united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

ICTP 40th Anniversary

SMR 1564 - 37

SPRING COLLEGE ON SCIENCE AT THE NANOSCALE
(24 May - 11 June 2004)

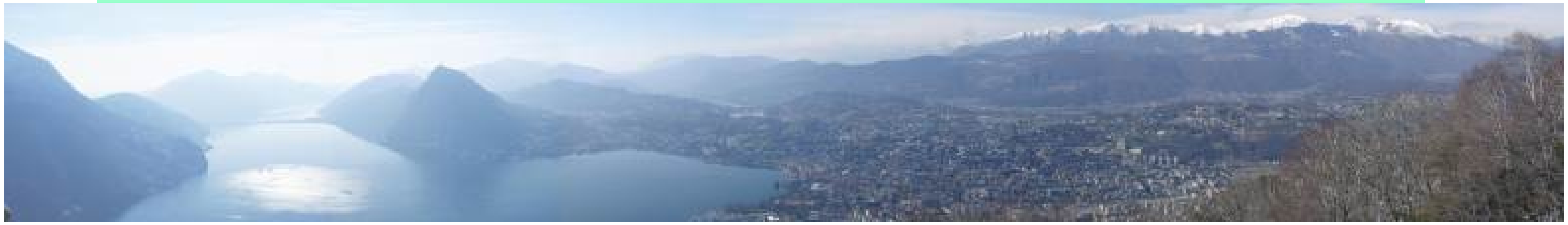
BIOMOLECULES; SIMULATION

Michele PARRINELLO
CSCS, ETH, Zürich, Switzerland

Pushing back the frontiers of computer simulations

Michele Parrinello

Department of Chemistry and Applied Biosciences
ETH
USI Campus, Lugano, Switzerland



Molecular dynamics

Given a potential energy surface:

$$\boldsymbol{U}\left(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_N\right)$$

The dynamics can be determined from Newton’s equation:

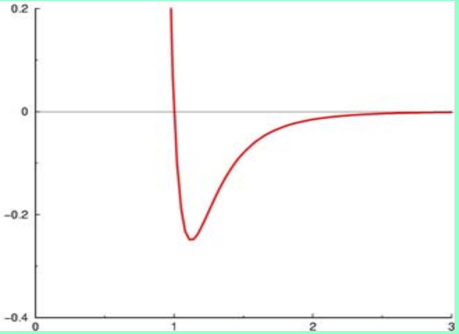
$$\boldsymbol{M}_I \ddot{\boldsymbol{R}}_I = -\nabla \boldsymbol{U}\left(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_N\right)$$

Empirical potentials

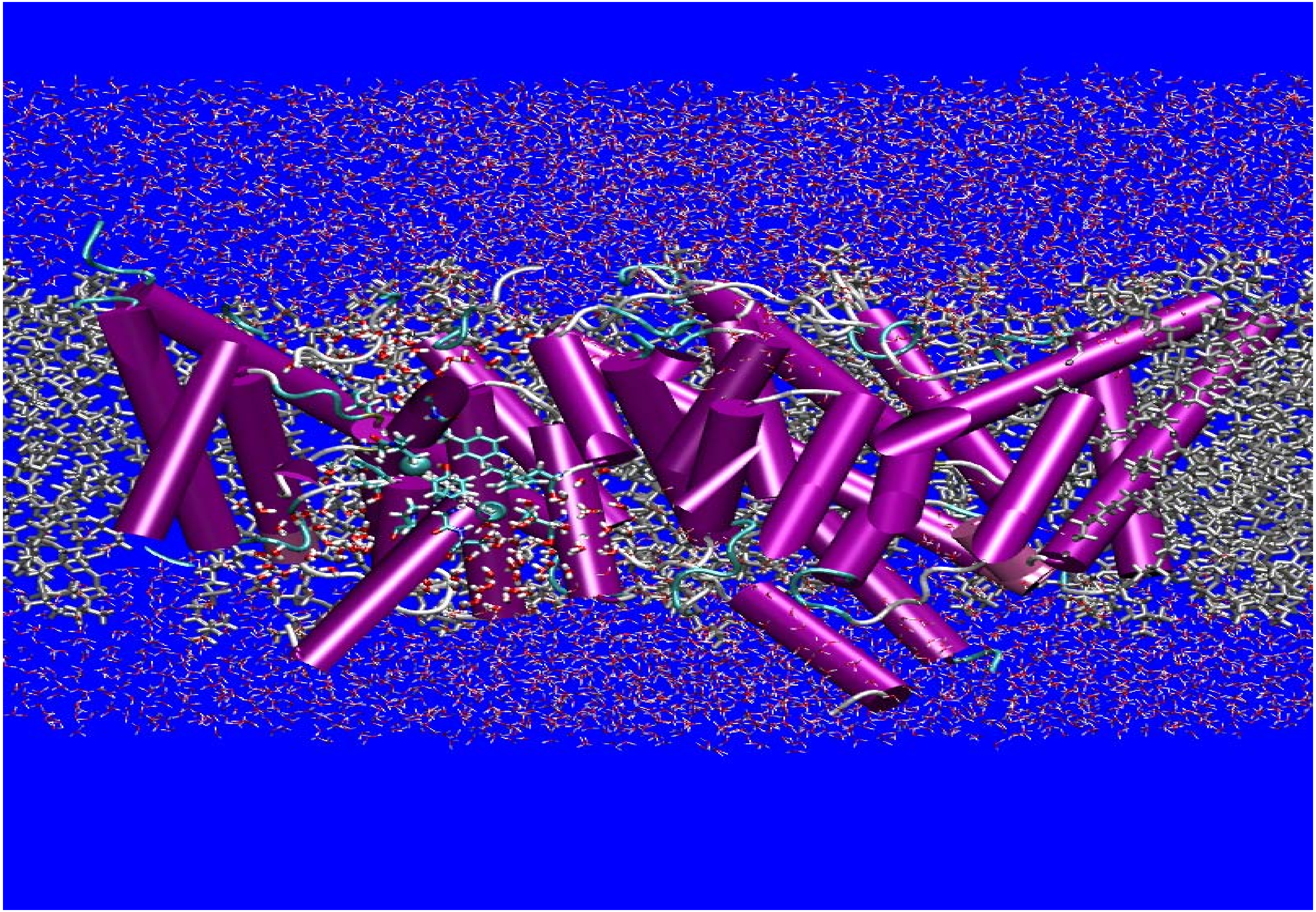
- Molecular mechanics:
  Intramolecular forces: bond stretch, bending, torsion

- Electrostatic interactions:
  Partial charges, dipoles, polarization

- Van der Waals interactions:
  Lennard-Jones, Buckingham potential



- Embedded-atom methods:
  Finnis-Sinclair, Glue model, Daw-Baskes



Pros and cons

PROS

- Efficient
- Accurate in specific cases

CONS

- Not transferable
- No chemistry

Dealing with the electrons

$$H\psi\left(r_1,\ldots,r_M;R_1,\ldots,R_N\right)=E\psi\left(r_1,\ldots,r_M;R_1,\ldots,R_N\right)$$

Hartree-Fock

$$\Psi(x_1,\dots,x_N)=\frac{1}{\sqrt{N!}}\begin{vmatrix}\varphi_1(x_1) & \cdots & \varphi_1(x_N)\\ \vdots & \ddots & \vdots\\ \varphi_N(x_1) & \cdots & \varphi_N(x_N)\end{vmatrix}$$

$$\left[-\frac{1}{2}\nabla+V_n(x)+\sum_{j\neq i}\int\frac{1}{|x-x'|}\left|\varphi_j(x')\right|^2dx'-J\right]\varphi_i(x)=\varepsilon_i\varphi_i(x)$$

Exchange operator:

$$J\varphi_i(x)=\frac{1}{2}\sum_j\int\varphi_j(x)\frac{1}{|x-x'|}\varphi_j^*(x')\varphi_i(x')dx'$$

Beyond Hartree-Fock

Adding correlations

🟥 Perturbation theory MP2, MP4, …	N^4, N^5
🟥 Configuration interaction	exp(N)
🟥 Coupled clusters	N^6, N^7

Small is beautiful!

Hohenberg-Kohn

The energy of the ground state of a many-body system is a unique functional of the electron density:

$$\boldsymbol{E} = \boldsymbol{E}\left[\rho_{\boldsymbol{e}}(\boldsymbol{r})\right]$$

The functional is minimum for the ground state density:

$$\boldsymbol{E} = \boldsymbol{E}\left[\rho_{\boldsymbol{e}}(\boldsymbol{r})\right] - \mu \boldsymbol{N}$$

$$\frac{\delta \boldsymbol{E}\left[\rho_{\boldsymbol{e}}(\boldsymbol{r})\right]}{\delta \rho_{\boldsymbol{e}}(\boldsymbol{r})} = \mu$$

Kohn-Sham

$$\rho_e(\mathbf{r}) = 2\sum_n \psi_n^*(\mathbf{r})\psi_n(\mathbf{r})$$

$$E = -\frac{1}{2}\sum_n \int d\mathbf{r}\,\psi_n^*(\mathbf{r})\nabla\,\psi_n(\mathbf{r}) + \int d\mathbf{r}\,\rho_e(\mathbf{r})V_{ext}(\mathbf{r})$$
$$+\frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\rho_e(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}\rho_e(\mathbf{r}') + E_{xc}[\rho_e(\mathbf{r})]$$

$$\boxed{\left(-\frac{1}{2}\nabla + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r})\right)\psi_n(\mathbf{r}) = \varepsilon_n\psi_n(\mathbf{r})}$$

$$V_H(\mathbf{r}) = \int d\mathbf{r}'\frac{1}{|\mathbf{r}-\mathbf{r}'|}\rho_e(\mathbf{r}') \qquad V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho_e(\mathbf{r})]}{\delta\rho_e(\mathbf{r})}$$

Born-Oppenheimer

The potential energy surface is defined by the instantaneous ground state electronic energy:

$$\phi(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_N) = \boldsymbol{E}_0(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_N)$$

But $$\boldsymbol{E}_0(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_N)$$

needs to be approximated. We shall choose a theory which has the right balance between accuracy and computational efficiency.

Ab-initio MD

$$L = \frac{1}{2}\mu \sum_{n} \int d\mathbf{r} \left|\dot{\psi}_{n}(\mathbf{r})\right|^{2} + \sum_{I} \frac{1}{2} M_{I} \dot{\mathbf{R}}_{I}^{2} - E_{KS}\left[\psi_{n}, \mathbf{R}_{I}\right]$$

$$+ \sum_{n,m} \Lambda_{n,m} \left(\langle \psi_{n} | \psi_{m} \rangle - \delta_{n,m} \right)$$

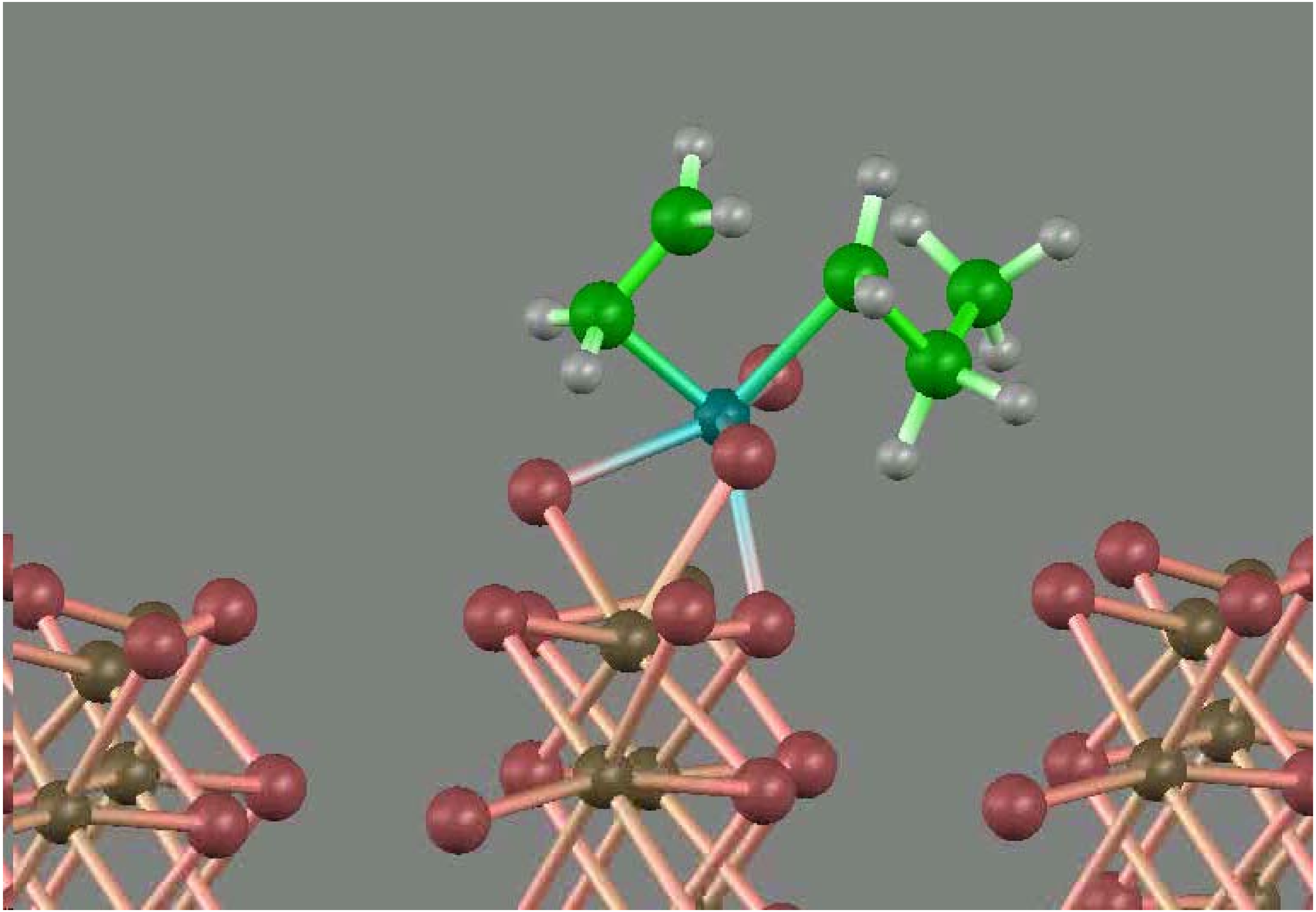
$$\omega_{e} \propto \sqrt{\frac{E_{g}}{\mu}} << \omega_{I}$$

Car-Parrinello molecular dynamics

$$\mu\ddot{\psi}_i = -\frac{\delta E^{DFT}}{\delta\psi_i^*} + \sum_j \Lambda_{ij}\psi_j$$

$$M_I\ddot{\mathbf{R}}_I = -\nabla_{\mathbf{R}_I} E^{DFT}$$

$$\mu_q\ddot{\alpha}_q = -\frac{\partial E^{DFT}}{\partial\alpha_q}$$



Some problems

- Size
- Time scale
- Accuracy
- Non Born-Oppenheimer

Going to larger systems

- New Gaussian-based code QUICKSTEP
- Field theoretical approach
- QM/MM
- DFT-based potentials

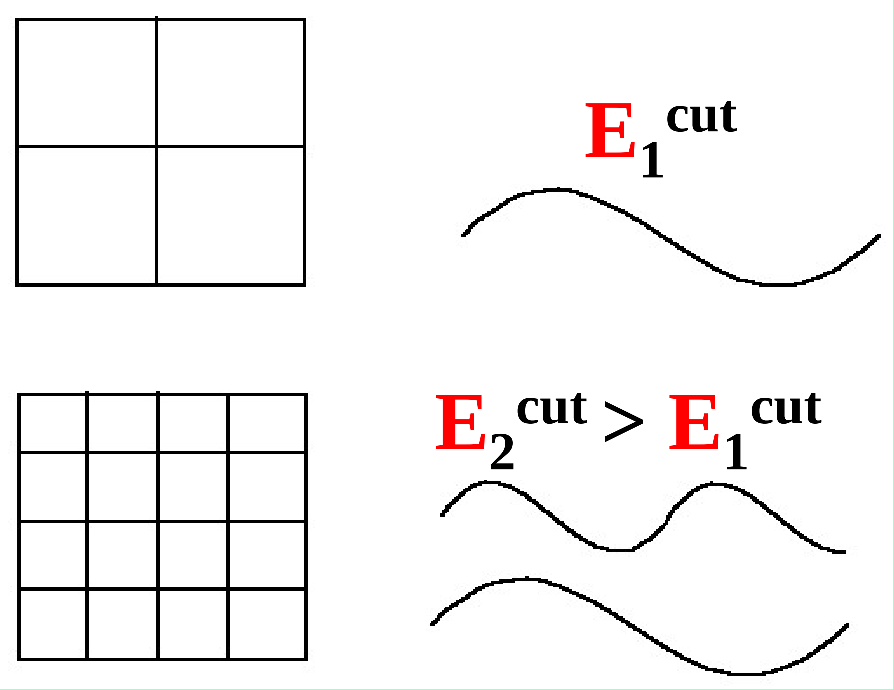
Plane wave basis set

Plane wave expansion: $\psi_i(\mathbf{x}) = \Sigma_{\mathbf{G}}\, c_i(\mathbf{G})\, e^{i\mathbf{G}\mathbf{x}}$

R space $\mapsto$ **G** space

G are the reciprocal space vectors.

The Hilbert space spanned by **PWs** is truncated to a cut-off $\mathbf{G}^2/2 < E^{\mathrm{cut}}$



Which basis set?

Plane Waves	**Gaussians**
• Orthogonal	• Non orthogonal
• No chemical input	• Chemical input
• Convergence easy to check	• Convergence less easy to check
• Simple algebra	• More complex
• Memory intensive	• Reduced memory
• Linear algebra and FFT	• Quantum Chemistry know-how
• No basis set superposition error	• Basis set superposition error
www.cpmd.org	cp2k.berlios.de

Basis set expansion

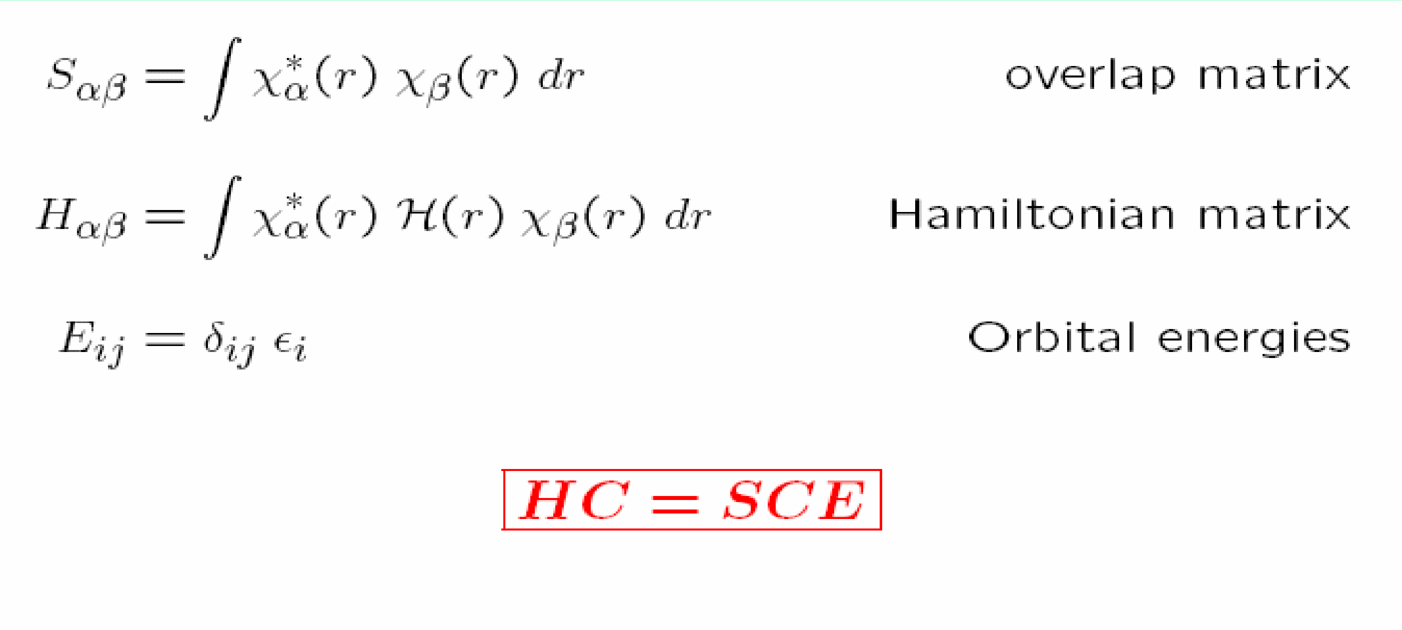
Orbitals Φ_i are expanded in a set of M basis functions $\{\chi_\alpha\}$

$$\Phi_i(r) = \sum_{i=1}^{M} c_{\alpha i}\, \chi_\alpha(r)$$

Basis functions

- Atomic orbital based (Gaussian, Slater, numerical)
- Plane waves
- Grid based, finite elements, wavelets

Basis set expansion



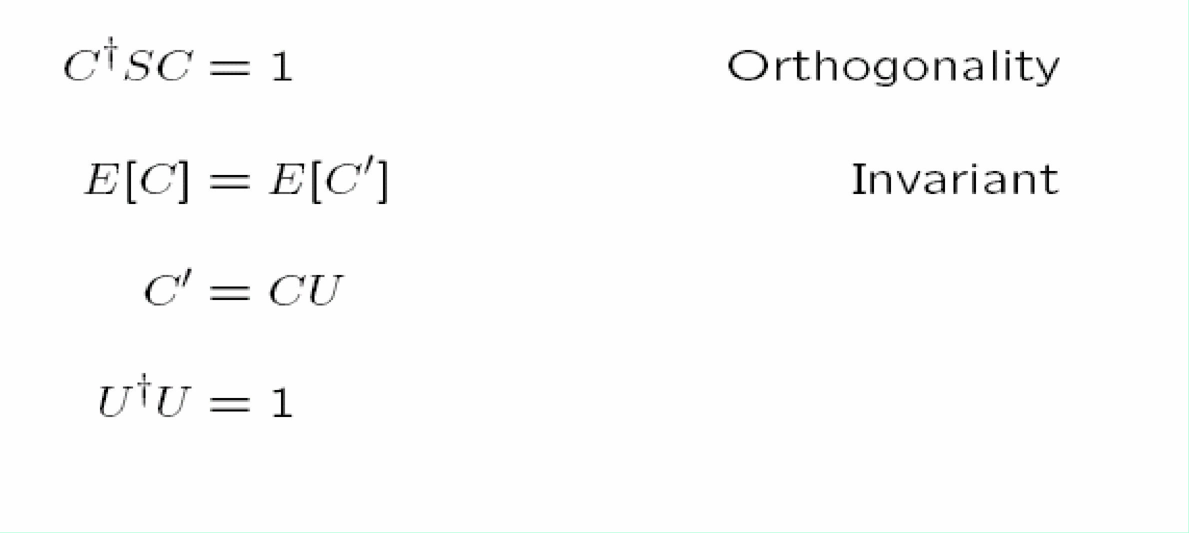
$$S_{\alpha\beta} = \int \chi_\alpha^*(r)\, \chi_\beta(r)\, dr$$ overlap matrix

$$H_{\alpha\beta} = \int \chi_\alpha^*(r)\, \mathcal{H}(r)\, \chi_\beta(r)\, dr$$ Hamiltonian matrix

$$E_{ij} = \delta_{ij}\, \epsilon_i$$ Orbital energies

$$\boxed{\boldsymbol{HC = SCE}}$$

Orbitals



$C^\dagger S C = 1$ Orthogonality

$E[C] = E[C']$ Invariant

$C' = CU$

$U^\dagger U = 1$

Density matrix

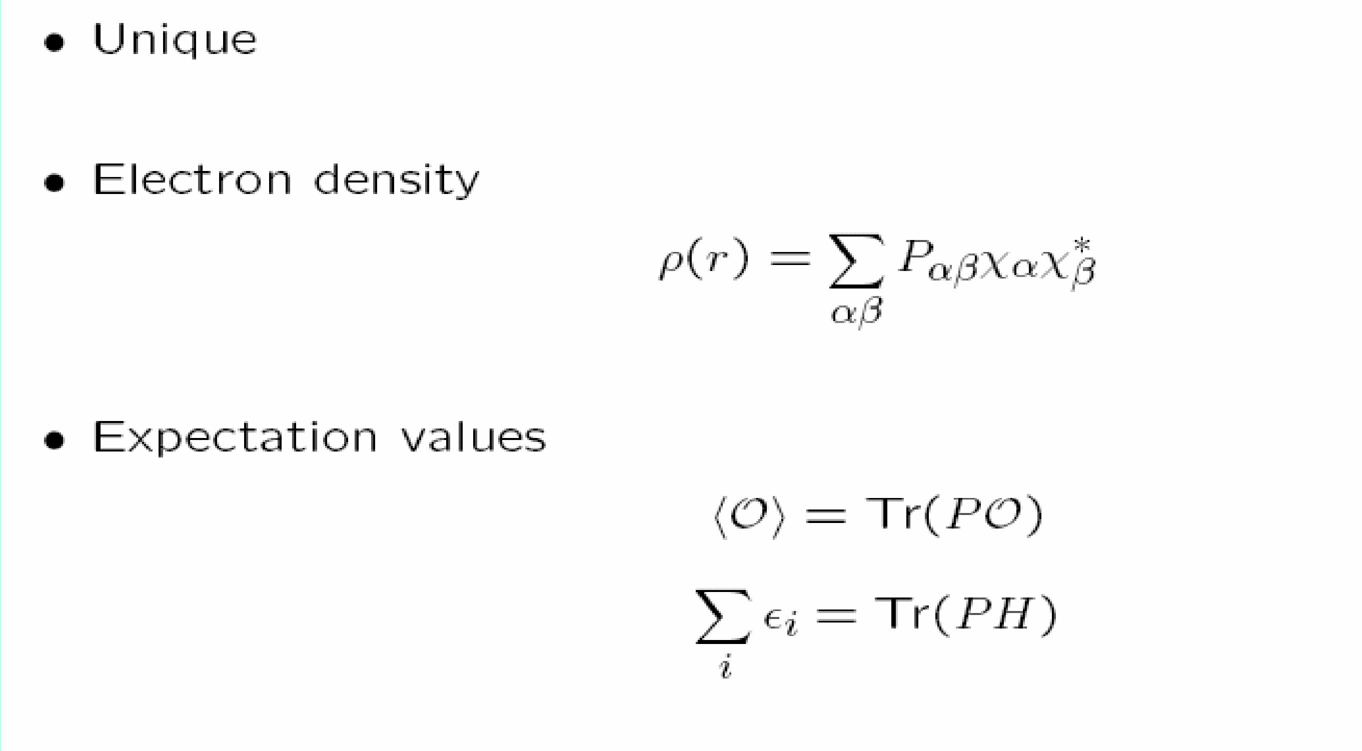
$$\boxed{P_{\alpha\beta} = 2 \sum_{i=1}^{N_e/2} c_{\alpha i}\, c^*_{\beta i}}$$

Properties of the density matrix

$$Tr(\boldsymbol{PS}) = \sum_{\alpha\beta}^{M} P_{\alpha\beta} S_{\alpha\beta} = N_e$$ normalisation

$$\boldsymbol{P} = \frac{1}{2}\boldsymbol{PSP}$$ idempotency

Density matrix



- Unique

- Electron density

$$\rho(r) = \sum_{\alpha\beta} P_{\alpha\beta}\chi\alpha\chi_{\beta}^{*}$$

- Expectation values

$$\langle\mathcal{O}\rangle = \mathrm{Tr}(P\mathcal{O})$$

$$\sum_{i} \epsilon_i = \mathrm{Tr}(PH)$$

Gaussian basis

Basis functions

$$\chi(r) = x^l y^m z^n \,\mathrm{exp}[-\alpha r^2]$$

Product of basis functions

$$\chi(r-A)\,\chi(r-B) = \tilde{\chi}(r-C)$$

Localization

$$\mathrm{exp}[-\alpha r^2] \rightarrow \mathrm{FFT} \rightarrow \mathrm{exp}[-\frac{G^2}{4\alpha}]$$

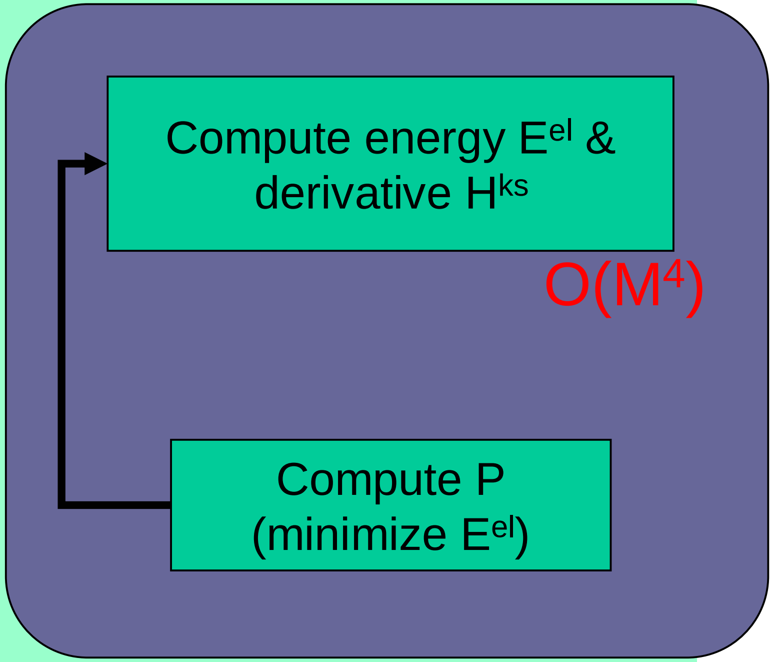
DFT with gaussians

$\phi_\mu(\boldsymbol{r})$ Atom centered Gaussians

$$n(\mathbf{r}) = \sum_{\mu\nu} P^{\mu\nu} \phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})$$

$$\begin{aligned}
E^{el}[P^{\mu\nu}] = {} & \sum_{\mu\nu} P^{\mu\nu} \int \phi_\mu(\mathbf{r})\left(-\frac{\Delta}{2}\right)\phi_\nu(\mathbf{r}) \\
& + \sum_{\mu\nu} P^{\mu\nu} \int\!\!\int \phi_\mu(\mathbf{r}) V^{PP}_{sep}(\mathbf{r},\mathbf{r}')\phi_\nu(\mathbf{r}') \\
& + \frac{1}{2}\int\!\!\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\
& + \int n(\mathbf{r})\varepsilon_{xc}[n](\mathbf{r})
\end{aligned}$$

$P^{\mu\nu}$ Density matrix

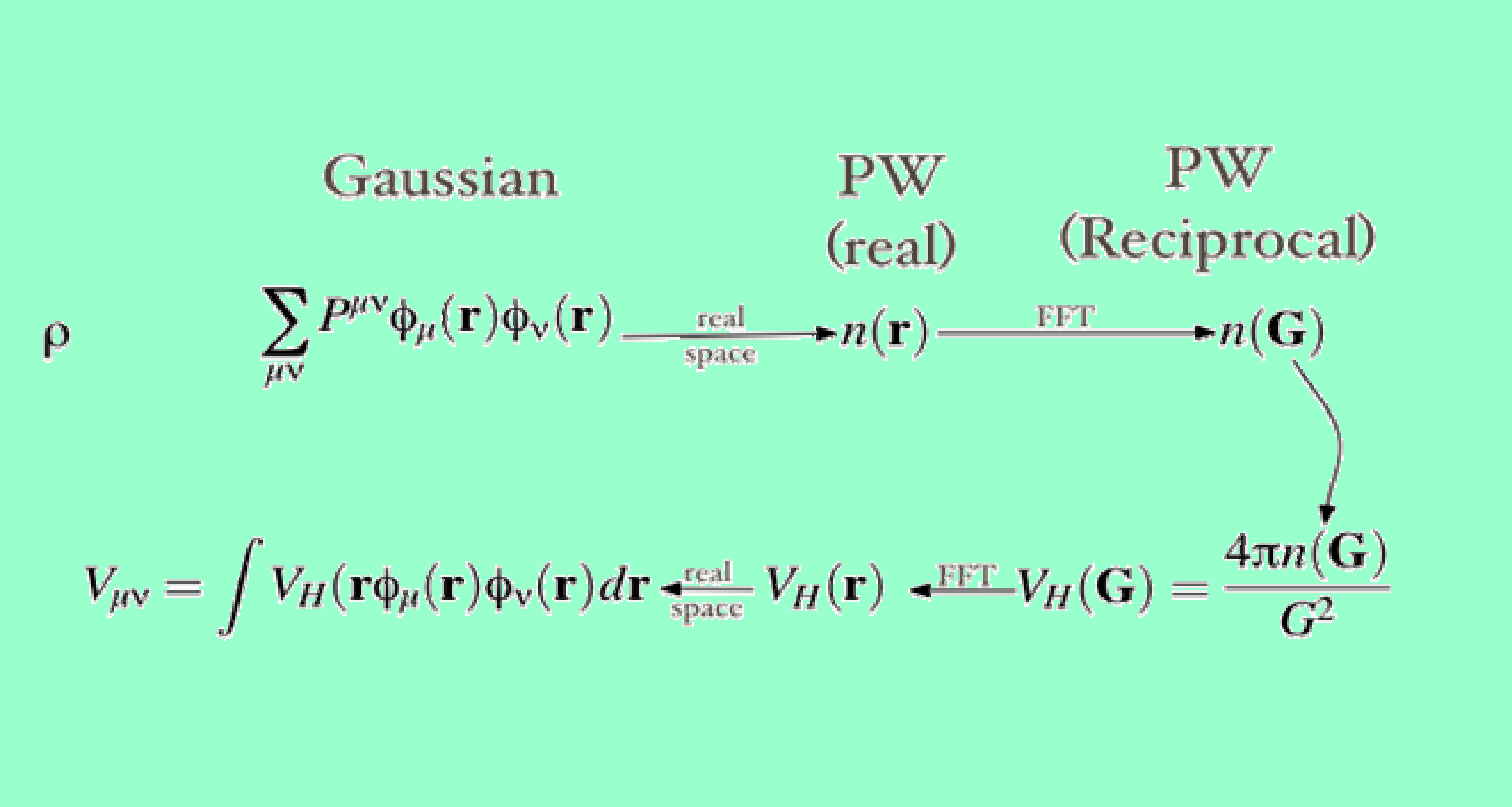


Compute energy E^{el} & derivative H^{ks}

$O(M^4)$

Compute P (minimize E^{el})

The best of both worlds



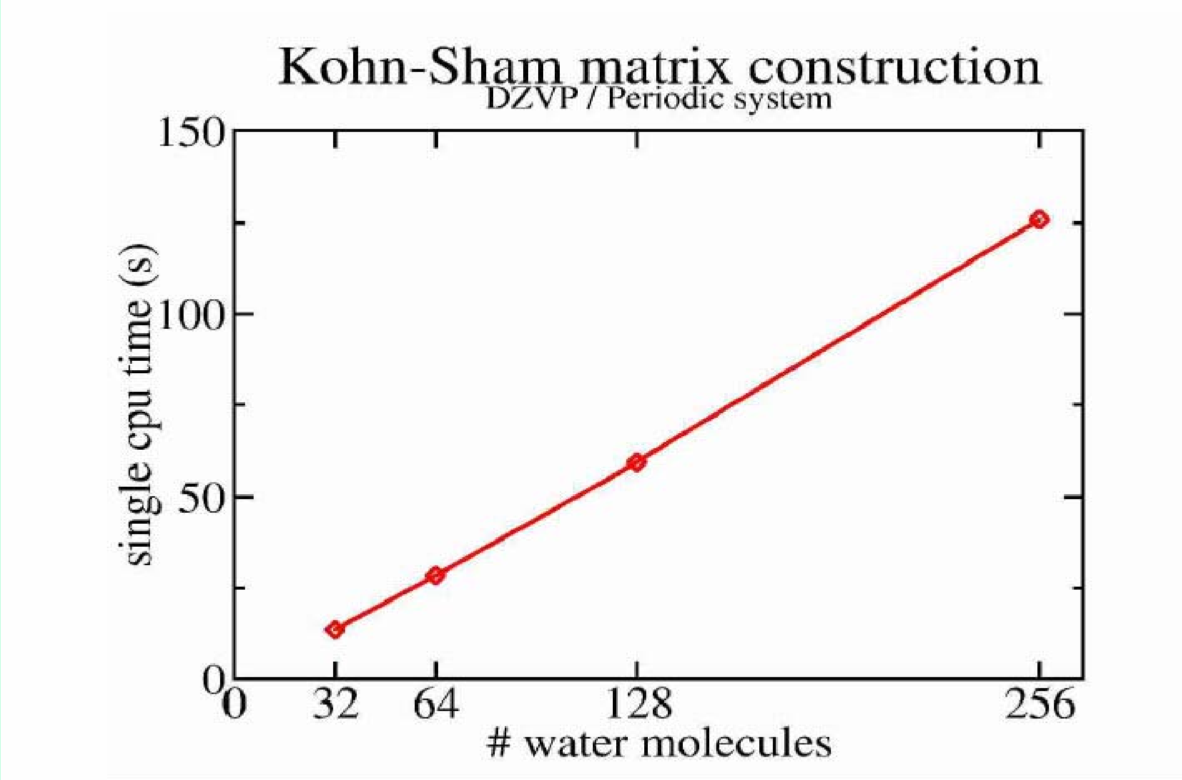
Gaussian

PW (real)

PW (Reciprocal)

$$\begin{tikzcd}
\rho & \sum_{\mu\nu} P^{\mu\nu}\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}) \arrow[r, "\text{real}", "\text{space}"'] & n(\mathbf{r}) \arrow[r, "\text{FFT}"] & n(\mathbf{G}) \arrow[d] \\
 & V_{\mu\nu} = \int V_H(\mathbf{r}\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})d\mathbf{r} & V_H(\mathbf{r}) \arrow[l, "\text{real}"', "\text{space}"] & V_H(\mathbf{G}) = \frac{4\pi n(\mathbf{G})}{G^2} \arrow[l, "\text{FFT}"']
\end{tikzcd}$$

Linear scaling



Standard approach

Solve by diagonalization

$$H_{\mu\nu} C^{\nu i} = S_{\mu\delta} C^{\delta i} \varepsilon_i$$

Construct P

$$P^{\mu\nu} = \sum_i C^{\mu i} C^{\nu i}$$

$$i = 1 \ldots N$$

$$\mu, \nu, \delta = 1 \ldots M$$

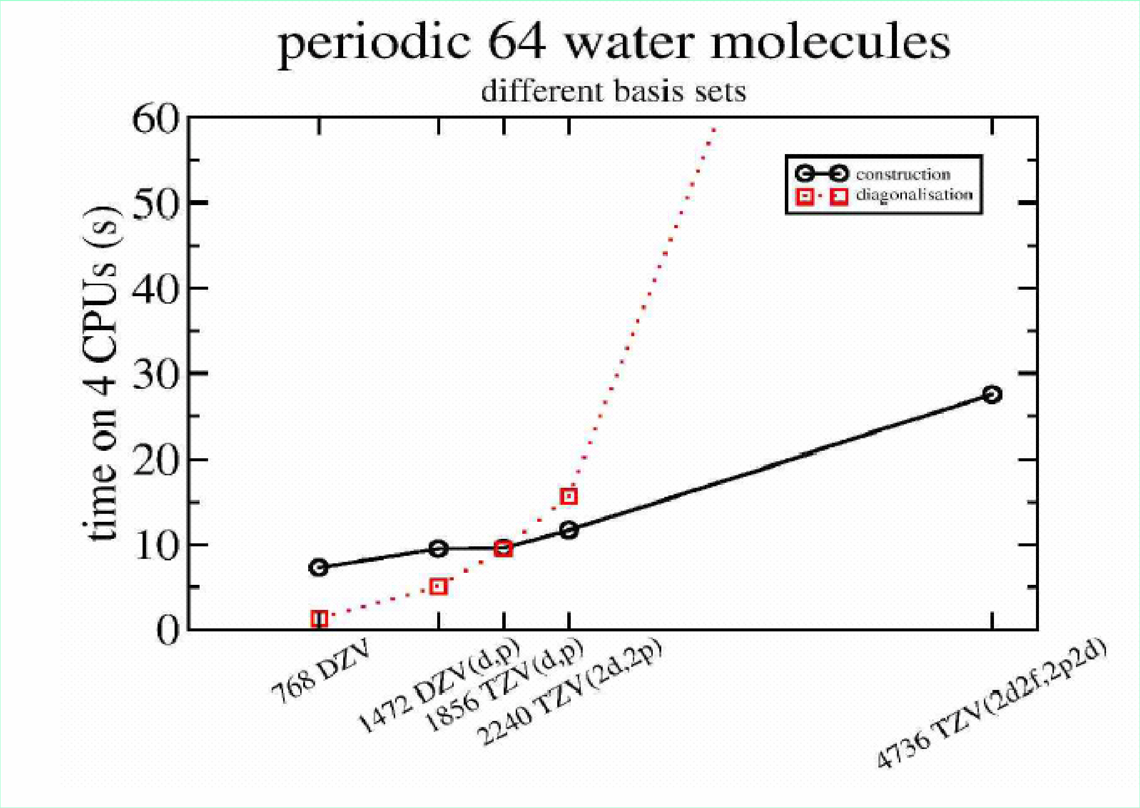
Orbital rotation

$$C(X) = C_0 \cos(\sqrt{X^T S X}) + X \frac{\sin(\sqrt{X^T S X})}{\sqrt{X^T S X}}$$

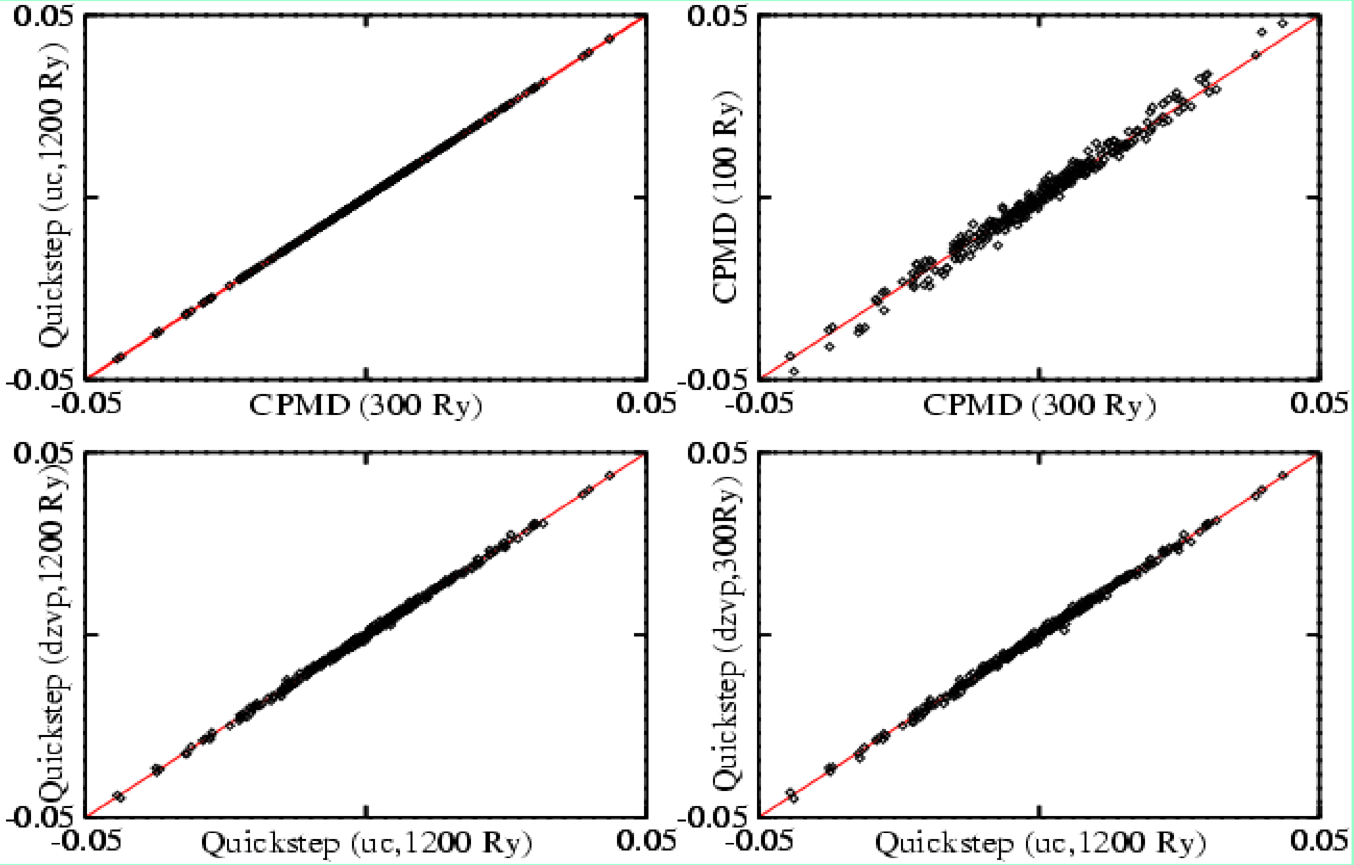
$$X^T S C_0 = 0 \qquad C(X)^T S C(X) = 1 \ \forall X$$

Minimize the energy with respect to X

Overall cubic scaling

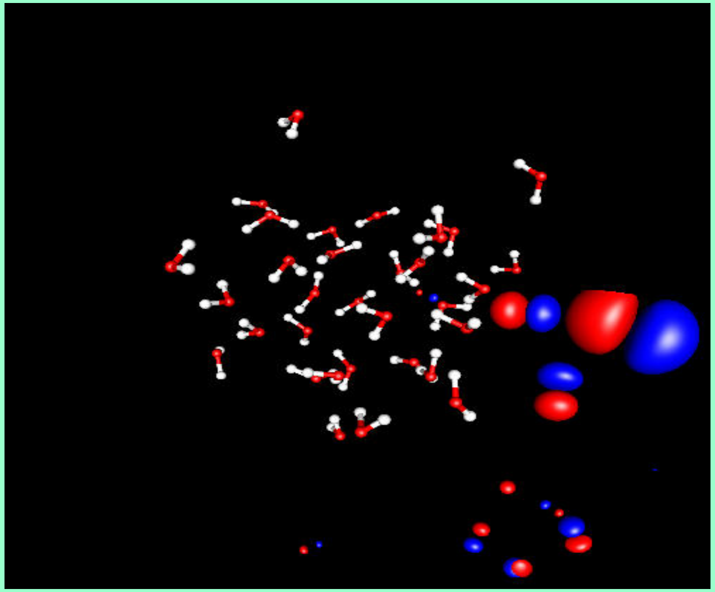


Accurate forces

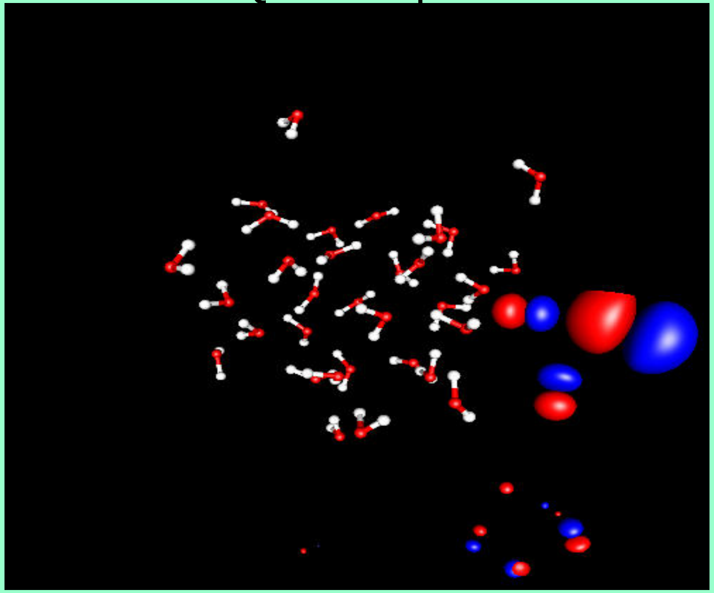


Checking the eigenstates

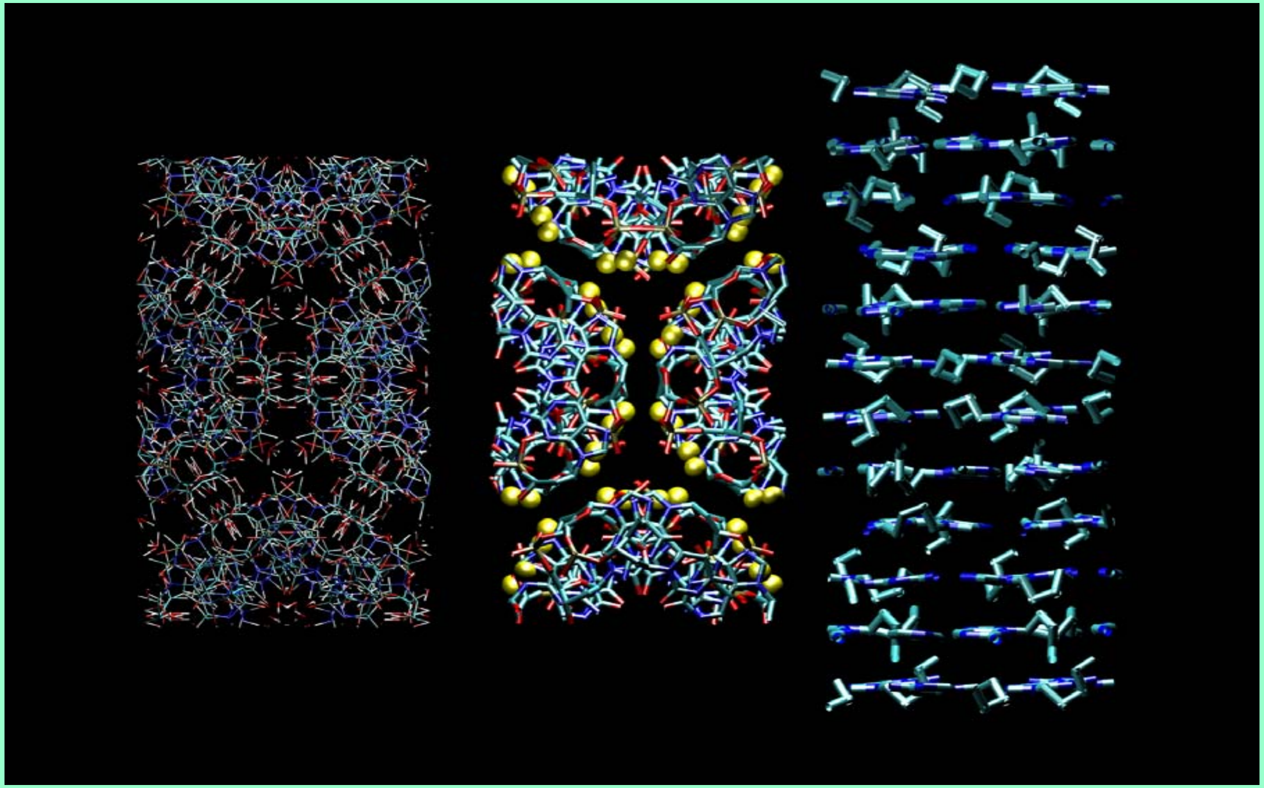
CPMD



Quickstep



DNA crystal



2388 atoms, 3960 orbitals

DZV(d,p) 22596, TZV(2d,2p) 38688

675, 1100 sec / line search (SP4-32-1.3G)

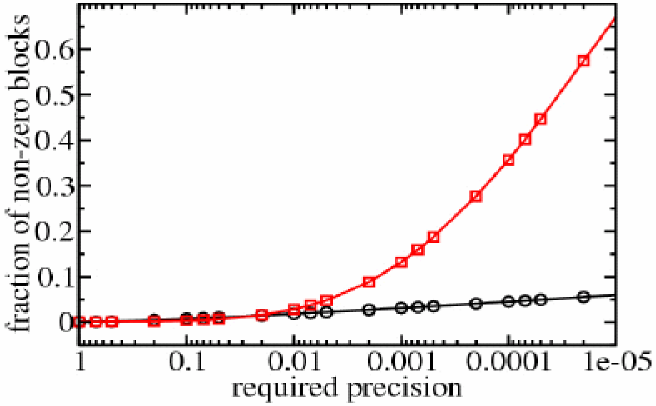
2.5, 5 h /total

Not yet fully cubic (45,43,8 % 3,2,1)

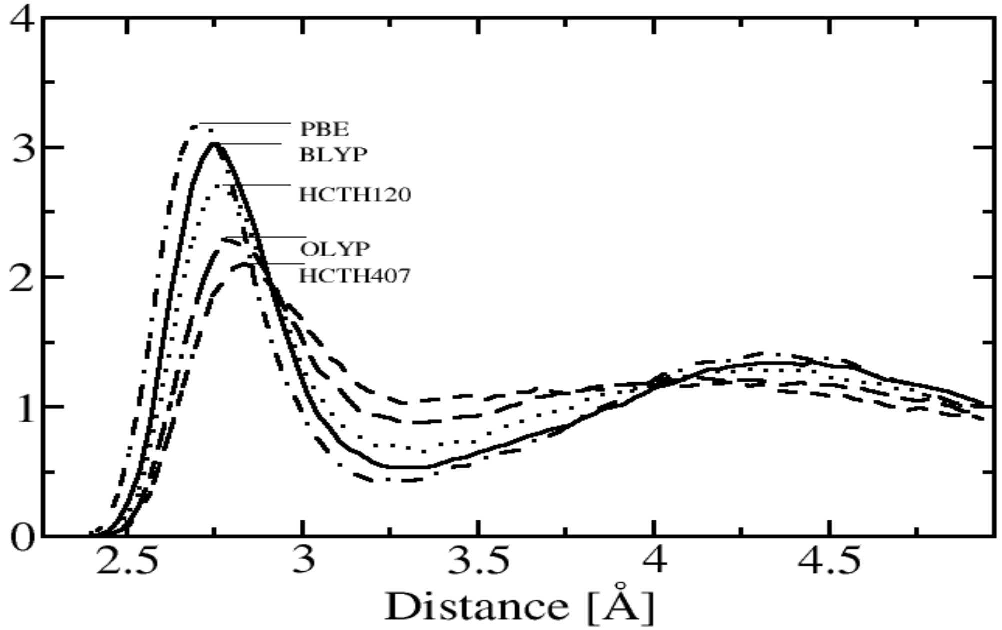
Not yet sparse

Example: DNA Crystal

2388 atoms, 3960 orbitals, 38688 BSF (TZV(2d,2p))
density matrix, overlap matrix



One order of magnitude better



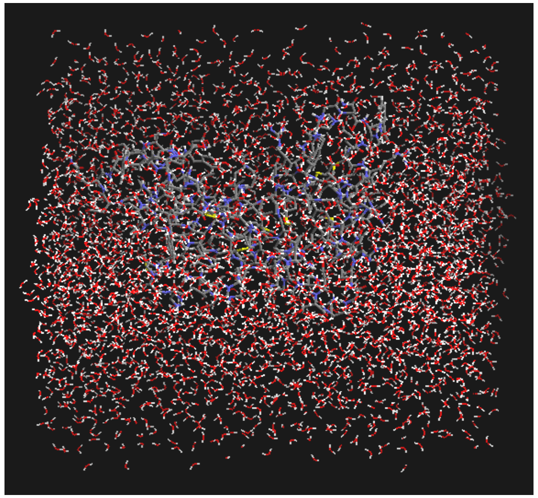
20 ps
QZV2P

PBE/BLYP overstructured
HCTH407/OLYP understructured

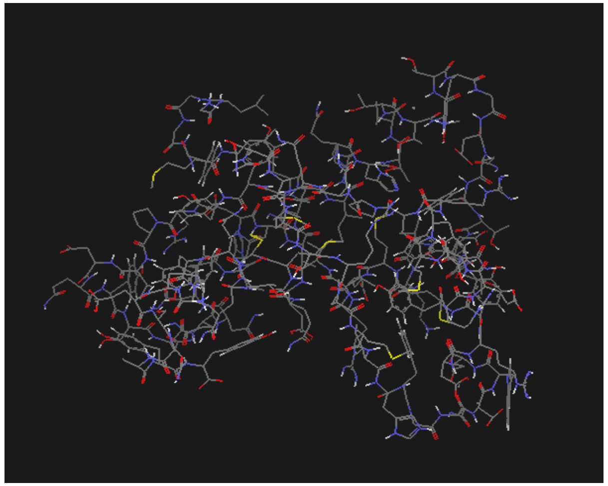
Significant influence of DFT

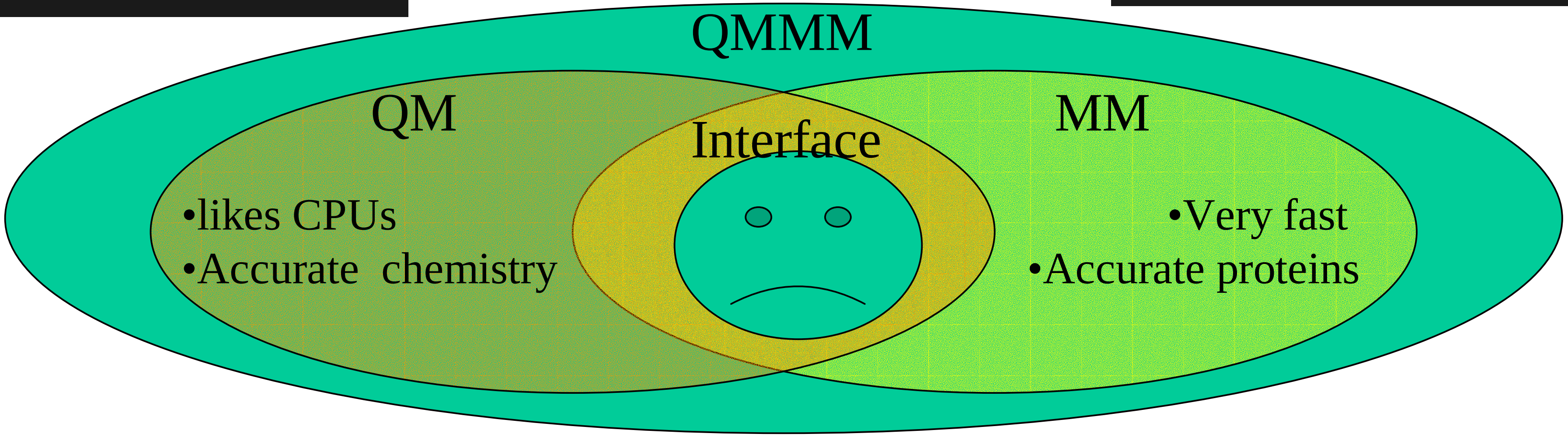
Why should one combine QM and MM ?

Bio-systems are typically very large and catalyse complicated reactions



- Proteins >1000 atoms
- Solvent >10000 atoms
- The active site ~ 100 atoms





- likes CPUs
- Accurate chemistry

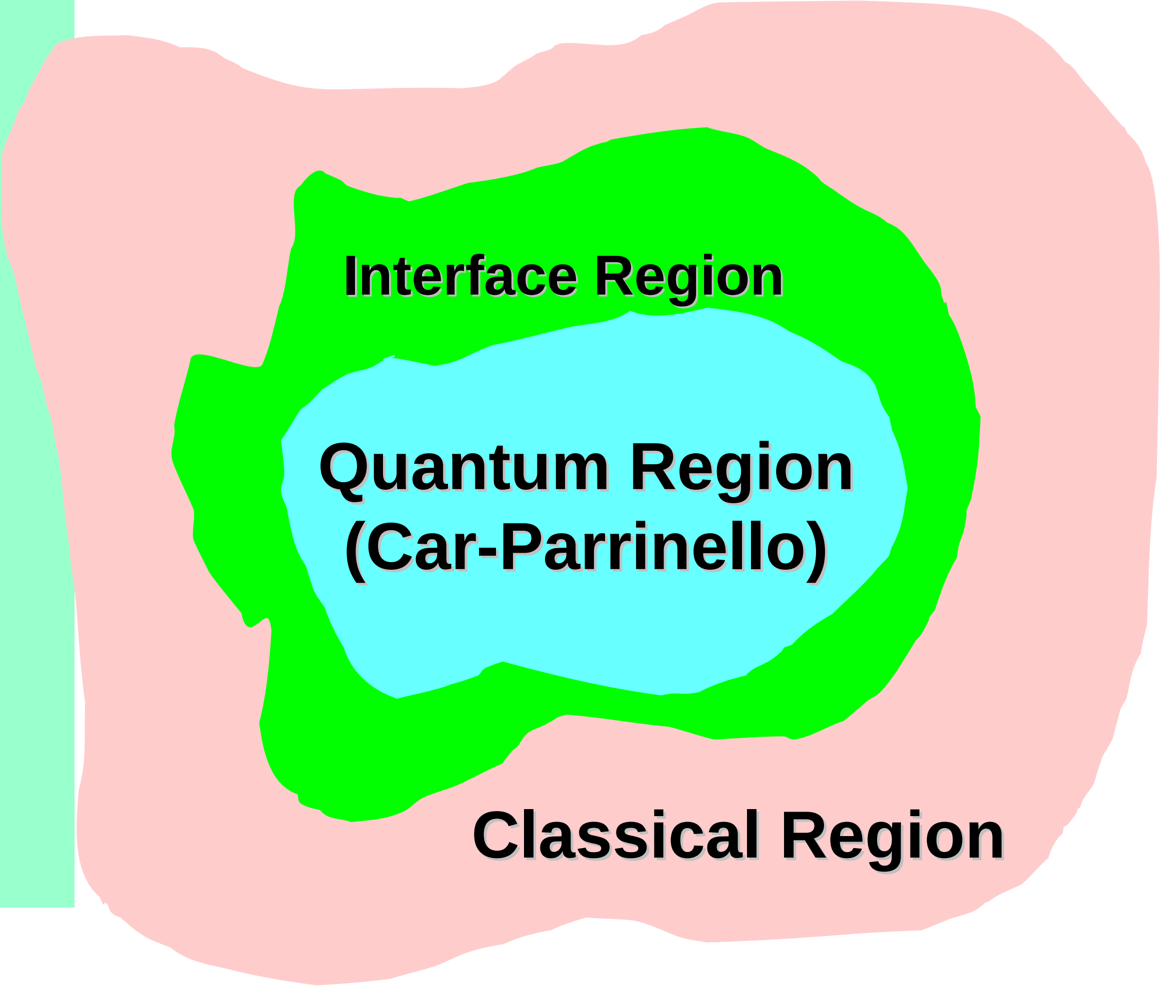
- Very fast
- Accurate proteins

Mixed Quantum-Classical QM/MM- Car-Parrinello Simulations

- highly parallel QM/MM Car-Parrinello hybrid code
- **Fully Hamiltonian**
- MD driver: CPMD

QM-Part: CPMD 3.3
(pbcs (2 boxes), plane waves,
pseudo potentials, GGAs:
BP86, BLYP, PW91, PBE...)
n-1 nodes

MM-Part: GROMOS96 + P3M,
AMBER)
1 node



Interface Region

Quantum Region (Car-Parrinello)

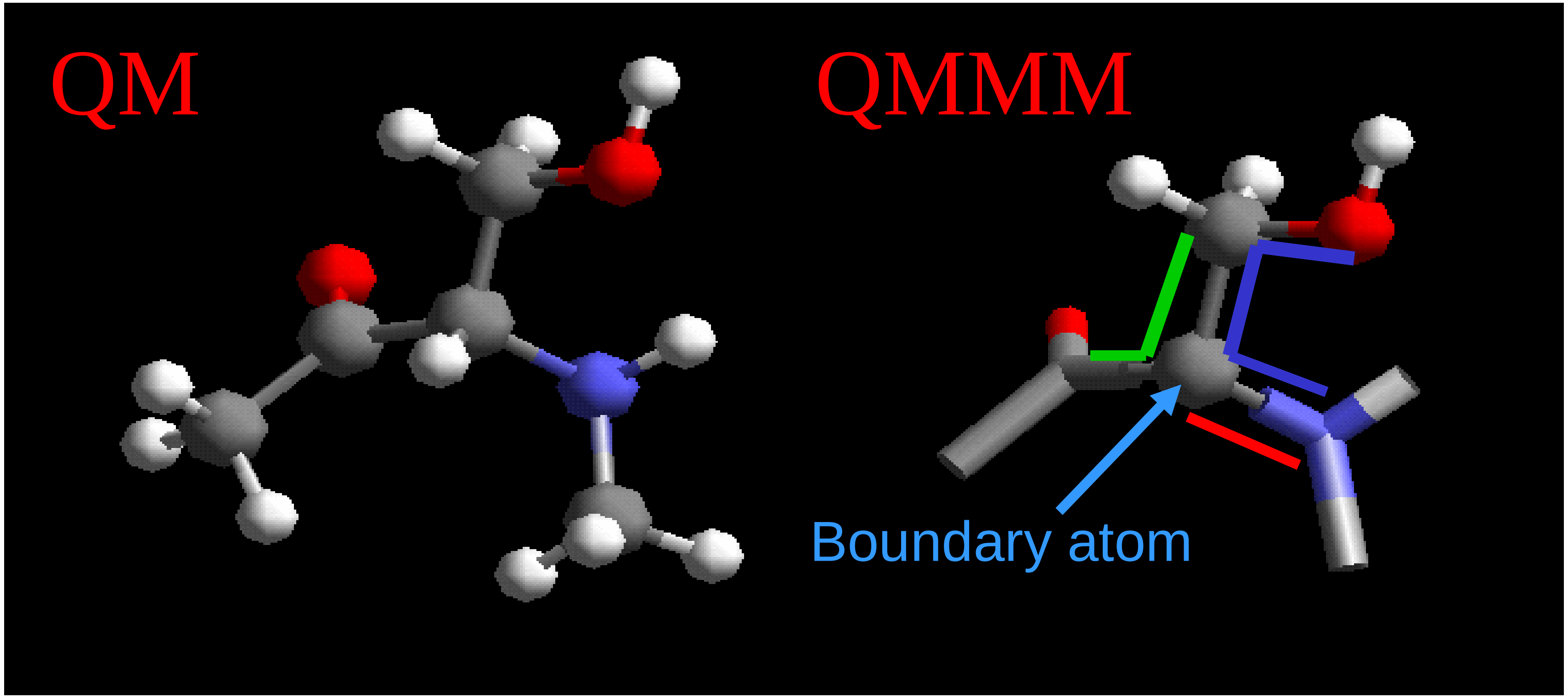
Classical Region

QM/MM- Car-Parrinello Simulations

- Development of improved QM/MM interfaces:

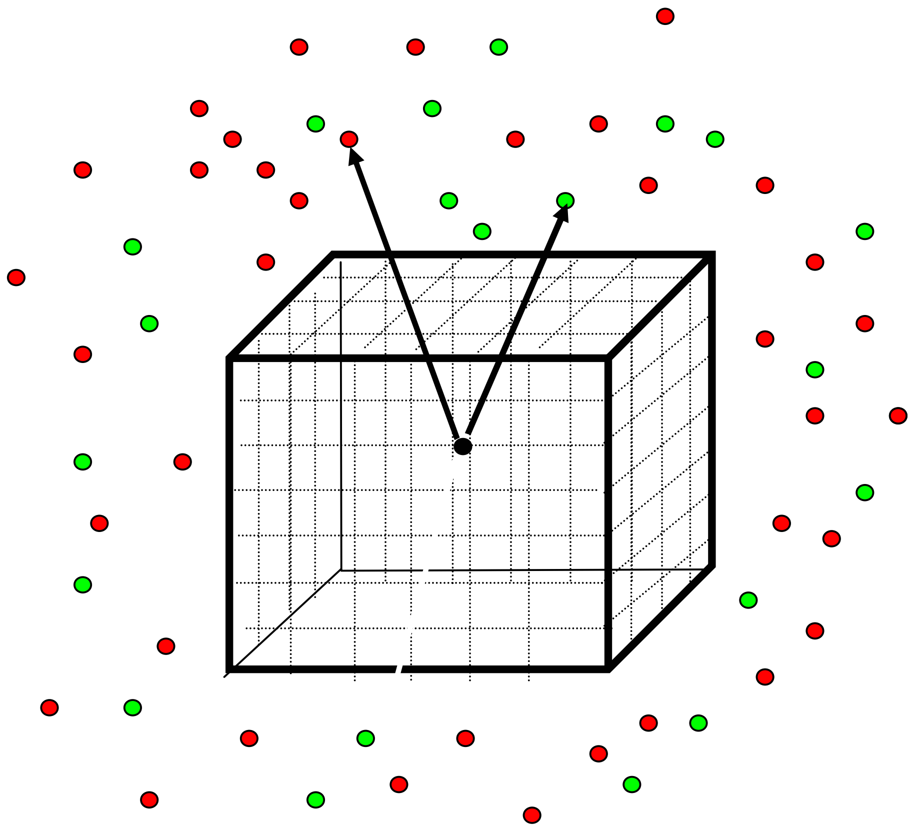
  - pseudo potentials for boundary atoms
  - efficient treatment of long-range electrostatics
  - electron spill out problem

The Bonded Part



- boundary atoms: monovalent pseudopotential
- distances - angles - torsions involving MM and QM atoms come from the force field

Mixed Quantum-Classical Simulations



Coulomb interaction:

$$H_{el} = \sum_{i \in MM} q_i \int dr \frac{\rho(r)}{|r - r_i|}$$

3D-grid: (NR1,NR2,NR3)   NR~100

MM atoms ~10000-100000

⇩

NR1*NR2*NR3*MM
DISTANCE CALCULATIONS!!!

“Variational” ESP charges

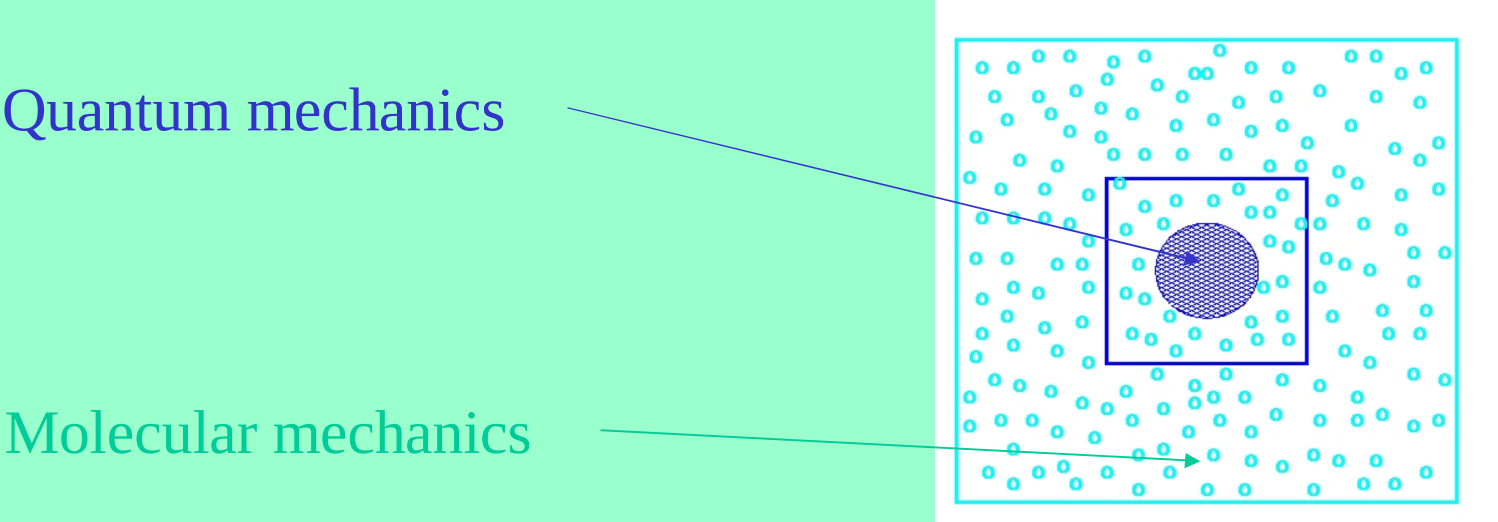
- charges, located on the QM atoms, are fitted to the electrostatic fields on the MM atoms due to the electronic charge distribution

$$\sum_{i\in QM} \frac{q_i^{ESP}}{r_{ij}} \overset{\text{(Least square)}}{=} \int dr\, \rho(r)\, v(|r - r_i|) \qquad \forall\ i\in NN$$

modified Coulomb

Explicit dependence of q^{ESP} on the positions of all the MM and QM atoms and on the electronic density

QM/MM



Quantum mechanics

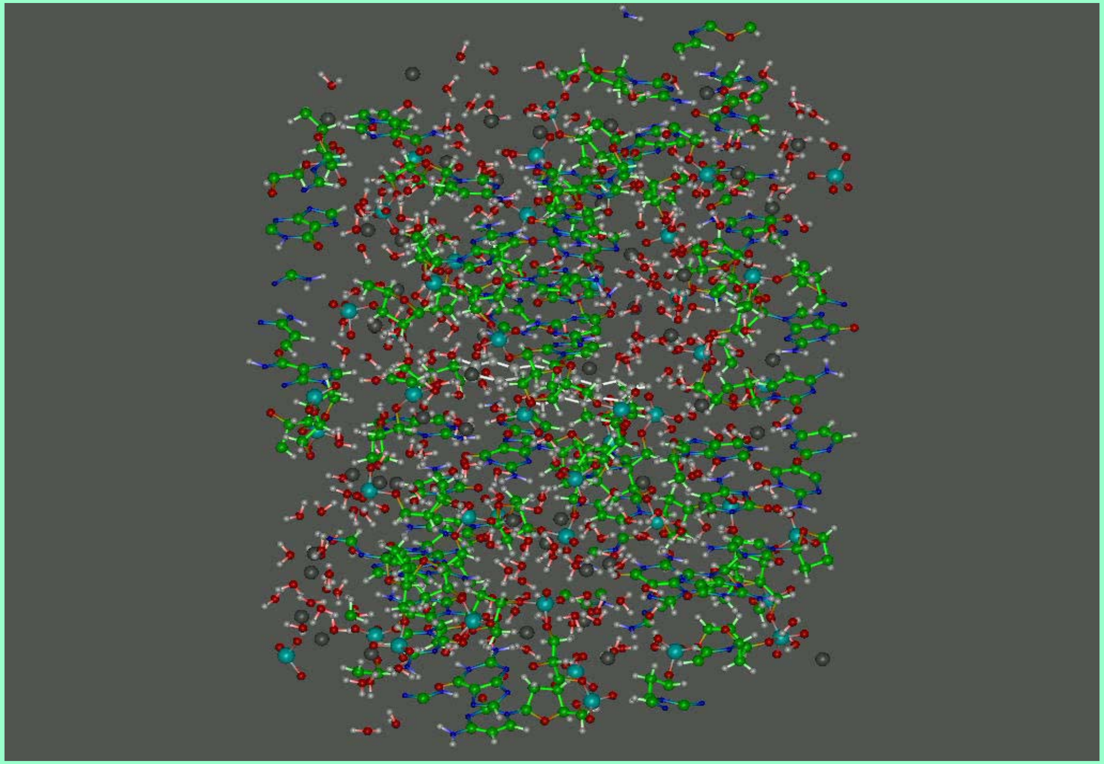
Molecular mechanics

$$L=\frac{1}{2}\sum_{i}\mu\int d\vec{r}\left|\dot{\psi}_{i}\right|^{2}+\frac{1}{2}\sum_{I}M_{I}\dot{\vec{R}}_{I}^{2}-E_{KS}\left[\left\{\psi_{i}\right\},R_{I}\right]+\sum_{i,j}\Lambda_{i,j}\left(\left\langle\psi_{i}\middle|\psi_{j}\right\rangle-\delta_{i,j}\right)$$

$$+\frac{1}{2}\sum_{I'}m\dot{\vec{R}}_{I'}^{2}-\sum_{I'<J'}V_{LL}\left(\left|\vec{R}_{I'}-\vec{R}_{J'}\right|\right)-\sum_{I,I'}V_{CL}\left(\left|\vec{R}_{I}-\vec{R}_{I'}\right|\right)$$

A. Laio, J. VandeVondele and U. Röthlisberger, Chem. Phys. ***116****, 6941(2002)*

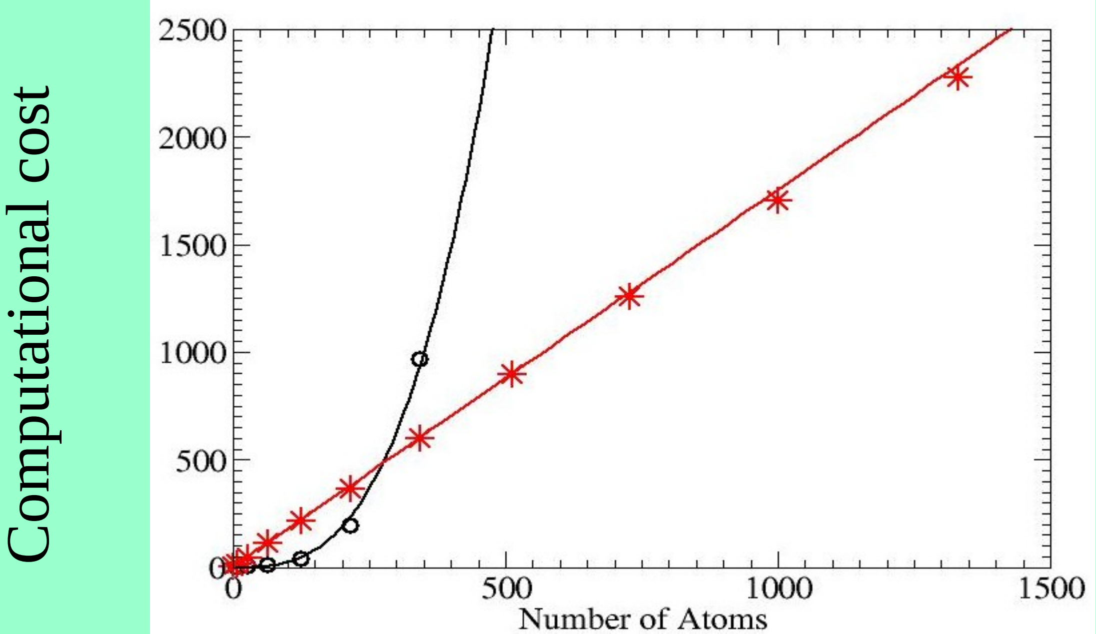
Capturing the complexity



DNA oxidation

The way of the future?

O(N) vs O(N3)



Highly parallel and local new field theoretical approach

The quickstep team

J. Hutter, University of Zurich

J. VandeVondele, University of Cambridge

W. I-Feng Kuo, LLNL

C. Mundy, LLNL

Fawzi Mohammed, ETH Lugano

M. Krack, ETH Lugano

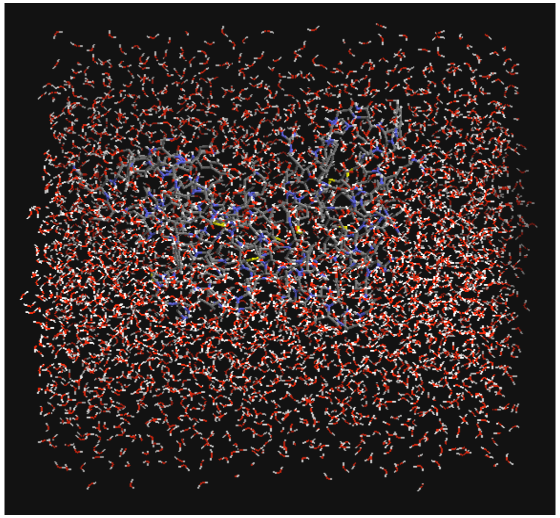
G. Lippert, BASF

M. McGrath, LLNL

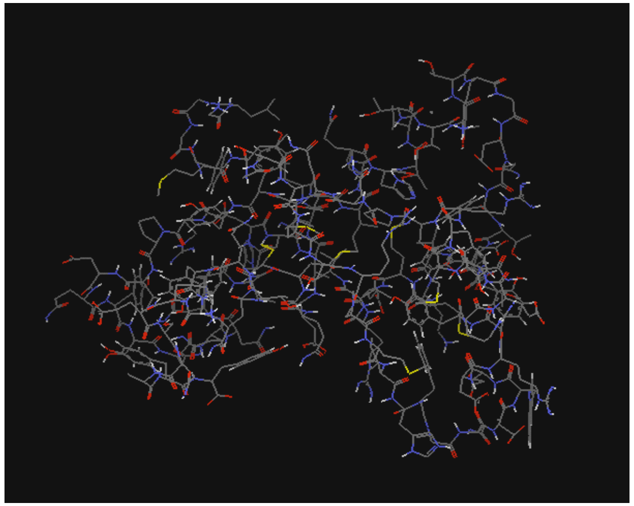
I. Siepman, LLNL

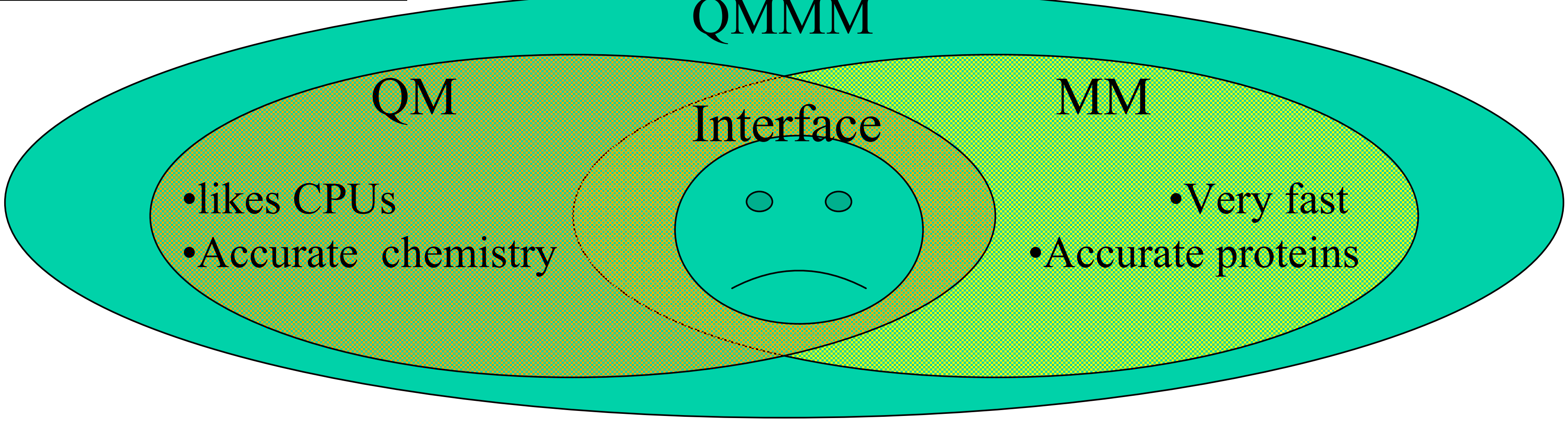
Why should one combine QM and MM ?

Bio-systems are typically very large and catalyse complicated reactions



- Proteins >1000 atoms
- Solvent >10000 atoms
- The active site ~ 100 atoms





QMMM

QM

- likes CPUs
- Accurate chemistry

Interface

MM

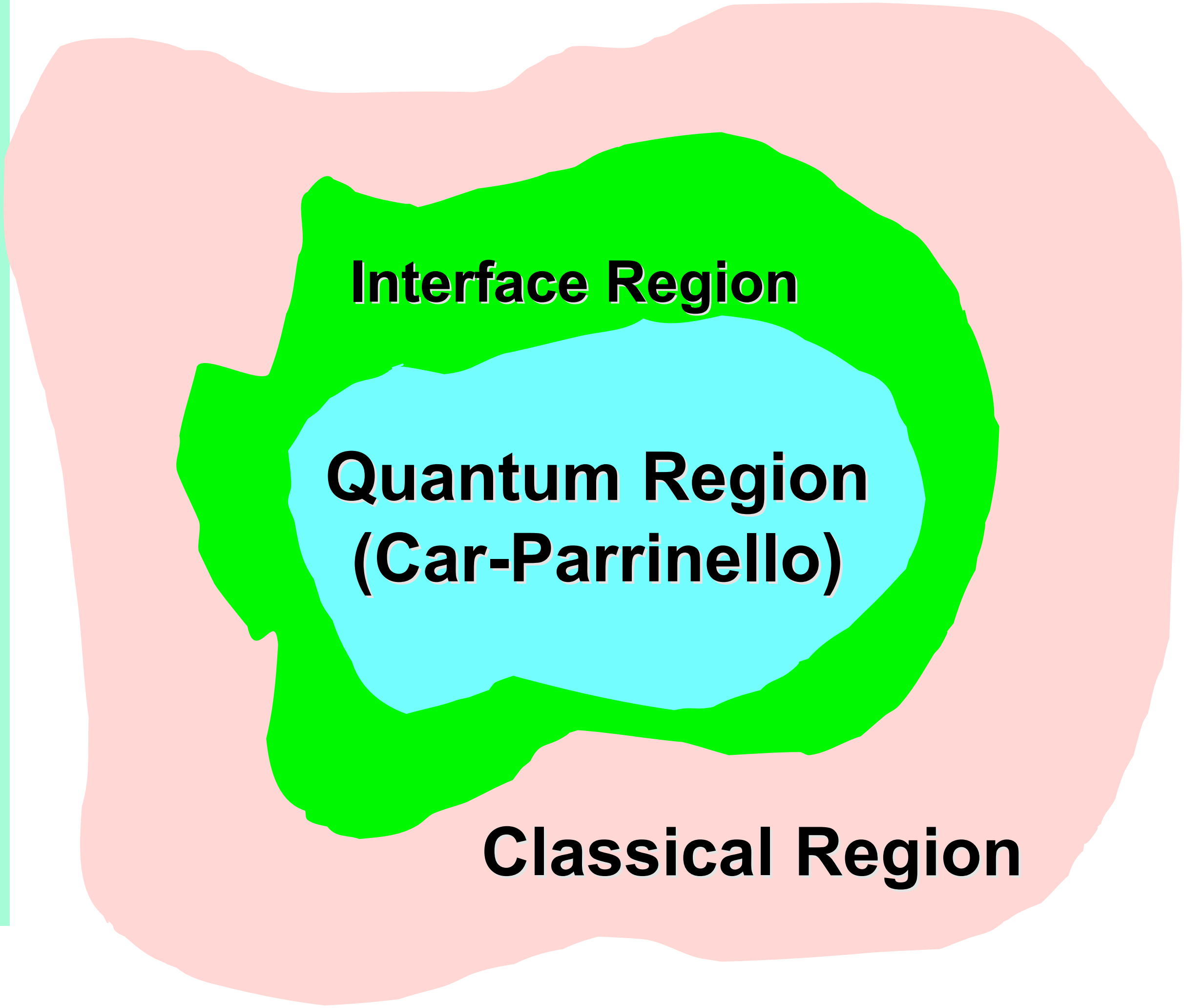
- Very fast
- Accurate proteins

Mixed Quantum-Classical QM/MM- Car-Parrinello Simulations

- highly parallel QM/MM Car-Parrinello hybrid code
- **Fully Hamiltonian**
- MD driver: CPMD

QM-Part: CPMD 3.3
(pbcs (2 boxes), plane waves, pseudo potentials, GGAs: BP86, BLYP, PW91, PBE...)
n-1 nodes

MM-Part: GROMOS96 + P3M, AMBER)
1 node



Interface Region

Quantum Region
(Car-Parrinello)

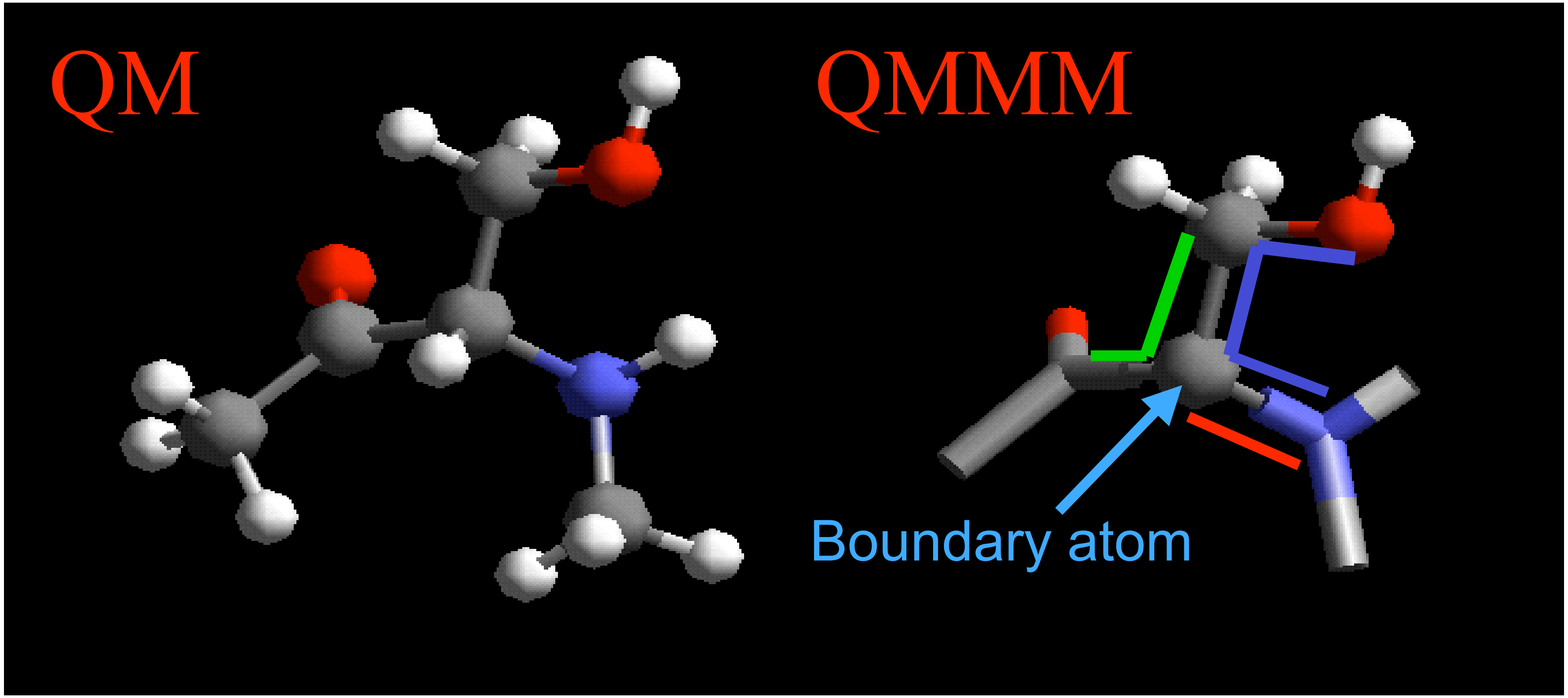
Classical Region

QM/MM- Car-Parrinello Simulations

- Development of improved QM/MM interfaces:

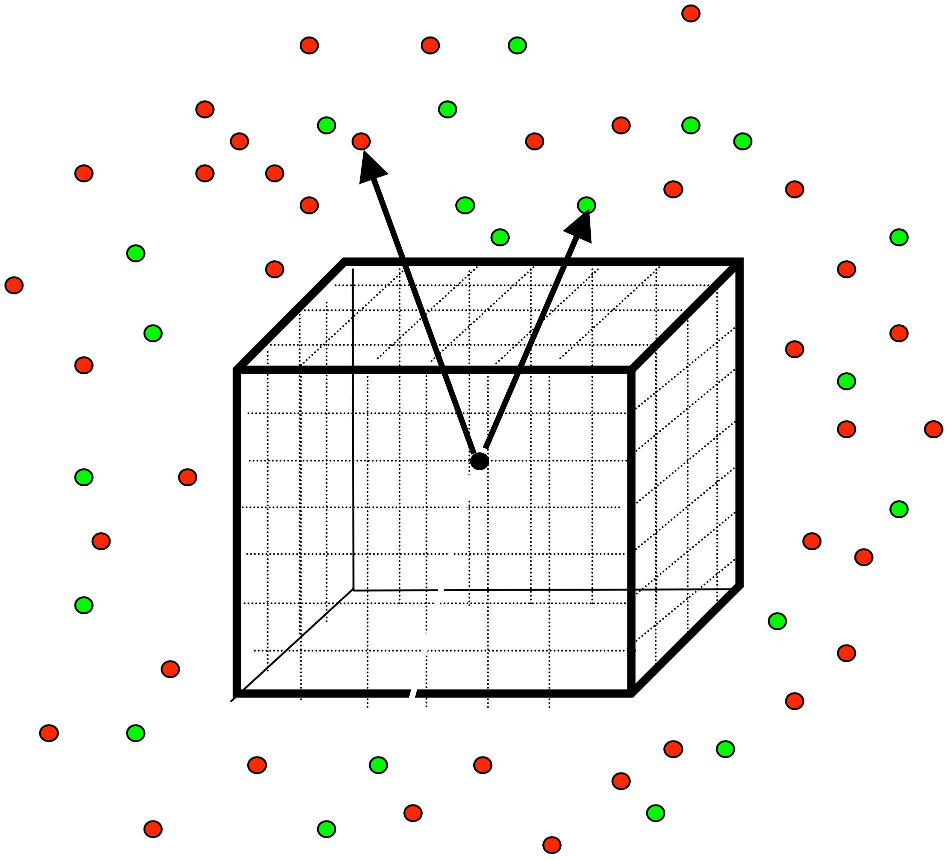
  - pseudo potentials for boundary atoms
  - efficient treatment of long-range electrostatics
  - electron spill out problem

The Bonded Part



- boundary atoms: monovalent pseudopotential
- distances - angles - torsions involving MM and QM atoms come from the force field

Mixed Quantum-Classical Simulations



ρ(nr1i,nr2j,nr3k)

Coulomb interaction:

$$H_{el} = \sum_{i \in MM} q_i \int dr \frac{\rho(r)}{|r - r_i|}$$

3D-grid: (NR1,NR2,NR3) NR~100

MM atoms ~10000-100000

⇩

NR1*NR2*NR3*MM
DISTANCE CALCULATIONS!!!

“Variational” ESP charges

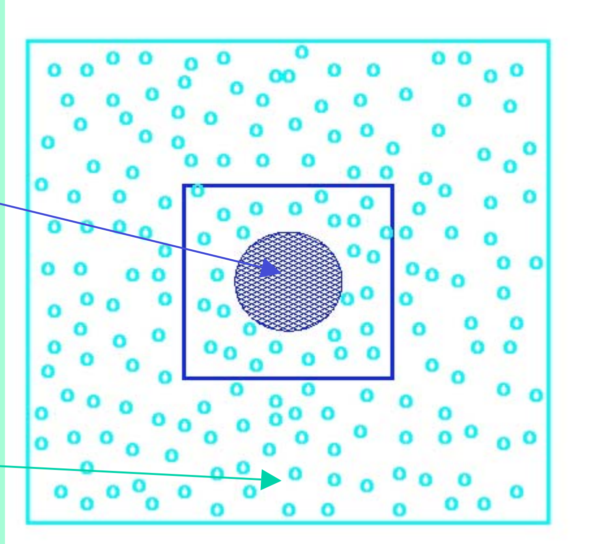
- charges, located on the QM atoms, are fitted to the electrostatic fields on the MM atoms due to the electronic charge distribution

$$\sum_{i\in QM} \frac{q_i^{ESP}}{r_{ij}} \overset{\text{(Least square)}}{=} \int dr\, \rho(r)\, v(|r - r_i|) \qquad \forall\ i\in NN$$

modified Coulomb

Explicit dependence of q^{ESP} on the positions of all the MM and QM atoms and on the electronic density

QM/MM



Quantum mechanics

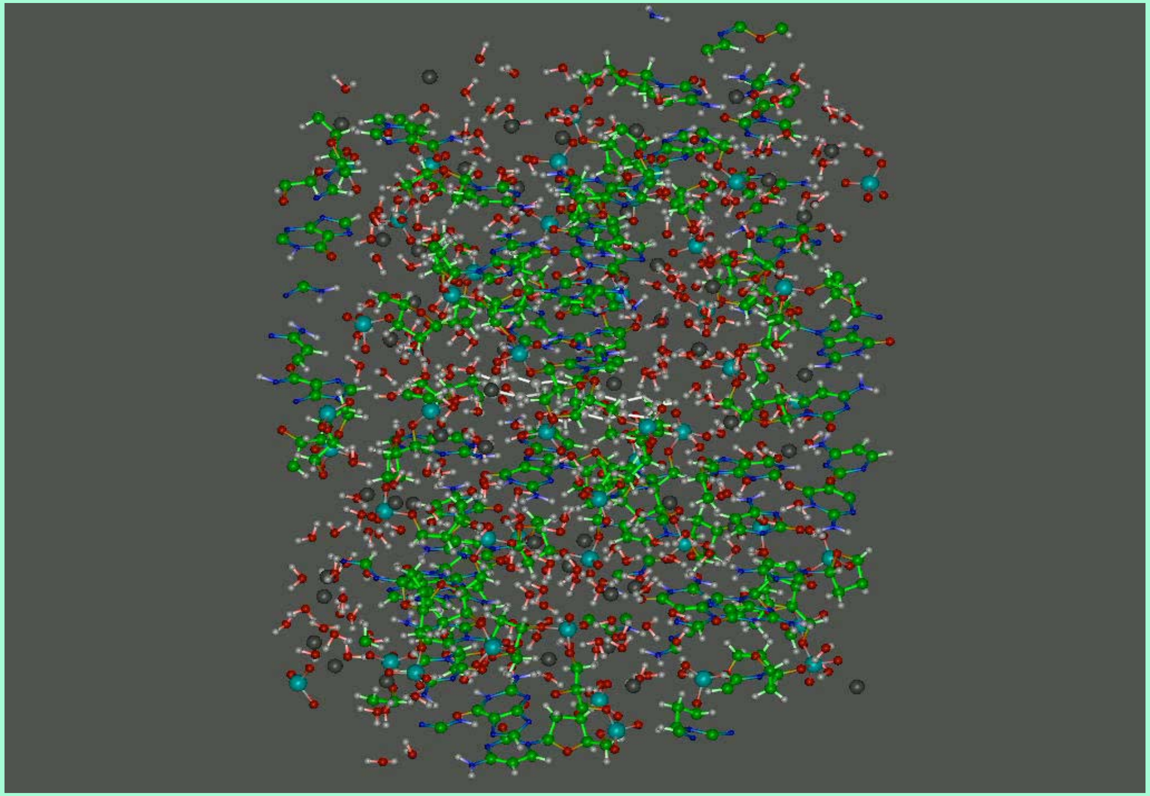
Molecular mechanics

$$L=\frac{1}{2}\sum_i \mu\int d\vec{r}\left|\dot{\psi}_i\right|^2+\frac{1}{2}\sum_I M_I\dot{\vec{R}}_I^2-E_{KS}\left[\{\psi_i\},R_I\right]+\sum_{i,j}\Lambda_{i,j}\left(\langle\psi_i|\psi_j\rangle-\delta_{i,j}\right)$$

$$+\frac{1}{2}\sum_{I'} m\dot{\vec{R}}_{I'}^2-\sum_{I'<J'}V_{LL}\left(\left|\vec{R}_{I'}-\vec{R}_{J'}\right|\right)-\sum_{I,I'}V_{CL}\left(\left|\vec{R}_I-\vec{R}_{I'}\right|\right)$$

A. Laio, J. VandeVondele and U. Röthlisberger, Chem. Phys. ***116****, 6941(2002)*

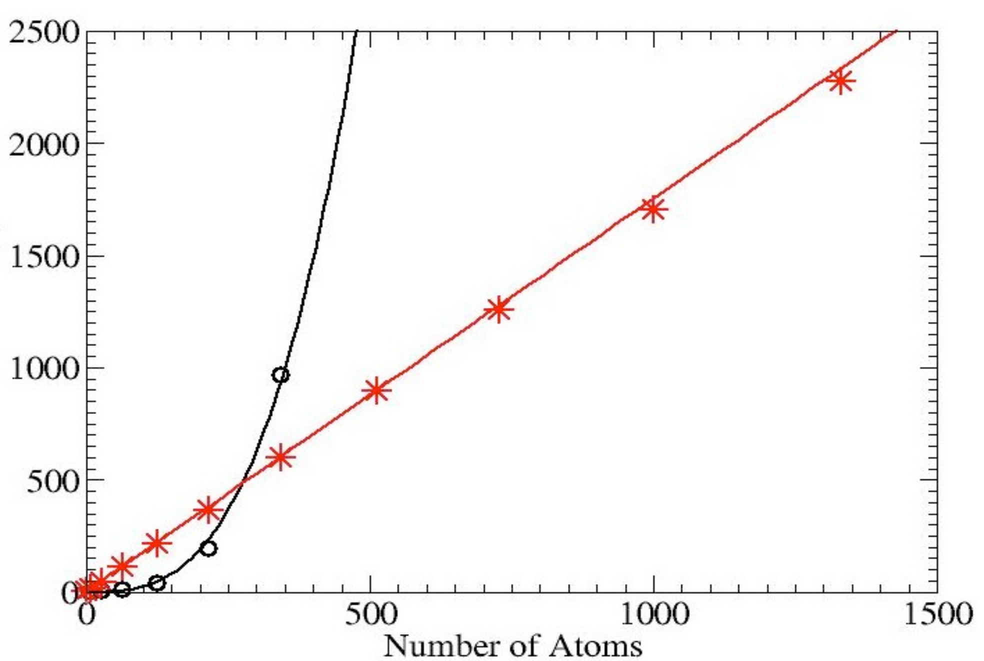
Capturing the complexity



DNA oxidation

The way of the future?

O(N) vs O(N^3)



Computational cost

Highly parallel and local new field theoretical approach

The time scale problem

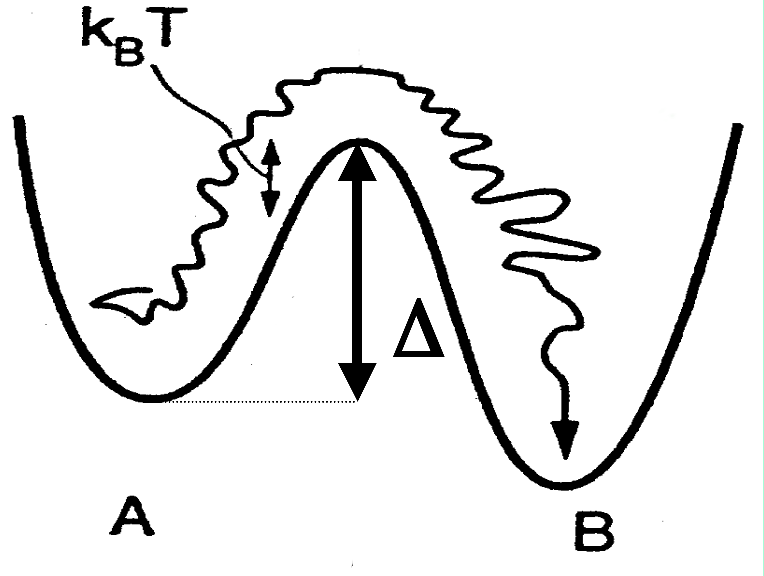
Direct simulation allows only very short runs:
~ 10 ps for ab-initio MD, ~10 ns for classical MD

Many relevant phenomena take place on a larger time scale:
chemical reactions, conformational changes, protein folding, etc.

Two-fold strategy:

a) Finding reactive paths
b) Exploring the free energy surface

Activated events



$$\tau \sim \tau_{mol} e^{\frac{F^*}{k_BT}}$$

$$F^* \sim \Delta$$

The quantum chemical approach

- Find the saddle point on the PES
- Use transition state theory
- Correct for zero point motion

Many different solutions proposed

- Thermodynamic integration
- "Flattening" the surface
  (hyperdynamics, puddle-skimming, umbrella sampling, etc.)
- Trajectory-based schemes
  (reaction path sampling, Lagrangian action minimization, nudged elastic band, etc.)
- Finding the saddle points
  (eigenvalue following, dimer method, hessian-based methods, etc.)
- Temperature enhanced sampling
  (histogram reweighting, parallel tempering,etc.)
- Etc. etc.

Driving the reaction

Suppose that the reaction coordinate ***q*** is known:

$$q(R_1, R_2, \ldots, R_N) = q$$

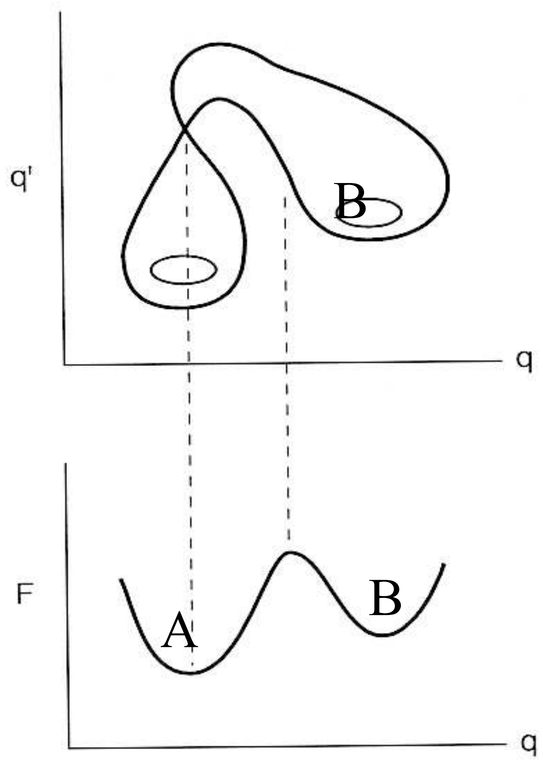
We force the reaction by adding a constraint term to the dynamics with a Lagrange multiplier:

$$\lambda\left(q(R_1, R_2, \ldots, R_N) - q\right)$$

$$\langle \lambda \rangle \approx \frac{\partial F}{\partial q}$$

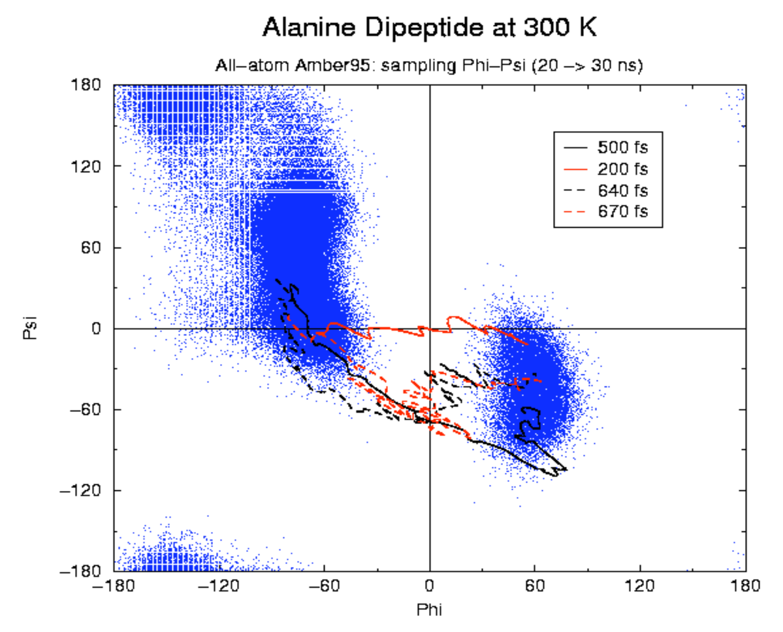
The activation energy is: $\Delta F = \int_{q_A}^{q^*} dq \langle \lambda \rangle_q$

The right reaction path?



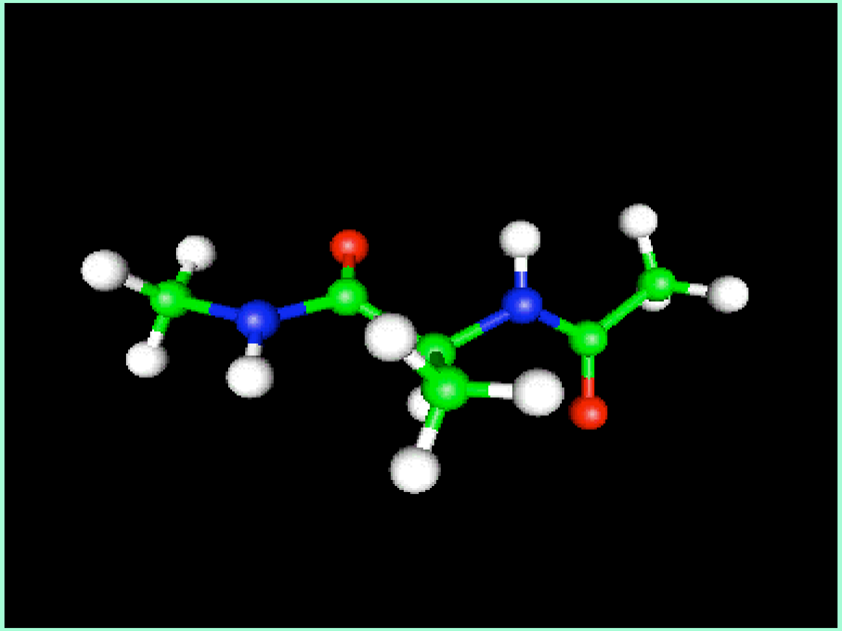
Activated events are often intrinsically multidimensional!!!

Life is complicated



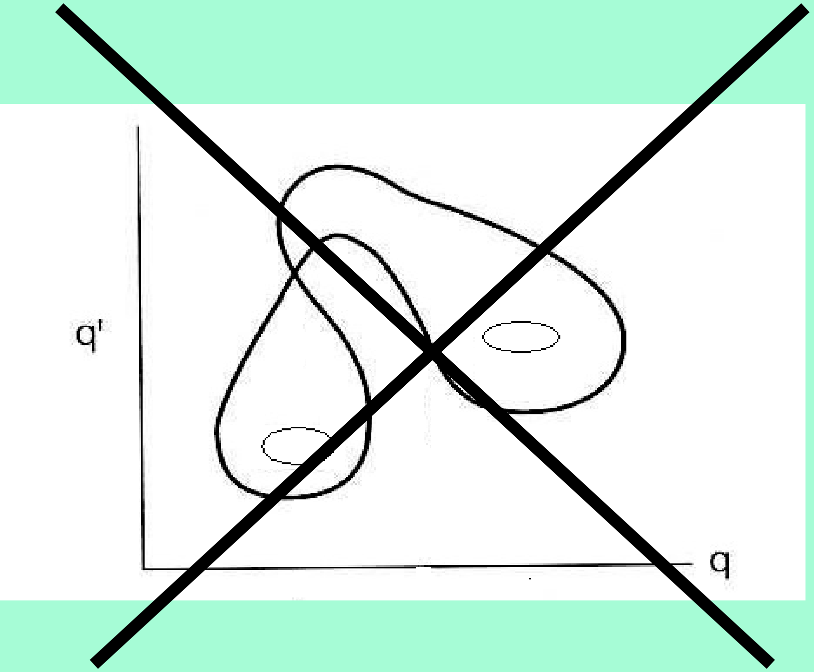
Alanine Dipeptide at 300 K

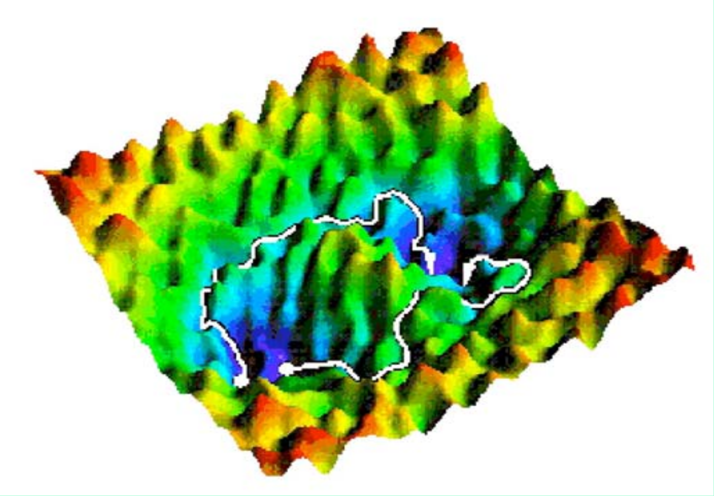
All–atom Amber95: sampling Phi–Psi (20 –> 30 ns)



Life is complicated

The potential energy surface of a complex system is rough





© *D. Chandler*

How to explore a multidimensional free energy surface?

Need to be able to escape free energy minima

Our solution:
Non-Markovian coarse-grained dynamics

A. Laio and M. Parrinello, PNAS

Collective variables

Choose a small set of slow collective variables:

$$s_i = s_i\left(\vec{R}_I\right) \qquad i=1,n$$

The $\mathbf{s_i}$:

- Discriminate between reactants and products
- Include all the relevant slow modes
- The reaction coordinate is a linear combination of the $\mathbf{s_i}$

Examples of collective variables

- Distances
- Angles: bending and torsional
- Coordination numbers: between individual atoms
- or between different species
- Local electric fields
- Number of n-fold rings
- Solvation energy
- Lattice vectors
- Energy
- Etc. etc.

Probability distribution

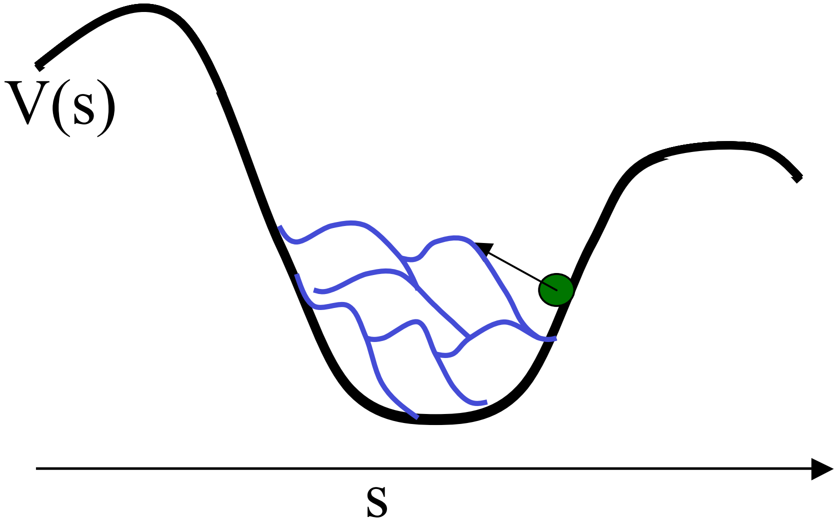
$$P(S_1,S_2\ldots,S_n)=\frac{\int d\vec{R}_I \prod_i^n \delta\left(S_i-s_i\left(\vec{R}_I\right)\right)\; e^{-\beta V\left(\vec{R}_I\right)}}{\int d\vec{R}_I e^{-\beta V\left(\vec{R}_I\right)}}$$

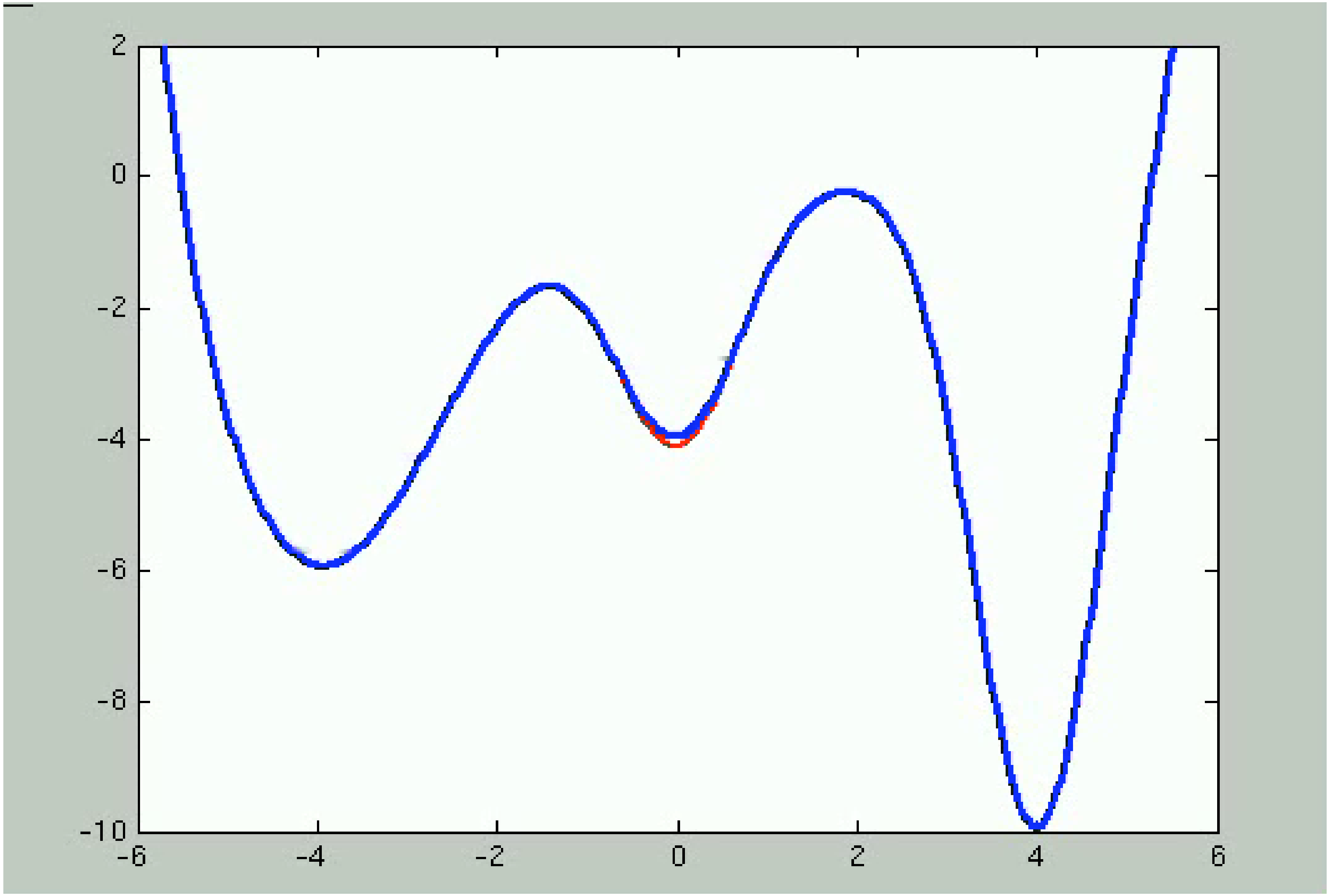
We want to study the free energy as a function of these variables:

$$\beta F(S_1,S_2,\ldots,S_n)=-\ln P(S_1,S_2,\ldots,S_n)$$

The algorithm:

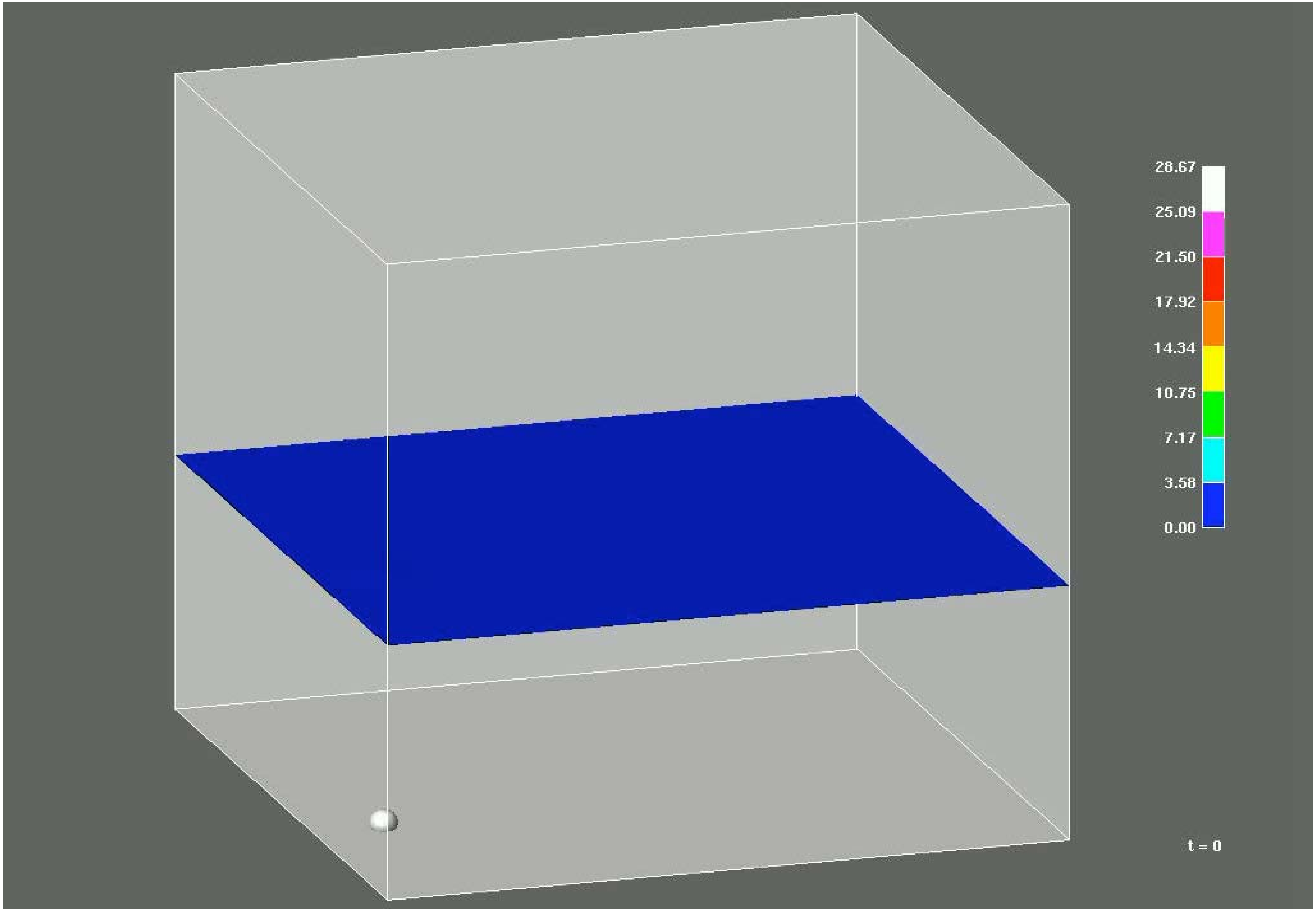
- Wherever you go put a “small” Gaussian
- Always move in the direction of the direction that minimizes the sum of V(s) and all the Gaussians



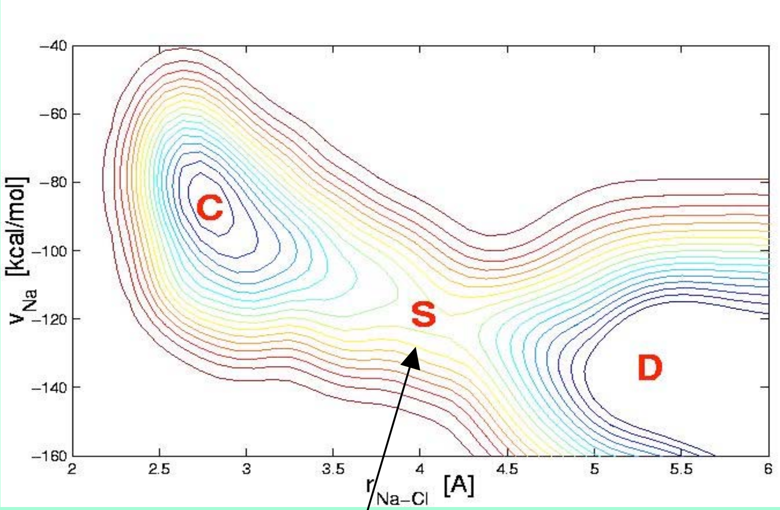

2
0
-2
-4
-6
-8
-10
-6
-4
-2
0
2
4
6

NaCl in water

Two minima:	Contact ion pair (metastable)
	Dissociated
Collective coordinates:	Electric field on Na^+
	Electric field on Cl^-
	Distance Na^+ Cl^-
Classical MD:	Amber force field

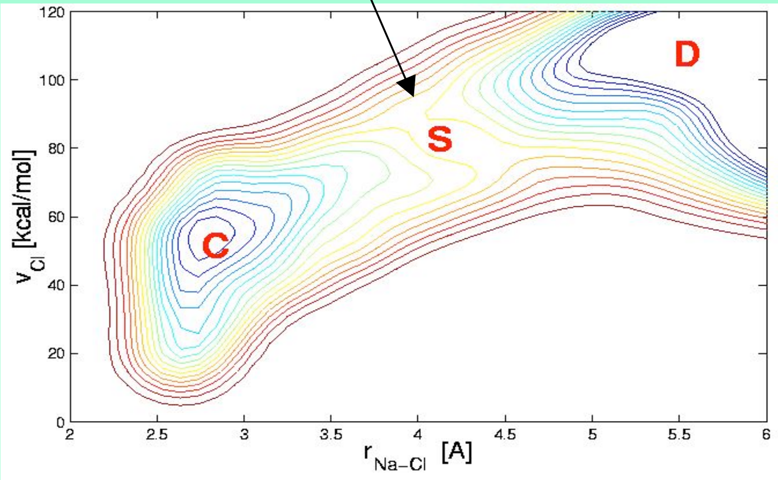

28.67
25.09
21.50
17.92
14.34
10.75
7.17
3.58
0.00
t = 0

Free energy surface



Transition state

Transition state

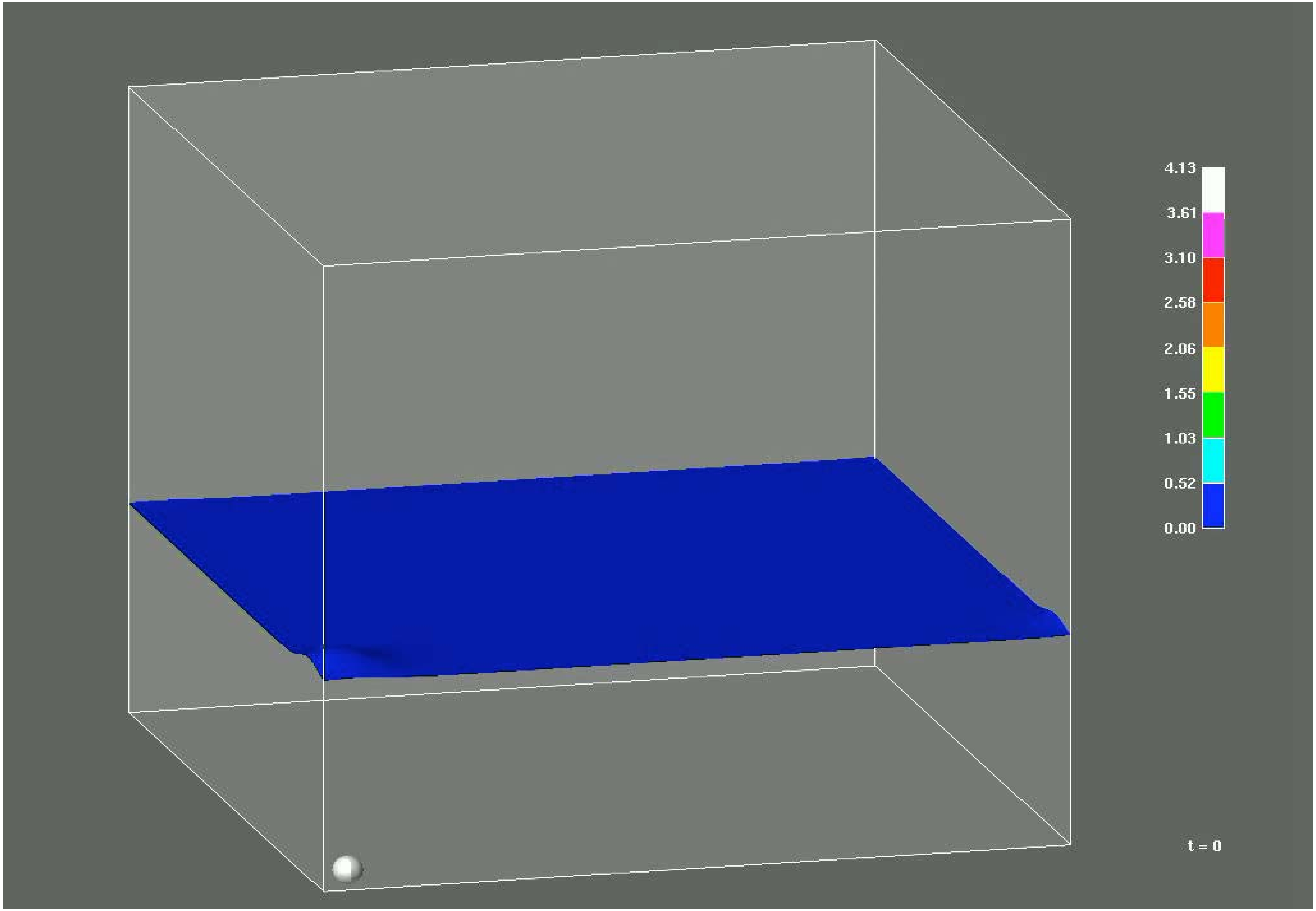


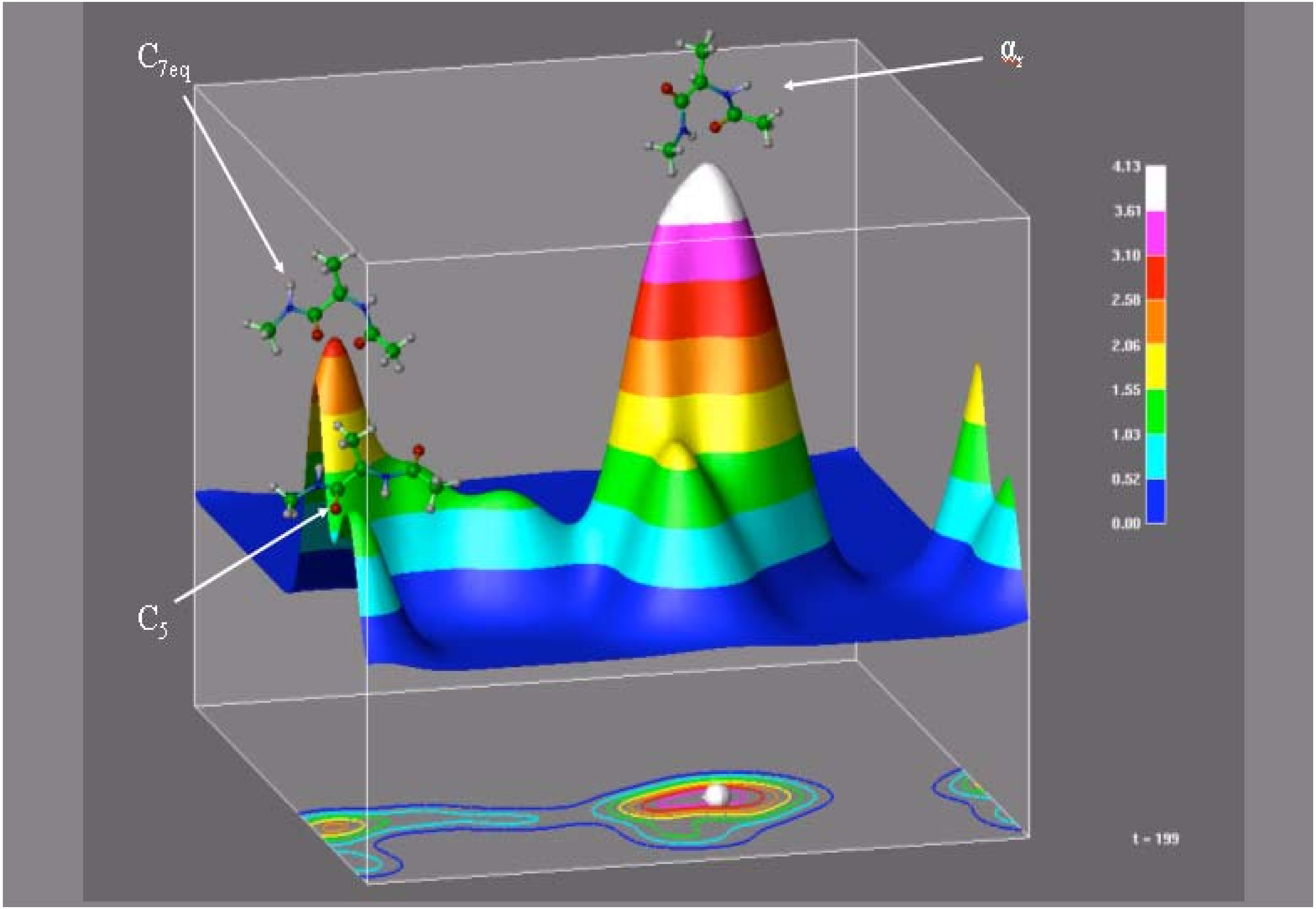
Dialanine in water

1 dialanine in 287 TIP3P water

AMBER95 force field

Collective coordinates: backbone dihedral angles Φ and Ψ


4.13
3.61
3.10
2.58
2.06
1.55
1.03
0.52
0.00
t = 0


C_{7eq}
α_R
C_5
4.13
3.61
3.10
2.58
2.06
1.55
1.03
0.52
0.00
t = 199

Stationary action principle

$$L = \frac{1}{2}\dot{q}^2 - V(q)$$

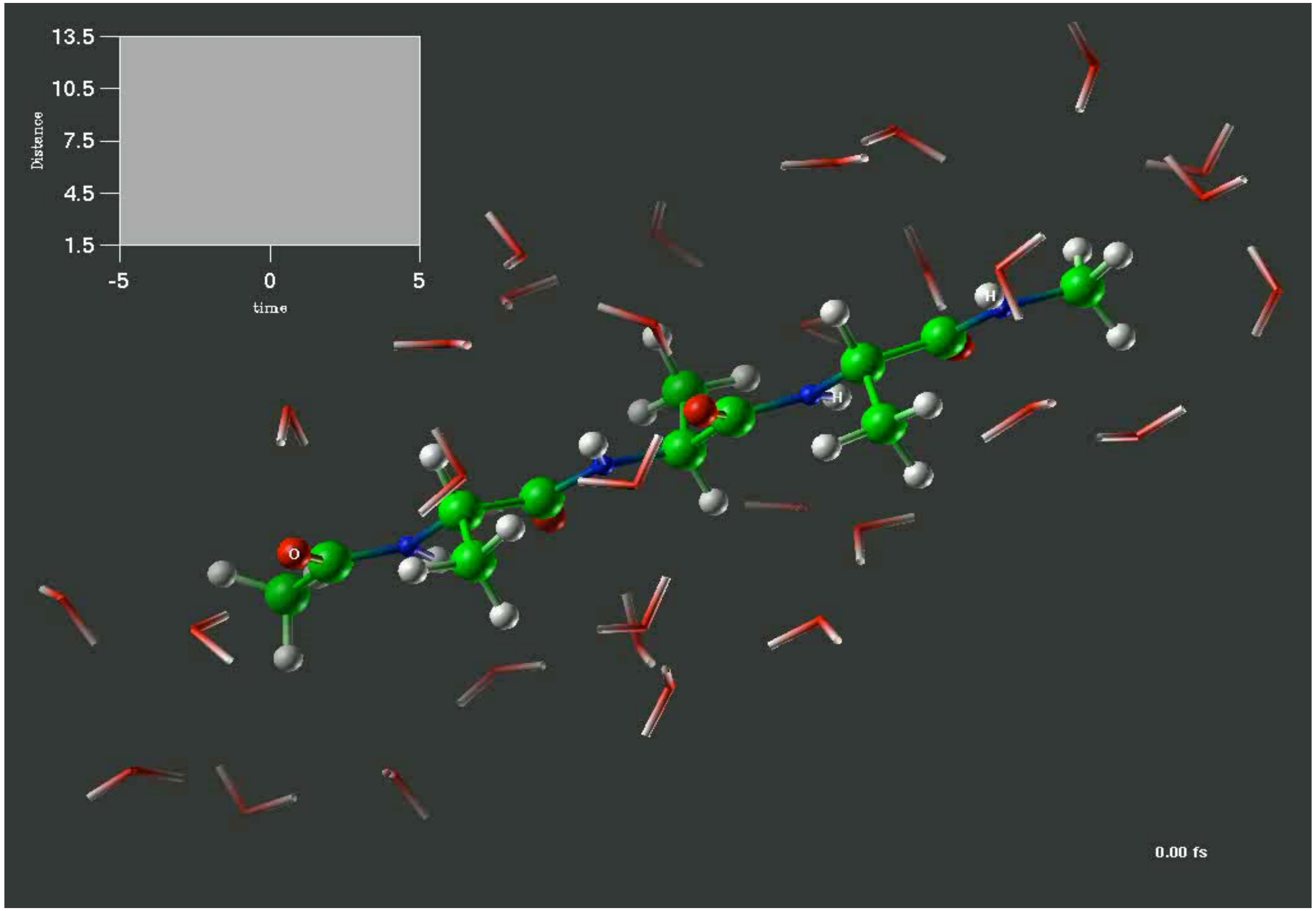
$$\delta S = \delta \int_{t_A}^{t_B} L(q(t), \dot{q}(t))\,\mathrm{d}t = 0$$

$$q(t_A) = q_A \qquad q(t_B) = q_B$$

Can we use this principle to find the trajectories?

Saddle point!

Folding a small peptide



Motivation and aims

Oxidatitive damage to DNA is common and has fatal consequences

Guanine, having the lowest oxidation potential among the DNA bases, plays a fundamental role

Does the structure of DNA funnel the reactions towards a unique product?

A continuous metadynamics

Fictitious kinetic energy

$$\mathcal{L} = \mathcal{L}_{CP} + \sum_{\alpha} \frac{1}{2} M_{\alpha} \dot{s}_{\alpha}^{2} - \sum_{\alpha} \frac{1}{2} k_{\alpha} \left(S_{\alpha}(\mathbf{R}) - s_{\alpha} \right)^{2} - V(\mathbf{s}, t)$$

Restrain potential

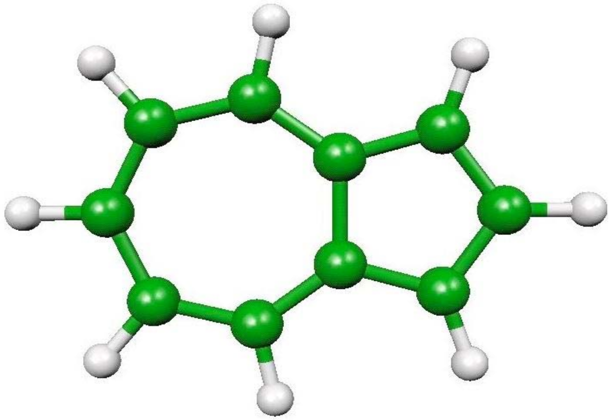
If √(k /M) « _I

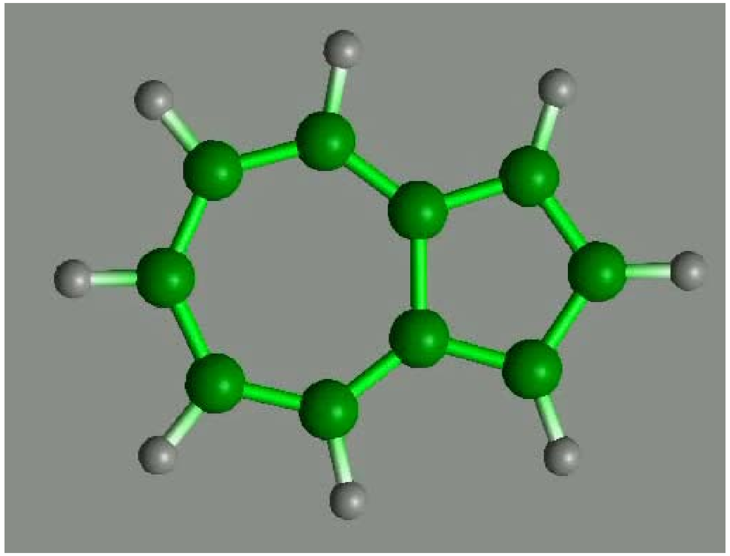
$$\frac{\partial \mathcal{F}}{\partial s_{\alpha}} \simeq \langle k_{\alpha} (S_{\alpha}(\mathbf{R}) - s_{\alpha}) \rangle$$

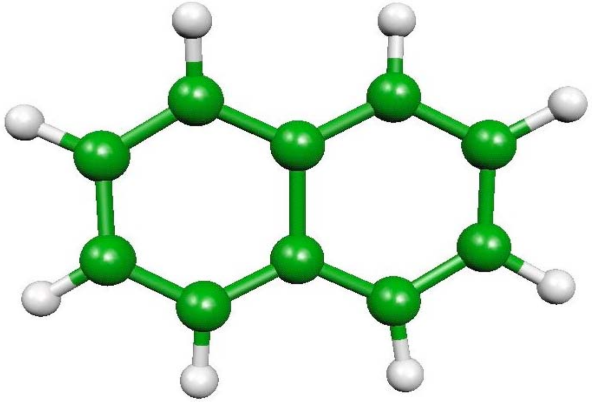
History-dependent potential

M. Iannuzzi, M. Laio and M. Parrinello, PRL 2003

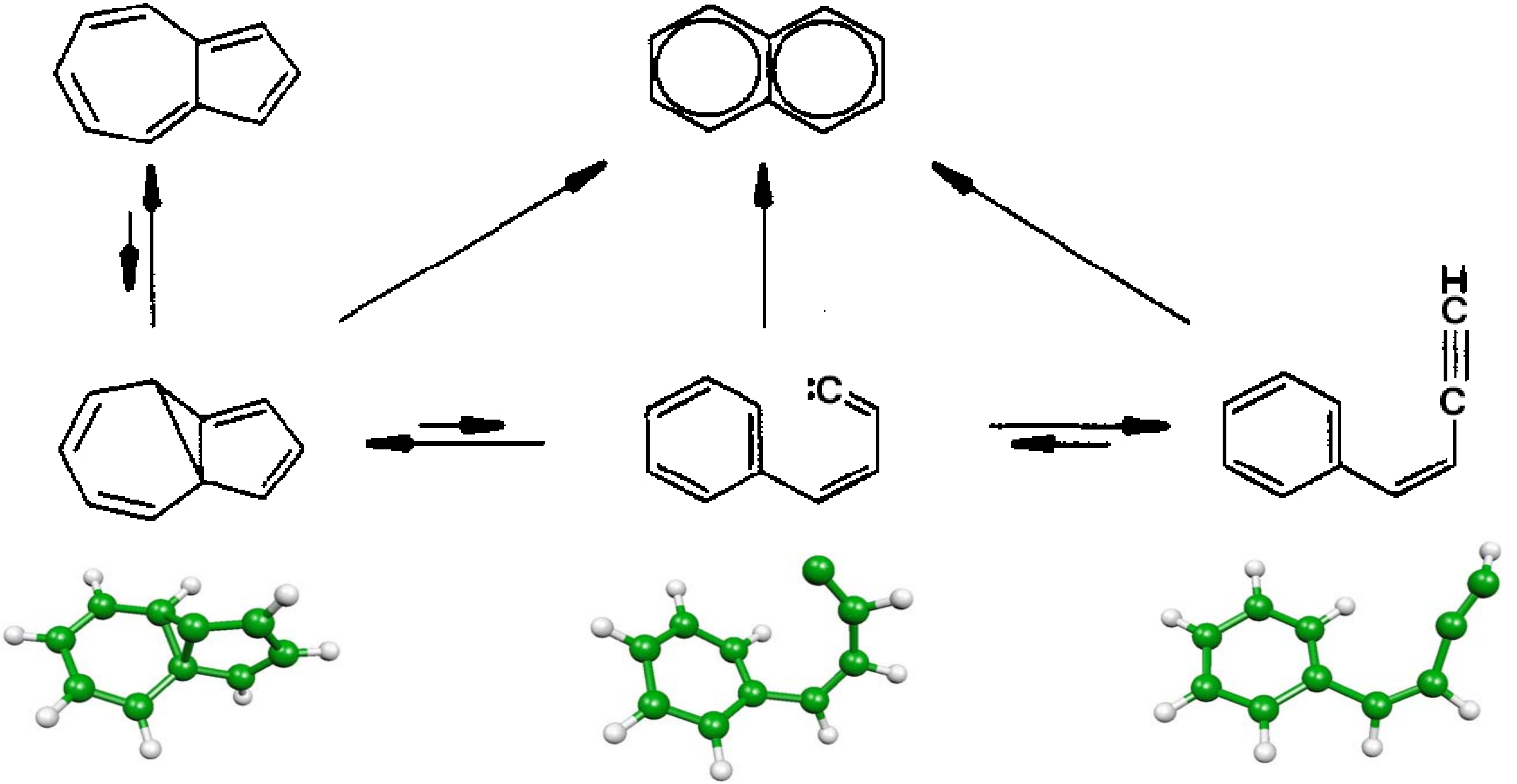
From azulene to naphthalene







Science fiction?

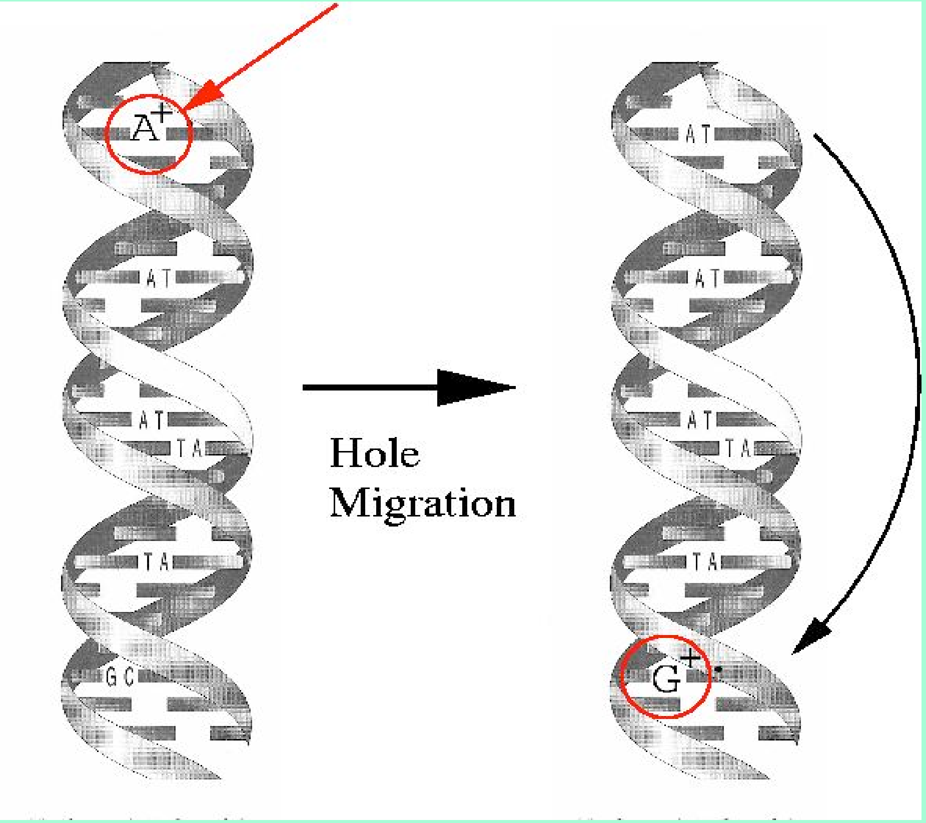


L.T. Scott, Acc. Chem. Res. ***52****, 15 (1982)*

G^+ localization how?

$G^{+\cdot}$ can be formed:

1. Directly
2. Via hole migration



Chem. Rev. ***98****, 1109-1151 (1998)*

Computational details

Model
G:C decamer, Z conformation

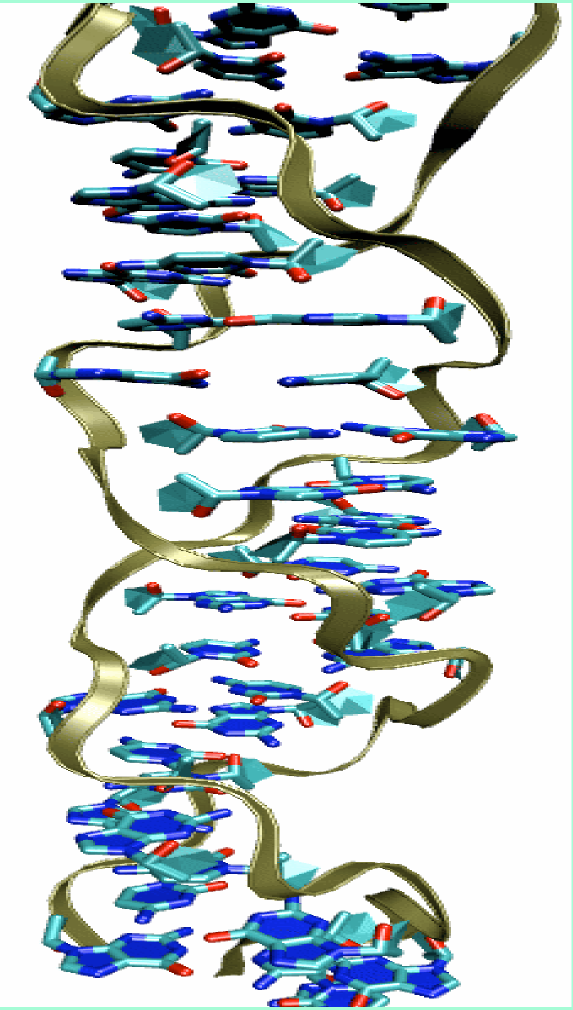
X-ray structure available

Rich in G and the smallest cell Electronic structure treated by the Kohn-Sham method

BLYP and HCTH functionals, plane waves (70Ry cut-off)

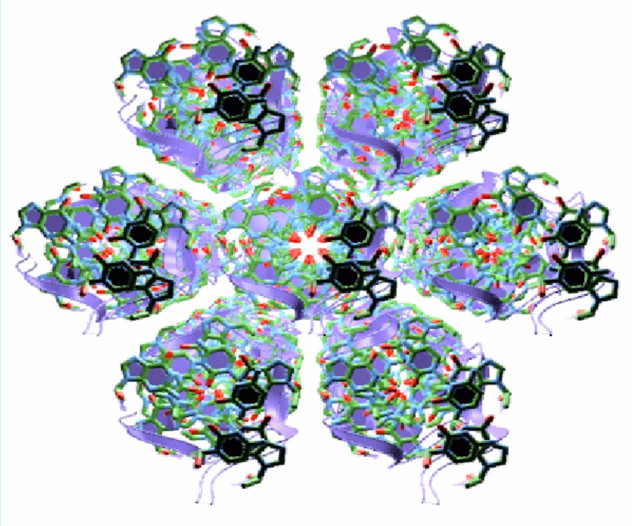
Martin-Troullier pseudopotentials

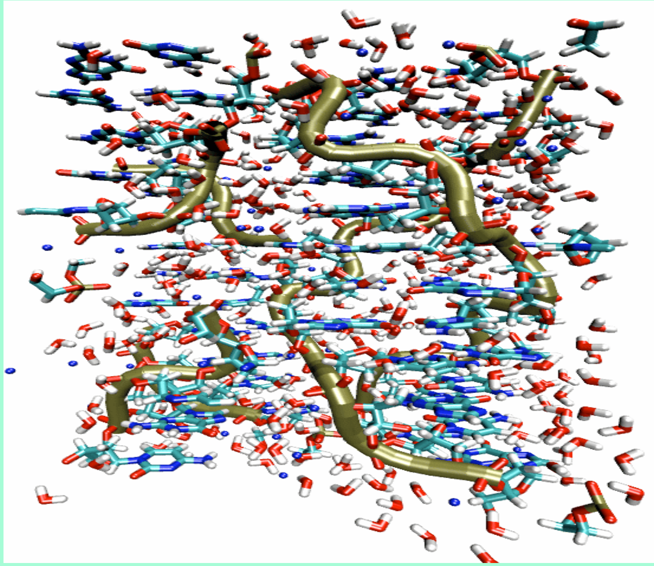
Car-Parrinello molecular dynamics, CPMD code



Computational details

The hole localizes on G$^+$ and the proton moves to C





The model includes water and counter ions

1194 atoms

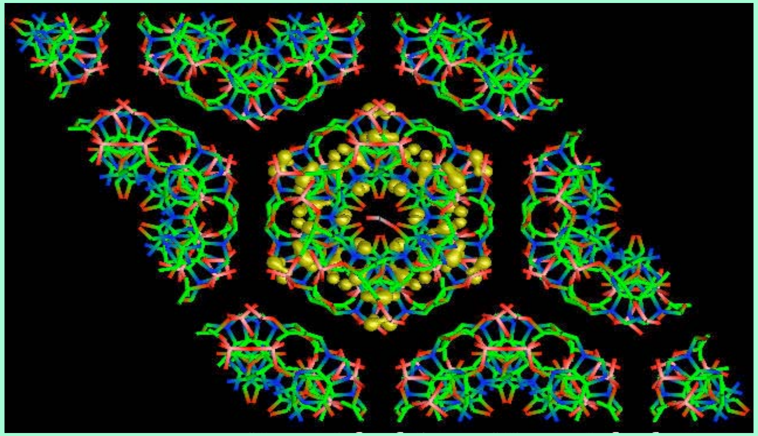
Full quantum: 3,960 valence electrons, 408,238 PW!

The HOMO of DNA

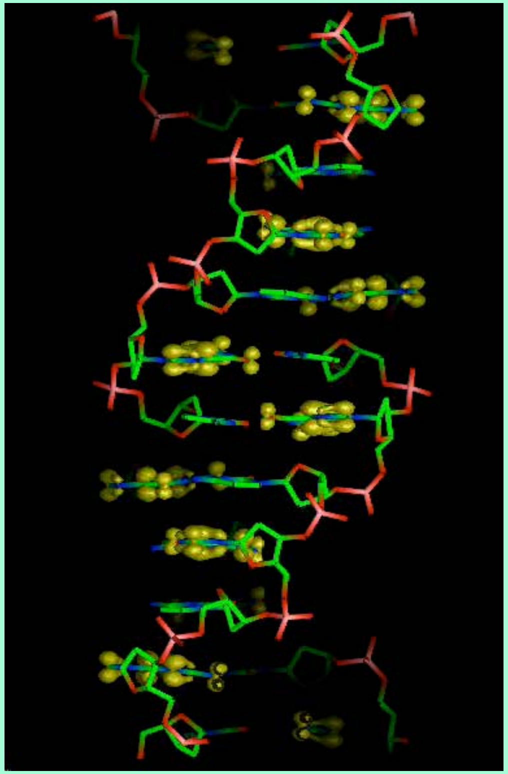
DNA laboratory specimen
The full Monty

1194 atoms 1980 states

408.238 PW 13 Gb



HOMO



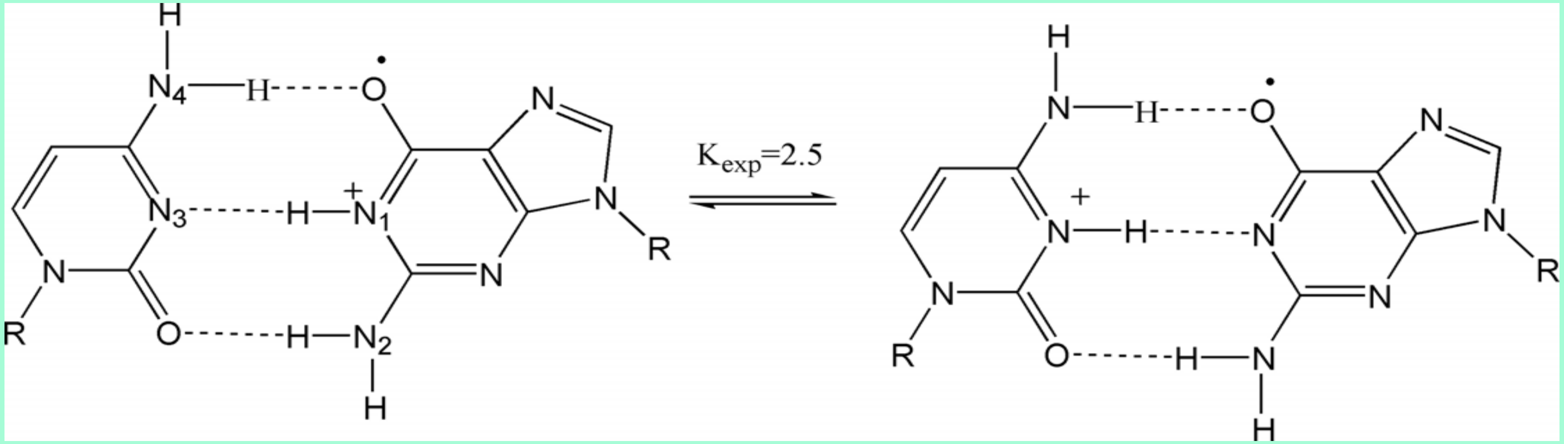
F. Gervasio, P. Carloni and M.P., PRL

Fate of $G^{+\cdot}$ in DNA

In duplex DNA the protonation is not clear

CH^+ has a pKa of 4.3 ($G^{+\cdot}$: pKa 3.9)

_ Sharing of proton expected



$K_{exp} = 2.5$ is predicted from experiments in water

Is this relevant for DNA?

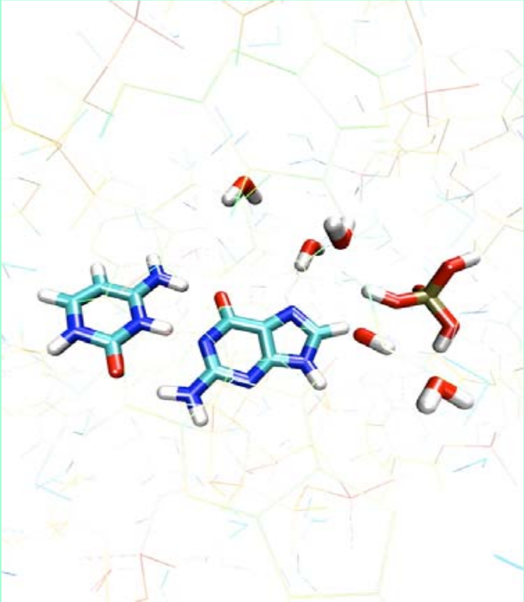
Bridging length scale

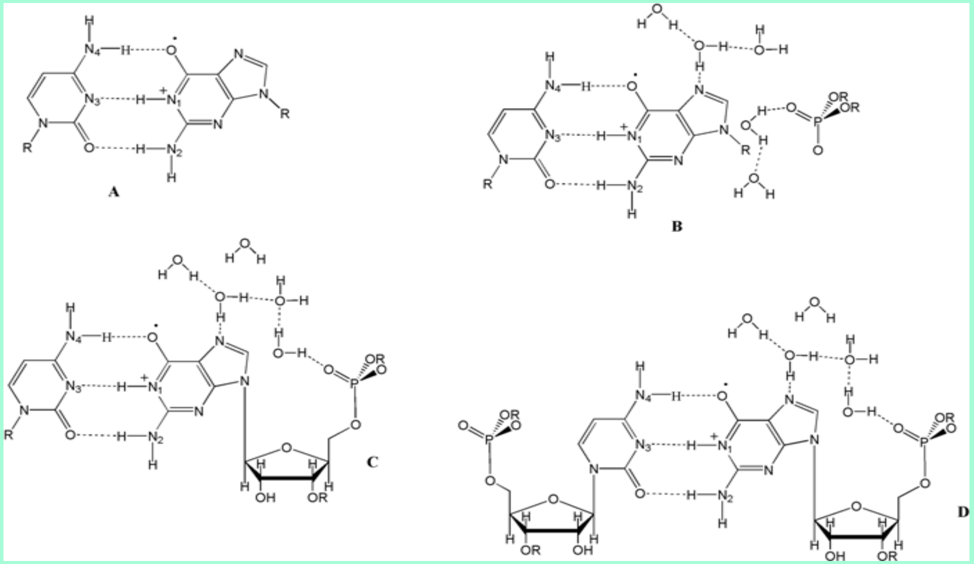
Most calculations were done in the QM/MM framework.

Laio A., VandeVondele J. and Roethlisberger U., J.Chem. Phys. ***116****, 6941(2002)*

Due to high polarity of bonds H-capping was used.

Added H were decoupled to MM.





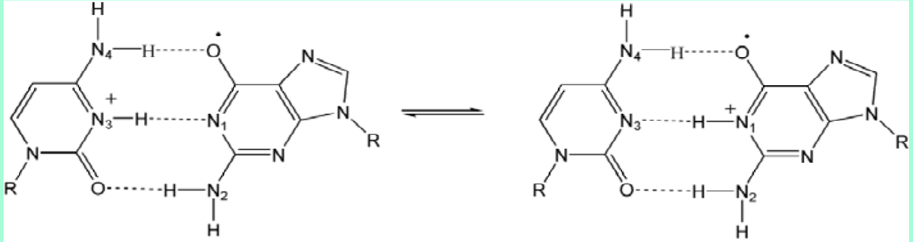
Quantum sub-systems of increasing size were used.

The biggest quantum model was the full DNA.

Protonation state of the bases: gas phase

Collective coordinate used: coordination number (C_{AB})

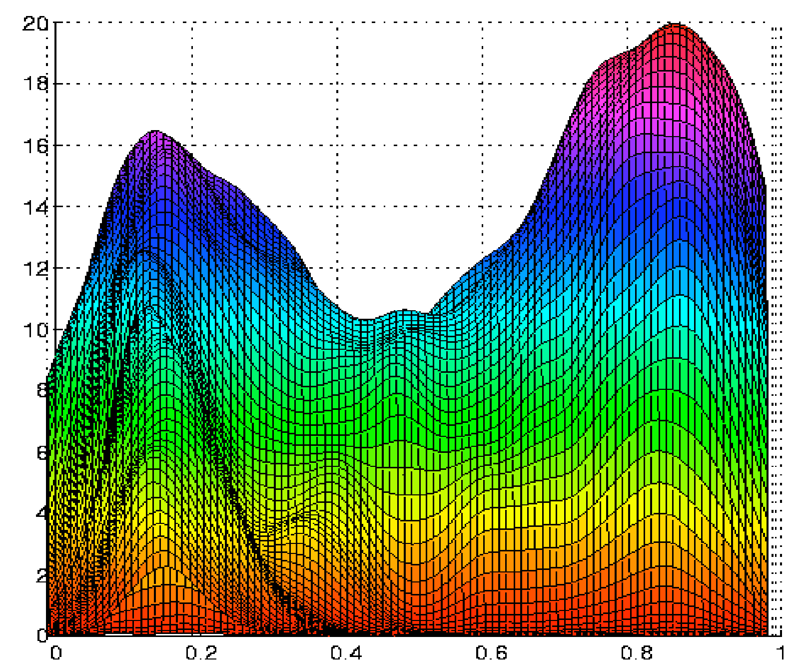
$$C_{AB} = \sum_{i=1}^{N_B} \frac{1-\left(\frac{r_{Ai}}{R_{AB}}\right)^n}{1-\left(\frac{r_{Ai}}{R_{AB}}\right)^m}$$



n=6, m=12

of N1, N2 and N1, N4

with respect to H



Gas Phase ΔE:

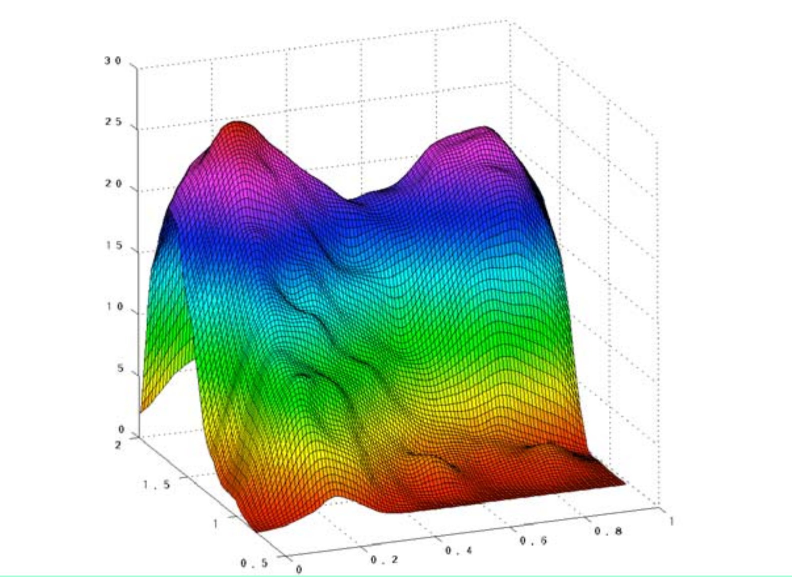
BLYP 1.9 kcal/mol

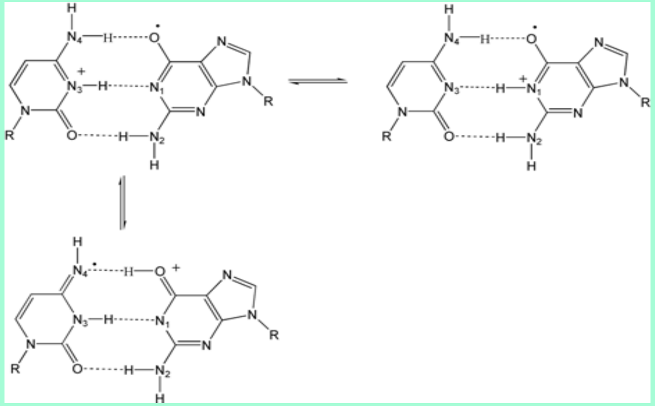
HCTH 2.03 kcal/mol

B3LYP/D95* 1.6 kcal/mol

Hutter and Clark, J.Am.Chem.Soc. ***118****, 7574*

Protonation state of the bases: DNA





DNA ΔE(kcal/mol):

Gas phase:	~2./3. (inversion)
QM/MM	–4.5
No charge	–1.2
Quantum PO_4^{3-}/full	–3.7 /–3.9
Charge on PO_4^{3-} and sugar	–5.3

ΔΔE ~ 7-8

Sources

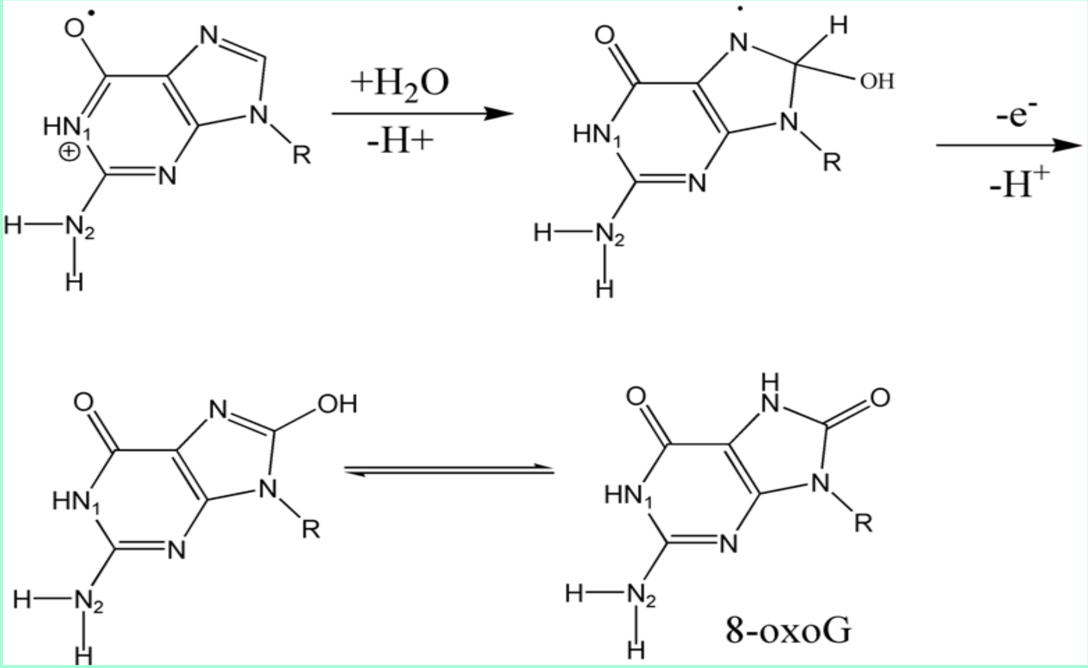
Backbone charges	~3
Error (QM/MM, exclusion)	<1
Geometry	3-4

Fate of $G^{+\cdot}$ in DNA

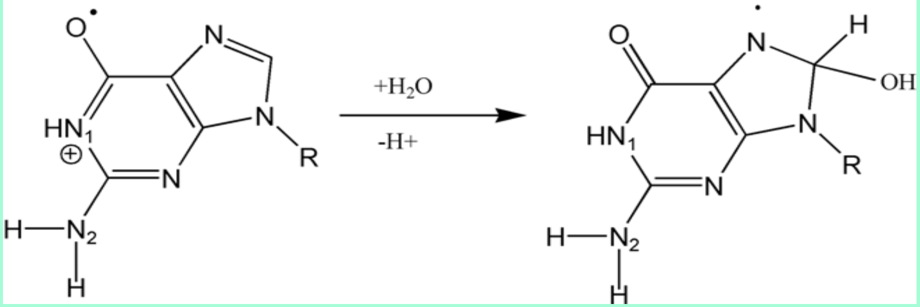
In duplex DNA the oxidation product is 8-oxo-guanine (8-oxoG)

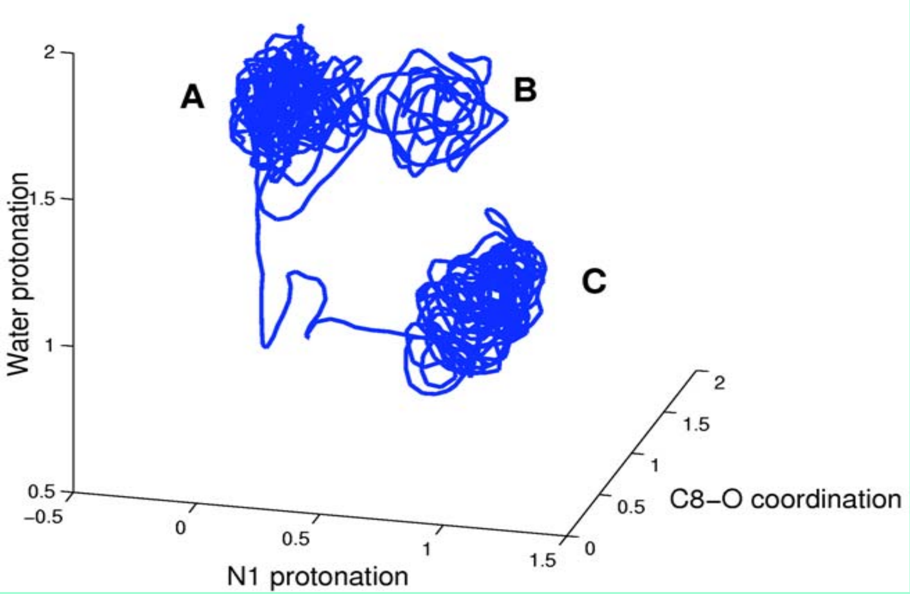
30,000 8-oxoG in human genome!
Methods Enzymol. ***186,*** *521 (1990)*

In water a variety of products are formed



Oxidation reaction 1





Meta coordinates used:
N1 and water oxygen H-coordination number
Carbon 8 O-coordination number

Rate limited by water autoprotolysis

Catalyzed by $R_2PO_4^-$

N1 protonation state matters

ΔE N1 deprotonation of 8-OH-G:
22 kcal/mol!

Oxidation reaction