

PCLAB 3 – MD Analysis

In this practical, you will learn how to analyze data produced by MD. We will use a 20ns molecular dynamics simulation trajectory generated from a mutated helix where a GLN was replaced by a PRO (*see practical 2*).

The softwares used will be:

- Gromacs analysis tools
- Xmgrace: software to open gromacs output graphics. For Windows and Mac user download QTGrace (<https://sourceforge.net/projects/qtgrace/>)

Files

Because of space limitation, not all the MD outputs are given in the archive.

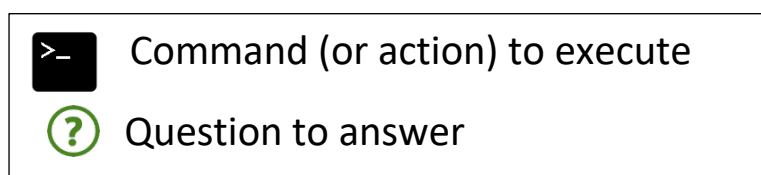
On the folder “**result**” you will find 4 folders:

- mini: output files from minimization
- nvt: output files from NVT equilibration
- npt: output files from NPT equilibration
- prod: output files from production (20ns)

info on file types:

- edr: contains the energy/pressure/temperature data during the MD.
- trr: trajectory file (text)
- xtc: trajectory file (binary and compressed)
- gro: structure in GROMACS format. Generally, it is the last frame
- tpr: Topology Portable Run file (contains all information about topology)

0. Before starting



Notes:

1. There is no breakline in commands line (commands are on 1 line)
2. There is no space between the “-“ and the argument letter (“- o” doesn’t work, “- o” works!)

1. Analysis convergence before production

During the previous practical, because of time, we didn't check if our system converged during minimization and equilibrations. This is a very important step: you have to check if your system has converged before running the next MD phase. So we will do that now 😊

1.1

A. Minimization:

For minimization, you can check if the system has converged by looking at the potential energy contained in the .edr file.

1. Go into the "mini" folder
2. Run this command:

```
>- gmx energy -s em.tpr -f em.edr -o potential.xvg
```
3. Choose the potential energy (write "11 0" then press 'enter')
4. Open the graphics with

```
>- xmgrace potential.xvg
```

❓ Q1 – Did the potential energy converged?

B. NVT Equilibration

Since in NVT equilibration we add velocities to the system by increasing the temperature, we want to be sure that we converge to our desired temperature (300K here)

1. Go into the "nvt" folder
2. Run this command:

```
>- gmx energy -s nvt_300.tpr -f nvt_300.edr -o temperature.xvg
```
3. Select the temperature: "16 0" then press Enter.
4. Open the graphics with

```
>- xmgrace temperature.xvg
```

❓ Q2 – Question on NVT Equilibration

- a. From the terminal output: What is the average temperature of the system?

Open the xvg file with xmgrace (xmgrace temperature.xvg).

- b. What is the temperature variation (minimum/maximum, approximately)?

- c. Do you think that the temperature is stable?

C. NPT Equilibration

Same as for NVT, we need to check the convergence of the pressure.

1. Go into the “npt” folder

2. Run this command:

```
>- gmx energy -s npt_ab.tpr -f npt_ab.edr -o pressure.xvg
```

3. Select the pressure: “17 0” then press Enter.

4. Open the graphics with

```
>- xmgrace pressure.xvg
```

Q3 – Question on NVT Equilibration

a. From the terminal output: What is the average pressure of the system?

Open the xvg file with xmgrace (xmgrace pressure.xvg).

b. What is the pressure variation (minimum/maximum, approximately)?

c. Do you think the pressure has converged ?

5. Analyzing Production

5.1 Cleaning trajectory

Files:

- md_helixPRO_prod.tpr
- md_helixPRO_prod_10ns.xtc
- md_helixPRO_prod.gro

The first thing you can do when the simulation is over, it's to visualize your MD.

1. Go into the "prod" folder.
2. Open the trajectory with vmd (md_helixPRO_prod.gro is the topology)

```
>_ vmd md_helixPRO_prod.gro md_helixPRO_prod_dt20.xtc
```

3. Write in the terminal "pbc box" to draw the simulation box
4. Show only the protein
Graphics (menu) → Representation → erase 'all' in the selection and write "not waters"
5. Play the trajectory (▶ icon in main windows, bottom right)

? Q4. Production:

- a. How is the helix during the trajectory (full, cut...?)

During the MD, your protein can "move". It is important to "wrap" the atoms together, to make the molecule "full". You can do that with gromacs tools. trjconv is a tool used to align, wrap, cut the molecule.

Close your terminal and write the command:

```
>_ gmx trjconv -s md_helixPRO_prod.tpr -f  
md_helixPRO_prod_dt20.xtc -o traj_wrapped.xtc -pbc mol
```

Gromacs will ask what do you want to save as output, press 0 then Enter to save the full system (with water)

If you visualize your trajectory (optional), you will see that the helix "jump" from one side to the other side of the box. We need to center the molecule inside the box and align the molecule to itself.

```
>_ gmx trjconv -s md_helixPRO_prod.tpr -f traj_wrapped.xtc  
-o traj_wrapped_aligned.xtc -fit rot+trans -center
```

GROMACS will ask 3 things:

- on what you want to align the system, choose backbone (4 then enter)
- then on what you want to center the simulation box, choose a protein (1 then enter)
- and finally, the output group, choose all the system (0 then enter)

One last thing, we will not use the water here, we can create a smaller trajectory file by removing the water molecules.

- ```
>_ 1. gmx trjconv -s md_helixPRO_prod.tpr -f
 traj_wrapped_aligned.xtc -o traj_wrapped_aligned_protein.xtc
```
2. write 1 then enter to select only the protein when prompted.

It is not finished... yet... To have a topology file that corresponds to your trajectory without water and ions, you need to generate a new topology file (tpr for gromacs, pdb or gro for VMD)

PDB file for VMD (or other softwares). Note: -b 0 -e 0 means that you start the trajectory at 0 and stop at 0 (only the first frame will be saved)

- ```
>_ 1. gmx trjconv -s md_helixPRO_prod.tpr -f
    traj_wrapped_aligned.xtc -o topology.pdb -b 0 -e 0
```
2. choose the **protein**

TPR file for VMD (or other softwares)

- ```
>_ 1. gmx convert-tpr -s md_helixPRO_prod.tpr -o protein.tpr
```
2. choose the **protein**

## 2.2 Analysis

### 2.2.1 - RMSD

1. Generate the RMSD of the helix simulation:

```
>_ gmx rms -s protein.tpr -f traj_wrapped_aligned_protein.xtc -o
 rmsd_protein.svg -tu ns
```

2. Choose the **backbone** for alignment **AND** calculation
3. Open the output file with xmgrace (xmgrace rmsd\_protein.svg)

#### Q5. RMSD

- b. has the system converged?

- c. From the curve, can you describe how the system behaves?

### 2.2.2 - RMSF

1. Generate the RMSF of the helix:

```
>- gmx rmsf -s protein.tpr -f traj_wrapped_aligned_protein.xtc -o rmsf_protein.xvg -res
```

2. Choose the **backbone**
3. Open the output file with xmgrace (xmgrace rmsf\_protein.xvg)

#### Q6. RMSF

- a. which amino acids move the most?

- b. *Remember, the mutation was in position 12.* Do the amino acids around this position fluctuate?

### 2.2.3 - H-BONDS

The structure of an alpha helix is maintained by a network of hydrogen bonds between the carboxyl oxygen of the backbone and amino group of 4<sup>th</sup> next amino acid (**n to n+4**)

We can follow the number of hydrogen bonds during the dynamic with the gromacs' H-bonds software.

Arguments:

- tu: time unit = ns
- num: total number of h-bonds
- hx: check only for n to n+1 to n+6

```
1. gmx hbond -s protein.tpr -f traj_wrapped_aligned_protein.xtc
>- -tu ns -num hbonds_num.xvg -hx hbonds_helix.xvg
```

2. choose the "mainchain+H" (7) for the two groups

3. Open the output with

```
>- xmgrace -nxy hbonds_helix.xvg
```

nxy is an argument to read a file that contains multiple variables, here we have n to n+1, n+2, n+3, n+4, n+5, n+6

- #### Q7: Only look the *n,n+4* H-bonds. Describe how many H-bonds are formed over the 20ns MD (do we gain interactions or do we lose them? What are the maximum and minimum values ?)

### 2.2.4 - SECONDARY STRUCTURE

Finally, we can check the secondary structure over time and see if the helix break and where. DSSP is a software (pre-installed on your computers) that calculate the secondary structure for proteins. Gromacs can be linked to DSSP and calculate the secondary structure on a trajectory and produce a secondary structure map for every amino acid over time. You need to execute those 3 commands :

1. `export DSSP=/usr/bin/dssp`  
*this tell to gromacs where is the DSSP software*
2. `gmx do_dssp -s protein.tpr -f traj_wrapped_aligned_protein.xtc -o helix_ss.xpm -ver 3 -tu ns -dt 0.05`  
*- ver is for the DSSP version, use 2 for the old version*  
*- tu: time unit in ns.*  
*- dt: timestep for SS calculation (every 0.05ns so every 50ps)*
3. Choose the protein
4. `gmx xpm2ps -f helix_ss.xpm -o helix_ss.eps -by 10 -bx 3`  
*convert the output xpm map to a readable file with legends, axis, title..... by and bx are the scale of the axis to have a "readable" figure.*

Q8. Secondary structure:

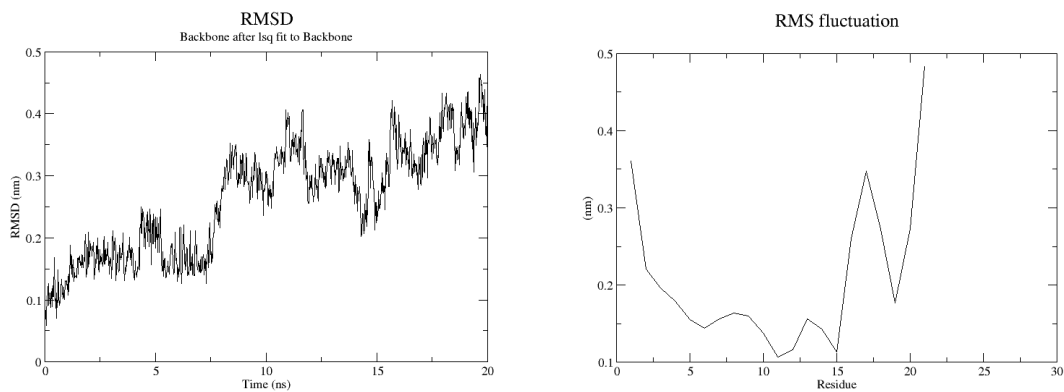
a. from when the helix "break"?

b. On which position the helix break?

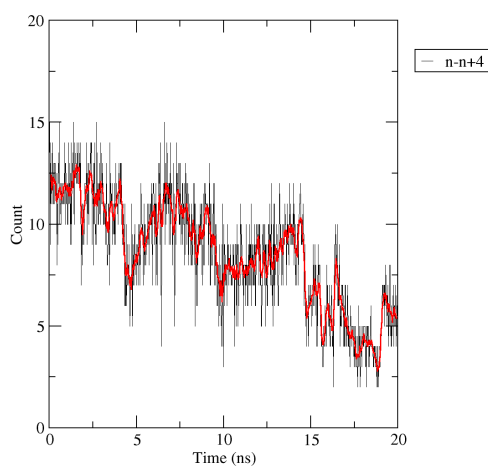
c. Regarding the position of the mutation (12), does it make sense to you?

## 4. Comparison with Wild Type (without mutation)

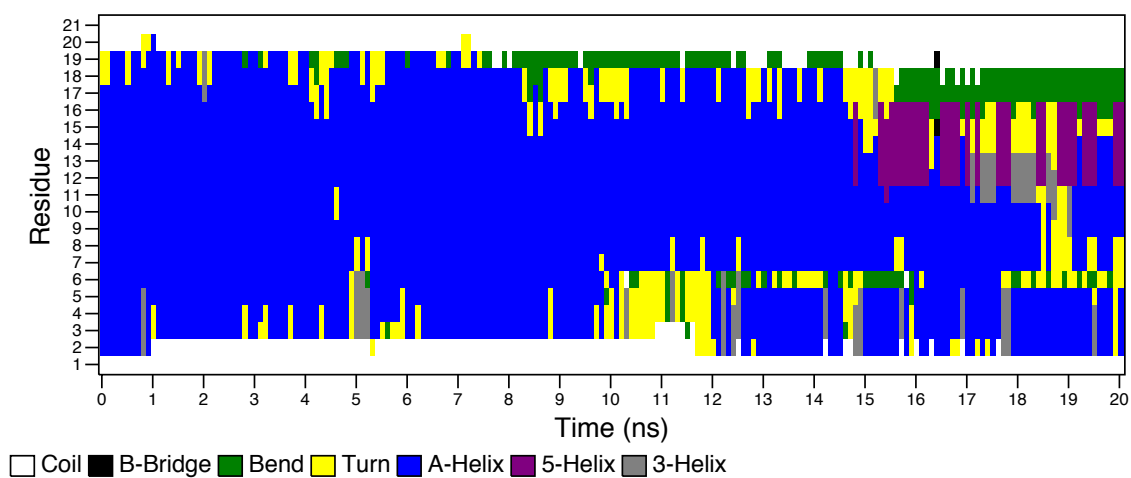
Just for comparison, here are the same results with the helix without the mutation.



Hydrogen Bonds



Secondary structure





Q9. Check the RMSD, RMSF, the number of H-BONDS for n+4 and the secondary structure map.

- a. Is the wild type helix (*ie* the helix without mutation) more stable?
- b. From all the analysis you did, can you explain why the proline is called an Alpha Helix breaker?