Introduction to Biochromophores and their environment

Xabier Lopez

Contents

i) Absorption in the visible region by **phrosthetic groups**:

- Basic Design
- π -> π^*
- Importance of the environment (protein):
  Photosynthesis and Vision

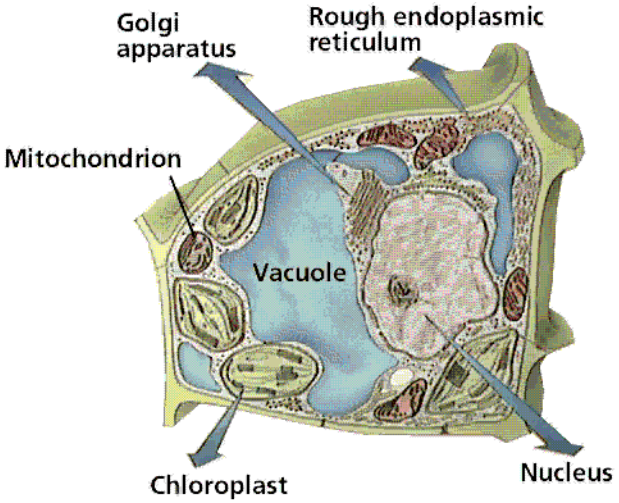
ii) Introduction to **Protein Structure**:

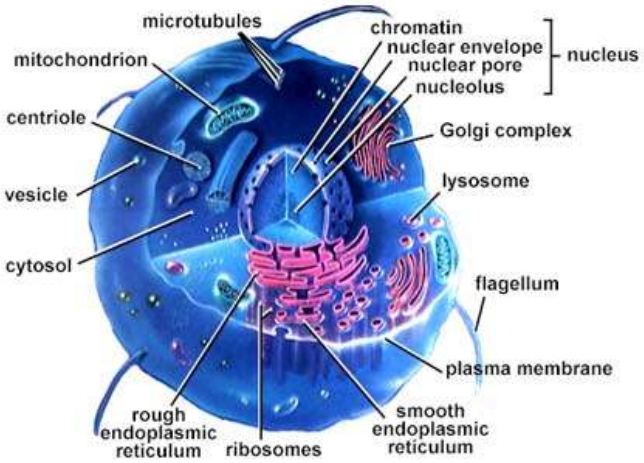
- From primary to quaternary structure
- Experimental Structure determination
- Protein as Chromophores
- Simulation of Proteins (FF's, MD and QM/MM)

iii) GFP/BFP as an **ilustrative example(s)**

Biochromophores:

Definition: Molecules that absorb light (near UV or visible) present in plant/animal cells.





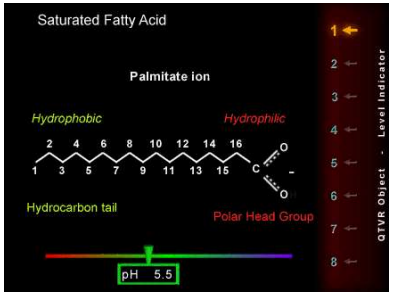
Motivation:

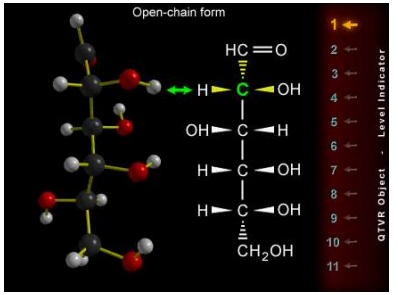
1) **Fundamental Processes** (Vision, Photosynthesis, Photoperiodism, Bioluminiscence,..., DNA Damage) are governed or triggered by photo absorption
2) **Analysis** of Biomolecular structures
3) Various technical **applications** (GFP)

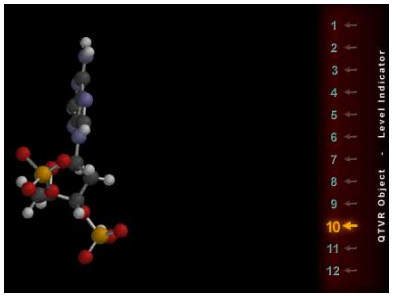
Main Biological Molecules

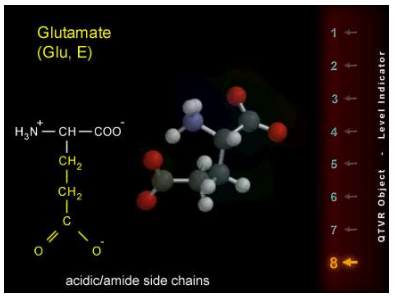
1. **Lipids** (Fatty Acids, Cholesterol, membrane
2. **Sugars** (mono/poly-saccharides)
3. **Nucleic Acids** (DNA, RNA) (UV)
4. **Proteins** (Aminoacid polymers) (UV)

Biochromophores that absorb in the visible range are formed by **Prosthetic** Groups bound **covalently** to a protein.

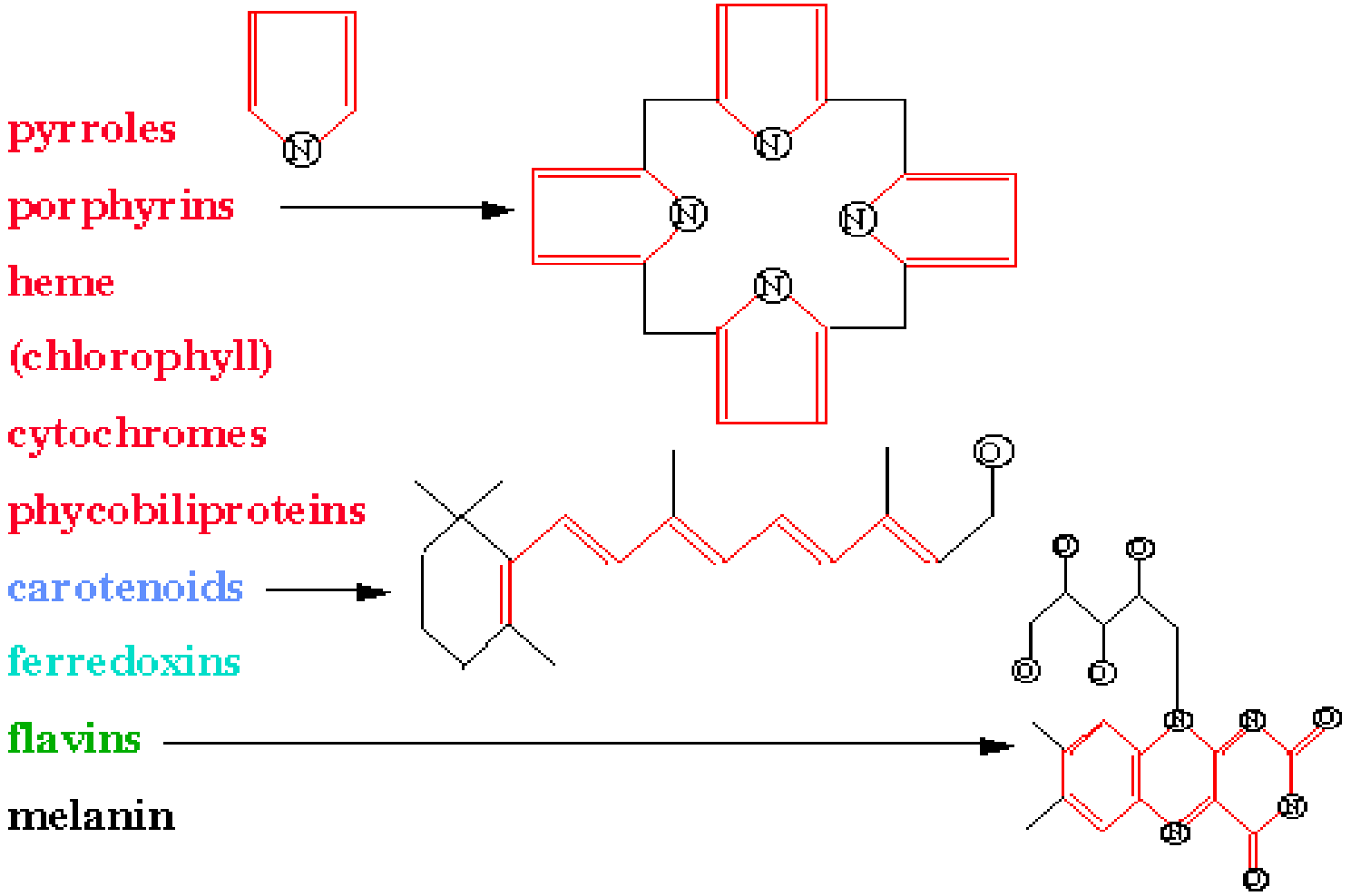








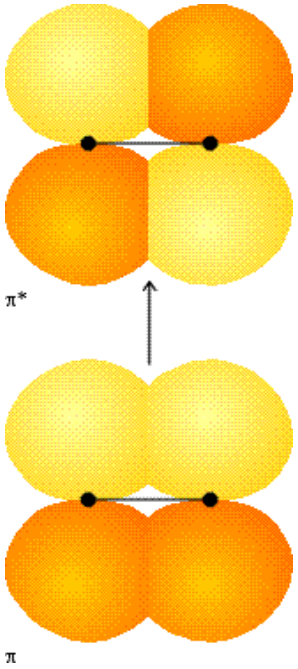
General Design of Prosthetic Chromophores (visible)



π –> π* Transitions

- Efficient transitions:

$$I = I_0 10^{-\epsilon(\nu)[J]l}$$



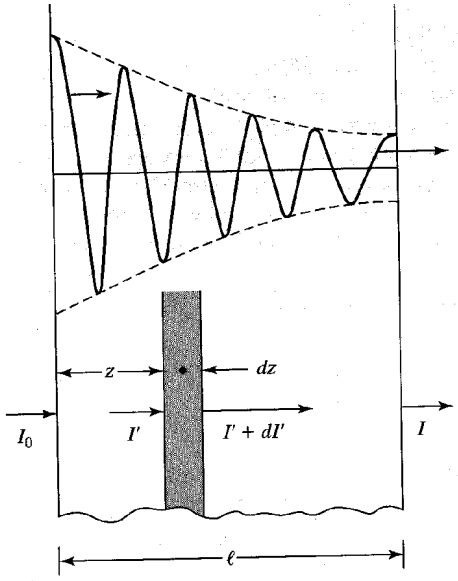


Table 17.2 Absorption characteristics of some groups and molecules

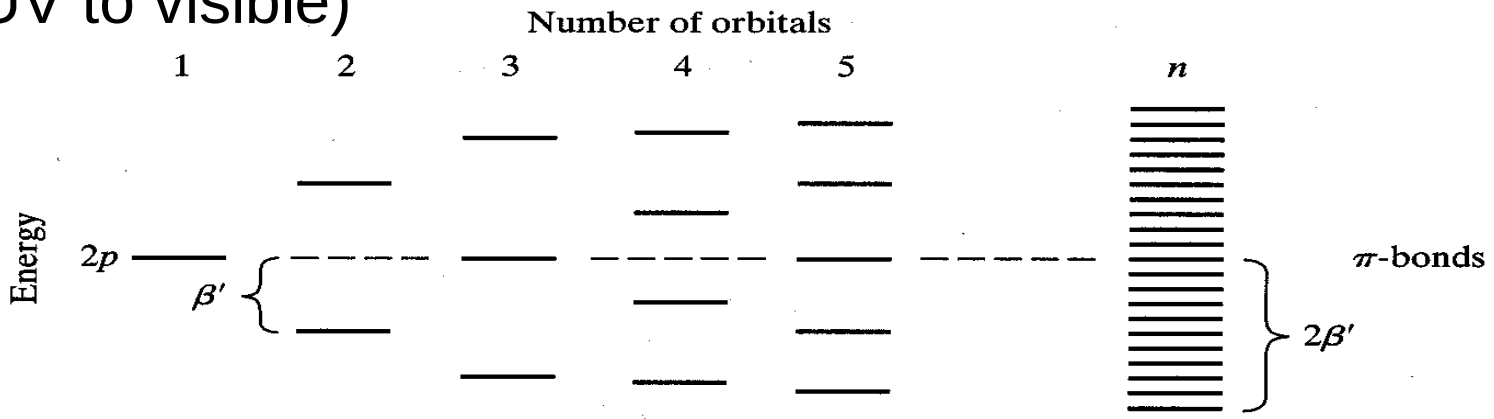
Group	$\tilde{\nu}_{max}/(10^4\ cm^{-1})$	λ_{max}/nm	$\varepsilon_{max}/(L\ mol^{-1}\ cm^{-1})$
C=C(π* ← π)	6.10	163	1.5×10^4
	5.73	174	5.5×10^3
C=O(π* ← n)	3.7–3.5	270–290	10–20
—N=N—	2.9	350	15
	> 3.9	< 260	Strong
—NO_2	3.6	280	10
	4.8	210	1.0×10^4
C_6H_5—	3.9	255	200
	5.0	200	6.3×10^3
	5.5	180	1.0×10^5
$[Cu(OH_2)_6]^{2+}$(aq)	1.2	810	10
$[Cu(NH_3)_4]^{2+}$(aq)	1.7	600	50
H_2O(π* ← n)	6.0	167	7.0×10^3

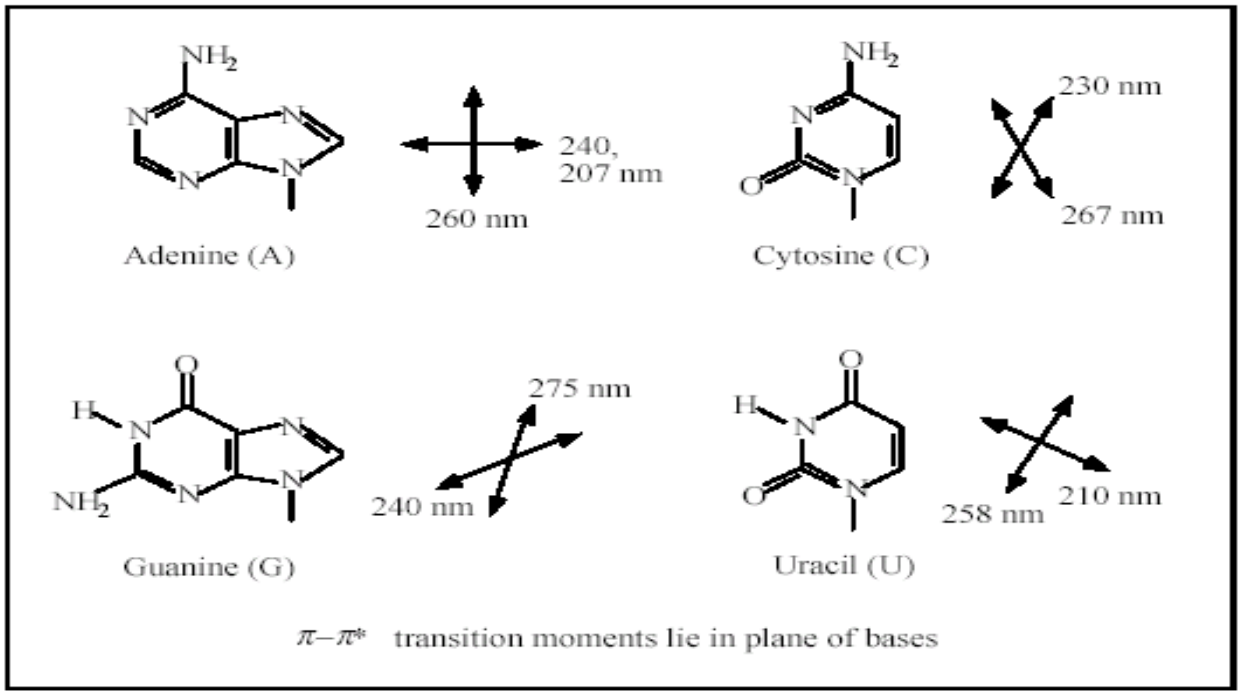
- Transition dipole along the bond axis

$$\vec{\mu}_{\pi\rightarrow\pi^*} = \vec{i}\int \psi_{\pi^*} x \psi_\pi d\tau + \vec{j}\int \psi_{\pi^*} y \psi_\pi d\tau + \vec{k}\int \psi_{\pi^*} z \psi_\pi d\tau$$

Conjugated π bonds

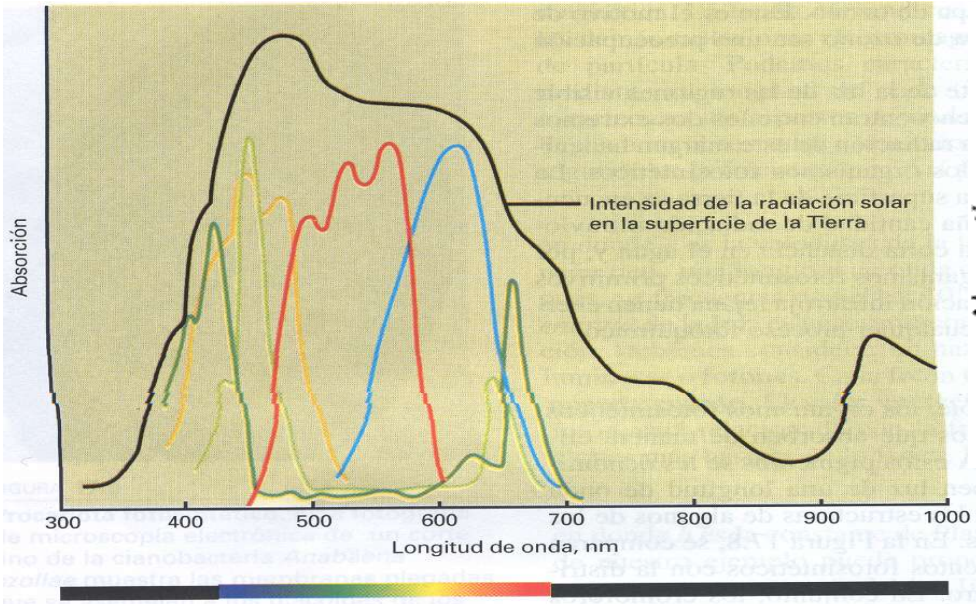
- Conjugation would lead to a shift towards lower wavelengths (from UV to visible)

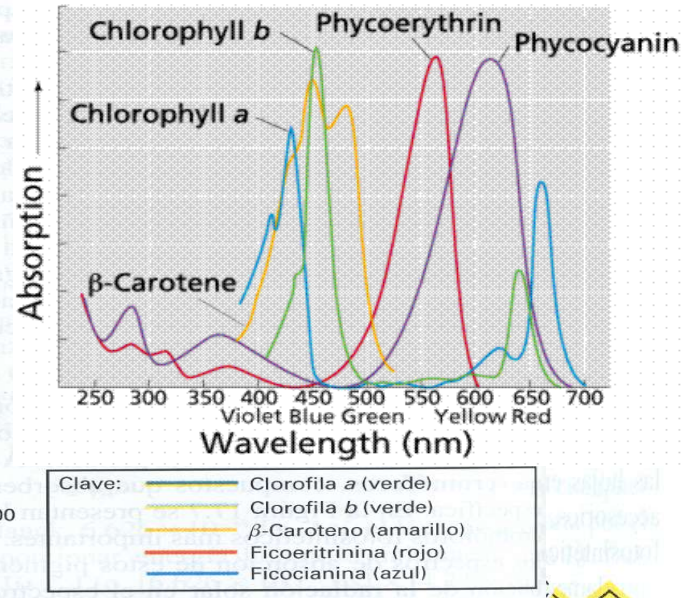




Adaptability of Biochromophores ...

...through chemical design that modulate the position of π bonds.





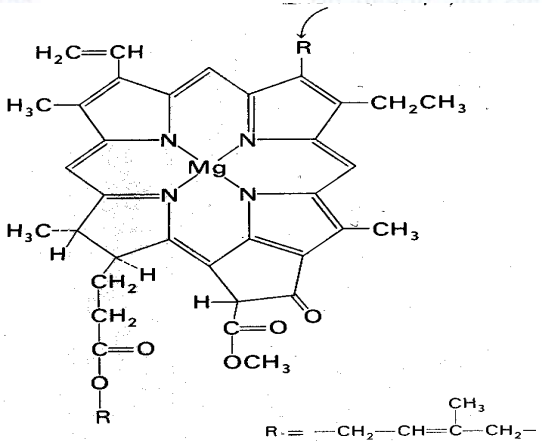
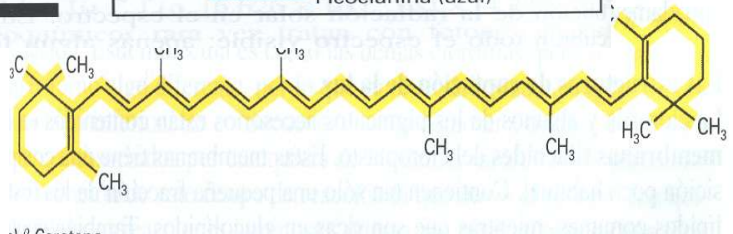
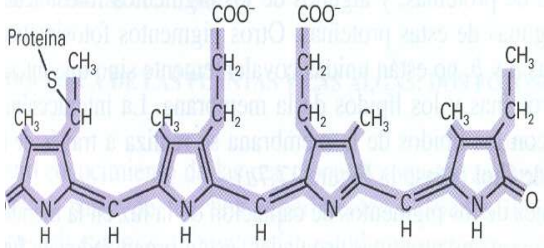


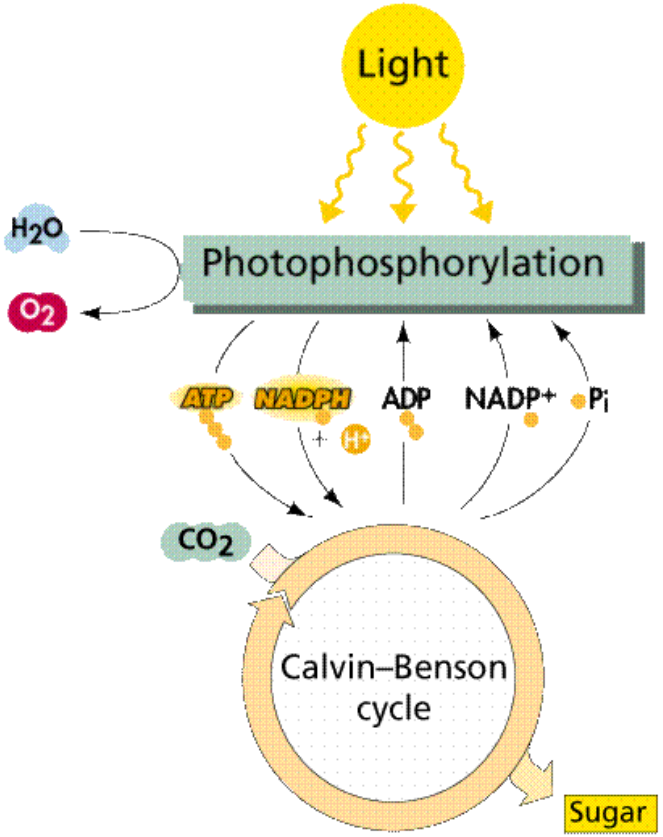
Figure 22-4
Formulas of chlorophylls *a* and *b*.

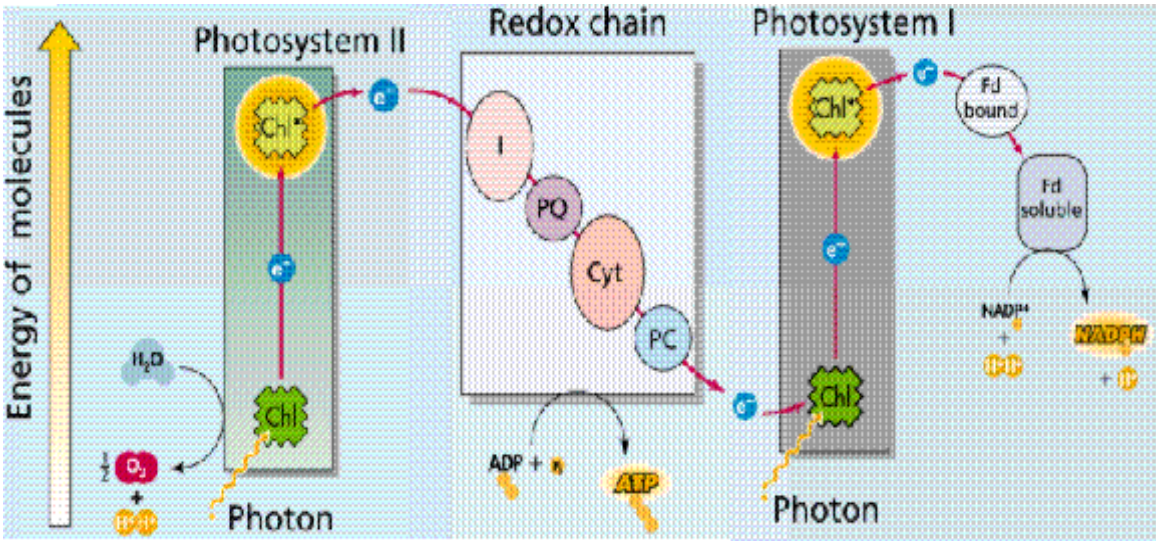




Biochromophores and Irati Forest

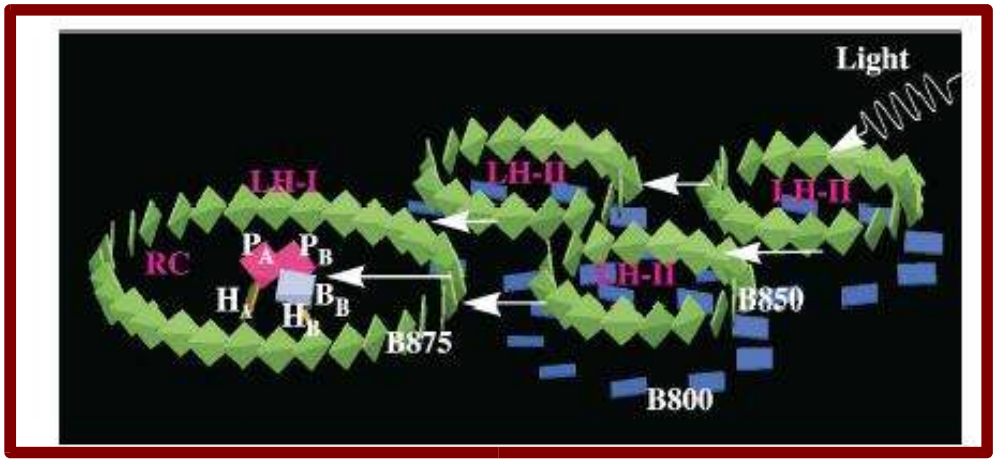
Photosynthesis. Process by which the plants cells use solar energy to produce ATP. The conversion of sunlight energy into usable chemical energy. Our atmosphere is oxidant thanks to this process.



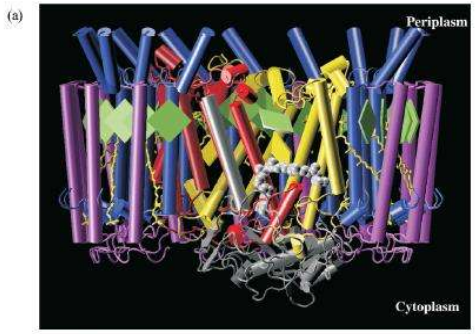


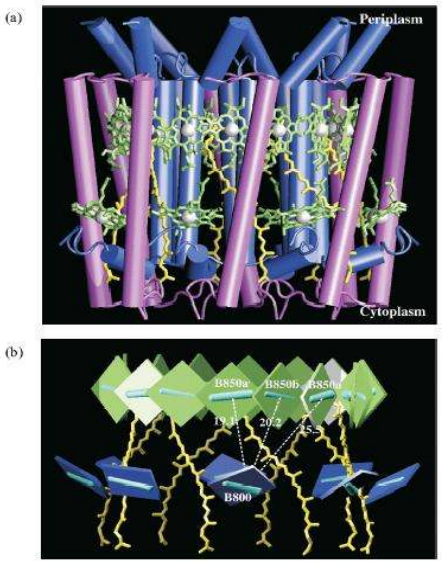
Chlorophylls are used to create a series of redox potentials that will give energy and e- to yield chemical energy

Photosyntetic Unit: Complex Structures



LH-I-RC

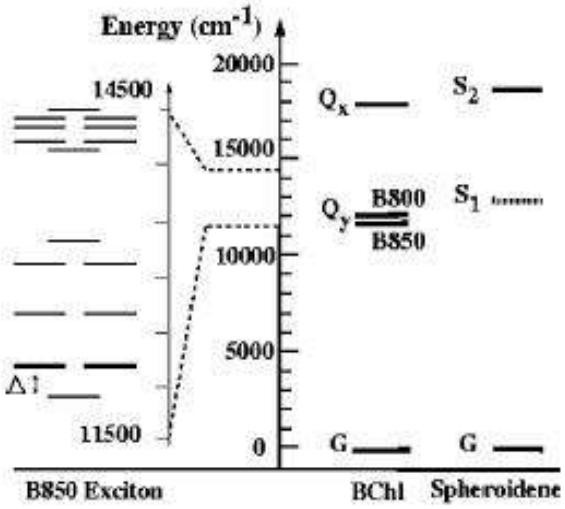


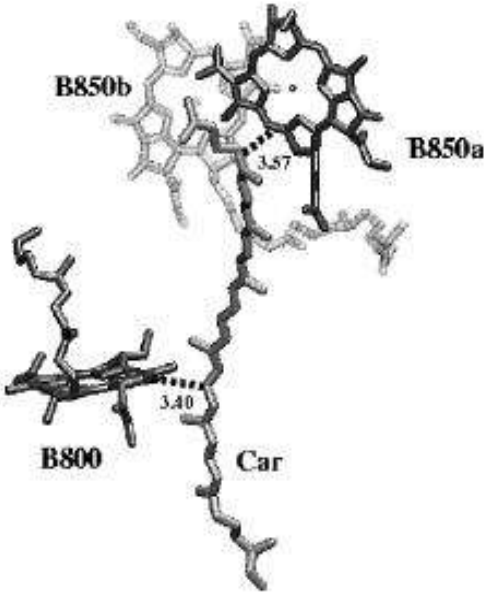


LH-II-RC

Hu X., Damjanovic A., Ritz T., Schulten K. PNAS, **1998**, *95*, 5935

Electronic excitations in Bacteriochloropylls

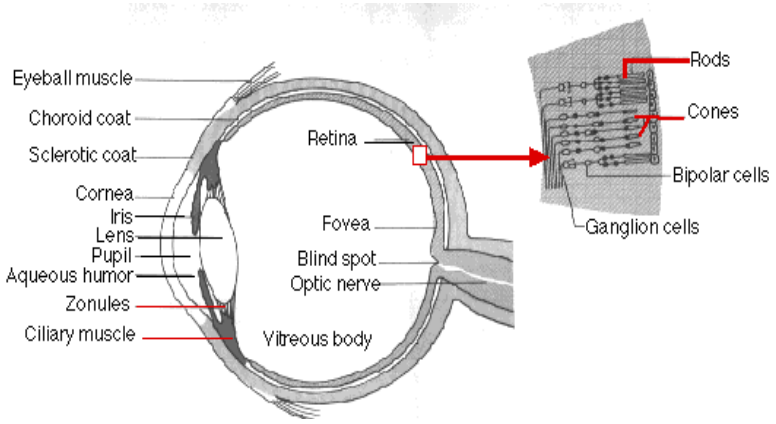


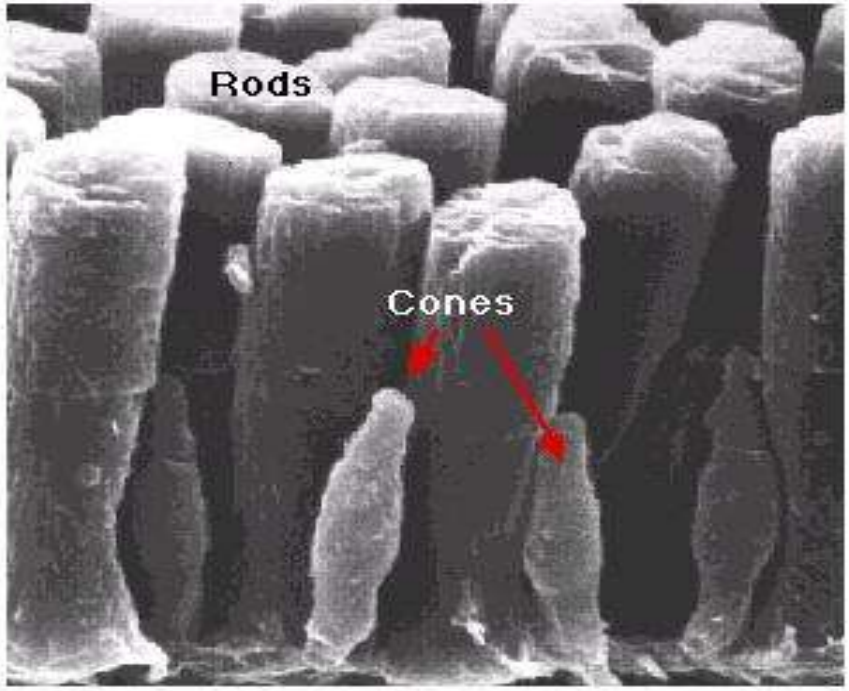


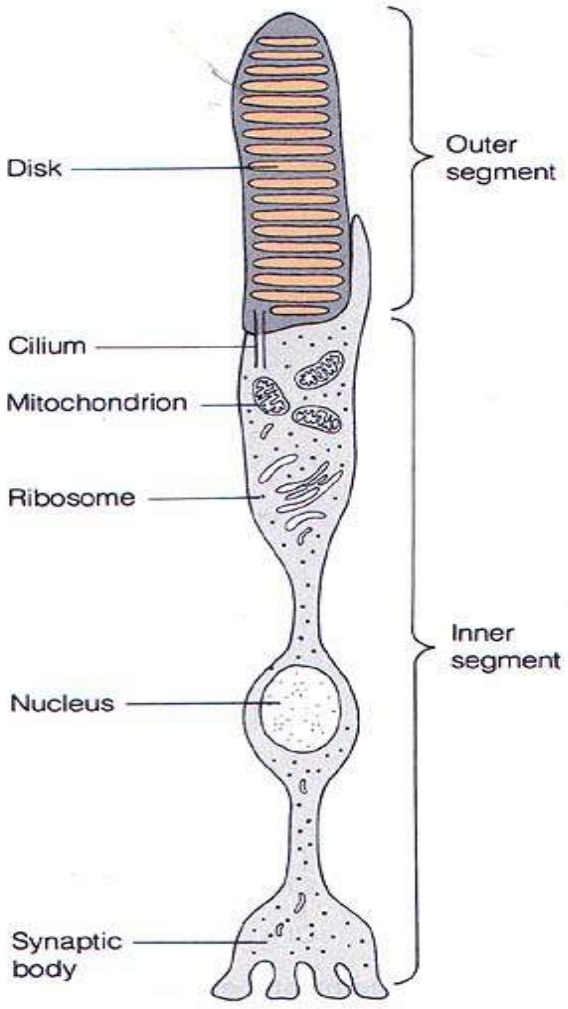
- INDO/s ZINDO/s + Effective Exciton Hamiltonians
- Why Mg? Fe, Zn
- Why the chlorins? Absorption of the Q band
- Why a dimer?
- Protein? Hole the structure + polarizable environment for stabilizing the charge transfer states

Zerner et al. J. Phys. Chem. B, **1998**, *102*, 7640

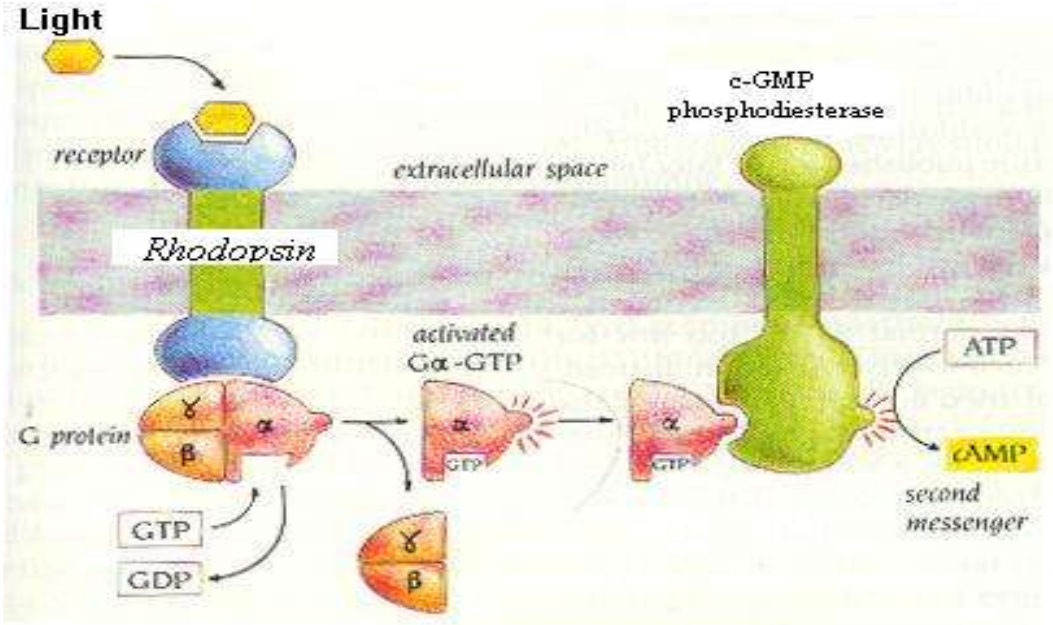
Vision: Light is used to trigger a conformational change in the chromophore

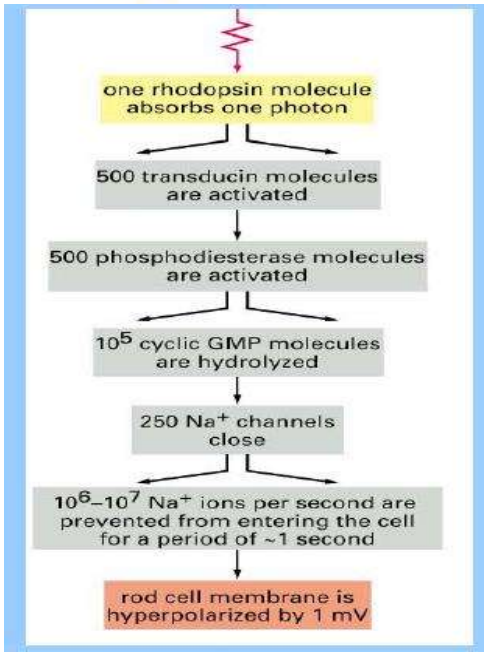






Rhodopsin: a G-Protein-Coupled receptor





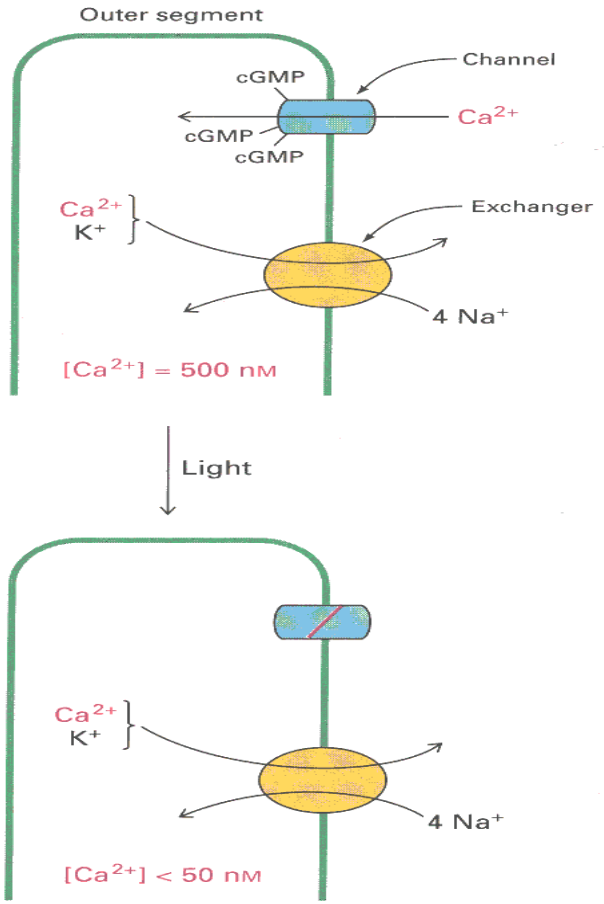
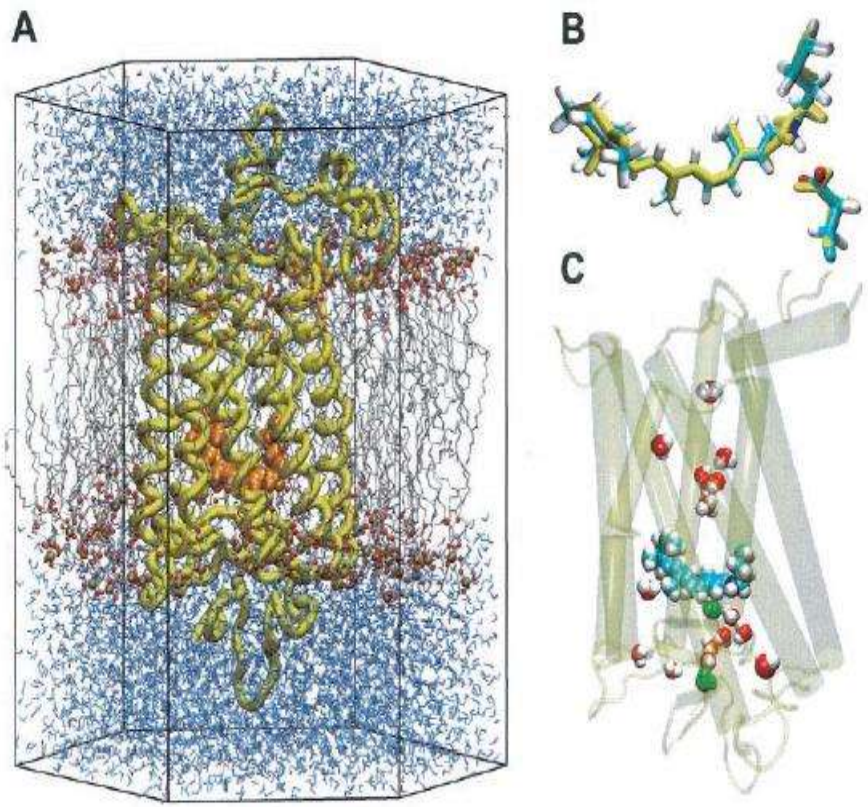


Figure 13-25
Light lowers the cytosolic calcium level in retinal rod cells by blocking the entry of Ca^{2+} through the cGMP-gated channel.

Rhodopsin X-Ray Data

- 7 α–Helixes
- Chromophore inserted at the middle of the protein structure
- Not only an scaffold. Chromophore structure significantly changed by the protein environment (non-planar and cis!!!). Trp86

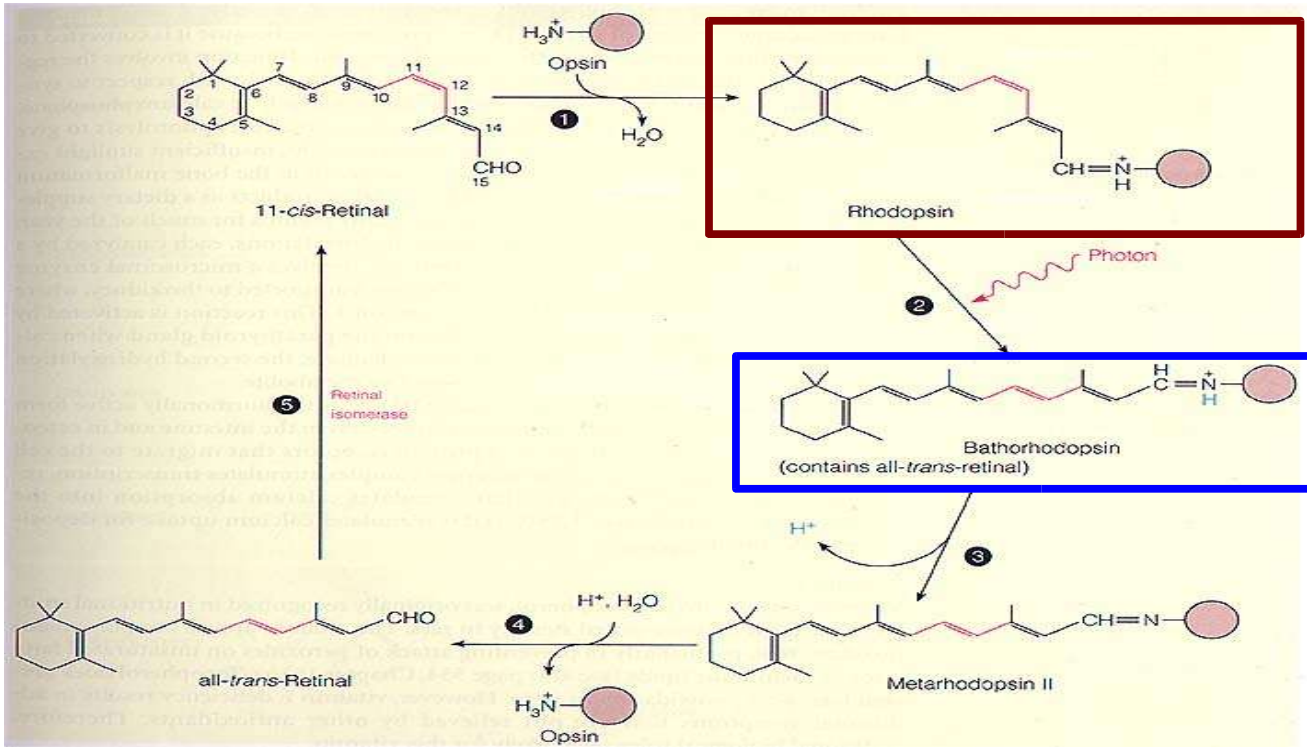


Pakzewski et al. Science **2000**, *289*, 739

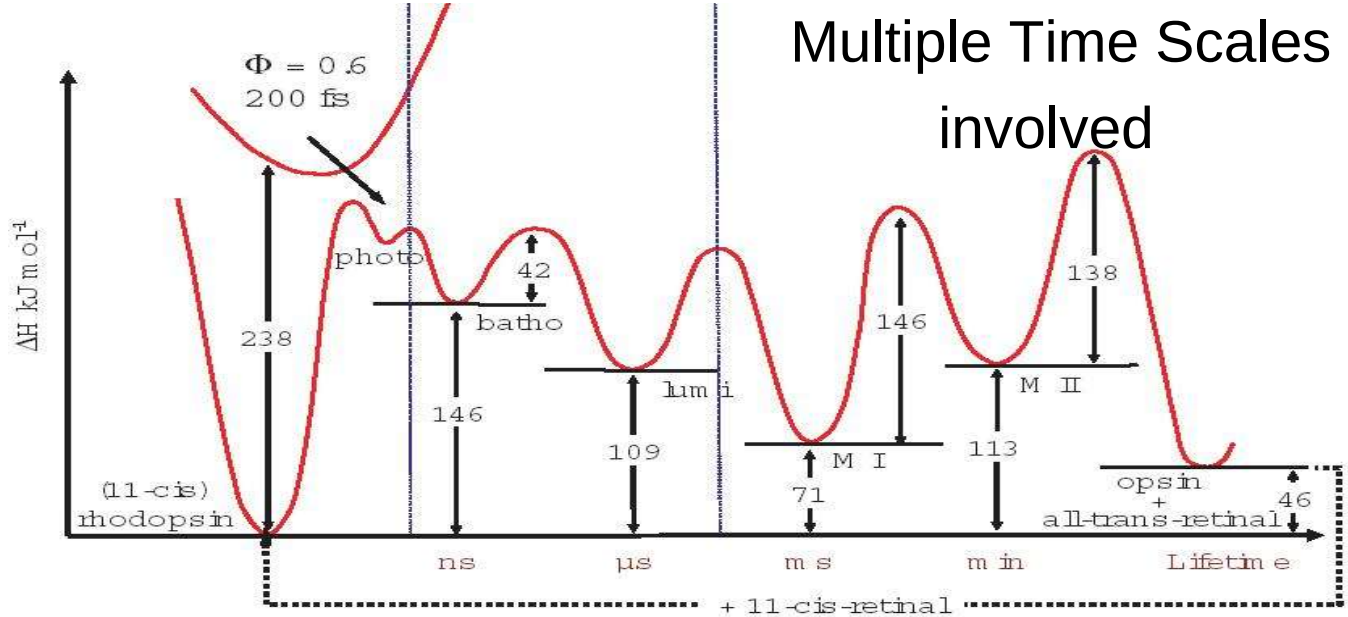
J. Saam J., E. Tajkhorshid, S. Hayashi, K. Schulten, Biophysical Journal, **2002**, *83*, 3097

E. Tajkhorshid, J. Baudry, K. Schulten, K. Suhai, Biophysical Journal, **2000**, *78*, 683.

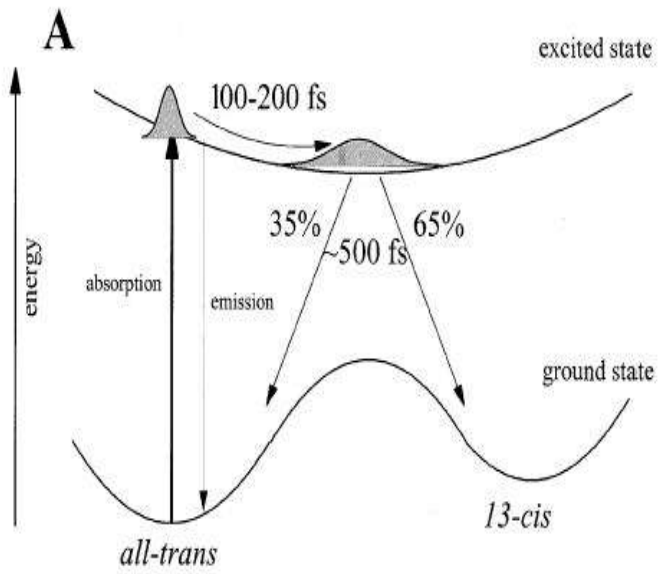
S. Hayashi, E. Tajkhorshid, K. Schulten, Biophysical Journal, **2003**, *85*, 1440.

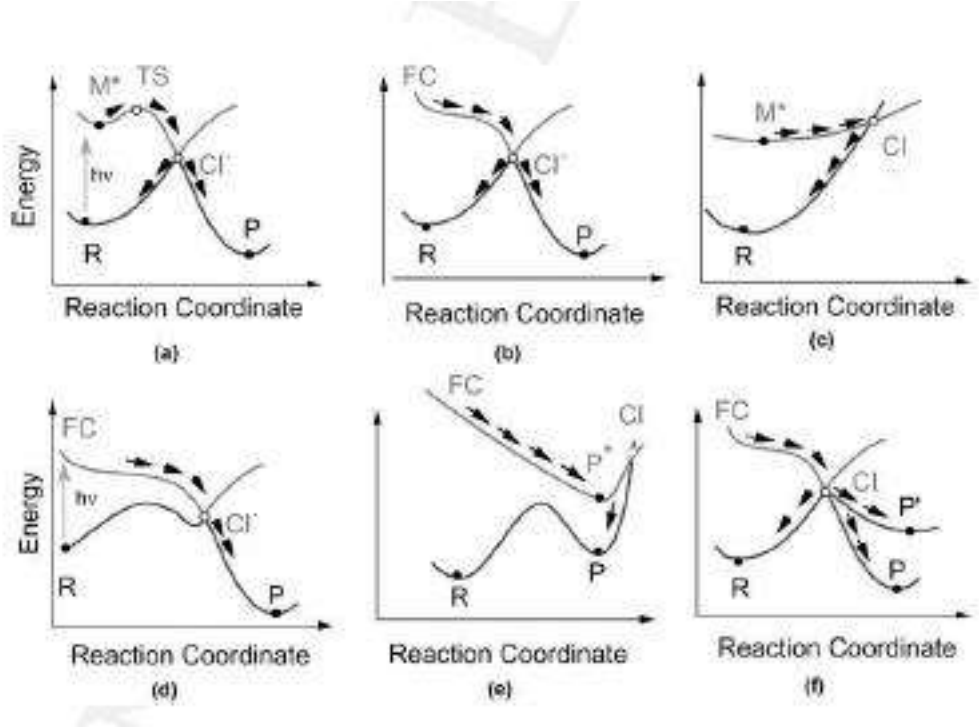


Chemical/Conformer Change



Multiple Time Scales involved





Robb and Olivucci Journal of Photochemistry and Photobiolo
2001, *5737*, 1

Study of “small molecules” to understand the photoisomerization in the gas phase but Protein Environment needed: stabilize cis conformer, non-planarity of the conformer, dynamics of the conformational change, but also ...

... modulates the absorption: Color vision Same cromophore but different response. Interactions with the protein environment are important(!)

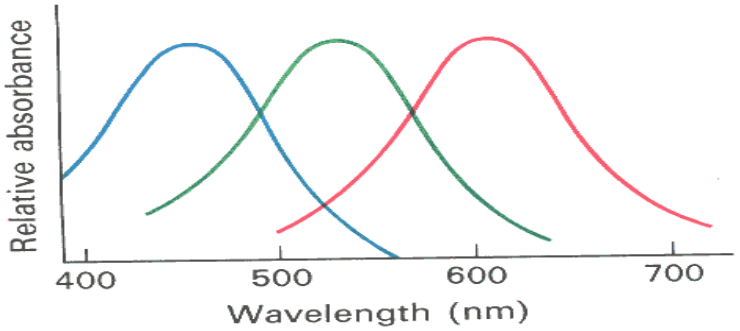


Figure 13-28
Absorption spectra of the three receptors mediating color vision in goldfish.

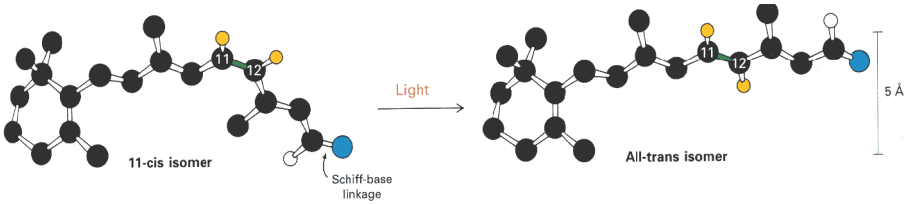
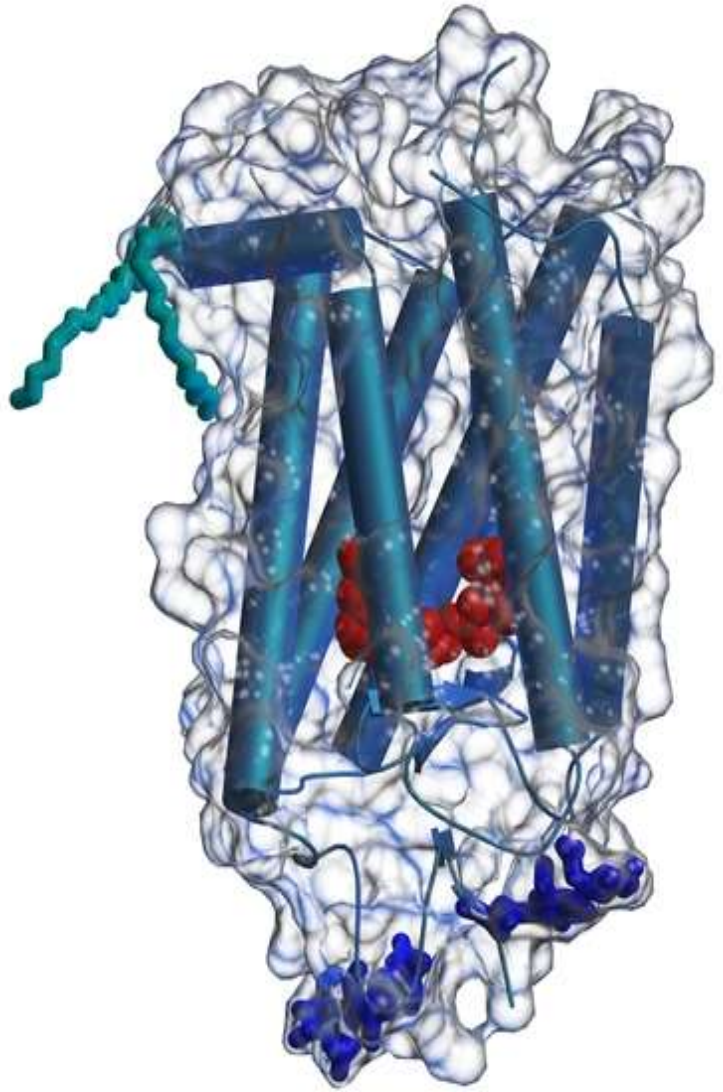
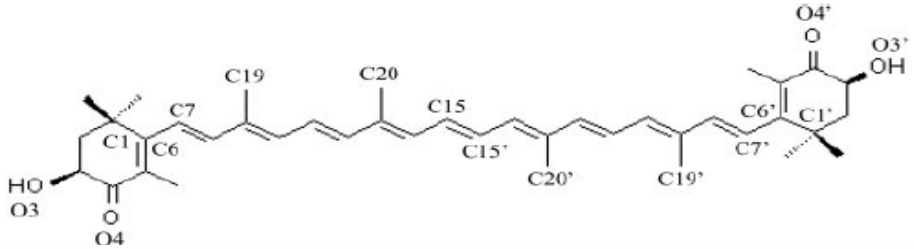
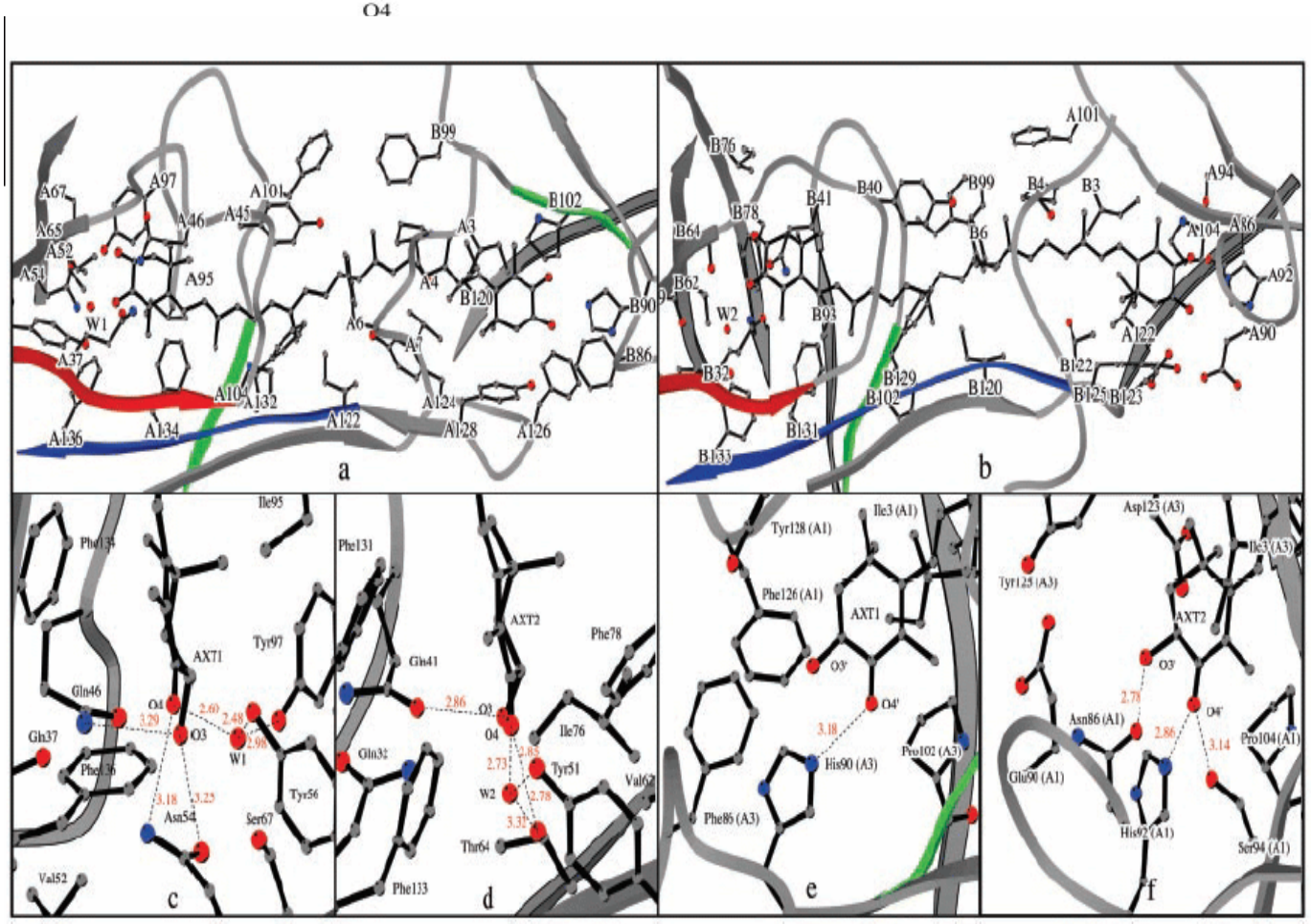


Figure 13-19
The primary event in visual excitation is the isomerization of the 11-cis isomer of the Schiff base of retinal to the all-trans form. The double bond between C11 and C12 is shown in green; hydrogen atoms attached to these carbons are shown in yellow.



Blue Colouration in Lobster Shell





Durbeej and Eriksson, Chem. Phys. Lett., **2003**, *375*, 30

Proteins (Aminoacid polymers)

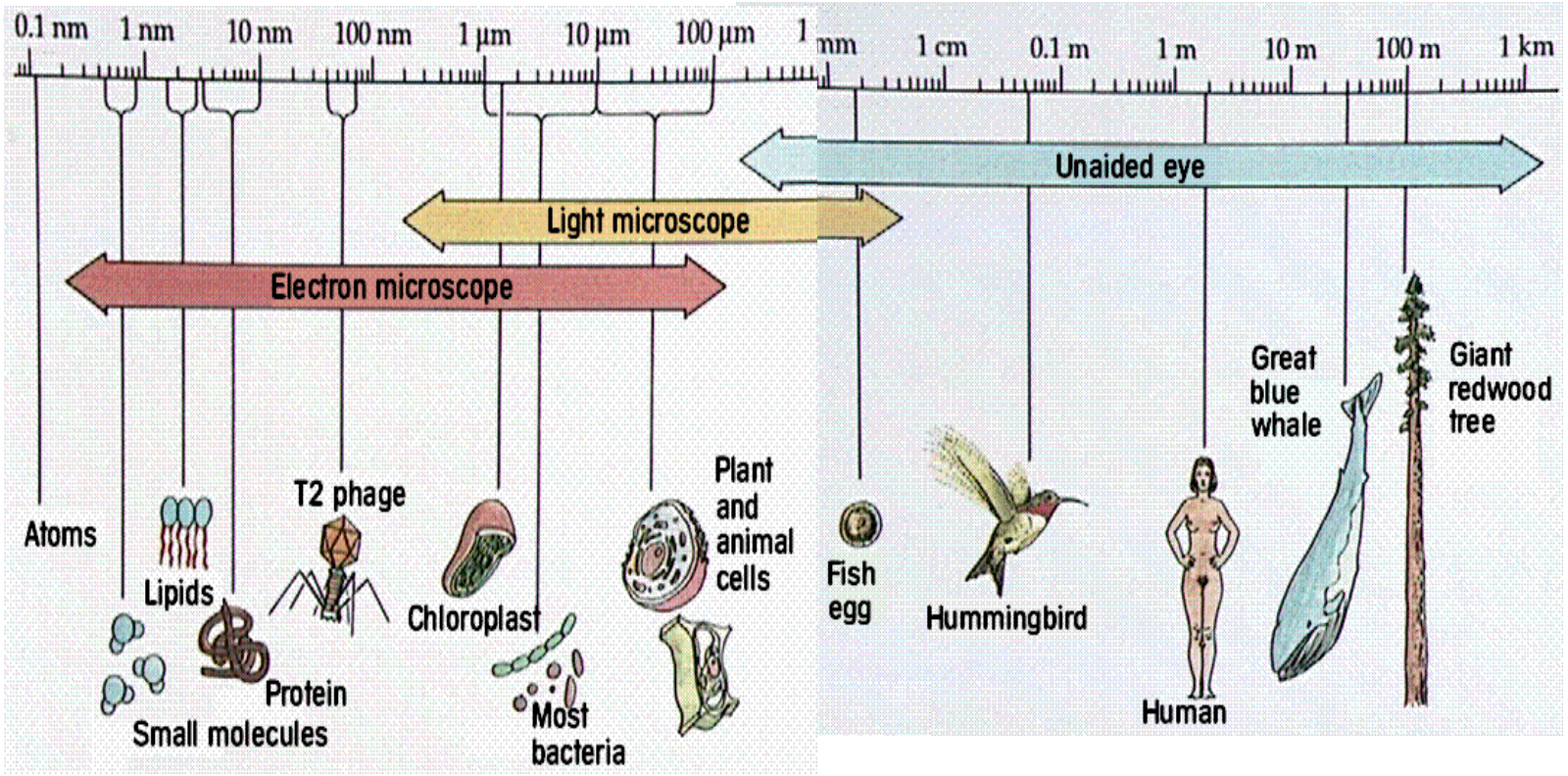
Protein structure will have a direct effect on the properties displayed by the chromophore:

Preferential Conformer (e.g. cis/trans in ret)

Structural Changes in the cromophore (non-planarity)

Chemical Changes of the chromophore

polarization of the electronic structure by Electrostatics/HB



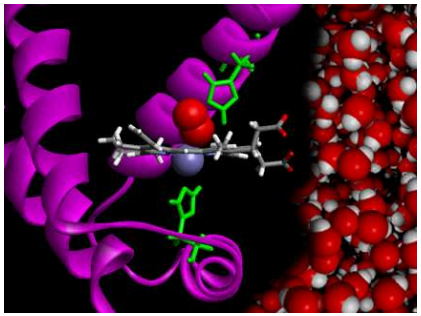
AminoAcids

Flexibility on the Physical Chemical Properties:

- Size/Shape
- Polar/Apolar AA
- +/- AA
- Hydrogen Bonds
- Reactivity

Big **variety** of Proteins with diverse functions

Structure-Function relationship



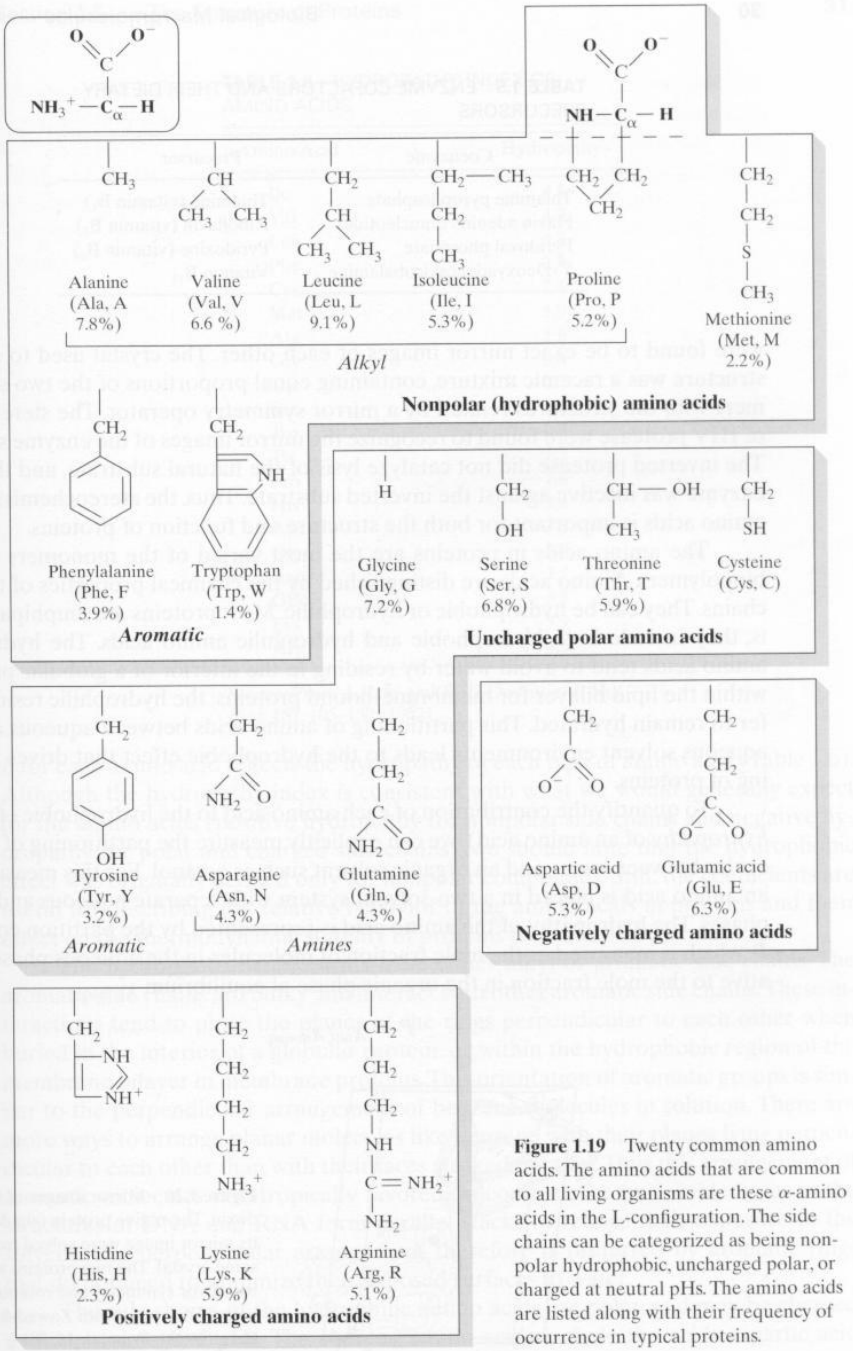
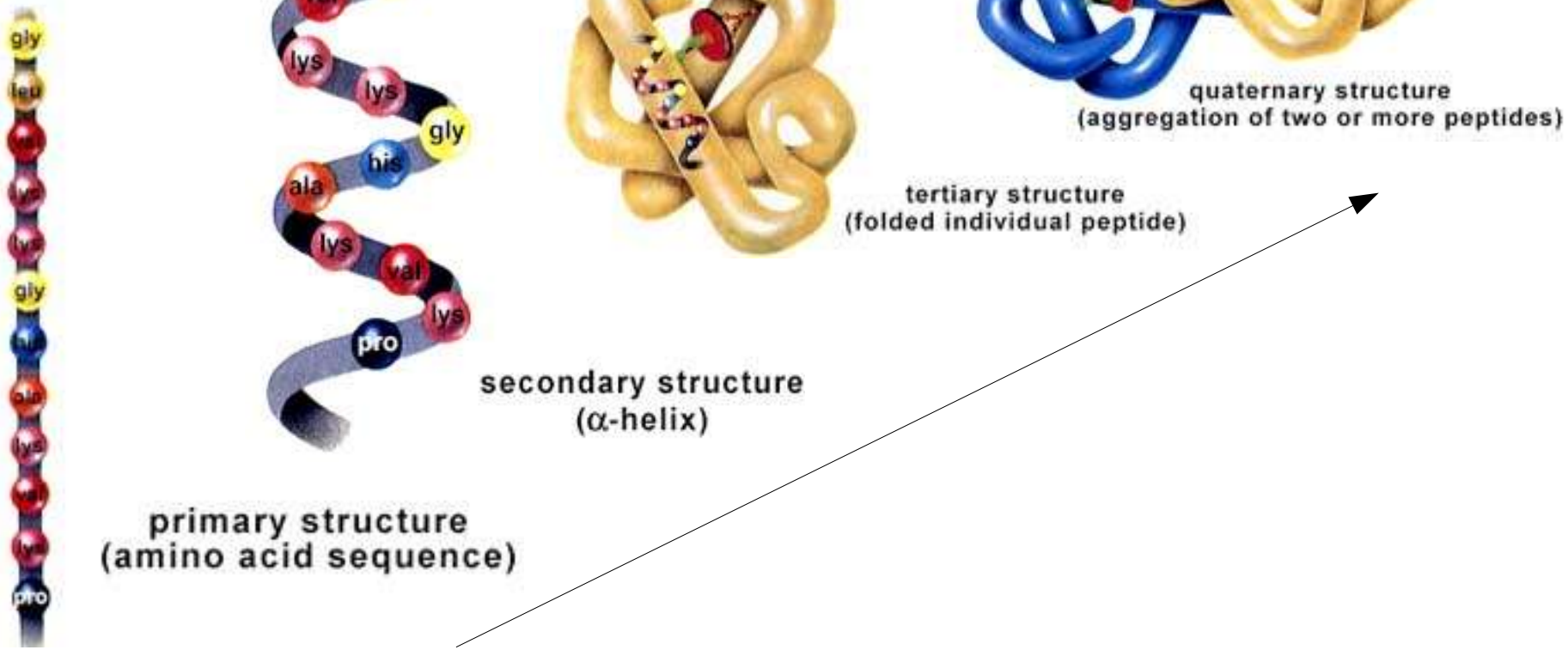


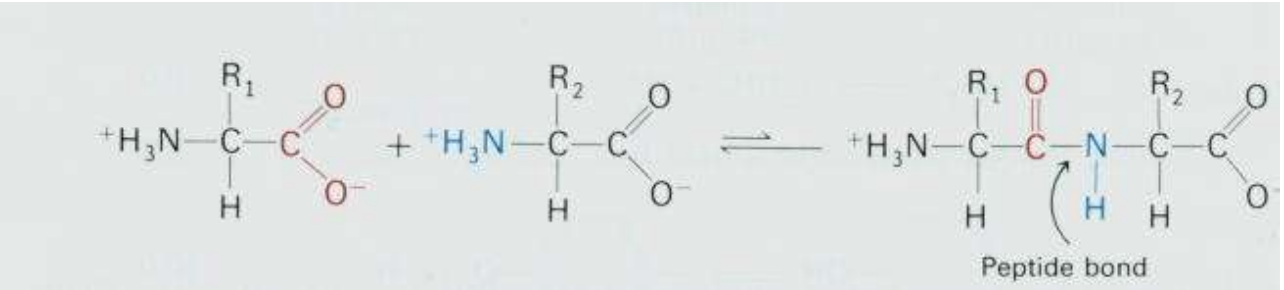
Figure 1.19 Twenty common amino acids. The amino acids that are common to all living organisms are these α-amino acids in the L-configuration. The side chains can be categorized as being non-polar hydrophobic, uncharged polar, or charged at neutral pHs. The amino acids are listed along with their frequency of occurrence in typical proteins.

Structural Levels



Level I: Peptide Bond

Aminoacid sequence assembled by peptide bonds.



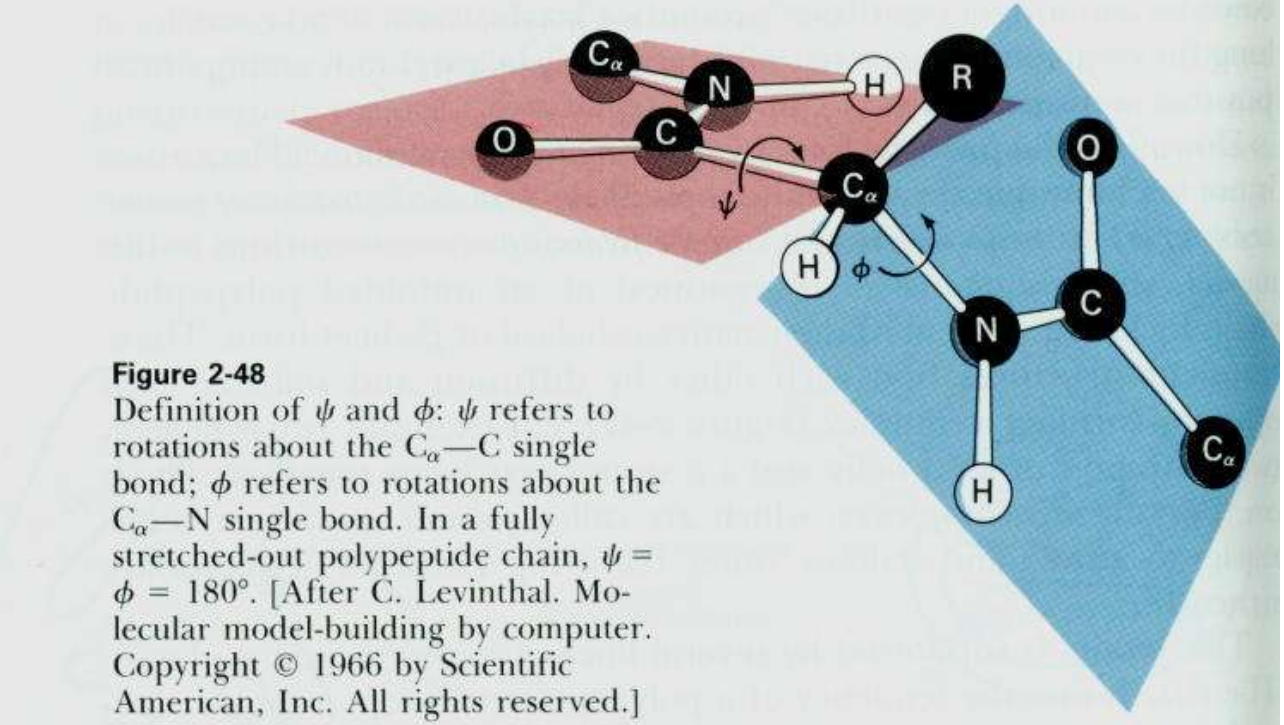
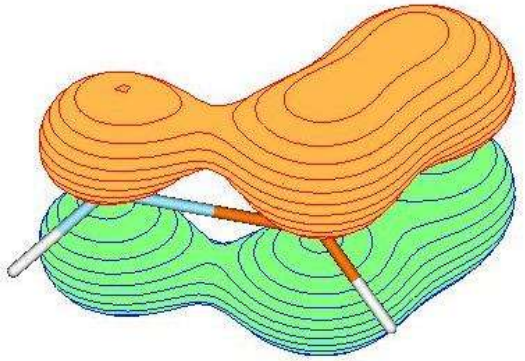
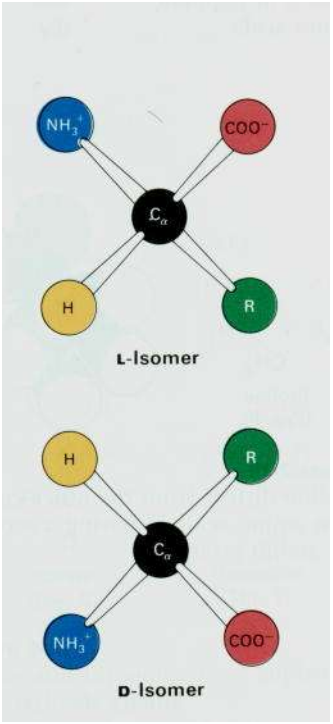


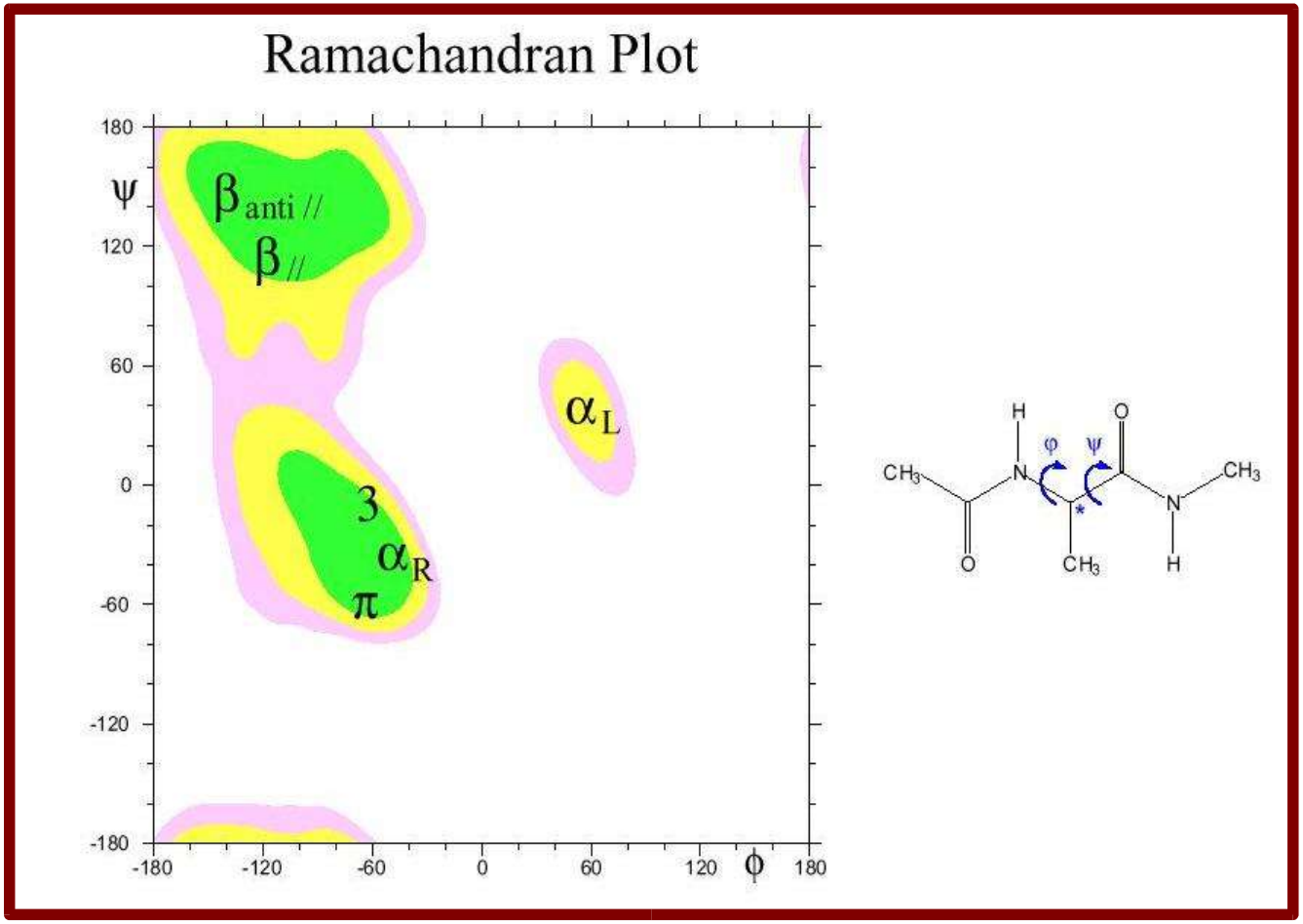
Figure 2-48
Definition of ψ and ϕ: ψ refers to rotations about the C_α—C single bond; ϕ refers to rotations about the C_α—N single bond. In a fully stretched-out polypeptide chain, $\psi = \phi = 180°$. [After C. Levinthal. Molecular model-building by computer. Copyright © 1966 by Scientific American, Inc. All rights reserved.]





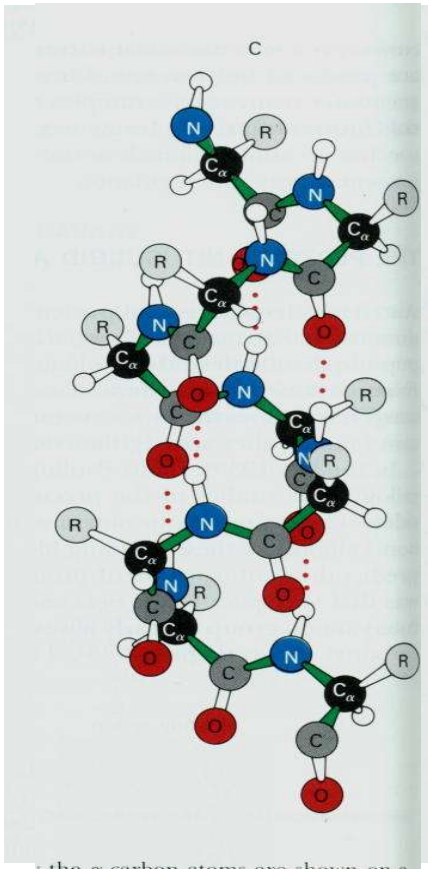
•Specific combinations of ψ,ϕ are repeated revealing that there is a local folding order due to the specific electronic characteristics of the peptide bond

•There repetitive motifs are called secondary structure

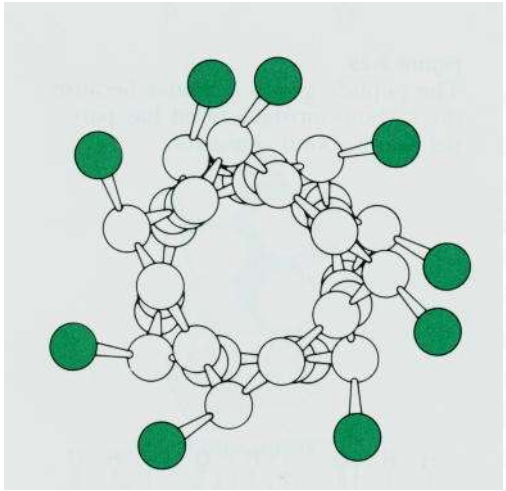


Secondary Structure

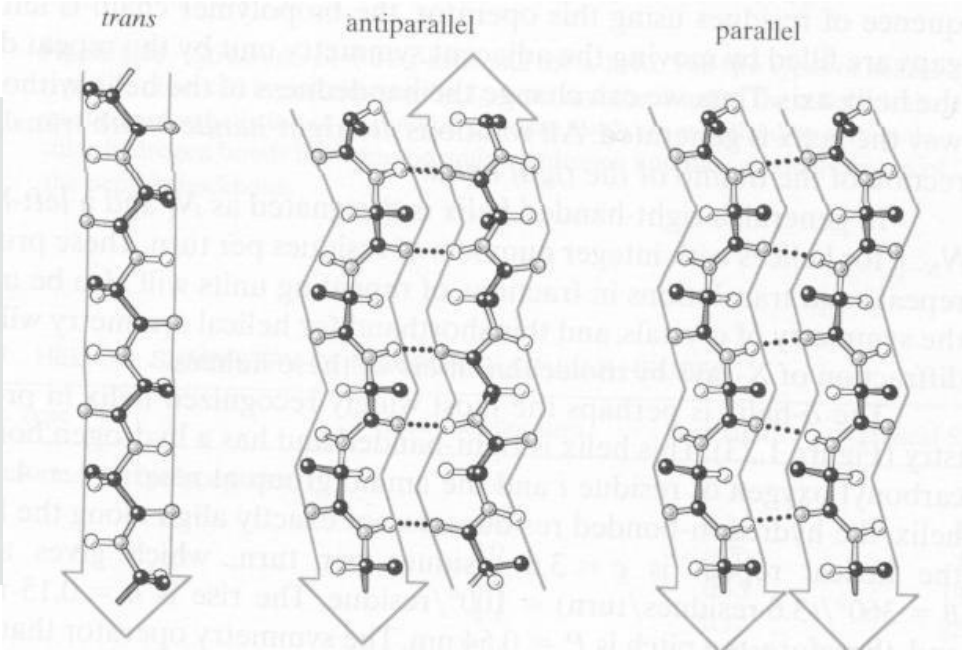
Local Structure: Main Types.



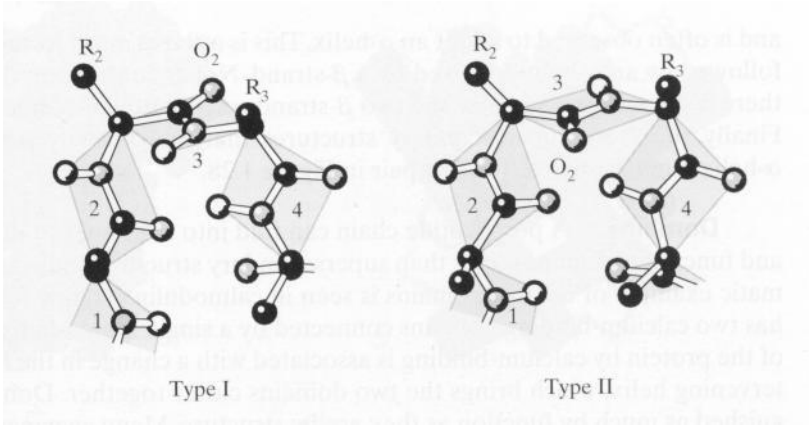
α Helix

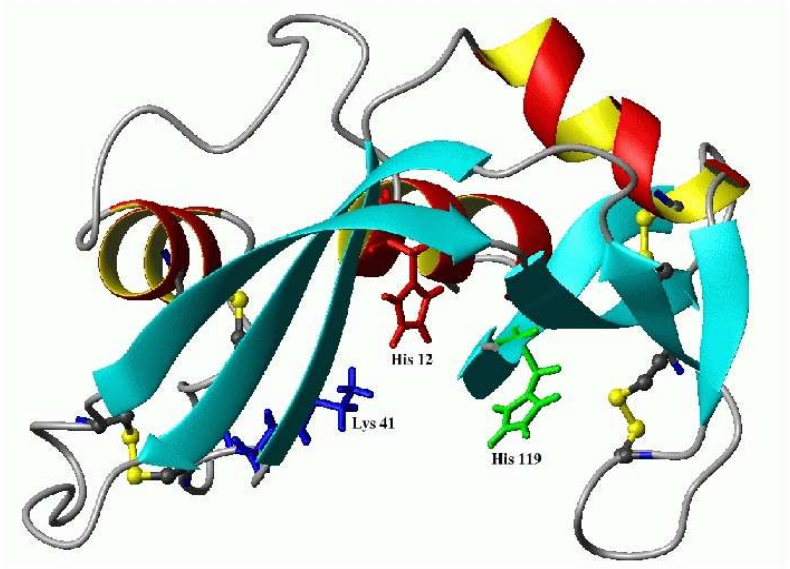


β Sheet



β Turns





Tertiary Structure

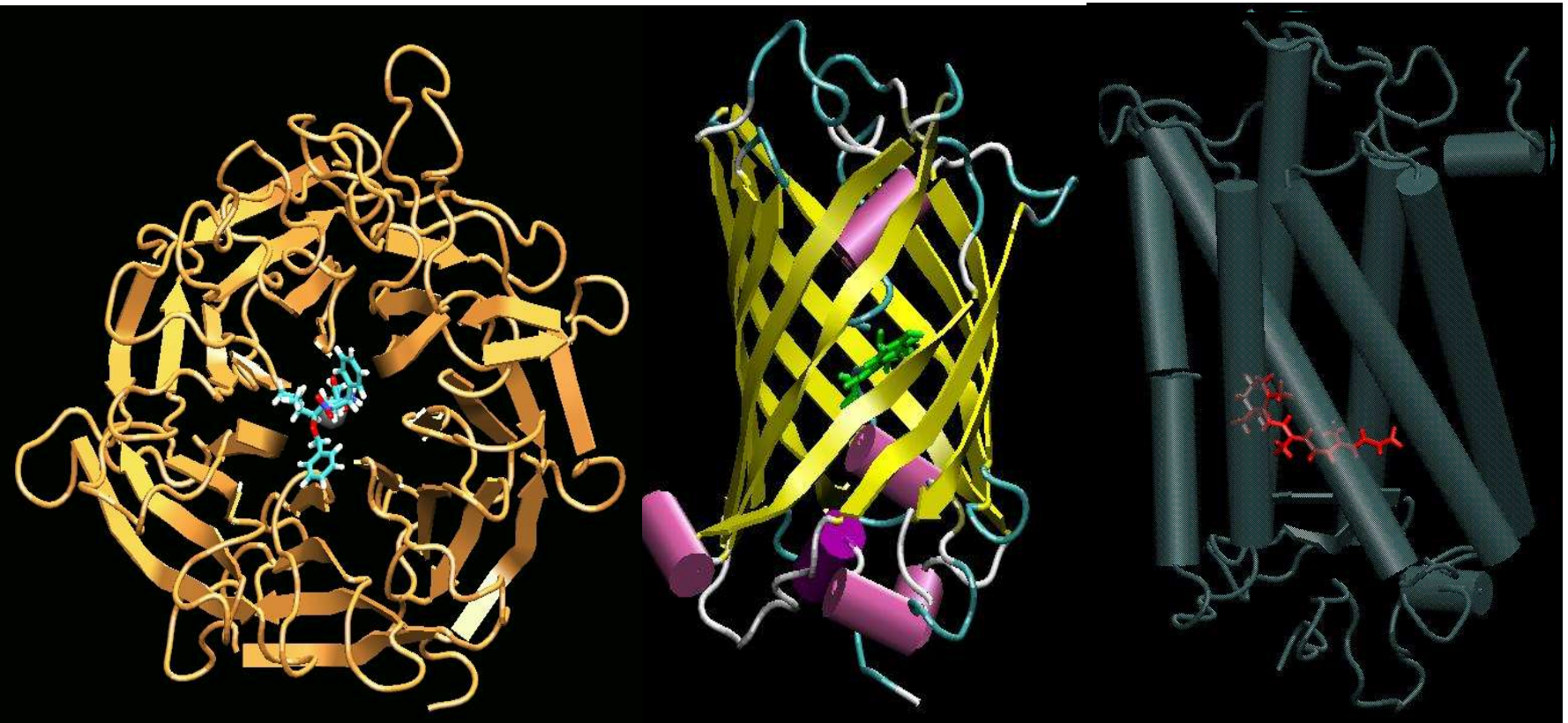
3D Structure of Proteins

Stabilized by:

HB/Electrostatic Interactions

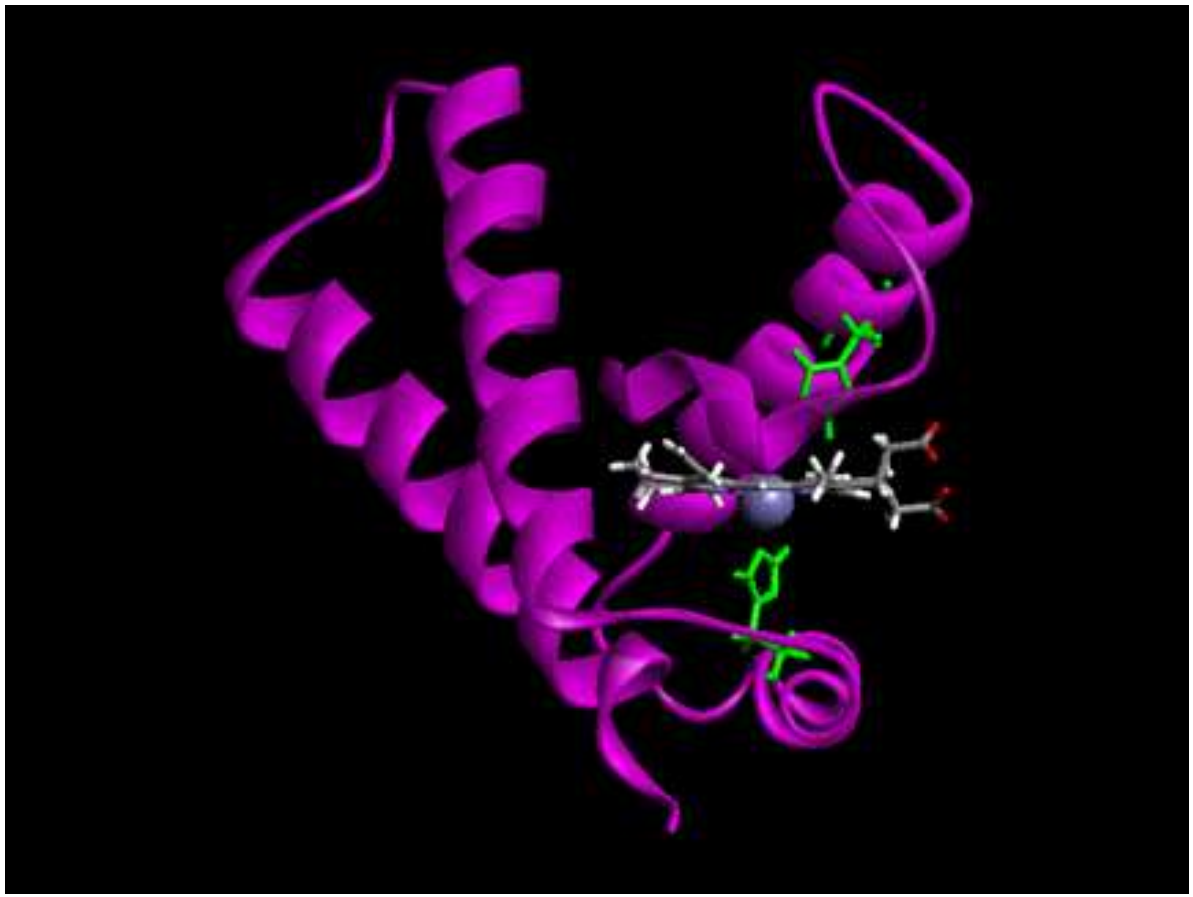
Covalent Bonds S-S

Solvent Effects



Quaternary

SupraMolecular Assamblies: Protein Polymer



Experimental Structure Determination

X-ray Crystallography

Main Steps:

* Expression of pure soluble protein in large amount in bacteria or other cells
* Vary solution properties (high concentrations) to induce crystal formation and growth
* X-ray diffraction patterns -> electron density map
* Fit and refine the polypeptide chain into the density (Simulated annealing, minimization, human intervention)

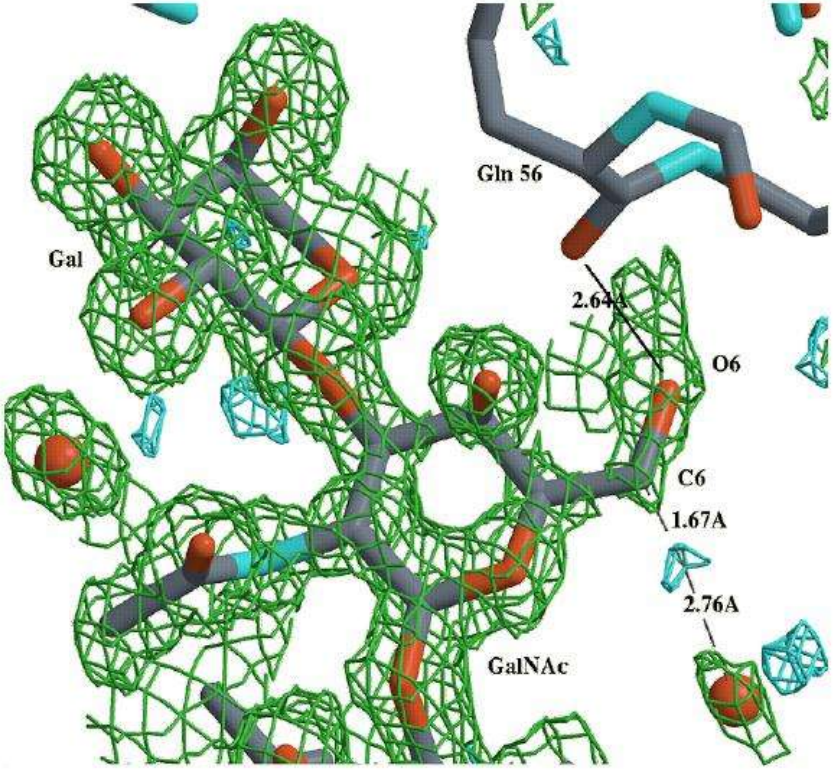


Figure 1: Electronic density map (shown as green lines) used to fit the heavy atoms positions.

Advantages:

* Fast, Large molecules
* Very low resolution i.e., up tp 0.54 A concentrations) to induce crystal formation and growth

Disadvantages:

* Requires crystal formation
* Not in physological conditions
*Hydrogens not resolved

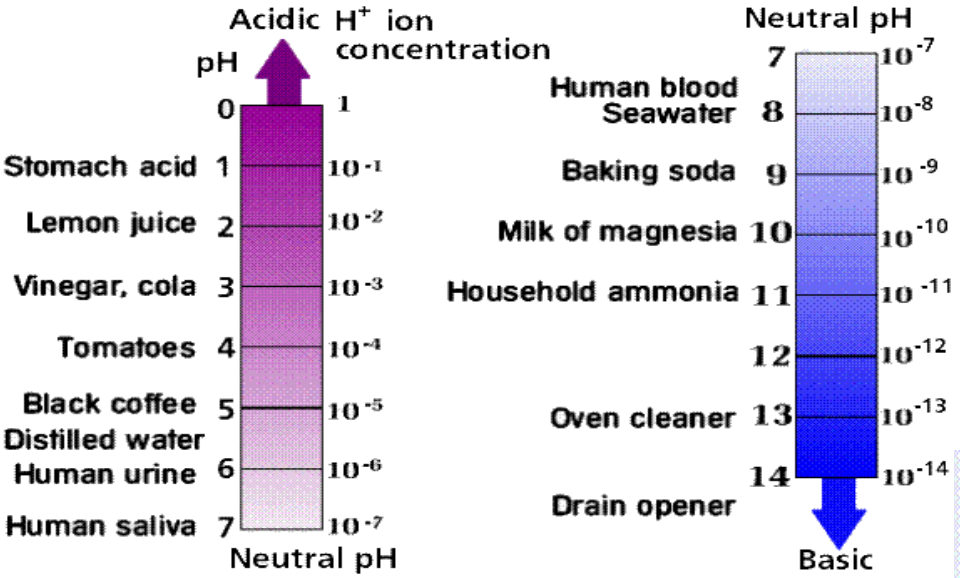
Effect of pH in AA's

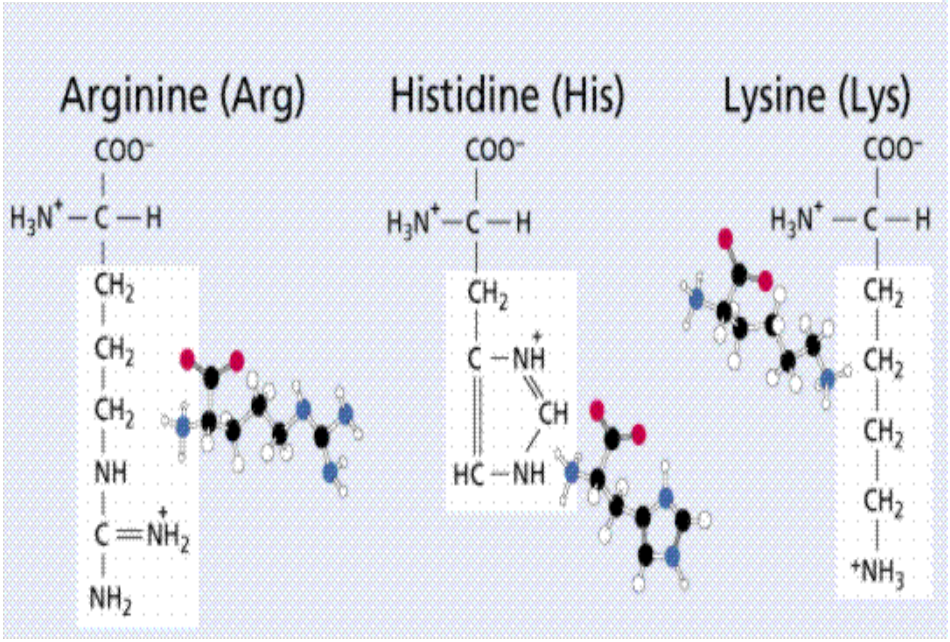
Influence on the Protonation State and therefore in Protein Structure

Information not available from X-Ray

Based on pKa values in aqueous solution we know that Arg and Lys are protonated, but His?

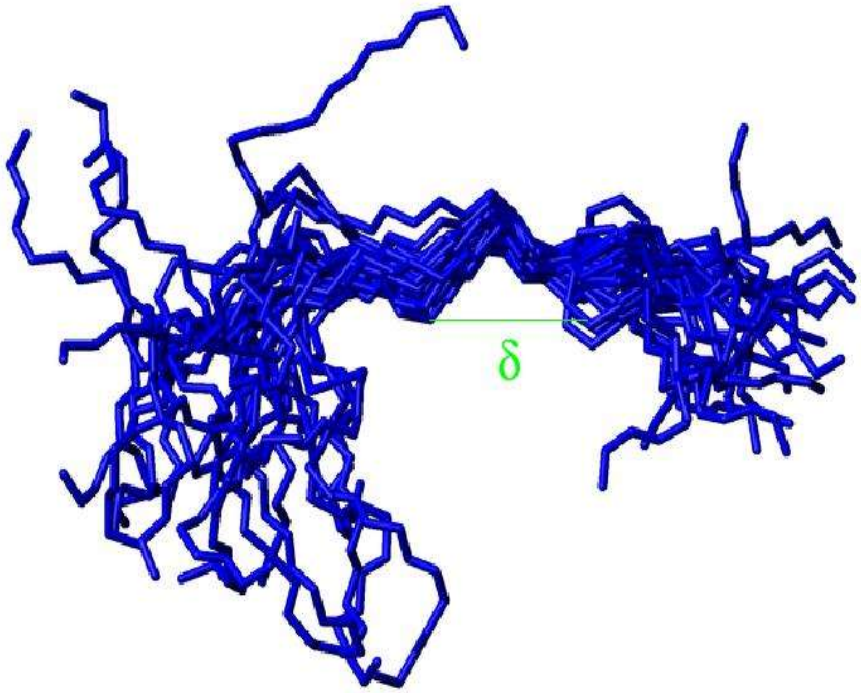
What about pKa inside the proteins?





Alternatives to X-Ray

NMR



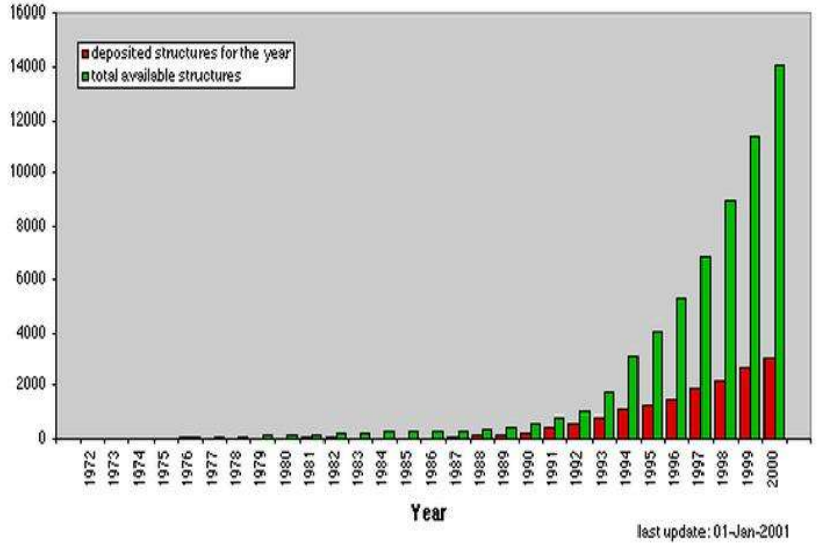
Homology Modeling

TABLE 4.2 PROPENSITIES OF AMINO ACIDS TO FORM α-HELICES ($\langle P_\alpha \rangle$) AND β-SHEETS ($\langle P_\beta \rangle$)

α-residues	$\langle P_\alpha \rangle$	α-assignment	β-Residues	$\langle P_\beta \rangle$	β-assignment
Glu	1.44 ± 0.06	H_α	Val	1.64 ± 0.07	H_β
Ala	1.39 ± 0.05	H_α	Ile	1.57 ± 0.08	H_β
Met	1.32 ± 0.11	H_α	Thr	1.33 ± 0.07	h_β
Leu	1.30 ± 0.05	H_α	Tyr	1.31 ± 0.09	h_β
Lys	1.21 ± 0.05	h_α	Trp	1.24 ± 0.14	h_β
His	1.12 ± 0.08	h_α	Phe	1.23 ± 0.09	h_β
Gln	1.12 ± 0.07	h_α	Leu	1.17 ± 0.06	h_β
Phe	1.11 ± 0.07	h_α	Cys	1.07 ± 0.12	h_β
Asp	1.06 ± 0.06	h_α	Met	1.01 ± 0.13	I_β
Trp	1.03 ± 0.10	I_α	Gln	1.00 ± 0.09	I_β
Arg	1.00 ± 0.07	I_α	Ser	0.94 ± 0.06	i_β
Ile	0.99 ± 0.06	i_α	Arg	0.94 ± 0.09	i_β
Val	0.97 ± 0.05	i_α	Gly	0.87 ± 0.05	i_β
Cys	0.95 ± 0.09	i_α	His	0.83 ± 0.09	i_β
Thr	0.78 ± 0.05	i_α	Ala	0.79 ± 0.05	i_β
Asn	0.78 ± 0.06	i_α	Lys	0.73 ± 0.06	b_β
Tyr	0.73 ± 0.06	b_α	Asp	0.66 ± 0.06	b_β
Ser	0.72 ± 0.04	b_α	Asn	0.66 ± 0.06	b_β
Gly	0.63 ± 0.04	B_α	Pro	0.62 ± 0.07	B_β
Pro	0.55 ± 0.05	B_α	Glu	0.51 ± 0.06	B_β

Listed are values compiled from the crystal structures of 64 proteins, and the assignments as former (*H*

Protein Data Bank



Location: http://www.rcsb.org/pdb/

Highlights:

***** Founded in 1971, freely available

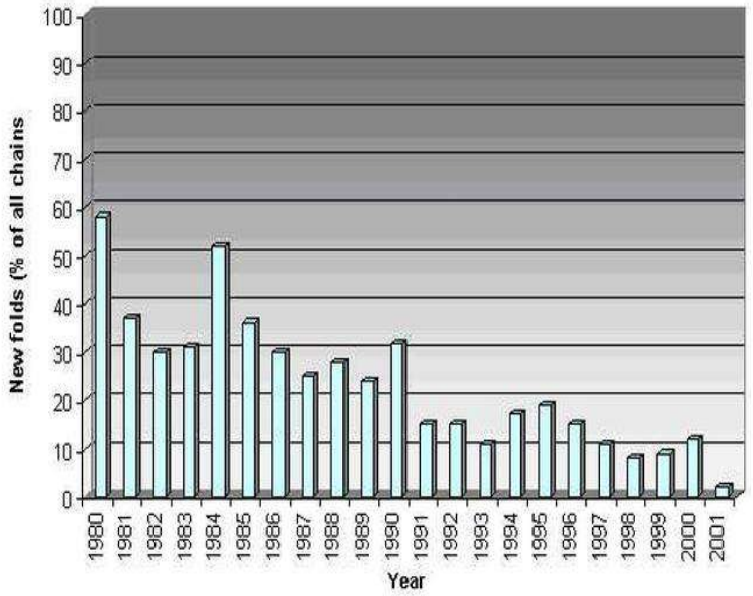
* Around 16000 in Jan 2002 (80 % X-ray)

***** Standard PDB format:

-Header: Experimental details, ref,etc.

-coordinates for each heavy atom

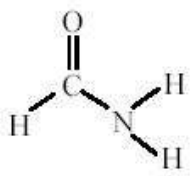
-B factors



Proteins as chromophores:

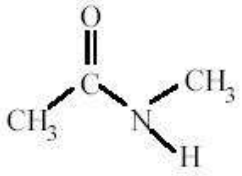
Peptide Bond

Typical models:



formamide

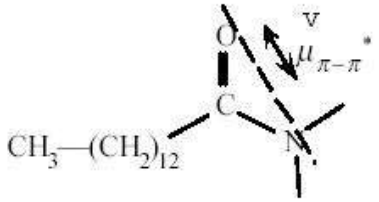
or



N-methyl acetamide

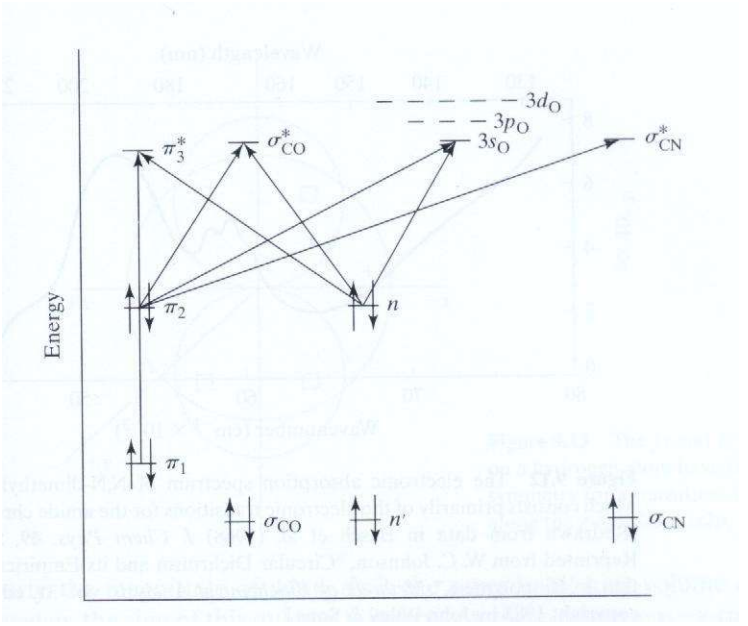
Peptide bond

$n \rightarrow \pi^*$ absorption	210-220 nm	ε_{max} ~ 100 weak
$\pi \rightarrow \pi^*$	190 nm	ε_{max} ~ 7000



myristamide

ν $\mu_{\pi-\pi^*}$ not along C═O but along a line between O and N in plane of peptide bond — determined by polarized absorption of single crystals



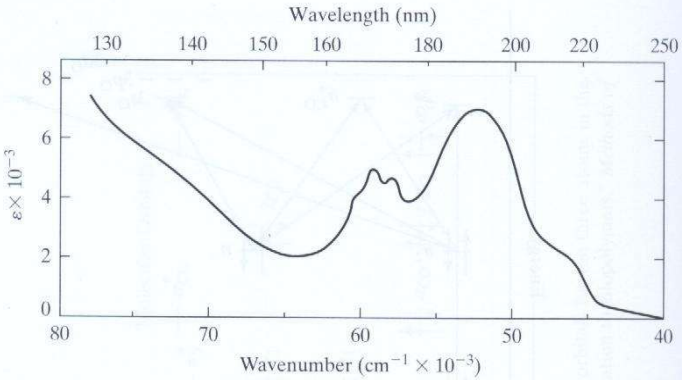
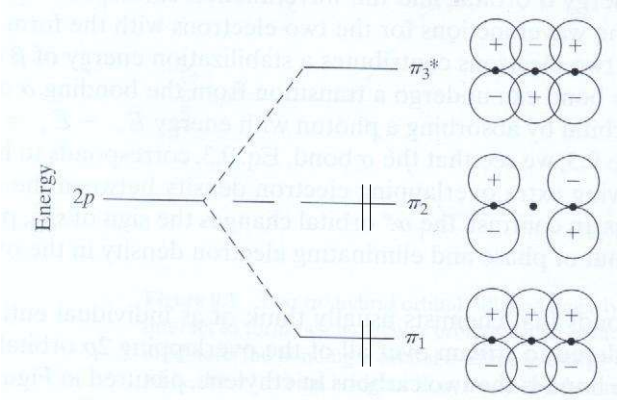
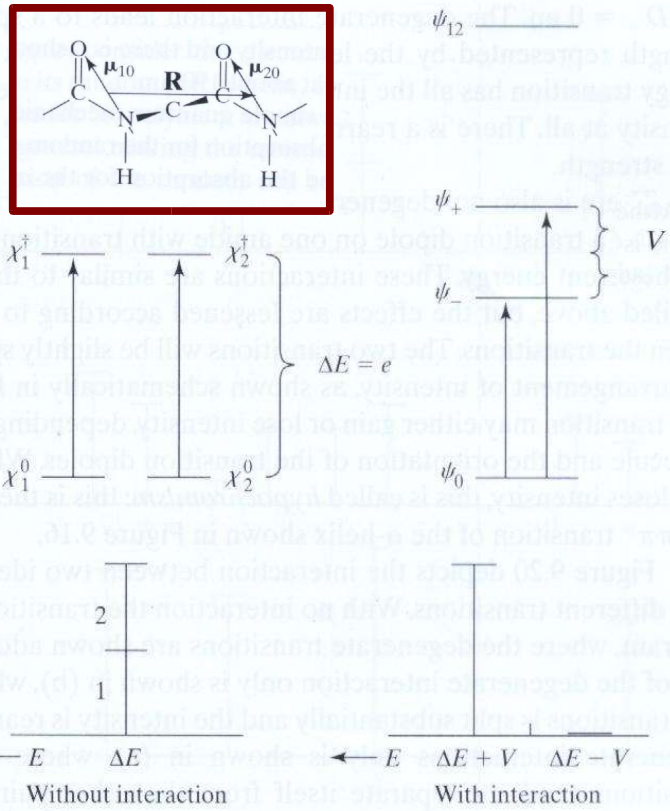


Figure 9.12 The electronic absorption spectrum of N,N-dimethylacetamide, which consists primarily of the electronic transitions for the amide chromophore. [Redrawn from data in Basch et al. (1968) *J. Chem Phys.* **49**, 5007–5018. Reprinted from W. C. Johnson, "Circular Dichroism and its Empirical Application to Biopolymers," *Methods of Biochemical Analysis,* vol. 31, ed. D. Glick, copyright 1985 by John Wiley & Sons.]



Interactions Between Amides:



$$V = \int \phi_1^* V_{12}\, \phi_2 d\tau \cong \frac{\boldsymbol{\mu}_{10} \cdot \boldsymbol{\mu}_{20}}{R^3} - 3\frac{(\boldsymbol{\mu}_{10} \cdot \mathbf{R})(\boldsymbol{\mu}_{20} \cdot \mathbf{R})}{R^5}$$

Figure 9.18 According to exciton theory, two electronic transitions on *different* chromophores that are degenerate without interaction will split in energy when their transition dipoles interact through Coulombic interaction.

$$E_0 = 0: \qquad \psi_0 = \phi_0$$

$$E_+ = e + V: \qquad \psi_+ = \frac{(\phi_1 + \phi_2)}{\sqrt{2}}$$

$$E_- = e - V: \qquad \psi_- = \frac{(\phi_1 - \phi_2)}{\sqrt{2}}$$

$$E_{12} = 2e: \qquad \psi_{12} = \phi_{12}$$

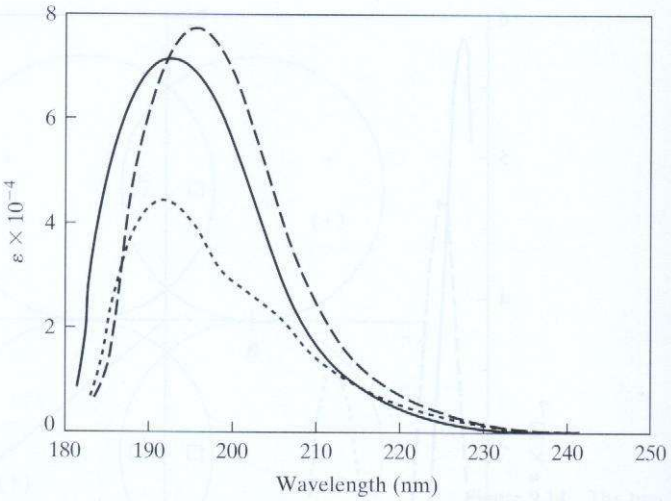


Figure 9.16 The electronic absorption spectra for poly-L-lysine hydrochloride in aqueous solution as a random coil at pH 6.0, 25°C (———); α-helix at pH 10.8, 25°C (- - -); β-strand at pH 10.8, 52°C (— —). [Adapted from K. Rosenheck and P. Doty (1961) *Proc. Natl. Acad. Sci. USA* **47**, 1775–1785.]

ips interacting with one another. Exciton theory was introduced briefly

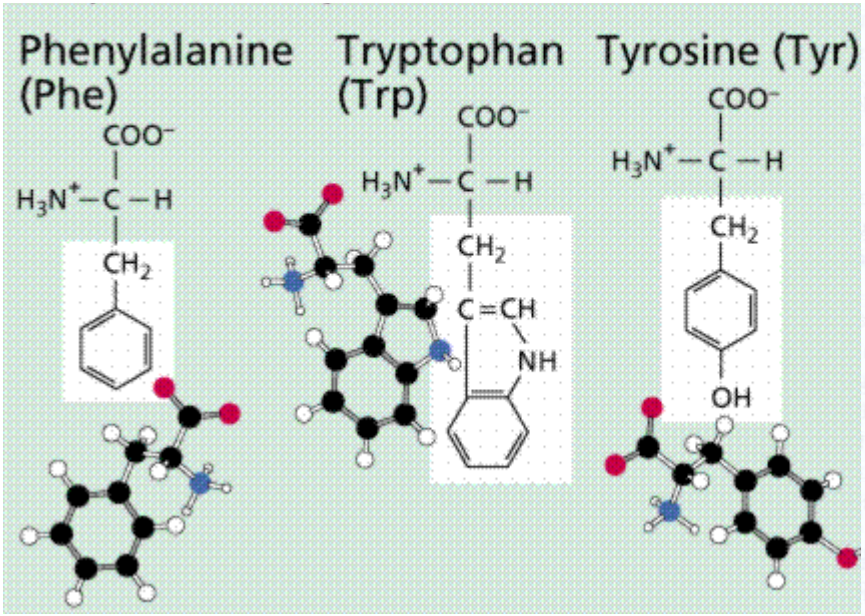
Proteins as chromophores (AA sidechains):

All the other AA's masked by the peptide bond, except for...

Amino Acid Side Chains

Table 11-1. Absorption of Amino Acid Side Chains

Trp (tryptophan); not present in large amounts in proteins	240-290 nm	most intense	absorption complex (3 transitions of indole ring)
Tyr (tyrosine)	274 nm	$\pi-\pi^*$ ($\varepsilon_{max} \sim 1400$)	analogous to 271 nm absorption in phenol
Phe (phenylalanine)	250 nm	weak $\pi-\pi^*$ symmetry forbidden	analagous to 256 nm absorption in benzene



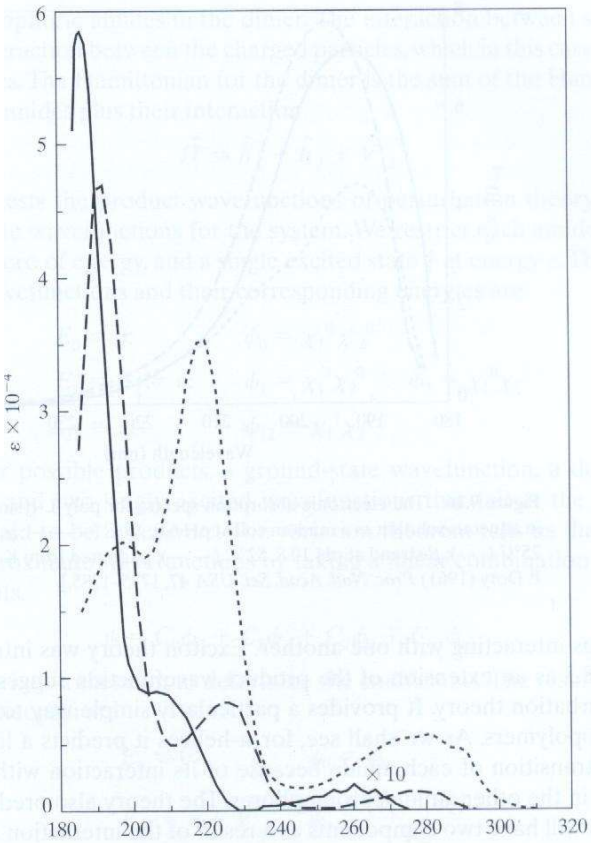


Figure 9.15 The electronic absorption spectra for the aromatic side chains, in aqueous solution, pH 5 to 7: phenylalanine (———); tyrosine (– – – –); tryptophan (····). [From data in *Practical Handbook of Biochemistry and Molecular Biology* (1989), pp. 81–83 (compiled by Elmer Mihalyi, G. D. Fasman, ed.), CRC Press; D. B. Wetlaufer (1962), "Ultraviolet Spectra of Proteins and Amino Acids," *Adv. Protein Chem.* **17**, 303–390; R. Sussman and W. B. Gratzer, personal communication.]

Simulation Methods

Thermodynamical Ensembles

- **Def:** A thermodynamical ensemble is a collection of microscopic states, i.e. of points in the phase space of the system, that all realize an identical macroscopic or thermodynamical state.
- Example of ensembles:
  - Microcanonic: fixed (N,V,E)
  - Canonic: fixed (N,V,T)
  - Constant P-T: fixed (N,P,T))
  - Grand Canonic: fixed (μ, P, T)

Expectation value

- The macroscopic value of a given observable is the average of the observable over all the microstates, weighted by the probability of the microstate. In the canonical ensemble

$$\mathcal{O} = <\mathcal{O}>_{Ensemble} = \frac{1}{Z}\sum_i \mathcal{O}_i e^{-\beta E_i}$$

where

$$Z = \sum_i e^{-\beta E_i}$$

- $P_i = 1/Z e^{-\beta Ei}$ is the probability of the microstate i
- *Cave:* These quantities are extremely expensive to compute

Ergodicity

- Time average (MD):

$$\mathcal{O} = <\mathcal{O}>_{Time} = \lim_{\tau\to\infty} \frac{1}{\tau}\int_{t=0}^{\tau} \mathcal{O}(\boldsymbol{r}(t), \boldsymbol{p}(t))dt$$

- Ergodic hypothesis:

$$<\mathcal{O}>_{Ensemble} = <\mathcal{O}>_{Time}$$

- here, the averages are not weighted by the probability of the microstate since the MD simulation will sample the states according to their probablity

Connection with thermodynamics

- The partition function serves as a bridge between the microstates of a macroscopic system and the thermodynamical properties of that system.
- Examples:
  - the Helmholtz free energy is expressed as

$$A(N, V, T) = -kT \ln Z$$

  - the average pressure is expressed as

$$\bar{p} = kT\left(\frac{\partial \ln Z}{\partial V}\right)_{N,T}$$

Molecular dynamics methods

Equations of motion

- The equation of motions for particle i (i=1,..,N) are given by

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d^2}{df^2} \vec{r}_i$$

- The force $\vec{F}_i$ acting on particle i is given by

$$\vec{F}_i = -\vec{\nabla} V_i$$

therefore

$$-\vec{\nabla} V_i = m_i \frac{d^2}{df^2} \vec{r}_i$$

Integration of the equations of motion (Verlet)

- For small δt increments, one can use a Tailor expansion of the positions $\vec{r}_i$

$$\vec{r}_i(t+\delta t) = \vec{r}_i(t) + \vec{v}_i(t)\delta t + \frac{1}{2}\vec{a}_i(t)\delta t^2$$

and

$$\vec{r}_i(t-\delta t) = \vec{r}_i(t) - \vec{v}_i(t)\delta t + \frac{1}{2}\vec{a}_i(t)\delta t^2$$

- summing the two equations, one gets

$$\vec{r}_i(t+\delta t) = 2\vec{r}_i(t) - \vec{r}_i(t-\delta t) + \vec{a}_i(t)\delta t^2$$

$\rightarrow$ knowing $\vec{r}_i(t-\delta t)$, $\vec{r}_i(t)$ and $\vec{a}_i(t)$ one can compute $\vec{r}_i(t+\delta t)$.

- Other algorithm are used: Leap-frog, Beeman's, Velocity Verlet.

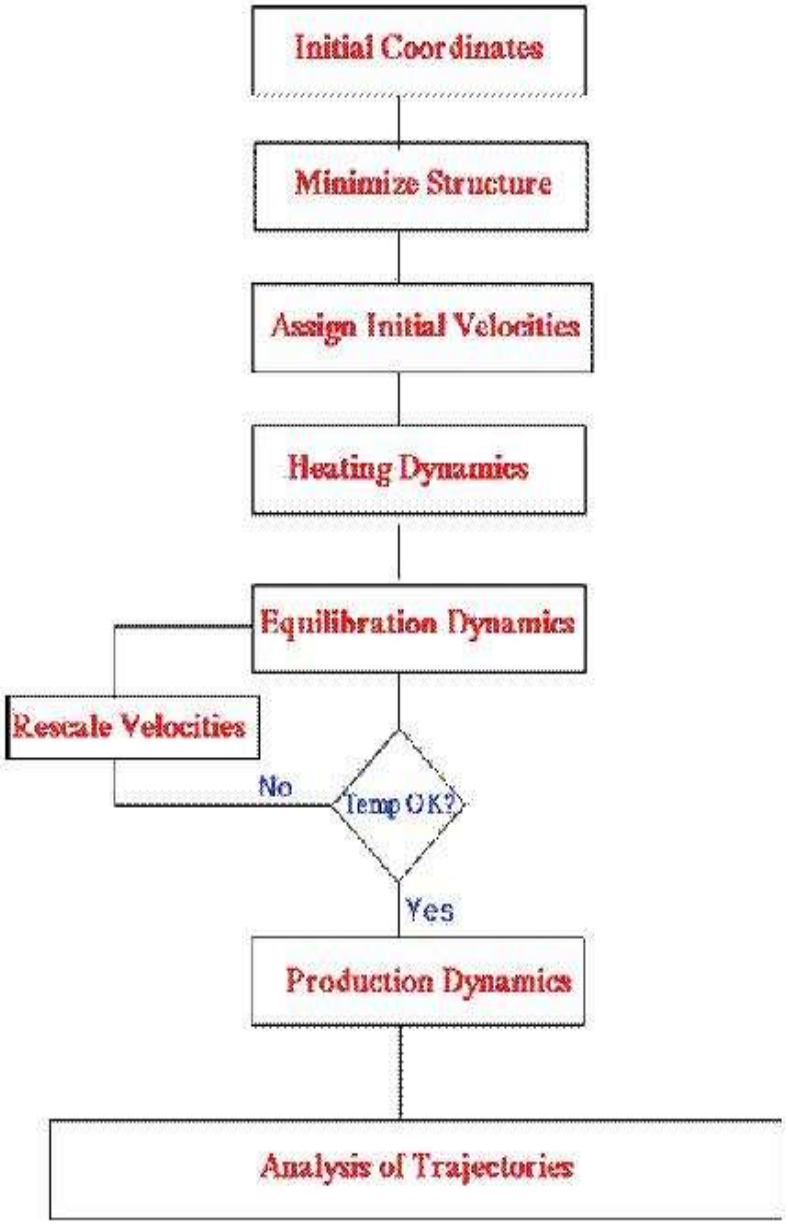


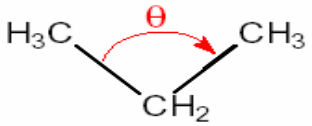
Figure 10:

Available Force-Fields

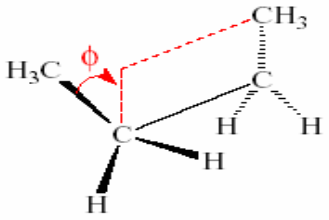
BondedTerms

$$\frac{1}{2}K_b(b - b_0)^2$$

$$\frac{1}{2}K_\theta(\theta - \theta_0)^2$$

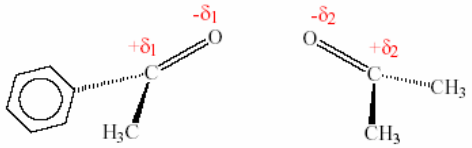


$$K_\phi[1 + \cos(n\phi - \delta)]$$

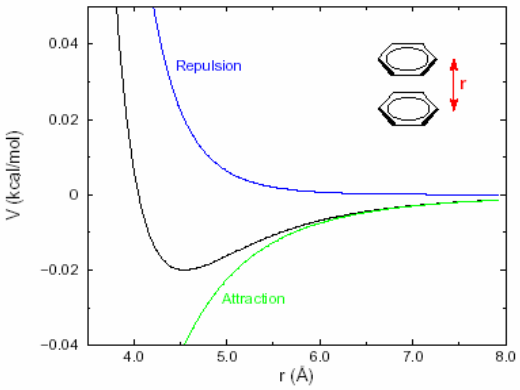


Non-Bonded Terms

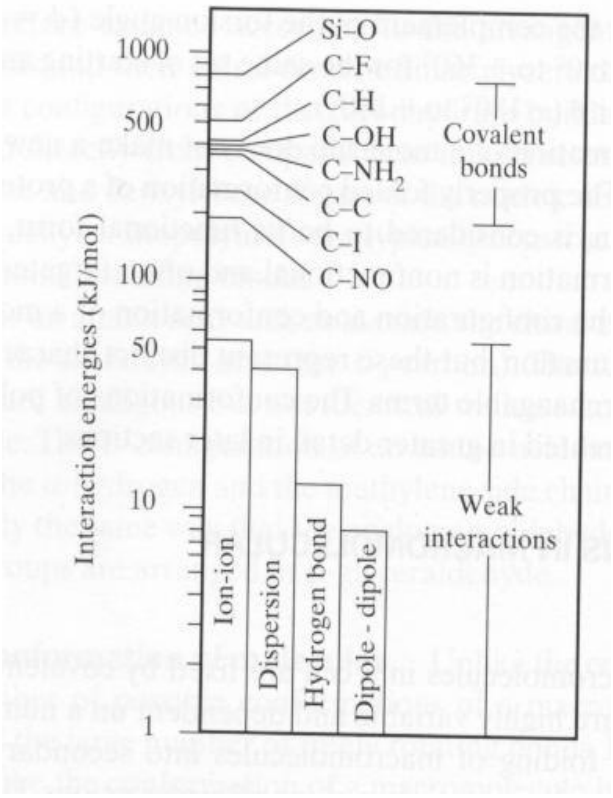
$$\frac{q_i q_j}{r \epsilon_R}$$



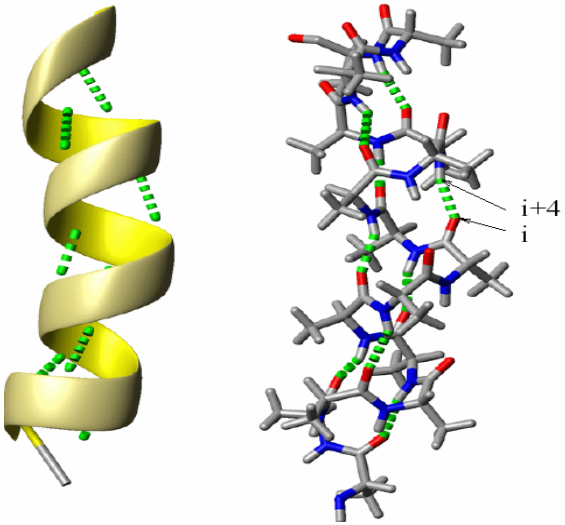
$$4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right]$$



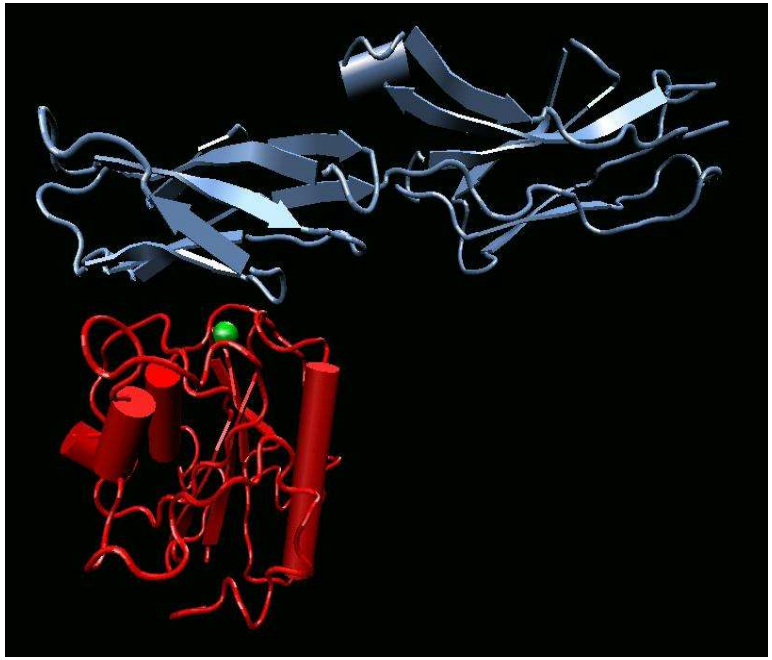
CHARMM
AMBER
GROMOS
OPLS



Structure



Recognition between proteins



Derived Interactions

The following are accounted for by the interactions described previously, i.e electrostatic and van der Waals

Hydrogen bonds

- Interaction of the type

$$—D^{-\delta}—H^{+\delta} \ldots {}^{-\delta}A—$$

- The nature of this interaction is a dipole-dipole interaction
  → already accounted for in the non-bonded energy terms
  → modern force-fields do not include a specific term for hydrogen bonds

Hydrophobic Effect

- This is a collective effect: contact between solvent and apolar is highly unfavorable (loss of solvent-solvent hydrogen bonds)
  → Apolar media reorganize to minimize the solvent exposed surface, e.g. oil drop in water.

Time Scales:

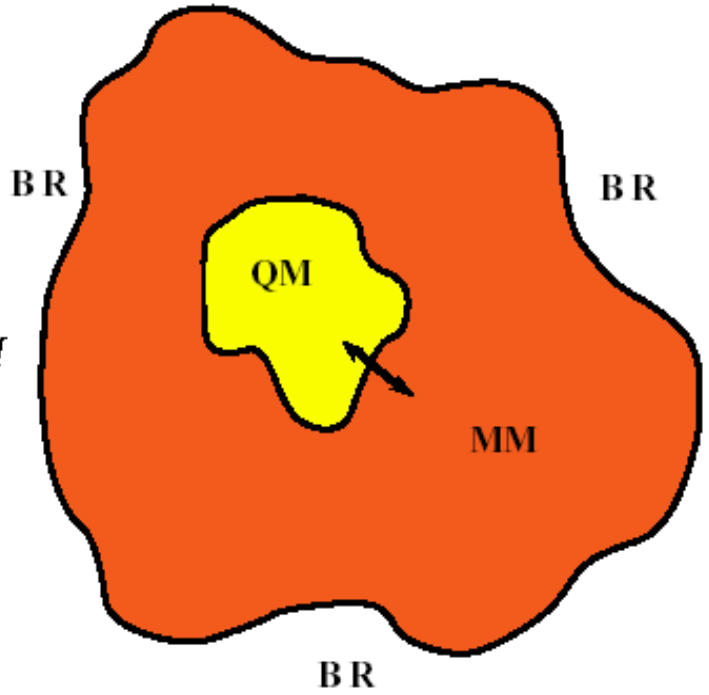
- Local Motions (0.01 to 5 Å, 10^{-15} to 10^{-1} s)
  - Atomic fluctuations
  - Sidechain Motions
  - Loop Motions
- Rigid Body Motions (1 to 10Å, 10^{-9} to 1s)
  - Helix Motions
  - Domain Motions (hinge bending)
  - Subunit motions
- Large-Scale Motions (> 5Å, 10^{-7} to 10^{4} s)
  - Helix coil transitions
  - Dissociation/Association
  - Folding and Unfolding

QM/MM Hamiltonian

For Local Phenomena

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{QM/MM} + \hat{H}_{MM} + \hat{H}_{boundary} + \hat{H}_{restraints}$$

$$\hat{H}_{QM/MM} = \hat{H}^{elec}_{QM/MM} + \hat{H}^{vdW}_{QM/MM} + \hat{H}^{bonded}_{QM/MM}$$



Warshel , A.; Levitt, M. J. Mol. Biol., **1976**, *103*, 227

Field, M.J.; Bash, P.A.; Karplus, M. J. Comp. Chem., **1990**, *11*, 700

Gao, J.; Xio, X. Science **1992**, *258*, 631.

Tuñón, I.; Martins-Costa, M.T.C.; Milloy, C.; Ruiz-López, M.F.; Rivail, J.-L.; J. Comp. Chem. **1996**, *17*, 19.

QM/MM Electrostatics

$$\hat{H}^{elec.}_{QM/MM} = -\sum_{i,M} \frac{q_M}{r_{iM}} + \sum_{\alpha,M} \frac{Z_\alpha q_M}{R_{\alpha M}}$$

SCF

1. Calculate Integrals needed to form Fock Matrix F
2. Calculate the overlap matrix S
3. Diagonalize S : $\mathbf{W}^{+}\mathbf{S}\mathbf{W} = \mathbf{D}$
4. Form $\mathbf{S}^{-1/2} = \mathbf{W}\mathbf{D}^{1/2}\mathbf{W}^{+}$
5. Form the F matrix
6. Form $\mathbf{F}' = \mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{-1/2}$
7. Diagonalize F' for the MO eigenvalues E: $\mathbf{V}^{+}\mathbf{F}'\mathbf{V} = \mathbf{E}$
8. Back transform V to obtain MO coefficients C, $\mathbf{C} = \mathbf{S}^{-1/2}\mathbf{V}$
9. Form the density matrix P
10. check for convergence in total energy and P

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda}\sum_{\sigma} P_{\lambda\sigma}[<\mu\nu|\lambda\sigma> - \frac{1}{2} <\mu\lambda|\nu\sigma>]$$

$$H_{\mu\nu} = <\mu\left|-\frac{1}{2}\nabla^2\right|\nu> - \sum_{B} Z_B <\mu\left|\frac{1}{r_{iB}}\right|\nu> - \sum_{M} q_M <\mu\left|\frac{1}{r_{iM}}\right|\nu>$$

Gaussian Integral
Numerical Integral
Semi-empirical Form

Semi-empirical QM/MM Electrostatics (NDDO)

Neglet of Differential Diatomic Overlap (MNDO, AM1, PM3, MNDO/d)

$$\chi_\mu^A(1)\chi_\nu^B(1)d\tau(1) \quad \rightarrow \quad \delta(A,B)\chi_\mu^A(1)\chi_\nu^B(1)d\tau(1)$$

Energy Terms and Interactions

1. $1C-1e^-$ energies $U_{\mu\mu}$ (as part of $H_{\mu\mu}$)
2. $1C-2e^-$ repulsion integrals $(\mu^A\nu^A,\lambda^A\sigma^A)$ (Coulomb and Exchange)
3. $2C-1e^-$ resonance integrals $\beta_{\mu\nu}=H_{\mu^A\nu^B}$
4. $2C-1e^-$ integrals $\boxed{(\mu^A\nu^A,B)}$ representing electrostatic core - electron attarctions
5. $2C-2e^-$ repulsion integrals $\boxed{(\mu^A\nu^A,\lambda^B\sigma^B)}$
6. $2C$ core-core repulsions $\boxed{E_{AB}^{core}}$

Thiel, W. "Perspectives on Semiempirical MO Theory" in *Advances in Chemical Physics*, V. XCIII, 1996. p. 703.

Electrostatic Interactions in NDDO

Multipole - Multipole Interactions

$$< \mu\nu | \lambda\sigma > \approx M_{\mu\nu} \Longleftrightarrow M_{\lambda\sigma}$$

Atomic Orbitals	Multipole distribution	# of charges
$< ss\|$	Monopole	1
$< sp\|$	Dipole	2
$< pp\|$	Monopole + linear quadrupole	4
$< pp'\|$	Square quadrupole	4

$$< ss|ss > = \frac{27.21}{\sqrt{(R + c_A + c_B)^2 + \frac{1}{4}(\frac{1}{A_A} + \frac{1}{A_B})^2}}$$

For R_{AB} -> 0 should go to the 1C-2e- Integral

To get a balance for the different electrostatic atractions and repulsions:

$$(\mu^A \nu^A, B) \longrightarrow (\mu^A \nu^A, s^B s^B)$$

$$E_{AB}^{core} \longrightarrow (s^A s^A, s^B s^B).$$

QM/MM Electrostatic Interactions in NDDO

1 e- Integrals

$$I_{\mu\nu} = -q_M < \mu_Q \nu_Q | s_M s_M >$$

Core /MM interaction = core-core interaction

$$E_{QM/MM}^{charge/core} = \sum_{QM/MM}^{charge/core} \frac{Z_\alpha q_M}{R_{\alpha M}} = \sum_{Q,M} < s_Q s_Q | s_M s_M > (1 + f(R_{QM}))$$

Scaling Functions:

$$E_N^{MNDO}(A,B) = Z_A Z_B < s_A s_A | s_B s_B > \left[1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}\right]$$

$$E_N^{AM1}(A,B) = E_N^{MNDO}(A,B) + \left(\frac{Z_A Z_B}{R_{AB}}\right)\left[\sum_k a_{kA} e^{-b_{kA}(R_{AB}-c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB}-c_{kB})^2}\right]$$

Van der Waals Interactions

$$\widehat{H}^{v.d.Waals}_{QM/MM} = \sum_{\alpha,M} \left[\frac{A_{\alpha M}}{R^{12}_{\alpha M}} - \frac{B_{\alpha M}}{R^{6}_{\alpha M}} \right]$$

Neccesary: Cl- vs Br-

Combination Rules (Atomic Terms)

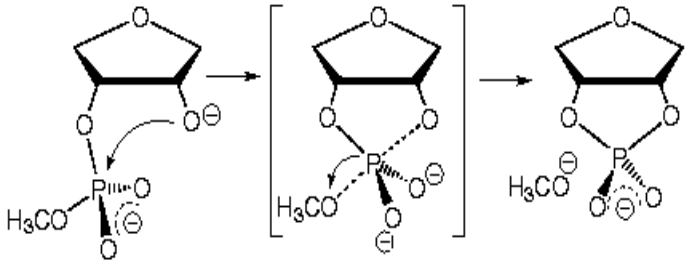
$$A_{ij} = \epsilon_{ij}\sigma_{ij}^{12} \quad ; \quad B_{ij} = 2\epsilon_{ij}\sigma_{ij}^{6}$$

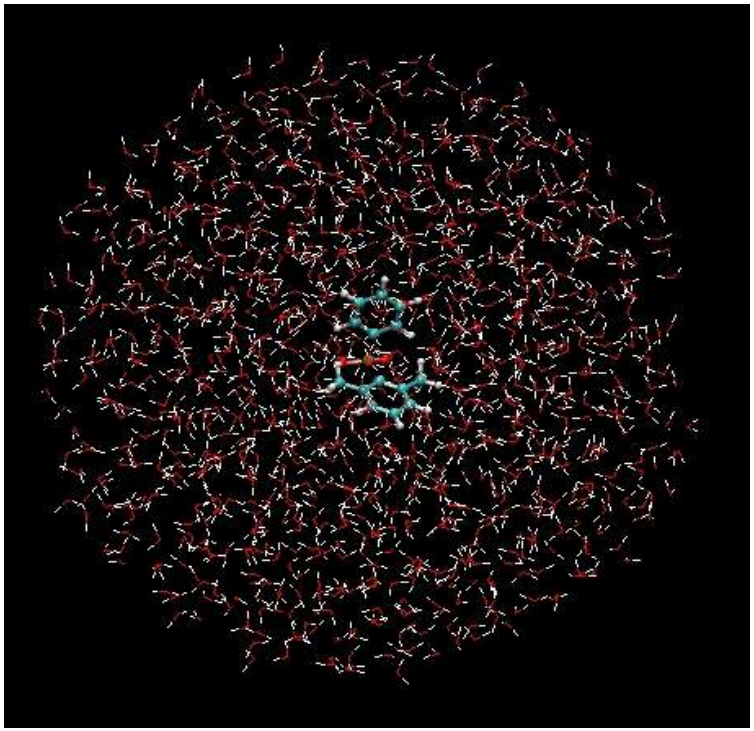
$$\epsilon_{ij} = (E_{min_i}E_{min_j})^{1/2} \quad ; \quad \sigma_{ij} = \frac{1}{2}(R_{min_i} + R_{min_j})$$

Classical Adjustable Term

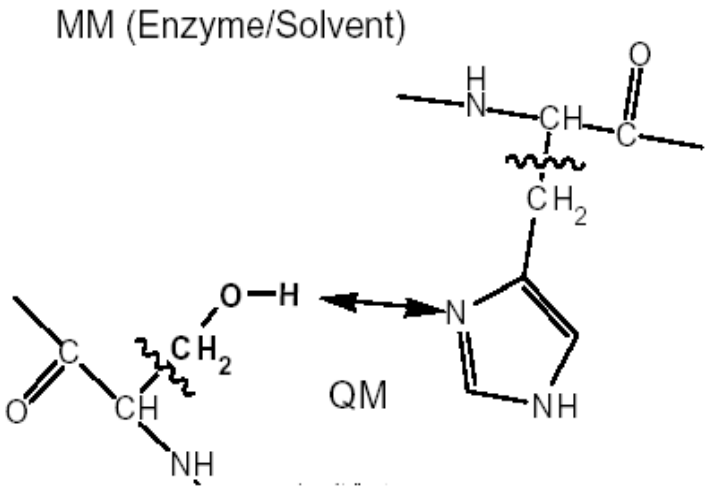
QM/MM Boundary

No Need to cut through Bonds





Need to cut through Bonds



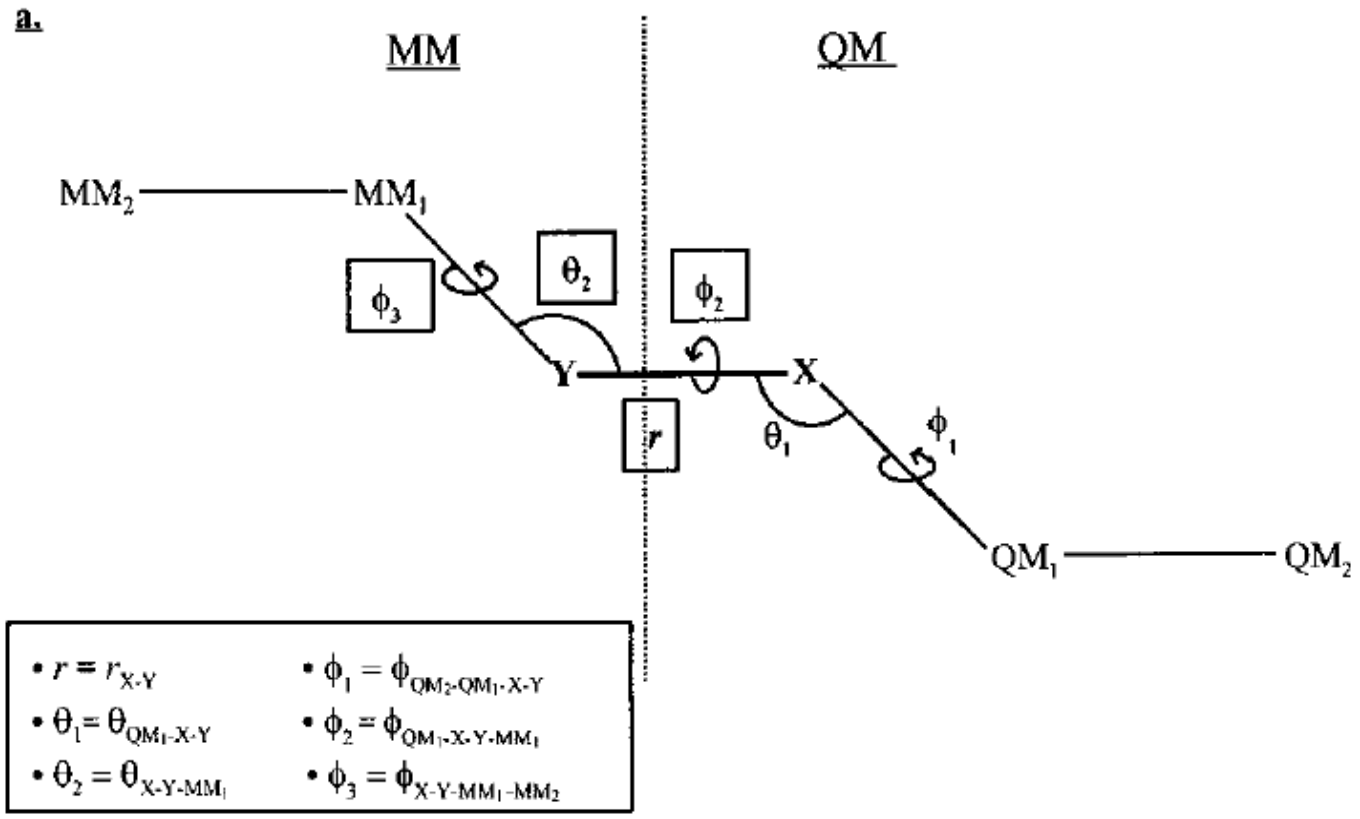
Gregersen G., Lopez X., York D.M., JACS, **2003**, *125*, 7178.
Gregersen G., Lopez X., York D.M., J. Am. Chem. Soc., **2004**, *126*, 7504

QM/MM Bonded Classical Terms

Bonded Classical terms for QM atoms are removed.

For MM atoms are kept

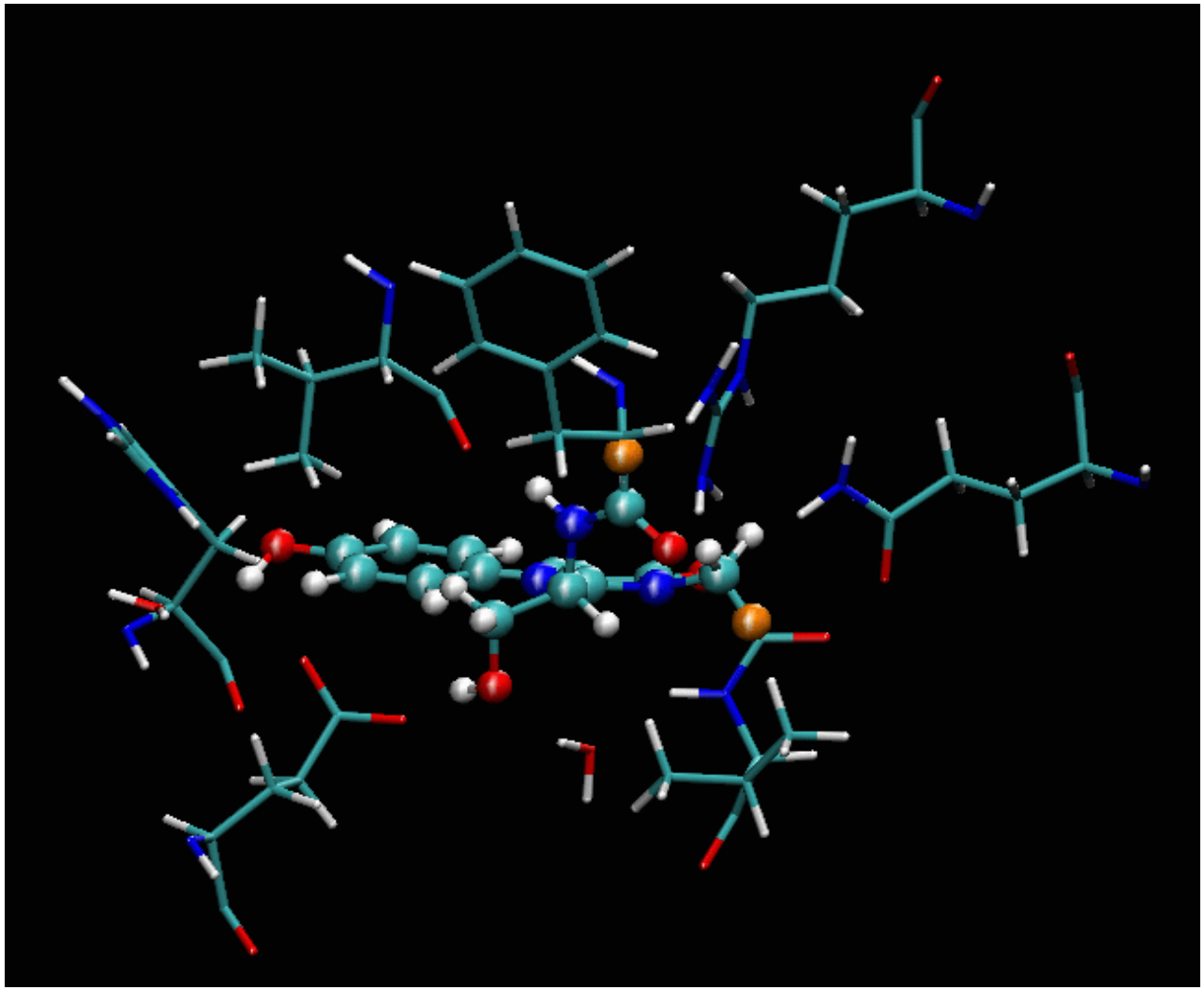
Mixed QM/MM Terms? = Also Kept.



Reuter, N.; Dejaegere, A.; Maigret B.; Karplus M.; J. Phys. Chem. A **2000**, *104*, 1720.

QM/MM Boundary

H-Link Atom



Part of the QM

Placed along Bond

QQ vs HQ

Capping-Atom

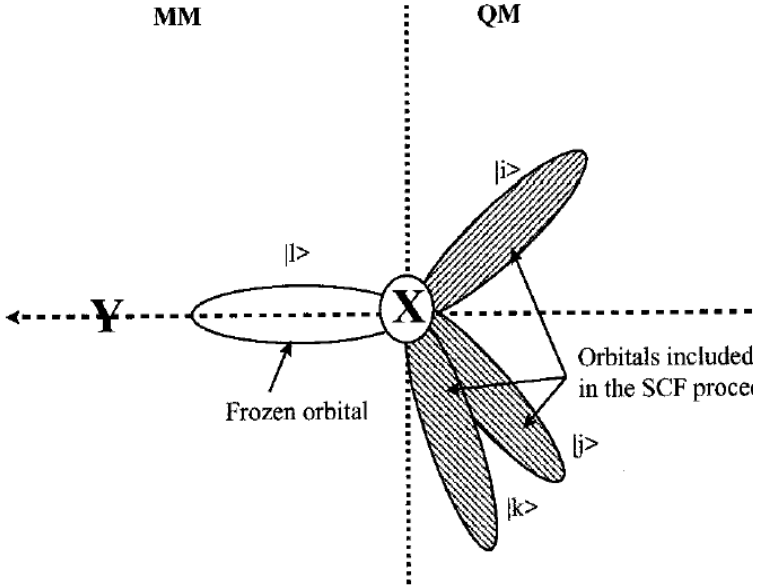
Reuter, N.; Dejaegere, A.; Maigret B.; Karplus M.; J. Phys. Chem. A **2000**, *104*, 1720.

Field, M.J.; Albe, M.; Bret C.; Proust-De Martin F.; Thomas, A. J. Comp. Chem. A **2000**, *21*, 1088.

QM/MM Boundary

Hybrid- or localized-orbital methods

$$[H] = \begin{bmatrix} a_{14} & 0 & 0 & -a_{11} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a_{11} & 0 & 0 & a_{14} \end{bmatrix}$$



$$\begin{bmatrix} |i\rangle \\ |j\rangle \\ |k\rangle \\ |l\rangle \end{bmatrix} = [H][R] \begin{bmatrix} |s\rangle \\ |x\rangle \\ |y\rangle \\ |z\rangle \end{bmatrix}$$

$$[R] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ 0 & -\sin\phi & \cos\phi & 0 \\ 0 & \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{bmatrix}$$

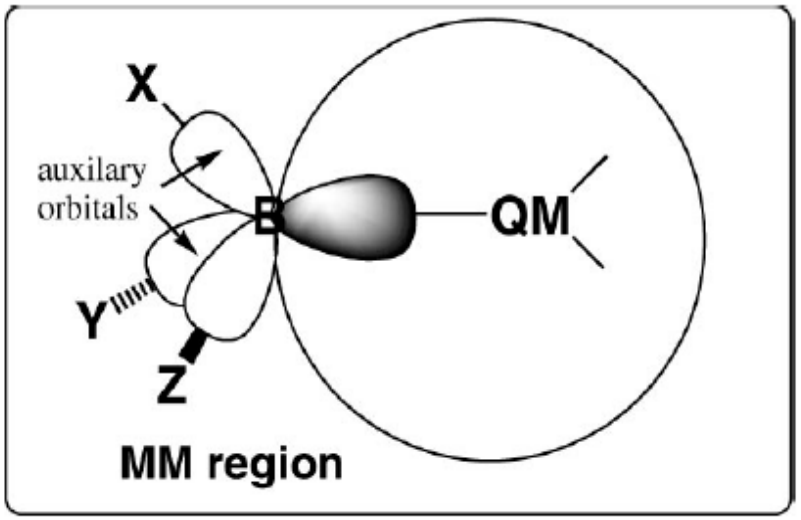
Construction of the Fock Matrix: P_{ll} frozen.

$$F^S_{\mu\nu} = H_{\mu\nu} + \sum_\lambda \sum_\sigma P_{\lambda\sigma}\left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu)\right] + \sum_l P_{ll}\left[(\mu\nu|ll) - \frac{1}{2}(\mu l|l\nu)\right]$$

$$= + \sum_M q_M(\mu\nu|s_M s_M)$$

(N-L)x(N-L) Fock Matrix Built and Diagonalize $\underline{\underline{F'_T}} = \underline{\underline{B}}^{-1}\underline{\underline{F'B}}$

Generalized-Hybrid Orbial (GHO)



Parameters for B adjusted

item	AM1 or CHARMM	GHO
β_s	−15.715 783	−5.500 524
β_p	−7.719 283	−14.666 638
U_{pp}	−39.614 239	−38.703 112
MM Bond Stretching Parameters (Å)		
$R_0(CT2-C_B2)^a$	1.530	1.485
$R_0(CT3-C_B2)^a$	1.528	1.478
$R_0(C_B-HA)^a$	1.111	1.091

Reuter, N.; Dejaegere, A.; Maigret B.; Karplus M.; J. Phys. Chem. A **2000**, *104*, 1720.

Théry , V.; Rinaldi D.; Rivail, J.-L., Maigret B. and Ferenczy G.G. Reuter, J. Comp. Chem. A **1994**, *15*, 269.

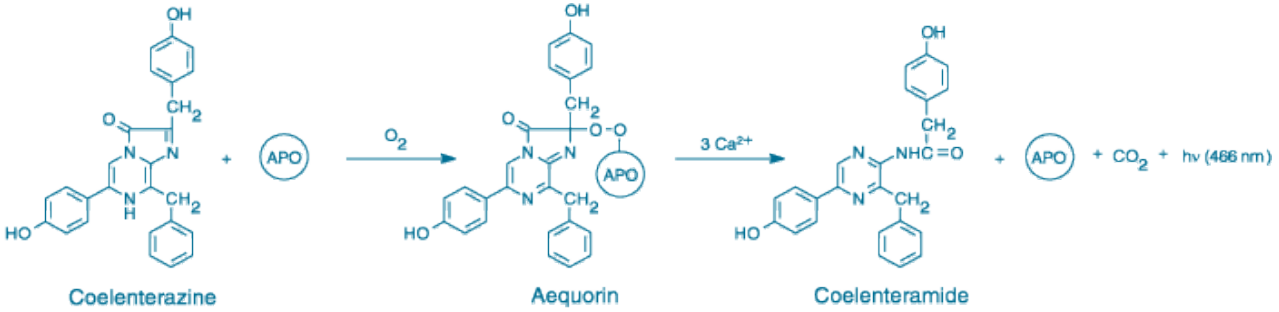
Assfeld, X.; Rivail, J.-L. Chem. Phys. Lett. **1996**, *263*, 100.

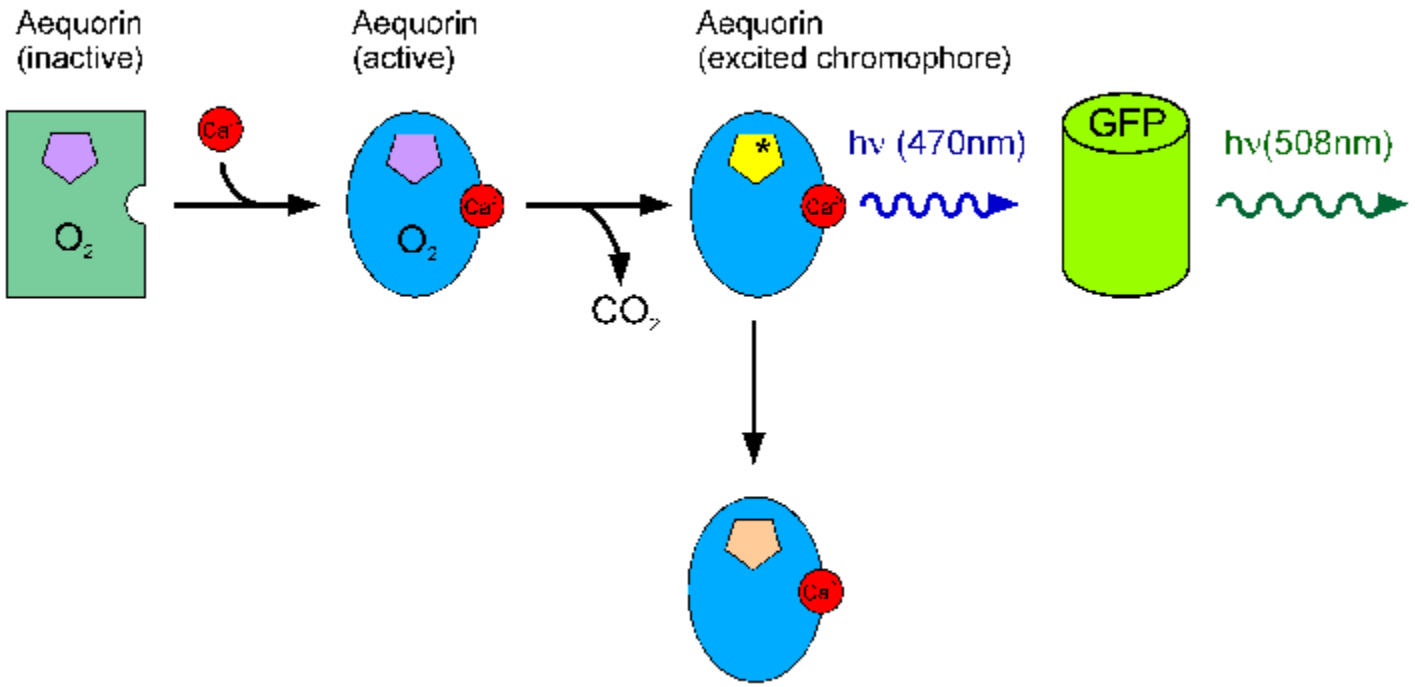
Ferré N.; Assfeld X.; Rivail, J.-L. J. Comp. Chem. A **2002**, *23*, 610.

Gao, J.; Amara, P.; Alhambra, C.; Field, M.J. J. Phys. Chem. A **1998**, *102*, 4714.

An Illustrative Example: Green Fluorescent Protein

Several bioluminescent coelenterates emit green light from a GFP upon mechanical stimulation.

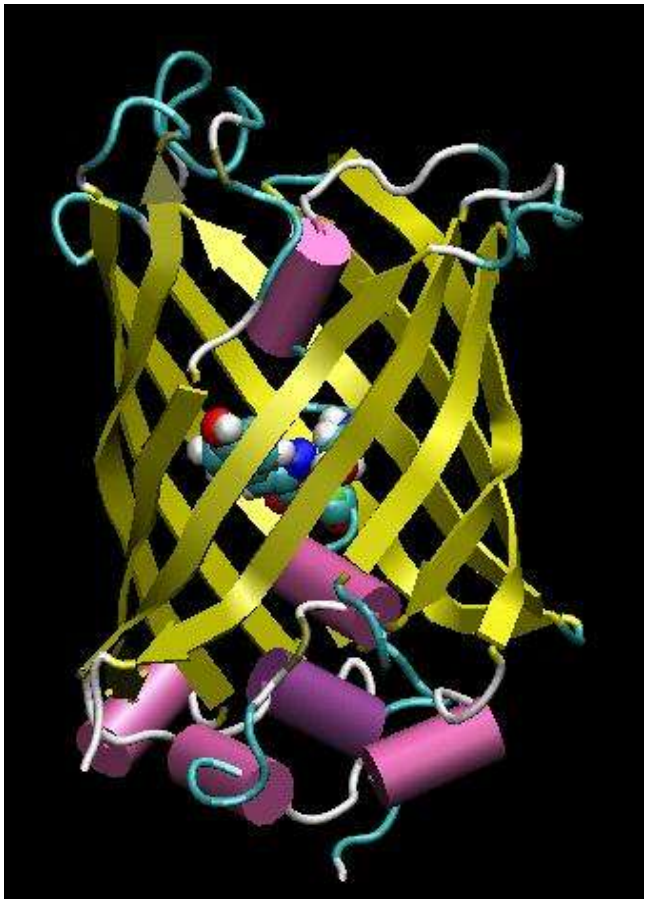




Green Fluorescent Protein

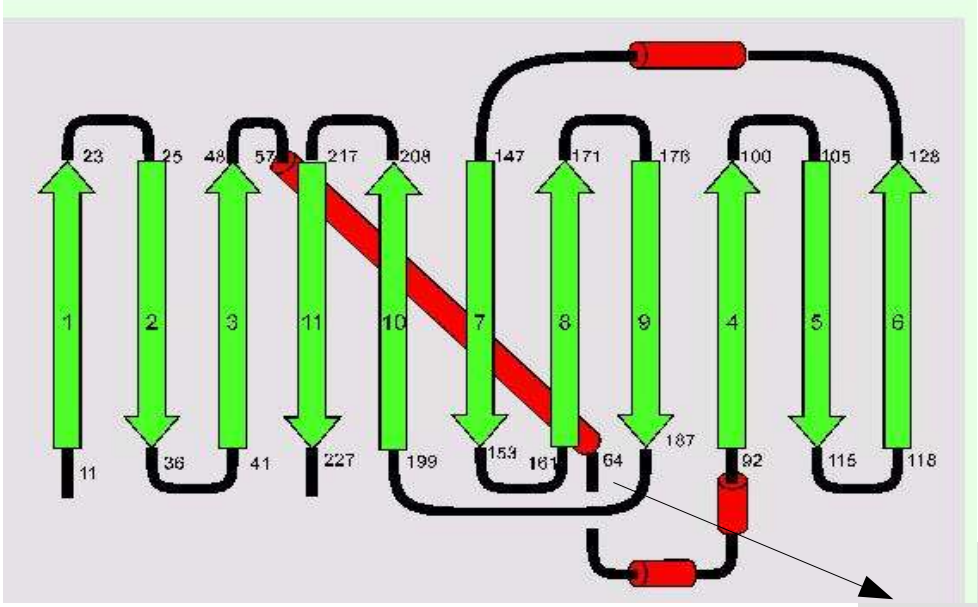
- 238 AA protein forming a β–barrel or β-can
- Chromophore located inside the β barrel (shielded)
- No phrosthetic group
- Info to create the chromophore contained entirely in the gene

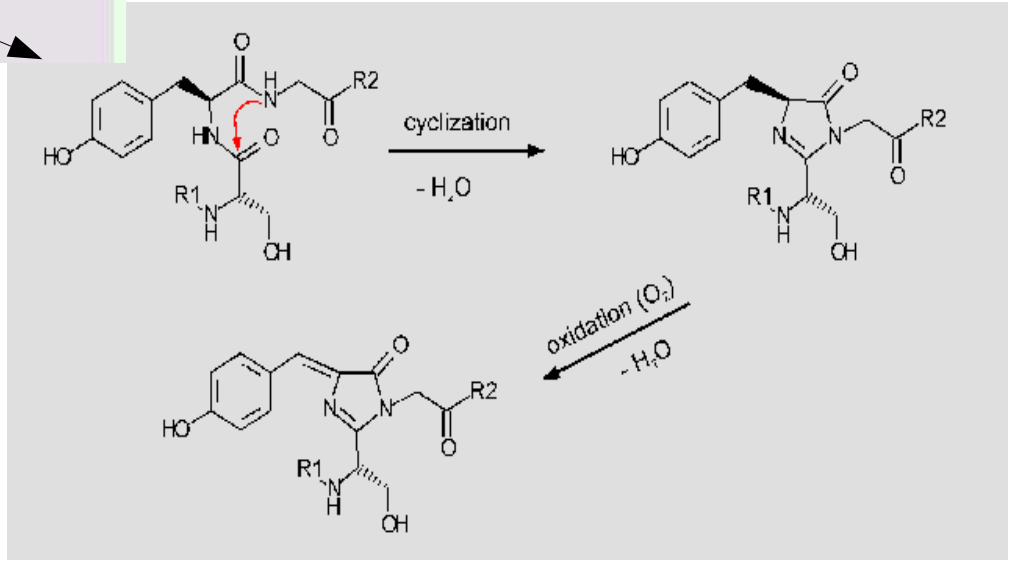
- High Stability: Wide pH, T, salt
- Long Half life: ~20 years
- Resistant to most Proteases
- Active after peptide fusions: reporter protein
- Availability of chromophores variants



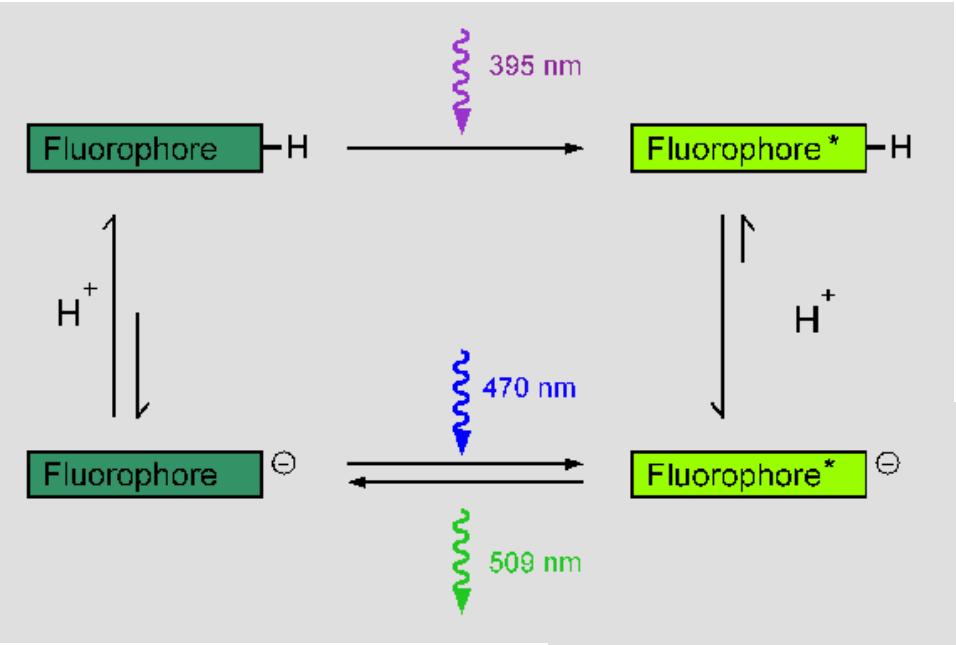
Green Fluorescent Protein

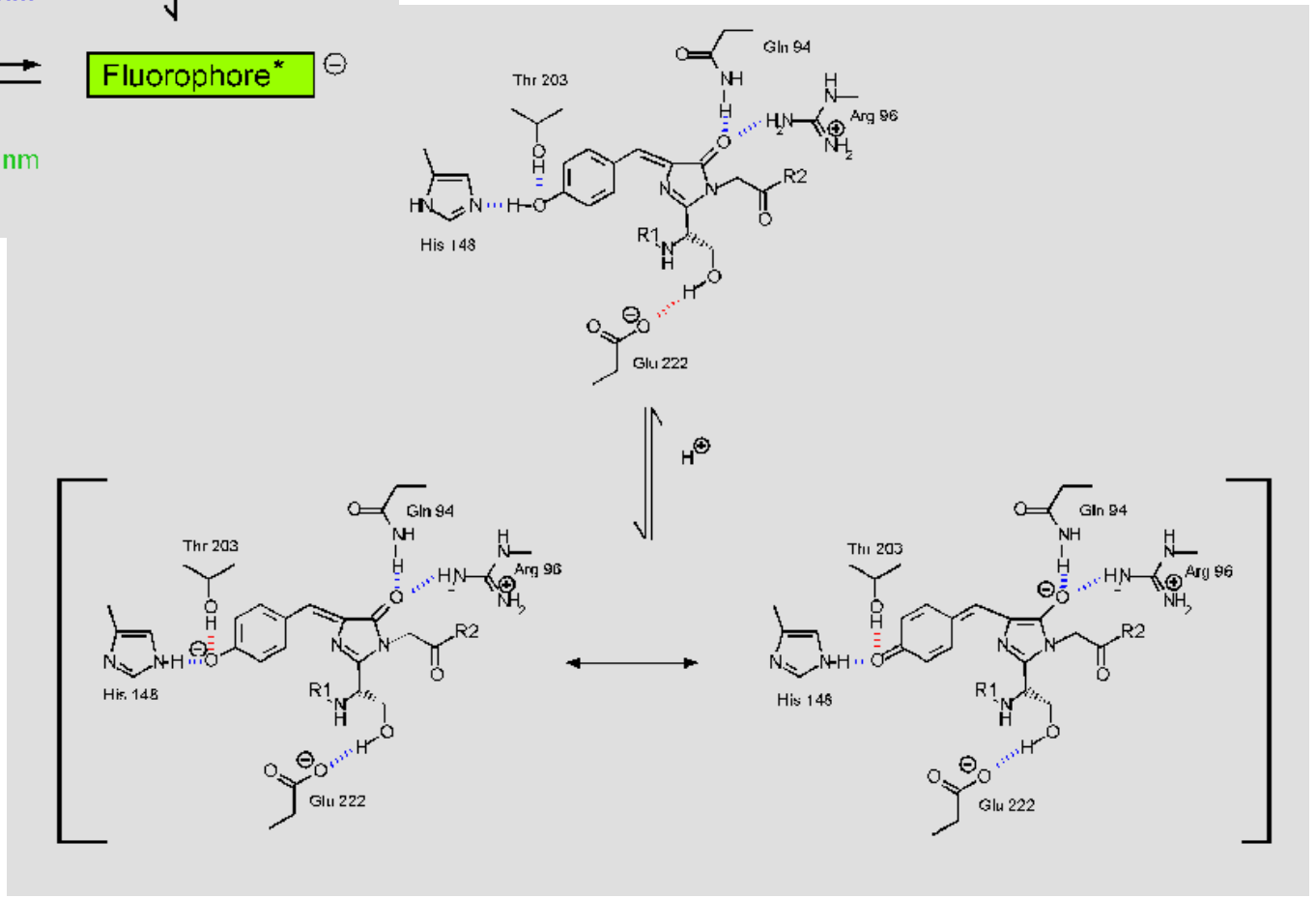
How can be that there is no a phrosthetic group? :GFP is also an **enzyme** that catalyzes the cyclization and oxidation of the Ser-Tyr-Gly triad.

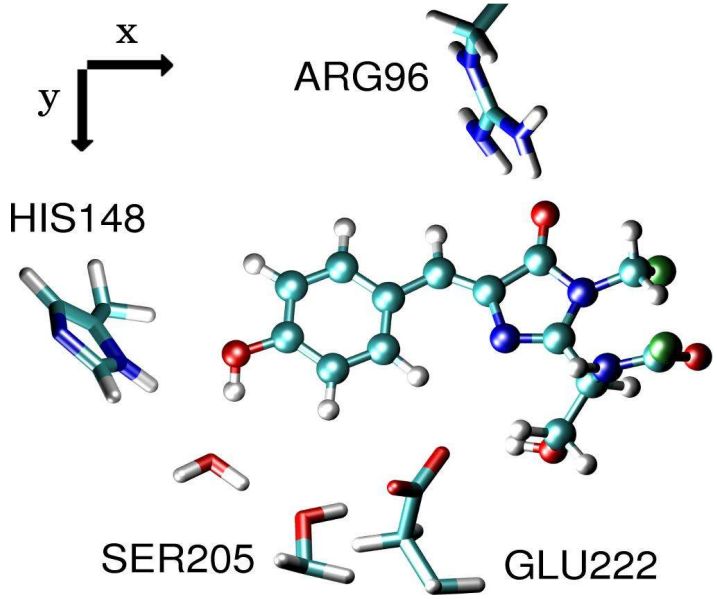


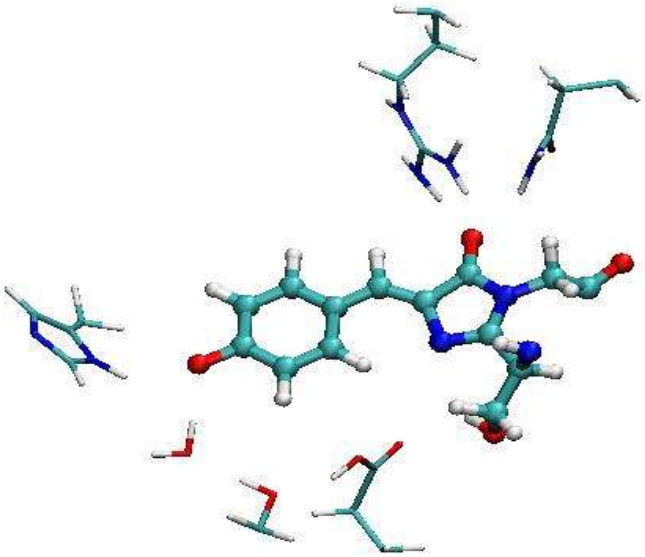


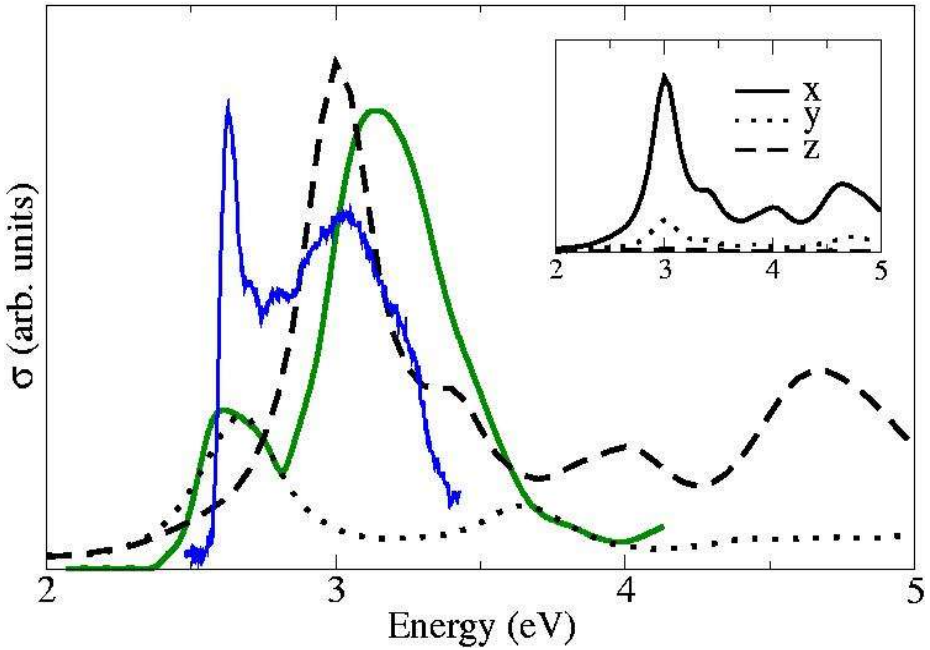
Foster Cycle







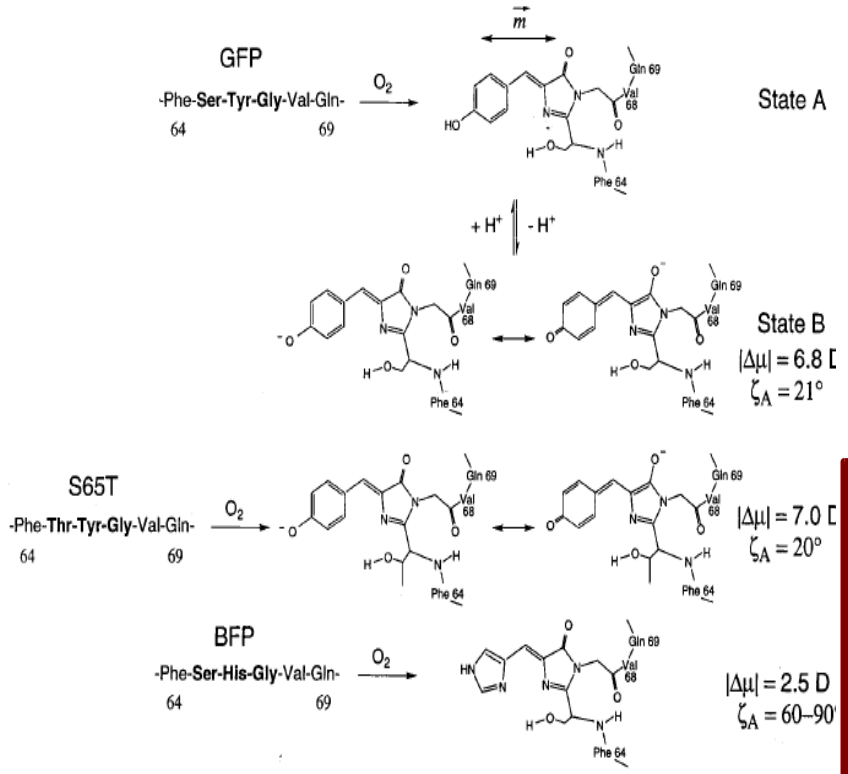


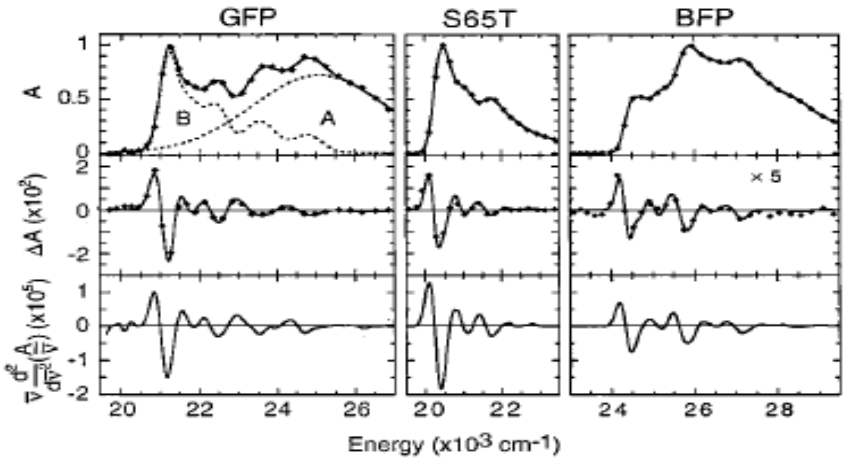


- QM/MM preparation of the structure + Octopus LR
- Different degree of non-planarity for neutral/anionic
- Anisotropy. Absorption along the X directio
- Excellent agreement with the peaks, which confirmed the anionic/neutral Foster cycle.

M.A.L. Marques, X. Lopez, D. Varsano, A. Castro and A. Rubio, *PRL*, **2003**, *90*, 258101

Mutants of the GFP





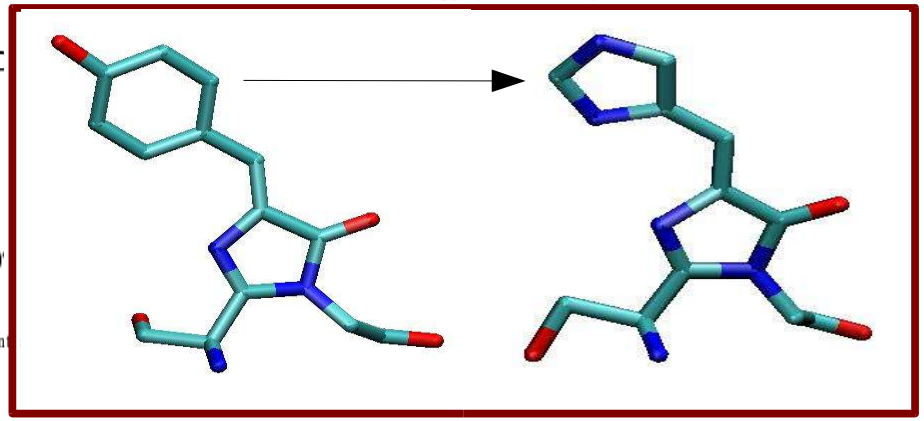


Figure 1. Primary sequences of the chromophore and nearby residues as well as the chromophore structures in GFP, the S65T mutant (Y66H/Y145F). The arrow above the GFP structure is only an approximation to the transition moment direction (*vida infra*).[13]

Chattong, King, Bublitz, Boxer, *PNAS*, **1996**, *93*, 8362

Watchter, King, Heim, Kallio, Tsien, Boxer, *Biochemistry*, **1997**, *36*, 9759

Bublitz, King, Boxer, J*. Am. Chem. Soc*., **1998**, *120*, 9370

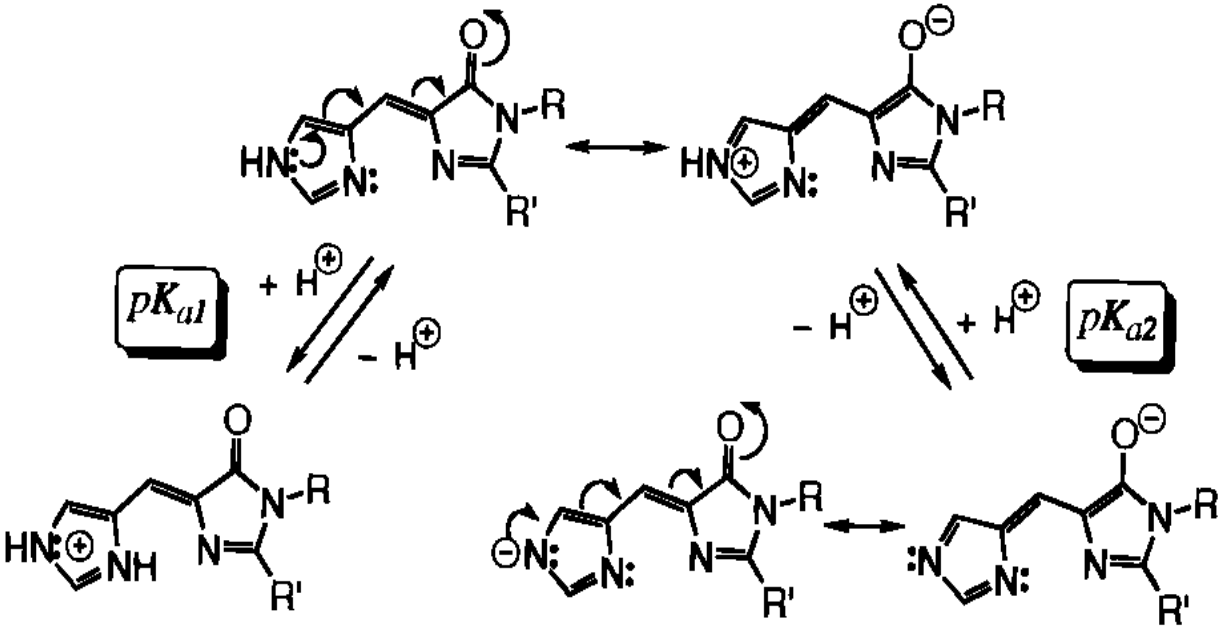
Hanson et al. *Biochemistry*, **2002**, *41*, 15477

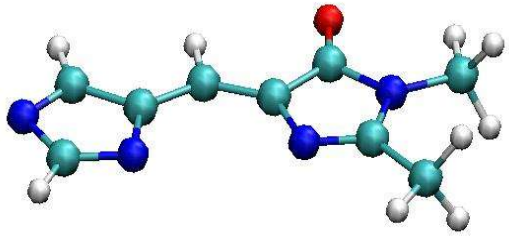
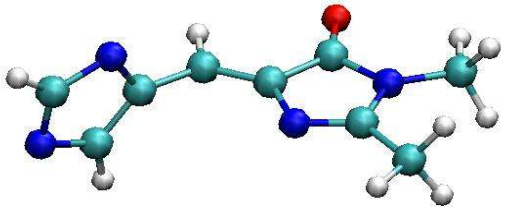
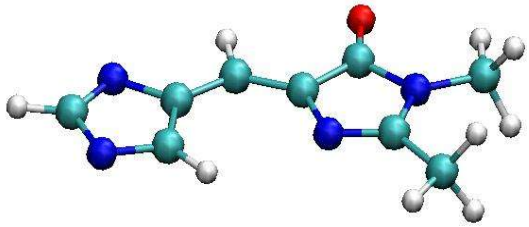
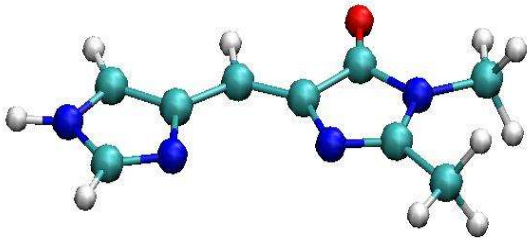
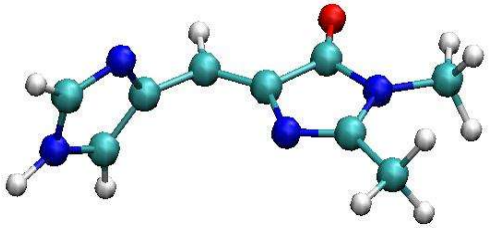
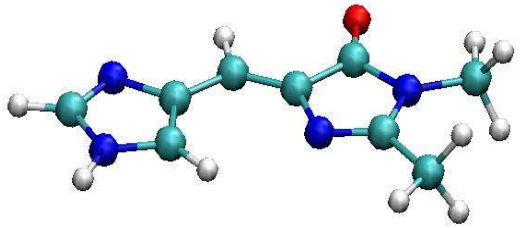
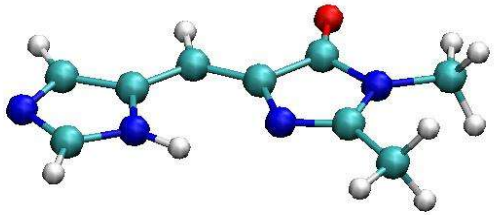
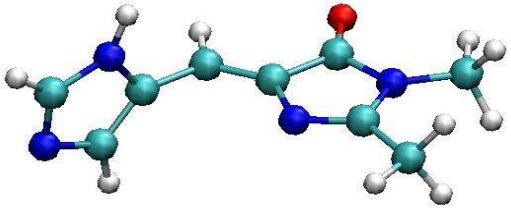
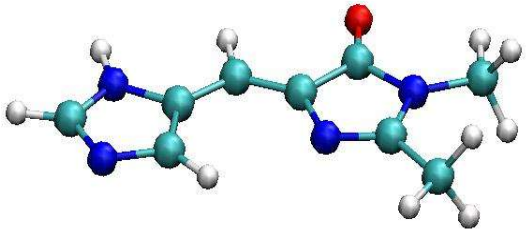
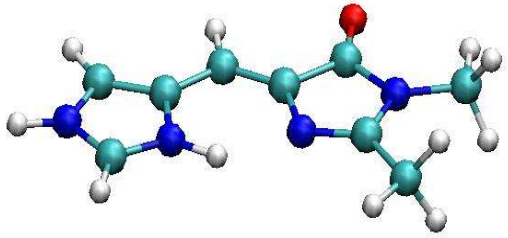
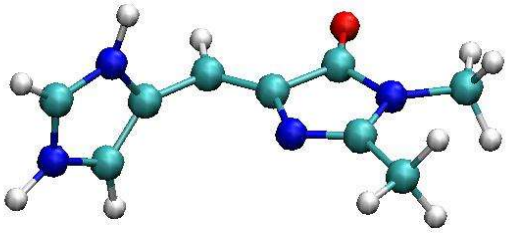
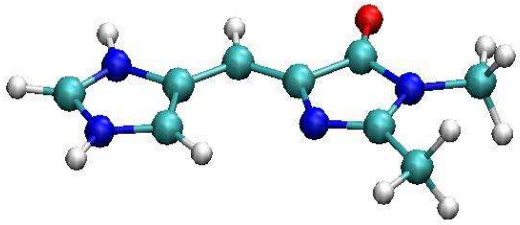
McAnaney et al. *Biochemistry*, **2002**, *41*, 15489

Blue Fluorescent Protein: -Ser-His-Gly-

- Added Complexity due to the higher number of possible protonation states. No proton information from X-Ray
- Experimental pKa's estimations suggest that HSA should be ruled out
- In addition, there are various conformers for each protonation state

Scheme 1: Charge States and Resonance Forms of the BFP Chromophore



	cis	TS	trans
HSA			
HSE			
HSD			
HSP			

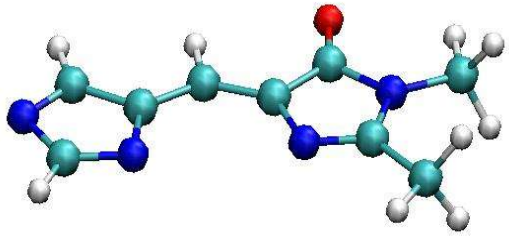
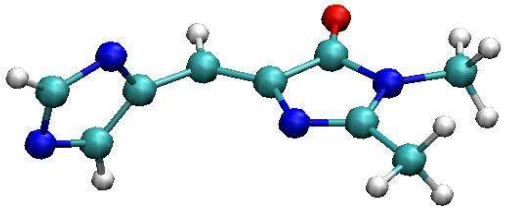
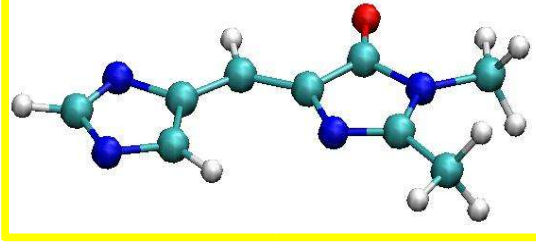
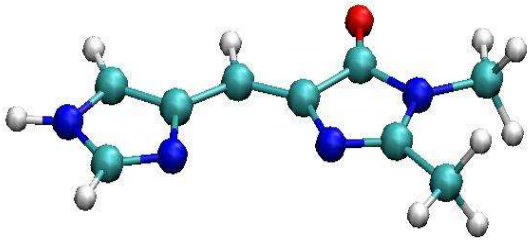
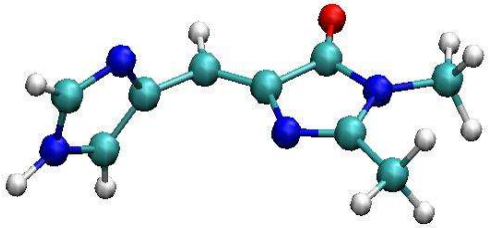
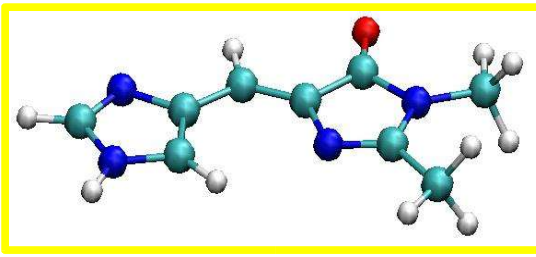
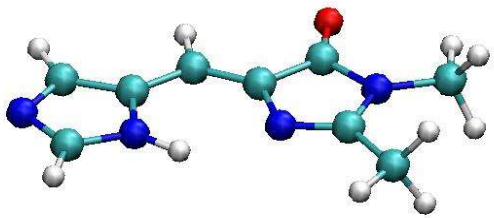
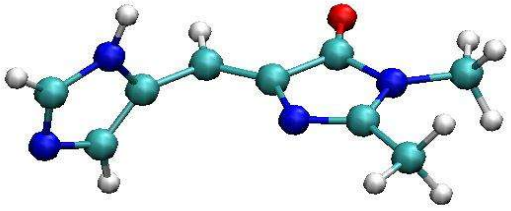
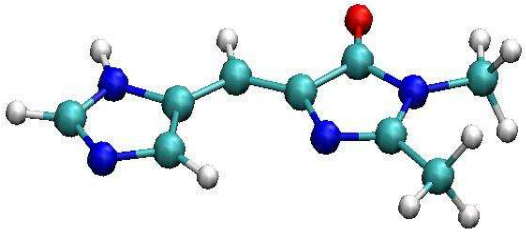
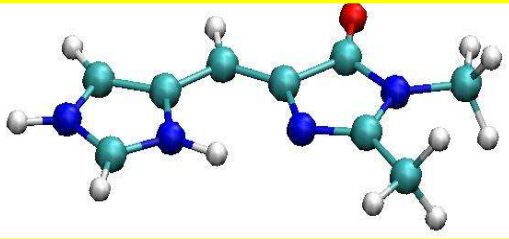
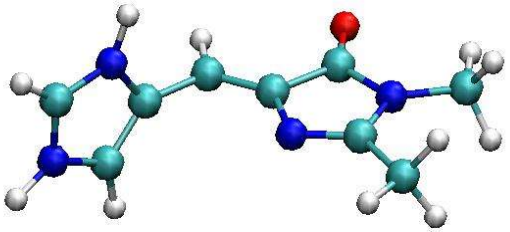
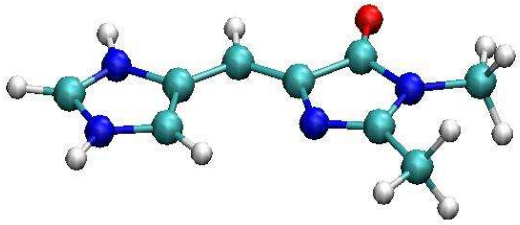
Relative Energies (kcal/mol) at B3LYP

	ΔE_e	ΔE_0	ΔH	$T \cdot \Delta$S	ΔG
Anionic HSA					
HSA cis	0.00	0.00	0.00	0.00	0.00
HSA TS	17.83	17.08	16.80	-0.58	17.38
HSA trans	-3.45	-3.36	-3.37	-0.06	-3.31
Neutral HSE/HSD					
HSE cis	0.00	0.00	0.00	0.00	0.00
HSE TS	5.34	5.05	4.62	-1.08	5.70
HSE trans	-4.76	-4.64	-4.69	-0.23	-4.46
HSD cis	-10.47	-10.01	-10.17	-0.57	-9.60
HSD TS	4.79	4.59	4.17	-1.00	5.17
HSD trans	-4.74	-4.59	-4.60	-0.10	-4.50
Cationic HSP					
HSP cis	0.00	0.00	0.00	0.00	0.00
HSP TS	13.70	13.43	13.16	-0.37	13.54
HSP trans	6.28	6.19	6.37	0.61	5.76

Geometrical Data

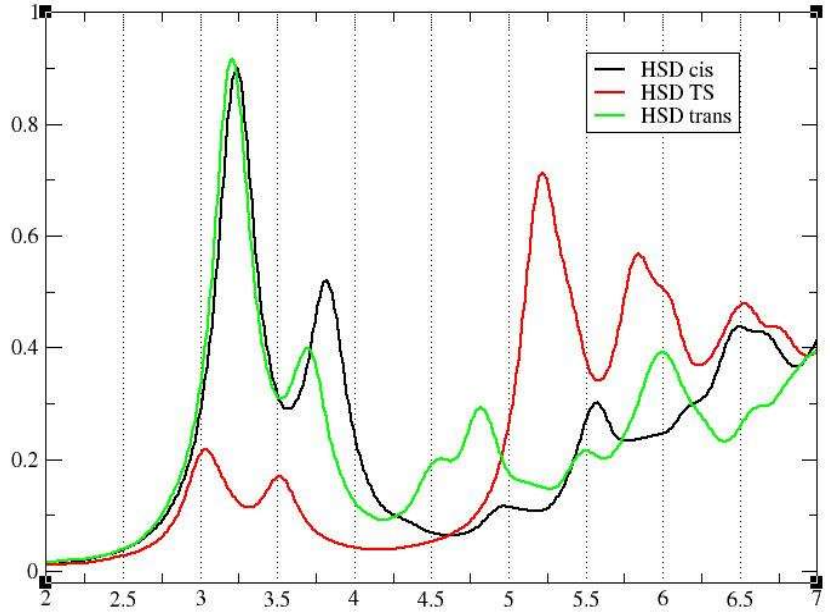
	Distances					Angles			Dipole
	$C_\alpha - C_\beta$	$C_\beta - C_\gamma$	X	$N\cdots X_\delta$	$N\cdots H_\delta$	$C_\alpha - C_\beta - C_\gamma$	DH1	DH2	
Anionic HSA									
HSA cis	1.383	1.409	N_δ	3.302	–	133.5	0.0	0.0	9.6
HSA TS	1.355	1.469				127.9	0.0	89.2	14.5
HSA trans	1.386	1.406	C_δ	3.231	2.737	130.4	0.0	180.0	10.3
Neutral HSE/HSD									
HSE cis	1.356	1.441	N_δ	3.171	–	130.4	0.0	0.0	2.6
HSE TS	1.345	1.476				124.9	1.4	83.6	4.4
HSE trans	1.357	1.439	C_δ	3.075	2.596	127.1	0.0	180.0	6.4
HSD cis	1.362	1.426	N_δ	2.853	2.165	125.6	0.0	0.0	6.0
HSD TS	1.347	1.471				124.8	-1.4	91.8	5.8
HSD trans	1.359	1.430	C_δ	3.182	2.687	127.1	0.0	180.0	1.9
Cationic HSP									
HSP cis	1.355	1.441	N_δ	2.696	1.913	122.2	0.0	0.0	9.1
HSP TS	1.348	1.469				121.9	-0.8	92.9	10.2
HSP trans	1.355	1.438	C_δ	3.013	2.523	123.5	0.0	180.0	9.9

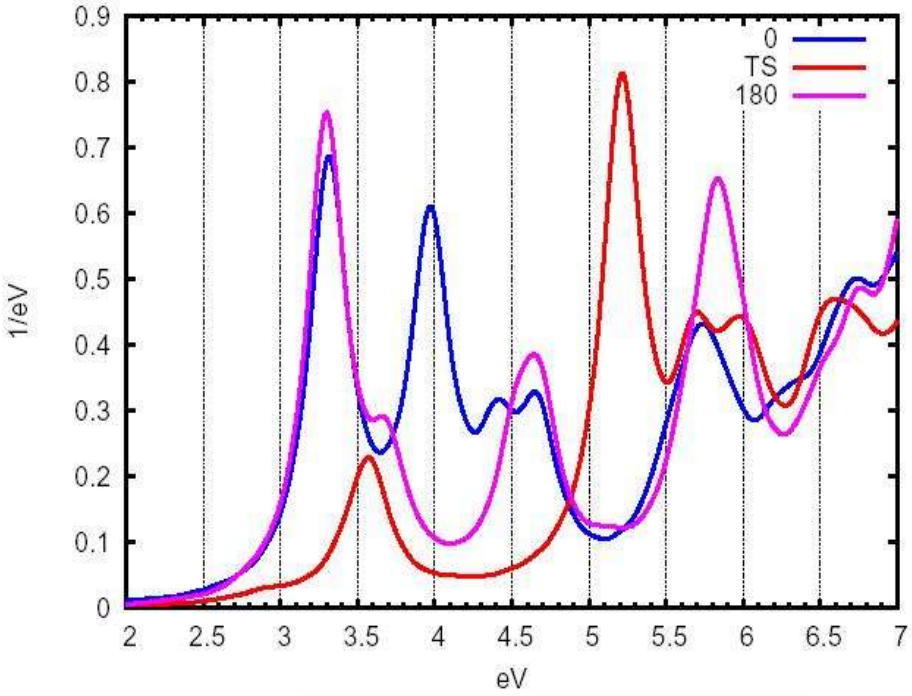
- Preference of Conformation cis for HSD and HSP, trans for HSA and HSE.
- Global minimum in Neutral State is cis-HSD
- Experimental pKa estimations would rule out HSA
- Planarity is observed for cis/trans isomers
- Cα-Cβ < Cβ-Cγ

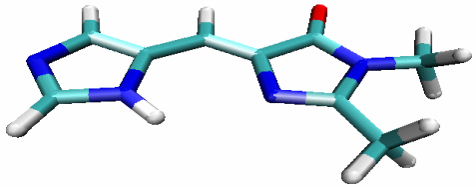
	cis	TS	trans
HSA			
HSE			
HSD			
HSP			

Effect of Conformers and Protonation on the Spectrum

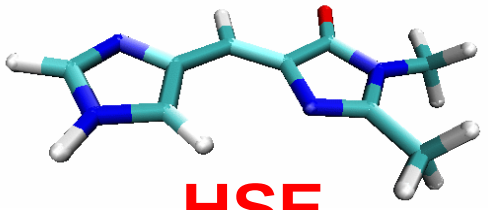
- Sensitive to conformational change
- Sensitive to Protonation State even for the same total charge







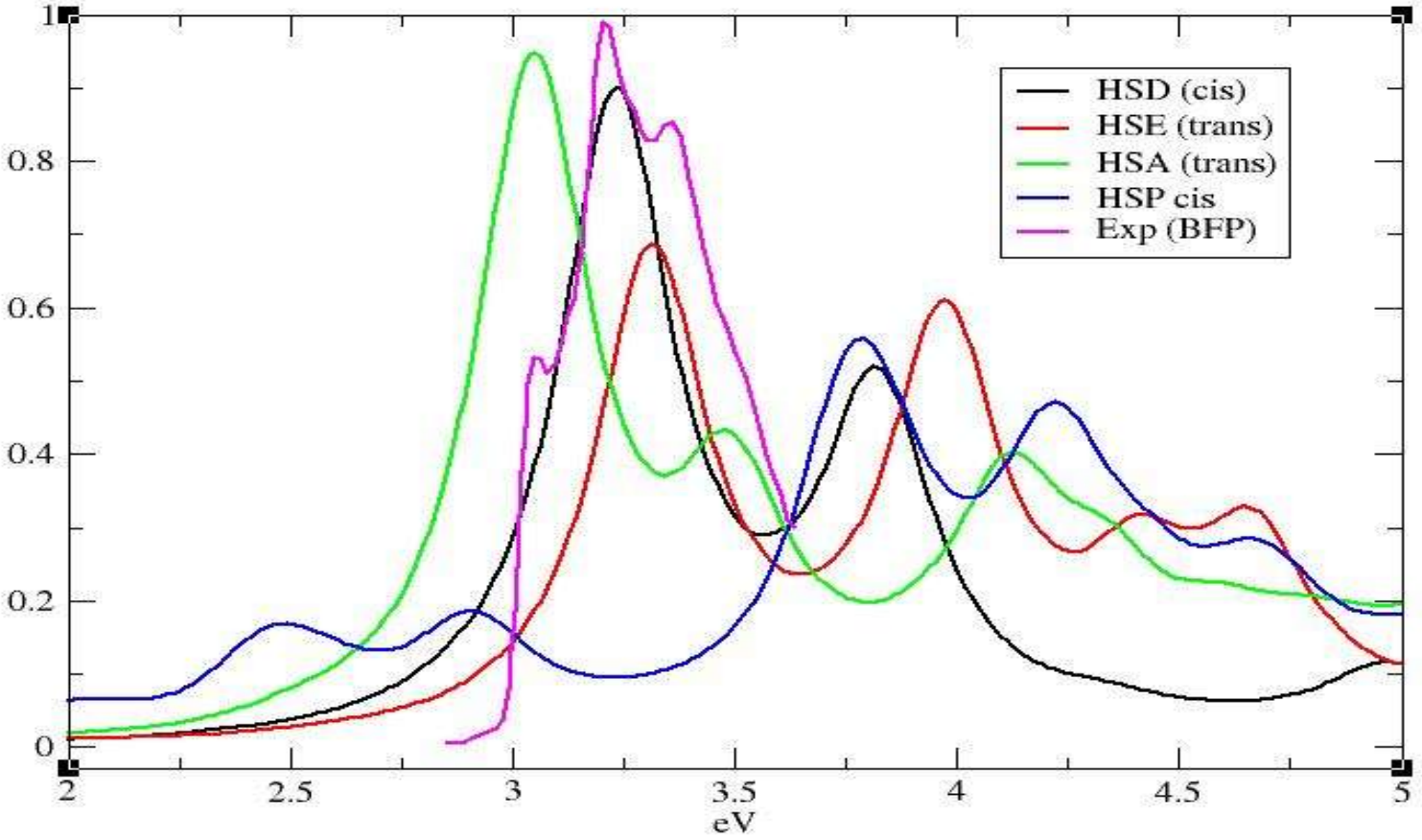
HSD

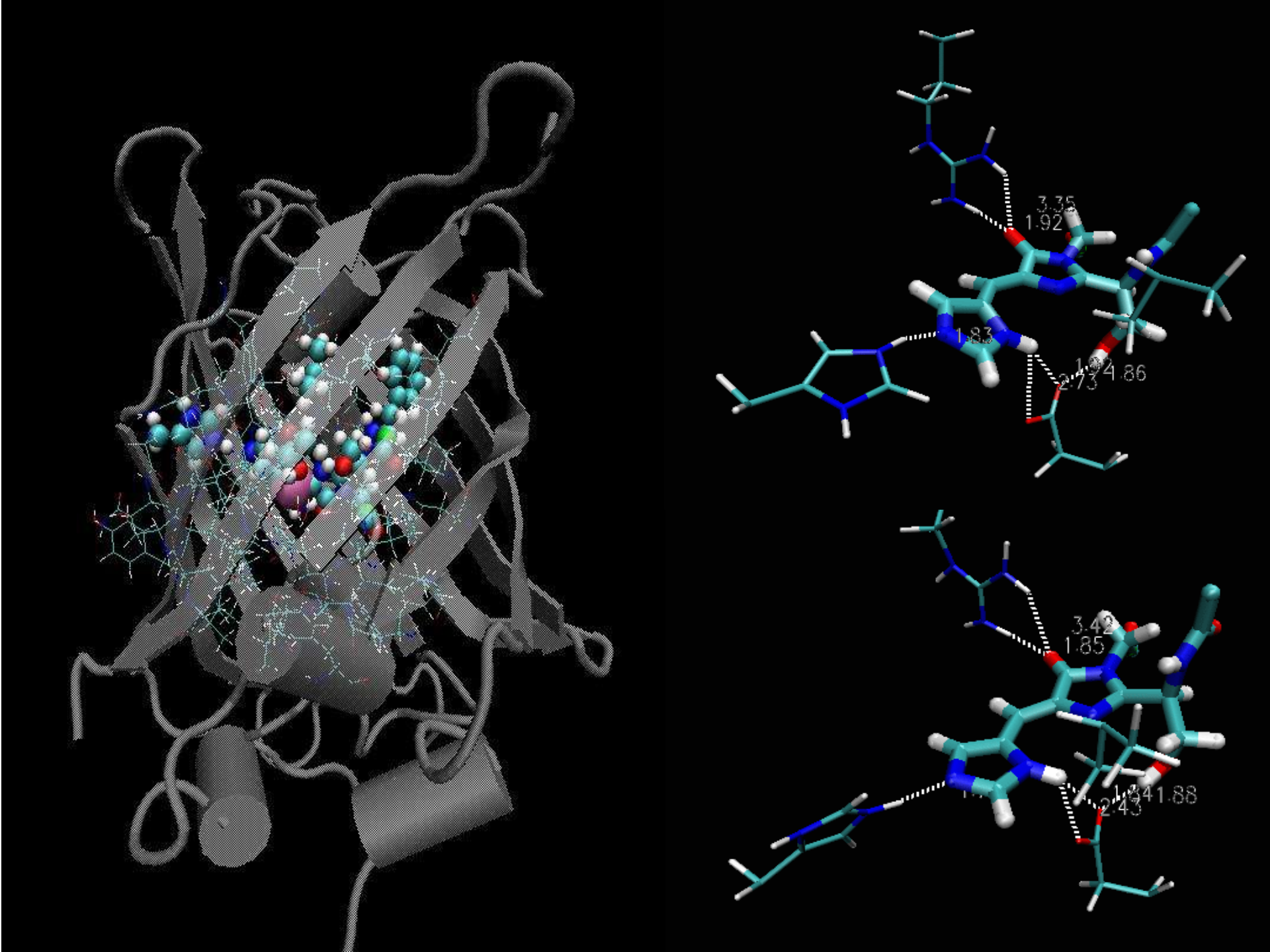


HSE

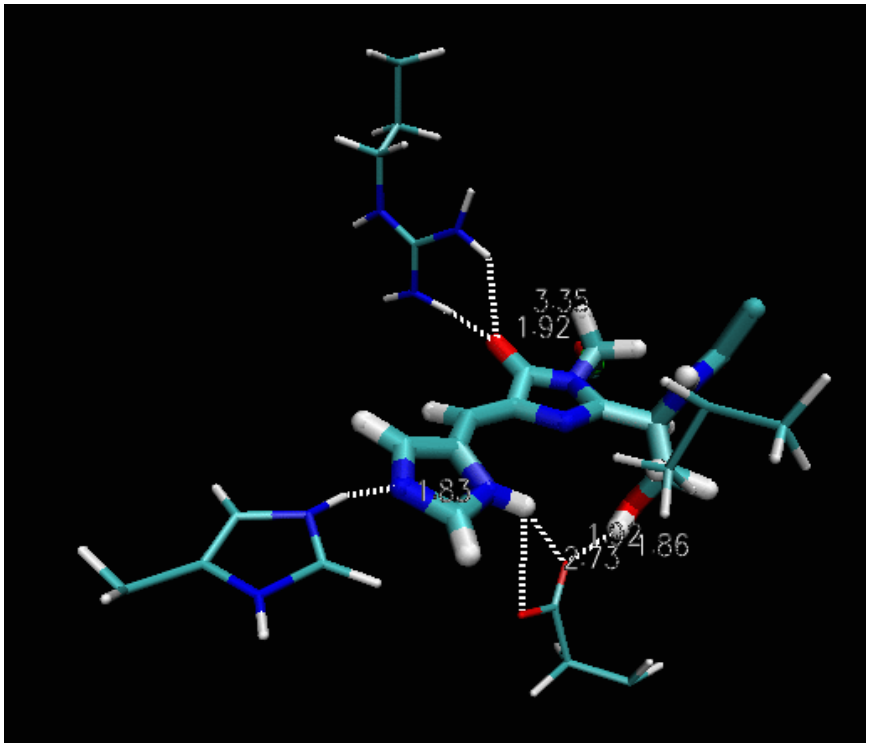
Effect of Conformers and Protonation on the Spectrum

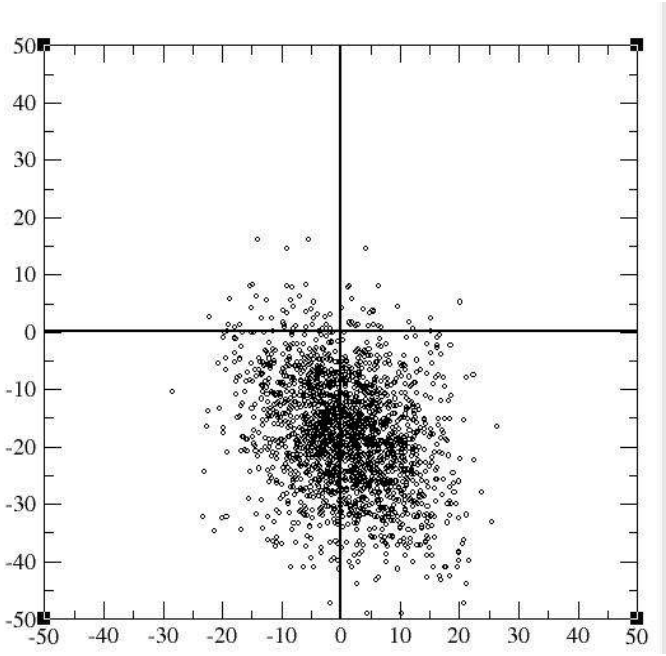
- Main peak associated with HSD(cis)
- HSA(pKa?) and HSE could also be involved
- Spectra rules out HSP
- Mixture of protonation states? Conformers? Peaks at 3.5-4
- What about the chromophore in the protein?

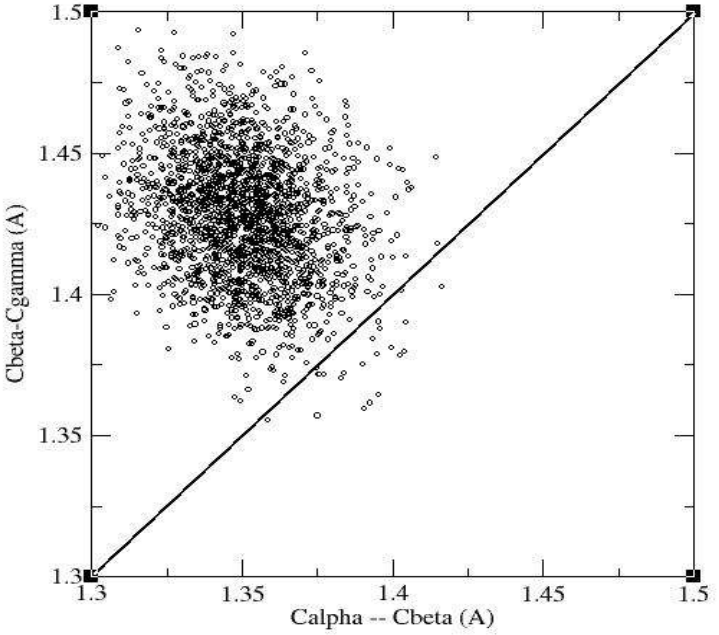



3.35
1.92
1.83
2.73
4.86
3.42
1.85
2.43
1.88

QM/MM dynamics of the ground state







- Cα-Cβ < Cβ-C
- Loss of planarity due to HB

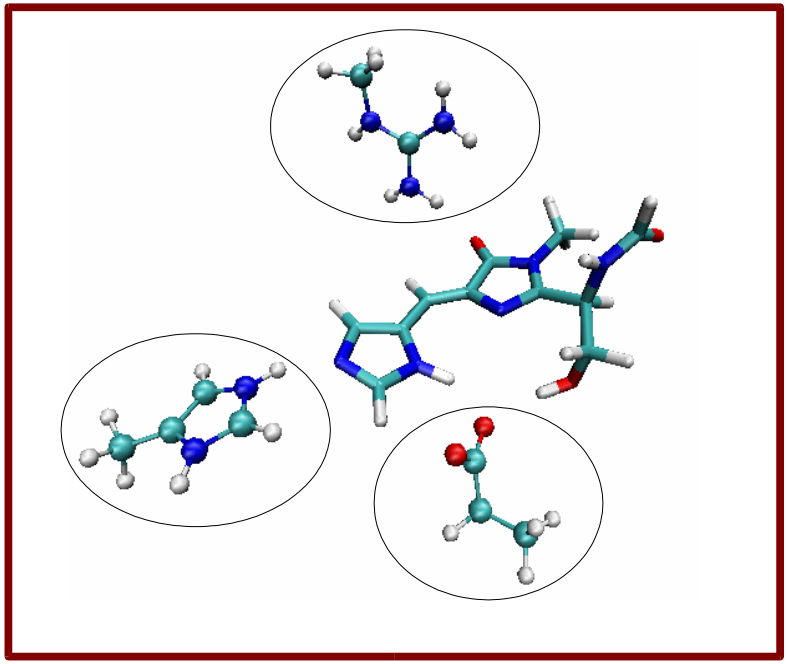
Inclusion of the environment in the spectrum

Ideally Full QM/MM: escaping charge problem in Octopus

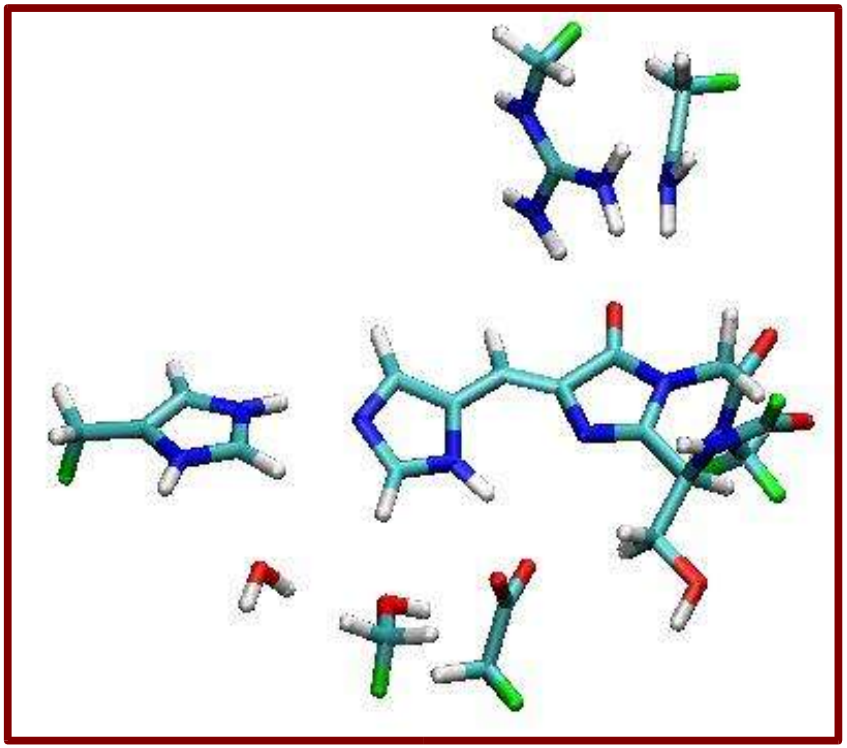
In practice: just include the closests residues.

Environment I:

Arg 96, His148, Glu 222



Environment II

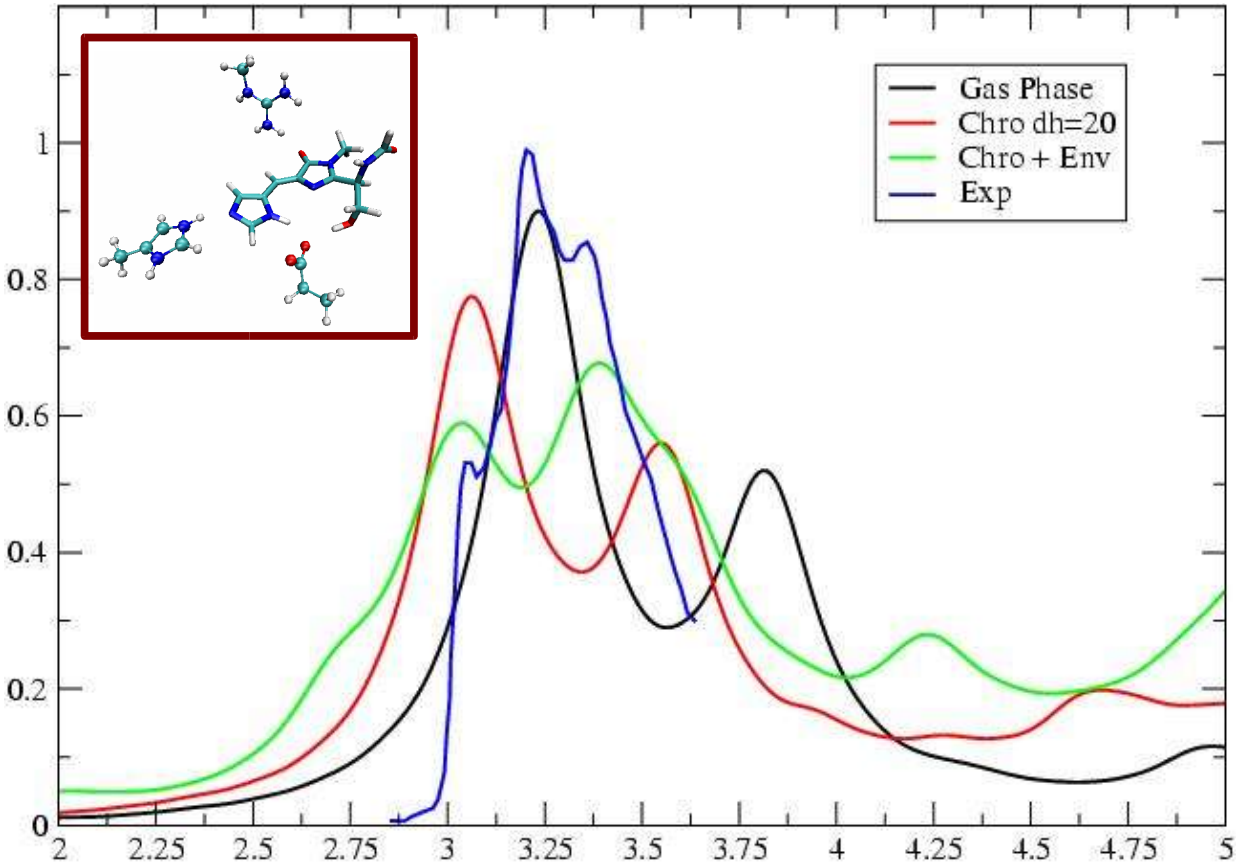


Environment and Excitation

•Structural effects shift the peaks towards lower energy but same shape

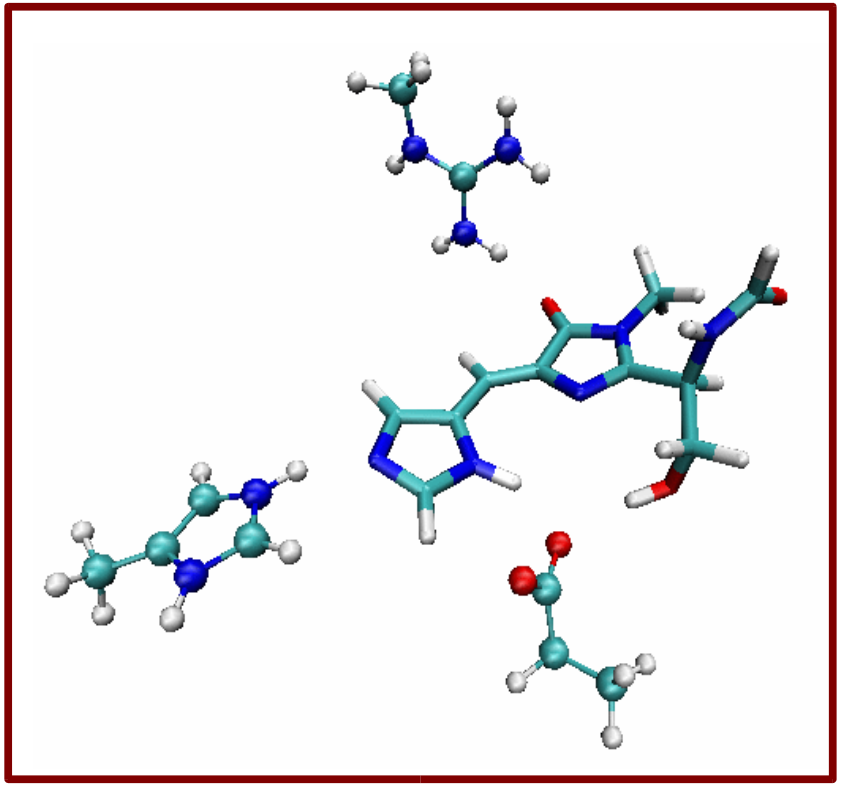
•Inclusion of the microenvironment shifts only one of the peaks and the complexity of the shape increases

•Less satisfactory results for the main exp, error of the calculation itself?



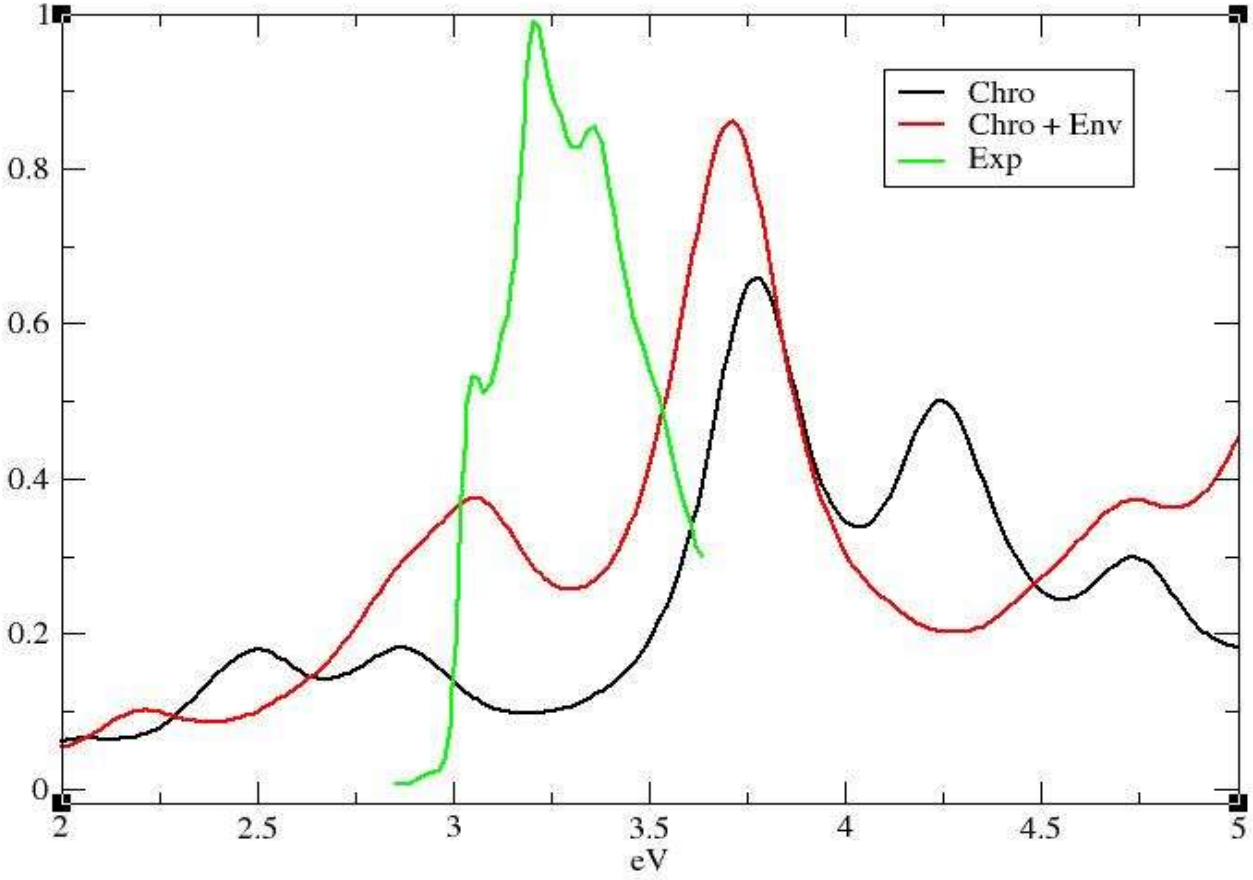
Environment			Chromophore	Peaks						
Arg 96	His148	Glu222								
+	HSP	Glu	cis-HSD	2.70(s)	**3.04**	**3.39**	3.61(s)	4.23		
+	HSP	GluP	cis-HSD	2.70(s)	**2.93**	3.14(s)	**3.56**	4.56(s)		
+	HSD	GluP	cis-HSE	**2.83**	3.02(s)	**3.56**	4.19	4.65		
+	HSD	Glu	cis-HSP	2.21	2.84(s)	3.05	**3.70**	4.73		
+	HSE	Glu	cis-HSD	2.41(s)	**2.99**	3.23(s)	3.44	3.91(s)	4.23(s)	4.58(s)

- Various possibilities. The most reasonable seems to be cis-HSD, which is in accord with gas phase calculations, protein residue distributions and pKa's
- HSP can be ruled out even with the env



Environment and Excitation for HSP

- Environment big effects
- Even if pKa values would allow this possibility, not from the spectrum

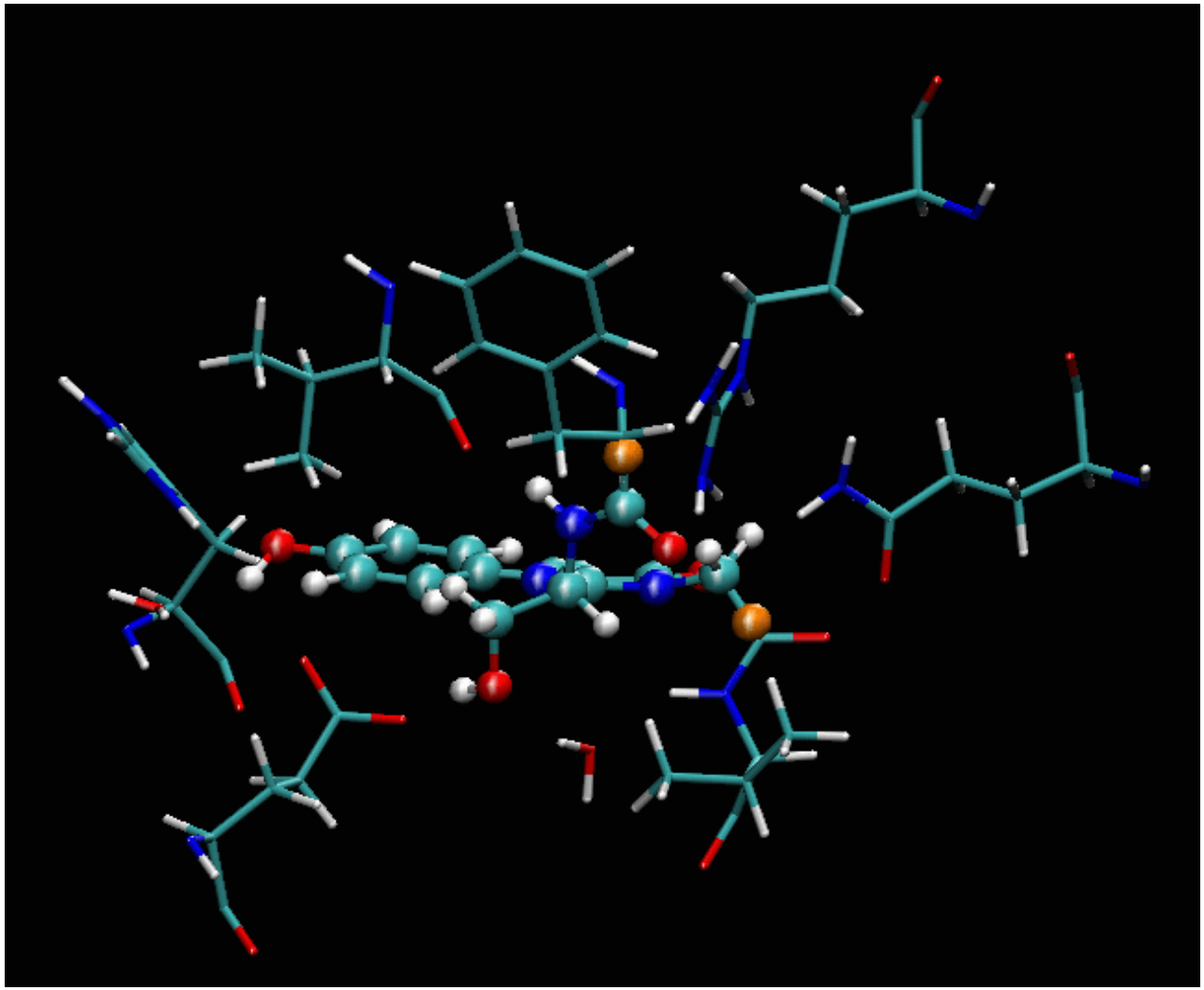


Summary

- Biochromophores and π-π^* transitions.
- Optical properties of chromophores are modulated by their protein environment (QM/MM): conformation, protonation state, polarization of the electronic structure, etc.
- In GFP/BFP, when we take into account the conformational change induced by the protein (GFP) + closest charged aminoacids (BFP), one is able to reproduce the main peaks in the optical absorption.
- This can help elucidating key chemical information such as the protonation state/conformation of the chromophore and give ideas to design new GFP-type pigments. Propose Mutations
- For Future work: QM/MM in TDDFT(OCTOPUS) needs improvements.

QM/MM Boundary

H-Link Atom



Part of the QM

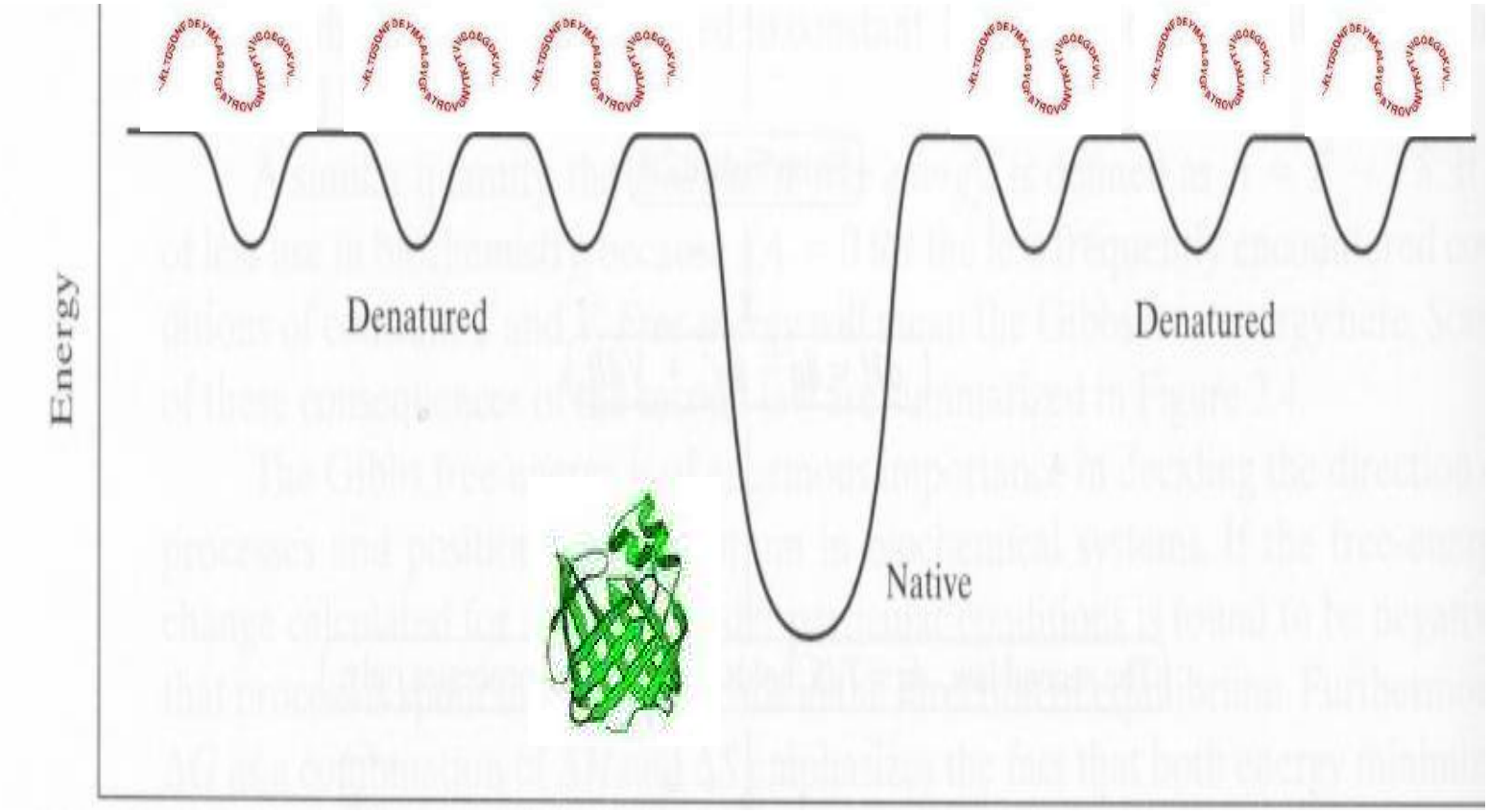
Placed along Bond

QQ vs HQ

Capping-Atom

7. Reuter, N.; Dejaegere, A.; Maigret B.; Karplus M.; J. Phys. Chem. A **2000**, *104*, 1720.
8. Field, M.J.; Albe, M.; Bret C.; Proust-De Martin F.; Thomas, A.
J. Comp. Chem. A **2000**, *21*, 1088.

Flexibitiy vs Rigid Structures



Enlace por Puente de Hidrogeno

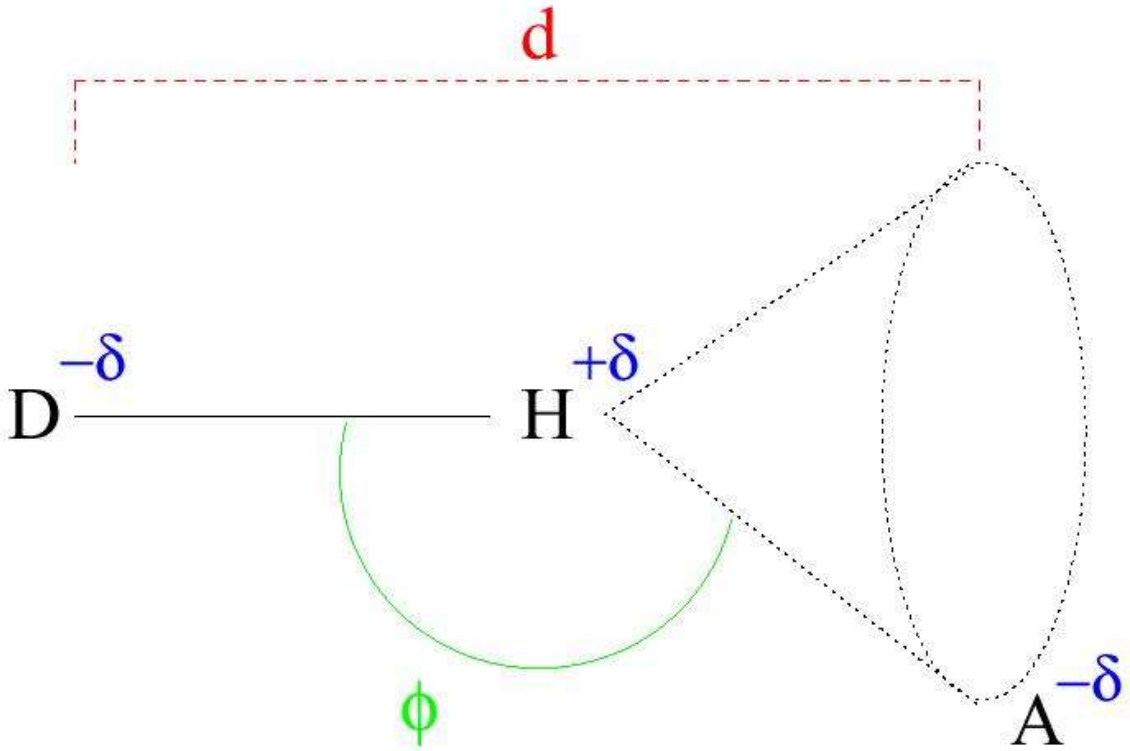
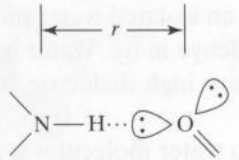
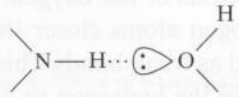
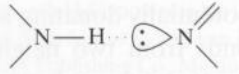
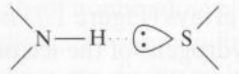
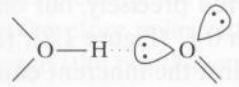
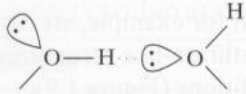
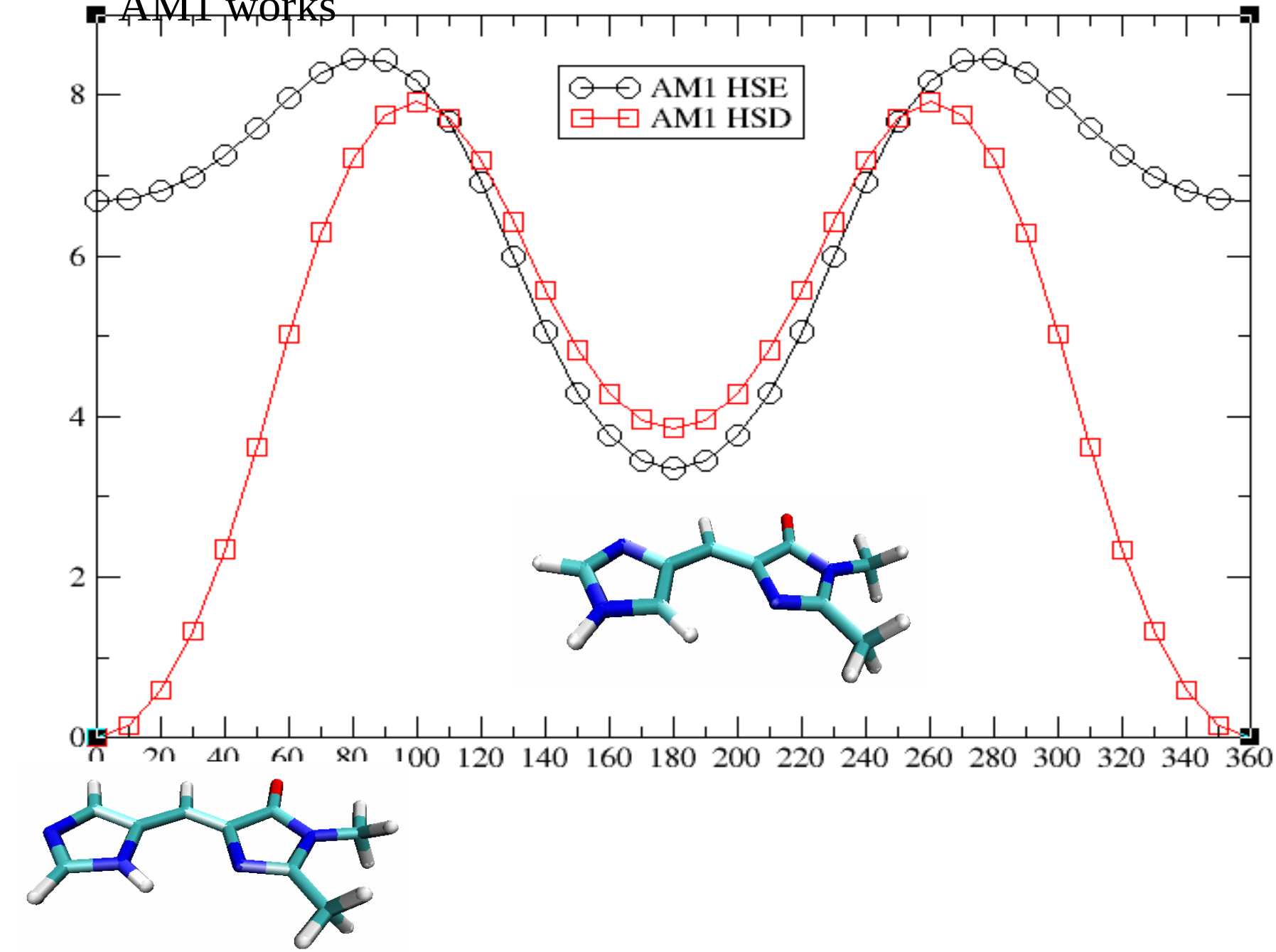
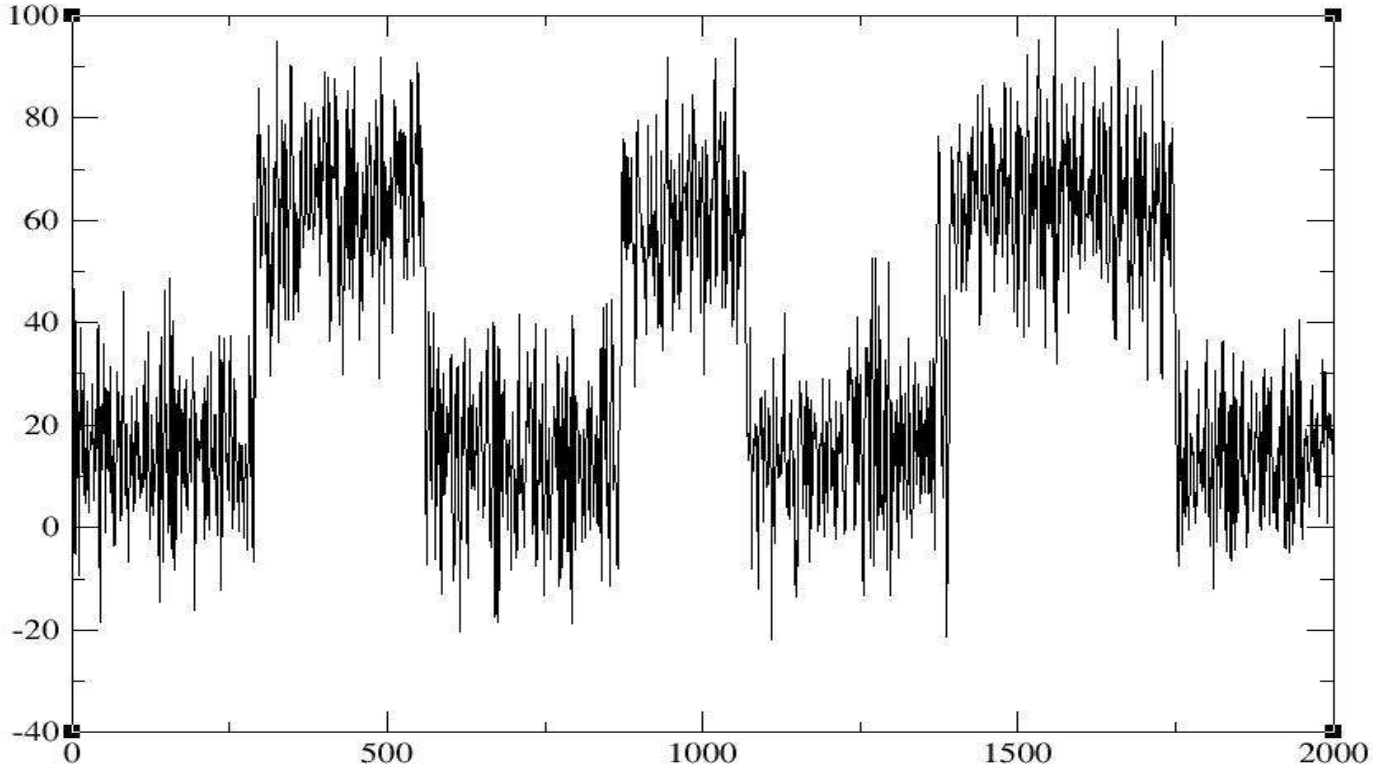


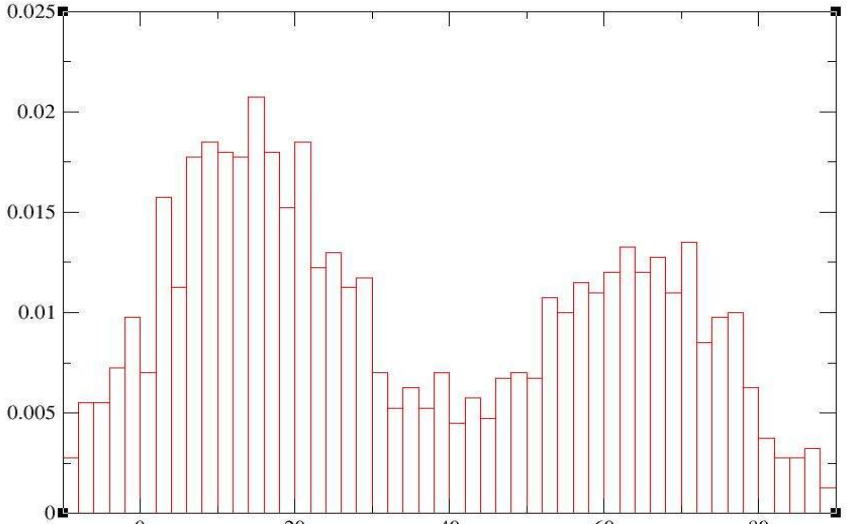
TABLE 1.3 HYDROGEN-BOND DONORS AND ACCEPTORS IN MACROMOLECULES

Donor	Acceptor	r (nm)
		0.29
		0.29
		0.31
		0.37
		0.28
		0.28

AM1 works


AM1 HSE
AM1 HSD
0
2
4
6
8
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360


100
80
60
40
20
0
-20
-40
0
500
1000
1500
2000


0.025
0.02
0.015
0.01
0.005
0
0
20
40
60
80

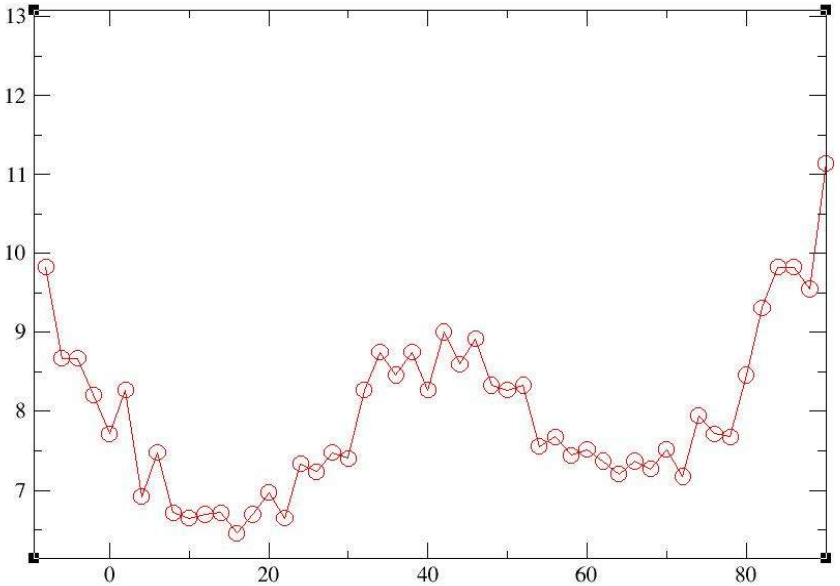
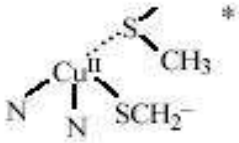

13
12
11
10
9
8
7
0
20
40
60
80

Table 11-2. Absorption of Prosthetic Groups

Protein	Prosthetic Group	Longest λ Absorption λ_{max} (nm)	$\varepsilon_{max} \times 10^{-4}$	2nd Longest λ Absorption λ_{max} (nm)	$\varepsilon_{max} \times 10^{-4}$
Amino acid oxidase (rat kidney)	FMN	455	1.27	358	1.07
Azurin, P. fluorescene, plastocyanin, spinach stellacyanin	 *	781	0.32	625**	0.35
Ceruloplasmin (human)	8 coppers type 1, 2, 3	794	2.2	610	1.13
Cytochrome *c* (reduced) (human)	Fe^{II}-heme	550	2.77		
Ferrodoxin	$2Fe^{III}$-$2S^{-}$cluster	421	0.98	330	1.33
Flavodoxin (C. pasteurianium)	FMN	443	0.91	372	0.79
pyruvate dehydrog-enase (E. coli)	FAD	460	1.27	438	1.46
Rhodopsin (bovine)	retinal-lys	498	4.2	350	1.1
Reubredoxin (M. aerogenes)	(Fe^{III},4 Cys) tetrahedra	570	0.35	490	0.76
Xanthine oxidase	Fe, Mo	550	2.2		
Threonine deaminase (E. coli)	4 pyridoxal phosphates	415	2.6		

*blue copper

** $Cu^{II} \leftarrow S^{-}$ charge transfer