

# Practical 1 – Visualization and output of energy minimization.

## This practical will be divided into two parts:

1. We will see together how to use VMD (~1h)
2. We will look at MD output files that I have generated, to compare different minimization algorithm.

## Part I: Visualization training.



First, make sure all the softwares are installed. Open a terminal and write

```
vmd
```

If it is not installed, download the file “install\_softwares\_LINUX.sh” on mittuib (folder “Misc”) and run it with the command



```
bash install_softwares_LINUX.sh
```

You will find a short tutorial of VMD made by Dr. Robert Johnson from the University of Pennsylvania on the “Misc” folder on mittuib as well.

If you are interested, you can also find a Pymol Tutorial on this same folder.

The idea of this session is just to make you practice VMD (and/or Pymol) on a protein: 1RVA: a protein dimer bound with a double DNA strand.

The idea of the exercise is to use VMD and/or Pymol :

- Observe what are the secondary structures.
- If the protein is “alone” (is there any other atoms / molecules that do not belong to the protein?)
- Generate a suitable picture for each topology (protein+molecule / dimer+RNA).

First: download each protein on the Protein DataBank (<https://www.rcsb.org/>) with the 4 letters code (2H4G, 1RVA) and then start practicing by changing the visualization/colour/selection...

## Part II: MD Output files

For all practicals, the MD software GROMACS (<http://www.gromacs.org/>) will be used for computation, VMD (<https://www.ks.uiuc.edu/Research/vmd/>) to visualize trajectories. You can also use PYMOL (<https://pymol.org/2/>) to visualize single structures (if you know already Pymol..)

We have prepared 4 different minimizations for a modified version of the hen egg-white Lysozyme (PDB ID: **1AKI**), a single protein made of 129 amino acids. This protein is a very good start for beginners because it is an easy system (not too big, only natural amino acids, monomeric) and it was well studied experimentally, thereupon this protein is often used to validate force field with experiments.

### A. Visualisation - Let's meet the system

It is important to know your simulation system (ie the protrein you simulate) before running any MD. This part is about to check the original egg-white lysozyme structure (PDB ID 1AKI) and observe differences with our modified version of this protein.



1. Download the egg white lysozyme on the PDB website and visualize it with your favorite molecular viewer. Change the representation to Cartoon. How many  $\alpha$ -helices and  $\beta$ -sheet there is?



2. Load our modified version of 1AKI: (1aki\_clash.pdb). Keep the representation as "lines" (to see bonds between atoms).
  - a. Do you see something strange (what seems to happen)?
  
  
  
  
  
  
  
  
  
  
  - b. What are the amino acids involved in this artifact?
  
  
  
  
  
  
  
  
  
  
  - c. Do you think the bonds are « real » ?



3. Load now `1aki_clash_conect.pdb` on Pymol and visualize it as '*lines*' or '*stick*'.
  - a. Is the artifact present in (2) is still here?

- b. Is the distance between the atoms the same?

- c. According to you, why the artifact disappeared (check the differences between the two PDB files with a text editor)?



4. Now open `1AKI_processed.gro`, `1AKI_solv.gro` and `1AKI_solv_ions.gro`. What are the differences between those 3 files (number of atoms, type of molecule...)?



5. According to you, which will be the file used for the minimization?

## B. Hands-on MD output files.

As you may saw, we introduced steric clashes in the original 1AKI PDB file. Sometime when you work on compound modeling, you may introduce strong clashes and you must know how to remove them. This practice will focus on analyzing outputs generated by GROMACS and in the next practical you will learn how to run an MD!

You will find 4 folders:

- **CG**
- **CGSTEEP**
- **STEEP**
- **L-BFGS**

And inside those folders 4 types of file (excepts for **CG**)

- .log → Log file containing all the information at each step of the minimization
- .edr → Energy file containing the energy (and other term values) at each step of the minimization
- .xtc → Trajectory file, containing the  $\begin{cases} x \\ y \\ z \end{cases}$  coordinates of each atom at each step of the minimization
- .pdb → PDB file of the topology.

**Note:** only the protein is kept for this practical.



1. Open the log file of each minimization methods check:
  - a. If the minimization has converged (if it has succeeded)
  - b. The time needed of the minimization (check the starting and ending hour)
  - c. The number of steps
  - d. The efficiency (the time per steps)
  - e. The final Potential Energy

Algorithm	Start	End	Time	Convergence	Steps	Efficiency (Time per step)	Epot_final
CG							
CG+STEEP							
STEEPEST							
L-BFGS							

2. In the log file, you can have all the energy terms for each minimization step.
- What is the Unit of energy?
  - How many terms do you have?
  - Do you recognize some terms (from the lectures)? Which ones ?
  - Do you find a mathematical relationship between all terms?
  - Which are the most important term (higher value)?

3. Now open the trajectories with VMD and Look the 2 amino acids clashing. Describe what you see (if the transitions are smooth, or with a lot of shaking, does it take time to reach the equilibrium,...?)

During the MD, temperature, pressure, energy and other terms are saved in an “EDR” file. Use “`gmx energy`” to extract the potential energy with the bonds, dihedral (proper and improper) LJ, Coulomb...

You can select several term by giving the number to gromacs before finishing with “0” (example 1 2 3 4 5 6 7 8 9 0)

4. Look at those graphics and:
- Describe what you see in the first 10 minimization steps.
  - One of the minimizations failed (CG), What is the difference between the CG and CG+STEEP?