

MC 3.1.8 Theoretische Chemie/ Quantenchemie Computerübungen im Wintersemester 2019/2020

Übung 8 : S_N2 Reactions With Molecular Metadynamics

2019-11-19

1 Introduction

Previously, we optimized and found the transition state for a S_N2 reaction using a static calculation by guessing a transition state manually or by using the quasi synchronous transit 2 method. While this works easily for simple reactions like these small S_N2 reactions, transition states of more complex reactions can be difficult to guess. Molecular Dynamics provides an unbiased methodology to study dynamical processes, such as reactions, especially if Ab Initio Molecular Dynamics is used.

1.1 Ab Initio Molecular Dynamics

In many cases the behaviour of chemical systems over time is of interest and dynamical processes cannot be neglected, such as:

- spectroscopy, which is sensitive to conformation of a molecule, such as NMR
- transfer processes, such as electron transfer dynamics
- proteine and enzyme folding
- molecular adsorption and filter processes, e.g. in zeolites
- ...

To describe the temporal evolution of a chemical system, MD treats the atomic positions classically by solving the newtonian equations of motion:

$$\vec{F}_A = m_A \ddot{\vec{R}}_A. \quad (1)$$

In Ab Initio MD the forces acting on a nucleus A are calculated by a quantum chemical method in each timestep, which makes this method quite expensive in terms of the required computational power. Most often DFT methods are used, but for electronically excited states even EOM-CC dynamics or calculations of the CAS-SCF type can be used. After the forces on

the current nuclear positions have been calculated, new positions and velocities are calculated by numerical integration, e.g. with the Leap-Frog algorithm:

$$\begin{aligned}\vec{R}_A(t + \tau) &= \vec{R}_A(t) + \tau \vec{v}_A(t + \tau/2) \\ \vec{v}_A(t + \tau/2) &= \vec{v}_A(t - \tau/2) + \frac{\tau}{M_A} \vec{F}_A(t),\end{aligned}\tag{2}$$

where τ is the propagation time step. Contrary to most force field based methods, which are very popular in biochemistry, AIMD allows to describe processes such as bond breaking and surface crossing. For basics of molecular dynamic see [3].

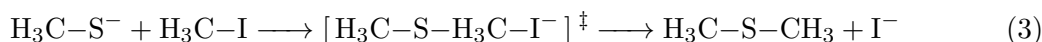
1.2 Metadynamics

For processes with high activation barriers, that can not easily be overcome within the affordable simulation time, one can employ Molecular Metadynamics. Building on the concept of normal MD, the energy surface is biased by repulsive potentials. In the original formulation of Metadynamics, repulsive Gaussian potentials with fixed width and height are spawned every n th timestep during an MD run. The flooding of potential wells is forcing the system to leave deep wells in the energy surface, where the trajectory resides for a long time. This also means, that many repulsive Gaussian potentials are spawned till the system leaves this well. Additionally to exploring more and more of the energy landscape over time, the second advantage is, that the negative sum of the repulsive Gaussian potentials converges to the Gibbs energy surface. Therefore, Metadynamics provides a method to explore even complex mechanisms relatively unbiased and to sample the Gibbs energy surface, which is more helpful for calculating rate constants and occupation distributions than the potential energy surface, as it includes all the other relevant nuclear degrees of freedom. [10]

Biasing the potential requires definition of so called collective variables, along which the repulsive potentials are spawned and the Gibbs energy surface is evaluated. They should describe the reaction/process, that you want to observe, well. Collective variables can be nearly any observable; arbitrarily defined distances, angles, dihedrals, coordination numbers, ionisation potentials, population numbers, so called path collective variables for reactions, which cannot easily be described by a small set of collective variables, and many more.

2 Calculation Details

We want to simulate a S_N2 reaction using molecular metadynamics including solvent effects. The reaction of methanesulfide with iodomethane is an easy example of such a reaction.



As synthesis is almost never done in the gas phase, we need to take solvent effects, such as cage effects, electrostatic and dispersion interactions into account. We will use ethanenitrile ($\text{H}_3\text{C-C}\equiv\text{N}$) as our solvent.

2.1 Generation of the Initial Configuration

Use *Avogadro* [5] to draw each reactand (ethanenitrile, methanesulfide, potassium as a counter ion) and the solvent (ethanenitrile). Optimize the drawn structures with the MMFF94 force field [4] and save the structures as PDB files. Open the PDB files with a text editor of your choice and remove the connectivity information. The UNL entries are the names of residues in a protein. Here we change them to a 3 letter code of our molecule, your choice. Be careful not to change indentation, the PDB format is sensitive with respect to columns! Also note the atom labels in column 3. Your final PDB file should look similar to this (for iodomethane):

Listing 1: PDB file for iodomethane obtained from *Avogadro* after modifications to the residue name (UNL changed to ICH)

```
      COMPND      IODOMETHANE
      AUTHOR      GENERATED BY OPEN BABEL 2.4.1
      HETATM      1  I  ICH      1      0.067 -0.751  0.151  1.00  0.00
I
      HETATM      2  C  ICH      1     -0.007  1.278 -0.429  1.00  0.00
C
      HETATM      3  H  ICH      1      0.554  1.445 -1.353  1.00  0.00
H
      HETATM      4  H  ICH      1      0.421  1.925  0.344  1.00  0.00
H
      HETATM      5  H  ICH      1     -1.039  1.601 -0.601  1.00  0.00
H
      END
```

Now that we have our molecules drawn, we prepare an initial configuration for our MD using the *Packmol* [13] software. The input format is very intuitive:

Listing 2: A *Packmol* input file for the solvation of iodomethane and methanethiole in ethanenitrile. Adapt to the names of your files as necessary.

```
#
# Solvate methanesulfide and iodomethane (reactands) in a box
# of 10 angstrom length in ethanenitrile.
#

# separate all molecules by at least 3.0 angstrom
tolerance 3.0

# read and write files in the PDB format
filetype pdb
output SN2ReactandSolvate.pdb

# 1 molecule iodomethane in a cubic simulation box of 8.5 angstrom
# length. This is reduced because of the periodic boundary conditions
# to avoid clashes between the atoms.
structure Iodomethane.pdb
number 1
inside cube 0.0 0.0 0.0 8.5
end structure

structure Methanesulfide.pdb
number 1
```

```
inside cube 0.0 0.0 0.0 8.5
end structure

structure Potassium.pdb
number 1
inside cube 0.0 0.0 0.0 8.5
end structure

structure Ethanenitrile.pdb
number 10
inside cube 0.0 0.0 0.0 8.5
end structure
```

Make sure you have the PDBs of your molecules and the *Packmol* input file in the same directory and then run *Packmol* by executing:

```
packmol < $PACKMOL_INPUT_FILE
```

You should now have a PDB file named *SN2ReactandSolvate.pdb*. You can have look at this structure with the *VMD* [7] program.

Questions

- Calculate the density of the solution in the box. Is this a physically meaningful density?
- In *Avogadro* we have used the MMFF94 force field. Have a quick look at the literature regarding this force field. Judge if this force field is a valid choice for obtaining preoptimized molecular structures for our individual molecules. Why is preoptimization necessary?

2.2 Unbiased Molecular Dynamics Run

We use the configuration obtained in the previous step to start an unbiased and therefore normal MD run. We will do this with the *CP2K* [8] program, which excels in DFT- and semiempiricism based molecular dynamics. *CP2K* is able to handle gas phase, liquid, soft matter and solid (crystalline and other condensed matter) systems in 0 to 3 dimensional periodicity. We use the Kohn-Sham DFT within the mixed Gaussian-Plane Wave scheme [12, 11]. The GPW scheme is most efficient for pure functionals without Hartree-Fock exchange. We use the reliable ab initio TPSS meta-GGA functional [14], with GTH-pseudopotentials [2, 6, 9] and basis sets of the Molopt type [15].

Write an input file, that resembles the following, commented example:

Listing 3: *CP2K* input file for an unbiased molecular dynamics run. The input file will be provided on the cluster computer.

```
&global
# Name of the project, which influences the names of output files,
# that are created
```

```
project SN2_DFT_MD

# Perform a molecular dynamics run (instead of pure energy evaluation,
# geometry optimization, ...)
run_type md
print_level low
&end global

# This section influences how energies, forces and properties are calculated
&force_eval
# Quickstep is the quantum module of CP2K. Fist would be for molecular mechanics
method quickstep

&dft
# Gaussian basis set and norm conserving pseudopotentials necessary for GPW DFT
basis_set_file_name BASIS_MOLOPT
potential_file_name POTENTIAL

# the MGRID section defines the number of plane waves and grid points in
# the GPW, GAPW and SIRIUS (PW) schemes
&mgrid
# The cutoff is the maximum energy of the plane waves in Rydberg.
# This normally needs to be converged to a sufficient value in
# a separate calculation
cutoff 250

# The number of differently spaced grids to map the density to.
# MOLOPT basis set profit from larger numbers of grids, while
# GTH basis sets or all electron basis sets are OK with 3 or 4 grids
ngrid5 5

# The distribution of hard and soft density domains over different grids
rel_cutoff 50
&end mgrid

&qscf
# The default screening threshold, below which matrix elements
# in many steps will be neglected. Influences many other thresholds
eps_default 1.0e-10

# Extrapolate the density matrix of the next MD step from
# the density matrix of previous steps by a polynomial fit.
# This greatly accelerates calculations
extrapolation aspc
extrapolation_order 5
&end qscf

# As we have bond breaking, use unrestricted orbitals
# to have some correlation present
uks .true.

# A fast convergence of the self consistent field is the most crucial step for
# fast MD simulations. We use the DIIS algorithm in the orbital transformation
# section, to avoid excessive matrix diagonalizations, which slow down the
# calculation. Preconditioning is also extremely important and the usage of
# outer scf cycles, which recalculates the preconditioner,
# accelerating convergence
```

```
&scf
scf_guess atomic
max_scf 20
eps_scf 1.0e-5
max_diis 15
&outer_scf
max_scf 20
eps_scf 1.0e-5
&end outer_scf
&ot
preconditioner full_all
energy_gap 0.001
algorithm irac
minimizer diis
n_history_vec 20
&end ot
&print
&restart
&each
md 0
&end each
&end restart
&end print
&end scf

# The choice of the exchange correlation functional
&xc
# the unmodified TPSS functional
&xc_functional tpss
&end xc_functional

# Grid smoothing for small cutoff values such as our
# 250 is very important to get accurate forces.
# GPW is NOT translational invariant.
&xc_grid
xc_deriv nn10_smooth
xc_smooth_rho nn10
&end xc_grid

# DFT is not very good at describing dispersion interactions.
# We use Grimme's empirical DFT-D3 correction with Becke-Johnson damping
&vdw_potential
dispersion_functional pair_potential
&pair_potential
type dftd3(bj)
parameter_file_name dftd3.dat
d3bj_scaling 1.000 0.4535 1.9435 4.4752
&end pair_potential
&end vdw_potential
&end xc

# As we have a liquid system, the electrostatic interactions
# need to be taken into account fully periodic
&poisson
periodic xyz
poisson_solver periodic
&end poisson
```

```
&end dft

# The SUBSYS section defines the chemical system and our simulation environment
&subsys
# Our simulation cell (cubic), which is fully periodic
&cell
abc 10.0 10.0 10.0
periodic xyz
&end cell

# Read atomic coordinates, atom types and so on from file.
# As we do quantum chemistry we are not interested in connectivities
&topology
coord_file_format pdb
coord_file_name SN2ReactandSolvate.pdb
conn_file_format off
&end topology

# Specify basis sets and pseudopotentials for each element
&kind H
element H
basis_set dzvp-molopt-sr-gth
potential gth-pbe
&end kind
&kind C
element C
basis_set dzvp-molopt-sr-gth
potential gth-pbe
&end kind
&kind N
element N
basis_set dzvp-molopt-sr-gth
potential gth-pbe
&end kind
&kind S
element S
basis_set dzvp-molopt-sr-gth
potential gth-pbe
&end kind
&kind K
element K
basis_set dzvp-molopt-sr-gth
potential gth-pbe
&end kind
&kind I
element I
basis_set dzvp-molopt-sr-gth
potential gth-pbe
&end kind

# Define collective variables for use in metadynamics later.
# They are not yet biased
# the S-C distance
&colvar # 1
&distance
atoms 6 2
&end distance
```

```
&end colvar

# the C-I distance
&colvar # 2
&distance
atoms 1 2
&end distance
&end colvar

# The angle of the reactive atoms S-C-I
# (should be around 180° in the transition state)
&colvar # 3
&angle
atoms 1 2 7
&end angle
&end colvar

# The number of carbons coordinating iodine
&colvar # 4
&coordination
atoms_from 1
atoms_to 2
r0 [angstrom] 2.1
nn 6
nd 12
&end coordination
&end colvar

# The coordination S to CI
&colvar # 5
&coordination
atoms_from 6
atoms_to 2
r0 [angstrom] 1.8
nn 6
nd 12
&end coordination
&end colvar
&end subsys
&end force_eval

# The MOTION section tells CP2K how to move atoms with respect to the
# forces calculated in the FORCE_EVAL section
&motion
# Parameters for our MD simulation
&md
# We want a constant temperature and volume simulation
ensemble nvt

# Over 10000 propagation steps
steps 10000

# With a time step of 0.5 fs
timestep 0.5

# At 375 K
temperature 375
```

```
# The temperature needs to be regulated by a thermostat, otherwise we would
# have a constant energy simulation
&thermostat
# Different types of thermostats exist. Nose-Hoover is a very good choice
type nose
&nose
# The time constant (50 fs) defines how strong the coupling is,
# that means how fast the system will equilibrate but also
# how heavily it is disturbed
timecon 50.0
&end nose
&end thermostat
&end md

# The section for metadynamics
&free_energy
method metadyn
&metadyn
# We are not doing metadynamics, yet. We only need this section for printing
do_hills .false.

# Even if we are not doing anything with the variables now,
# we need to reference them here, so that they get printed
&metavar
width 1.0
colvar 1
&end metavar
&metavar
width 1.0
colvar 2
&end metavar
&metavar
width 1.0
colvar 3
&end metavar
&metavar
width 1.0
colvar 4
&end metavar
&metavar
width 1.0
colvar 5
&end metavar

&print
&colvar
common_iteration_levels 1
&each
md 1
&end each
&end colvar
&end print
&end metadyn
&end free_energy

&print
```

```
&restart
backup_copies 0
&each
md 1
&end each
&end restart
&restart_history
&each
md 0
&end each
&end restart_history
&end print
&end motion
```

Run this MD for some time. As this takes quite some time, we will provide the output files. You will have the following files:

SN2_DFT_MD.out Output file of *CP2K*, containing informations about SCF convergence, calculation details, timings, ...

SN2_DFT_MD-pos-1.xyz Trajectory of this MD run in cartesian coordinates. This can be visualised using *VMD*

SN2_DFT_MD-COLVAR-1.metadynLog Behaviour of the collective variables over time

SN2_DFT_MD-1.ener Characteristics of the MD run, such as potential energy, kinetic energy, temperature, ...

Questions

- Plot the collective variables. Describe their evolution and judge if the selected variables are useful for a metadynamics run, in which we want to force the S_N2 reaction to occur. Do you recognize something about the C–I distance and the C–I coordination number? How is the coordination number calculated?
 - Discuss the choice of the time constant for the Nosé-Hoover thermostat. What is a good choice for a MD simulation? Take the evolution of the temperature over the MD and the influence on the simulation behaviour into account.
 - Have a look at the kinetic energy of the particles in the simulation. What is a typical activation barrier for a S_N2 reaction? Do you expect your reactands to overcome this barrier in this MD run, based on the kinetic energies?
 - Why did we choose 0,5 fs as the value for our timestep? Would it be desirable to choose larger timesteps for faster propagation?
-

2.3 Molecular Metadynamics Run

Now we choose our collective variables to be biased during our MD run. The C–I and S–C distances are promising and biasing both would give us a 2 dimensional Gibbs energy surface. Run the metadynamics with this two collective variables biased. We spawn repulsive Gaussian potentials every 50th step with a height of $0,0025 E_h$ and a width of the curve of 50 pm. Modify the input file for CP2K in the `subsys` and the `motion` sections.

Listing 4: Input for the `subsys` section of *CP2K*. This *replaces* the definition of the old collective variables

```
# the S-C distance
&colvar # 1
&distance
atoms 6 2
&end distance
&end colvar

# the C-I distance
&colvar # 2
&distance
atoms 1 2
&end distance
&end colvar
```

Listing 5: Input for the `motion` section of *CP2K*.

```
&motion
&md
ensemble nvt

# increase the number of MD steps
steps 100000
timestep 0.5
temperature 375
&thermostat
type nose
&nose
timecon 50.0
&end nose
&end thermostat
&end md

&free_energy
method metadyn
&metadyn
# bias the MD now
do_hills .true.

# spawn gaussian potentials every 50th step
nt_hills 50

# the gaussians have a height of 0.0025 Hartree
ww 0.0025

# and build up over nt_hills steps, instead of being
```

```
# spawned instantaneously
slow_growth .true.

&metavar # S-CI distance
width [angstrom] 0.5
colvar 1

# insert a wall for the collective variable at 8 angstrom distance
&wall
position [angstrom] 8.0

# the wall is a very large and narrow gaussian
type gaussian
# but we could also have used inversion of the velocities
&reflective
direction wall_minus
&end reflective
# or a quadratic potential
&quadratic
direction wall_minus
k 0.1
&end quadratic
&gaussian
sigma [angstrom] 0.025
ww [hartree] 0.5
&end gaussian
&end wall
&end metavar
&metavar # I-CI distance
width [angstrom] 0.5
colvar 2
&wall
position [angstrom] 8.0
type gaussian
&reflective
direction wall_minus
&end reflective
&quadratic
direction wall_minus
k 0.1
&end quadratic
&gaussian
sigma [angstrom] 0.025
ww [hartree] 1.0
&end gaussian
&end wall
&end metavar

# write informations about gaussian hills to file
&print
&colvar
common_iteration_levels 1
&each
md 1
&end each
&end colvar
&hills
```

```
common_iteration_levels 1
filename =SN2_DFT_MetaD-HILLS.dat
&each
md 1
&end each
&end hills
&end print
&end metadyn
&end free_energy

&print
&restart
backup_copies 0
&each
md 1
&end each
&end restart
&restart_history
&each
md 0
&end each
&end restart_history
&end print
&end motion
```

Questions

- Discuss the frequency, width and height with which Gaussian potentials are spawned. Take typical bond strengths, activation barriers and the simulation time into account.
 - Calculate and visualize the a) spatial distribution function of the hydrogen atoms of the methyl group of the ethanenitrile b) of the potassium ion with respect to the the ethanenitrile c) of the potassium atom with respect to the methanesulfide d) the radial distribution function of iodine and potassium with *Travis* [1]. Be careful in the first steps of the interactive *Travis* session to define your molecules properly. It is likely that *Travis* is making some mistakes for this trajectory. Discuss these distributions functions.
 - Visualize the behaviour of the two collective variables over time. The x axis should be the S–C distance and the y axis the I–C distance. Colour the trajectory by time. Discuss the behaviour of the collective variables.
 - Calculate the Gibbs energy surface of the metadynamics run using the `graph.psm` tool. Plot the Gibbs energy surface, discuss, where reactands and products can be found, where the transition state is, draw the minimum energy path and highlight the transition state. Discuss the results and possible artifacts.
-

Literatur

- [1] Martin Brehm und Barbara Kirchner. "TRAVIS - a free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories." In: *Journal of chemical information and modeling* 51.8 (08/2011), S. 2007–23. DOI: 10.1021/ci200217w.
 - [2] S. Goedecker, M. Teter und J. Hutter. "Separable dual-space Gaussian pseudopotentials". In: *Physical Review B* 54.3 (01/1996), S. 1703–1710. DOI: 10.1103/PhysRevB.54.1703.
 - [3] Reinhold Haberlandt u. a. *Molekulardynamik: Grundlagen und Anwendungen*. Vieweg+Teubner Verlag, 01/1995, S. 252. DOI: 10.1007/978-3-322-90870-4.
 - [4] Thomas A. Halgren. "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94". In: *Journal of Computational Chemistry* 17.5-6 (01/1996), S. 490–519. DOI: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.
 - [5] Marcus D Hanwell u. a. "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform". In: *Journal of Cheminformatics* 4.1 (01/2012), S. 17. DOI: 10.1186/1758-2946-4-17.
 - [6] C. Hartwigsen, S. Goedecker und J. Hutter. "Relativistic separable dual-space Gaussian pseudopotentials from H to Rn". In: *Physical Review B* 58.7 (01/1998), S. 3641–3662. DOI: 10.1103/PhysRevB.58.3641.
 - [7] William Humphrey, Andrew Dalke und Klaus Schulten. "VMD: Visual molecular dynamics". In: *Journal of Molecular Graphics* 14.1 (01/1996), S. 33–38. DOI: 10.1016/0263-7855(96)00018-5.
 - [8] Jürg Hutter u. a. "cp2k: atomistic simulations of condensed matter systems". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 4.1 (01/2014), S. 15–25. DOI: 10.1002/wcms.1159.
 - [9] M. Krack. "Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals". In: *Theoretical Chemistry Accounts* 114.1-3 (01/2005), S. 145–152. DOI: 10.1007/s00214-005-0655-y.
 - [10] Alessandro Laio und Michele Parrinello. "Escaping free-energy minima." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (10/2002), S. 12562–6. DOI: 10.1073/pnas.202427399.
 - [11] Gerald Lippert, Jürg Hutter und Michele Parrinello. "The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations". In: *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* 103.2 (01/1999), S. 124–140. DOI: 10.1007/s002140050523.
 - [12] Gerald Lippert, Michele Parrinello und Jurg Hutter. "A hybrid Gaussian and plane wave density functional scheme". In: *Molecular Physics* 92.3 (01/2010), S. 477–488. DOI: 10.1080/002689797170220.
 - [13] L Martínez u. a. "PACKMOL: a package for building initial configurations for molecular dynamics simulations." In: *Journal of computational chemistry* 30.13 (10/2009), S. 2157–64. DOI: 10.1002/jcc.21224.
-

- [14] Jianmin Tao u. a. “Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids.” In: *Physical review letters* 91.14 (10/2003), S. 146401. DOI: 10.1103/PhysRevLett.91.146401.
 - [15] Joost VandeVondele und Jürg Hutter. “Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases.” In: *The Journal of chemical physics* 127.11 (09/2007), S. 114105. DOI: 10.1063/1.2770708.
-