

## Chemistry block (exercise 1)

In this exercise, students will be introduced how to perform simple quantum – chemical calculations.

- Input files for Gaussian09.
- Output file structure.
- Geometry optimization, Frequency analysis, Electron density distribution, Implicit solvation methods (PCM).

In order to perform quantum-chemical calculations we need to define geometry of the system. There are several way:

### □ XYZ file format

XYZ file describes molecular geometry

Component from input files

First line – number of atoms in the molecule.

Second line – comment.

Cartesian coordinates of each atom.

Widely used in field of computational chemistry for description of the molecular geometry. Default units are [Ångströms](#).

Example:

```
5
methane molecule (comment)
C    0.000000    0.000000    0.000000
H    0.000000    0.000000    1.089000
H    1.026719    0.000000   -0.363000
H   -0.513360   -0.889165   -0.363000
H   -0.513360    0.889165   -0.363000
```

### □ Protein Data Bank (pdb) file format

Text file describes the tridimensional structure of molecules (usually proteins) from Protein data bank

The Protein Data Bank (PDB) format provides a standard representation for macromolecular structure data derived from X-ray diffraction and NMR studies. This representation was created in the 1970's and a large amount of software using it has been written.

pdb contains quite large number of lines, depending on the structure of the system.

### □ Z - matrix

Describes molecular geometry in terms of internal coordinates.

Internal coordinates.

Interatomic distance (between two atoms)

Angle (between three atoms)

Dihedral angle (four atoms)

Widely used in computational chemistry.

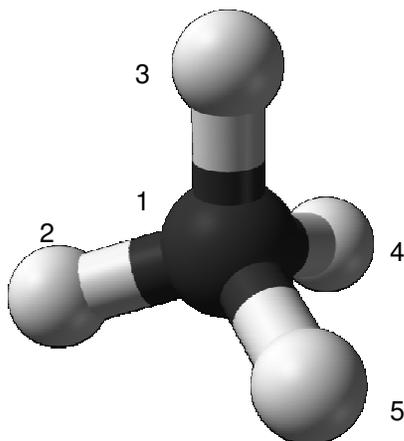
Easy to apply symmetry

Easy to introduce geometry constraints

Easy to create (from small molecules – by hand)

### Example: Z - matrix for methane molecule

```
1 C
2 H 1 1.089000
3 H 1 1.089000 2 109.4710
4 H 1 1.089000 2 109.4710 3 120.0000
5 H 1 1.089000 2 109.4710 3 -120.0000
```



Input file for G09 program contains several sections:

- Link 0: scratch files, checkpoint files.
- Route section (#): define the calculation protocol.
- Title section: description of the task, maximum 5 lines.
- Molecule specification: description of the molecular geometry

Simple example is :

```
%Chk=heavy           Link 0 section
# HF/6-31G(d) Opt    Route section
```

```
Opt job              Title section
0 1                  Charge and multiplicity of the system
section atomic coordinates (XYZ format or Z-matrix)
```

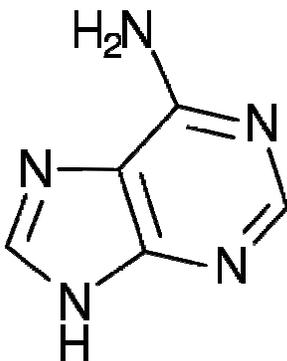
Input file is Free format, non sensitive to capital letters. Keywords must be separated by one of the following: blank space, tab, coma. Arguments to the keywords can be introduced in several formats:

```
keyword = option
keyword(option)
keyword=(option1, option2, ...)
keyword(option1, option2, ...)
```

**Task:** Relative stability of different tautomers of Adenine

- Geoemtry optimization of adenine in gas phase  
# B3LYP/6-31+G\* opt
- Relative stability of different tautomers of adenine

- Vibrational analysis and Natural bond orbital analysis of different tautomers of adenine.
- # B3LYP/6-31+G\* freq pop=nbo
- Difference in free energy (Gibbs free energy) between different tautomer forms.
- Solvation energy of adenine in framework of PCM model.
- # B3LYP/6-31+G\* opt scrf(pcm, solvent=water)



## Chemistry block (exercise 2)

### Perform ad-initio MD simulations with CP2K code

Molecular dynamics (MD) is a computer simulation of physical movements of atoms and molecules in the context of N-body simulation. The atoms and molecules are allowed to interact for a period of time, giving a view of the motion of the atoms. The approximations are based on theoretical considerations, not on empirical fitting. *Ab-initio* calculations produce a vast amount of information that is not available from empirical methods, such as density of electronic states or other electronic properties. A significant advantage of using *ab-initio* methods is the ability to study reactions that involve breaking or formation of chemical bonds, which correspond to multiple electronic states.

**CP2K** is a [freely available](#) (GPL) program, written in Fortran 95, to perform atomistic and molecular simulations of solid state, liquid, molecular and biological systems. It provides a general framework for different methods such as e.g. **density functional theory** (DFT) using a mixed Gaussian and plane waves approach (GPW), and **classical pair and many-body potentials**.

Necessary files to run a simulation are:

```
force_eval.inc
md_nvt.inp, md_npt.inp, md_nve.inp
Subsis.inc
```

Before the input is parsed, the input is run through a simple internal preprocessor. The preprocessor recognizes the following directives independent of capitalization: **@INCLUDE 'filename.inc**

'The file referenced by *filename.inc* is included into the input file and parsed. Recursive inclusions are not allowed and the files have to exist in the current working directory. There can be only one @INCLUDE statement per line. Single or double quotes have to be used if the filename contains

#### **@SET VAR value**

Assigns the text *value* to the preprocessing variable *VAR*. *value* is the text following *VAR* with the outer whitespace removed. The variable can be recalled with a  $\${VAR}$  (or  $\$VAR$ ) statement. There can be only one @SET statement per line.

#### **$\${VAR}$ or $\$VAR$**

Expand the variable *VAR*. The text  $\${VAR}$  (or  $\$VAR$ ) is replaced with the value assigned to *VAR* in the last @SET directive. There can be multiple variable statements per line. The expansion process is repeated until no more variables are found.

#### **@IF / @ENDIF**

Conditional block. The text from the @IF line up to the next line with a valid " @ENDIF is skipped, if the expression following @IF resolves to *false*. Available expressions are lexical comparisons for equality '=' or inequality '/='. If none of the two operators are found, a '0' or whitespace resolves to *false* while any text resolves to *true*. @IF/@ENDIF blocks cannot be nested and cannot span across files. There can be only one test (== or /=) per @IF statement.

Input file of CP2K code contains separate sections, each section starts with & and finish with &END. Each section contain (if necessary) subsection defined in the same way. The name of the sections, subsections and keywords are not key sensitive. Unit dimensions can be defined simply bay giving the proper unit in [] after the value. For example 1.2 [m] means 1.2 meters instead of default value Angstrom.

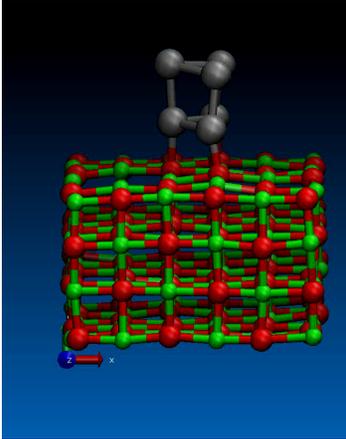
Examples for MD simulations with CP2K:

- *ssh sc\_student@62.44.102.111*
- */varna/sc\_student/test/cp2k*
- *force\_eval.inc, md\_nvt.inp, \*.xyz u md.jcf*

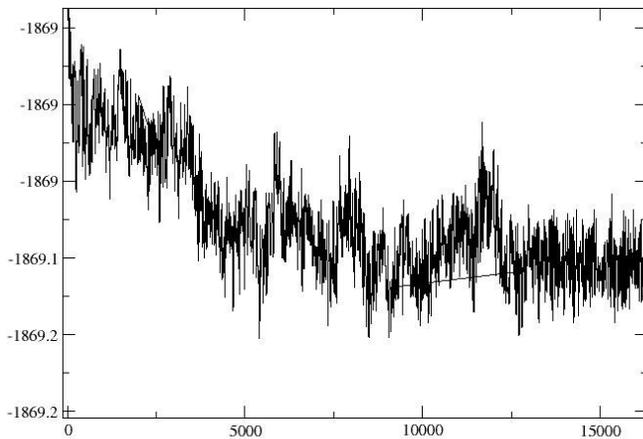
Important output files:

- ***MD-1.ener*** – file, in it you can find the following columns: ***md\_step, time[fs], e\_kin [hartree], temp[K], e\_pot [hartree], e\_tot [hartree], cpu\_time\_per\_step [s]***. From this file one can easily extract the data how the simulation parameters changes during the run and can analyze them.
- ***MD-1.restart*** – file necessary for further restart of the simulation.
- ***MD.out*** – main output file.
- ***MD-pos-1.xyz*** – file contains the trajectories of the particles in the simulated system during the run.

**Examples 1:** MD of Rh<sub>4</sub> cluster on defected MgO cluster

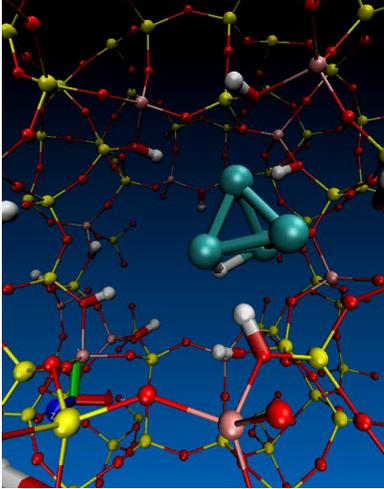


This movie shows the evolution of the Rh<sub>4</sub> cluster after formation of defect on the surface of MgO cluster. Since the Rh cluster is an electron deficient system it goes and occupies the place of defect where are located the electrons from the missing O from MgO surface.

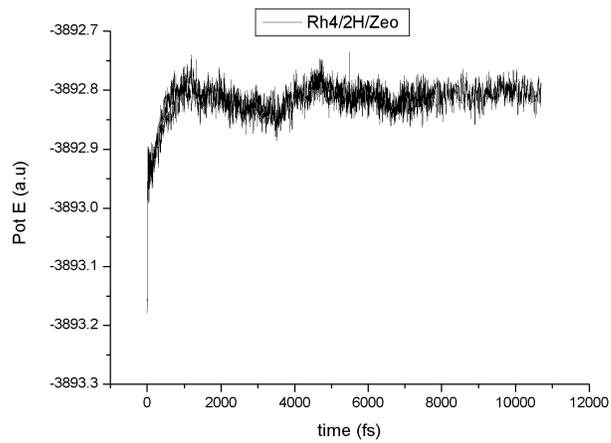


On the figure above is shown the variation of the potential energy of the system during the MD run. At the equilibrium state,  $E(\text{pot})$  starts to oscillate around average value and then we are able to start the productive run of the simulation.

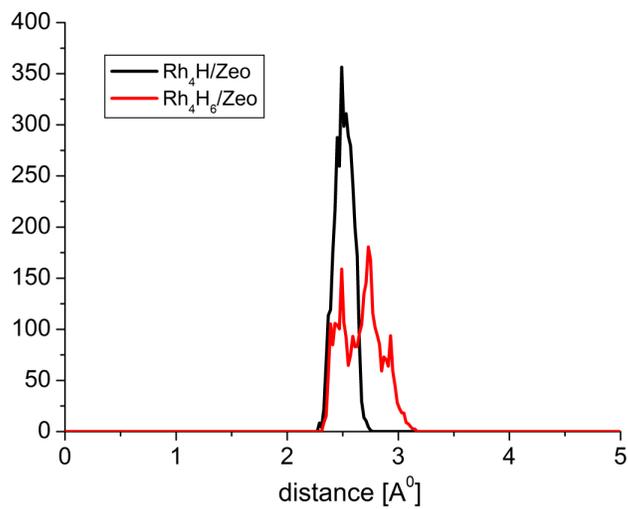
**Examples 2:** MD of Rh<sub>4</sub> cluster inside the zeolite channels and the process of reversed hydrogen spillover.



This movie shows the process of spontaneous reversed hydrogen spillover from the OH groups in the zeolite framework to the cluster.



Potential energy vs. simulation time shows stable behavior after ~10 ps of the simulation.



Radial pair distribution function for Rh atoms in the cluster is shown on the figure above. From the position of the maximum in RDF we are able to conclude that loading of the cluster with H leads to enhancement of the interatomic distances.

Task: Perform ab-initio MD simulation of dissociative adsorption of  $H_2$  on  $Rh_6$  cluster in NVT ensemble at 300K, time step of 1 fs, using DFT formalism with PBE exchange correlation functional.

