Multiscale Methods for the Description of Chemical Events in Biological System

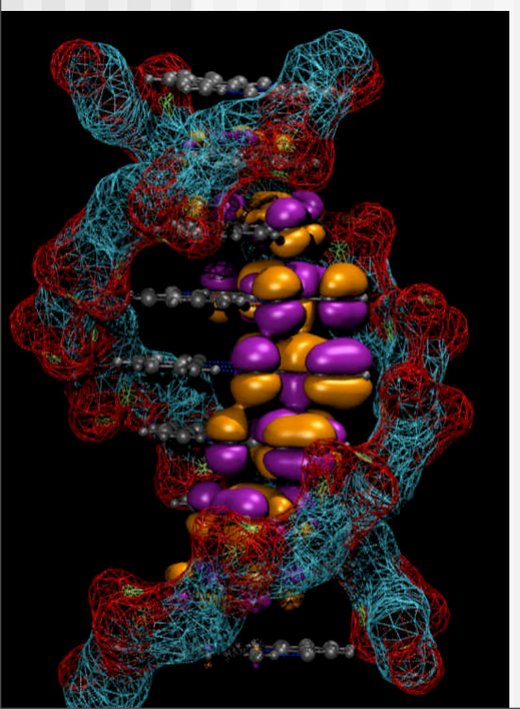
Marcus Elstner

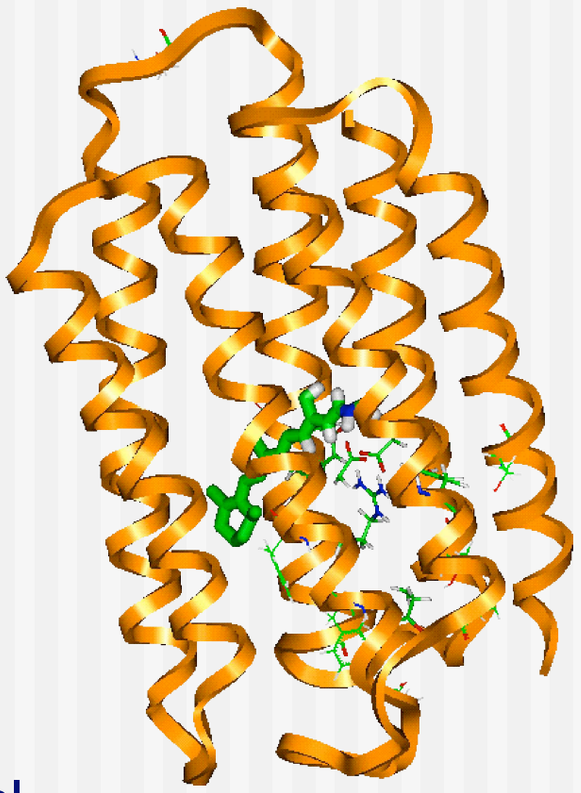
Institute of Physical Chemistry

Karlsruhe Institute of Technology (KIT)

-> May 2009: TU Braunschweig

www.tu-bs.de/pci/forschung/theorie





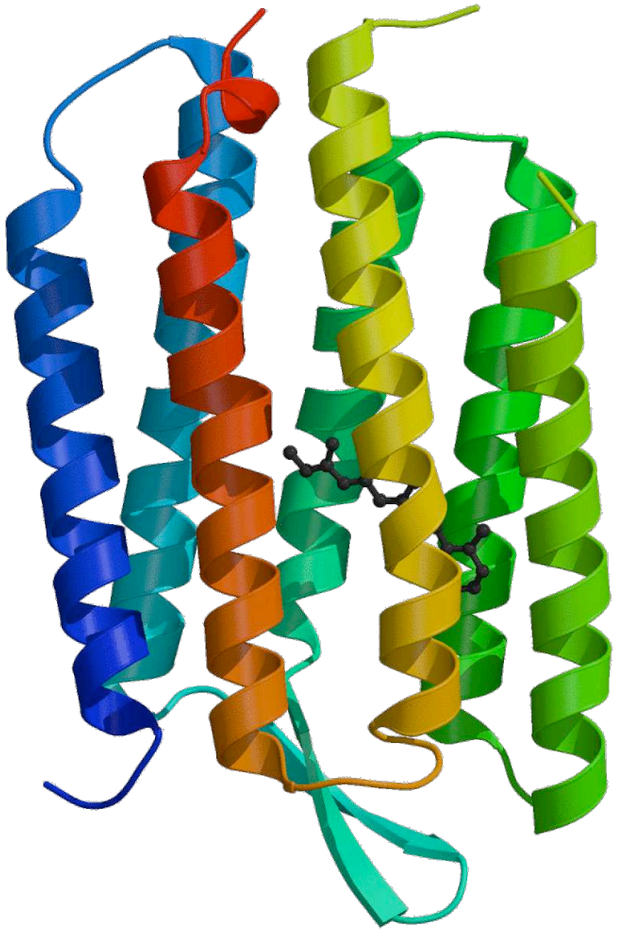
recent review:

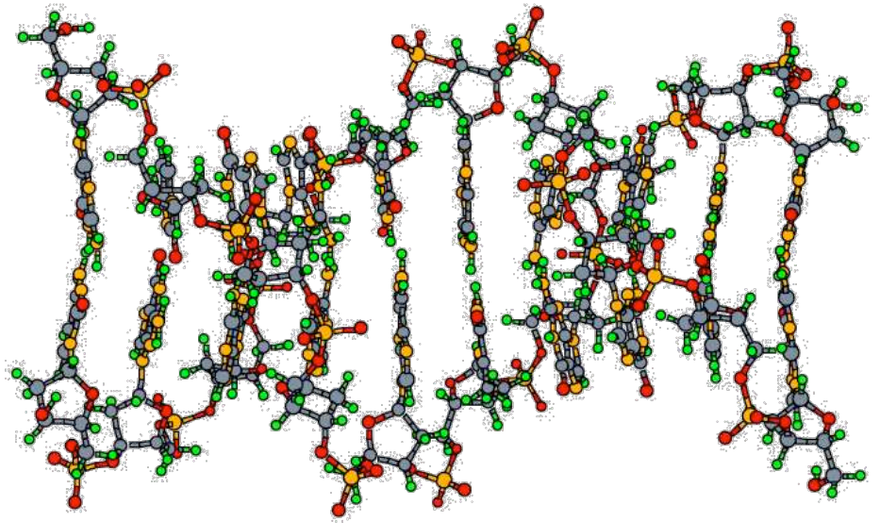
NIC Series Volume 42

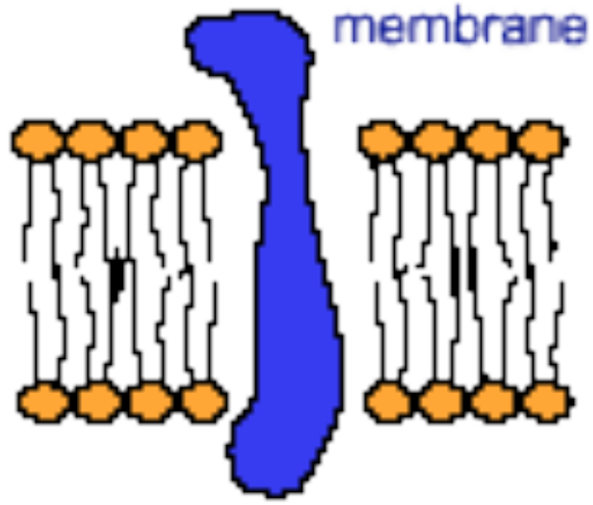
Multiscale Simulation Methods in Molecular Sciences

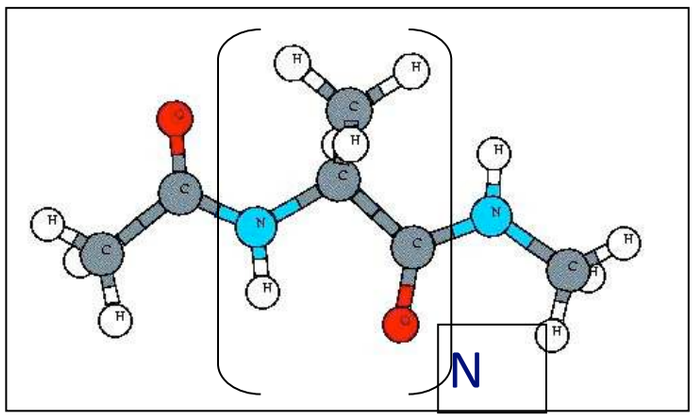
http://www.fz-juelich.de/nic-series/volume42/volume42.html

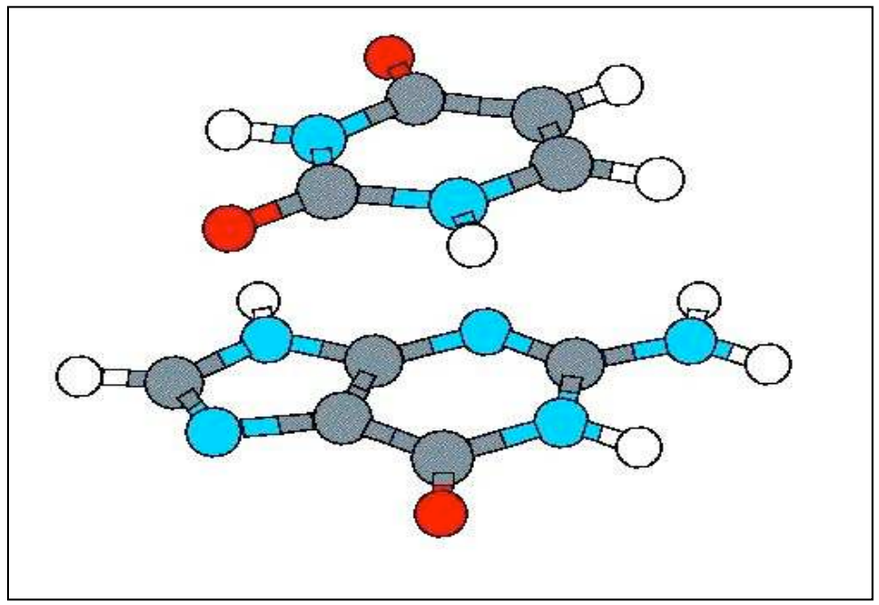
Biological structures: proteins, DNA, lipids







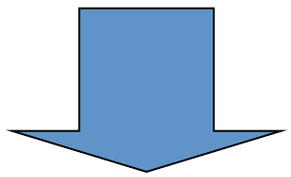




Understanding biological processes

Different length- and timescales are relevant

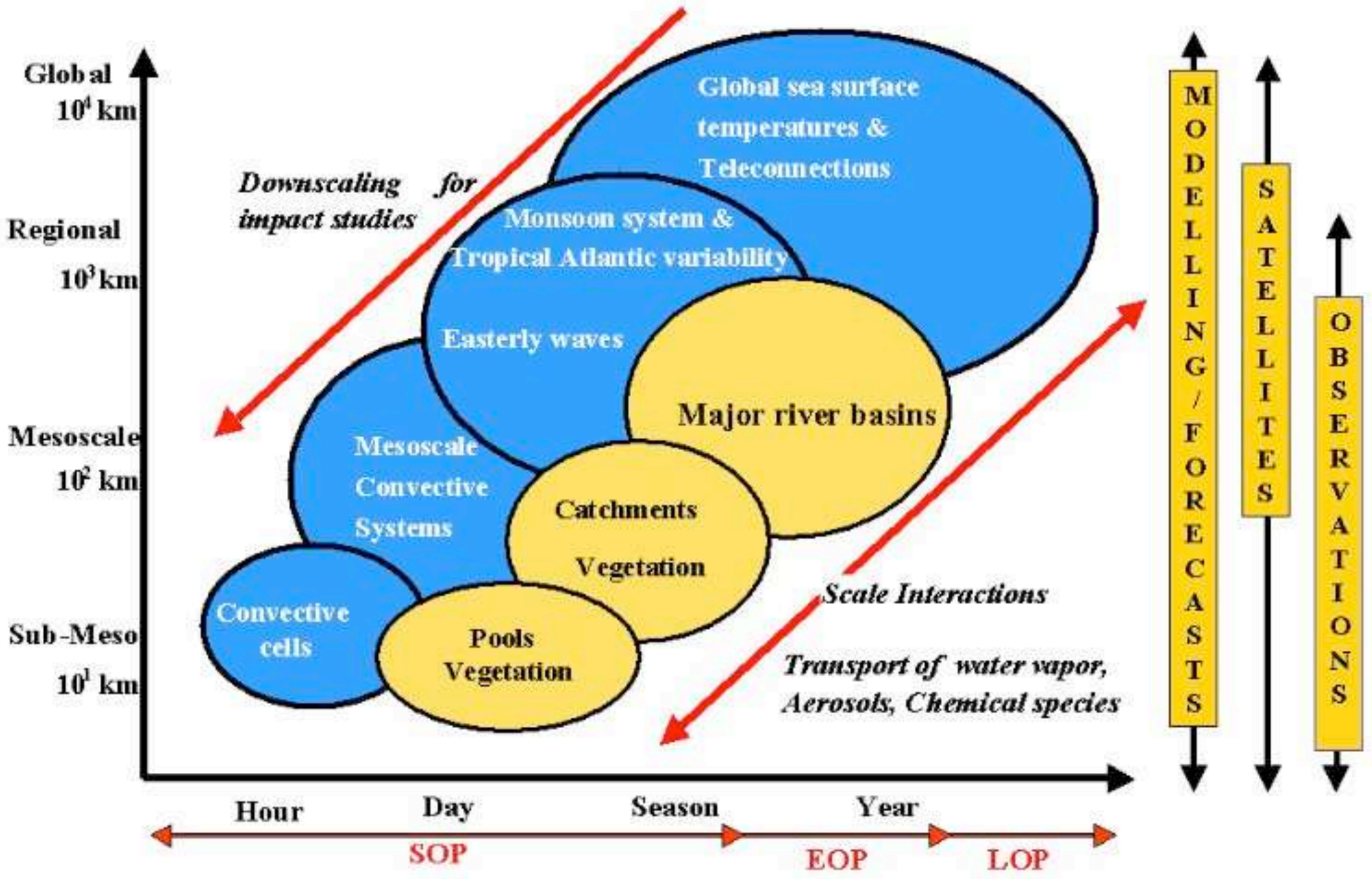
- atomistic: equations of motion for coupled N-body problem (classical/quantum mechanical)
- coarse grained simulations: include several atoms into ‘superatom’
- continuum: electrostatic and mechanical properties
- rate & transport equations, stochastic models etc.: phenomenological



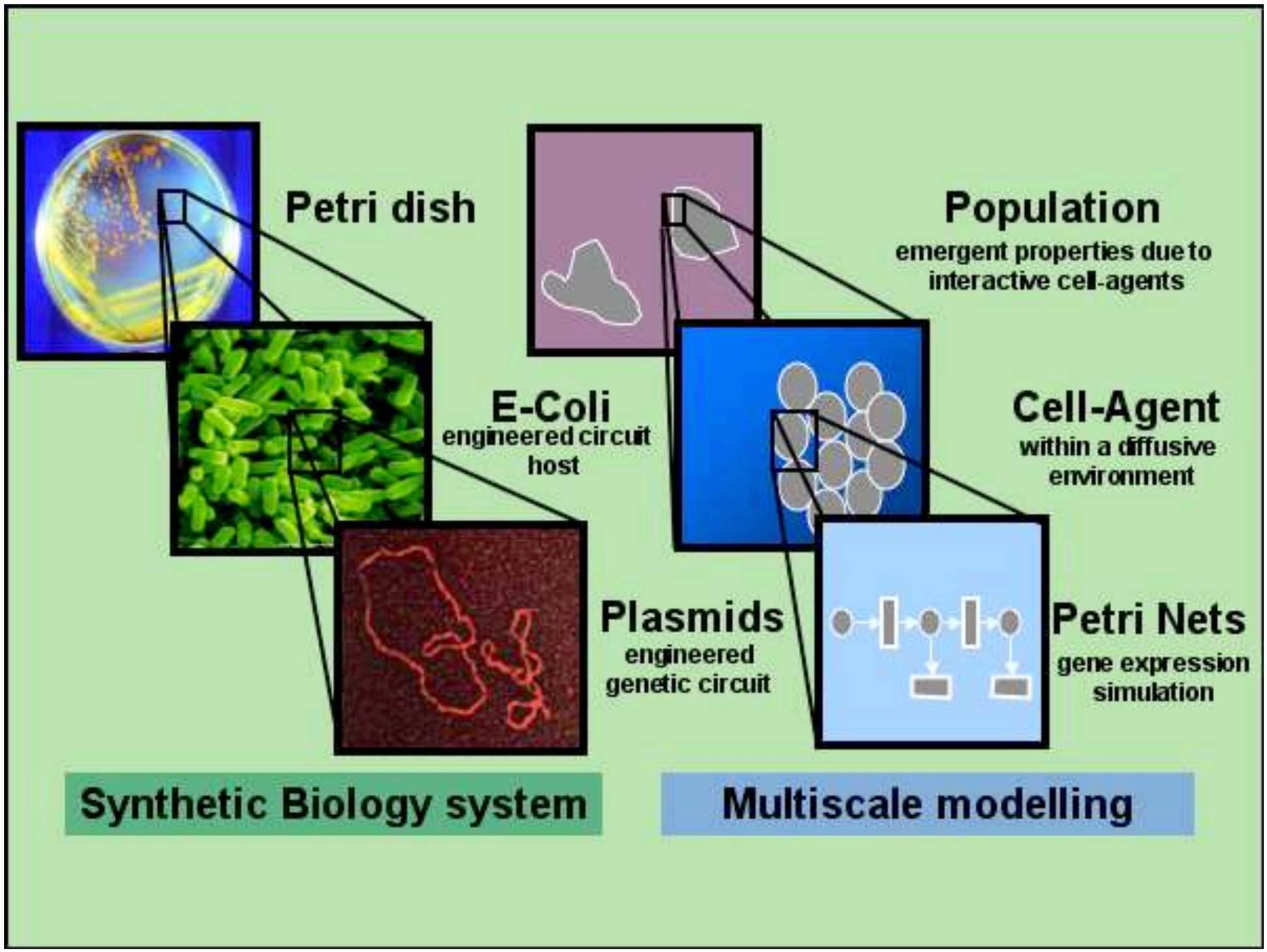
- very different theoretical models
- combination (within limits): “Multi-scale modeling”

Multi-scale methods: used in different areas

AMMA: African Monsoon Multidisciplinary Analysis

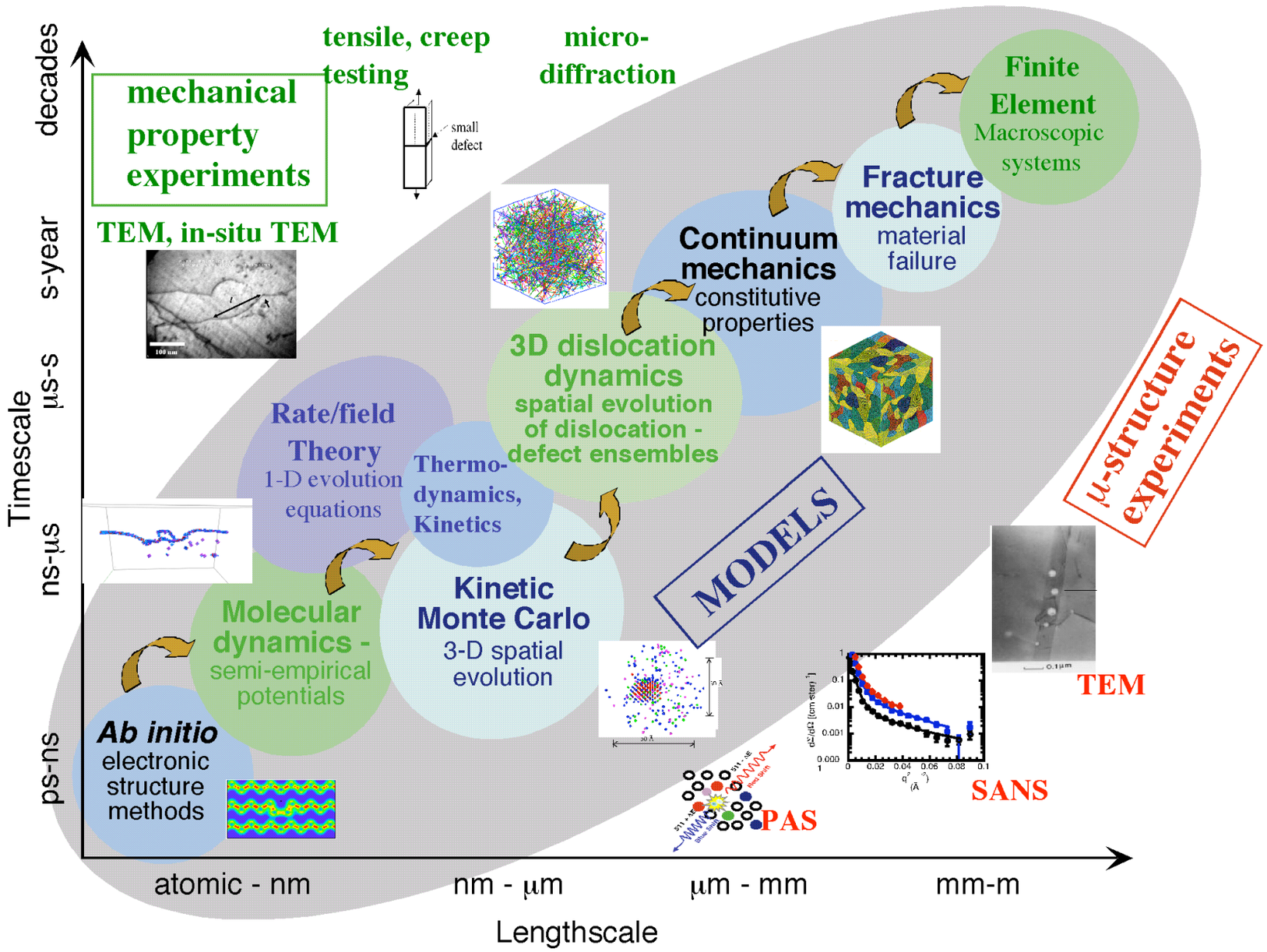


Multi-scale methods: used in different areas



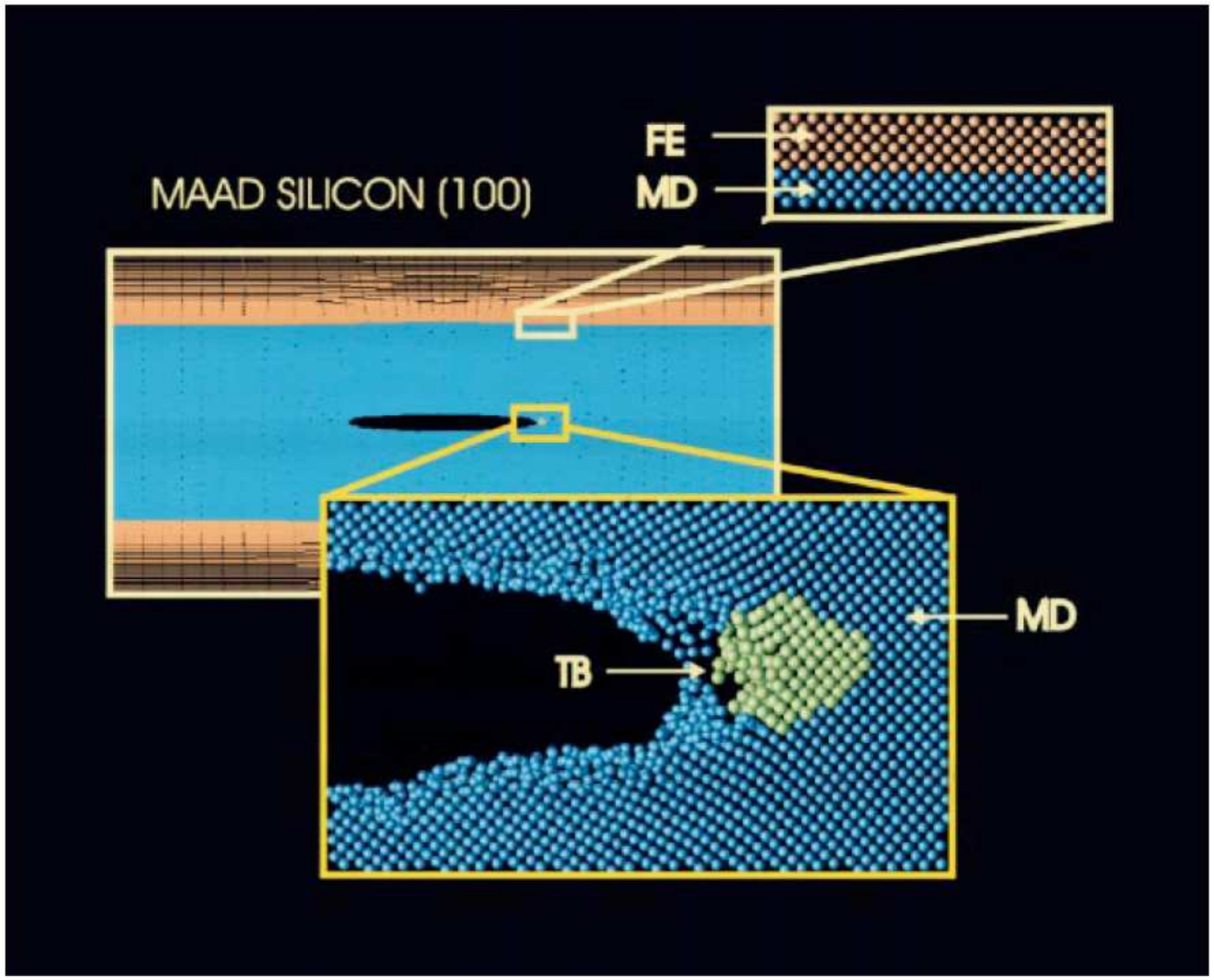
V. Rouilly. Imperial College

Multi-scale methods in computational materials science



B. Wirth, Berkeley

Crack propagation in silicon

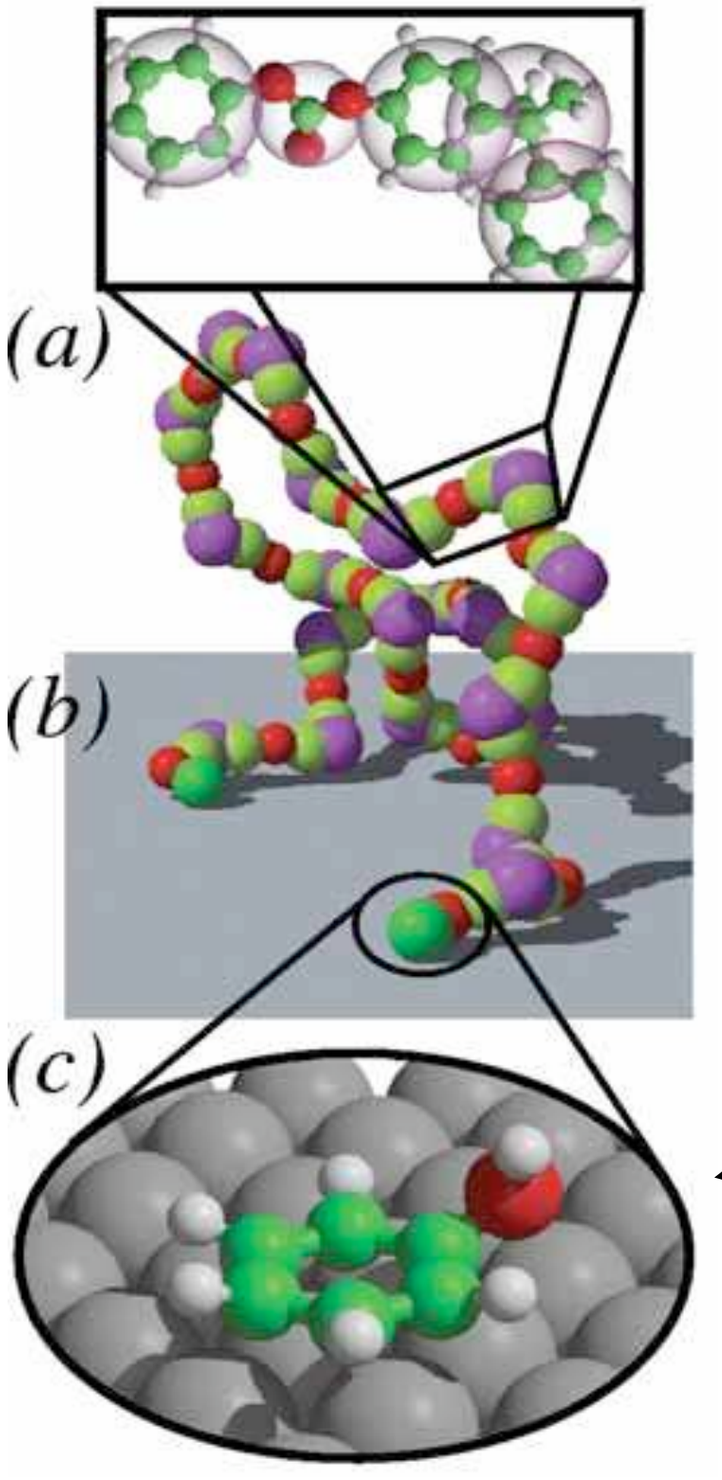


- quantum mechanics
- empirical force fields
- finite elements

Broughton et al PRB 60, 2391

‚local ' information required

Polymers on metal surfaces



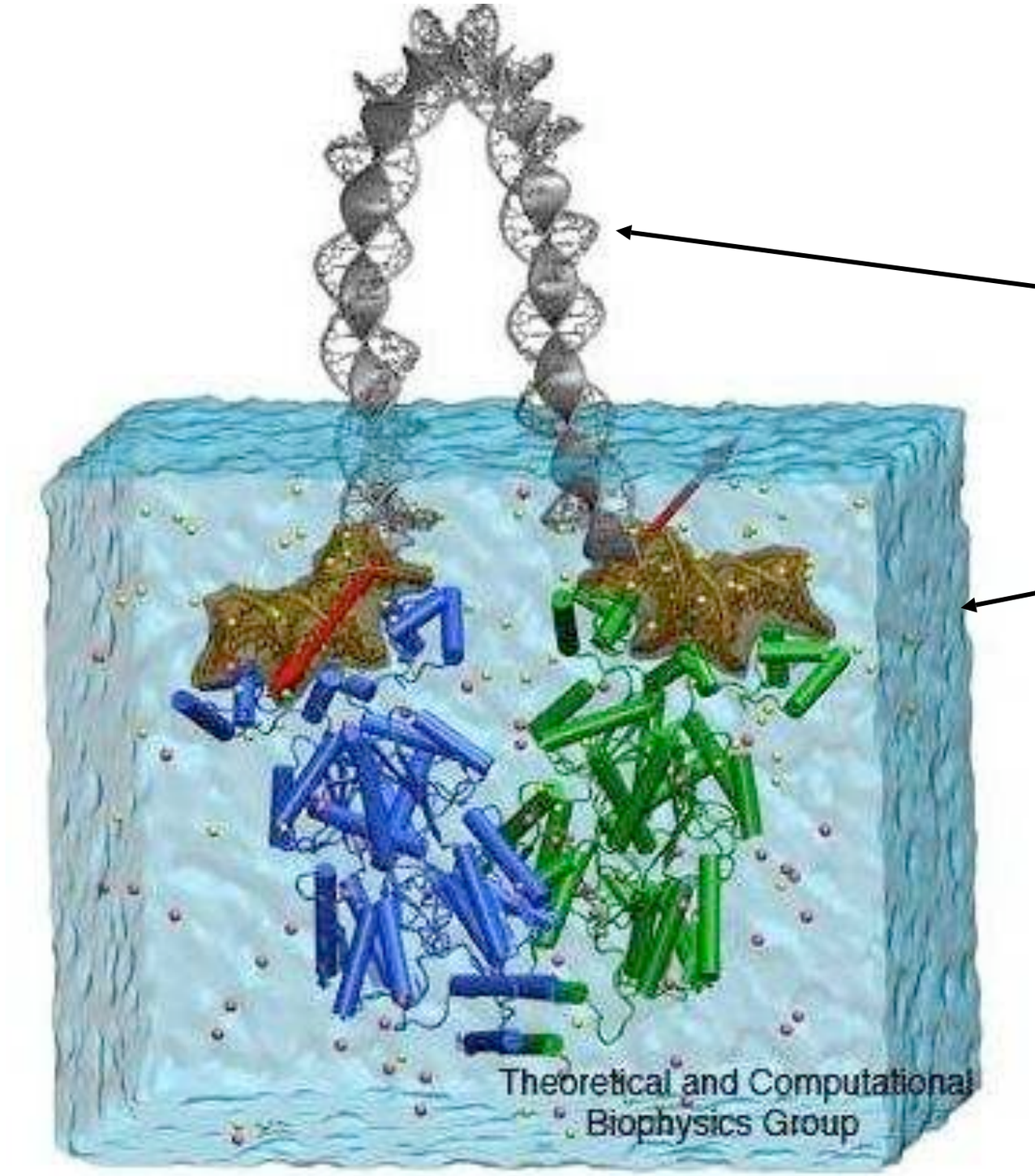
Bisphenol-A-Polykarbonat (BPA-PC) on Ni-surface.

Delle Site, Kremer, MPI Mainz

- ‚Coarse grained'
- quantum chemistry (DFT)

parametrization

Biophysics: DNA-protein interaction

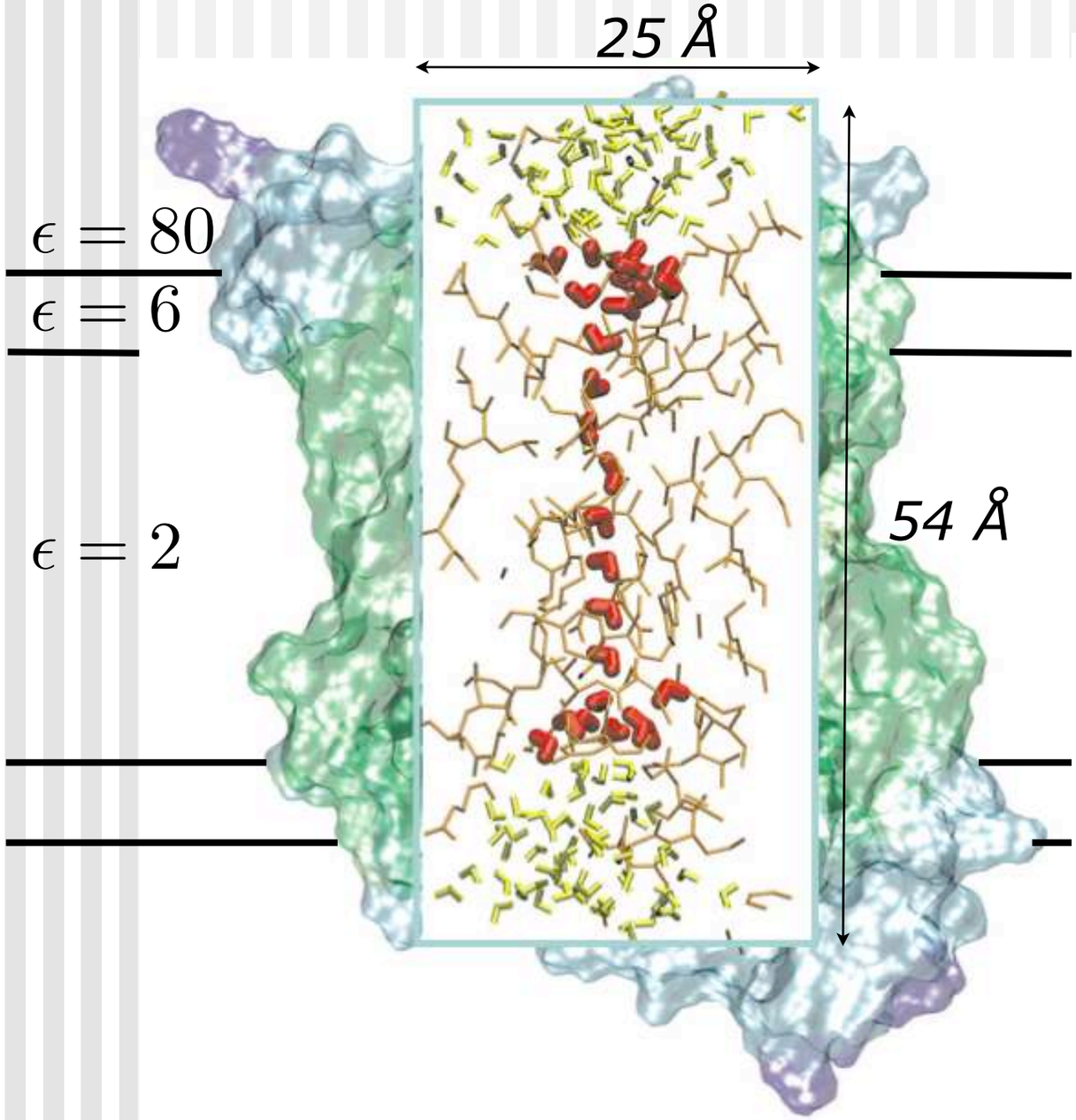


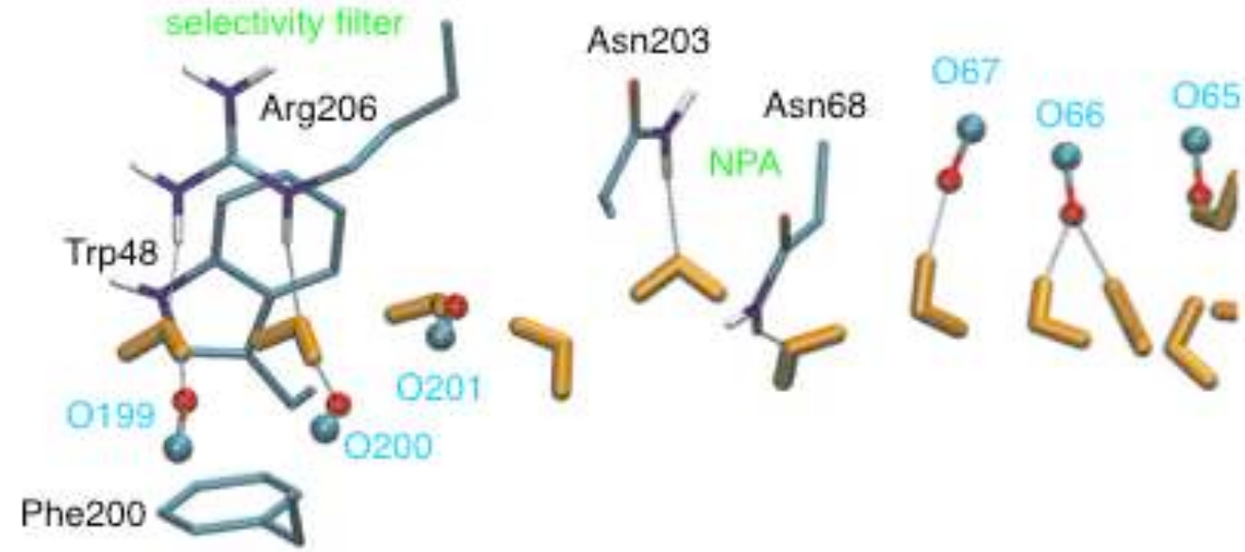
- **continuum description : elastic band for DNA**
- **empirical force field: water, protein**

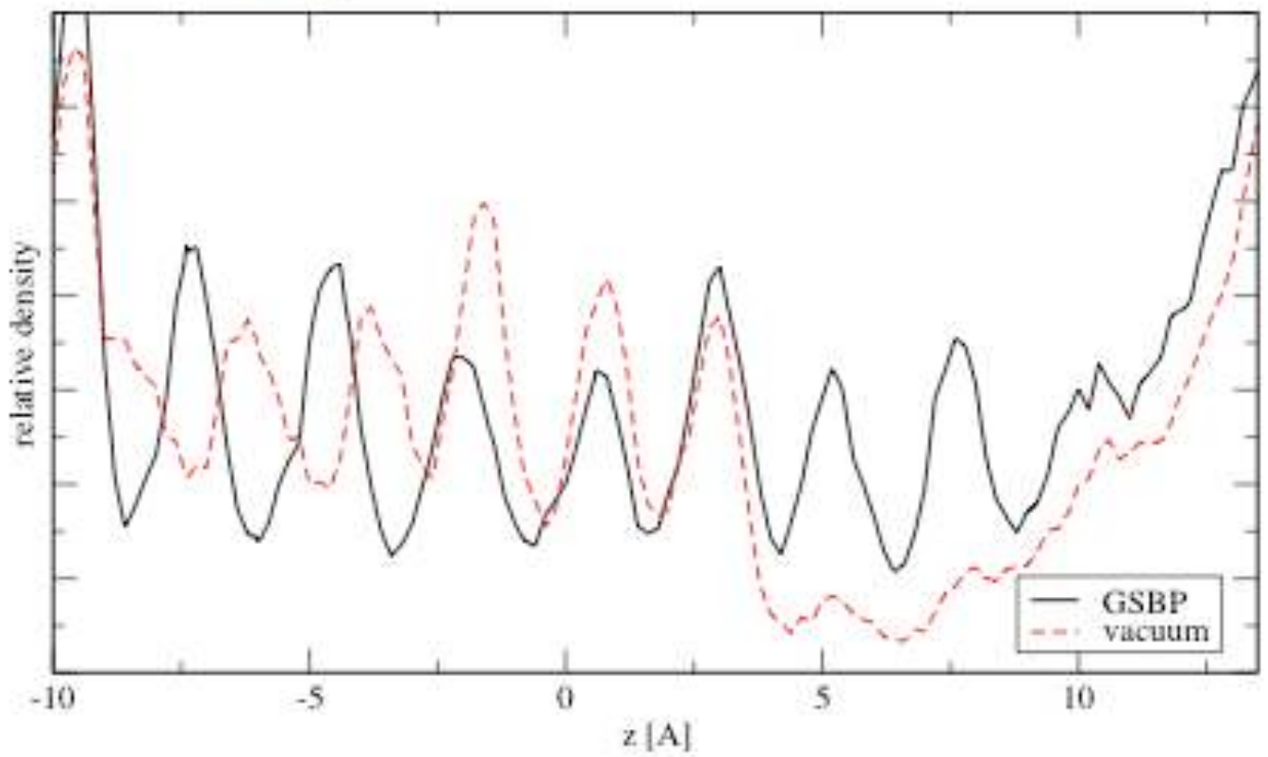
Lac repressor protein

Villa et al., PNAS 102 6783

Membrane systems

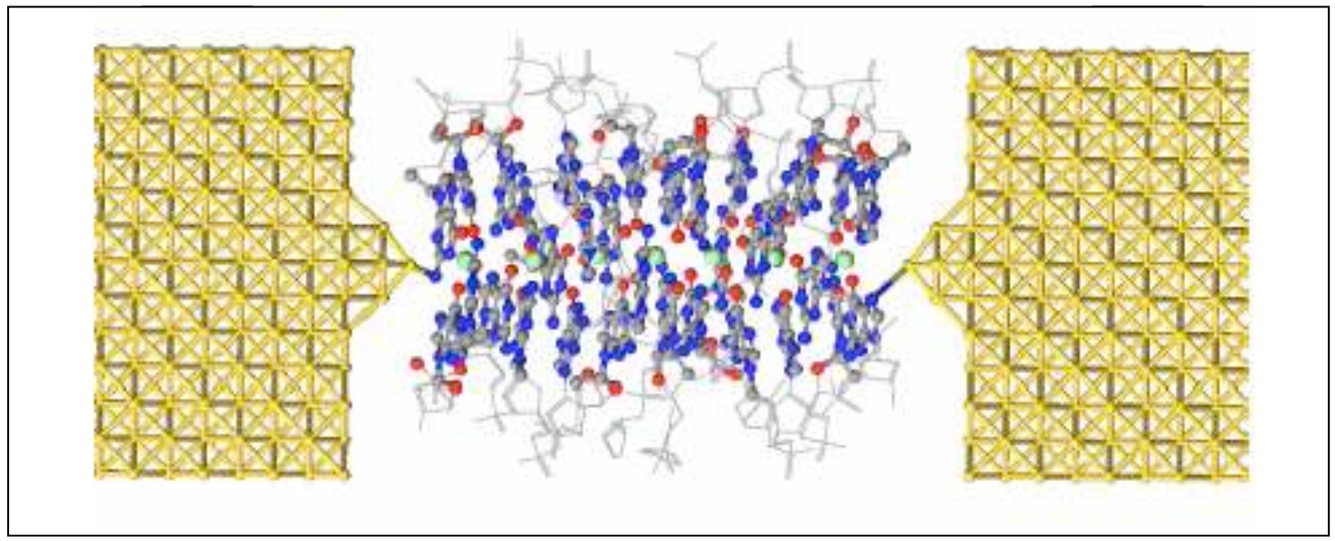






P. Konig, N. Ghosh, M. Hoffman, M. Elstner, E. Tajhorshid, Th. Frauenheim, QC, *J. Phys. Chem. A* Trhular Issue, 110, 548-563 (**2006**)

Charge transfer through DNA

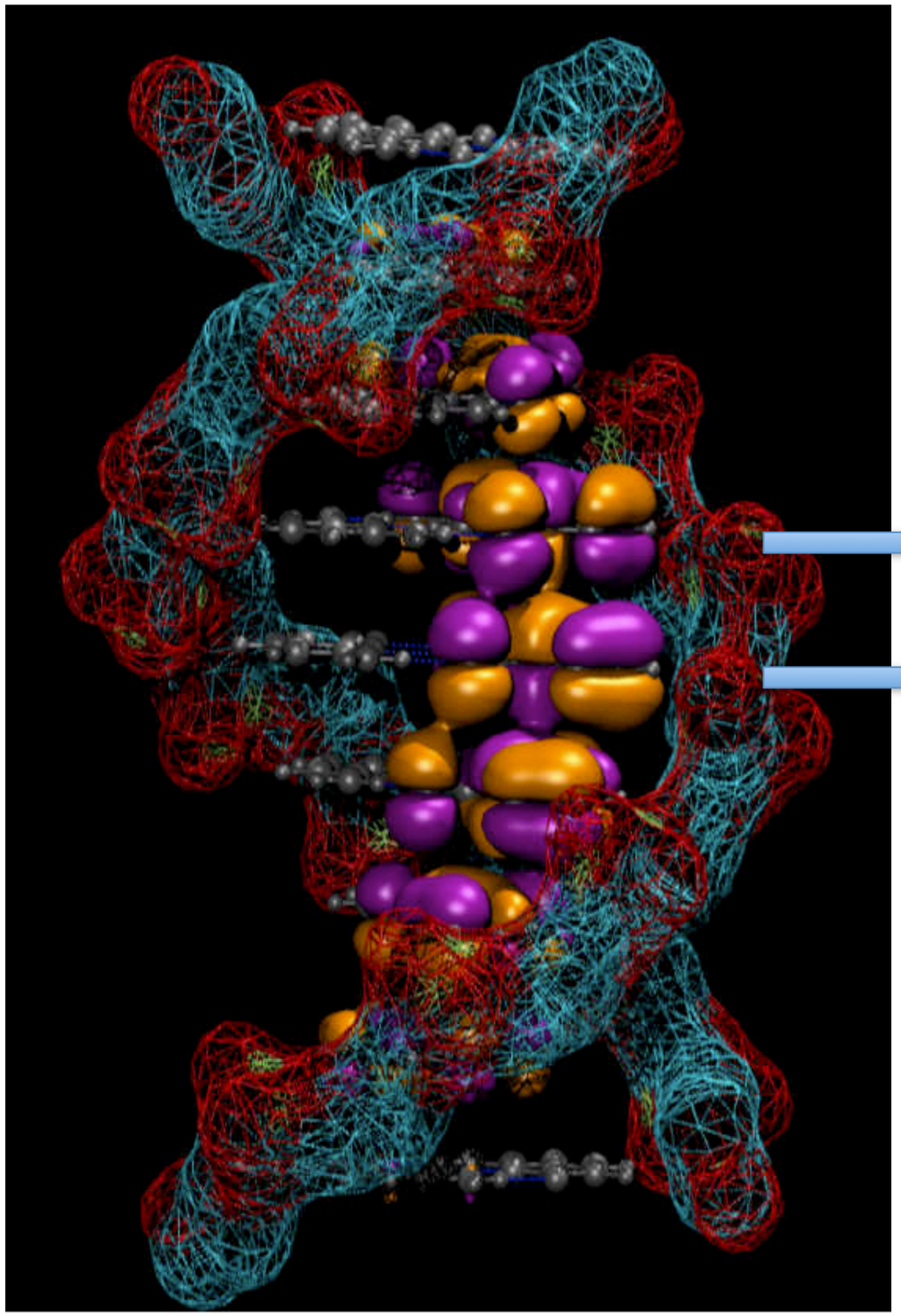


- system very large: 1000 atoms in DNA
- fluctuations important: MD for ns
- solvent explicitly required: put another 5000 atoms

Need QM description: NOT POSSIBLE

➔ Coarse graining of the electronic problem

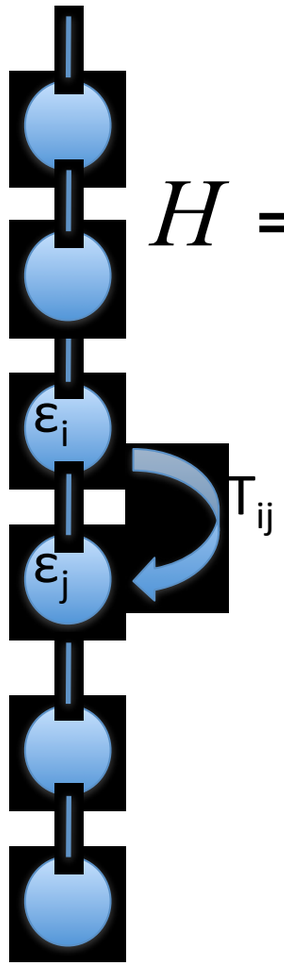
Charge transfer through DNA



site i

site j

Coarse grained
Hamiltonian



$$H = \sum_i \varepsilon_i a_i^+ a_i + \sum_{ij} T_{ij} a_i^+ a_j$$

Time dependent parameters $\varepsilon_i(t)$ and $T_{ij}(t)$ contain dynamical and solvation effects

Multiscale modelling

sequential: simulation with only one method

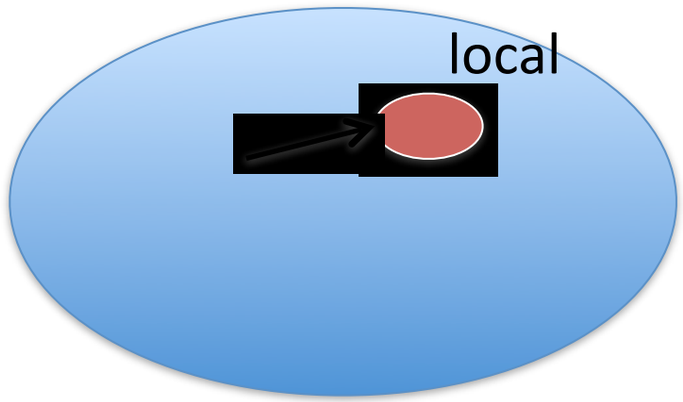
a) Get parameters : ‚bottom up' parametrization

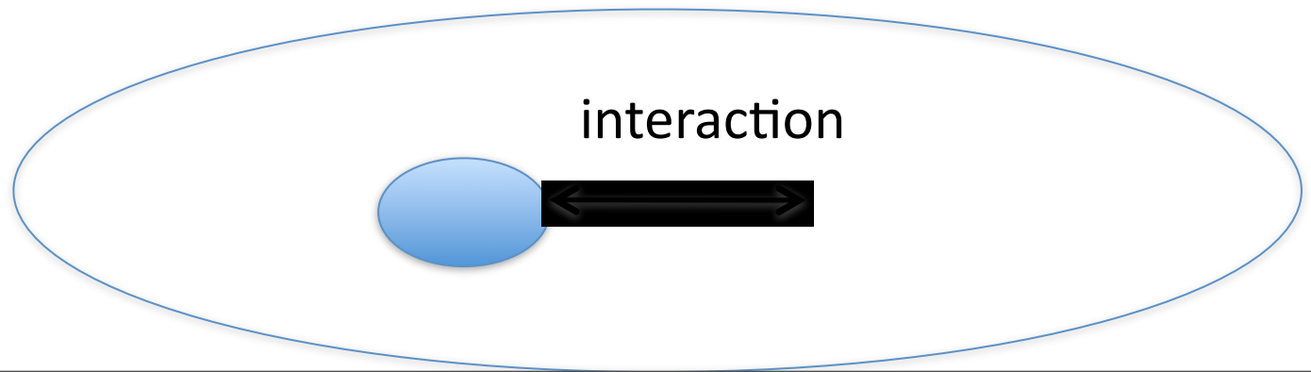
integrated : several methods combined

b) Even in a good model, often more accurate information is needed locally

e.g. crack propagation

c) atomistic simulations : long-range interactions





Computational Biophysics: atomistic simulations

empirical potentials, statistical mechanics

I. Dynamics of complex structures

- protein folding
- molecular motors
- protein-DNA complexes

II. Transport: water, ions, protons, …

III. Electron transfer

IV. Enzymes

- catalysis
- photochemistry

I. Dynamics of complex structures

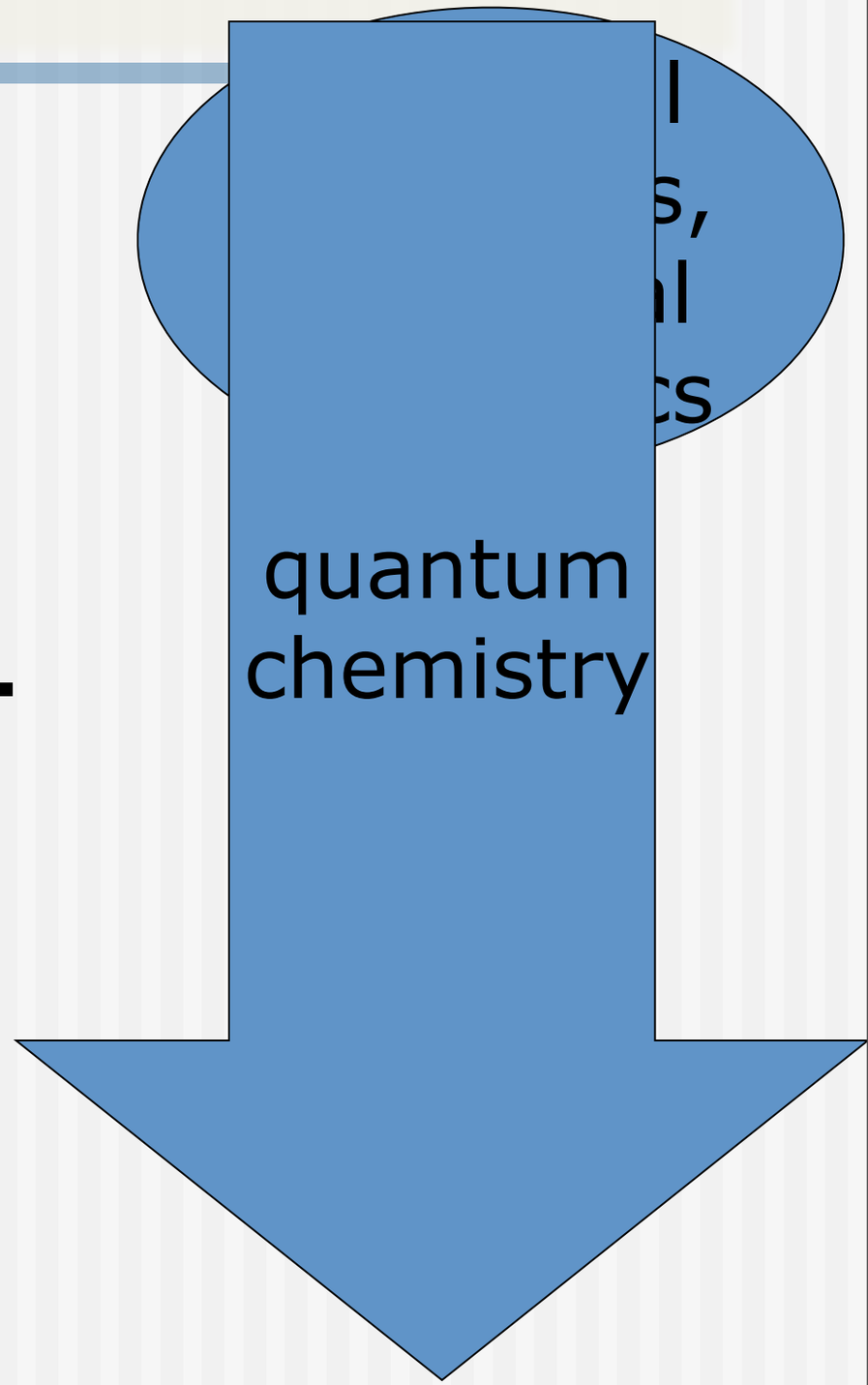
- protein folding
- molecular motors
- protein-DNA complexes

II. Transport: water, ions, protons, …

III. Electron transfer

IV. Enzymes

- catalysis
- photochemistry



I. Dynamics of complex structures

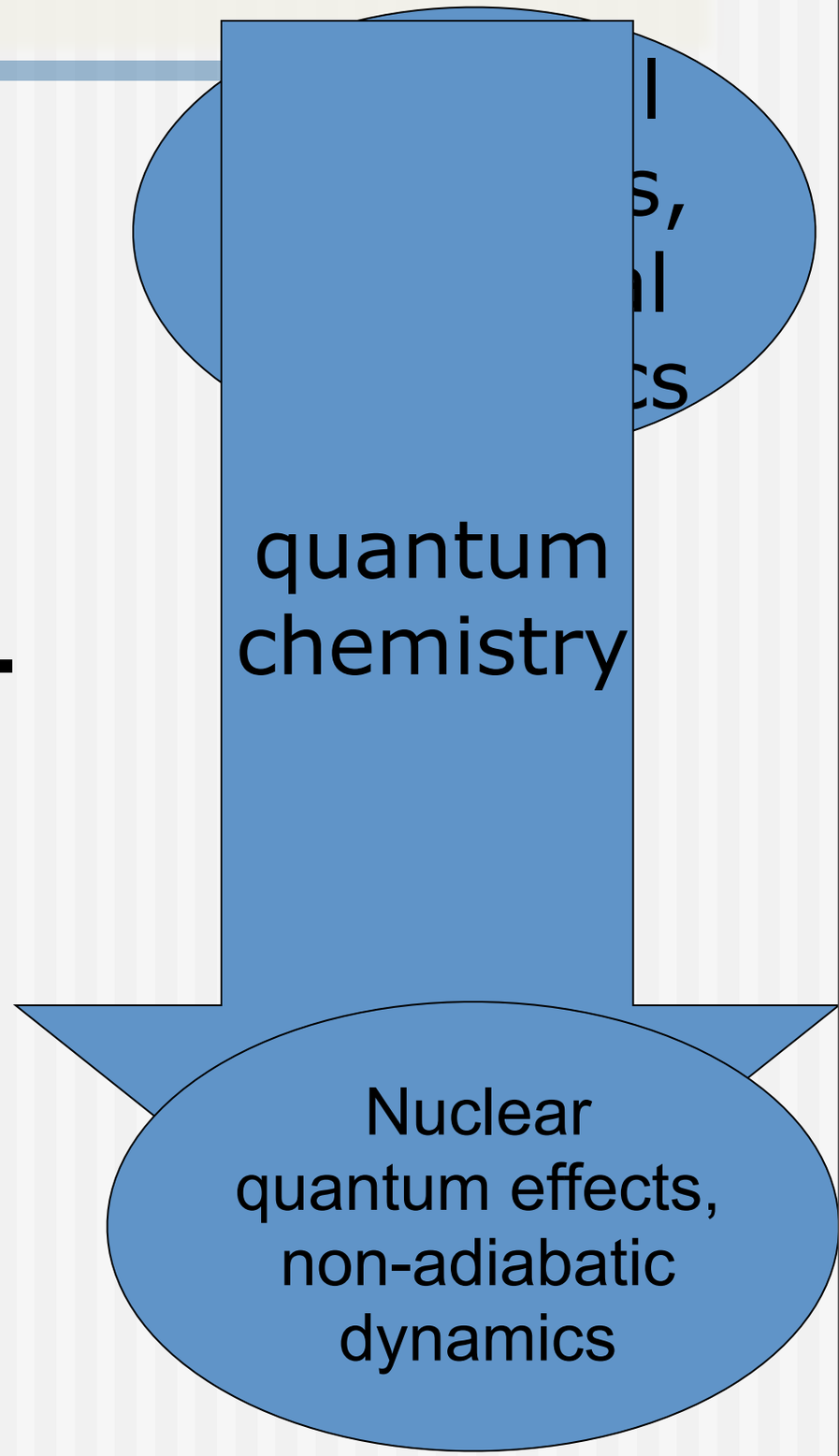
- protein folding
- molecular motors
- protein-DNA complexes

II. Transport: water, ions, protons, …

III. Electron transfer

IV. Enzymes

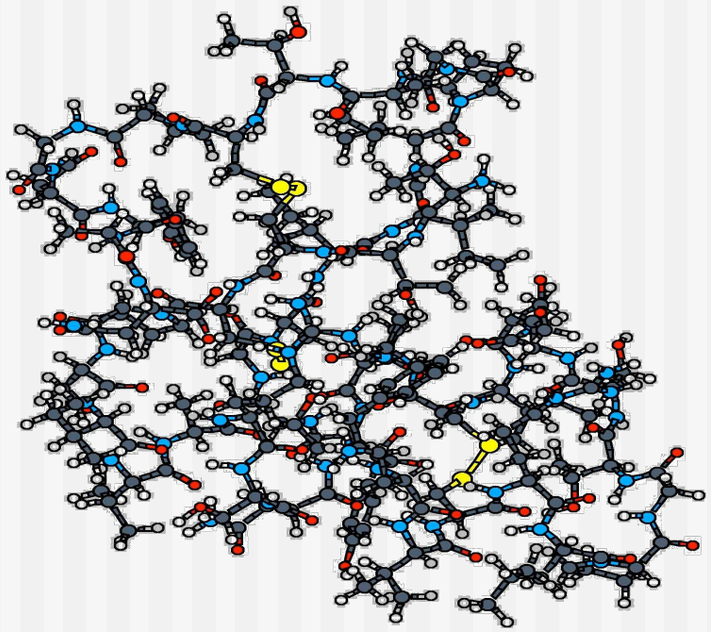
- catalysis
- photochemistry



Characteristics of biological matter

1. Although looking chaotic, well ordered structure in terms of electrostatic interactions
2. Long range electrostatic forces: not easy to truncate the system
3. Dynamics often very important
4. Chemical event often localized
5. Electronic structure often complex: high level methods necessary (e.g. DFT fails)

1. Although looking chaotic, well ordered structure in terms of electrostatic interactions



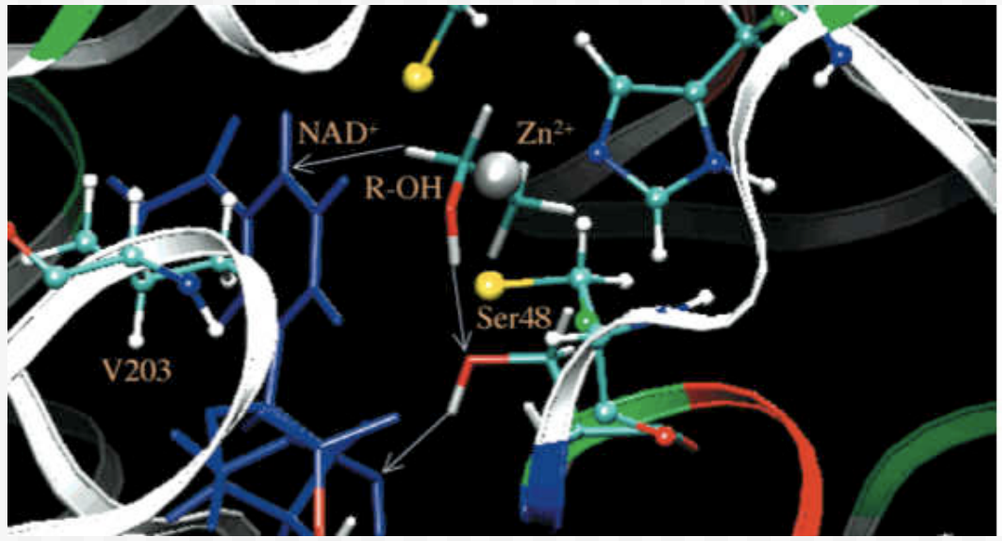
Understanding the action of enzymes

(Warshel, Annu. Rev. Biophys. Biomol. Struct. 2003. 32:425–43**)**

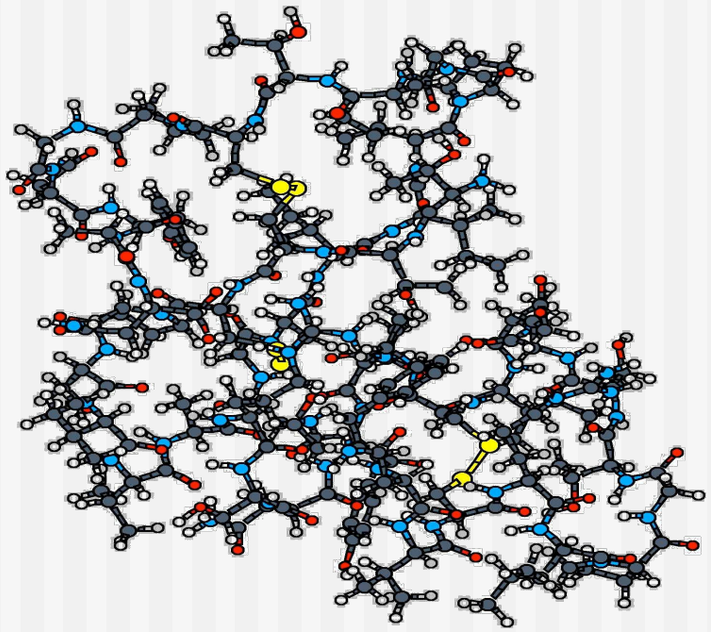
• in most proteins: catalytic effect due to electrostatic interaction with protein environment!

less important:

- ‚desolvation‘
- steric effects
- ‚near attac conformation‘ (NAC)
- ‚coherent dynamics‘



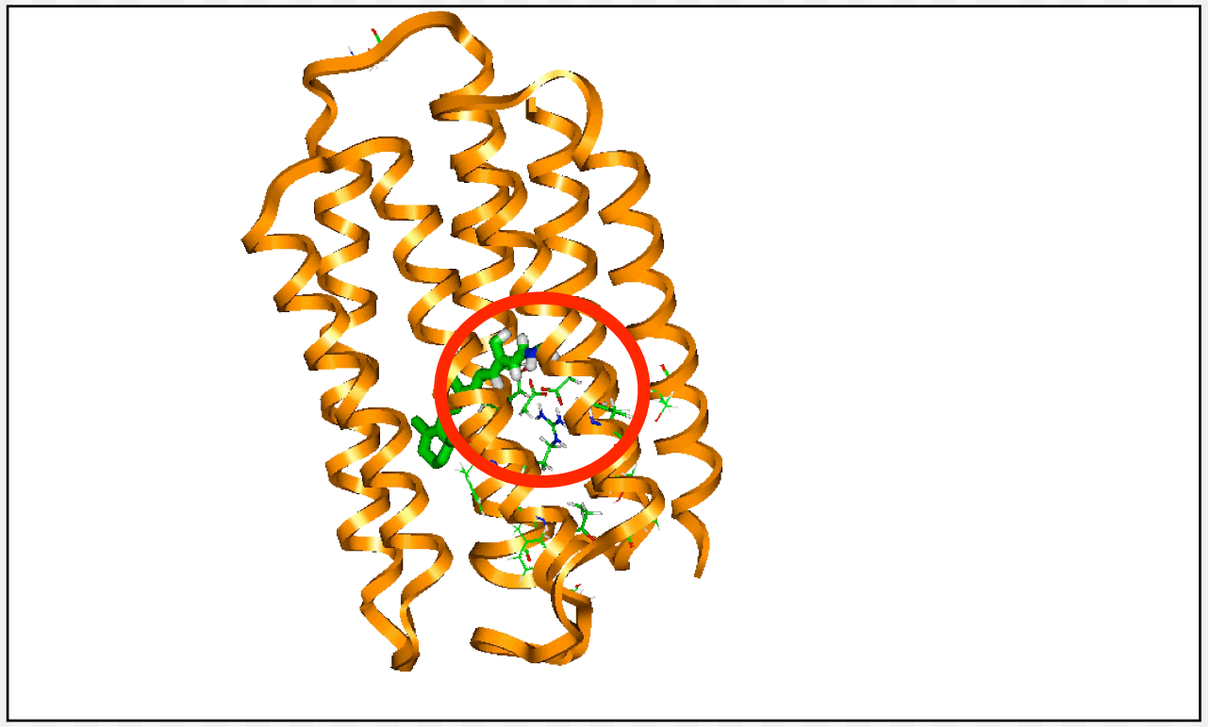
1. Although looking chaotic, well ordered structure in terms of electrostatic interactions



Understanding the action of enzymes

(Warshel, Annu. Rev. Biophys. Biomol. Struct. 2003. 32:425–43**)**

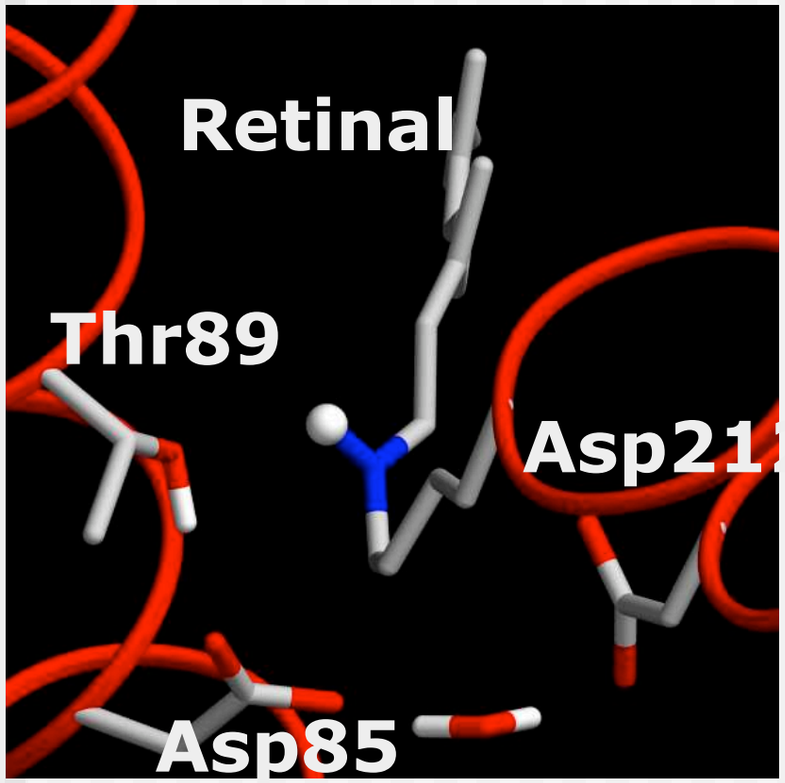
• in most proteins: catalytic effect due to electrostatic interaction with protein environment!

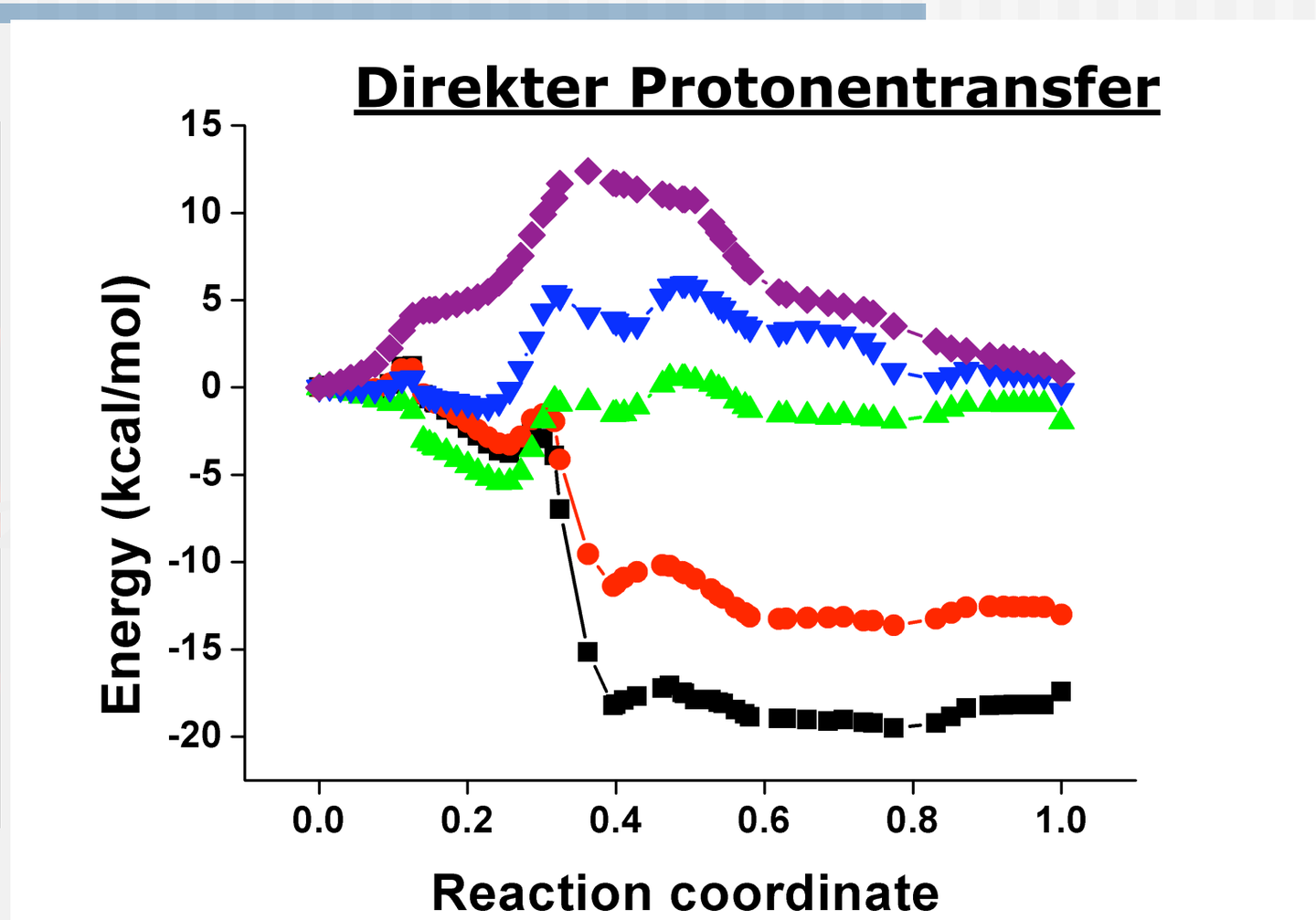


less important:

- ‚desolvation‘
- steric effects
- ‚near attac conformation‘ (NAC)
- ‚coherent dynamics‘

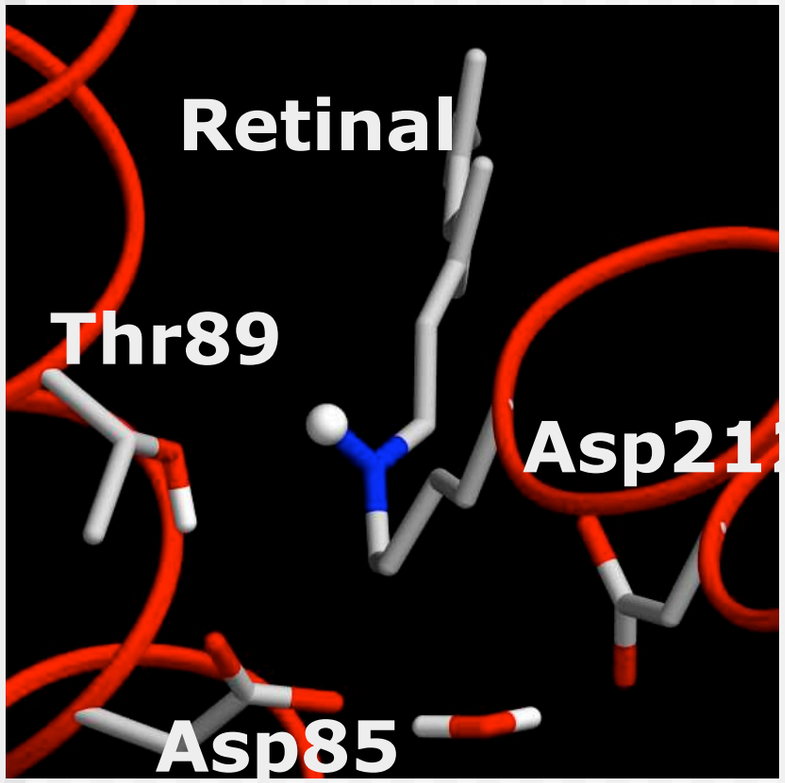
1. Although looking chaotic, well ordered structure in terms of electrostatic interactions

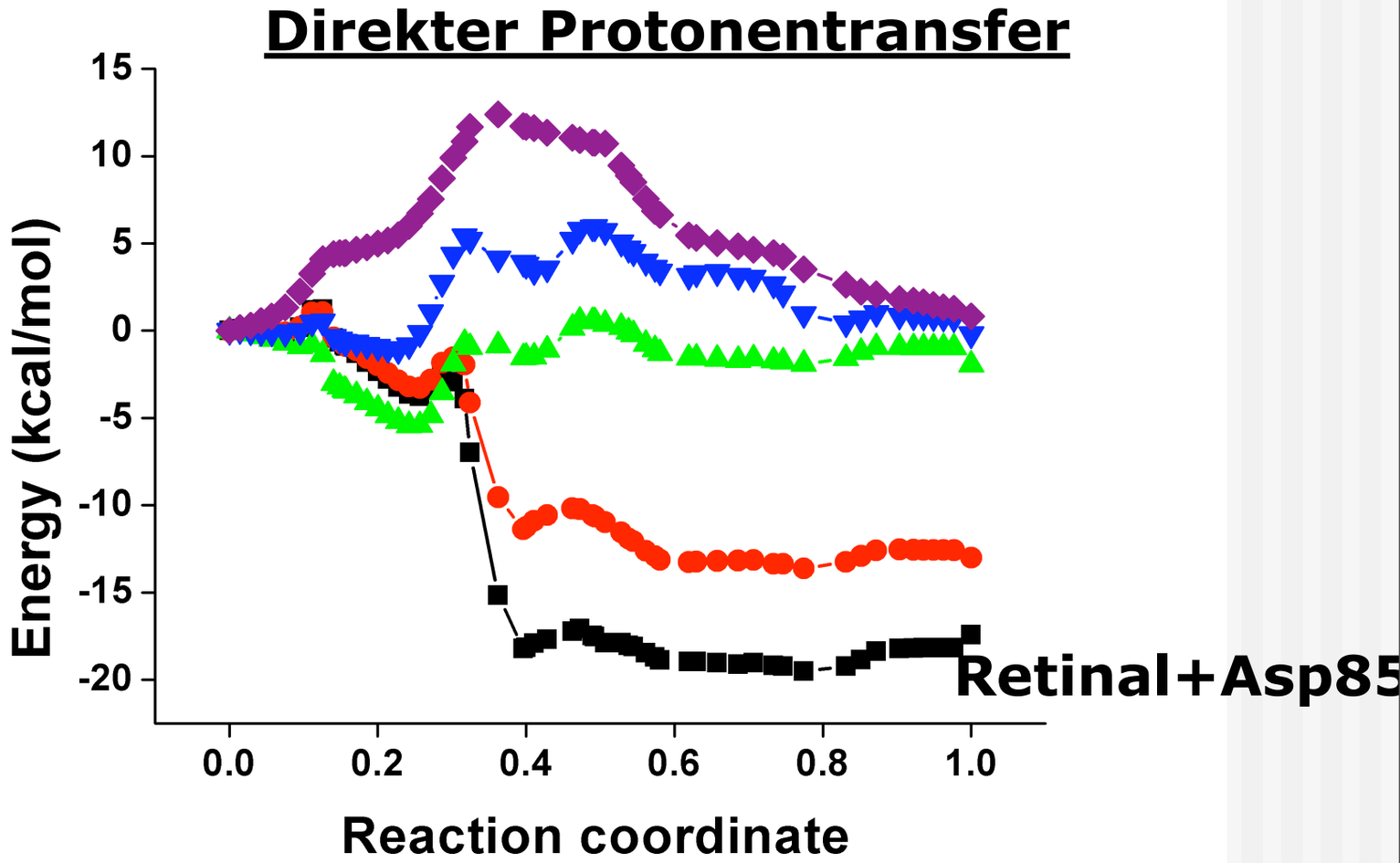




- whole protein contributes to reaction barrier
- ‘special design’ in order to provide specific function

1. Although looking chaotic, well ordered structure in terms of electrostatic interactions

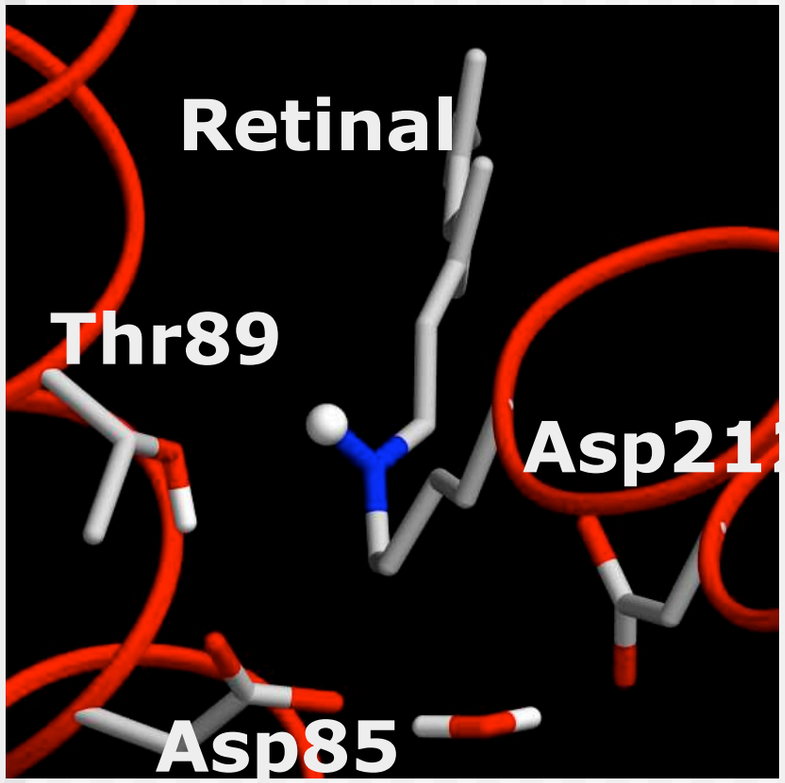


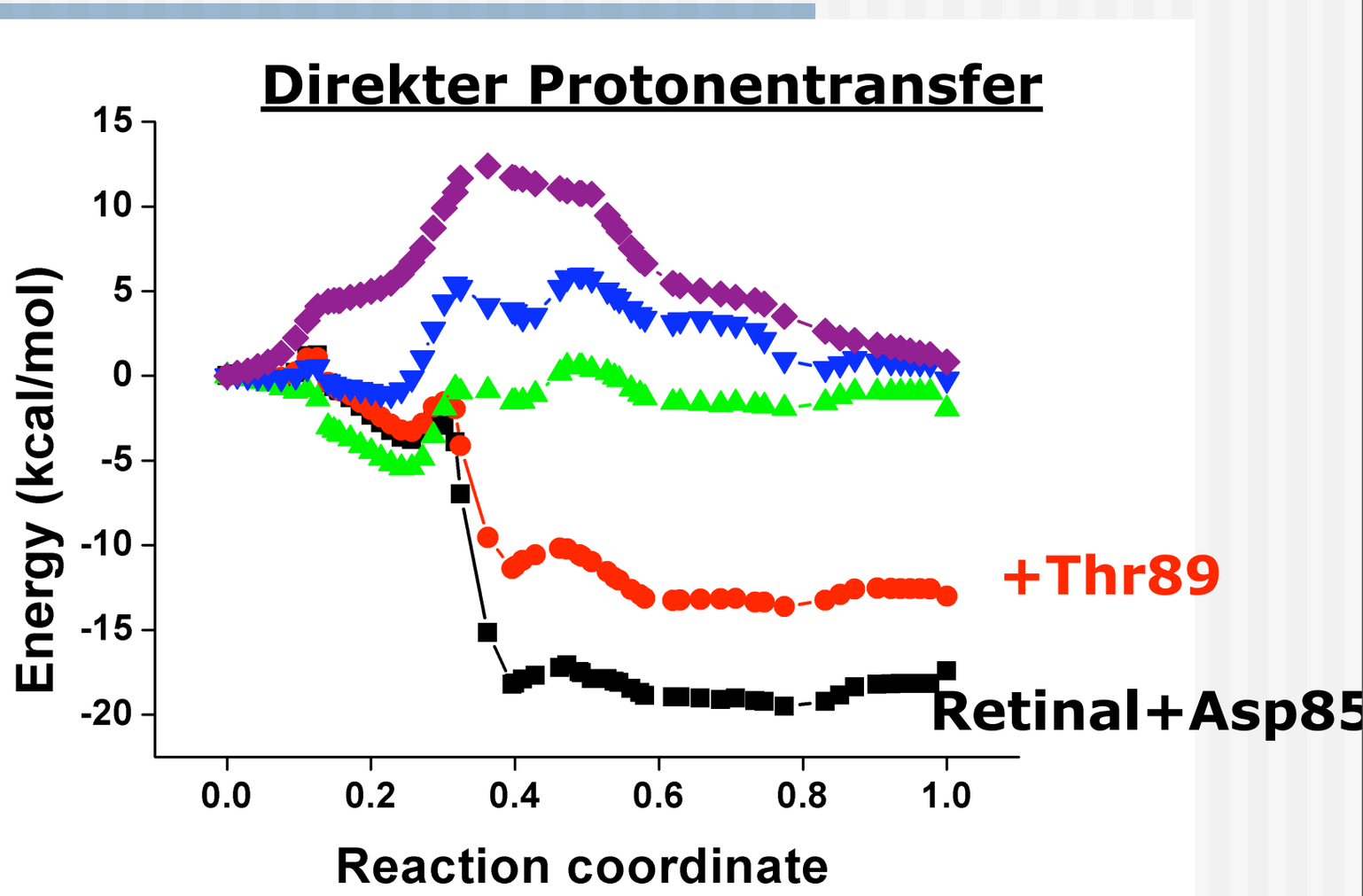


- whole protein contributes to reaction barrier
- ‘special design’ in order to provide specific function

active

1. Although looking chaotic, well ordered structure in terms of electrostatic interactions

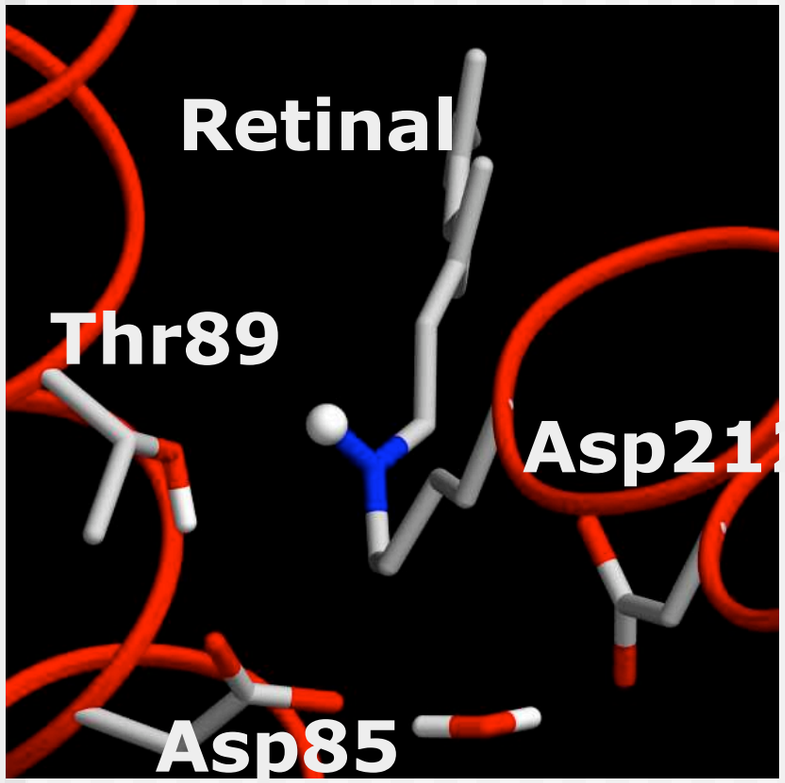


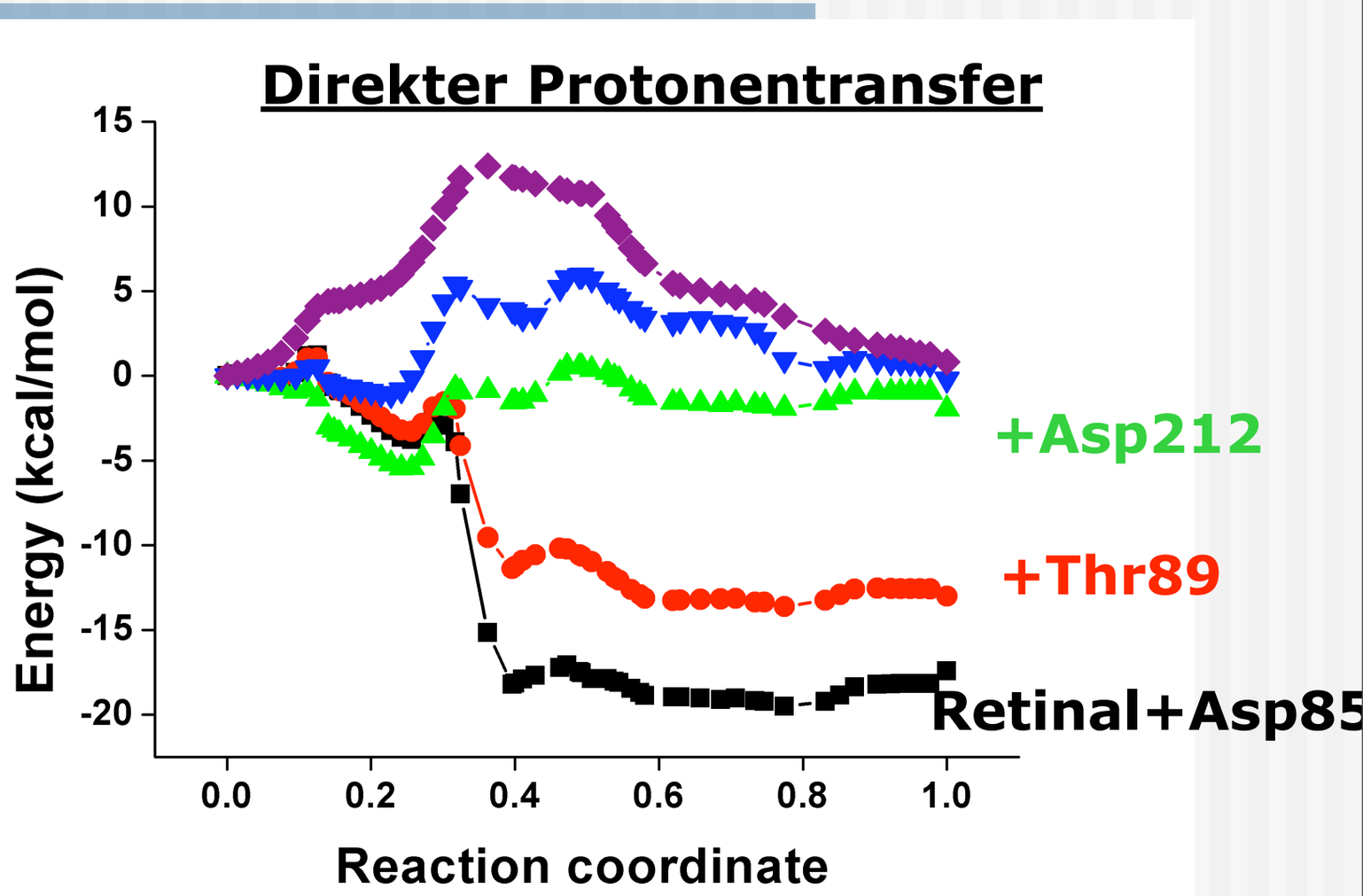


- whole protein contributes to reaction barrier
- ‘special design’ in order to provide specific function

active

1. Although looking chaotic, well ordered structure in terms of electrostatic interactions

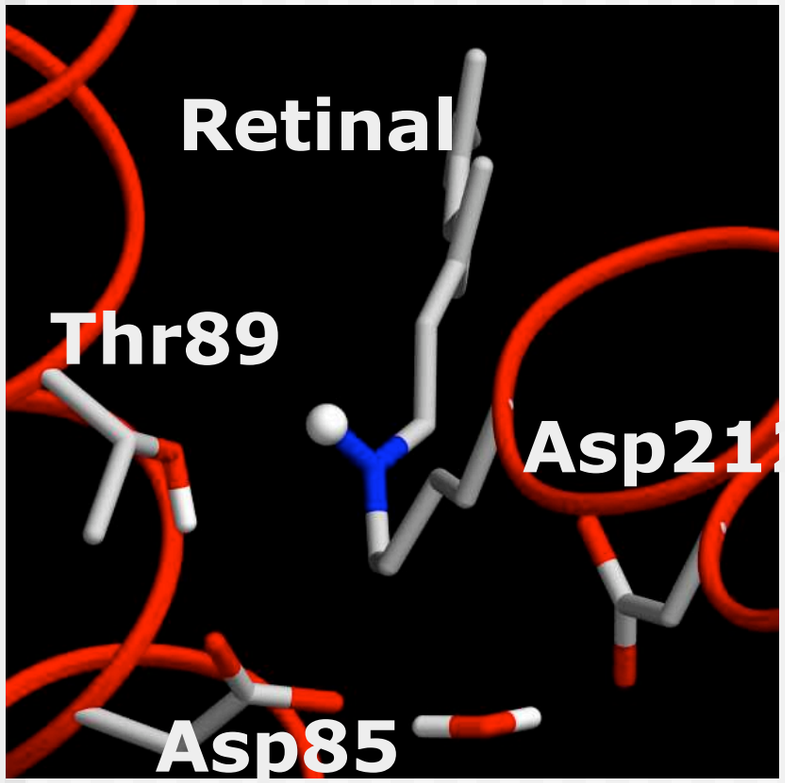


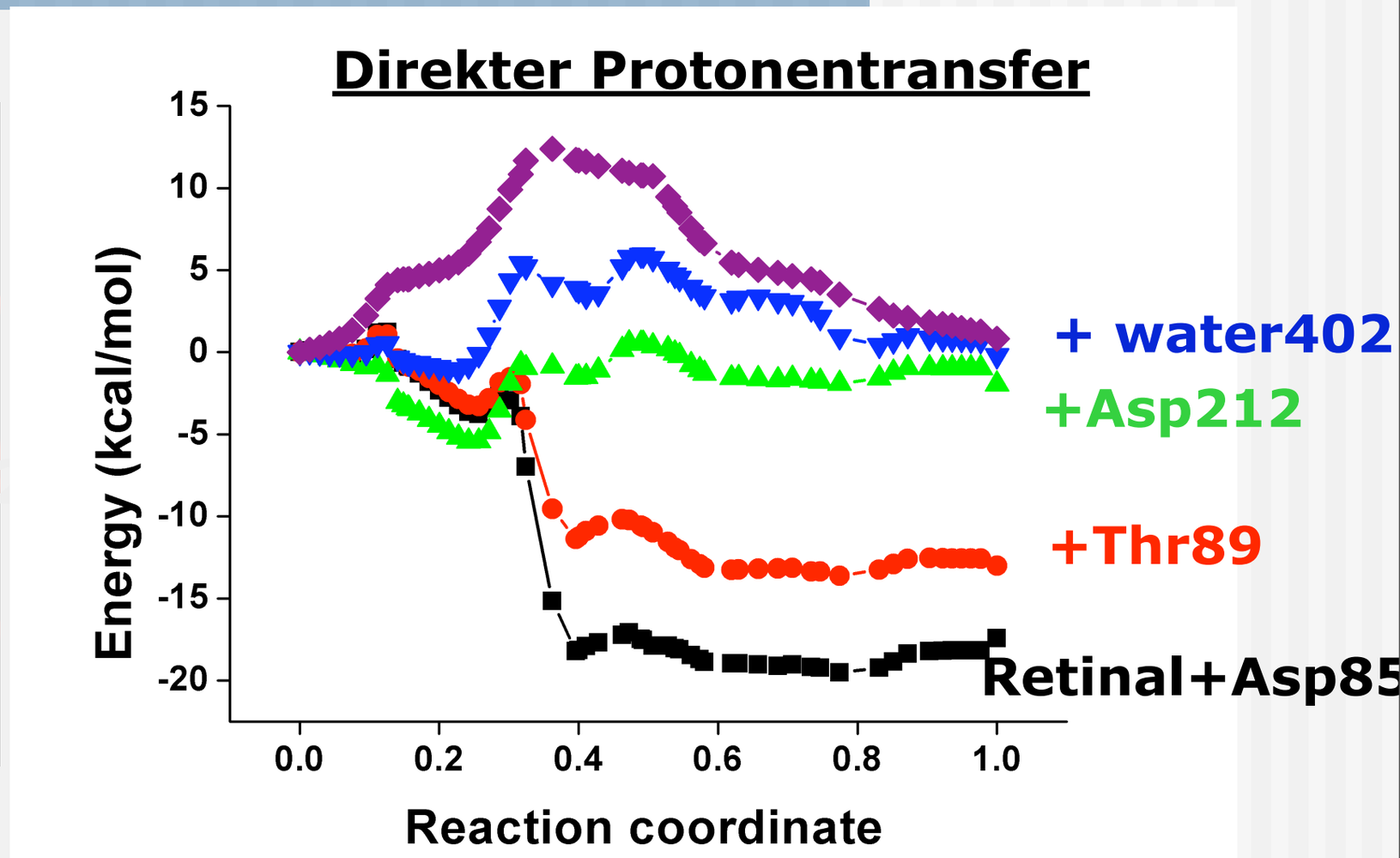


- whole protein contributes to reaction barrier
- ‘special design’ in order to provide specific function

active

1. Although looking chaotic, well ordered structure in terms of electrostatic interactions

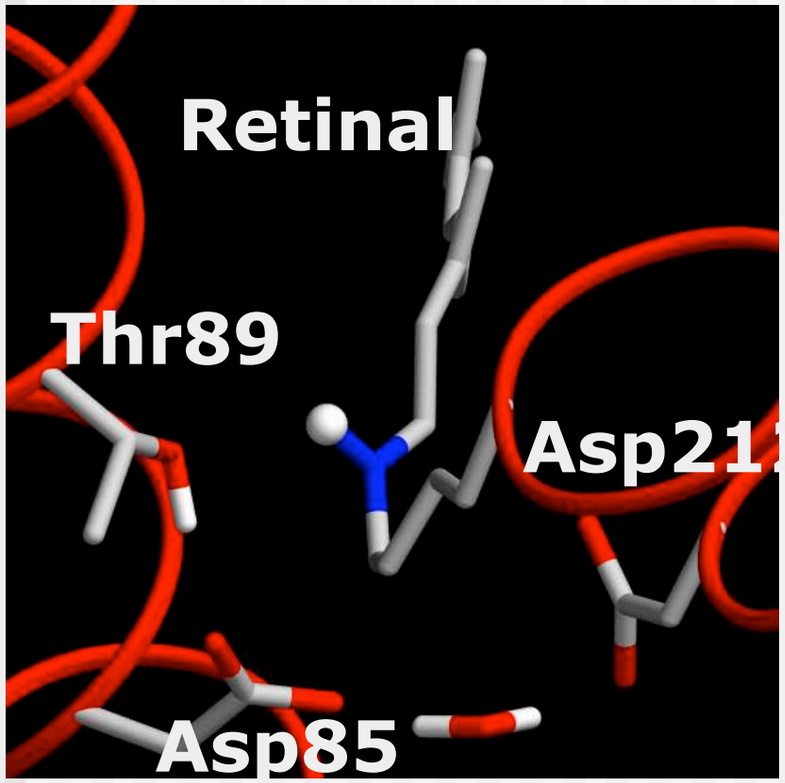


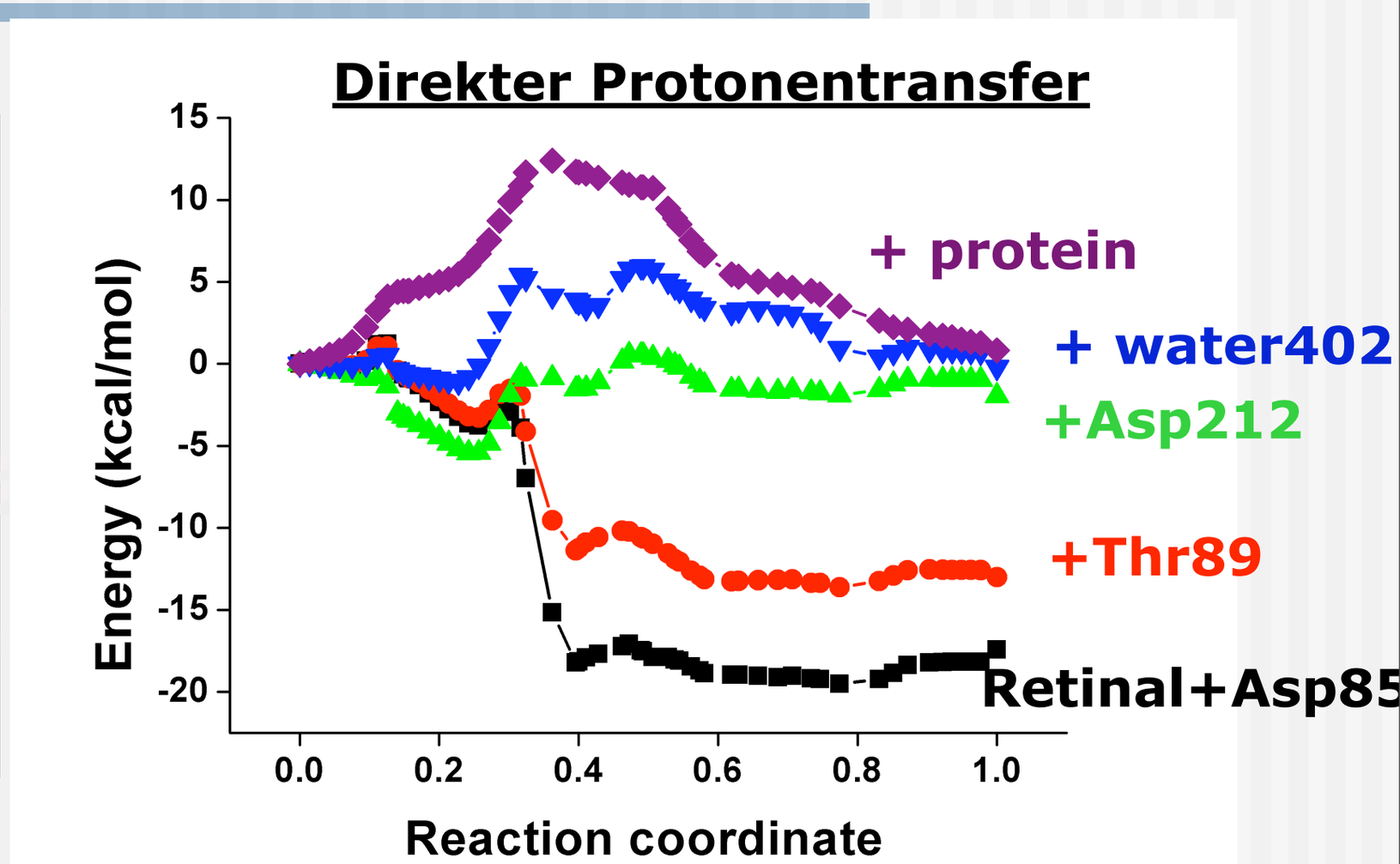


- whole protein contributes to reaction barrier
- ‘special design’ in order to provide specific function

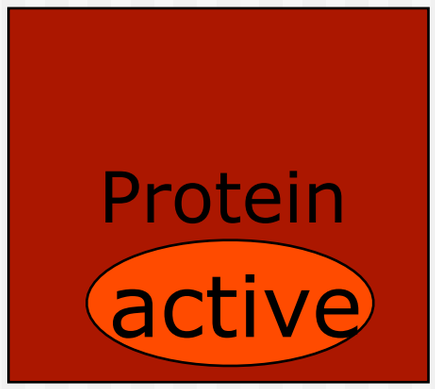
active

1. Although looking chaotic, well ordered structure in terms of electrostatic interactions

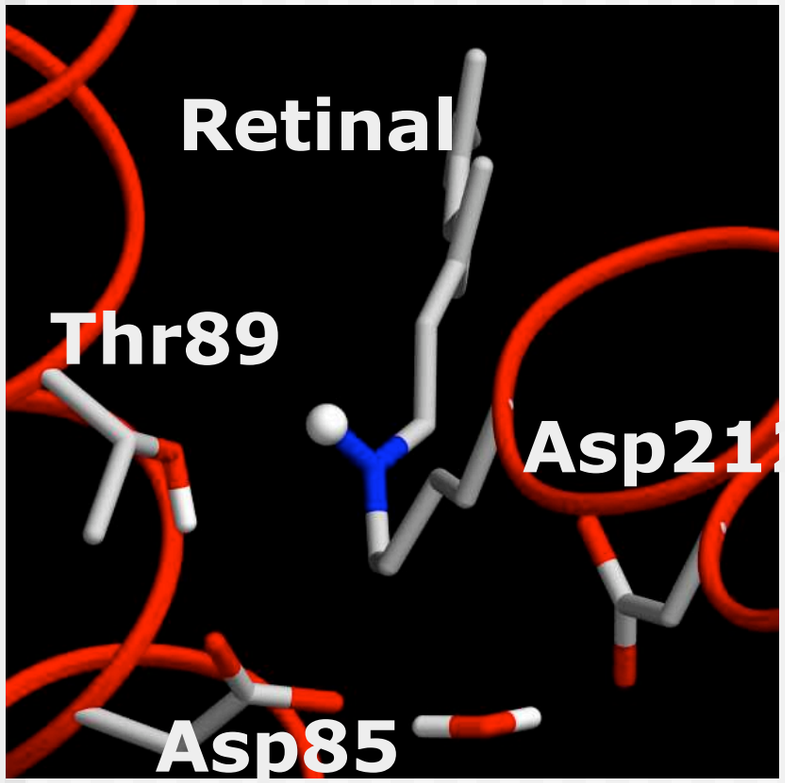


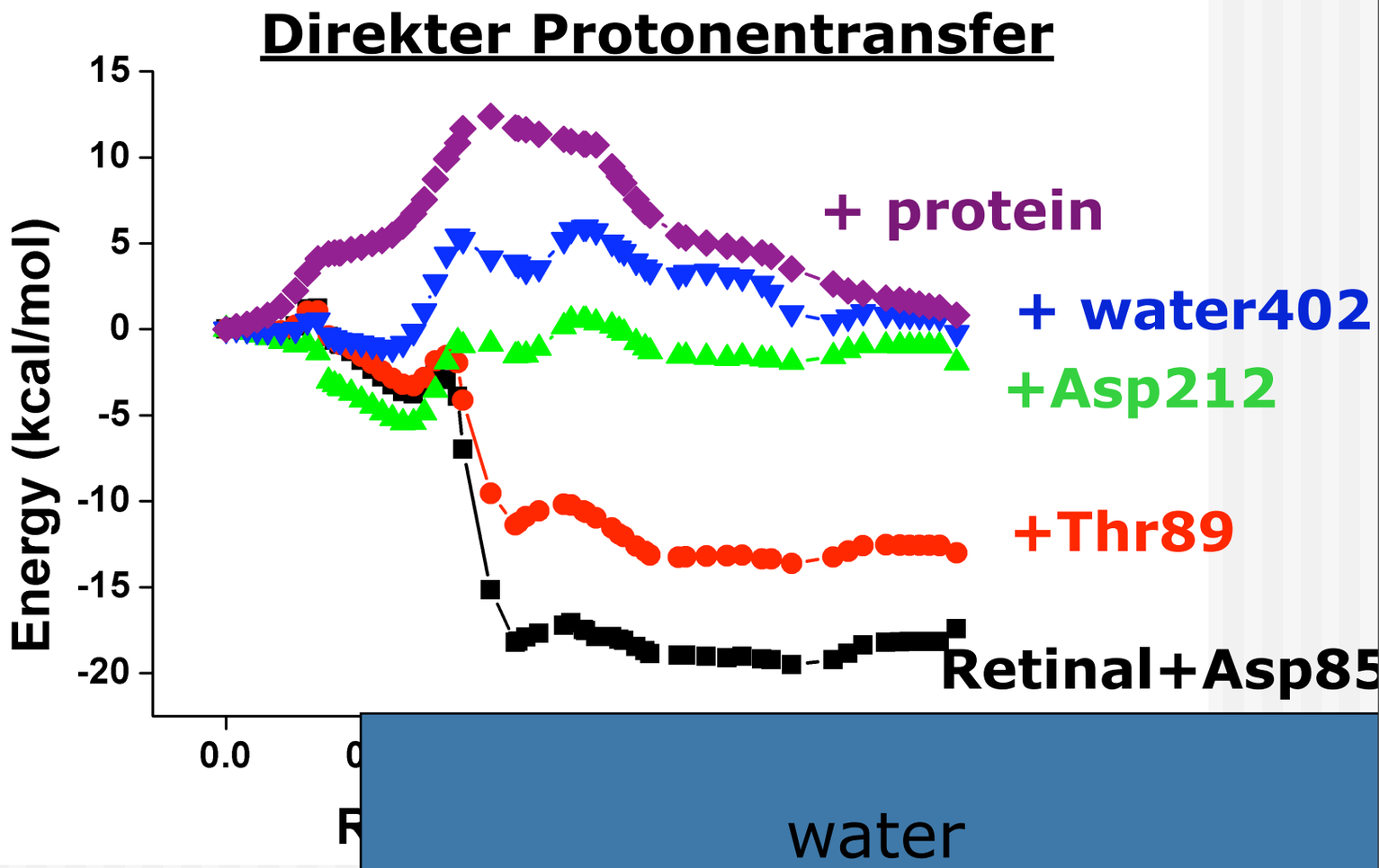


- whole protein contributes to reaction barrier
- ‘special design’ in order to provide specific function



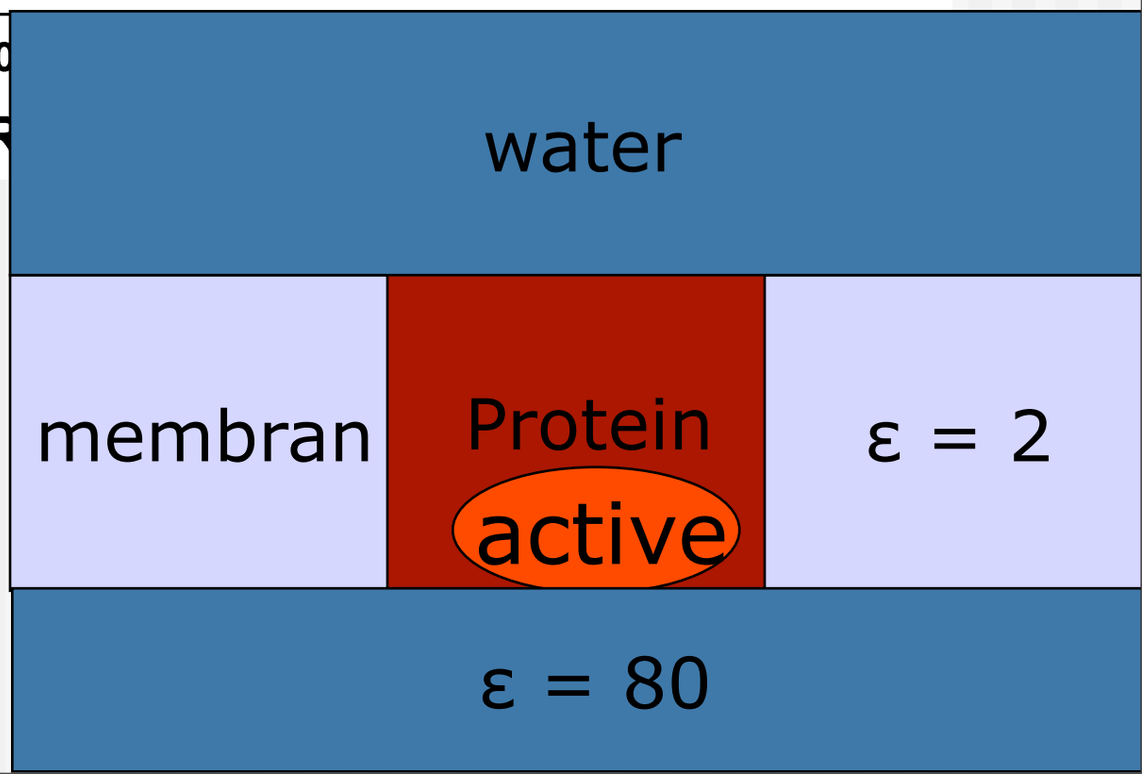
1. Although looking chaotic, well ordered structure in terms of electrostatic interactions



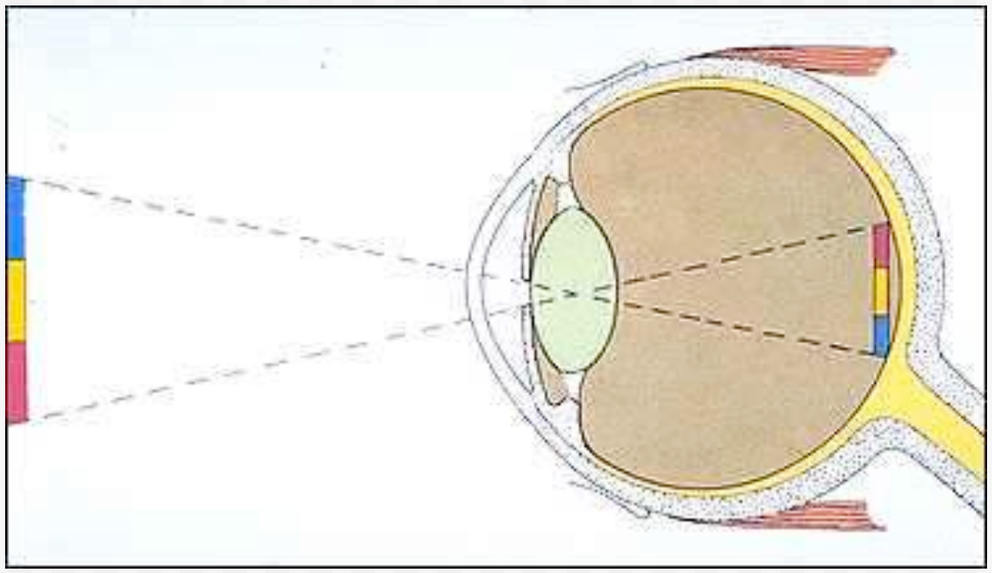


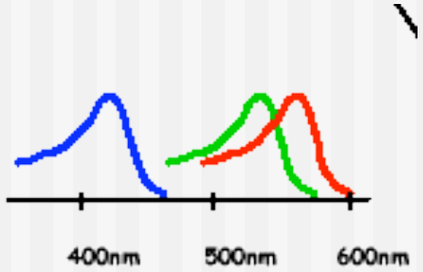
- whole protein contributes to reaction barrier
- ‘special design’ in order to provide specific function

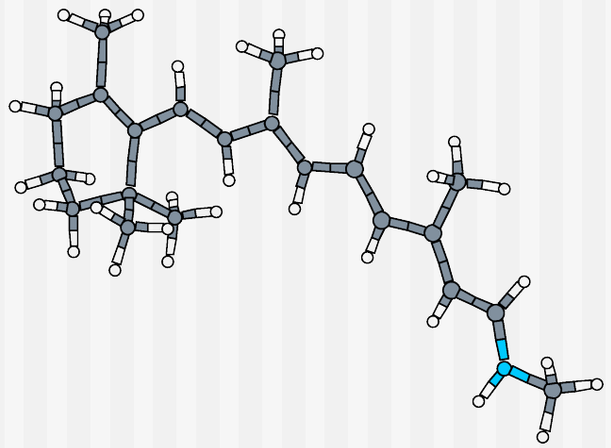
- often even water environment of importance

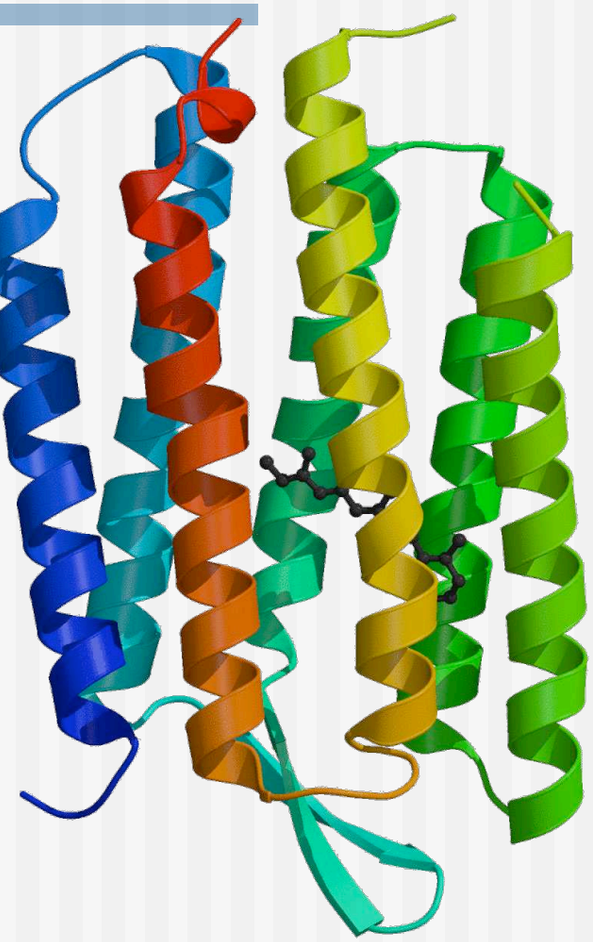


Process of vision









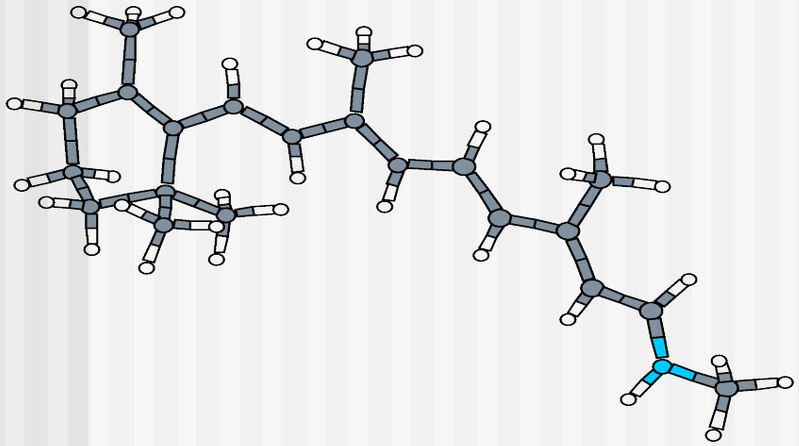
three color pigments, same chromphor:

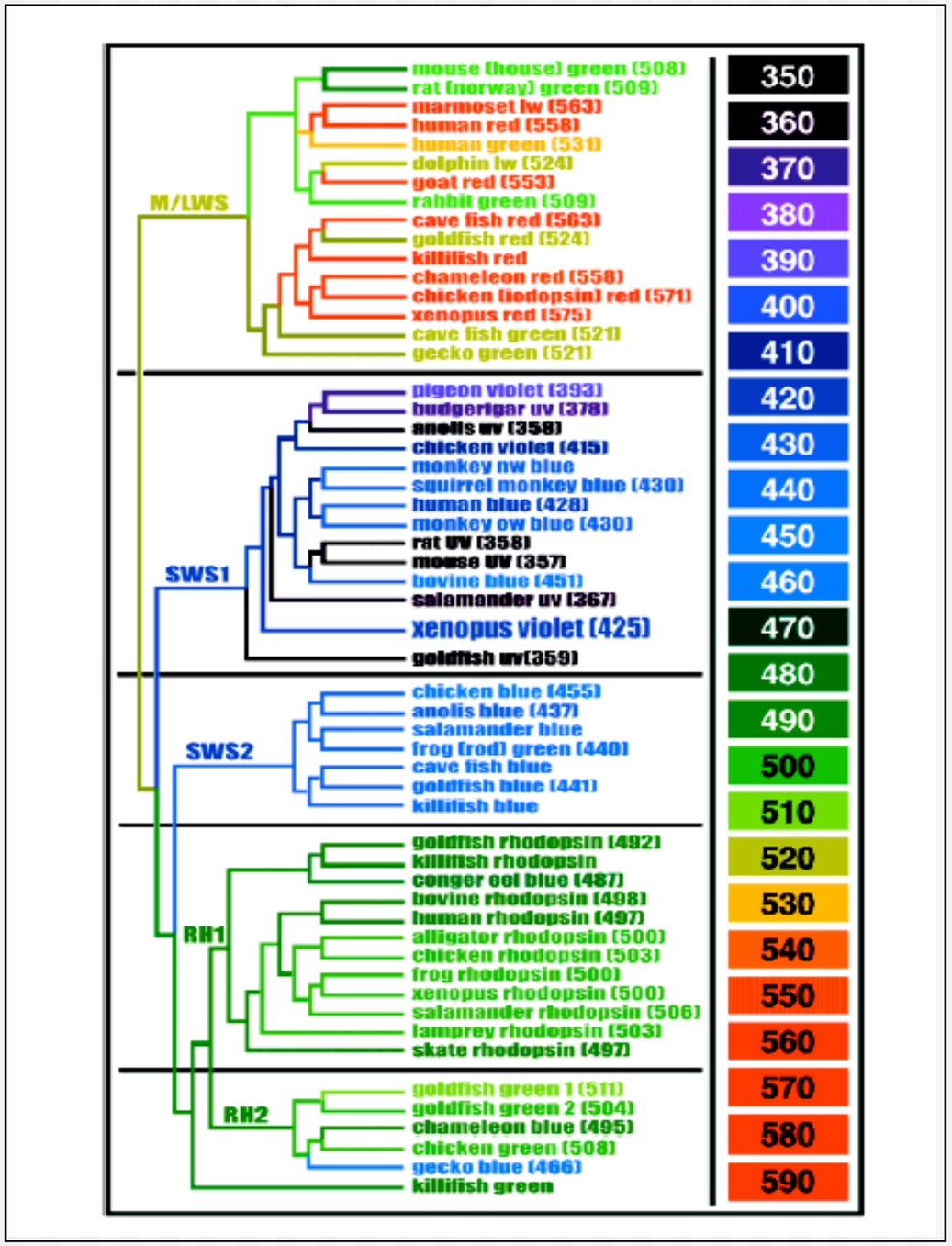
what determines the absorption maximum?

‘Spectral tuning’

Absorption over 300 nm

“Tuning” due to protein environement (opsin-shift)

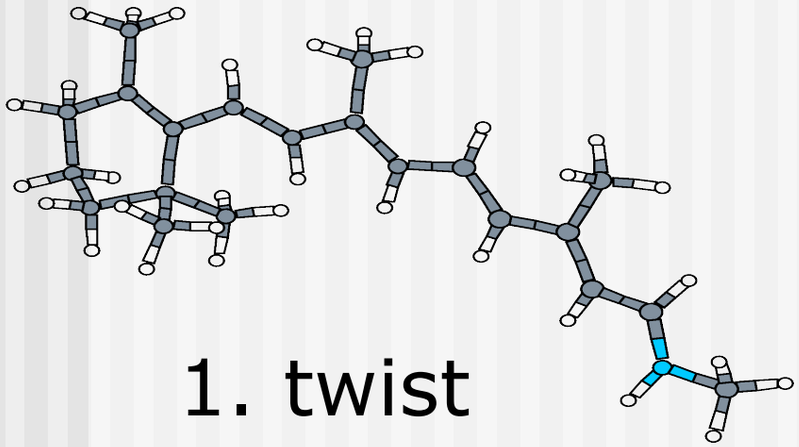




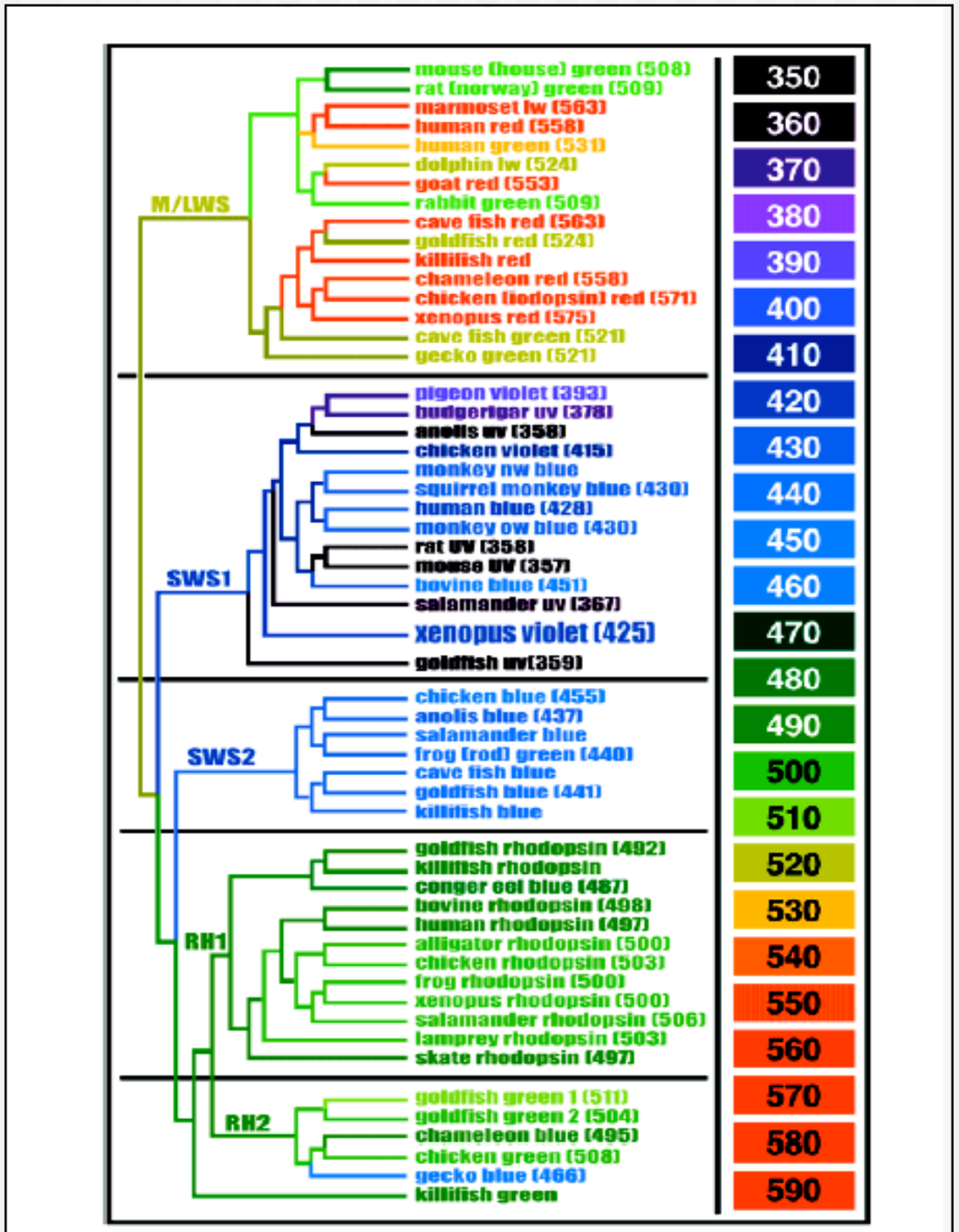
‘Spectral tuning’

Absorption over 300 nm

“Tuning” due to protein environement (opsin-shift)



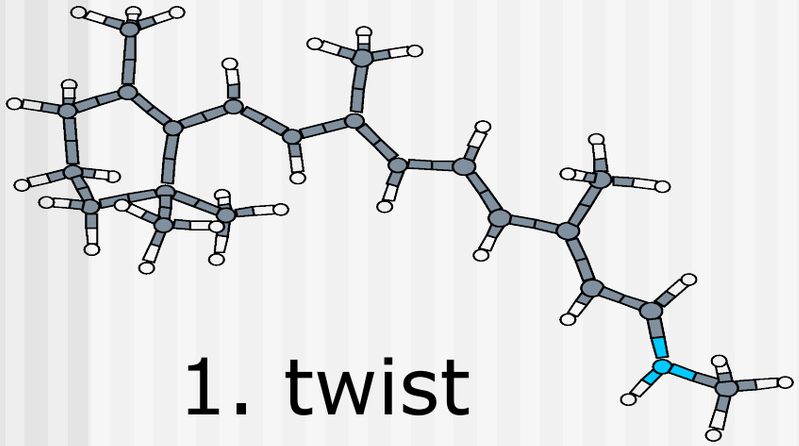
1. twist



‘Spectral tuning’

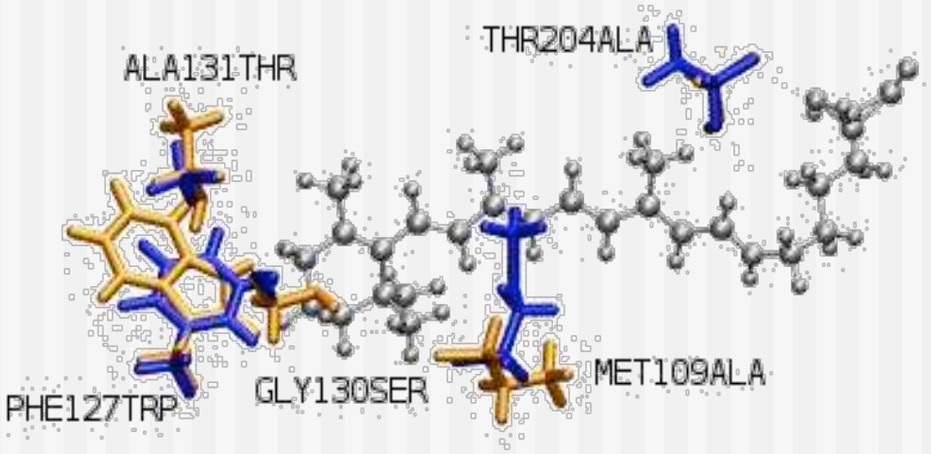
Absorption over 300 nm

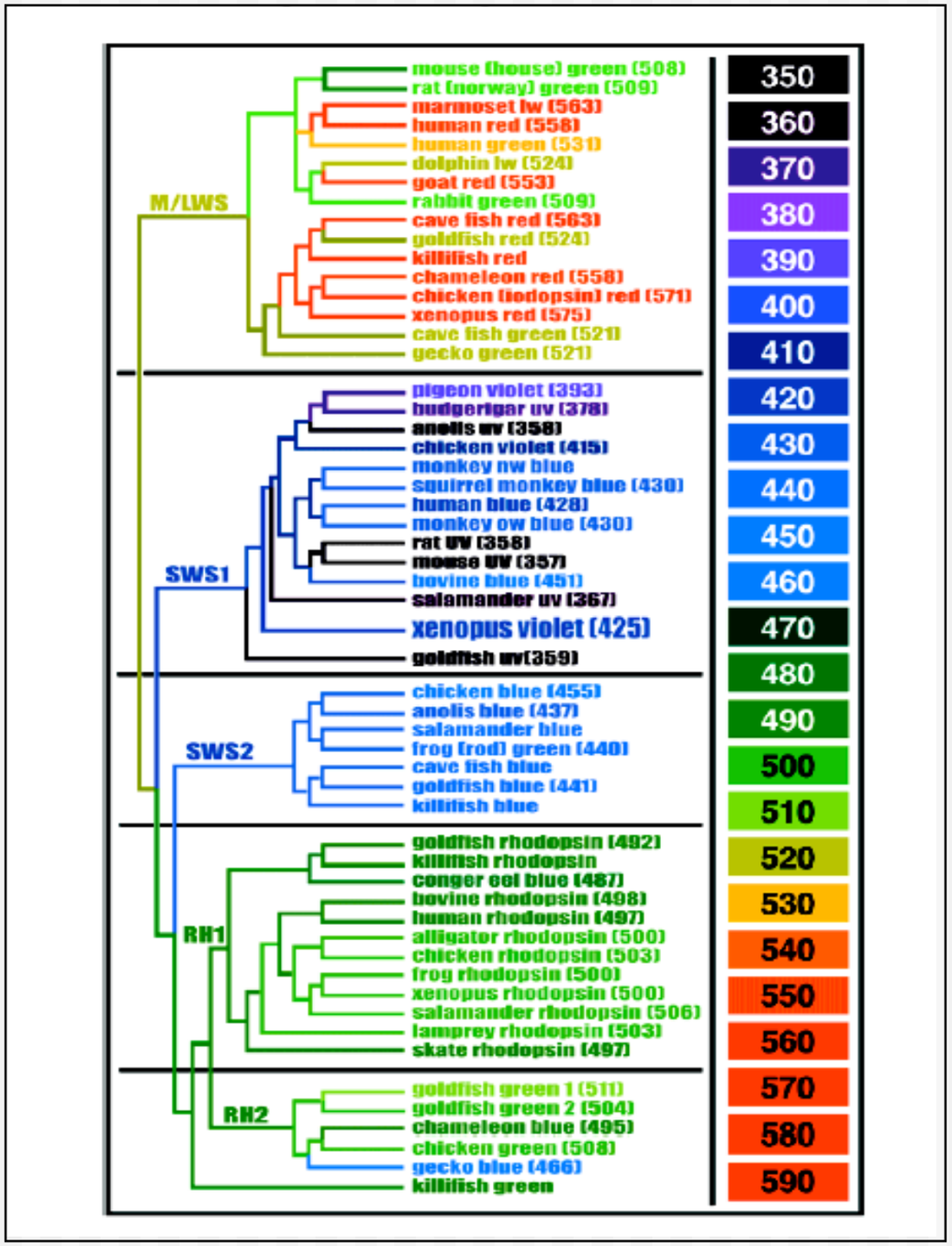
“Tuning” due to protein environement (opsin-shift)



1. twist

2. interaction with polar/ charged groups

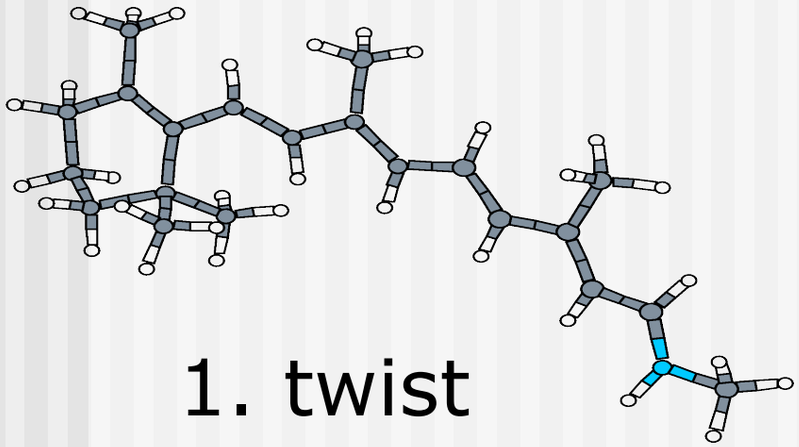




‘Spectral tuning’

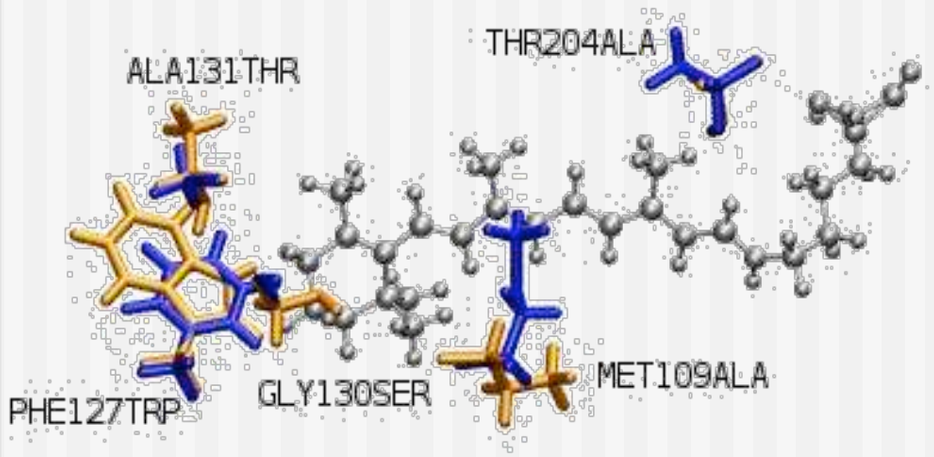
Absorption over 300 nm

“Tuning” due to protein environement (opsin-shift)

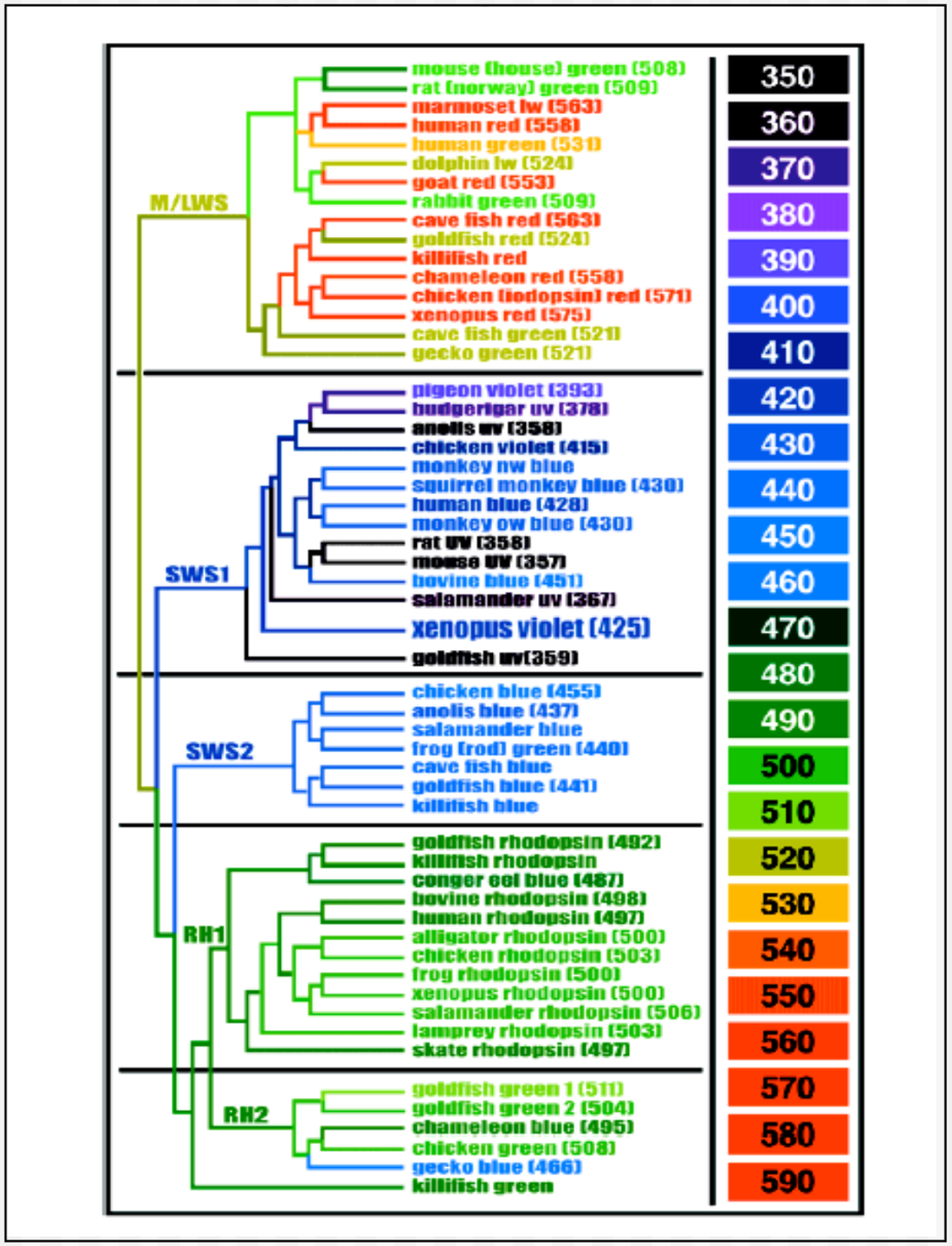


1. twist

2. interaction with polar/ charged groups



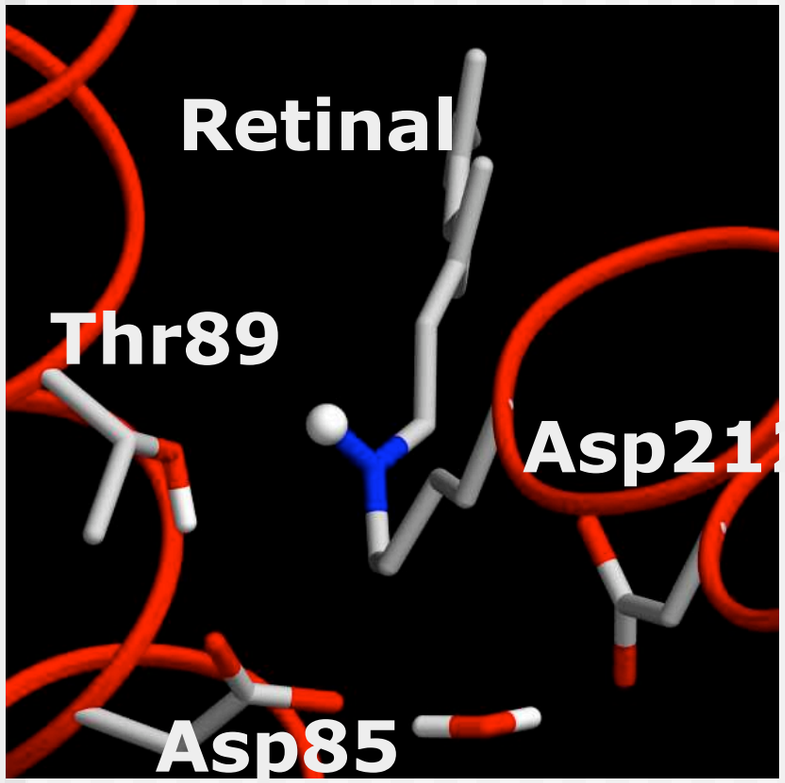
=> ‘predefined’ electrostatic interactions determine function

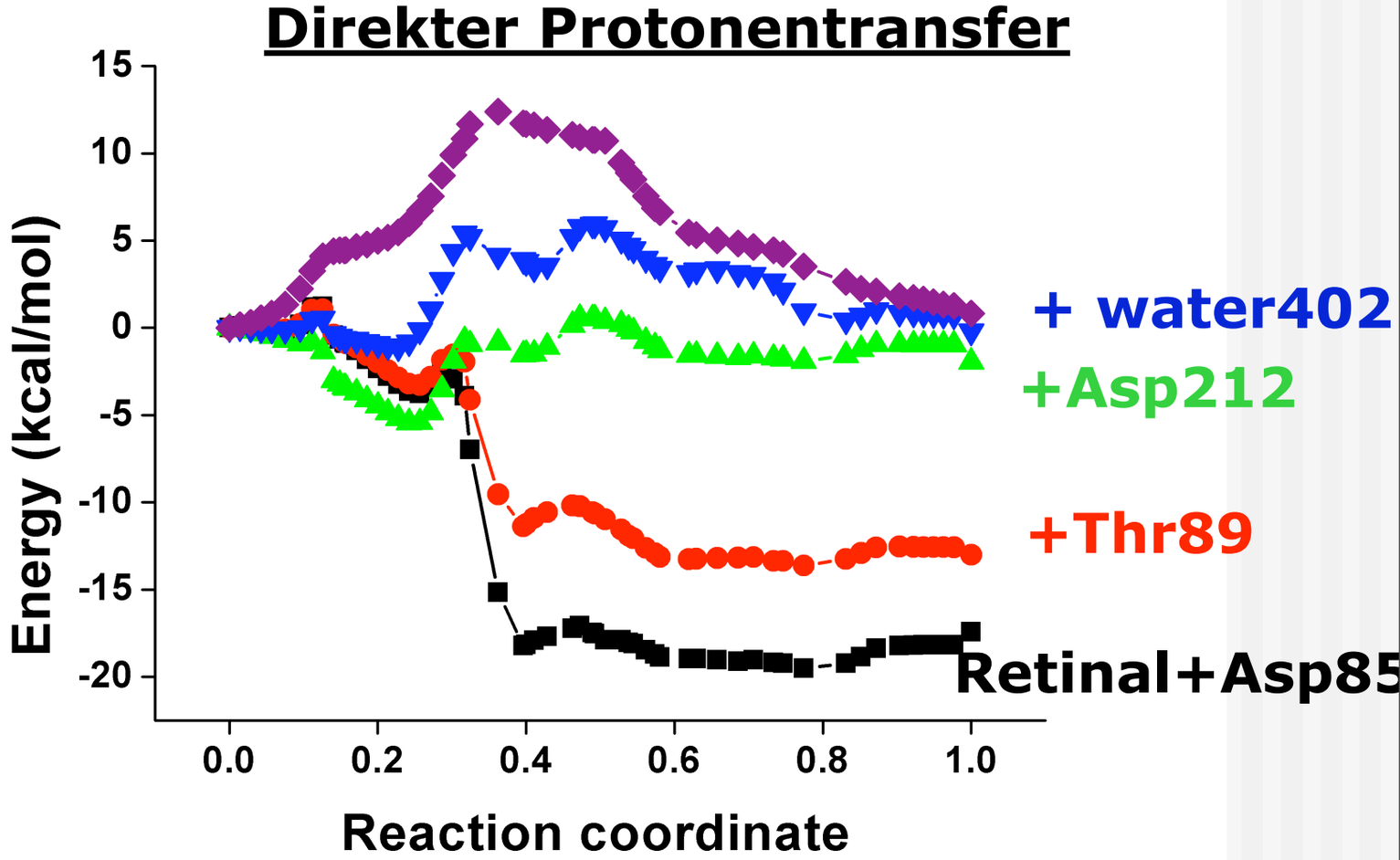


Characteristics of biological matter

1. Although looking chaotic, well ordered structure in terms of electrostatic interactions
2. Long range electrostatic forces: not easy to truncate the system
3. Dynamics often very important
4. Chemical event often localized
5. Electronic structure often complex: high level methods necessary (e.g. DFT fails)

3. Dynamics often very important





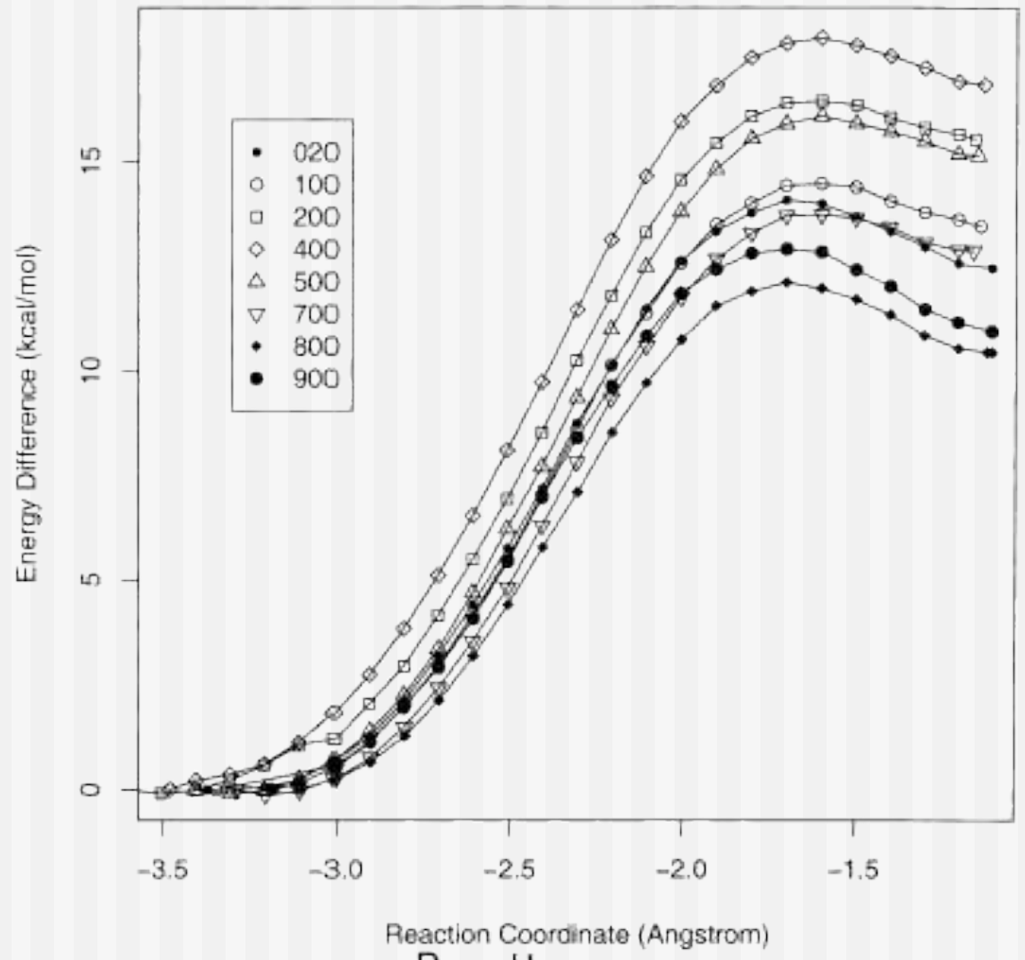
This is the total (potential) energy for one protein structure, but:

- the protein ‘moves’
- entropy

‘Problem’ of total energy

different energy-profiles

for different conformations of the surrounding protein



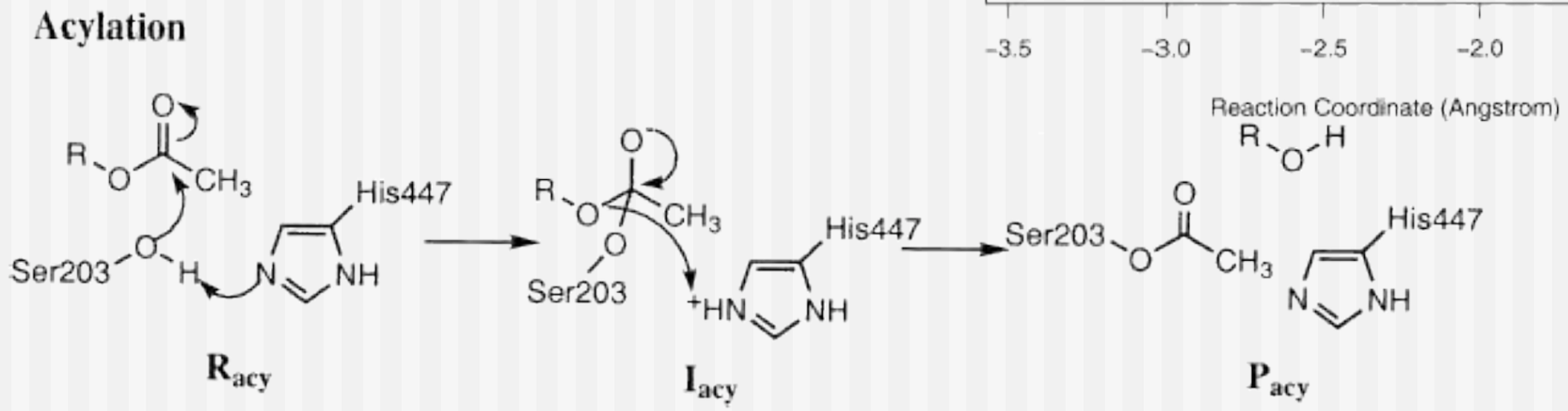
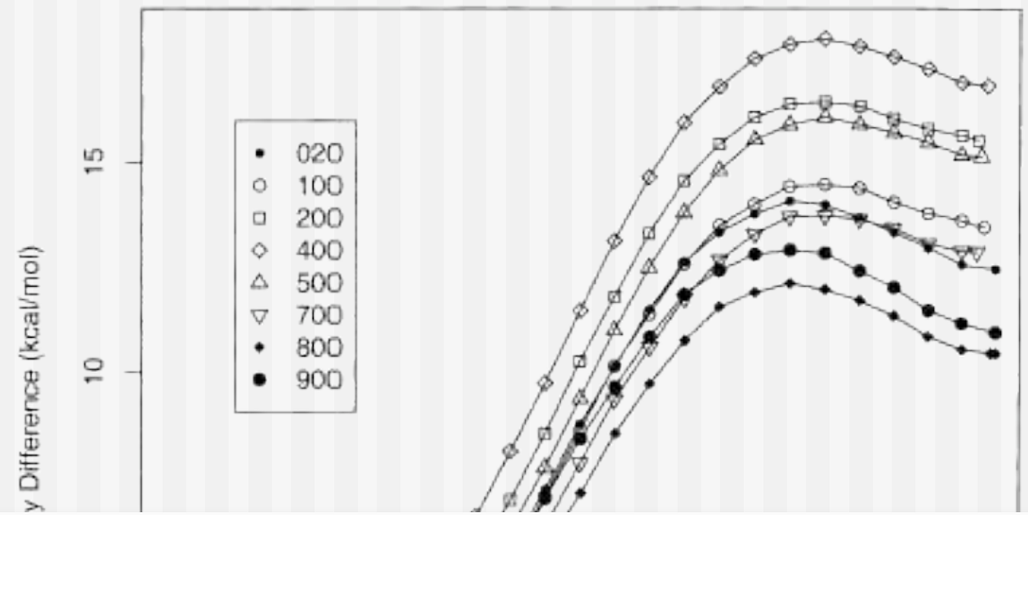


Figure 1. Acylation reaction mechanism of acetylcholine catalyzed by AChE.

Zhang et al JPCB 107 (2003) 44459

different energy-profiles

for different conformations of the surrounding protein



A) one always has to ‘average’ (sample) over acessible protein conformations :

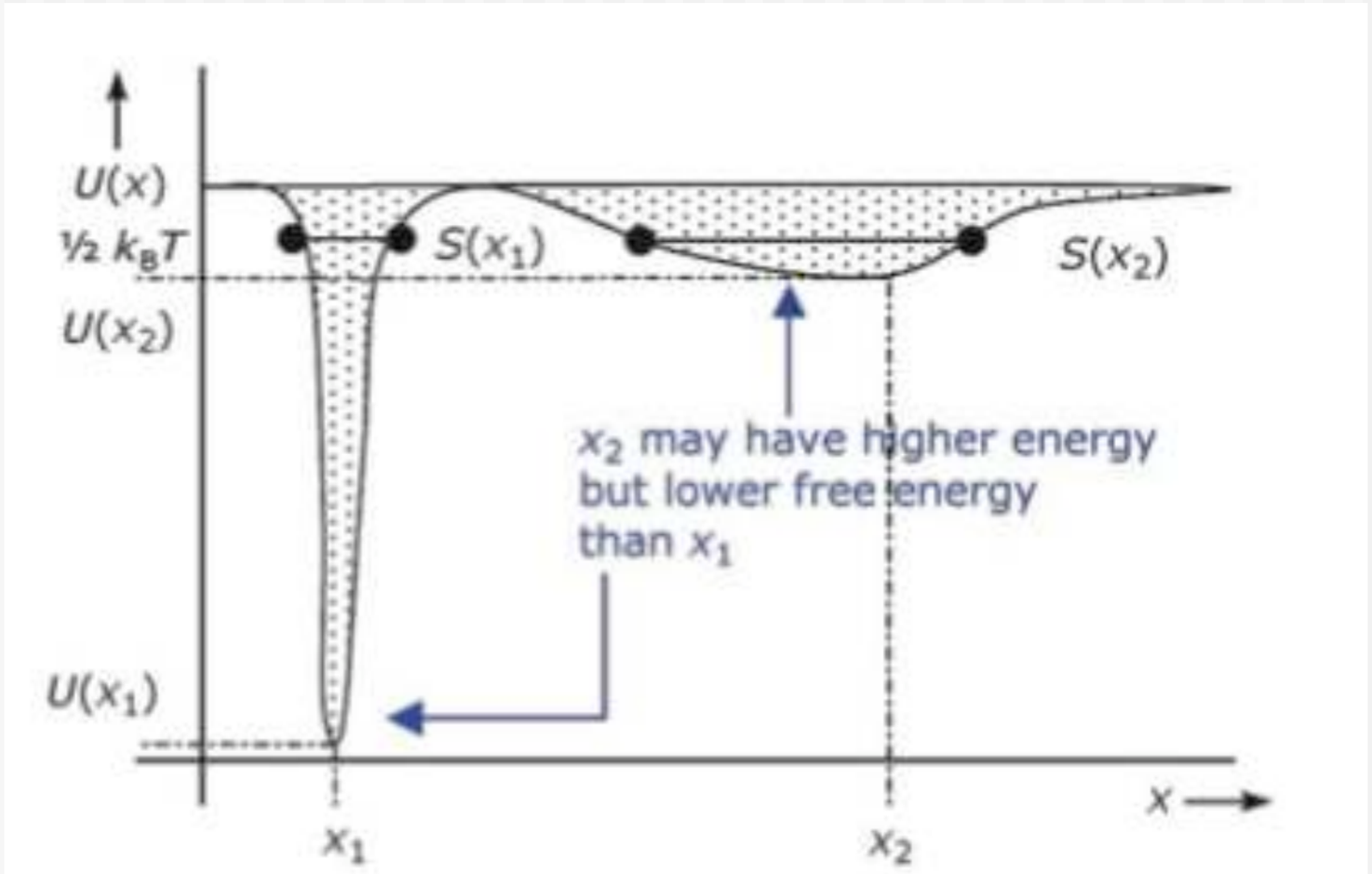
total energy➔ inner energy

E➔ U

B) entropy is often as important as accurate total enery E:

U➔ F

‘Problem’ of potential energy



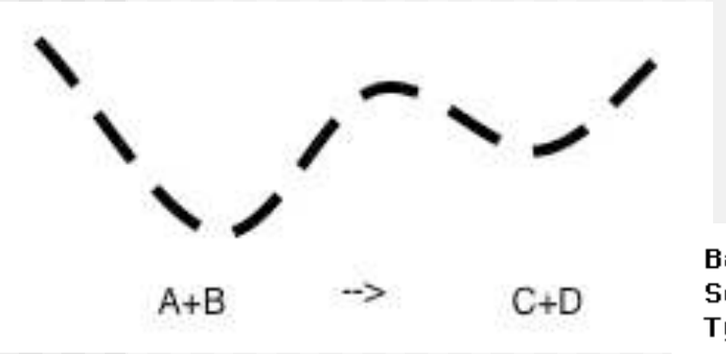
van Gunsteren AC 2006

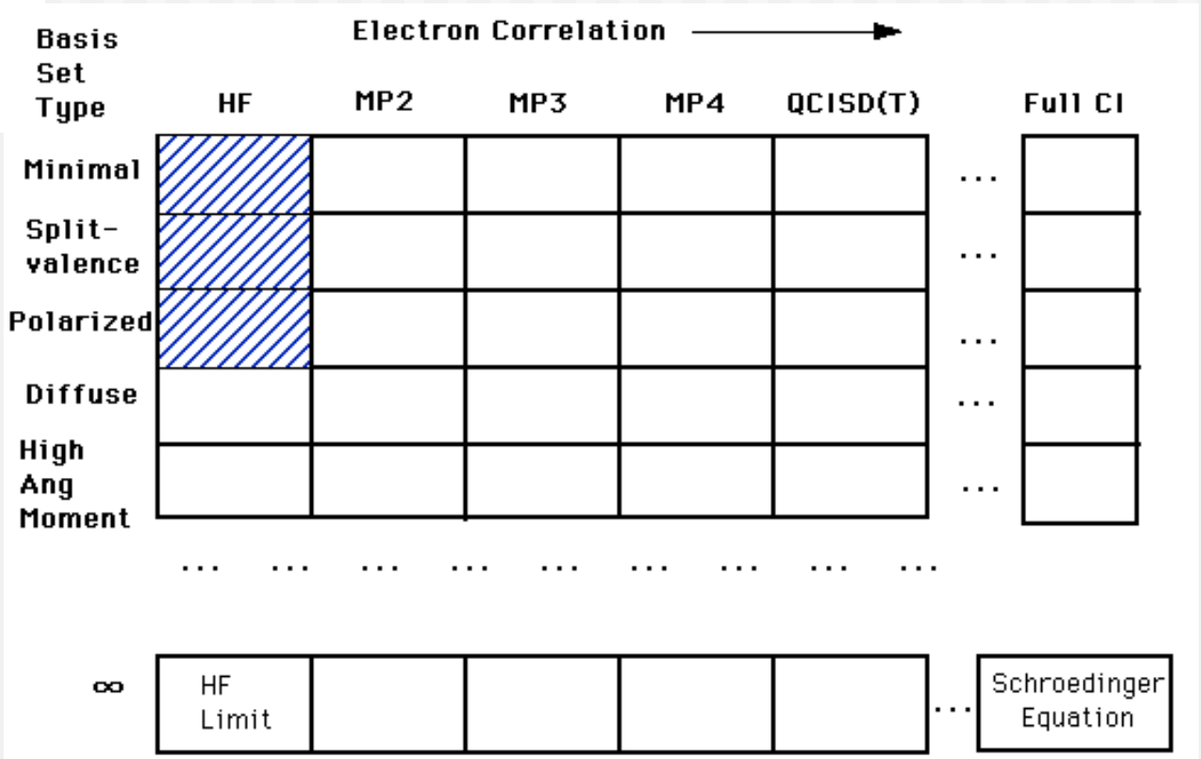
Two key problems

- include large part of system by treating some part at accurate QM level:
‘multiscale issue’
  - combine different methods
  - quantum chemistry problem: what QM level?

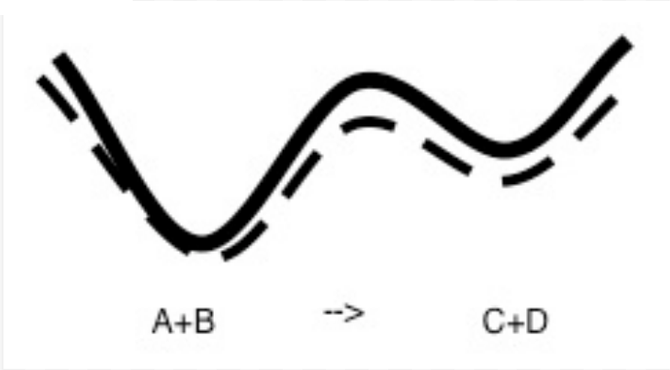
- find reaction pathway in complex environment, do the averaging and include entropic contributions
‘sampling issue’
(same as in MM MD)

Hirachy of methods in theoretical chemistry

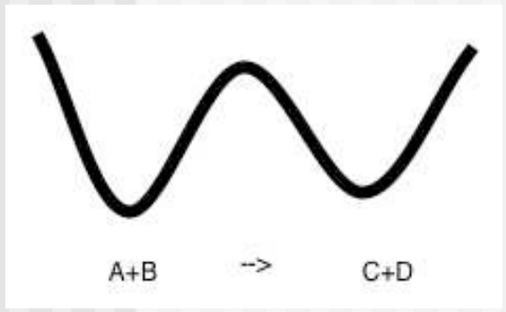




computational models and basis sets

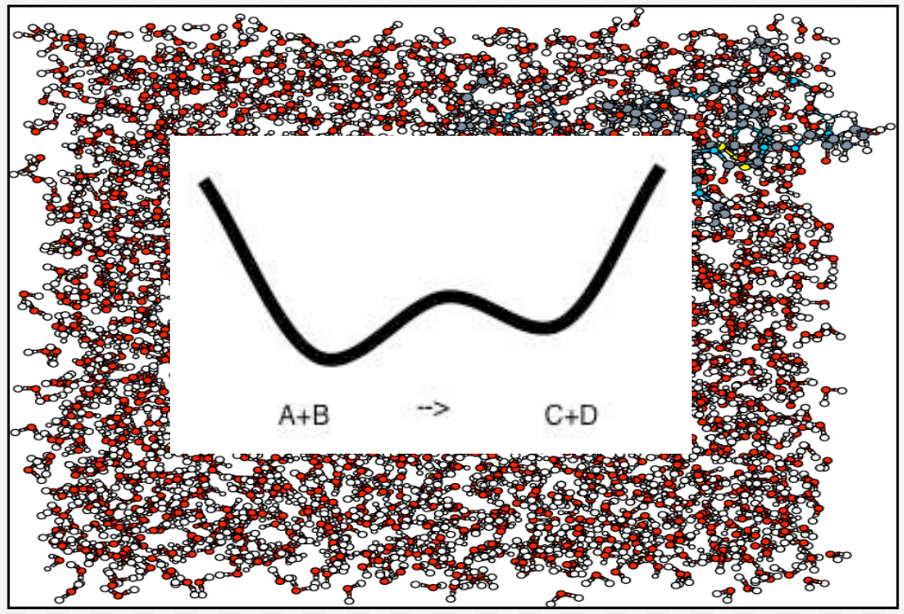


additional problems: environment and entropy

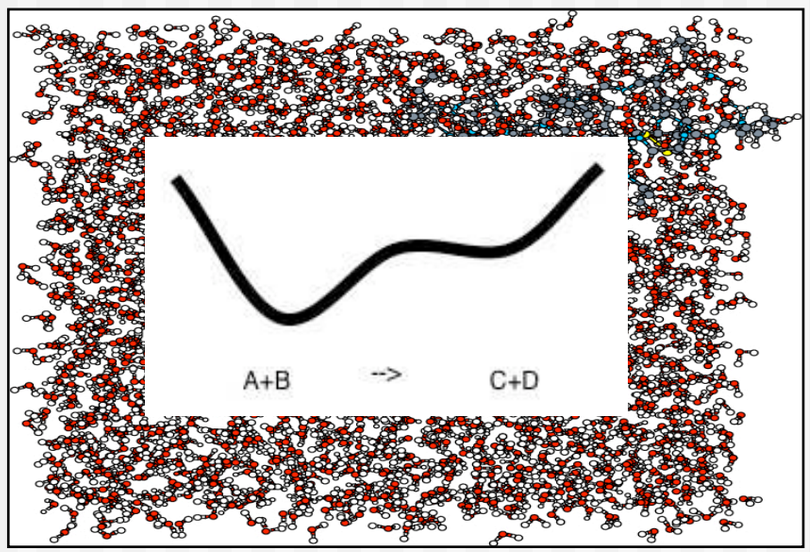


start:
QC in gas phase

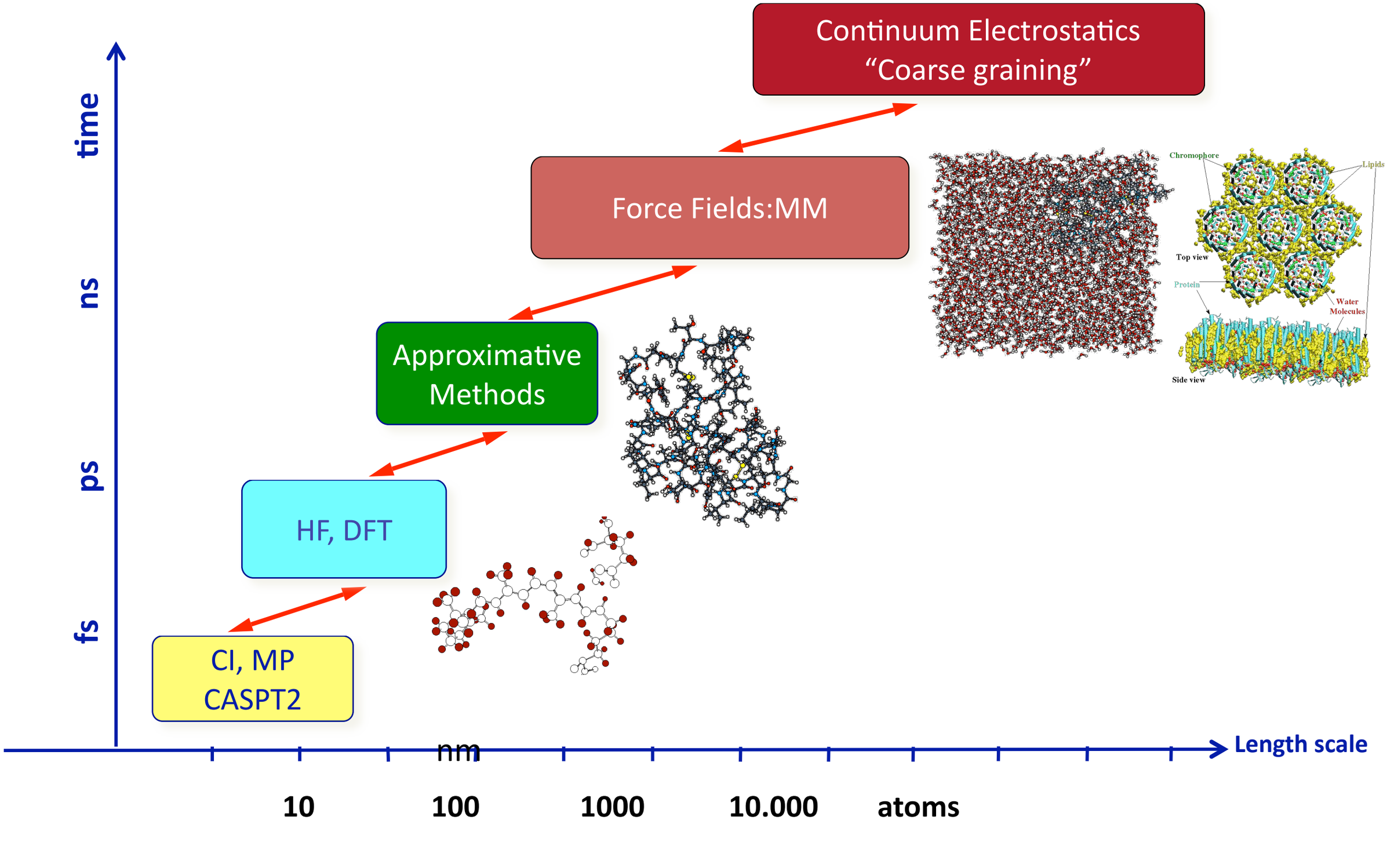
environment:
multiscale methods



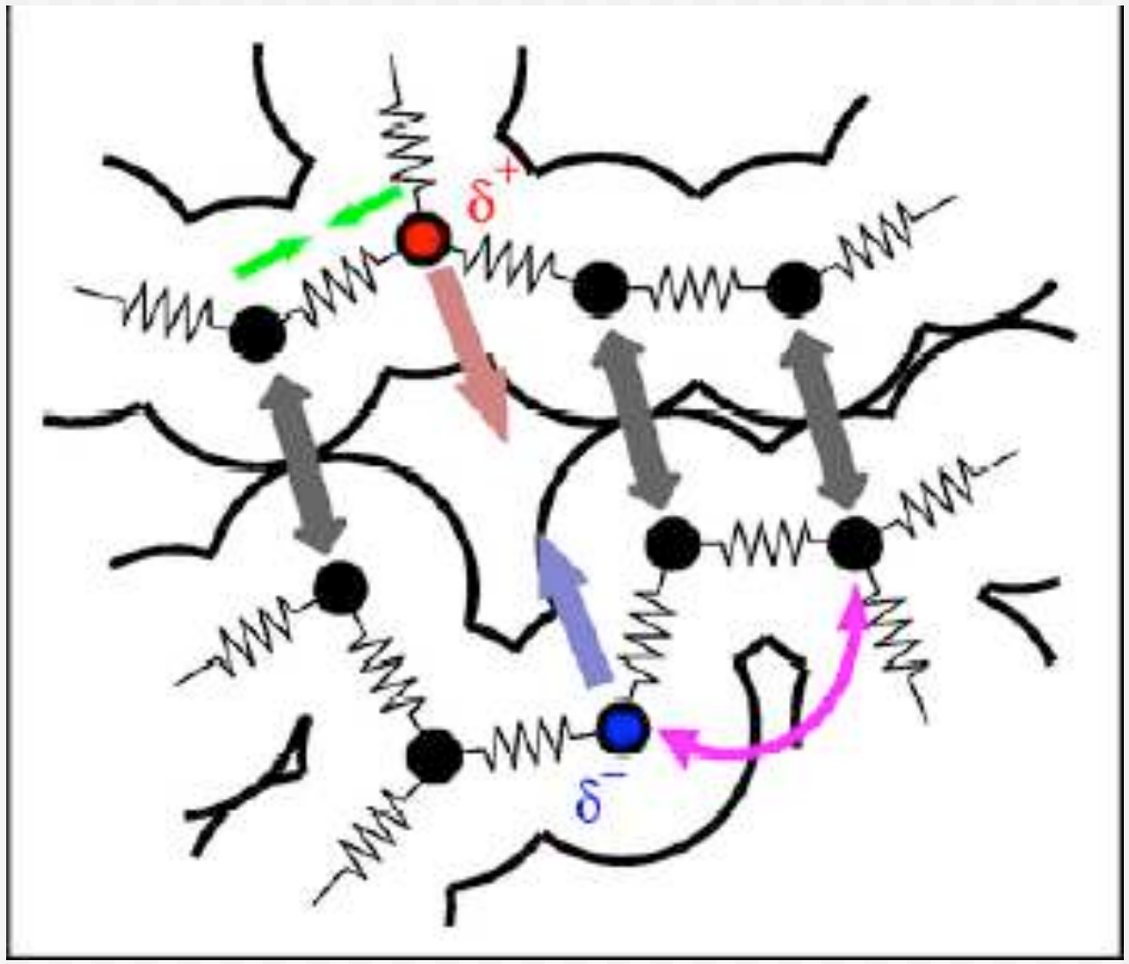
entropy:
'sampling'



Multi-scale models in theoretical biophysics

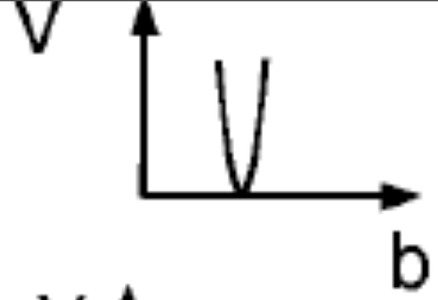


Empircal Force Fields: Molecular Mechanics

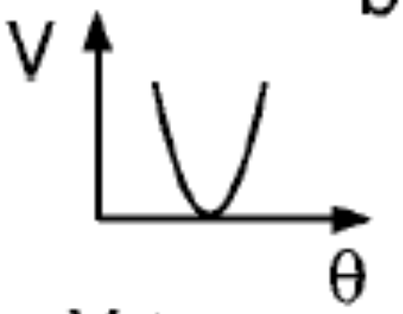


© Grubmüller

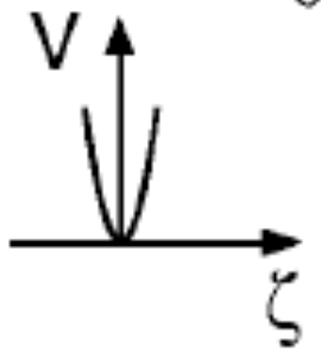
$$V_B = \sum_{\text{Bindungen}} \frac{1}{2} K_b (b - b_0)^2$$

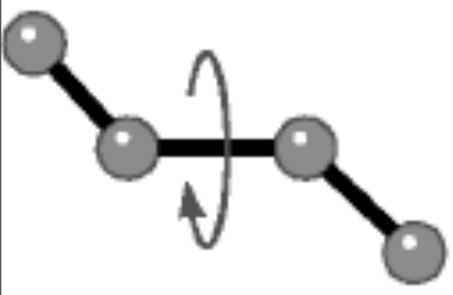


$$V_a = \sum_{\text{Winkel}} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

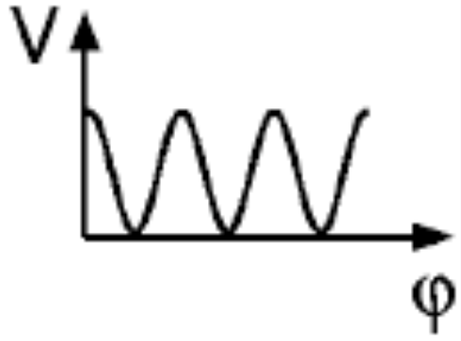


$$V_{imp} = \sum_{\text{Extraplanarwinkel}} \frac{1}{2} K_\zeta (\zeta - \zeta_0)^2$$

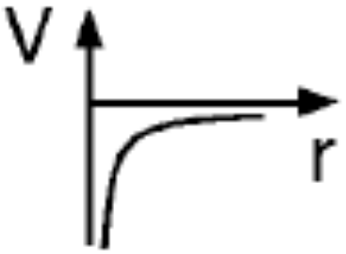




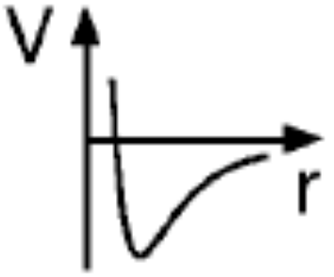
$$V_D = \sum_{\text{Dihedralwinkel}} K_\varphi [1 + \cos(n\varphi - \delta)]$$



$$V_q = \sum_{\text{Paare(i,j)}} q_i q_j / (4\pi\epsilon_0\epsilon_r r_{ij})$$



$$V_{vdW} = \sum_{\text{Paare(i,j)}} C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6$$



source: Grubmüller
MPI Göttingen

Molecular Mechanics (MM)

$$V = \sum_{bonds} k_b \left(b - b_0\right)^2 + \sum_{angles} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{dihedrals} \sum_{n=1}^{N} k_\phi^{(n)} \left[1 + \cos\left((n\phi - \delta\right)\right] + \sum_{impropers} k_\omega \left(\omega - \omega_0\right)^2$$

$$+ \sum_{i,j} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] + \sum_{i,j} \left(\frac{q_i q_j}{D r_{ij}} \right)$$

For Protein- and DNA ok!

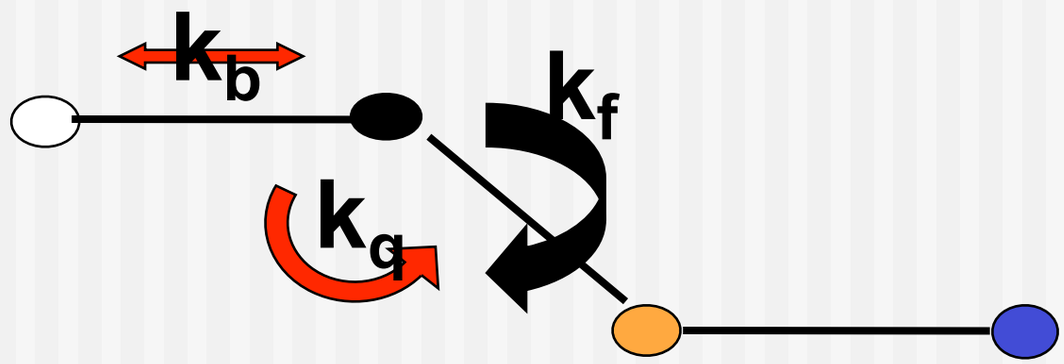
Problems.:

- **fixed charges:**

  no polarization

  no charge transfer

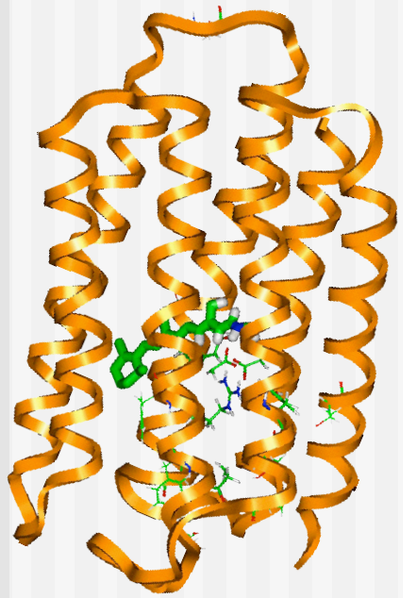
- **no reactions!**

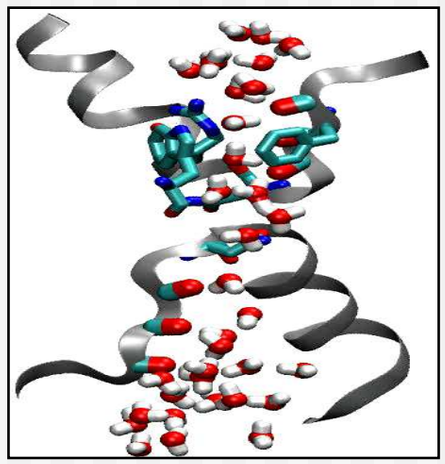


How to treat chemical reactions in proteins?

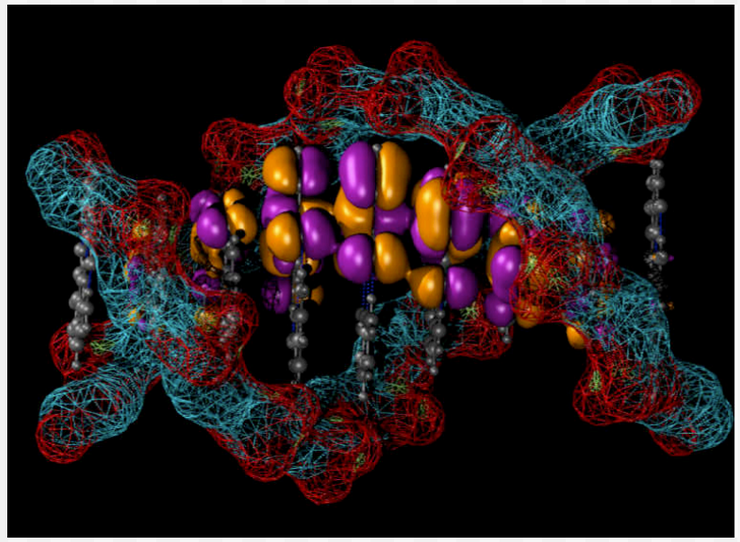
QM description needed

bioenergetics: proton transport

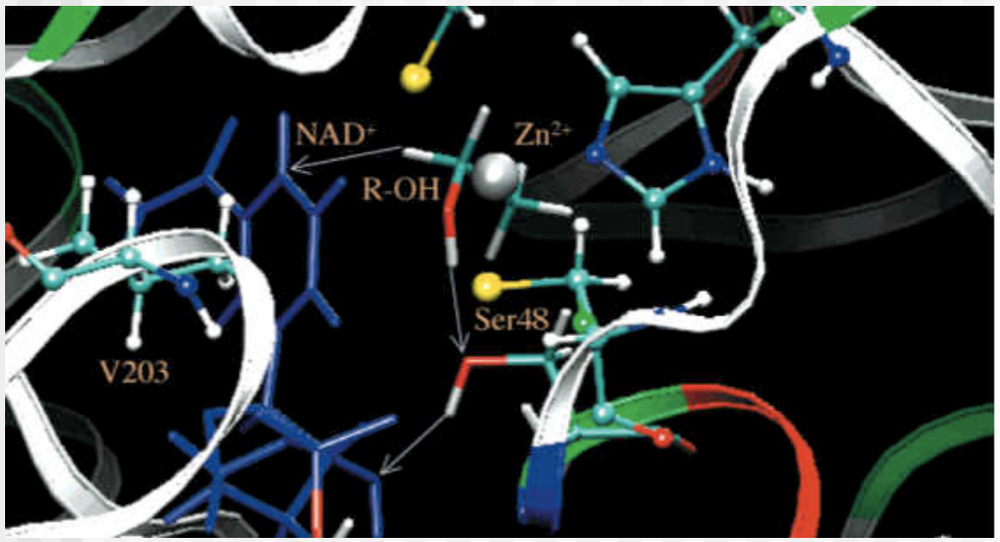




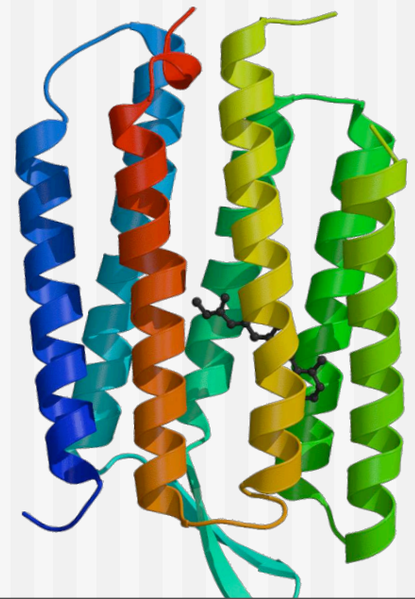
electron transport in DNA

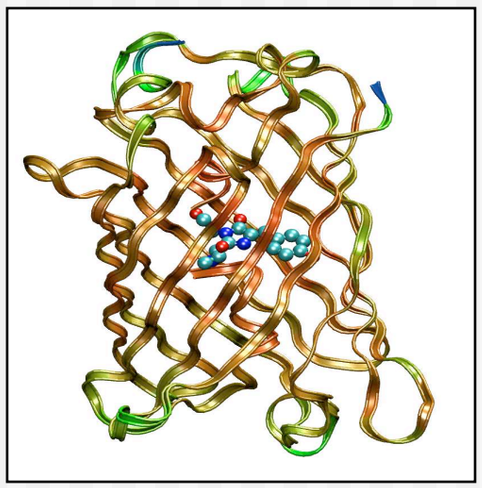


biocatalysis: alcohol dehydrogenase

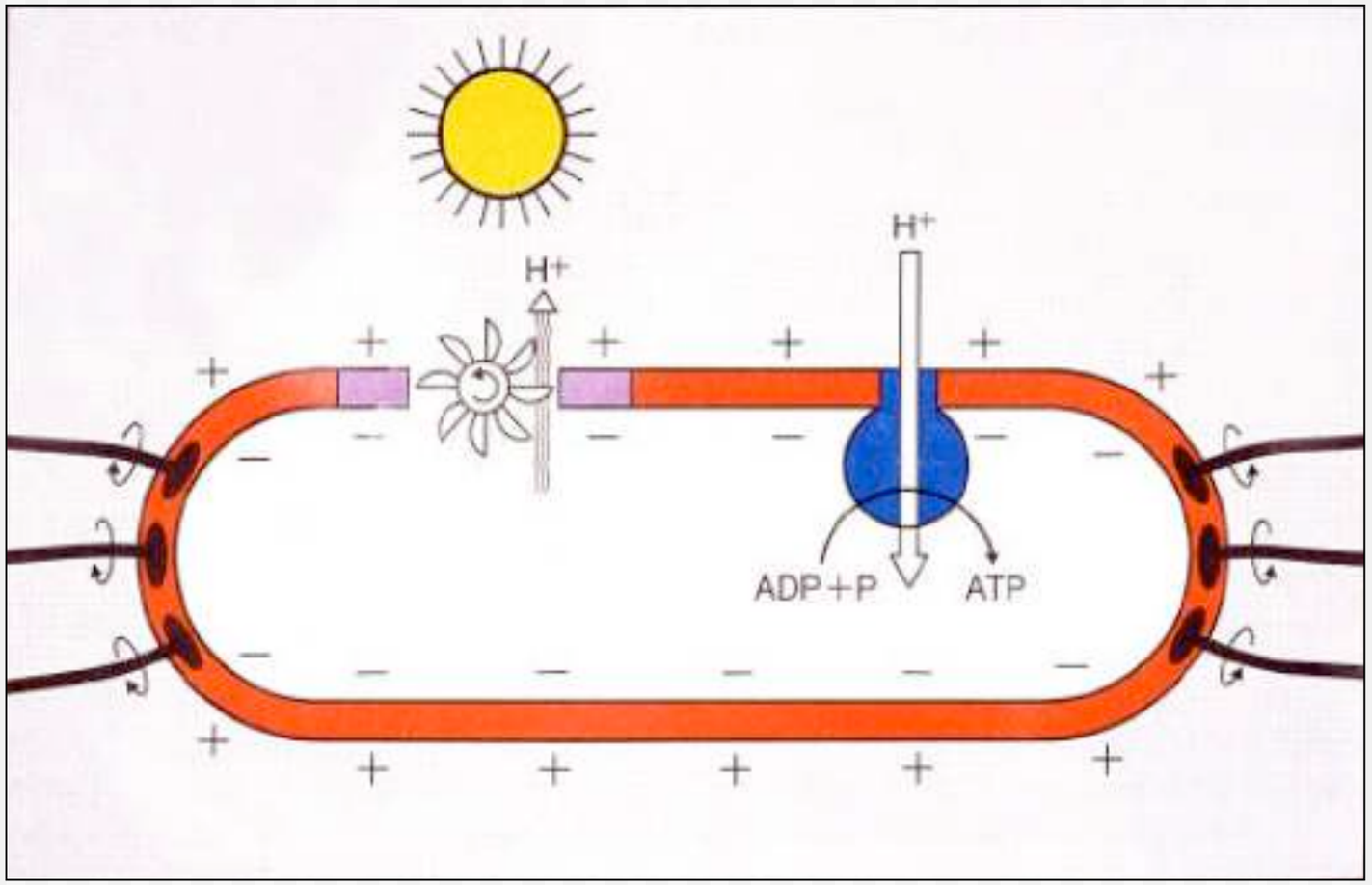


optical properties



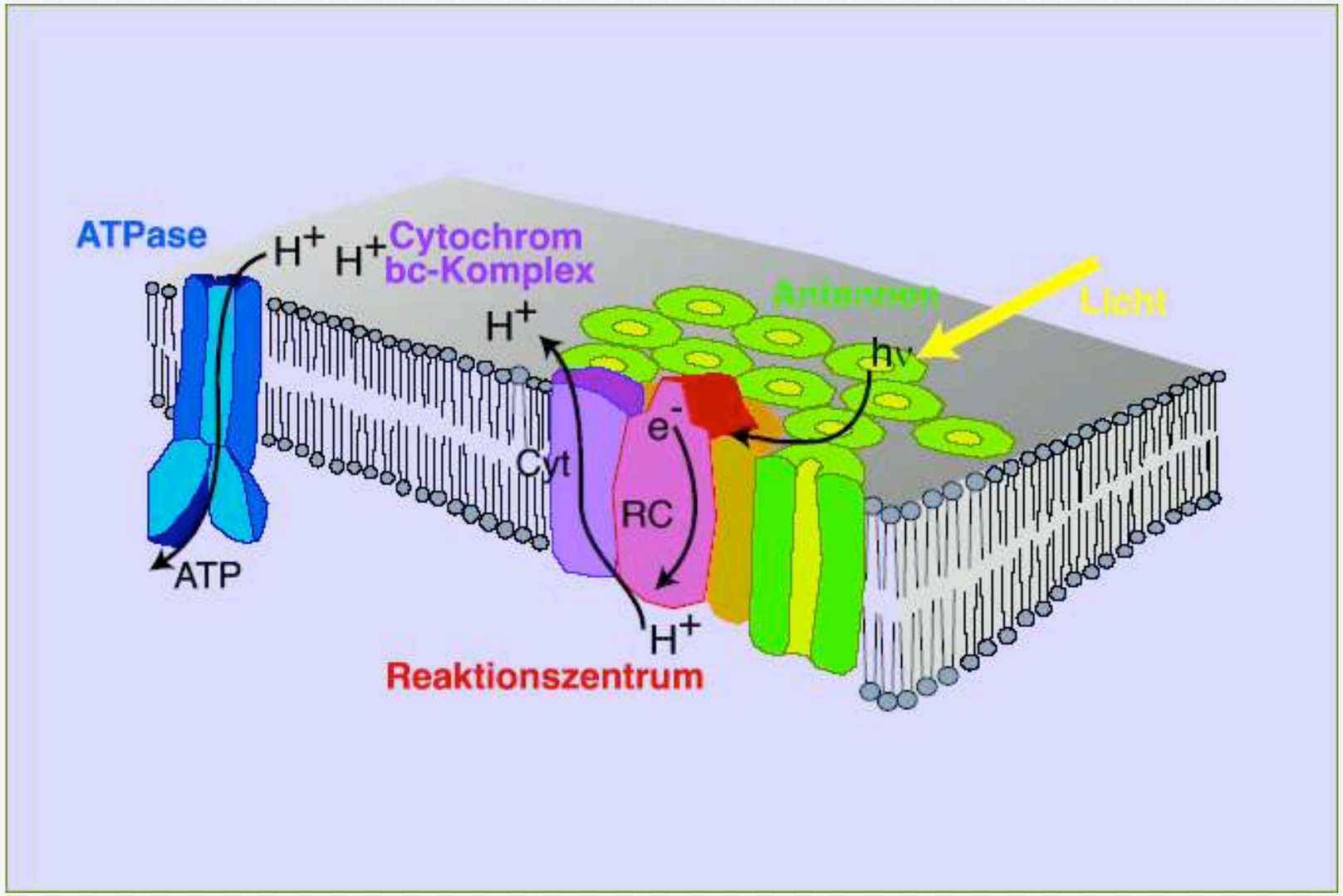


Bioenergetics: bacterial photosynthesis



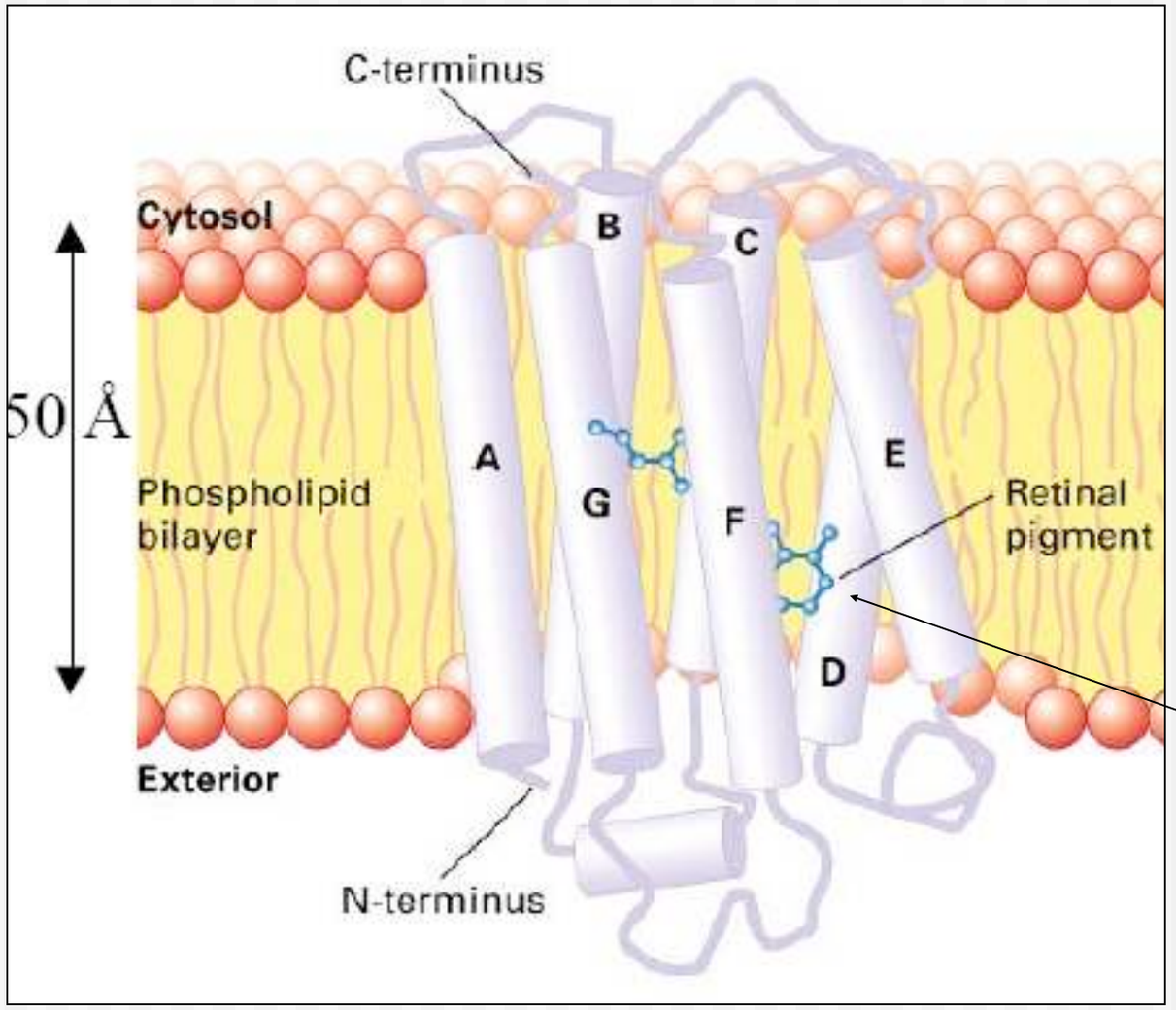
1) light absorption
2) proton transfer
3) ATP synthesis

Bacterial Reaction Center

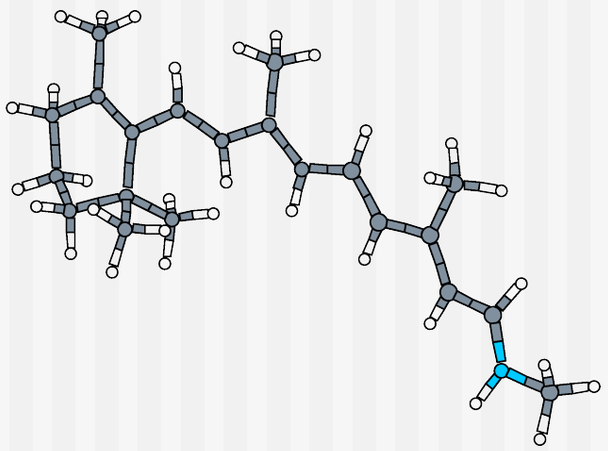


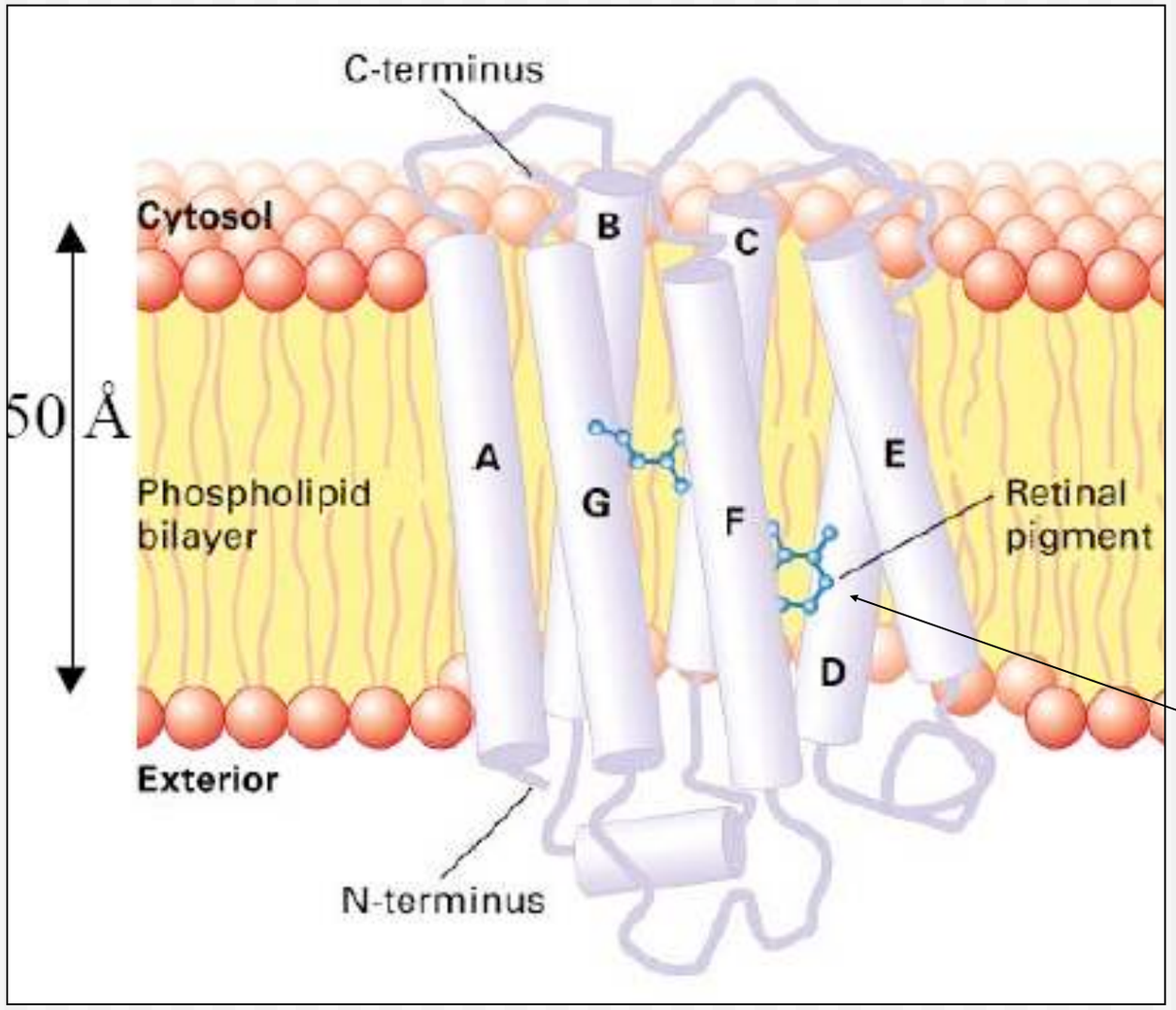
- photon absorption
- energy transfer
- electron transfer
- proton transfer
- Q_B movement:
  large structural transitions

Bacteriorhodopsin

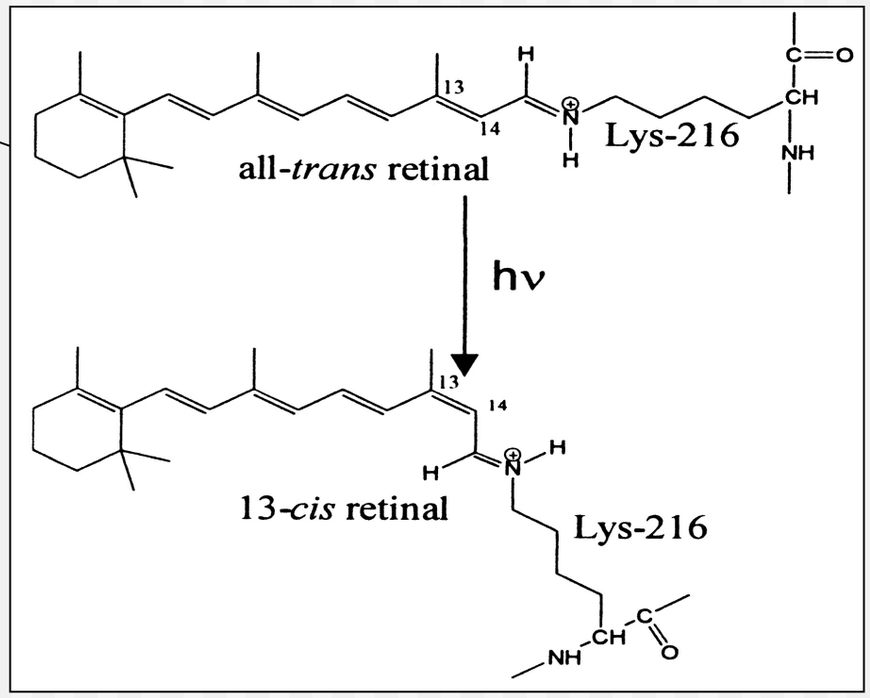


- Transmembrane protein
- 7 α-helices
- retinal chromophor

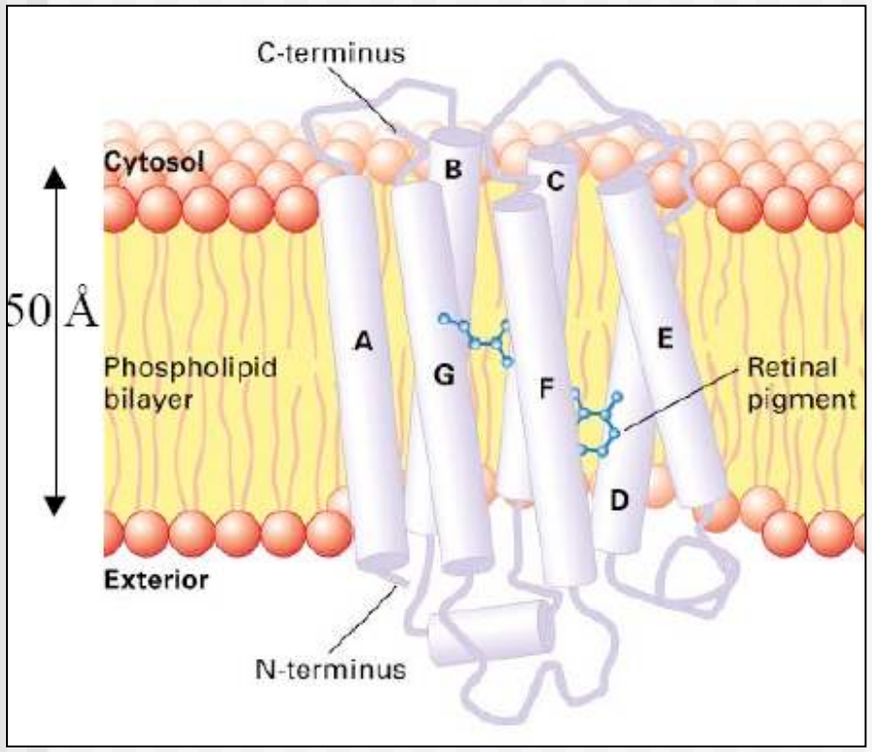


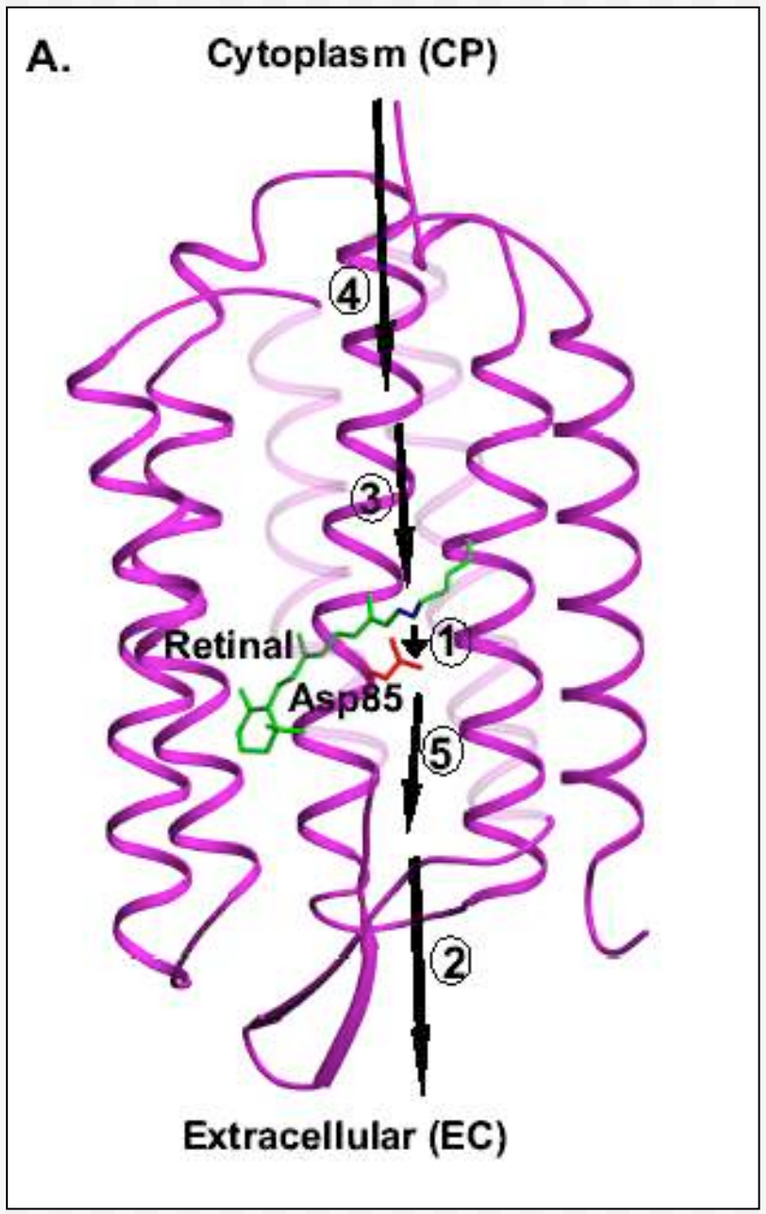


- Transmembrane protein
- 7 α-helices
- retinal chromophor



Bacteriorhodopsin

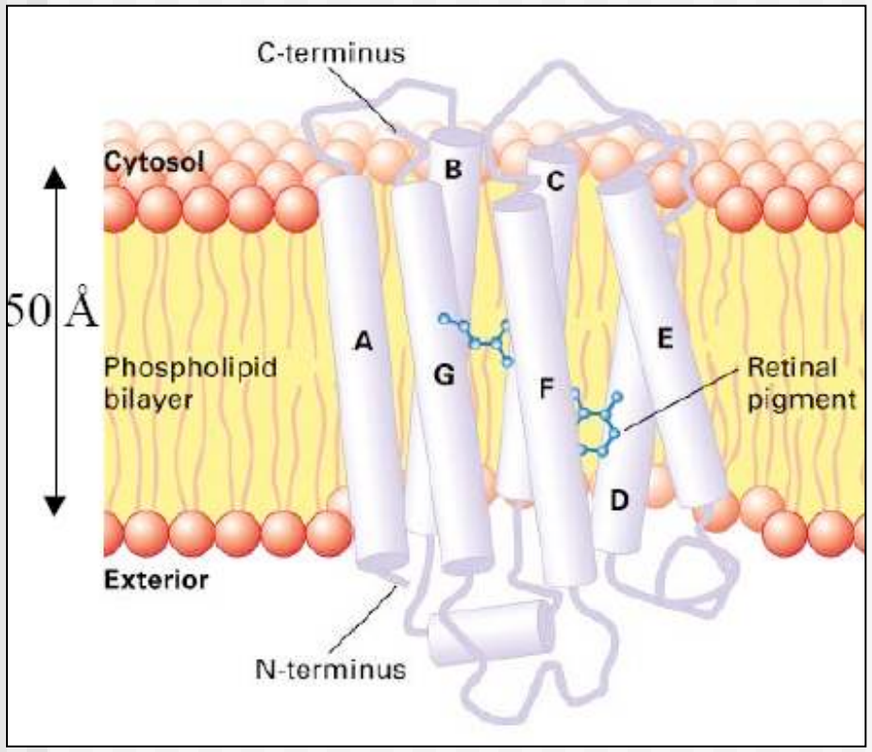


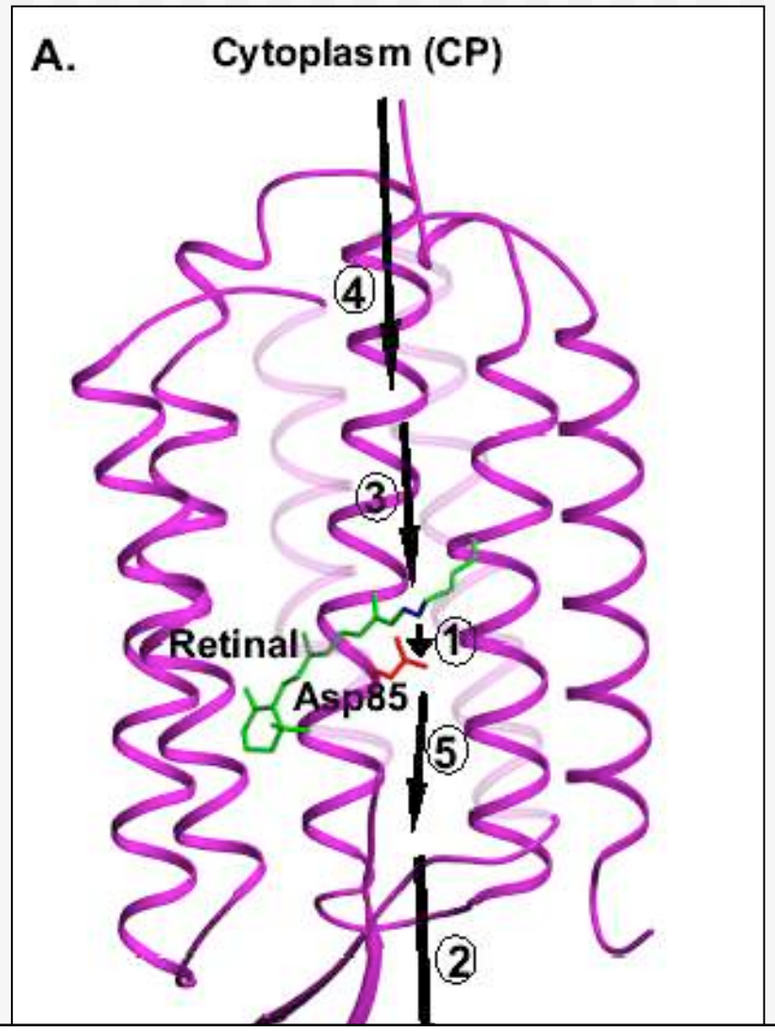


-pumps proton in 5 steps along photocycle

- structural information from x-ray
- IR/Raman/NMR spectra

However:
Complete bio-physical picture still missing





-pumps proton in 5 steps along photocycle

- structural information from x-ray
- IR/Raman/NMR spectra

However:
Complete bio-physical picture still missing

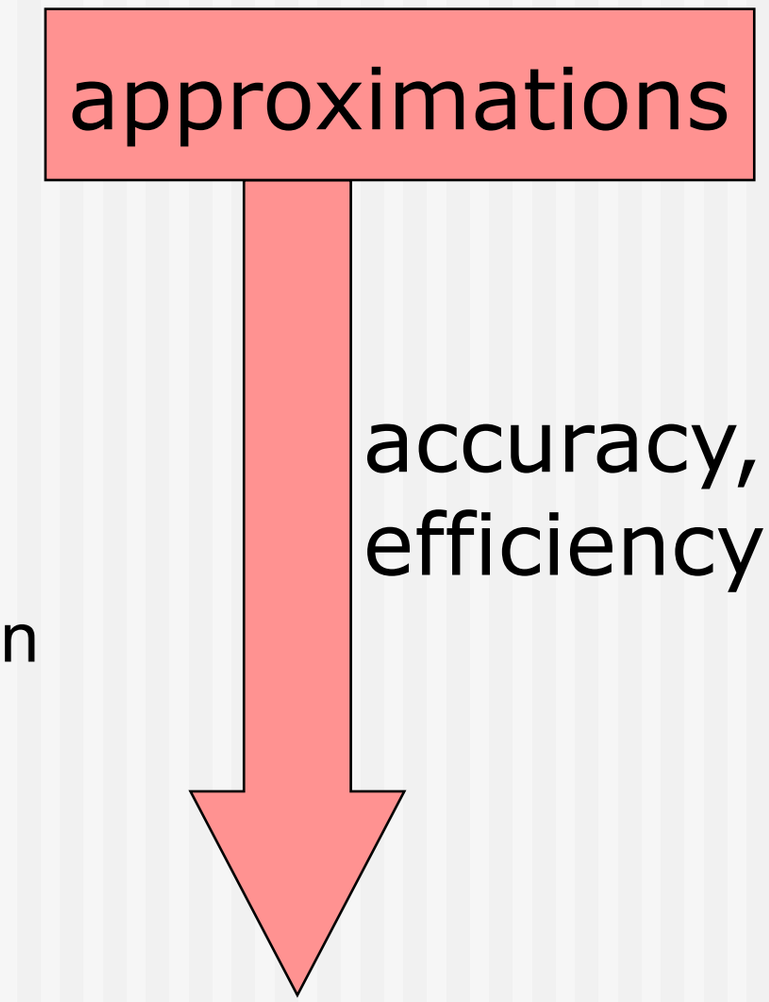
excited states, proton transfer: need QM

Quantum Chemistry (QC)

$$H\Psi = E\Psi$$

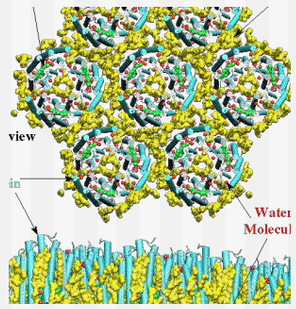
- Schrödinger equation 1926:
- Heitler and London 1927: H2
- Hund and Mulliken 1929: MO theory
- 1930 Hartree-Fock (HF)
- since 1950: use of computers
- 1965 Density Functional Theory (DFT)
- 1998 Nobel price for Chemistry: Pople & Kohn

- many developments:



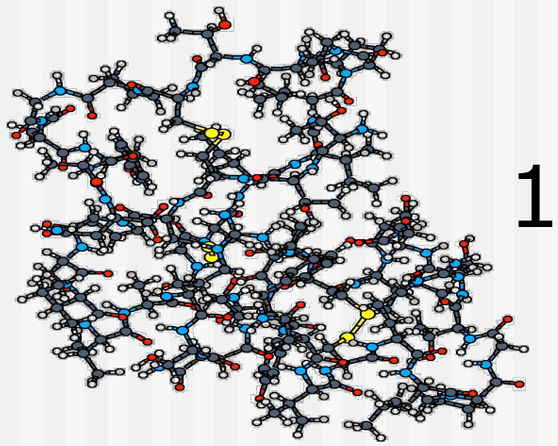
Methods

Molecular Mechanics (MM)



100k atoms

Semi-empirical methods



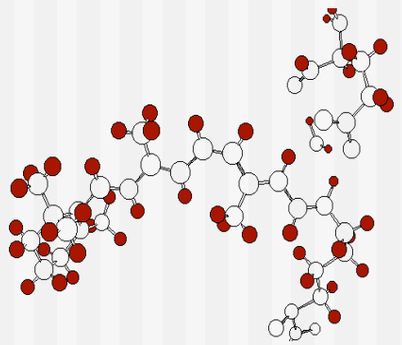
1000 atoms

DFT

Hartree-Fock (HF),

100 atoms

post-Hartree-Fock:
MP2,CC, CI, MRCI ...



10-50 atoms

speed

N-particle problem: wavefunction based methods

$$E = \left\langle \Psi_N \middle| \hat{H} \middle| \Psi_N \right\rangle$$

Problem: representation of N-electron wavefunction

- Hartree: $\Psi_H(r_1,\ldots r_N) = \phi(r_1)\phi(r_2)\ldots\phi(r_N)$
- Hartree-Fock $\Psi_{HF}(r_1,\ldots r_N) = \det[\phi(r_1)\phi(r_2)\ldots\phi(r_N)]$

single particle theories: effective one-electron Hamiltonian

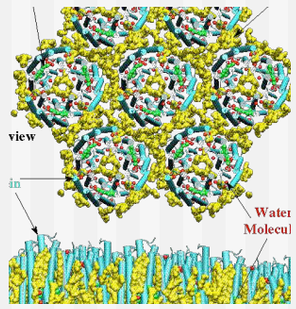
- Configuration interaction (CI): $\Psi_{CI}(r_1,\ldots r_N) = \sum_k C_k \Psi_S^k$
- perturbation theory (MP): $\left\langle \Psi_S \middle| r_{12}^{-1} \middle| \Psi_S \right\rangle$

Problem: electron correlation required!

Is this possible in the framework of effective one-electron theory?

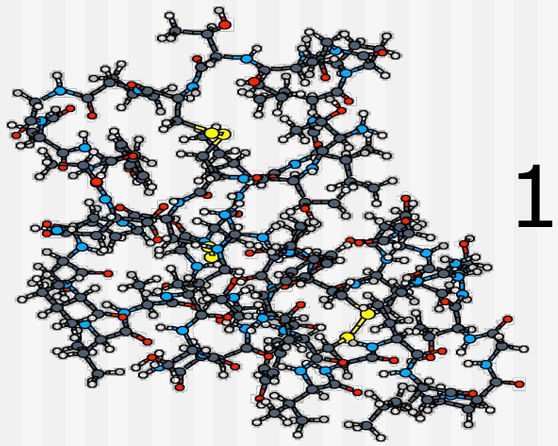
Methods

Molecular Mechanics (MM)



100k atoms

Semi-empirical methods



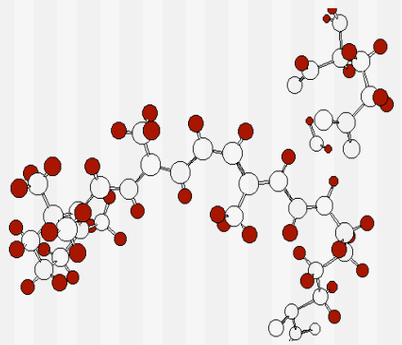
1000 atoms

DFT

Hartree-Fock (HF),

100 atoms

post-Hartree-Fock:
MP2,CC, CI, MRCI ...



10-50 atoms

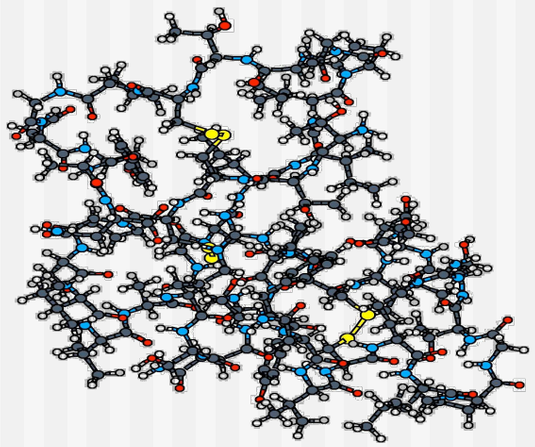
speed

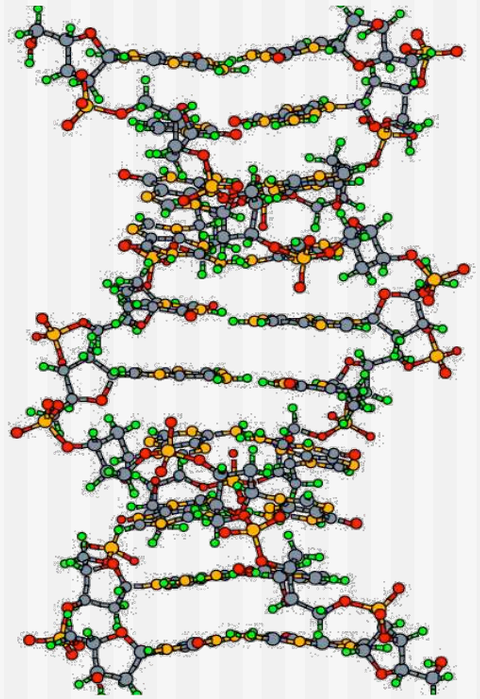
the big promise of DFT in the 1990's

- accounts for all important quantum effects: exact in principle

- faster than HF

=> apply for for large systems and long time-scales





Density Functional Theory (DFT)

consider : N electrons in the potential of nuclei α

$$\upsilon\left(\vec{r}\right) = -\sum_{\alpha}\frac{Z_{\alpha}}{\left|\vec{R}_{\alpha}-\vec{r}\right|}$$

Hohenberg & Kohn (1965)

- $\upsilon\left(\vec{r}\right) \leftrightarrow \rho\left(\vec{r}\right)$, **$\upsilon$-representability of electron density ρ**
- $E_0 \le E\left[\tilde{\rho}\right]$, **variational principle**

$$E\left[\rho\right] = T\left[\rho\right] + \int \upsilon\left(\vec{r}\right)\rho\left(r\right)d\vec{r} + \frac{1}{2}\iint\frac{\rho\left(\vec{r}\right)\rho\left(\vec{r}'\right)}{\left|\vec{r}-\vec{r}\right|}d\vec{r}d\vec{r}' + E_{xc}\left[\rho\right] + E_K$$

LDA, GGA: approximations of $E_{xc}\left(\rho\right), \upsilon_{xc} = \frac{\delta E_{xc}}{\delta\rho}$

Density Functional Theory (DFT)

Kohn & Sham (1966):

non-interacting electron gas in effective potential $\nu_{eff}\,[\rho]$

$$\left[-\frac{1}{2}\nabla^2+\nu_{eff}\left[\rho\right]\right]\Psi_i=\varepsilon_i\Psi_i,\qquad \rho=\sum_i^{occ}\left|\Psi_i\right|^2$$

$$\upsilon_{eff}\left[\rho\right]=\upsilon\left(r\right)+\int\frac{\rho\left(r'\right)}{\left|r-r'\right|}dr'+\upsilon_{xc}\left[\rho\right]$$

~ 100 atoms, ~ ps MD

todays view on DFT

Still most important method and widely applied, however:

- too slow for many interesting problems:

100 atoms
10 ps

- too inaccurate for many interesting problems:

VdW interactions
electronic excited states
reaction energies (e.g. PT)
...

todays view on DFT

Still most important method and widely applied, however:

- too slow for many interesting problems:

100 atoms
10 ps

- too inaccurate for many interesting problems:

VdW interactions
electronic excited states
reaction energies (e.g. PT)
...

to model the variety of biological processes, one needs the WHOLE toolbox of QC, i.e.

faster AND more accurate methods

Methods

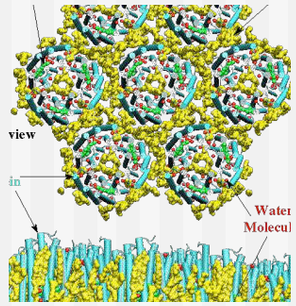
accuracy

empirical force fields

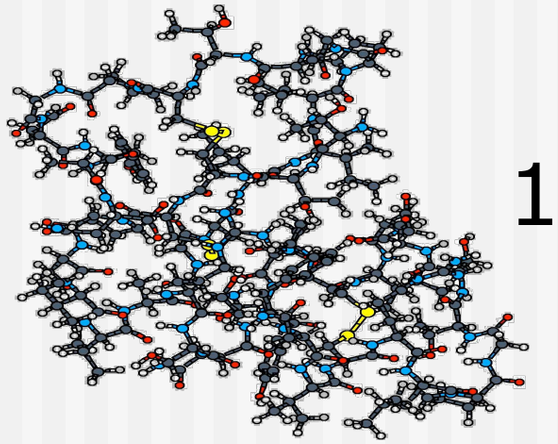
Semi-empirical methods
(DFTB, MNDO etc.)

**Hartree-Fock (HF),
Density Functional
Theory (DFT)**

post-Hartree-Fock:
MP2,CC, CI, MRCI ...

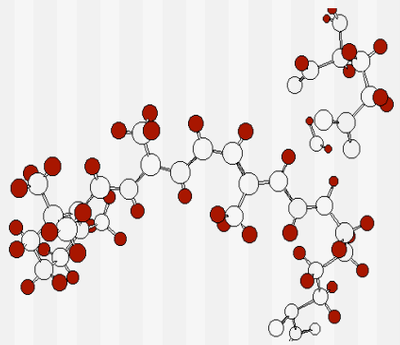


100k atoms



1000 atoms

100 atoms



10-50 atoms

speed

Methods in the QC toolbox

accuracy

Classical force fields
Molecular Mechanics (MM)

empirical potentials
fit to experiment

Semi-empirical methods

approximations
fit to exp. data

Hartree-Fock (HF),
Density Functional
Theory (DFT)

perturbation theory
'better wave-function'

post-Hartree-Fock:
MP2,CC, CI, MRCI ...

Semi-empirical /approximate methods

approximation, neglect and parametrization of interaction integrals from ab-initio and DFT methods

-HF-based:

CNDO, INDO, MNDO, AM1, PM3, MNDO/d, OM1,OM2

-DFT-based:

SCC-DFTB,

~ 1000 atoms, ~ ns MD

Approximate density-functional theory:SCC-DFTB

Self consistent - charge density functional tight-binding

www.dftb.org

- Seifert (1980-86): Int. J. Quant Chem., **58**, 185 (1996).
  O-LCAO; 2-center approximation: *approximate DFT*
  http://theory.chm.tu-dresden.de

- Frauenheim et al. (1995): Phys. Rev. B **51**, 12947 (1995).
  efficient parametrization scheme: *DFTB*
  www.bccms.uni-bremen.de

- Elstner et al. (1998): Phys. Rev. B **58**, 7260 (1998).
  charge self-consistency: *SCC-DFTB*
  www.tu-bs.de/pci

→ *approximate DFT*

SCC-DFTB

Second order expansion of the DFT total energy functional with respect to the charge density fluctuations $\delta\rho$ around a given reference density ρ_0 ($\rho_0' = \rho_0(\vec{r}')$, $\int' = \int d\vec{r}'$):

$$
\begin{aligned}
E &= \sum_i^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \int\int' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\, \delta\rho'} \right|_{n_0} \right) \Delta\rho\, \Delta\rho'. \\
&- \frac{1}{2} \int\int' \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0] n_0 + E_{cc}
\end{aligned}
$$

Reference density: $\rho_0 = \sum \rho_0^\alpha$
Superposition of atomic densities

SCC-DFTB total energy

$$E = \sum_{i}^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \iint' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\, \delta\rho'} \right|_{n_0} \right) \Delta\rho\, \Delta\rho'.$$

$$- \frac{1}{2} \iint' \frac{\rho'_0 \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0] n_0 + E_{cc}$$

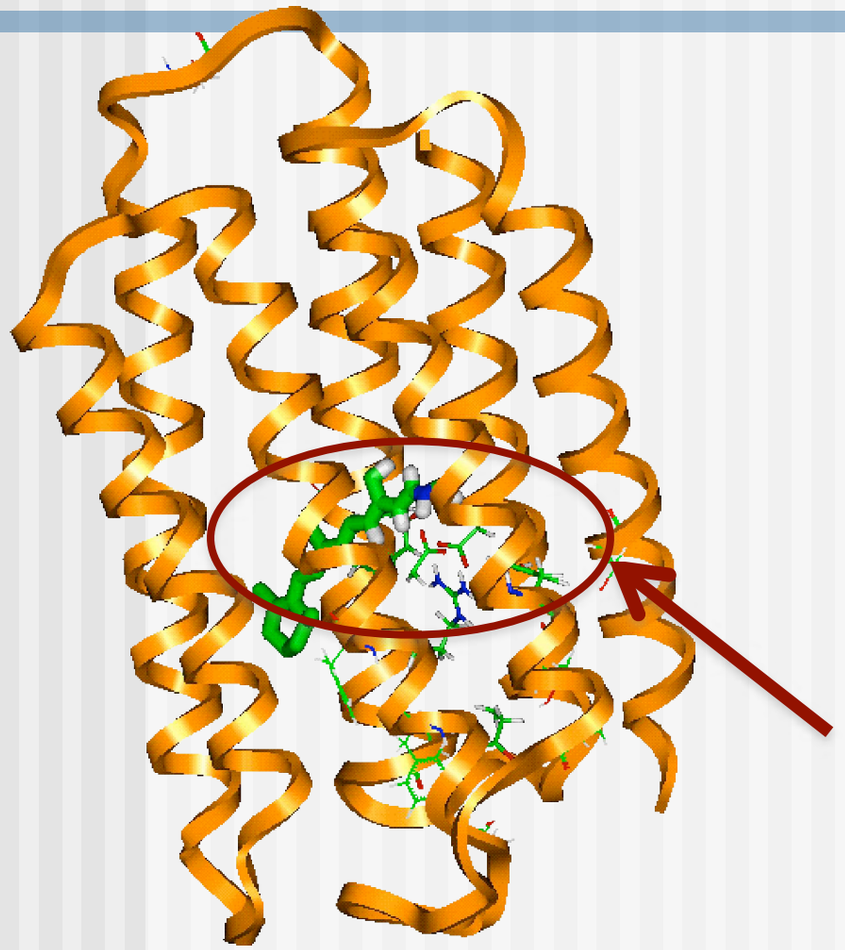
- minimal basis
- neglect of crystal field and three-center terms
- initial density fixed

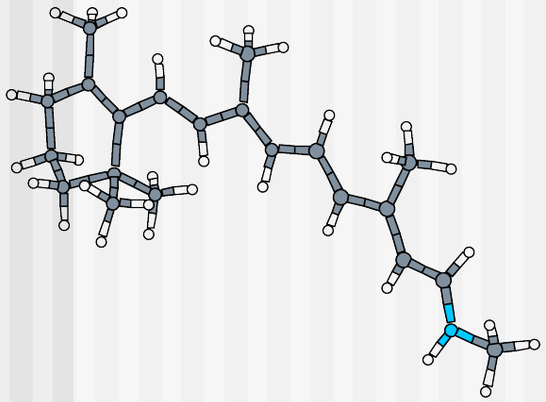
- second order expansion
- monopole approximation
- gamma

- two-body approximation
- fit procedure

$$E = \sum_i \sum_{\mu\nu} c^i_\mu c^i_\nu H^0_{\mu\nu} + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + \sum_{\alpha\beta} U_{\alpha\beta}$$

Computational problem

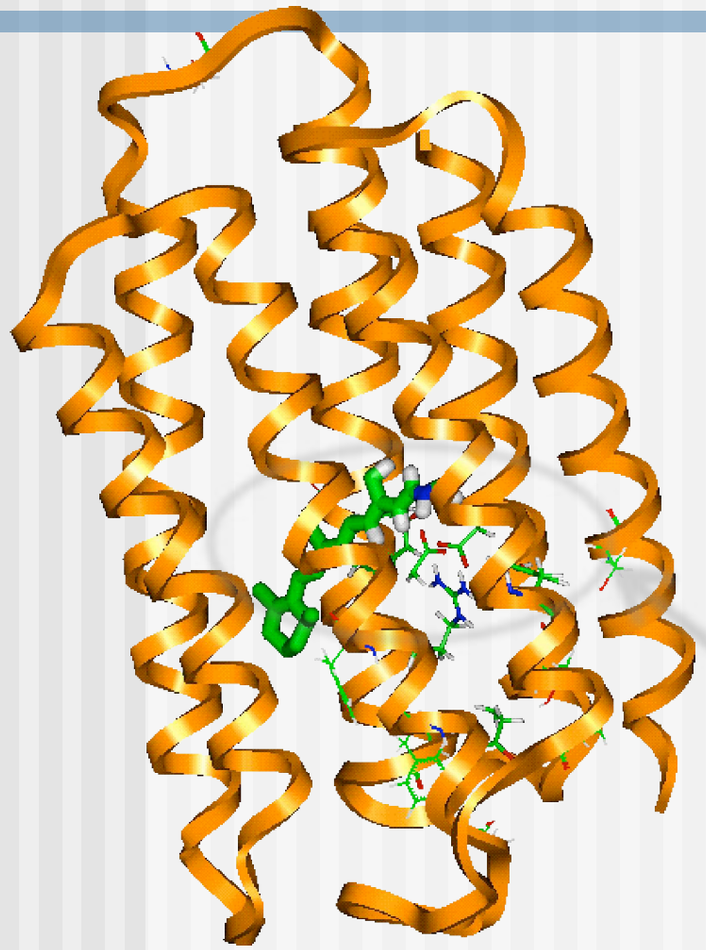


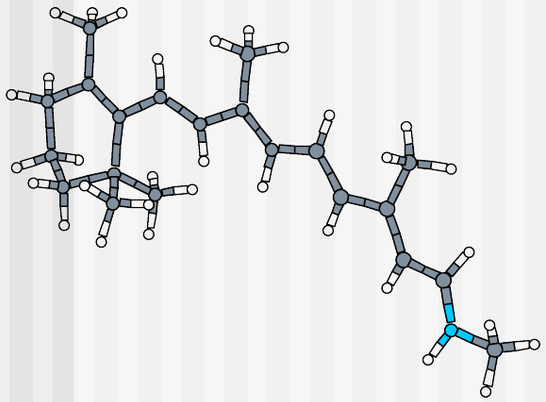


1) large systems: 1.000-100.000 atoms

2) need quantum chemical description

even DFT/DFTB much too slow





1) large systems: 1.000-100.000 atoms

2) need quantum chemical description

even DFT/DFTB much too slow

however: active site often localized

=> solution ‘of the early days’:

treat only part of the system with QM

‘speeding up QM’

Semi-empirical methods

Hartree-Fock (HF), Density Functional Theory (DFT)

integrals

solution of linear equations

post-Hartree-Fock: MP2,CC, CI, MRCI ...

‘speeding up QM’

Semi-empirical methods

Hartree-Fock (HF), Density Functional Theory (DFT)

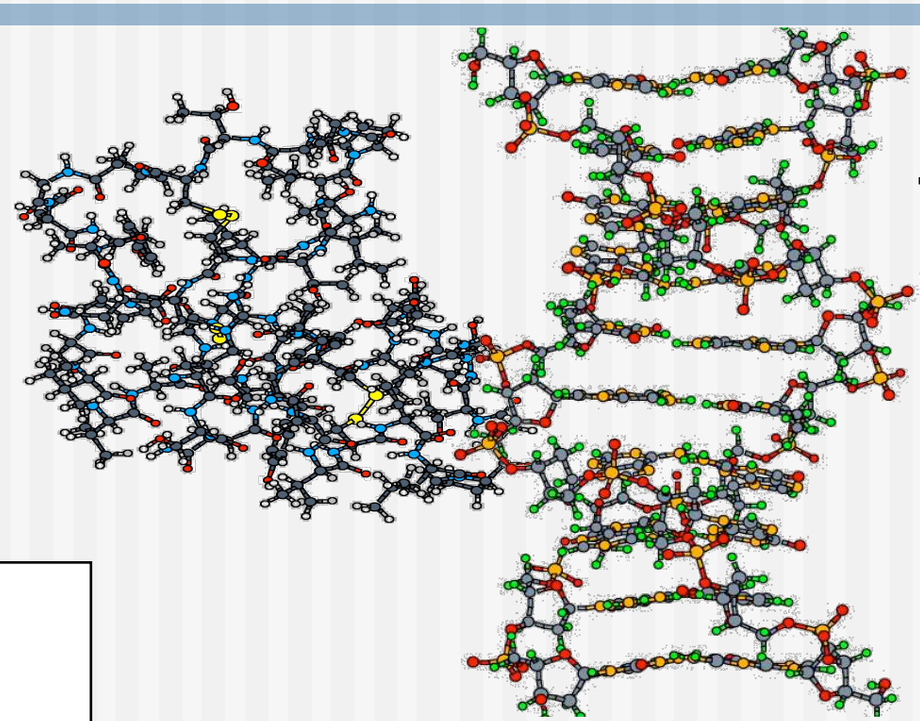
post-Hartree-Fock: MP2,CC, CI, MRCI ...

integrals → **integral approximations**

solution of linear equations → **linear-scaling**

parallelization

‘speeding up QM’



treatment of 1000 atoms with DFT/MP2 possible: e.g.

Siesta

Turbomole

…

Semi-empirical methods

Hartree-Fock (HF), Density Functional Theory (DFT)

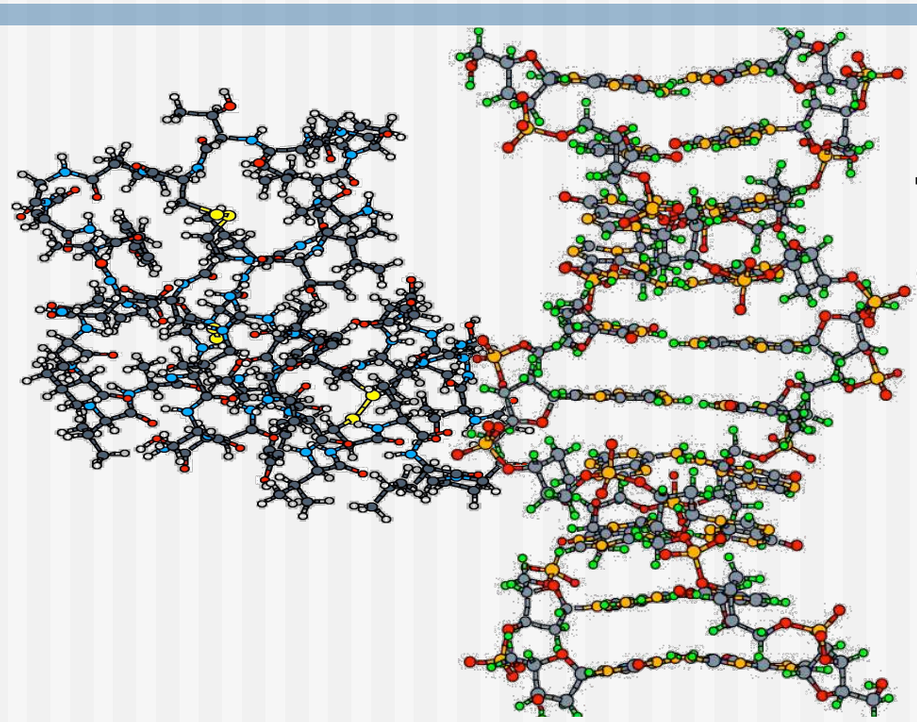
post-Hartree-Fock: MP2,CC, CI, MRCI …

integrals → **integral approximations**

solution of linear equations → **linear-scaling**

parallelization

‘speeding up QM’



treatment of 1000 atoms with DFT/MP2 possible: e.g.

Siesta

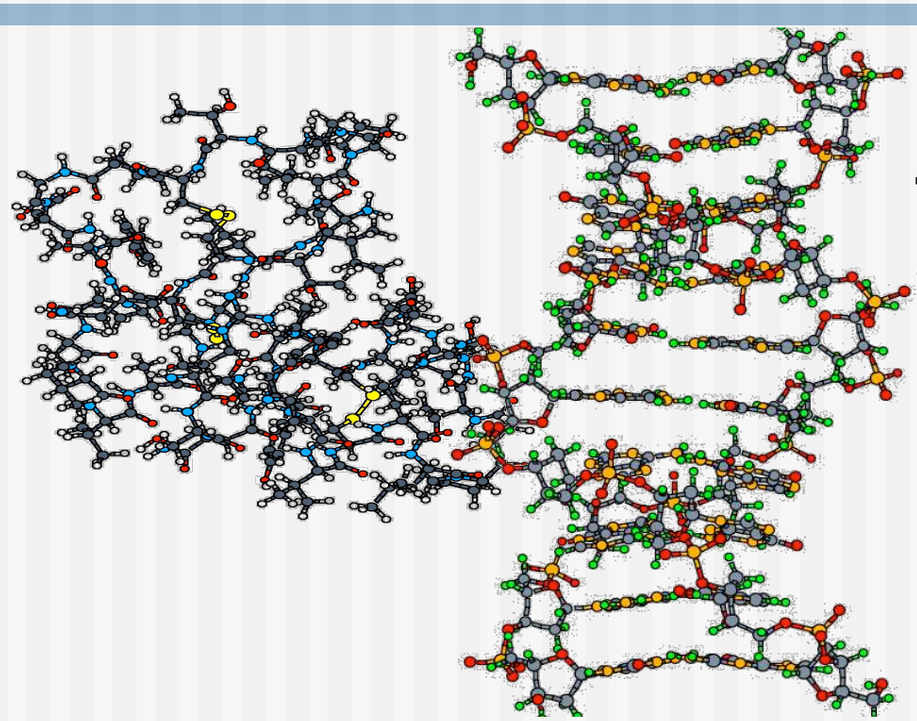
Turbomole

…

problem:

only ‘one’ (or few) structures

‘speeding up QM’



treatment of 1000 atoms with DFT/MP2 possible: e.g.

Siesta

Turbomole

…

problem:

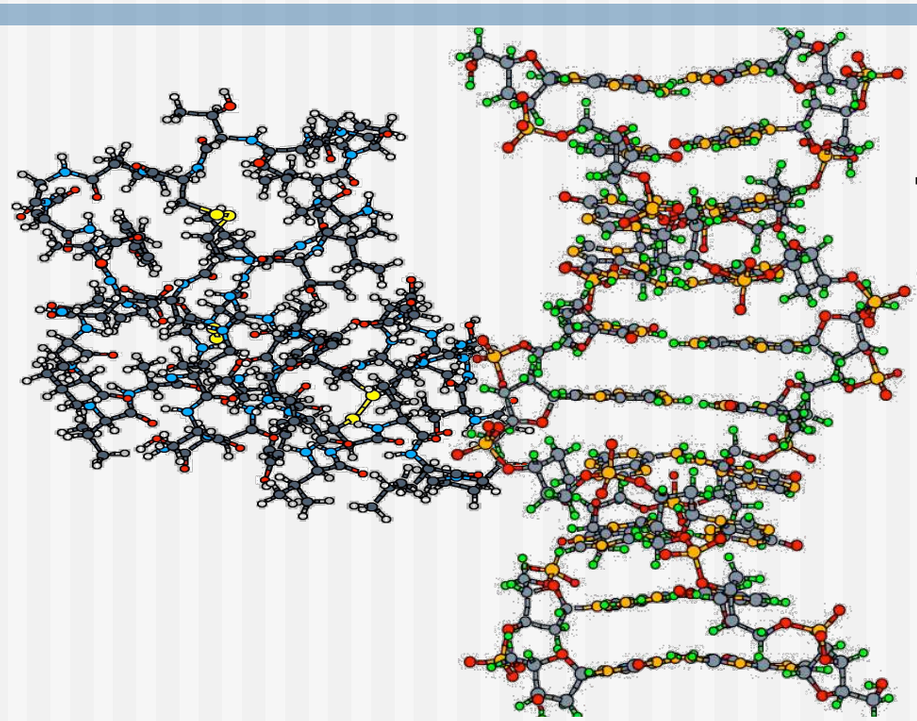
only ‘one’ (or few) structures

NEGLECTED:

- dynamics

- free energy vs potential energy

‘speeding up QM’



treatment of 1000 atoms with DFT/MP2 possible: e.g.

Siesta
Turbomole
…

problem:

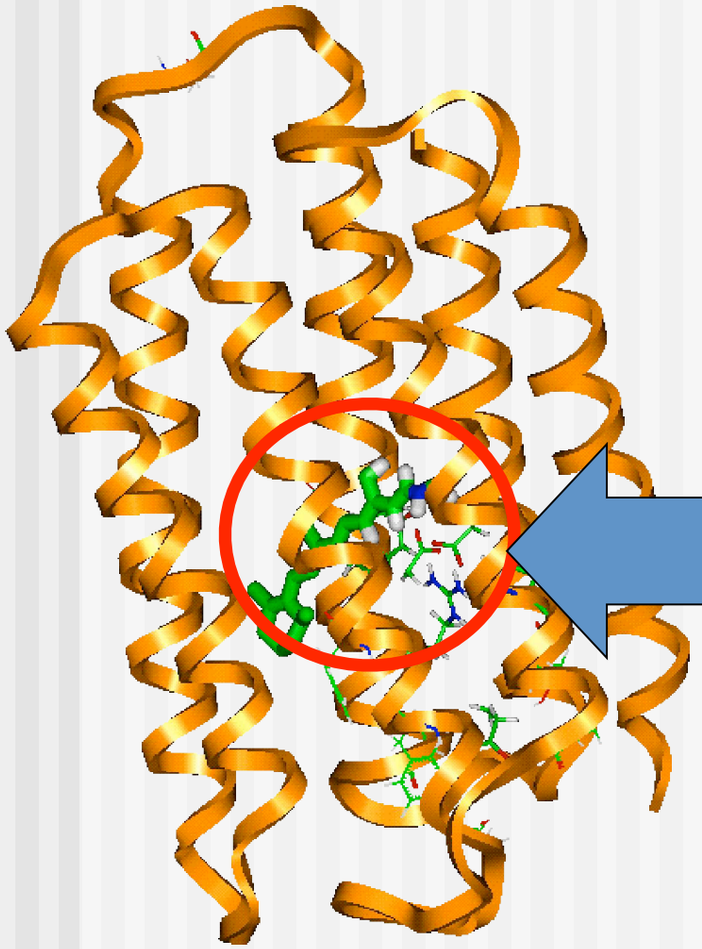
only ‘one’ (or few) structures

NEGLECTED:

- dynamics
- free energy vs potential energy

can be even more important than accurate total energy!

Combined QM/MM methods



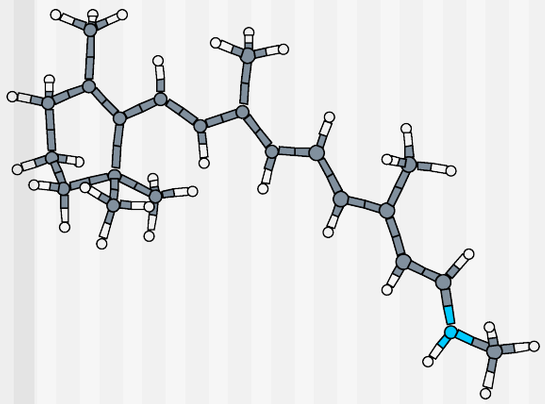
~ 1.000-100.000 atoms

~ ns MD simulations

(MD, umbrella sampling)

- chemical reactions

- excited states, spectroscopy



In many cases, the site of interest is localized

➔ apply QM locally

Recent review: Senn & Thiel, Top Curr Chem (2007) 268: 173

Combined QM/MM methods

1976 Warshel und Levitt

1986 Singh und Kollman

1990 Field, Bash und Karplus

QM

- semi-empirical methods
- quantum chemistry : DFT, HF, MP2, LMP2
- DFT ‘plane wave‘ codes: CPMD

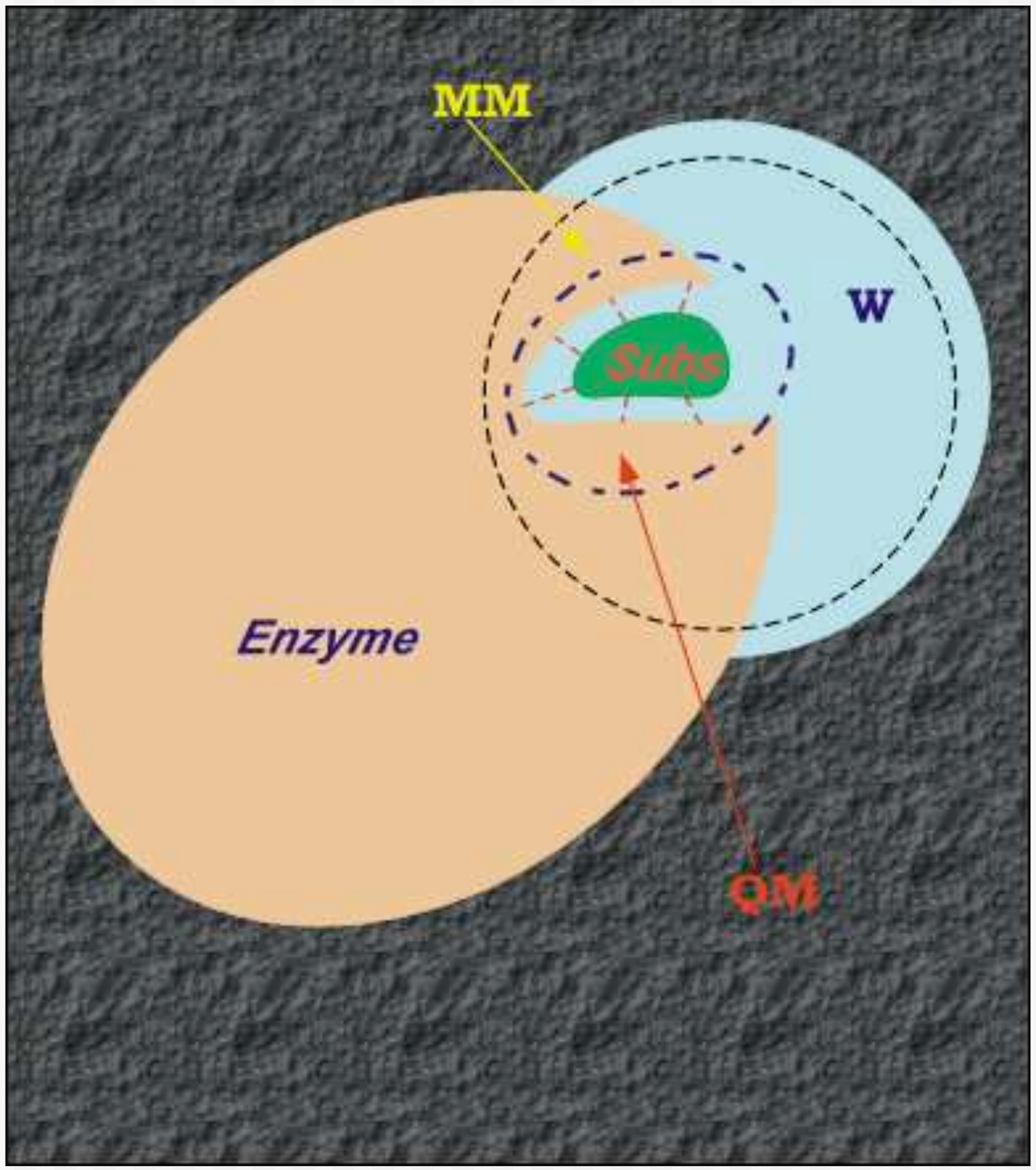
MM

- CHARMM, AMBER, GROMOS, SIGMA,TINKER, ...

Recent review: Senn & Thiel, Top Curr Chem (2007) 268: 173

Combined QM-MM methods

$\varepsilon = 80$



-QM region

- Molecular Mechanics (MM) region

Effects:

- **steric interactions**: keep the active site in place:

- **electrostatic interaction**: polarization of QM region due to MM

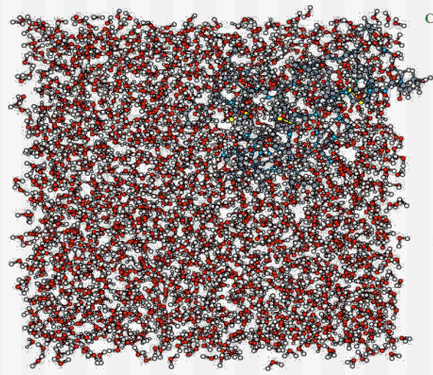
Recent review: Senn & Thiel, Top Curr Chem (2007) 268: 173

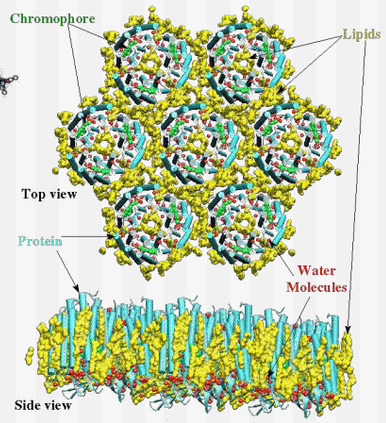
Spectrum of methods

Size and simulation time lime each other

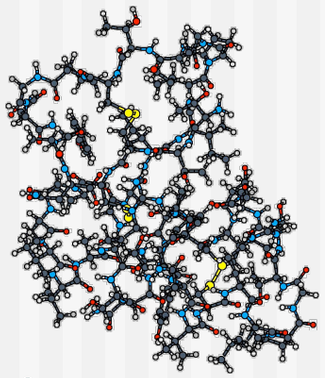
contiuum methods
"Coarse graining"

MM

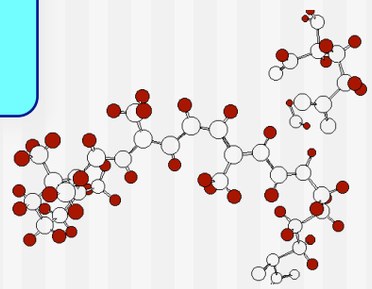




SE



HF, DFT



CI, MP
CASPT2

ns

ps

fs

time

nm

10

100

1000

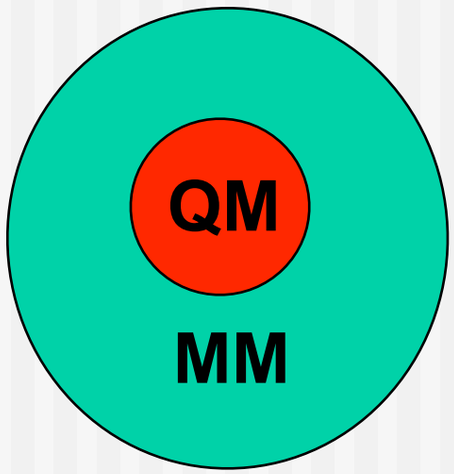
10.000

atoms

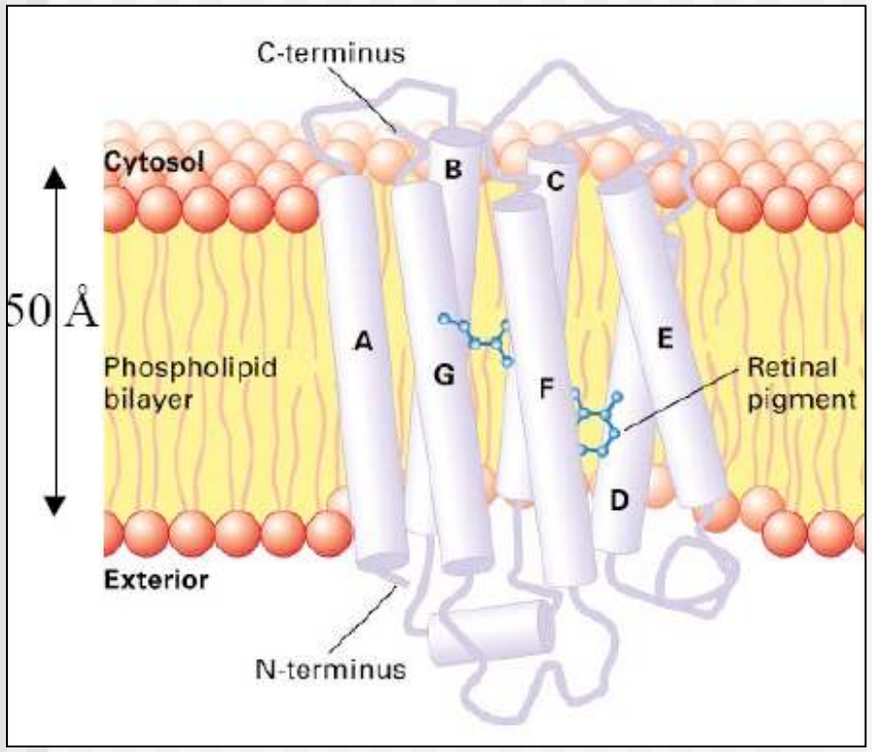
Length
scale

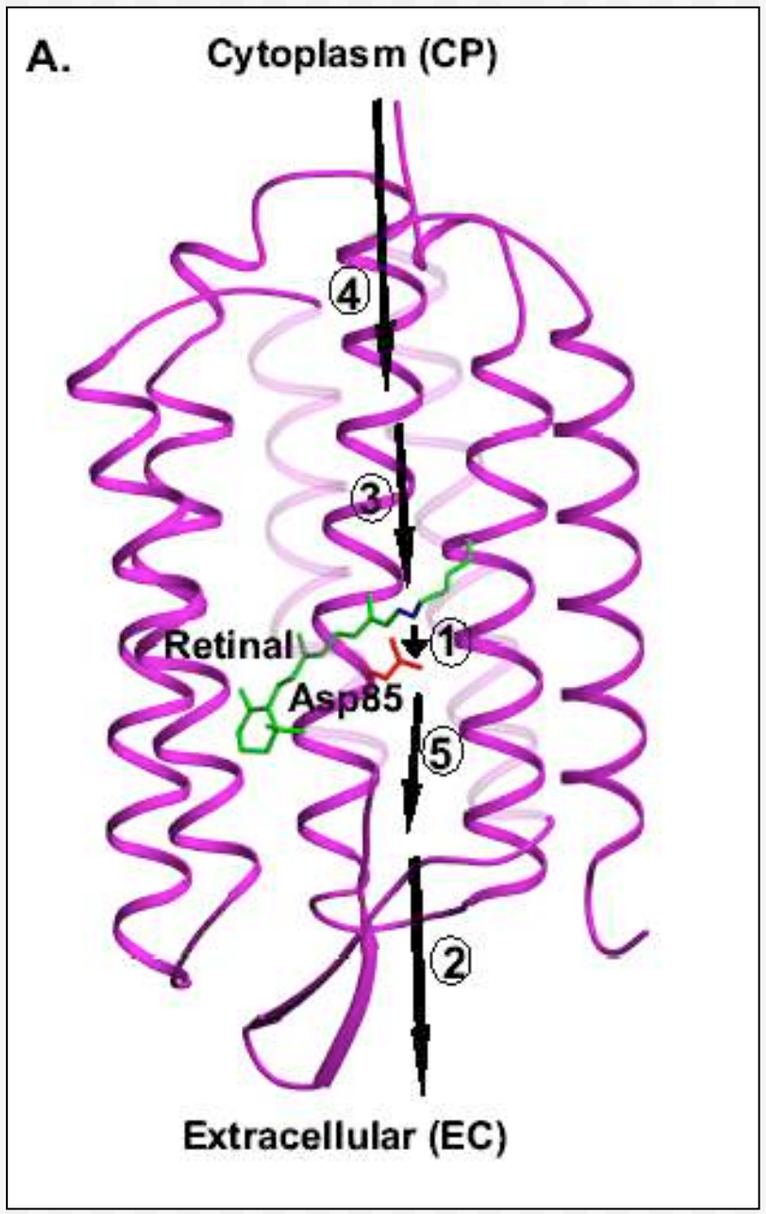
Main distinction between QM/MM methods

- additive vs. subtractive methods
- embedding: mechanic, electrostatic or polarizable
- treatment of the boundary:
  - link atom, pseudo atom, hybrid orbitals
  - electrostatics



Bacteriorhodopsin



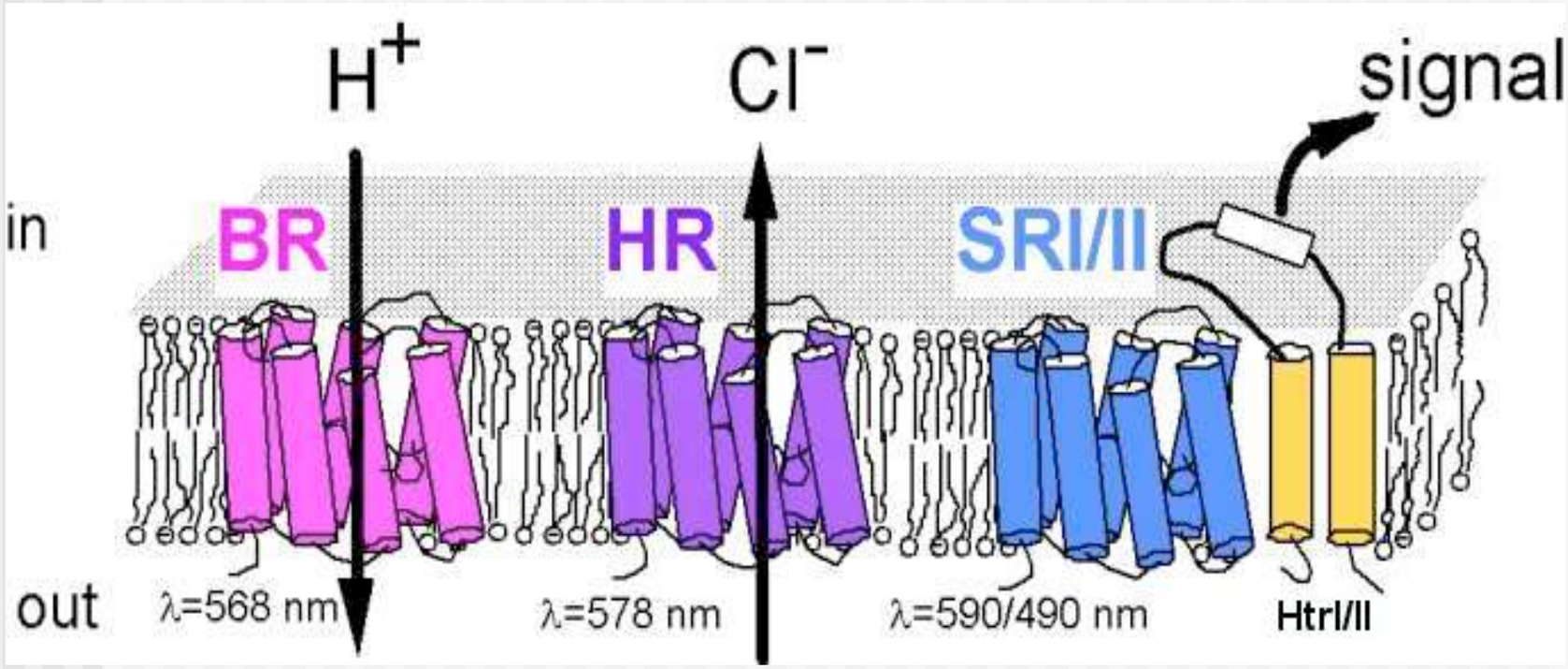


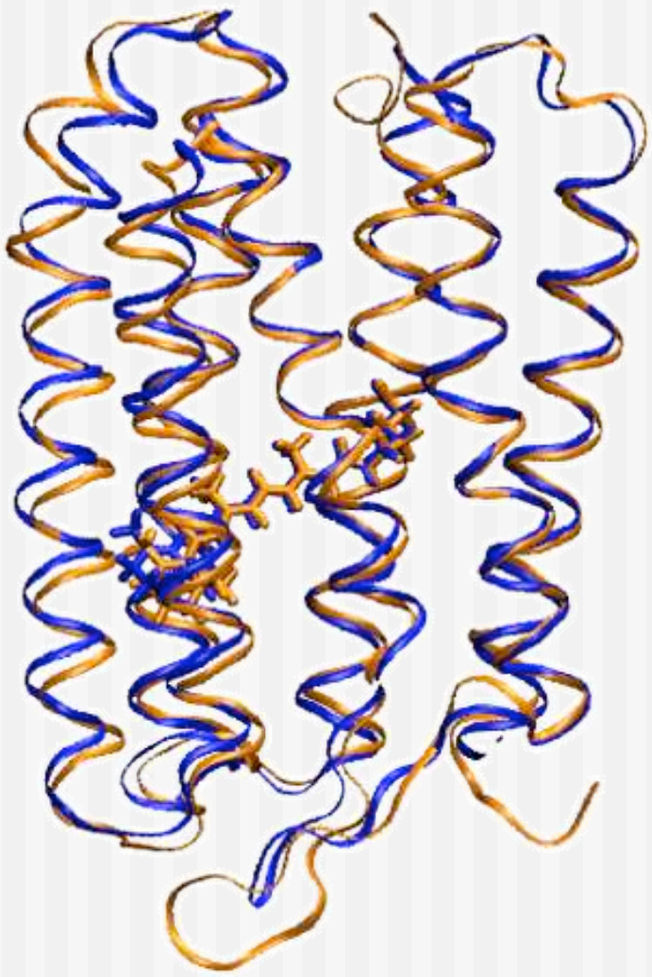
-pumps proton in 5 steps along photocycle

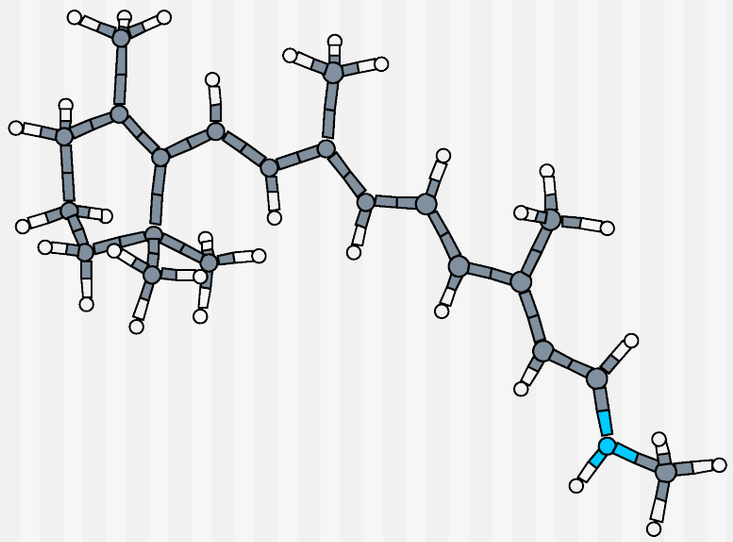
- structural information from x-ray
- IR/Raman/NMR spectra

However:
Complete bio-physical picture still missing

bR vs SRII

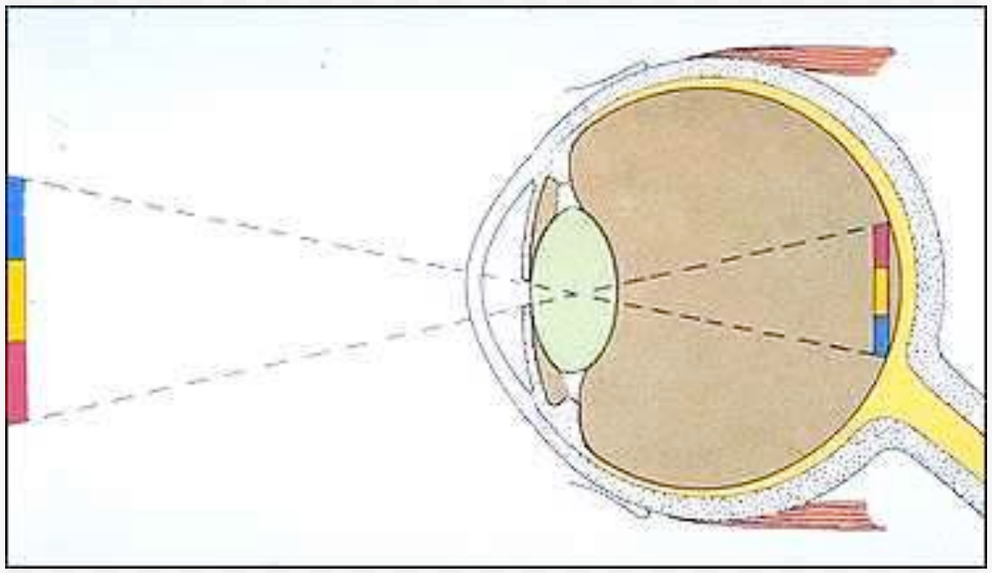


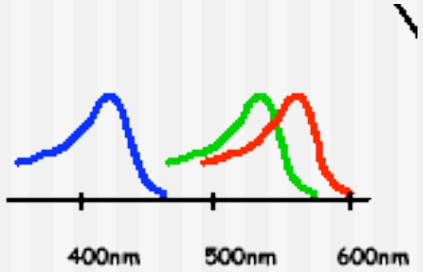


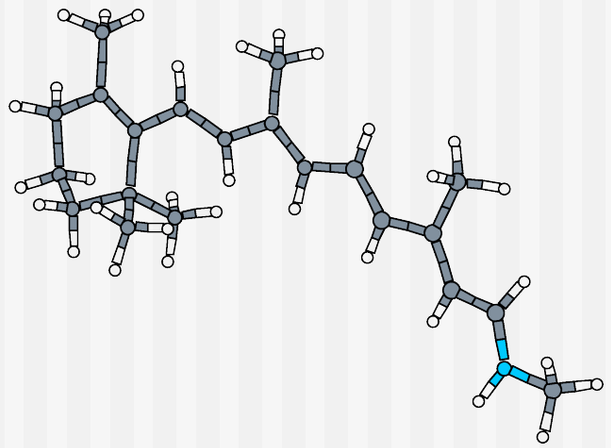


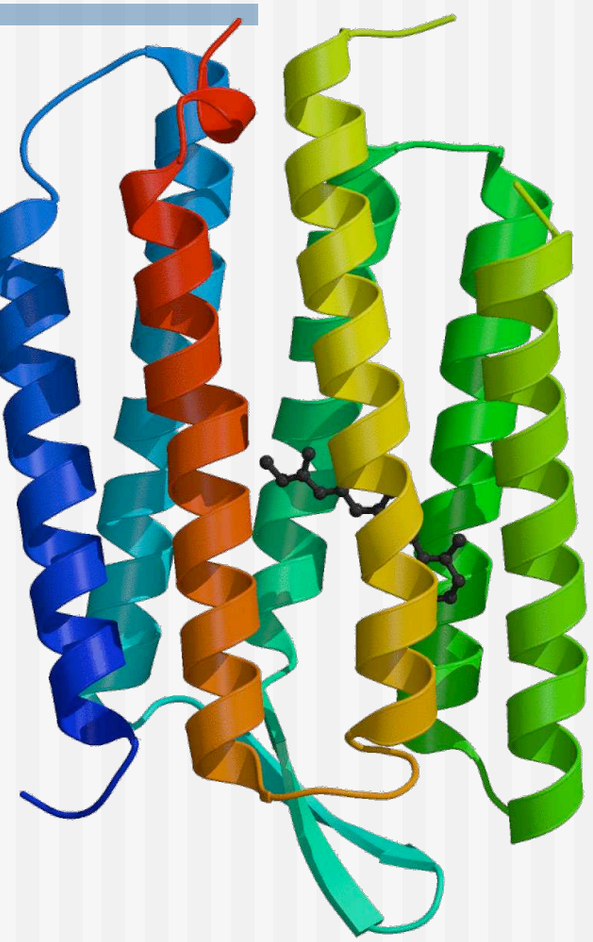
Hoffmann et al. 2006 JACS 108 10808.

Process of vision









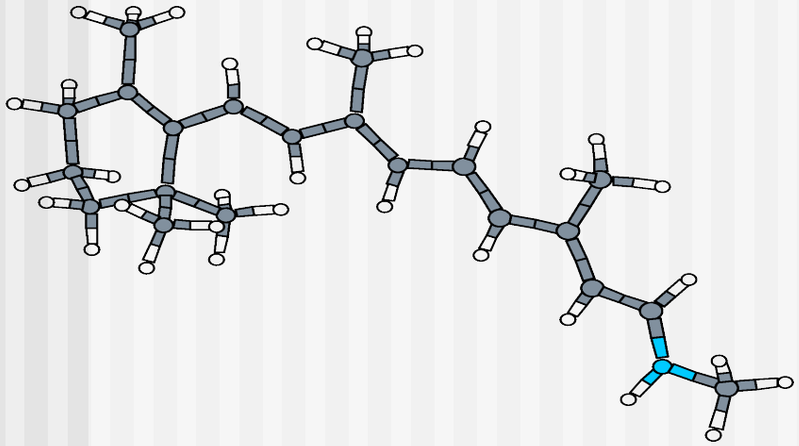
three color pigments, same chromphor:

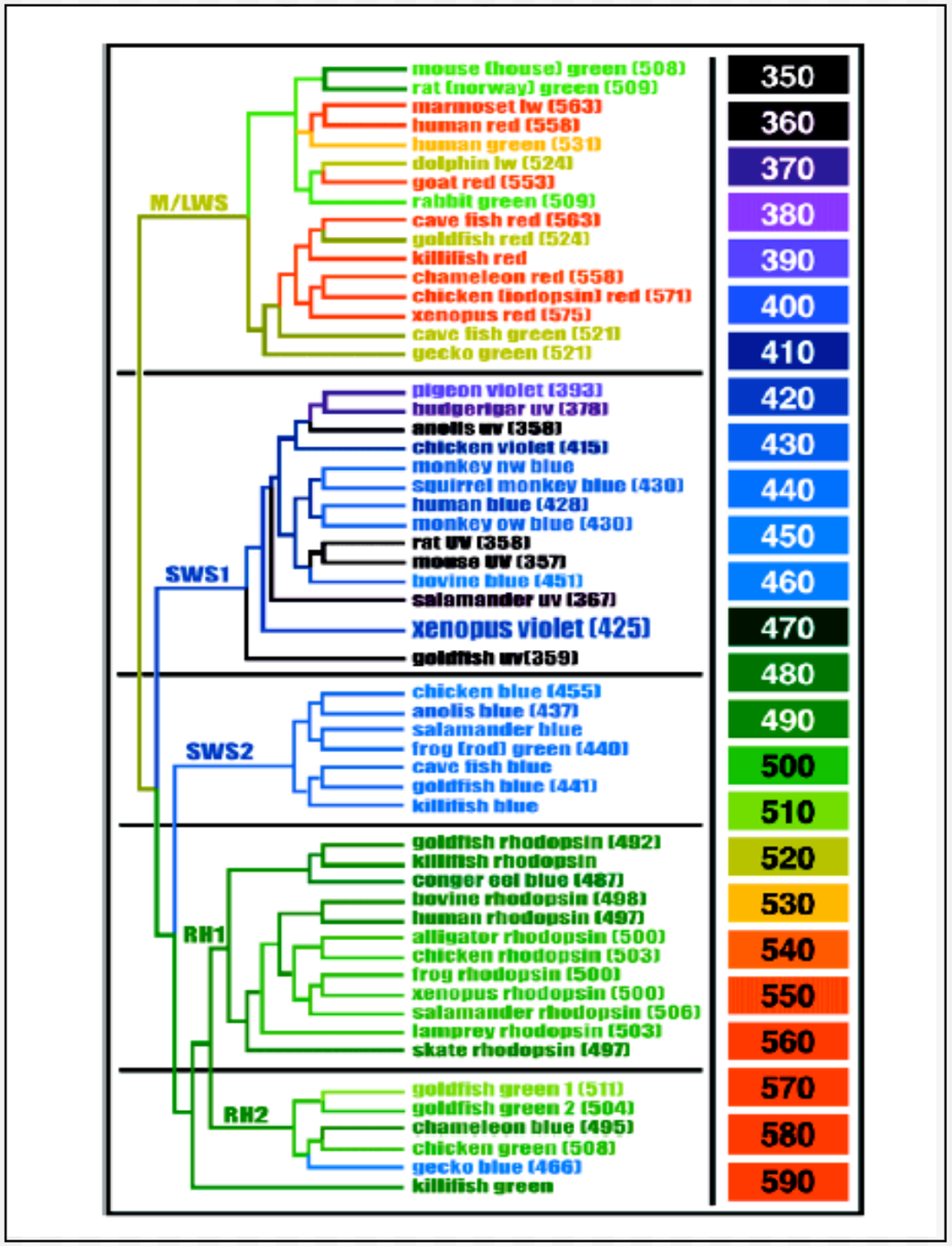
what determines the absorption maximum?

‘Spectral tuning’

Absorption over 300 nm

“Tuning” due to protein environement (opsin-shift)

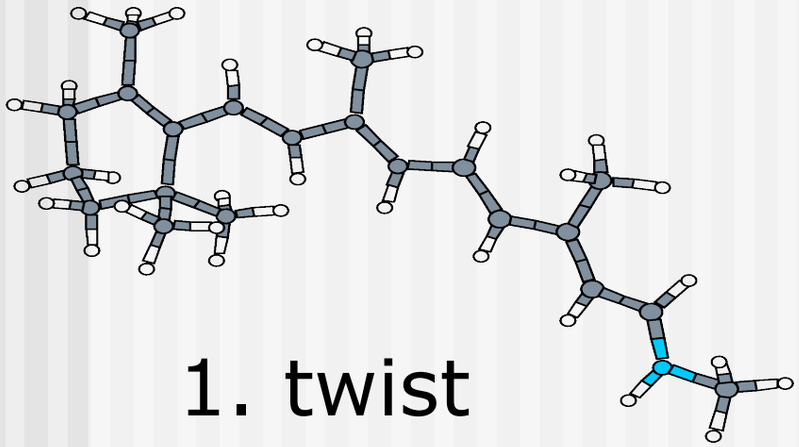




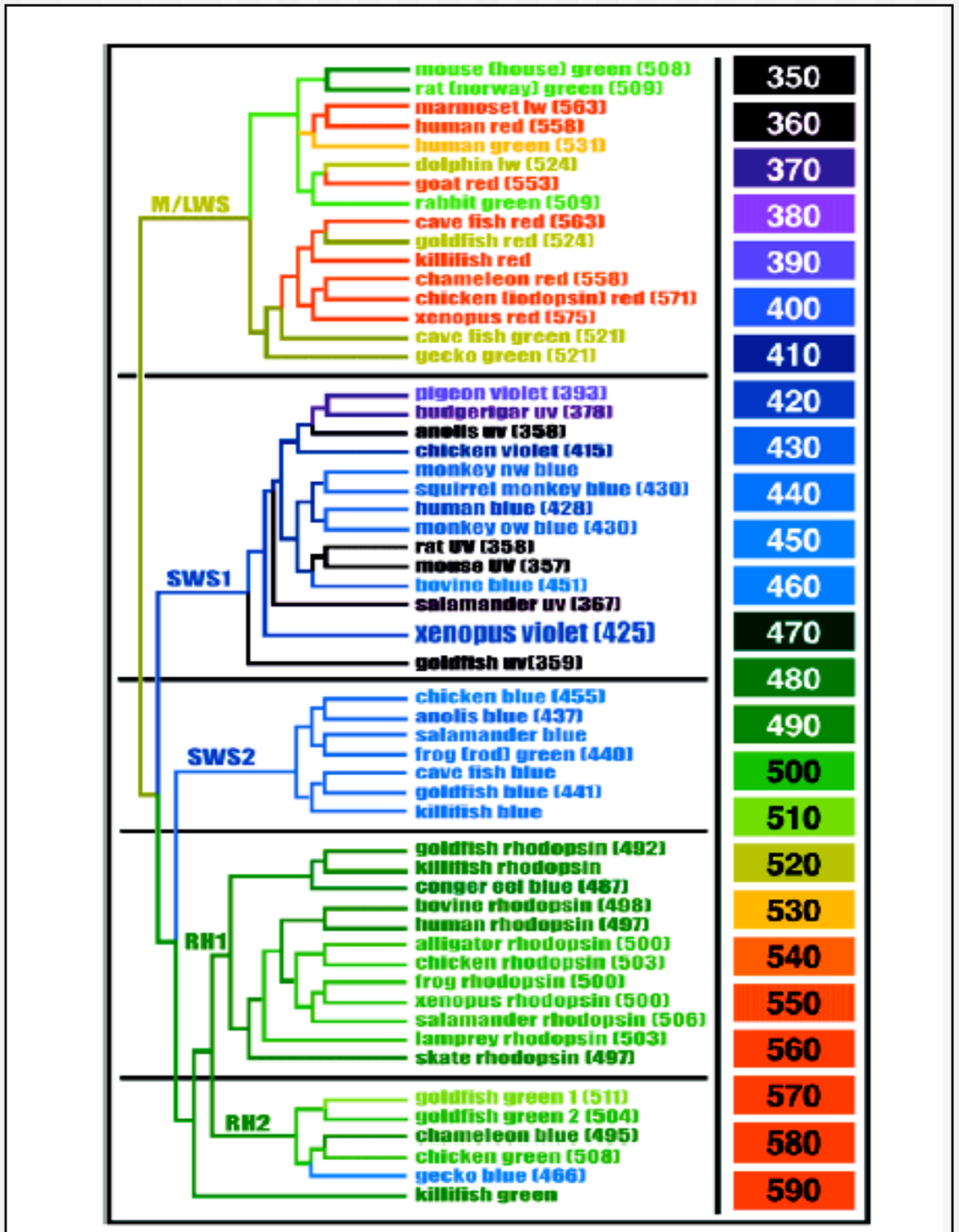
‘Spectral tuning’

Absorption over 300 nm

“Tuning” due to protein environement (opsin-shift)



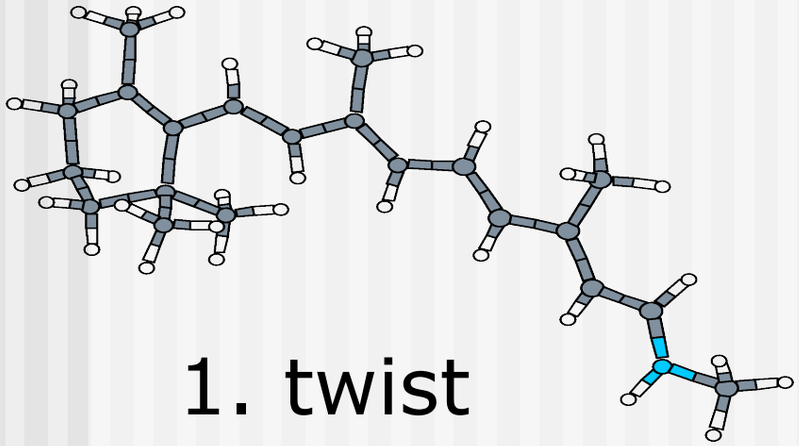
1. twist



‘Spectral tuning’

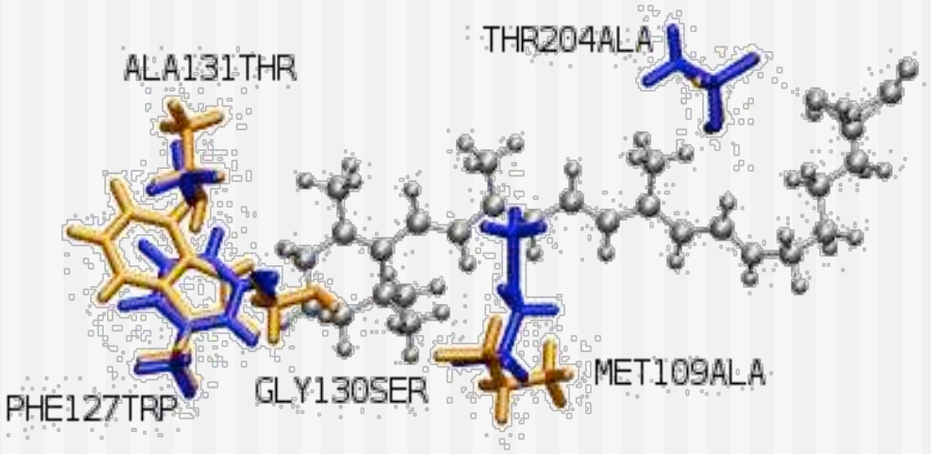
Absorption over 300 nm

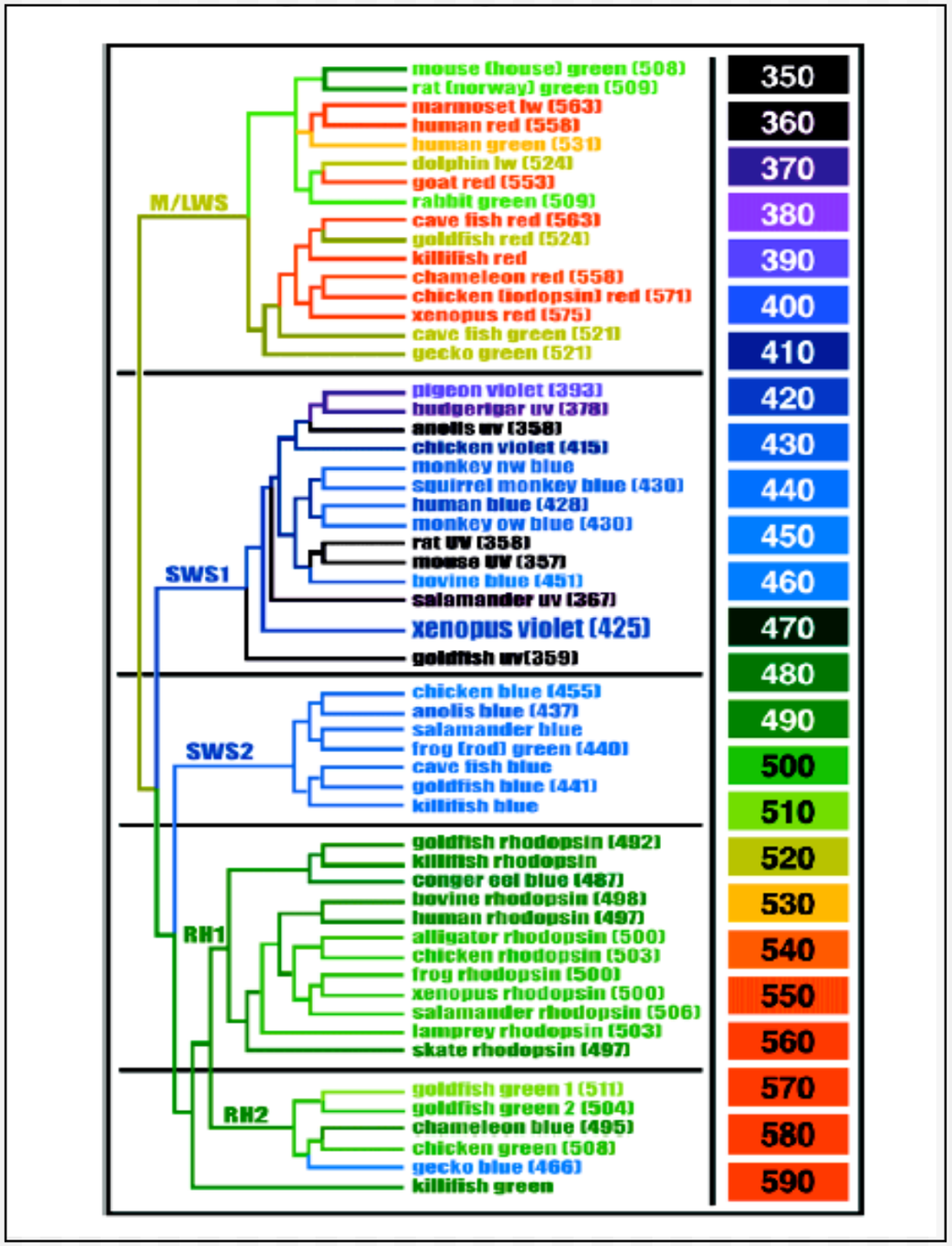
“Tuning” due to protein environement (opsin-shift)



1. twist

2. interaction with polar/ charged groups

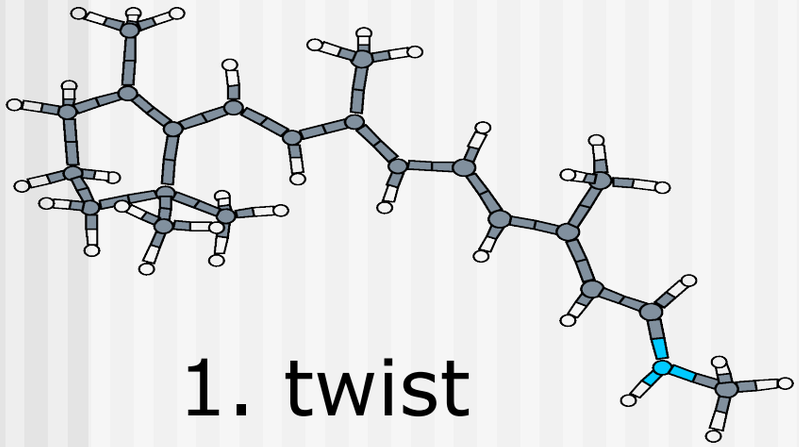




‘Spectral tuning’

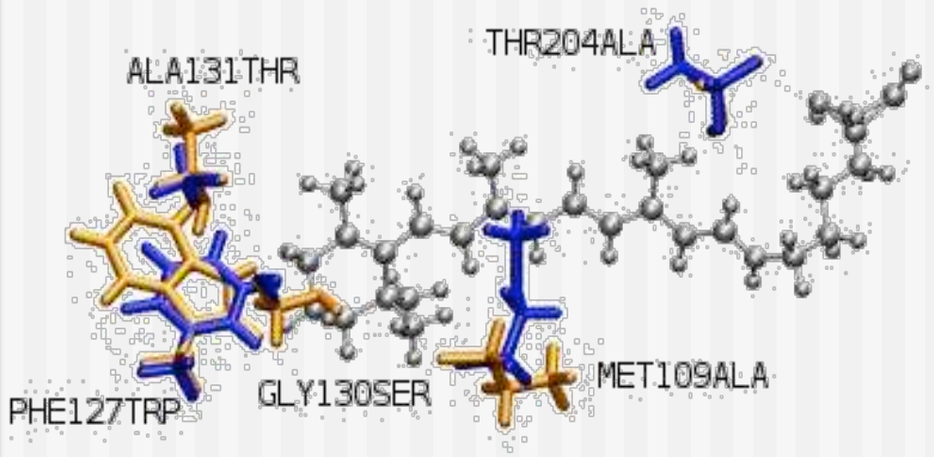
Absorption over 300 nm

“Tuning” due to protein environement (opsin-shift)

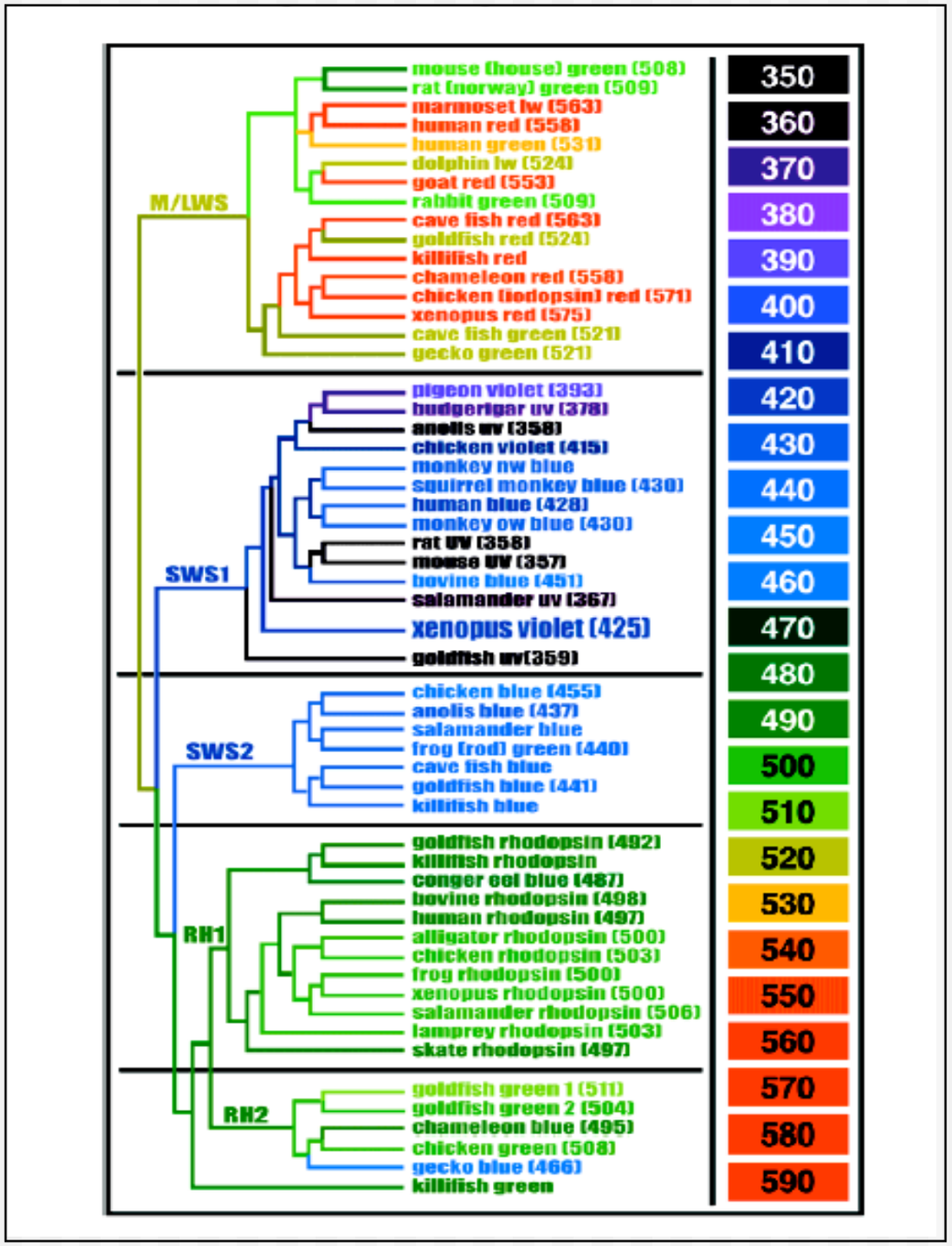


1. twist

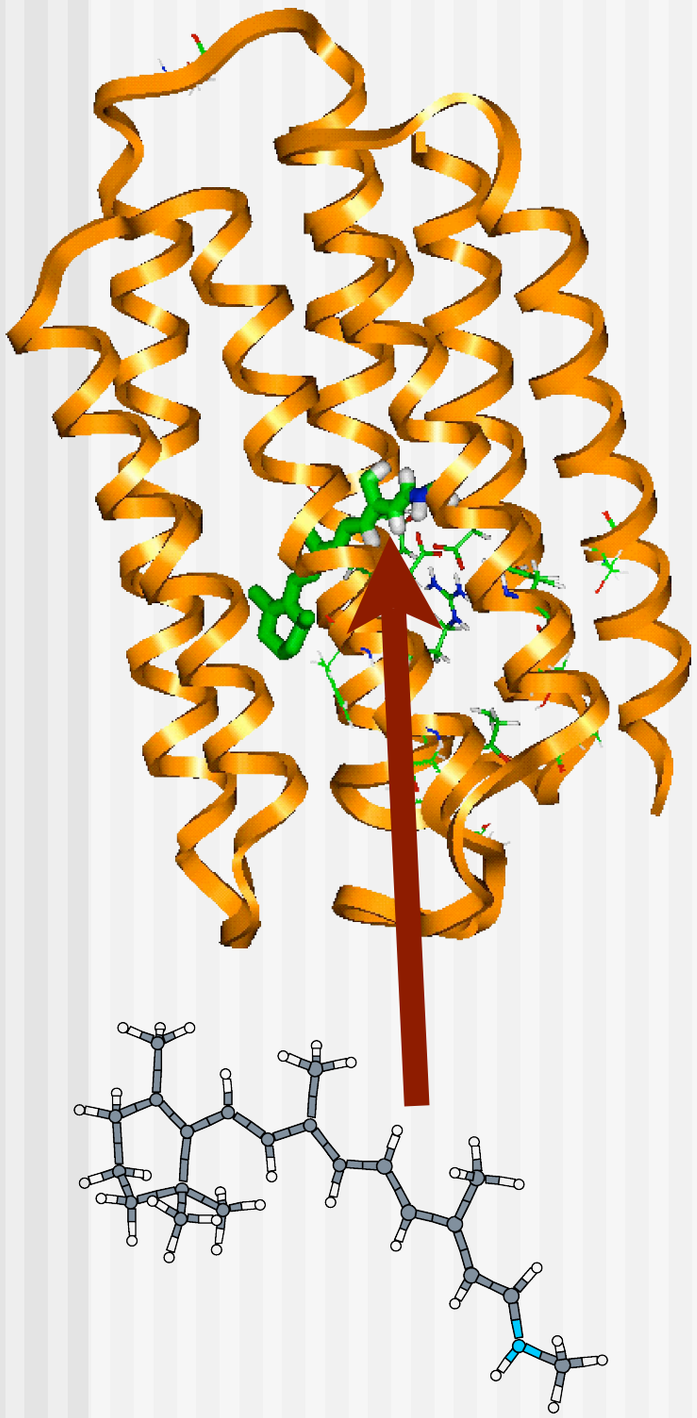
2. interaction with polar/ charged groups



=> ‘predefined’ electrostatic interactions determine function



QM/MM for excited states



Issues:

1) QM methods

a) for ground state: HF, CASSCF, DFT, SCC-DFTB

b) excited state: CI, TD-DFT, CASPT2, MRCI (SORCI)

2) QM/MM coupling: force field electrostatics

a) different MM

b) polarization

c) QM size: CT and dispersion

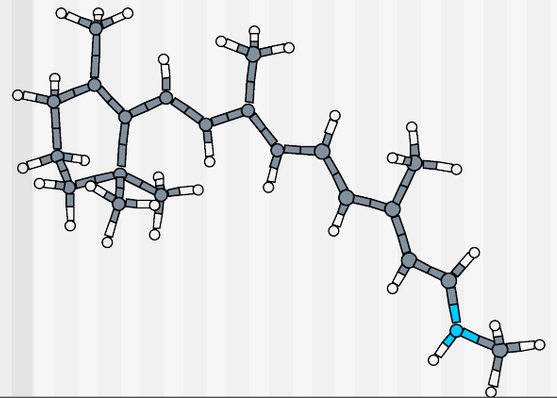
3) solvation: PBC, charge scaling or nothing

4) sampling

QM description of ground state

The problem of the bond length alternation (BLA)

	CASSCF[a]	HF	BH-LYP	B3LYP	DFTB	BLYP
bond length alternation[b] (Å)	0.100	0.069	0.035	0.028	0.025	0.023
average bond length (Å)	1.397	1.388	1.378	1.396	1.399	1.406
TD-BP86	2.00	2.16	2.37	2.33	2.29	2.31
TD-DFTB	1.76	1.95	2.22	2.19	2.18	2.18
TD-B3LYP	2.21	2.36	2.49	2.43	2.40	2.40
OM2/CIS	2.59	2.53	2.40	2.32	2.28	2.27
HF/CIS	3.25	3.18	3.04	2.94	2.89	2.87
OM2/MRCI	2.17	2.22	2.15	2.07	2.03	2.01
SORCI	2.10	2.07	2.04	1.95	1.91	1.89



Wanko et al, JPC B 109 (2005) 3606.

QM description of excited state

- complete failure of TD-DFT
- CIS based methods and CASSF are not accurate enough

=>
- CASPT2

- SORCI (F. Neese)
- OM2/MRCI (W.Thiel)

TD-DFT is color blind

	TD-DFT (B3LYP)	exp.
bR	2.57 eV	2.18 eV (570nm)
SRII	2.58 eV	2.48 eV (500nm)
Rh	2.52 eV	2.49 eV (498nm)

JCTC 3 (2007) 605

JPCB 112 2007 6814

Theor Chem Acc (2003) 109:125

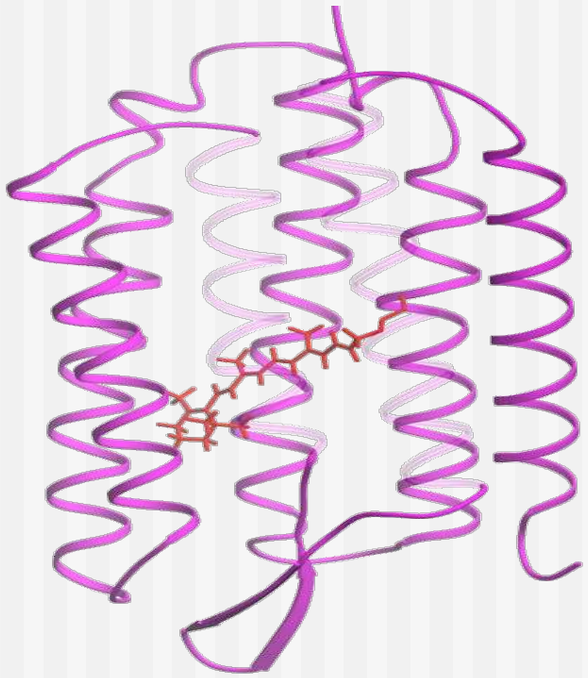
1) too weak response to external field
bR and SRII have same chromophor geometry=> same excitation energy

2) Rh chromophor much more twisted => excitation energy lower

Absolute excitation energies: bR

	S_1 excitation energy (eV)						
	exp	TD-B3LYP[1]	TD-DFTB	OM2/CIS	CASSCF[2]	OM2/MRCI	SORCI
bR (QM:RET)	2.18	2.53	2.21	2.54	3.94	2.66	2.32

[1]Vreven[2003] [2] Hayashi[2000]



Wanko et al, JPCB 109 2005 3606

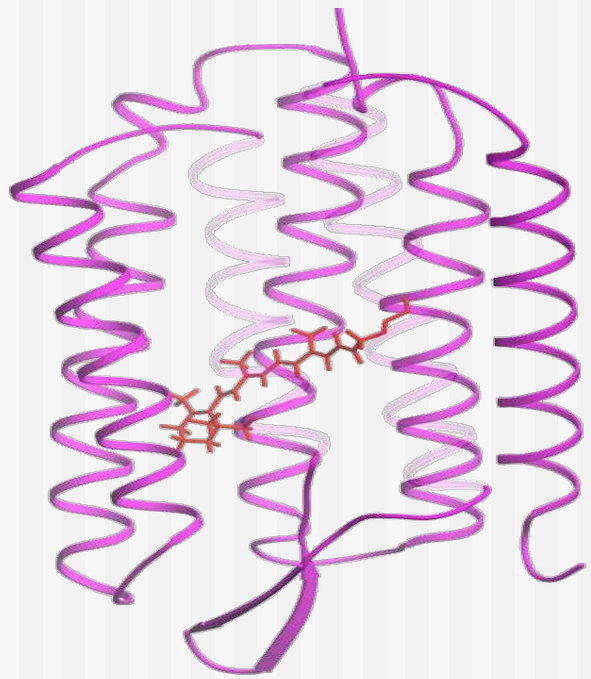
Absolute excitation energies

	S_1 excitation energy (eV)						
	exp	TD-B3LYP[1]	TD-DFTB	OM2/CIS	CASSCF[2]	OM2/MRCI	SORCI
vacuum		2.42	2.14	2.34	2.86	2.22	1.89
bR (QM:RET)	2.18	2.53	2.21	2.54	3.94	2.66	2.32
		0.1		**0.2**	**1.0**	**0.4**	

[1]Vreven[2003] [2] Hayashi[2000]

- TDDFT nearly zero
- CIS shifts still too small ~50%
- OM2/MRCI compares very well
- OM2: consistent blue shift

➔ TDDFT, CIS and CASSCF not applicable for color tuning!



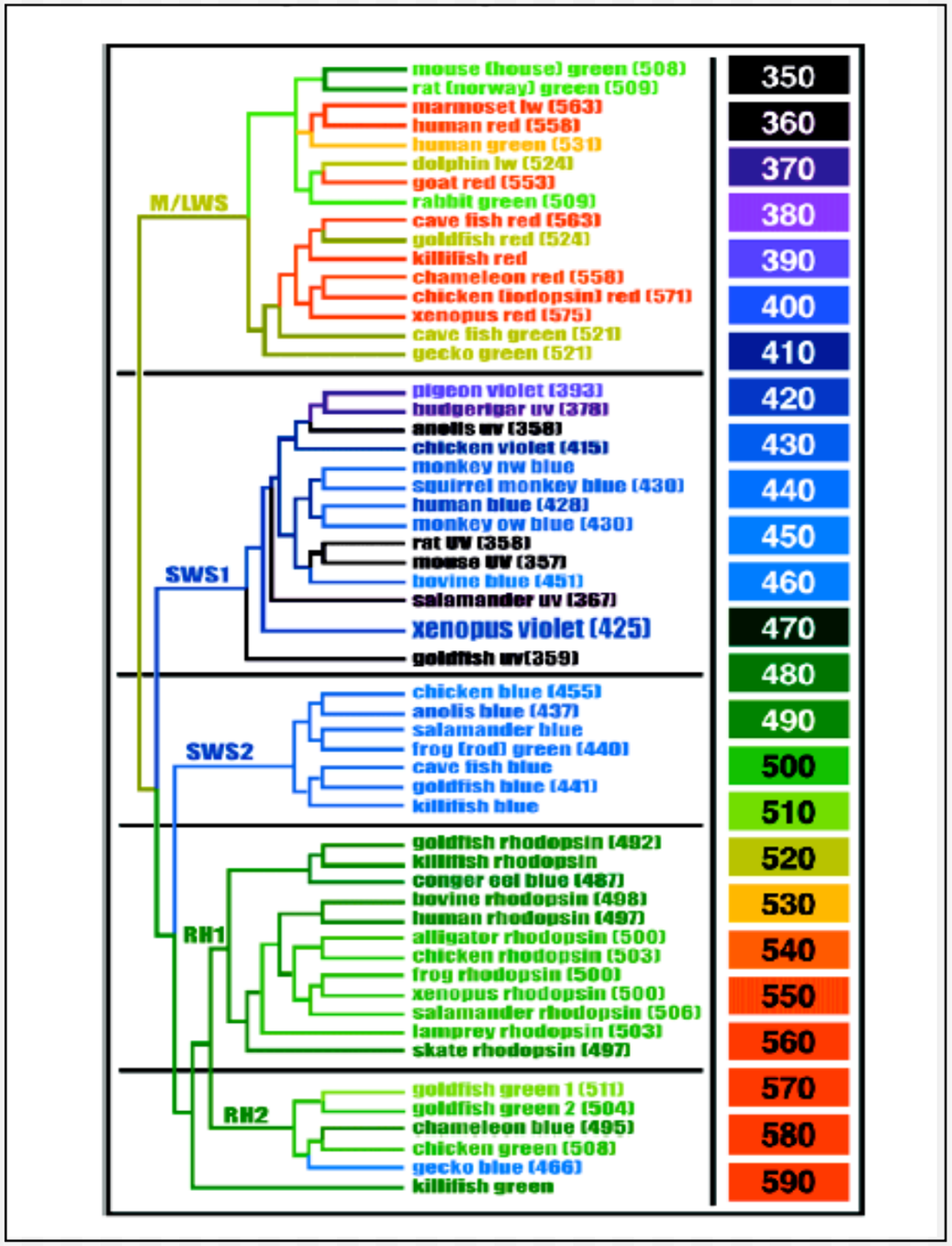
Wanko et al, JPCB 109 2005 3606

Color tuning

Spectral tuning over 300 nm
Mechanism of color tuning:

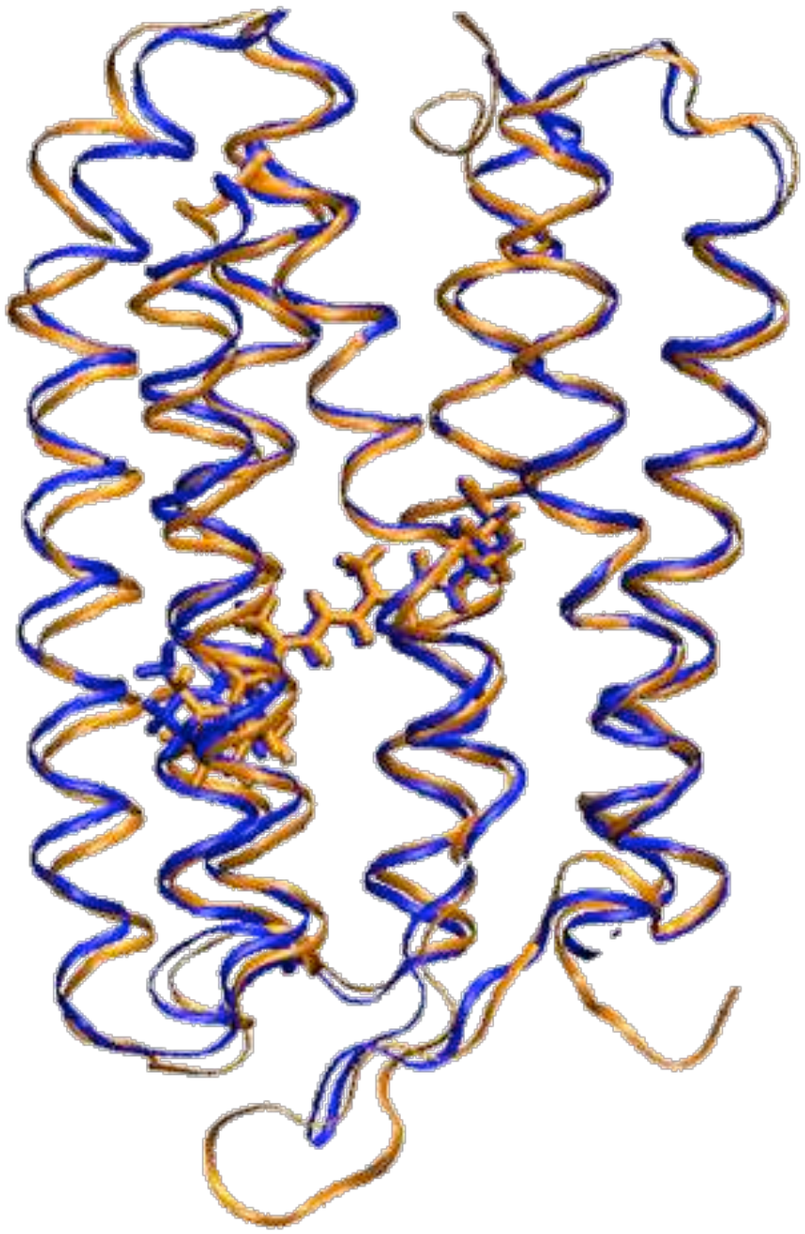
- retinal twist
- interaction with polar/charged residues
- interaction with the counterion(s)

can we understand the mechanisms of color tuning from theory?



from Kusnetzow et al.
Biochemistry 2001, 40, 7832

bR vs sRII (ppR): relate mechanism of color tuning to structural basis (X-ray)



- sRII UV sensor λ_{max} ~ 500 nm
- bR proton pump λ_{max} ~ 570 nm
- sRII spectrum is blue-shifted by 0.32 eV
- nearly identical 3d structure

- calculated: 0.31 eV

Molecular mechanism for this spectral difference?

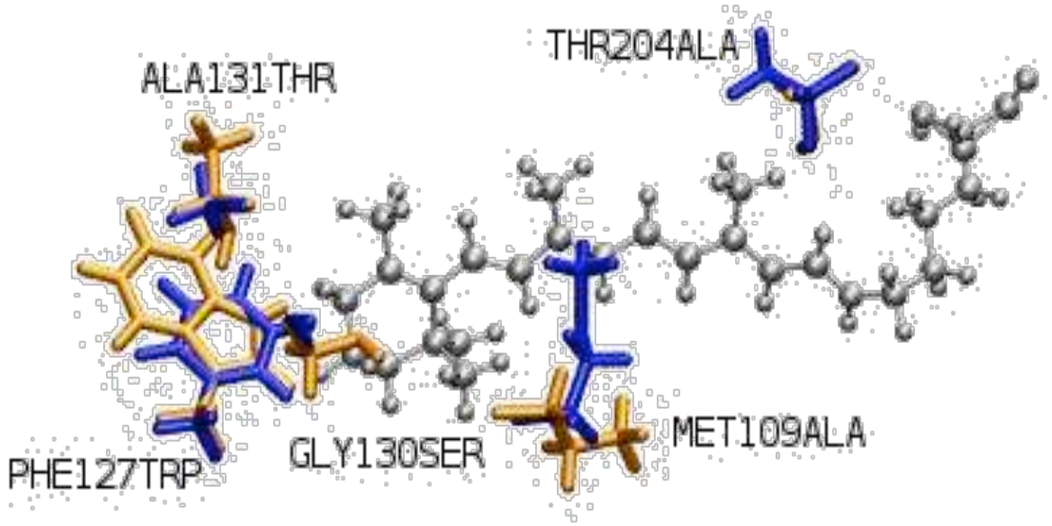
a)same retinal geometry.

b)different AAs in binding pocket.

c)counterion distance.

Hoffmann et al. 2006 JACS 108 10808.

Mutation experiments: sRII



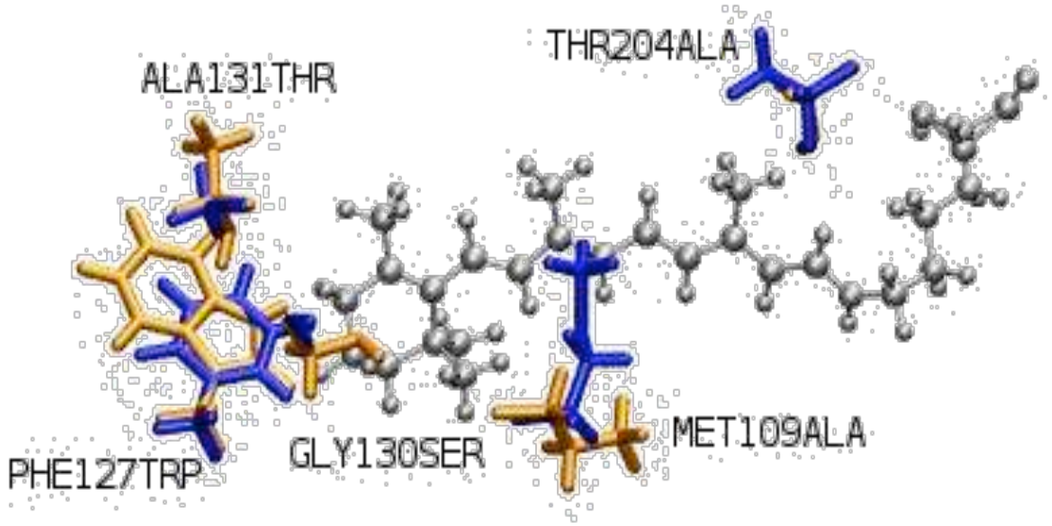
	exp.	local relax.		full relax.	
		OM2	SORCI	OM2	SORCI
V108M	-0.02	0.01	0.00	-0.02	-0.05
G130S	-0.02	-0.03	-0.02	-0.06	0.04
T204A	-0.04	-0.07	-0.04	-0.06	-0.04
G130S/T204A	-0.07	-0.11	-0.03	-0.13	-0.05
V108M/G130S	-0.05	-0.02	0.00	-0.07	-0.03
V108M/T204A	-0.05	-0.06	-0.04	-0.08	-0.06
V108M/G130S/T204A	-0.08	-0.09	-0.02	-0.14	-0.10

	exp.	local relax.	full relax.
Ile43Val	-0.02	0.00	-0.01
Ile83Leu	-0.01	0.00	0.00
Asn105AspH	0.00	-0.00	0.01
Met109Ile	0.01	0.02	0.03
Ala131Thr	-0.02	-0.01	-0.01
Phe127Trp	0.01	-0.00	0.01
Phe134Met	-0.01	-0.01	0.00
bR/ppR’	-0.05	-0.04	-0.06
bR/ppR	-0.12	-0.13	-0.16

bR/ppR:
Mutant identical to bR in binding pocket (Shimono 2001,2003)

~40% of total shift

Hoffmann et al. 2006 JACS 108 10808.



	exp.	local relax. OM2	local relax. SORCI	full relax. OM2	full relax. SORCI
V108M	-0.02	0.01	0.00	-0.02	-0.05
G130S	-0.02	-0.03	-0.02	-0.06	0.04
T204A	-0.04	-0.07	-0.04	-0.06	-0.04
G130S/T204A	-0.07	-0.11	-0.03	-0.13	-0.05
V108M/G130S	-0.05	-0.02	0.00	-0.07	-0.03
V108M/T204A	-0.05	-0.06	-0.04	-0.08	-0.06
V108M/G130S/T204A	-0.08	-0.09	-0.02	-0.14	-0.10

	exp.	local relax.	full relax.
Ile43Val	-0.02	0.00	-0.01
Ile83Leu	-0.01	0.00	0.00
Asn105AspH	0.00	-0.00	0.01
Met109Ile	0.01	0.02	0.03
Ala131Thr	-0.02	-0.01	-0.01
Phe127Trp	0.01	-0.00	0.01
Phe134Met	-0.01	-0.01	0.00
bR/ppR’	-0.05	-0.04	-0.06
bR/ppR	-0.12	-0.13	-0.16

bR/ppR:
Mutant identical to bR in binding pocket (Shimono 2001,2003)

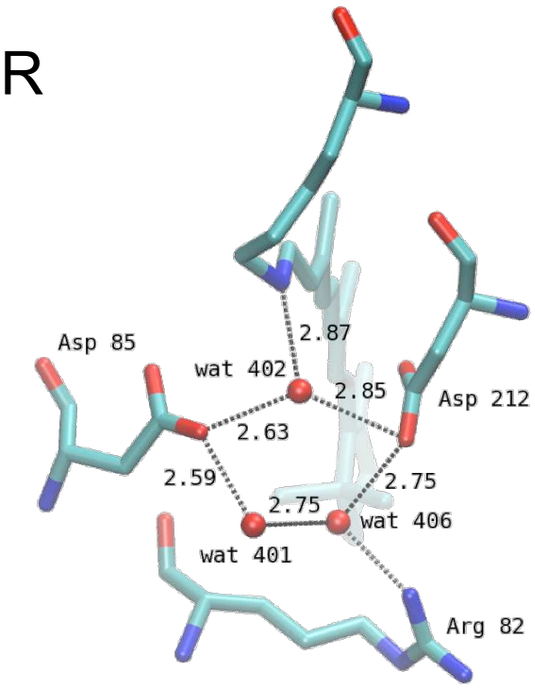
~40% of total shift

Calc. overestimate effect of mutation slightly: ~ 50% of total shift

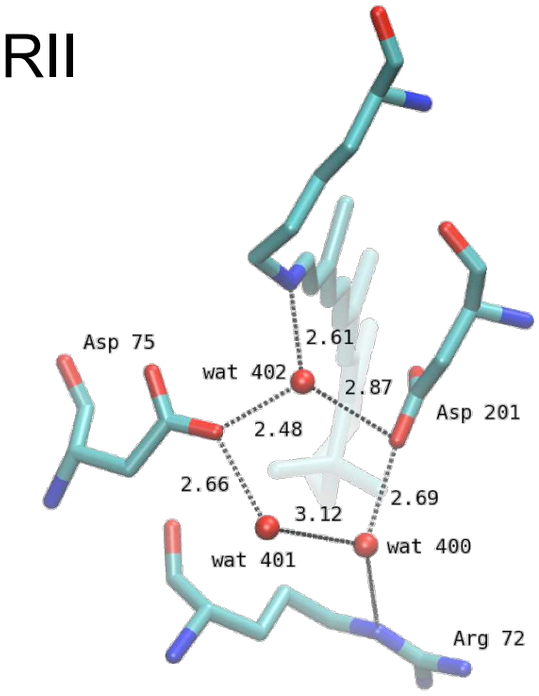
Hoffmann et al. 2006 JACS 108 10808.

Color tuning due to different hydrogen bonding pattern?

bR



sRII

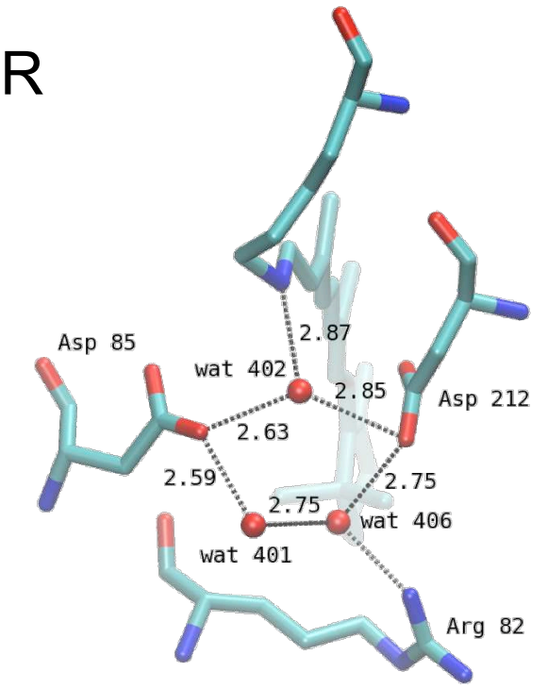


H-bonded network (HBN) stronger in SRII (Kandori 2003)

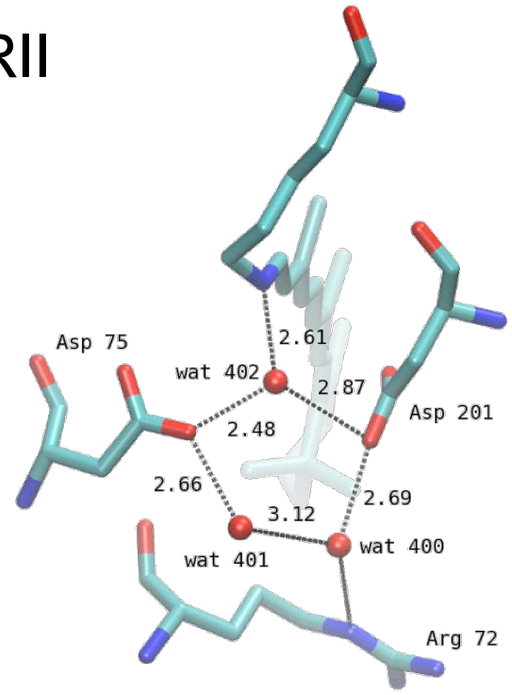
From FTIR: N-D mode shift 33 cm^{-1}

Hoffmann et al. 2006 JACS 108 10808.

bR



sRII



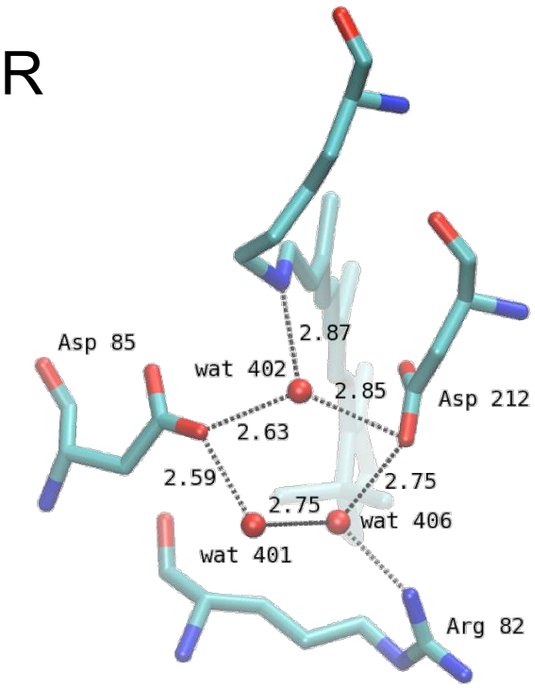
H-bonded network (HBN) stronger in SRII (Kandori 2003)

From FTIR: N-D mode shift 33 cm^{-1}

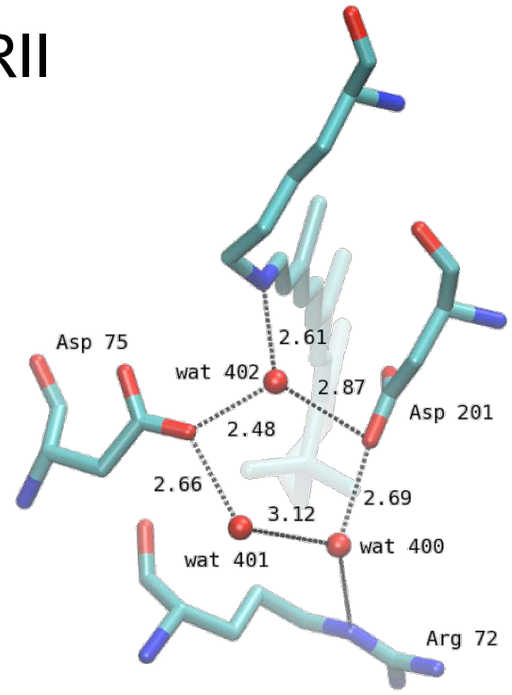
1) Calculated: 20 cm^{-1}

Hoffmann et al. 2006 JACS 108 10808.

bR



sRII



H-bonded network (HBN) stronger in SRII (Kandori 2003)

From FTIR: N-D mode shift 33 cm^{-1}

1) Calculated: 20 cm^{-1}

2) Contribution of hydrogen bonded network to shift:

~0.1eV ➔ 30%

Hoffmann et al. 2006 JACS 108 10808.

Further residues: perturbation analysis; mutate every residue to Gly

bR/ppR	bR	ppR	ΔE_{ppR-bR}	position	bp
		counterion residues			
Asp85/75	−0.39	−0.39	0.00	H−C	+
Asp212/201	−0.37	−0.32	0.05	H−G	+
Arg82/72	0.01	−0.02	−0.03	H−C	+
		water molecules in the HBN[b]			
W401/W401	0.02	0.02	0.00	−	+
W402/W402	−0.09	−0.08	−0.01	−	+
W406/W400	0.03	0.02	0.01	−	+
		remaining charged residues			
Glu194/Pro183	0.08	0.00	−0.08	H−F	−
GluH204/Asp193	0.01	0.04	0.03	H−G	−
		conserved residues			
Trp182/171	−0.01(8)	−0.01(4)	0.00	H−F	+
Tyr185/174	−0.02(5)	−0.01(2)	0.01	H−F	+
Trp189/178	0.02	0.01	−0.01	H−F	+
Tyr57/51	0.02	0.02	0.00	H−B	−
Tyr83/73	0.01(8)	0.00(4)	−0.01	H−C	+
Trp86/76	0.03	0.03	0.00	H−C	+
Thr89/79	−0.03	−0.03	0.00	H−C	+
Thr90/80	0.04	0.03	−0.01	H−C	+

Spectral tuning: several reasons

2 main factors: - H bonding network (HBN)
- polar residues in binding pocket

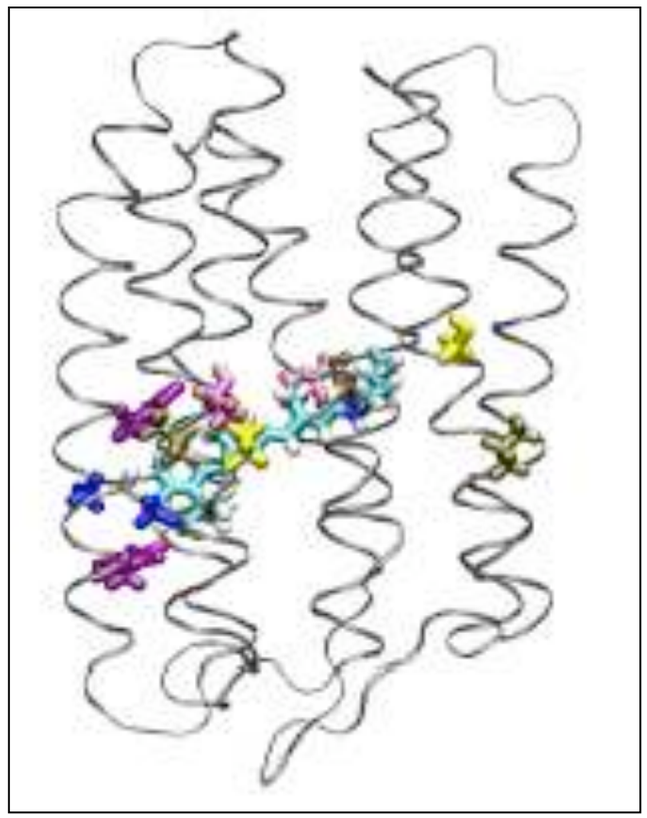
Impact of individual residues is small! 0.05 eV (10 nm)

AS in binding pocket: ~40%

AS more distant: ~10%

H-bonding: ~40%

Aromatic residues: ~10%

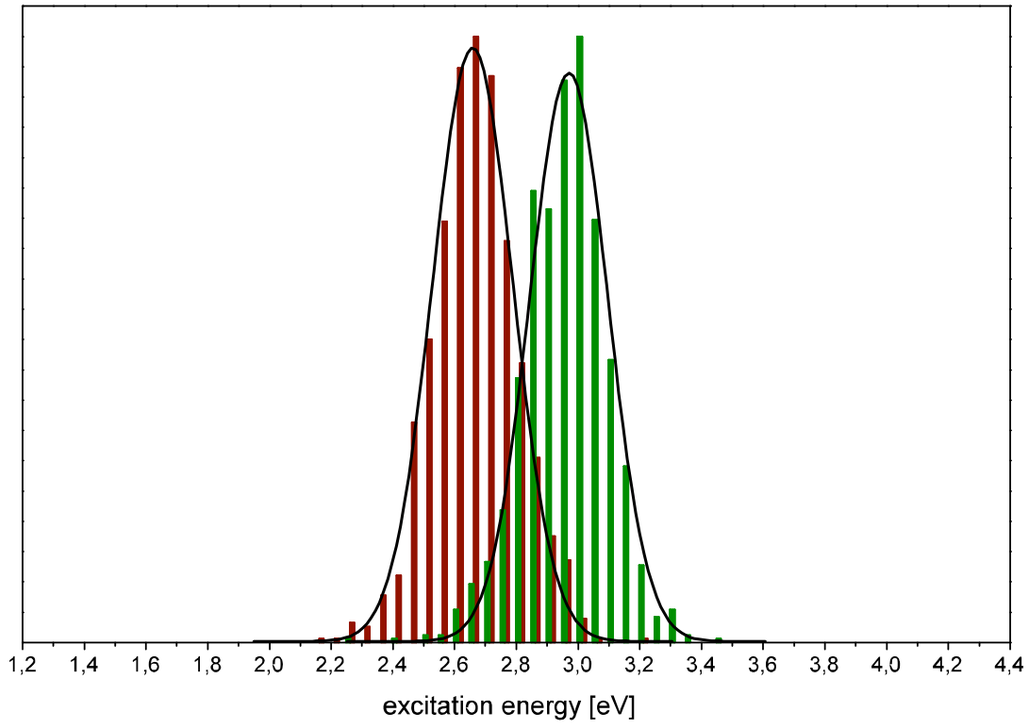


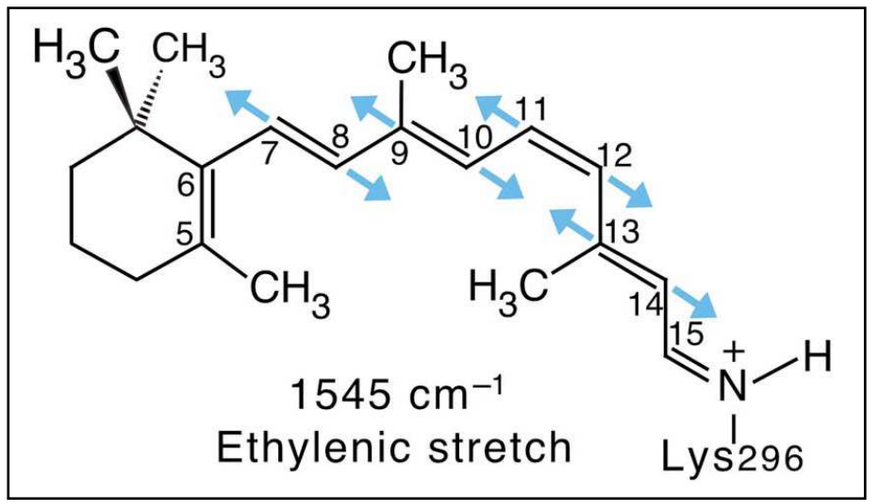
Hoffmann et al. 2006 JACS 108 10808.

Dynamical effects

Do we need sampling?

QM/MM MD: optical spectra with OM2/MRCI





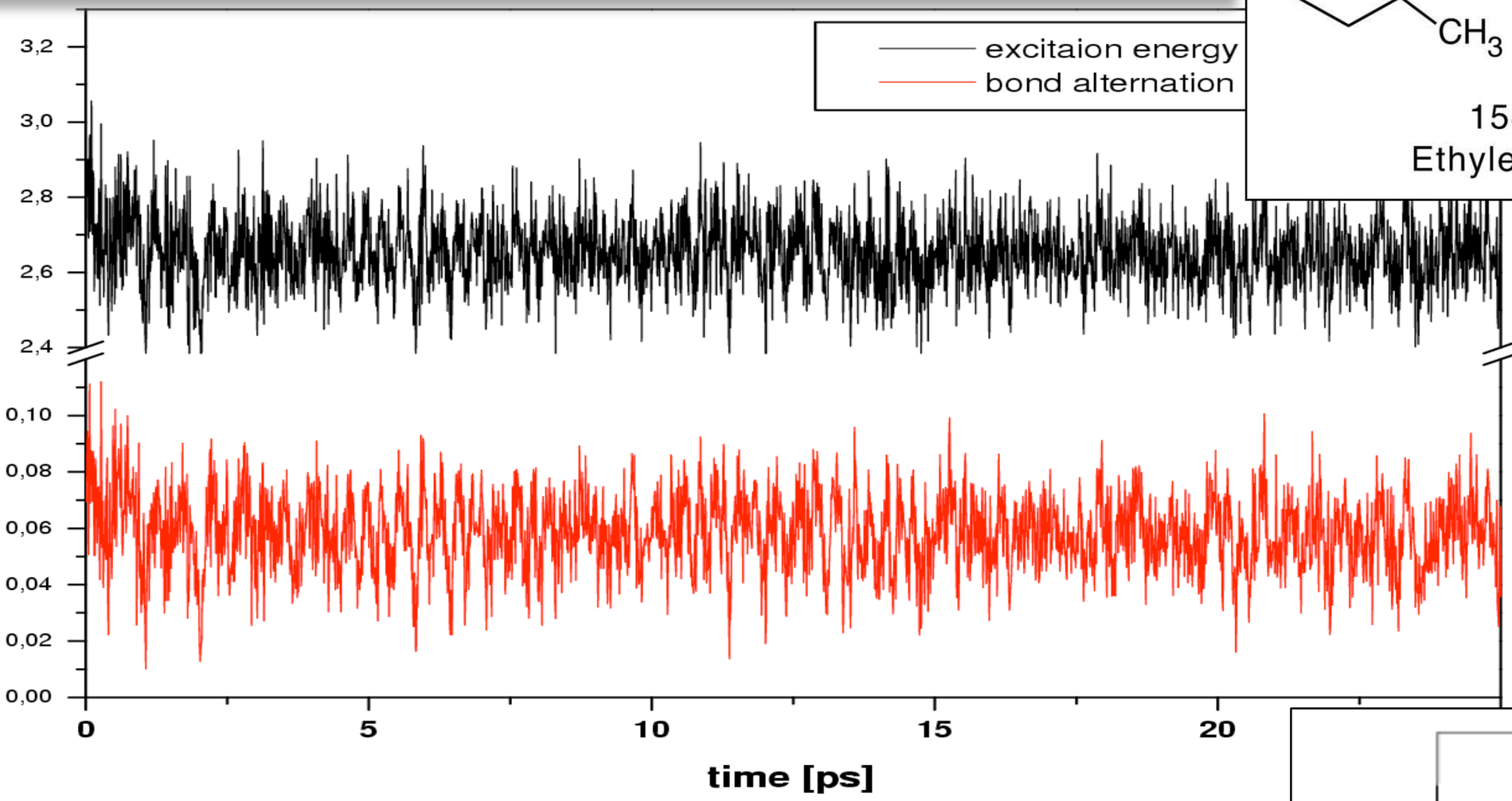
a) Calculated: 0.31 eV (exp. 0.32 eV)

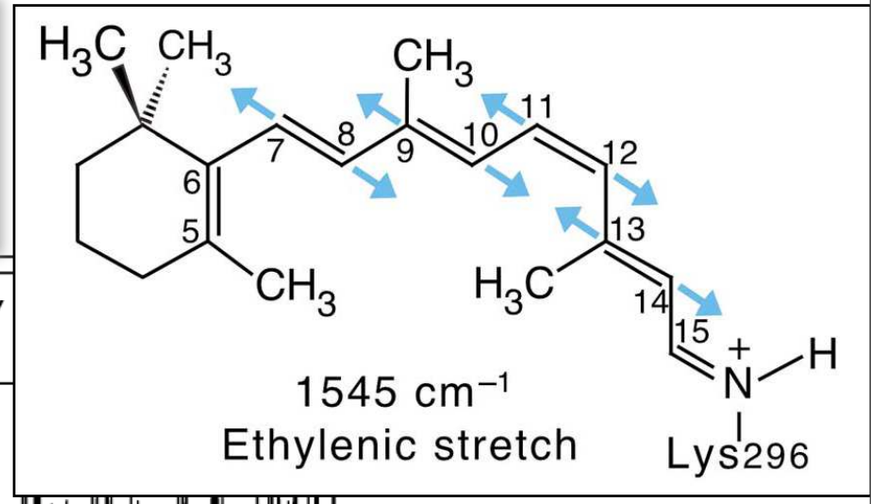
b) Spectral width in good agreement with experiment

Correlation of bond alternation and excitation energy: r=0.8

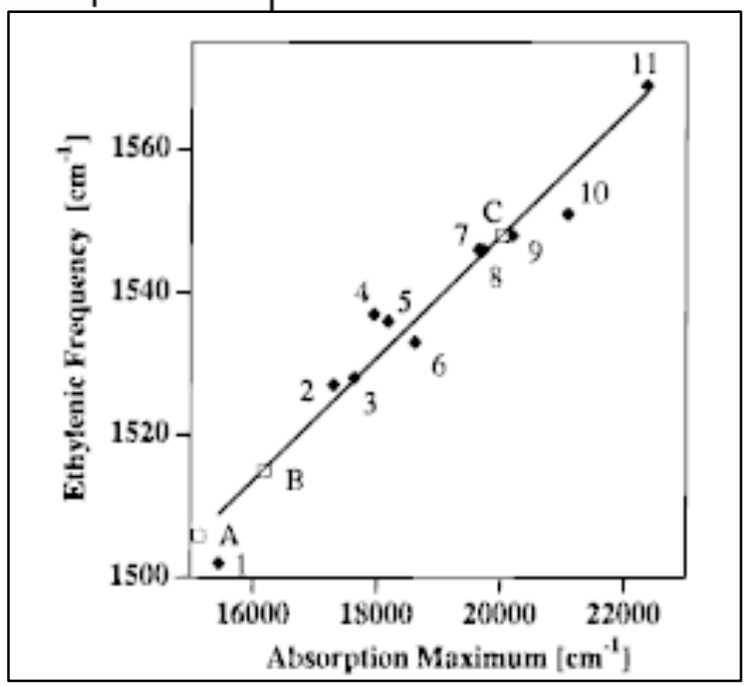
➔ Fluctuations in C=C mode responsible for line-width

QM/MM MD over 2ns: bondlength alternation and excitation energy

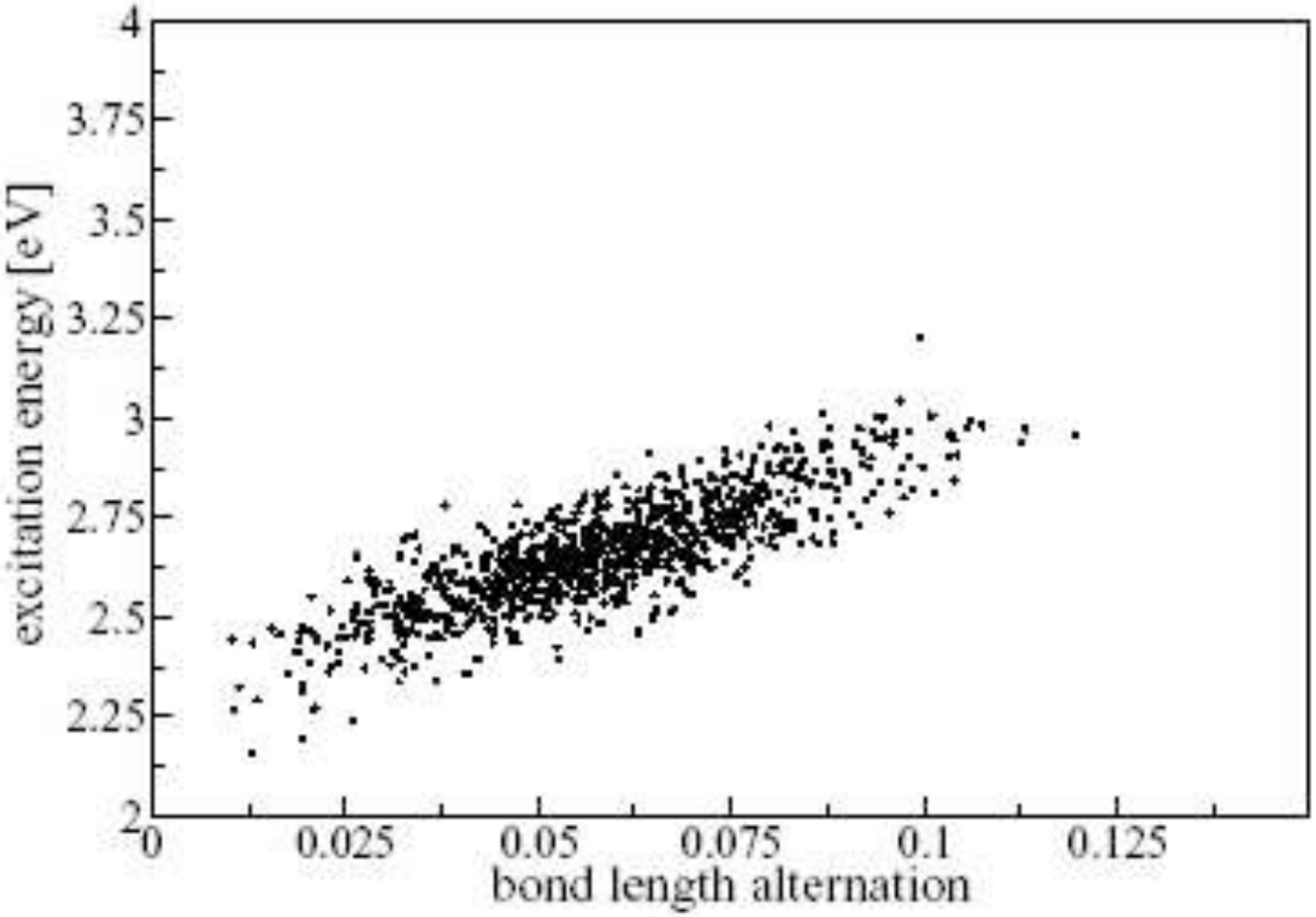


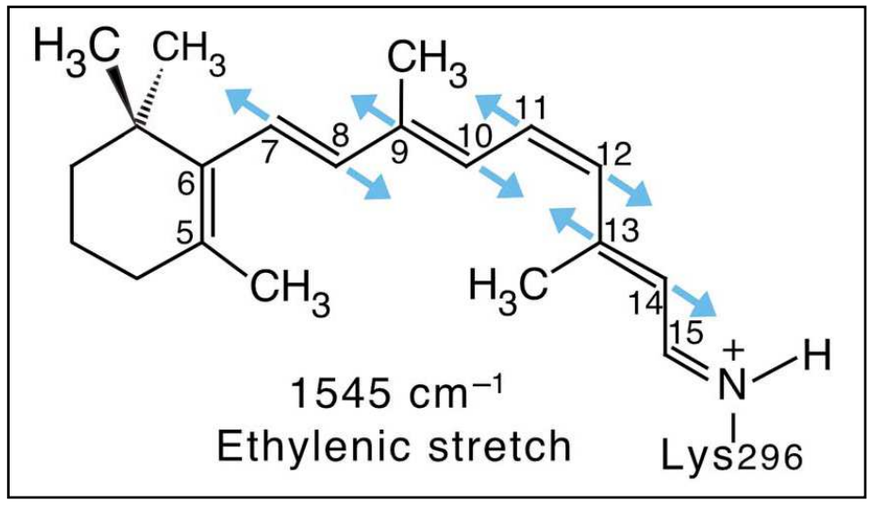


- BLA as a measure of C=C frequency
- Linear relationship of C=C stretch and excitation energy (Kochendoerfer et al 1997)



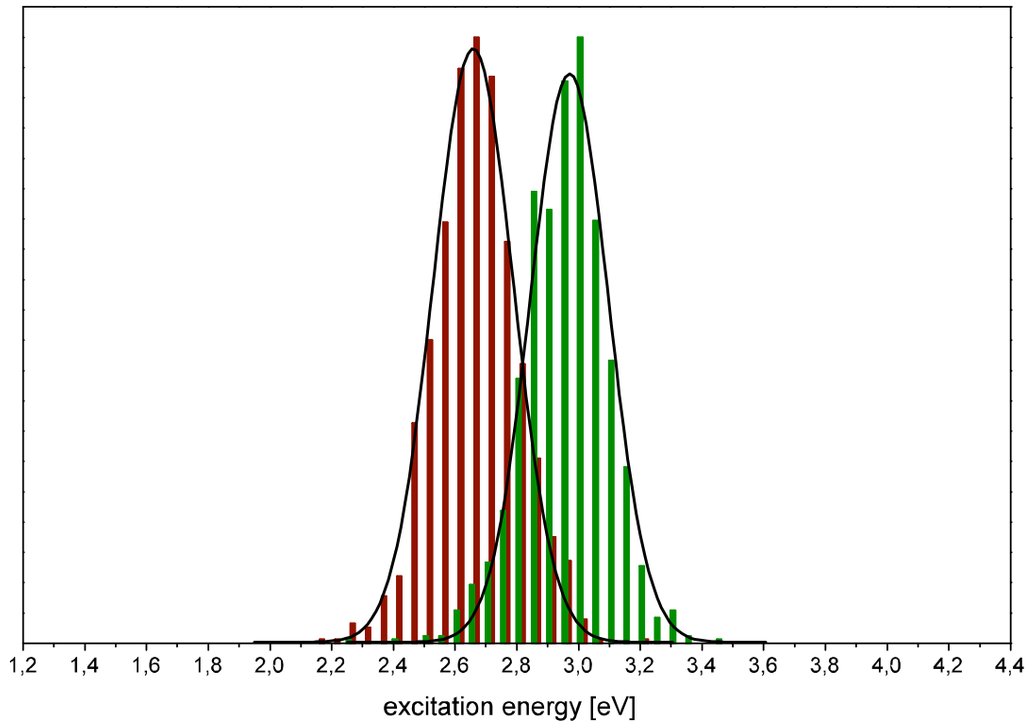
QM/MM MD over 2ns: bondlength alternation and excitation energy

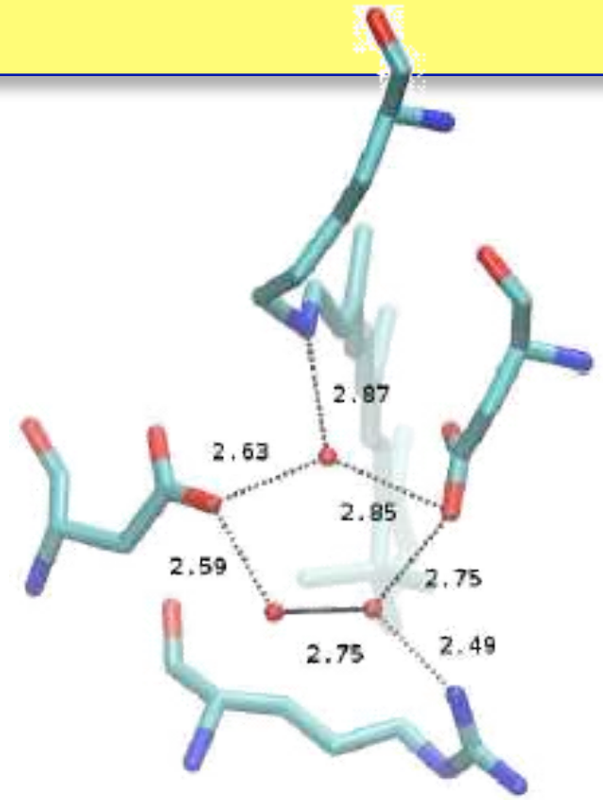




Correlation of bond alternation and excitation energy: r=0.8

Impact of fluctuations





Particular lucky case:

- excitation energy at optimized structure and maxima coincide

=> use of optimized structures meaningful

strong hydrogen bonded network keeps things in place

Towards absolute excitation energies

- Size of QM region
  dispersion and charge transfer

- MM charges
- solvent effects
- MM polarizability

Effects missing in QM/MM

Experiment	SORCI/CHARMM
2.18	2.32

- Though QM methods achieve accuracies of 0.1 eV and below, the bR absorption maximum is still overestimated

- Inclusion of CT effects ⇒ blue shift

- Houjou, Birge, and Warshel have suggested red shifts of 0.2-0.3 eV for bR due to polarization

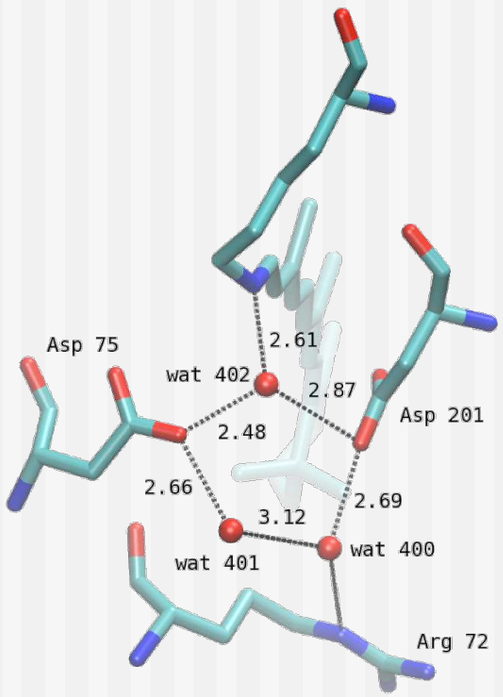
Effects missing in QM/MM

Experiment	SORCI/CHARMM
2.18	2.32

- Though QM methods achieve accuracies of 0.1 eV and below, the bR absorption maximum is still overestimated

- Inclusion of CT effects ⇒ blue shift

- Houjou, Birge, and Warshel have suggested red shifts of 0.2-0.3 eV for bR due to polarization



Effects missing in QM/MM

Experiment	SORCI/CHARMM
2.18	2.32

- Though QM methods achieve accuracies of 0.1 eV and below, the bR absorption maximum is still overestimated

- Inclusion of CT effects ⇒ blue shift

- Houjou, Birge, and Warshel have suggested red shifts of 0.2-0.3 eV for bR due to polarization

Effects missing in QM/MM

Experiment	SORCI/CHARMM
2.18	2.32

- Though QM methods achieve accuracies of 0.1 eV and below, the bR absorption maximum is still overestimated

- Inc

- Ho
ha
of 0.2-0.3 eV for bR
due to polarization

In particular retinal proteins:

•extended charge transfer (Dm=12 debye) due to S_0-S_1 excitation

•significant polarization of nearby aromatic residues

Polarizable force field for environment

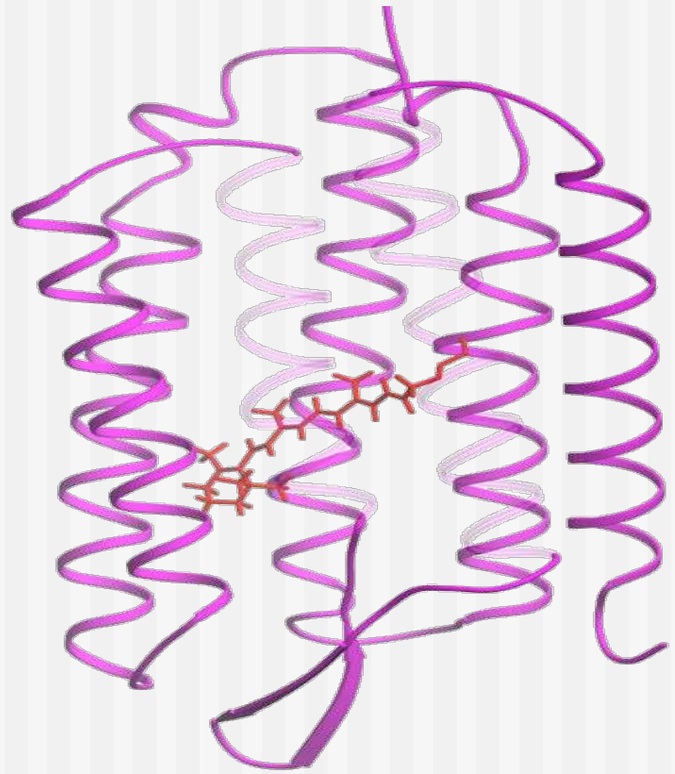
- 0.2 eV [Warshel and Chu 2000]
- 0.36 eV [Houjou et al. 2001]
- 0.15 eV [Ren et al 2001] includes dispersion of small active site

- MM charges
- MM polarization

➔ RESP charges for residues in gas phase

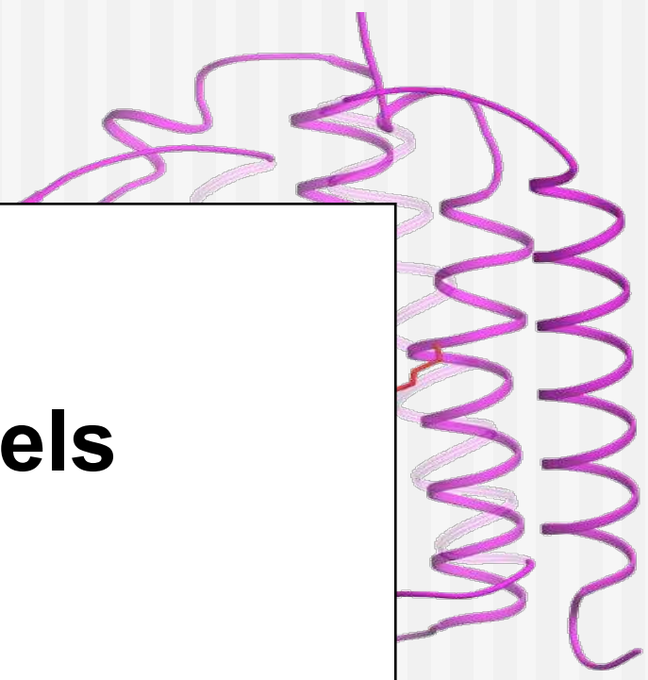
➔ atomic polarizabilities: $\mu = \alpha E$

➔ Polarization red shift of about 0.14 eV:



Polarizable force field for environment

- 0.2 eV [Warshel and Chu 2000]
- 0.36 eV [Houjou et al. 2001]
- 0.15 eV [Ren et al 2001] includes dispersion of small active site



- MM charges
- MM polarization
-
-
-

How reliable are MMpol models for that purpose?

Limits of standard QM/MM models

Issues:

- MM charges ➔ 0.05 eV (e.g. AMBER vs CHARMM)
- charge scaling ➔ 0.05 eV
- MM polarization ➔ 0.05 eV
- charge transfer ➔ 0.05 –0.1 eV
- dispersion ➔ 0.05-0.1 eV

Effect of MM charges and charge scaling

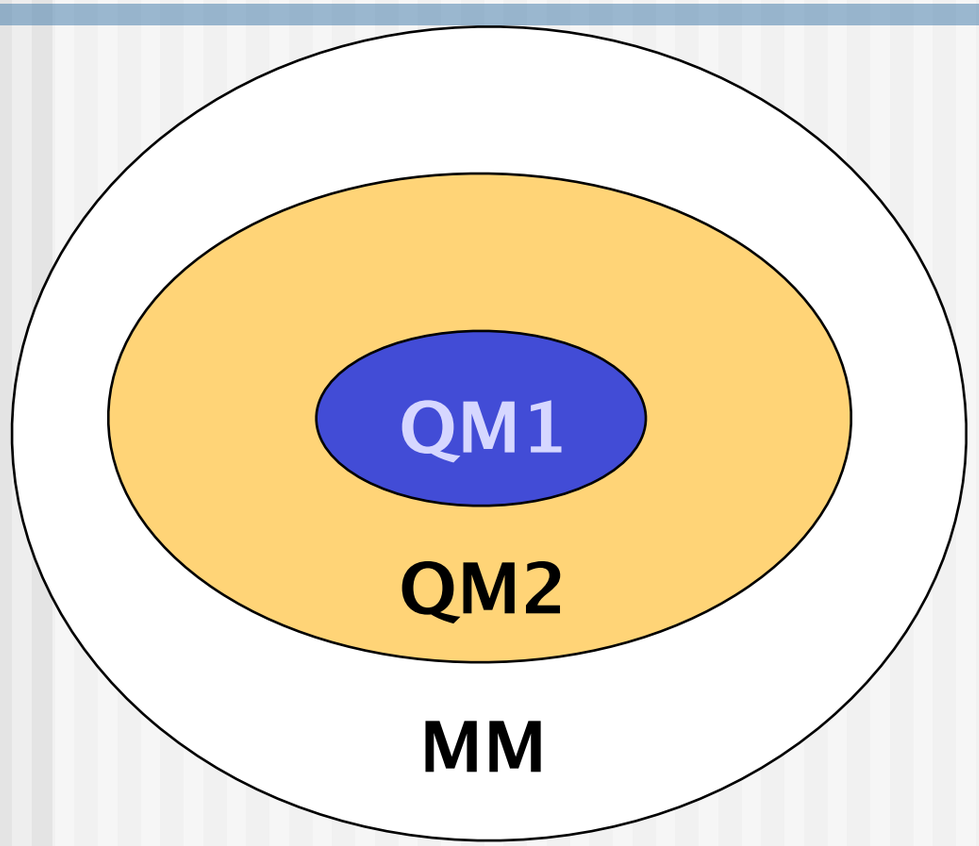
TABLE 6: SORCI QM/MM S_1 Excitation Energies (eV)[a]

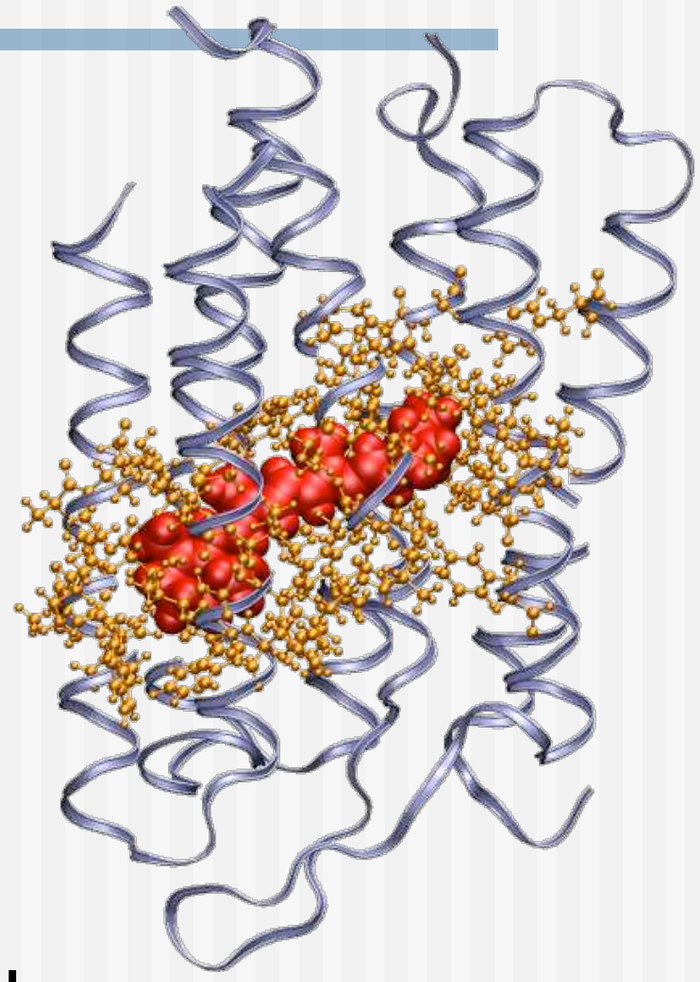
	bR	psRII	Rh(u)	Rh(p)
CHARMM27	2.32	2.56	2.53	2.42
amber ff03	2.32	2.49	2.61	2.45
charge scaling				
CHARMM27[c]	2.28	2.62	2.50	2.34

- MM charges ➔ 0.05 eV (e.g. AMBER vs CHARMM)
- charge scaling ➔ 0.05 eV

Wanko et al. JPCB 2008 112, 11462.

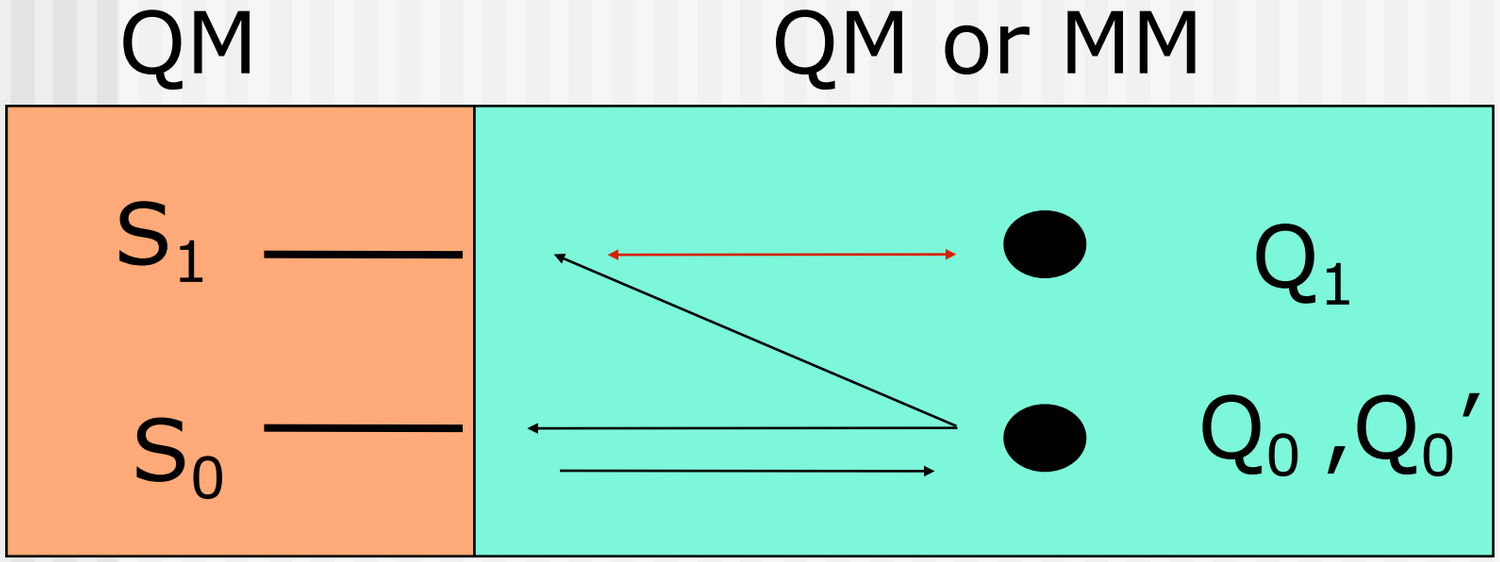
Polarization of Protein with QM/QM/MM:





QM1: retinal OM2, SORCI
QM2: 300 atoms (DFT, PBE0/SVP and DFTB)
MM: rest of protein (CHARMM, AMBER)

Wanko et al. JPCB 2008 112, 11462.



Q_0 : MM charges

Q_0': S_0 polarized charges

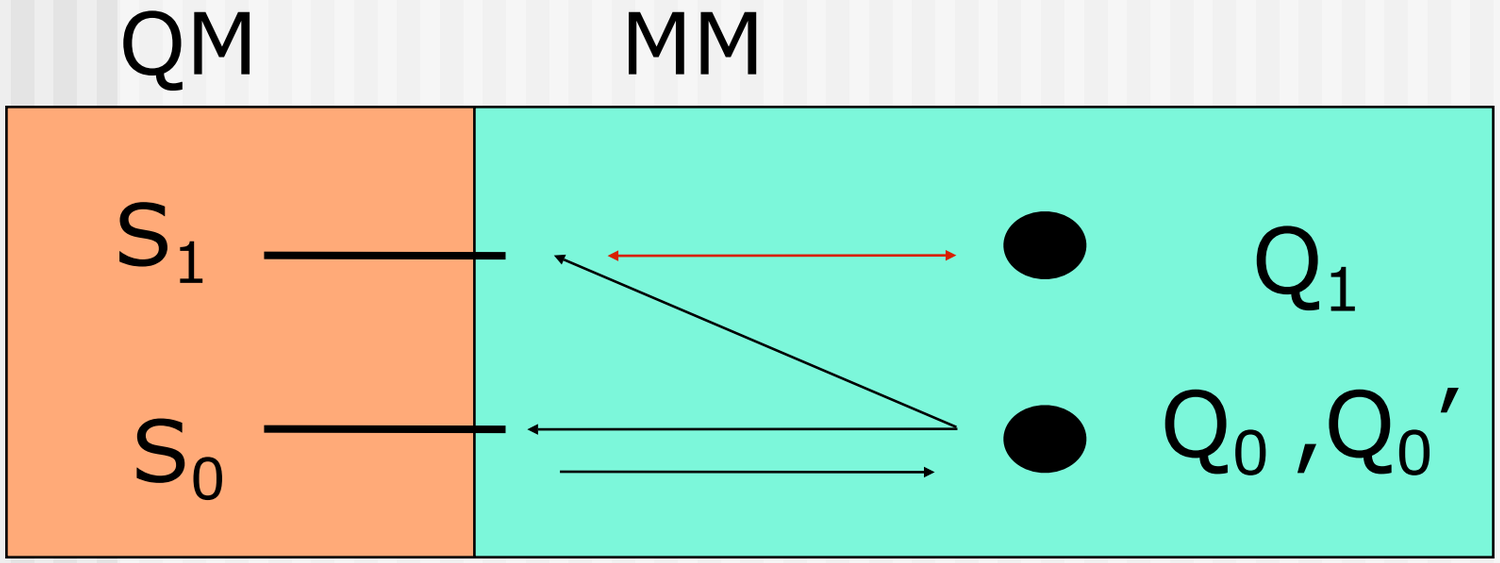
Q_1 : S_1 polarized charges

2 Effects:

a) Q_0 ➔Q_0' effect of ground state charges

b) Q_0' ➔Q_1 Polarization

Wanko et al. JPCB 2008 112, 11462.



Q_0 : MM charges

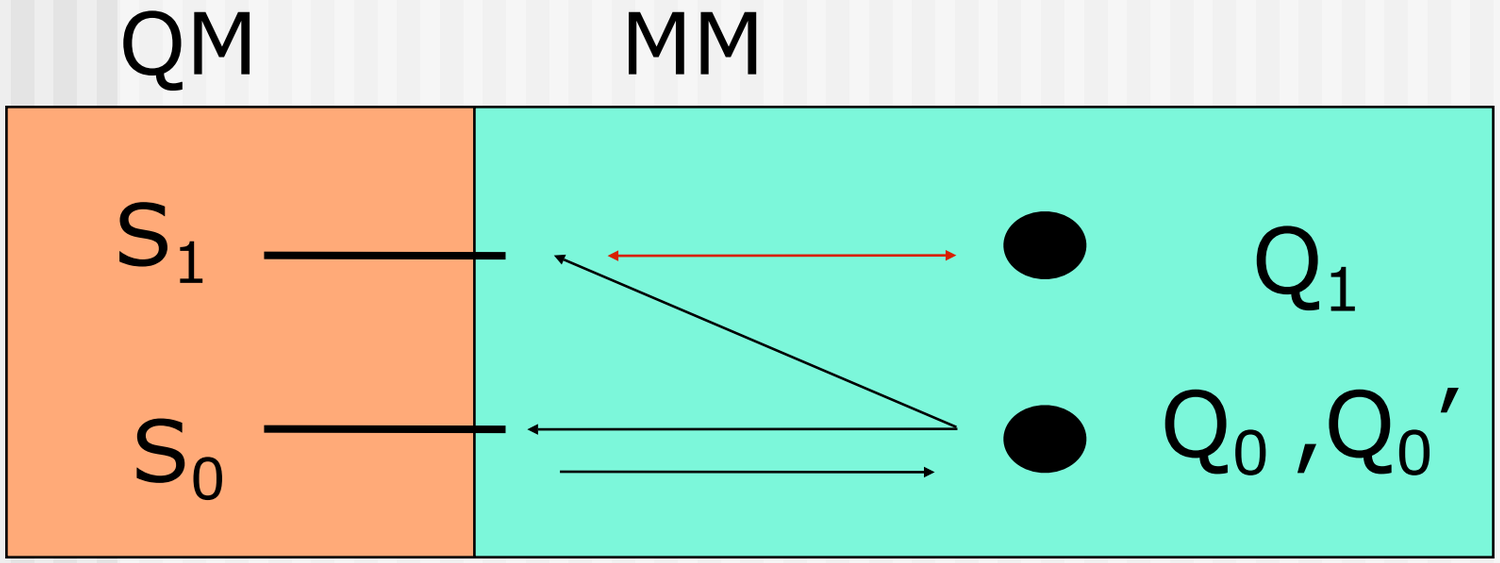
Q_0’: S_0 polarized charges

Q_1 : S_1 polarized charges

2 ways to calculate:

a) $Q_1 \longleftrightarrow S_1$ self-consistent for each state

$Q_0' \longleftrightarrow S_0$ (S_0 and S_1 not orthogonal)



Q_0 : MM charges

Q_0’: S_0 polarized charges

Q_1 : S_1 polarized charges

2 ways to calculate:

a) $Q_1 \longleftrightarrow S_1$ self-consistent for each state

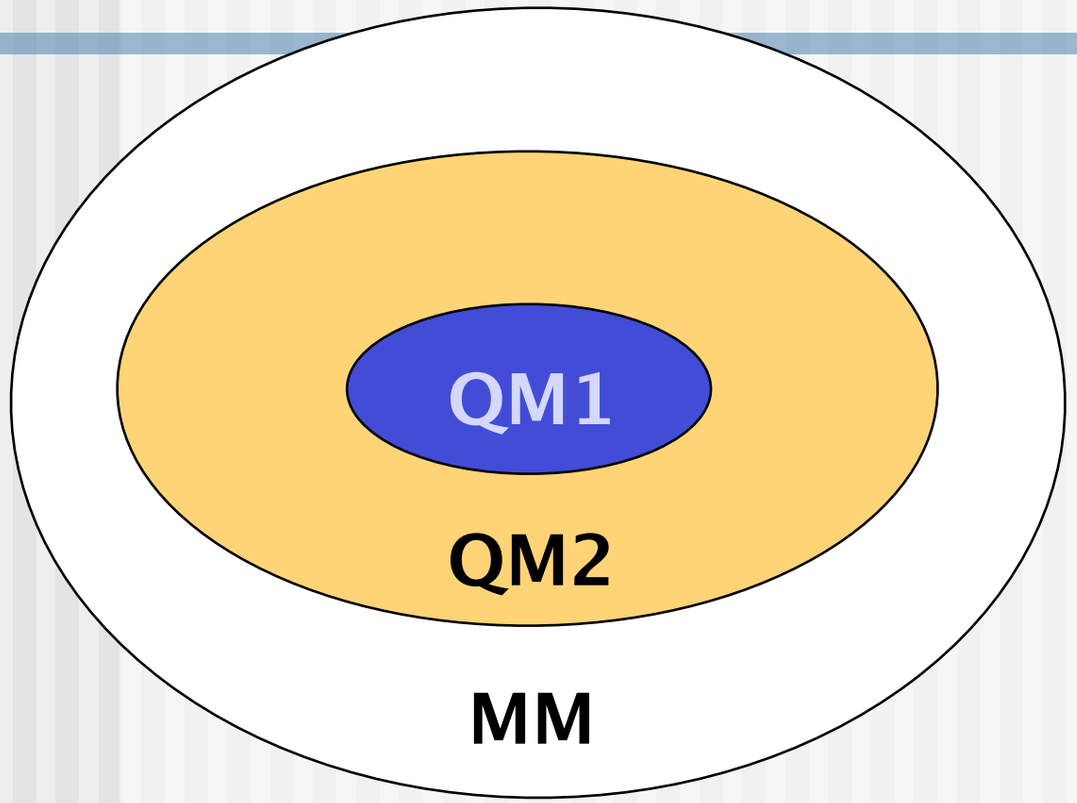
$Q_0' \longleftrightarrow S_0$ (S_0 and S_1 not orthogonal)

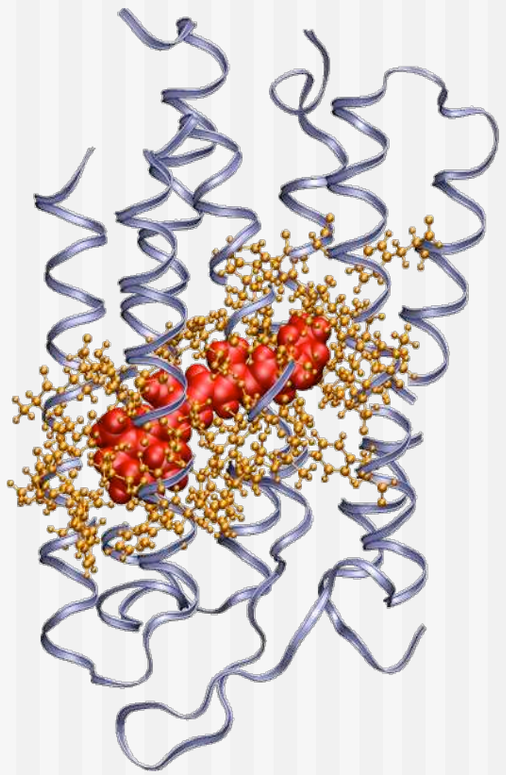
b) $Q_0' \longleftrightarrow S_0$ self-consistent for S_0

$Q_0' \rightarrow S_1$

$S_1 \rightarrow Q_1$ (S_1 not Q_1 polarized)

Polarization of Protein with QM/QM/MM:





only 300 atoms
polarizable region

2 Effects:

a) $Q_0 \rightarrow Q_0'$ effect of ground state charges **-0.05 eV**

b) $Q_0' \rightarrow Q_1$ Polarization **-0.05 eV**

total: -0.1 eV

Wanko et al. JPCB 2008 112, 11462.

Explicit Polarization Models

- fluctuating (point) charge models (FQ)
  - QM SCF $\longrightarrow$ ρ/point charges/multipoles
  - Chemical hardness models (e.g. SCC-DFTB, CHARMM-FQ)

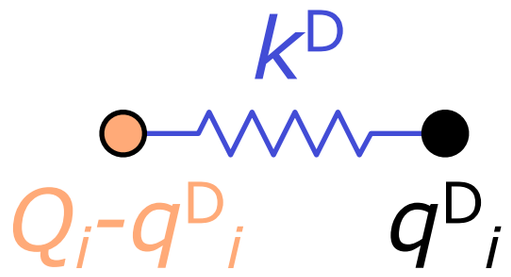
$$E^{\mathrm{ES}}(Q) = -\sum_i \mu_i Q_i + \tfrac{1}{2}\sum_{ij} \eta_{ij}(R_{ij},\eta_i,\eta_j) Q_i Q_j$$

- induced (atomic) dipole models
  - additive
  - interactive

$$\mu_i^{\mathrm{ind}} = \alpha_i \xi_i \left(M, \underline{\mu^{\mathrm{ind}}}\right)$$

$$E^{\mathrm{ES}}(M,\mu^{\mathrm{ind}}) = -\tfrac{1}{2}\sum_{i\neq j} M_i T_{ij} M_j - \sum_{i\neq j} \mu_i^{\mathrm{ind}} T_{ij}\left(M_j + \underline{\tfrac{1}{2}\mu_j^{\mathrm{ind}}}\right) + \tfrac{1}{2}\sum_i \frac{1}{\alpha_i}\left|\mu_i^{\mathrm{ind}}\right|^2$$

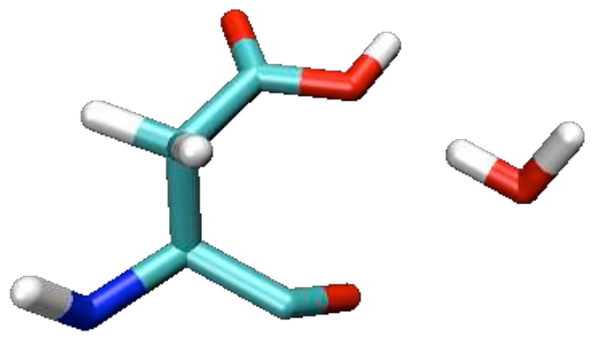
- Drude oscillator model



$$\alpha = \frac{q_{\mathrm{D}}^2}{k_{\mathrm{D}}}$$

Going Beyond Conventional Force Fields

Fixed point charges of conventional force fields
incorporate polarization *implicitly*:



- Fit to ESP of HF/6-31G* (Amber)
- Fit to water (TIP3P)—peptide interaction energies (CHARMM)

⇒bond dipoles larger than in gasphase
⇒balanced solvation, liquid phase

Problem: averaging over conformations and solvent polarities!

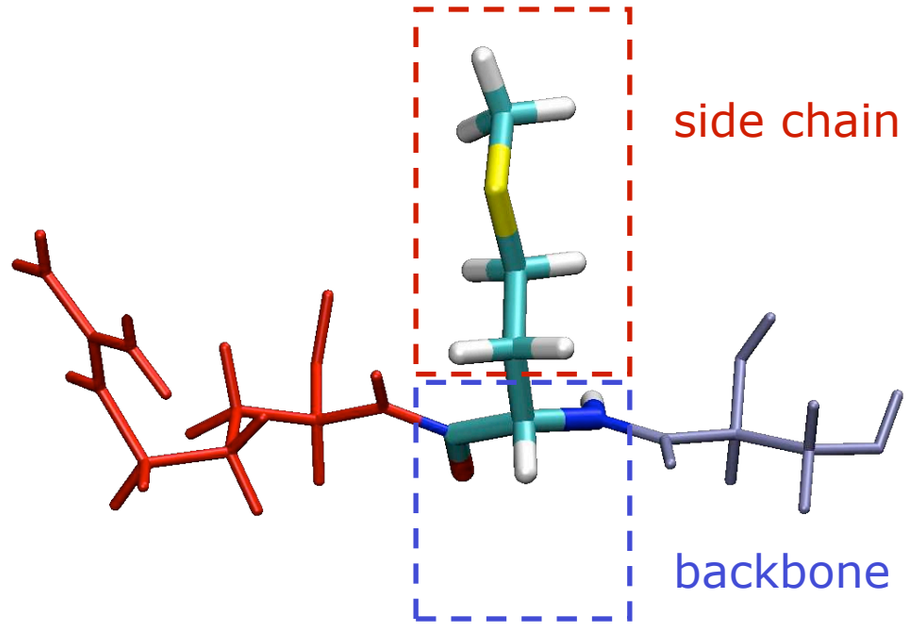
In particular retinal proteins:

- extended charge transfer ($\Delta\mu$=12 debye) due to S_0-S_1 excitation
- significant polarization of nearby aromatic residues

Induced Dipole Model: "polar.h"

- Splitting into polarization groups

- Obtain "unpolarized" charge model to avoid "double counting"
  - side chains: restraint ESP fit (RESP) to B3LYP/6-311G(2d,2p)
  - backbone: multi-configurational RESP fit ≅ CHARMM charges



- Interactive model

- Thole´s short-range damping scheme

$$\rho = \frac{3a}{4\pi}\exp\left(-au^3\right) \qquad u = R_{AB}/(\alpha_A\alpha_B)^{1/6}$$

  => damping parameter *a*=0.39

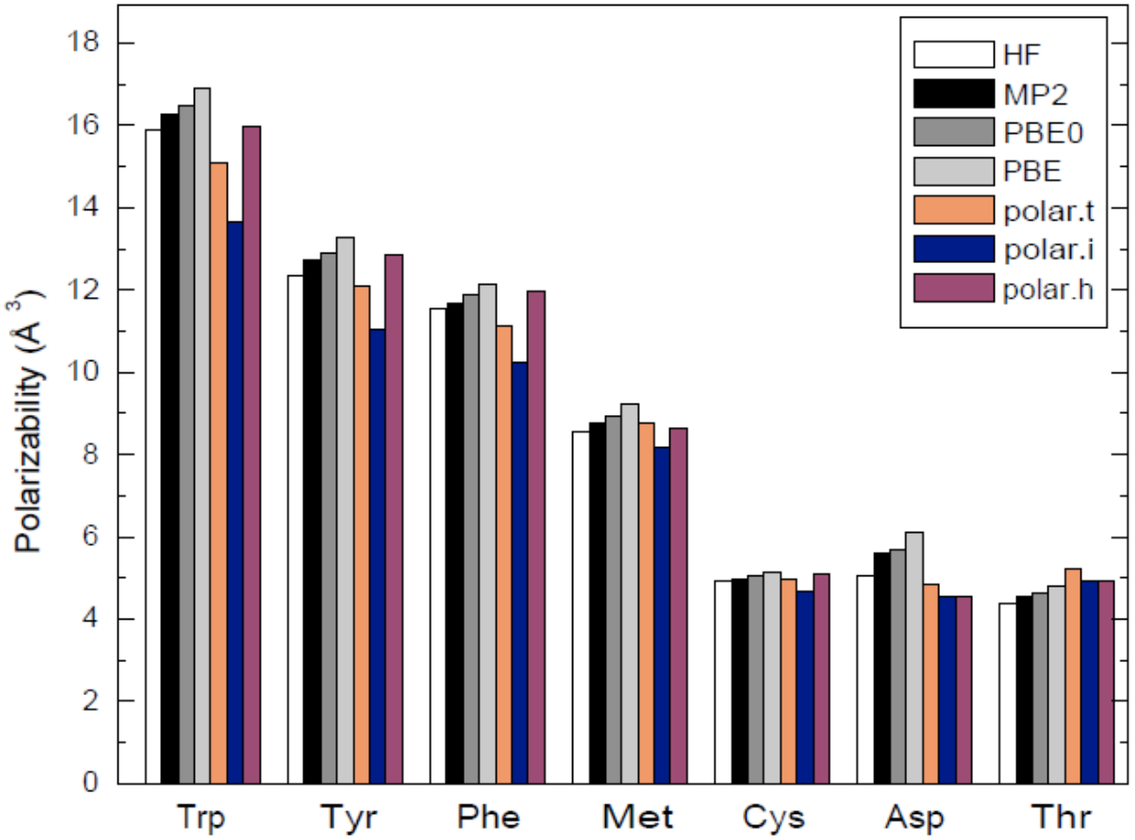
- Iterative relaxation of QM density and μ^{ind}

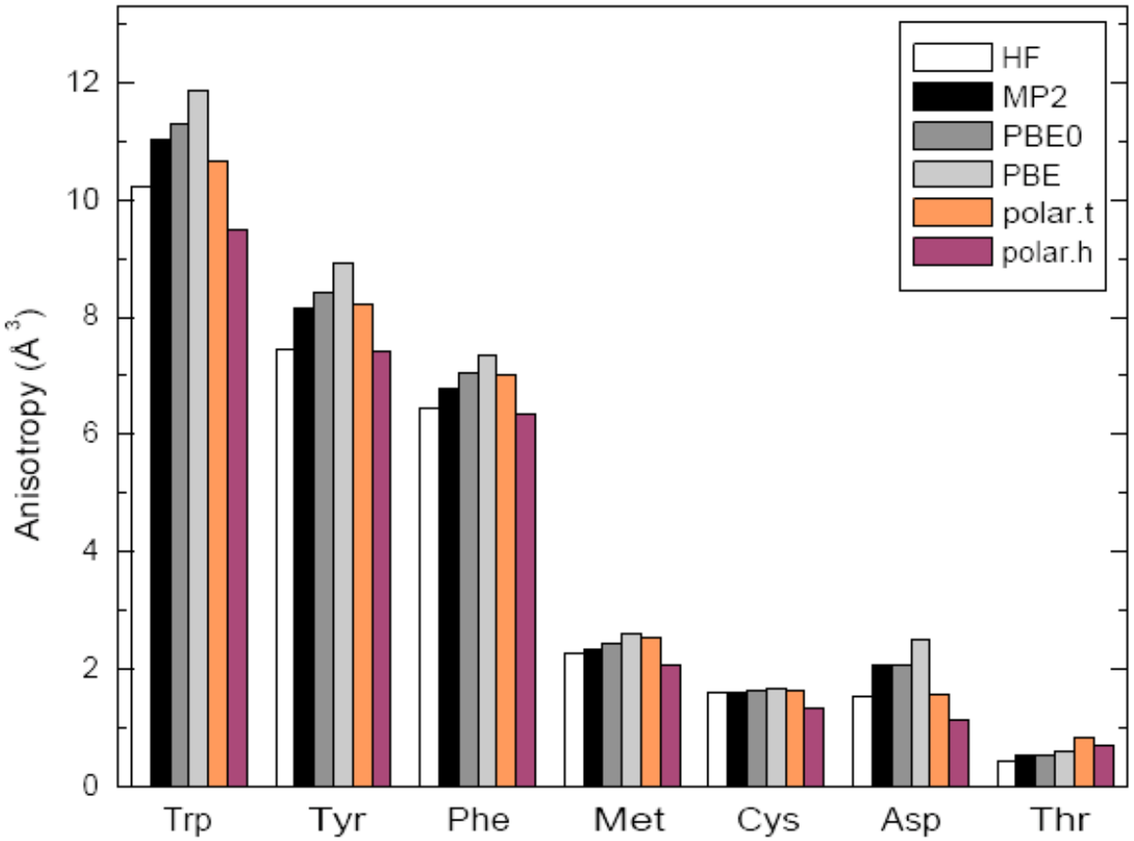
Atomic polarizability parameters (Å^3)

C	1.334/1.720[a]
N	1.073
O	0.837
S	2.440
H	0.496

[a] sp2 carbon in Trp, Tyr

Protein Polarization Models



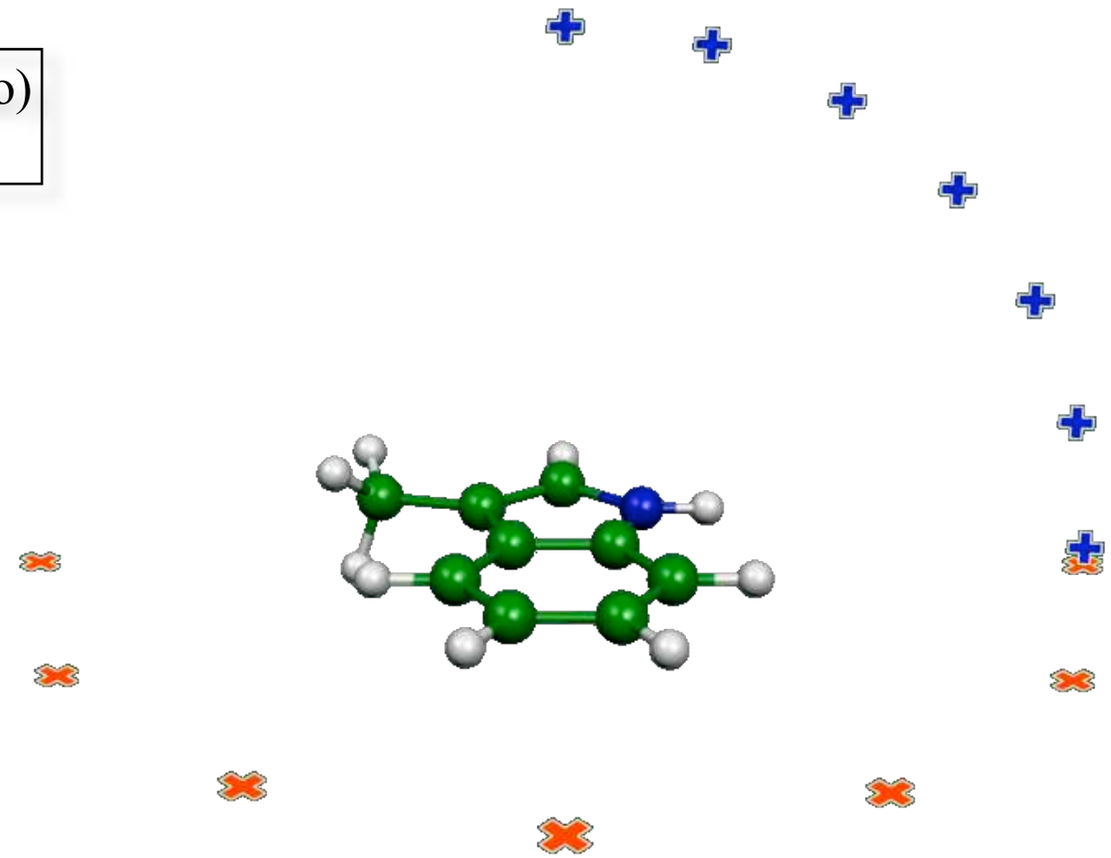


Mean α (17 AA)	PBE0	polar.t	polar.i	polar.h
RMSD rel. dev. (%)	1.6	9.5	9.5	7.3
~ (neutral only)	1.7	6.9	7.6	3.0
MSD rel. dev. (%)	1.5	3.3	-3.6	-0.3

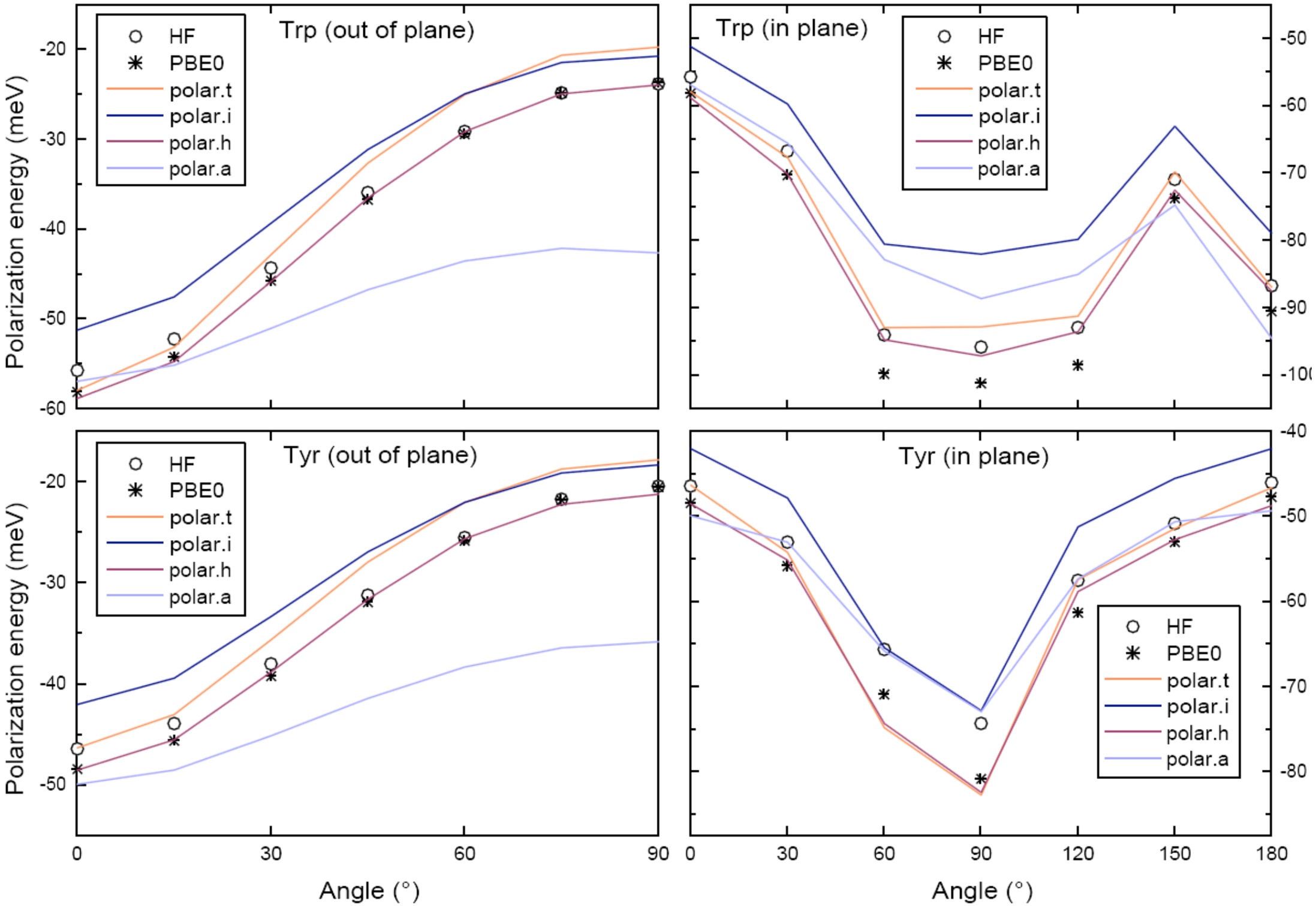
Reference: MP2/cc-pVQZ

Polarization Energies: QM vs. polar

$$E_{\mathrm{pol}} = E_{\mathrm{tot}}^{(\mathrm{PC})} - E_{\mathrm{tot}}^{(\mathrm{vacuo})} - q_{\mathrm{PC}}\Phi^{(\mathrm{vacuo})}$$



Polarization Energies: QM vs. polar



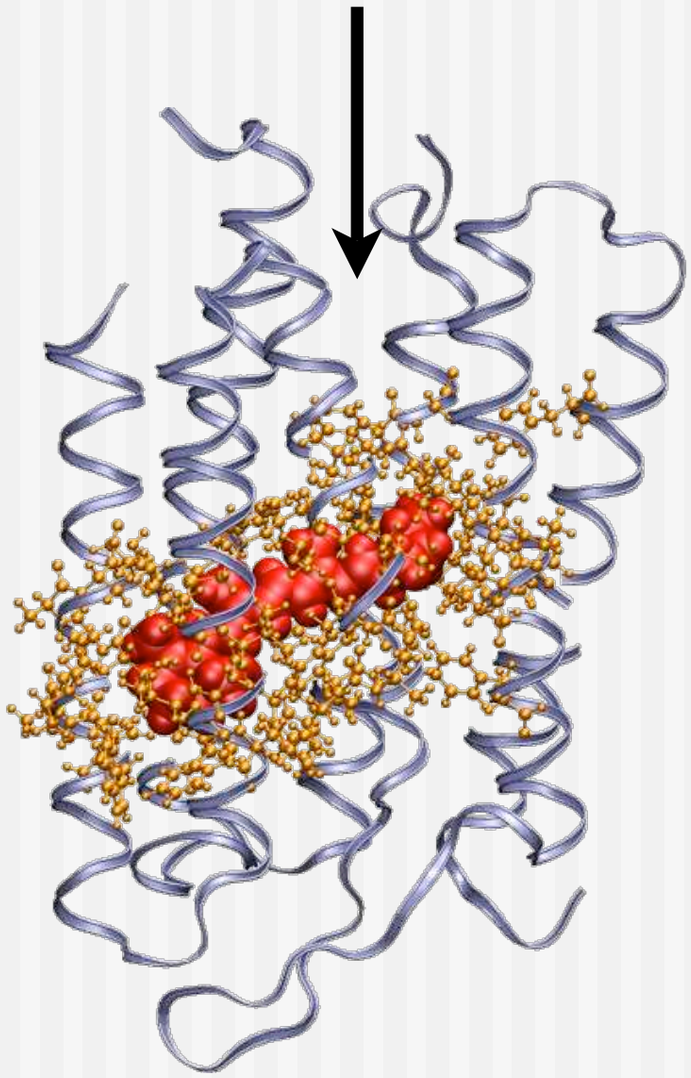
Comparison of QM/QM/MM and QM/MMpol

	QM1	QM2	ΔE^{QM1}	$\Delta E^{(ii)}$
bR				
	SORCI	polar.h	2.26	2.21
	SORCI	PBE0[b]	2.27	2.23
psRII				
	SORCI	polar.h	2.51	2.47
	SORCI	PBE0[b]	2.50	2.48

ground state polarized charges

excited state polarized

only 300 atoms
polarizable region

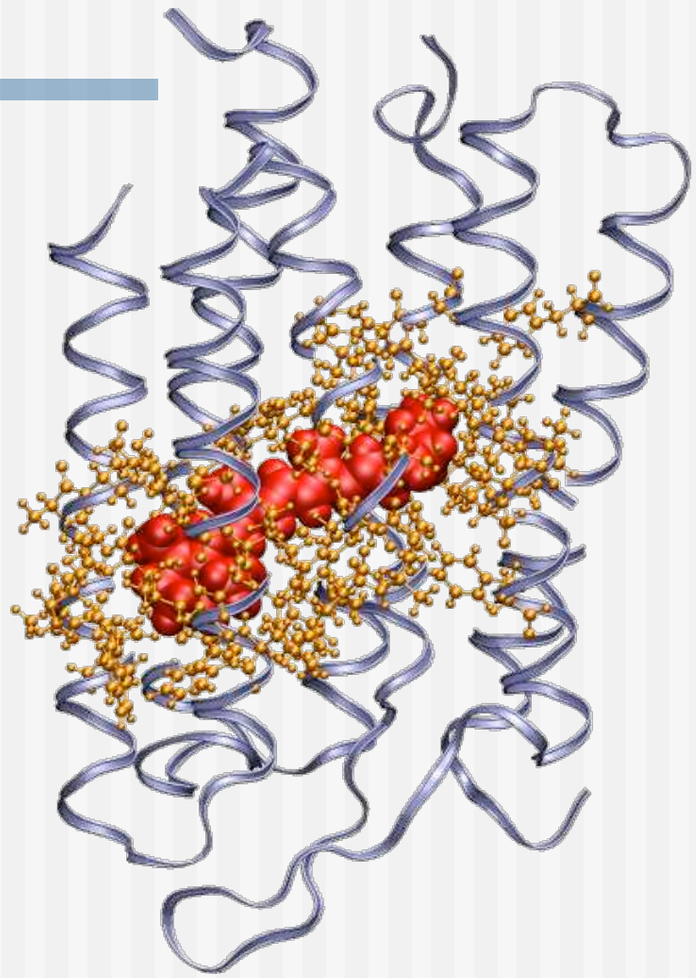


- good agreement of QM/QM with MMpol
- polarization red-shift of about 0.04 eV

Wanko et al. JPCB 2008 112 11468.

QM/MMpol for full protein

	state	MM	ΔE^{QM1}	$\Delta E^{(ii)}$	$\Delta E^{(ii)}_{pol}$
bR					
	S1	polar.h	2.24	2.16	−0.08
	S2	polar.h	2.47	2.45	−0.01
psRII					
	S1	polar.h	2.48	2.42	−0.07
	S2	polar.h	2.53	2.52	−0.01
Rh(u)					
	S1	polar.h	2.43	2.36	−0.07
	S2	polar.h	2.60	2.60	−0.01
Rh(p)					
	S1	polar.h	2.32	2.26	−0.06
	S2	polar.h	2.58	2.57	−0.01



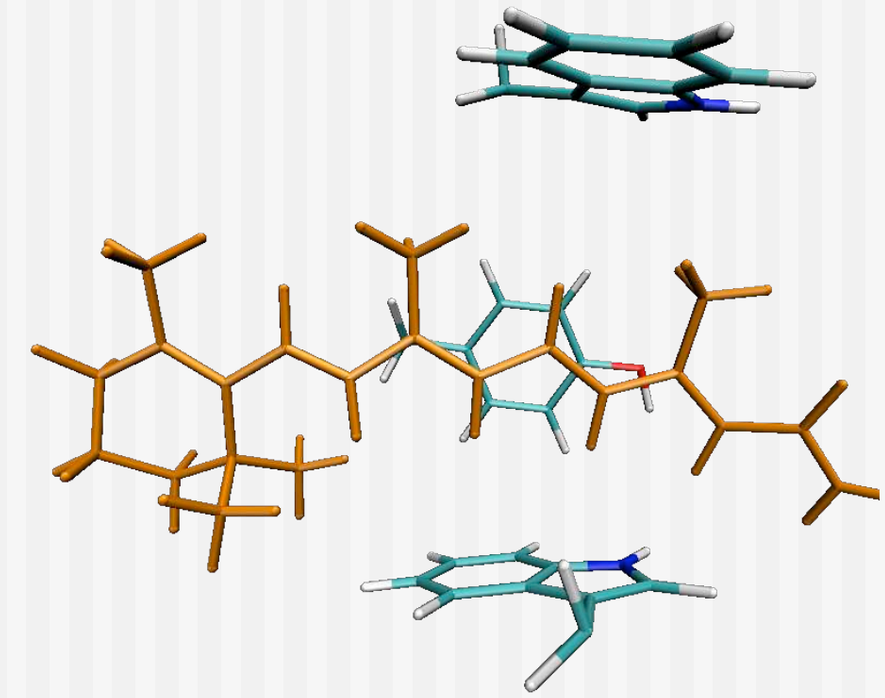
- ground state MMpol effect: -0.07
- polarization red-shift of about -0.07 eV
- **total effect of polarization: -0.14 eV**

Wanko et al. JPCB 2008 112 11468.

Role of dispersion

SORCI dispersion red shift (eV) in bR

arenes	protein	$\omega_{vertical}$	ΔE_{tot}	shift
CHARMM	CHARMM	2.32		
polar.h	CHARMM	2.22	2.19	
SORCI	CHARMM	2.12	-0.07	



Effects not covered in standard QM/MM:

optimal ground state charges: 0.07
effect of polarization : 0.07
Dispersion: 0.07

total redshift: 0.2 eV

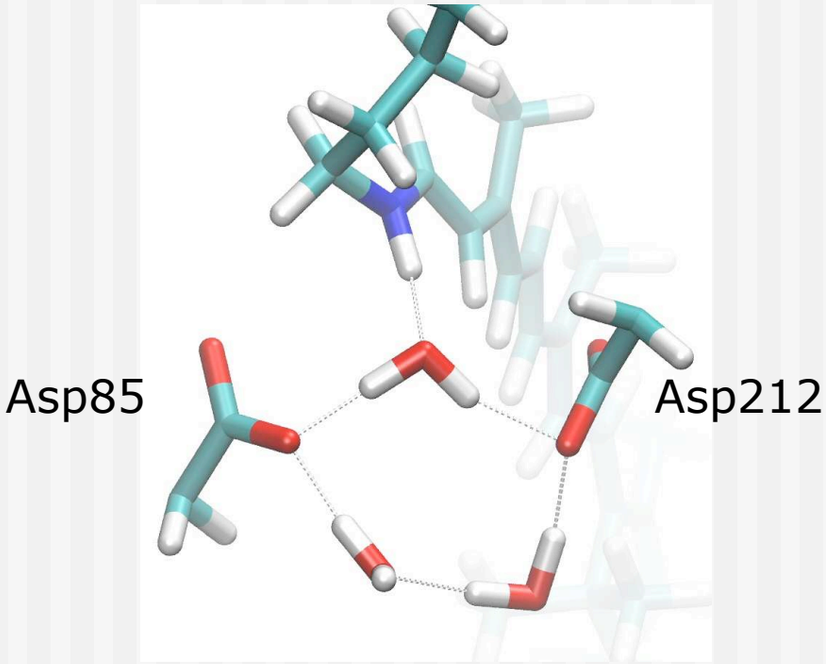
Dispersion contribution:

0.07 eV in bR/SRII

0.01 eV in Rh

Extending the QM Zone

bR QM zone (add.)	S_1 (eV)
Ret	2.32
+Asp85,W2	2.48
+Asp212,W1,W6	2.48

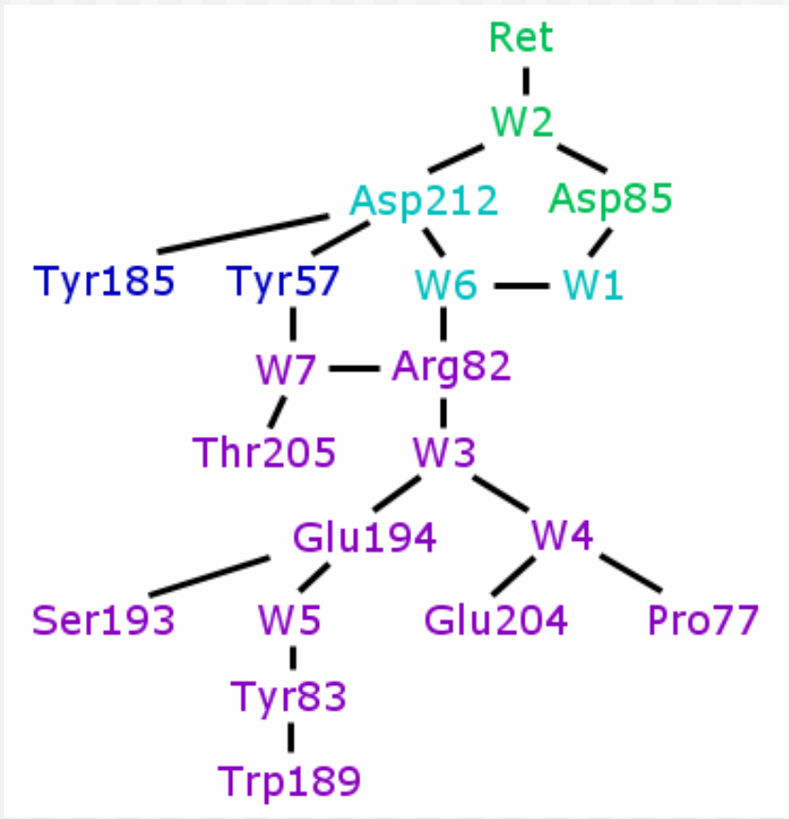


What is the optimal QM-zone?

- HF NPA charges: net charge on Ret
  → converged (qm2 or qm4)
  → 0.16 eV blue shift (Rh: only 0.03 eV)

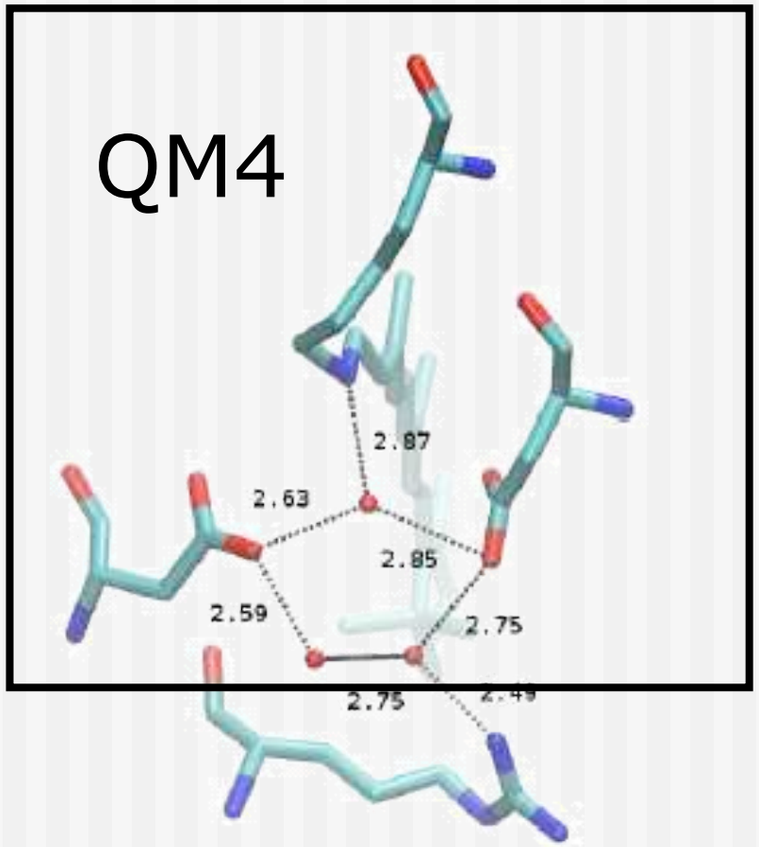
- ESP at Ret atoms (conjugated backbone)
  → no convergence (polarization!)

	qm2	qm4	hbn3	hbn8
Ret	0.956	0.949	0.946	0.942
Asp75/85	-0.94	-0.89	-0.85	-0.85
Asp201/212		-0.93	-0.84	-0.84



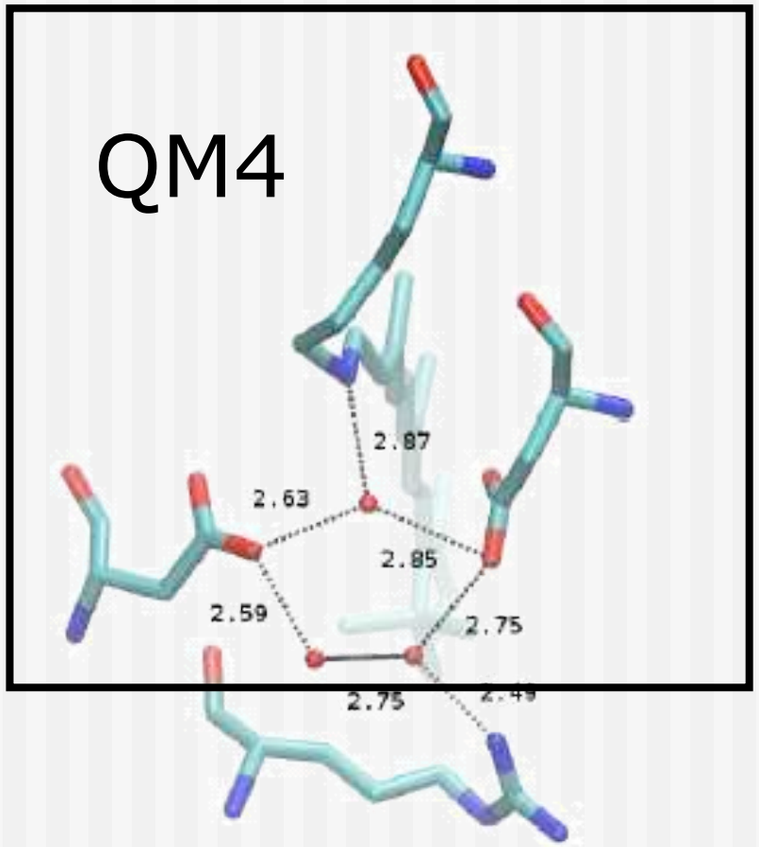
‘post-QMM’ corrections for bR

	calc		exp
vacuum	1.89		2.0
MM-charges	2.32	+0.43	
MMpol		-0.14	
CT		+0.16	
QM4-pol	2.23		
disp	2.16	-0.07	2.18



‘post-QMM’ corrections for bR

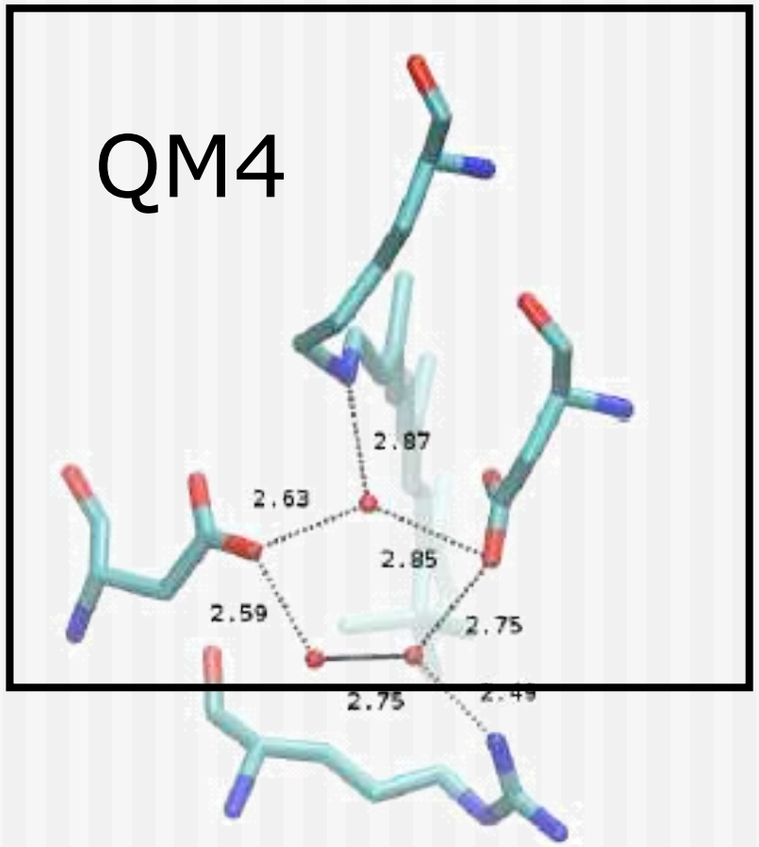
	calc		exp
vacuum	1.89		2.0
MM-charges	2.32	+0.43	
MMpol		-0.14	
CT		+0.16	
QM4-pol	2.23		
disp	2.16	-0.07	2.18



NOTE: in different proteins (bR, Rh, SRI)
- polarization effects quite similar
- CT shifts very different
- dispersion shifts different (Rh: -0.01)

‘post-QMM’ corrections for bR

	calc		exp
vacuum	1.89		2.0
MM-charges	2.32	+0.43	
MMpol		-0.14	
CT		+0.16	
QM4-pol	2.23		
disp	2.16	-0.07	2.18



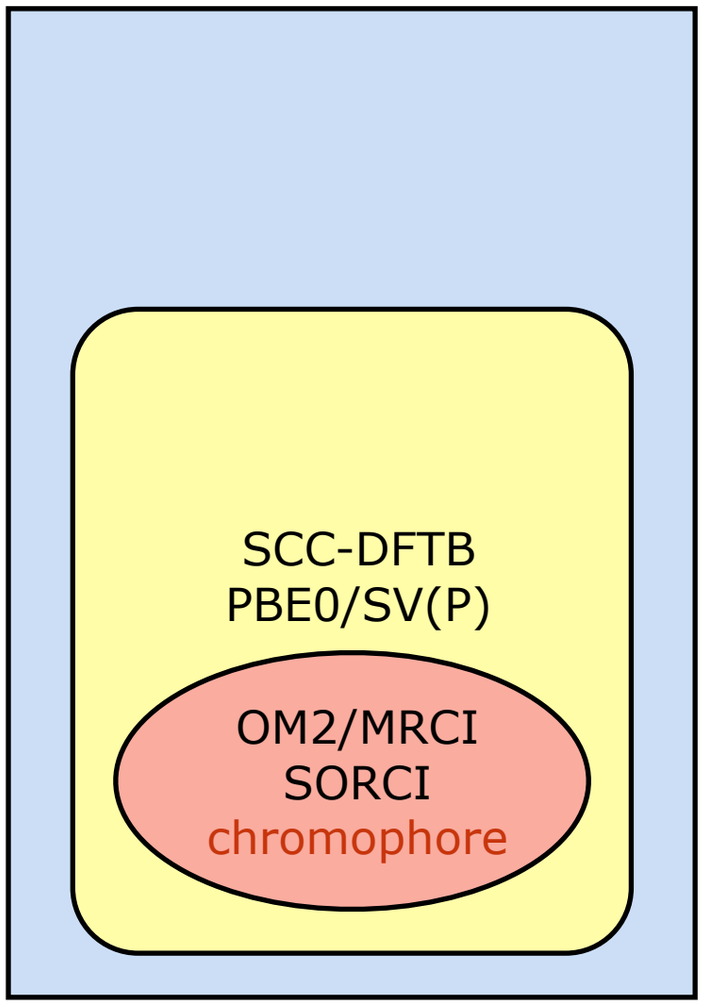
NOTE: in different proteins (bR, Rh, SRI)
- polarization effects quite similar
- CT shifts very different
- dispersion shifts different (Rh: -0.01)

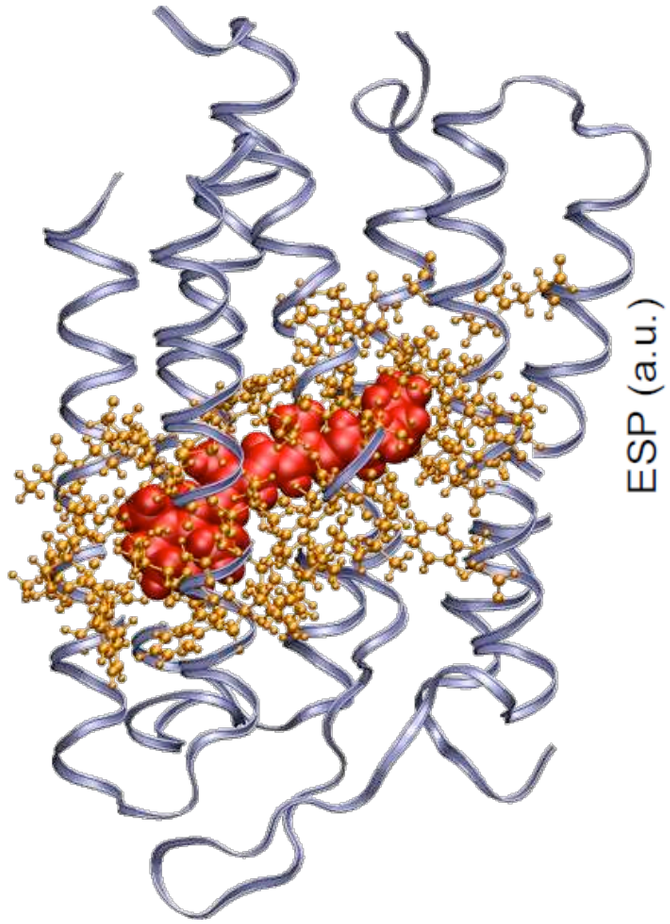
TODO:
- calculations are based on SCC-DFTB geometries
  => bond length alternation underestimated

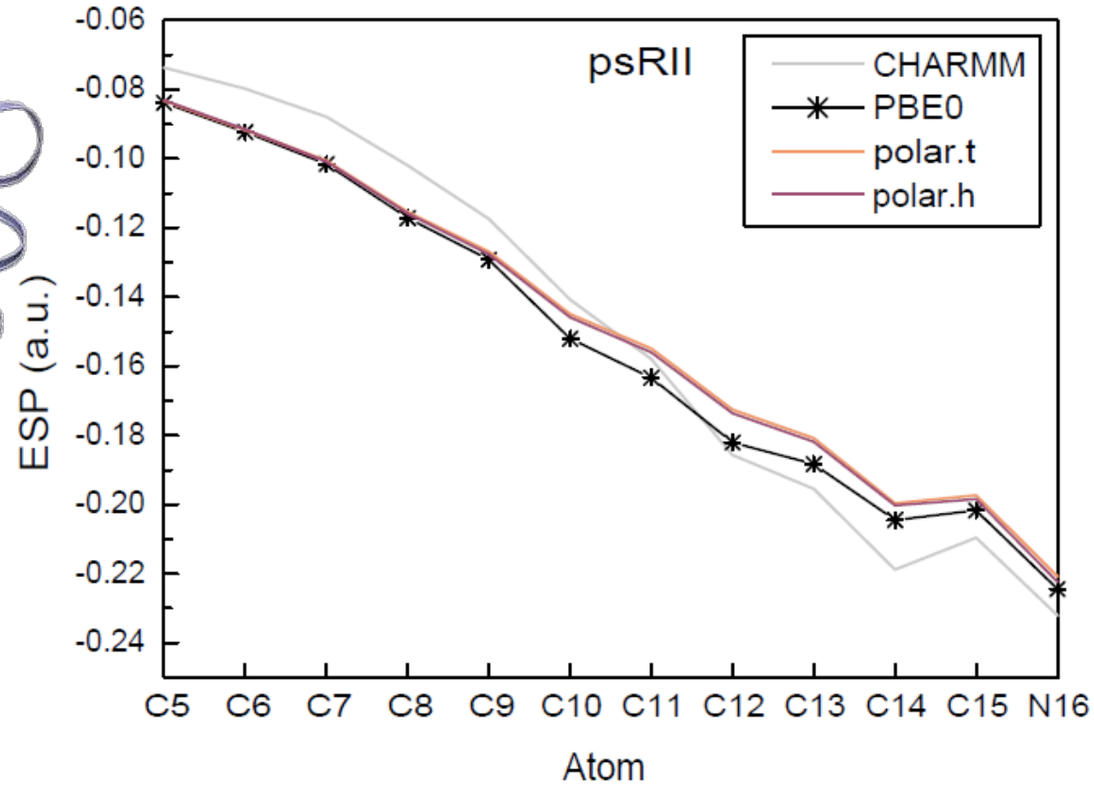
=> blue shift

Protein Polarization Models

QM/QM/MM model
(iterative solution)







Ret	Binding pocket	Protein (rest)	bR	psRII
SORCI	CHARMM	CHARMM	2.32	2.56
SORCI	PBE0	CHARMM	2.24	2.47
SORCI	polar.h	CHARMM	2.22	2.47

$$|\Psi\rangle = |\Psi^{(QM1)}\rangle|\Psi^{(QM2)}\rangle$$

$$E_I^{\mathrm{tot}} = E_I^{\mathrm{QM1}}[\rho_2] + E_0^{\mathrm{QM2}}[\rho_1] - V_{QM1-QM2}^{\mathrm{ES}}$$

Basic literature:

- Jensen: Introduction to Computational Chemistry, Wiley and Sons.
- Koch, Holthausen: A Chemist's Guide to Density Functional Theorey, Wiley.
- Parr, Yang: Density Functional Theory of Atoms and Molecules, Oxford.

The wavefunction is the central concept of quantum mechanics, since it describes the quantum mechanical state.

- it determines all other properties through the calculation of expectation values.
- However, it is not a measurable quantity.
- It is a function of 3 N coordinates. Is such a detailed information required or is this an 'information overkill'?.
- It becomes an increasingly complex task to construct better wavefunctions:
  orbital $\phi(r) \rightarrow$ product Ansatz $\rightarrow$ Slater determinant $\rightarrow$ CI!

In contrast, the electron density

$$\rho(r)$$

- is an observable, can be determined e. g. by X-ray.
- is a function of three coordinates (x,y,z).
- it can be shown, that the information of 3N coordinates is NOT required to calculate the desired expectation values.

Density and wavefunction

$$\rho(r_1) = N \int |\Psi(r_1 ... r_N)|^2 dV_2 ... dV_N$$

However, this is not the way to go, since the determination of the true N-particle wf is the complicated task!

- Can we determine the density directly?
- Can we get an energy depending on the density only $E[\rho]$?

==> this would be a Density-Functional: 'Function of a function'

==> how to determine? Need energy functional and then 'minimize' as in HF: Variational principle

==> but most important question: is the density an unique feature of a certain system? I.e., are the densities coming from different external potentials (= core potentials in QC) different? Only then, the energy of a system can be uniquely determined by the density!

Hohenberg and Kohn (HK) Theorems

HK1: the map G: v(r) $\longrightarrow \rho(r)$ is invertible.

I.e. there is a one-to-one correspondence of potential and density, therefore, it is uniquely defined through the external potential. Since the potential uniquely determines the wf and the wf the expectation values, this theorem assures that any quantum mechanical observable is completely determined by the density.

Hohenberg and Kohn (HK) Theorems

HK2: There exists a functional $E[\rho]$ with (ρ_0: ground state density):

$$E[\rho] \geq E_0,$$

$$E[\rho_0] = E_0$$

Therefore, the derivative:

$$\frac{\delta E[\rho]}{\delta \rho} = 0$$

results in an equation, from which the ground state density can be determined.

Total energy functional

$$E[\rho] = T[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho],$$

- $J[\rho]$ Hartree energy.

$$J[\rho] = \frac{1}{2} \int \frac{\rho(x_1)\rho(x_2)}{x_1 - x_2} dx_1 dx_2$$

und

$$E_{en}[\rho] = \sum_{\alpha} \int \frac{Z_\alpha \rho(x)}{R_\alpha - x} dx$$

- $E_{xc} = E_x + E_c$: exchange-correlation (XC) energy functional.
- T: kinetic energy functional

Thomas and Fermi (1927)

$$T_{TF}[\rho] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho^{5/3}(x)dx \tag{1}$$

von Weizäcker (1935)

$$T[\rho] \approx T_{TF}[\rho] + \frac{1}{81} \int \frac{|\nabla\rho(x)|^2}{\rho(x)} dx \tag{2}$$

Accuracy of kinetic energy functionals: introduce orbitals again and evaluate kintic energy as:

$$T_s = -\frac{1}{2}\sum_i < \phi_i|\nabla^2|\phi_i > . \tag{3}$$

==> will be evaluated from Slater determinant as in HF
difference of exact T and T_s 'moved' into E_{xc}
DFT is a single determinal method: fails in multi-reference cases

Kohn-Sham (KS) Theorem: non-interacting electrons

Let ρ_0 be the true ground state density of the interacting electrons.

- Then there exists a potential $v_{eff}[\rho_0]$ for the non-interacting electrons, leading to the same density ρ_0 via solution of the KS equations:

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}[\rho]\right]\phi_i = \epsilon_i\phi_i, \qquad \rho_0(r) = \sum_i |\phi_i|^2$$

- KS effective potential: $v_{eff}[\rho] = \frac{\delta E_{pot}}{\delta \rho}$,

$$v_{eff}[\rho] = \sum_\alpha \frac{Z_\alpha}{R_\alpha - r} dr + \frac{1}{2}\int \frac{\rho(r')}{r - r'} dr' + v_{xc}[\rho]$$

- XC-potential: $v_{xc}[\rho] = \frac{\delta E_{xc}}{\delta \rho}$

XC functionals

- **LDA** (Local Density Approximation: from electron gas):

$$E_x = C \int \rho^{4/3}(r) d^3 r$$

- **GGA** (Generalized Gradient Approximation):

$$E_x = C \int \rho^{4/3}(r) F(s) d^3 r, \qquad s = \frac{\nabla \rho}{\rho^{4/3}}$$

  Various approximations for X and C:
  BP, BLYP and PBE being the most popular.

- **Hybride Functionals**:

$$E_x^h = (1 - c) E_x^{GGA} + c E_x^{HF}$$

  Usually, 20-30% HF-X work well (B3LYP: c=0.2)

Performance

- LDA, GGA ...:
  Accuracy for Geometries, vib. frequencies quite good.
- Energies: LDA, some GGA's show severe overbinding $\rightarrow$ hybrid functionals.

See Koch/Holthausen for more details.

Probems due to the approximate nature of the functionals:

- Self-interaction error (SIC)
- asymptotics of v_{xc}
- 'near-sightedness' of E_{xc}

Asymptotics of v_{xc}: Fast decay of LDA (GGA...) exchange potential

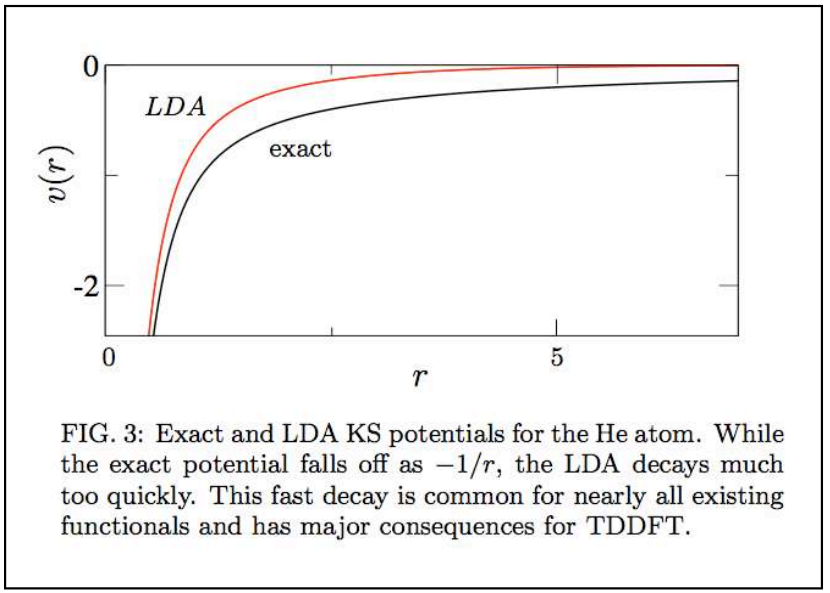


FIG. 3: Exact and LDA KS potentials for the He atom. While the exact potential falls off as $-1/r$, the LDA decays much too quickly. This fast decay is common for nearly all existing functionals and has major consequences for TDDFT.

Figure: Elliot, Burke, Furche: arXiv:cond-mat/0703590v1 2007

Eigenvalue spectrum quantitatively incorrect:

- Ionization threshold too low.
- Rydberg states unbound (underestimated).

LDA and GGA are local functionals,

however, should be non-local as e.g. HF exchange:

- Locality: consider two weakly interacting fragments:

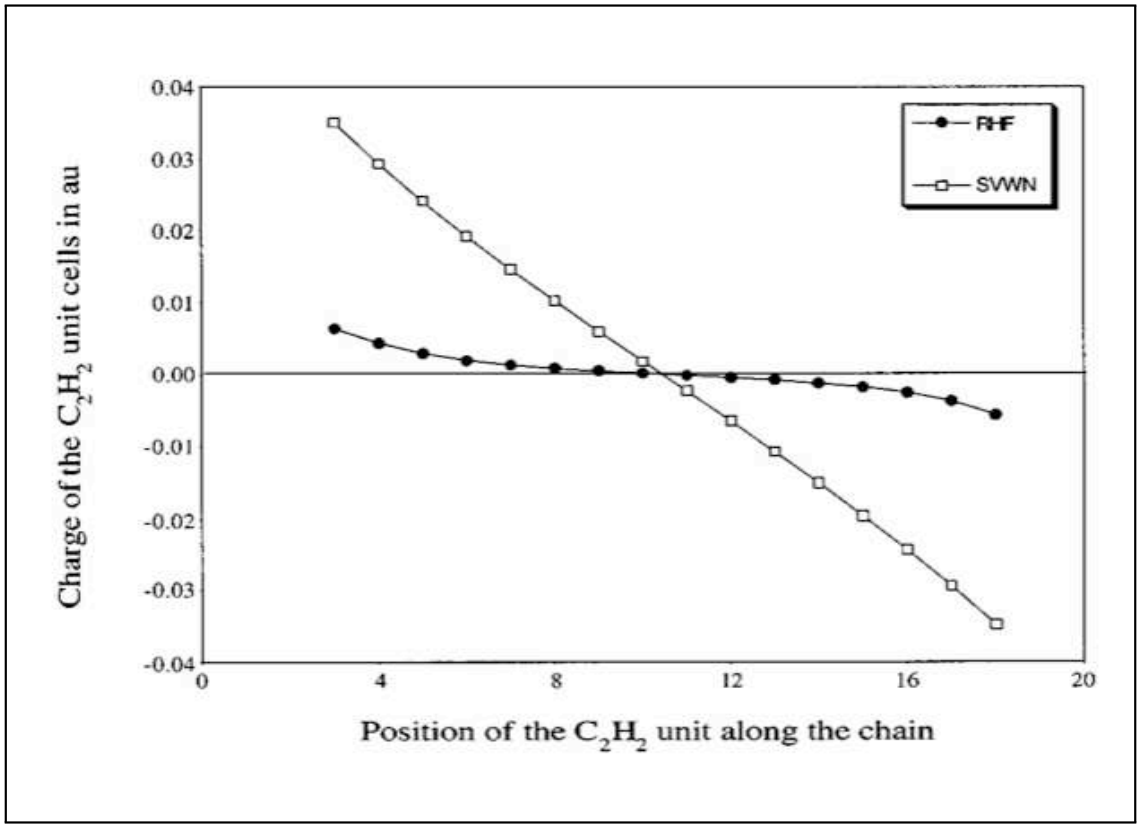
  $$\rho = \rho_1 + \rho_2$$

  Then, if the densities do not overlap, the local functionals vanish:

  $$E_{xc} = 0$$

- This is in particular a problem for VdW interactions, which DFT-GGA is not able to handle.
  (hybrides change only E_x, but E_c is the problem here!)
  (see e.g. JCP114 (2001) 5149)

Overestimation of polarization in extended conjugated chains (Champagne et al. JCP 109, 10489) due to 'short-sightedness' (locality) of E_x.



Overestimation of polarizabiliy (Champagne et al. JCP 109, 10489) due to 'short-sightedness' of E_x.

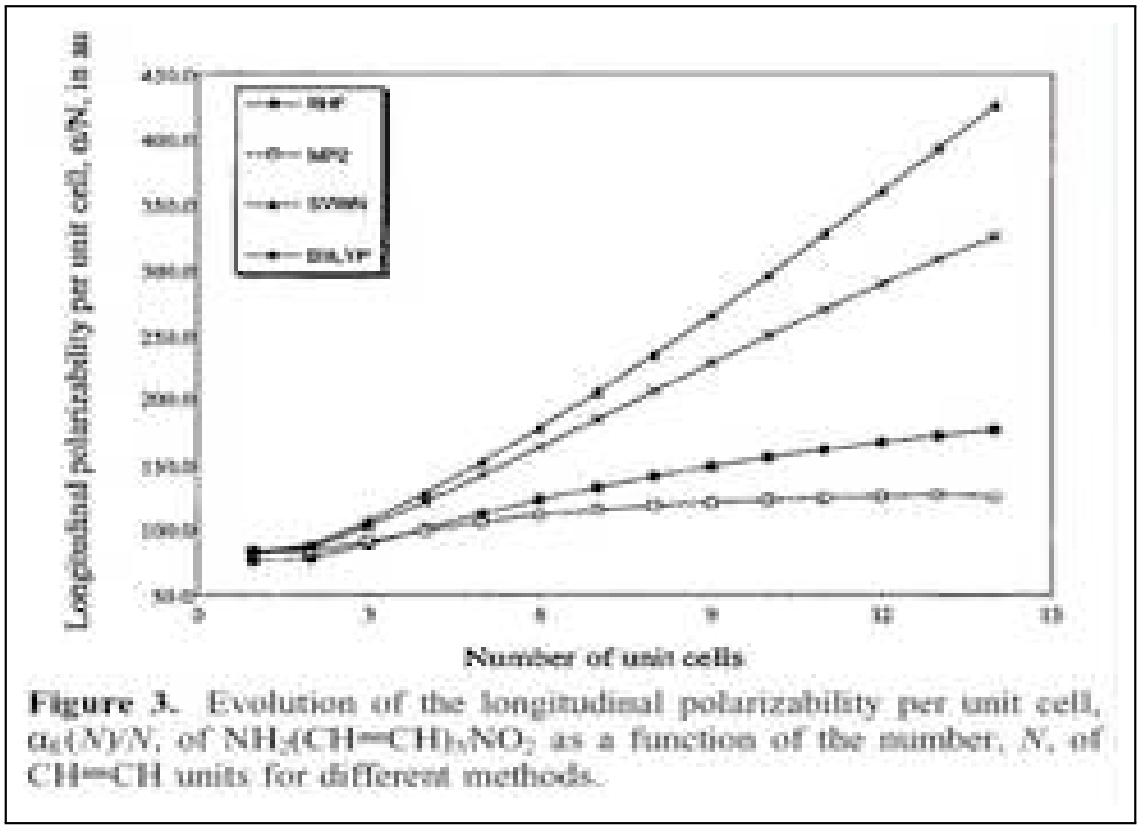
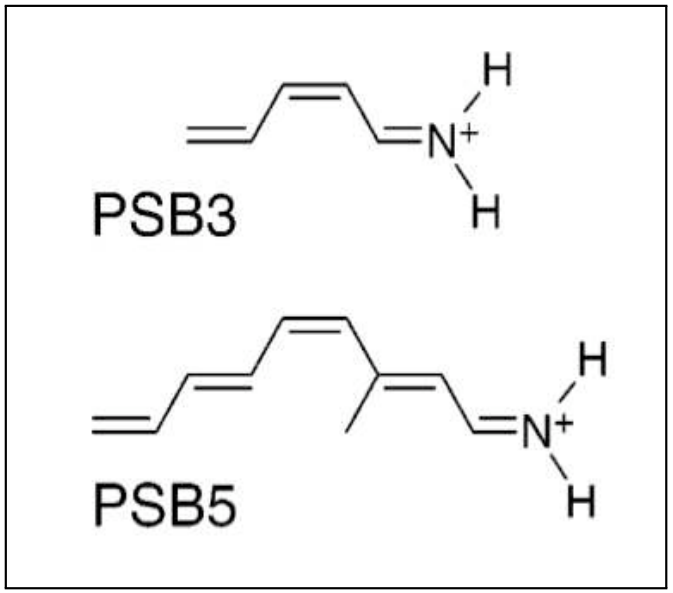
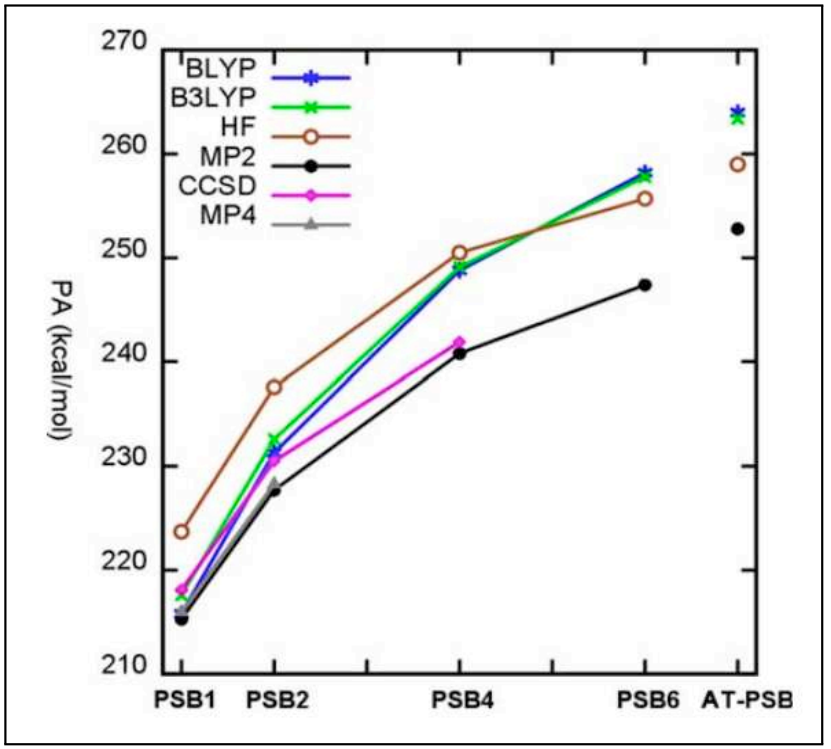


Figure 3. Evolution of the longitudinal polarizability per unit cell, $\alpha_L(N)/N$, of $NH_2(CH{=}CH)_NNO_2$ as a function of the number, N, of CH=CH units for different methods.

Figure: Gritsenko, Champagne,Gisbergen, Baerends

This has severe implications for many properties, e.g. proton affinities:
(J. Computer-Aided Mol. Design, **20** (2006) 511)





And will be of particular relevance for charge transfer excitations in TDDFT.

- Phys. Rev. B 51, (1995) 12 947 (DFTB)
- Phys. Rev. B 58 (1998) 7260. (SCC-DFTB)
- J. Chem. Phys. 122 (2005) 114110, J. Phys. Chem. A (2006), 110, 13551, J. Phys. Chem. A 2007, 111, 5751 (performance)
- J. Chem. Phys. 114 (2001) 5149 (VdW interactions in DFT)
- Phys. Rev. B 63 (2001) 5108 (TD-DFTB)
- J. Phys. Chem. B 105 (2001) 569 (QMMM)
- PROTEINS 44 (2001) 484 (O(N))
- phys. stat. sol. (b) 217 (2000) 41 and 357, J. Phys. : Condens. Matter 14 (2002) 3015. (reviews)
- J. Phys. Chem. B 110 (2006) 6458. ('multiscale')
- J. Phys. Chem. A, 111 (2007) 5655. (third order)
- Theor Chem Acc (2006) 116: 316 (bio-review)

SCC-DFTB: self-consistent charge density -functional tight-binding

Second order expansion of the DFT total energy functional with respect to the charge density fluctuations $\delta\rho$ around a given reference density ρ_0 ($\rho_0' = \rho_0(\vec{r}')$, $\int' = \int d\vec{r}'$):

$$
\begin{aligned}
E &= \sum_{i}^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \iint' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\, \delta\rho'} \right|_{n_0} \right) \Delta\rho\, \Delta\rho'. \\
&- \frac{1}{2} \iint' \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0] n_0 + E_{cc}
\end{aligned}
$$

Reference density: $\rho_0 = \sum \rho_0^\alpha$
Superposition of atomic densities

Matrix elements

LCAO basis $\Psi_i = \sum c_\mu^i \eta_\mu$:

$$\langle \Psi_i | \hat{H}^0 | \Psi_i \rangle = \sum c_\mu^i c_\nu^i H_{\mu\nu}^0$$

$$H_{\mu\nu}^0 = \begin{cases} \epsilon_\mu & : \quad \mu = \nu \\ \langle \phi_\mu | H_{KS}[\rho_A + \rho_B] | \phi_\nu \rangle & : \quad \mu \in A, \nu \in B \\ 0 & : \quad \text{otherwise} \end{cases}$$

- 'Special' minimal basis set ϕ_μ and initial densities from atomic KS eqs.
- $H_{\mu\nu}^0$ and $S_{\mu\nu}$ calculated and stored $\rightarrow$ no integral evaluation during program runtime.

Second order terms

$$\frac{1}{2}\iint' \left(\frac{1}{|\vec{r}-\vec{r}'|} + \left.\frac{\delta^2 E_{xc}}{\delta\rho\,\delta\rho'}\right|_{n_0} \right) \Delta\rho\,\Delta\rho'.$$

- Monopole approximation: $\Delta\rho = \sum_\alpha \Delta\rho_\alpha \approx \sum_\alpha \Delta q_\alpha F_{00} Y_{00}$
- second derivative $\longrightarrow \gamma_{\alpha\beta}$

$$\frac{1}{2}\sum_{\alpha\beta} \gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta$$

- This approximation will also be used for TD-DFTB

Repulsive energy term

$$-\frac{1}{2}\iint' \frac{\rho_0'\rho_0}{|\vec{r}-\vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]n_0 + E_{cc} \rightarrow \sum_{\alpha\beta} U_{\alpha\beta}$$

SCC-DFTB total energy:

$$E = \sum_i \sum_{\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2}\sum_{\alpha\beta} \gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta + \sum_{\alpha\beta} U_{\alpha\beta}$$

Semi-empirical /approximate methods

approximation, neglect and parametrization of interaction integrals from ab-initio and DFT methods

-HF-based:

CNDO, INDO, MNDO, AM1, PM3, MNDO/d, OM1,OM2

-DFT-based:

SCC-DFTB,

DFT- 3-center- tight binding (Sankey)

Fireballs --- > Siesta DFT code

~ 1000 atoms, ~ ns MD

Approximate density-functional theory: SCC-DFTB

Self consistent - charge density functional tight-binding

www.dftb.org

- Seifert (1980-86): Int. J. Quant Chem., **58**, 185 (1996).
  O-LCAO; 2-center approximation: *approximate DFT*
  http://theory.chm.tu-dresden.de

- Frauenheim et al. (1995): Phys. Rev. B **51**, 12947 (1995).
  efficient parametrization scheme: *DFTB*
  www.bccms.uni-bremen.de

- Elstner et al. (1998): Phys. Rev. B **58**, 7260 (1998).
  charge self-consistency: *SCC-DFTB*
  www.tu-bs.de/pci

→ *approximate DFT*

DFTB is derived from DFT

inherits the problems of DFT:

- VdW interactions => empirical dispersion
- TD-DFT failures => limited use of TD-DFTB
- overpolarizability
- overbinding
- single reference method
- ...

DFTB is derived from DFT

inherits the problems of DFT:

- VdW interactions => empirical dispersion
- TD-DFT failures => limited use of TD-DFTB
- overpolarizability
- overbinding
- single reference method
- ...

but also the strenghts of DFT

- conceptual simplicity
- geometries
- vib. frequencies
- ...

=> MNDO-type and DFTB methods complement each other

SCC-DFTB

Second order expansion of the DFT total energy functional with respect to the charge density fluctuations $\delta\rho$ around a given reference density ρ_0 ($\rho_0' = \rho_0(\vec{r}')$, $\int' = \int d\vec{r}'$):

$$
\begin{aligned}
E &= \sum_i^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \int\int' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\, \delta\rho'} \right|_{n_0} \right) \Delta\rho\, \Delta\rho'. \\
&- \frac{1}{2} \int\int' \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0] n_0 + E_{cc}
\end{aligned}
$$

Reference density: $\rho_0 = \sum \rho_0^\alpha$
Superposition of atomic densities

Hamilton matrix elements

LCAO basis $\Psi_i = \sum c_\mu^i \eta_\mu$:

$$\langle \Psi_i | \hat{H}^0 | \Psi_i \rangle = \sum c_\mu^i c_\nu^i H_{\mu\nu}^0$$

$$H_{\mu\nu}^0 = \begin{cases} \epsilon_\mu & : \quad \mu = \nu \\ \langle \phi_\mu | H_{KS}[\rho_A + \rho_B] | \phi_\nu \rangle & : \quad \mu \in A, \nu \in B \\ 0 & : \quad \text{otherwise} \end{cases}$$

- 'Special' minimal basis set ϕ_μ and initial densities from atomic KS eqs.
- $H_{\mu\nu}^0$ and $S_{\mu\nu}$ calculated and stored $\longrightarrow$ no integral evaluation during program runtime.

Second order contributions

$$\frac{1}{2}\iint' \left(\frac{1}{|\vec{r}-\vec{r}'|} + \left.\frac{\delta^2 E_{xc}}{\delta\rho\,\delta\rho'}\right|_{n_0} \right) \Delta\rho\,\Delta\rho'.$$

- Monopole approximation: $\Delta\rho = \sum_\alpha \Delta\rho_\alpha \approx \sum_\alpha \Delta q_\alpha$
- second derivative $\longrightarrow \gamma_{\alpha\beta}$

$$\frac{1}{2}\sum_{\alpha\beta} \gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta$$

- This approximation will also be used for TD-DFTB

Repulsive energy terms

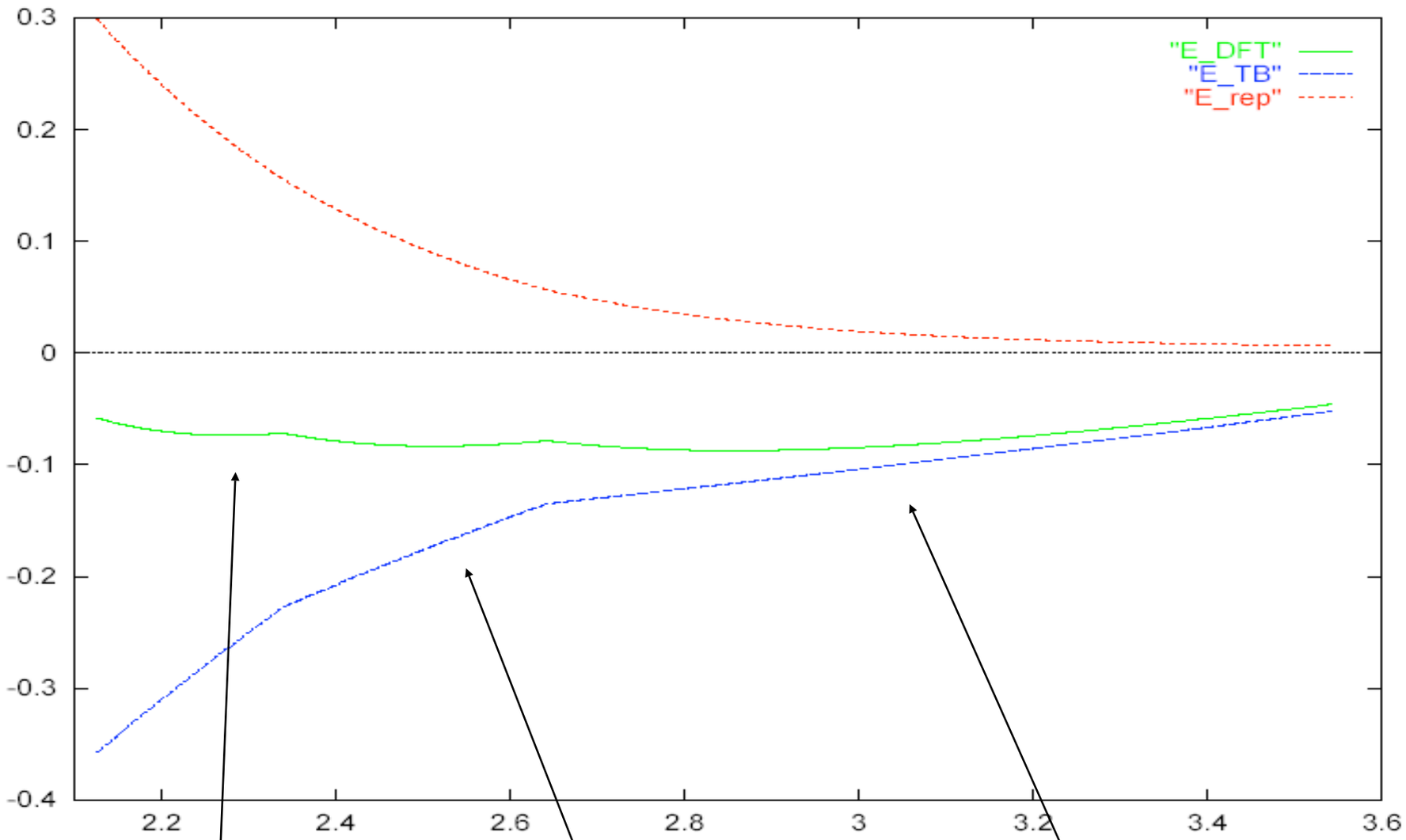
$$-\frac{1}{2}\iint' \frac{\rho_0'\rho_0}{|\vec{r}-\vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]n_0 + E_{cc} \rightarrow \sum_{\alpha\beta} U_{\alpha\beta}$$

SCC-DFTB total energy:

$$E = \sum_i \sum_{\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2}\sum_{\alpha\beta}\gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta + \sum_{\alpha\beta} U_{\alpha\beta}$$

Determination of the repulsive potential

$$E_{rep}(\mathbf{R}_{\alpha\beta}) = E^{DFT}(\mathbf{R}_{\alpha\beta}) - \left[\sum_{i}^{occ}\sum_{\mu\nu} c_{\mu}^{i} c_{\nu}^{i} H_{\mu\nu}^{0} + E^{2}\right](\mathbf{R}_{\alpha\beta})$$



HC-CH H_2C-CH_2 H_3C-CH_3

SCC-DFTB total energy

$$E = \sum_{i}^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \iint' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\, \delta\rho'} \right|_{n_0} \right) \Delta\rho\, \Delta\rho'.$$

$$- \frac{1}{2} \iint' \frac{\rho'_0 \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0] n_0 + E_{cc}$$

- minimal basis
- neglect of crystal field and three-center terms
- initial density fixed

- second order expansion
- monopole approximation
- gamma

- two-body approximation
- fit procedure

$$E = \sum_i \sum_{\mu\nu} c^i_\mu c^i_\nu H^0_{\mu\nu} + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + \sum_{\alpha\beta} U_{\alpha\beta}$$

SCC-DFTB Tests 1: Elstner et al., PRB 58 (1998) 7260

Performance for small organic molecules

(mean absolut deviations)

- Reaction energies[a)]: ~ 5 kcal/mole
- Bond-lenghts[a)] : ~ 0.014 A°
- Bond angles[b)]: ~ 2°
- Vib. Frequencies[c)]: ~6-7 %

a) J. Andzelm and E. Wimmer, J. Chem. Phys. **96**, 1280 1992.
b) J. S. Dewar, E. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 1985.
c) J. A. Pople, et al., Int. J. Quantum Chem., Quantum Chem. Symp. **15**, 269 1981.

SCC-DFTB Tests 2: T. Krueger, et al., J. Chem. Phys. 122 (2005) 114110.

	DFTB	cc-pVDZ	cc-pVTZ	G2
$H_2+CH\equiv CH \rightarrow CH_2=CH_2$	−38.8	−41.0	−37.9	−40.1
$H_2+CH_2=CH_2 \rightarrow CH_3-CH_3$	−37.0	−28.5	−26.5	−30.5
$3H_2+HCN \rightarrow NH_3+CH_4$	−47.9	−47.1	−48.5	−53.7
$H_2+CO \rightarrow H_2CO$	+10.9	+0.1	+0.7	+3.1
$2H_2+CO \rightarrow CH_3OH$	−7.8	−12.6	−13.1	−15.5
$H_2+CH_3OH \rightarrow CH_4+H_2O$	−25.7	−20.6	−24.6	−26.2
$2H_2+N_2 \rightarrow NH_2-NH_2$	+32.7	+31.6	+31.9	+30.7
$H_2+NH_2-NH_2 \rightarrow 2NH_3$	−47.4	−38.3	−43.1	−46.7
$H_2+H_2O_2 \rightarrow 2H_2O$	−81.7	−63.3	−71.1	−82.8
$2H_2+CO_2 \rightarrow H_2O+H_2CO$	+15.6	+26.6	+21.9	+14.5
$CH_4+CO \rightarrow CH\equiv CH+H_2O$	+55.9	+51.5	+43.7	+44.3
$CH_4+H_2CO \rightarrow CH_2=CH_2+H_2O$	+6.2	+10.4	+5.2	+1.1
$CH_4+CH_3OH \rightarrow CH_3-CH_3+H_2O$	−12.1	−5.4	−7.6	−10.8
$2CH_4+N_2 \rightarrow NH_2-NH_2+CH_2=CH_2$	+83.3	+75.4	+75.4	+76.6
$CH_4+H_2O_2 \rightarrow CH_3OH+H_2O$	−56.0	−42.6	−46.4	−56.6
$2NH_3+CH\equiv CH \rightarrow NH_2-NH_2+CH_2=CH_2$	+8.7	−2.8	+5.1	+6.7
$2NH_3+CH_2=CH_2 \rightarrow NH_2-NH_2+CH_3-CH_3$	+10.4	+9.7	+16.6	+16.2
$NH_3+HCN \rightarrow CH_4+N_2$	−33.1	−40.5	−37.4	−37.6
$NH_3+CO \rightarrow HCN+H_2O$	+14.3	+13.8	+10.8	+12.0
$2NH_3+H_2CO \rightarrow NH_2-NH_2+CH_3OH$	+28.8	+25.5	+29.3	+28.2
$H_2+H_2CO \rightarrow CH_3OH$	−18.7	−12.8	−13.8	−18.6
$Oxirane+H_2O \rightarrow OH-CH_2-CH_2-OH$	−30.9	−20.5	−16.7	−19.6
$Oxirane+NH_3 \rightarrow NH_2-CH_2-CH_2-OH$	−33.1	−20.4	−17.5	−22.6
$HNCO+H_2O \rightarrow NH_2-COOH$	−4.3	−17.3	−11.1	−16.1
$CH_2=NH+CH_4+NH_3 \rightarrow 2CH_3NH_2$	−2.7	+1.1	+4.3	−0.6
$H_2CO+CH_4+H_2O \rightarrow 2CH_3OH$	+7.1	+7.9	+10.9	+7.6
$HCN+2CH_4+2NH_3 \rightarrow 3CH_3NH_2$	+17.4	+14.1	+21.4	+14.6
$CO+2CH_4+2H_2O \rightarrow 3CH_3OH$	+43.6	+28.6	+36.2	+36.8

SCC-DFTB Tests 2: T. Krueger, et al., J. Chem. Phys. 122 (2005) 114110.

	DFTB	cc-pVDZ	cc-pVTZ	G2
$H_2+CH\equiv CH \rightarrow CH_2=CH_2$	−38.8	−41.0	−37.9	−40.1
$H_2+CH_2=CH_2 \rightarrow CH_3-CH_3$	−37.0	−28.5	−26.5	−30.5
$3H_2+HCN \rightarrow NH_3+CH_4$	−47.9	−47.1	−48.5	−53.7
$H_2+CO \rightarrow H_2CO$	+10.9	+0.1	+0.7	+3.1
$2H_2+CO \rightarrow CH_3OH$	−7.8	−12.6	−13.1	−15.5
$H_2+CH_3OH \rightarrow CH_4+H_2O$	−25.7	−20.6	−24.6	−26.2
				+30.7
				+46.7
				+82.8
				+14.5
				+44.3
				+1.1
				+10.8
2CH				+76.6
				+56.6
$2NH_3+CH\equiv$				+6.7
$2NH_3+CH_2$				+16.2
				+37.6
$NH_3+CO \rightarrow HCN+H_2O$	+14.3	+13.8	+10.8	+12.0
$2NH_3+H_2CO \rightarrow NH_2-NH_2+CH_3OH$	+28.8	+25.5	+29.3	+28.2
$H_2+H_2CO \rightarrow CH_3OH$	−18.7	−12.8	−13.8	−18.6
$Oxirane+H_2O \rightarrow OH-CH_2-CH_2-OH$	−30.9	−20.5	−16.7	−19.6
$Oxirane+NH_3 \rightarrow NH_2-CH_2-CH_2-OH$	−33.1	−20.4	−17.5	−22.6
$HNCO+H_2O \rightarrow NH_2-COOH$	−4.3	−17.3	−11.1	−16.1
$CH_2=NH+CH_4+NH_3 \rightarrow 2CH_3NH_2$	−2.7	+1.1	+4.3	−0.6
$H_2CO+CH_4+H_2O \rightarrow 2CH_3OH$	+7.1	+7.9	+10.9	+7.6
$HCN+2CH_4+2NH_3 \rightarrow 3CH_3NH_2$	+17.4	+14.1	+21.4	+14.6
$CO+2CH_4+2H_2O \rightarrow 3CH_3OH$	+43.6	+28.6	+36.2	+36.8

With respect to G2:
mean ave. dev.: 4.3 kcal/mole
mean dev.: 1.5 kcal/mole

SCC-DFTB Tests 3: Sattelmeyer & Jorgensen

J. Phys. Chem. A **2006,** *110,* 13551

Mean Absolute Errors in Calculated Heats of Formation for Neutral Molecules Containing the

Elements C, H, N and O (kcal/mol).

	N	AM1	PM3	PDDG/PM3	SCC-DFTB
Hydrocarbons	254	5.6	3.6	2.6	4.8
All Molecules	622	6.7	4.4	3.2	5.9
Training Set	134	6.1	4.3	2.7	7.0
Test Set	488	6.8	4.4	3.3	5.6

SCC-DFTB Tests 3: Sattelmeyer & Jorgensen

J. Phys. Chem. A **2006,** *110,* 13551

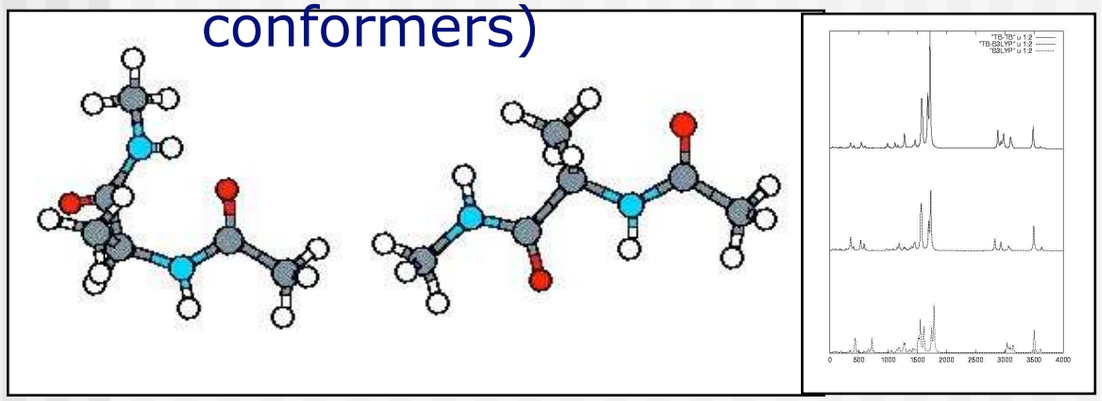
Absolute Errors for Additional Molecular Properties of CHNO-containing Species.

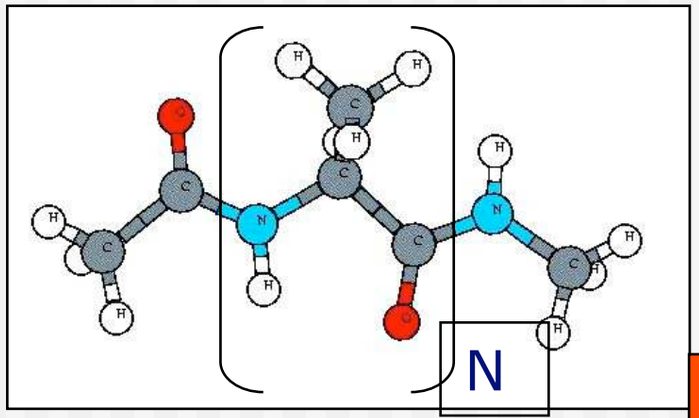
	N	AM1	PM3	PDDG/PM3	SCC-DFTB
Bond lengths (Å)	218	0.017	0.012	0.013	0.012
Bond angles (deg.)	126	1.5	1.7	1.9	1.0
Dihedral angles (deg.)	30	2.8	3.2	3.7	2.9
Dipole moments (D)	47	0.23	0.25	0.23	0.39

Secondary-structure elements for Glycine und Alanine-based polypeptides

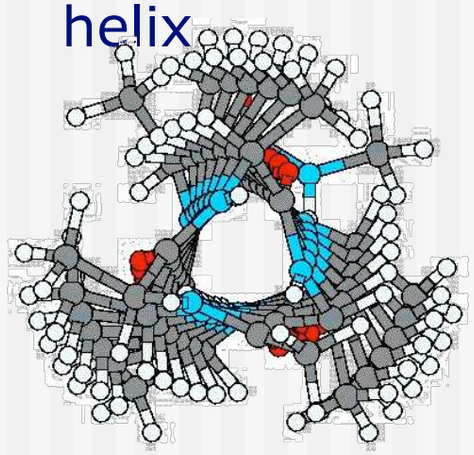
Elstner, et al.. Chem. Phys. 256 (2000) 15

N = 1 (6 stable conformers)

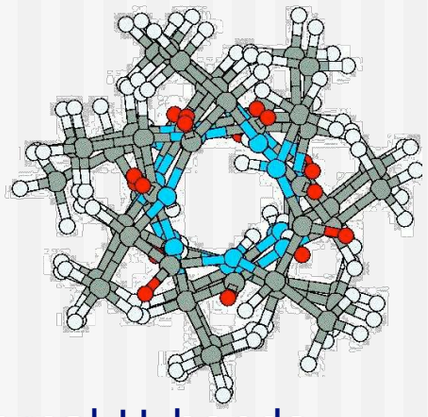




3_{10} - helix



α_R-helix



stabilization by internal H-bonds

between i and i+3

between i and i+4

DFTB very good for:

- relative energies
- geometries
- vib. freq. o.k.!

➔main problem for DFT(B): dispersion!

➔ AM1, PM3, MNDO quite bad

➔ OM2 much improved (JCC 22 (2001) 509)

Secondary-structure elements for Alanine-based polypeptides

Otte, Scholten & Thiel JPCA 111, 5753

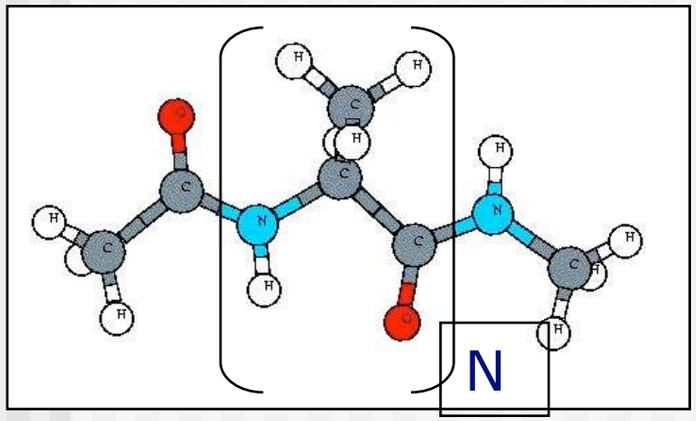


TABLE 5: Mean Absolute Deviations for the Peptide Test Set[a]

	N	AM1	OM2	DFTB
relative energies (kcal/mol)	22	2.0	1.7	1.1
backbone H-bond lengths (Å)	67	0.22	0.34	0.26
backbone dihedral angles (deg)	190	17.0	12.0	9.0

Hydrogen bonding

Otte, Scholten & Thiel JPCA 111, 5753

Hydrogen bonds of 57 complexes

	AM1	OM2	SCC-DFTB
E	2.8	1.5	2.7
R	0.25	0.20	0.08
θ	33.7°	12.1°	6.2°

- DFTB scatters around B3LYP values
- AM1 0.12 A too long
- OM2 0.14 A too short