CDCS
CENTER FOR DATA AND COMPUTING
IN NATURAL SCIENCES

OPENING SYMPOSIUM 2022

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DESY.

TUHH

Contribution ID: **59** Type: **Poster**

Dynamic structure investigation and spectra prediction of near edge X-ray absorption spectroscopy fine structure (NEXAFS) implementing supervised and unsupervised machine learning techniques

The phenomenal growth of computing capabilities have accelerated the ability to combine chemistry, physics and Machine Learning (ML), as a true symbiosis, so as to precisely model and understand complex biomolecular processes at the atomistic scale. However, complexities of proteins and high computational costs of quantum mechanics methods for large systems impose a great challenge in obtaining insight into inherent properties of these biomolecules. To overcome this challenge, including data-driven ML models into the simulator's toolbox (molecular dynamics (MD) simulations) ease the path to perform large-scale simulations and understand the complex interplay of interatomic and intermolecular interactions. Examples of this approach can broadly fall into two categories, namely unsupervised and supervised ML techniques. In unsupervised ML, the conformational space of the biomolecules is explored based on estimating the probability distribution of relevant variables, describing the data of MD simulations. For instance, clustering and dimensionality reduction techniques are of particular interest for exploring datasets which are too high-dimensional to be visualised and understood. A supervised ML approach could be used to predict specific properties of the biomolecules based on so-called user-defined collective variables or reaction coordinates in which the simplest case is to investigate the atomization energy of the system or the force acting on the atoms during the simulation. In this project, we have implemented intricate replica exchange molecular dynamics (REMD) on different peptides in order to thoroughly screen the potential energy surface (PES). Unsupervised ML are applied to predict the spectroscopic data, e. g. infrared(IR) and X-ray absorption spectra. Moreover, we used supervised ML, such as graph neural networks (GNN) to predict more sophisticated chemical properties to explain the interplay between the biomolecules'conformations and spectroscopic data.

Primary authors: KOTOBI, Amir (FS-SCS (Strukturdynamik Chemischer Systeme)); Dr MEISSNER, Robert (amburg University of Technology, Institute of Advanced Ceramics); Dr BARI, Sadia (DESY)

Presenter: KOTOBI, Amir (FS-SCS (Strukturdynamik Chemischer Systeme))

Session Classification: Poster session with buffet

Track Classification: CDL2 (Photon Science)