MD05 - 5 Week Plan

Topic: Next-Gen Drug Discovery: Generative and Reinforcement Learning for Designing Anticancer Molecules

Week 1:

A master Github will be created with all of you as contributors, so please be thorough with all the version control commands using github and setup vs code if not done until now.

- Dataset collection or creation from scratch (1 Student)
- RL, LLM to be known to everyone
- As a part of exercise try to run the PaccMann RL model on a small dataset to show a reward based learning in action
- Document each and every thing **To Be Submitted (5 July) EOD.**

Week 2:

- 1 day for Week 1 pending work
- Design a pipeline for the end to end data input to the training with all the parameters and distribute the work amongst yourselves for each segment
- Identify the importance of each task and start working on it immediately as next steps might be awaiting results from the previous one
- Learn about creating a package for the functions and also decide what functionalities will you be providing by the end, make it exhaustive as not everything can be achieved.
- Finalise the loss functions, rewards functions and any optimisation techniques to be kept at back of mind while coming up with all these things
- **Submission for this week** will be a presentation containing all the above

Week 3:

- 1 day for any pending work
- Should be following up for the mathematical formulation of the functions decided above and research over any positive optimisations if can be achieved
- Try to learn about the Scaffold technique of molecule generation and try to implement(1 Student)
- Be ready with either of the code to generate molecules using Scaffold and previously decided loss function(Recommended) or the PaccMann RL agent reward based feature understanding.
- **Submission for this week** will be a methodology report comprising the architecture dependencies and follow up techniques for the mathematical modelling of the method and the result/demo for the running code

Week 4:

- Important week as we will be able to see something by the end of this week
- Rigorous coding in this week, 2-3 days for the model training code and learning outcomes
- Analyse the loss functionality and trends and pitfalls and resolve them in next 2-3 days
- **Submission for this week** will be a semi functional code, at least on a small dataset such that it can be used for questioning and peer reviews, for this you might also be asked to make a presentation or a report to document the functionality of the working project

Week 5:(Can be extended)

- Work on any pending presentations and report
- Review carefully the working issues, analyse the peer reviews and try to improve the functionality
- Finally we will be compiling the code into a public library or package and make a documentation for the provided use cases and essential function arguments.
- **Submission for this week** might be your acceptance to come up for a full fledged treat (based on Anirban and Sir's approval 😀)