Max Planck Institute
of Colloids and Interfaces

MSC PROJECT: MACHINE LEARNING FOR BIOMOLECULAR SIMULATIONS

BACKGROUND: PHYSICS, COMPUTER SCIENCE, BIOINFORMATICS, ENGINEERING, MATHEMATICS

This project explores the potential of neural networks in building biomolecular models for molecular dynamics simulations.

At its best, molecular dynamic (MD) simulations provide realistic 3D videos (with atom-level resolution) on the functioning of biomolecules. This makes MD an impactful tool in biosciences, used by thousands of scientists and facilitating discoveries like the Tamiflu drug resistance of the H1N1 Swine Flu. At the core of every MD simulation is a force field—a model used to describe the interactions between the atoms. The correctness of the MD simulation relies entirely on the performance of the chosen force field.

Unfortunately, the current force fields are somewhat flawed, and their improvement is held back by the complexity of the problem and old-fashioned approaches. Development of more efficient, high-throughput tools for building biomolecular force fields is desperately needed.

In this project, you will achieve this by harnessing the methods used in big data for biomolecular model development.

A force field is generically defined through a set of parameters: a "parameter vector". Similarly, the biomolecular structures resulting from the MD simulation can be described as a "result vector". Your starting point will be to treat the problem by using a neural network that connects the "result vector" to the "parameter vector". Once trained, the network will be used to predict the "parameter vector" that is able to reproduce our experimentally measured structural data. Based on the performance of this initial approach, other machine learning methods may also be explored to improve accuracy and reduce computational cost.

Initially, you will work on lipid systems—an important class of biomolecules that play a central role in cell structure and signaling. The approach, however, is generalizable to any biomolecule.

You should have solid programming skills, strong interest in model building, and an enthusiastic attitude towards learning. Good working knowledge of mathematics, and experience in biomolecular simulations or machine learning will be considered a plus. Suitable backgrounds for the project include physics, engineering, bioinformatics, computer science, and mathematics. The working language will be English.

The project will be carried out in Max Planck Institute of Colloids and Interfaces as part of a joint collaboration between Dr. Markus Miettinen, Dr. Angelo Valleriani, and Dr. Hanne Antila. Interested candidates should send a CV, a letter of motivation, and a transcript of university record via email to hanne.antila@mpikg.mpg.de by Oct. 1st, 2019. Please include "*Machine learning for biomolecular simulations*" in the email title.

The tentative starting date of the project is 15.11.2019.