Project plan

Project title:

Implementing the string method with swarms of trajectories using the newly developed Gromacs API

February 2020

Student: Marko Petrović

Supervisor: Prof.dr. Lucie Delemotte

Co-Supervisor: PhD. candidate Oliver Fleetwood

Version: 1

Contents

1 Introduction

Welcome to the plan specifying the work on my masters thesis: *Implementing the string method with swarms of trajectories using the newly developed Gromacs API*. Where an application will be constructed in the Python3 programming language to implement the string method and provide an easier to use user interface for researchers.

2 Objectives

The main objective is to construct an easy to use interface for the Gromacs molecular dynamics simulation software, focused on the string method.

2.1 Setup

Install Gromacs and it's API locally on my personal computer. Get simulations to run locally and on calculation cluster using Gromacs API

2.2 Create implementation

Construct an implementation capable of iterating through the running of multiple simulations and doing subsequent refinement calculations for the following iteration.

2.3 Create user interface

2.3.1 Results presentation

My implementation should be able to create visualisations of collective variables for use in articles and reports.

2.3.2 User experience

This might not actually be an objective but rather a goal. The implementation should have a graphical user interface for setting up and initiating simulation runs.

3 Background

Simulating molecular and atomic systems at atomic resolution is a practice known as molecular dynamics (MD). A MD simulation is run one time step at a time. In each time step, the forces acting on each atom is calculated and used for calculating the atoms acceleration and subsequent movement for that time step. As the relative speed of the atoms and molecules is quite large compared to both the volume of the simulated system and distance between atoms and molecules, each time step is limited to 1-2fs. MD can be used for very large systems, beyond 50000 atoms. The combination of very large systems and very short time steps results in huge amounts of calculations for simulations spanning a only few μs setting very narrow limits of which biological processes can be simulated.

Trying to use conventional MD simulations to analyse conformational changes in proteins and other macro-molecules would not be feasible, *"Conformational changes in large biomolecules are complex and slow processes taking place on timescales that are beyond the reach of brute force molecular dynamics simulations."*[1] The string method with swarms of trajectories can be used to alleviate the calculational load by setting a best guess path from one state to the next. Using snapshots equidistantly placed along that path as virtual initial states (VIS) the string method runs several short MD simulations for each VIS, a swarm of trajectories. The assumption made is that the general drift in each swarm would be towards a conformation of lower energy. Each VIS is updated according to a mean of its swarm of trajectories and then also reparametrised to keep the VISs equidistant along the path.[1]

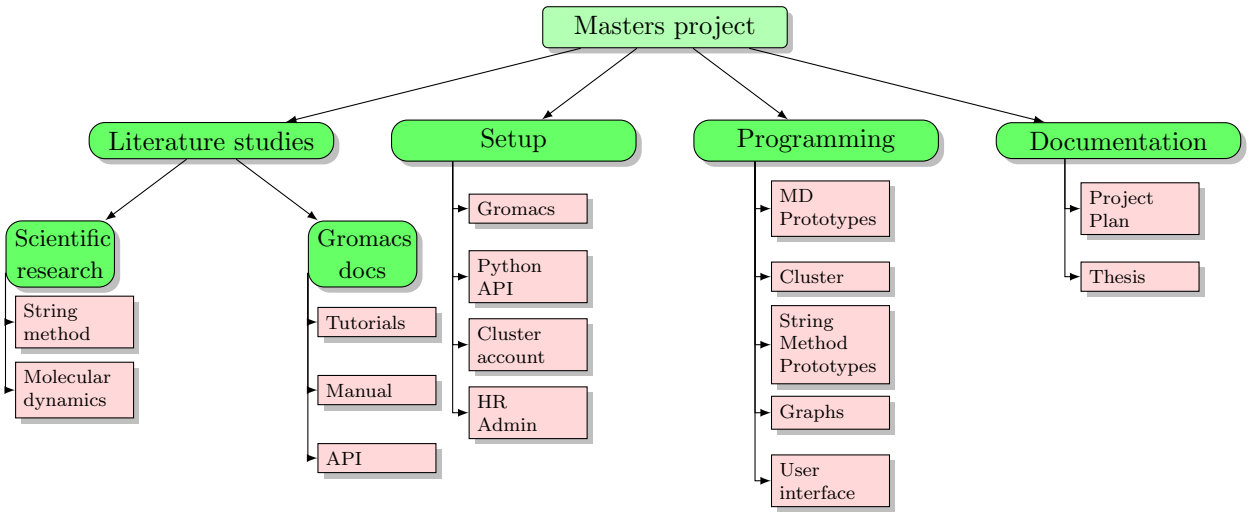
Gromacs is mainly a software package created to run one singular simulation at a time. In the string method, many iterations of many simulations need to be run with some calculations between iterations in

order to set the initial conditions for the next iteration. A lot of the calculations are similar or even the same and can easily be automated. With a good structure of automation the researcher workload can be focused not on repetitive tasks but rather interpretations and analysis.

Computer software ranges from games, through creative aids to scientific calculations and record keeping. As novel scientific software often is created by the researcher in need of it, the user interface is commonly kept as simple as possible. In the case of Gromacs, the main user interface is through the command line. Over time, Gromacs has grown and evolved updating it's calculation algorithm but also adding an application programmers interface (API). As Gromacs has evolved and grown, so has it's user base. With a wider user base, a command line interface for the program (and usually needing scripting skills for the operating system) creates a skill barrier for new users. The end result of one MD simulation is a trajectory file, this file then needs interpreting. Depending on the purpose of the simulation, the interpreting can be a 3D visualised model or movie, an atom to atom distance or binding angles.

In order to automate the iterations of MD runs and VIS updates a software needs to be created. If this software can lower the computer skill threshold for users and also automate some of the interpretations of simulation, the process of researching protein state transitions using MD will be significantly simplified.

4 Work breakdown structure



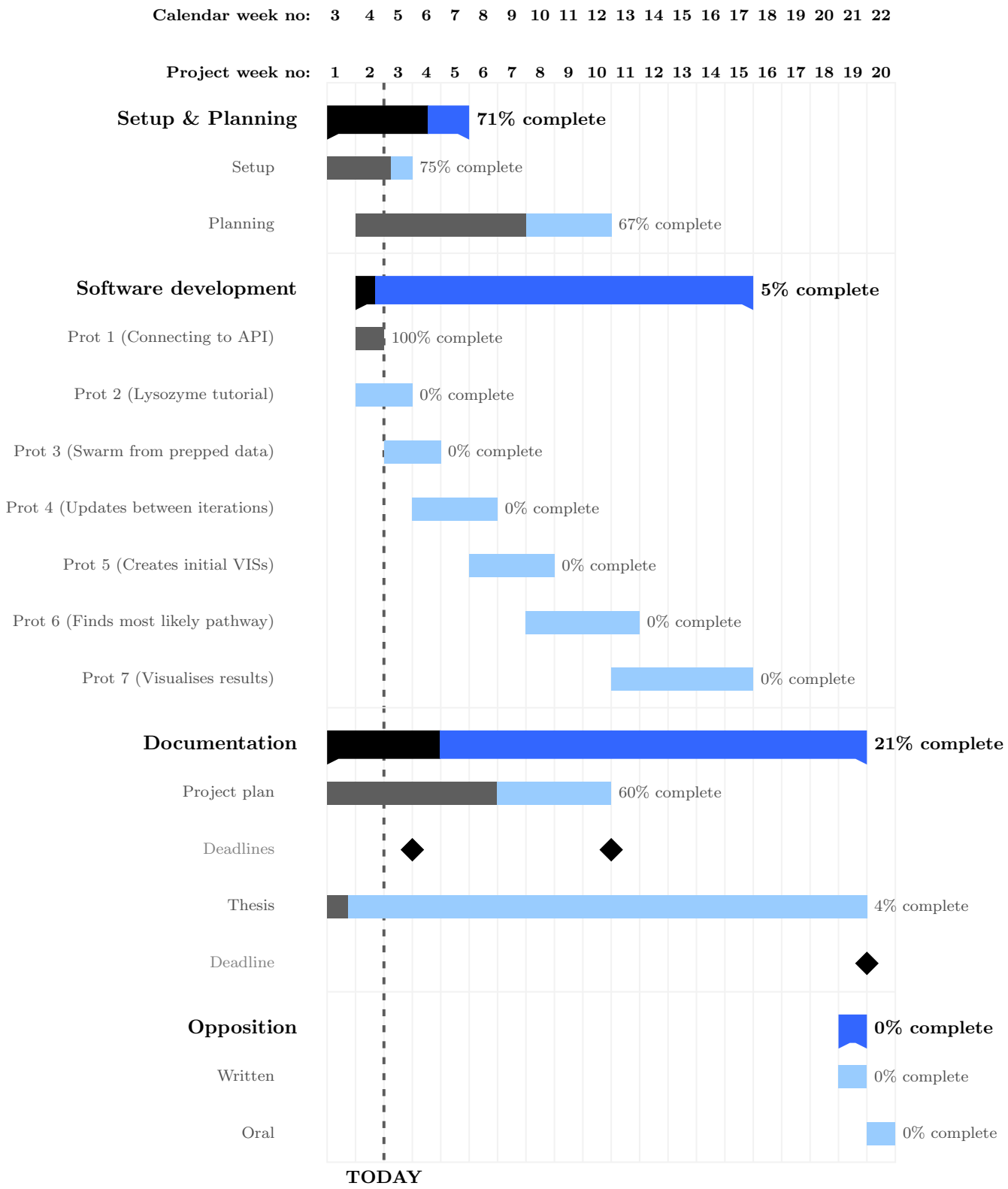
5 Milestones

The milestones of the project in intended order of completion:

1. Prepared Gromacs simulations can be launched from Python.
2. The entire Lysozyme Gromacs tutorial is automated from a Python program.
3. Initial project plan finished.
4. Running MD simulations on cluster.
5. Given a set path of initial VISs, the software can do one iteration of MD simulations.
6. The software can use the results from one iteration of the String method to prepare the next.
7. With hardcoded collective variables, a path of VISs is calculated.
8. The software can iteratively improve upon a path to produce a final most likely transition pathway.
9. The software can set up and run a complete transition pathway generation from known start and end states, and preselected collective variables.

10. The software can create graphs of collective variable and transition path iterations.

6 Time plan


Calendar week no: 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Project week no: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Setup & Planning
71% complete
Setup
75% complete
Planning
67% complete
Software development
5% complete
Prot 1 (Connecting to API)
100% complete
Prot 2 (Lysozyme tutorial)
0% complete
Prot 3 (Swarm from prepped data)
0% complete
Prot 4 (Updates between iterations)
0% complete
Prot 5 (Creates initial VISs)
0% complete
Prot 6 (Finds most likely pathway)
0% complete
Prot 7 (Visualises results)
0% complete
Documentation
21% complete
Project plan
60% complete
Deadlines
Thesis
4% complete
Deadline
Opposition
0% complete
Written
0% complete
Oral
0% complete
TODAY

7 Specification

As the objective of the project is to construct software the specification will be focused on the software.

7.1 Must

- Run Gromacs MD simulations for the user.
- Implement the string method with swarms of trajectories using Gromacs and the Gromacs API.
- Track and log the collective variables through a series of iterations of the string method.
- Update/create new topology files between runs.
- Work on both personal(UNIX/LINUX) computer and calculation cluster.

7.2 Should

- Prepare for the string method from pdb files and defined collective variables.
- Visualise the pathway in the collective variable space (graphs).

7.3 Could

- Identify collective variables.
- Automate 3D visualisations.
- Record movies.
- Assist in selecting MD parameters.

7.4 Won't

What will be left for future endeavors is not yet decided.

8 Stakeholder analysis

9 Business case

10 SWOT analysis

11 Risk evaluation assessment

12 Summary

References

[1] Pan, A.C., Sezer, D. and Roux, B. 2008. Finding transition pathways using the string method with swarms of trajectories. The Journal of Physical Chemistry. B 112(11), pp. 3432–3440.