QM/MM vs. implicit solvation methods
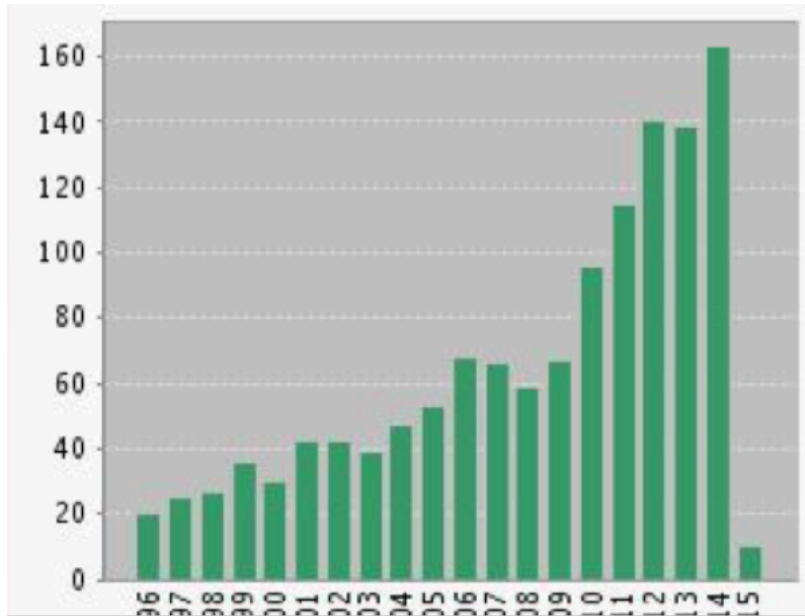
Elise Dumont, ENS de Lyon

Winter 2021
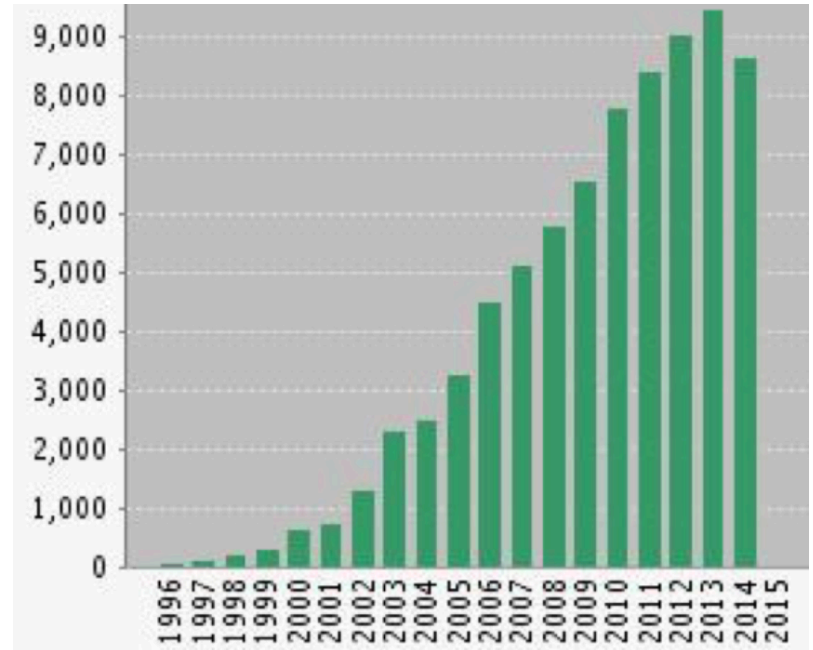
Useful ?

Number of citations of PCM

Number of citations of QM/MM

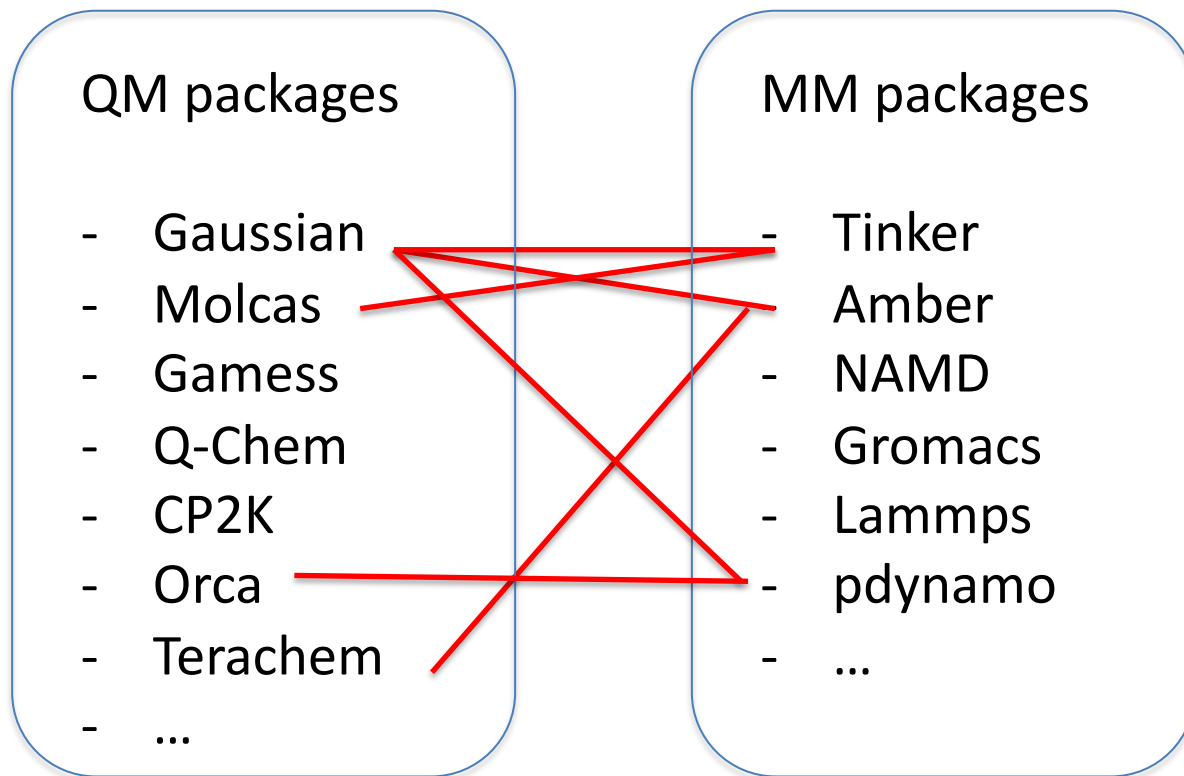More with other schemes
(COSMO, homemade codes…)

↓

Nobel Prize

Outline

1. Examples of QM/MM application: vision, magnetism, …

1. Merging MM and QM…

2. Situating the implicit models for solvation

3. Summary: Implicit vs. QM/MM

Utility ?

QM packages

- Gaussian
- Molcas
- Gamess
- Q-Chem
- CP2K
- Orca
- Terachem
- …

MM packages
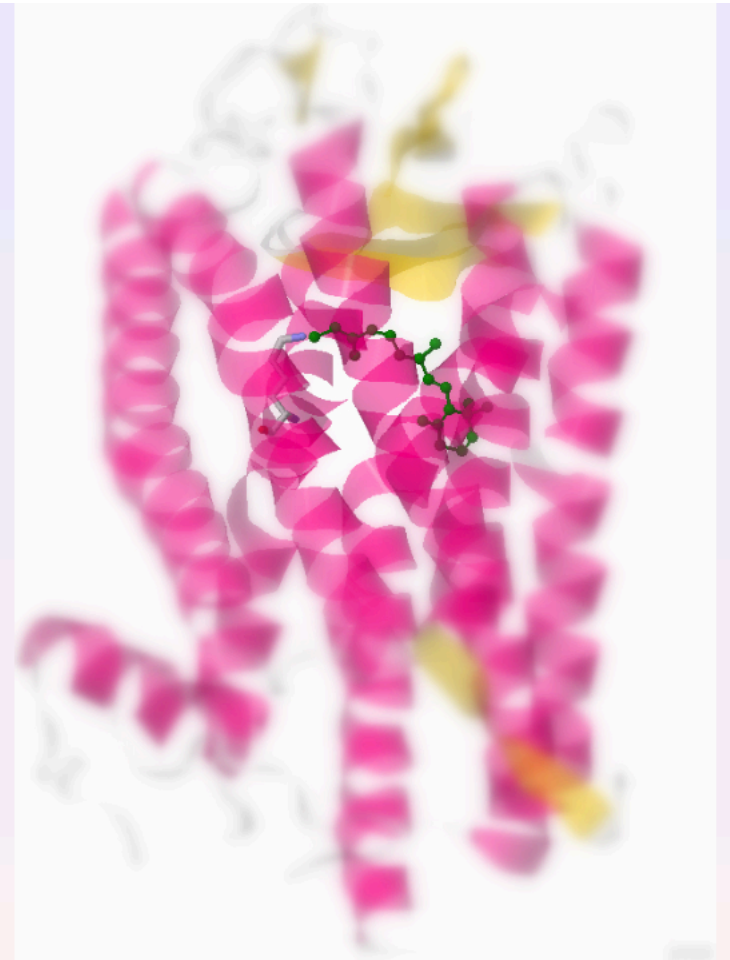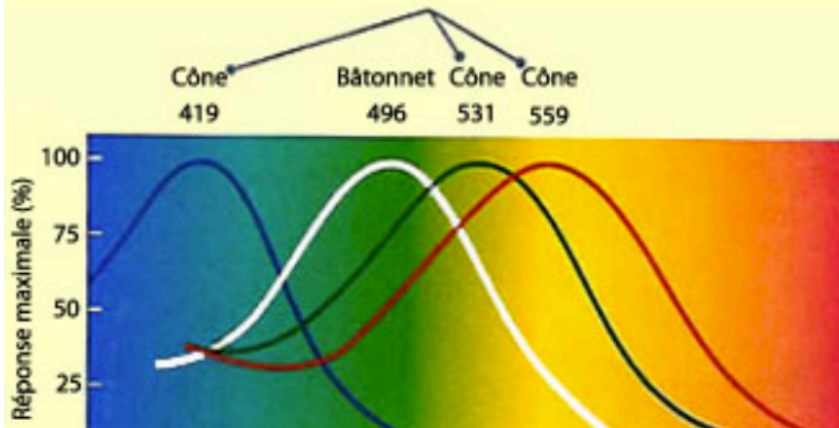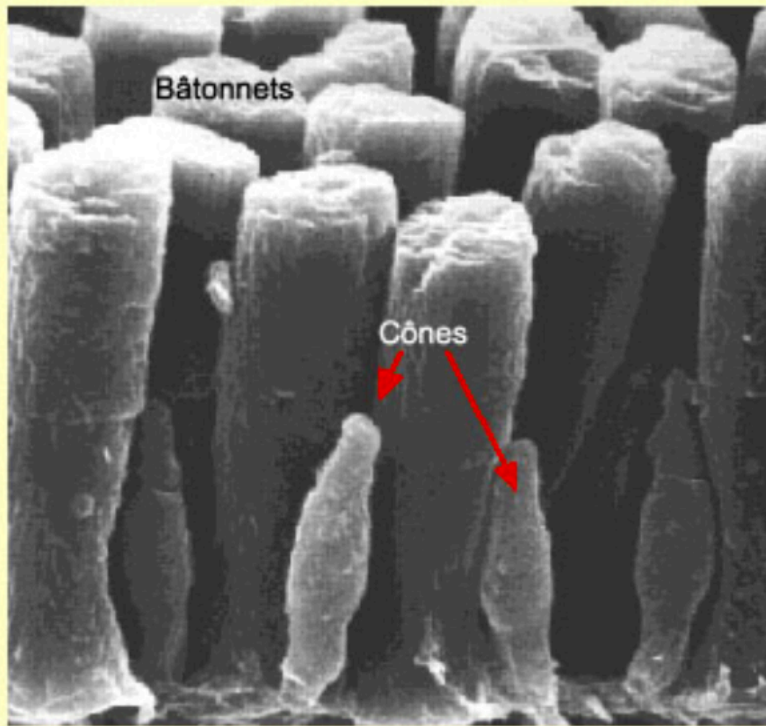
- Tinker
- Amber
- NAMD
- Gromacs
- Lammps
- pdynamo
- …

And your own flavor because it is relatively easy to implement…

Virtually all QM and MM packages now include a QM/MM coupling and PCM. Dedicated interfaces (scripts) also exist (Oniom only): ChemShell, Pupil, ComQum, …
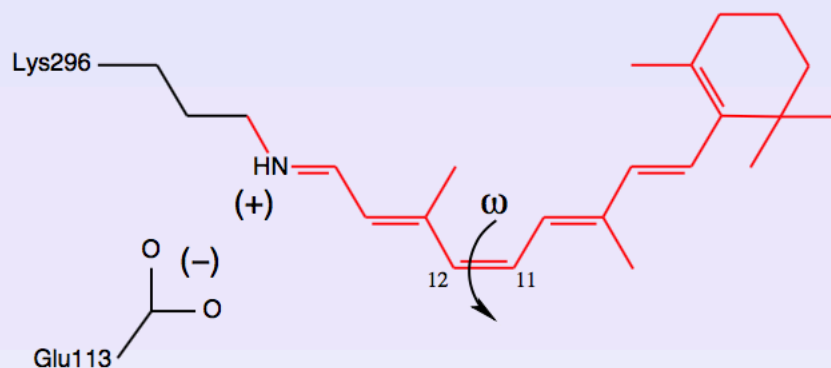
Example 1: the retinal chromophore within rhodopsin

The macromolecular environment strongly tunes the photophysics

The protein matters…

	methanol	rhodopsin	vacuum
λ_{max} (nm)	445	498	610
Time scale	> 2 ps	$\simeq$ 200 fs	> 2 ps
Quantum yield	24 %	67 %	20 %

The protein tunes finely the optical properties:
how to treat the solvent and the proteic embedding ?
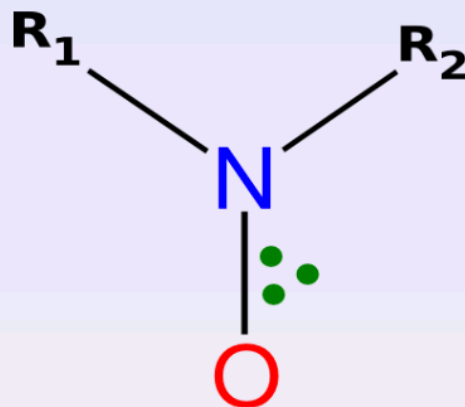
→ 3 π electrons: spin polarisation contributes to the N spin density

A subtle system… needs for polarizable ff
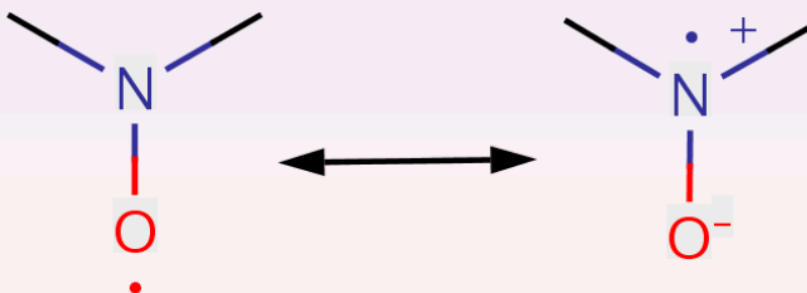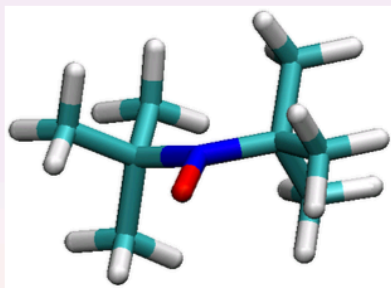
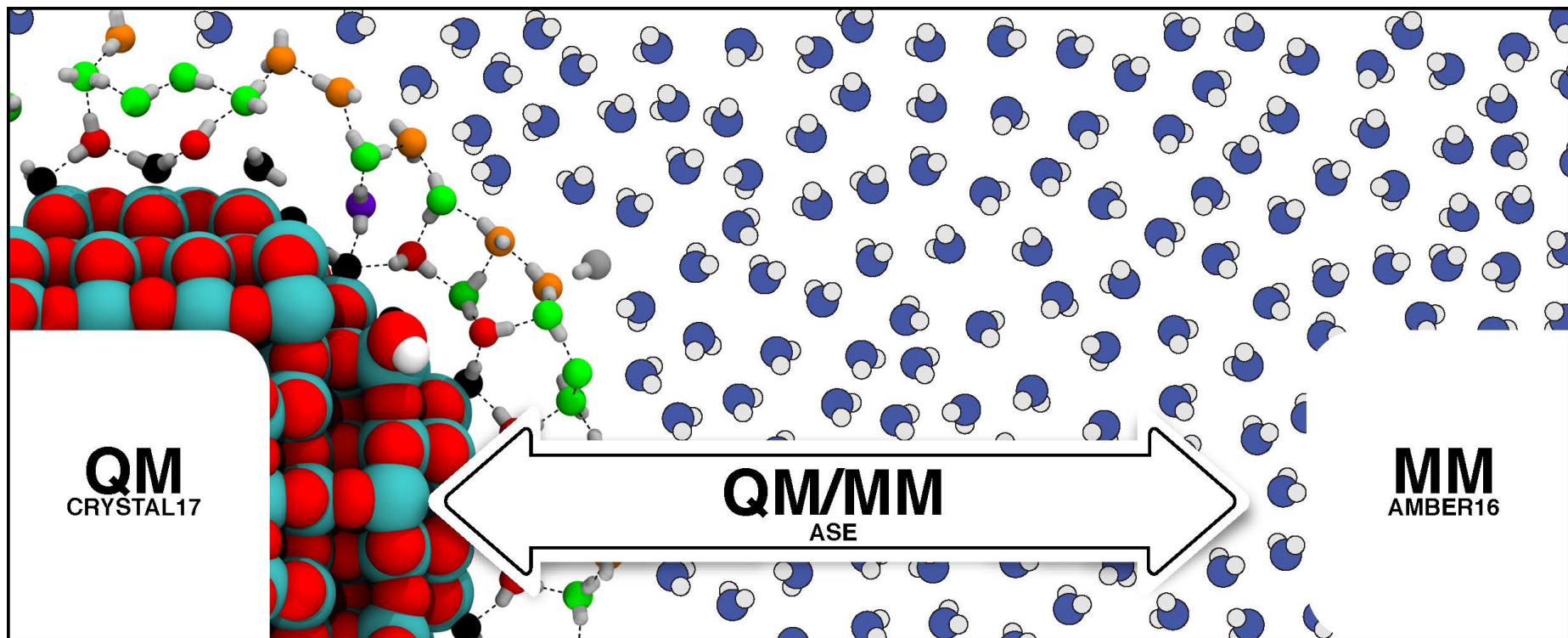|	hexane	water
a_N	14,9 G	17,2 G
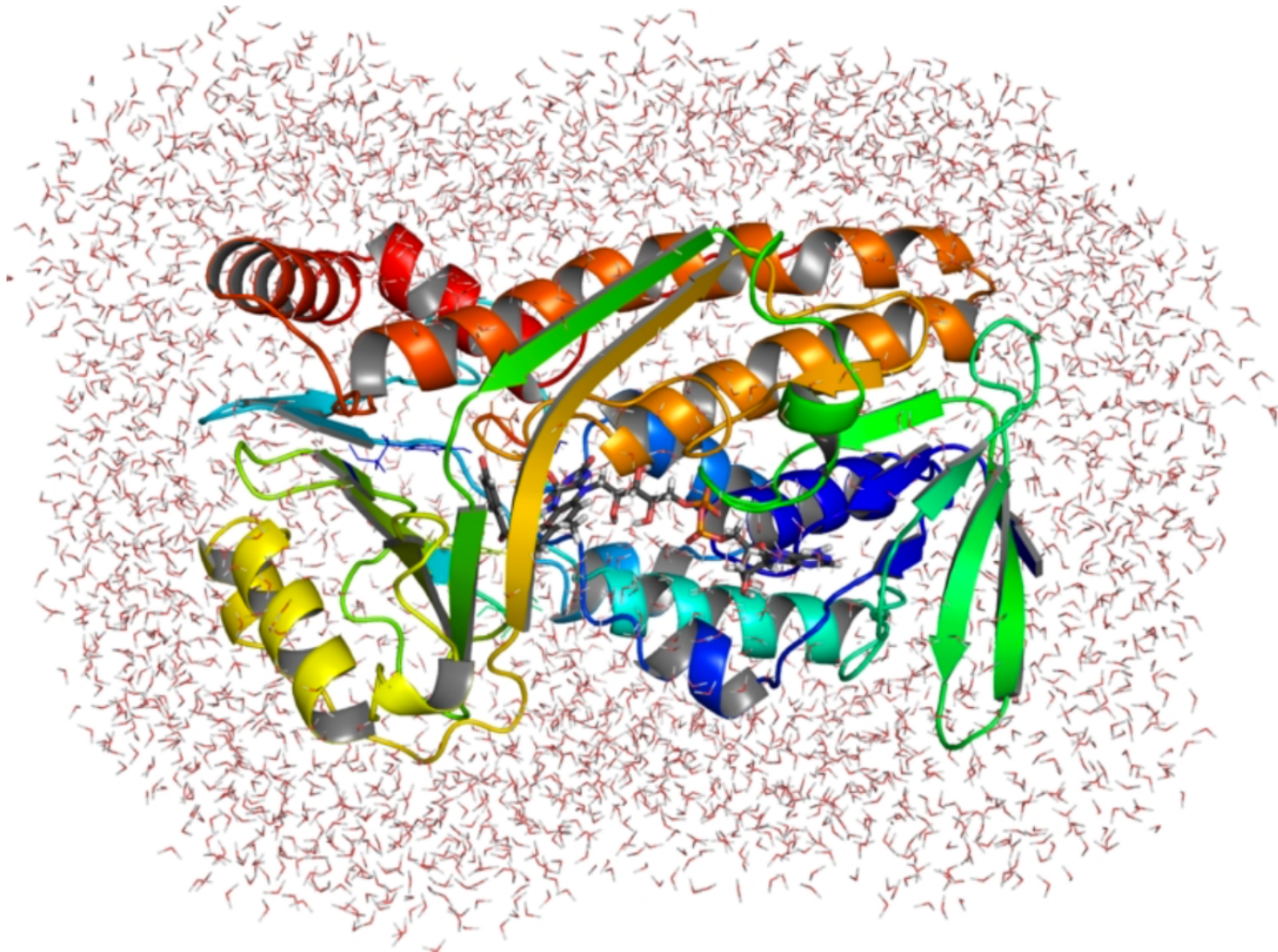
QM/MM approaches

Full DFT molecular dynamics for such a system would not maintain the structure. Force fields parameters offer more accuracy.

QM/MM approaches for metallic interfaces

Interface ? How to describe the solvent ?

QM and MM meet
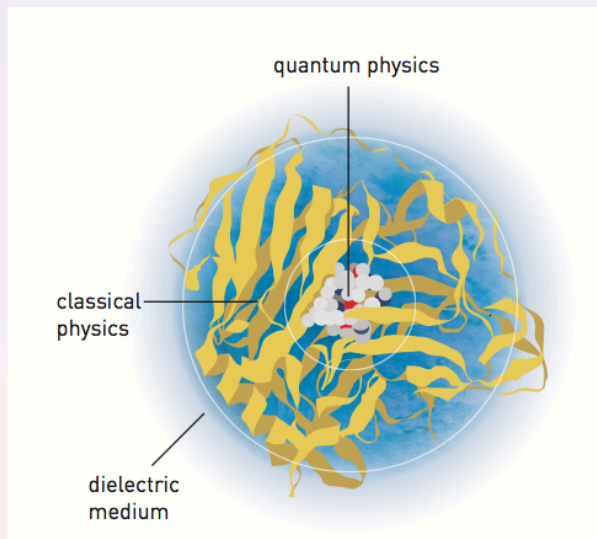
QM and MM meet
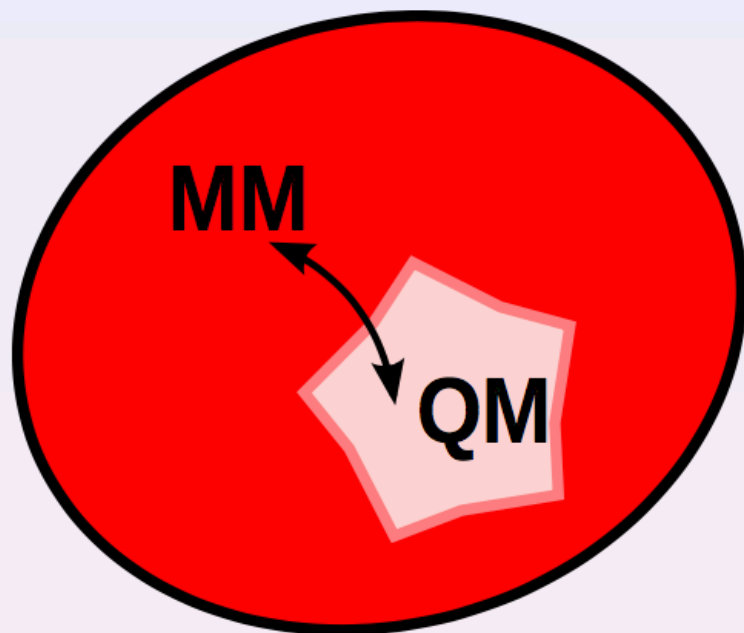
The Nobel Prize in Chemistry 2013

Photo: A. Mahmoud
Martin Karplus
Prize share: 1/3

Photo: A. Mahmoud
Michael Levitt
Prize share: 1/3

Photo: A. Mahmoud
Arieh Warshel
Prize share: 1/3

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*.

Photos: Copyright © The Nobel Foundation

QM and MM try to talk one each other
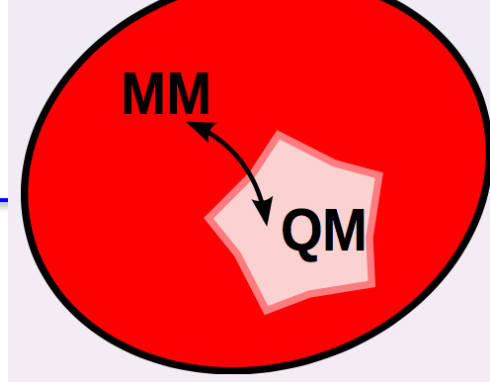
The active site (high-level region):
few degrees of freedom described with an accurate QM model

The surroundings:
a lot of degrees of freedom described with an approximate MM model

How can we couple them ?

- steric interactions mainly
→ mechanical embedding
- electrostatic interactions mainly
→ electrostatic embedding or polarization embedding

QM and MM try to talk one each other

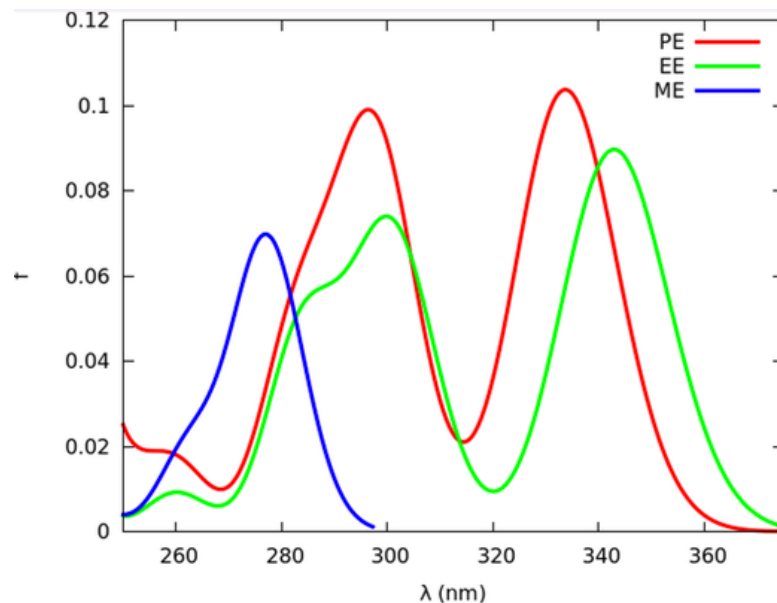$$E = \left\langle \Psi \middle| H \middle| \Psi \right\rangle = \left\langle \Psi \middle| H_{QM} + H_{MM} + H_{QM/MM} \middle| \Psi \right\rangle$$

Energy of the isolated QM system

Most of the interactions don't depend explicitely on the electronic degrees of freedom

Second approximation « à la MM »:
 only the QM/MM electrostatic interactions depend on the electrons

QM and MM try to talk to each other: substractive scheme
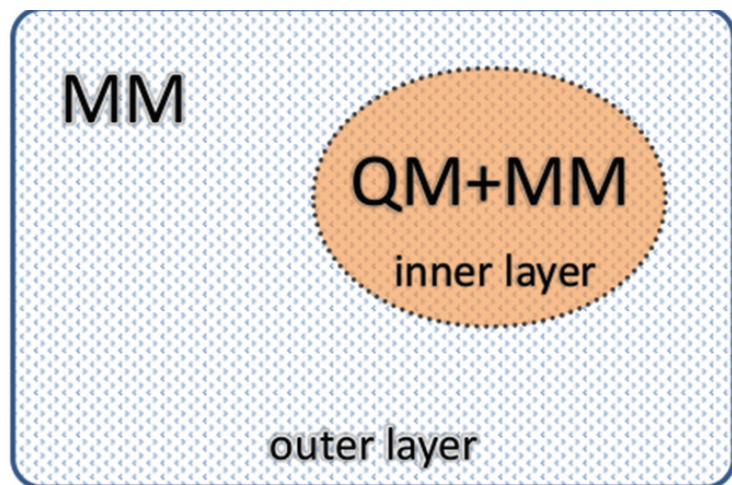
$$E^{MM}(in+out) - E^{MM}(in) + E^{QM}(in)$$

Ease of implementation, typical of ONIOM…
Easy to implement for more layers (multi-scale QM'/QM/MM)

But need to assign force fied parameters for the QM region

QM <-> fixed-charge force fields

Additive scheme: more robust and no need to define parameters for the QM region

$$E^{QM/MM}[\rho] = \boxed{E^{QM}[\rho]} + \boxed{E^{MM}} + E^{QM*MM}[\rho]$$

$$E^{QM}[\rho] = T_s[\rho] + \int v_{ext}\rho(\boldsymbol{r})d\boldsymbol{r} + J[\rho] + E_{xc}[\rho] + E_{NN}$$

$$E^{MM} = E^{bonded} + E^{non-bonded}$$

$$E^{bonded} = E^{bond} + E^{angle} + E^{torsion} + E^{improper}$$

$$E^{non-bonded} = E^{elec} + E^{LJ}$$

Classical description within QM/MM

Each and every MM force field includes a large set of parameters, fitted to reproduce experimental data (free energies) and/or QM calculations (torsions profile)

Rather clearly, one relies on a fortitious cancellation of errors

We assume a transferability of force fields when we do a QM/MM calculation but with no proof for it

One simplifies the wavefunction (density) into a force field
-> RESP procedure

Done a priori once-for-all, yet the QM part can undergo significant changes...
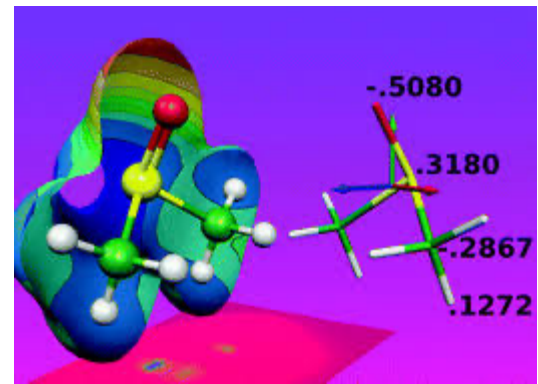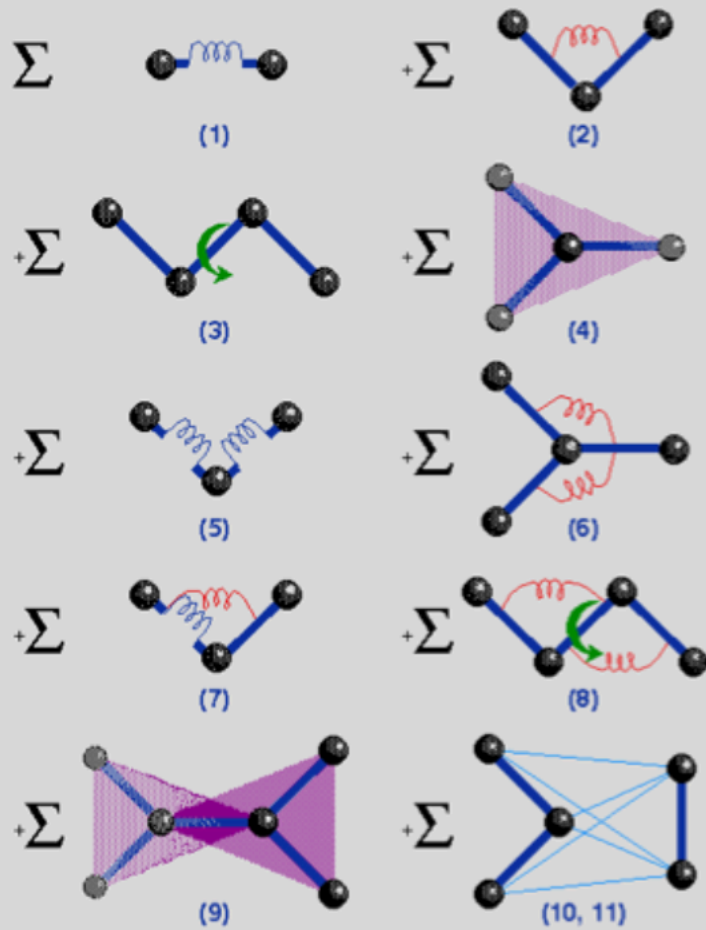
Inside a force field…

Figure 3-2. Graphic Illustration of Terms in CVFF

- $E_b(A-B) = \frac{k_{XY}}{2}(d_{AB} - d_{AB}^0)^2$

- $E_a(\angle A-B-C) = \frac{k_{ABC}}{2}(\theta_{ABC} - \theta_{ABC}^0)^2$

- $E_t(\angle ABC/BCD) = \sum_n \frac{V_n}{2}(1 + \cos(n\varphi - \gamma))$

- $E_{vdw}(A\cdots B) = 4\varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{d_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{d_{AB}}\right)^6\right]$

- $E_q(q_A\cdots q_B) = \frac{q_A \cdot q_B}{4\pi\varepsilon_0 d_{AB}}$

- $E_{ind}(q_A\cdots \vec{\mu}_B) = \frac{q_A \cdot \alpha_B \vec{E}_B}{d_{AB}^2}$

  $\vec{E}_B = \left(\vec{E}_B^0 + \sum_{C\neq B} T_{BC}\vec{\mu}_C\right)$

The coupling term $E^{QM/MM}$

Additive scheme: more robust and no need to define parameters for the QM region

$$E^{QM/MM}[\rho] = E^{QM}[\rho] + E^{MM} + \boxed{E^{QM*MM}[\rho]}$$

$$E^{QM}[\rho] = T_s[\rho] + \int v_{ext}\rho(\boldsymbol{r})d\boldsymbol{r} + J[\rho] + E_{xc}[\rho] + E_{NN}$$

$$E^{QM*MM}[\rho] = E^{\rho q} + E^{Zq} + E^{LJ}$$

$$E^{QM*MM}[\rho] = -\sum_{K}^{MM} \int \frac{q_K \cdot \rho(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{R}_K|}d\boldsymbol{r} + \sum_{A}^{QM}\sum_{K}^{MM}\left[\frac{q_K \cdot Z_A}{|\boldsymbol{R}_{AK}|} + \varepsilon_{AK}\left[\left(\frac{\boldsymbol{r}_{AK}^{min}}{\boldsymbol{R}_{AK}}\right)^{12} - 2\left(\frac{\boldsymbol{r}_{AK}^{min}}{\boldsymbol{R}_{AK}}\right)^{6}\right]\right]$$
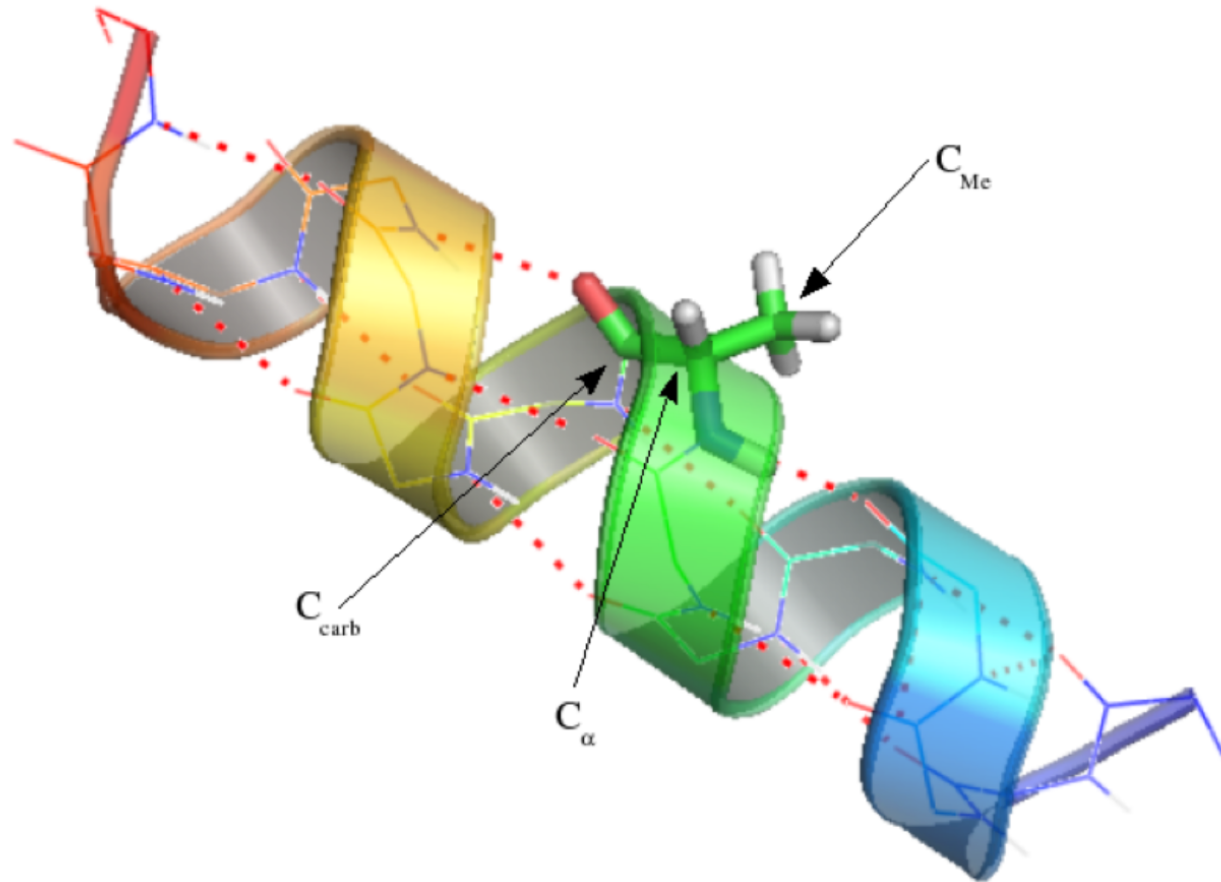
Need of additional constraints…

Most QM/MM scheme are not compatible with periodic boundary conditions.

Applications toward material science are more recent and delicate :

- many covalent bonds to mimick
- polarization effects…
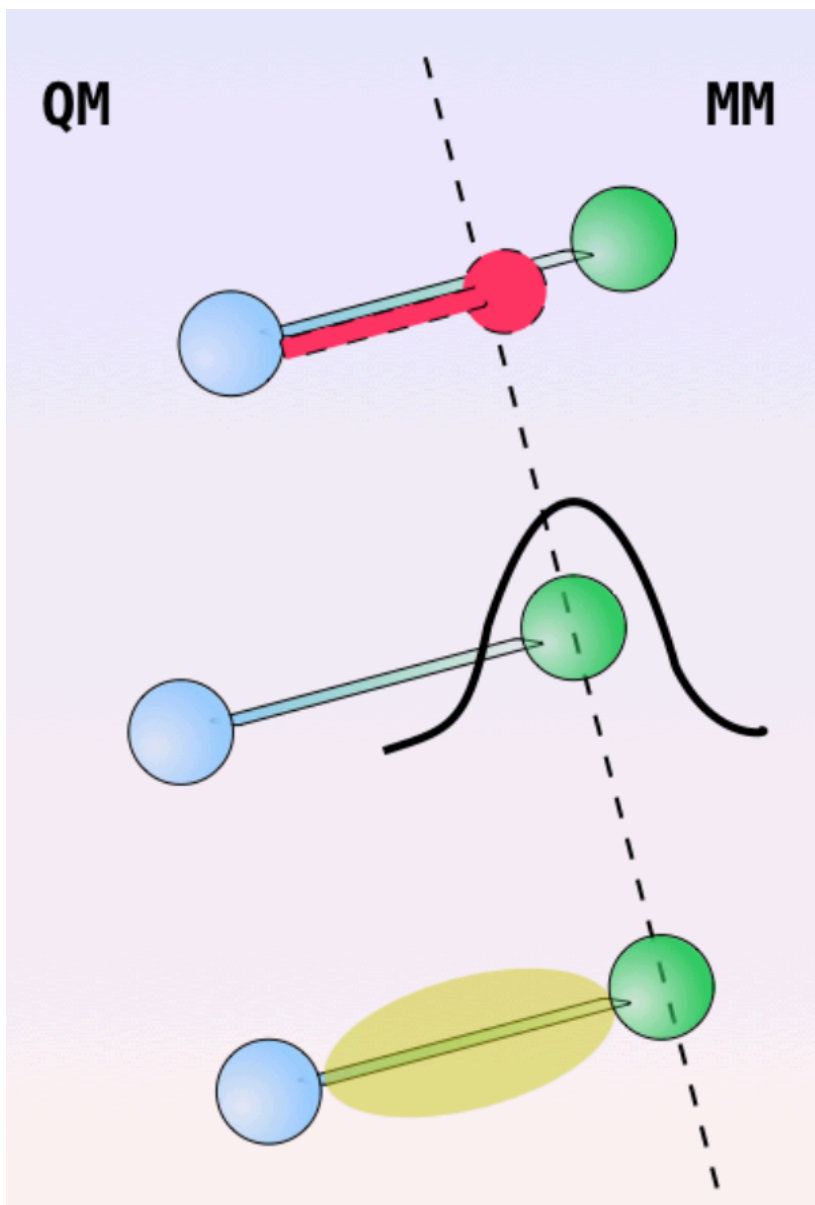- possibly more difficult to redistribute charges…

If possible avoid cutting… otherwise a single carbon-carbon bond (for peptides and proteins)

How to cut ?

Link-atom

Pseudo-potential
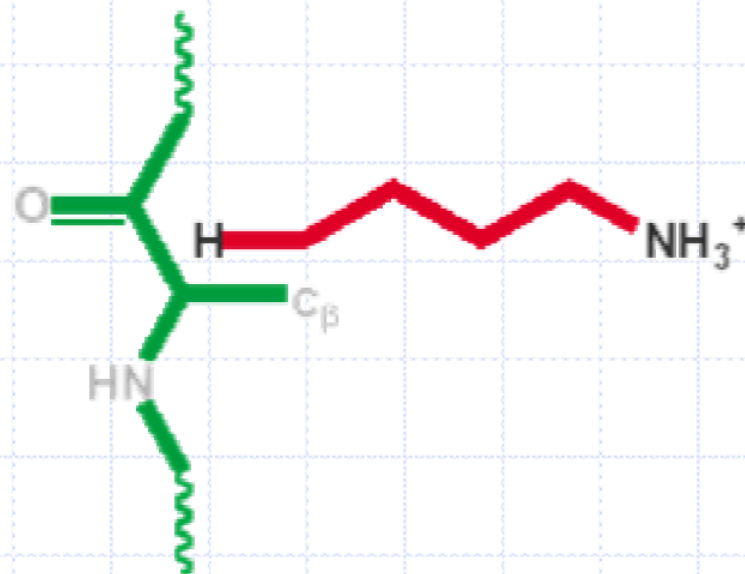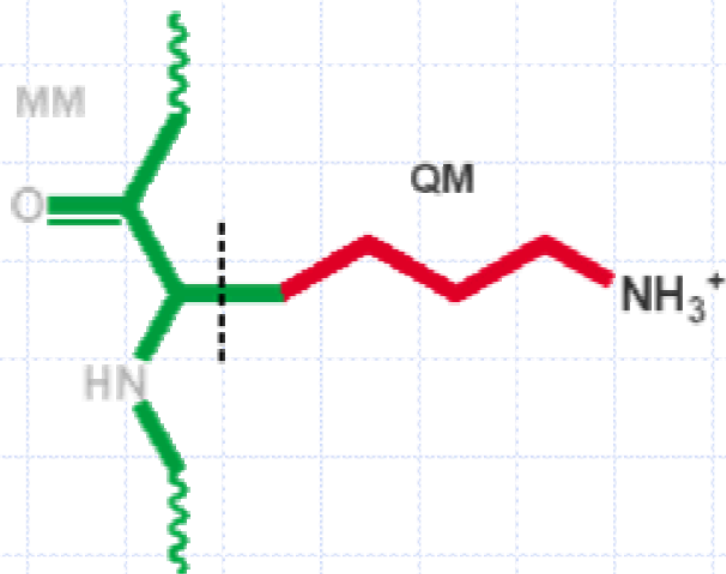
Frozen orbital

The easiest solution (Link atom)

If possible avoid cutting… otherwise a single carbon-carbon bond (for peptides and proteins) ; also avoid several cuts

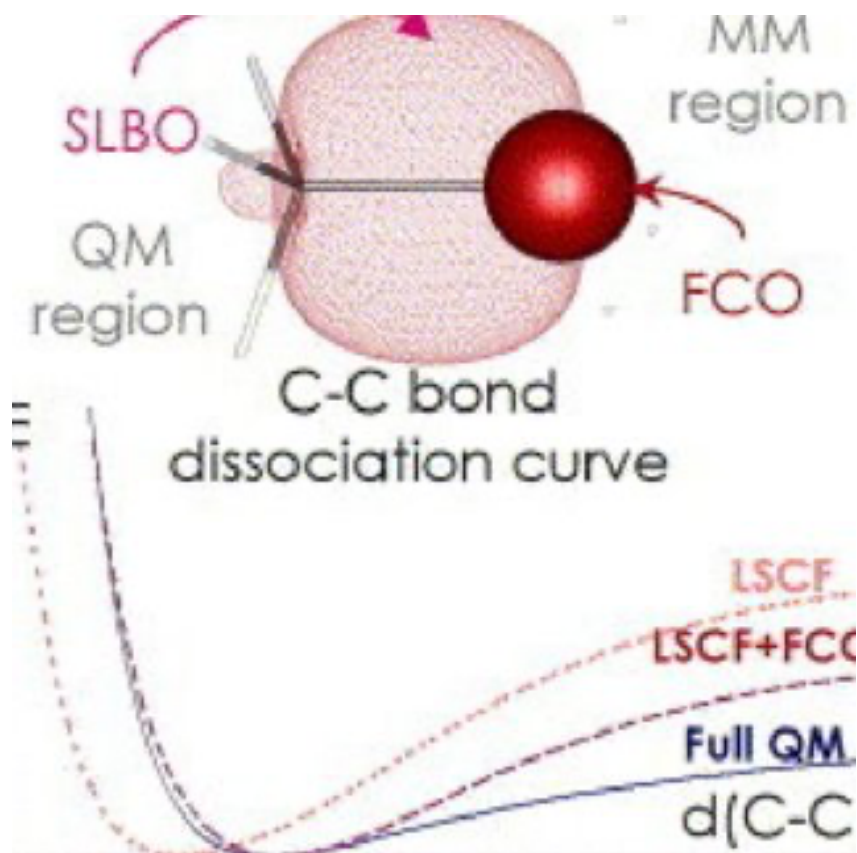Link atom: the capping hydrogen may be free to move or not

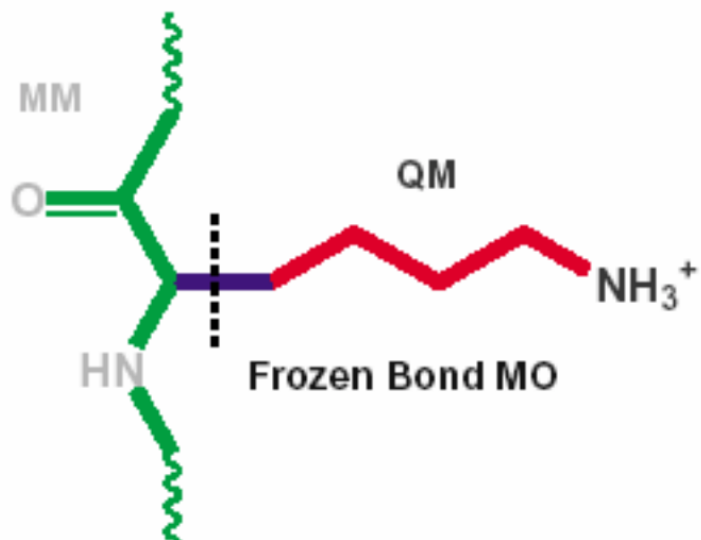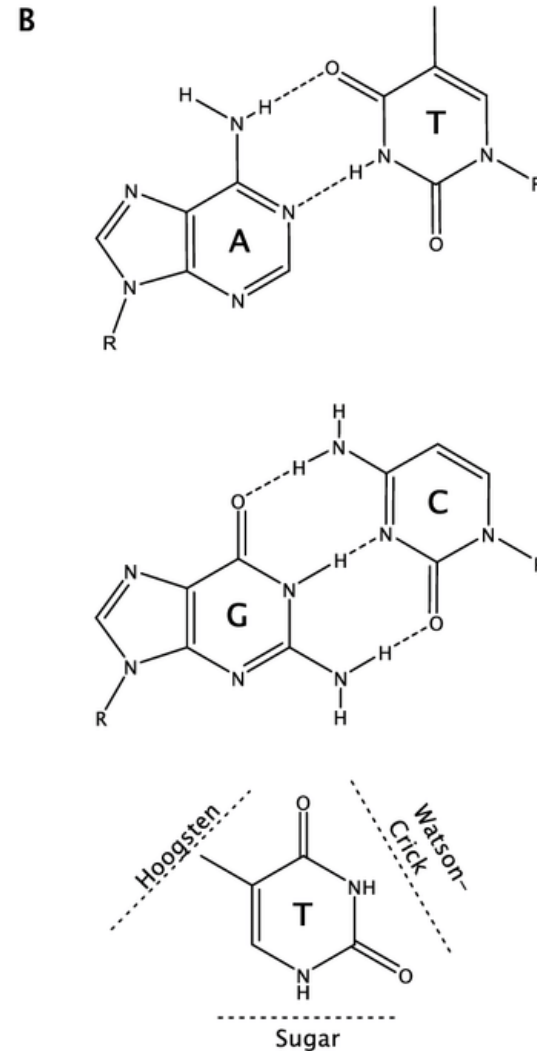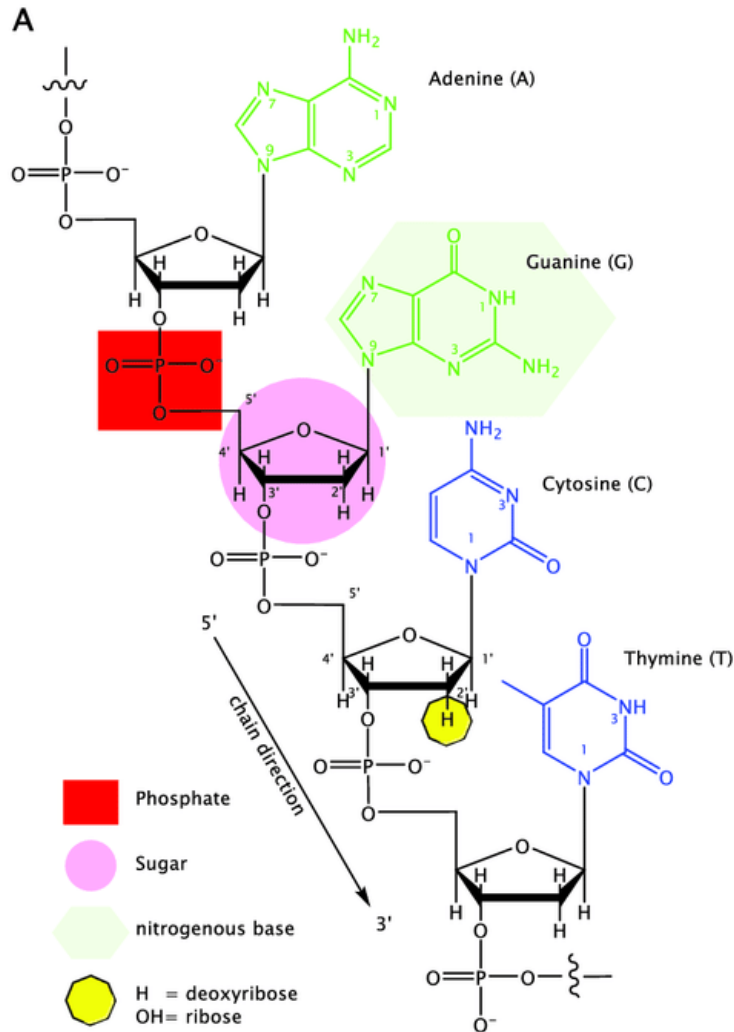# More sophisticated schemes…

Essentially needed to cut along polarized bonds

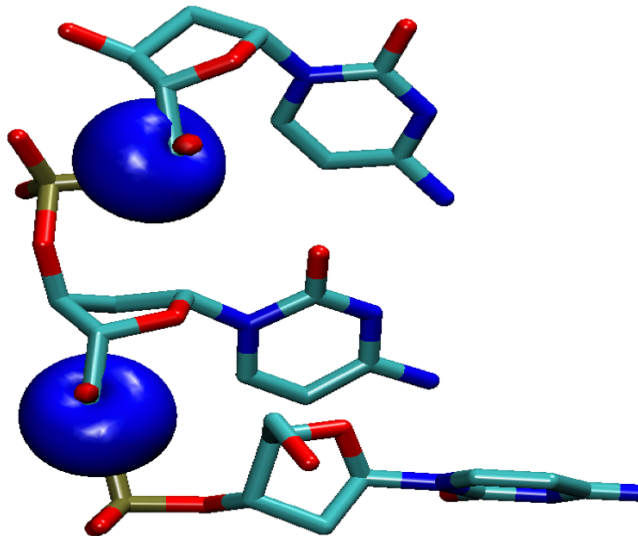# More sophisticated schemes…

Essentially needed to cut along polarized bonds

More sophisticated schemes…

Optimization of the wavefunction under constraints. Only for σ orbitals.

F.C = S.C.E + S.C.Λ. (second term frozen)

Drawback of being close to a phosphate…

Partie quantique   Partie classique
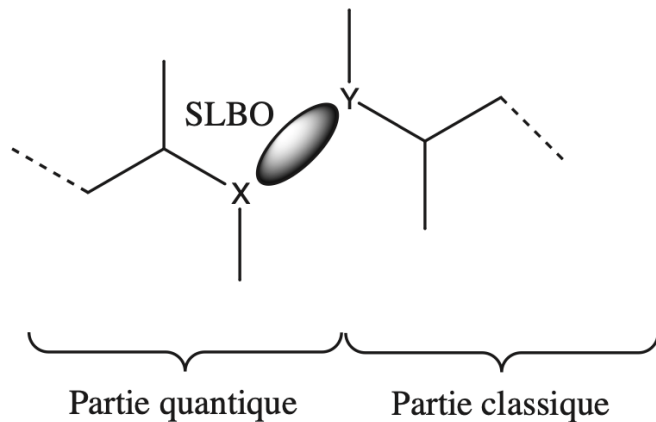
More sophisticated schemes…

Essentially needed to cut along polarized bonds
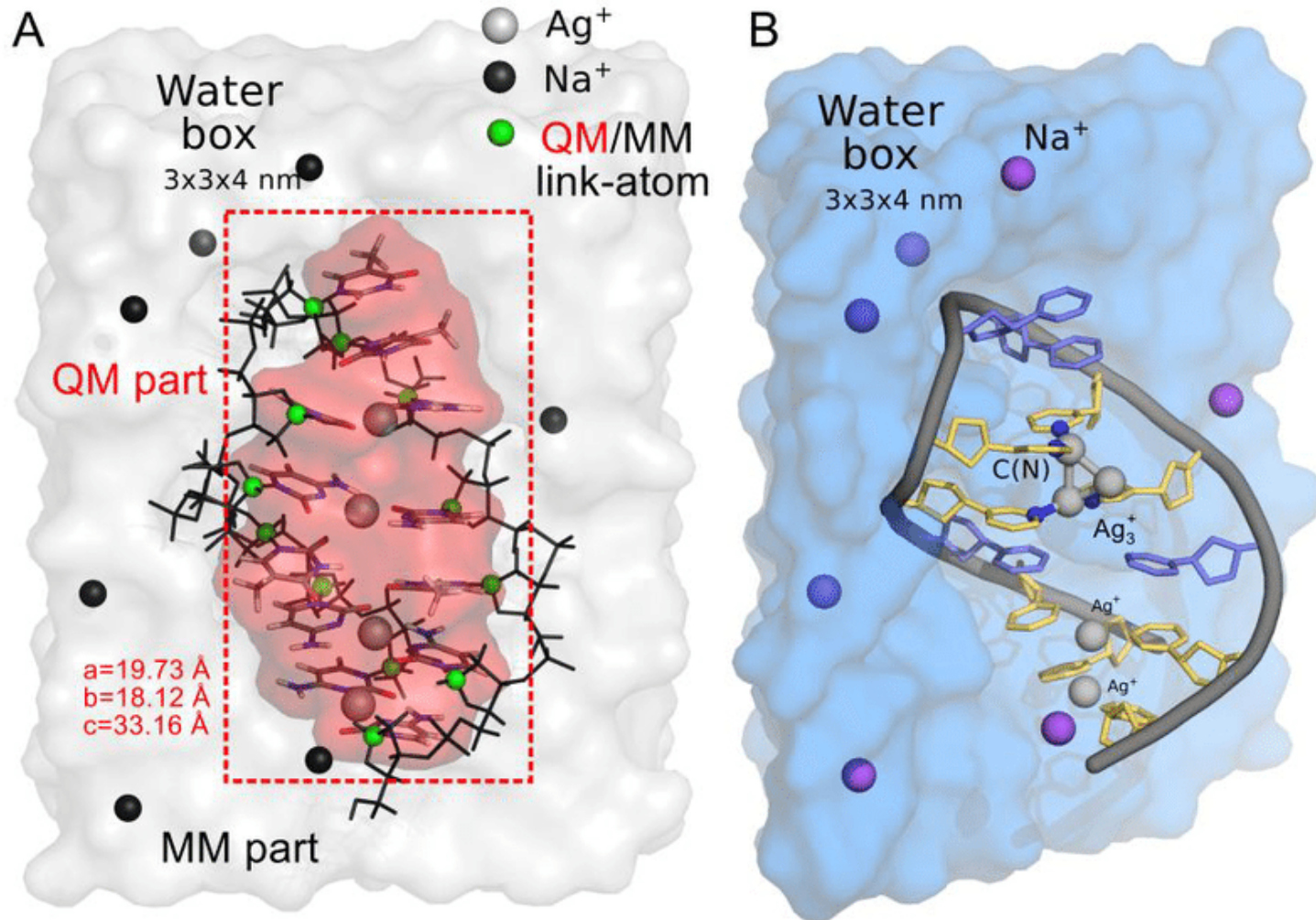
Hands-on for guanine charge transfer

Before to treat the PDB, an MD exploration is required (selecting the last frame or several representative ones).
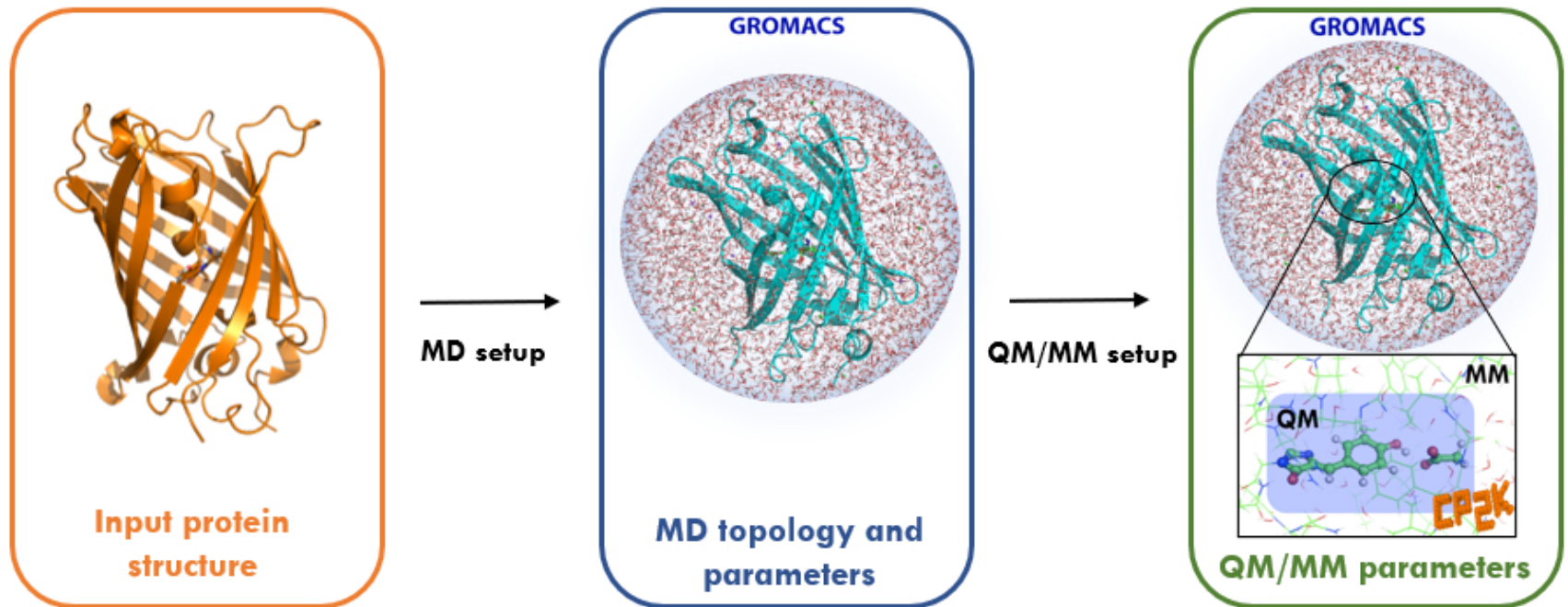
PDB 143D

Need to set up a frozen layer

Hands-on for guanine charge transfer

Constrained DFT for two π-stacked guanines
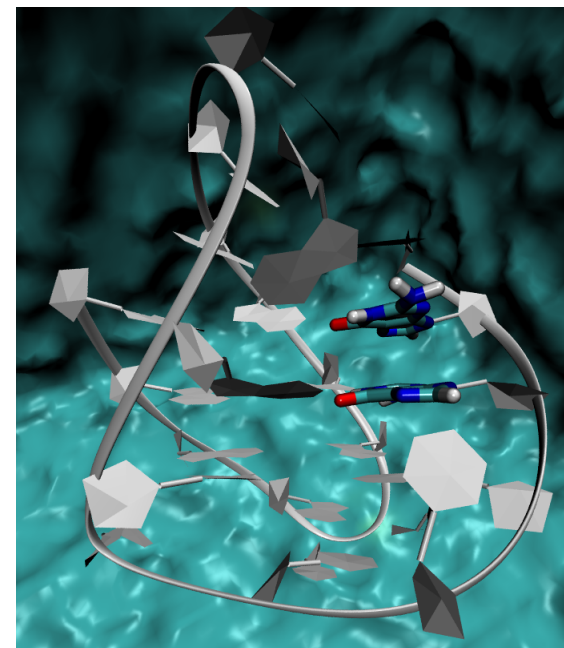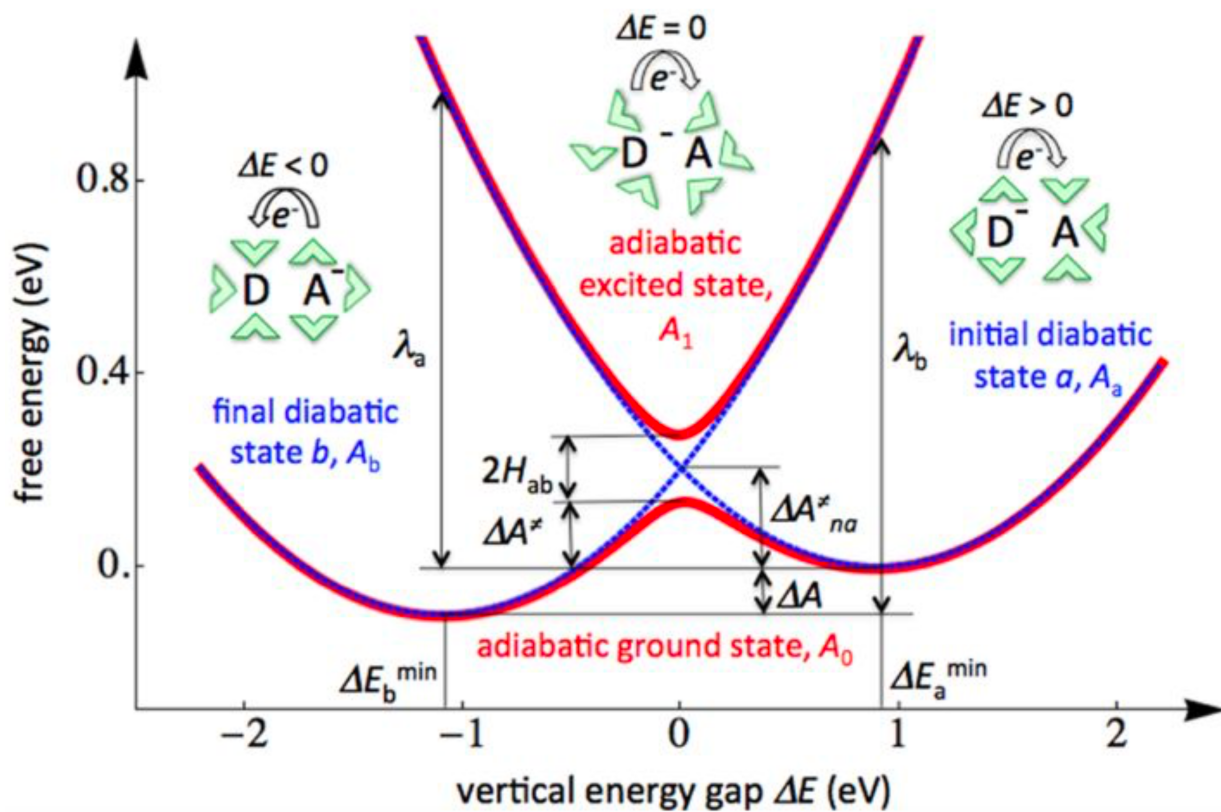
G quadruplexes

How to treat a solvent ?
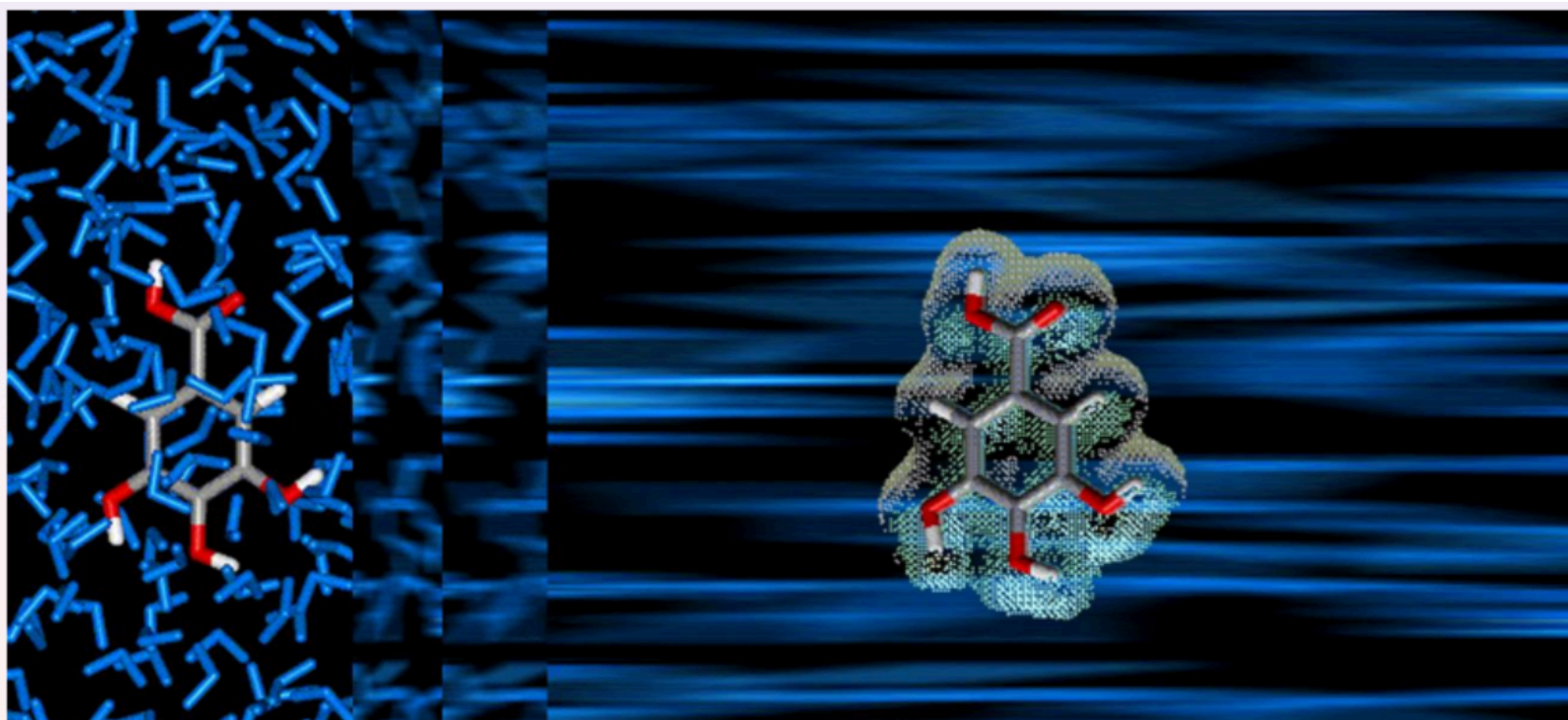
Explicit treatment
(with dynamics...) → Implicit treatment
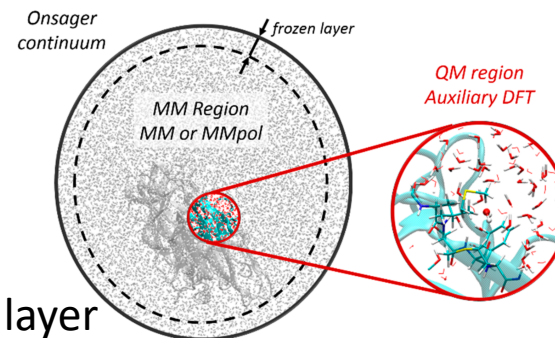
QM/MM-MD

QM/MM cost is such that on can afford to run them along time

QM/MM-MD simulations time from several fs (radiation chemistry) to 1 ns (either with CASSCF or DFT or approximate DFT scheme or semi-empirical methods)

-> compromise between electronic accuracy and sampling

Not straightforward to reach energy conservation… + problem of "cold" QM region
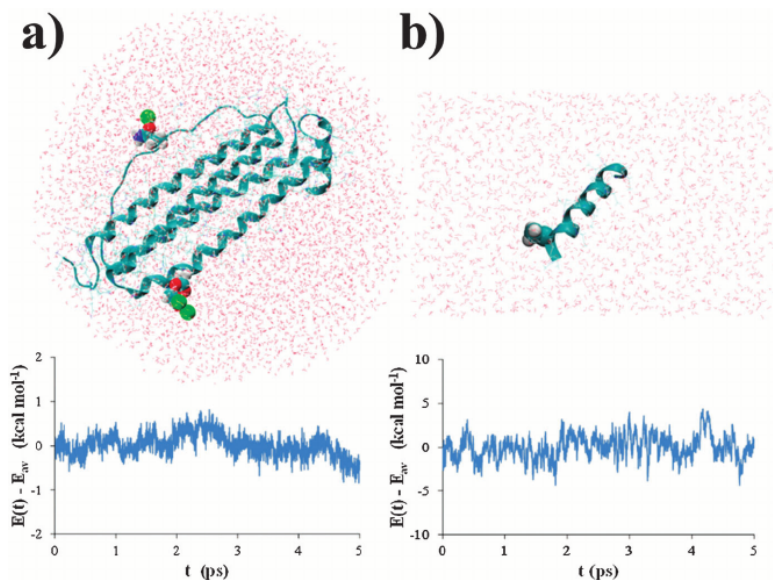
Torras, PCCP 2015, 17:9959

2 examples of QM/MM-MD : spectroscopy

A

PBE0

Static
Dynamic

B
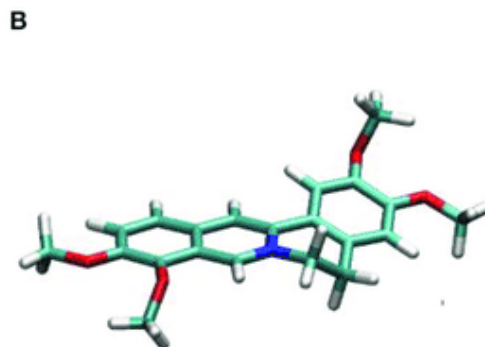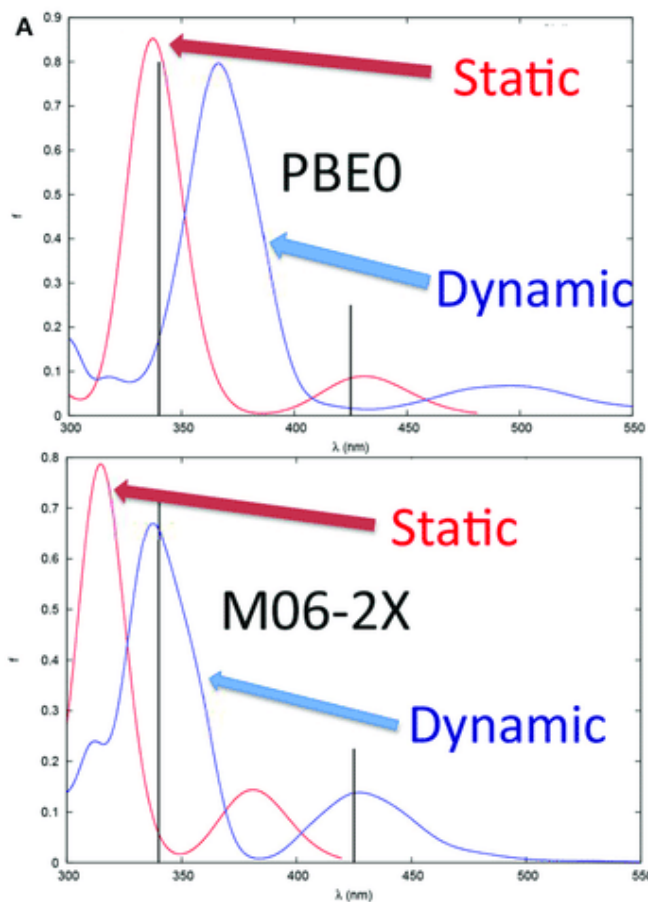
Palmatine is charged +1.
A bit of luck here…

M06-2X

Static
Dynamic
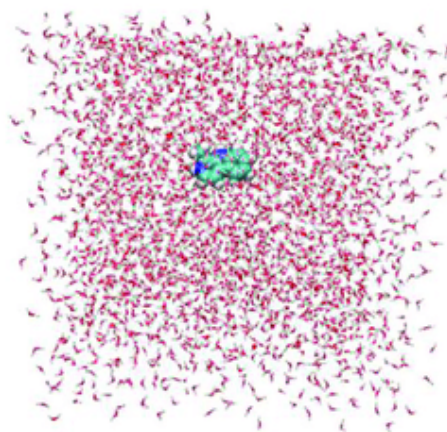
TDDFT/MM spectroscopy

M. Marazzi et al., Front. Chem. 2018, 6:86
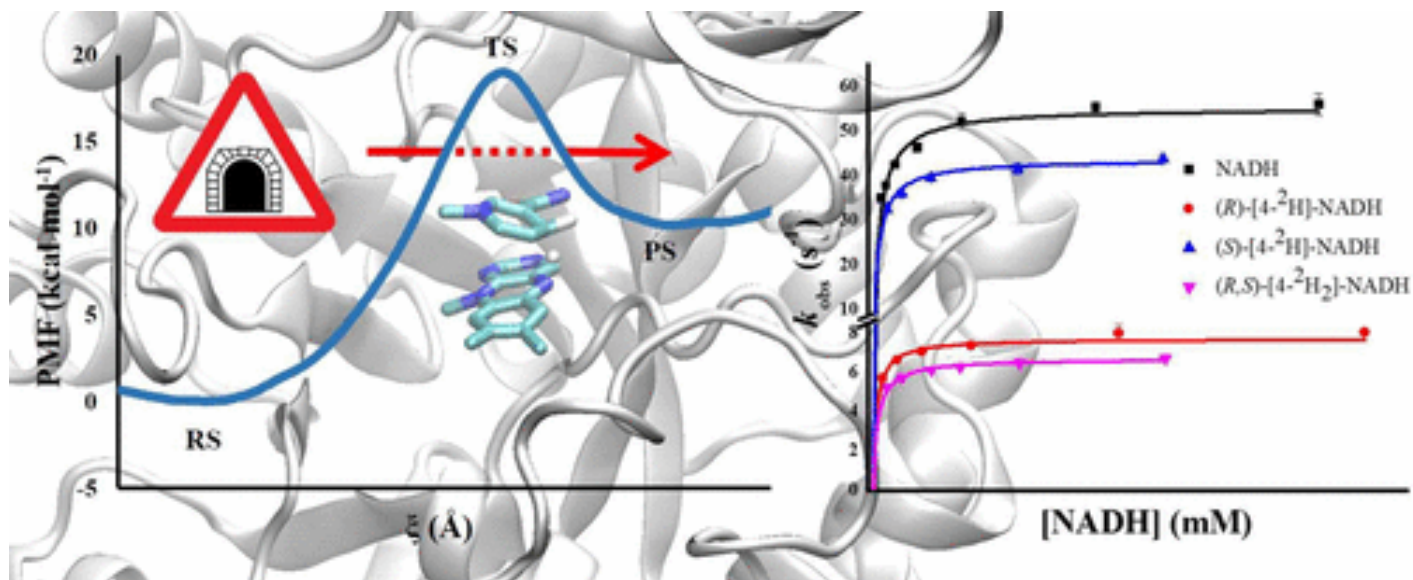
Tunon and coworkers, ACS catalysis, 2017, 7:3190 (hydride transfer)

"AM1"(M062X-recalibrated)/QM-MD potential mean force

Sometimes not trivial to find the right coordinate…

Adaptative QM/MM-MD
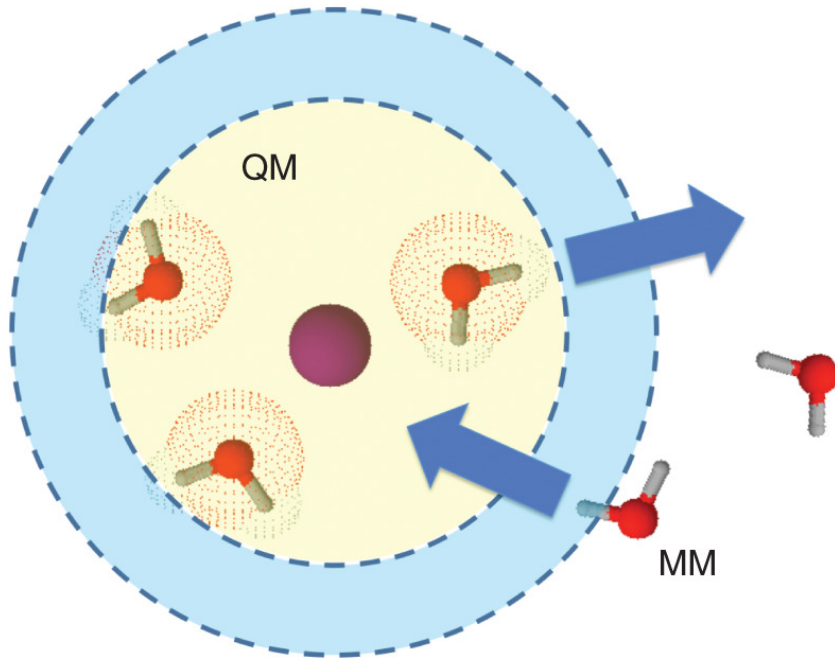
Account for dynamic exchange of water molecules along time

-> problem of energy conservation

Need to define a buffer region with a smoothing function

Also 10 times more expensive than QM/MM, but needed for proton transfer

Outline

1. Examples of QM/MM application: vision, magnetism, …

2. Merging MM and QM…

3. Situating the implicit models for solvation

4. Summary: implicit vs. QM/MM

Free solvation energies

- Thermodynamics

The *absolute* free energies of hydration are difficult to estimate

$$\Delta G^{solv} = \boxed{\Delta G^{cav}} + \boxed{\Delta G^{vdw}} + \boxed{\Delta G^{elec}}$$

Energetic penalty to create cavity

- In modeling, we have a microscopic point of view:
  solute-solute interactions + modified solvent-solvent interactions

Accurate estimation of ΔG^{solv} is very important in docking,
supramolecular chemistry…

How to treat a solvent ?

One needs to take into account the solvation for :
- Accurate conformations of molecules (whose potential energies surfaces are modified)
- Acid/base properties…
- Spectroscopic properties : solvatochromism
- Dipole moment are modified: $\mu(H_2O)$ = 1.8 D -> 2.5 D from gas phase to bulk

But the complete description by DFT is delicate… $\Delta\lambda$ if electrostatics dominates, $\Delta\lambda$ if van der Waals dominates

But DFT is calibrated (only) on static properties…

It is difficult to treat dispersion. For instance, *ab initio* molecular dynamics struggle for water description whereas it is straightforward with a force field

QM/MM approaches

Take-home message: solute-solvent interactions are not always isotropic

Generalization: interactions between the active site of a system and its surroundings are not always isotropic

Classical (Newton) molecular mechanics force fields are (generally):
- anisotropic
- empirical (highly parametrized)
- most often additive

$\Rightarrow$ looks good for qualitatively-correct interactions

Recall chemical transformation are most of the time localized

A smart coupling between quantum mechanics and molecular mechanics, aka QM/MM , may be thought as a good solution?

Implicit vs. explicit models
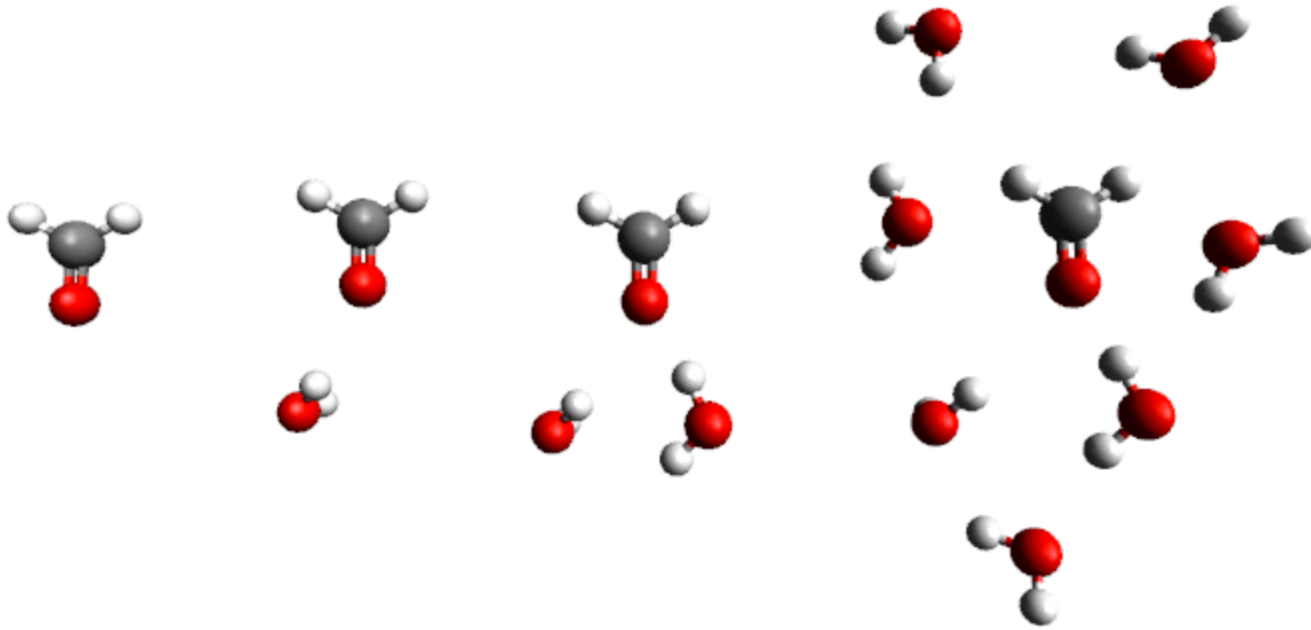
QM model	MM model
Micro-solvation	Infinite solvent (PBC)
Electron-based process	MD ; statistical properties
Solvent degrees of freedom are ignored	Electron degrees of freedom are ignored

Working with implicit solvent

Main idea: replace the (explicit) solvent degrees of freedom with a continuum

Application to electrostatics: first hald 20th century (Born, Onsager, Kirkwood)

Poisson equation for a charge distribution in vacuum:

$$-\varepsilon_0 . \Delta V(\vec{r}) = \rho(\vec{r})$$

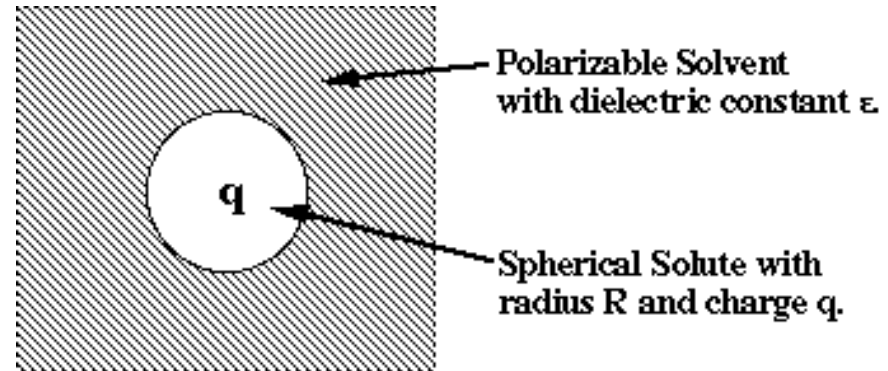Generalisation to an isotropic continuum with a dielectric constant ε

$$-\nabla(\varepsilon . \Delta V(\vec{r})) = 4\pi\rho(\vec{r})$$

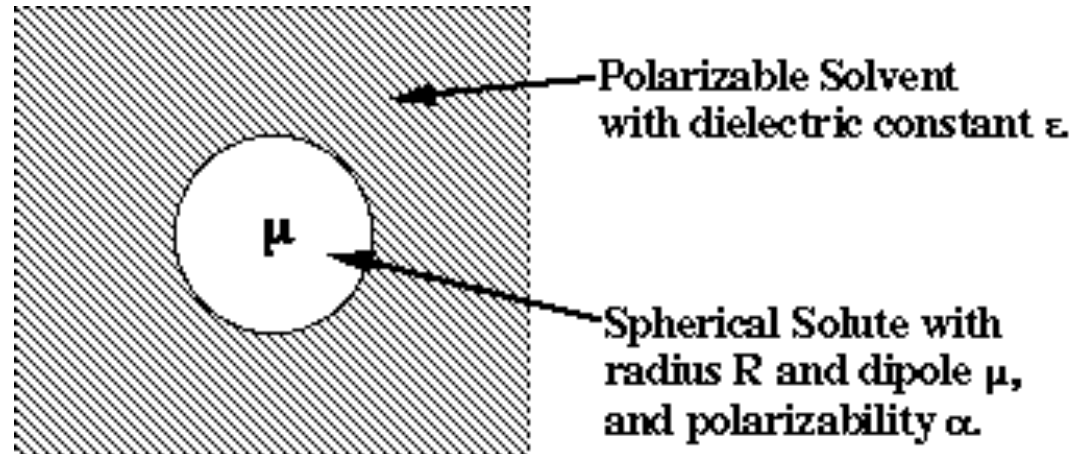In CGS units; these are non-linear equation, usually solved numerically

Born, Onsager model… (even if not so used)

Born

$$\Delta G_{elec} = -\frac{\varepsilon - 1}{2\varepsilon} \frac{q^2}{R}$$

Polarizable Solvent with dielectric constant ε.

Spherical Solute with radius R and charge q.

Onsager (including dipole)

Polarizable Solvent with dielectric constant ε.

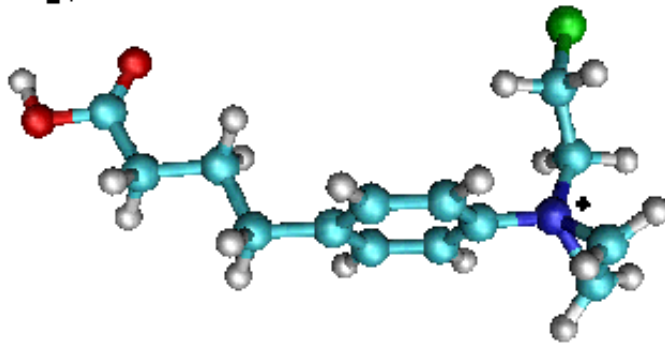Spherical Solute with radius R and dipole μ, and polarizability α.

$$\Delta G_{elec} = -\frac{\varepsilon - 1}{2\varepsilon + 1} \frac{\mu^2}{R^3} [1 - \frac{\varepsilon - 1}{2\varepsilon + 1} \frac{2\alpha}{R^3}]^{-1}$$

Pitfalls of Onsager model

The Onsager method can fail in case the electron distribution is poorly described by the dipole moment

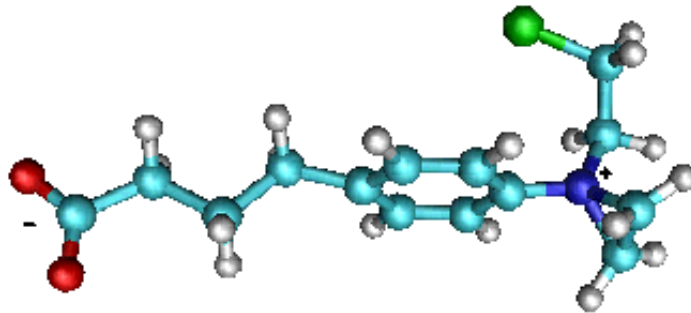**Chlorambucil Metabolite (Anticancer Drug):**
  Pk$_a$ predicted to be −153

Dipole = 18.0
Onsager DG$_{solv}$ = 19.3 kcal/mol

Dipole = 60.0
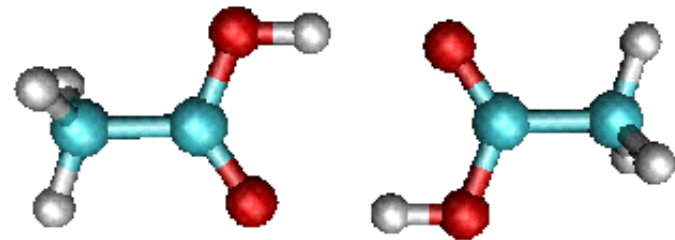Onsager DG$_{solv}$ = 190.2 kcal/mol

Carboxylate Dimer:
No Onsager Energy because dipole ⁓0 by symmetry.
Est. Solvation Energy (BEM) = −15 kcal/mol.

The continuum model

The hamiltonian of the solute (r_1) and the solvent (r_2)

The solvent model is a continuum characterized by its dielectric constant ε, polarizing the solute and polarized by the solute

$$H(r_1, r_2) = H^1(r_1) + \cancel{H^2(r_2)} + \boxed{H^{\text{int}}(r_{,1}, r_2)}$$

neglected

Replaced by a response function

$$V^{\text{int}}(r_1, Q(\vec{r}, \vec{r}\,'))$$

The continuum model: principles

1. The solute is described with any QM model
2- Solute-solvent interactions are mainly electrostatics

In reality, the other interactions may be important as well

3- The solution is highly diluted
4- The solvent is isotropic, in thermal equilibrium
5- We are only interested in the ground state
6- No dynamical effects
7- The solute lies in an empty cavity $\varepsilon=1$

The continuum model: the cavity
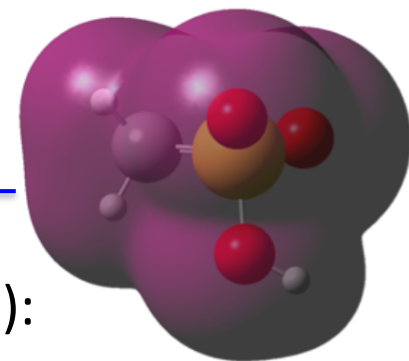
1. Defined using the electron density (from a DFT calculation): isodensity surface (constant or self-consistent)

2. Or defined by superposing spheres, using radii slighlty larger than the van der Waals ones (Pauling, Bondi, UFF…)

One cannot compare results obtained with different cavity models…

Solvent accessible surface (SAS)
vs. solvent excluded surface (SES)
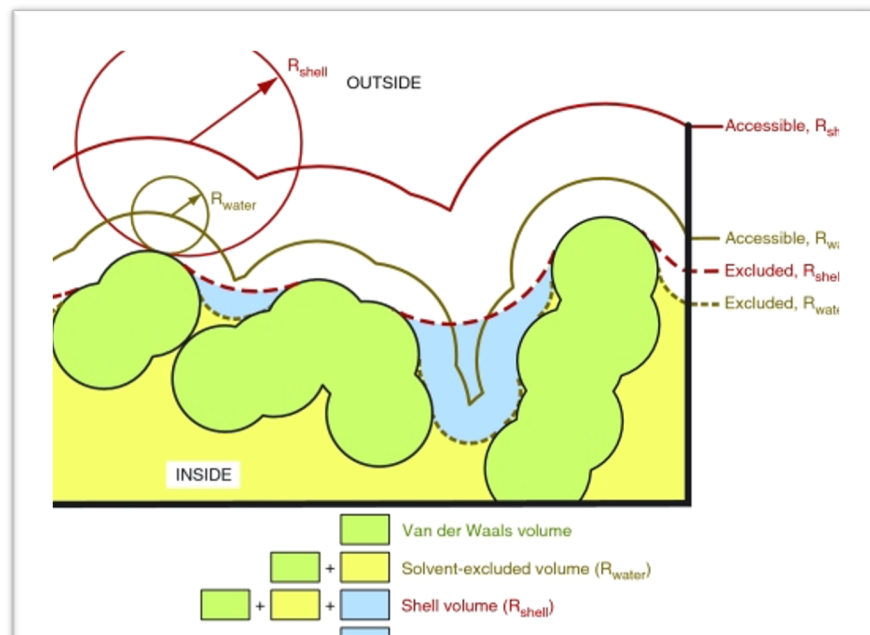
Solving the electrostatic equation
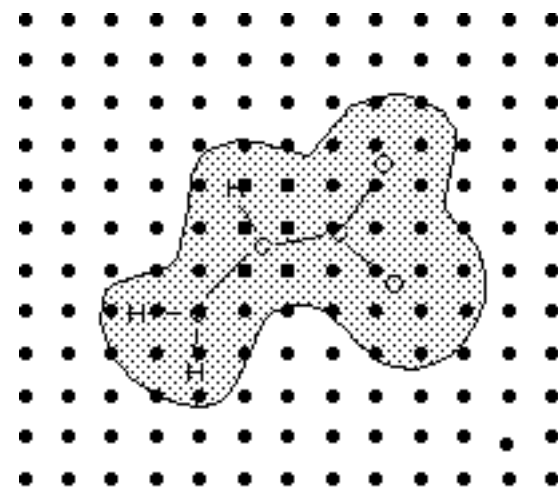
ρ   is the solute charge density (nuclei + electrons)

ρ   polarizes the continuum, which polarizes in turn ρ etc...
     -> non-linear problem, self-consistent reaction field (SCRF)
        iterative solution (that sometimes does not converge)

Poisson's equation

$$-\nabla^2 V_{in}(\vec{r}) = 4\pi\rho(\vec{r})$$ inside

$$-\varepsilon\nabla^2 V_{in}(\vec{r}) = 0$$ outside

Hypothesis: the charge outside the cavity is zero...

Solving the electrostatic equation

One imposes constraints of having:

$$\lim_{r \to \infty} r . V(\vec{r}) = \alpha$$

$$\lim_{r \to \infty} r^2 . V(\vec{r}) = \beta$$

α and β having finite values and
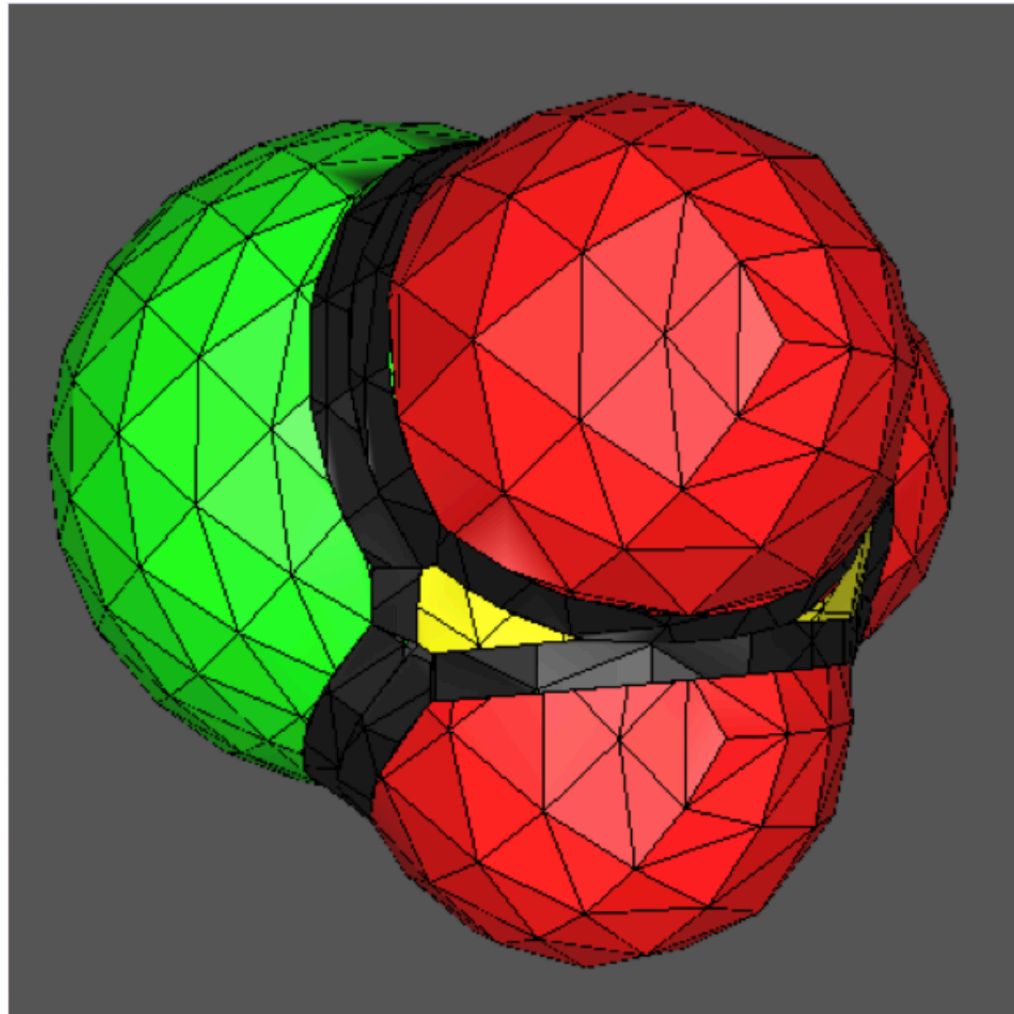α equality of V_{in} and V_{out} at the surface

One needs to discretize the electric field discontinuty by adding apparent surface charges (ASC)

$$V_R(\vec{r}) = V_\sigma(\vec{r}) = \int_r \frac{\sigma(\vec{s})}{|\vec{r} - \vec{s}|} d^2 s$$

$$V_\sigma(\vec{r}) \cong \sum_k \frac{\sigma(\vec{s}_k) A_k}{|\vec{r} - \vec{s}_k|} = \sum_k \frac{q_k}{|\vec{r} - \vec{s}_k|}$$
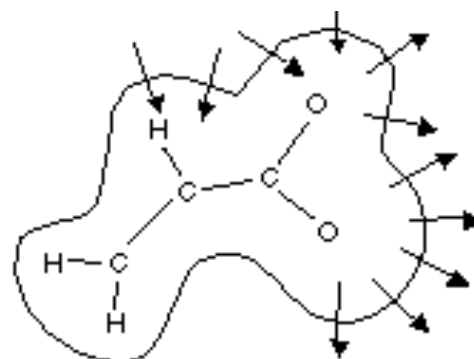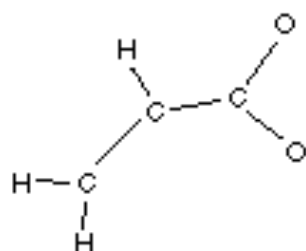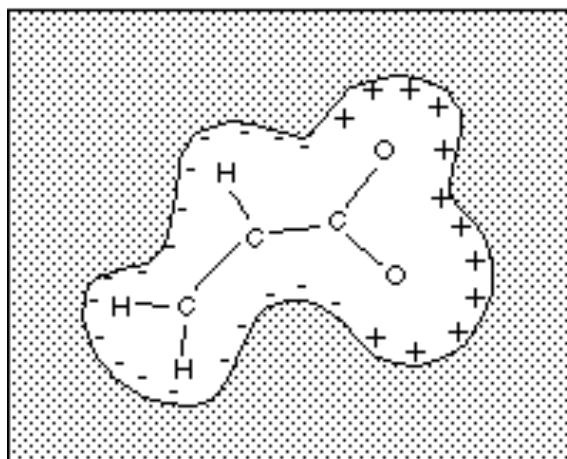
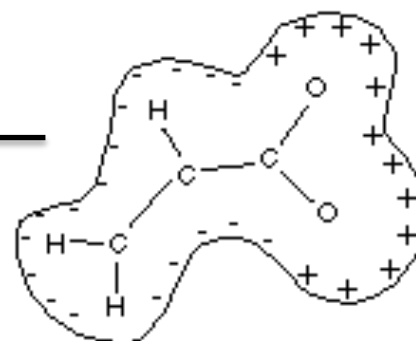# Tesserae

In a scheme...

1. Calculate the *in vacuo* QC wave function for the solute.

2. Calculate electric field normal to the SAS at grid points on the surface.

3. Calculate image charges on the SAS generated in dielectric continuum.

4. Add image charges to electronic Hamiltonian. Go to step 1..

The polarization vector is defined in each region I as:

$$\vec{P}_i(\vec{r}) = -\frac{\varepsilon_i - 1}{4\pi}\nabla V(\vec{r})$$

At the frontier between two regions i and j, an ensemble of ASC σ_{ij} are defined:

$$\sigma_{ij} = -(\vec{P}_j - \vec{P}_i).\vec{n}_{ij}$$

PCM case: $\varepsilon_i=1$ and $\varepsilon_j=\varepsilon$

$$\sigma(\vec{s}) = \frac{\varepsilon - 1}{4\pi\varepsilon}\frac{\partial}{\partial\vec{n}}V_{in}(\vec{s})$$

Other models

PCM: the most employed "thanks" to Gaussian

ICPM and SCI-PCM based on isodensity surfaces also possible

Conductor-like screening Model (COSMO): $\varepsilon \rightarrow \infty$
- The continuum is assimilated to a conductor $V(\overline{s}) = 0$
  $\sigma(\vec{s}) = f(\varepsilon)\sigma^*(\vec{s})$ avec $f(\varepsilon) = \dfrac{\varepsilon - 1}{\varepsilon + k}$
- k value ?
- works nicely for solvent with a large ε

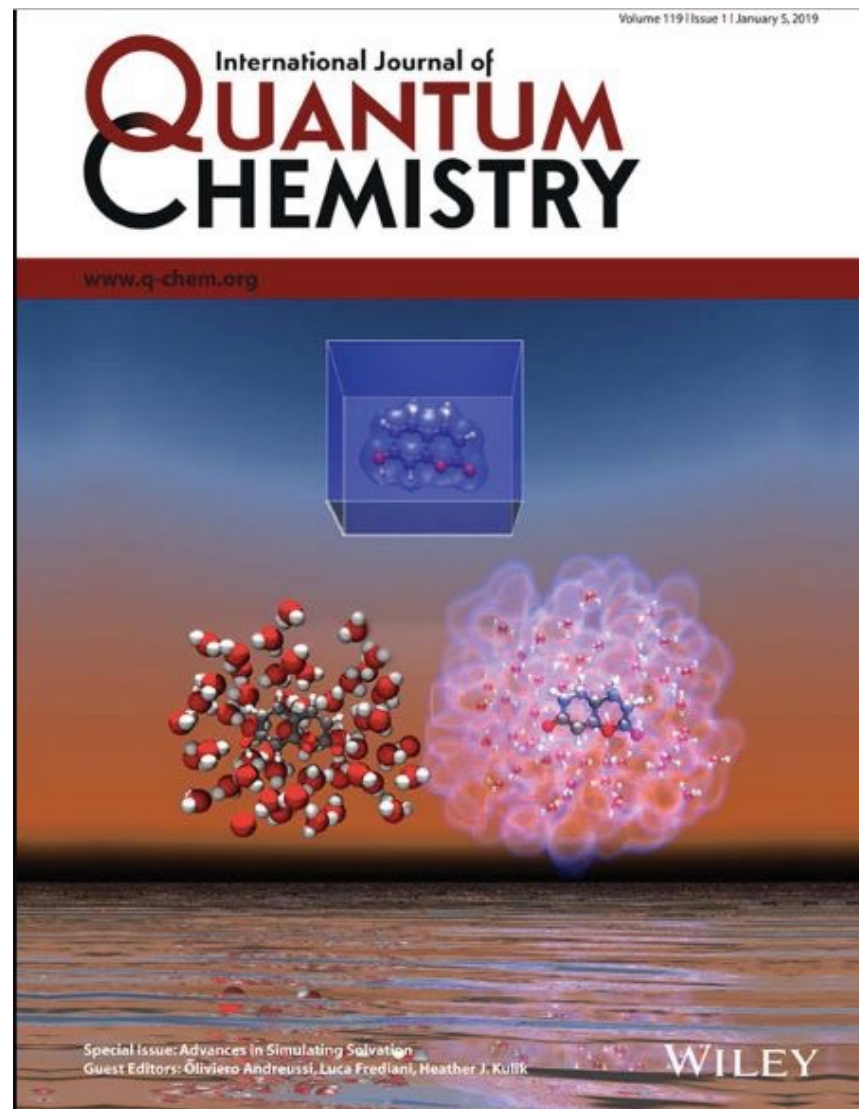PCM changed to integral equation formalism (IEFPCM) for all kinds of solvent
- Based on the Green function for a potential at position x induced by a charge at position y
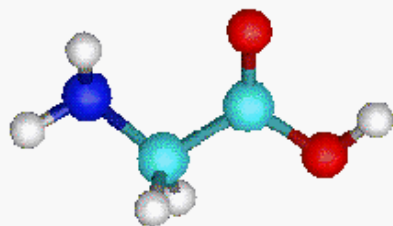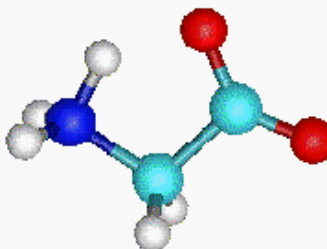
How to treat a solvent ?

Many different « flavors »
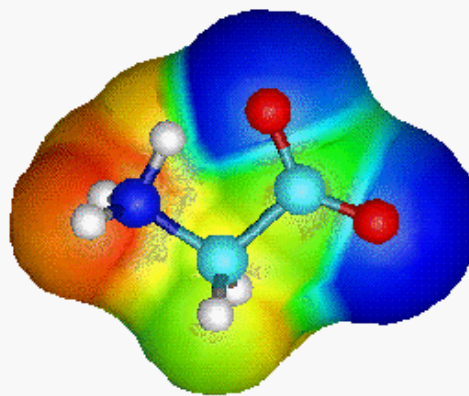
Polarizable Continuum Solvation Model
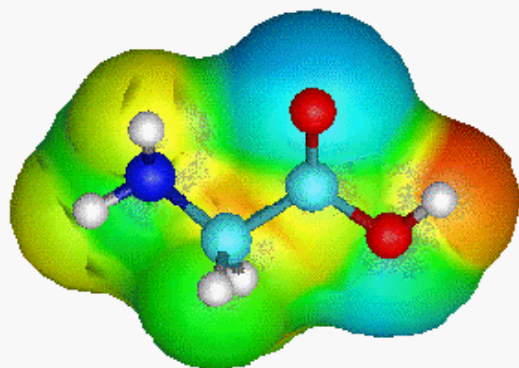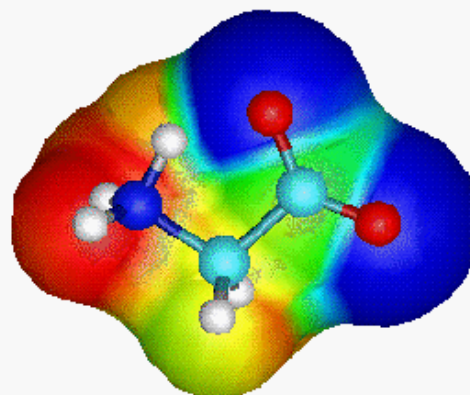
Neutral

Zwitterion

Relative Energies
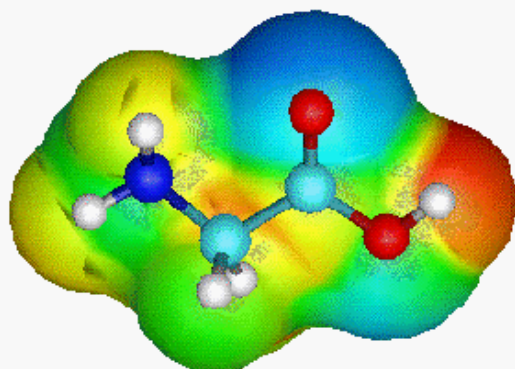neutral − zwitterion
(kcal/mol)

Gas Phase: −22.0

Zeroth order
reaction field: −4.9
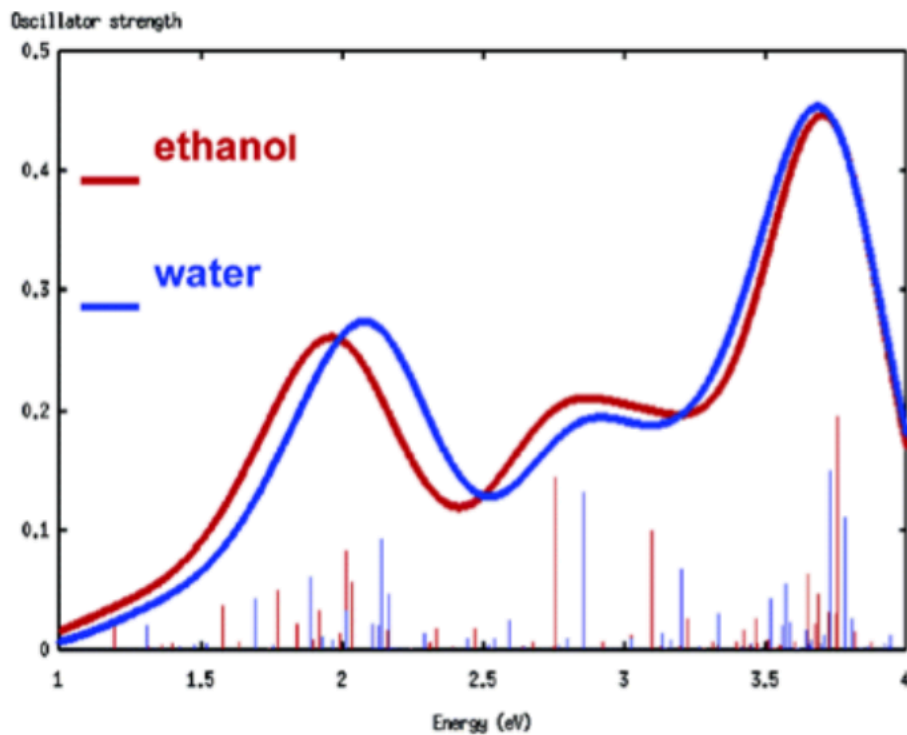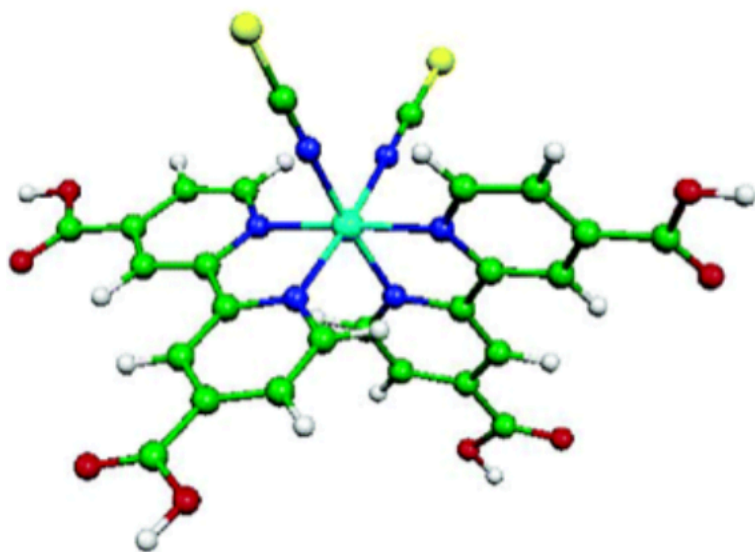
Self−consistent
reaction field: +13.5

Comparison of induced surface charges on glycine neutral and zwitterion. Relaxation of the wavefunction lowers the zwitterion solvation energy by nearly 20 kcal/mole relative to the neutral. The increased charge separation is evident in the surface charges of the zwitterion.

Useful for UV-Vis spectroscopy

Fantacci et al., JACS, 2003, 125, 4381

Comparison QM/MM+continuum vs. SCRF

Compound	QM/MM	SCRF	Expt.
Ethane	0.0	0.0	0.0
Benzene	0.3	-1.7	-2.6
Water	-8.3	-7.5	-8.1
Methanol	-6.2	-7.0	-6.9
Methylamine	-4.0	-6.4	-6.4
Acetic acid	-8.4	-8.9	-8.5
Acetate Ion	-80	-76	-79
Imidazole	-7.2	-12.2	-12,-6

SCRF does not describe explicitly HB, one relies on a fortitious compensation of errors

One can add a « certain » number of water molecules… why not many ?

(Data from Table 2, J. Gao, Reviews in Computational Chemistry, Vol. 7, pp. 119-185 (1996).)

Limitations of continuum approaches

« For systems as large as protein or a DNA fragment, the original assumption that continuum model provide a cheap way to introduce solvent effects is no longer valid, as the computational cost associated with the solution of the polarization equations, and in particular with the linear system obtained with some discretization techniques, can become unaffordable. »

Lipparini et al., J. Chem. Phys., 2014, 141, 184108

Supercomplex approaches

Adding explicit solvent molecules in the QM part to account for anisotropy (HB... but also charge transfer)

But one time we add a water molecule, the energy surface (PES or FES) becomes less smooth.

The energy minimization becomes meaning less at some point and the initial placement of the water molecules arbitrary.

-> a molecular dynamics scheme is needed...