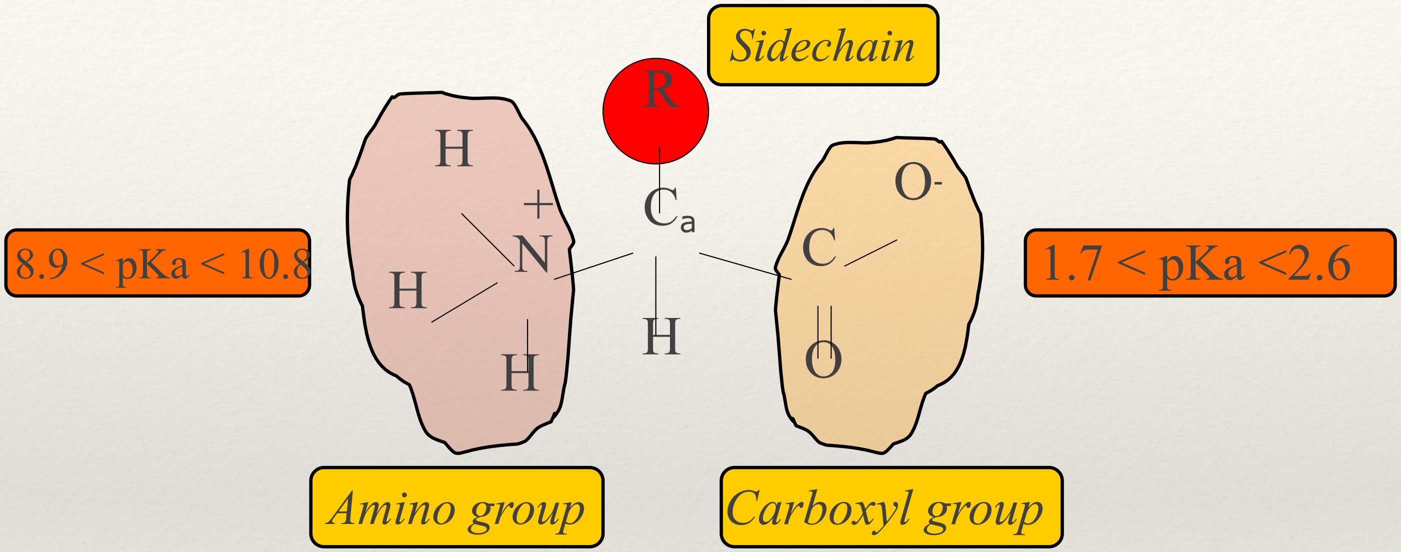
$\mathrm{BOH} \longrightarrow \mathrm{B}^+ + \mathrm{OH}^-$

$$\boxed{\begin{gathered} K_B = \frac{[B^+][OH^-]}{[BOH]} \\ pK_B = -\log(K_B) \end{gathered}}$$

For an (acid,base) pair:

$$\boxed{pK_A + pK_B = 14}$$

The Basic Block: Amino Acid



Sidechain

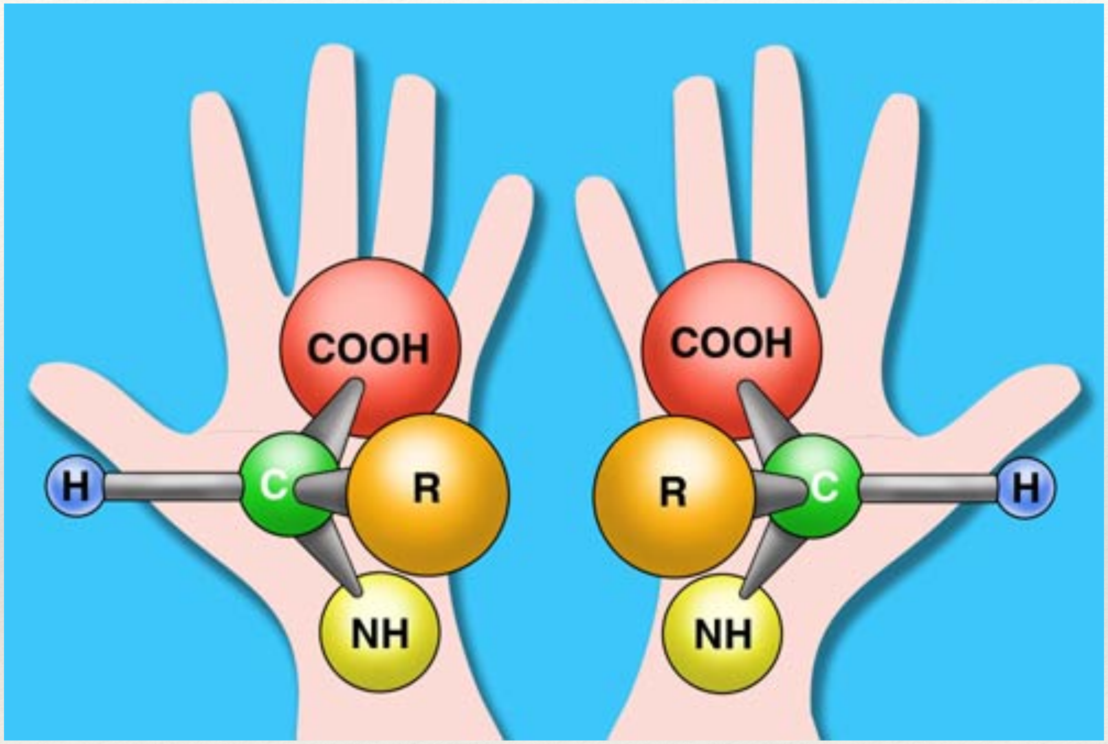
8.9 < pKa < 10.8

1.7 < pKa <2.6

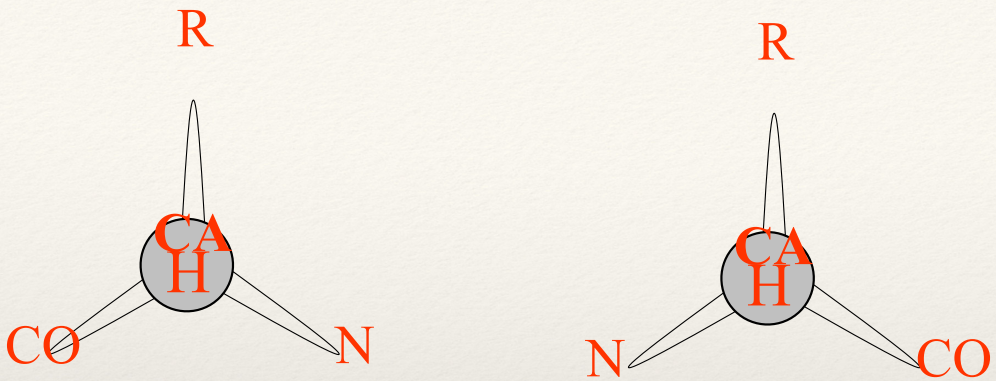
Amino group

Carboxyl group

“zwitterion”


COOH
H
C
R
NH
COOH
R
C
H
NH

Amino Acid Chirality



L-form
(CORN rule)

D-form

Amino acids in proteins are in the L-form

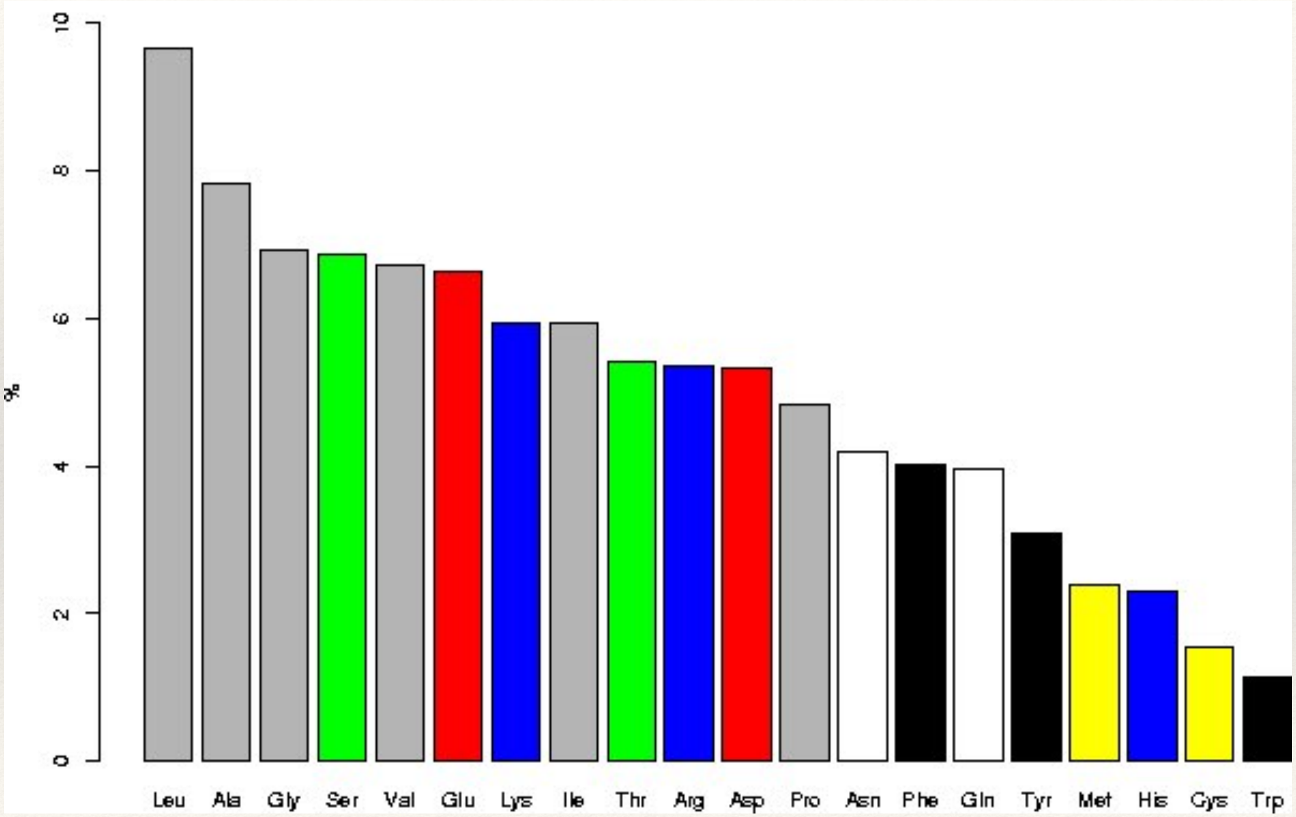
Threonine and Isoleucine have a second optical center which is also identical in all natural amino acids.

The 20 amino acids

1-letter	3-letter	Amino acid
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine

1-letter	3-letter	Amino Acid
M	Met	Methionin
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonin
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Amino Acids: Usage



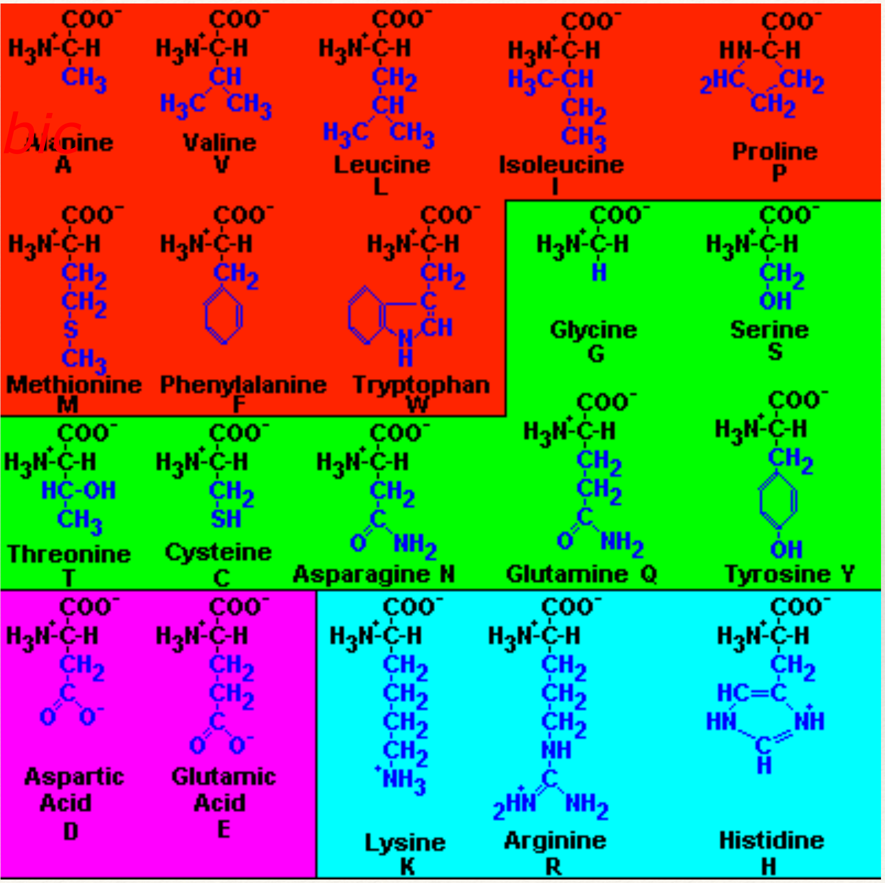
The 20 amino acids

Hydrophobic

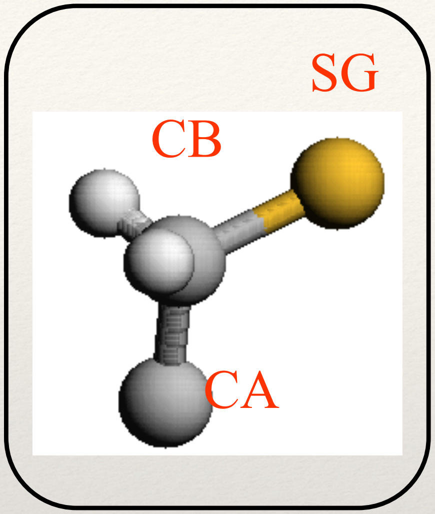
Polar, neutral

Acidic

Basic



Polar Amino acids: Cysteine



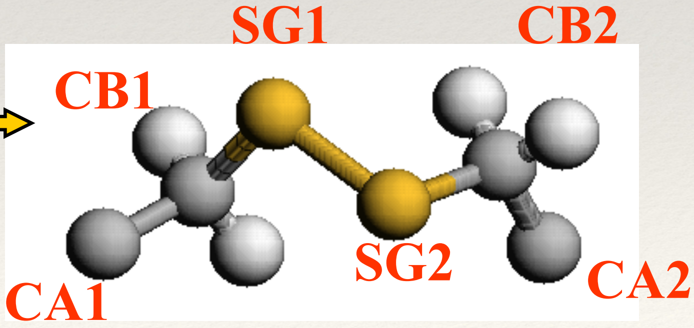
S
|
CH2
|
C

Names: Cys, C

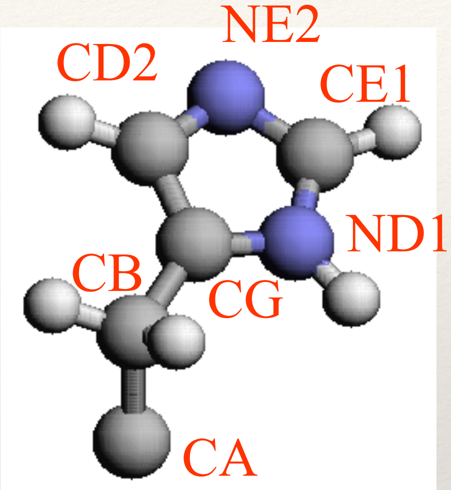
Occurrence: 1.8 %

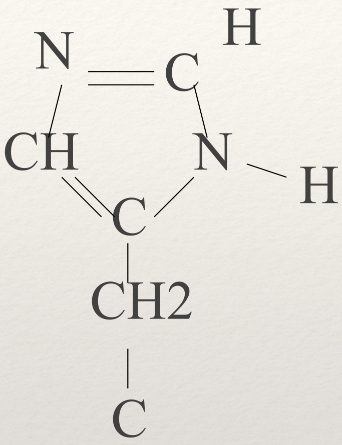
pKa sidechain: 8.3

Can form disulphide bridges in proteins



Polar Amino acids: Histidine



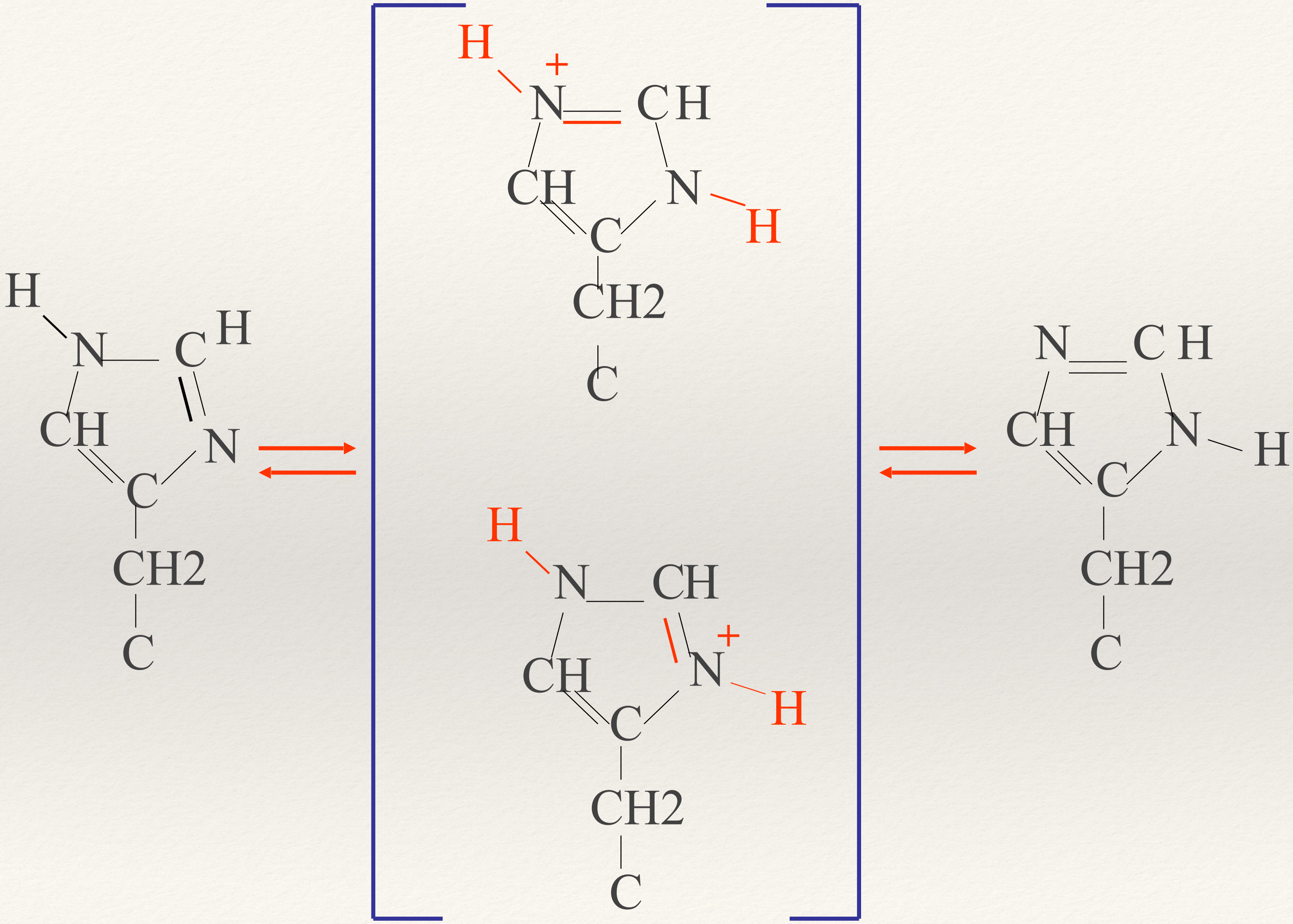


Name: His, H

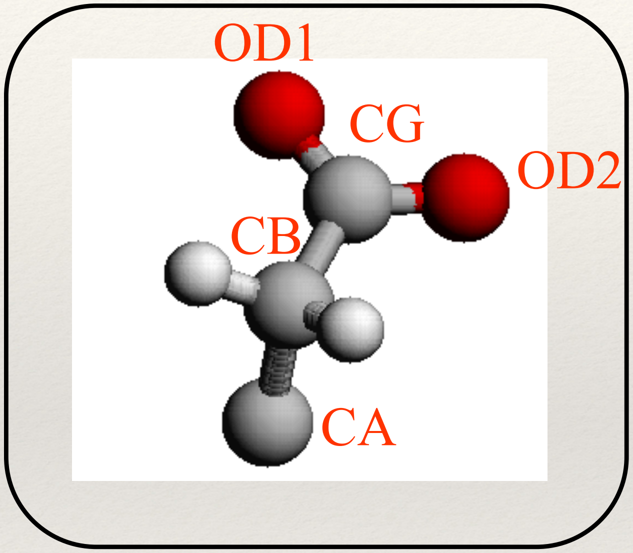
Occurrence: 2.2 %

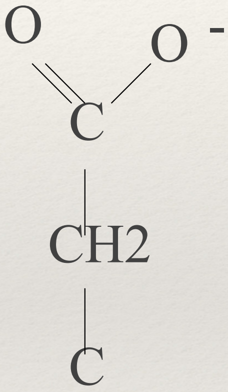
pKa sidechain: 6.04

Different ionic states of Histidine



Charged Amino acids: Aspartic Acid



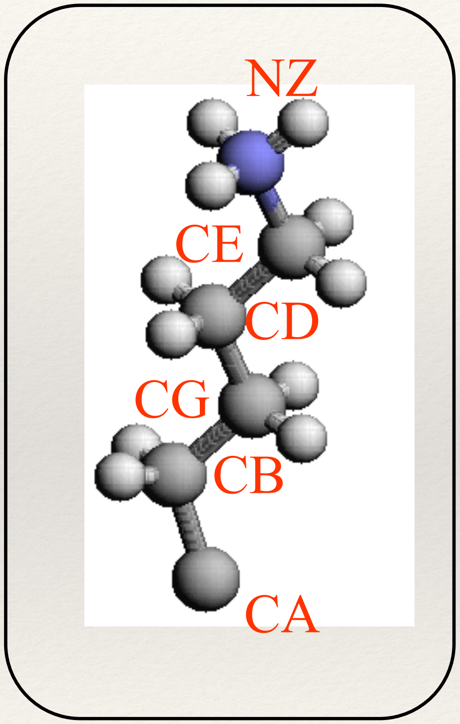


Names: Asp, D

Occurrence: 5.2 %

pKa sidechain: 3.9

Charged Amino acids: Lysine



NH3$^+$
|
CH2
|
CH2
|
CH2
|
CH2
|
C

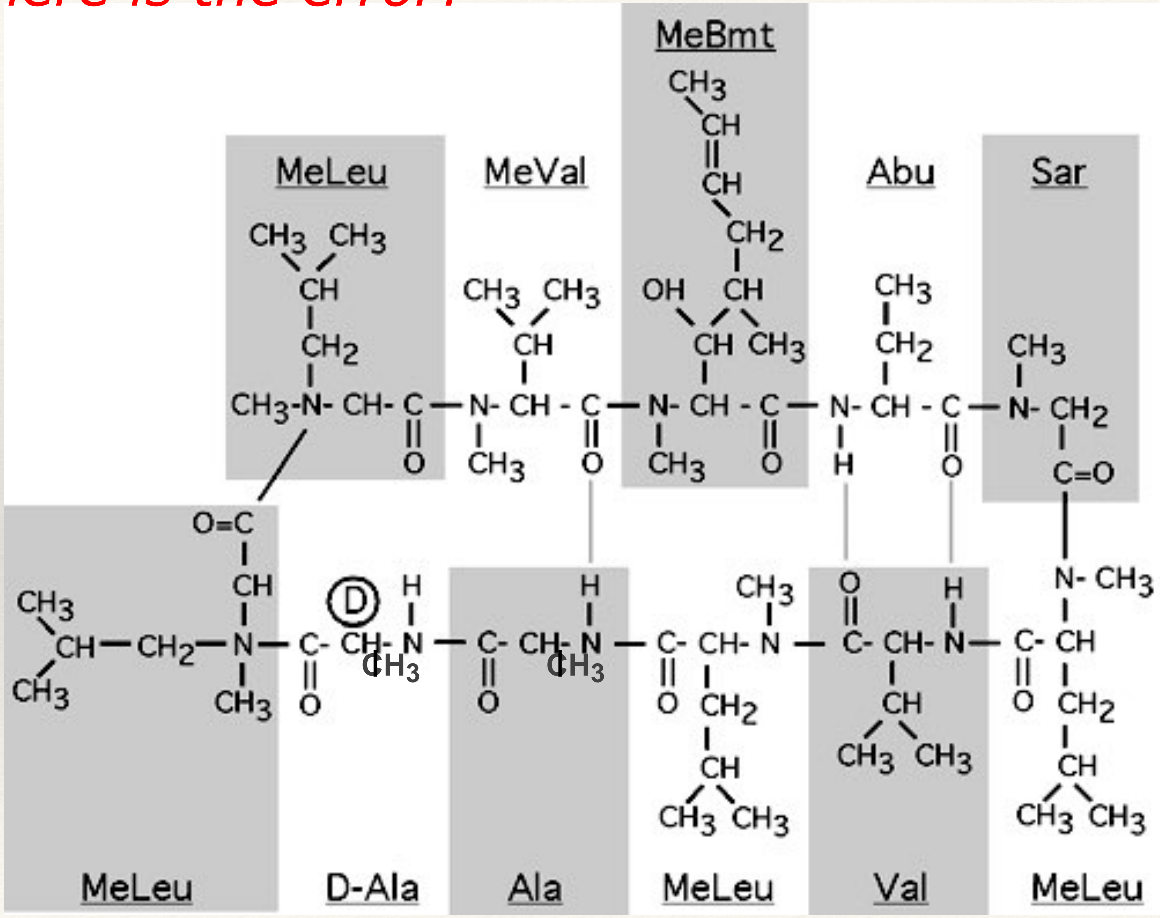
Names: Lys, K

Occurrence: 5.8 %

pKa sidechain: 9.2

Unusual Amino Acids: Cyclosporin

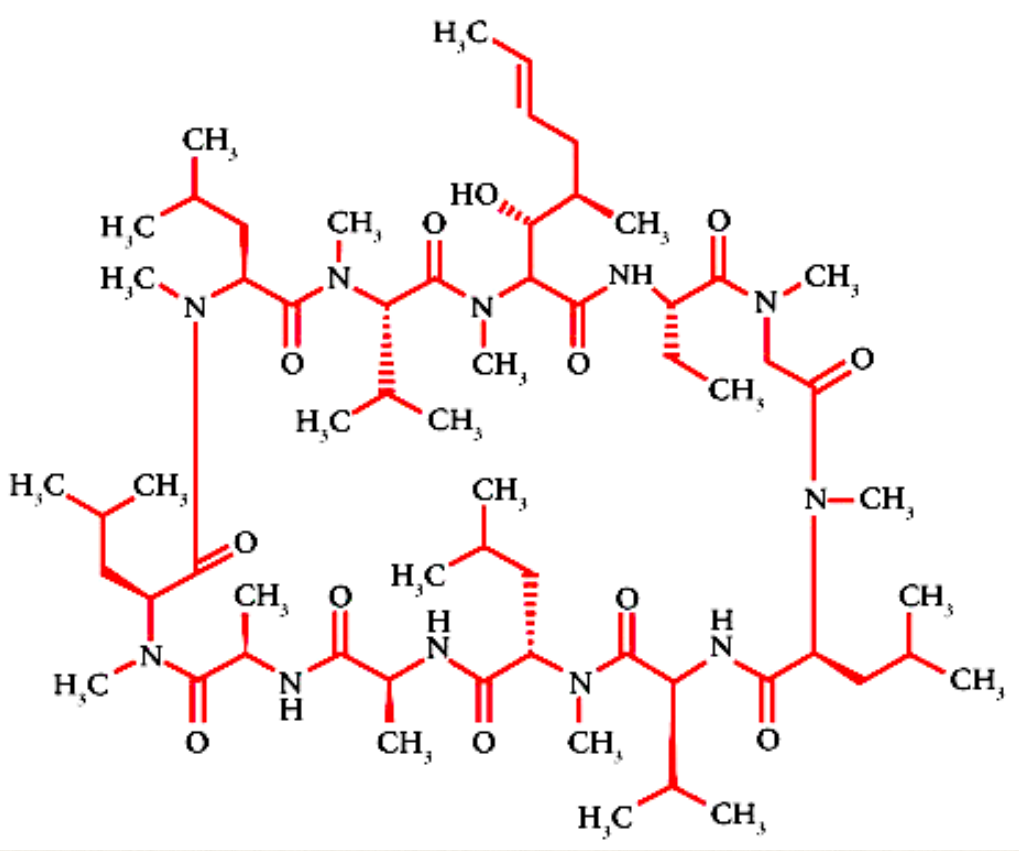
Where is the error?



http://purefixion.com/attention/2006_03_26_archive.html

Unusual Amino Acids: Cyclosporin

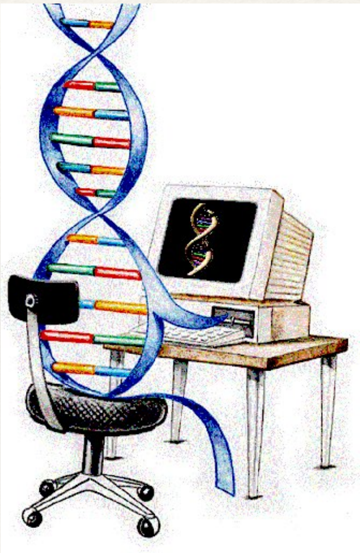
Correct!!



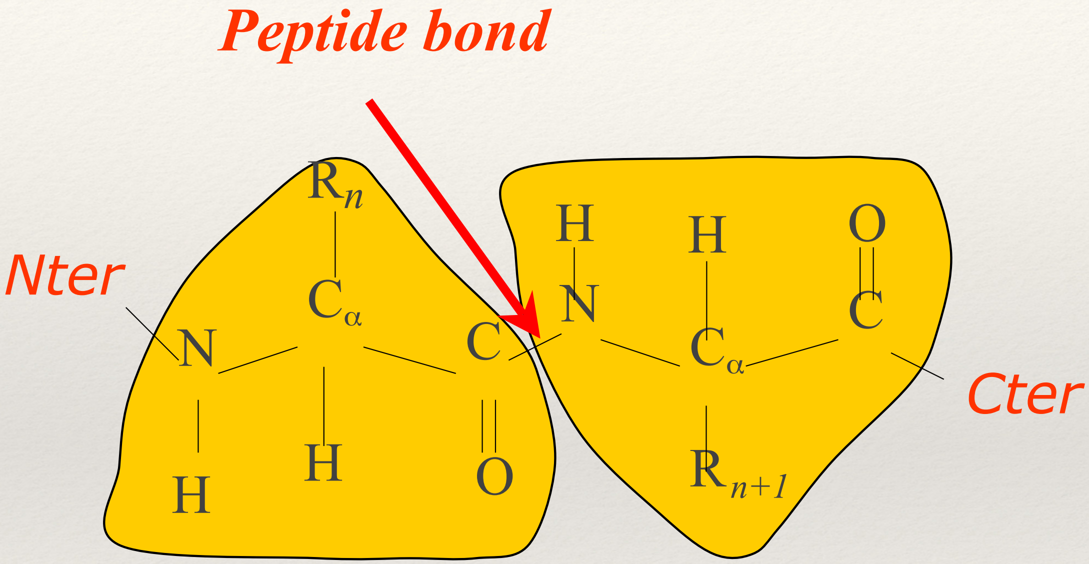
http://www.cellsignal.com/products/9973.html

Proteins

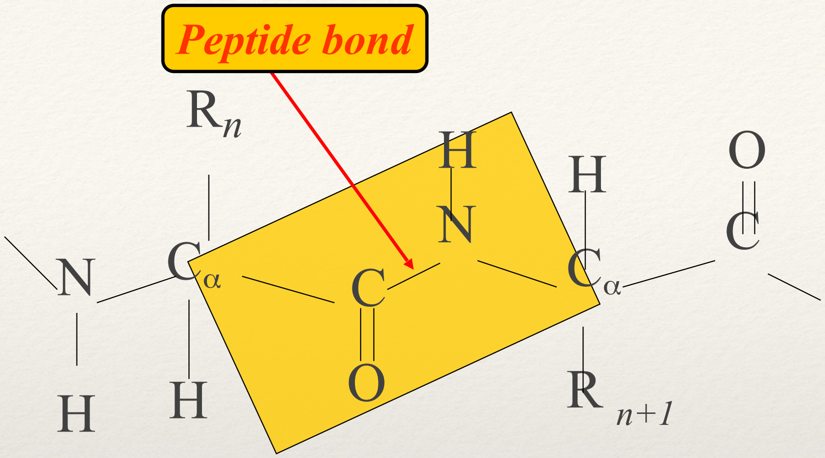
- Proteins: The Molecule of Life
- Proteins: Building Blocks
- Proteins: Secondary Structures
- Proteins: Tertiary and Quartenary Structure
- Proteins: Geometry



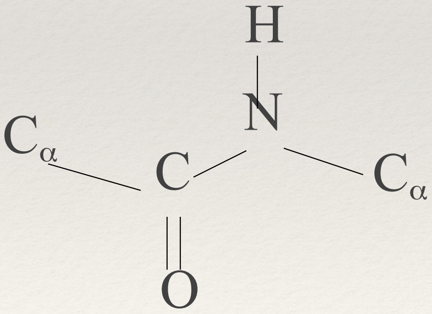
The Protein: A polymer of Amino acids



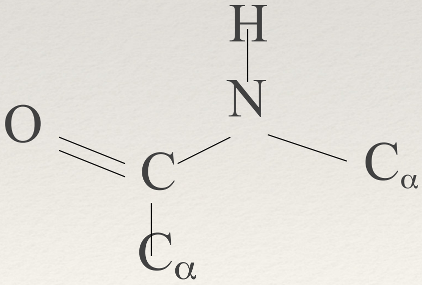
The Peptide Bond



The peptide bond is planar

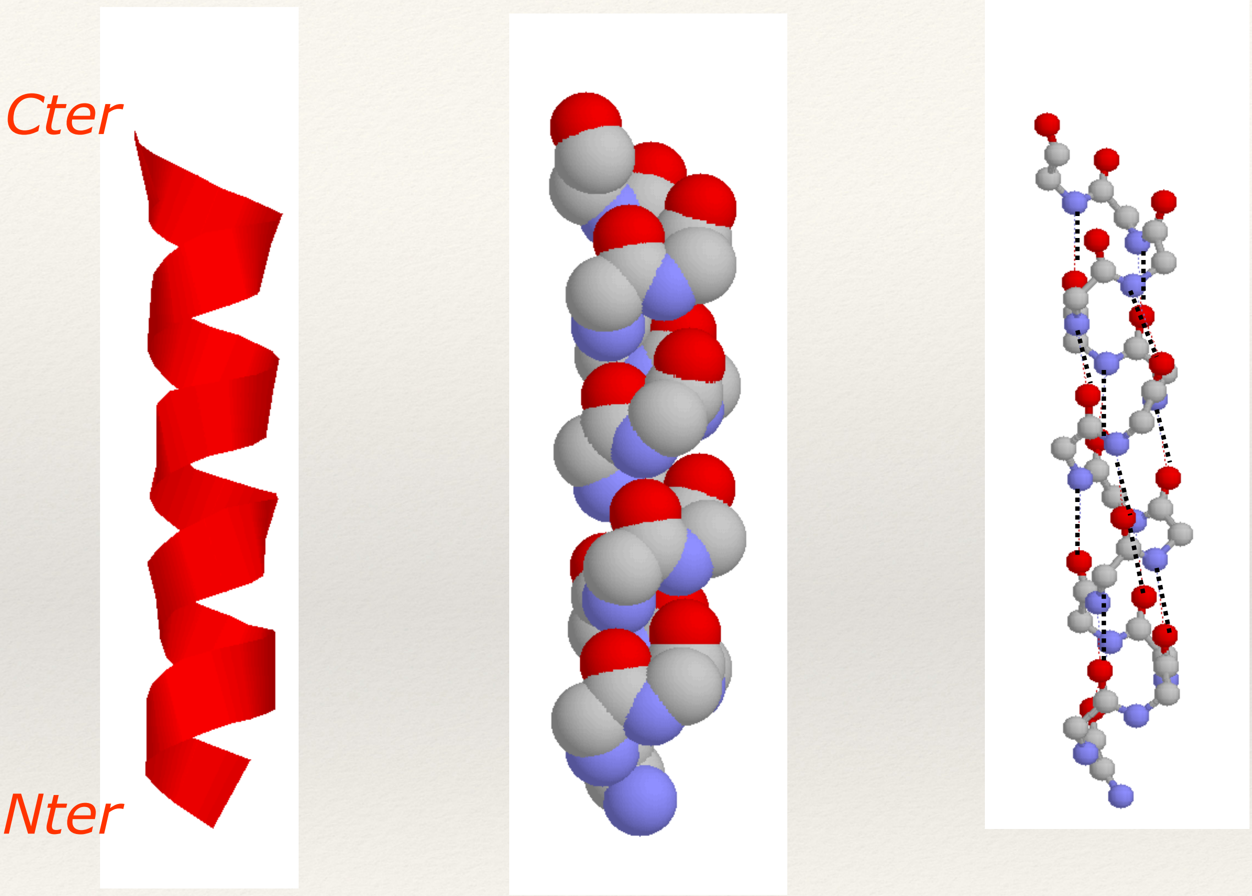


Conformation “Trans”



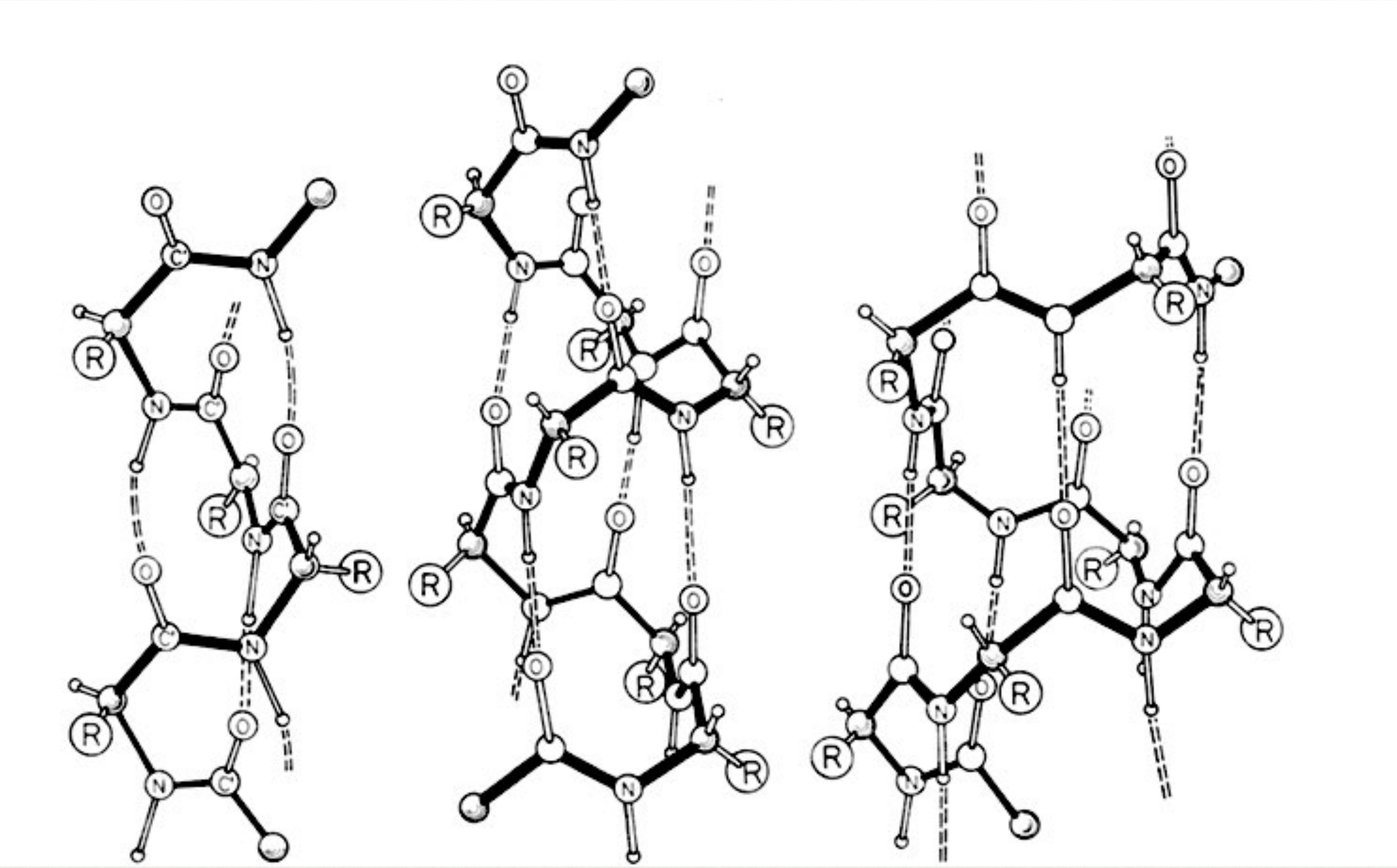
Conformation “Cis”

Helices



Hydrogen bonds: O (i) <-> N (i+4)

Helices



3_{10} helix

α-helix (4_{13})

π-helix (5_{16})

Helices

3_{10} helix

"Thin"; 3.0 residues /turn; ~ 4 % of all helices

π-helix (5_{16})

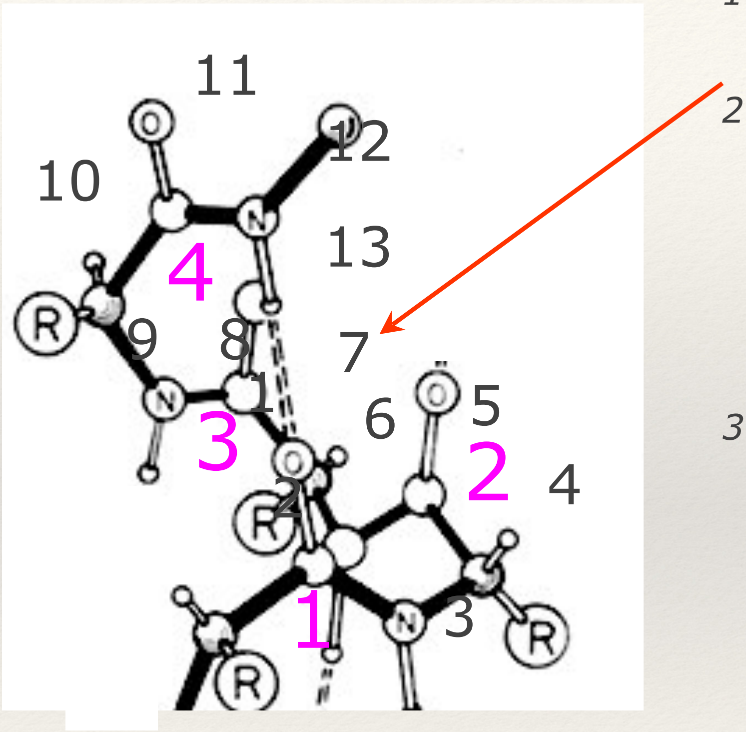
"Fat"; 4.2 residues /turn; instable

α-helix (4_{13})

"Right"; 3.6 residues /turn; 5.4 Å /turn;

most helices

Identify Helix Type



1. *Find one hydrogen bond loop*

2. *Count number of residues*

   (by number of C atoms in the loop).

   Here : *4*

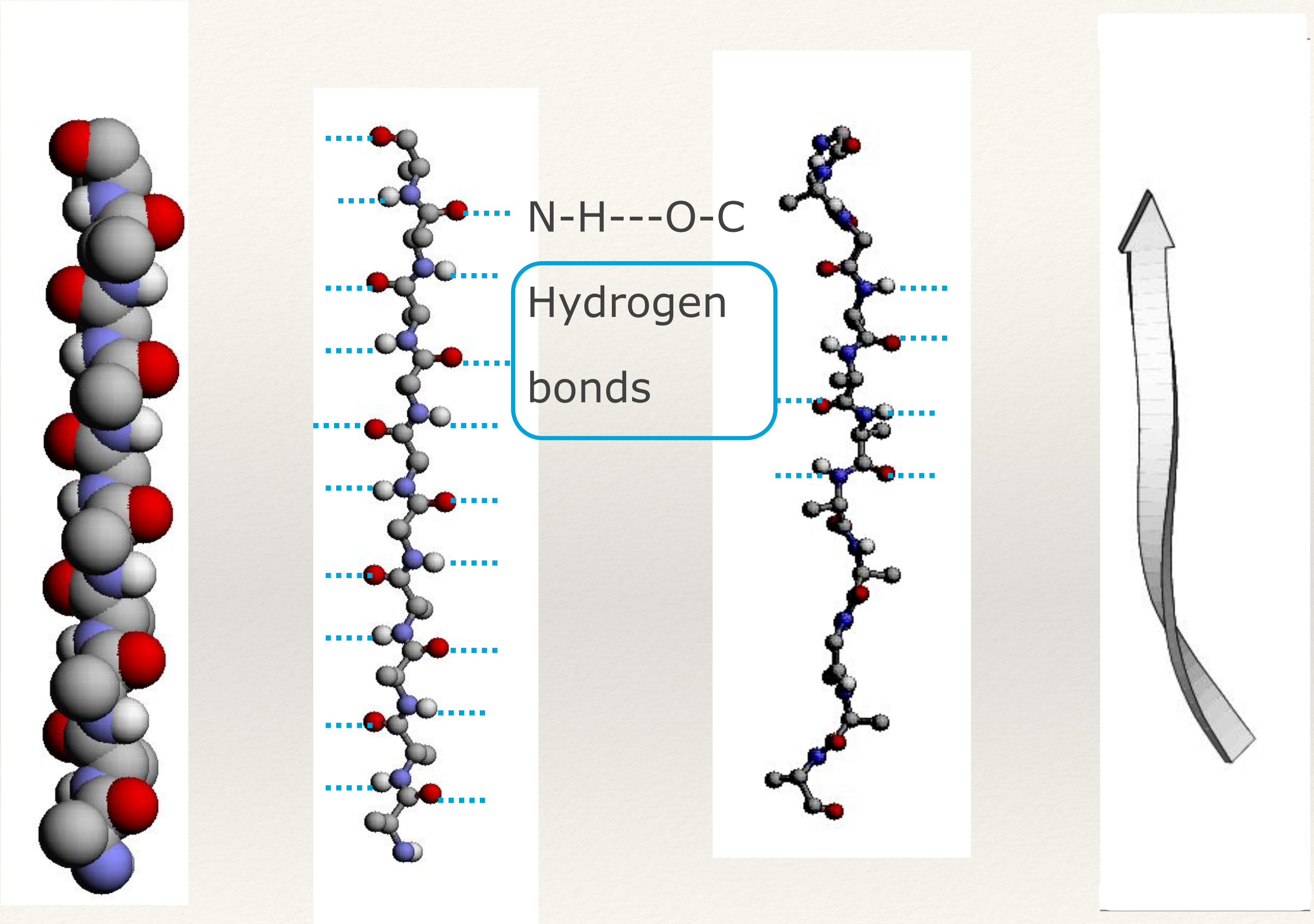
3. *Count number of atoms in the loop*

   (including first O and last H). Here:

   13

→ 4_{13} helix = α-helix

The β-strand



N-H---O-C

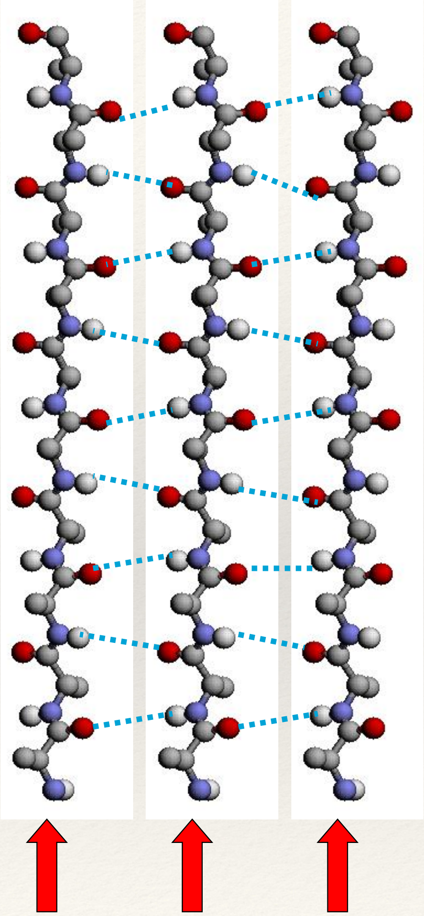
Hydrogen bonds

Extended chain is flat

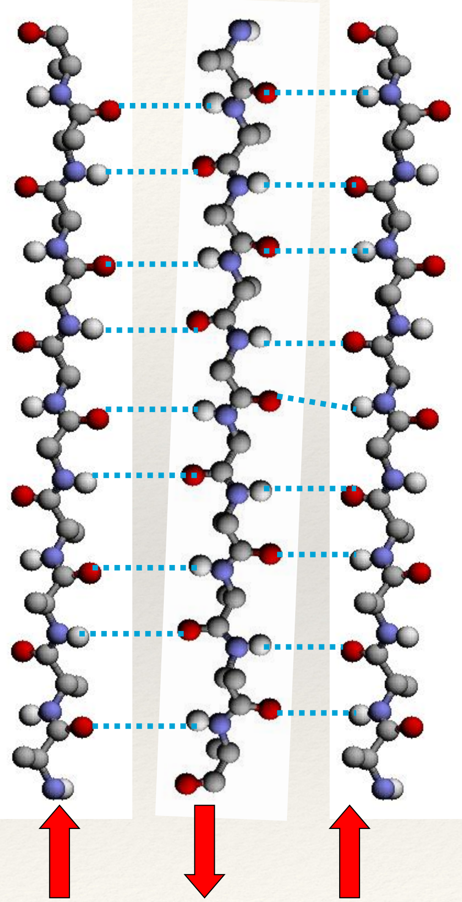
"Real β-strand is twisted"

Two types of β-sheets

Parallel



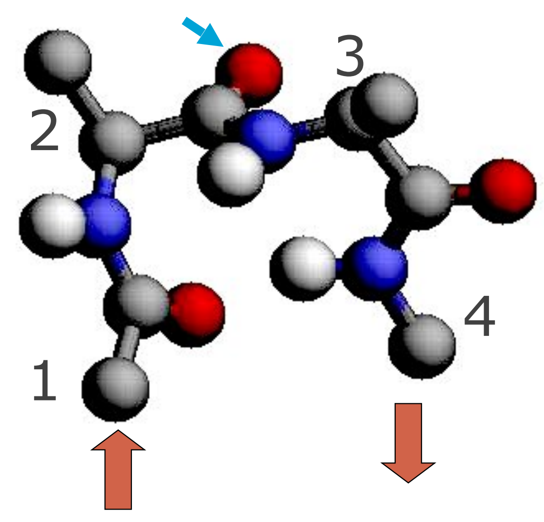
Anti-parallel



β-turns

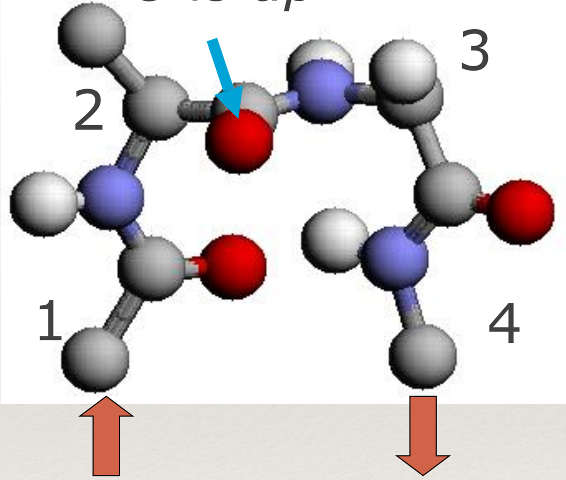
Type I

O is down



Type II

O is up

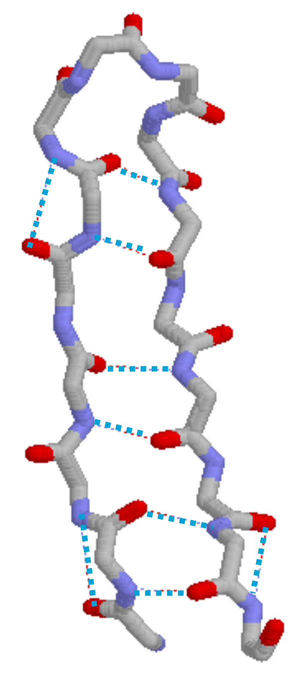


The chain changes direction by 180 degrees

Favorable /Unfavorable Residues In Turns

Turn	1	2	3	4
I	Asp, Asn, Ser, Cys	Pro	Pro	Gly
II	Asp, Asn, Ser, Cys	Pro	Gly, Asn	Gly

The β-hairpin



Proteins

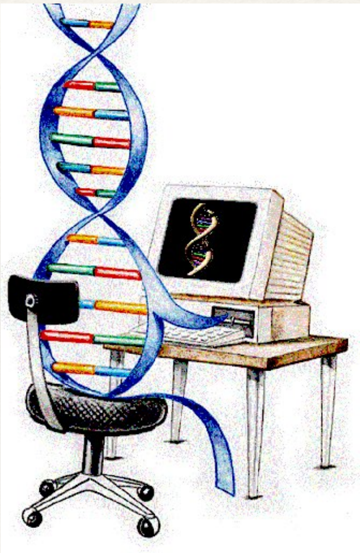
Proteins: The Molecule of Life

Proteins: Building Blocks

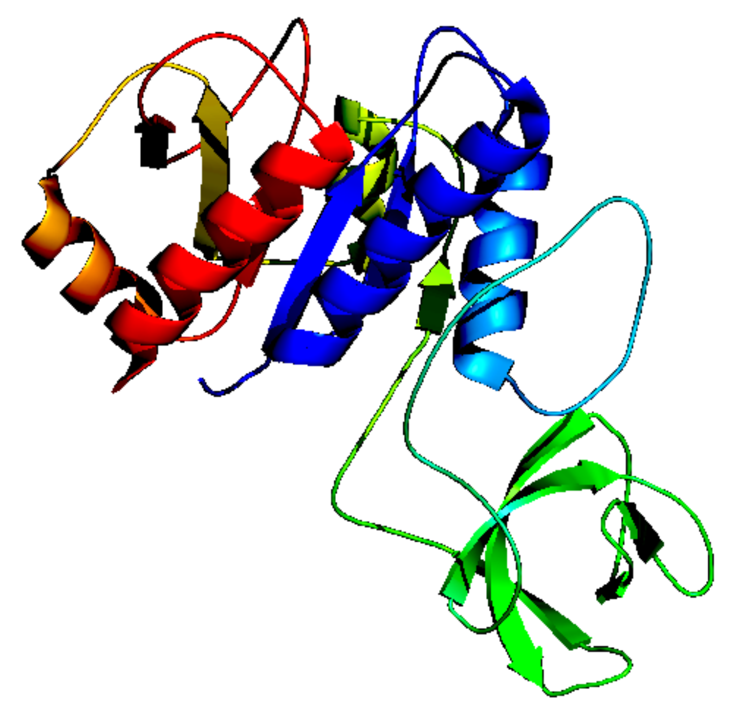
Proteins: Secondary Structures

Proteins: Tertiary and Quartenary Structure

Proteins: Geometry



Protein Tertiary Structure



- All α proteins
- All β proteins
- Alpha and beta proteins:
  - α/β proteins (alternating α and β)
  - α + β proteins

All-Alpha topologies

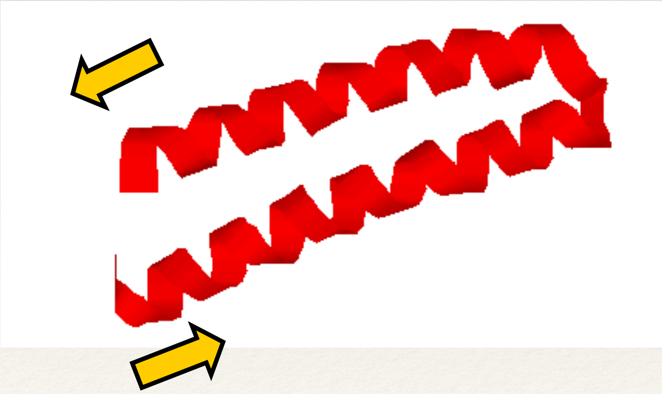
- The lone helix

Glucagon (hormone involved

Is regulating sugar metabolism)

PDB code : 1GCN

- The helix-turn-helix



ROP: RNA-binding Protein

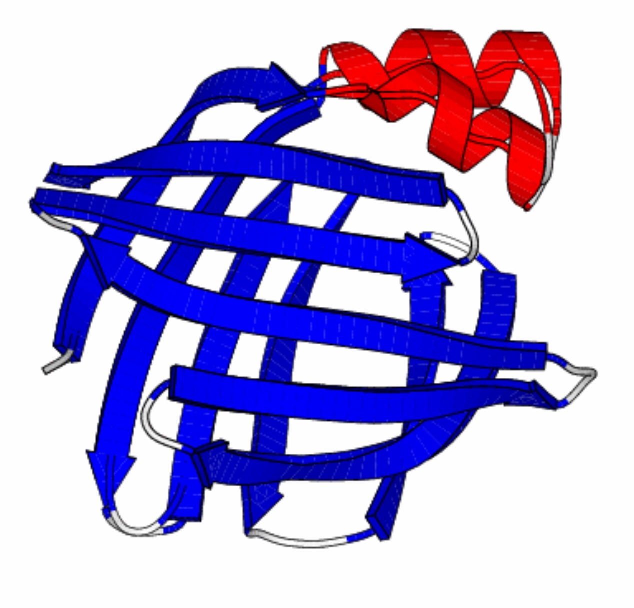
PDB code: 1ROP

The 2 helices

are twisted

All Beta Topology

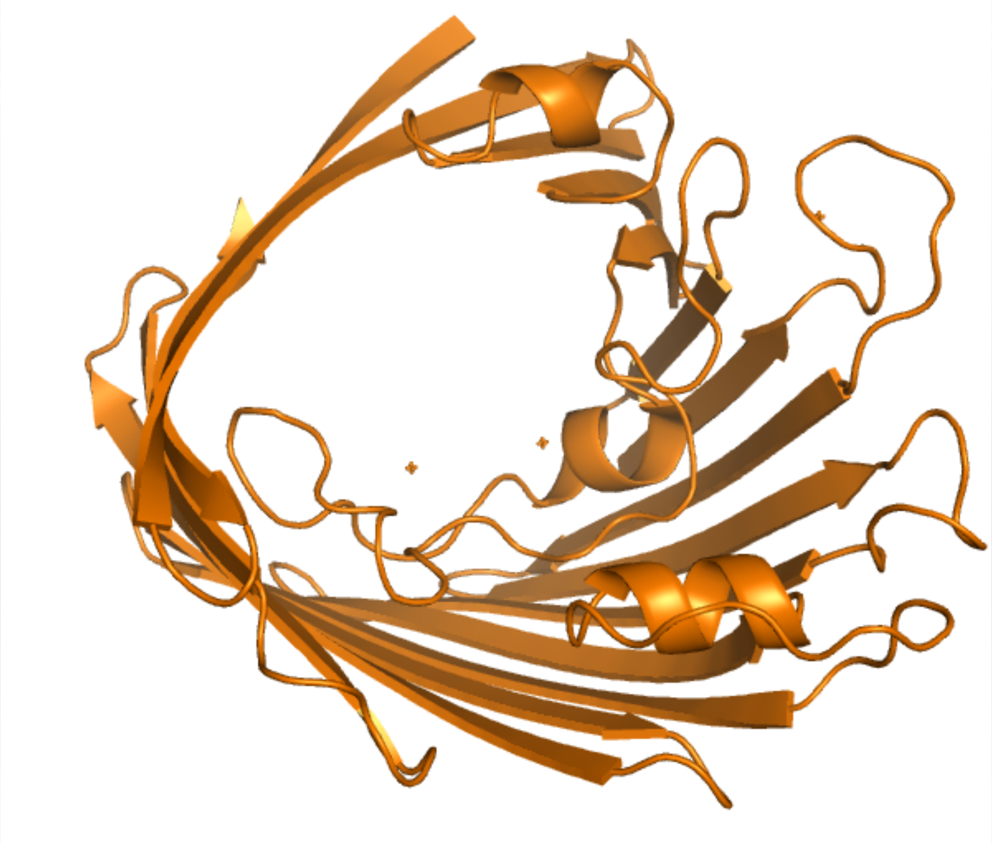
Beta sandwiches:



Fatty acid binding protein

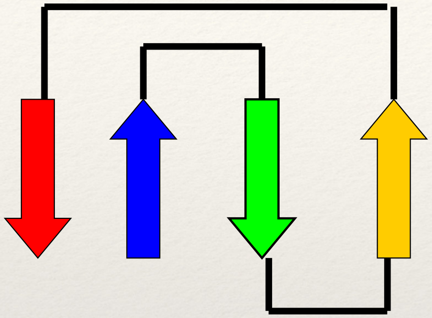
PDB code: 1IFB

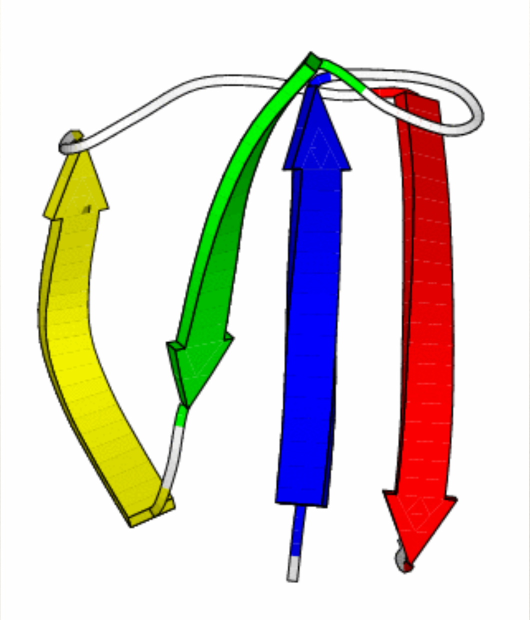
Closed Beta Barrel



PDB file: 2POR

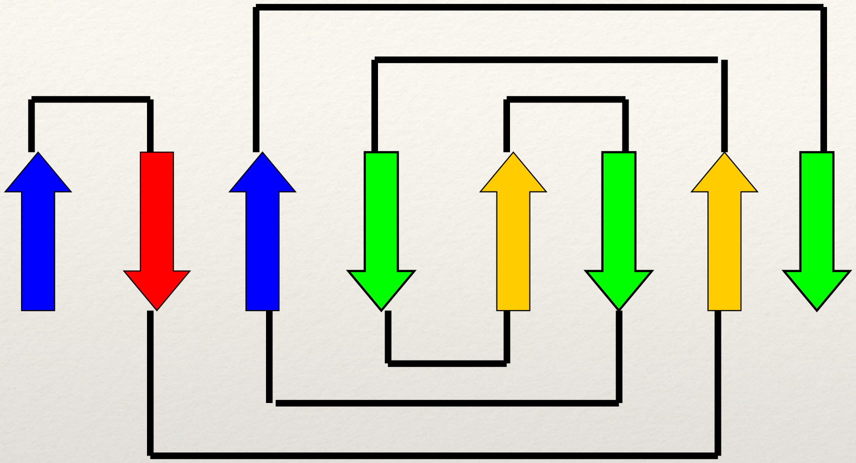
The Greek Key Topology





Folds including the Greek key topology include 4 to 13 strands.

The Jellyroll Topology

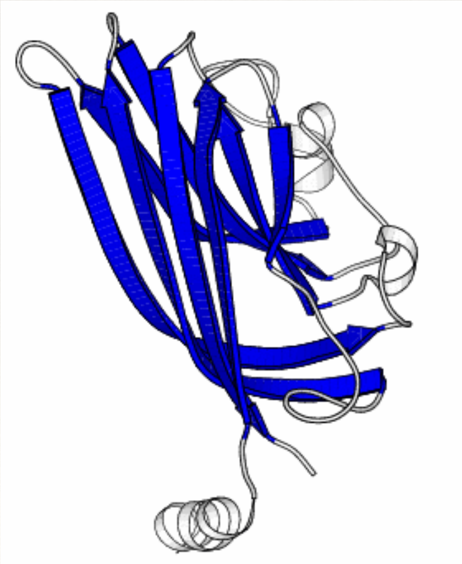


A Greek key with an

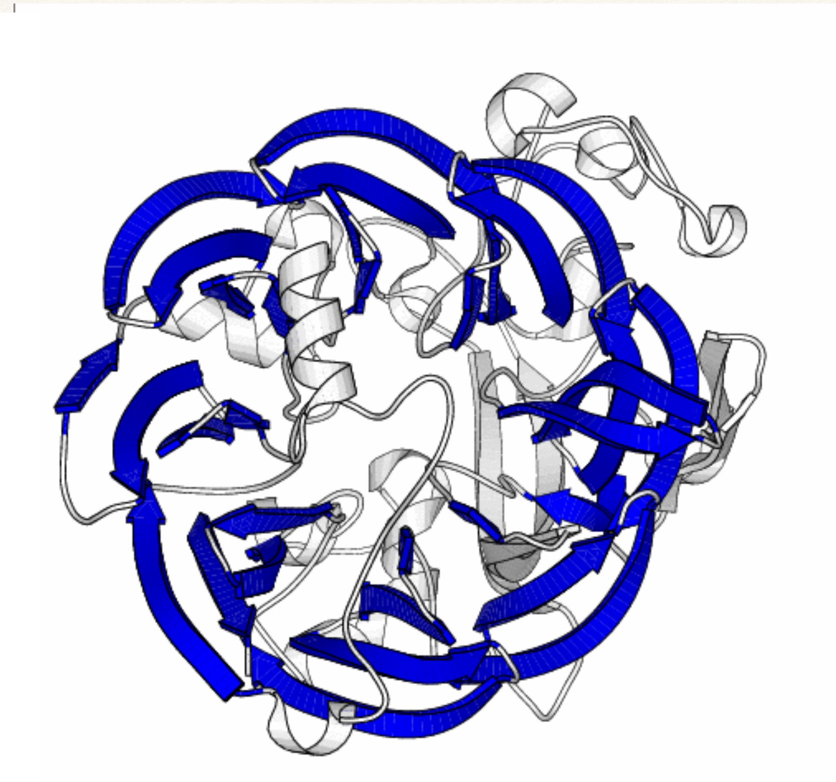
extra swirl

PDB code 2BUK

(coat protein of a virus)



The Beta Propellor



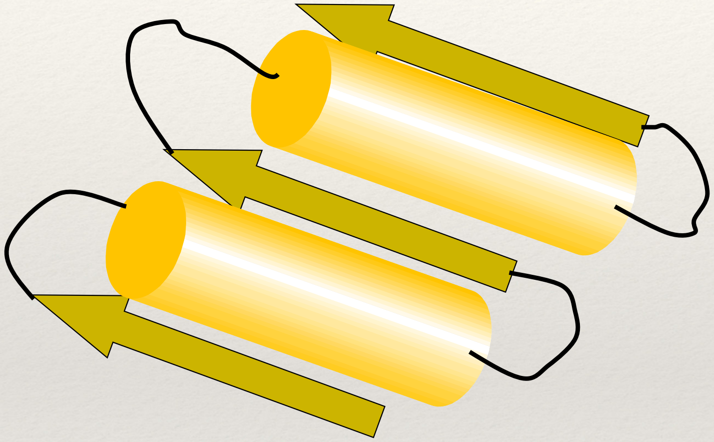
Eight-plated propellor:

Each plate is a four-stranded anti-parallel sheet

PDB code 4AAH

Alpha- Beta Topology

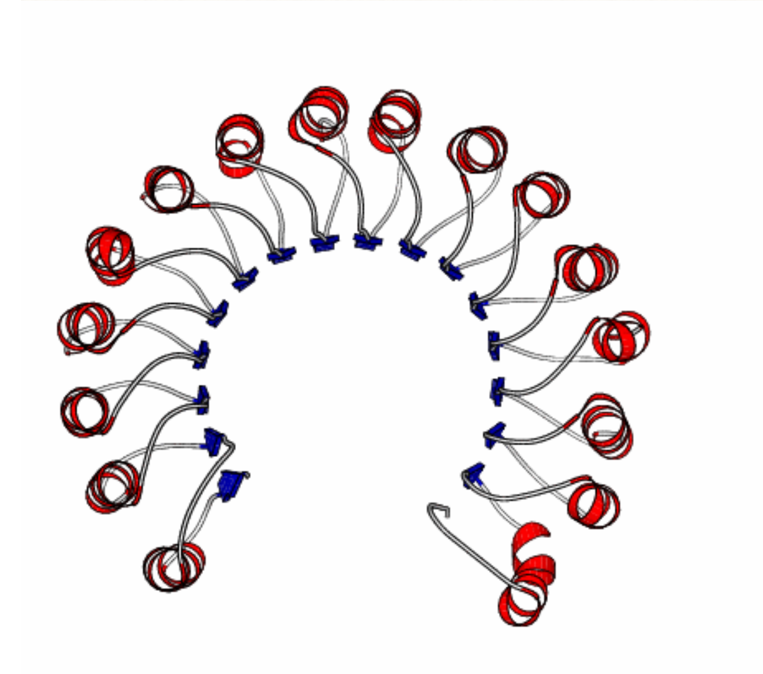
The Rossman fold:



Alternate beta / alpha motif

Always right handed

The Horseshoe



PDB code: 2BNH

The alpha/beta barrel

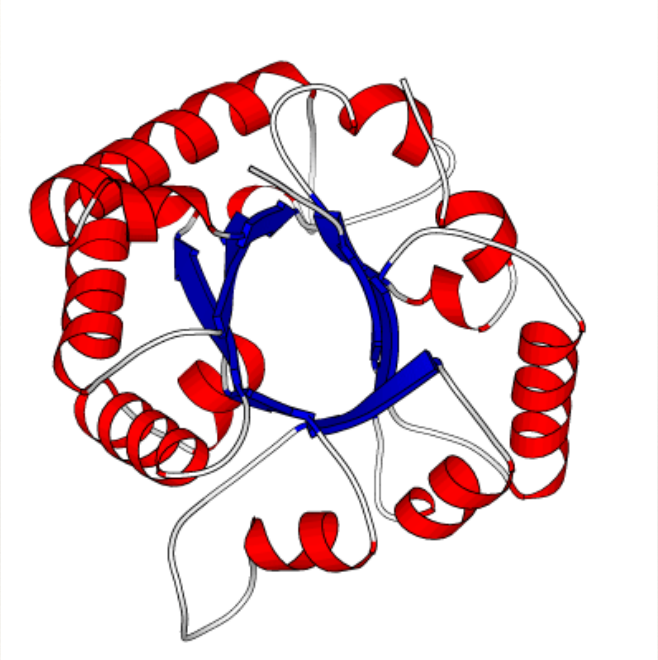
motifs, if the first strand

connects to the last,

then the structure resembles a

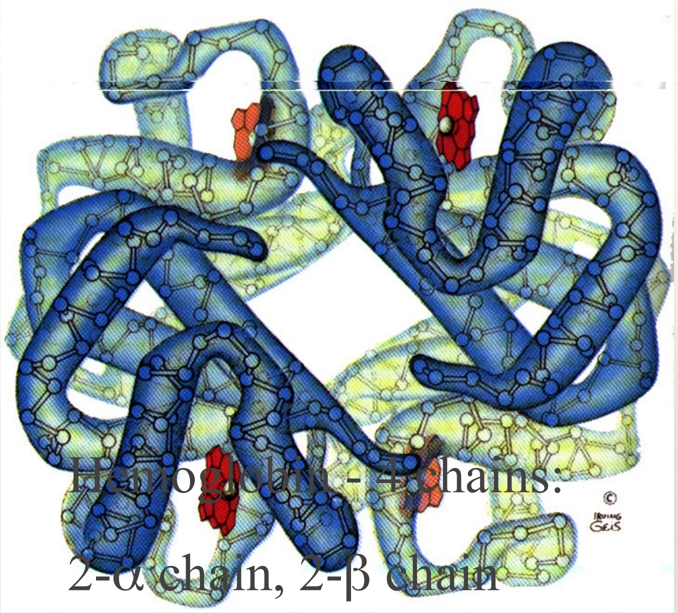
Barrel.

PDB code : 1TIM



Quaternary Structures

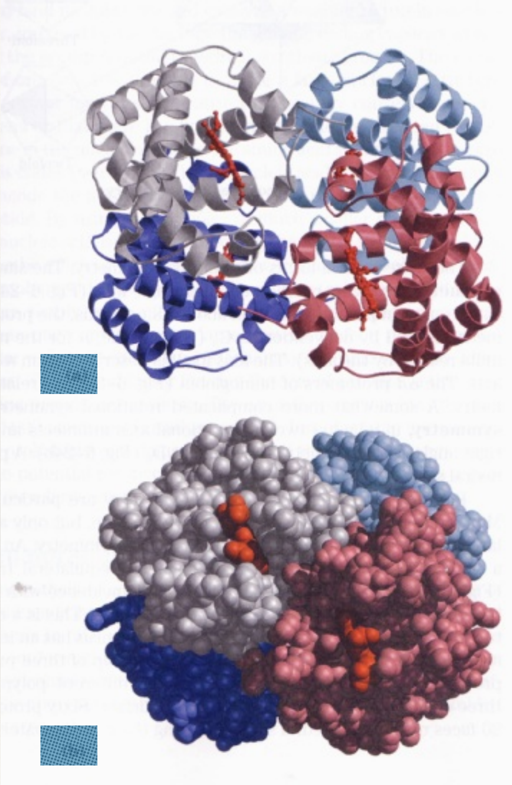
Assemblies of Protein Chains



Hemoglobin - 4 chains:

2-α chain, 2-β chain

(Heme- four iron groups)

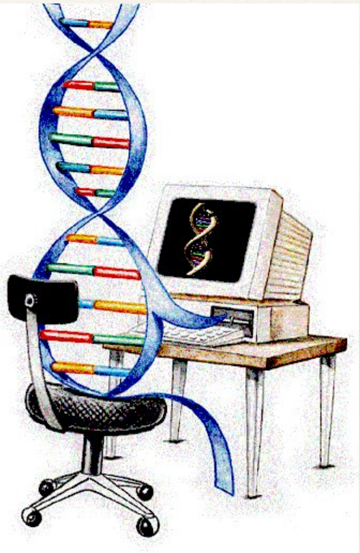


Structural Bioinformatics: Proteins

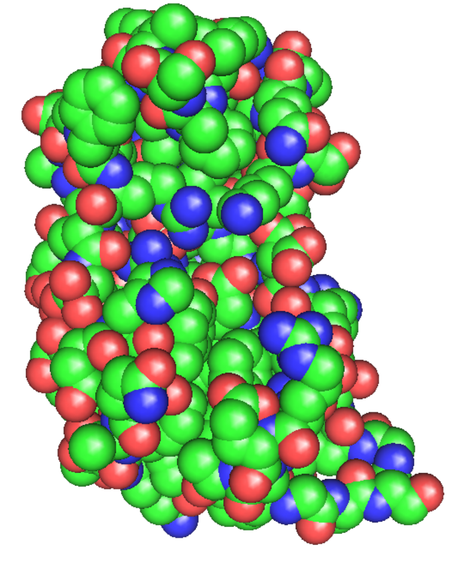
Proteins: Secondary Structures

Proteins: Tertiary and Quartenary Structure

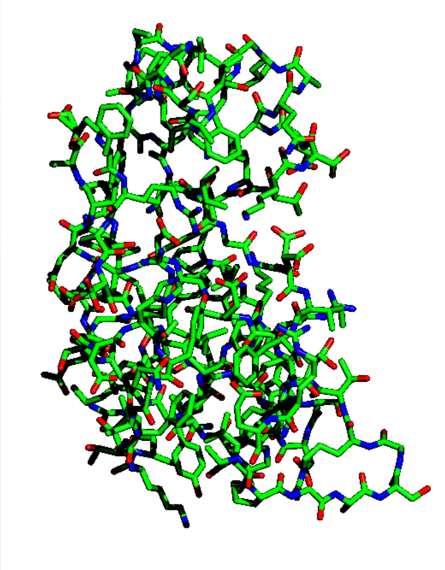
Proteins: Geometry



Protein Structure Representation



CPK: hard sphere model

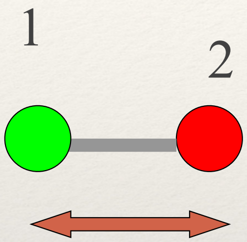


Ball-and-stick

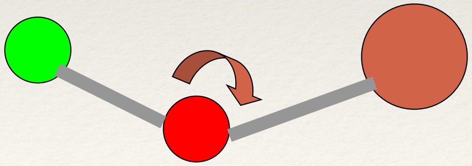
Cartoon

Degrees of Freedom in Proteins

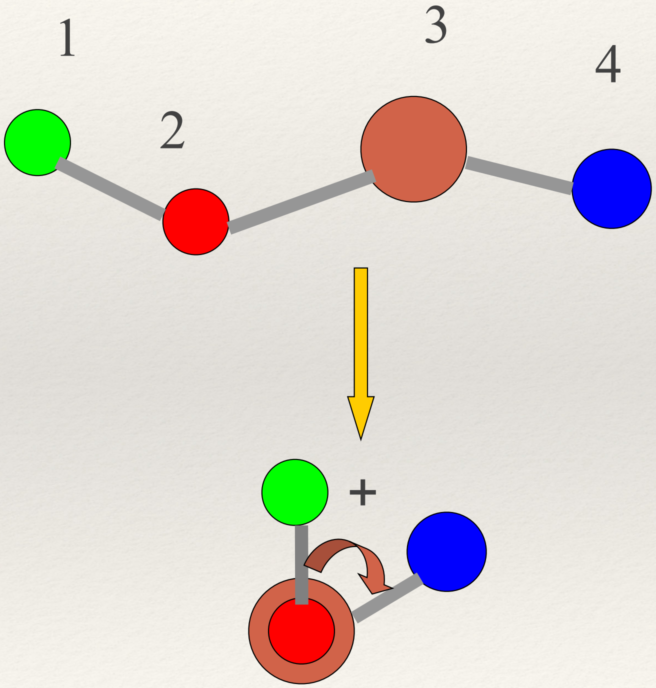
Bond



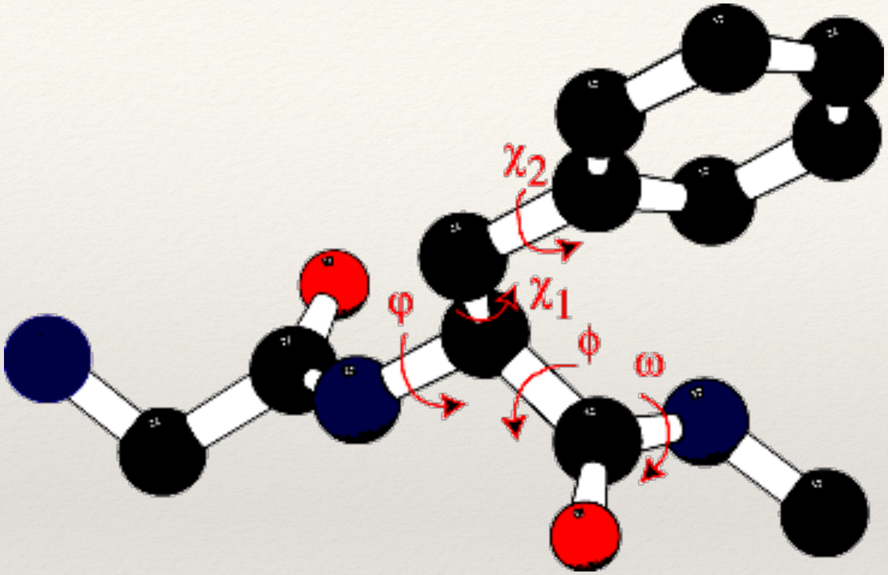
Angle



Dihedral angle



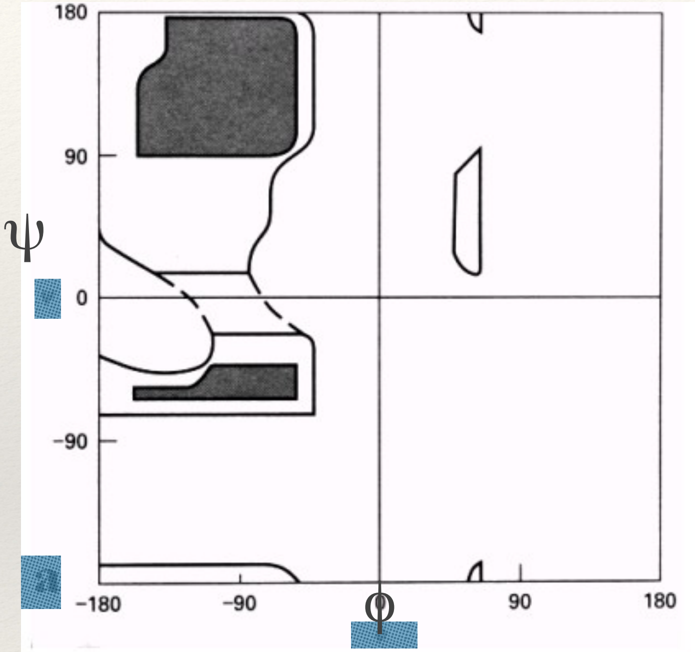
Protein Structure: Variables



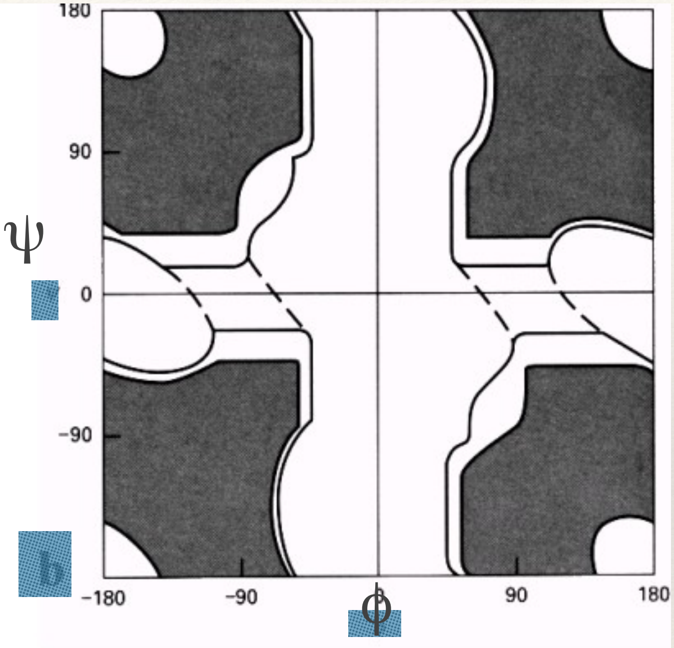
Backbone: 3 angles per residue : φ, ϕ and ω

Sidechain: 1 to 7 angles, χ; each χ has 3 favored values: 60^o, -60^o, 180^o.

Ramachandran Plots



All residues, but glycine



Glycine

Acta Cryst. (2002). D**58**, 768-776

What have we learnt?

- All proteins are polymers built up from 20 amino acids.
- All 20 amino acids have a similar structure: they all have a main-chain, consisting of an amino group and an acidic group, attached to a central carbon, named CA; the remaining atoms form the side-chain, that can be hydrophobic, polar or charged (acid or basic).
- The conformation of the backbone of amino acids is restricted, except for glycine that does not have a sidechain.
- There are 3 main graphical representations of proteins: space-filling, wireframe and cartoon.

What have we learnt?

- There are 3 major types of secondary structures: α-helices, β-sheets and β-turns.
- Most helices are α-helices, stabilized through a network of CO (i) --- HN (i+4) hydrogen bonds
- There are two types of β-sheets: parallel and anti-parallel
- β-turns correspond to 180 change in the backbone direction.

What have we learnt?

- There are three main classes of proteins: all Alpha, all Beta and Alpha + Beta. The latter can be divided in two, considering the alternating alpha/beta proteins as defining their own class.
- Bundles are common alpha-proteins
- Common beta folds include the greek key and the sandwiches. Immuno-globulins adopt a beta fold.
- The Rossman fold (alternating alpha/beta) is a common motif in proteins. It is found in the horseshoe, as well as in the TIM fold.