LOOPER: Machine learning for druggable protein conformations

Dr Antonia Mey, University of Edinburgh, EaStCHEM School of Chemistry
Dr Matteo Degiacomi, Durham University, Department of Physics

Any organism, from a simple bacterium up to a human body, contains millions of proteins with roles as diverse as sensing, transport, offence, motility or structural support. Proteins are biopolymers carrying out their biological task by folding in specific three-dimensional shapes, which in turn determine their capacity of interacting with designated binding partners such as DNA, drugs or other proteins. While these interactions are key for life as we know it, they are also associated with diseases and disorders. Indeed, viruses or bacteria use proteins for their reproduction, while misfolded proteins and their aggregates can lead to complex diseases. Given the crucial link between proteins' structure and their function, X-ray crystallography or cryo-electron microscopy techniques are often applied to unveil their structure at the atomic scale, most recently aided by AlphaFold2. To date, >180k structures have been obtained and made freely available online on the Protein Data Bank. Structural information is crucial in small molecule drug design, however, the drug discovery process is often complicated by the fact that most proteins are flexible, and many feature highly dynamic regions called intrinsically disordered regions (IRDs). These regions play a central role in modulating their interactions.

This project aims at combining molecular dynamics (MD) simulation and machine learning (ML) techniques to predict how proteins move, and to improve the way drug-like molecules can target flexible protein regions. This builds upon the expertise of the Degiacomi group on generative neural networks (GNNs) and that of the Mey group in method development for MD simulations and drug design. The student will address the questions of: can a GNN be used for generating biologically relevant protein conformations? How can the ensemble of structures then be used for improved design of drug-like molecules that specifically target their IDRs? These questions are central to understanding neurodegenerative diseases such as Parkison's and Alzheimer's, as well as the molecular mechanisms behind antimicrobial resistance. A three-month placement at Redesign Science (www.redesignscience.com) will enable the student to apply their acquired MD and ML techniques to get insights into an active drug discovery project, and learn how to streamline their work in accordance with industry timescales.