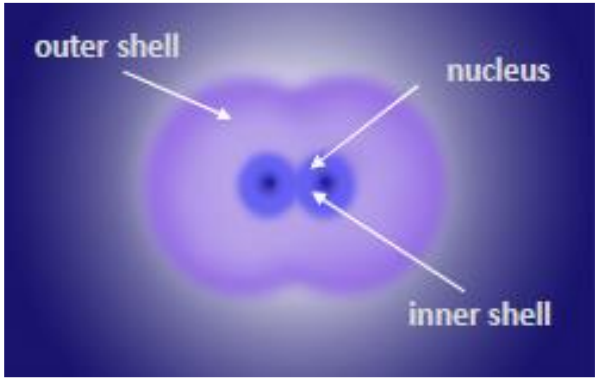
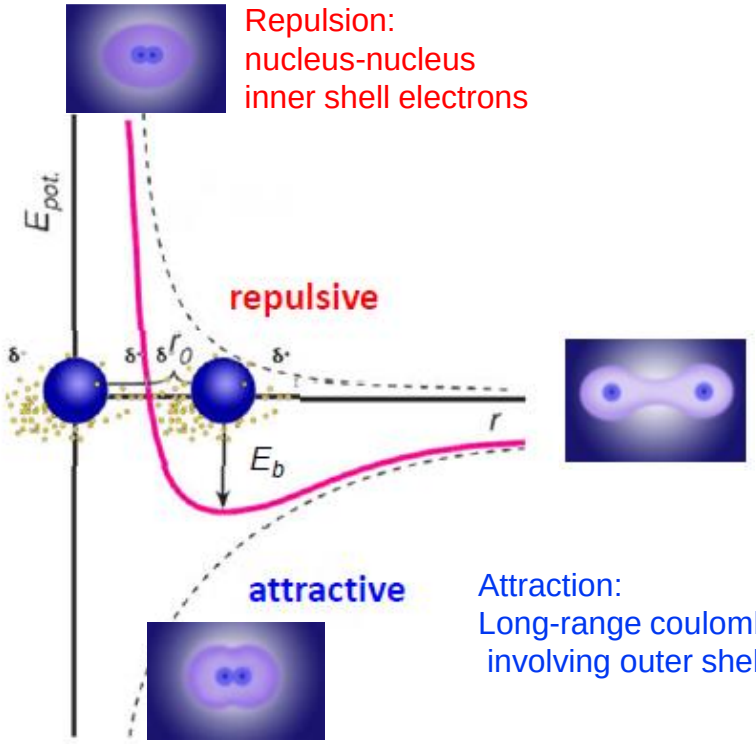
Structural hierarchy of proteins

- G.Schay
- Semmelweis University
- Dept. Biophysics and Radiation Biology

Proteins are LARGE structures consisting thousands of atoms.

The atomic interaction rules apply here too.



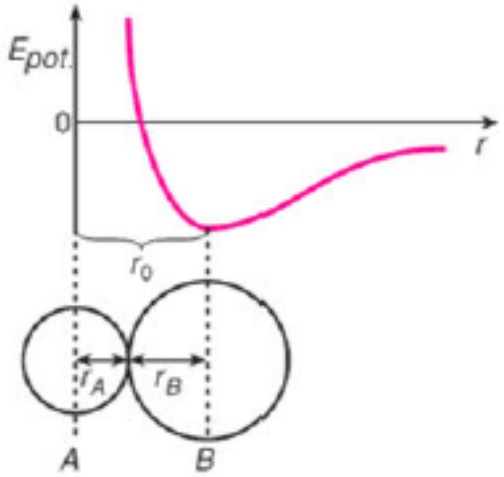


$$E_{pot} = \frac{-A}{r^n} + \frac{B}{r^m}$$

Power law

Example: 6-12 potential

Van der Waals radius: r_0



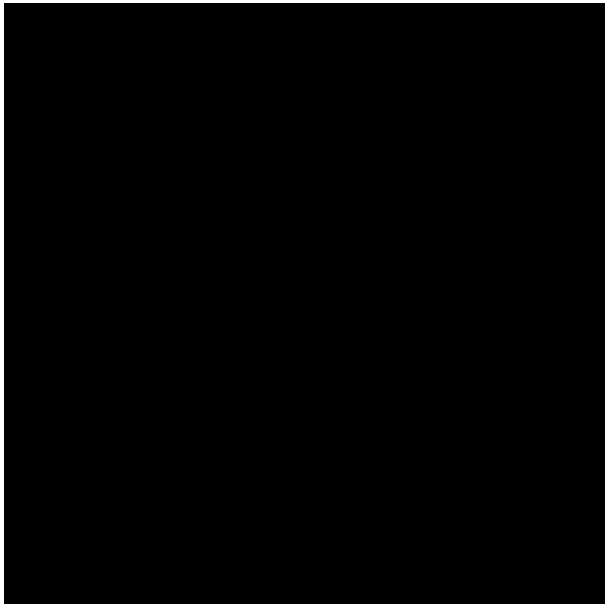
Primary bonds: E_b>1eV

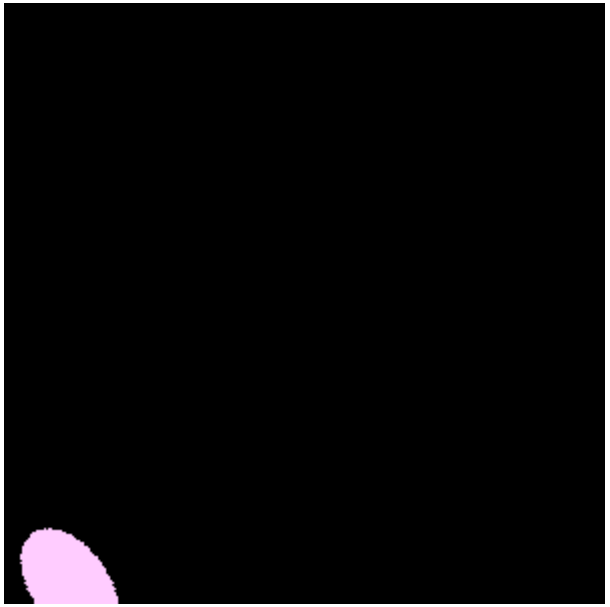
- **covalent:** common electron state around the participating nuclei
- **metallic bond:**multi-atomic system
- **ionic bond:** Coulomb-forces between ions

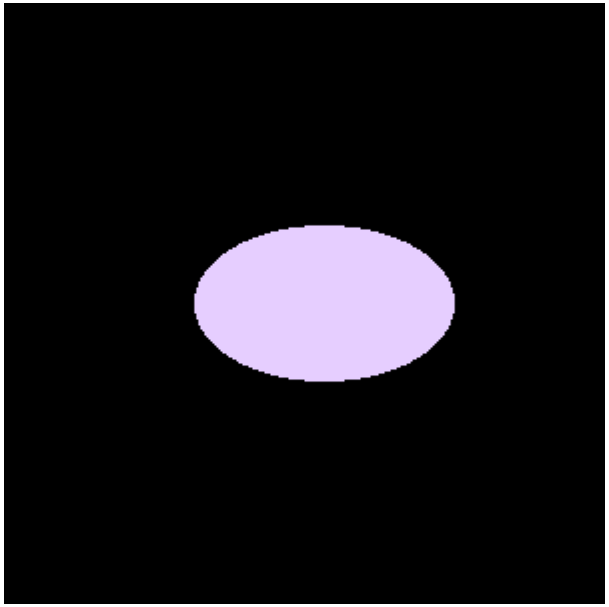
Secondary bonds: **inter**molecular bonds, weaker than **intra**molecular

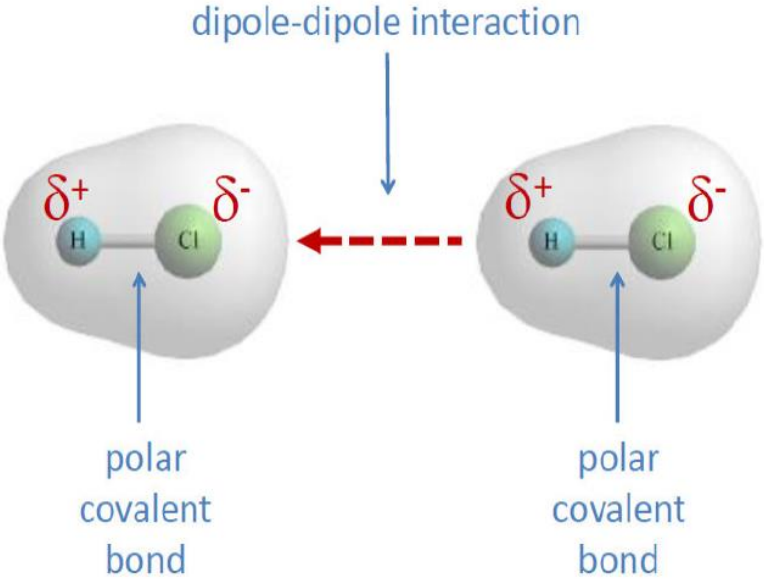
- **Van der Waals:** between two apolar atoms (without permanent dipole moment) where a temporarily created dipole interacts with an apolar molecule or atom thus converting it into a dipole **(induced dipole)**
- Van der Waals radius of atoms can be used (-> “spacefill models, Solvent Accessible Surface Area”)
- Weak: (E_b ~ 0,02 eV)

1eV=1.9·10^{-19} Joule

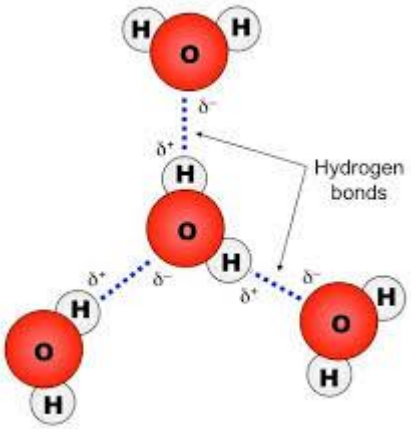


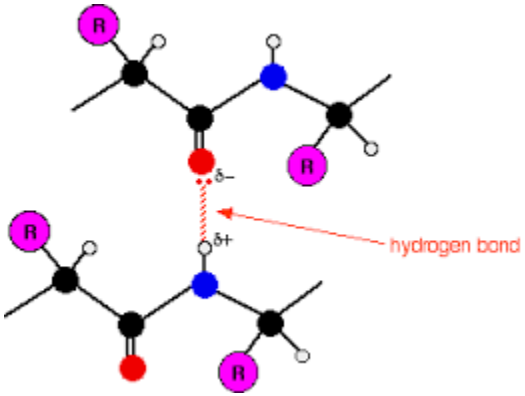






$E_{binding}$ ~ 0.003 – 0.02 eV





$E_{binding}$ ~ 0.2 eV

Classification of proteins

According to the biological function

- Enzymes (E.g.: trypsin, cytochrome-c…)
- Transport proteins (E.g.: hemoglobin...)
- Chaperone proteins
- Toxins (E.g.: ricin, snake venom…)
- Hormones (E.g.: insulin, growth hormone…)
- Contractile proteins (E.g.: myosin, actin…)
- Structure proteins (E.g.: collagen, elastin …)
- Storage proteins (E.g.: ovalbumin, casein, ferritin…)
- Others (E.g.: histone proteins…)

Classification of proteins

According to shape and solubility

- Fibrous proteins (E.g.: collagen…)

  These proteins have a rod like structure. They are not soluble in water.

- Globular proteins (E.g.: hemoglobin, myoglobin...)

  These proteins more or less spherical in nature. Due to their distribution of amino acids (hydrophobic inside, hydrophilic outside) they are very soluble in aqueous solution.

- Membrane proteins (E.g.: rhodopsin…)

  These are protein which are in association with lipid membranes. Those membrane proteins that are embedded in the lipid bilayer have extensive hydrophobic amino acids that interact with the non-polar environment of the bilayer interior. Membrane proteins are not soluble in aqueous solution. Rhodopsin is an integral membrane protein and is embedded in the bilayer.

Classification of proteins

According to secondary structure

- All Alpha helical (E.g.: myoglobin…)

  Proteins that contain only (in some exceptional instances there may be isolated beta-sheets) alpha helical secondary structure.

- Alpha/beta (E.g.: Triose phosphate-isomerase…)

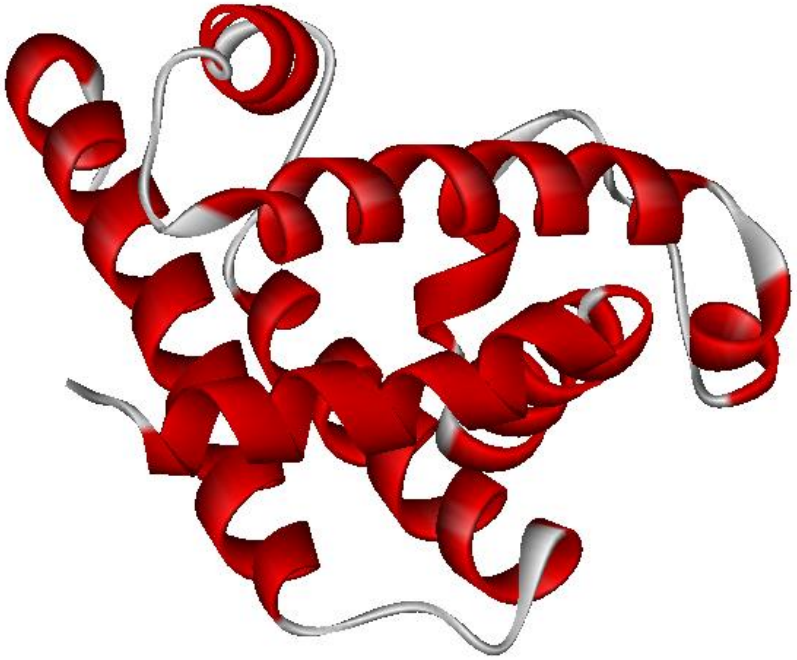
  Proteins that contain alternating alpha-helical and beta-sheet secondary structure elements. The alternating helical and sheet segments may form a closed barrel like structure.

- Alpha+beta (E.g.: ribonuclease…)

  In these proteins the alpha helical and beta sheet regions occur in independent regions of the molecule.

- Beta sheet (E.g.: tenascin…)

  Protein that contain only (in some exceptional instances there may be isolated alpha helices) beta-sheet secondary structure.

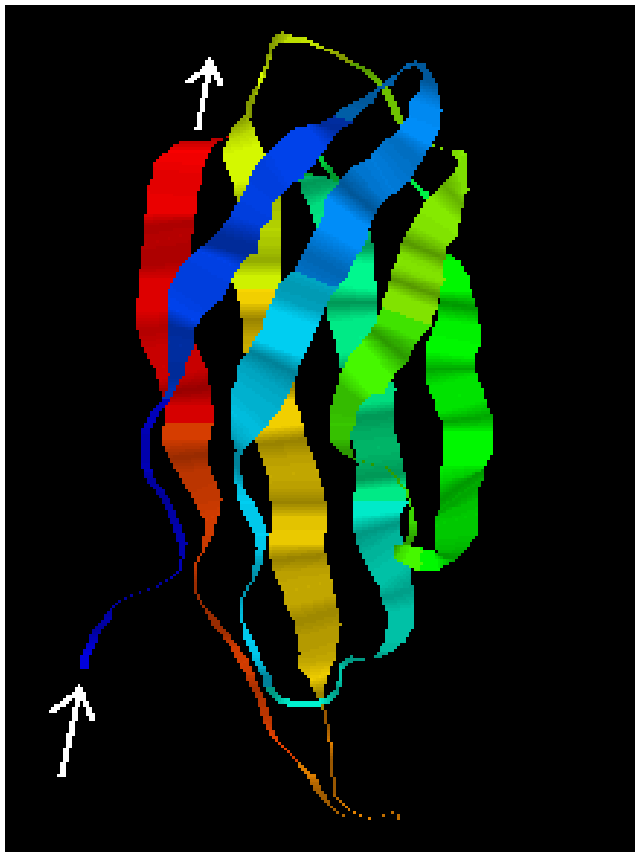


Myoglobin

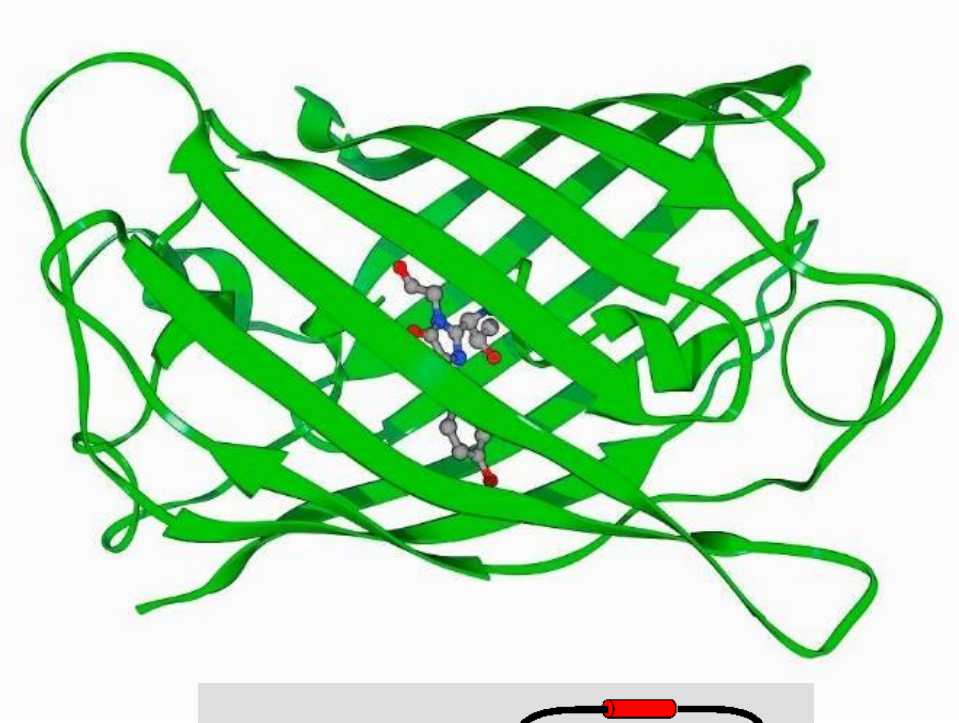
All alpha

Tenascin

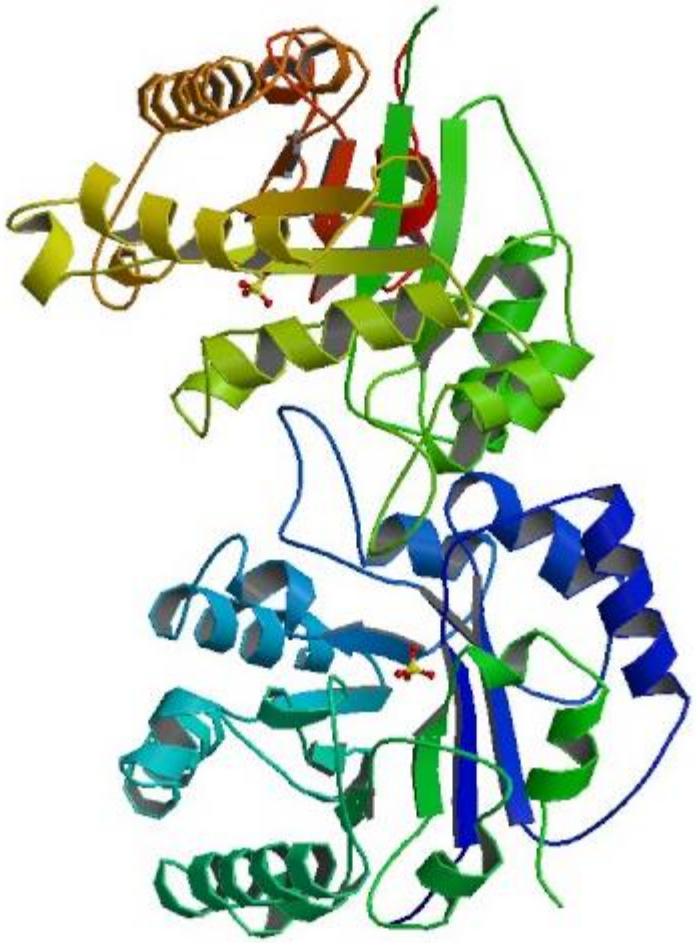
Beta

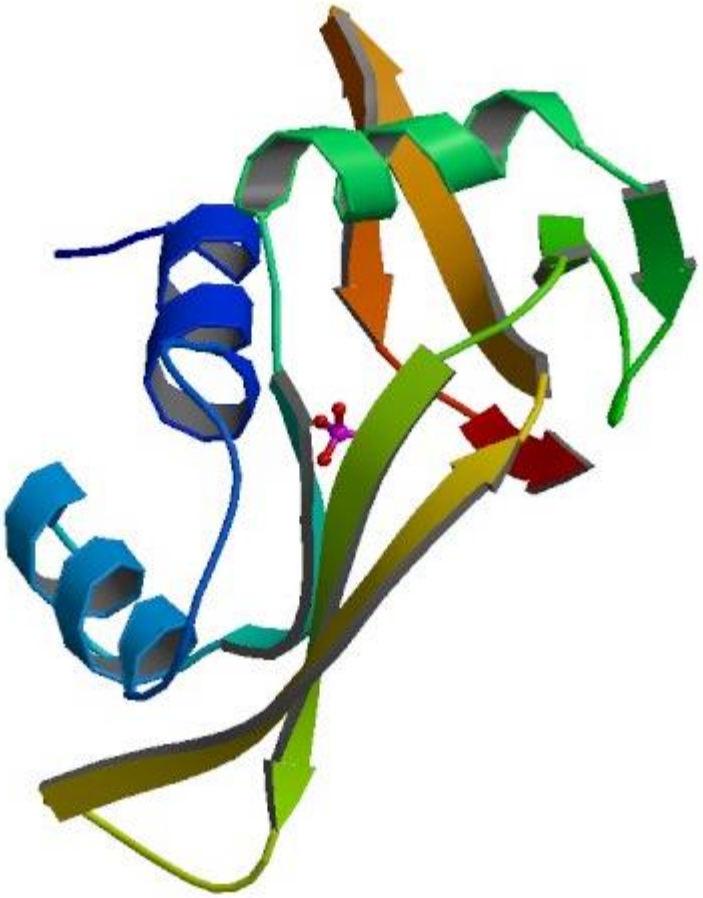


Beta sandwich



Beta barrel (GFP)





Alpha/beta structure (Triose phosphate isomerase)

Alpha+beta (ribonuclease)

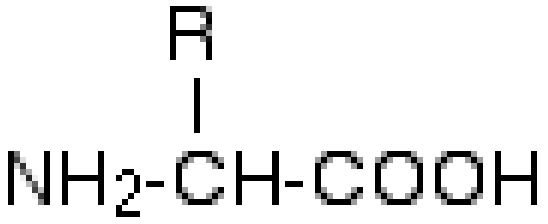
Structural hierarchy

- Primary structure
- Secondary structure
- Tertiary structure
- Quaternary structure

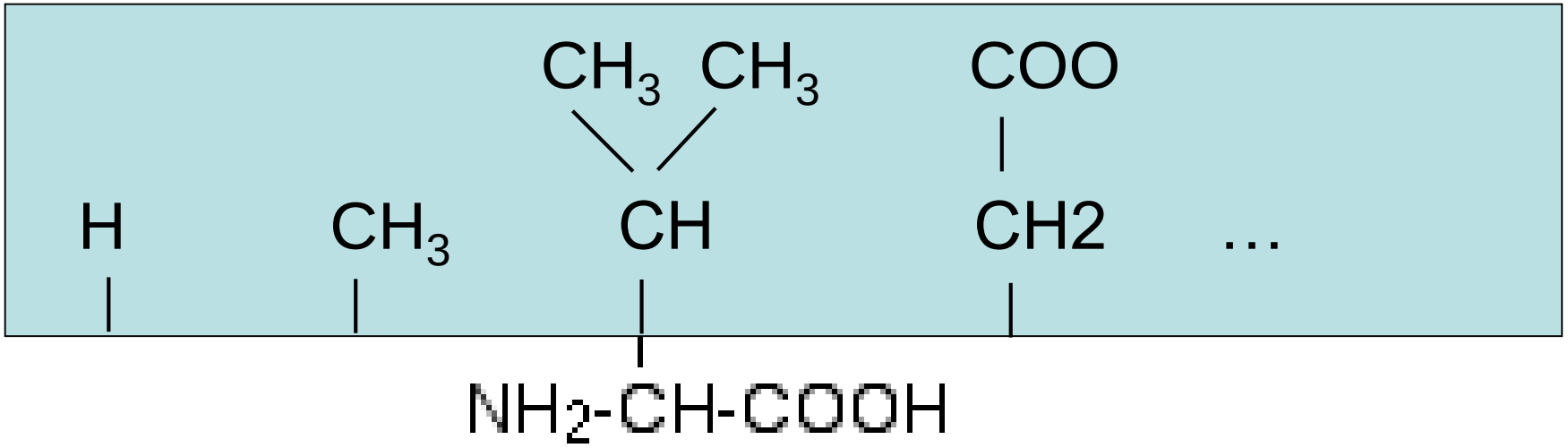
Supramolecular organizations

Amino acids: the building blocks of proteins

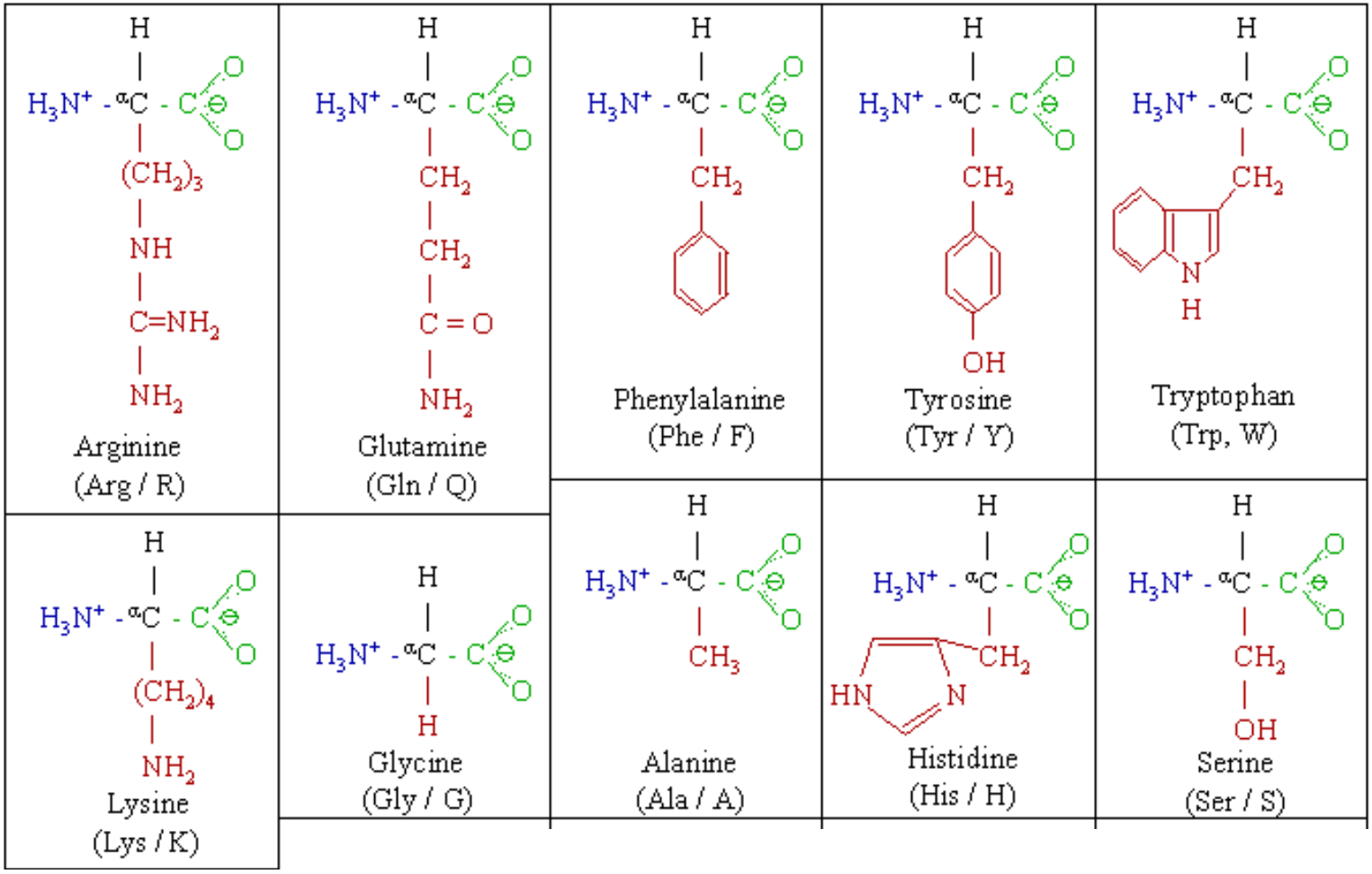
General structure of amino acids:



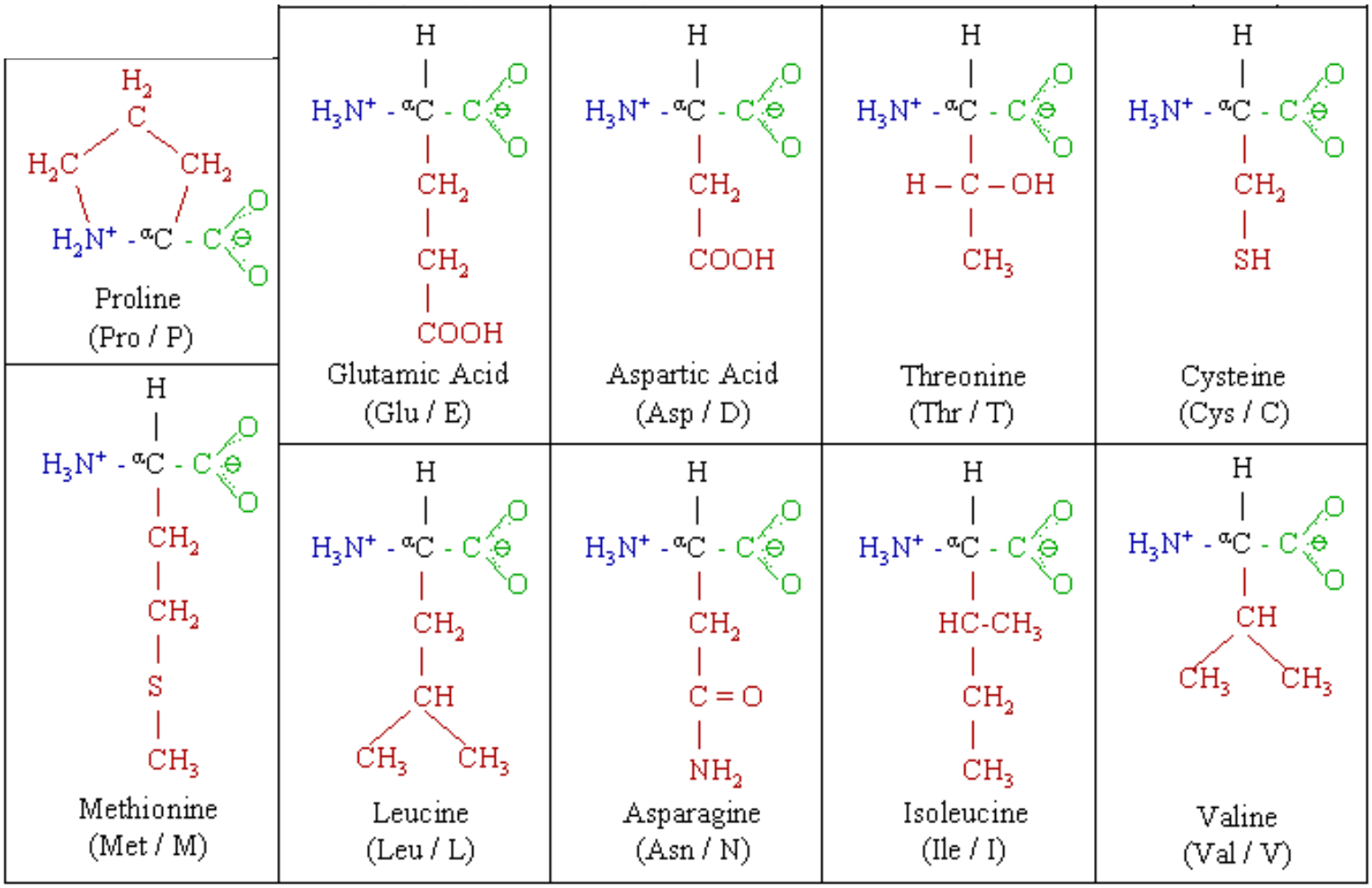
Structural variability:



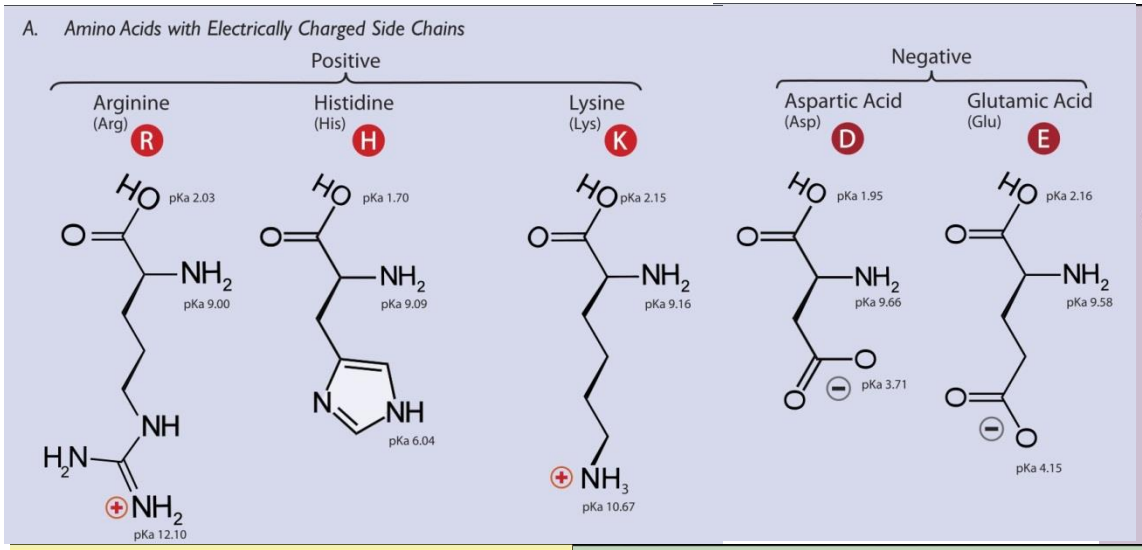
Amino acids of the proteins

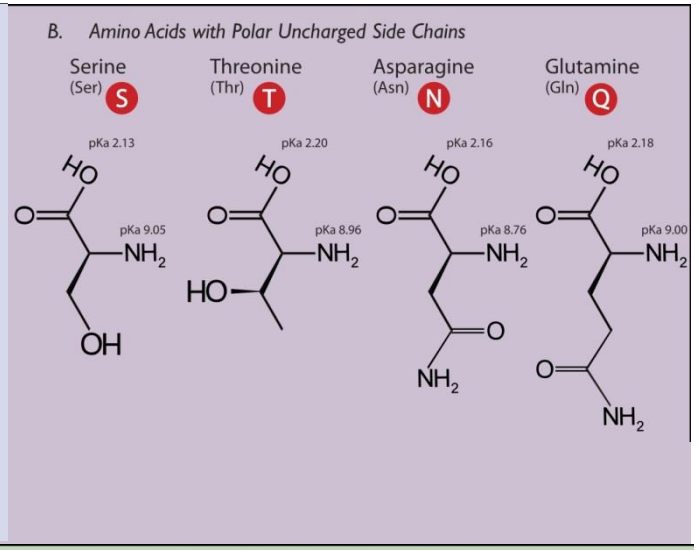


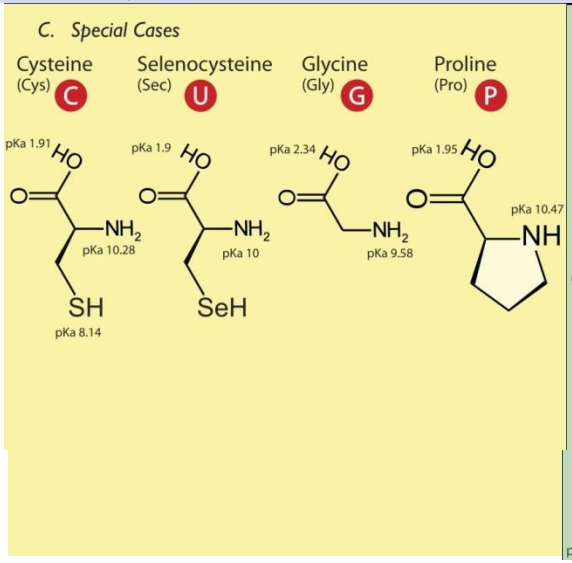
Amino acids of the proteins

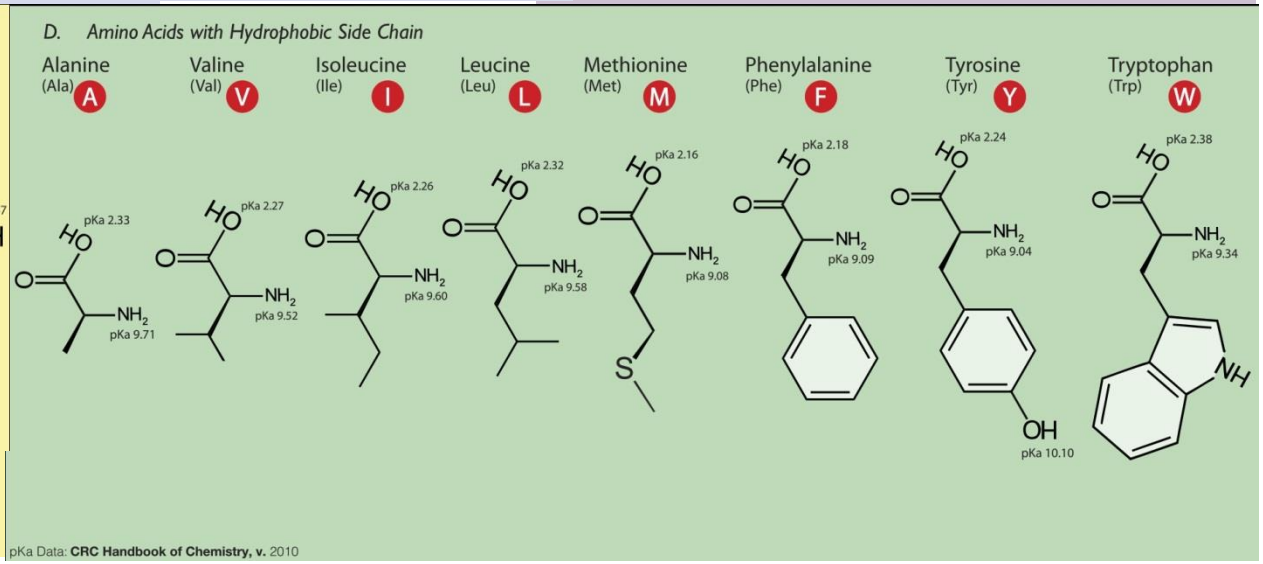


Properties of amino acids



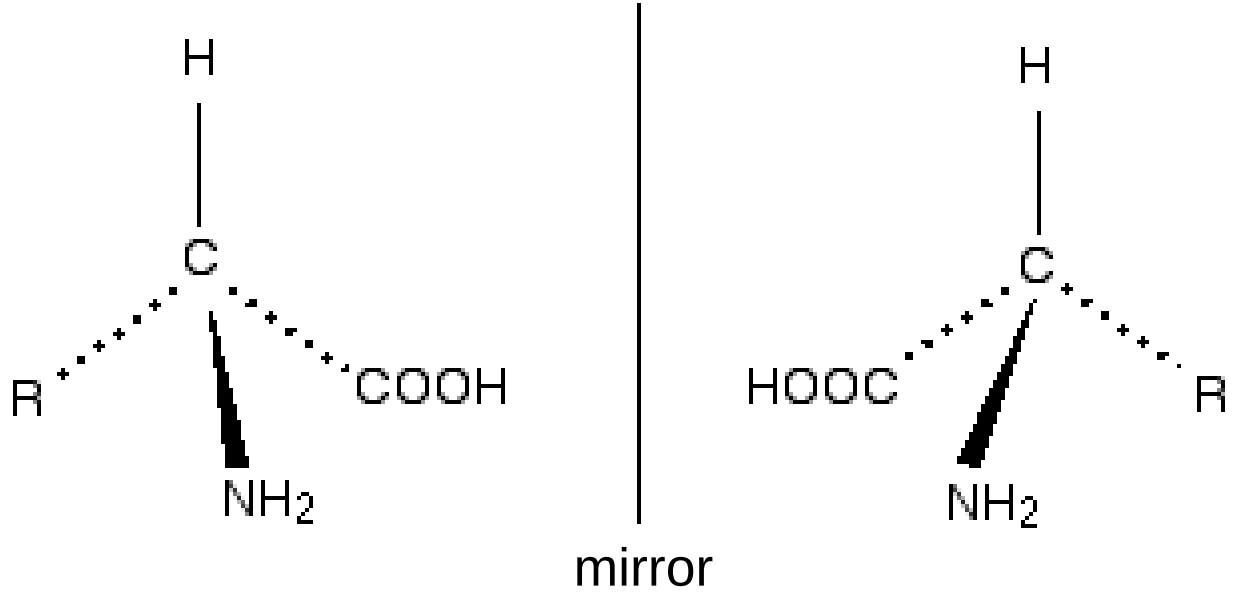






pKa Data: **CRC Handbook of Chemistry, v.** 2010

Chirality

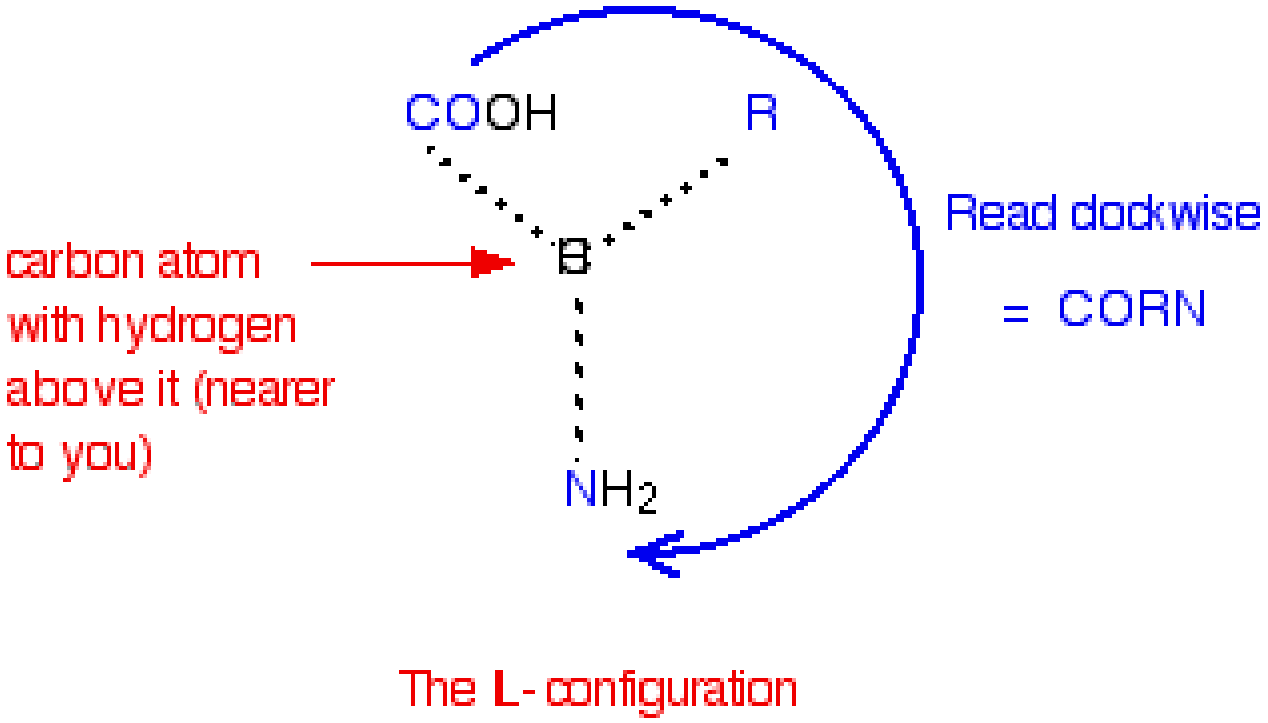


Chirality center:
a carbon atom, connected to four different atoms.

Hand:

Optical activity
(rotation of the plane of polarization of linearly polarized light)

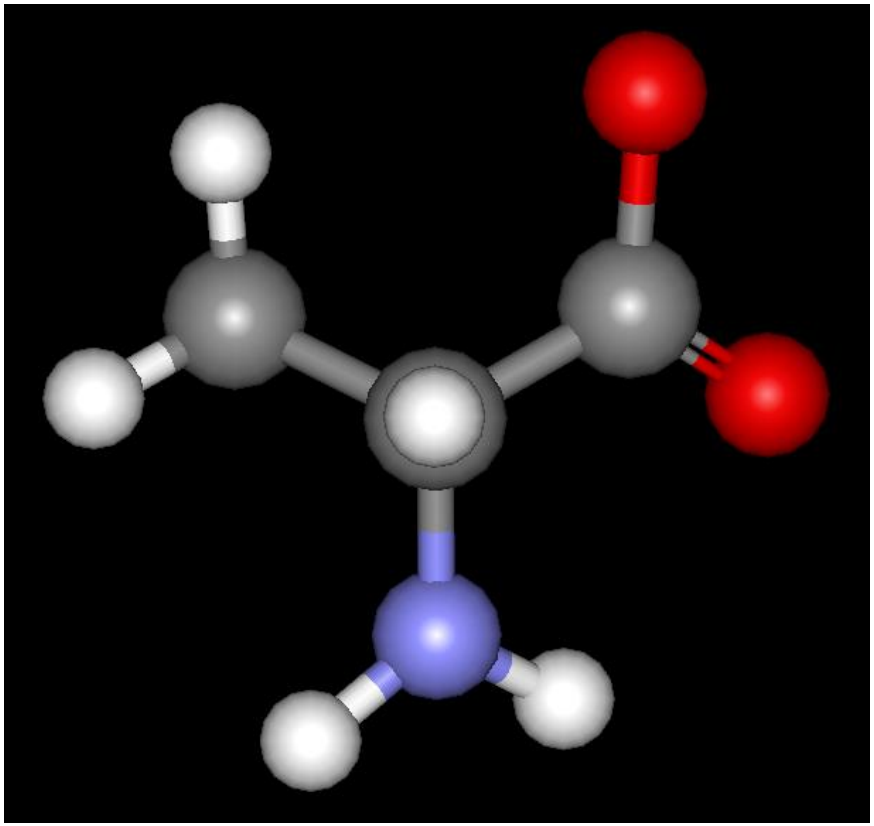
D and L enantiomers



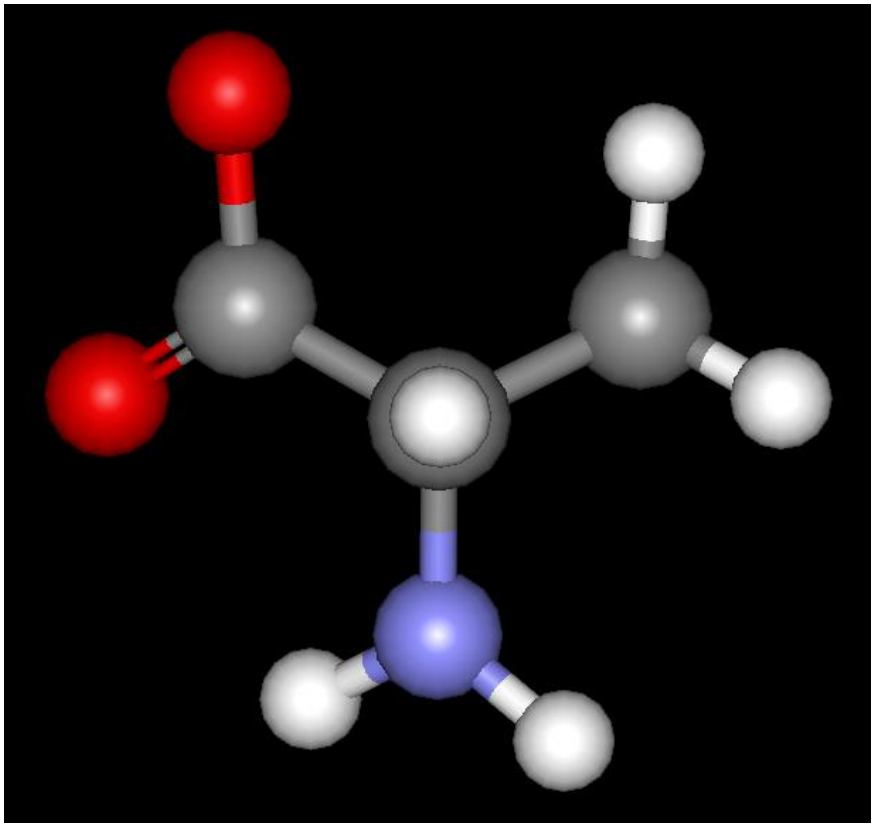
In living systems: L type !

No direct connection between the L-D enantiomer forms an the direction of the rotation of the polarization plane.
E.g.: (+)alanine (-)cysteine (-)tyrosine (+)valine

D and L enantiomers of alanine

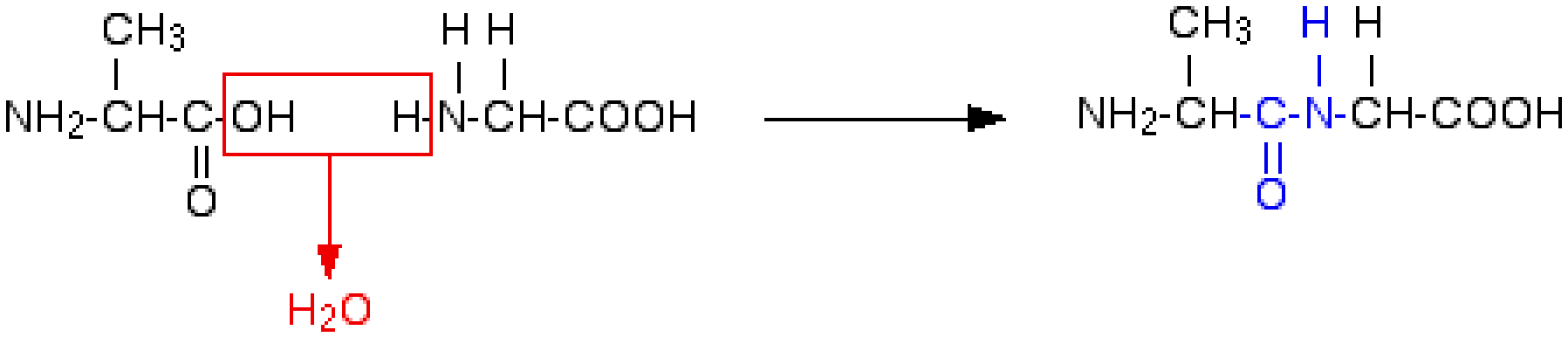


D



L

Connection of the amino acids: the peptide bond



Peptide 2.. c.a 20 amino acids

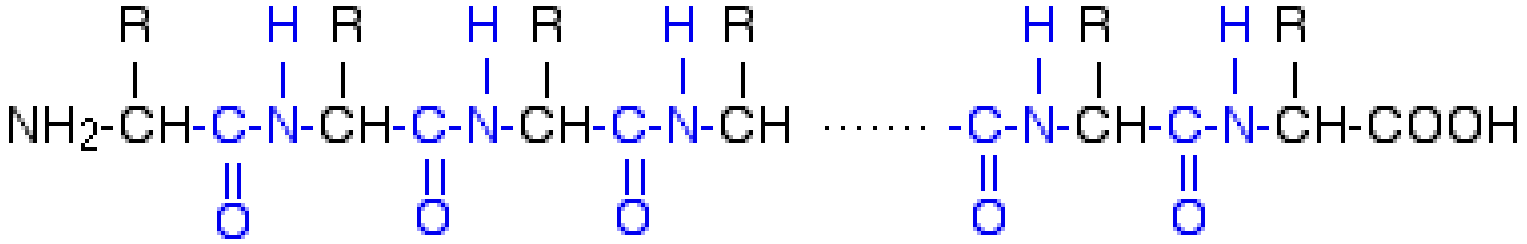
Protein: more than 20 amino acids

The primary structure

Primary structure:

the amino acid sequence in the polypeptide chain

Which direction? N terminal -> C-terminal



E.g.: (myoglobin, 1YMB)

GLY LEU SER ASP GLY GLU TRP GLN GLN VAL LEU ASN VAL…

... ALA LYS TYR LYS GLU LEU GLY PHE GLN GLY

Example: Myoglobin

Primary structure with 3-letter code (153 aa.):

GLY LEU SER ASP GLY GLU TRP GLN GLN VAL LEU ASN VAL
TRP GLY LYS VAL GLU ALA ASP ILE ALA GLY HIS GLY GLN
GLU VAL LEU ILE ARG LEU PHE THR GLY HIS PRO GLU THR
LEU GLU LYS PHE ASP LYS PHE LYS HIS LEU LYS THR GLU
ALA GLU MET LYS ALA SER GLU ASP LEU LYS LYS HIS GLY
THR VAL VAL LEU THR ALA LEU GLY GLY ILE LEU LYS LYS
LYS GLY HIS HIS GLU ALA GLU LEU LYS PRO LEU ALA GLN
SER HIS ALA THR LYS HIS LYS ILE PRO ILE LYS TYR LEU
GLU PHE ILE SER ASP ALA ILE ILE HIS VAL LEU HIS SER
LYS HIS PRO GLY ASP PHE GLY ALA ASP ALA GLN GLY ALA
MET THR LYS ALA LEU GLU LEU PHE ARG ASN ASP ILE ALA
ALA LYS TYR LYS GLU LEU GLY PHE GLN GLY

Example: Myoglobin

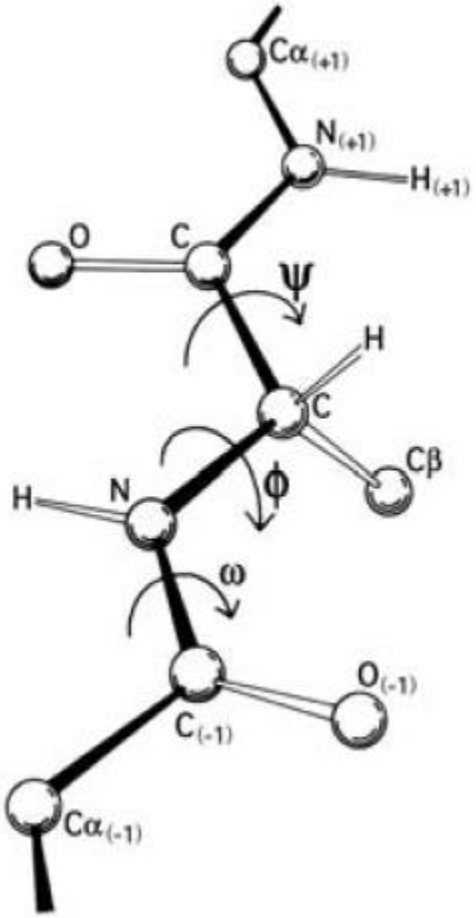
Primary structure with one letter code (153 aa.):

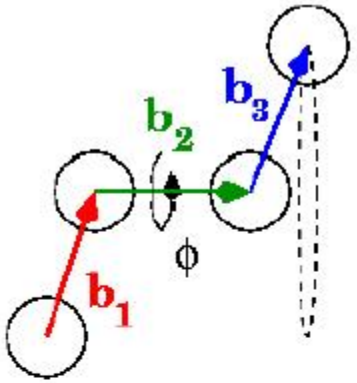
```
>1YMB:A|PDBID|CHAIN|SEQUENCE
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAE
MKASEDLKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFIS
DAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG
(FASTA format)
```

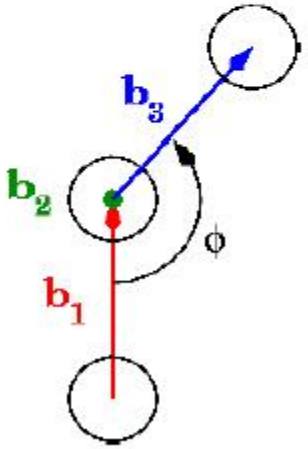
Rotation around the chemical bonds of the backbone

3 backbone bonds for each AA

- 1 rigid (delocalization)
- 2 rotation possible:
  - Φ, Ψ dihedral angles







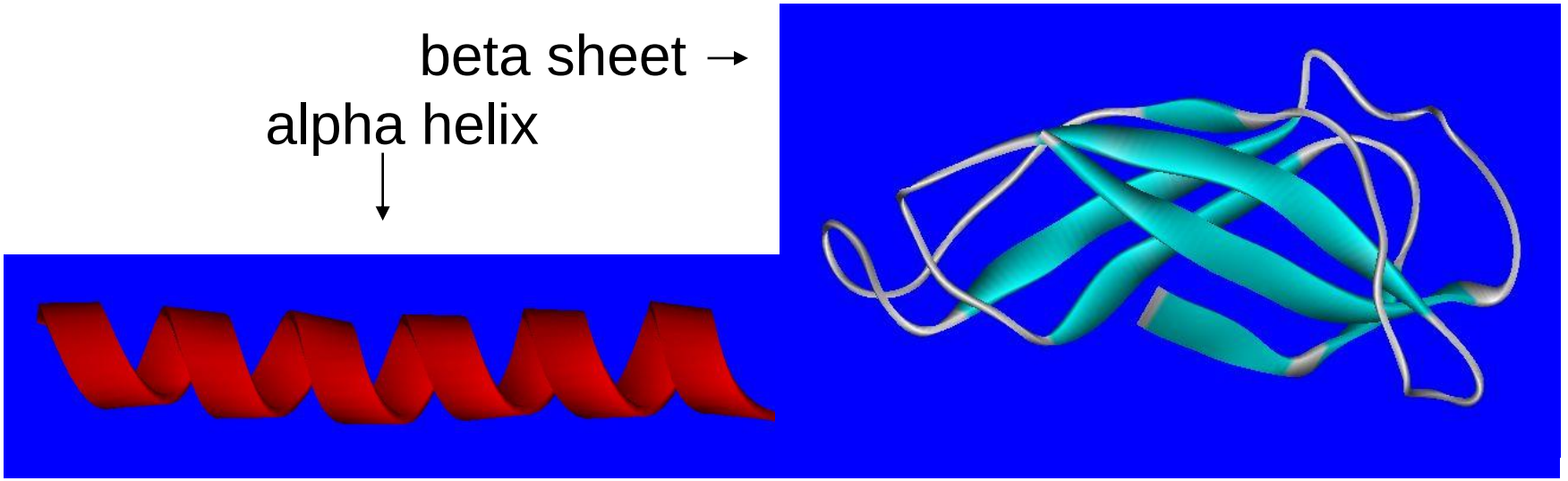
2N rotational degree of freedom

Secondary structure

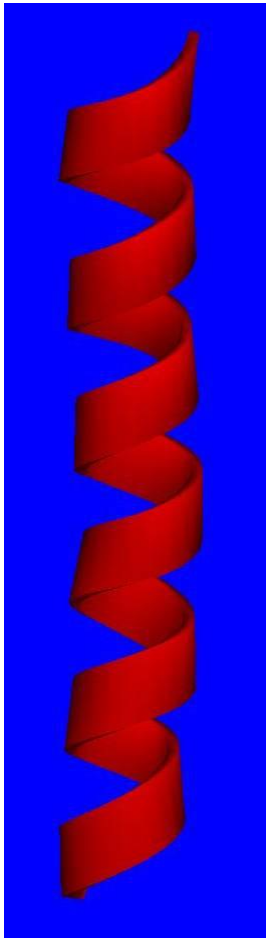
Regular local structures of linear segments of polypeptide chains.

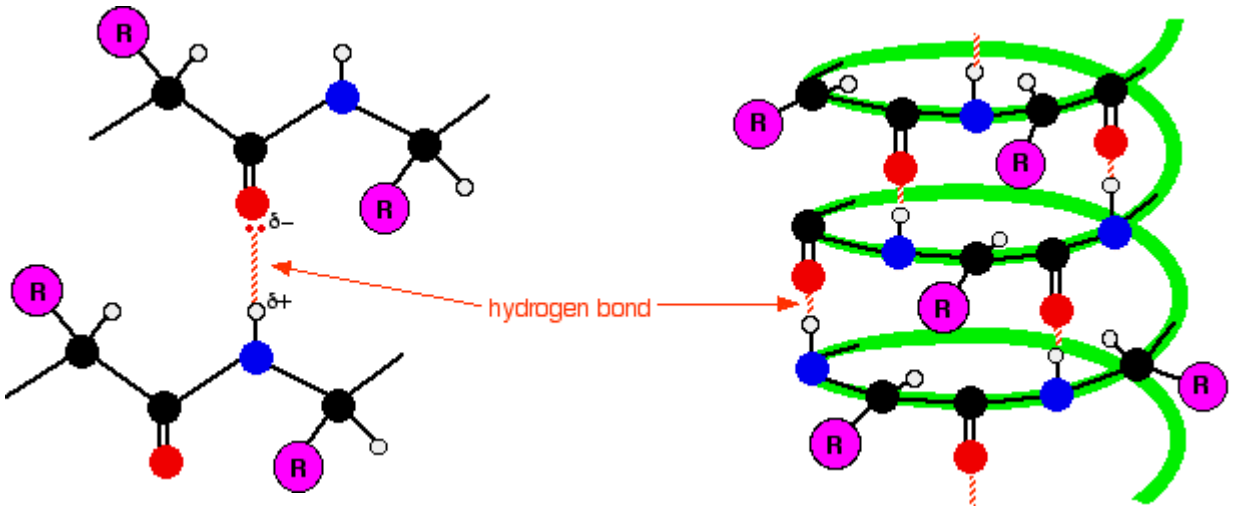
They are stabilized by H-bonds

Typical forms:

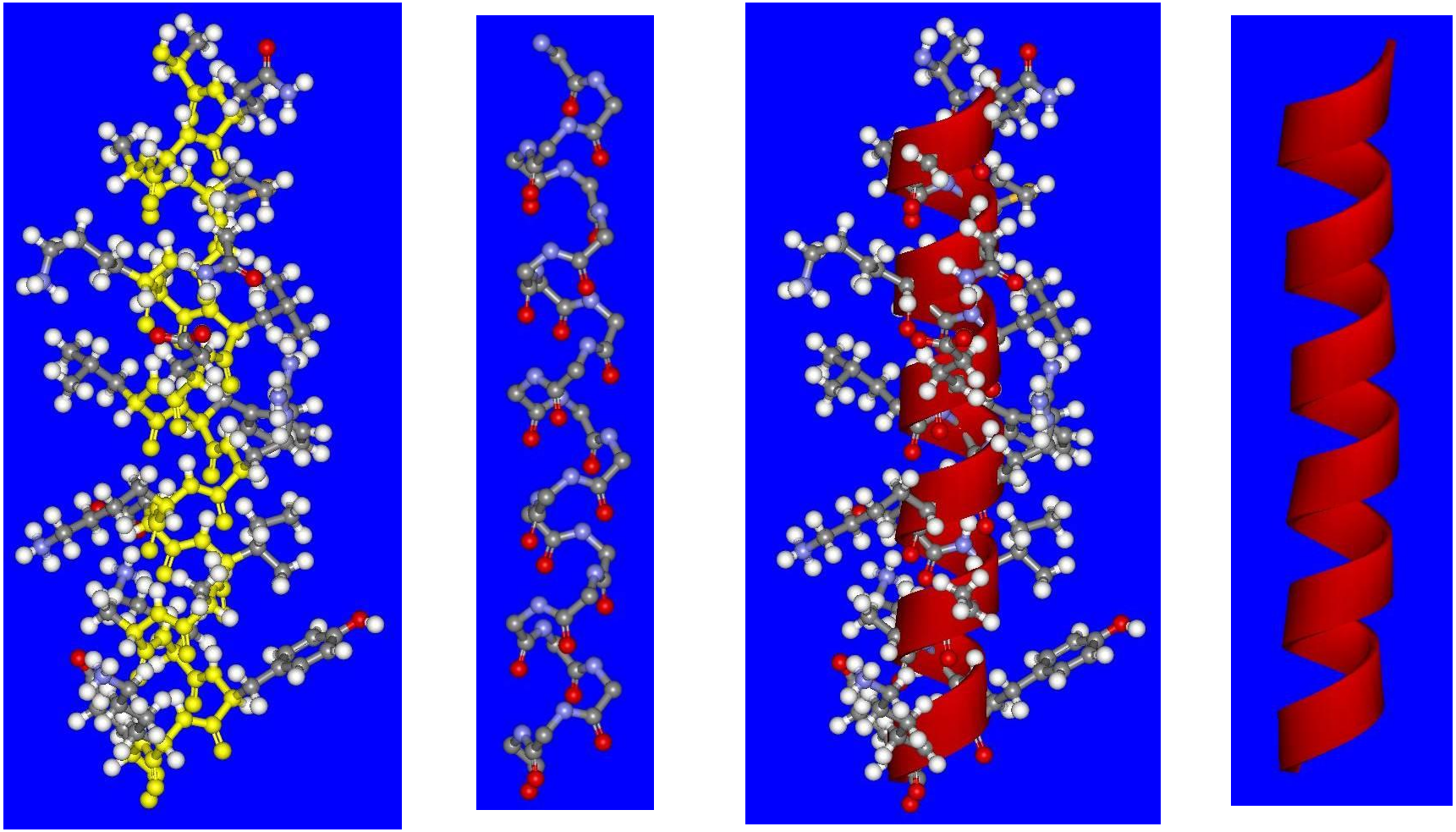


Alpha helix

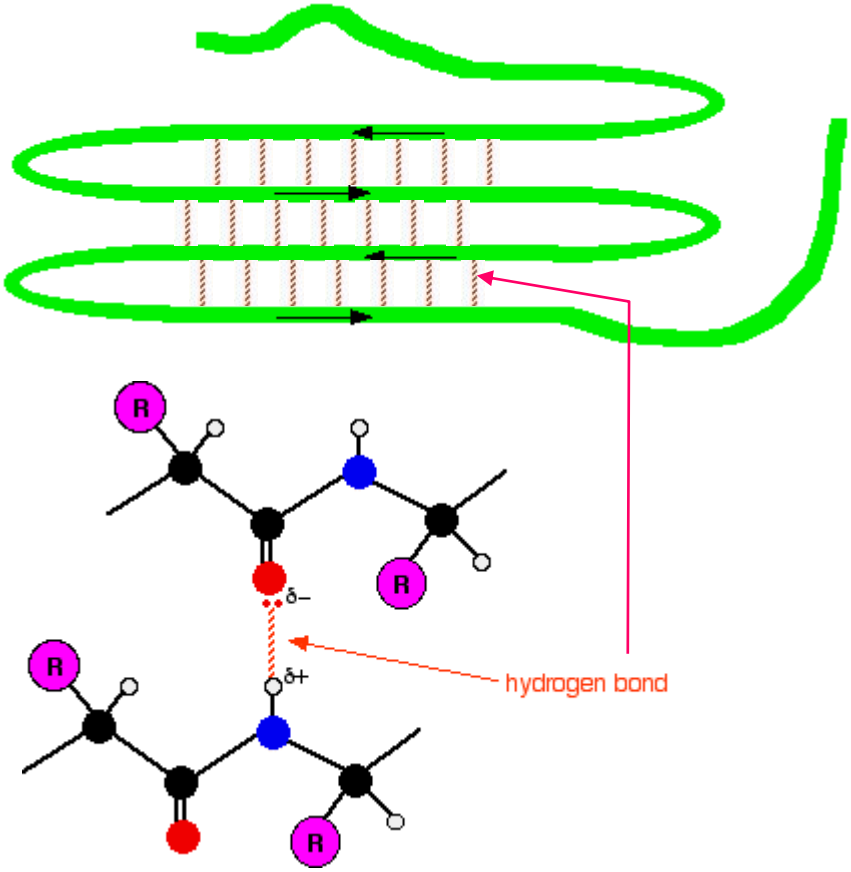


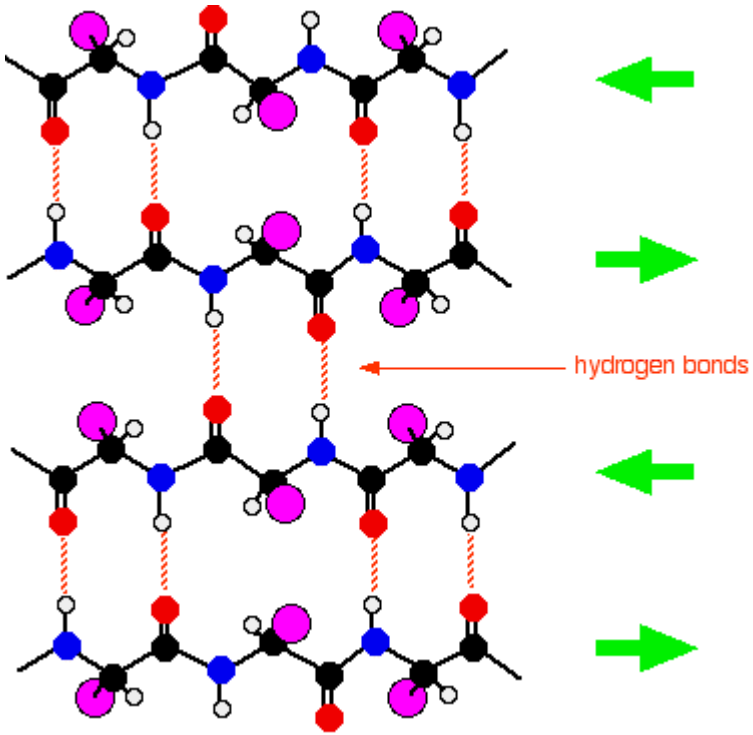


Alpha helix



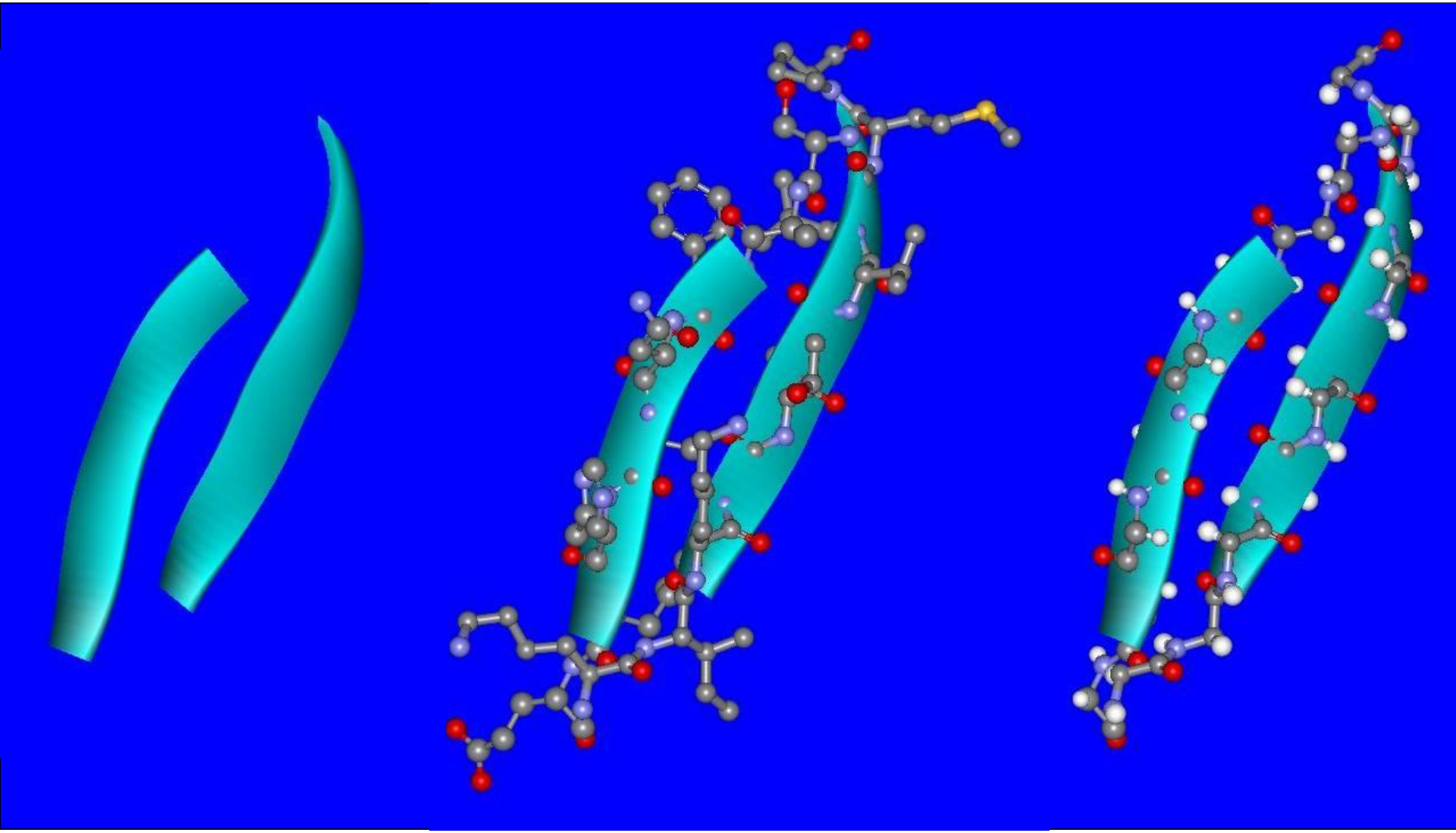
Beta sheet



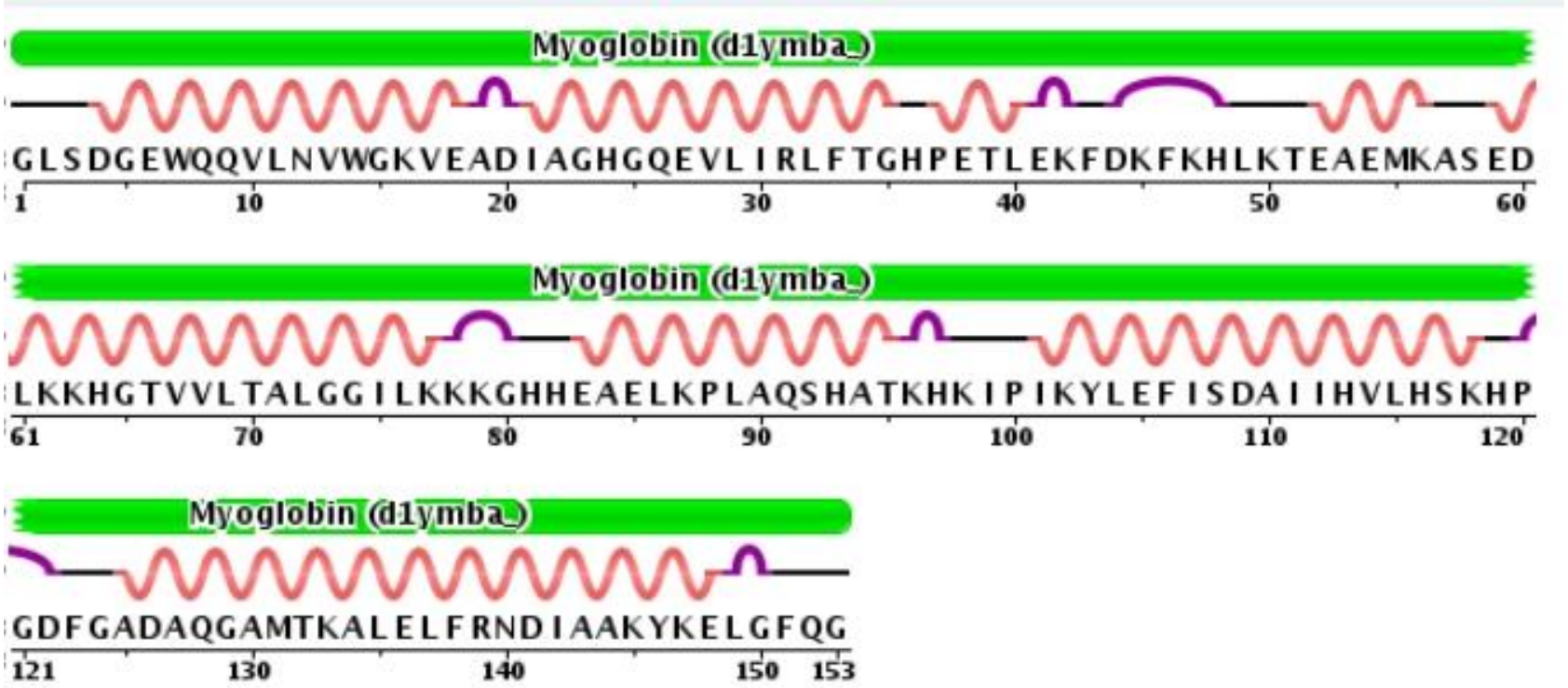


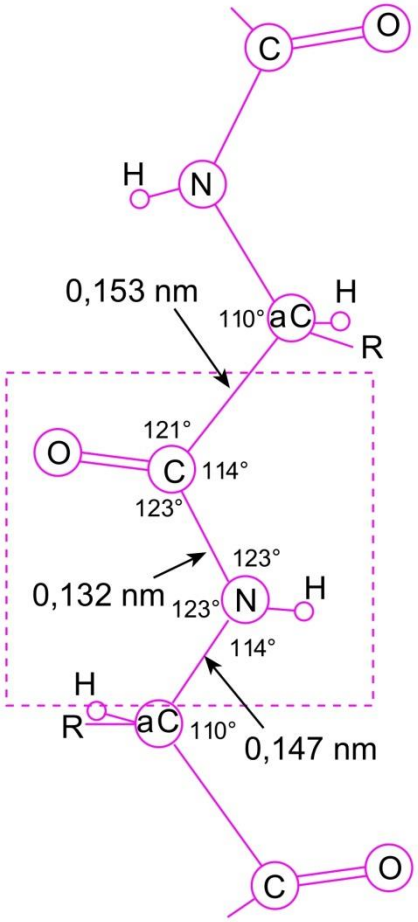
antiparallel

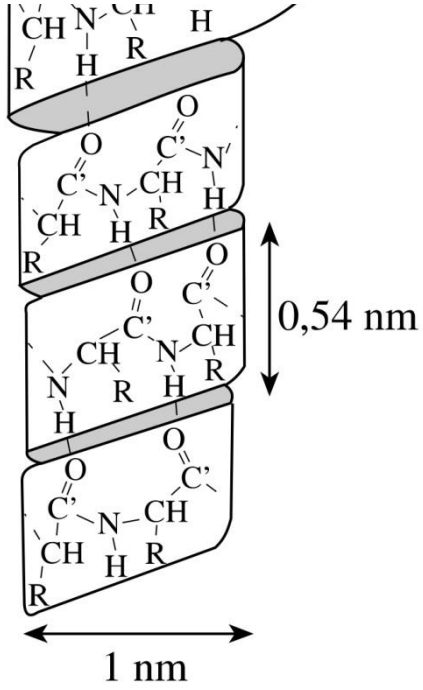
Beta sheet



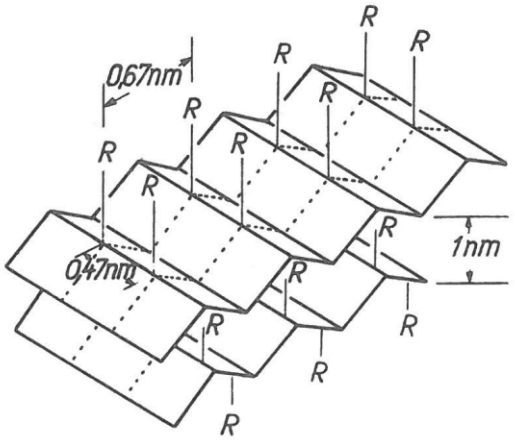
One dimensional representation of the secondary structure

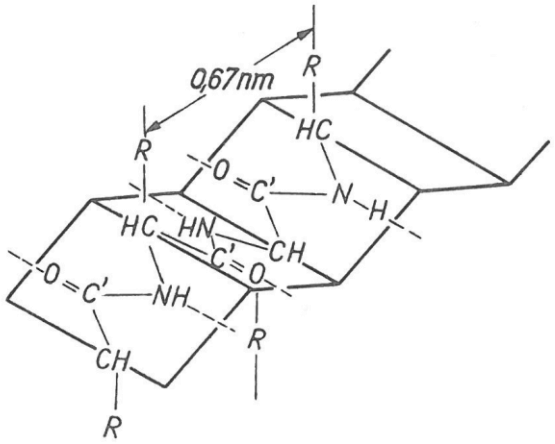






3,6 AA/turn
i -> i+4





Stabilization by H-bonds

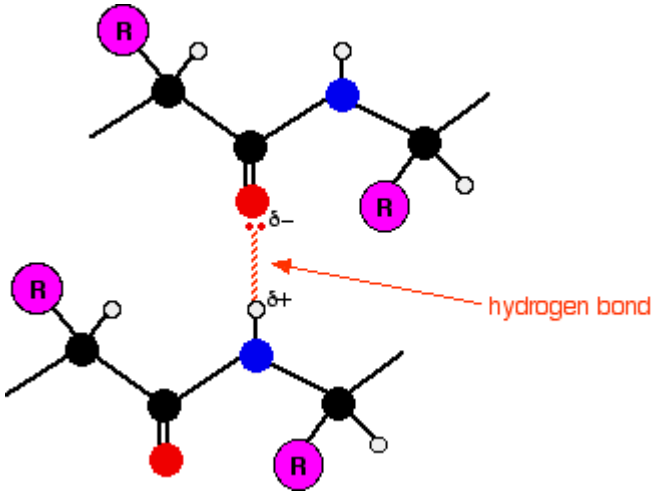
12-30 kJ/mol

Cf: Covalent bond: 200 kJ/mol

van der Wals:1-2 kJ/mol

thermal energy (RT):

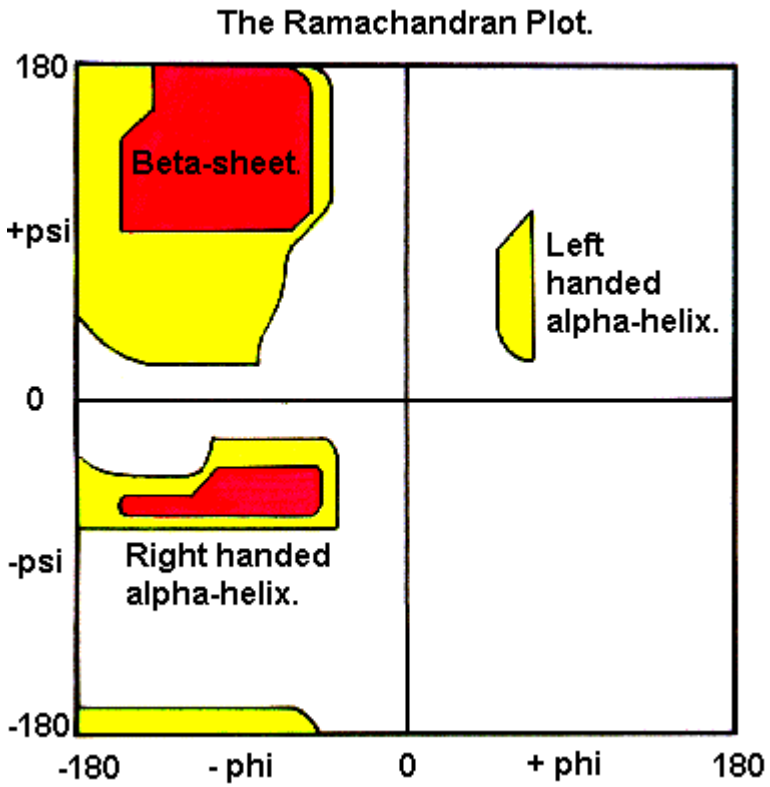
2.5 kJ/mol (T=300K)

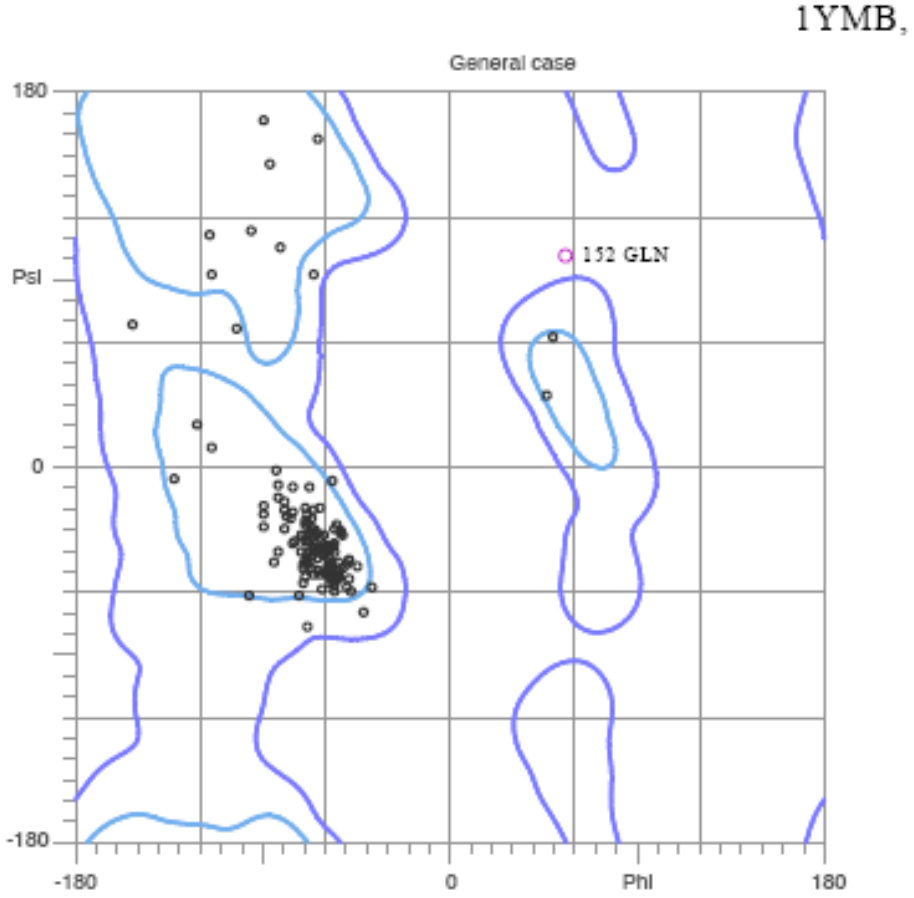


Boltzmann factor: $e^{-\frac{\Delta E}{RT}} = 0.000335 = \frac{1}{2981} \approx \frac{1}{3000}$

(ΔE=20kJ/mol)

Ramachandran plot





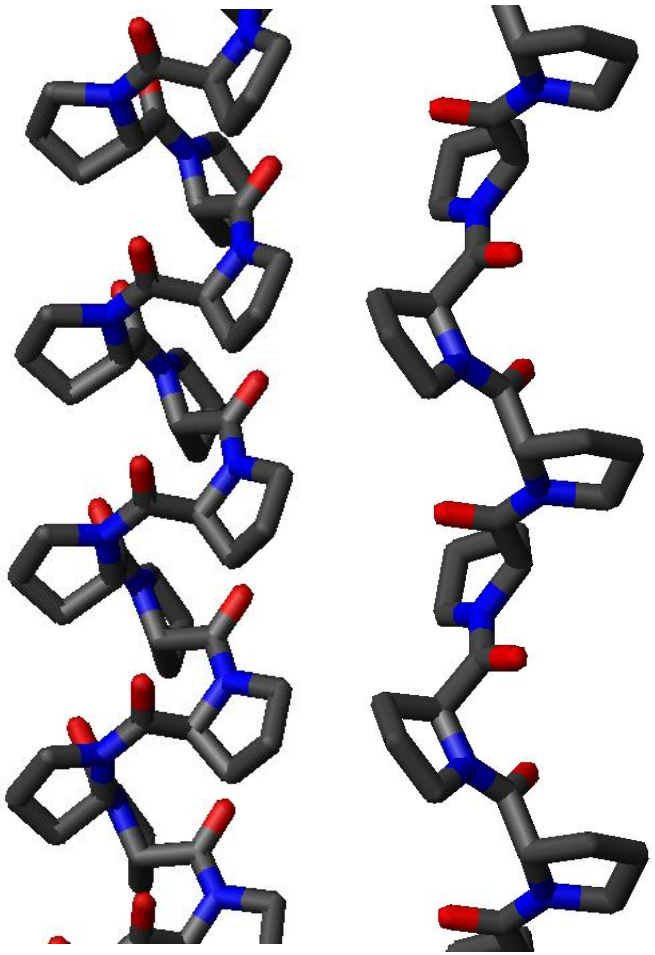
Special helices

3_{10}-helix* i -> i+3 (10 atom)

π-helix i -> i+5*

Polyproline I helix cis

Polyproline II helix** trans



Polyproline

I II

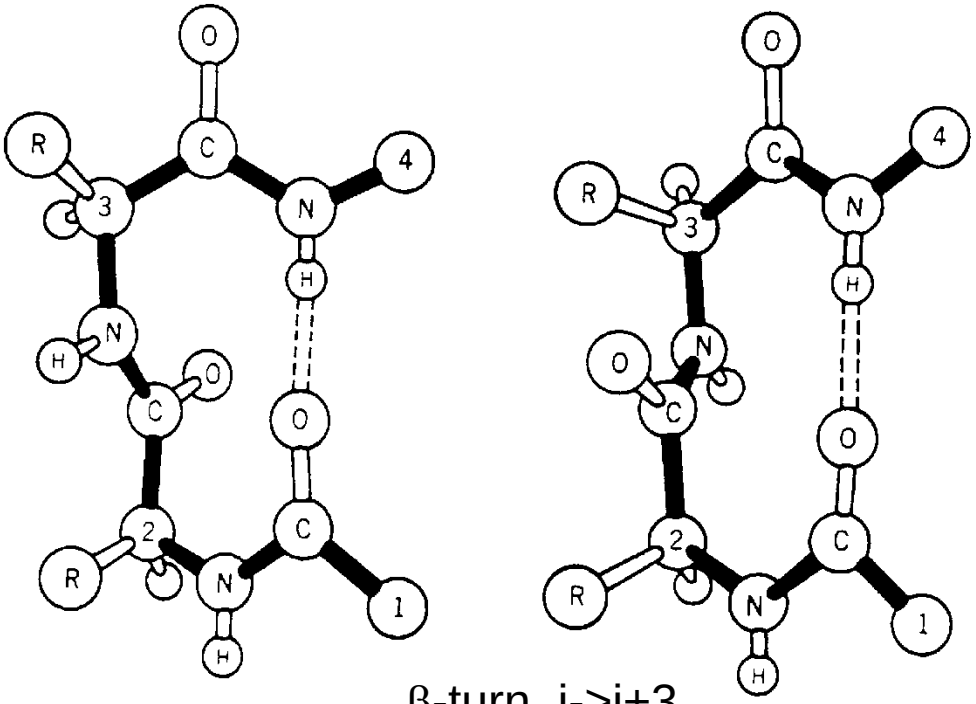
* α-helix: i->i+4 $3,6_{16}$ helix

** in water

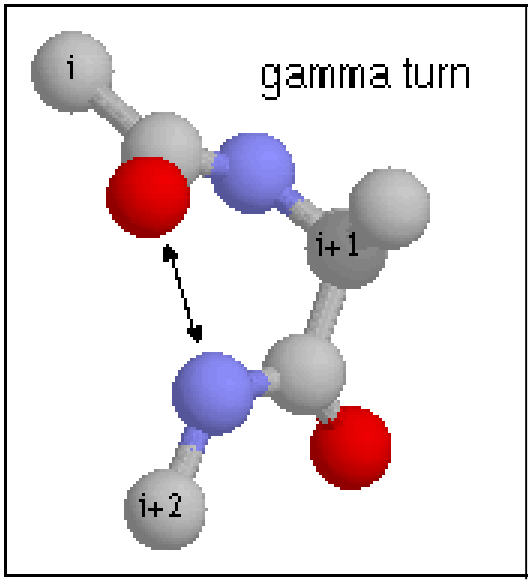
Other nonhelical structures

Loops and turns

(loop) (turn)



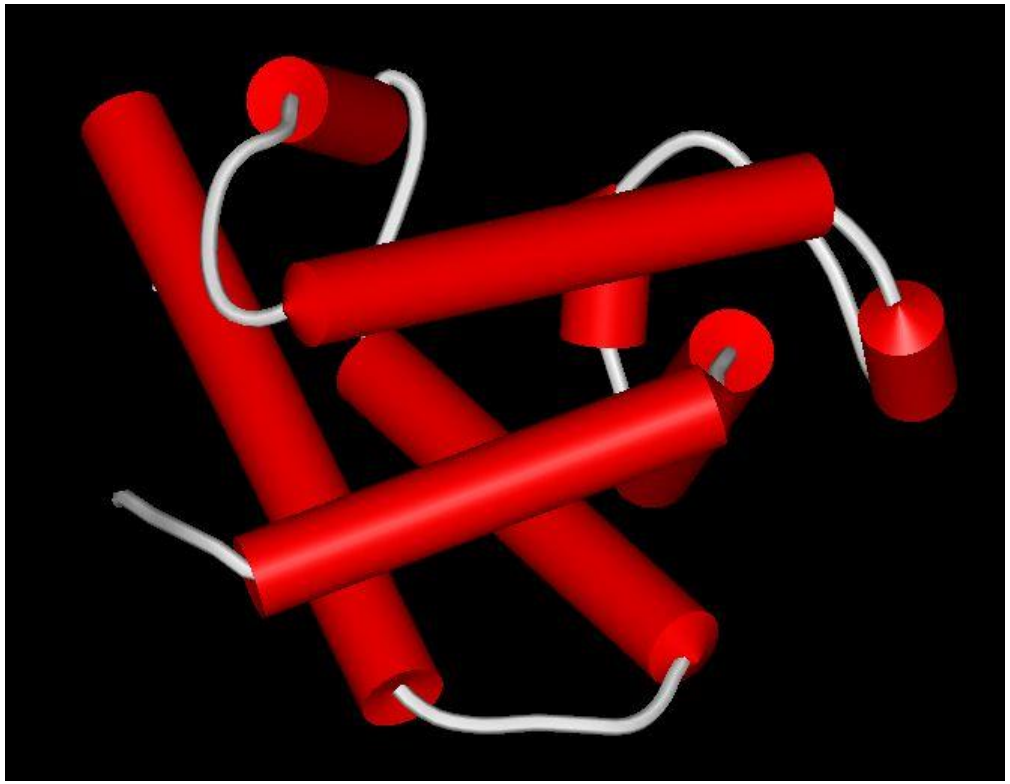
β-turn i->i+3



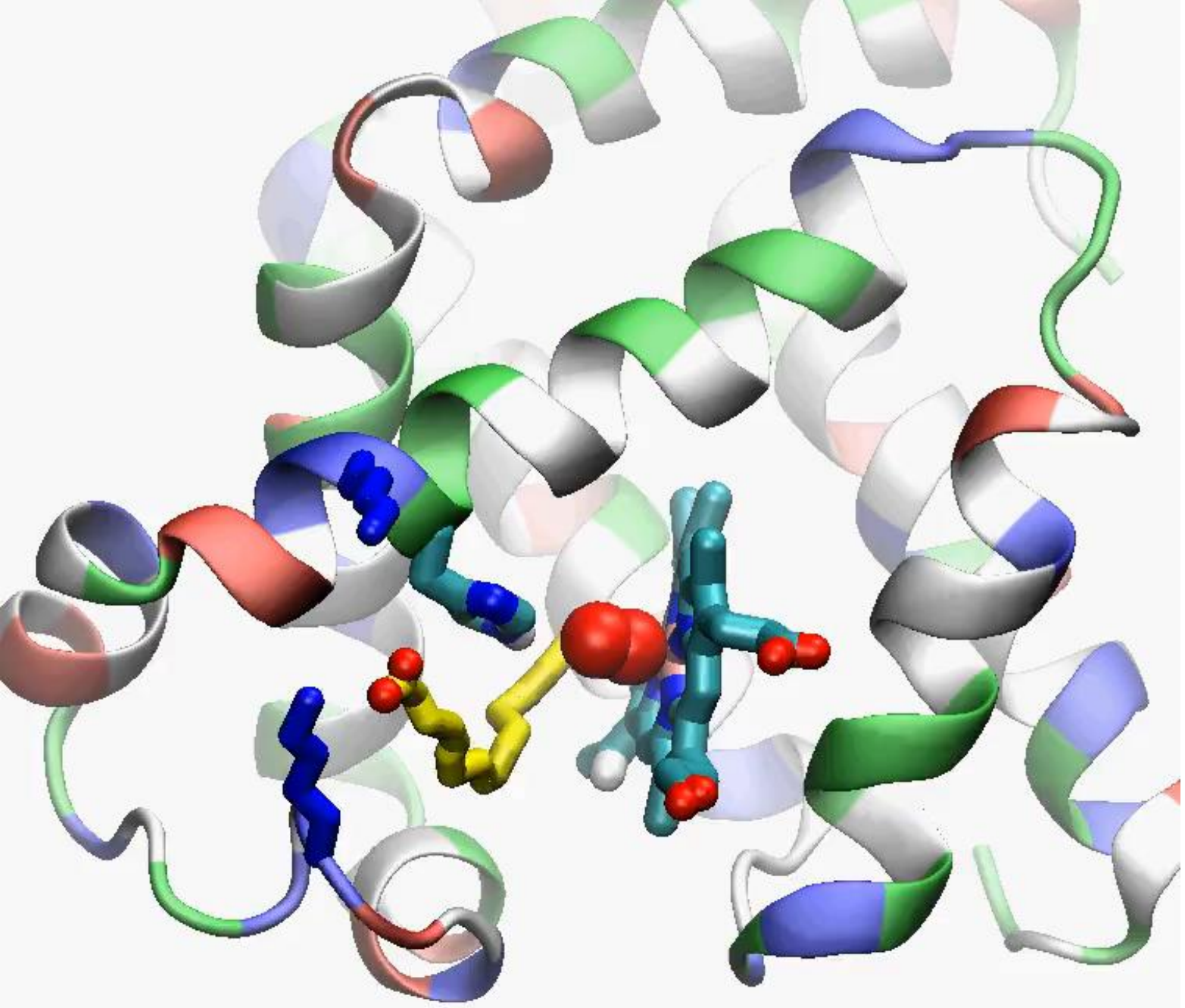
γ-turn i->i+2

Tertiary structure

Overall topology of the folded polypeptide chain (Organization of the secondary structure elements)



Myoglobin



Oxy-Myoglobin
+
Palmitic acid

100ns
simulation
slow-motion
video

Hierarchy of time-scales in motions

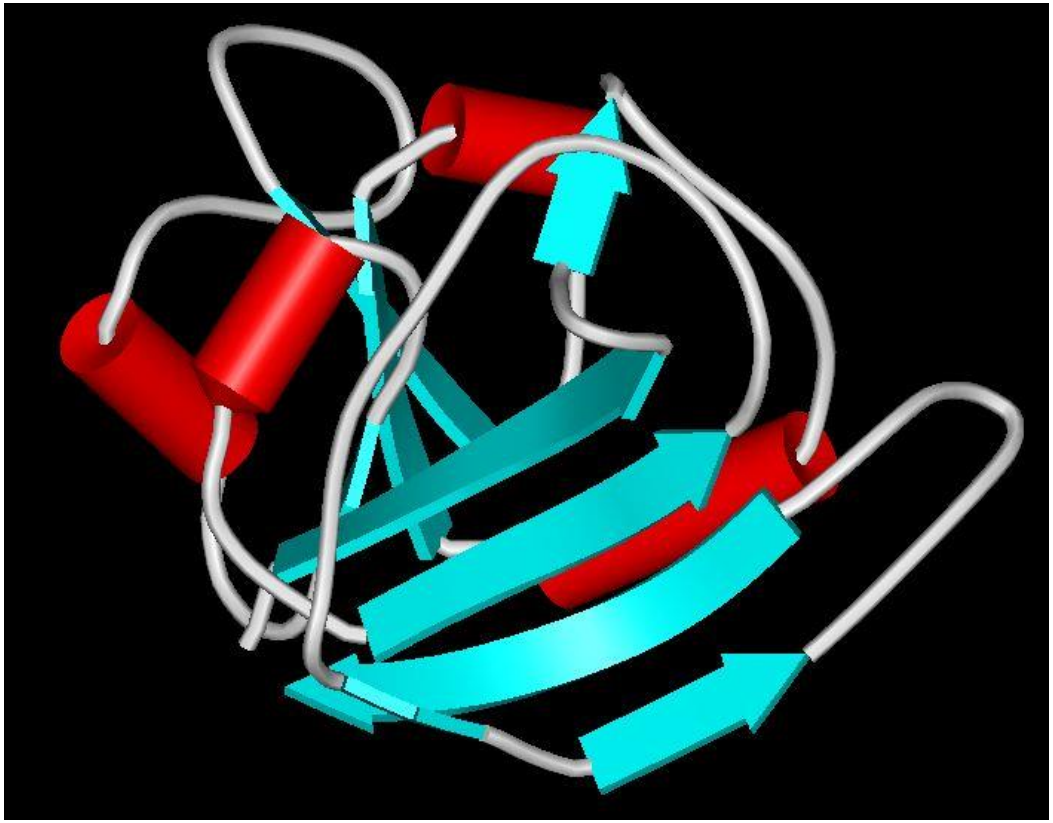
ps
ns
µs

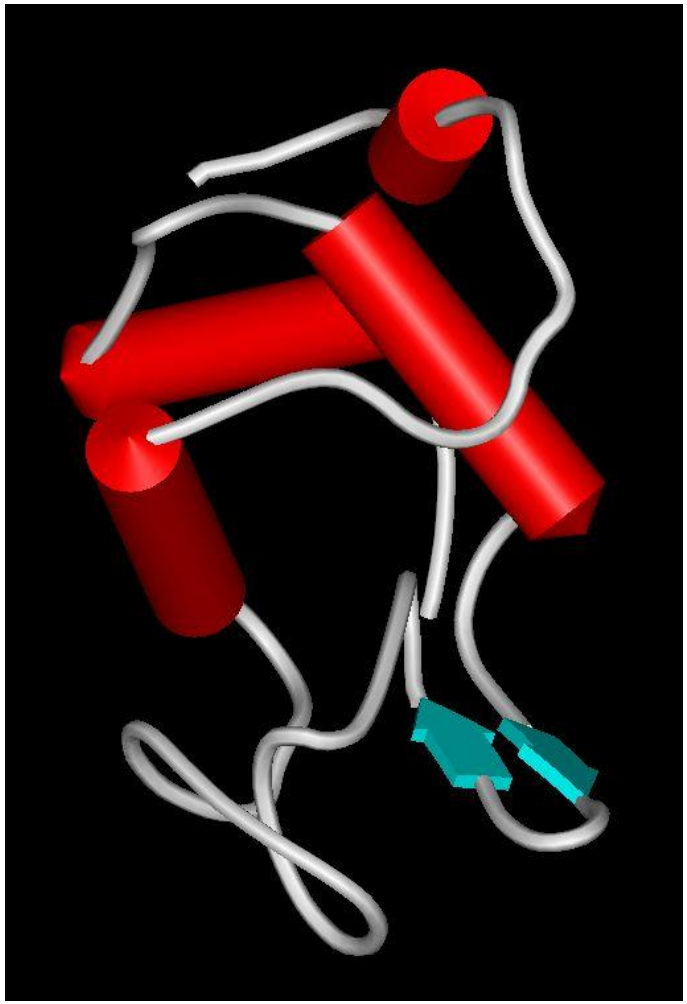
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128496#sec012

Examples

Lyzozime (HEW)

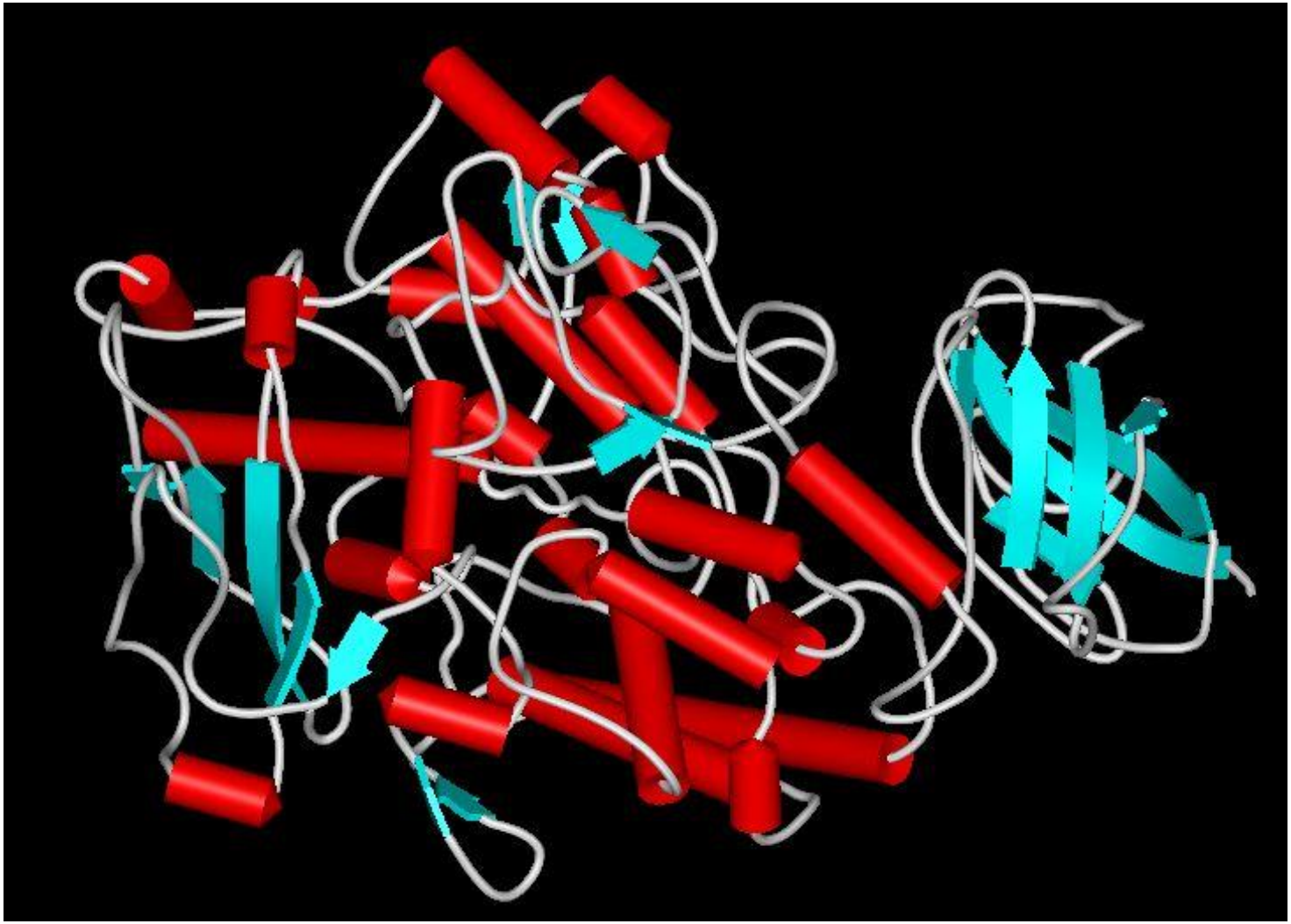
Dihydrofolate reductase



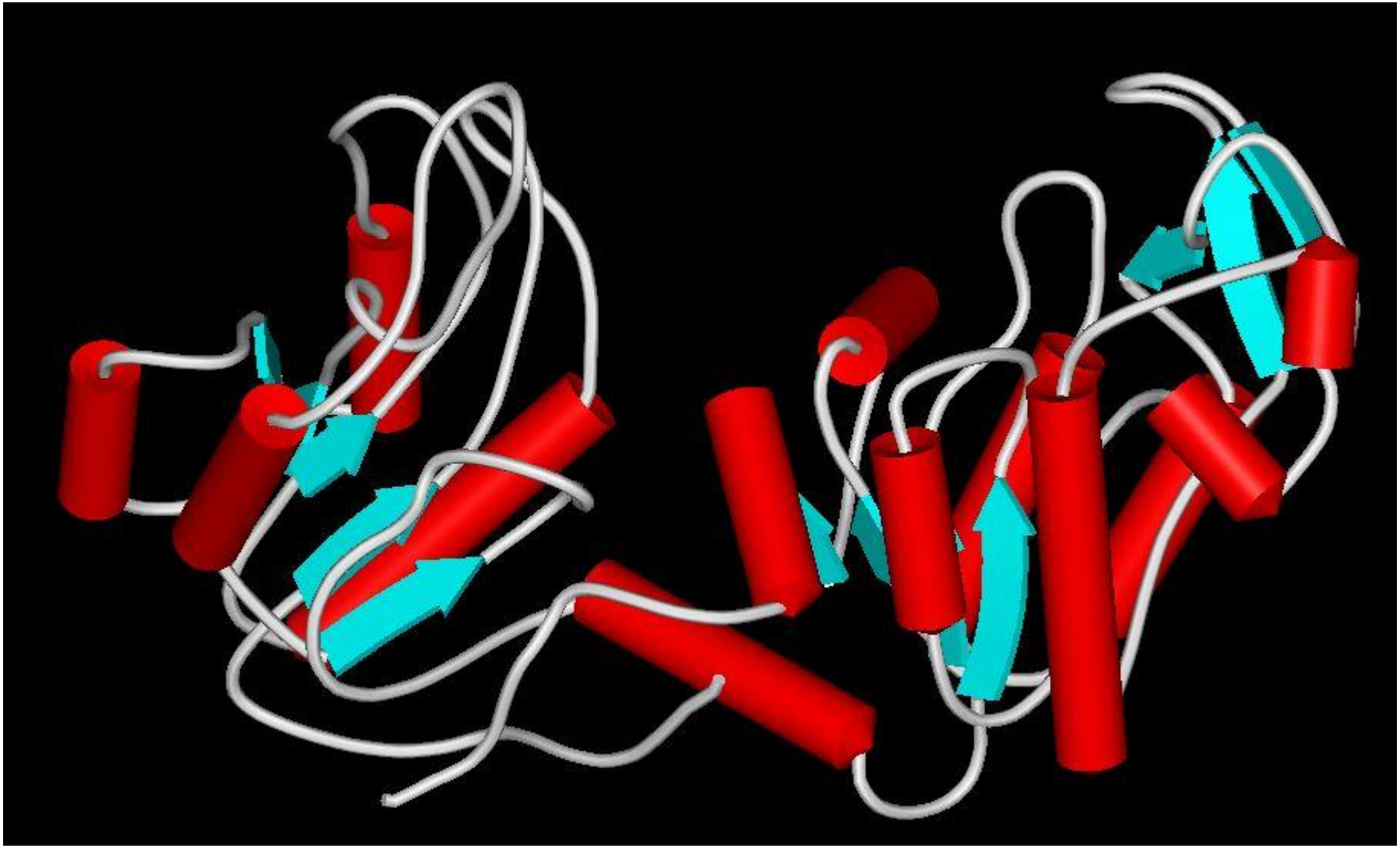


Examples

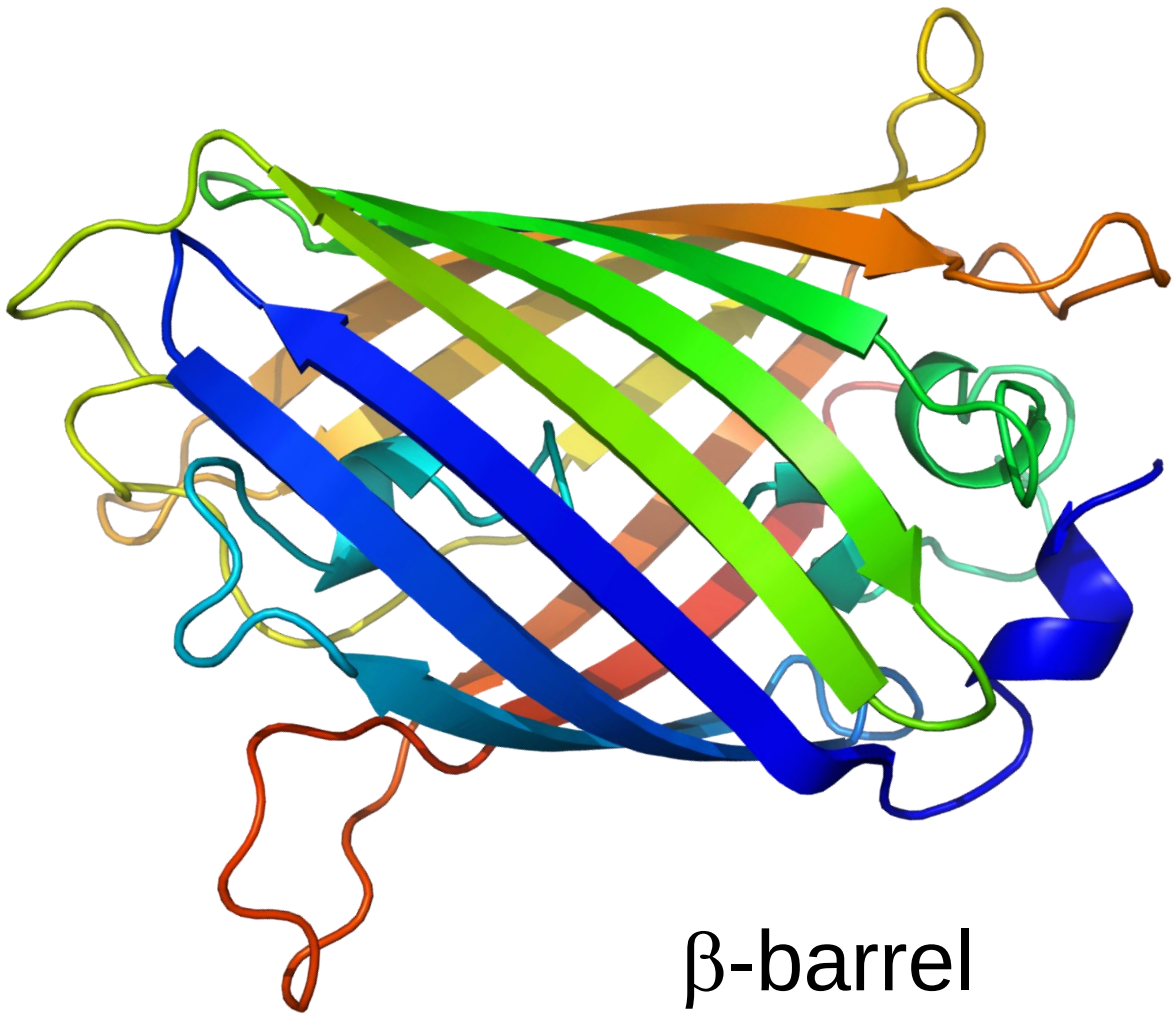
Lypoxigenase



Examples: Phosphogycerate-kinase (PGK)



Examples: GFP

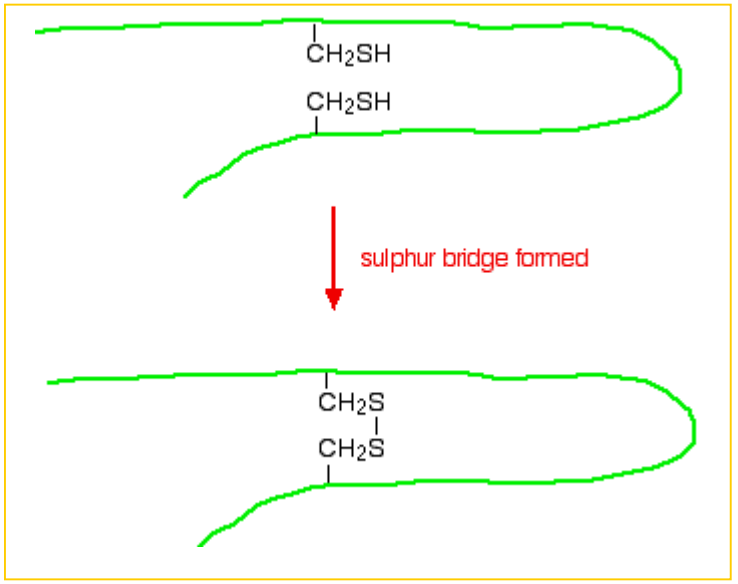


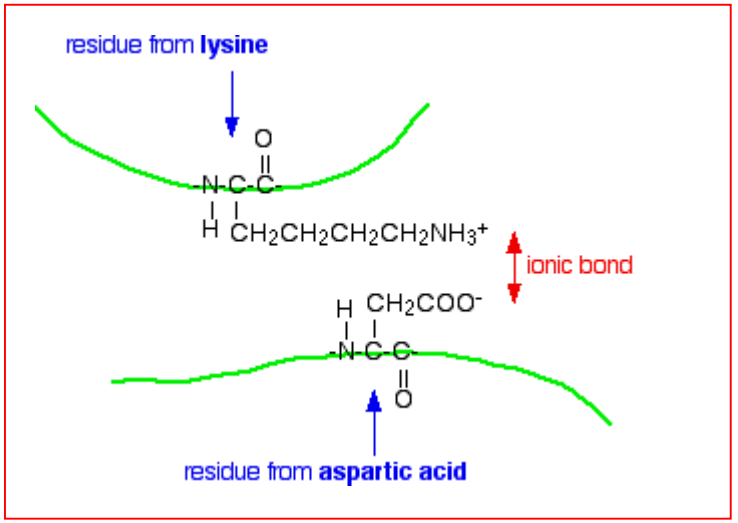
β-barrel

Stabilization of the tertiary structure

Between the side chains:

- disulfide bond
- ionic bonds
- H-bond
- Van der Waals int.

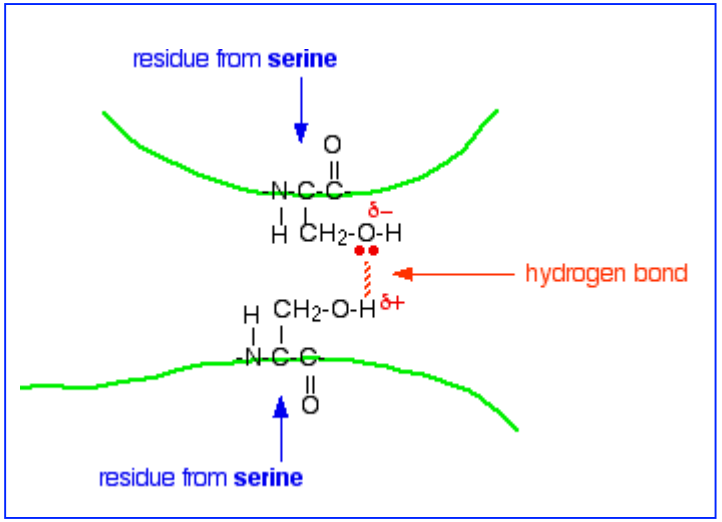


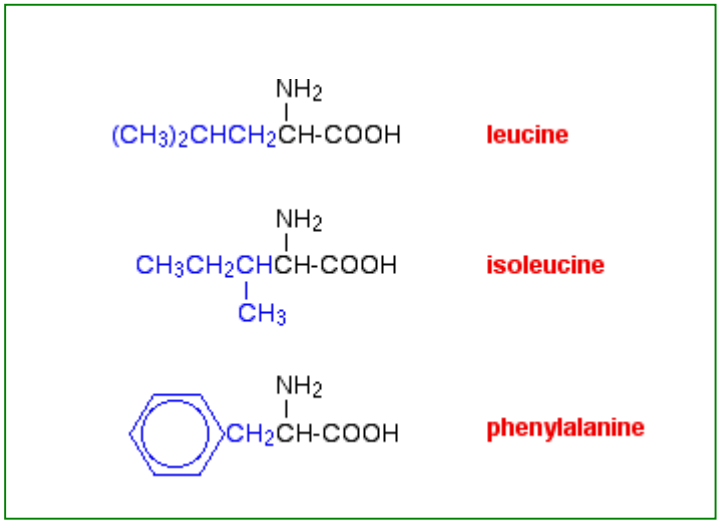


Stabilization of the tertiary structure

Between the side chains:

- disulfide bond
- ionic bonds
- H-bond
- Van der Waals int.



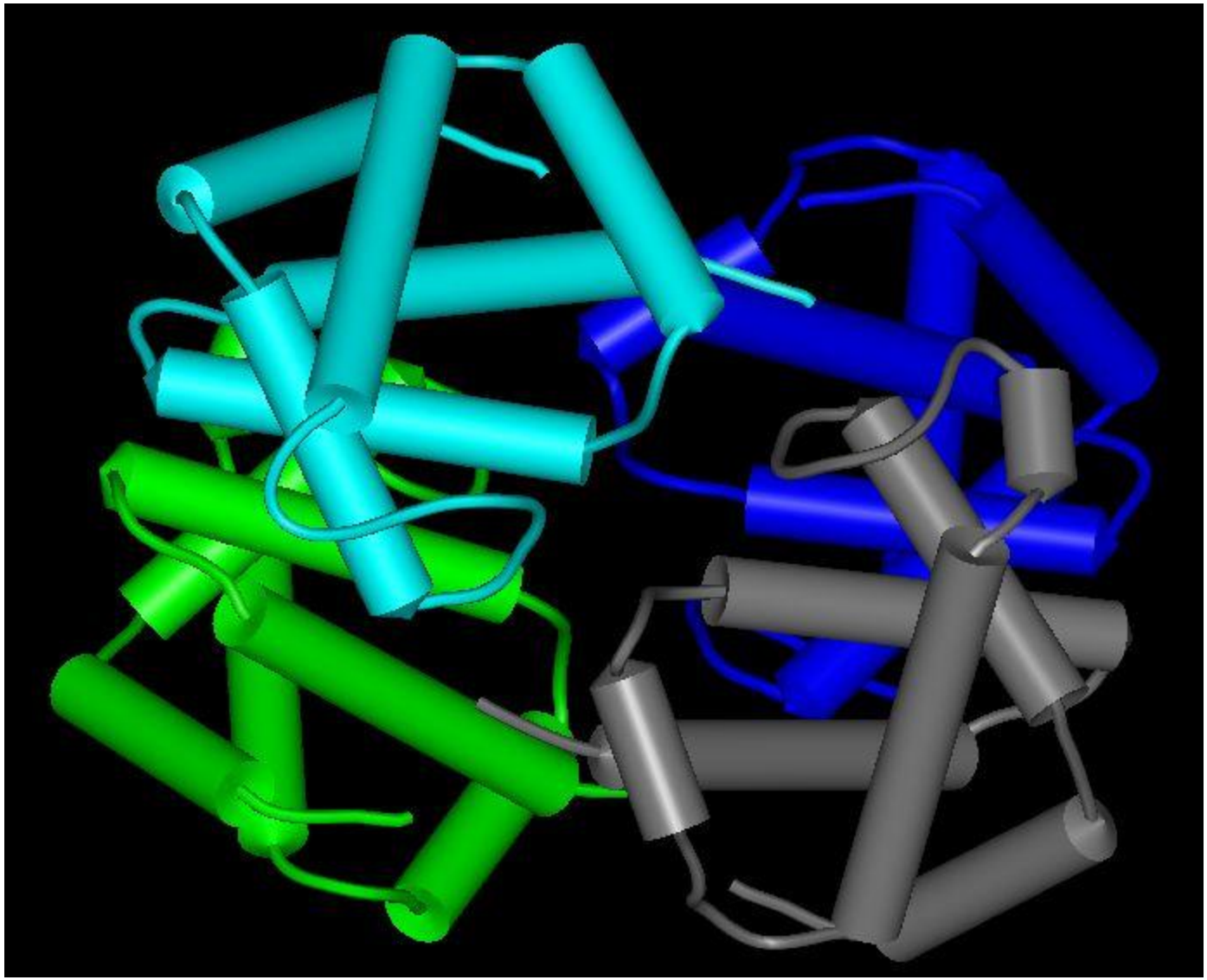


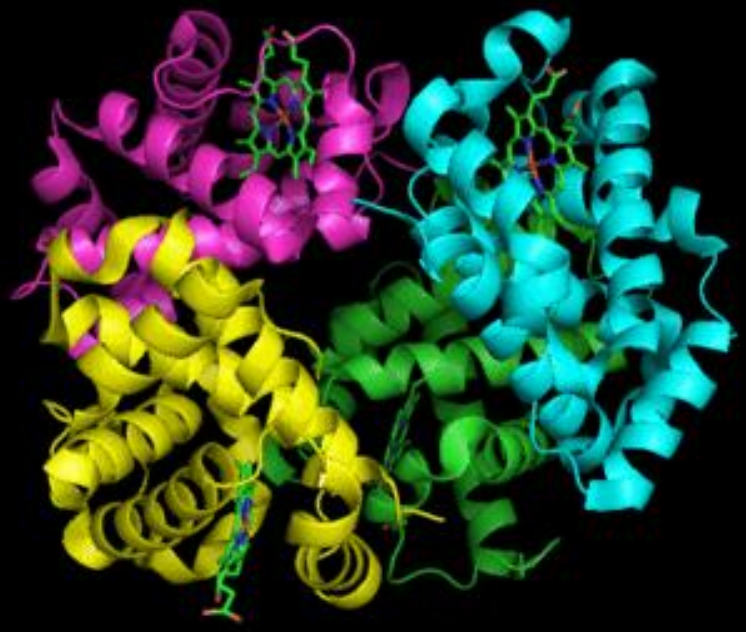
Quaternary structure

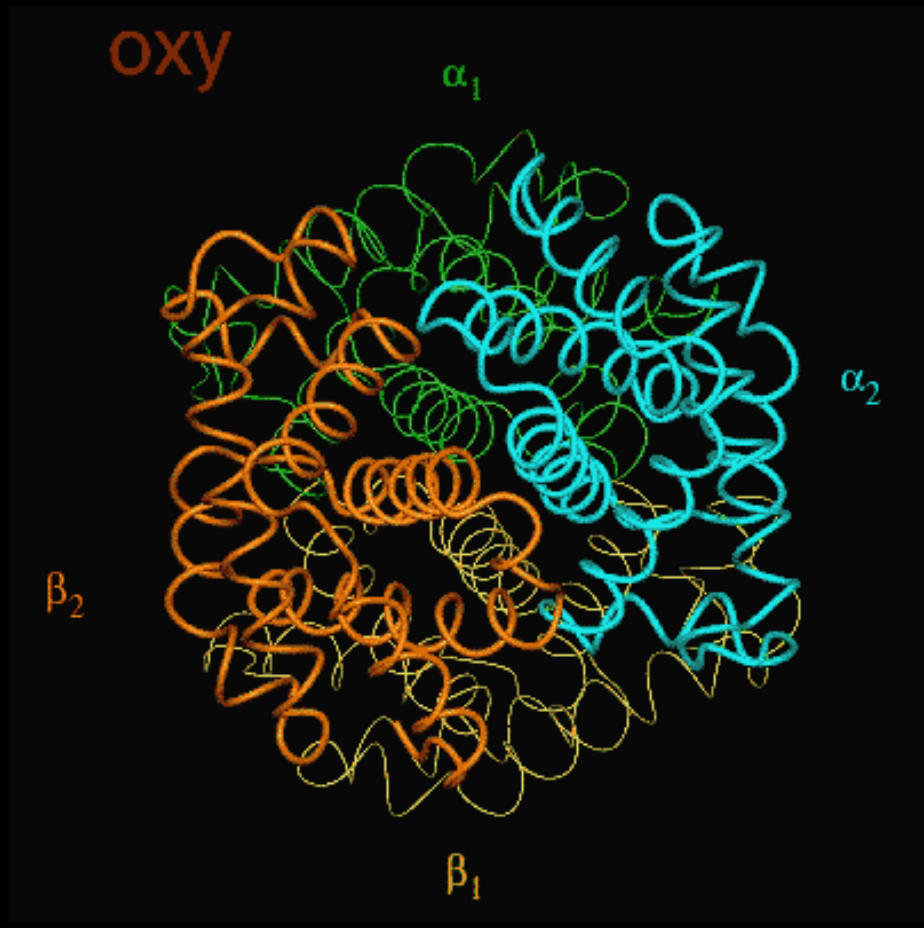
only for proteins with more than one polypeptide chains.

E.g.:

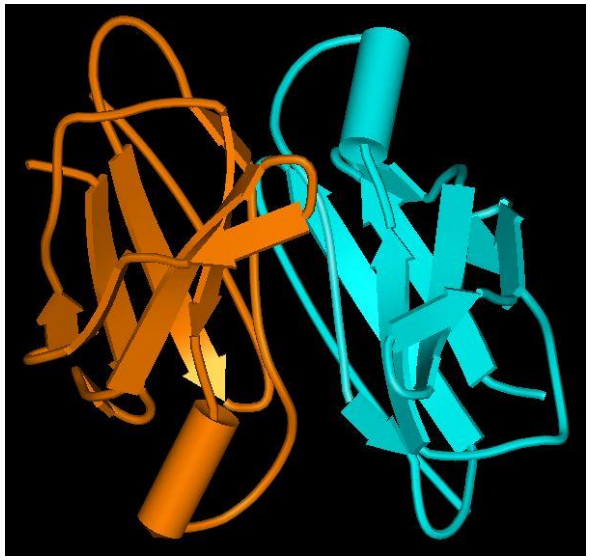
Hemoglobin tetramer



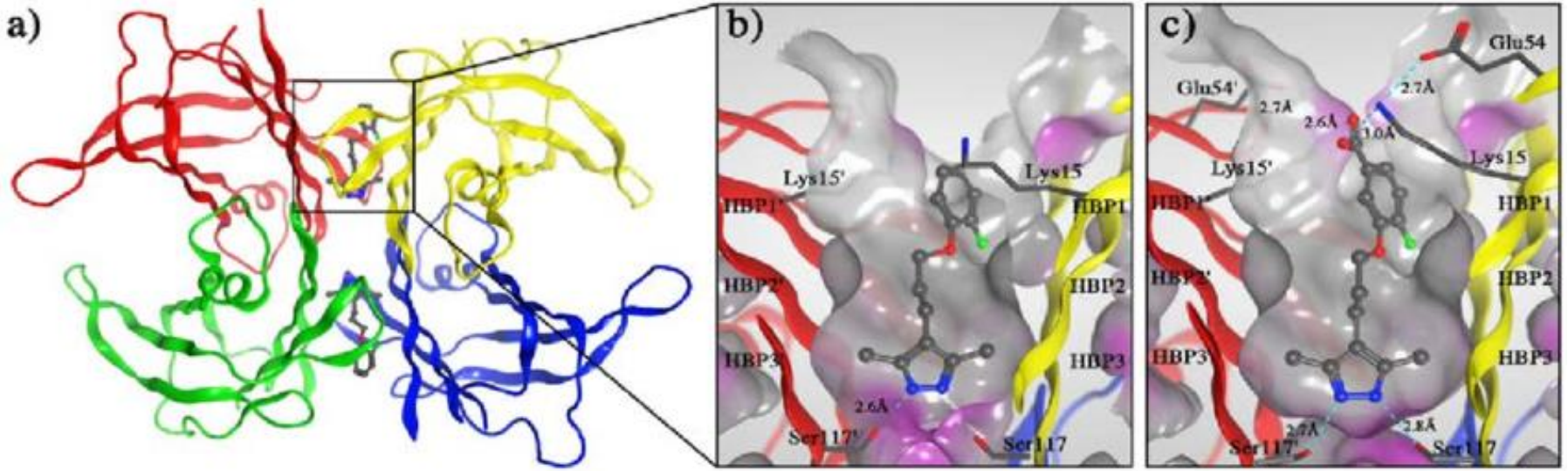



oxy
α_1
α_2
β_2
β_1

Further examples: Transthyretin

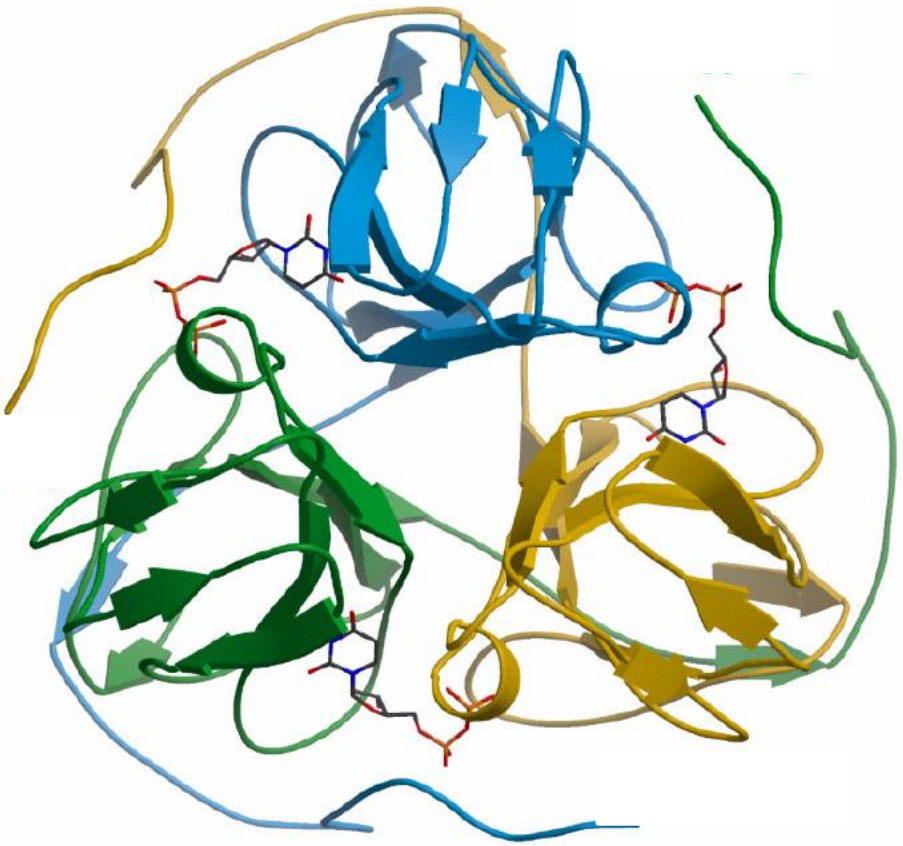


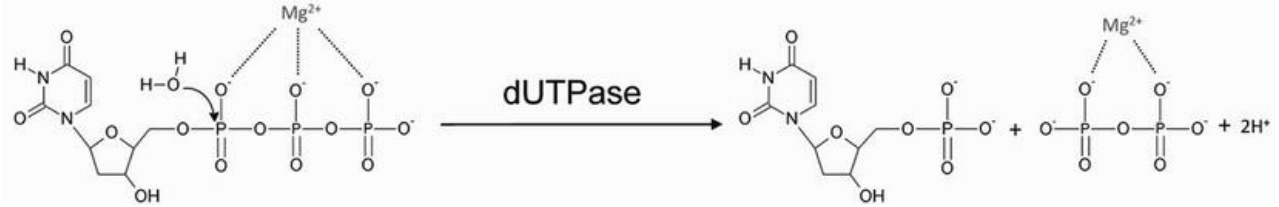
four binding sites;
two for thyroxine and two
for retinol-RBP complex



Further examples: DUTPase

3 subunits





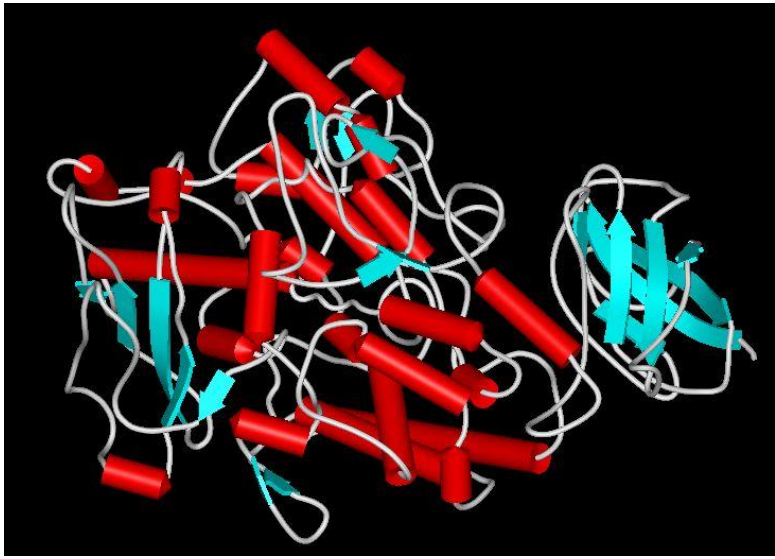
from: http://www.enzim.hu/~vertessy/kovari_phd.pdf

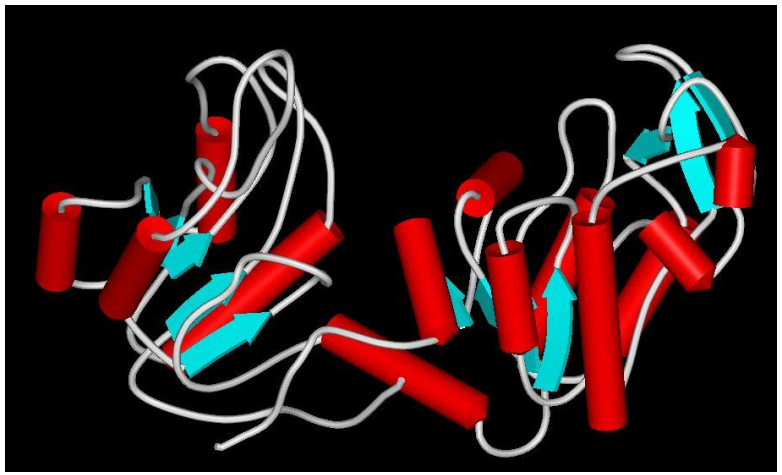
Important further aspects of the protein structure

- Domain
- Prosthetic group
- Posttranslational modifications
- Active site
- Pocket

Domain

Part of the protein, which can fold into an ordered structure. Its structure is stable, it can function without the presence of the rest of the protein. The different domains of a protein may have different functions: e.g.: ATP binding domain, etc.

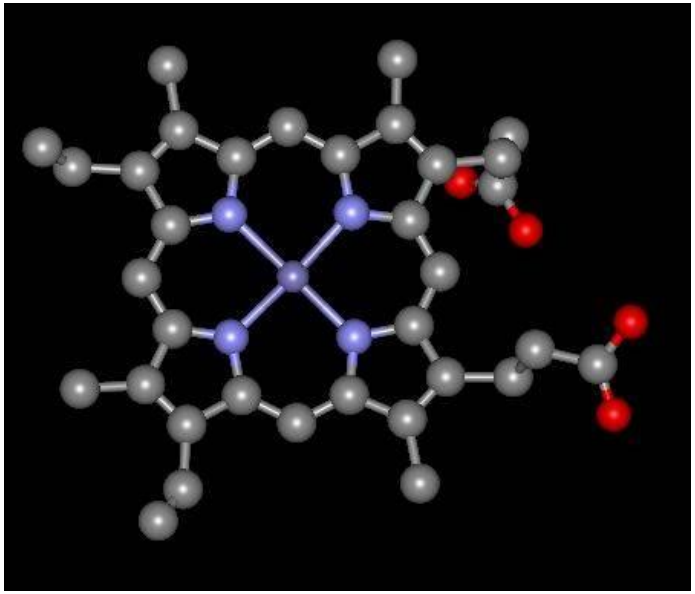


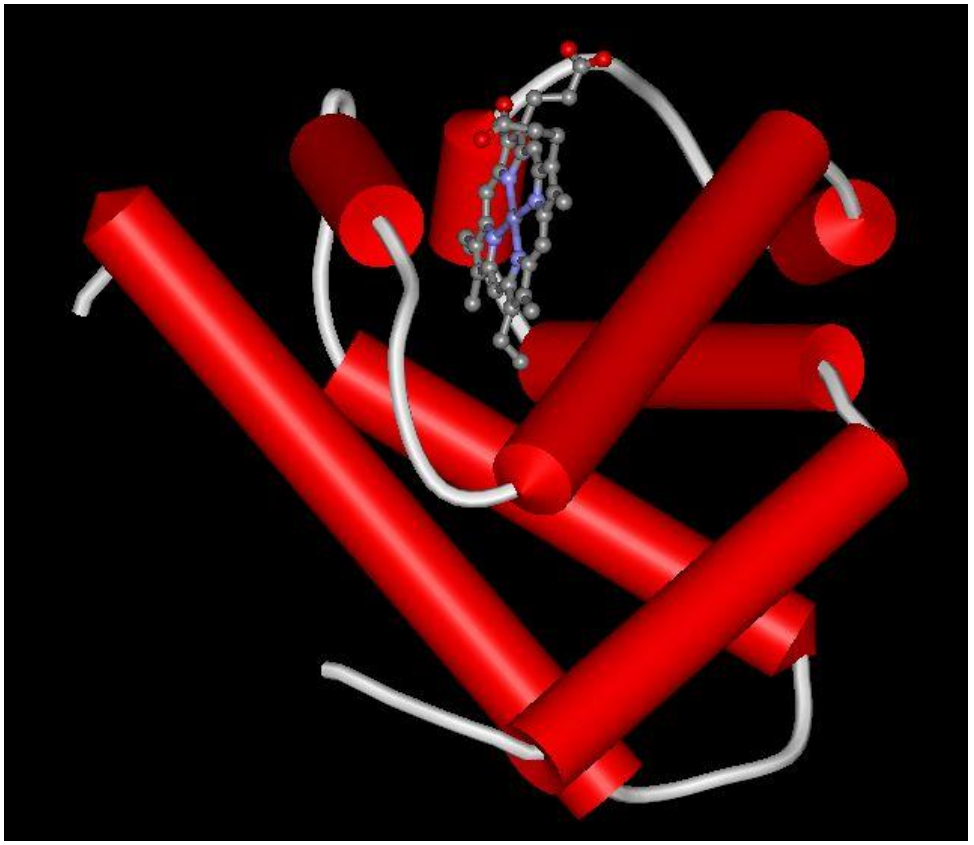


Prosthetic group

A non-protein chemical compound that is required for an enzyme's activity. They are bound strongly to the protein.

E.g.: hem group

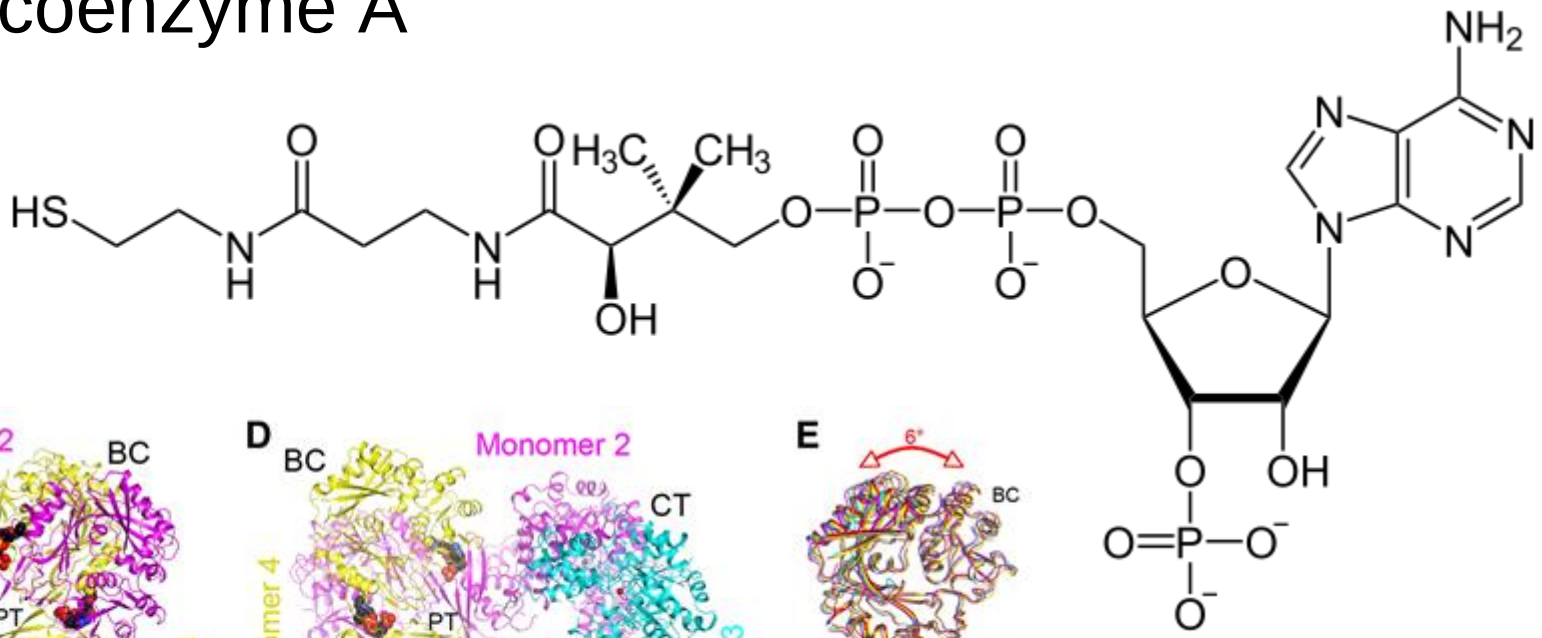


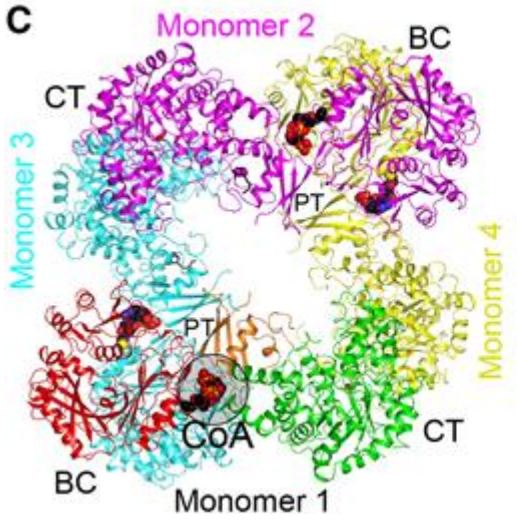


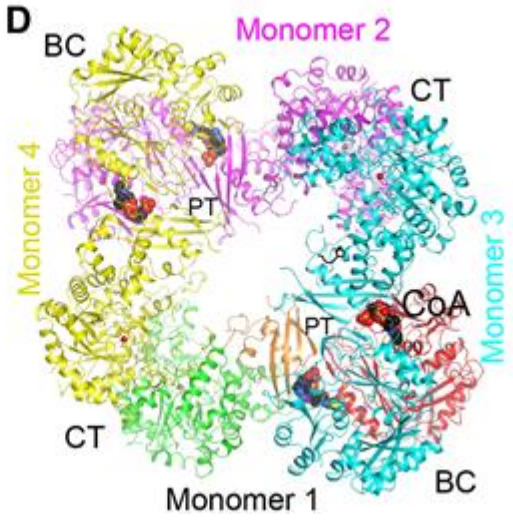
Coenzymes

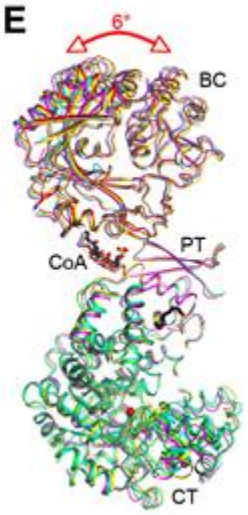
A cofactor is a non-protein chemical compound that is required for an enzyme's activity. They bind weakly and reversible.

example: coenzyme A





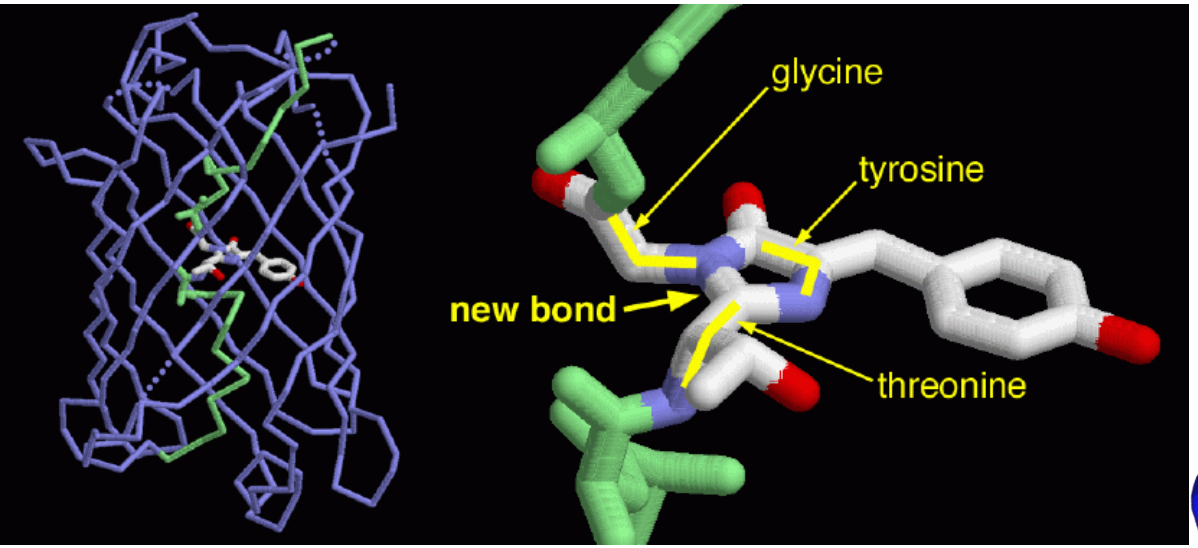


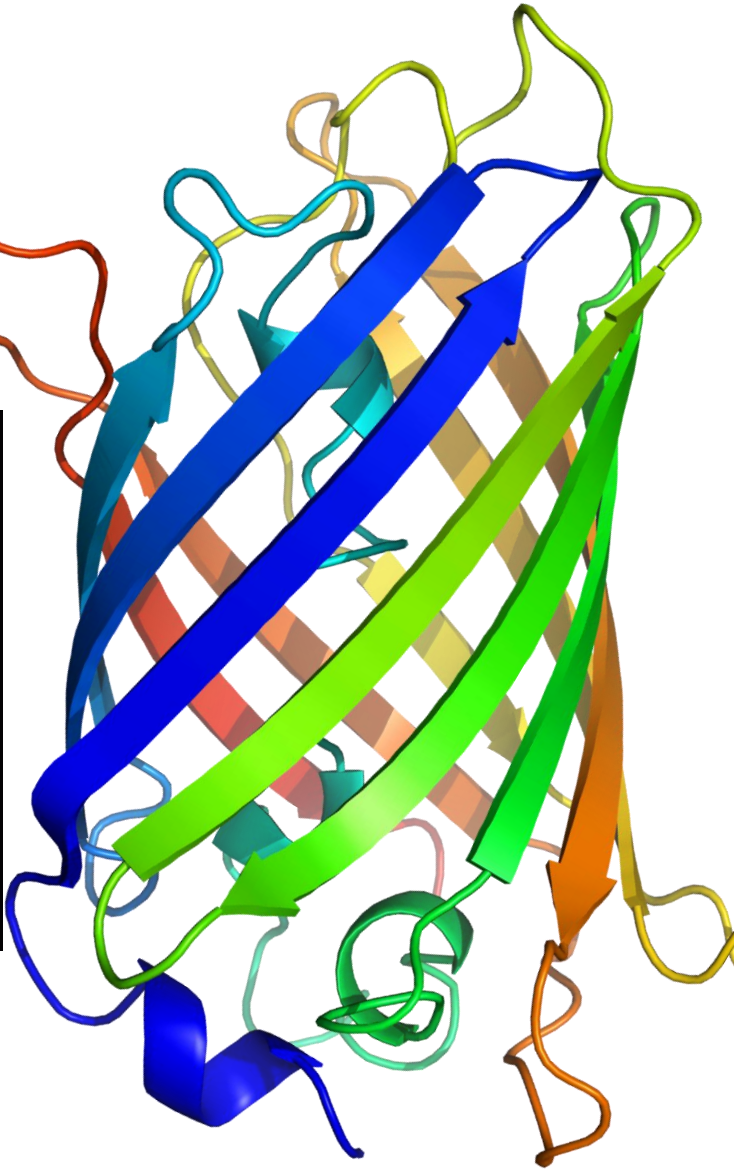


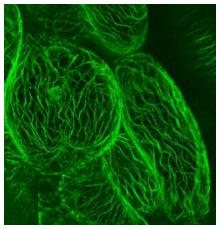
Symmetrical Tetramer for S. aureus Pyruvate Carboxylase in Complex with Coenzyme-A Structure , Vol.17 Issue 6 Pages 823-832 (2009)

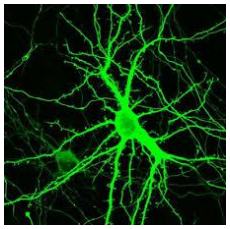
Posttranslational modifications

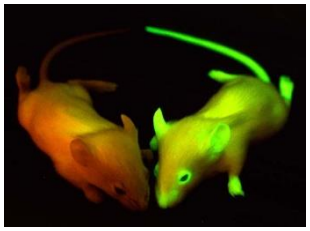
E.g.: formation of the chromophore in GFP









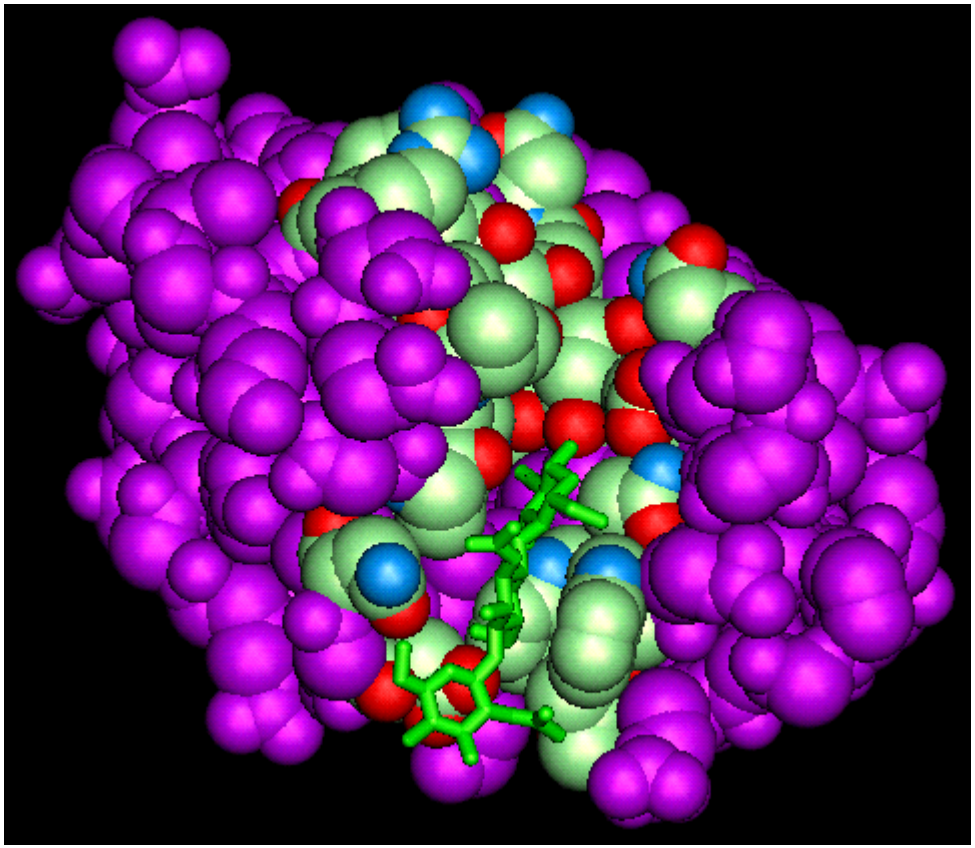


Active site

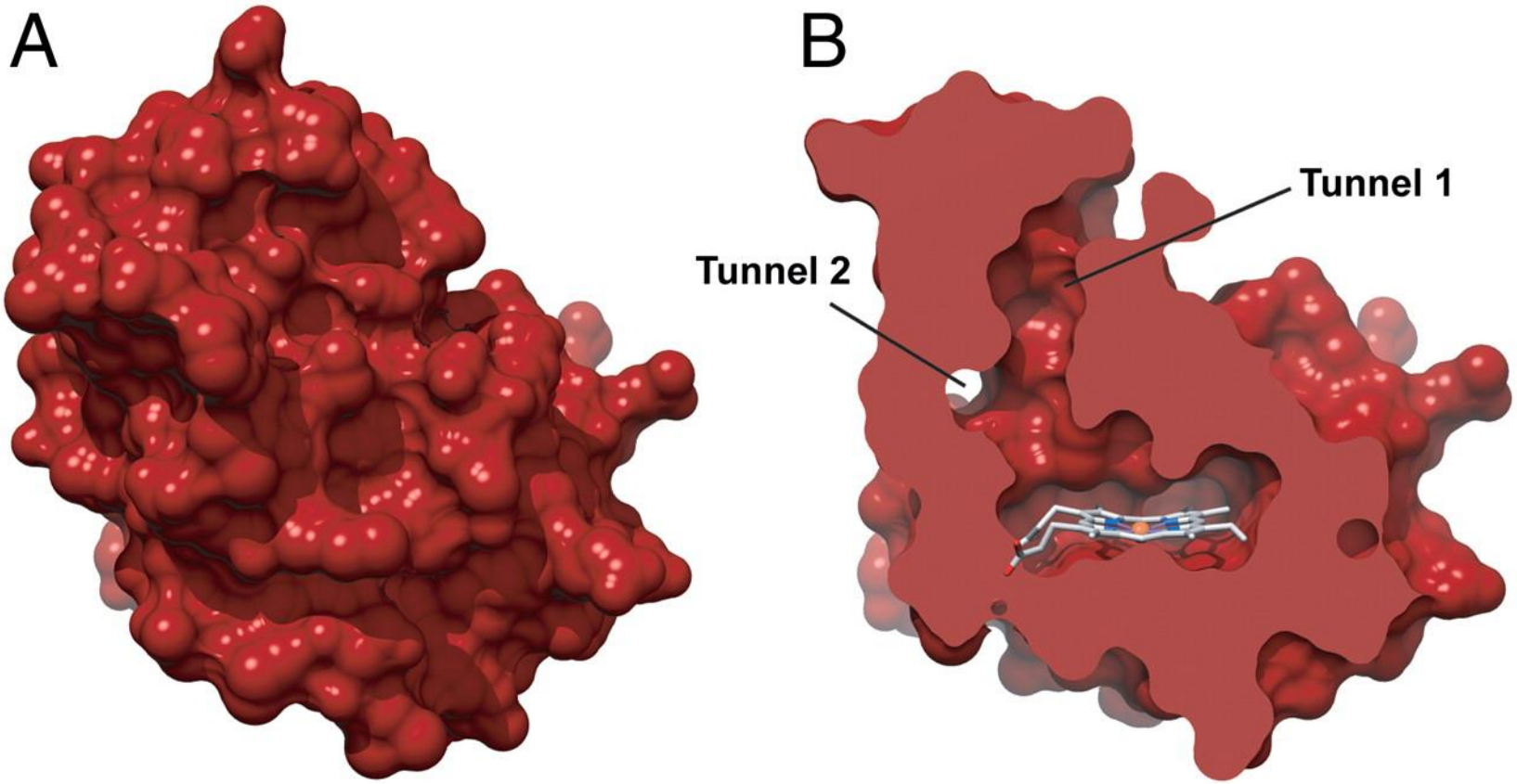
Active site is the region of an enzyme where substrate molecules bind and undergo a chemical reaction

Binding site

Catalytic site



Hem pocket



hem nitric oxide/oxygen binding (H-NOX) domain

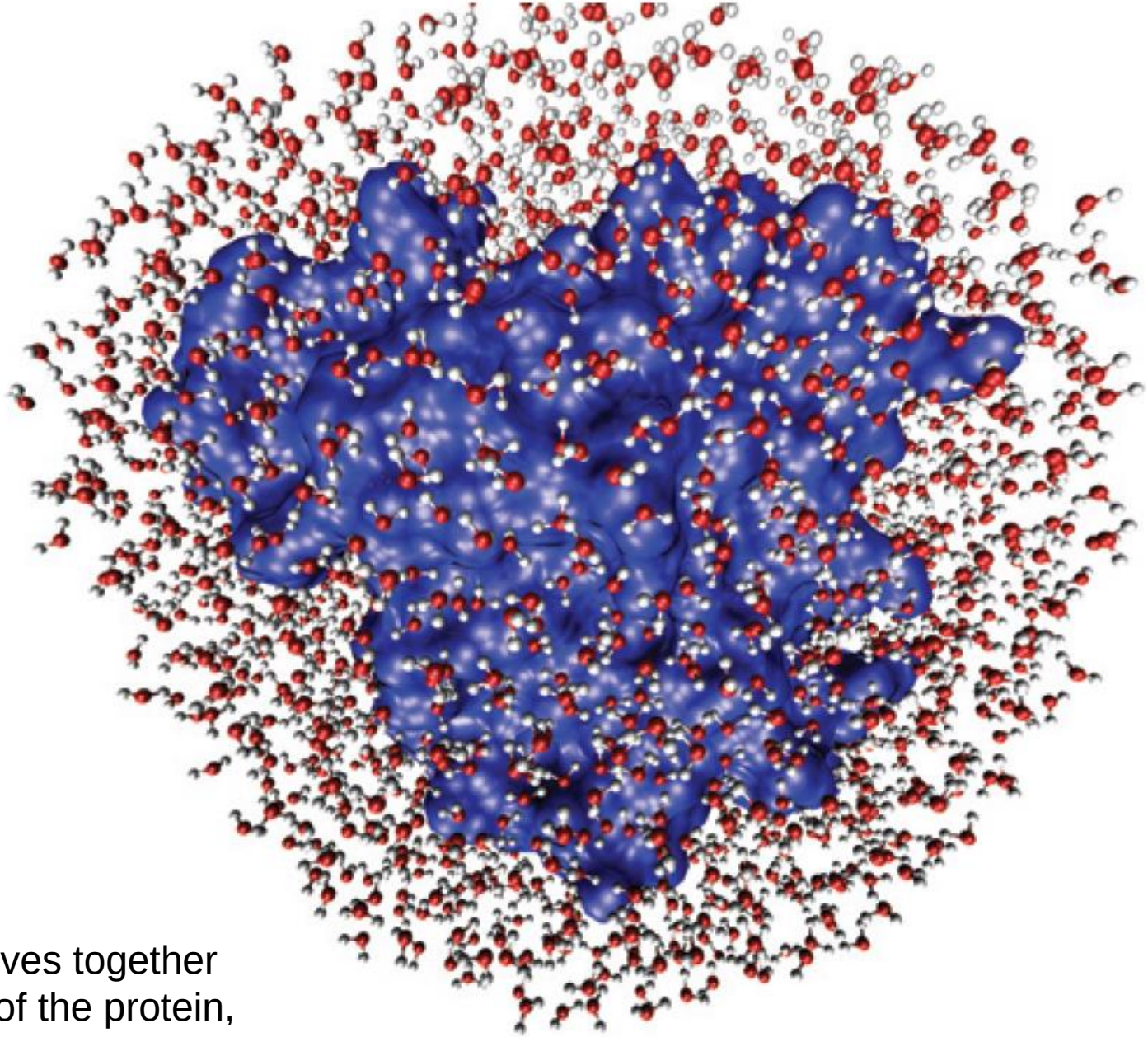
Winter M B et al. PNAS 2011;108:E881-E889

©2011 by National Academy of Sciences

Role of the water

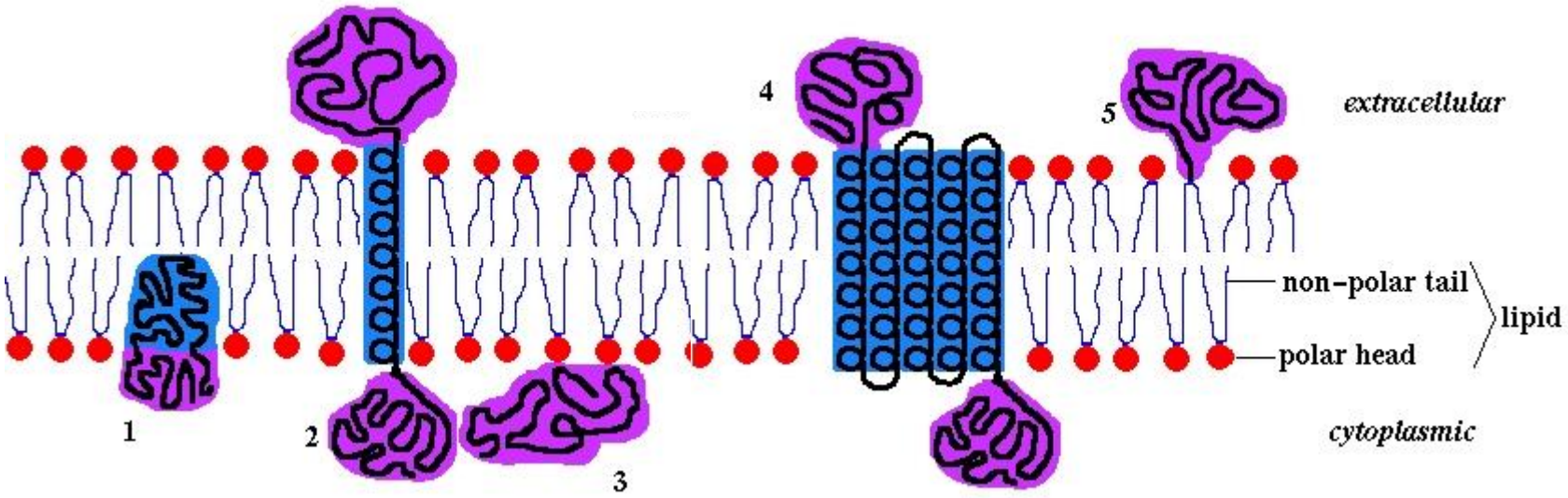
Hydration layer:

2-3 Water molecules



This somewhat moves together with the dynamics of the protein, couples to the solvent

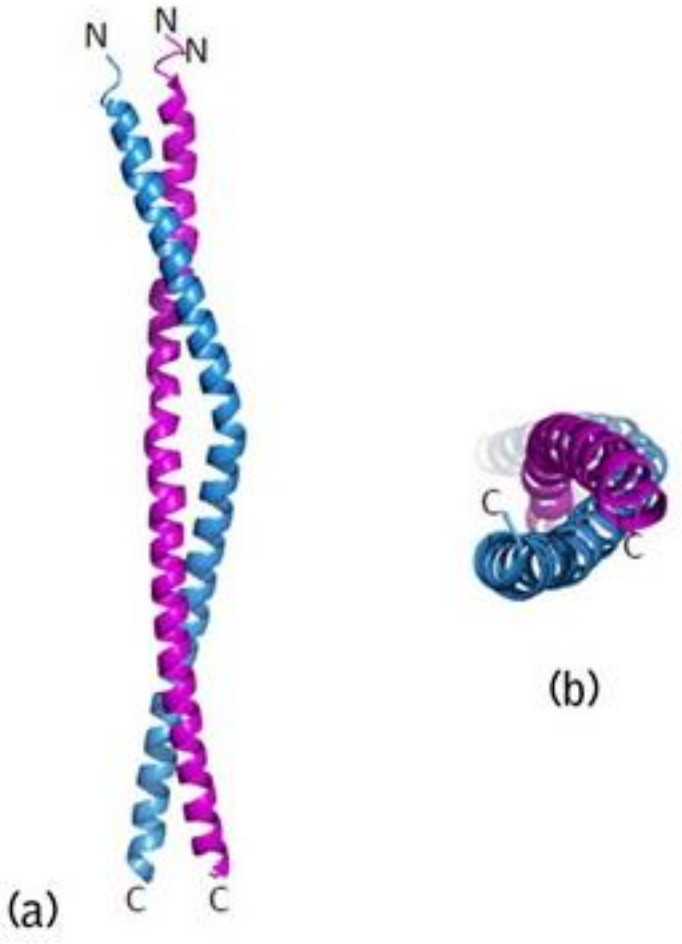
Membrane proteins

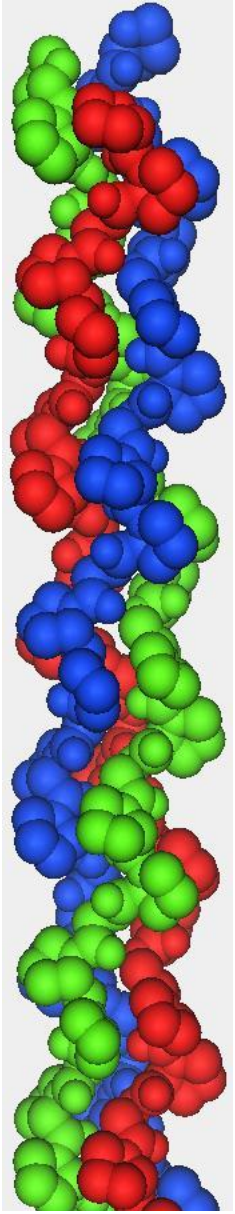


Domains with **hydrophobic** **hydrophilic** surfaces

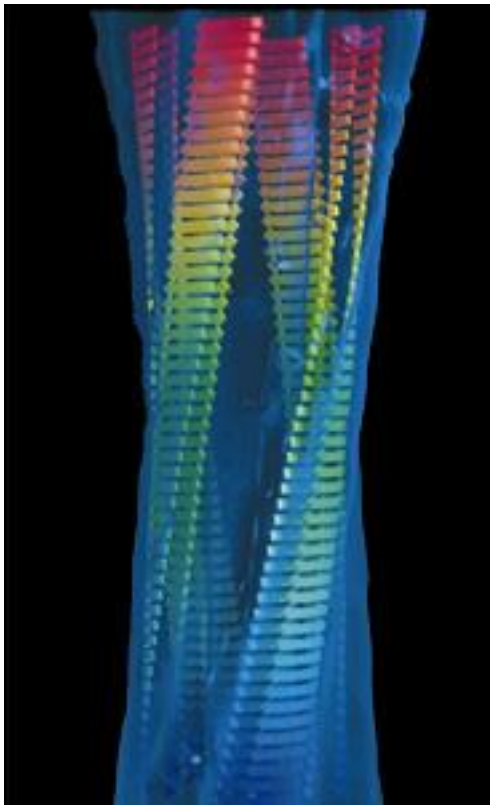
Supramolecular organizations

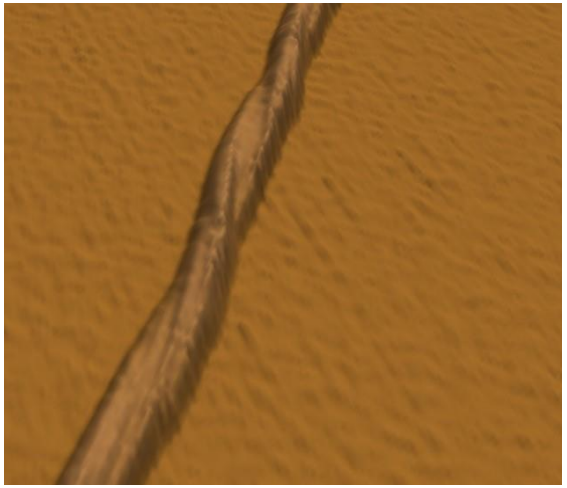
- Coiled coil
- Collagen
- Fibrillar structures





Collagen





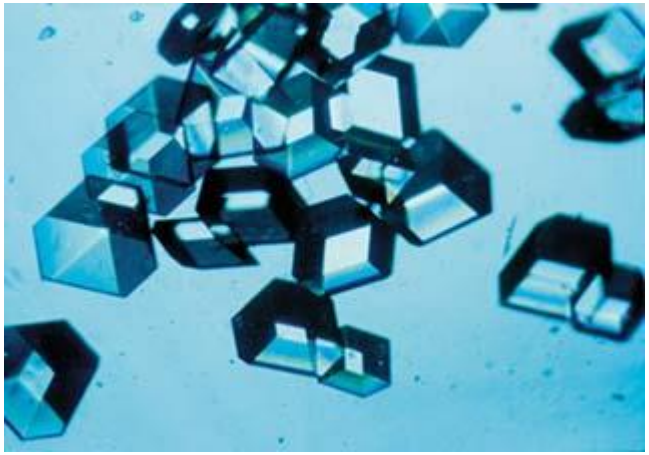
Fibrillar organization

Some methods for determination of the 3D structure of proteins

X-ray crystallography

NMR

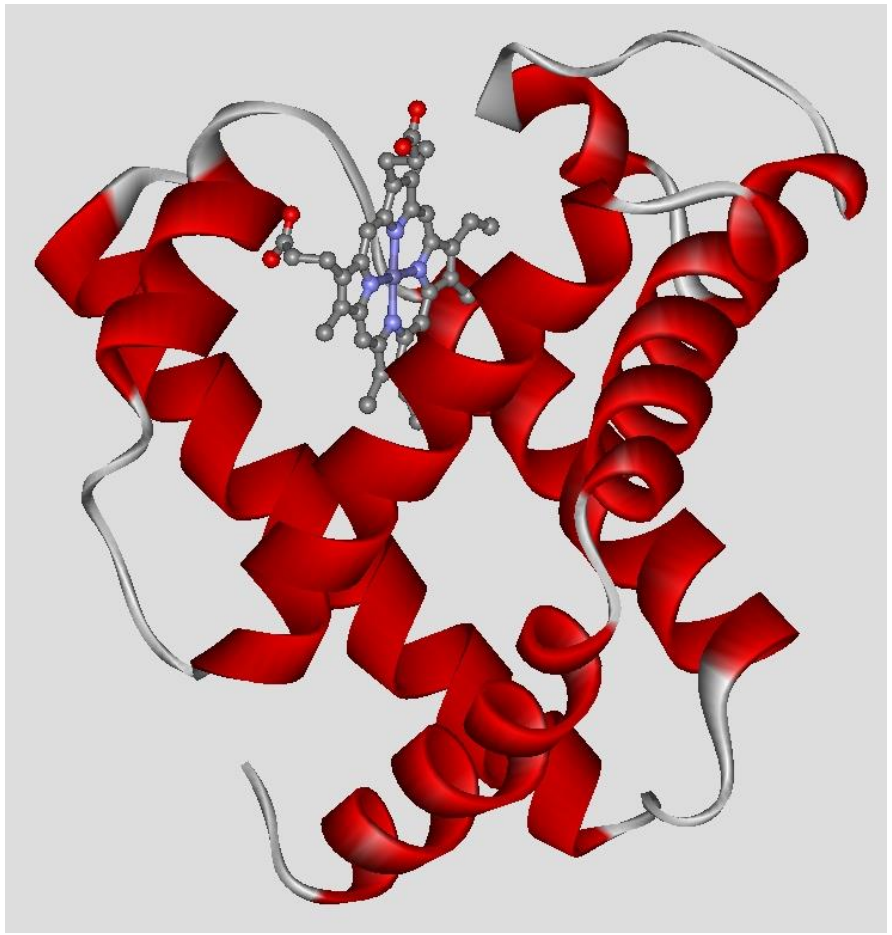
Prediction (homology modelling)

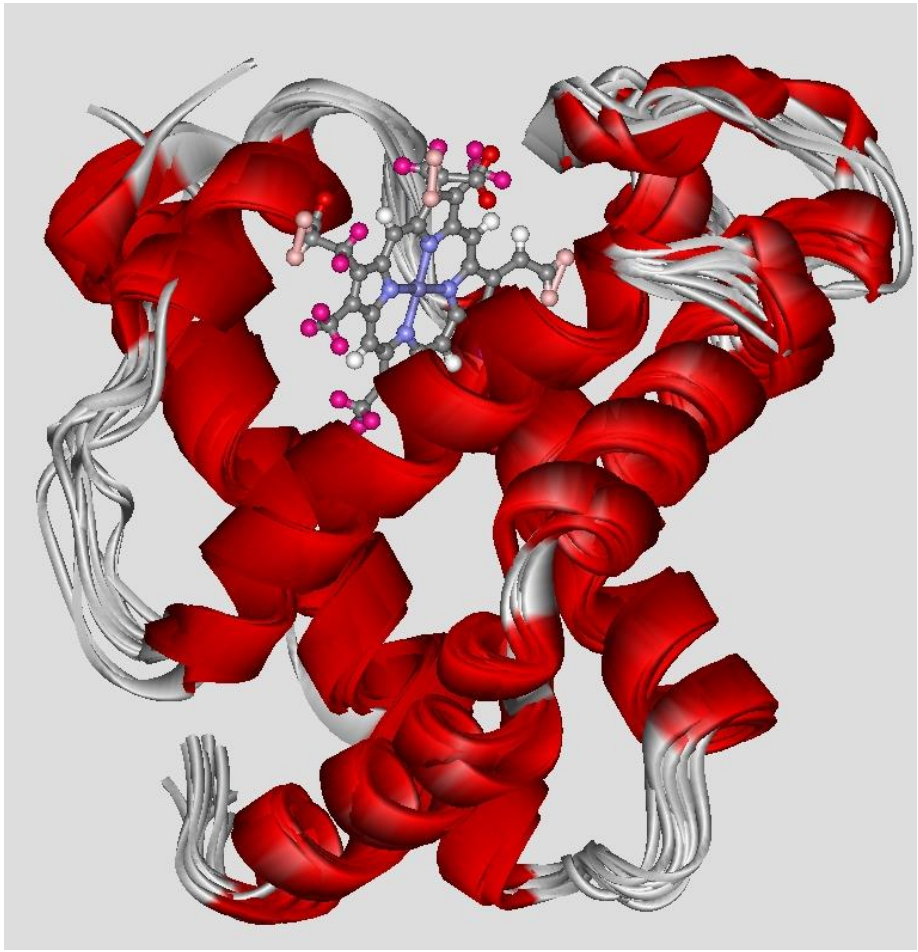


Spectroscopic methods sensitive to the changes of the protein structure

- Circular dichroism (CD)
- Infrared spectroscopy (IR, FTIR)
- Luminescence spectroscopy
- UV absorption spectroscopy
- …

Crystallography <-> NMR





myoglobin

Protein databases

- PDB
  - Protein Data Bank
    - 3D Structures (c.a.150 000) from
      - X-ray and
      - NMR experiments

Swiss-prot
  - Protein sequences
  - Proteomics software
    - Structure prediction (homology modeling)
    - Calculation/Estimation of the chemical parameters (e.g. isoelectric point…)
    - Comparison of the sequences…

https://www.rcsb.org

RCSB PDB | Deposit | Search | Visualize | Analyze | Download | Learn | More | Documentation | MyPDB

RCSB PDB PROTEIN DATA BANK

174507 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Enter search term(s)

Advanced Search | Browse Annotations | Help

PDB-101 | Worldwide PDB | EMDataResource | NDB Nucleic Acid Database | Worldwide Protein Data Bank Foundation

Celebrating 50 YEARS OF Protein Data Bank

- Welcome
- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

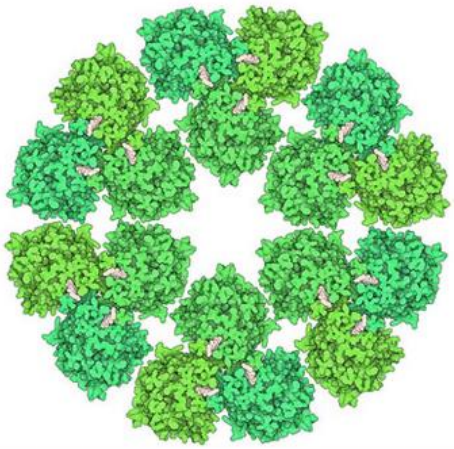
A Structural View of Biology

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.

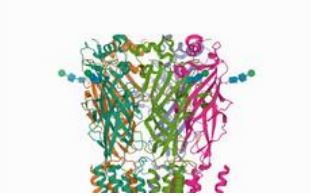
February **Molecule of the Month**



Cellulose Synthase

Latest Entries

As of Tue Feb 09 2021



Features & Highlights

IQB and ERN: Electron Microscopy Community Voice of the Customer

Register for the online February 11 workshop that will solicit feedback from microscopists and facility managers about IT challenges

News | Publications

PDB50: Submit Posters by March 15

Join the wwPDB May 4-5 for a symposium of speakers from around the world who have made tremendous advances in structural biology and bioinformatics » *02/15/2021*

26 Feb 2019: 149174 -> 15 Feb 2021: 174507

References

http://www.molecularmodels.ca/molecule/modelfiles/jb16alan.html

http://www.chemguide.co.uk/organicprops/aminoacids/background.html#top

http://www.chemguide.co.uk/organicprops/aminoacids/proteinstruct.html

http://www.enzim.hu/~vertessy/kovari_phd.pdf

http://www.pdb.org/

T. E Chreighton: Proteins, Freeman and Company, New York

Orvosi Biofizika Szerk: Damjanovich, Fidy, Szőllősi

Tarján Imre: A biofizika alapjai

Elődi Pál: Biokémia

http://www.pnas.org/content/108/43/E881.full

http://imtech.res.in/raghava/

http://www.cryst.bbk.ac.uk/PPS2/course/
http://www.cryst.bbk.ac.uk/PPS95/course/

http://mkk.szie.hu/dep/aeet/tanweb/Fogalomtar/index_fogalom.htm

http://www.med.upenn.edu/shorterlab/research.html

Különböző szerkezetvizsgáló módszerek a szerkezet eltérő nívóira érzékenyek

CD, IR: másodlagos szerkezeti arányok

Fluoreszcencia:

- trp: lokális környezet hidrofobicitása, harmadlagos struktúra változása, kitekeredés
- fluoreszcens jelzők: energiatranszfer kötődés kitekeredés,
- quenching: szerkezet kompaktsága (KI, akrilamid)
- ANS kötés: hidrofób felszínek
- …

UV absz.: aromás aminosavak környezete, kitekeredés, harmadlagos szerkezet

PDB file

```
HEADER    OXYGEN TRANSPORT                         27-SEP-93   1YMB      1YMB   2
COMPND    METMYOGLOBIN (HORSE HEART)                                     1YMB   3
SOURCE    HORSE (EQUUS CABALLUS)                                         1YMB   4
AUTHOR    S.V.EVANS,G.D.BRAYER                                           1YMB   5
REVDAT   1   31-JAN-94 1YMB    0                                         1YMB   6
JRNL        AUTH   S.V.EVANS,G.D.BRAYER                                 1YMB   7
JRNL        TITL   HIGH RESOLUTION STUDY OF THE THREE-DIMENSIONAL       1YMB   8
...
SEQRES   1    153  GLY LEU SER ASP GLY GLU TRP GLN GLN VAL LEU ASN VAL  1YMB  38
SEQRES   2    153  TRP GLY LYS VAL GLU ALA ASP ILE ALA GLY HIS GLY GLN  1YMB  39
...
HET    HEM    154       43     PROTOPORPHYRIN IX CONTAINS FE(II)        1YMB  50
...
HELIX    1   A SER      3  GLU     18  1                                1YMB  55
...
ATOM      1  N   GLY     1      -3.224  16.888  15.241  1.00 24.67      1YMB  74
ATOM      2  CA  GLY     1      -3.120  15.454  14.959  1.00 24.55      1YMB  75
ATOM      3  C   GLY     1      -2.299  14.719  16.007  1.00 24.06      1YMB  76
ATOM      4  O   GLY     1      -2.669  13.528  16.285  1.00 24.75      1YMB  77
ATOM      5  N   LEU     2      -1.312  15.335  16.627  1.00 23.47      1YMB  78
ATOM      6  CA  LEU     2      -0.482  14.622  17.641  1.00 21.68      1YMB  79
ATOM      7  C   LEU     2      -0.991  14.649  19.051  1.00 21.75      1YMB  80
...
END
```

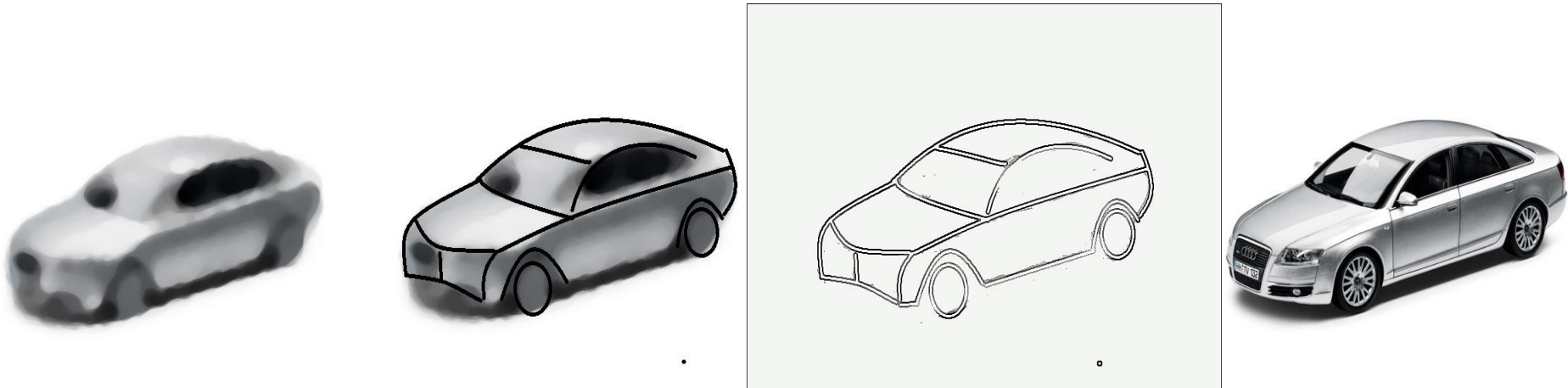
PDB file

Definició:

http://www.pdb.org/pdb/static.do?p=file_formats/index.jsp
http://www.wwpdb.org/docs.html

ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v33_A4.pdf

Felbontás: 1,5-2 Å (1 Å=10^{-10} m)



Fehérjevizualizáló programok

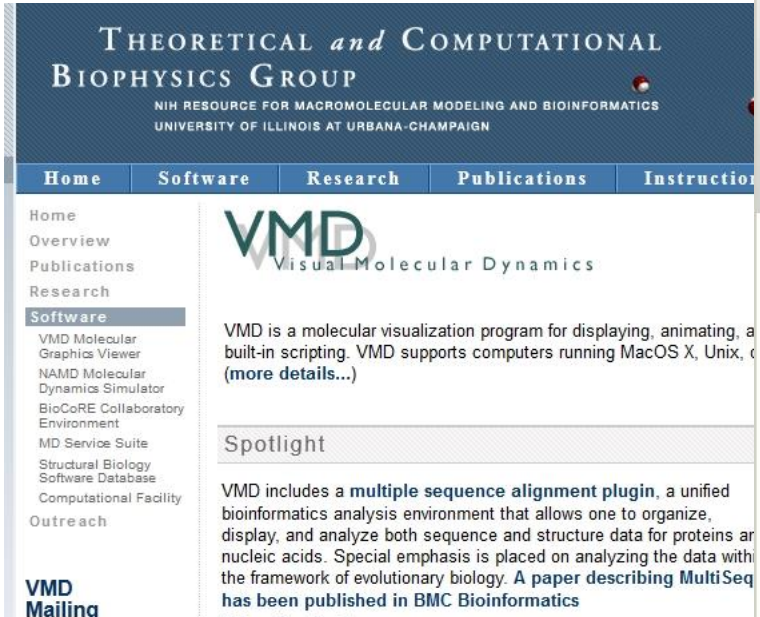
Rasmol (ingyenes, már nem fejlesztik)

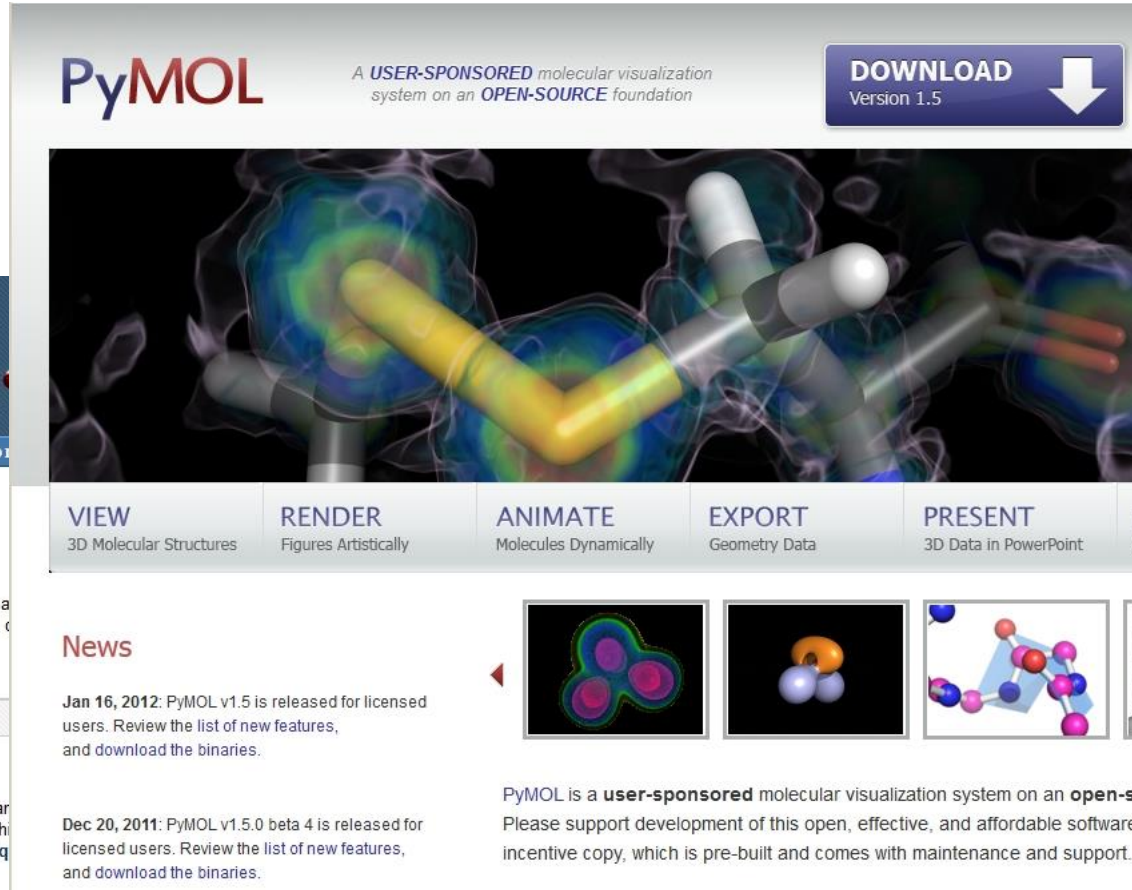
ViwerLite (ingyenes, már nem fejlesztik) csak fizetős verziója van

VLMD

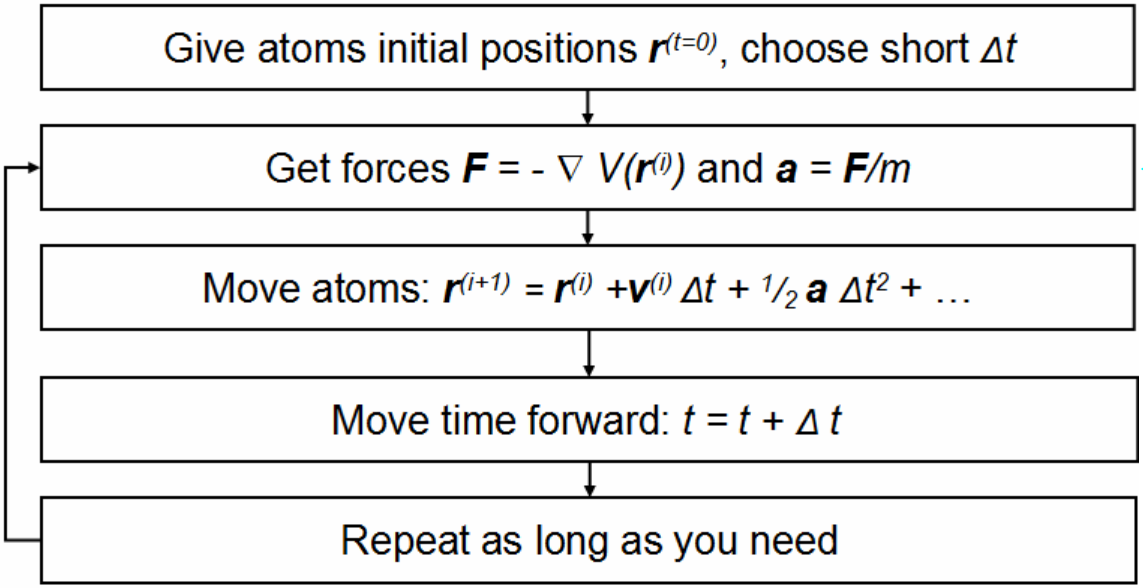
PyMOL

JMol (Java) www.jmol.org





Fehérjekonformáció és dinamika számolások:Molekuláris dinamika (MD)



PDB

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}$$

$$E_{nonbonded} = E_{electrostatic} + E_{vanderWaals}$$

Classification of Proteins According to Shape and Solubility

Proteins can be broadly classified into three groups, based on their shape and solubility.

Fibrous proteins: these proteins have a rod like structure. They are not soluble in water. Collagen is an example of a fibrous protein.

Globular proteins: these proteins more or less spherical in nature. Due to their distribution of amino acids (hydrophobic inside, hydrophillic outside) they are very soluble in aqueous solution. Myoglobin is an example of a globular protein.

Membrane proteins: these are protein which are in association with lipid membranes. Those membrane proteins that are embedded in the lipid bilayer have extensive hydrophobic amino acids that interact with the non-polar environment of the bilayer interior. Membrane proteins are not soluble in aqueous solution. Rhodopsin is an example of a membrane protein. Note that rhodopsin is an integral membrane protein and is embedded in the bilayer. The lipid membrane is not shown in the structure presented here.

Classification of Globular Proteins According to Secondary Structure

Proteins are also classified based on the type of secondary structure present.

All alpha: Proteins that contain only (in some exceptional instances there may be isolated beta-sheets) alpha helical secondary structure. Myoglobin is an example of an all alpha protein.