



Protein Structure and Visualization

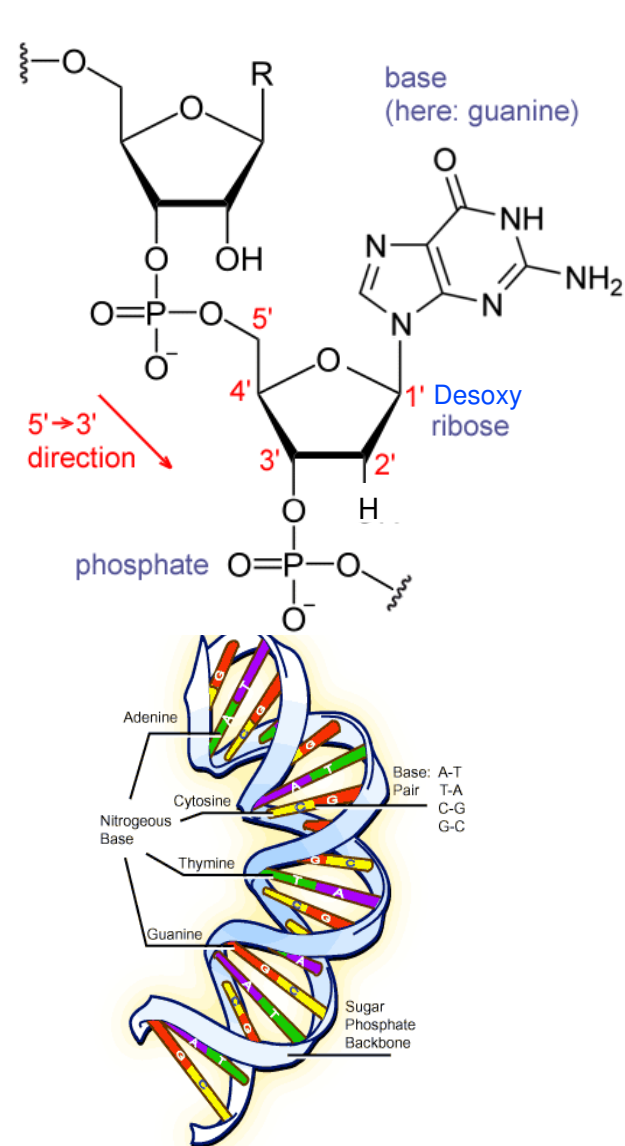
Peter Schellenerg

Friday 27.10.2023

The Molecular basis of nature

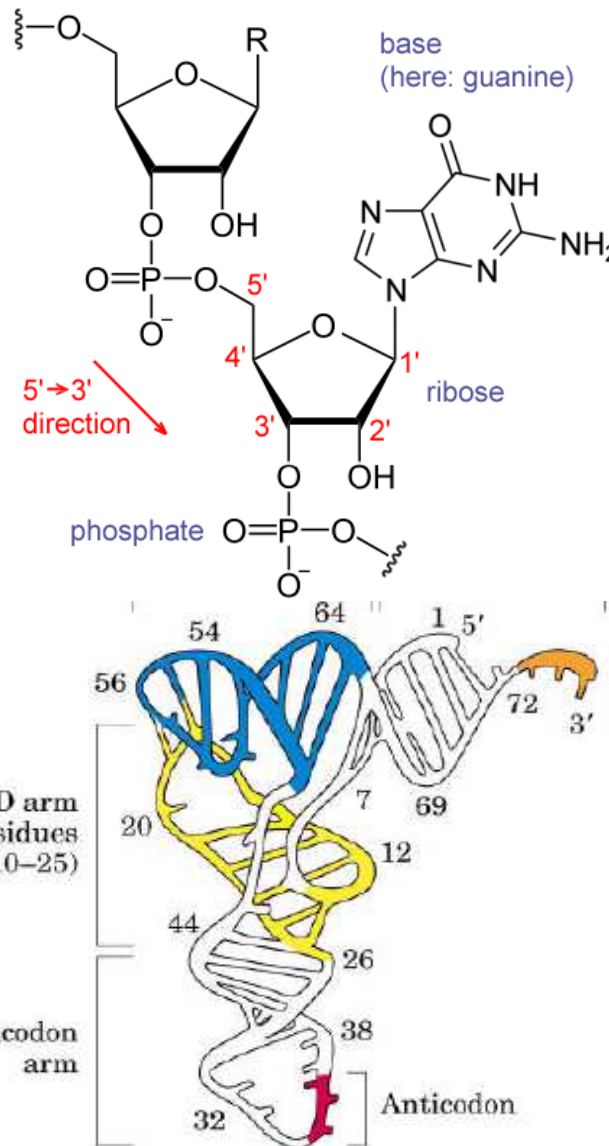
Information Storage

Deoxyribonucleic Acid



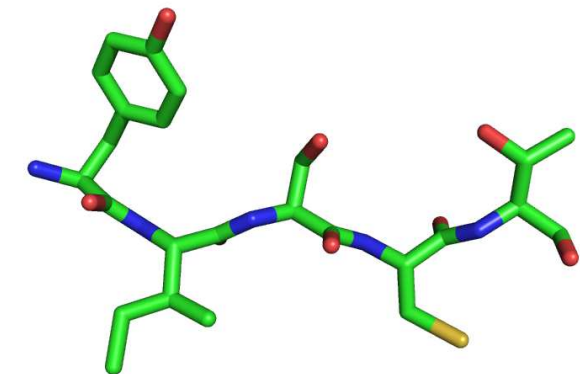
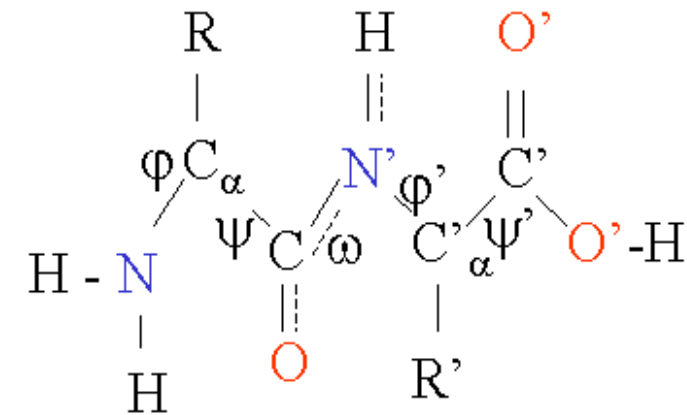
Information Transfer

Ribonucleic Acid



Functional Units

Proteins



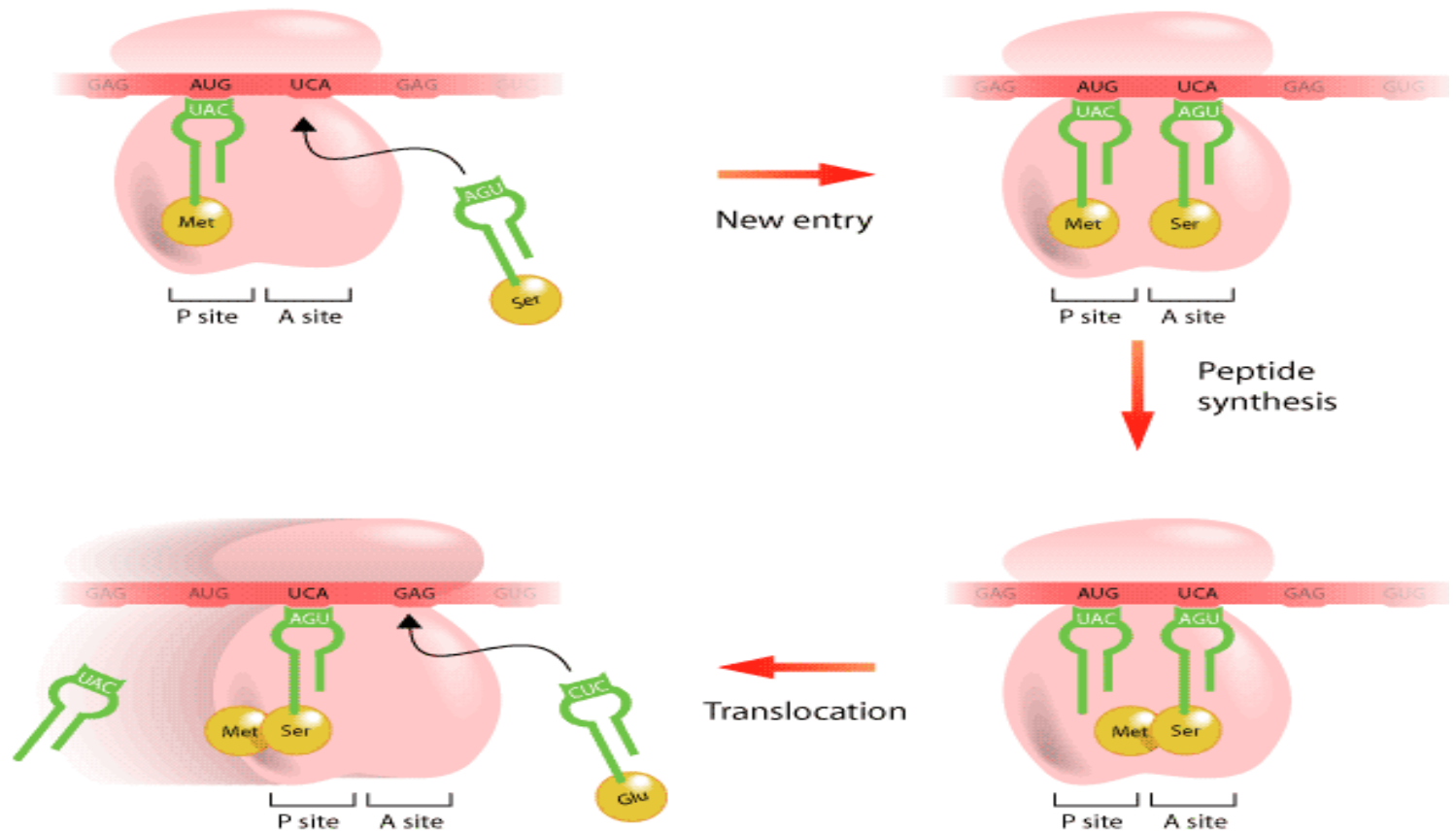
Here: TYR-ILE-SER-CYS-THR

Translation from RNA to Protein

a) Initiation

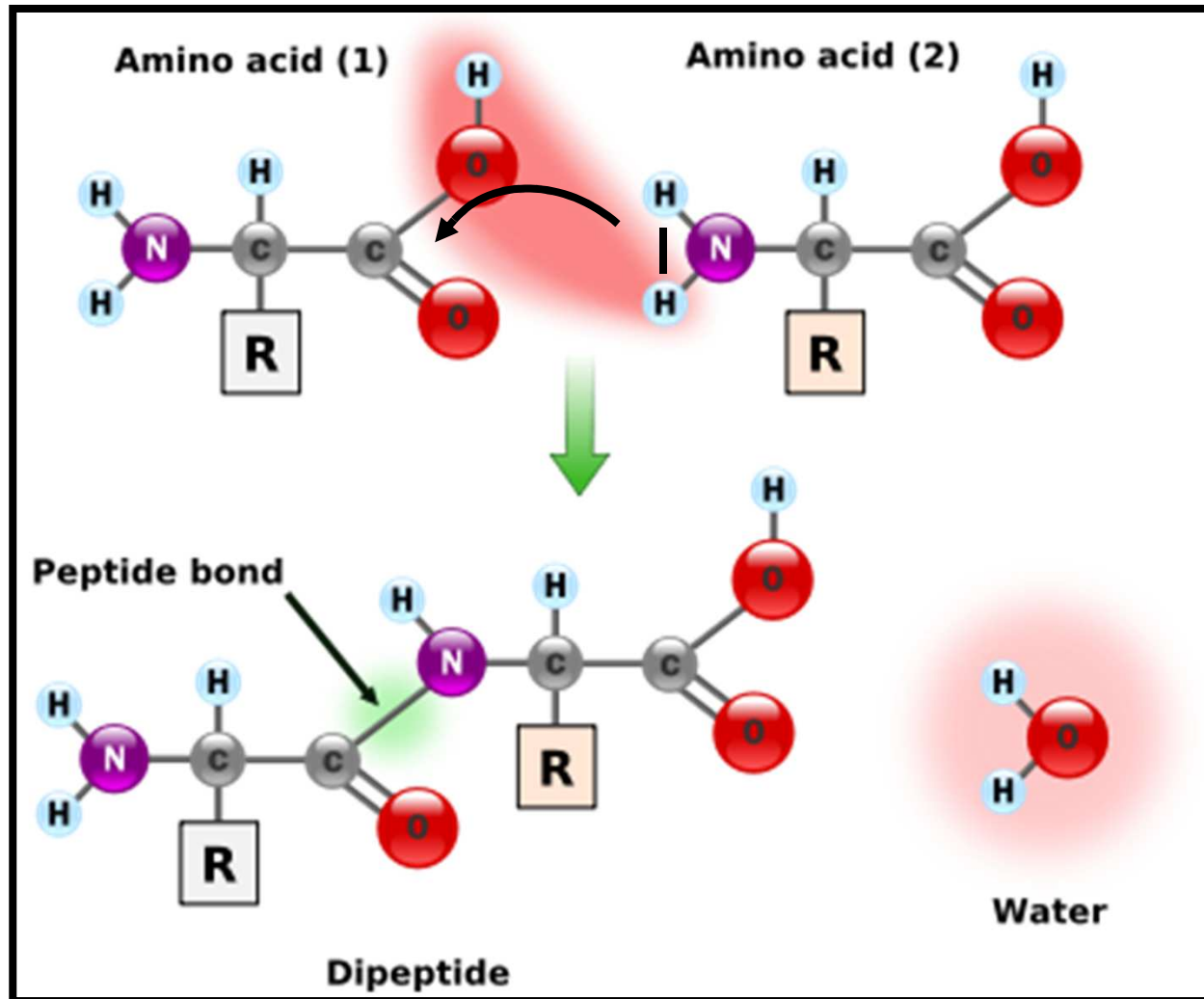


b) Elongation



Peptide Bond Formation

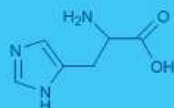

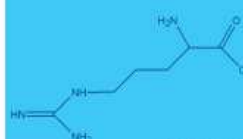
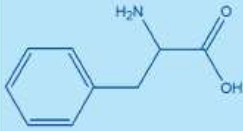
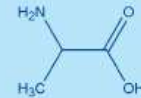

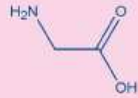
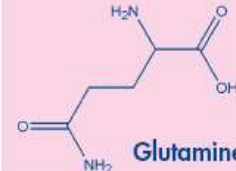
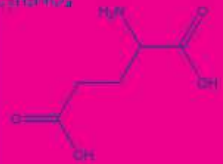

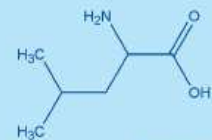
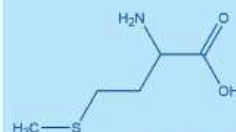
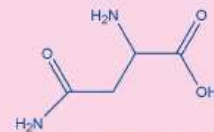
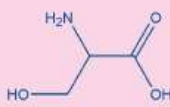

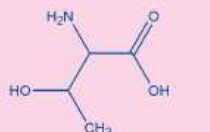
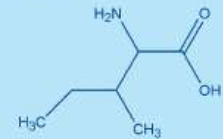

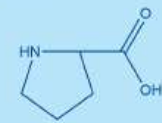
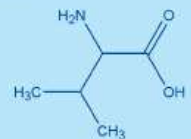
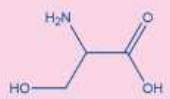
Nucleophilic attack of the **Nitrogen** electron lone pair on the positivited **Carbon**



Natural amino acid classification

Periodic Chart of Amino Acids

www.bachem.com

| | | | | | | |
|--|---|--|---|---|---|--|
| <div><div>H</div><div>155.16 137.14 C₉H₉N₃O₂</div><div></div><div>Histidine</div></div> | <div>Periodic Chart of Amino Acids</div> <div>www.bachem.com</div> | | | | <div><div>B</div><div>133.10 115.09 C₄H₇NO₄</div><div></div><div>Aspartic Acid</div></div> | |
| <div><div>R</div><div>174.20 156.19 C₆H₁₄N₄O₂</div><div></div><div>Arginine</div></div> | <div><div>F</div><div>165.19 147.18 C₉H₁₁NO₂</div><div></div><div>Phenylalanine</div></div> | <div><div>A</div><div>89.09 71.08 C₃H₇NO₂</div><div></div><div>Alanine</div></div> | <div><div>C</div><div>121.16 103.14 C₃H₇NO₂S</div><div></div><div>Cysteine</div></div> | <div><div>G</div><div>75.07 57.05 C₂H₅NO₂</div><div></div><div>Glycine</div></div> | <div><div>Q</div><div>146.15 128.13 C₅H₁₀N₂O₃</div><div></div><div>Glutamine</div></div> | <div><div>E</div><div>147.17 129.11 C₅H₉NO₄</div><div></div><div>Glutamic Acid</div></div> |
| <div><div>K</div><div>146.19 128.17 C₆H₁₄N₂O₂</div><div></div><div>Lysine</div></div> | <div><div>L</div><div>131.17 113.16 C₆H₁₃NO₂</div><div></div><div>Leucine</div></div> | <div><div>M</div><div>149.21 131.20 C₅H₁₁NO₂S</div><div></div><div>Methionine</div></div> | <div><div>N</div><div>132.12 114.10 C₄H₈N₂O₃</div><div></div><div>Asparagine</div></div> | <div><div>S</div><div>105.09 87.08 C₃H₇NO₃</div><div></div><div>Serine</div></div> | <div><div>Y</div><div>181.19 163.17 C₉H₁₁NO₃</div><div></div><div>Tyrosine</div></div> | <div><div>T</div><div>119.12 101.10 C₄H₉NO₃</div><div></div><div>Threonine</div></div> |
| <div><div>I</div><div>131.18 113.16 C₆H₁₃NO₂</div><div></div><div>Isoleucine</div></div> | <div><div>W</div><div>204.23 186.21 C₁₁H₁₂N₂O₂</div><div></div><div>Tryptophan</div></div> | <div><div>P</div><div>115.13 97.12 C₅H₉NO₂</div><div></div><div>Proline</div></div> | <div><div>V</div><div>117.15 99.13 C₅H₁₁NO₂</div><div></div><div>Valine</div></div> | <div><div><div><div></div>Basic</div><div><div></div>Nonpolar (hydrophobic)</div><div><div></div>Polar, uncharged</div><div><div></div>Acidic</div></div><div><div>1-Letter Amino Acid Code</div><div>Molecular Weight</div><div>MW-H₂O</div><div>Molecular Formula</div></div><div><div>3-Letter Amino Acid Code</div><div>Chemical Structure</div><div>Chemical Name</div></div></div> <div><div><div>S</div><div>105.09 87.08 C₃H₇NO₃</div><div></div><div>Serine</div></div></div> | | |

If Amino Acids can not do the task required, additional non-coded Cofactors are employed

In many cases these Cofactors are Chromophores and result in Chromoproteins

Light reception

Polyene:
-Carotenoids
-Rhodopsin

linear Tetrapyrrols

Cinnamic acid

Flavins

Photosynthesis

Cyclic Tetrapyrrols
-Chlorophylls
-Pheophytins

Open Tetrapyrrols

Carotenoids

Miscellaneous

Light absorption
not primary function!

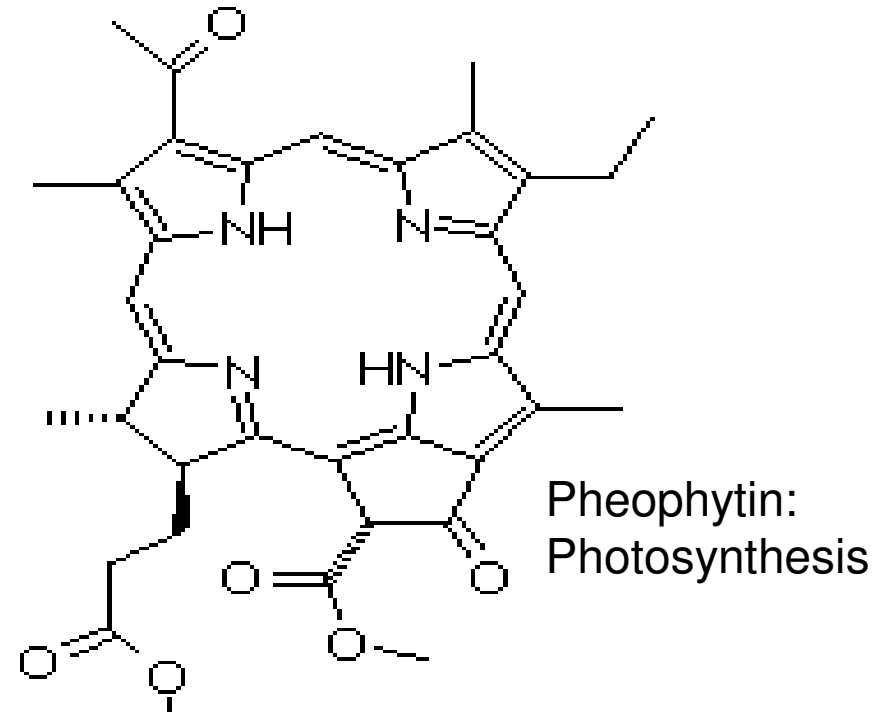
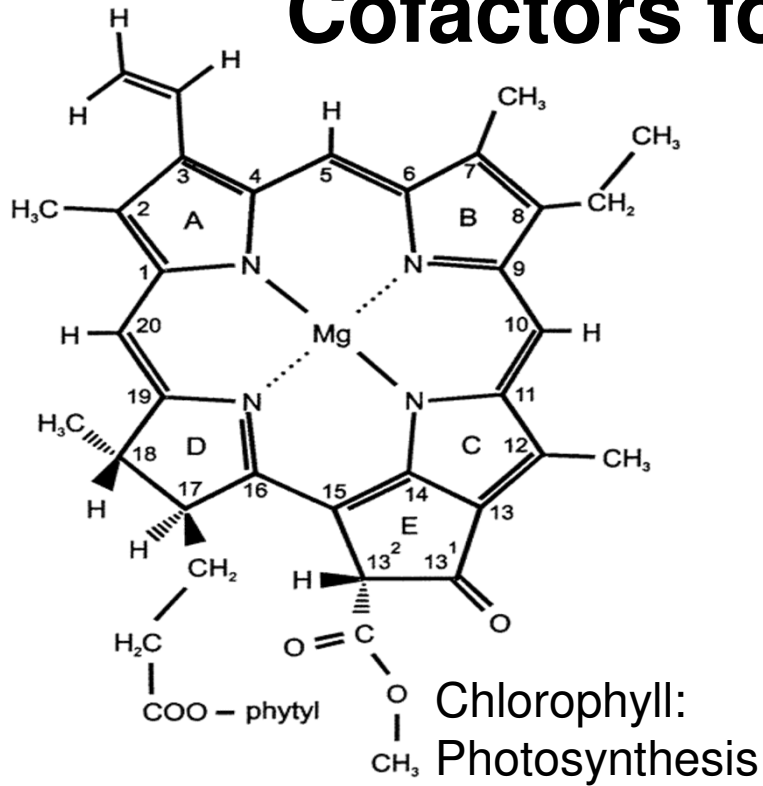
Hems

Flavins

Metalloproteins

GFP, XFP

Cofactors for light absorption –type I



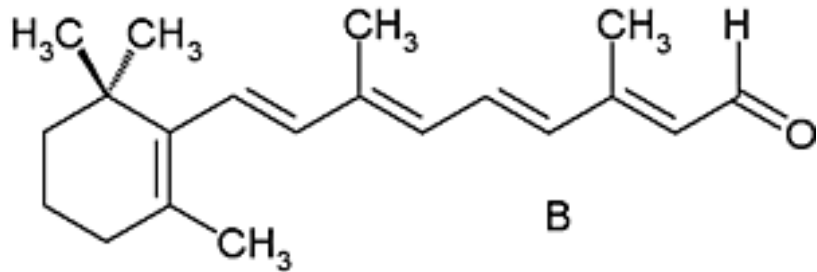
Large Frank Condon factor for 0-0 transitions:

- weak electron phonon coupling**
- small rearrangement of molecule and solvent (protein) cage upon excitation**

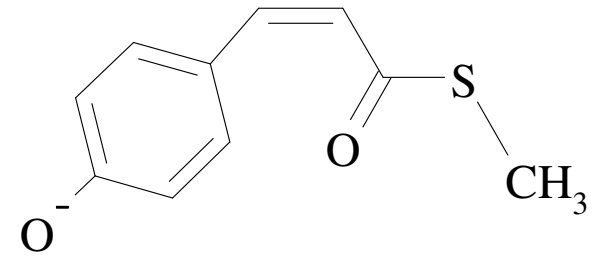
Therefore: small energy losses in the excitation process

very suitable for light to energy conversion !

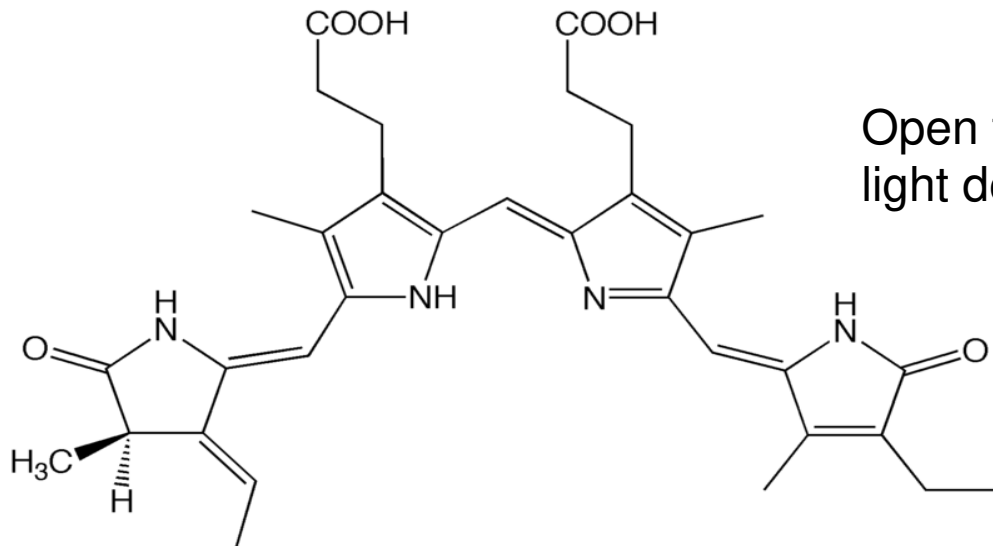
Cofactors for light absorption –type II



Retinal: light detection, photosynthesis,
Singlet-Oxygen Quencher



Photoactive Yellow Protein (PYP)
Light detection



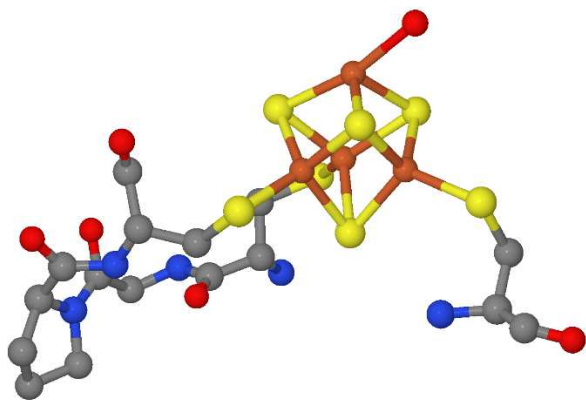
Open tetrapyrroles:
light detection, Photosynthesis

Large Frank Condon factor for vibrational modes due to cis-trans isomerization:

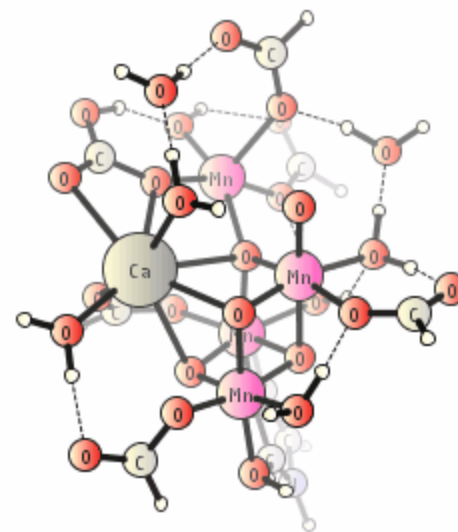
- strong electron phonon coupling**
- Large rearrangement of molecule and solvent (protein) cage upon excitation**

Therefore: high energy losses in the excitation process but large ‘signal’ you can hear it!

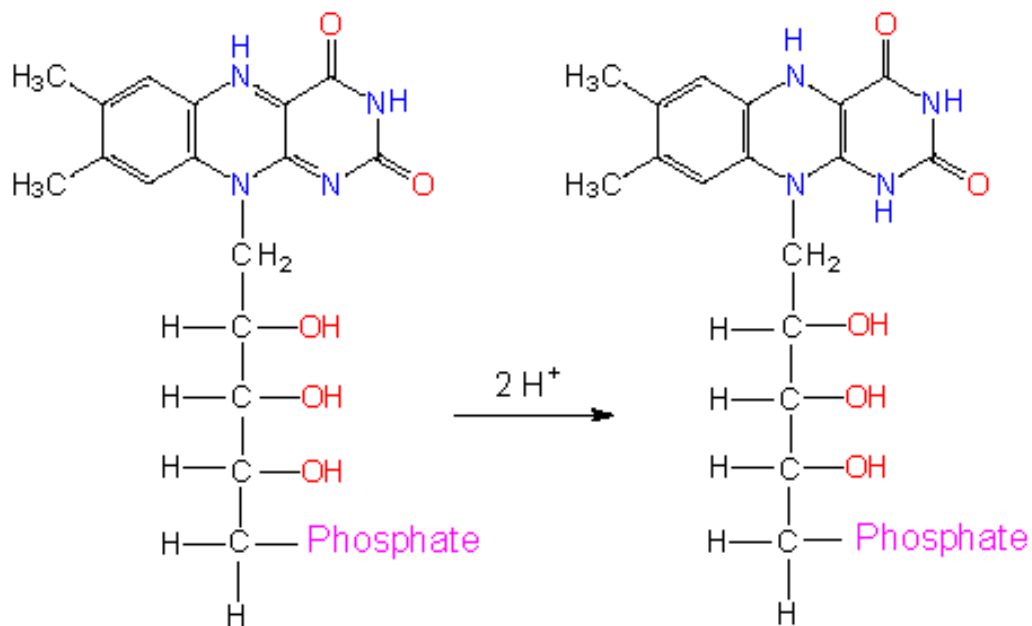
Cofactors for Redox Reactions



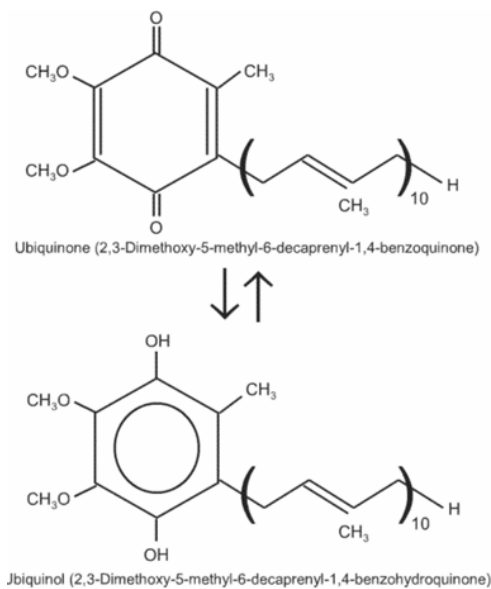
Iron-Sulfur-Cluster (4,4-Cluster)



Mn-O-Ca-Cluster (PSII)

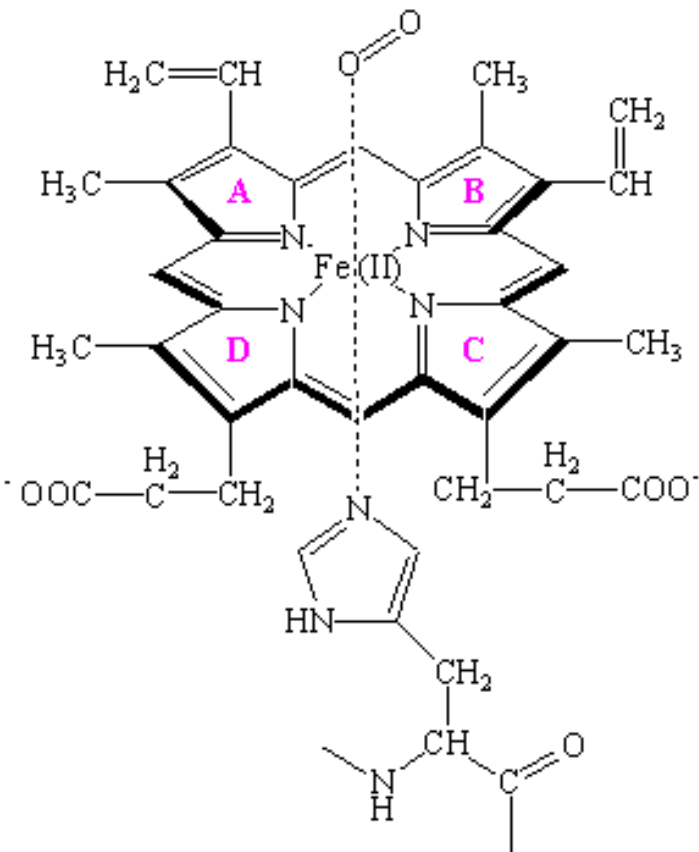


Riboflavin

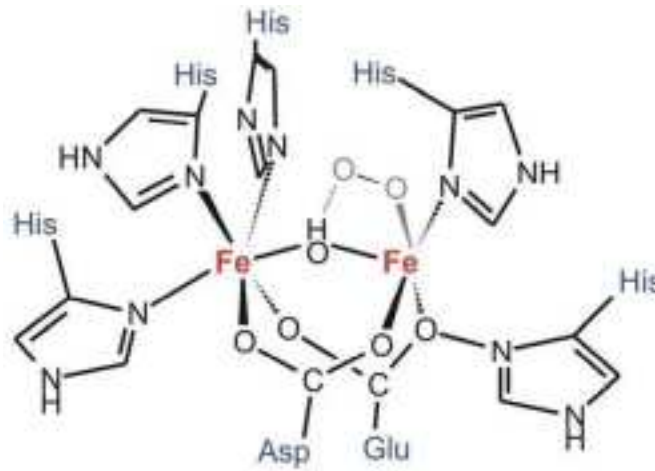


Ubichinone

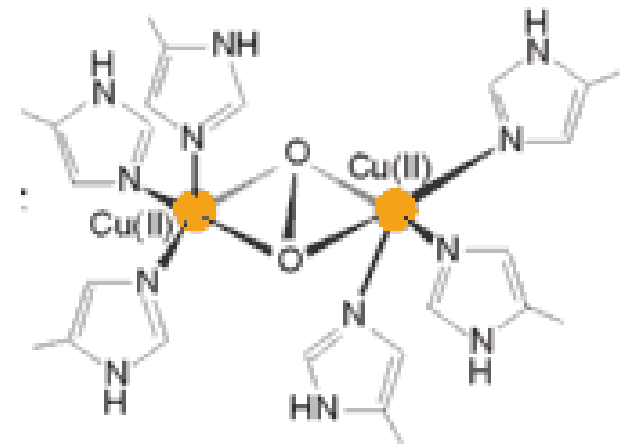
Cofactors for oxygen binding



HEM



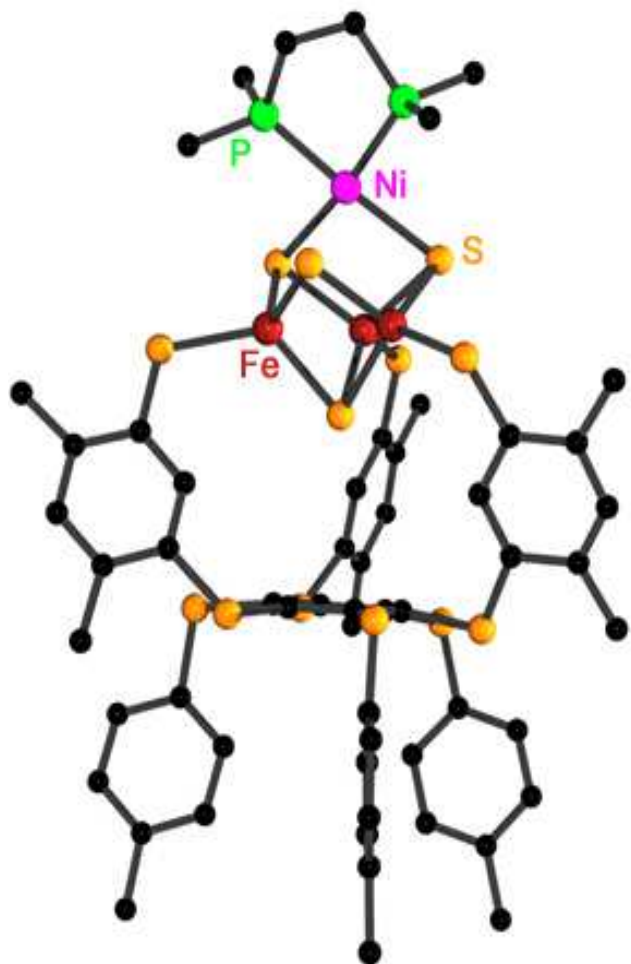
Hemerythrin
(marine invertebrates,
some worms, corals)



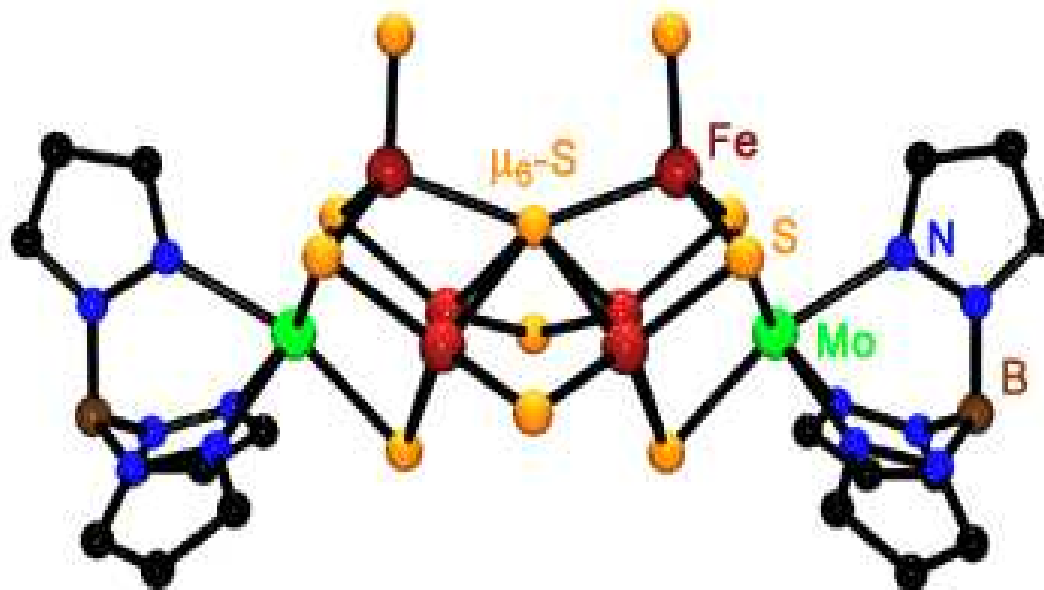
Hemocyanin
(from Squids, spiders
Arthropods, Mr. Spock)

Cofactors for Binding and Catalysis / Formation of Complex Compounds

Carbon monoxide dehydrogenases (CODH) catalyze the reaction
 $\text{CO} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + 2\text{H}^+ + 2\text{e}^-$.



The nitrogenase enzyme catalyzes the key reductive step of dinitrogen to ammonia in the global biological nitrogen cycle. Shown is the catalytic center (cofactor).



Chemical formula of a protein

C₇₇₄ H₁₂₂₄ N₂₁₀ O₂₂₂ S₅ Fe

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

How does this protein look like ?



The Protein Data Bank (PDB)

52

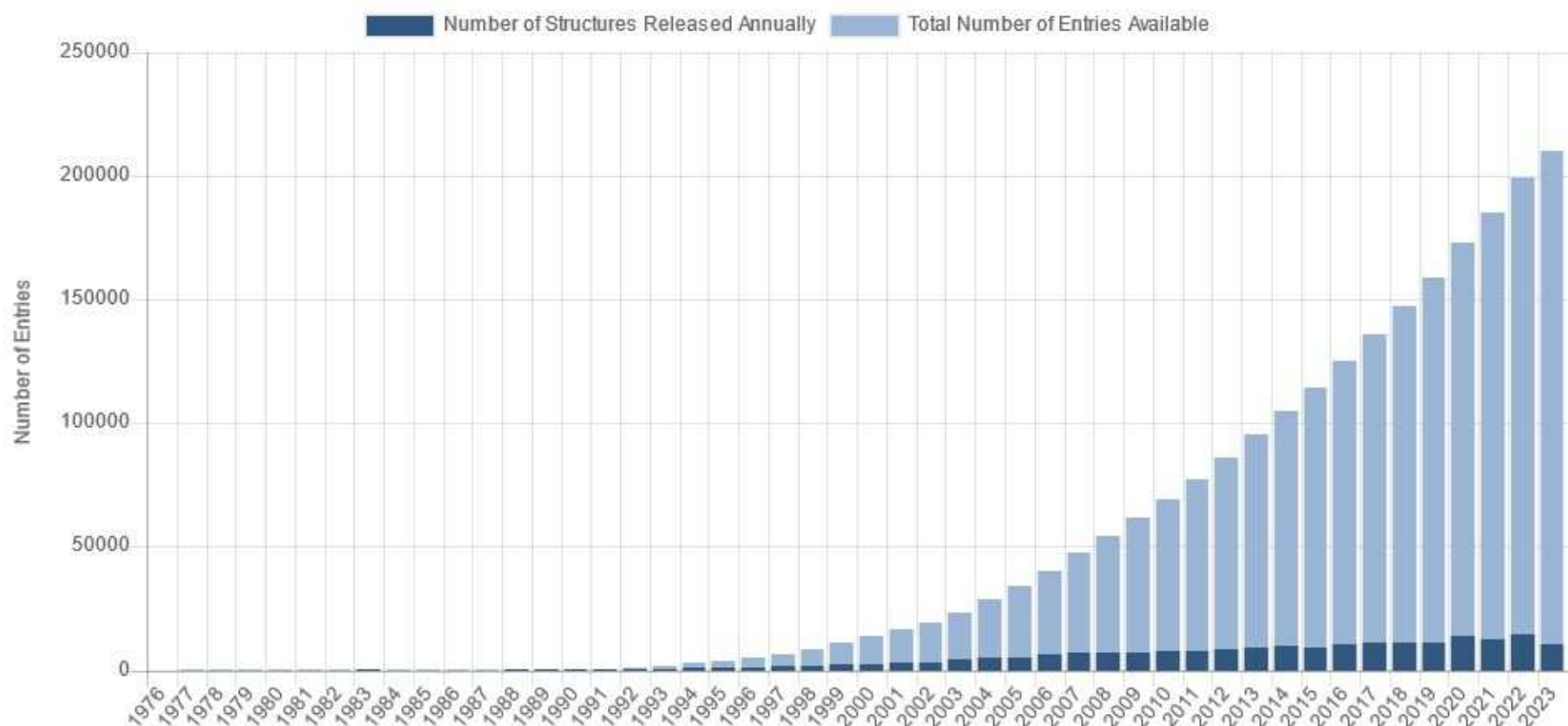
Happy Birthday PDB! Today 20th of Oct. is the 52nd anniversary of the founding of the Protein Data Bank

- PDB: publicly available archive of macromolecular structure data
>200000 deposited structures !
- Link: <http://www.rcsb.org>
- Reference: H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne.
The Protein Data Bank. Nucl. Ac. Res. **28**, 235-242 (2000).

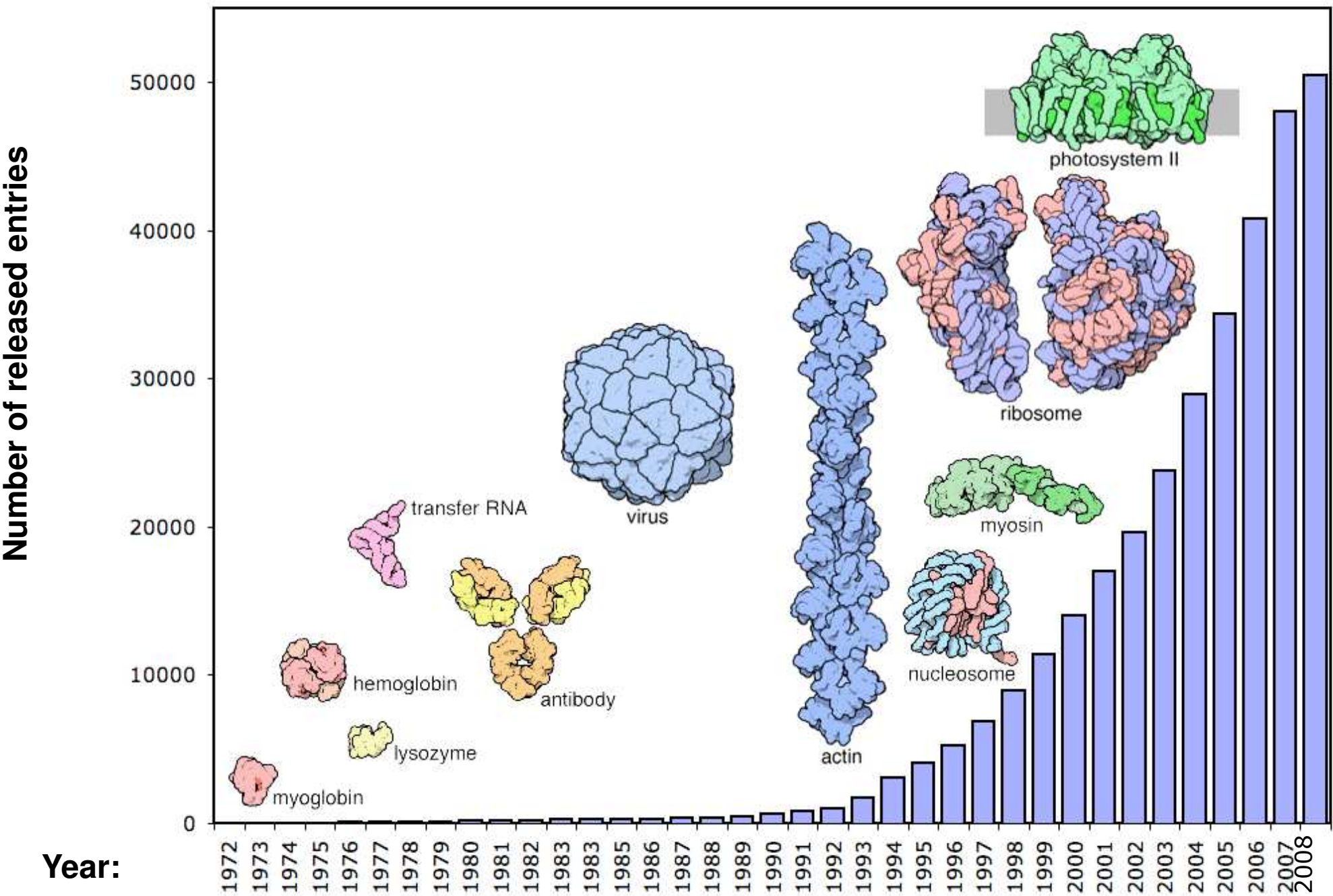
Deposited structures in the Protein Data Bank

PDB Statistics: Overall Growth of Released Structures Per Year

All Statistics



Also increase in complexity of the determined structures:



Methods utilized for structure determination

RCSB PDB PROTEIN DATA BANK

210,342 Structures from the PDB
1,068,577 Computed Structure Models (CSM)

3D Structures Include CSM ☐

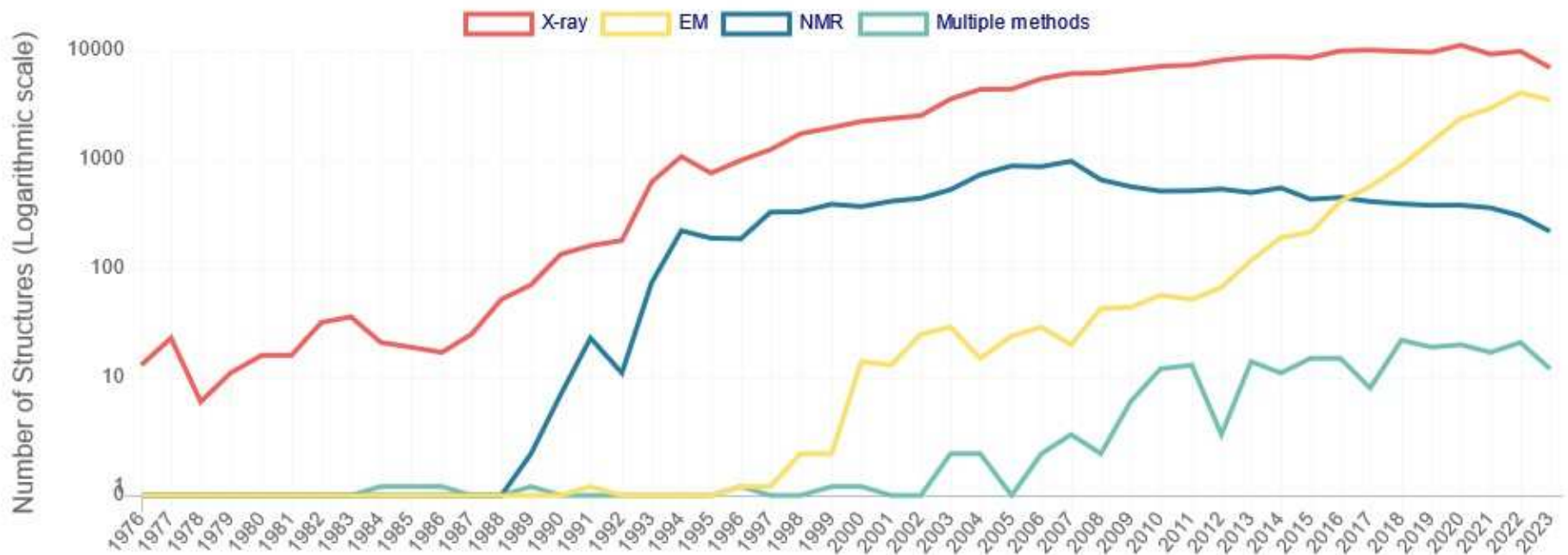
[Advanced Search](#) | [Browse Annotations](#) [Help](#)

PDB-101 PDB EMDDataResource NAKB wwPDB Foundation PDB-Dev

[f](#) [t](#) [v](#) [y](#)

Number of Released PDB Structures per Year

All Statistics



Structure file from the protein databank

HEADER DE NOVO PROTEIN 11-SEP-03 1QYS
TITLE CRYSTAL STRUCTURE OF TOP7: A COMPUTATIONALLY DESIGNED
TITLE 2 PROTEIN WITH A NOVEL FOLD
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: TOP7;
COMPND 3 CHAIN: A;
COMPND 4 ENGINEERED: YES
SOURCE MOL_ID: 1;
SOURCE 2 ORGANISM_SCIENTIFIC: COMPUTATIONALLY DESIGNED SEQUENCE;
SOURCE 3 EXPRESSION_SYSTEM: ESCHERICHIA COLI;

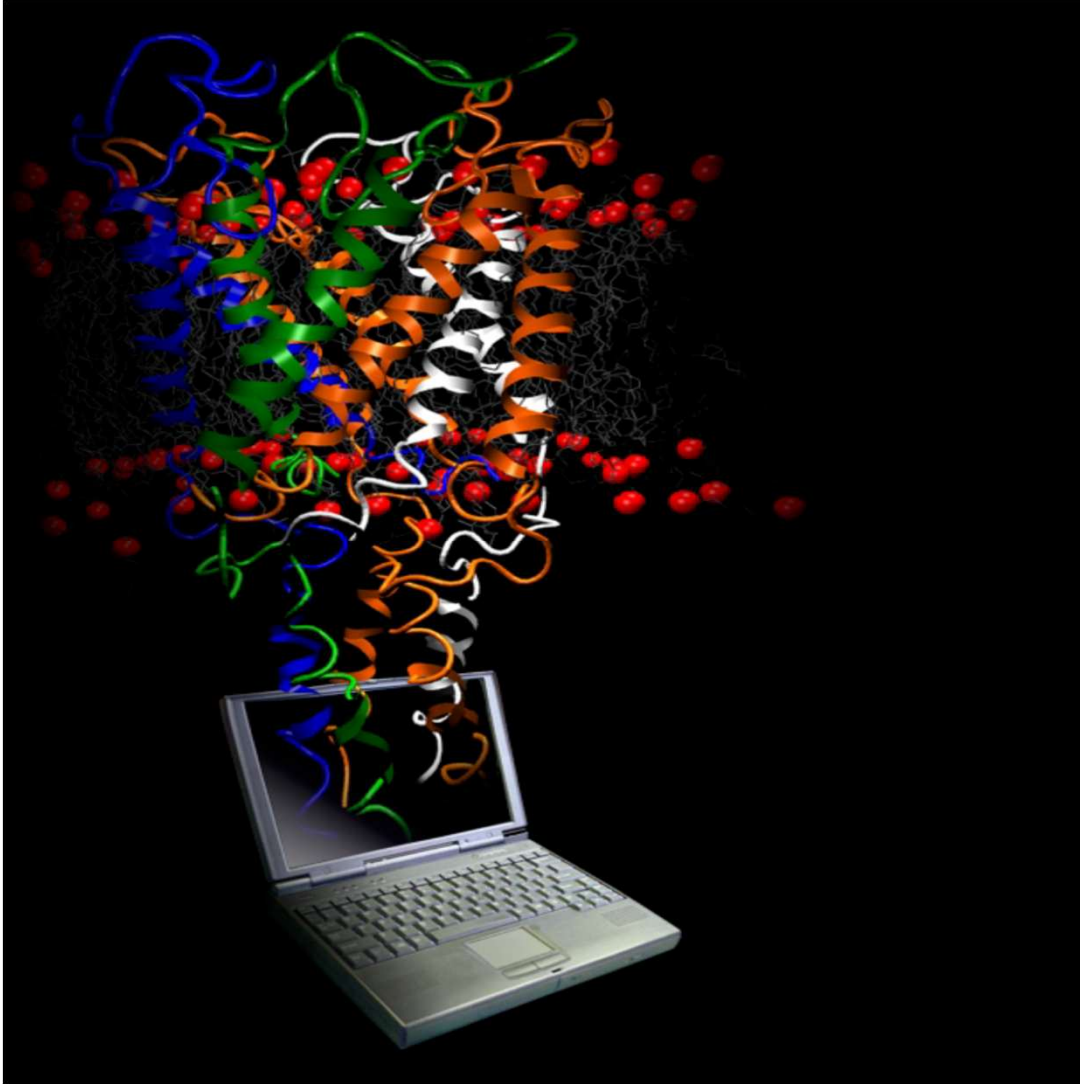
...

...

...

| | | | | | | | | | | |
|------|----|-----|-----|---|---|--------|--------|--------|------------|---|
| ATOM | 1 | N | ASP | A | 3 | -4.522 | 18.306 | 17.409 | 1.00174.51 | N |
| ATOM | 2 | CA | ASP | A | 3 | -3.061 | 18.228 | 17.122 | 1.00174.51 | C |
| ATOM | 3 | C | ASP | A | 3 | -2.664 | 16.993 | 16.324 | 1.00174.51 | C |
| ATOM | 4 | O | ASP | A | 3 | -3.515 | 16.306 | 15.754 | 1.00174.51 | O |
| ATOM | 5 | CB | ASP | A | 3 | -2.261 | 18.246 | 18.422 | 1.00 83.84 | C |
| ATOM | 6 | CG | ASP | A | 3 | -1.658 | 19.600 | 18.711 | 1.00 83.84 | C |
| ATOM | 7 | OD1 | ASP | A | 3 | -1.169 | 20.249 | 17.760 | 1.00 83.84 | O |
| ATOM | 8 | OD2 | ASP | A | 3 | -1.654 | 20.007 | 19.892 | 1.00 83.84 | O |
| ATOM | 9 | N | ILE | A | 4 | -1.360 | 16.714 | 16.297 | 1.00 57.73 | N |
| ATOM | 10 | CA | ILE | A | 4 | -0.823 | 15.562 | 15.568 | 1.00 57.73 | C |
| ATOM | 11 | C | ILE | A | 4 | -0.721 | 14.309 | 16.433 | 1.00 57.73 | C |
| ATOM | 12 | O | ILE | A | 4 | 0.091 | 14.222 | 17.355 | 1.00 57.73 | O |
| ATOM | 13 | CB | ILE | A | 4 | 0.555 | 15.888 | 14.980 | 1.00 57.14 | C |
| ATOM | 14 | CG1 | ILE | A | 4 | 0.425 | 17.108 | 14.058 | 1.00 57.14 | C |
| ATOM | 15 | CG2 | ILE | A | 4 | 1.097 | 14.674 | 14.218 | 1.00 57.14 | C |
| ATOM | 16 | CD1 | ILE | A | 4 | 1.737 | 17.831 | 13.766 | 1.00 57.14 | C |
| ATOM | 17 | N | GLN | A | 5 | -1.567 | 13.342 | 16.105 | 1.00 49.21 | N |
| ATOM | 18 | CA | GLN | A | 5 | -1.632 | 12.078 | 16.814 | 1.00 49.21 | C |
| ATOM | 19 | C | GLN | A | 5 | -0.907 | 10.950 | 16.066 | 1.00 49.21 | C |
| ATOM | 20 | O | GLN | A | 5 | -1.306 | 10.567 | 14.965 | 1.00 84.12 | O |
| ATOM | 21 | CB | GLN | A | 5 | -3.095 | 11.691 | 17.017 | 1.00 87.59 | C |
| ATOM | 22 | CG | GLN | A | 5 | -3.284 | 10.472 | 17.883 | 1.00 87.59 | C |
| ATOM | 23 | CD | GLN | A | 5 | -2.740 | 10.671 | 19.282 | 1.00 87.59 | C |

Protein visualization programs



Computer:

Rasmol (historical)

Pymol

UCSF Chimera

Swiss PDB -viewer

VMD

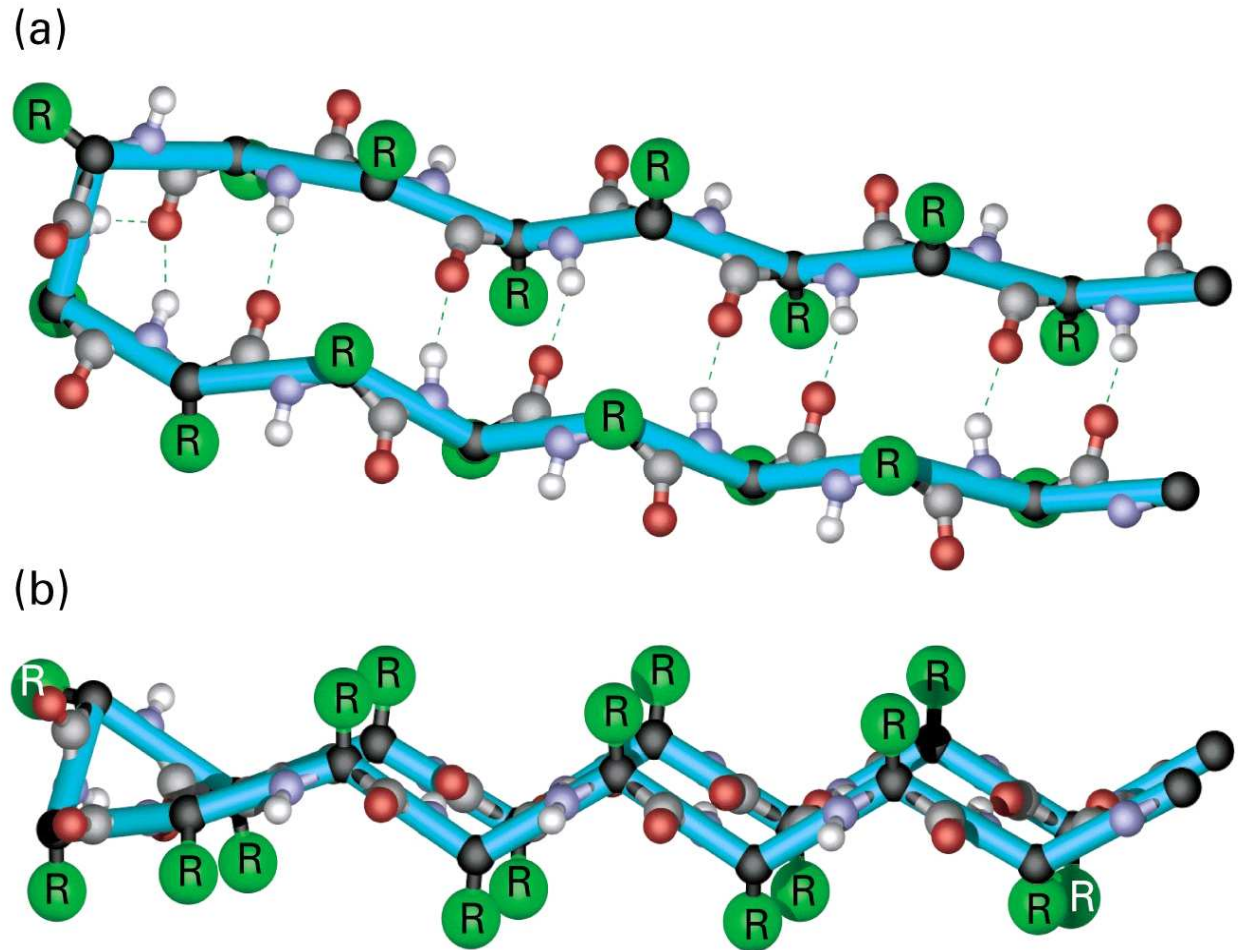
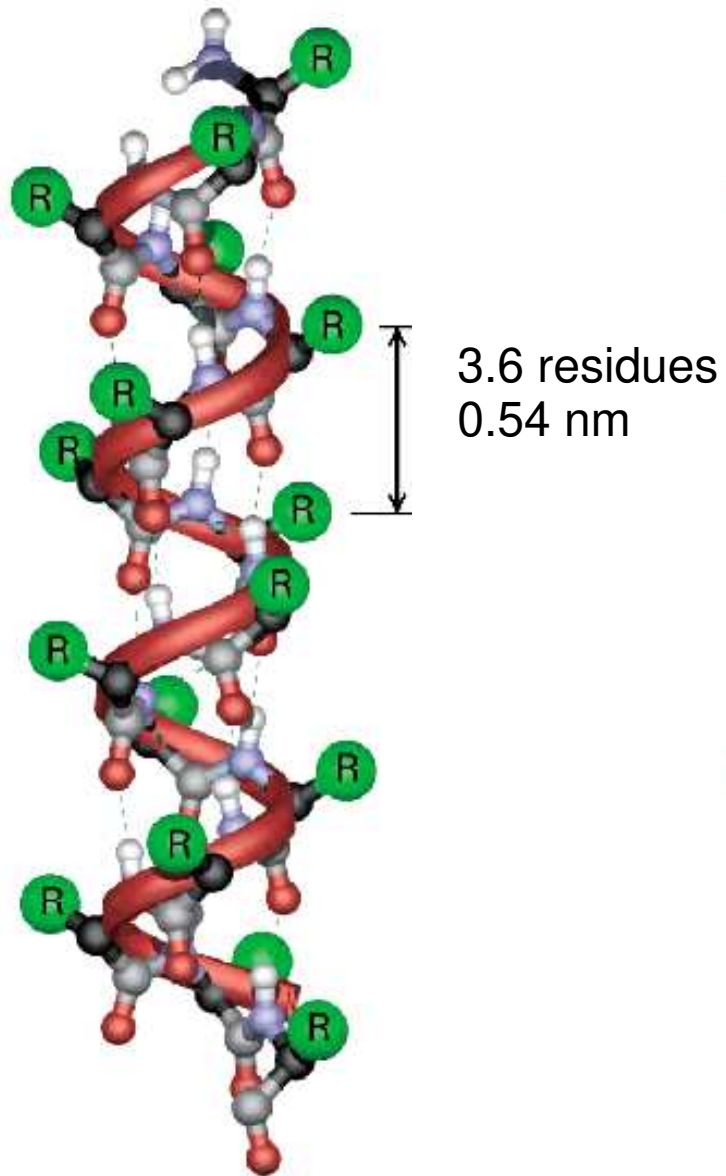
Browser:

Jmol

Mol* 3D Viewer

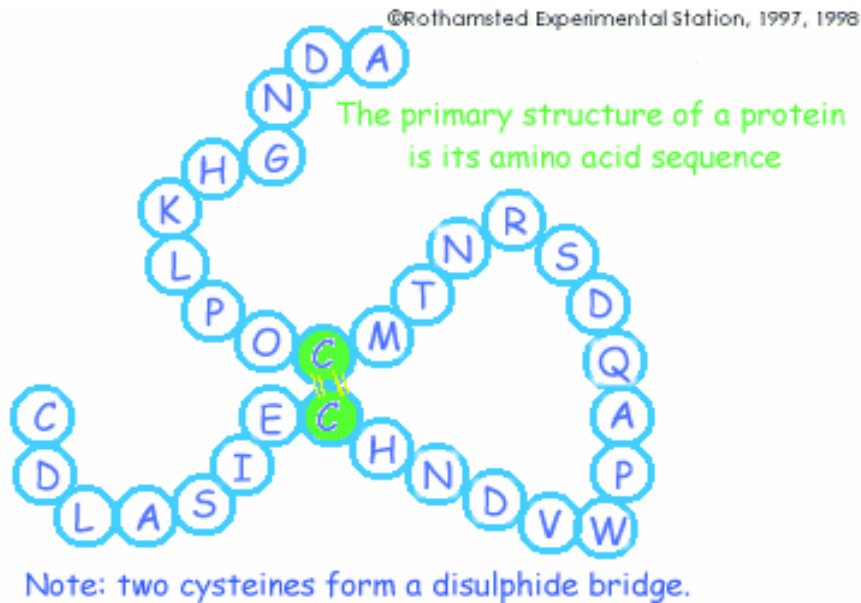
Chime (historical)

The protein folds in an ordered fashion supported by hydrogen bonding: -The secondary structure

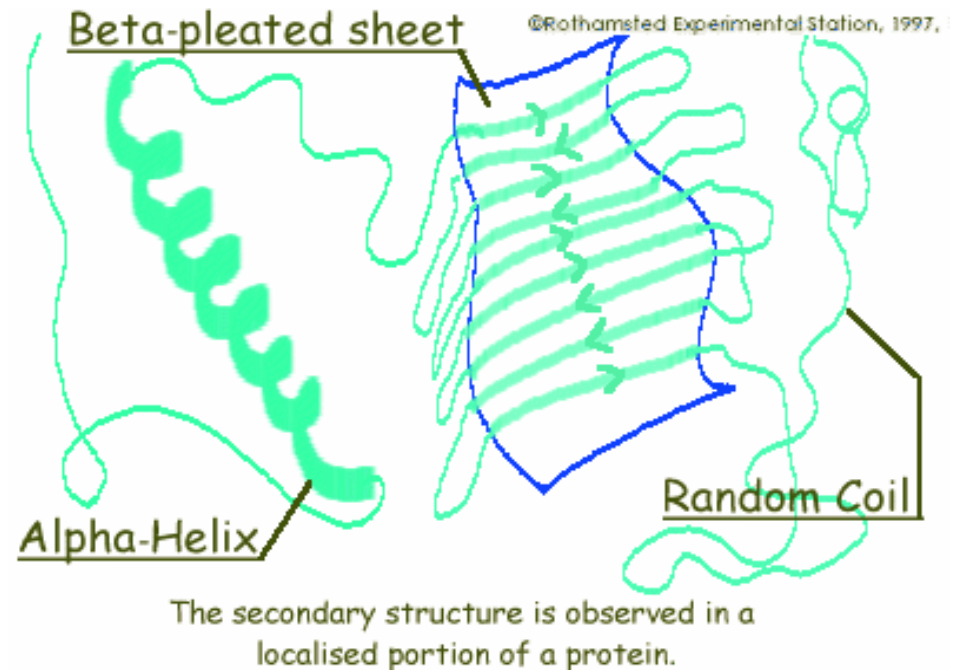


Hierarchy of Protein Structure Features

Primary structure

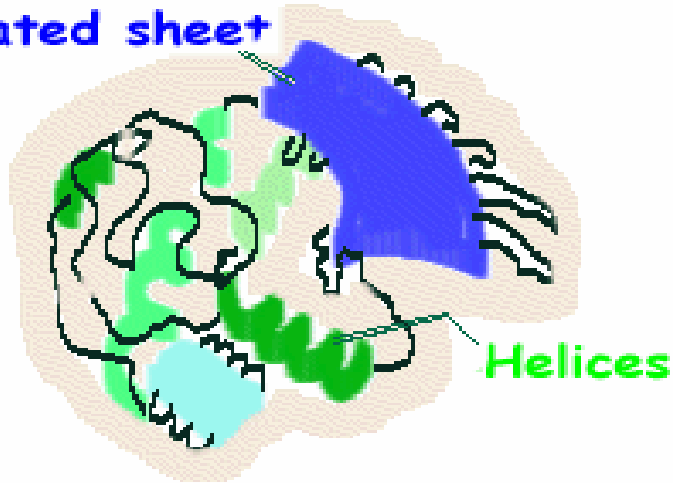


Secondary structure



Tertiary structure

Pleated sheet

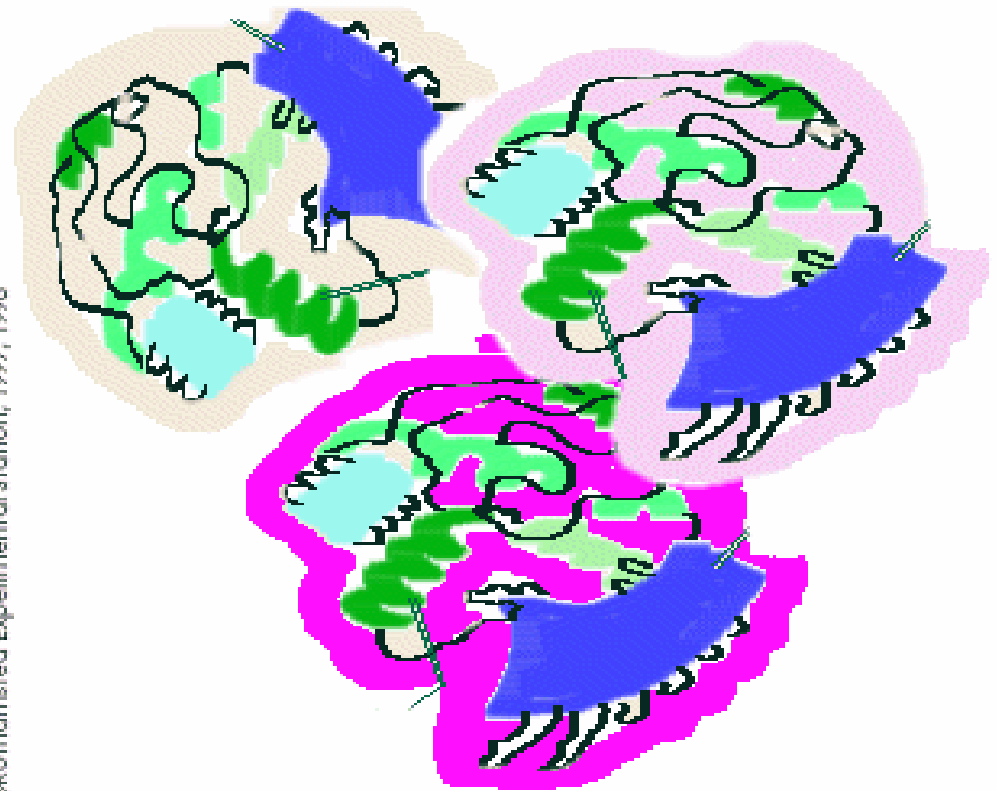


The tertiary structure is the way the secondary structures fold onto themselves to form a protein or a subunit of a more complex protein.

©Rothamsted Experimental Station, 1997, 1998

Quaternary structure

Only proteins with more than one chain have a quaternary structure



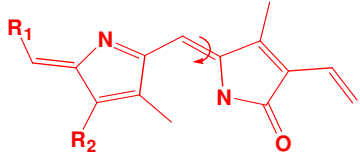
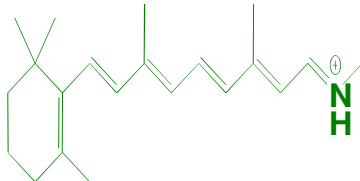
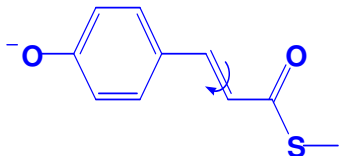
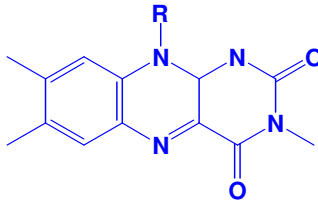
©Rothamsted Experimental Station, 1997, 1998

Additonal possible discussion

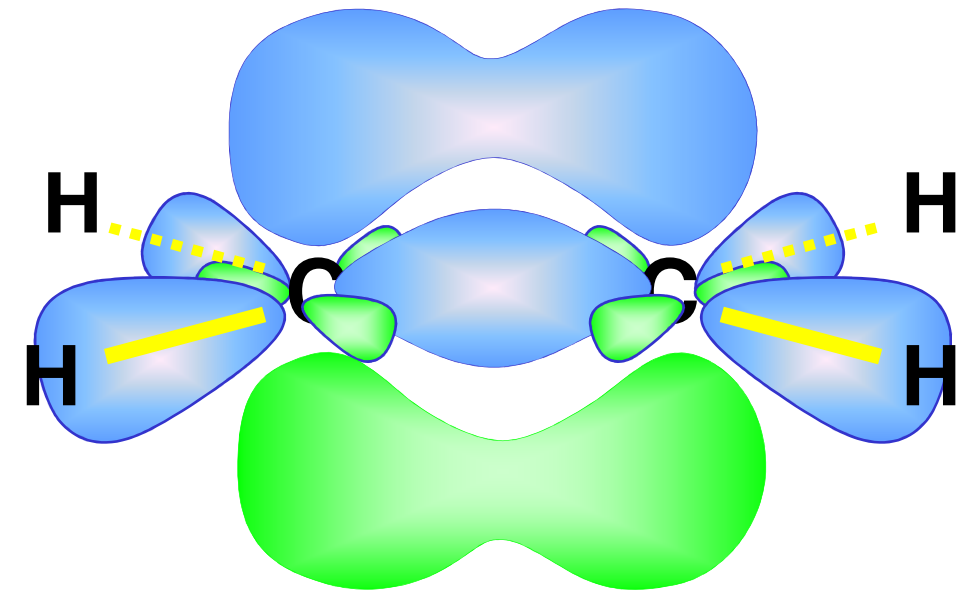
Molecular mechanism of cis –trans Isomerization in solution

-Orbitals , hyperpotential surfaces and Jablonski -Diagram

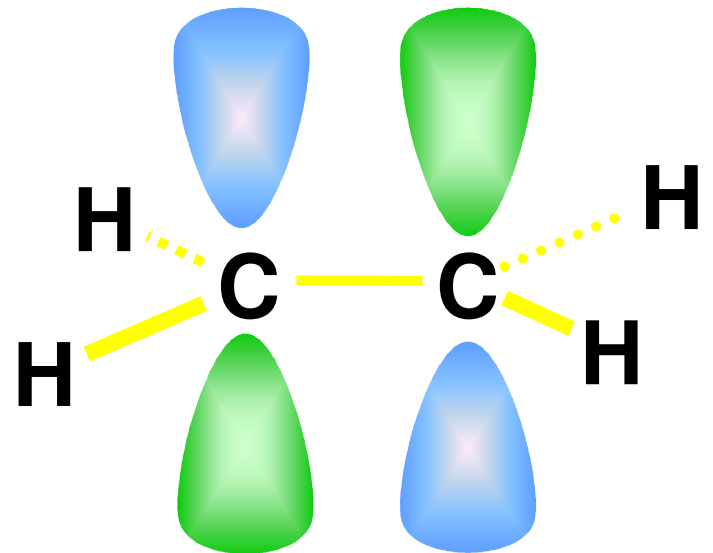
PHOTOSENSORY PROTEIN FAMILIES & their photochemistry:

| Photoreceptor family | Chromophore structure | Example | Primary photochemistry |
|----------------------|--------------------------------|---|------------------------------|
| Phytochromes | linear tetrapyrrole |  | cis <-> trans isomerization |
| Rhodopsins | retinal (<i>i.e.</i> polyene) |  | cis <-> trans isomerization |
| Xanthopsins | 4-OH-cinnamic acid |  | cis <-> trans isomerization |
| Cryptochromes | Flavin (FAD; plus pterin) |  | electron transfer? |
| Phototropins | Flavin (FMN) | | cysteinylyl adduct formation |

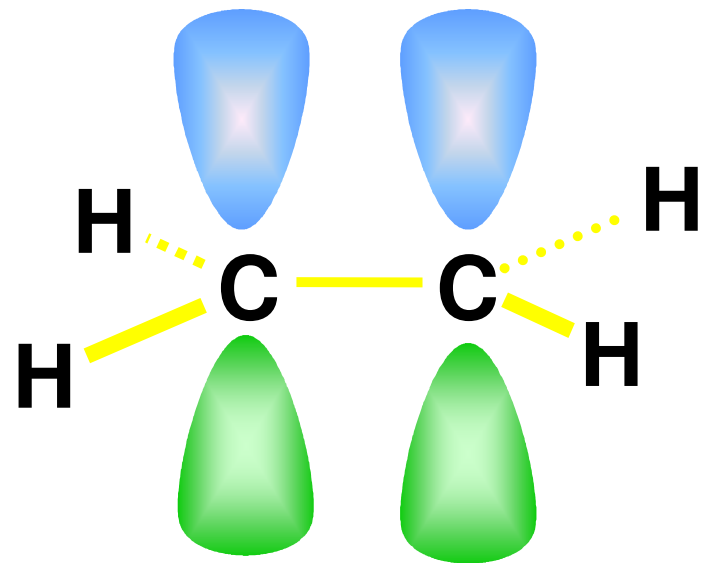
MO Schema von Ethene



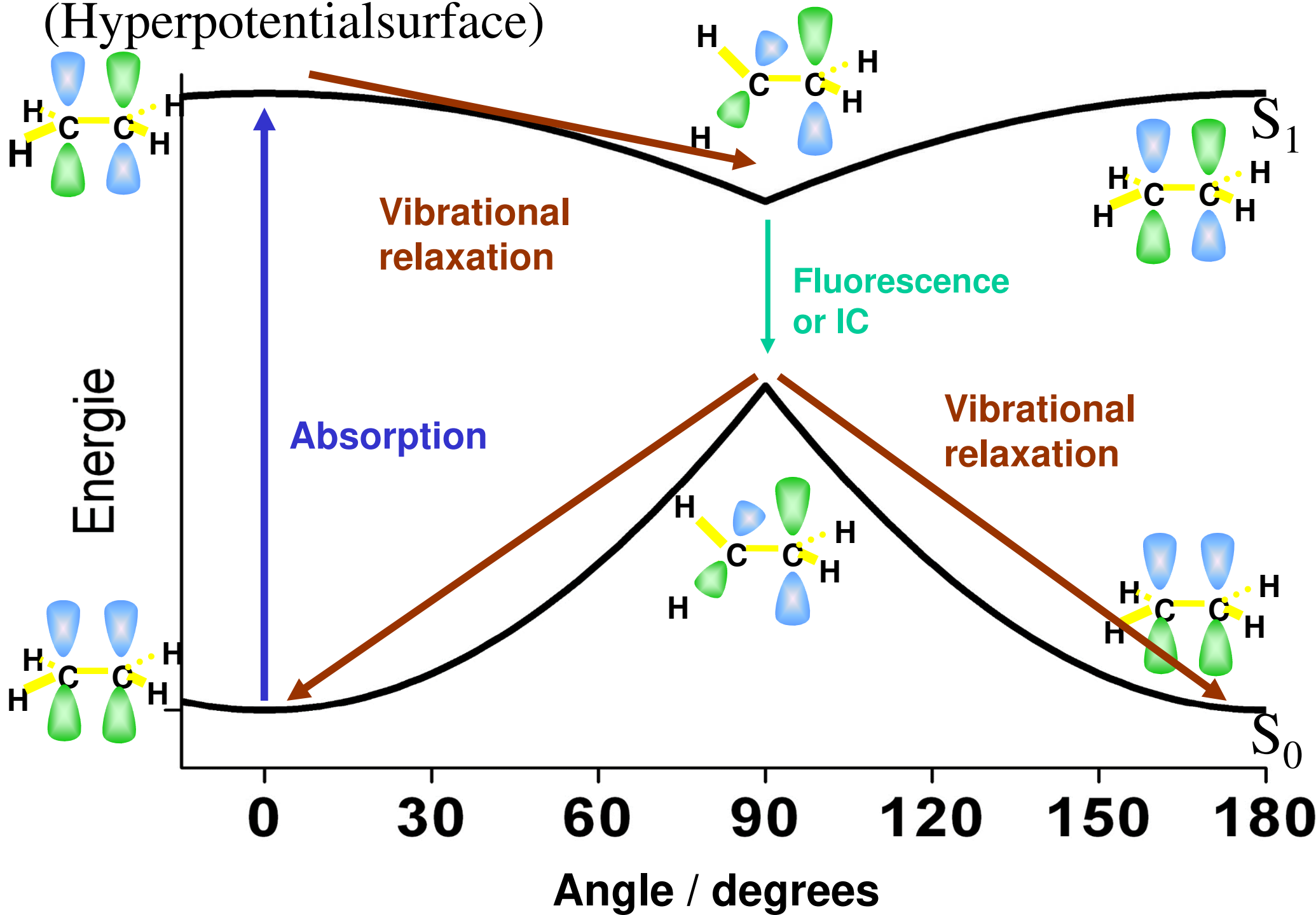
LUMO



HOMO

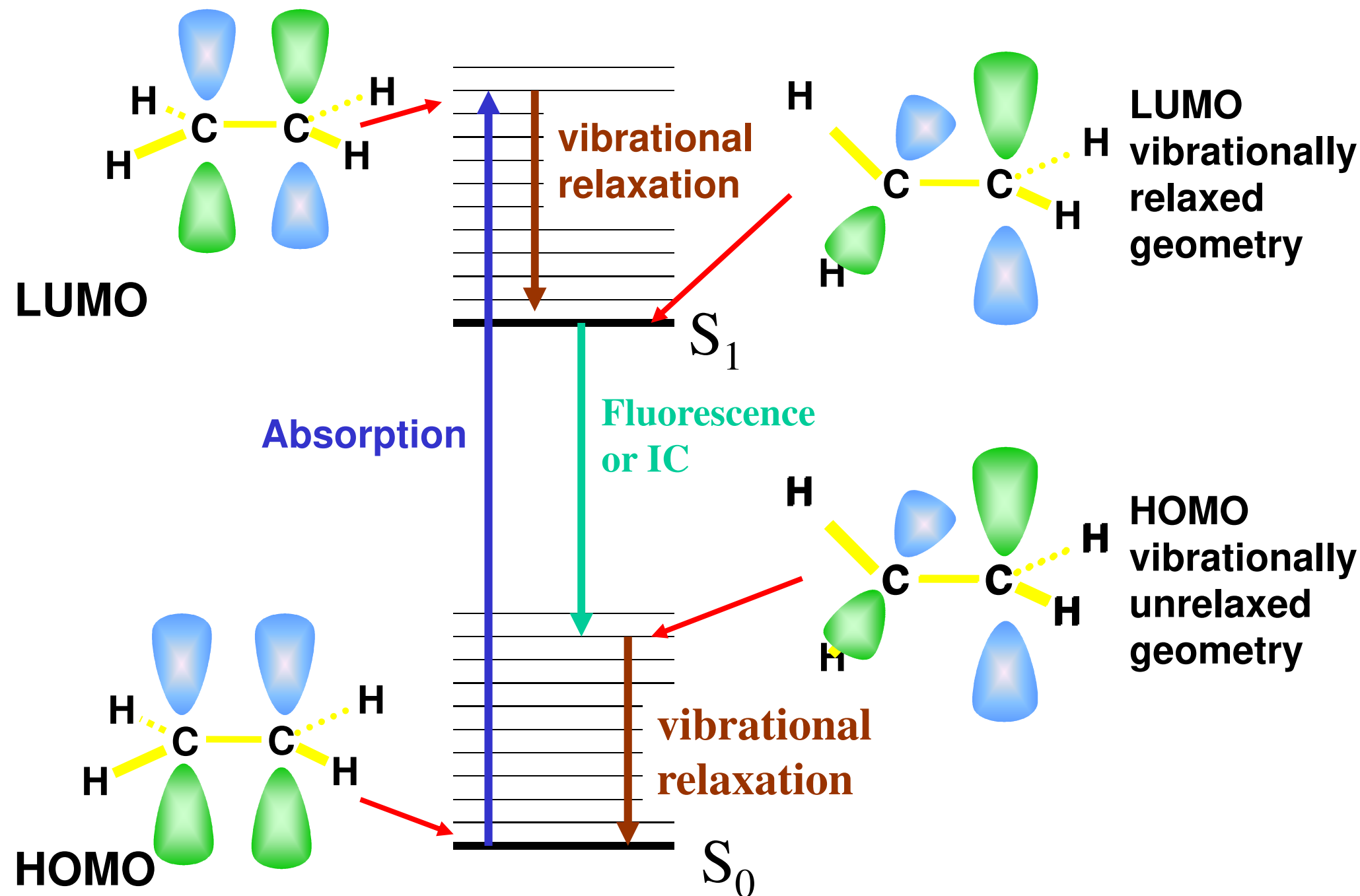


Mechanism of cis –trans Isomerisation of Ethene (Hyperpotential surface)



Mechanism of the cis –trans isomerisation

Example: ethene (Jablonski-Diagramm)



Cis –trans Isomerization as a decisive process in light detection

