

# Programming for Data Management

## Introduction to Linux, Bash, and Python

Seminar



Dr. Joel Mieres-Perez  
Dr. Yasser Almeida-Hernández  
Lehrstuhl für Computational Bioengineering  
[professors.cbe.bci@tu-dortmund.de](mailto:professors.cbe.bci@tu-dortmund.de)

1. Create a main bash script that control all the steps. Create all additional necessary scripts for the rest of the tasks.
2. Create folders run1, run2, and run3, and move the files to their respective folders. Add checks and warnings to tell the user if the folders already exist. If the folders exist, ask the users if wants to be overwritten.
3. Write the files with the first line that contains the date, but without the rest of the lines starting with “#” or “@”.
4. For every run folder, create 2 files that contains one X column (Time (ps)) and all the forces, and the same for all the distances:  
# Time (ps) TAB frameA TAB frameB TAB frameC  
... TAB ... TAB ... TAB ...
5. For each run, create files that contains the average force and distance with their standart deviation for each frame:  
# Frame TAB Distance (nm) TAB STD TAB Force (kJ/mol/nm) TAB STD  
frameA TAB ... TAB ... TAB ... TAB ...  
...
6. Use the previous files to create a plot of Distance  $\pm$  SD vs. Force  $\pm$  SD for each run. Use a proper labeling for the plots.
7. With the files from step 3, create plots of superimposed frequency histograms of the distance and the force for all the frames.

8. Write a report file with the following header and information:

```
# REPORT
# Created on <DATE> by <NAME>
#
# Run TAB Force variation(kJ/mol/nm) TAB Distance variation(nm)
Run_1          ...          ...
Run_2
Run_3
```

\*You can calculate the Force and Distance variation from the data in steps 5 and 6.

9. Write information about each step in a log file.

10. Comment/document the scripts appropriately.