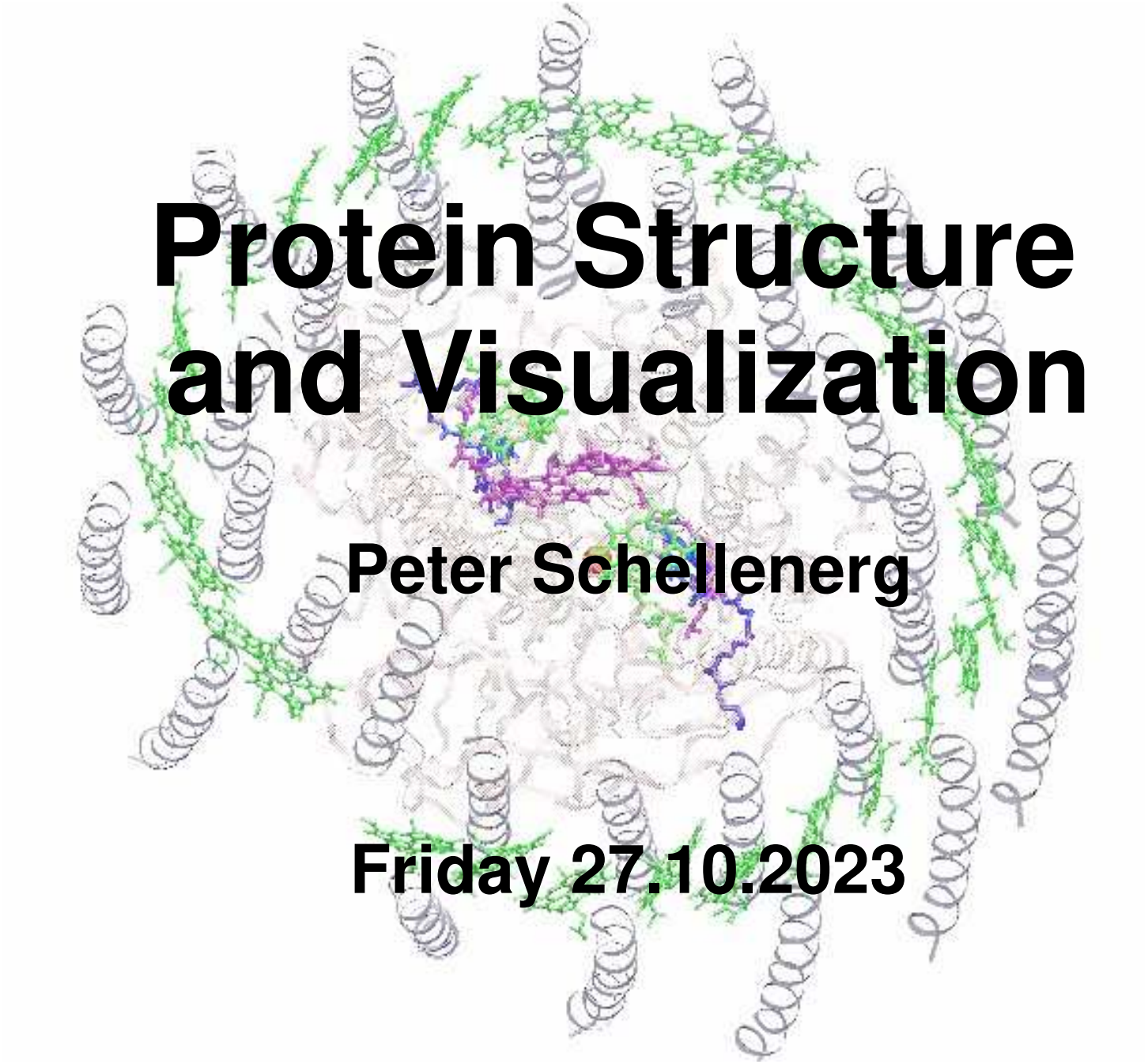


Protein Structure and Visualization

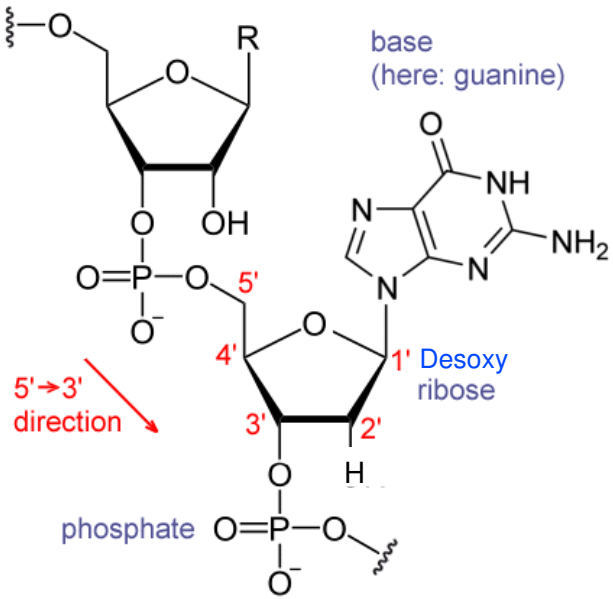
Peter Schellenerg

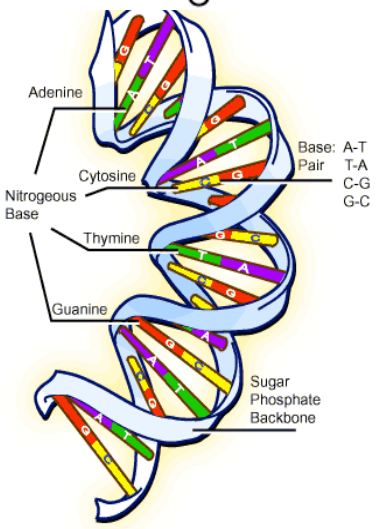
Friday 27.10.2023

The Molecular basis of nature

Information Storage

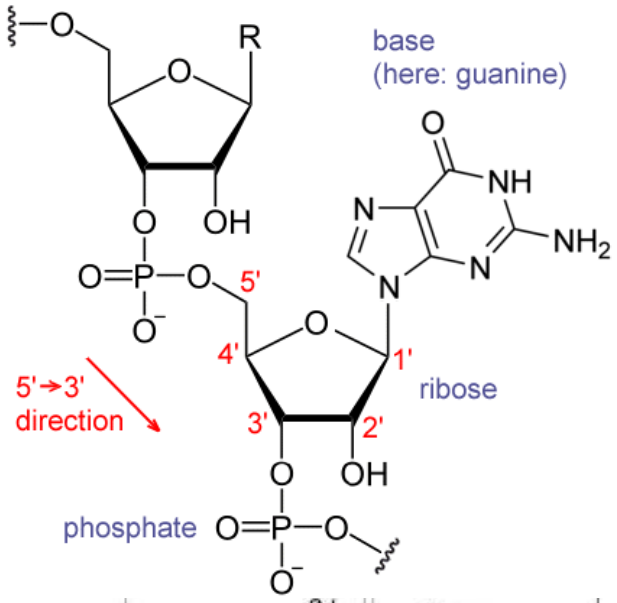
Deoxyribonucleic Acid

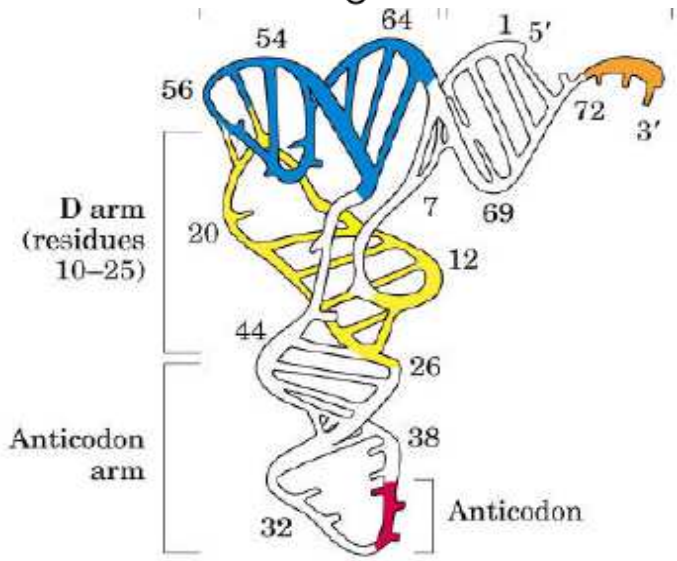




Information Transfer

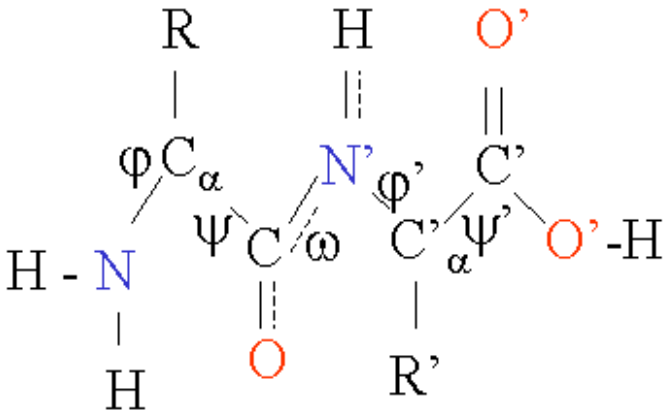
Ribonucleic Acid

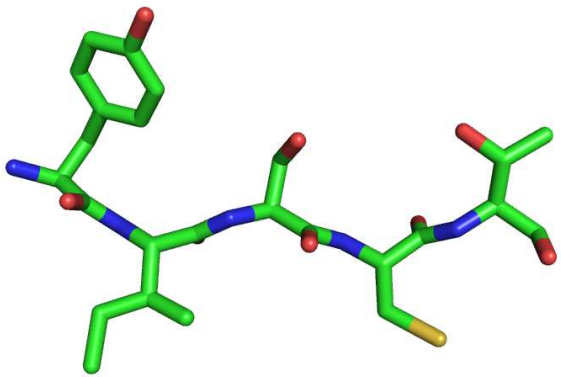




Functional Units

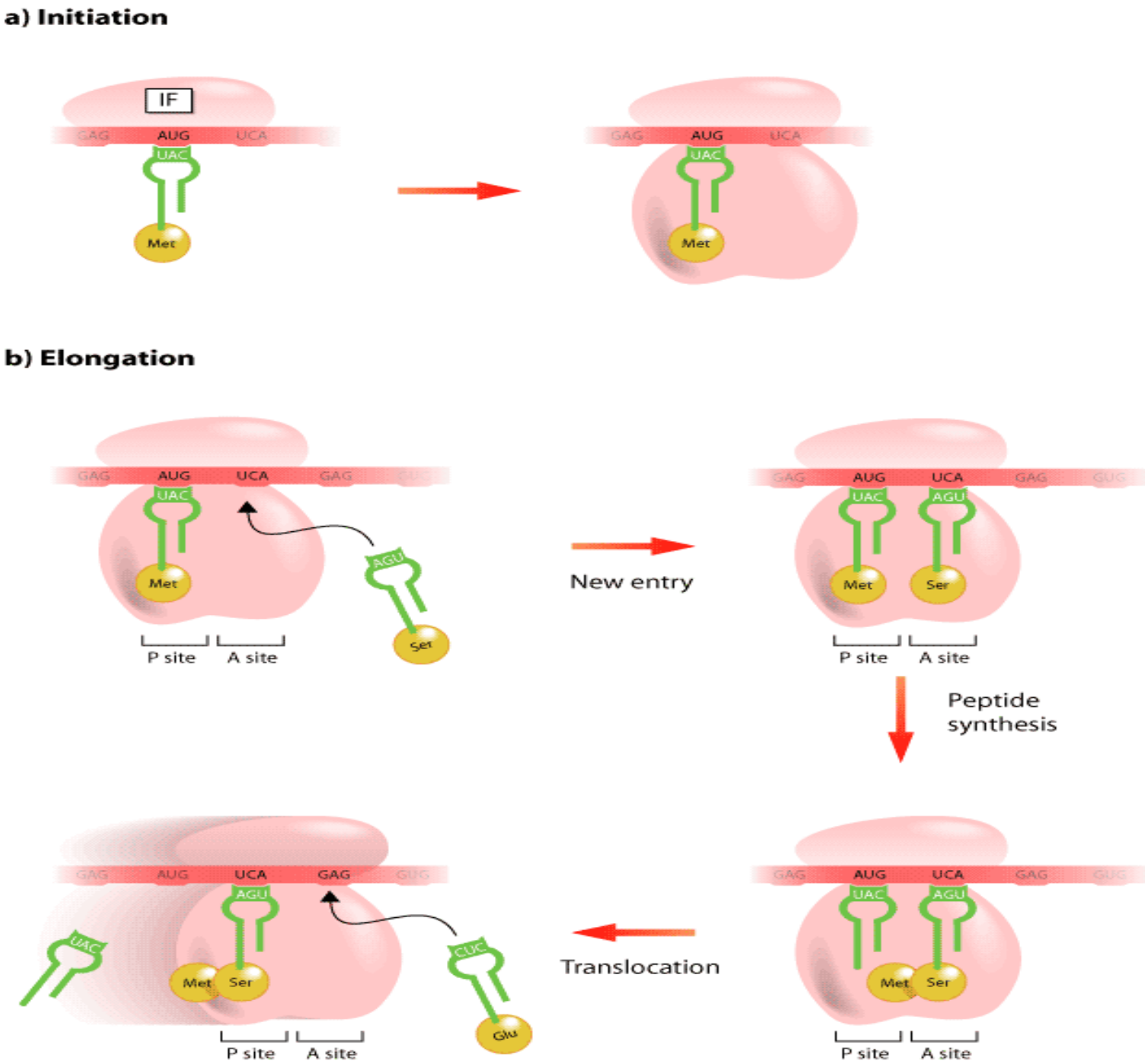
Proteins





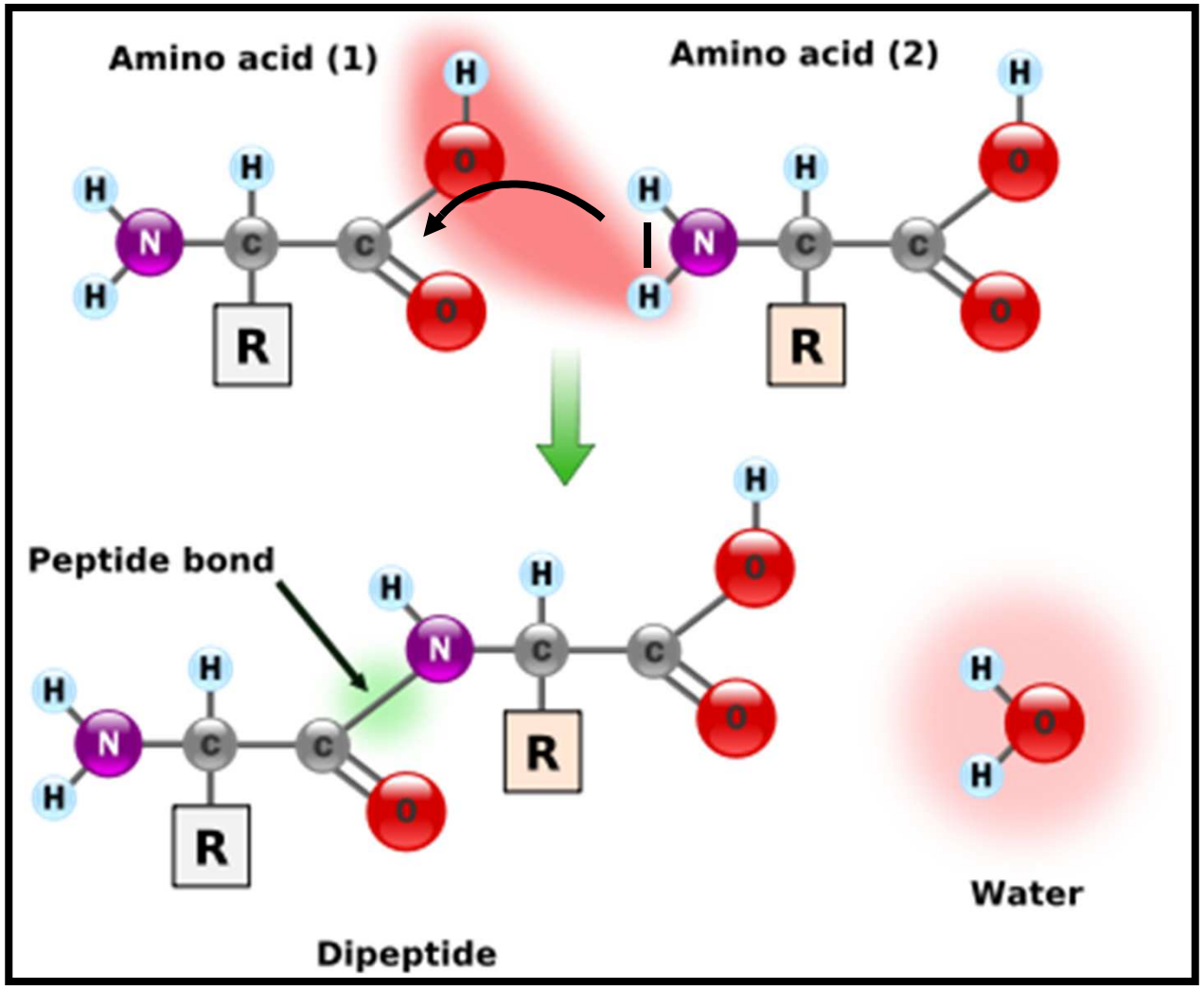
Here: TYR-ILE-SER-CYS-THR

Translation from RNA to Protein

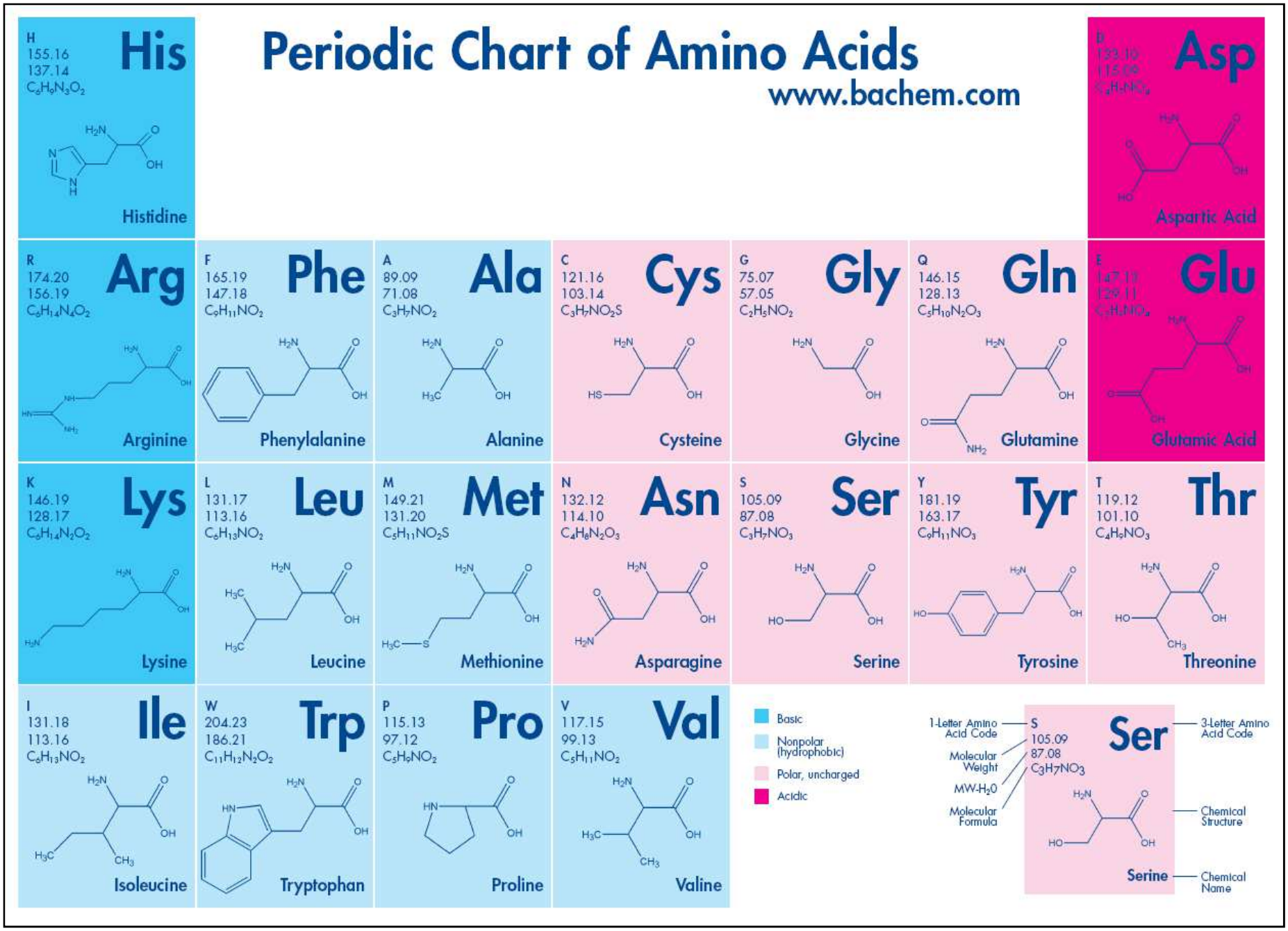


Peptide Bond Formation

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



Natural amino acid classification

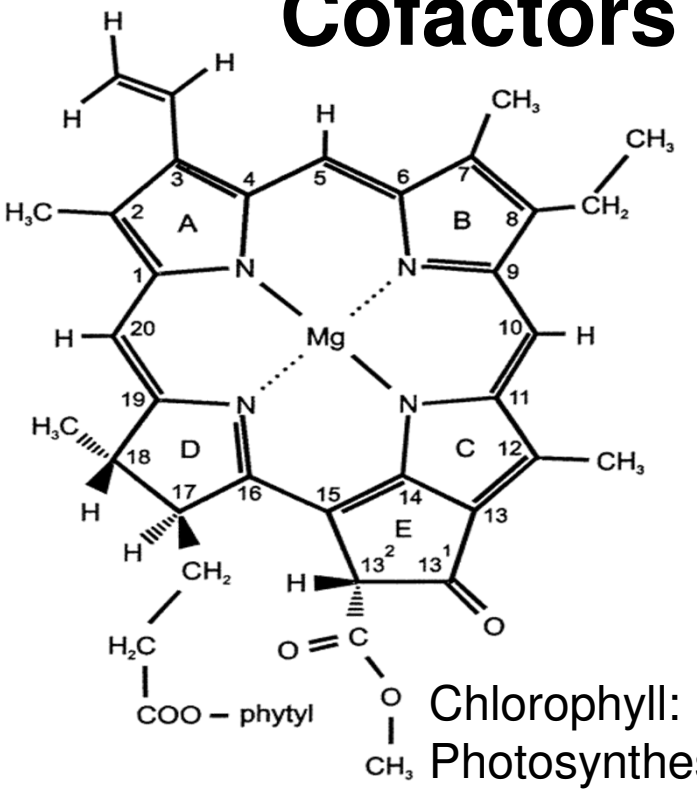


If Amino Acids can not do the task required, additional non-coded Cofactors are employed

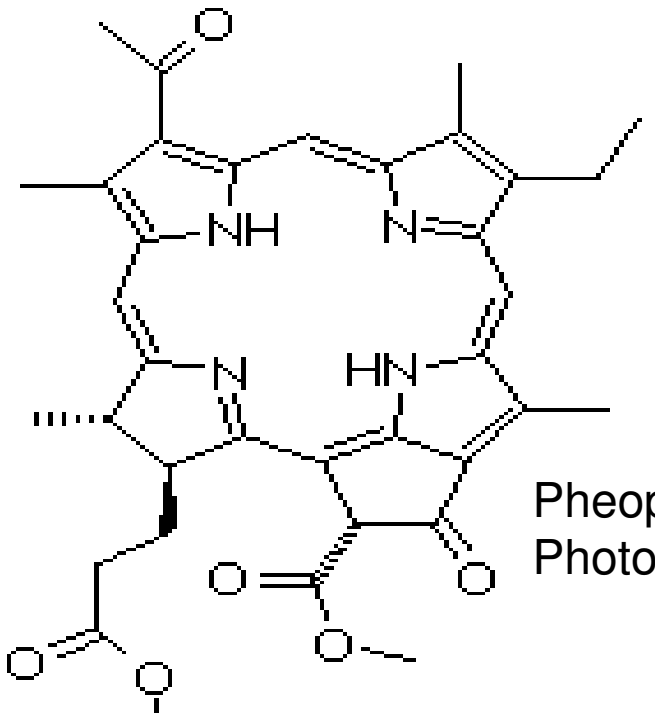
In many cases these Cofactors are Chromophores and result in Chromoproteins

Light reception	Photosynthesis	Miscellenous
Polyene: -Carotenoids -Rhodopsin	Cyclic Tetrapyrolls -Chlorophylls -Pheophytins	Light absorption not primary function!
linear Tetrapyrolls	Open Tetrapyrolls	Hems Flavins Metalloproteins
Cinnamic acid	Carotenoids	GFP, XFP
Flavins		

Cofactors for light absorption –type I



Chlorophyll:
Photosynthesis



Pheophytin:
Photosynthesis

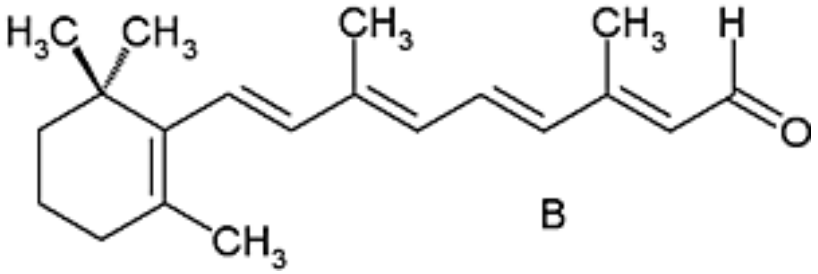
Large Frank Condon factor for 0-0 transitions:

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

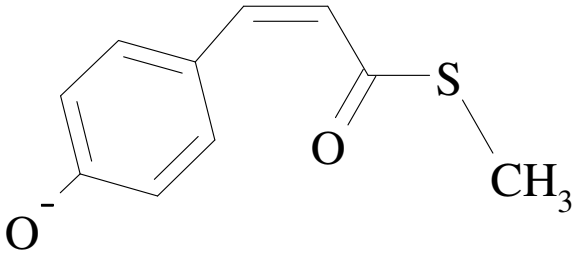
Therefore: small energy losses in the excitation process

very suitable for light to energy conversion !

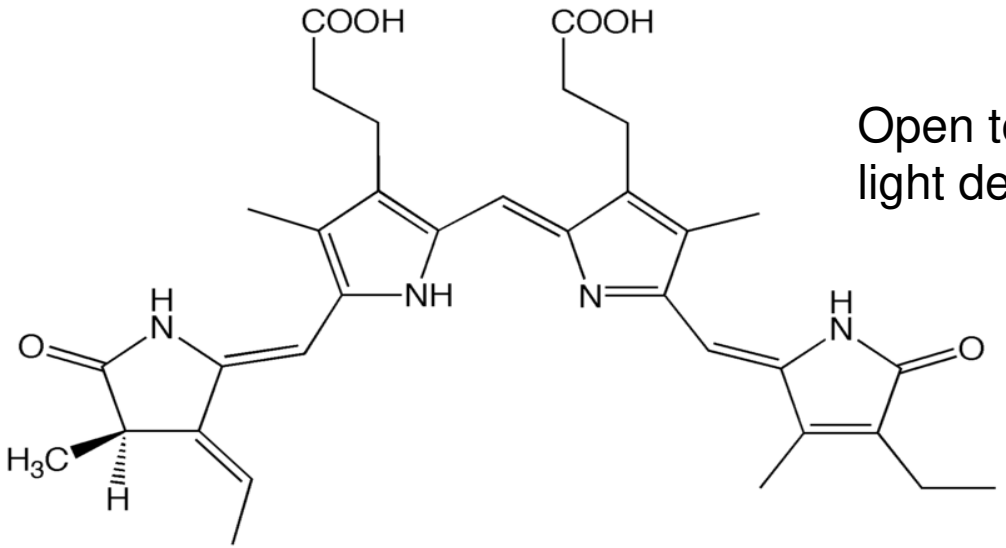
Cofactors for light absorption –type II



Retinal: light detection, photosynthesis,
Singulet-Oxygen Quencher



Photoactive Yellow Protein (PYP)
Light detection



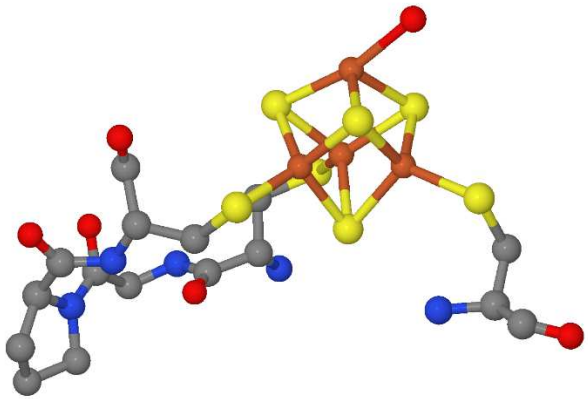
Open tetrapyrolls:
light detection, Photosynthesis

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

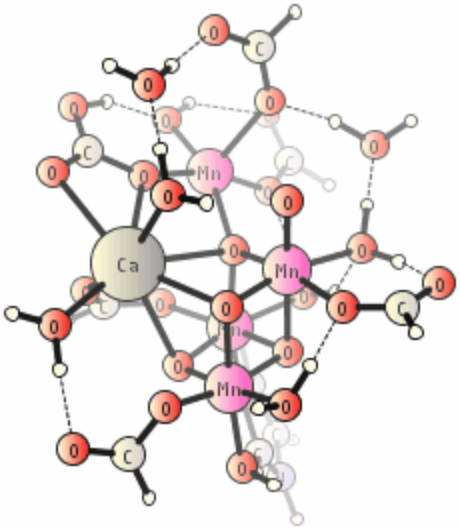
-strong electron phonon coupling
- Large rearrangement of molecule and solvent (protein) cage upon excitation

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

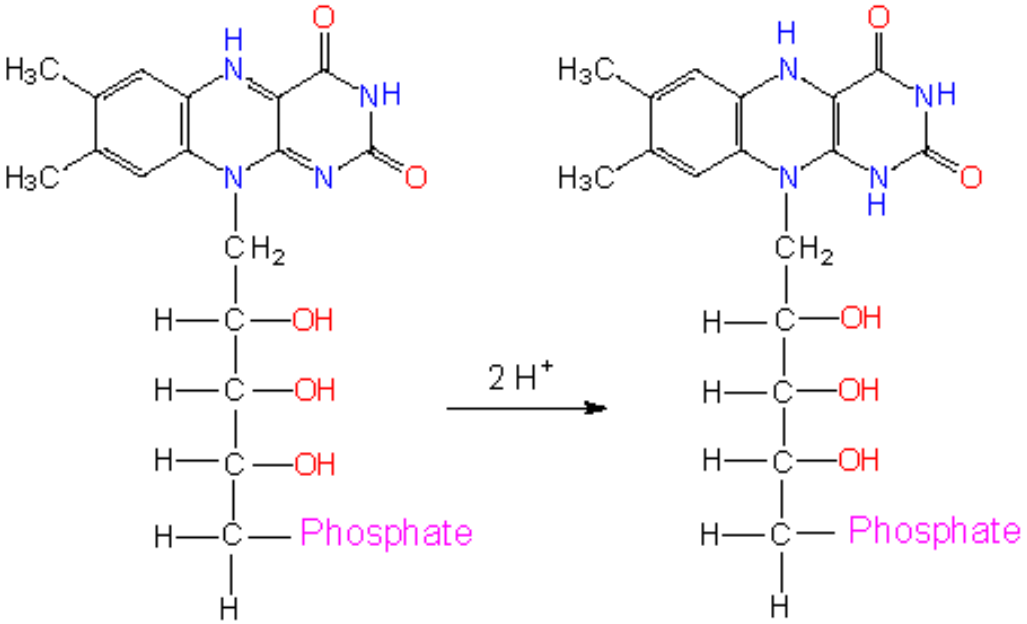
Cofactors for Redox Reactions



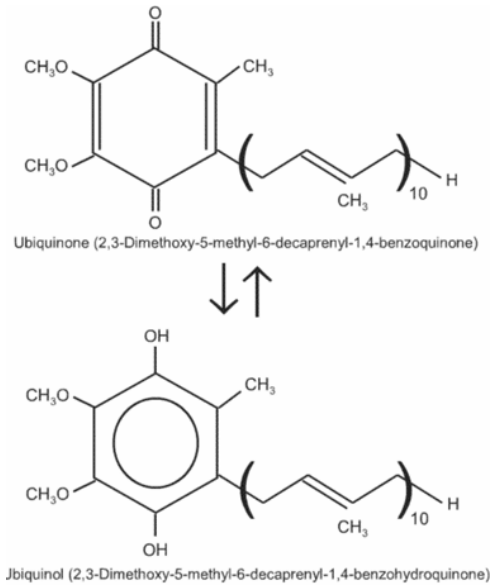
Iron-Sulfur-Cluster (4,4-Cluster



Mn-O-Ca-Cluster (PSII)

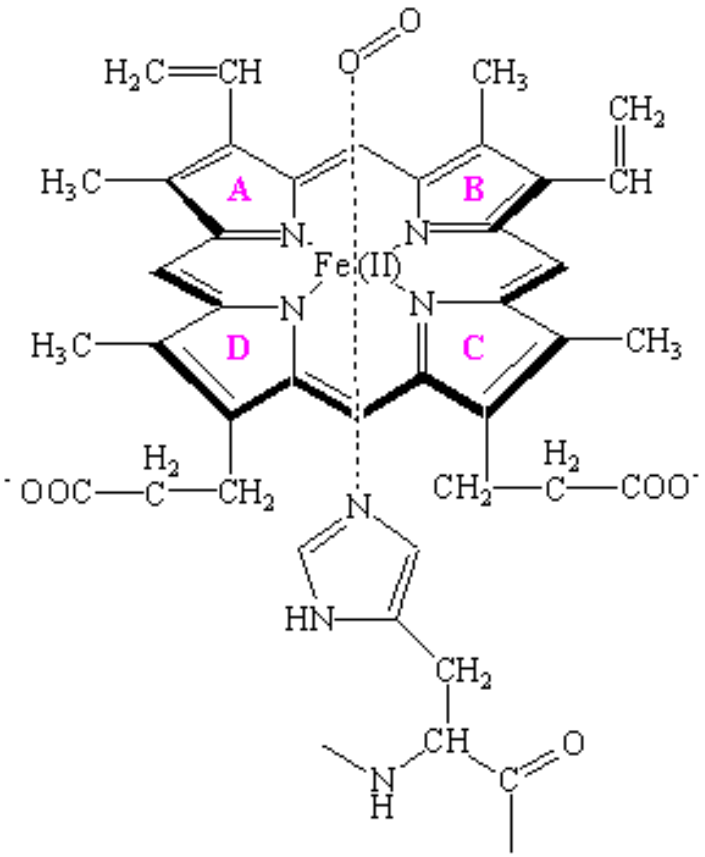


Riboflavin

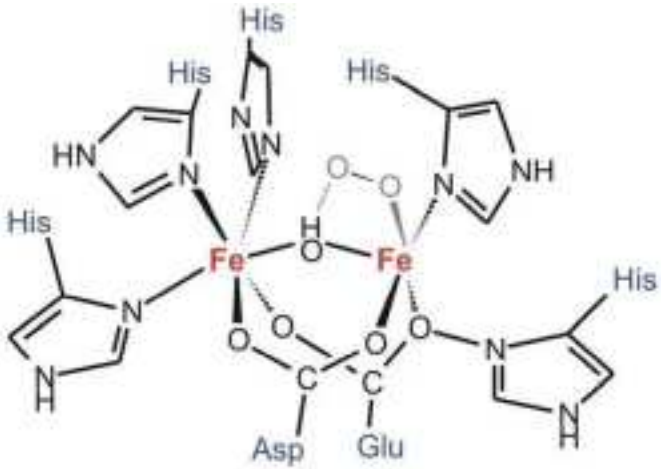


Ubichinone

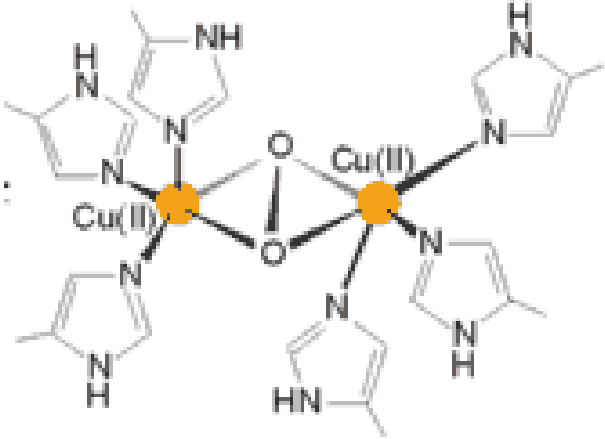
Cofactors for oxygen binding



HEM



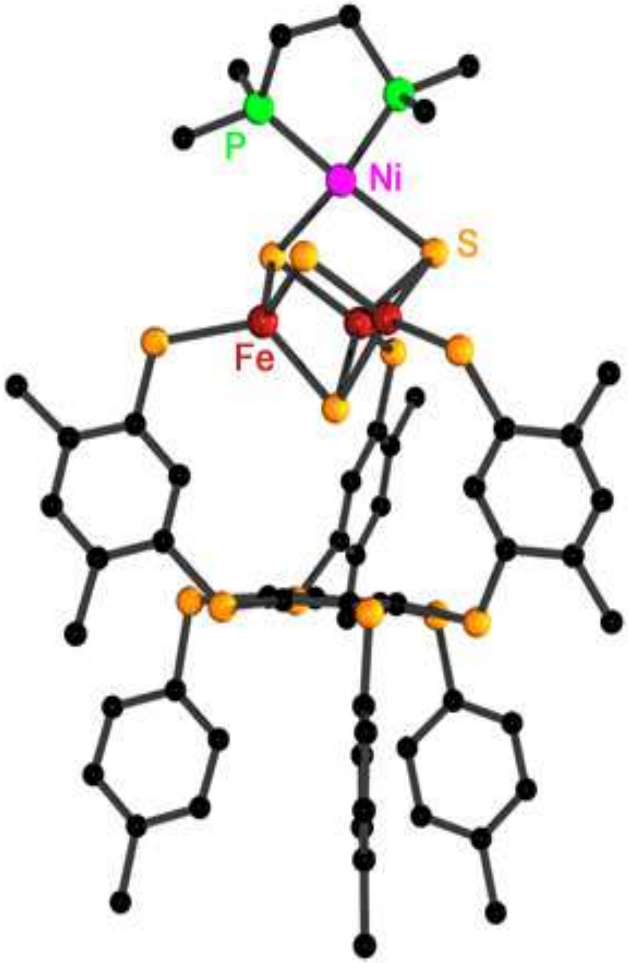
Hemerythrin
(marine invertebrates,
some worms, corals ....)



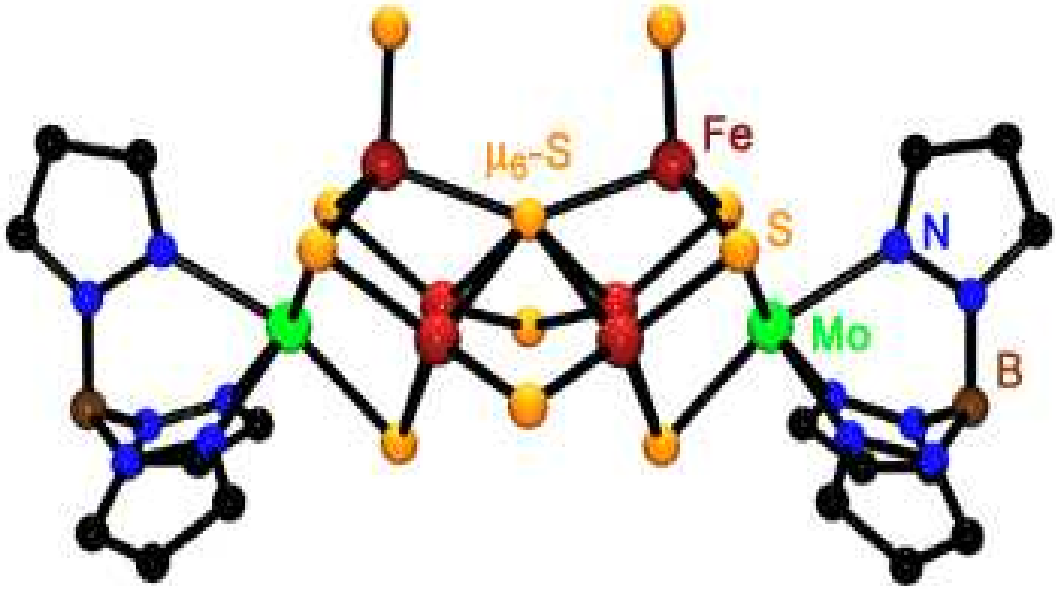
Hemocyanin
(from Squids, spiders
Athropods, Mr. Spock ....)

Cofactors for Binding and Catalysis / Formation of Complex Compounds

Carbon monoxide dehydrogenases (CODH) catalyze the reaction $CO + H_2O \rightarrow CO_2 + 2H^+ + 2e^-$.



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

$C_{774}\ H_{1224}\ N_{210}\ O_{222}\ S_5\ Fe$

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)

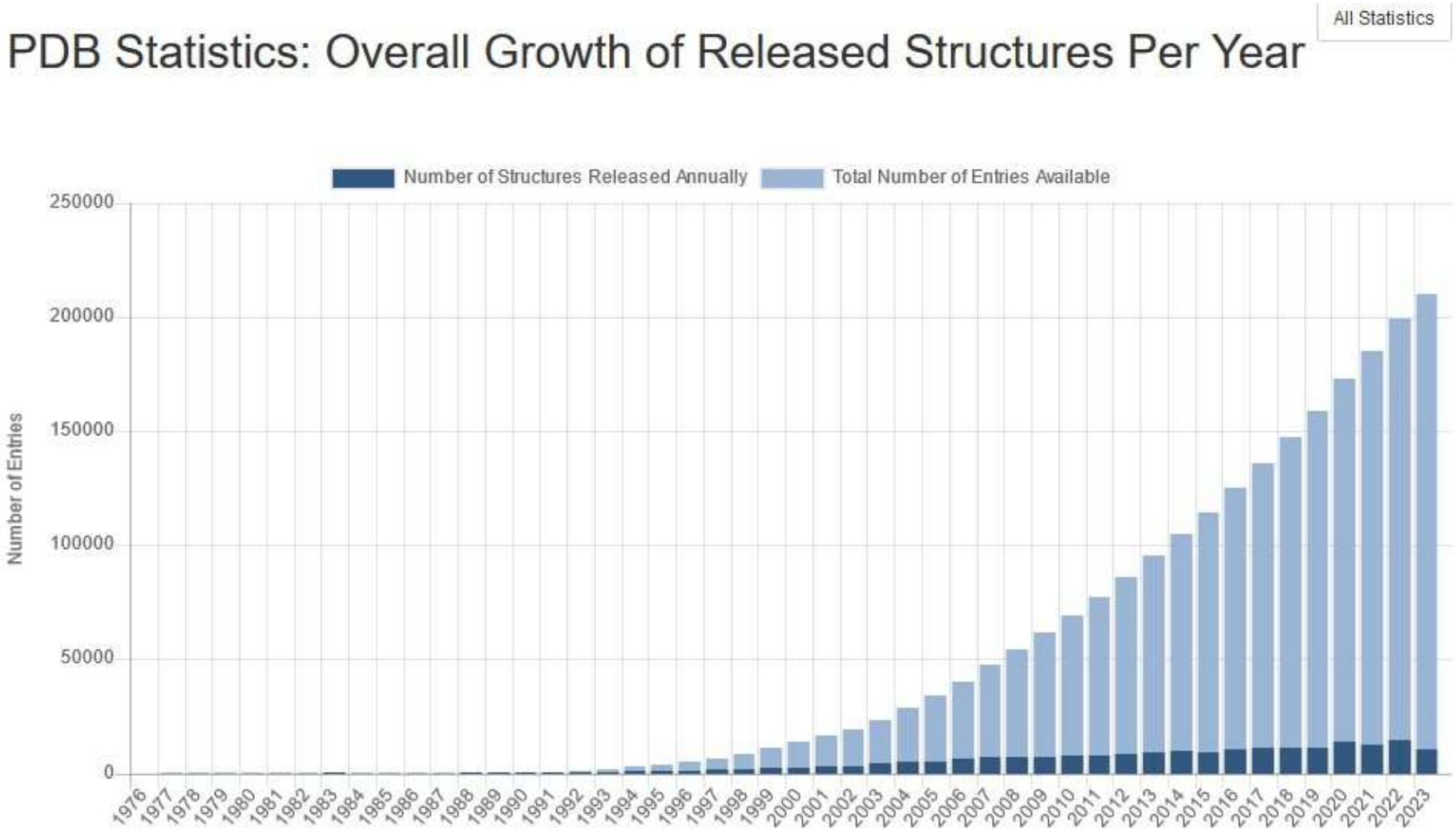
Happy Birthday PDB! Today 20th of Oct. is the 52nd anniversary of the founding of the Protein Data Bank

- PDB: publicly available archive of macromolecular structure data
  >200000 deposited structures !

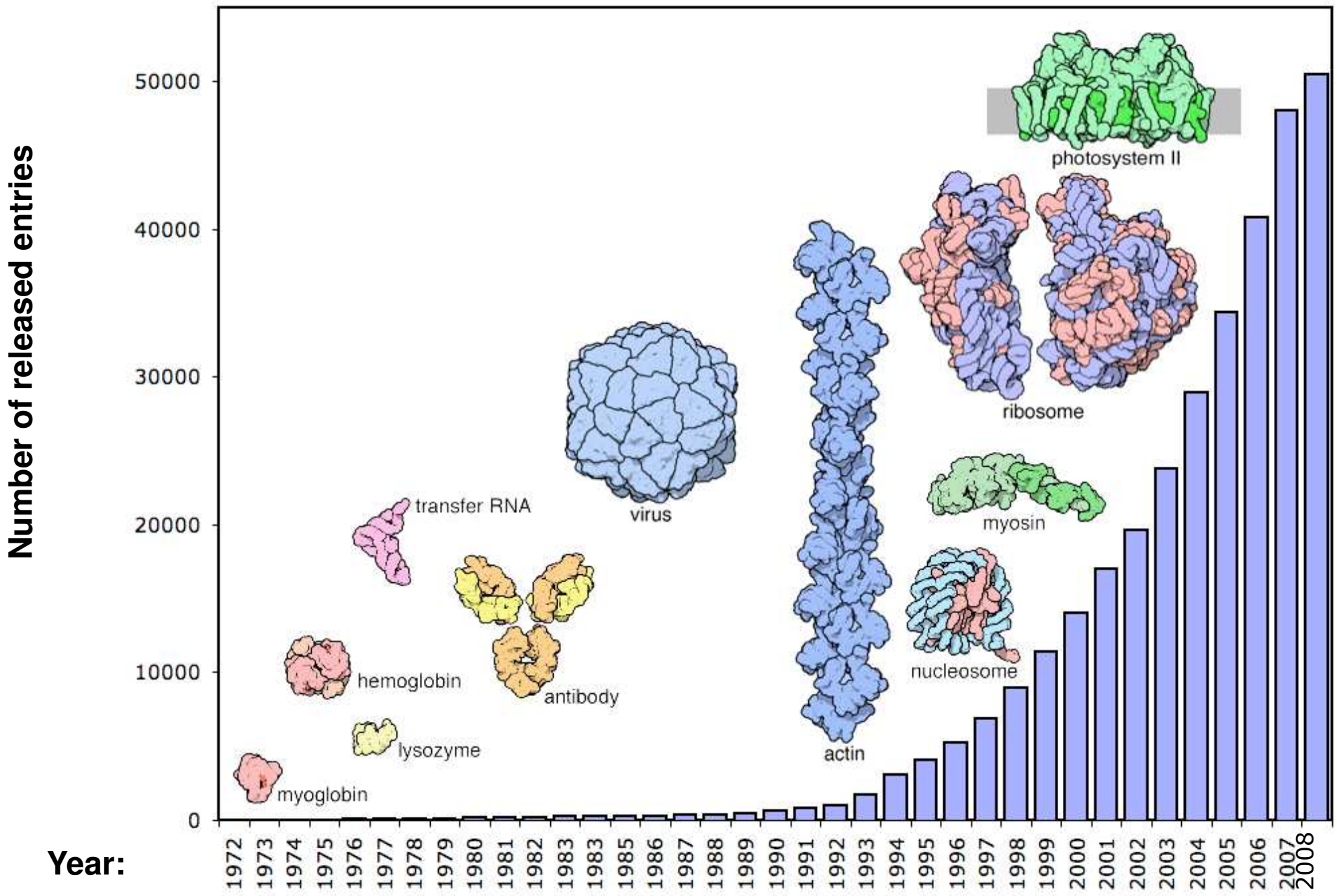
- Link: **http://www.rcsb.org**

- 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).

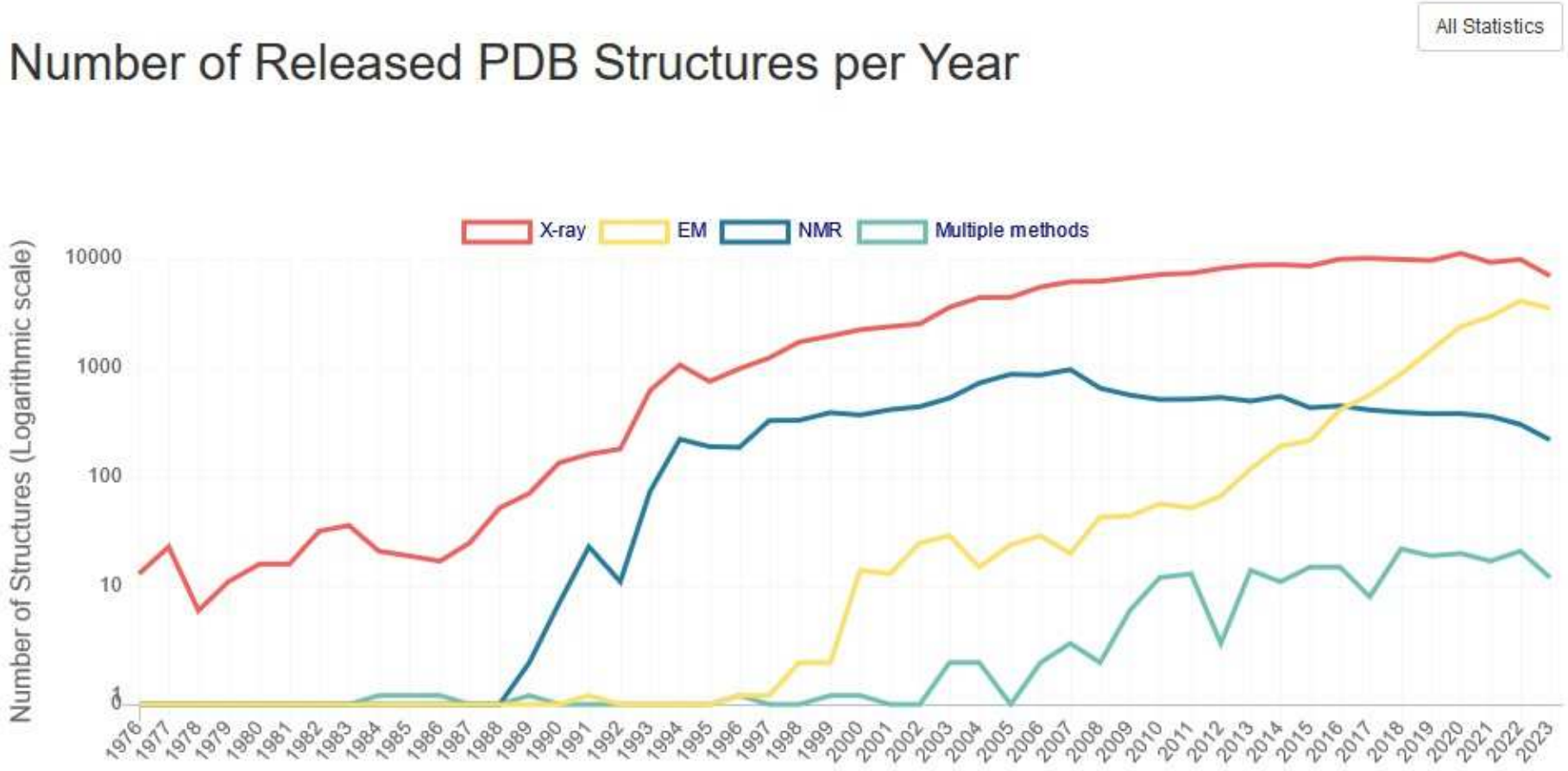
Deposited structures in the Protein Data Bank



Also increase in complexity of the determined structures:



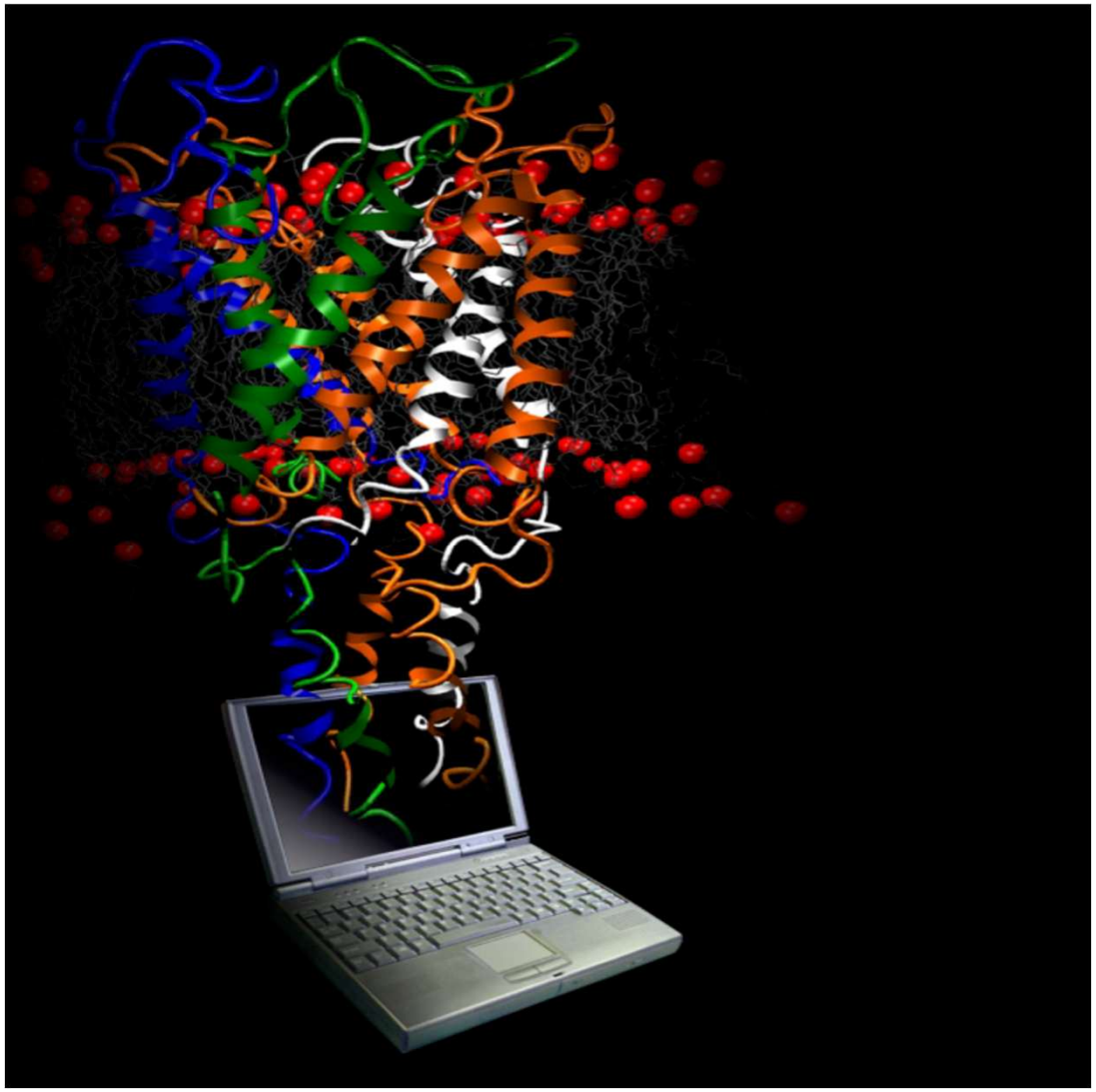
Methods utilized for structure determination



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

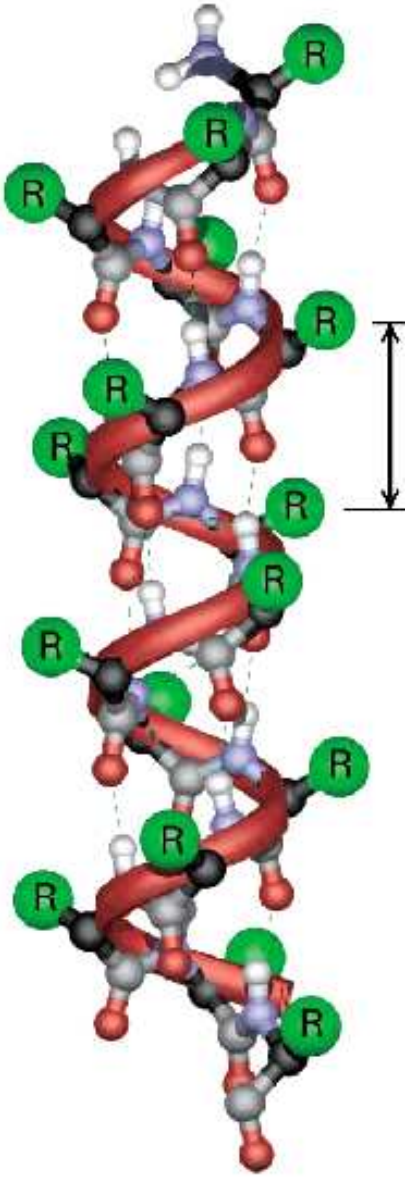


Computer:

Rasmol (historical)
Pymol
UCSF Chimera
Swiss PDB -viewer
VMD

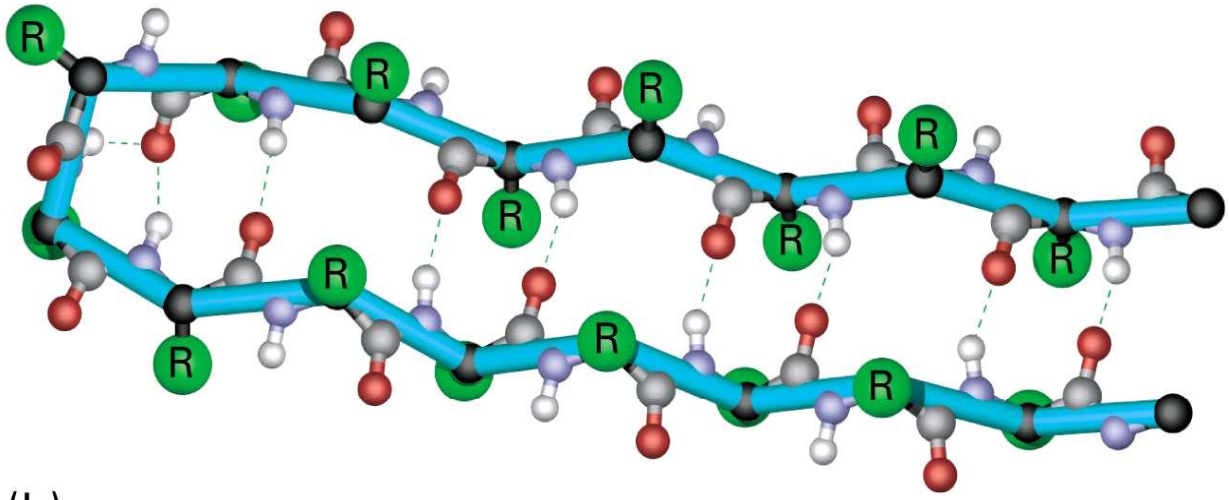
Browser:
Jmol
Mol* 3D Viewer
Chime (historical)

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

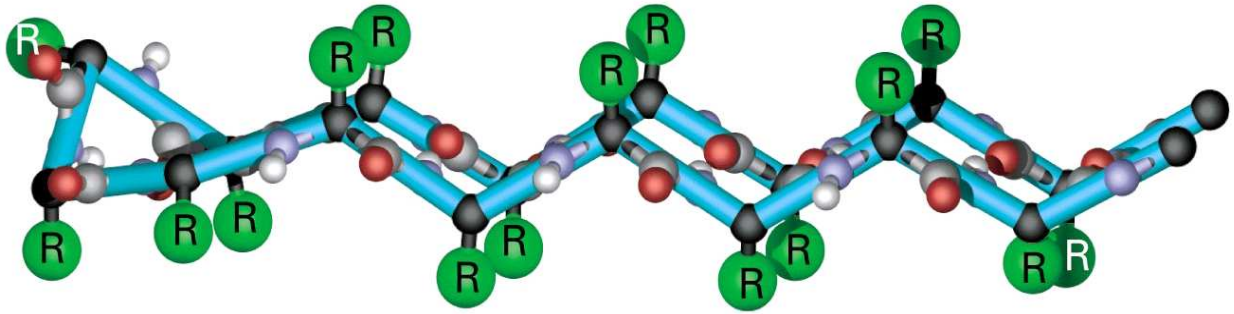


3.6 residues
0.54 nm

(a)

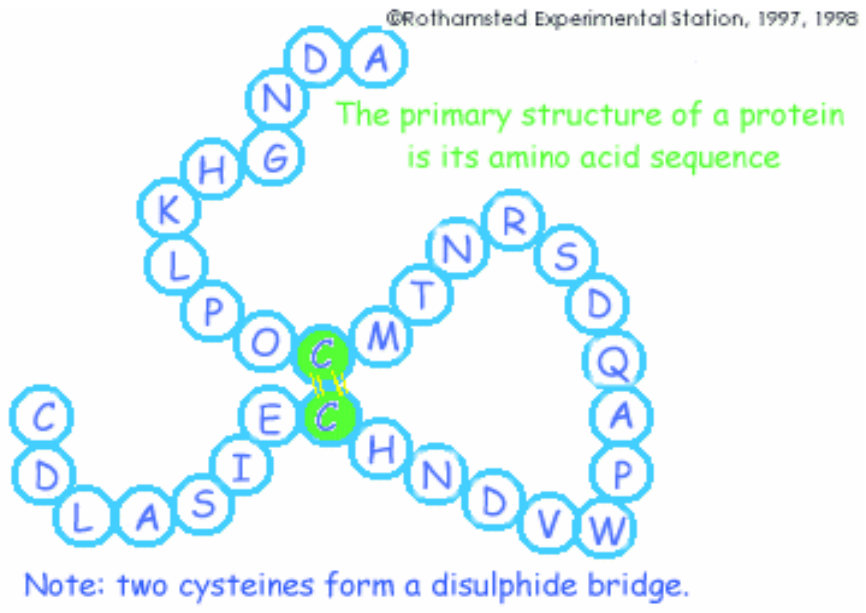


(b)

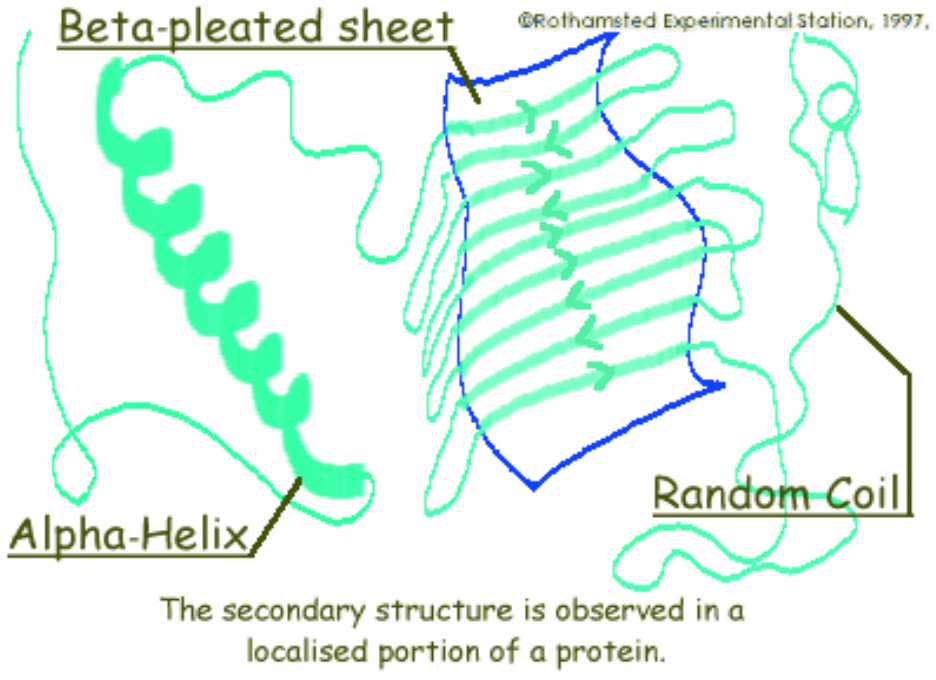


Hierachy of Protein Structure Features

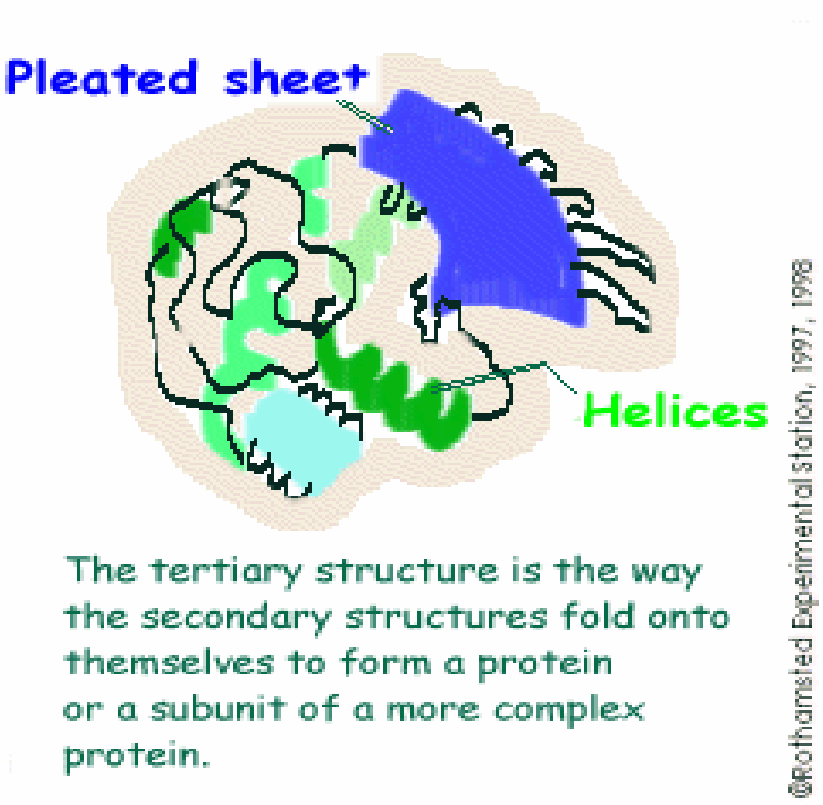
Primary structure



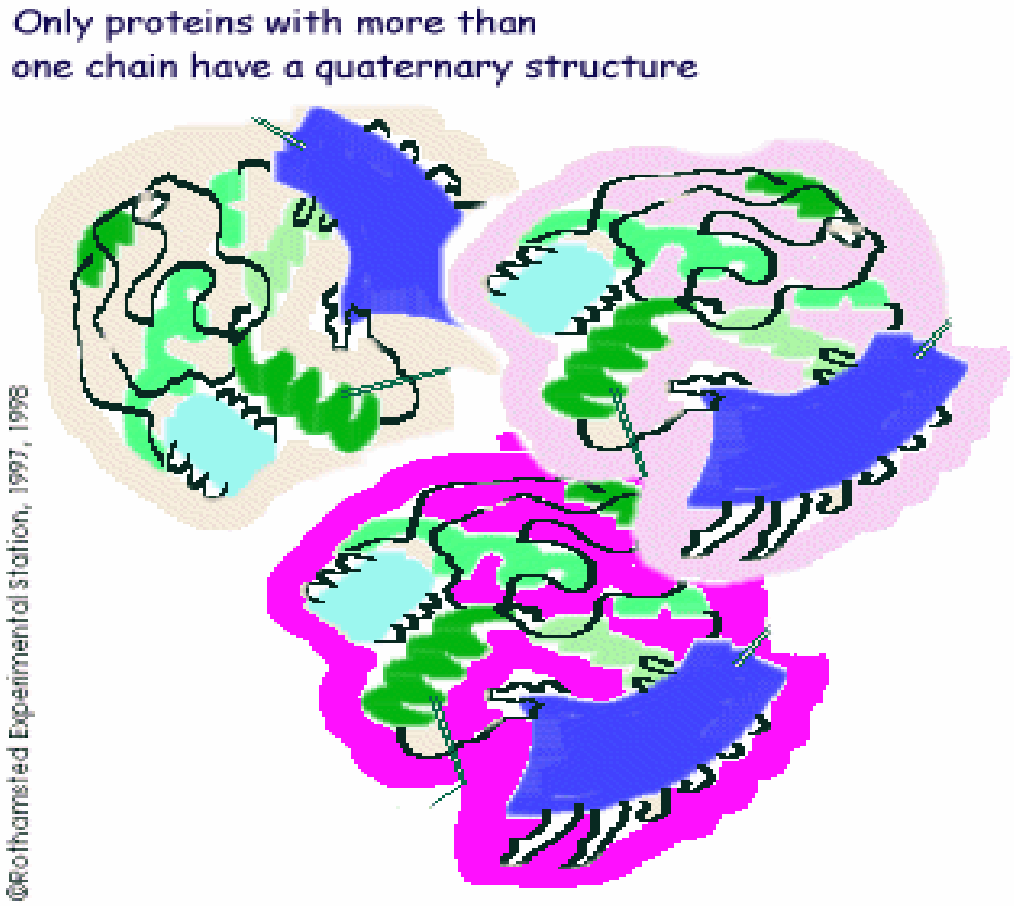
Secundary structure



Tertiary structure



Quaternary structure

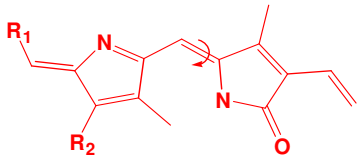
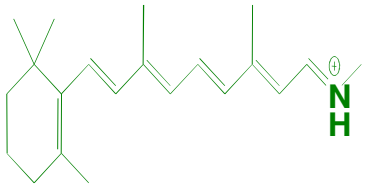
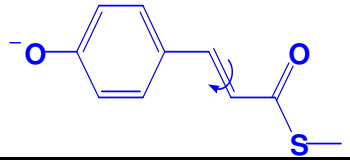
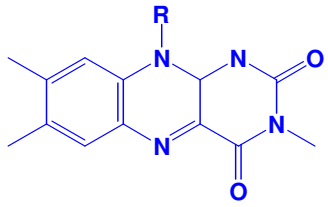


Additonal possible discussion

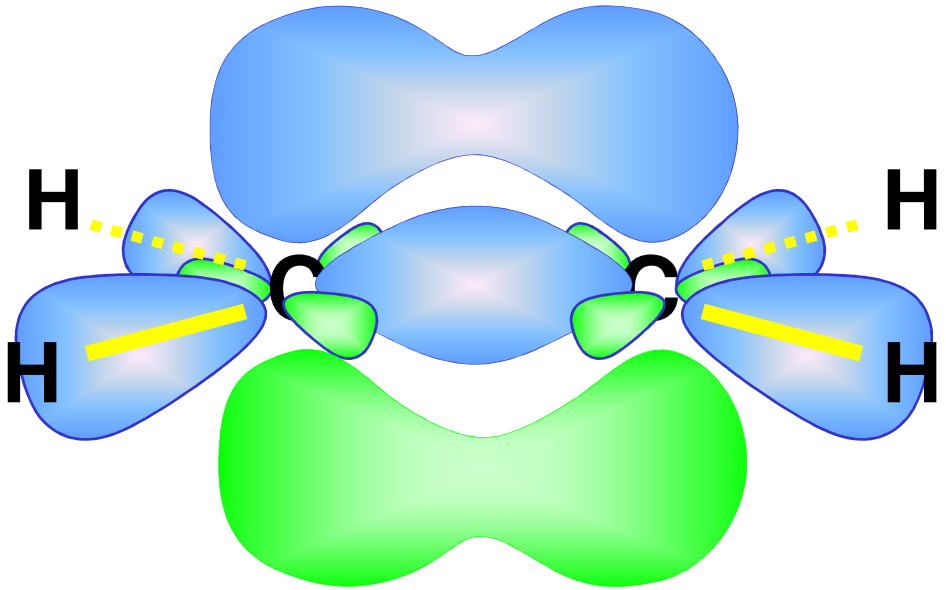
Molecular mechanism of cis –trans Isomerization in solution

-Orbitals , hyperpotential surfaces and Jablonski -Diagram

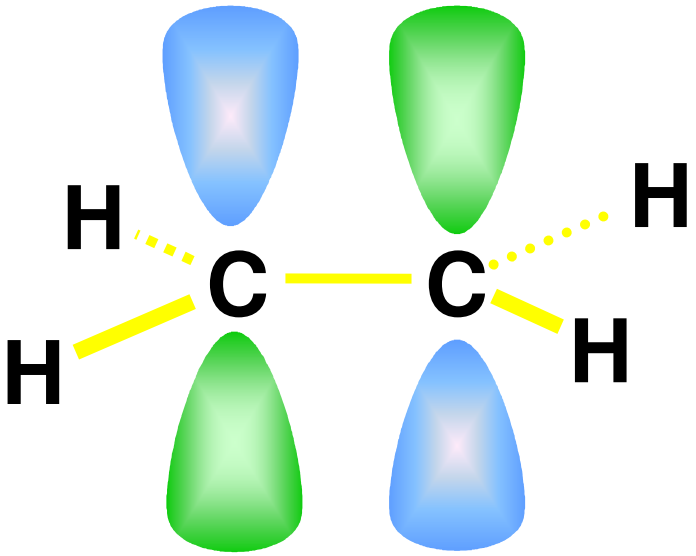
PHOTOSENSORY PROTEIN FAMILIES & their photochemistry:

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

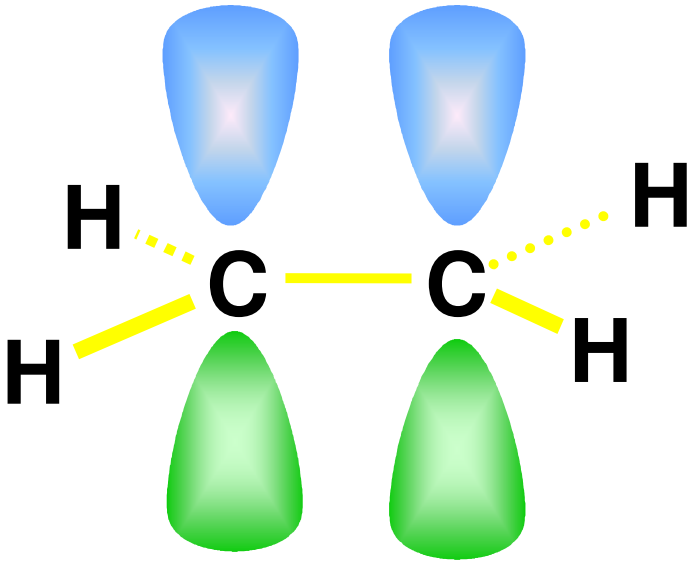
MO Schema von Ethene



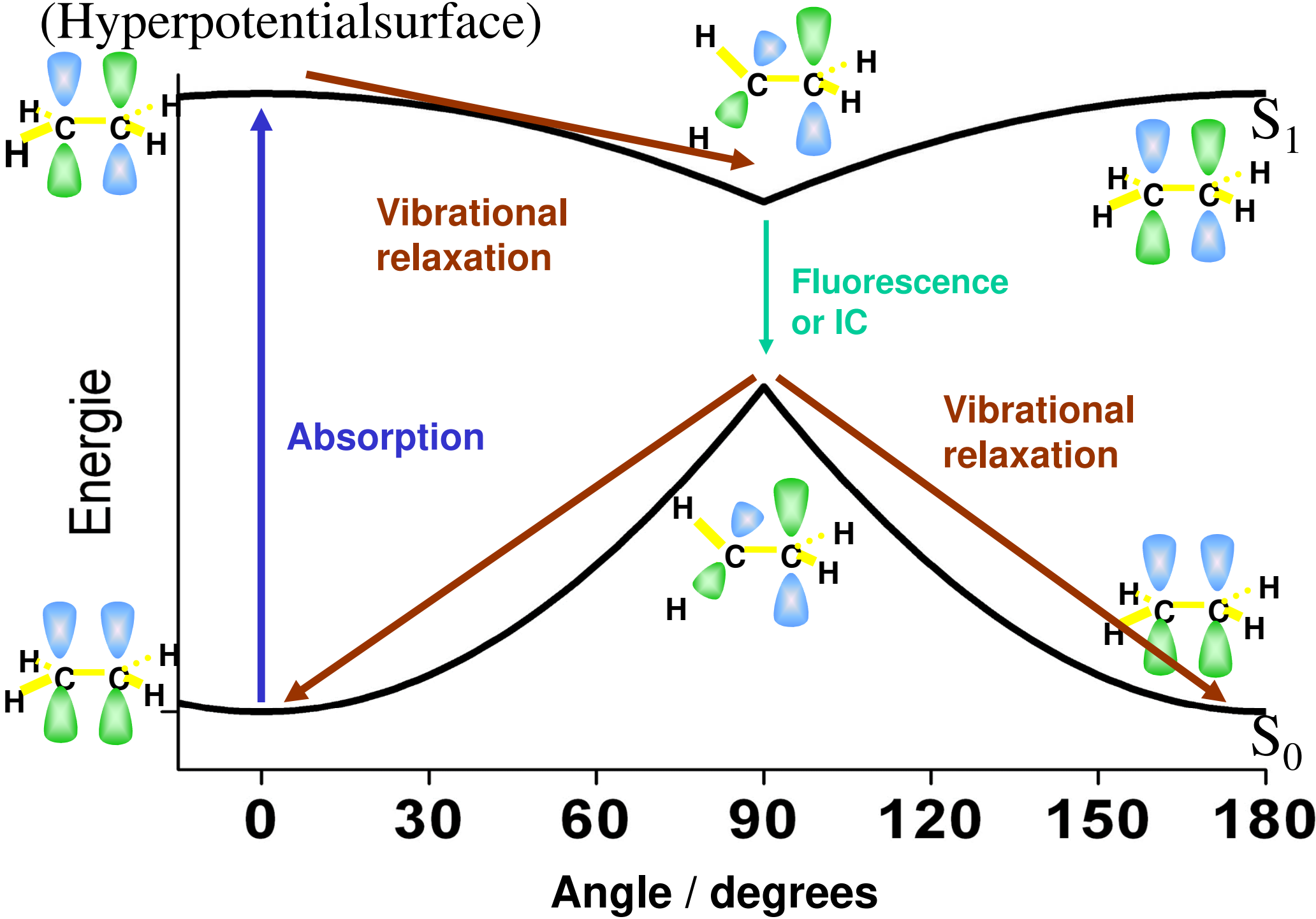
LUMO



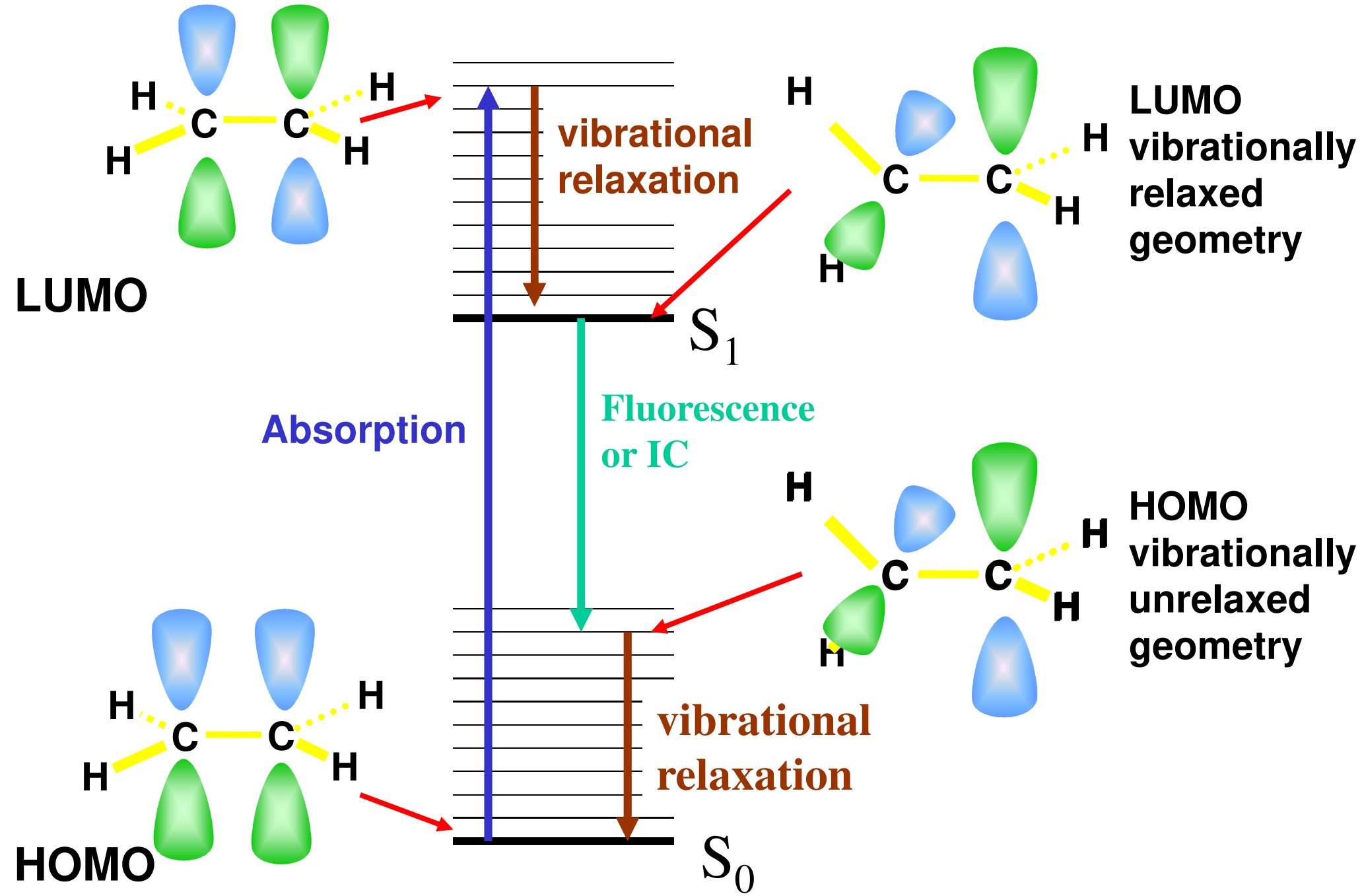
HOMO



Mechanism of cis –trans Isomerisation of Ethene (Hyperpotentialsurface)



Mechanism of the cis –trans isomerisation
Example: ethene (Jablonski-Diagramm)



Cis –trans Isomerization as a decisive process in light detection