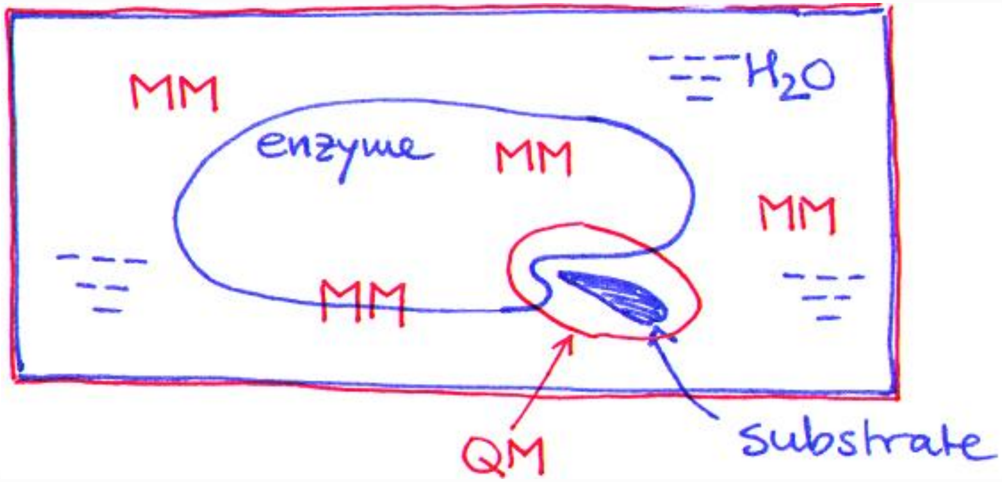
QM/MM

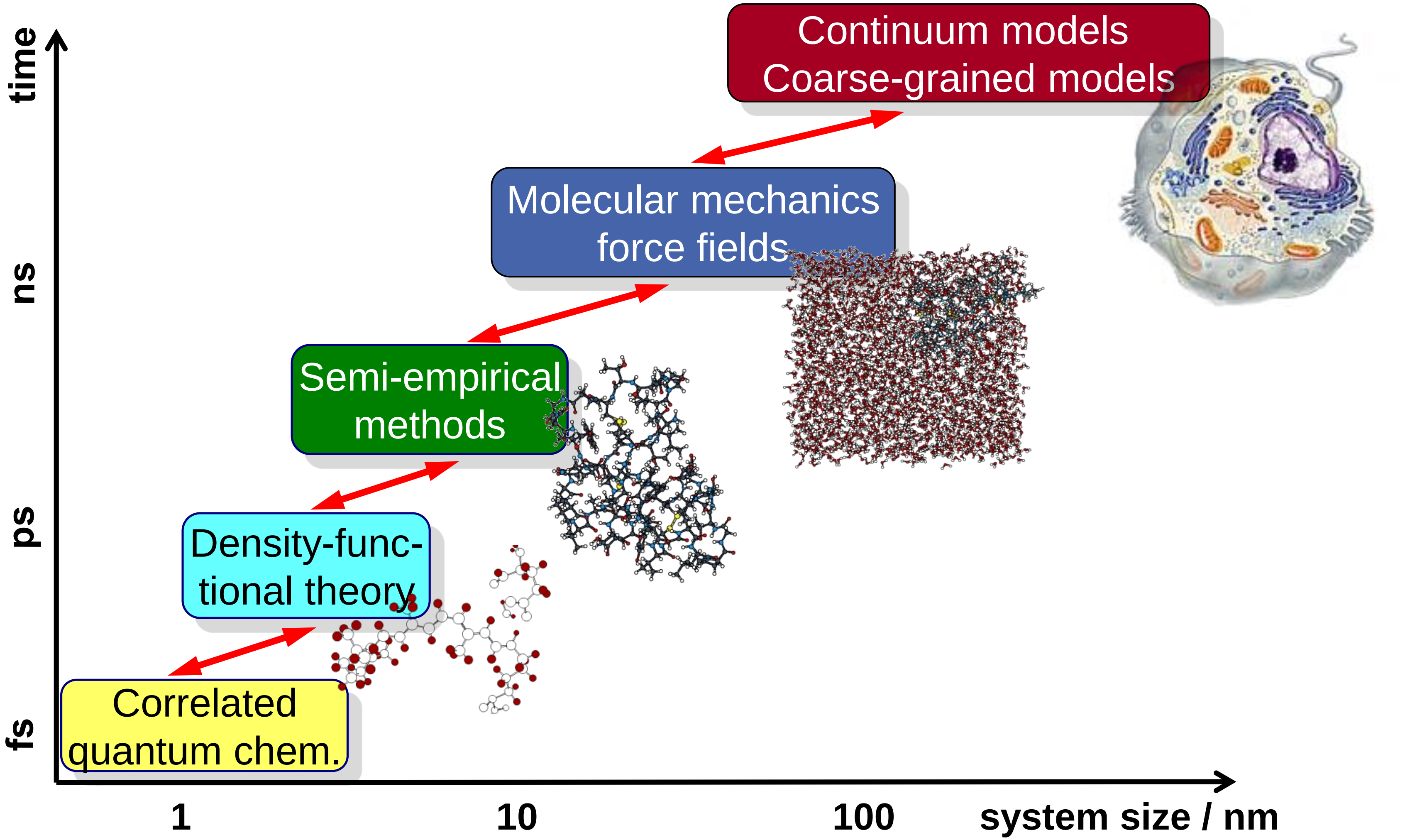
Marcus Elstner and Tomáš Kubař

June 26, 2018

Institute of Physical Chemistry & Center for Functional Nanostructures



Computational chemistry



Correlated quantum chem.

Density-functional theory

Semi-empirical methods

Molecular mechanics force fields

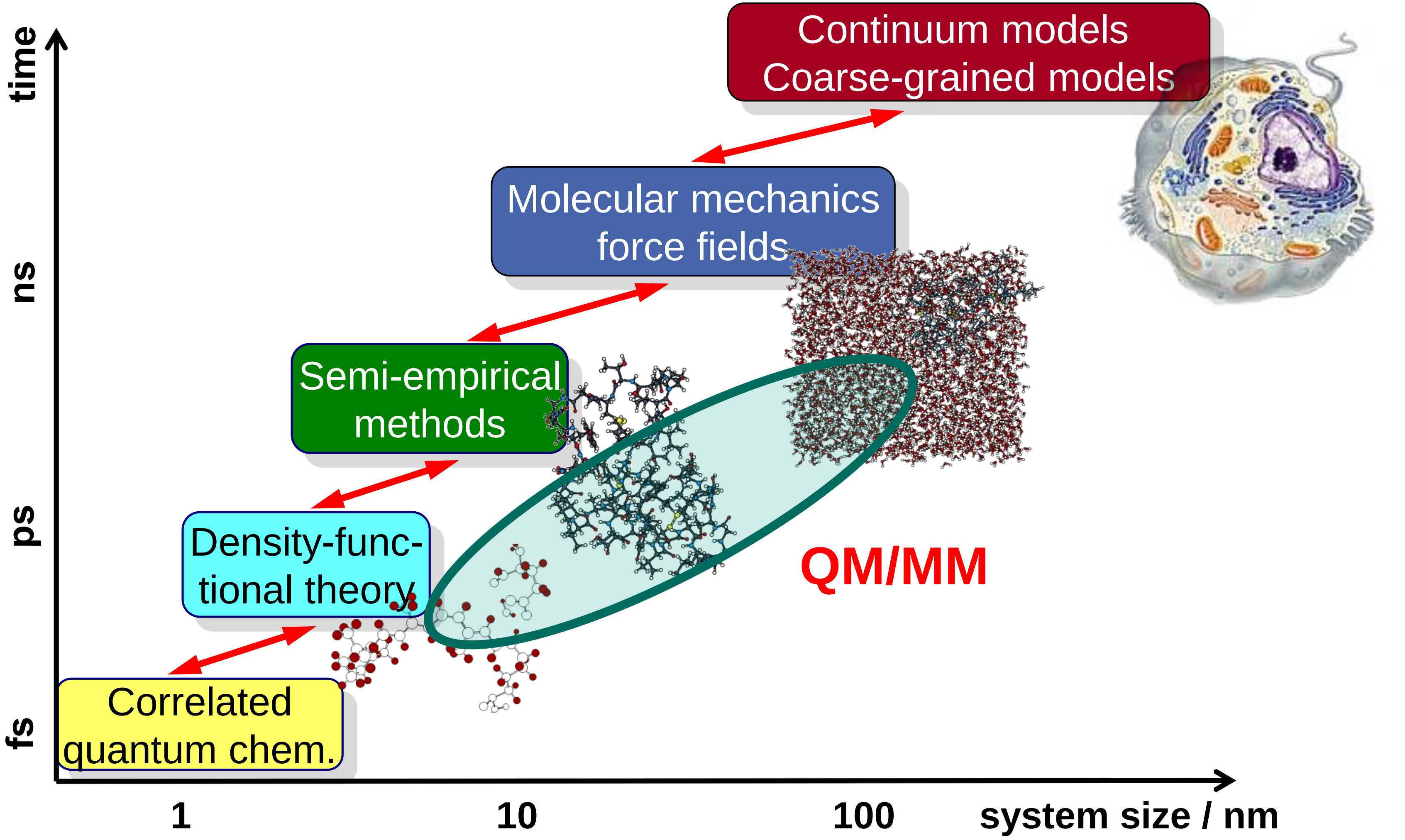
Continuum models
Coarse-grained models

Computational chemistry

- Molecular mechanics (force fields)
  - Large efficiency
  - Crude approximations necessary
  - Constant electron density → covalent bonds cannot be created / broken
  - Chemical reactions cannot be described
  - Problems also if electron density differs for different conformations

- Quantum chemistry
  - No or little assumptions → large transferability
  - Description of chemical reactions possible
  - Larger computational cost
  - Semi-empirical methods – involve parameters, compromise (decreased transferability, increased efficiency)

Computational chemistry



Continuum models
Coarse-grained models

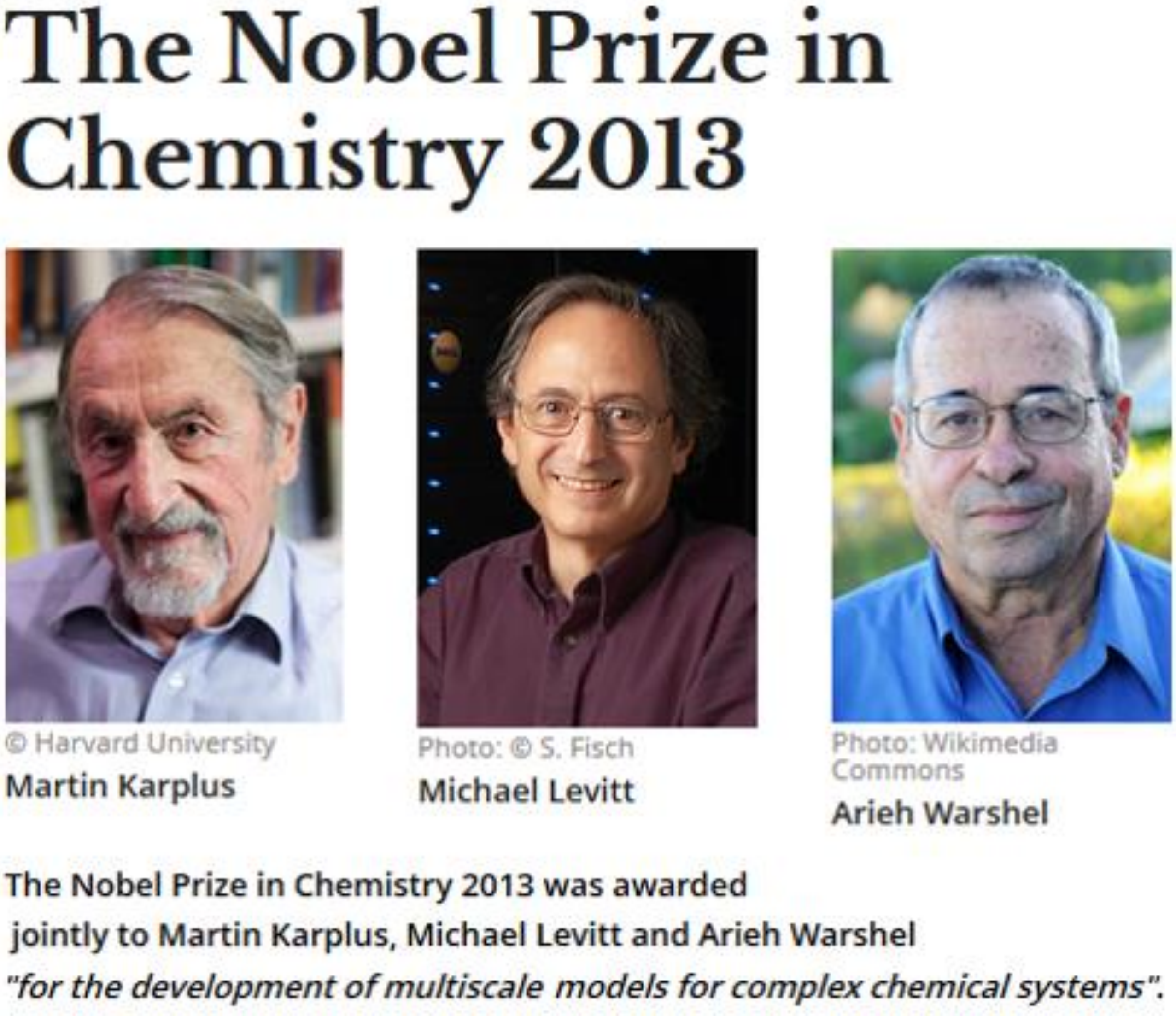
Molecular mechanics
force fields

Semi-empirical
methods

Density-functional theory

Correlated
quantum chem.

QM/MM



The Nobel Prize in Chemistry 2013

© Harvard University
Martin Karplus

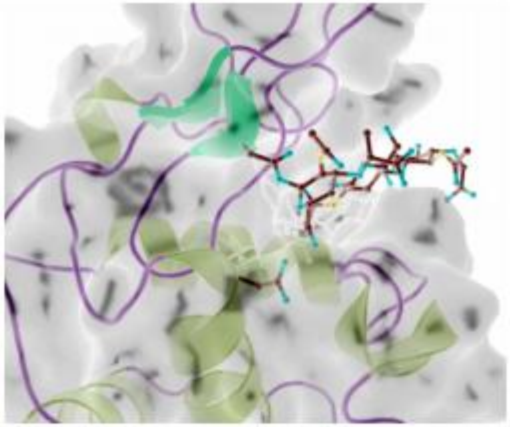
Photo: © S. Fisch
Michael Levitt

Photo: Wikimedia Commons
Arieh Warshel

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".***

Hybrid QM/MM methods

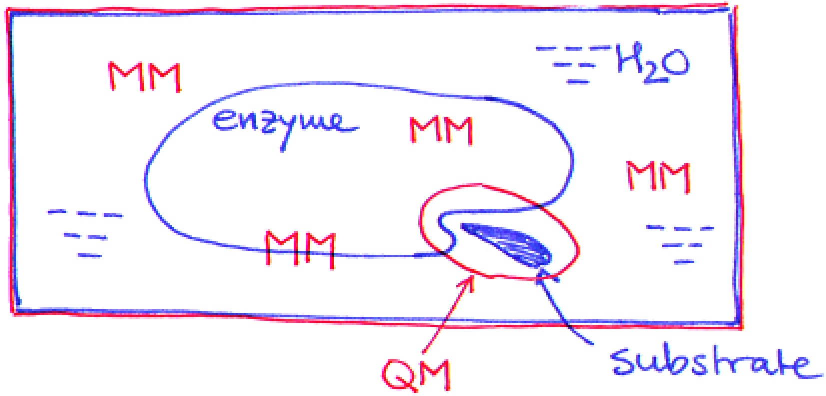
- Assumption – changing electron density localized
  in a small part of the molecular system
  – e.g. the binding site of a ligand in an enzyme



- Only a small number of atoms are involved in a chemical reaction
- The large remainder of the system stays outside of the process,
  but still may interact by way of non-covalent interactions

Hybrid QM/MM methods

- Assumption – changing electron density localized
  in a small part of the molecular system
  – e.g. the binding site of a ligand in an enzyme



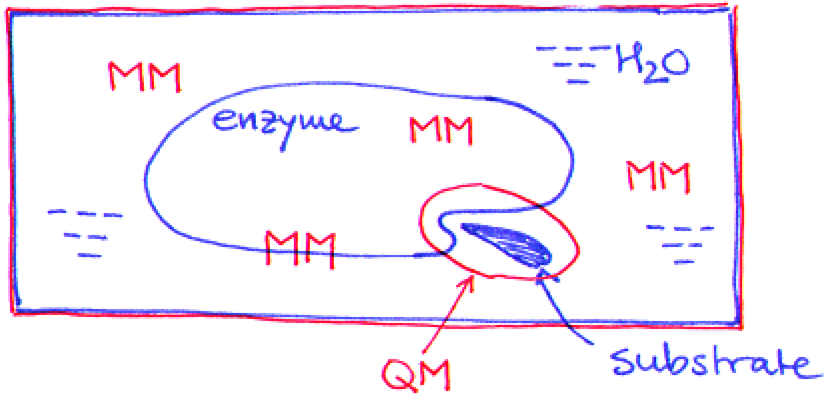
- Region where chemical reaction occurs – quantum chemistry (QM)
- The large remainder – MM force field
- Interaction of both regions must be accounted for – QM/MM / embedding

$$E_{\text{total}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}}$$

Hybrid QM/MM methods

- Interaction of both regions must be accounted for – QM/MM / embedding

$$E_{\mathrm{total}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}}$$



- The QM and MM calculations – performed in the usual way
- Key issue – how to couple the QM and MM regions?

  Goal – to obtain $E_{\mathrm{QM/MM}}$

Embedding schemes

- Many various schemes proposed
- Differences – how much of the QM–MM interaction is neglected

In the order of increasing complexity / completeness:

- Mechanical embedding
- Electronic embedding
- Polarized embedding

Mechanical embedding

- Unpolarized interactions

- Entirely independent QM and MM calculations
- Interaction between the QM and MM regions – with a force field

$$E_{\mathrm{QM/MM}} = \sum_i^{\text{QM atoms}} \sum_m^{\text{MM atoms}} \left(\frac{q_i^{\text{Mull}} \cdot q_m}{r_{im}} + 4\varepsilon_{im} \left(\frac{\sigma_{im}^{12}}{r_{im}^{12}} - \frac{\sigma_{im}^{6}}{r_{im}^{6}} \right) \right)$$

- QM atoms have to be assigned Lennard-Jones parameters and their quantum-chemical charges are used (e.g. Mulliken charges)

- Specific combinations of force fields and quantum-chemical methods lead to very good results for specific classes of molecules / reactions
- Generally, care must be taken

Mechanical embedding

- Problem – the electron density in the QM region
  is unaffected by (‘does not feel’) the MM region
  – therefore ‘unpolarized’ scheme

- As if the QM region was calculated ‘in vacuum’
  – this may deviate from reality considerably
  – imagine that the reaction center of the enzyme is surrounded
  by charged amino acid side chains – affects the reaction strongly

Electronic embedding

- Polarized QM / unpolarized MM

- Electrostatic interaction with the MM atomic charges
  is included in the QM Hamiltonian
- Works with any electronic structure method
  – HF, DFT, correlated wave-function, semi-empirical…

Electronic embedding

Interaction of QM electrons with MM point charges

- moves from the QM/MM term (described with a force field)
  to the quantum chemical contribution
- the charge density interacts with a set of point charges
- *little* increase of the computational cost

$$\hat{H}'_{\mathrm{QM}} = \hat{H}_{\mathrm{QM}} - \sum_{j}^{\text{QM electrons}} \sum_{m}^{\text{MM atoms}} \frac{q_m}{r_{jm}}$$

Interaction of QM nuclei with MM point charges may remain in QM/MM

$$E'_{\mathrm{QM/MM}} = \sum_{i}^{\text{QM atoms}} \sum_{m}^{\text{MM atoms}} \left(\frac{Z_i \cdot q_m}{r_{im}} + 4\varepsilon_{im} \left(\frac{\sigma_{im}^{12}}{r_{im}^{12}} - \frac{\sigma_{im}^{6}}{r_{im}^{6}} \right) \right)$$

Electronic embedding

- A good approach with chances for acceptable accuracy
- Possible imbalance
  - – QM region is polarized by MM
  - – MM region cannot be polarized
    because of the constant atomic charges

Polarized embedding

- Fully polarized interactions

- Involves a polarizable MM force field
  - Every MM atom is assigned a polarizability α
  - Induced point dipole on the atom obtained from the field intensity $\vec{E}$:

$$\vec{\mu}^{\mathrm{ind}} = \alpha \cdot \vec{E}$$

  - To be calculated iteratively until self-consistence
  - Possible problems – computational cost, difficult convergence of dipoles

- The induced dipoles interact with MM point charges
  as well as with the QM electron density and nuclei
- Both QM and MM calculations are iterative
  → convergence of both is required!

Polarized embedding

- Formally the most complete approach
- Dramatically increased computational cost

- Question: Does the improvement of the quality of results justify the extreme cost of calculation?
- Electronic embedding may be sufficient in many cases

QM/MM study – work plan

1. Define the QM region
   – shall involve all of the interactions for which MM is insufficient

2. Choose the QM (and MM) methods
   – QM is the crucial and by far most consuming part of calculation
   – accuracy as well as efficiency shall be judged

3. Determine the type of embedding (perhaps electronic)
   – also introduce LJ parameters for QM atoms (usual MM, or special)

4. Simulate
   – each MD step involves 1 QM calculation (and 1 MM, 1 QM/MM)
5. Analyze

Covalent bond crossing the QM/MM boundary

- What shall be done if a QM atom and an MM atom are bound covalently?
- Usually inevitable in studies of proteins…

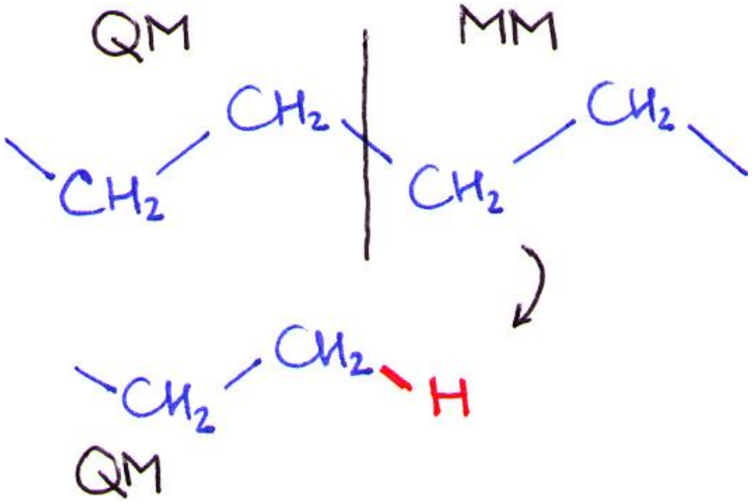
- Special treatment of the boundary is necessary
  (because the QM region is no real ‘molecule’ etc.)

- Linear combination of molecular fragments
- Link atoms (additive or subtractive coupling)
- Frozen orbitals
- other, further, current development…

Link atom approach

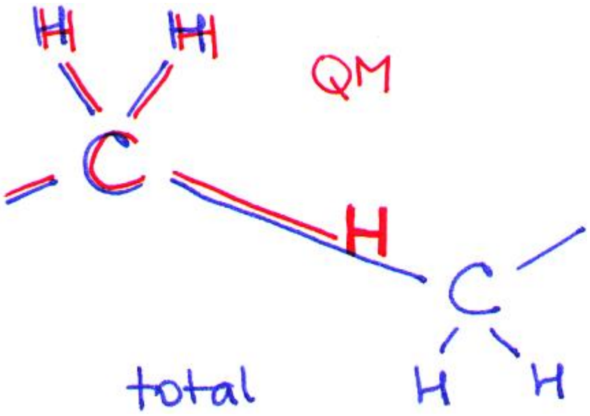
- We need a description suitable even if there is QM–MM electrostatics
- Typically in proteins – polar and charged amino acids in the MM region
  – polarization of the electron density of the QM region
- Missing in the electronic embedding
  – description of covalent bonds crossing the QM–MM boundary

- Link atom – replaces the covalently bonded MM system at the boundary



Link atom approach

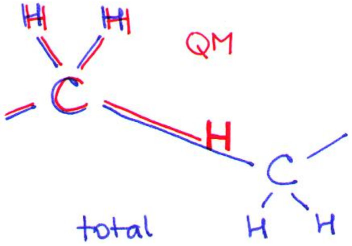
- Usually:
  - bonds between two sp^3-carbon atoms chosen to be cut
  - hydrogens used as link atoms (similar electronegativity of C and H)
    – also methyl ‘atoms’ may be used with some quantum chem. methods
- Desirable – define the QM region such that
  the bonds to be cut are as unpolar as possible (to minimize errors)
- Link atom – placed on the original C–C bond in a typical C–H distance



Link atom approach

Additive coupling

- Bonded interactions crossing the QM–MM boundary calculated with the MM force field
- Sometimes omitted – angles involving 2 QM atoms and 1 MM atom, dihedrals with 3 QM atoms and 1 MM atom



- MM atoms situated close to the QM system
  - almost overlap with the link atom
  - unphysically large influence on the electronic structure of the QM system
  - their charges may be scaled down or even removed
  - good alternative 1: replace point charges by gaussian charge distributions
  - good alternative 2: divide the charge among not-so-close MM atoms (motivation: the total charge should remain, and should not move far)

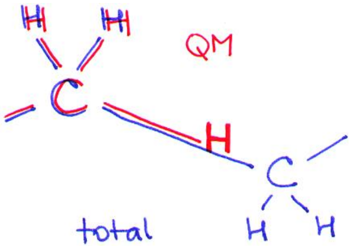
Link atom approach

Subtractive coupling

- Two MM calculations: one for the entire system, and one for the QM region including the link atom(s) – the same as for the QM calculation

$$E_{\text{total}} = E_{\text{MM}}^{\text{large}} + \left(E_{\text{QM}}^{\text{small+L}} - E_{\text{MM}}^{\text{small+L}}\right)$$

(large = entire system, small+L = QM region)

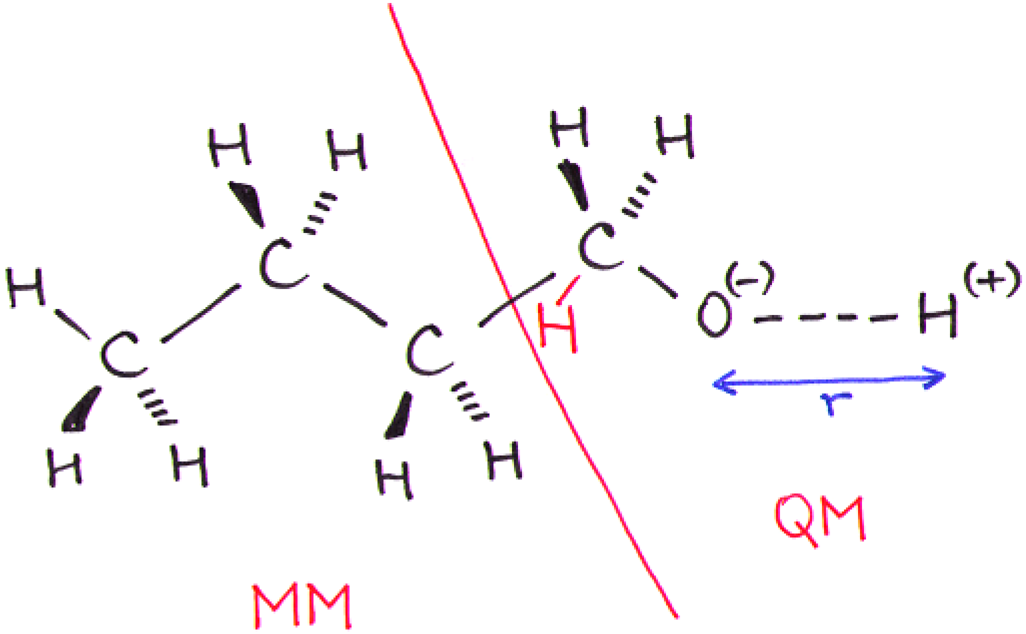


- May be deficient in some applications…

Link atom approach

Deficiency of subtractive coupling – example

- Reaction $$\mathrm{CH_3CH_2CH_2CH_2OH \rightarrow CH_3CH_2CH_2CH_2O^- + H^+}$$
- QM region – methanol fragment, MM region – remainder
- Goal – evaluate the energy as a function of the O–H distance *r*



Link atom approach

Deficiency of subtractive coupling – example

- Reaction $$\mathrm{CH_3CH_2CH_2CH_2OH \rightarrow CH_3CH_2CH_2CH_2O^- + H^+}$$
- Assume – parameters for methanol and butanol differ
  only in the force constant *k* for the O–H bond;
  all remaining terms give a constant independent of *r*

$$
\begin{aligned}
E^{\mathrm{QM/MM}}_{\mathrm{butanol}}(r) &= E^{\mathrm{QM}}_{\mathrm{methanol}}(r) + \left(\frac{1}{2}k^{\mathrm{OH}}_{\mathrm{butanol}} \cdot (r - r_0)^2 - \frac{1}{2}k^{\mathrm{OH}}_{\mathrm{methanol}} \cdot (r - r_0)^2\right) + \mathrm{const.} \\
&= E^{\mathrm{QM}}_{\mathrm{methanol}}(r) + \frac{1}{2}\left(k^{\mathrm{OH}}_{\mathrm{butanol}} - k^{\mathrm{OH}}_{\mathrm{methanol}}\right) \cdot (r - r_0)^2 + \mathrm{const.}
\end{aligned}
$$

- MM energy – retains the form of r^2
- QM energy – for large *r*, proportional to 1/*r* due to Coulomb’s law

Link atom approach

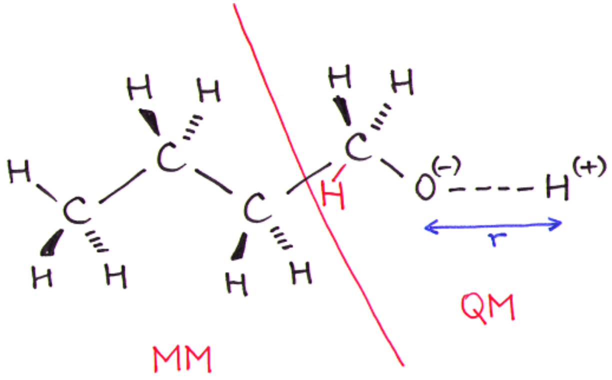
Deficiency of subtractive coupling – example

- Reaction $$\mathrm{CH_3CH_2CH_2CH_2OH} \rightarrow \mathrm{CH_3CH_2CH_2CH_2O^-} + \mathrm{H^+}$$
- Asymptotic behavior of total energy:

$$\lim_{r\to\infty} E^{\mathrm{QM/MM}}_{\mathrm{butanol}}(r) = \lim \left(-\frac{1}{r} + \frac{1}{2} k \cdot r^2 \right) = \lim r^2 = \infty$$

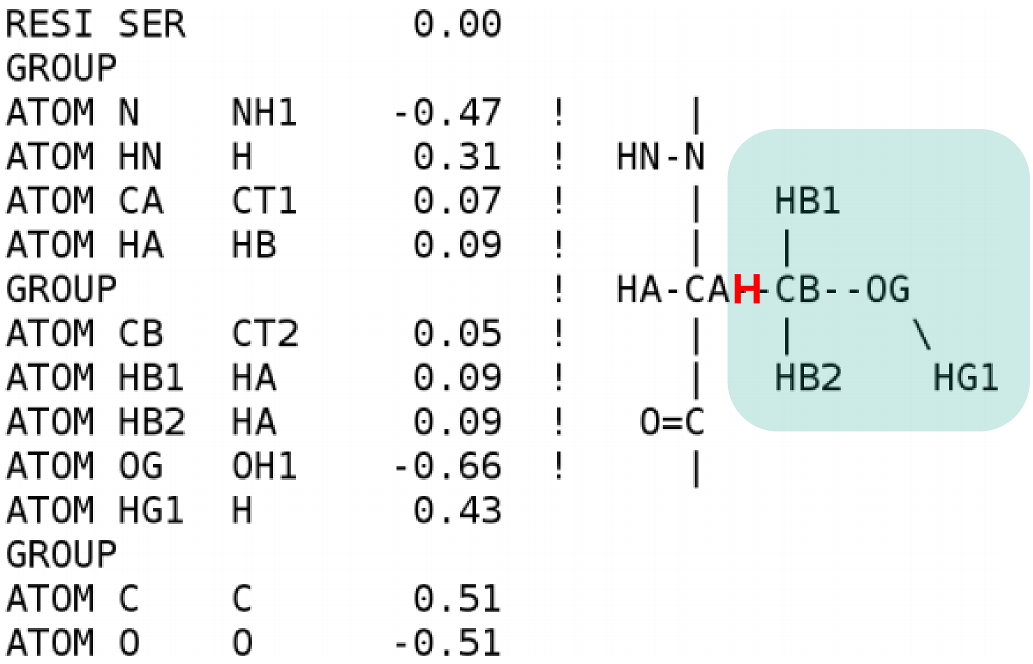
- Slightly different parameter values (really possible)
  → total energy grows over all limits
  – absolutely useless calculation

- Additive coupling – no such problem
  - methanol would be calculated with QM
  - propyl would be calculated with MM
    – MM parameters for the O–H bond not required!



Link atom approach

- Example – QM is an amino-acid side chain (starting from CB)



```
RESI SER          0.00
GROUP
ATOM N    NH1    -0.47 !     |
ATOM HN   H       0.31 !  HN-N
ATOM CA   CT1     0.07 !     |   HB1
ATOM HA   HB      0.09 !     |   |
GROUP                  !  HA-CAH-CB--OG
ATOM CB   CT2     0.05 !     |   |     \
ATOM HB1  HA      0.09 !     |   HB2    HG1
ATOM HB2  HA      0.09 !   O=C
ATOM OG   OH1    -0.66 !     |
ATOM HG1  H       0.43
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
```

- CA is very close to the link atom (which is between CA and CB)
- The point charge on CA would disturb the QM region drastically
- Simply removing CA would break electroneutrality

Link atom approach

Possible solutions to the problem:

- ‘exgroup’ – remove all charges of the charge group close to QM region
  - Here – CA, HA, N, H
  - Drastic – many important electrostatic interactions lost (hydrogen bonds…)

- ‘div’ – divide the charge of CA among the remaining atoms in the group
  - Here – q: CA = 0.07
  - After re-distribution: CA=0, HA=0.11, N=−0.44, HN=0.33
  - Strictly possible only if QM–MM boundary lies between two charge groups
  - The idea may be adapted to each specific case

Link atoms or something else?

- Link atoms – may cause problems with non-bonded interactions
  - possibly extremely close to nearby MM atoms
  - representation of charge densities by MM point charges
    may lead to inaccuracy / computational instability

- Alternative – introduce no new *atoms*,
  rather treat the *orbitals* on the QM/MM boundary in a special way
- Shape of these orbitals can be held constant during the simulation
  → 'frozen orbitals'

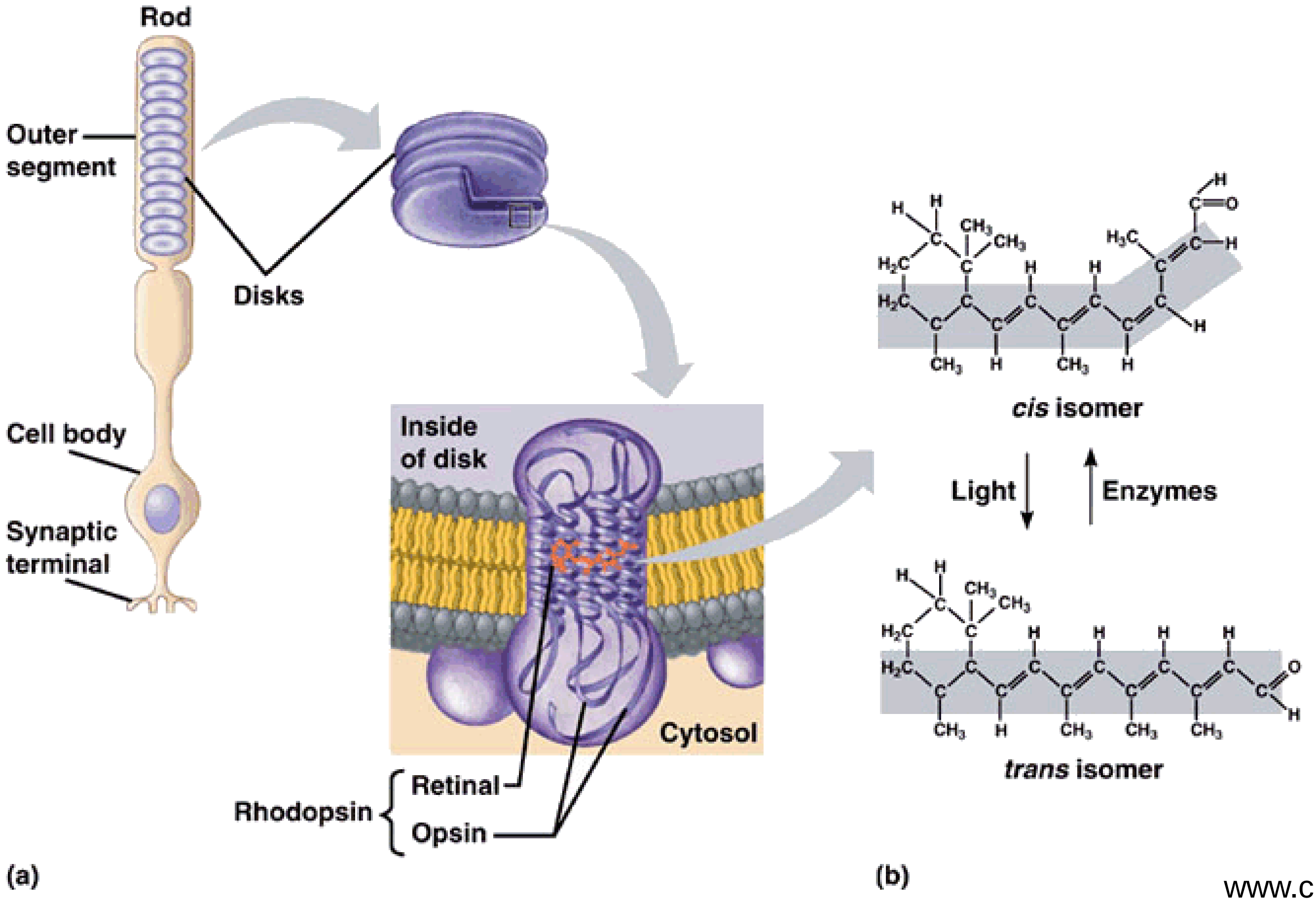
Beyond QM/MM

- Divide the system even more – 3 disjunctive regions, QM1/QM2/MM
- QM1 – advanced, expensive electronic structure method
  (highly correlated like CC, CAS…)
- QM2 – more approximative and efficient electronic structure method
  (DFT or even semi-empirical)
- MM

- Example – ONIOM in the Gaussian package

Choice of the QM region – illustration

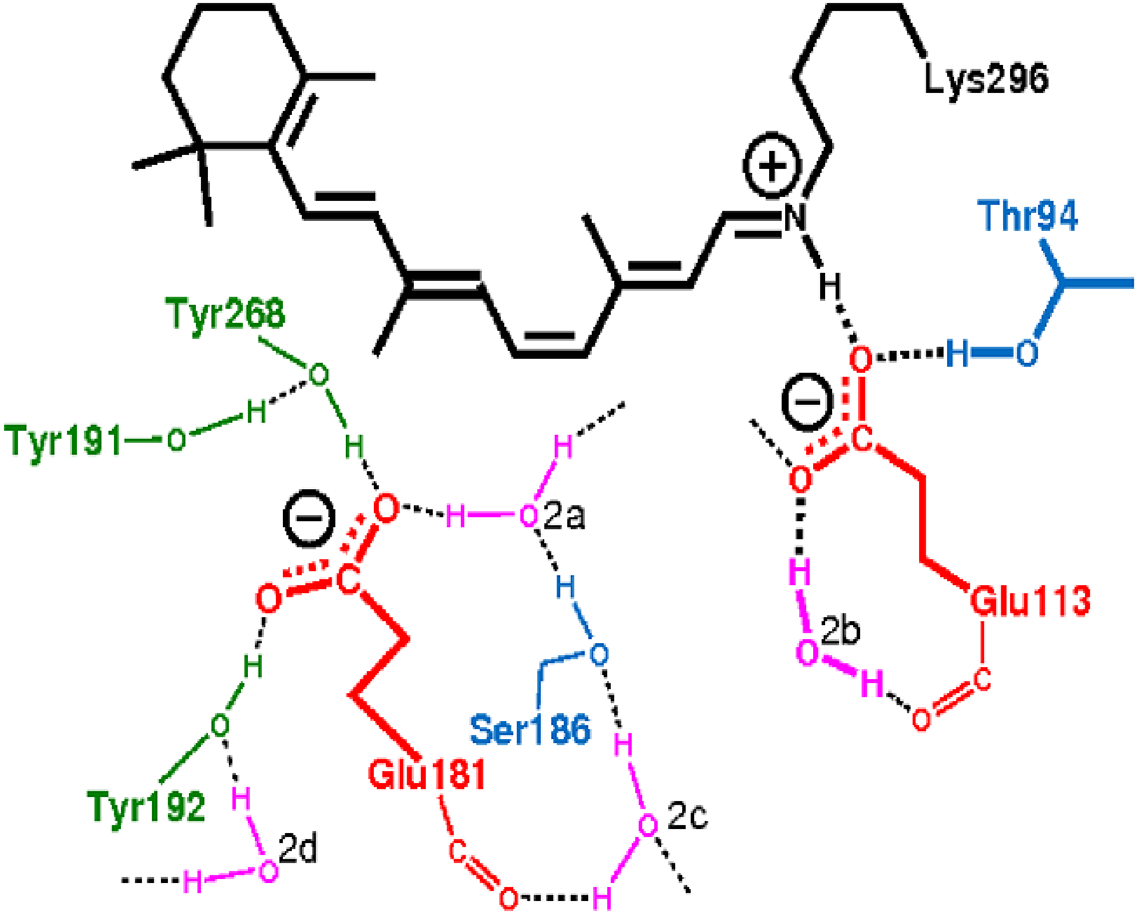
- Retinal – polyene bound covalently via a protonated Schiff base to a lysine side chain in a protein from the rhodopsin family



www.chm.bris.ac.uk

Choice of the QM region – illustration

- Many different choices possible



- polyene (ring to NH):
  bad, boundary cuts a polar bond
- retinal + CH2
  bad, link atom too close
  to the important region
- retinal + sidechain to CB
  fair, but no charge transfer
  to Glu113 possible
- + counterion Glu113
  better, but no charge transfer
  to Wat possible
- + Wat2b+Thr94
  good, but no polarization
  at Glu181
- + Glu181
  very good, but…

Adaptive QM/MM

- Problematic situation – molecules entering and leaving QM zone
- Exchange of QM and MM characters desirable
- Naïve exchange impossible due to discontinuities:
- Spatial – structure changes abruptly at the QM–MM boundary
- Temporal – forces change abruptly upon an exchange of characters

- 1st gen. – alleviates spatial discontinuity, temporal discontinuity persists
- 2nd gen. – several partitionings, they however contain different numbers of QM molecules (QM core region, and transition region), and there is a certain required number of partitionings

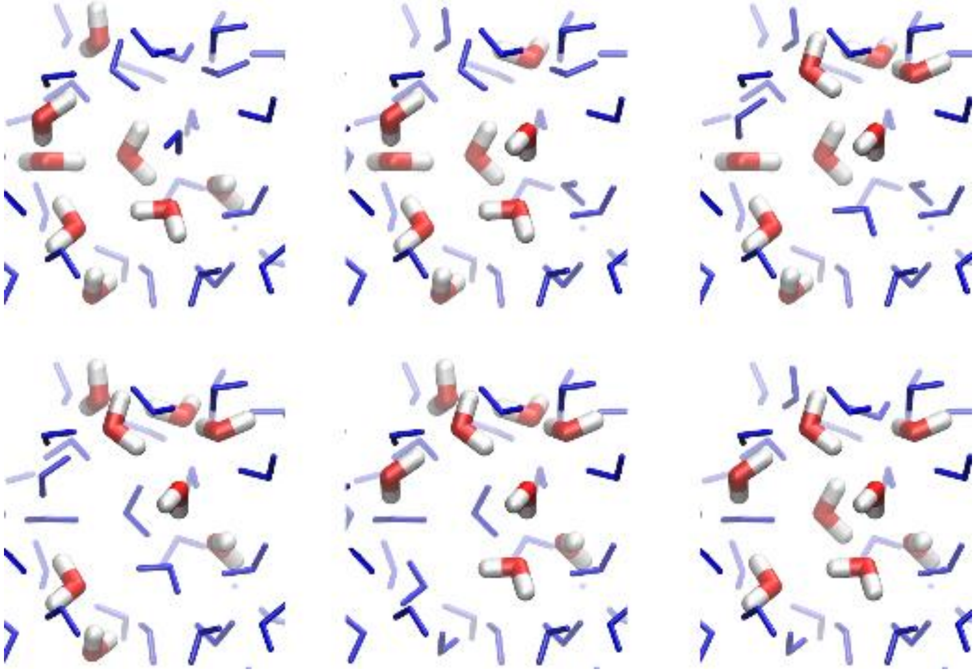
- Size-consistent multi-partitioning adaptive QM/MM
  – removes discontinuities, avoids artificial forces
  author: Hiroshi C. Watanabe (JCTC 10, 4242 (2014))

Adaptive QM/MM – SCMP

- Forces obtained by averaging from several QM/MM calculations

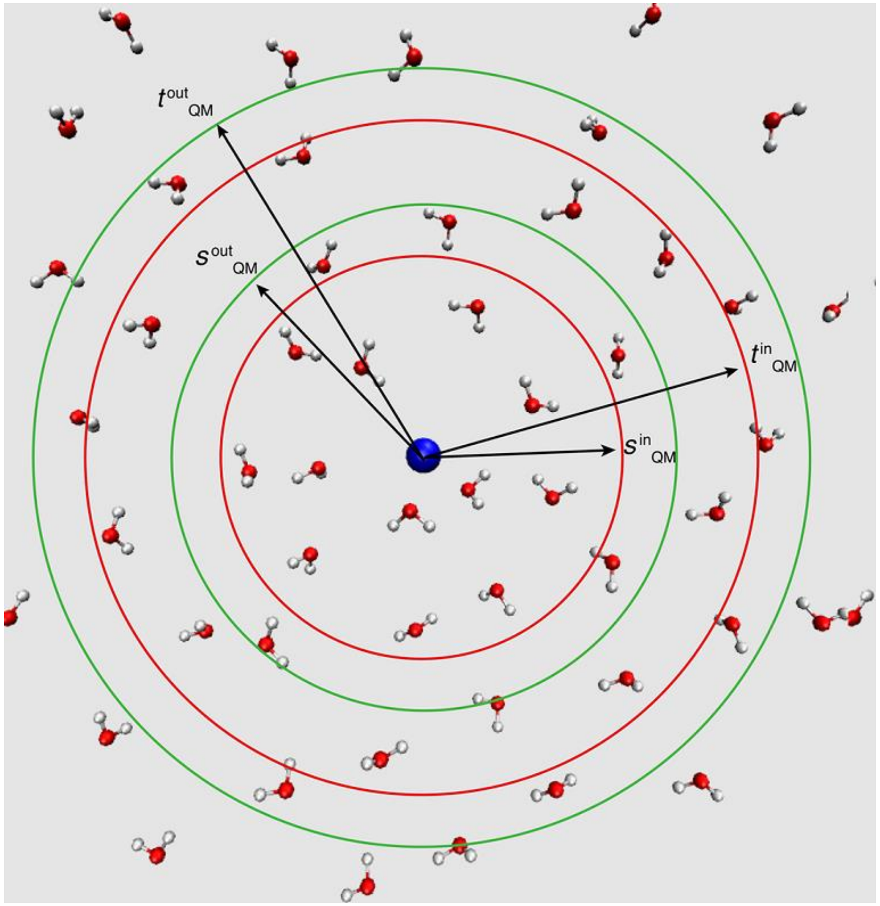
$$\mathbf{F}_j^{\mathrm{eff}}(\mathbf{r}) = \sum_n^N \sigma^{(n)}(\mathbf{r})\cdot\mathbf{f}_j^{(n)}(\mathbf{r})$$

- Partitionings with identical sizes of the QM region are considered



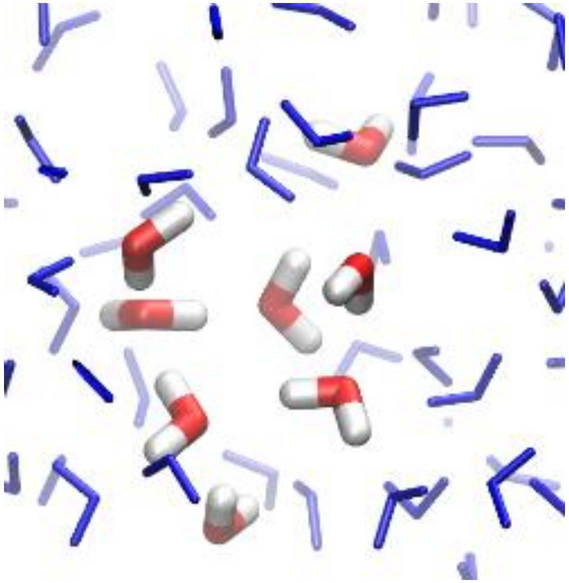
Adaptive QM/MM – SCMP

- Weight function σ
  – depends on distances of QM and MM molecules from the center



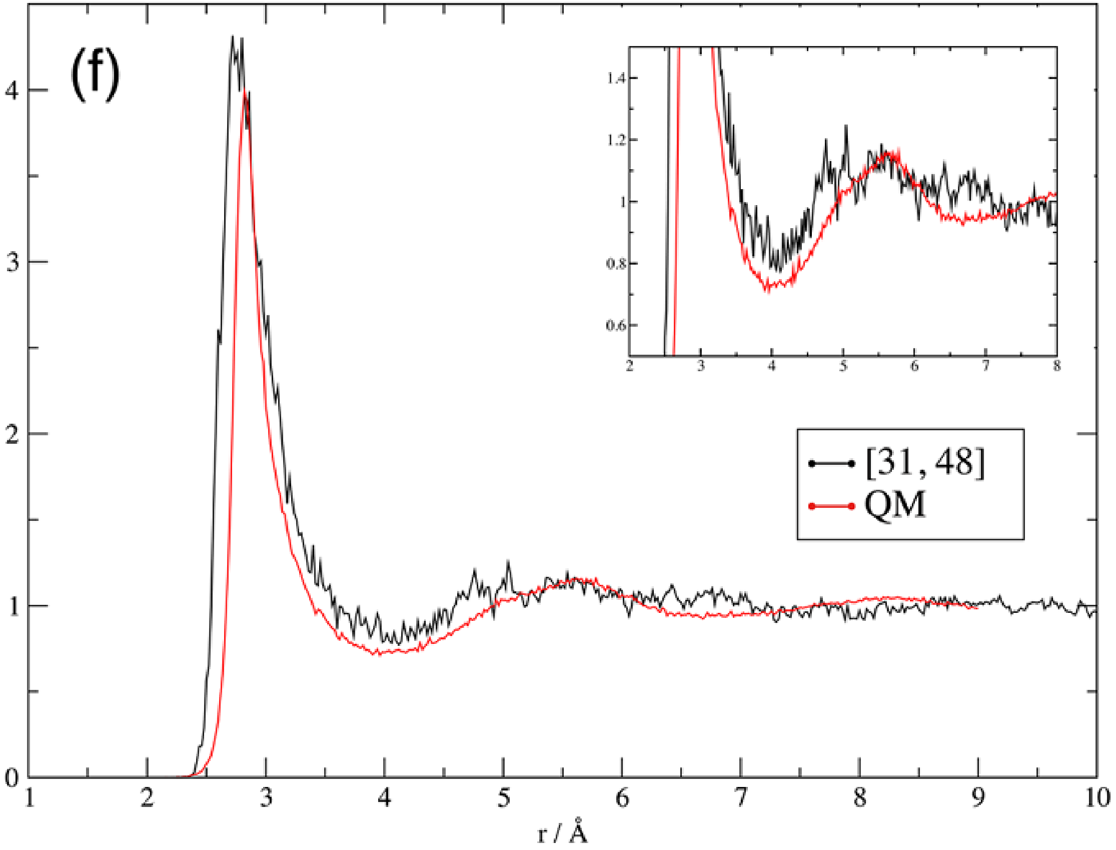
Adaptive QM/MM – SCMP

- Weight function σ
  – vanishes for a partitioning with fragmented QM zone
  – also vanishes for the most compact partitioning,
- Is σ of a partitioning lower than threshold?
  replace by the most compact, so-far not considered partitioning
  – zero weight of both – discontinuities avoided



Adaptive QM/MM – SCMP

- Example: RDF of “QM water in MM water”
  – 31 QM water molecules, 48 partitionings considered



Adaptive QM/MM – SCMP

- Continuous forces (no artificial forces on the QM–MM boundary)
- Continuous switching of properties at the boundary
- Improvement upon many previous multi-partitioning methods
  – applicability for large QM core regions
  (no required number of partitionings)
  – conserved energy independent of the number of partitionings
- Efficient parallel implementation
  – identical sizes of QM regions – similar computational cost

QM/MM in GROMACS

- Interface by G. Groenhof
- Communicates with external QM software
  (Gaussian, Mopac, Orca, Gamess UK)
- Additive scheme, electrostatic embedding

- DFTB3 – ‘semi-empirical density-functional theory’
  – newly implemented and fully integrated within Gromacs
- No file-based communication
- Gromacs PME routines used for electric potential – efficiency
  (PME used also for QM–QM interaction, and not full Ewald)

- Free energies – internal Gromacs functionality
  or available external tools (PLUMED)

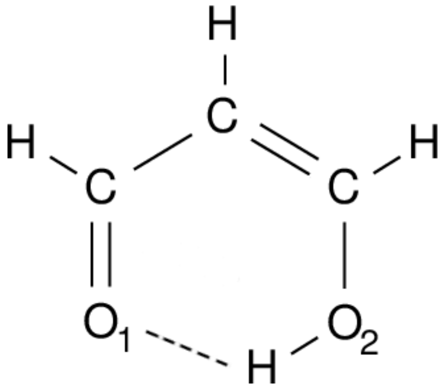
QM/MM in GROMACS

```
; OPTIONS FOR QMMM calculations
QMMM                        = yes
QMMM-grps                   = MAL
QMMMscheme                  = normal
QMcharge                    = 0
QMmult                      =
MMChargeScaleFactor         = 1

QMdftbsccmode               = 3
QMdftb-telec                = 10.
QMdftb-slko-path            = /home/tkubar/DFTB/3ob/
QMdftb-slko-separator       =
QMdftb-slko-lowercase       = yes
QMdftb-slko-suffix          = -c.spl
QMdftb-partial-pme          = 1
QMdftb-dispersion           = 1
```

Application 1 – Proton Transfer

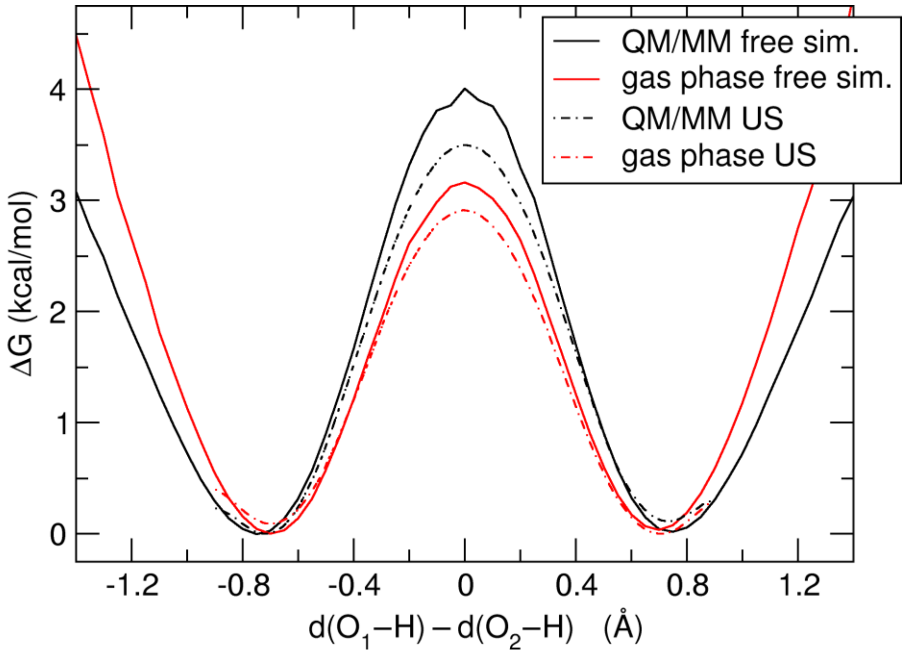
- Malonaldehyde – ultra-fast intramolecular PT



- Free energy barrier of few kcal/mol – rate of several times per ns
- Popular test case for QM/MM

Application 1 – Proton Transfer

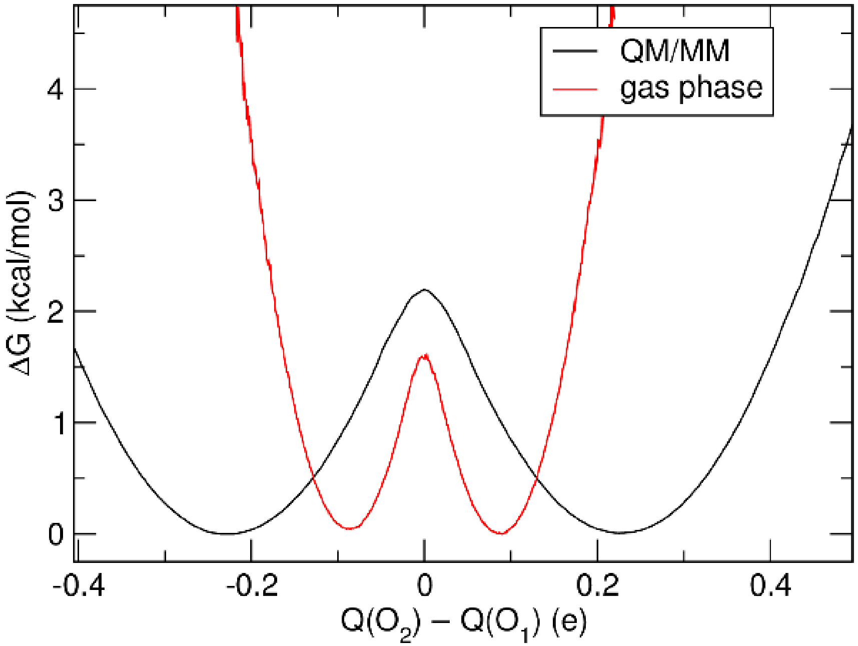
- With QM/MM and in the gas phase
  reaction coordinate: difference of O–H distances



- Barrier from free simulation slightly higher than from umbrella sampling
  (possibly undersampled?)

Application 1 – Proton Transfer

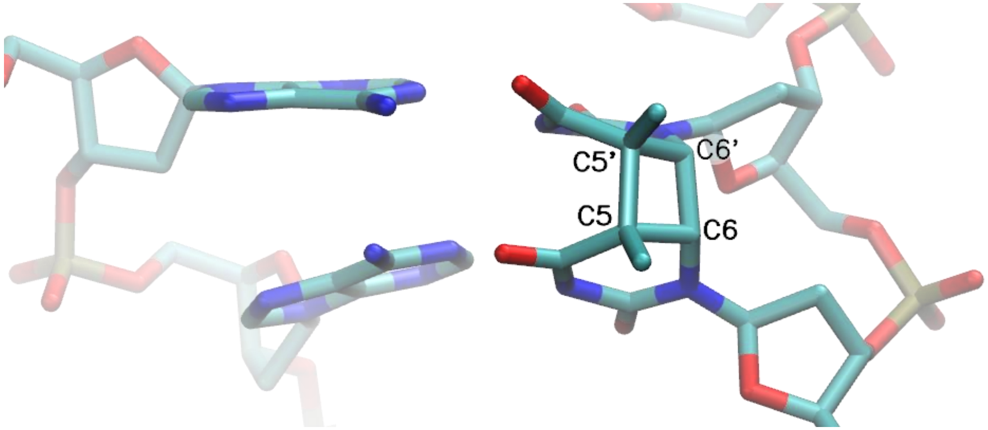
- How to choose the reaction coordinate?
  (generally non-trivial for PT with intermediate molecules)



- Perhaps non-geometrical RC – difference of Mulliken charges
- Disadvantage: biased sampling not easily possible

Application 2 – DNA Lesion

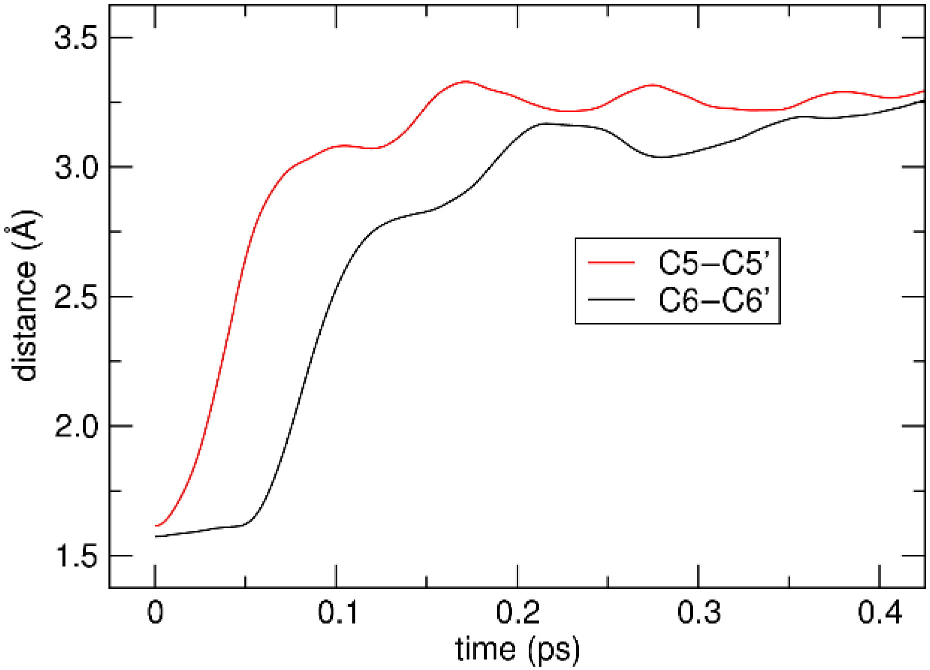
- Cyclobutane pyrimidine dimer



- Stable in the electro-neutral state
- Uptake of excess electron leads to disintegration
- Mechanism debated (experimental and QM/MM studies)

Application 2 – DNA Lesion

- Free simulation of the lesion
  after the electron uptake – radical anion

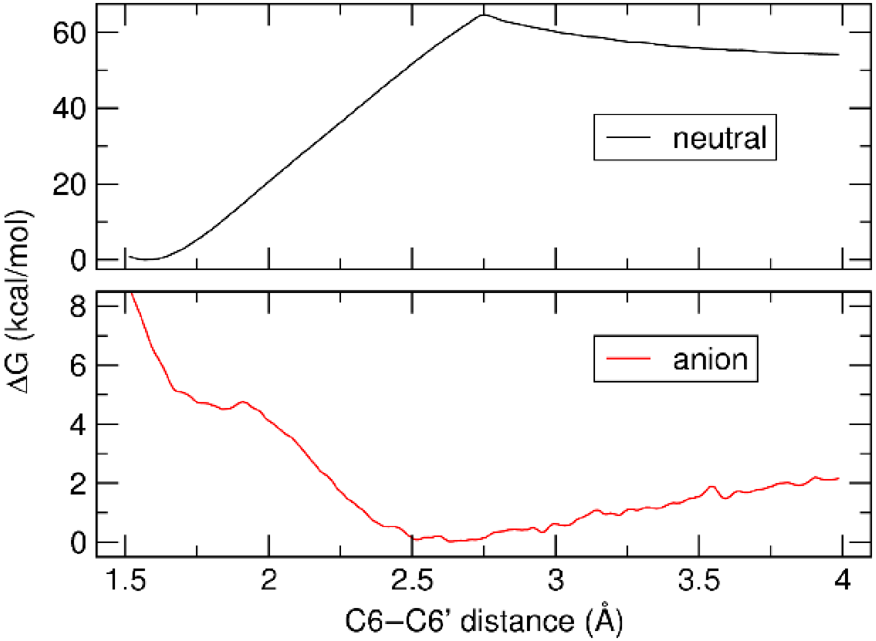


- The C5 – C5' bond breaks first, immediately
- The C6 – C6' bond follows

(averages from 100 simulations)

Application 2 – DNA Lesion

- Umbrella sampling – stretching of the C6–C6' bond
  – CPD electro-neutral or after the electron uptake (radical anion)



- Meta-stable state with C5–C5' broken and C6–C6' intact?
- Tiny barrier for complete disintegration
  (NB: improved sampling and DFTB benchmark necessary)