cecam
Centre Européen de Calcul Atomique et Moléculaire

Hybrid Quantum Mechanics / Molecular Mechanics (QM/MM) Approaches
- QM/MM border: link atoms (LA), frontier orbitals (FO), optimized effective core potentials (OECP), scaled-position link atom method (SPLAM)

Mauro Boero

Institut de Physique et Chimie des Matériaux de Strasbourg
University of Strasbourg - CNRS, F-67034 Strasbourg, France

Computational Materials Science Initiative
計算物質科学イニシアティブ

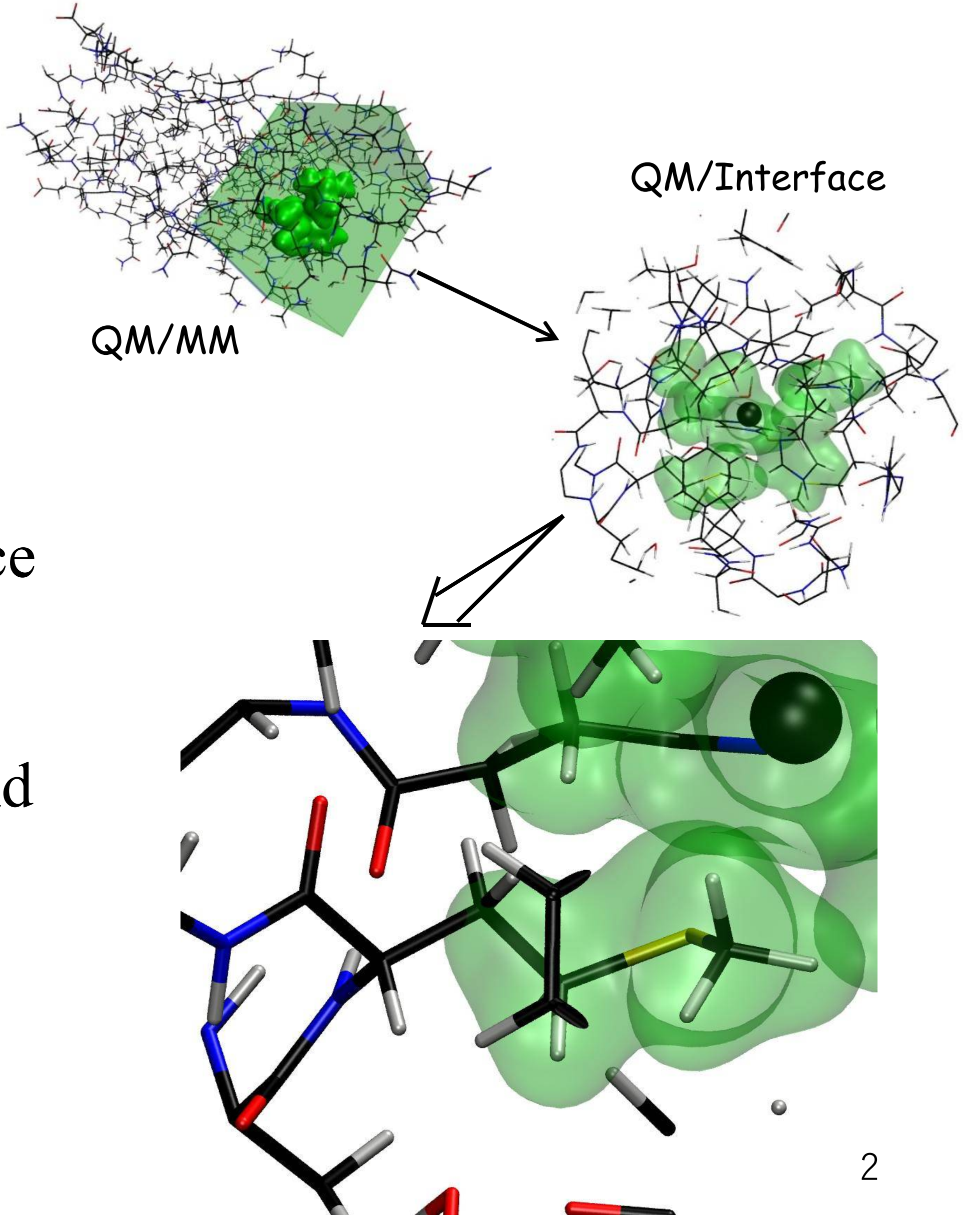
名古屋大学 *@Institute of Materials and Systems for Sustainability,*
Nagoya University - Oshiyama Group, Nagoya Japan

Partitioning the system: Shopping List

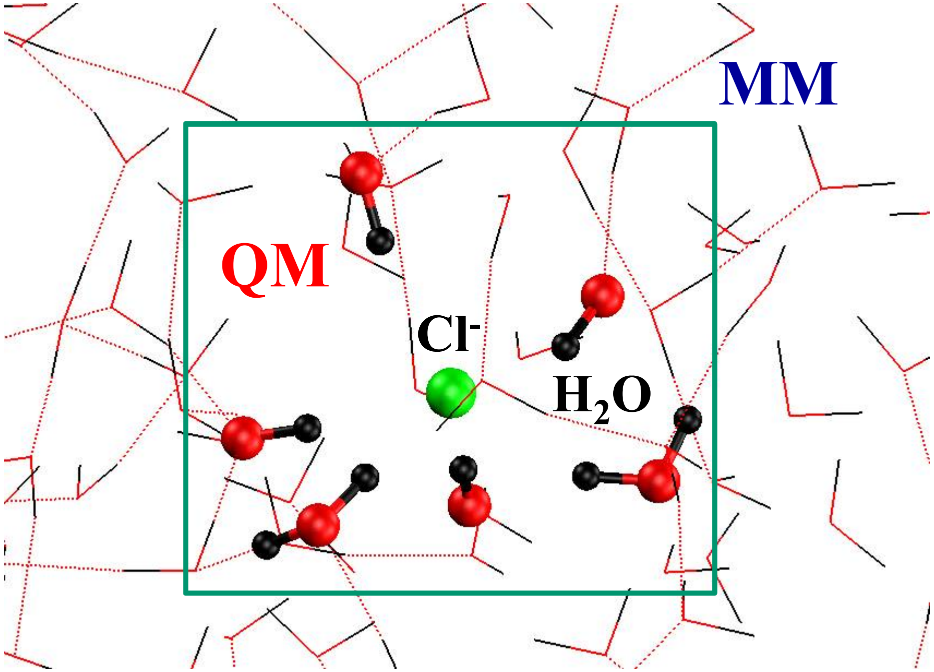
1. chemical active part treated by QM methods

2. large environment that is modeled by a classical force field (MM)

3. Interface between QM and classical parts



- In the easiest (lucky) case, QM atoms interact with the MM atoms via:
  - H-bonds
  - Non-bonding interactions (e.g. Coulomb or van der Waals)



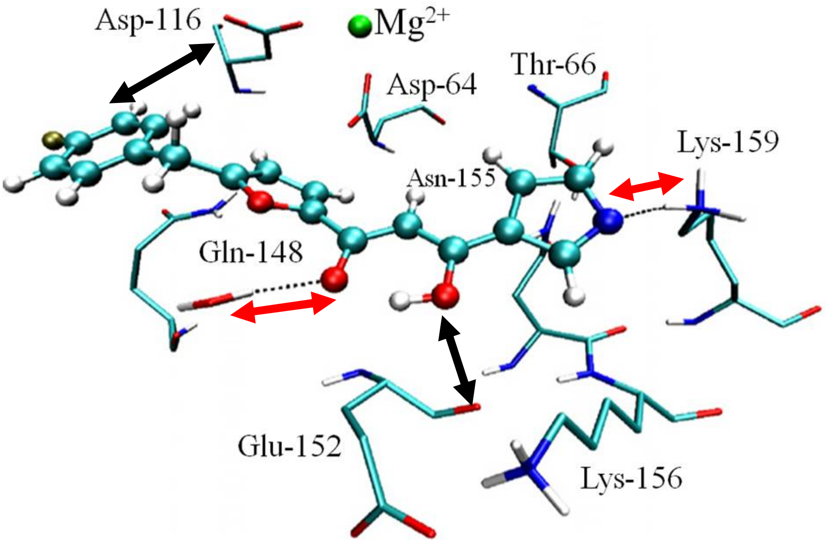
QM and MM atoms are **not** chemically bonded.

- In the case of QM atoms and MM atoms not chemically bonded, selection of the QM/MM frontier does not pose particular difficulties. ***But*** due to the weak interaction QM atoms ***can escape*** from the QM box upon long dynamics (watch out !)

Examples: (*i*) QM solute surrounded by MM water molecules, or ligand-protein interacting via non-bonding forces, e.g. complex HIV-1 integrase and its inhibitor S-13

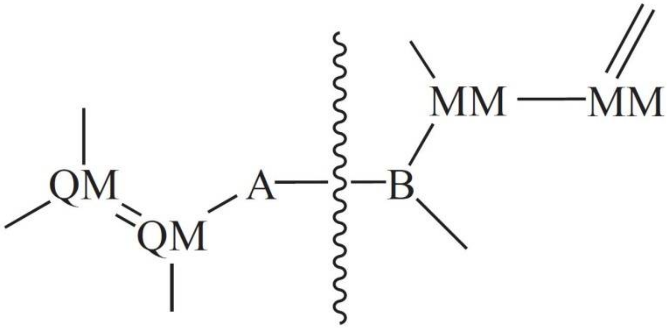
↔ **H-bond**

↔ **Non-bond**



(C. N. Alves et al. *Bioorg. Med. Chem*. **15**, 3818 (2007))

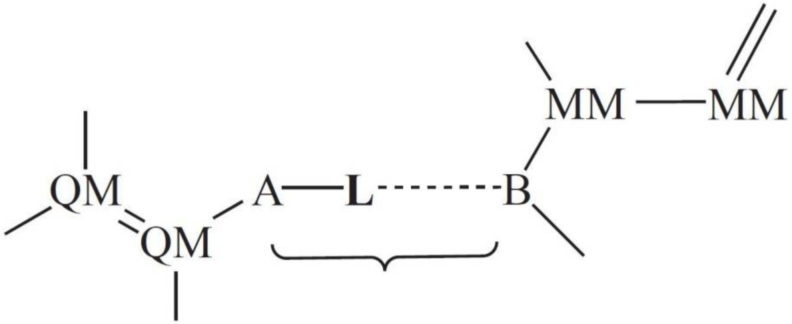
- In most of the cases, the QM/MM frontier passes across a (covalent) chemical



- Suitable termination of the boundary is required in order not to create artificial *dangling bonds*.
- To this aim, the methods proposed in the literature can be classified into three groups:

1. Link atoms,
2. Frontier orbitals
3. Optimized effective pseudopotentials

1. Link atoms (L)



Link (L) atoms are additional monovalent hydrogen-like atoms added to the QM subsystem to saturate the cut covalent bonds.

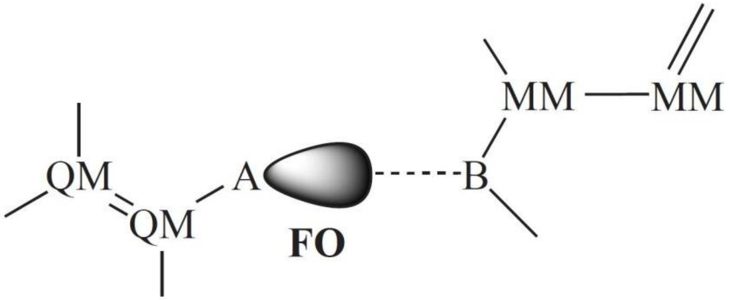
- L atoms are generally invisible to the MM atoms
- L atom should reproduce the local chemical environment (e.g. *sp*^{3}, *sp*^{2}, etc.)
- They are preferentially placed far from each other to avoid spurious interactions

(Singh and Kollman, J. Comp. Chem. **7**, 718 (1986); Field *et al.* J. Comp. Chem. **11**, 700 (1990))

1. Link atoms (L) – continue

- Beside monovalent H-like L-atoms, F or CH_3 (Adjusted Connection Atom) can be used. Anes & Thiel, *J. Phys. Chem. A* **103**, 9290 (1999)
- L-atoms, generally invisible to the MM atoms, interact via the force field directly with the border QM atoms to ensure that the QM-MM covalent bond are not affected by the frontier passing across these chemical bonds.
- There are cases in which L-atoms must be kept into account also from the MM side, e.g for C species in which non-negligible polarization effects occur.
- Polarization of L-atom – C bonds could bias the results if L atoms are neglected in the calculation of the MM interactions. N. Reuter et al. *J. Phys. Chem. A* **104**, 1720 (2000)

2. Frontier Orbitals (FO)



The unsaturated covalent bond of a border QM atom is compensated by an additional localized orbital $\psi_{FO}(\mathbf{x}-\mathbf{R}_A)$ treated as frozen during the calculation.

Note: the freezing of FOs can give problems in variational approaches in which wavefunctions or the charge density are used as dynamical variables.

(Assfel and Rivail, Chem. Phys. Lett. **263**, 100 (1996); Gao *et al.* J. Phys. Chem. A **102**, 4714 (1990))

2. Frontier Orbitals (FO) - continue

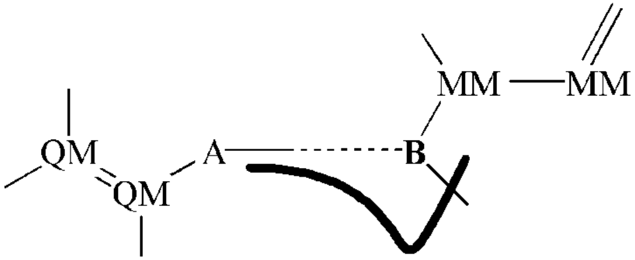
- Frozen FOs work well in self-consistent field optimization
- However, contributions to the forces can result in spurious components that can bias the dynamics

One of the most recent (and remarkable) applications is the study of H transfer by tunneling to the active site catalyzed by coenzyme B_{12}-dependent methylmalonyl-CoA mutase. QM subsystem = 45 atoms, including the ligand and a portion of the methylmalonyl-CH_2- substrate.

FO = at the carbon atoms C_2 of the b-mercaptoethylamine part of the CoA.

Dybala-Defratyka, et al. *Proc. Nat. Acad. Sci. USA* **104**, 10774(2007)

3. Optimized Effective Core Pseudopotentials (OECP)



Border PP written as a sum of a local and a non-Local part

$$V_I^{\mathrm{OECP}}(\mathbf{r},\mathbf{r}') = V^{loc}(\mathbf{r})\,\delta(\mathbf{r}-\mathbf{r}') + \sum_l V_l^{NL}(\mathbf{r},\mathbf{r}')$$

$\mathbf{r} = \mathbf{x} - \mathbf{R}_I$, being $\mathbf{R}_I$ the a capping atom at the QM/MM interface. All the PP parameters are optimized by minimizing iteratively the differences in electron density between the QM subsystem and a full QM reference configuration including atoms beyond the QM/MM boundary

(DiLabio *et al*. J. Chem. Phys. **116**, 9578 (2002), von Lilienfeld *et al*. J. Chem. Phys. **122**, 014133 (2005))

3. Optimized Effective Core Pseudopotentials (OECP) - continue

- Local part:

$$V^{loc}(\mathbf{r}) = -\frac{Z_I}{r}\operatorname{erf}\left(\frac{r}{r_0\sqrt{2}}\right) + e^{-(r/r_o)^2/2}\left[c_1 + c_2\left(\frac{r}{r_0}\right)^2 + c_3\left(\frac{r}{r_0}\right)^4 + c_4\left(\frac{r}{r_0}\right)^6\right]$$

- Non-local part:

$$V_l^{NL}(\mathbf{r},\mathbf{r}') = \sum_{m=-l}^{l} Y_{lm}^*(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}}) \sum_{i,j=1}^{3} p_{lj}(r) h_{lji} p_{li}(r')$$

where $p_{ij}(r)$ = const $r^{l+2(h-1)}$ exp(-0.5 r^2/r_l^2) and Y_{lm} are the spherical harmonics.

- All the parameters $\{r_0, c_1, c_2, c_3, c_4, h_{lji}, r_l\}$ are optimized by minimizing iteratively the differences in electron density between the QM subsystem and a full quantum reference configuration including atoms beyond the QM/MM boundary.

3. Optimized Effective Core Pseudopotentials (OECP) – continue

- We remark that the dimensionality of the parameter space is determined by the maximum angular momentum in the non-local part of the OECP.
- In practical applications (von Lilienfeld et al. 2005) it has been shown that a maximum value l = s or, rarely, l = p is enough to achieve a good optimization for oxygen in water or carbon in acetic acid.
- OECPs are particularly suitable in the cases in which the QM subsystem embedded in the MM environment is characterized by the presence of highly ionic species.
- **Warning**: OECP can affect other nearby bonds !

Scaled Position Link Atom Method (SPLAM)

1. Molecular oscillations could be partly biased by the presence of monovalent L-atoms.
2. L-atoms, chemically bonded to QM atoms, are subject to dynamical fluctuations during the simulation.
3. In general, they do not reproduce the correct bond length of the MM atom that they replace.

Proposed solution: SPLAM, Echinger et al. *J. Chem. Phys*. **110**, 10452 (1999),

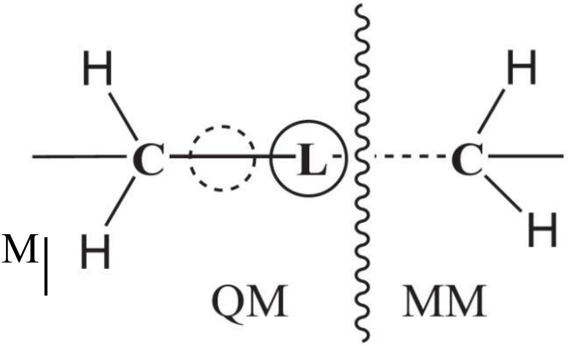
- Focusing specifically on C-C bonds:

  Non-polar carbon single bonds joining CH_2 groups are ubiquitous and their cut represent one of the best choices to terminate a QM region.

Scaled Position Link Atom Method (SPLAM)

- continue

The position of the monovalent saturating H-like L atoms is rescaled from the artificial C-H bond length to that of the original C-C bond distance.



If :

- C-C equilibrium distance = r_{CC}^{0}
- actual bond length $r_{CC} = |\mathbf{r}_C^{QM} - \mathbf{r}_C^{MM}|$
- H-like L-atom $r_{CH} = |\mathbf{r}_C^{QM} - \mathbf{r}_L|$

then the scaled position becomes $r_{CL} = r_{CH} + \dfrac{k_{CC}}{k_{CH}}\left(r_{CC} - r_{CC}^{0}\right)$

where k_{CC} and k_{CH} are deduced from the corresponding bond stretching.

Scaled Position Link Atom Method (SPLAM)

- continue

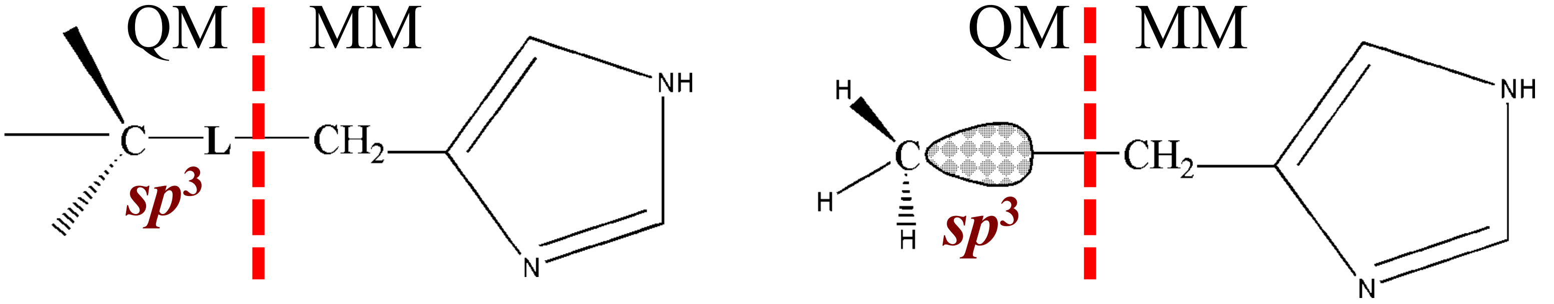
However:

1. SPLAMs are somehow an artificial way of elongating a C-H chemical bond
2. Has the drawback of introducing spurious force components that in some cases can affect the dynamics of the system and lead to inconsistent results.
3. In any case, an energy correction is required, and this is written as written as a harmonic term

$$\Delta E_{\mathrm{stretch}}(\mathbf{r}_{\mathrm{C}}^{\mathrm{QM}}, \mathbf{r}_{\mathrm{C}}^{\mathrm{MM}}) = k_{\mathrm{CC}}\left(1 - \frac{k_{\mathrm{CC}}}{k_{\mathrm{CH}}}\right) \cdot (r_{\mathrm{CC}} - r_{\mathrm{CC}}^{0})^{2}$$

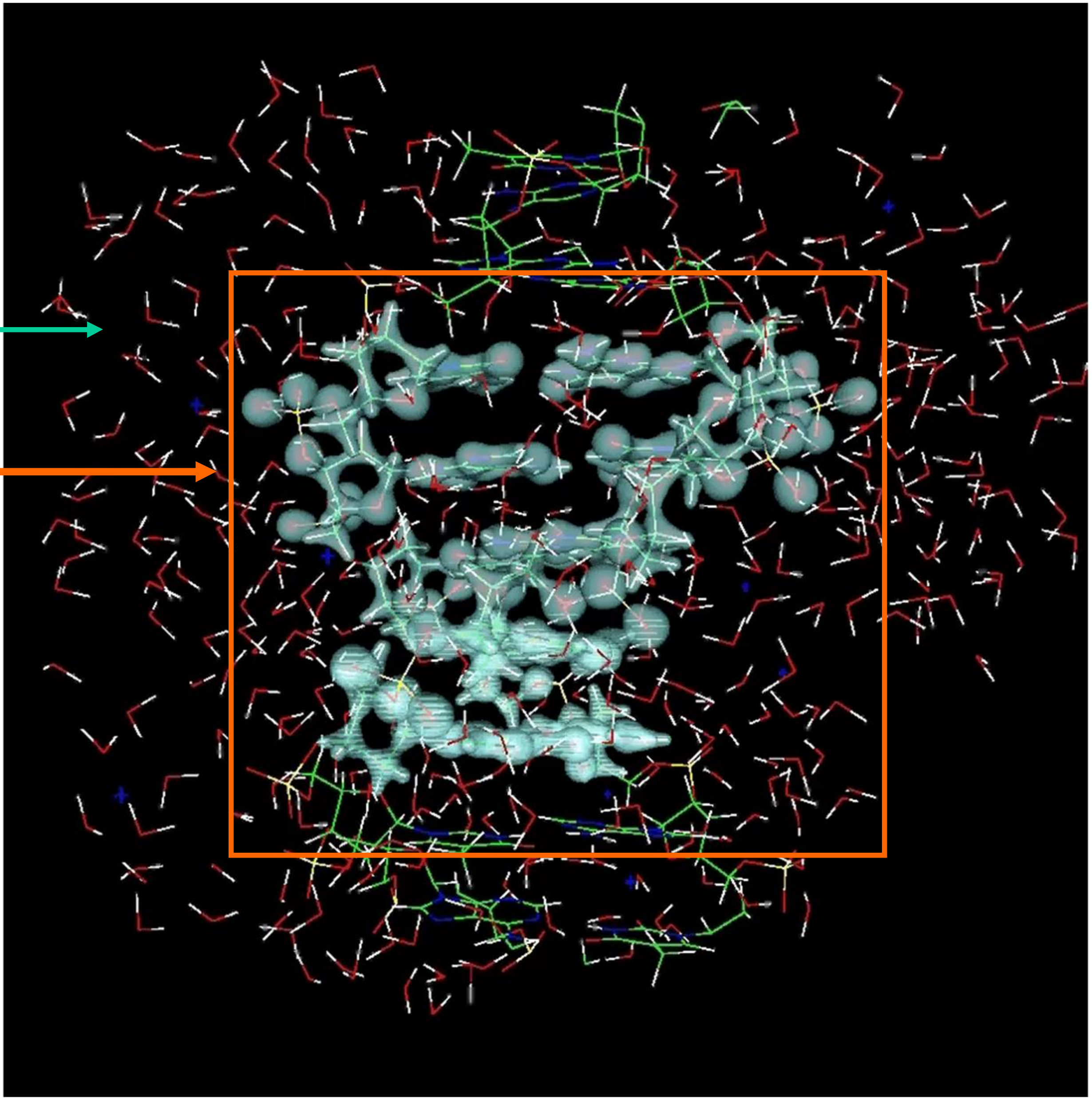
Where are we supposed to put a Link Atom (or a Frozen Orbital) ?

- Try to place the L-Atom or FO at an aliphatic C (CH_4-like)



- This has in general the smallest possible charge distribution at the frontier (MacKerell @ www.psc.edu/general/software/packages/charmm/tutorial mackerell/QMMM_00.pdf)

Example of L-atoms in CPMD: DNA



MM subsystem
(250000 atoms)

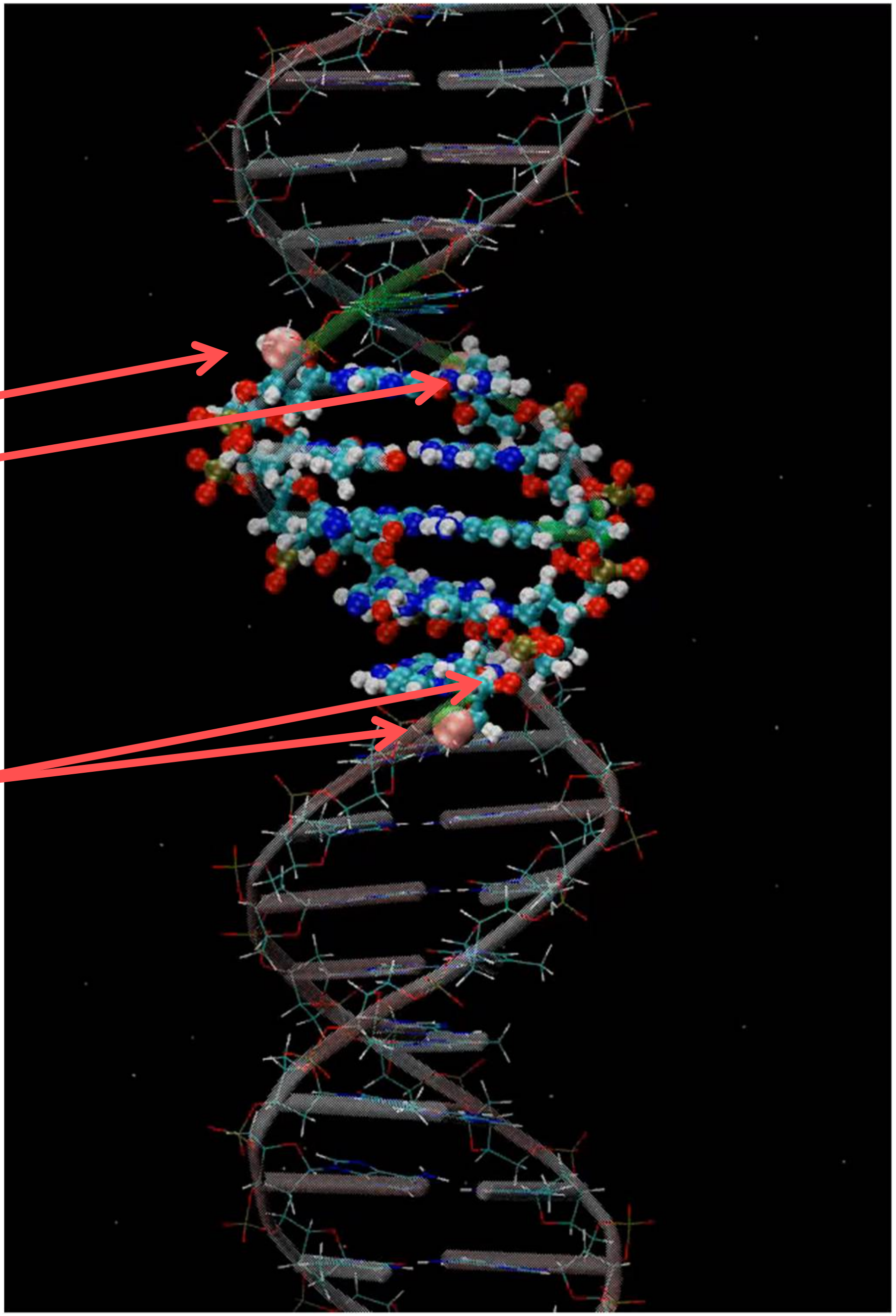
QM subsystem
(500 atoms)

Earth Simulator –
8 nodes x 8 CPU
CPU time for 1 iteration:

QM	**14.81 s**
QM/MM_Int	**48.13 s**
MM	**5.76 s**

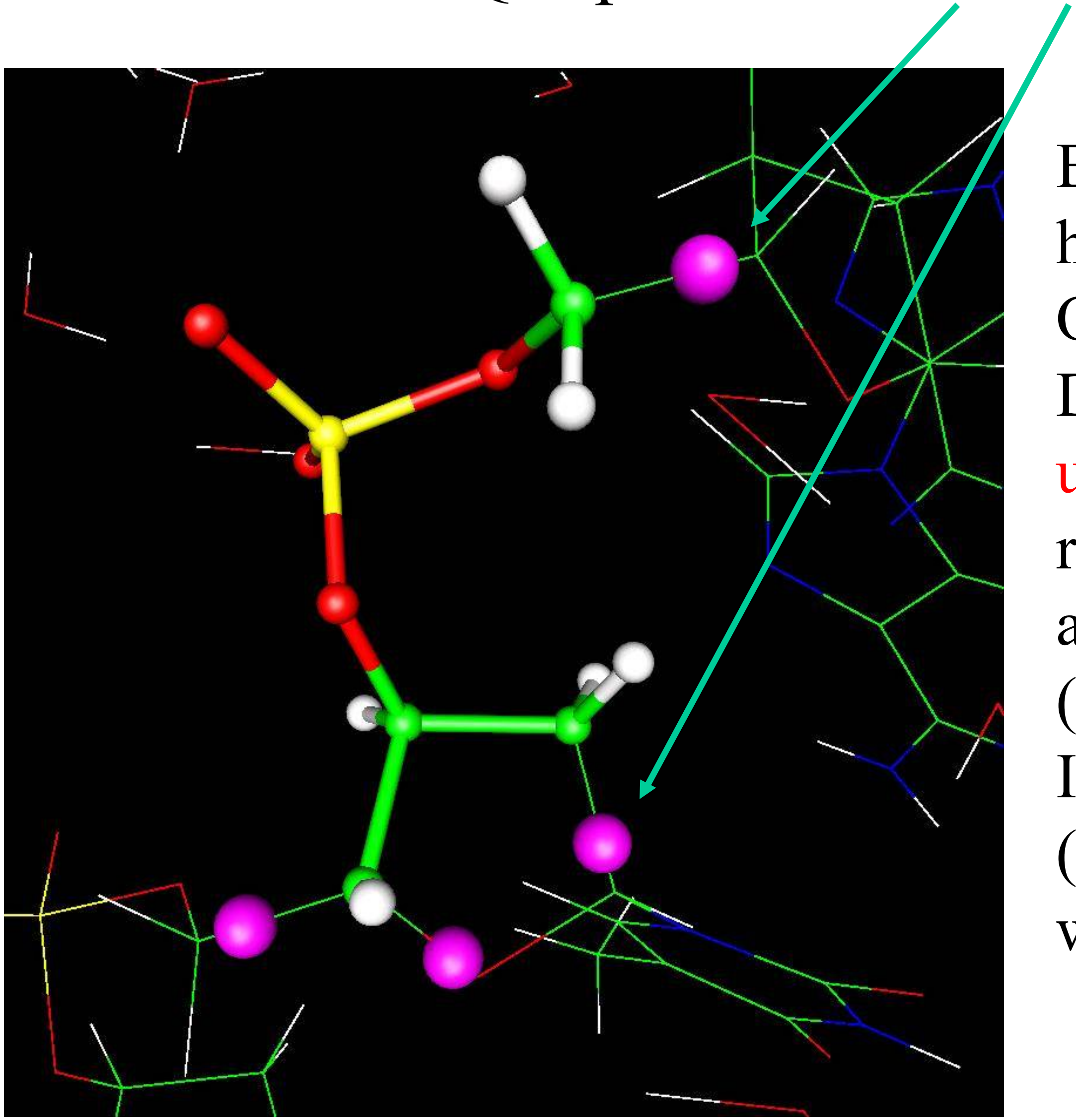
Angew. Chem. Int. Ed. **45**, 5606 (2006)

Example of L-atoms in CPMD: DNA



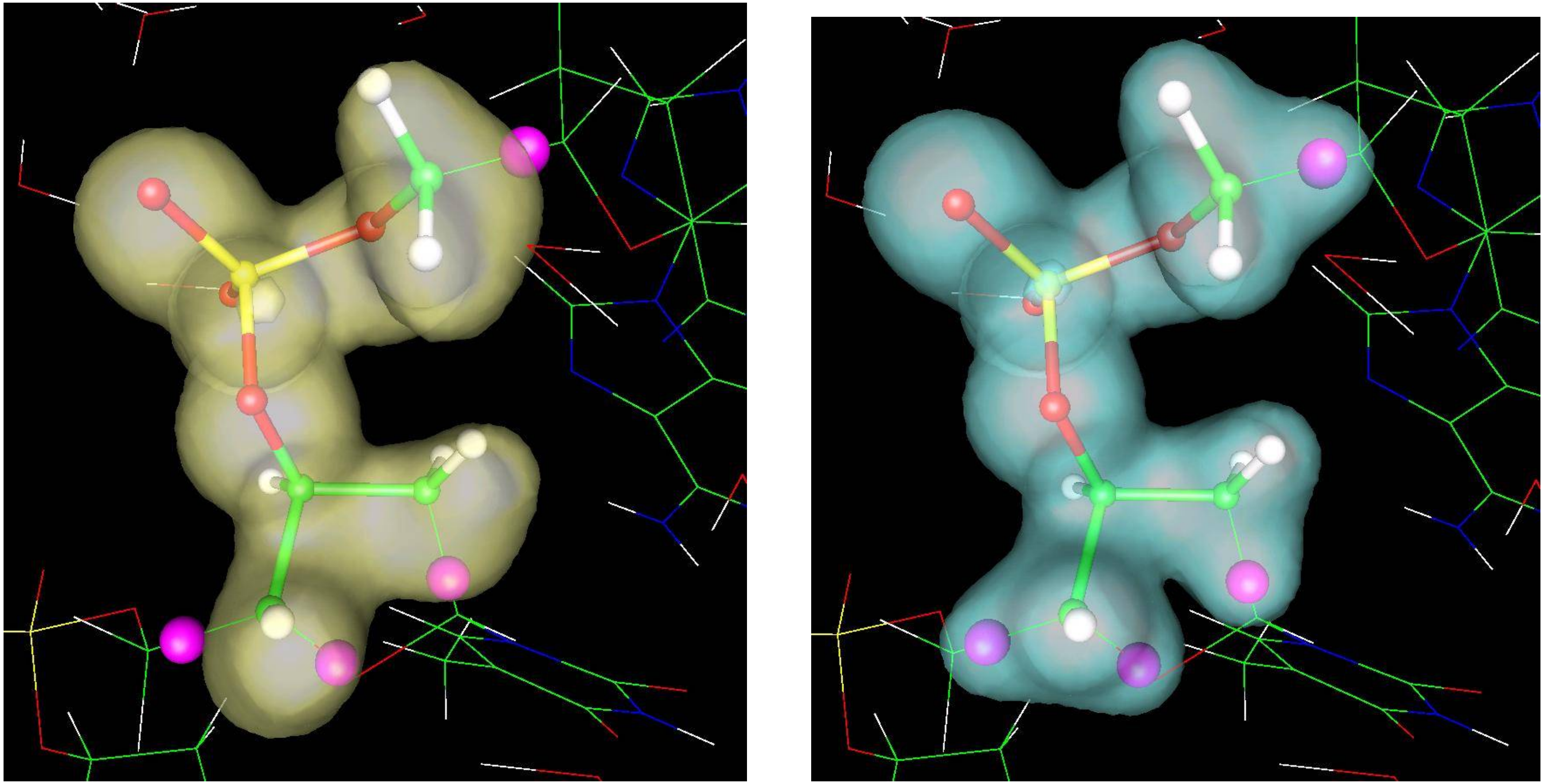
4 pink atoms along the phosphate backbone

Warning: saturate dangling bonds (DB) between MM and QM parts with link atoms (H-like)



Example: small QM of hydrated DNA.
Omitting the capping of DB can originate a large unbalanced charge, redistributed in an arbitrary way on the (MM) atoms around.
If this charge is large (roughly > $5x10^{-2}$) we can be in trouble !

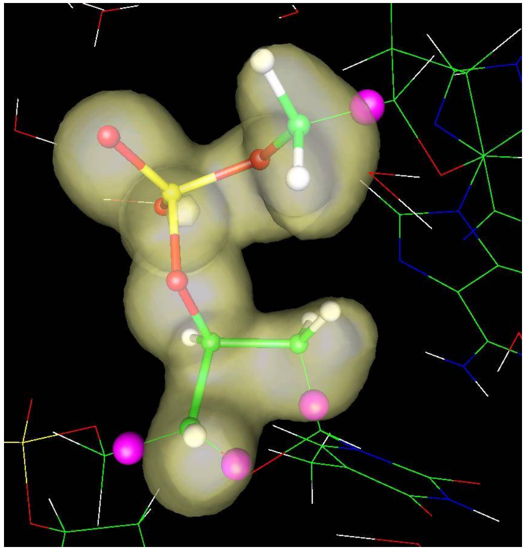
QM/MM Dangling bonds: influence on the local electronic structure (isosurface at $4 \times 10^{-2}\ e/\text{A}^3$)



No inclusion of link atoms

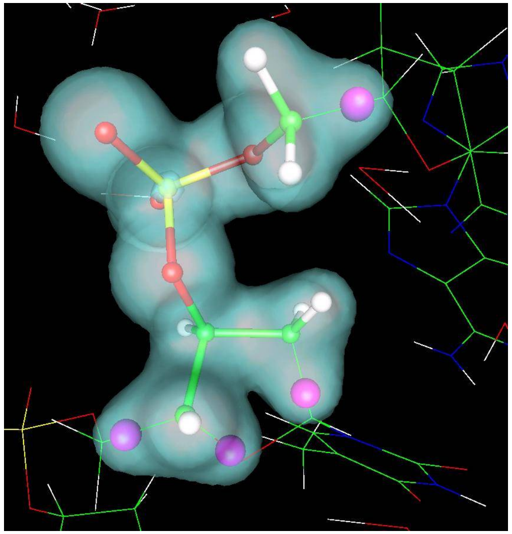
Inclusion of link atoms

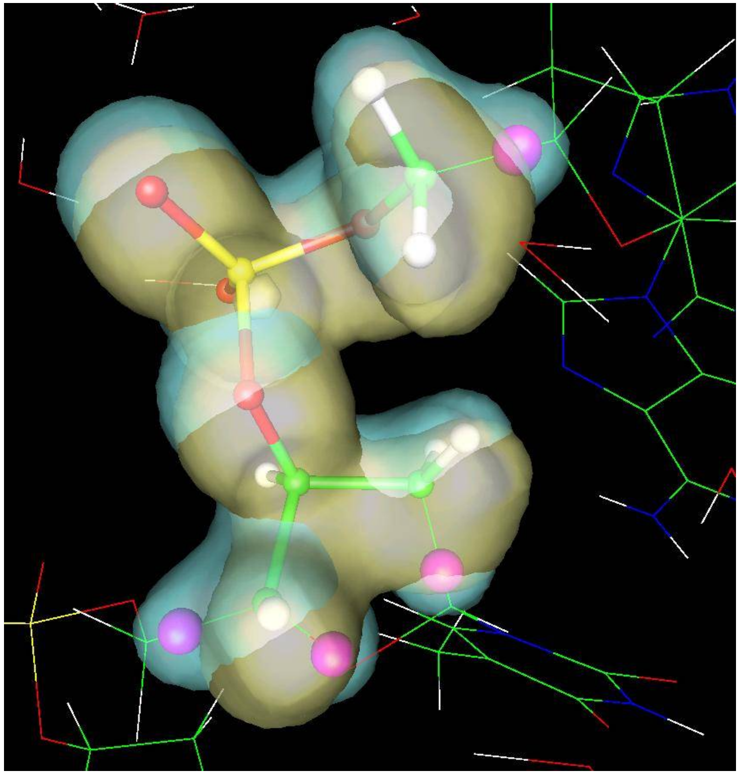
QM/MM Dangling bonds: influence on the local electronic structure (isosurface at 4×10^{-2} e/A^3)



No H-capping

H-Capping





General Warning about Link Atoms / Capping Atoms (and not only)

- L-atoms must not be too close to each other to avoid spurious ***link atom-link atom*** interactions. Remember that they carry a wavefunction $\psi_{link}(\mathbf{x})$ that in a DFT-like scheme enters in the total electron density $\rho\,(\mathbf{x})$ as

$$\rho(\mathbf{x}) = \sum_{i=1}^{QM} f_i \left|\psi_i(\mathbf{x})\right|^2 + \sum_{link=1}^{LINK} f_{link} \left|\psi_{link}(\mathbf{x})\right|^2$$

with all the related consequences on the Kohn-Sham Hamiltonian and potential. For instance the Coulomb interaction

$$\frac{\left|\psi_{link}(\mathbf{x})\right|^2 \left|\psi_{link'}(\mathbf{y})\right|^2}{\left|\mathbf{x} - \mathbf{y}\right|} \underset{\left|\mathbf{x}-\mathbf{y}\right|\to\infty}{\longrightarrow} 0$$

How does CPMD look like once unpacked :

- Versions 3.* no longer developed/supported since November 2013
- **New release (ongoing) 4.3 available since 2019**
- Fortran 90/95 (c/c++ @ sysdepend.c) / Fortran 2003/2008/2018 + CUDA
- Organized in modules & structure of the code slightly changed (more rational):

```
-|---  CPMD ---|---- src
       |       |---- configure
       |       |---- doc
       |       |---- scripts
       |       |---- modules -|-MM_Interface
       |                      |-Gromos/Amber
       |---- addons--|cpmd2cube
       |             | plumed_for_CPMD
       |---- regtests
```

- Code compilable via a (linked) configure.sh script in the CPMD directory.

A few remarks before practice: How does CPMD look like once unpacked

- ./CPMD/src …the main directory is the QM code
- ./CPMD/modules/MM_Interface …in this sub-directory you can find all the routines of the QM/MM interface needed to run the code in hybrid mode.

  Routines are named mm_*.mod.F90 in ./CPMD/src

  mm_*.F in ./CPMD/modules/MM_Interface

  Note: mm_*. mod.F90 routines in the main ./CPMD/src directory are *partly* compiled (skipping all QM/MM options) also in the full QM code
- ./CPMD/modules/Gromos …in this sub-directory we have all the classical force field(s) routines, i.e an AMBER and a GROMOS force field (rewritten by MB).

Files in input:

- The usual files of CPMD, i.e. the standard input and the pseudopotentials
- The COORDINATES, TOPOLOGY, and INPUT files of the classical force field

Files in output:

- The standard output + ENERGIES, TRAJECTORY, etc.
- The interacting part (TRAJECTORY_INTERACTING), the total electrostatic potential (EL_ENERGY, ESP), the MULTIPOLE moments, charges (CHJ), etc…

Structure of the CPMD input file

In the section &CPMD … &END we simply add the new keyword **QMMM**, e.g.

```
&CPMD
 MOLECULAR DYNAMICS
 RESTART WAVEFUNCTIONS COORDINATES VELOCITIES LATEST
 RESTART ACCUMULATORS nosep nosee cell LATEST
 QMMM
 ODIIS
 4
 …
&END
```

The sections &DFT … &END, &PROPERTIES … &END, etc.. do not change.

Structure of the CPMD input file

A new section &QMMM … &END must be included

```
&QMMM
 TOPOLOGY
 topology.top
 COORDINATES
 coordinates.crd
 INPUT
 input.inp
 AMBER                 -> specify here the force field AMBER/GROMOS
 RCUT_NN               -> set the cut-off radii for the 3 regions
 10.
 RCUT_MIX
 15.
 RCUT_ESP
 25.
 CAP_HYDROGEN         -> add monovalent H-like atoms on cut bonds
 ARRAYSIZES           -> set some array size for dynamical allocation of
 MAXATT 28               the classical force field
 …
&END
```

Structure of the CPMD input file

The section &ATOMS … &END looks a bit different

```
&ATOMS
CONSTRAINTS
FIX MM                     -> Fix the whole classical part if neded
END CONSTRAINTS
*O_MT_HCTH.psp  KLEINMAN-BYLANDER  RAGGIO=0.9
 LMAX=P  LOC=P
 40
 421  435  441  443  444  445 -> give the atom number as listed
 …                               in the coordinates.crd file
*C_MT_HCTH.psp  KLEINMAN-BYLANDER  RAGGIO=1.0
 LMAX=P  LOC=P
 75
 419  422  425  427  429  434  436  438  446  449
 …
&END
```

and all the rest is what you know from ordinary full QM CPMD…