Metadynamics & *CPMD*

Petr Kulhánek

kulhanek@chemi.muni.cz

National Centre for Biomolecular Research, Masaryk University
Faculty of Science, Kotlářská 2
CZ-61137 Brno

Contents

- **Free Energy Calculations**
  - Sampling Problem
  - Potential of Mean Force Methods
    - Metadynamics
    - Constrained Dynamics
    - Adaptive Biasing Force
    - Multiple Walkers Approach
- **Molecular Dynamics and Reactions**
  - Empirical Valence Bond
  - CPMD versus BOMD

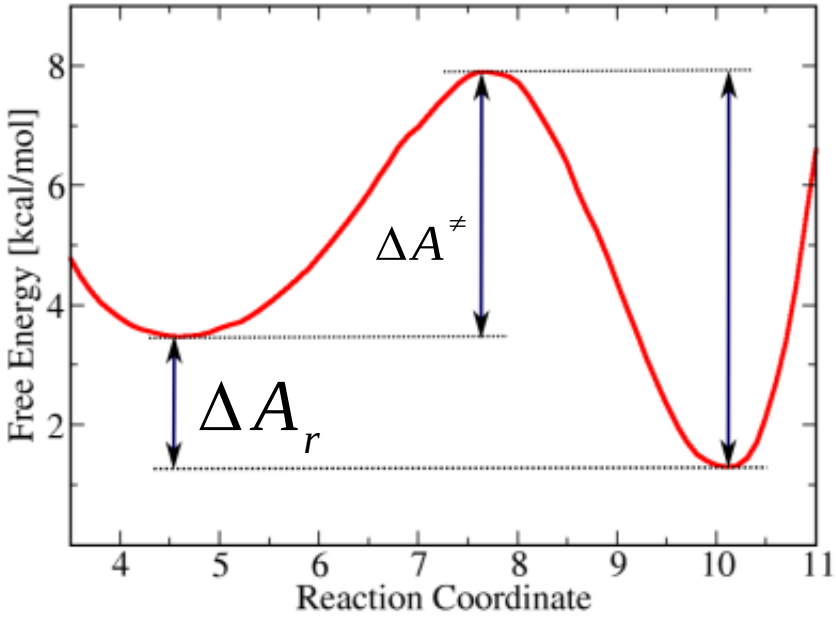
Free Energy Calculations

Free Energy

Free energy is related to **equilibrium** and **rate** constants.

$$\Delta A_r = -RT \ln K$$

$$k_1 = \kappa \frac{k_B T}{h} e^{-\frac{\Delta A^{\neq}}{RT}}$$



The free energy forms a bridge between a theory and an experiment.

Knowledge of **free energy** allows to quantify:

- chemical reactivity (e.g. enzymatic activities)
- thermodynamics (e.g. binding affinities)

Free Energy Calculations

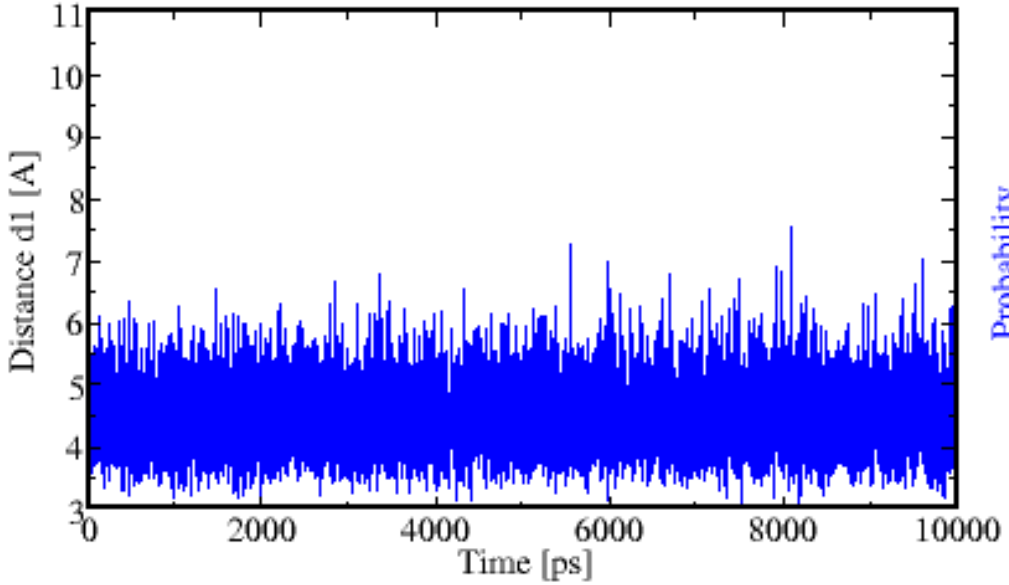
$$\Delta A_{1\to 2} = -RT \ln \frac{\sigma_2}{\sigma_1}$$

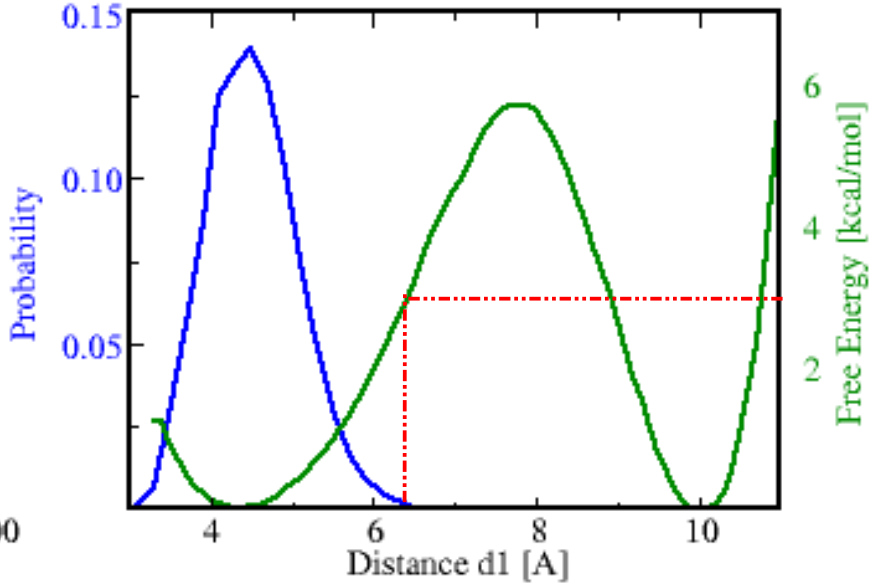
Density of state (probability)
It can be calculated from molecular dynamics or Monte Carlo simulations, **but …**

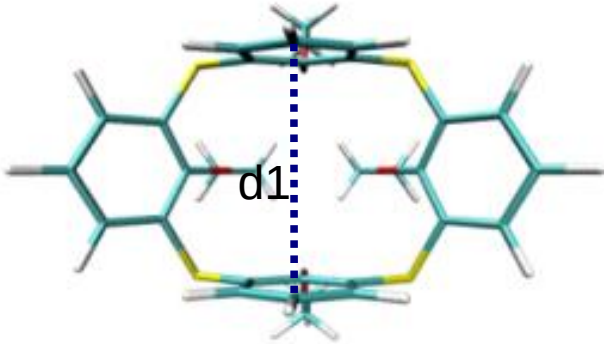
$$A(\xi) = -RT \ln \sigma(\xi) + A_0$$

reaction coordinate
collective variables

Sampling Problem







10 ns long simulation is able to discover free energy landscape with depth only about **3 kcal/mol**.

Free Energy Calculations

A system has to be **biased** achieving efficient sampling in the region of interest. We need to know how to obtain the **unbiased free energy** from such biased simulation.

Available methods:

- ***constrained dynamics***

  system is biased by constraining reaction coordinate

- ***adaptive biasing force***

  system is biased by force which is opposite to potential of mean force

- ***umbrella sampling***

  system is biased by restraining reaction coordinate

- ***metadynamics***

  system is biased by Gaussian hills, which fill free energy landscape

Free Energy Calculations

➢ Alchemical Transformation

one system is slowly changed to another one (changes are very often unrealistic, atoms are created and/or annihilated)

what: mostly *changes* in binding free energies:

how: thermodynamic integration (TI), free energy perturbation (FEP)

➢ Potential of Mean Force

system is changed along reaction coordinate

what: free energy of conformation changes, **reaction free energies**

how: constrained dynamics, adaptive biasing force, umbrella sampling, metadynamics, steered dynamics

➢ *End-points Methods*

free energy of every state is calculated independently

what: mostly binding free energies

how: MM/XXSA; XX=PB, GB, LRA

Metadynamics

Implemented in CPMD

Metadynamics, theory

Free energy landscape is filled by Gaussian hills.

Equations of motion

$$m_i \frac{d^2 x_i(t)}{dt^2} = -\frac{\partial V(x)}{\partial x_i}$$

MTD history potential

$$V_h(s,i) = \sum_{t=1}^{i} H_t \exp\left(-\frac{(s-s_t)^2}{2\sigma^2}\right)$$

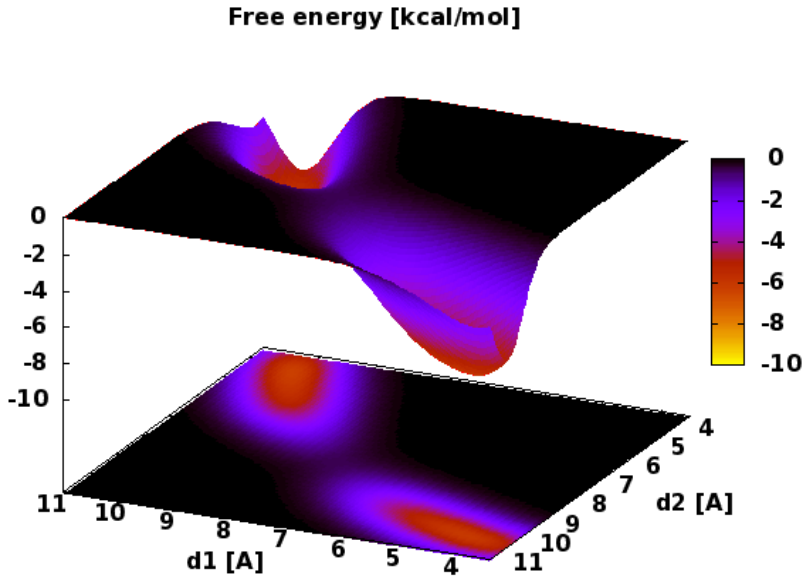
Equations of MTD motion (direct approach)

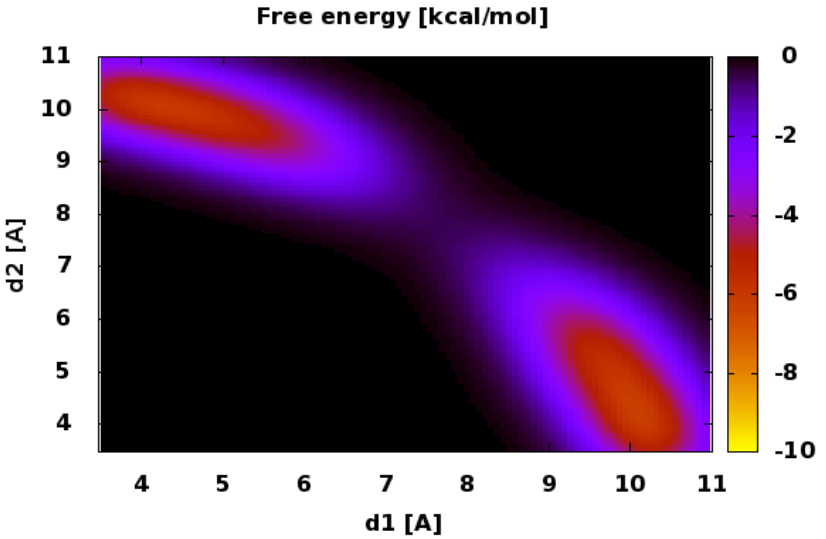
$$m_i \frac{d^2 x_i(t)}{dt^2} = -\frac{\partial}{\partial x_i}\{V(x) + V_h(x,i)\}$$

History-dependent term converges to FES

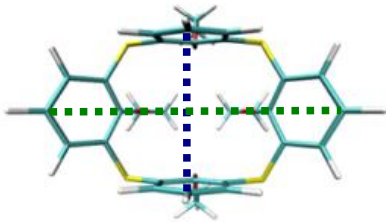
$$A(s) = \lim_{i \to \infty} -V_h(i,s)$$

Metadynamics, example



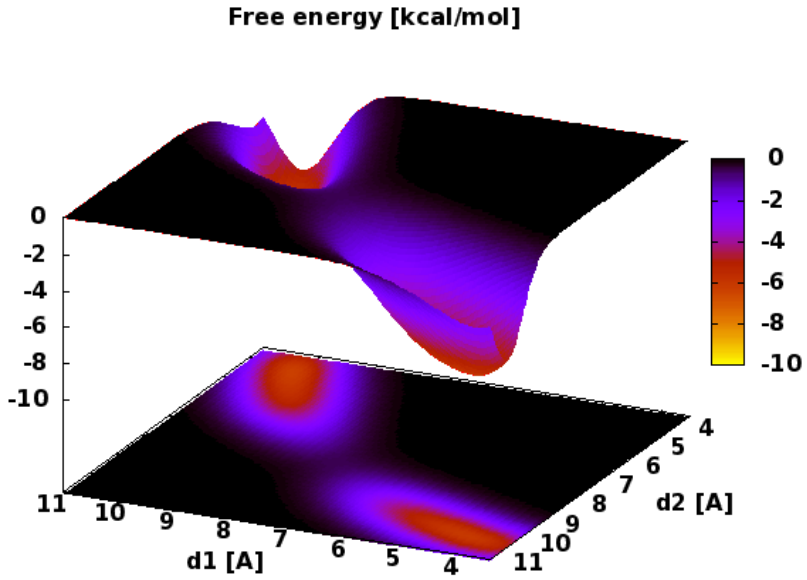


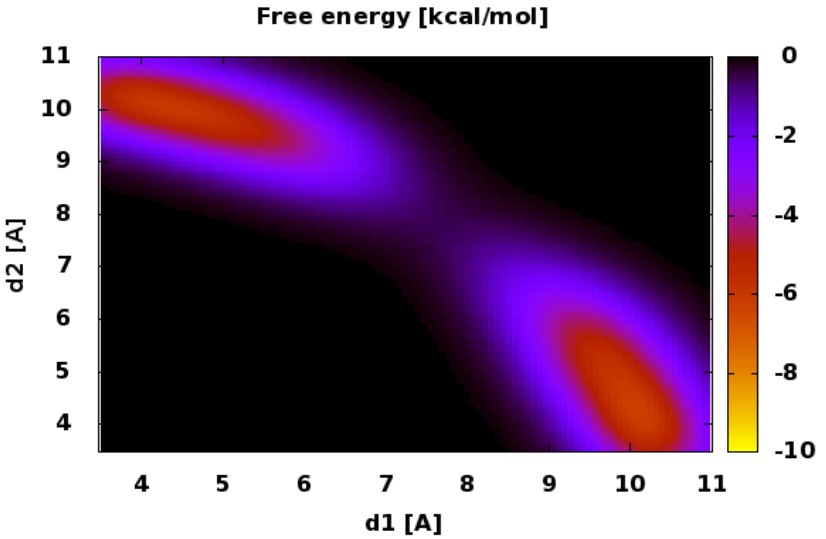
hill height 0.01 kcal/mol, width 0.5 x 0.5 Å
MTD frequency 500 fs
2 ns long simulation
300 K, vacuum, GAFF force field, time step 0.5 fs



DIS (distance)

Metadynamics, example

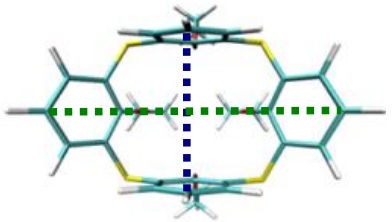




hill height 0.01 kcal/mol, width 0.5 x 0.5 Å
MTD frequency 500 fs

2 ns long simulation

300 K, vacuum, GAFF force field, time step 0.5 fs



DIS (distance)

Constrained Dynamics

Implemented in CPMD

Constrained Dynamics, theory

Reaction coordinate is fixed (constrained) at the value of interest.

Equations of motion

$$m_i \frac{d^2 x_i(t)}{dt^2} = -\frac{\partial V(x)}{\partial x_i}$$

Constraint condition

holonomic constraint

$$\sigma(x) = \xi(x) - \xi_0 = 0$$

method of Lagrange multipliers

Equations of constrained motion

$$m_i \frac{d^2 x_i(t)}{dt^2} = -\frac{\partial}{\partial x_i}\left\{ V(x) + \sum_k \lambda_k(t)\sigma_k(x) \right\}$$

λ_k = Lagrange multipliers

Constrained Dynamics, theory

Derivative of **unbiased free energy** is also given by (concise formulation):

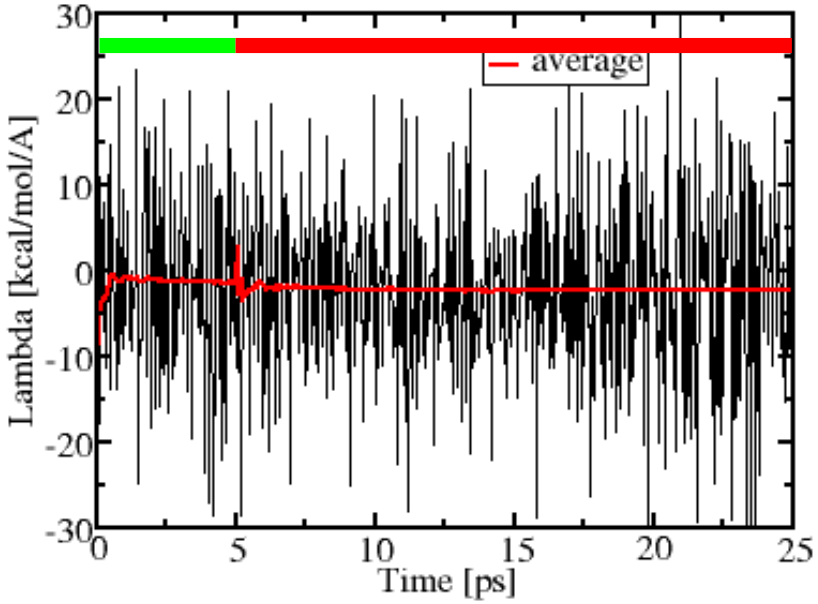
$$\frac{dA^{uc}(\xi)}{d\xi} = \frac{dA^{c}(\xi)}{d\xi} + \frac{dA^{c\to uc}(\xi)}{d\xi} = \left\langle -\lambda \right\rangle_{\xi_0} - RT\,\frac{d}{d\xi}\ln\left\langle Z^{-1/2}\right\rangle_{\xi_0}$$

second derivatives are not required

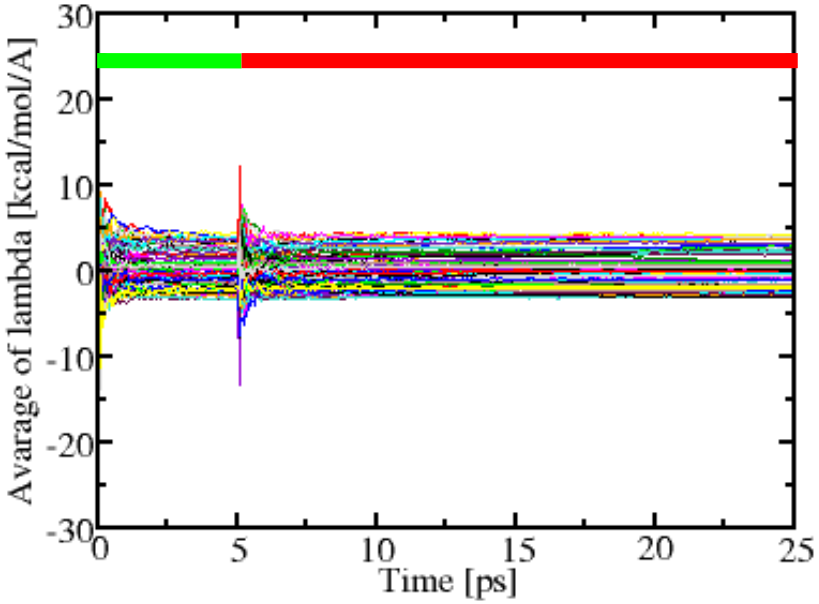
Final free energy is obtained by numerical **integration**:

$$\Delta G = \int_{\xi_1}^{\xi_2} \frac{dA^{uc}(\xi)}{d\xi} d\xi$$

Constrained Dynamics, example

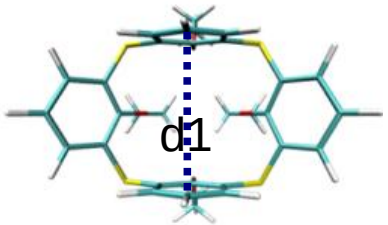


70 points



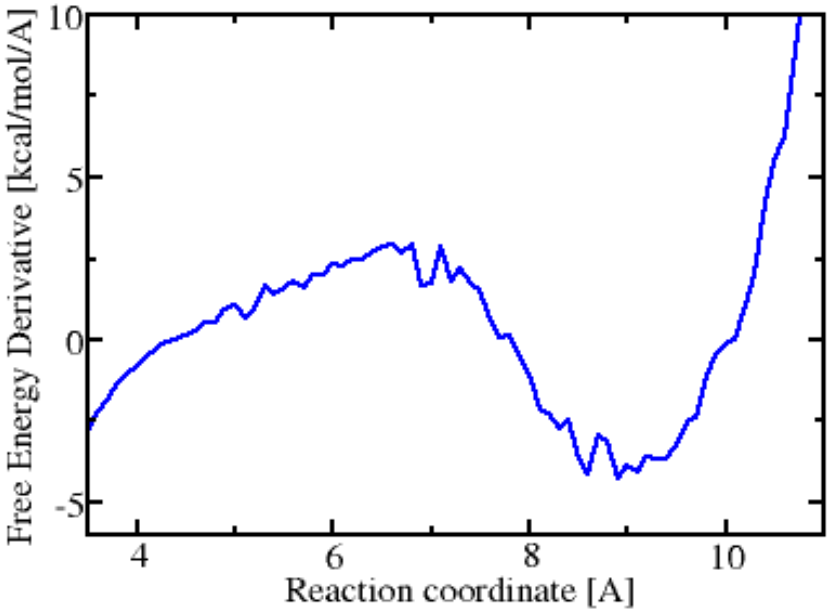
Small numbers are calculated by averaging big numbers.

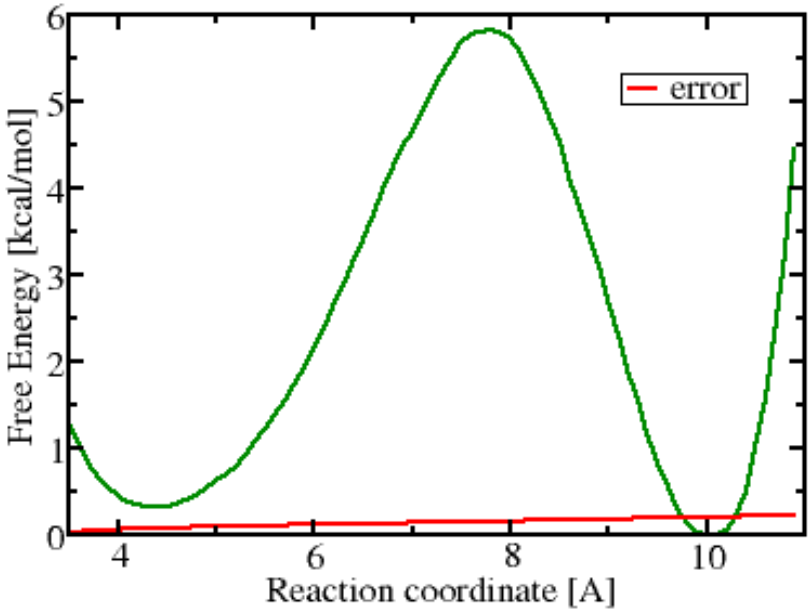
77 points, $\Delta\xi$ 0.1 Å
Method B: 5 ps shift, 5 ps equilibration, 20 ps production
300 K, vacuum, GAFF force field, time step 1 fs / 0.5 fs



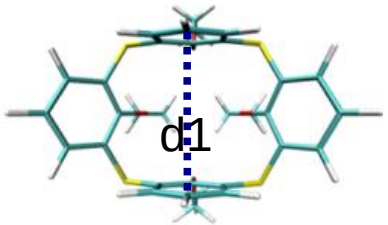
DIS (distance)

Constrained Dynamics, example



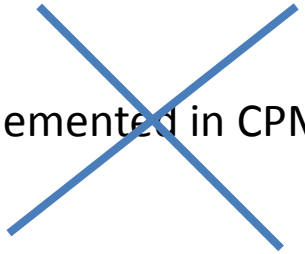


77 points, $\Delta\xi$ 0.1 Å
Method B: 5 ps shift, 5 ps equilibration, 20 ps production
300 K, vacuum, GAFF force field, time step 1 fs / 0.5 fs



DIS (distance)

Adaptive Biasing Force



ABF, theory

Movement along reaction coordinate is the subject of diffusion process.

Equations of motion

$$m_i \frac{d^2 x_i(t)}{dt^2} = -\frac{\partial V(x)}{\partial x_i}$$

Free energy and force along RC

$$F_{ABF}(x) = -\frac{dA(\xi)}{d\xi}\frac{d\xi}{dx}$$

force along reaction coordinate is subtracted from the system

Equations of ABF motion

$$m_i \frac{d^2 x_i(t)}{dt^2} = -\frac{\partial V(x)}{\partial x_i} + F_{ABF}(\xi)$$

ABF, theory

Free energy is given by:

$$\frac{\partial A}{\partial \xi} = -\left\langle \frac{d}{dt}\left(\frac{1}{Z_\xi} \frac{d\xi}{dt} \right) \right\rangle_\xi$$

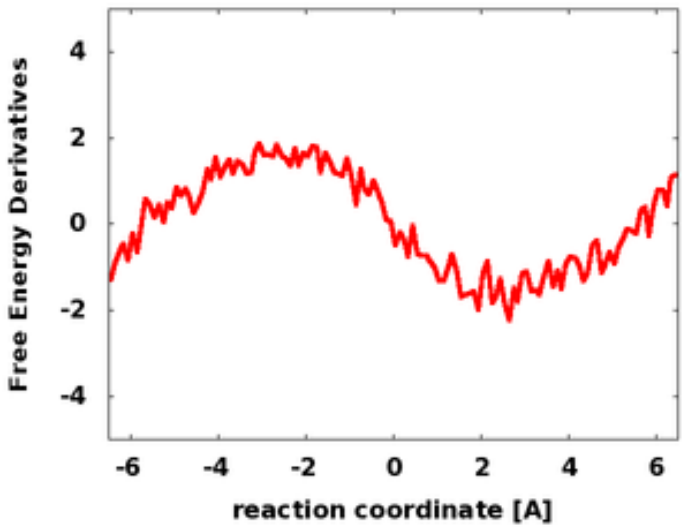
it contains the second derivatives of reaction coordinate if treated analytically

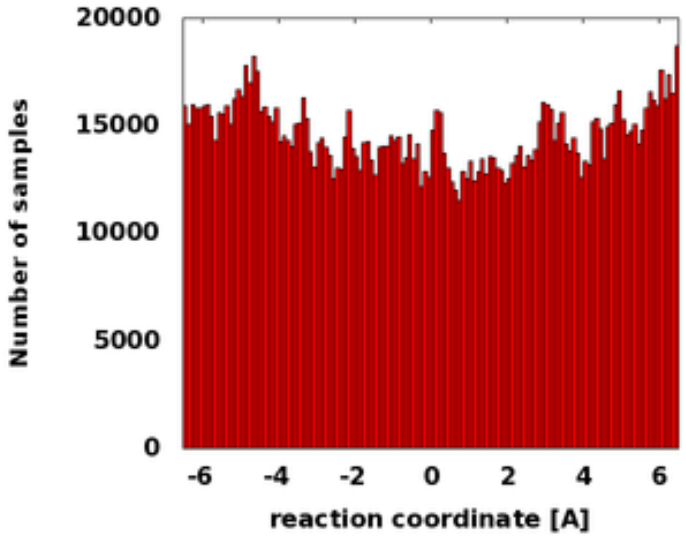
equation is solved **numerically**

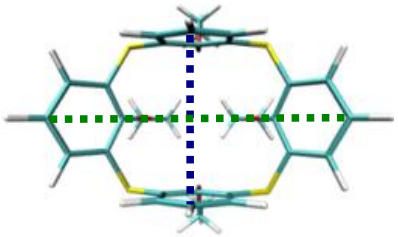
Procedure:

- range of reaction coordinate is divided into bins
- a value of reaction coordinate determine a bin
  - a contribution to the derivative of free energy is accumulated into a bin
  - ABF force calculated from accumulated free energy derivative is applied to the system
- accumulated free energy derivative very rapidly converges

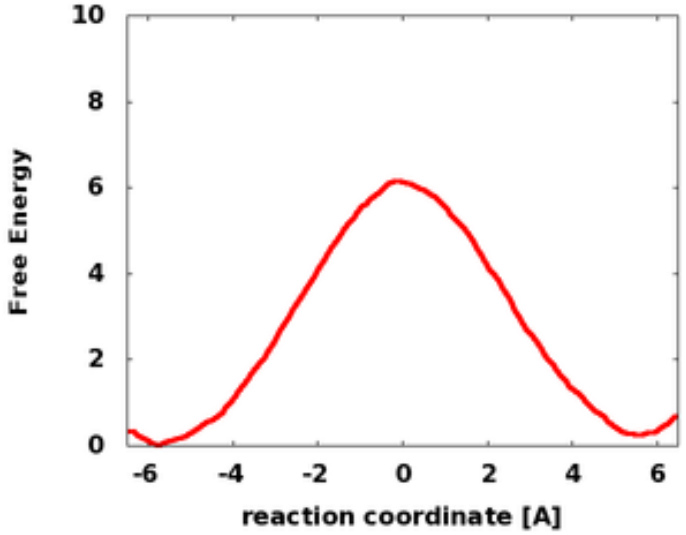
ABF, example







DD
(difference of distances)

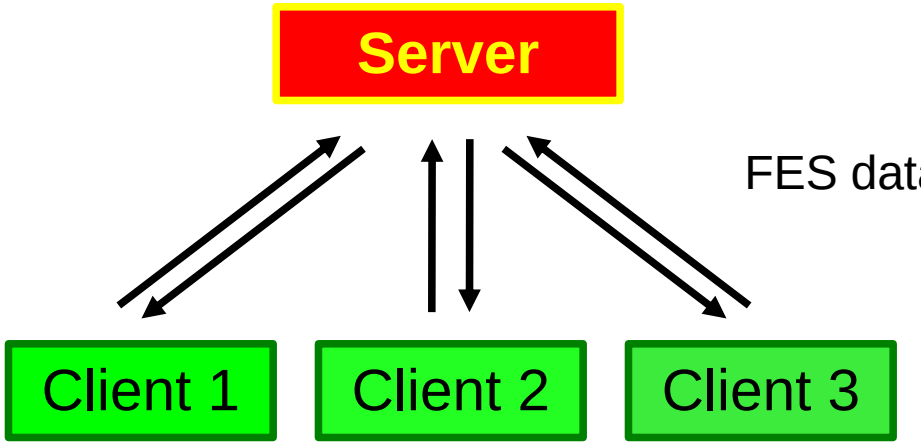


Show movie?

ξ range from -6.5 to 6.5 Å, 130 bins
2 ns long simulation
300 K, vacuum, GAFF force field, time step 1 fs

Multiple Walkers Approach

server collects information about free energy surface (FES) and redistributes it among clients



FES data are exchanged every $\mathbf{N_{update}}$ MD steps

N_{update} ~ 500-1000 steps

Applicable to:

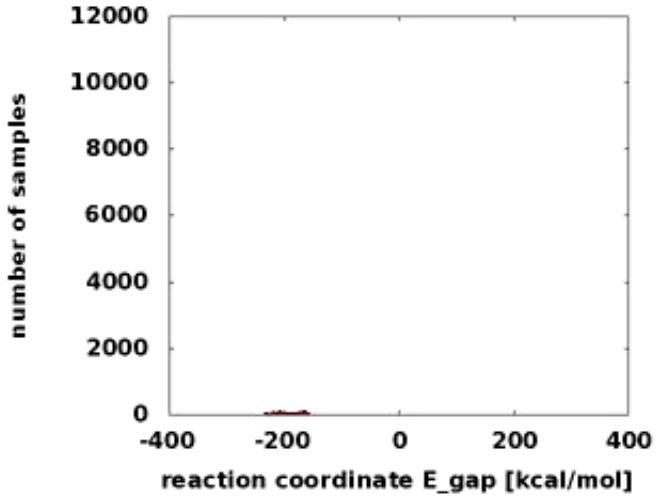
- Metadynamics
- Adaptive biasing force

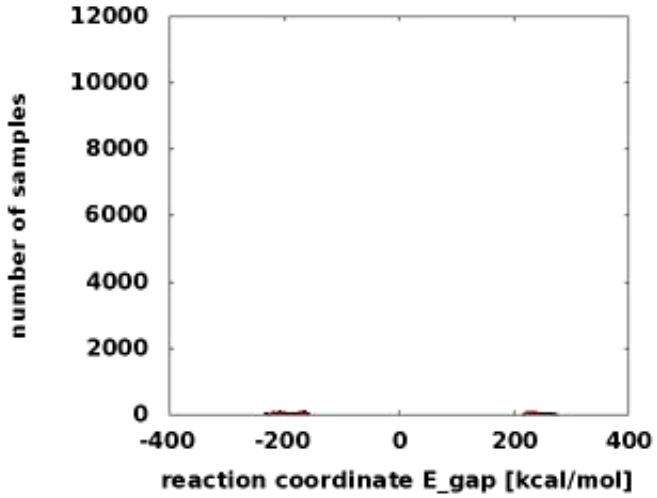
Advantages:

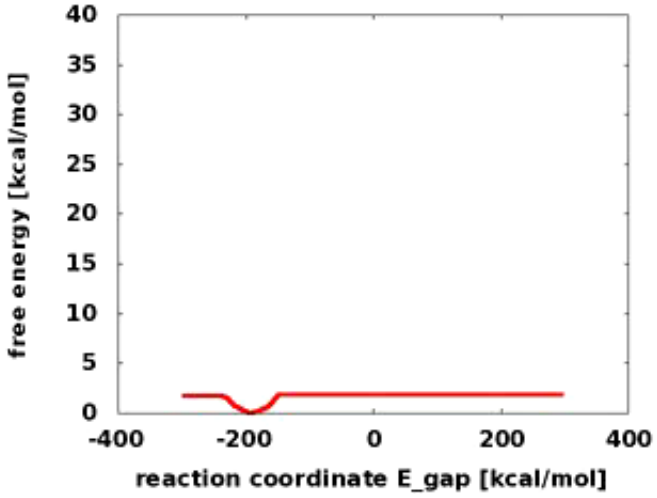
- Faster convergence
- Easy to implement
- Parallel scaling is almost linear

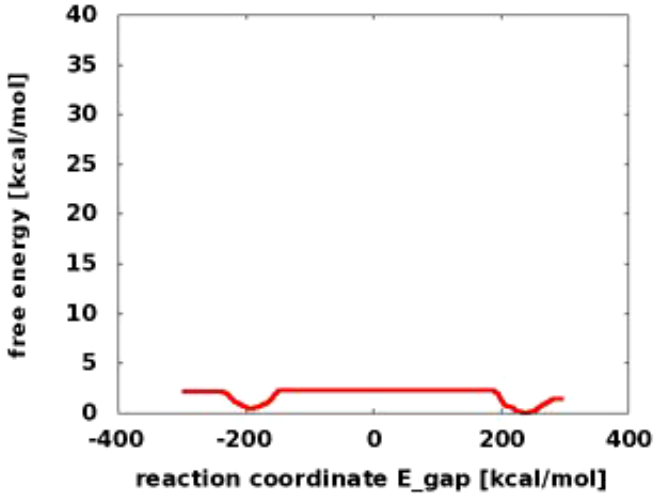
Multiple Walkers Approach

Nucleophilic substitution reaction (test case)









One walker
(from reactants)

Two walkers
(from reactants and products)

PMFLib

A Toolkit for Free Energy Calculations

developed by
Petr Kulhánek

PMFLib, functionality

Implemented methods:

- Constrained dynamics (BM)
- Adaptive biasing force (ABF)
- Metadynamics (MTD)
- Umbrella sampling
- Multiple walker MTD
- Multiple walker ABF
- Replica Exchange Dynamics
- String Method

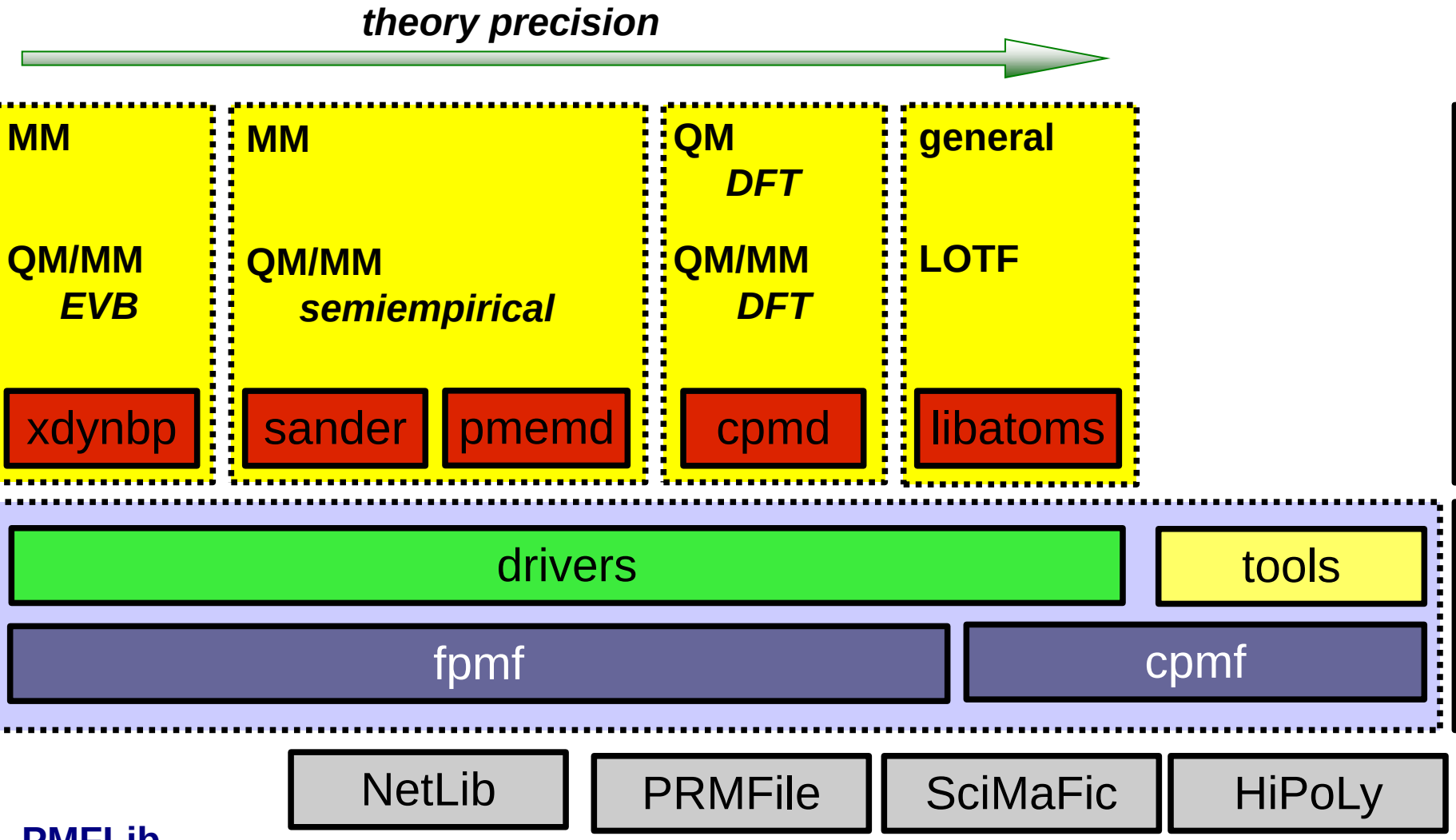
Reaction coordinates:

- DIS, DS
- DD, DDS
- ODISM, ODSM
- ANG, ANGM, CANG, CANGM
- DIH, DIHM
- AC, GC
- RGYR
- RMSDT, RMSDL
- EPOT, EGAP and variants

Tools:

- MTD energy
- BM integration
- ABF integration
- Multiple walker MTD – server and administration client
- Multiple walker ABF – server and administration client

PMFLib, design

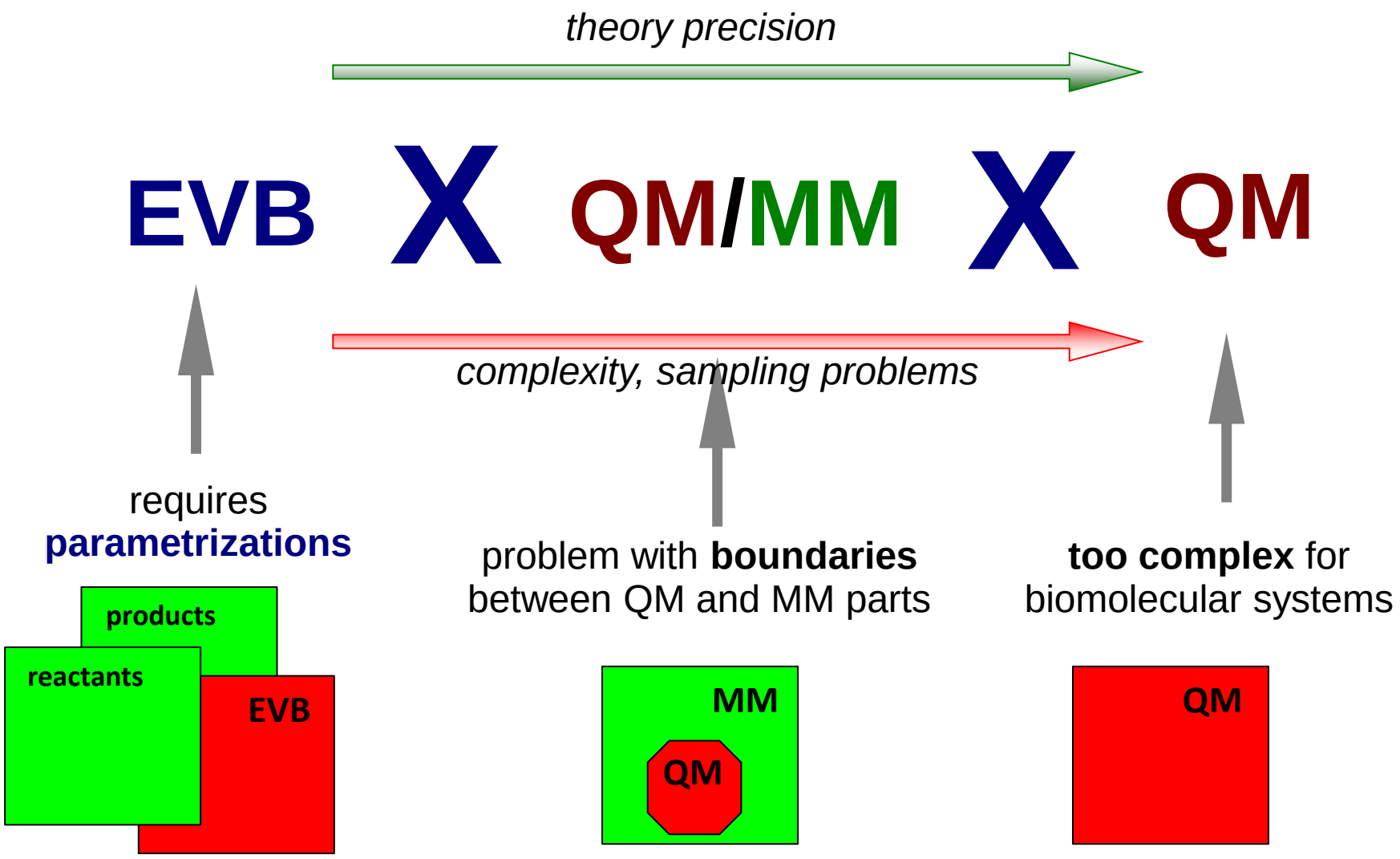


PMFLib

- Fortran 90 (162 files / ~25000 lines)
- C/C++ (178 files / ~24000 lines)

Molecular Dynamics and Reactions

Description of Chemical Reactions



CPMD
Car-Parrinello Molecular Dynamics

Method

$$\mathcal{L}_{\mathrm{CP}}[\mathbf{R}^N, \dot{\mathbf{R}}^N, \{\Phi_i\}, \{\dot{\Phi}_i\}] = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \mu \left\langle \dot{\Phi}_i \middle| \dot{\Phi}_i \right\rangle - \mathcal{E}^{\mathrm{KS}}\left[\{\Phi_i\}, \mathbf{R}^N\right]$$

Equations of motion:

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E^{\mathrm{KS}}}{\partial \mathbf{R}_I} + \sum_{ij} \Lambda_{ij} \frac{\partial}{\partial \mathbf{R}_I} \langle \Phi_i \mid \Phi_j \rangle$$ ions

$$\mu \mid \ddot{\Phi}_i \rangle = -\frac{\delta E^{\mathrm{KS}}}{\delta \langle \Phi_i \mid} + \sum_j \Lambda_{ij} \mid \Phi_j \rangle$$ wavefunction

fictitious mass of wavefunction
(ca 300-700 a.u. , **typical value is about 600 a.u.**

constraints due to
orthonormality of wavefunction

CPMD versus BOMD

CPMD

- no SCF procedure
- motion of ions in time
- motions of wavefunction in time
- time step ~ 0.1 fs (5 a.u.)

BOMD

- SCF procedure
- motion of ions in time only
- wavefunction follows ions by SCF (BO)
- time step max ~ 1 fs
- gradients require very tight convergence of wavefunction optimization

- DFT only (in CPMD)
- hybrid functional possible but very slow
- dispersion correction available

- planewaves wavefunction (periodicity!}
- wavefunction quality is determined by cutoff (single value)
- pseudopotentials required (core electrons)

Practicals …

- read manual (it was very improved in the last version) !
- read two chapters from NIC books about CPMD (about 150 pages), freely downloadable
- be veeeery patient

Typical setup:

- time step 5 a.u.
- WF mass 600 a.u.
- pseudopotentials: Troulier-Martins normconserving
- WF cutoff: 70 Rydbergs
- charged/isolated systems (also in QM/MM calculations)
  - add 2-3A to box dimensions, molecule has to be centered!!!
  - Poission solver: Tuckerman
- heating: Berendsen thermostat
- production: Nose-Hoover thermostat
- ultrasoft Vanderbilt pseudopotential are problematic for Mg

Practicals in smaller group.

Acknowledgements

- Igor Tvaroška, Stanislav Kozmon
- Jaroslav Koča, Zora Střelcová, Jakub Štěpán
- Monika Fuxreiter
- POSTBIOMIN (financial support)