Worksheet 3: properties and fitting of atomistic water models and ab initio molecular dynamics of water monomer/dimer

Miriam Kohagen, Frank Uhlig, Maria Fyta

Institute for Computational Physics, University of Stuttgart

May 15, 2018

Contents

General remarks

- Deadline for the report is **Monday, 30th May 2018, 13:00 p.m.**

- In this worksheet, you can achieve a maximum of 20 points.

- The report should be written as though it would be read by a fellow student who attends the lecture, but doesn't do the tutorials.

- To hand in your report, send it to your tutor via email.

    - Miriam (`mkohagen@icp.uni-stuttgart.de`)

- Please attach the report to the email. For the report itself, please use the PDF format (we will *not* accept MS Word doc/docx files!). Include graphs and images into the report.

- The report should be 5–10 pages long. We recommend using LaTeX. A good template for a report is available online.

- The worksheets are to be solved in **groups of two or three** people.

1   Introduction

In this worksheet, you will first tackle several theoretical tasks related to water models. In the second part, you will create your own water model using the open-source Molecular Dynamics software package GROMACS and you will run various analyses on the

simulation data and investigate the properties of your "optimal" water model. In the third part, you will run ab initio molecular dynamics simulations of a water monomer and a water dimer and examine their vibrational spectra, and some other properties. The simulations will be performed using the software package CP2K.

All files required for this tutorial can be found in the archive `templates.zip` which can be downloaded from the lecture's homepage.

2 Short questions - short answers (4 points)

Have a look at `http://www.edinformatics.com/interactive_molecules/water.htm` and at `http://www1.lsbu.ac.uk/water/water_models.html`.

Task (4 points)

Answer the following questions:

- Why is a water molecule polar?

- What is a hydrogen bond?

- What are the values for the H-O-H angle in water and the typical distance of a hydrogen bond (*not* O-H bond)?

- What are the main differences between various atomistic water models?

Hints

- The first website provides some nice applets to play with. Anyway: Don't trust the results of the measurements you can make there! The second website gives an overview for various water models.

3 Atomistic water simulations with GROMACS (8 points)

In this exercise, you will perform several simulations to optimize a given water model. During the original optimization procedure of this model, someone forgot to include the polarization energy correction. You will re-optimize the water model taking this correction into account. The optimization procedure focuses on the partial-charge distribution of the water model. You will run several simulations using different parameter sets and based on the heat of vaporization and other data optimize the initially given water model.

All simulations are performed with GROMACS. GROMACS is a freely available Molecular Dynamics software package available at `www.gromacs.org`. You can either use the computers in the ICP CIP pool or install GROMACS on your own computer. Please note, that since Gromacs 5.0.1 the structure of the input files changed and the input files (at least some of them) provided here will not work. The simulations focus on improving the SPC model.

Download the archive `templates.zip` from the lecture website. After unpacking, you will find different files and directories in the directory called `gromacs`:

- `gen.sh`: (incomplete) bash script to help setup calculations

- `setup/`: setup files for simulation

    - `conf.gro`: pre-equilibrated water structure with 216 solvent molecules
    - `grompp.mpd`: parameters for the simulation
    - `topol.top`: topology file for the water simulations
    - `index.ndx`: file needed for the analyses
    - `tut.itp`: topology file of unoptimized water model

- `reference_data/`: reference data

    - `rdf_oo.dat`: experimental oxygen-oxygen radial pair-distribution function

3.1 Visualizing the system

You can visualize the system using `vmd` (Visual Molecular Dynamics):

```
$> vmd conf.gro
```

In the vmd console, you can use the following command to show the simulation box:

```
> pbc box
```

Hint

- By default, all molecules are represented by lines only. To get a nicer looking visualization, use the menu in the "VMD Main" window and click on *Graphics > Representations...*. In the "Graphical Representations" dialog that will pop up, use the *Drawing Method* drop-down menu and select *CPK*.

- You can also save or load a "visualization state" in the drop-down menu under "File", that will contain information on the trajectory files you loaded and the graphical representations you used to visualize them. This allows you to easily load and visualize data according to your preferences.

3.2   Running the simulations

It is a good idea to run the different simulations in separate directories. Create directories for each partial charge set that you want to investigate. Copy the files from the archive into each of these directories and change them accordingly. The archive also has an incomplete script to help you setup your calculations. Some minor fixes are needed before you can actually use it.

First, we will start a simulation using the unoptimized `TUT` water model. In order to have the required GROMACS commands available in your terminal, execute the following command:

```
$> source /group/sm/2016/tutorial_02/gromacs-4.6.7/bin/GMXRC
```

To prepare the first simulation, change to a directory you just created (e.g., `tut`) and call the GROMACS preprocessor `grompp`:

```
$> cd tut
$> grompp
```

If `grompp` completed without errors, the simulation can be started with

```
$> mdrun -v
```

The system will then be simulated for 500 ps with a time step of 2 fs at 300 K.

Task   (2 points)

Perform one simulation for the unoptimized TUT water model and for each of your charge sets.

Hints

- To tell GROMACS about the different charge sets you have to change the charges in the `tut.itp` file

- Use the commands `grompp` and `mdrun` for each simulation as described above.

- reasonable values for the partial charges on the oxygen atom are between -0.8 and -0.9 e.

3.3 Analysis

In the following tasks, you will analyze the simulation data in various ways. Luckily, GROMACS comes with a variety of easy-to-use analysis tools.

Radial distribution function

The radial distribution functions (RDFs) give a first hint on the local structure of the systems. To compute the RDFs, you can use the command

```
$> g_rdf -n index
```

Task (1 point)

- Compute the RDFs of the systems you have simulated.

- Compare the RDFs of the different water models with each other and the given reference and interpret the results (peaks, distances between peaks, differences between water models,...)

Hint

- You can use the program `xmgrace` to plot the output `rdf.xvg` files.

Mean square displacement

In three-dimensional space, the diffusion coefficient D can be calculated from the mean square displacement (MSD) function

$$\langle \Delta r^2(t) \rangle = 6Dt \tag{1}$$

for large enough t. You can calculate the MSD using GROMACS with

```
$> g_msd -n index
```

Task (2 points)

- Compute the MSD for each of the different water models.

- Have a look at the output files `msd.xvg` for the different water models and compare them. What are the differences? Can you identify the diffusional ($\langle \Delta r^2(t) \rangle \propto t$) and the ballistic regime?

- Determine the fitting range that GROMACS uses and check if it is a sensible choice for you analysis. If not, adapt the options to the g_msd command.

Heat of vaporization

The heat of vaporization H_{vap} provides information on the binding strength in the liquid phase compared to the gas phase. It is commonly expressed in units of kJ/mol. The necessary energies can be extracted from the simulations using GROMACS:

```
$> g_energy
```

Further input then needs to be supplied on the console to determine which properties to calculate. For automatization you can also input the latter data input using a command similar to this:

```
$> echo <num1> [num2,...] | g_energy
```

where num1, num2, ... are integers identifying quantities you want to calculate (careful, they depend on the settings of your simulation).

Task (3 points)

- Compute the heat of vaporization for your model systems, including polarization energy correction. Determine the optimal water model from this data (experimental value: $H_{\text{vap}} = 44$ kJ/mol).

- Compare your choice to what you would have determined as optimal water model from diffusivity and the radial distribution functions.

- What other properties have been affected during the parametrization and what physical effects have been neglected in the simulation.

Hints

- the equation for H_{vap} including polarization correction, E_{pol} is as follows:

$$H_{\text{vap}} = -\left[\frac{E_{\text{pot}}}{N} + E_{\text{pol}}\right] + R \cdot T \tag{2}$$

$$E_{\text{pol}} = \frac{1}{2} \cdot \frac{\mu^2}{\alpha} \tag{3}$$

- E_{pot} is the potential energy, N is the number of water molecules, R is the ideal gas constant, T is the temperature, α is the polarizability and μ the induced dipole moment

4 Atomistic water simulations with CP2K (8 points)

In this exercise, you will perform simulations of a water monomer and a water dimer. You will use the density-functional tight-binding approximation to calculate forces for the molecular dynamics integrator.

All simulations are performed with CP2K. CP2K is an open-source ab initio molecular dynamics software package available at `https://www.cp2k.org/`. You can either use the computers in the ICP CIP pool or install CP2K on your own computer.

Check the archive `templates.zip` from the lecture website. You will find a directory called `cp2k/`. Here, you will find the files needed for the simulations:

- `setup/`: directory containing input file templates for the monomer and dimer including the equilibrated input structures

  - `md.inp`: general input file to be used for molecular dynamics
  - `1.xyz and 2.xyz`: equilibrated input structures

- `data/`: directory containing parameter files needed for simulations

The data directory contains parameters for the DFTB calculations. This directory (or a symbolic link to it) needs to be present in the directory where you run the calculations. Symbolic links can be created by:

```
$> ln -s <TARGET> <LINK_NAME>
```

The online, input reference for CP2K is available at `https://manual.cp2k.org/`. Please note that this reference manual merely contains the available input options, but not detailed information on the involved algorithms. The relevant references are, however, linked in the corresponding sections. All input is structured into sections that start with &SECTIONNAME and end with an &END. All sections can contain further subsections and keywords.

4.1 Installing CP2K

In case you are doing these exercises in the computer pool of the Institute for Computational Physics, skip this section and use the pre-installed version of CP2K. Otherwise, you can download CP2K from its git-repository:

```
$> git clone https://github.com/cp2k/cp2k
```

You can build an OpenMP-parallelized version of CP2K as follows:

```
$> cd cp2k/cp2k/arch
$> wget goo.gl/M5JYhr -O myCP2K.sopt
$> cd ../makefiles
$> make ARCH=myCP2K VERSION=sopt
```

This assumes that you have the GNU compiler collection (gcc/gfortran), and basic linear algebra packages installed (LAPACK, BLAS). If everything goes well, the final executable `cp2k.sopt` will be located under `cp2k/exe/myCP2K`.

Hint

- You can supplement the make command with the `-j <NCPUS>` option, where <NCPUS> is to be substituted by the number of parallel processes used to build CP2K.

- If you desire advanced features, examine the other architecture files included in the CP2K distribution. They are however not necessary for this tutorial.

4.2 Running the simulations

On the desktop PCs in the ICP CIP pool you can add CP2K to your local environment by issuing the following command:

```
$> export OMP_NUM_THREADS=2
$> export PATH=/group/sm/2017/tutorial_06/cp2k/exe:$PATH
```

Afterwards, you should be able to simply run the `cp2k.sopt` executable. You can run any CP2K input file by issuing a command like:

```
$> cp2k.sopt <INPUT_FILE> &> <OUTPUT_FILE>
```

where already the redirection of standard and error output to the <output> file is included. Please note, that you do not have to redirect the output, but you will get a lot of screen output if you do not.

First you should inspect the input structures for the monomer and the dimer you can find in the `templates.tar.gz` archive.

After that, inspect the input files. Note that algorithms and their parameters in the input files do not need to be changed. The only input parameters you are supposed to change are the temperature used for thermostatting and the total number of integration steps (for the production run).

The timestep for the molecular dynamics is set to 0.5 fs which is small enough to allow accurate integration of Newton's equations of motion and large enough to achieve reasonable sampling.

Task (2 points)

- Run simulations of the water dimer at 100 K and 300 K. Select a temperature for your final production simulation. Explain your choice.

- Run a simulation for the monomer at the same temperature.

Hints

- You can use programs like Molden or Avogadro to build and visualize your systems. A copy of Molden is also installed on the CIP pool computers in the CP2K directory.

- For the production runs, increase the run time at least ten-fold compared to the values in the to you provided input files. This should suffice to achieve reasonable sampling.

4.3   Analysis of the ab initio molecular dynamics simulations

In the following tasks, you will analyse the simulation data from your production runs in various ways. Only use the part of the simulation that is equilibrated. You can judge the relaxation by trends seen in the potential energy of the simulation.

Energy and geometry

After running the molecular dynamics simulation, CP2K will have generated several output files. `MD-pos-1.xyz` contains the corresponding trajectory, `MD-1.ener` the relevant energies of your system during the MD. The latter file contains a header line denoting the contents of the individual columns.

Task (2 points)

- Determine the mean, potential energy of the hydrogen bond. Do the same for only the input structures and compare the two values.

- Determine and compare structural properties of the water dimer at finite temperature to the one given as input structure.

Hints

- You can use VMD to analyse structural properties, like bond lengths, angles, and dihedrals. Labels for the latter quantities are created by enabling different selection modes for the mouse in the VMD main window under `Mouse`→`Label`. The so created labels can be analysed under `Graphics`→`Labels`→`Graph`.

Infrared spectroscopy

The infrared, vibrational spectrum can be obtained from molecular dynamics simulations. The necessary information is the total dipole moment of the simulated system. Your CP2K simulations should have produced a corresponding output file called `DIPOLE.out`. The classical approximation to the infrared absorption cross-section α is:

$$\alpha(\nu) = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{j=x,y,z} \left| \int_0^\tau dt \exp\left(-i\nu t\right) \frac{d\mu_j}{dt} \right|^2 , \tag{4}$$

where μ_j are the cartesian components of the system's dipole moment.

Task (4 points)

- Determine the infrared spectra of the water monomer and dimer. Plot the absorption cross-section α in dependence of the wavenumber in cm^{-1}.

- What is the origin of the individual vibrations?

- Explain the differences in the obtained monomer and dimer spectra and their origins.

Hints

- You can perform numeric differentiation and Fourier transforms efficiently with NumPy.