Deep Learning Applied for 3D Molecular Interactions

Project Overview:

Molecules constantly interact with each other, maintaining integrity and functionality of living organisms. Broken molecular interactions typically lead to serious diseases, such as cardiovascular, neurodegenerative diseases, cancer, and others. Nowadays, there is a great need in more efficient, specific and safer drugs that modulate molecular function via atomic interaction. In order to design a chemical with desired modulation activity one needs to known molecular interactions in atomic details. Experimental techniques, such as X-ray crystallography, NMR, and Cryo-electron microscopy are used to determine spatial structures of molecular complexes on sub-nanometer scale and, hence, allow to look at molecular interactions in atomic details. This information is typically used as the starting point for virtual ligand screening (VLS), where drug candidates from the large chemical libraries (~10^6-10^8 chemical compounds) are docked to the target molecule one by one, and then top ~10^3-10^4 chemicals are selected as the most promising ones. The success rate of VLS critically depends on the scoring function. The ideal scoring function determines the value of the binding affinity of the chemical compound and the pharmaceutical target, based on the atomic structure of the molecular complex. There are already powerful methods for modelling of molecular complexes, but the existing scoring functions in general poorly correlate with the binding affinity. With the growing number of structural information about molecular complexes and powerful computational resources, statistical scoring functions developed with machine learning, become superior, as compared to the classical scoring functions. This project aims to develop new generation of scoring functions for computer-aided drug discovery using state-of-the-art machine learning methods.

PhD Objectives:

A PhD candidate will start with pruning and processing of protein and chemical databases using molecular modelling techniques in order to prepare the training and test sets. These include data analysis, homology modelling, molecular docking, conformational ensemble generation. As the next step, we will develop efficient voxel representation of the collected molecular sets and use the obtained representation as the input for convolutional neural networks (CNN). Different voxel channel attributes as well as different CNN architectures will be examined in order to solve two important problems in computational chemistry: 1) given a set of molecular complexes, identify the native one 2) given a receptor and a set of ligands, estimate affinity of each ligand to this receptor. The developed methods will be extensively benchmarked on the well established protein-ligand benchmarks and will be rigorously applied in world-wide docking and scoring competitions. Finally, the developed methods shall be applied to the relevant pharmacological targets, such as G protein-coupled receptors, in order to identify novel drug candidates.

Requirements:

We are looking for creative, passionate and hard-working graduates from all over the world - whether you hold a M.Sc. degree in computer science, computational biology, chemistry, physics, or a related field, as long as you have a passion for working at the interface of data science and biology, this may be the place for you.
The ideal candidate has i) solid background in molecular modelling of protein complexes or computational chemistry, ii) programming experience in C/C++/Python, iii) experience in machine

learning and data analysis, and iv) excellent communication skills in English, and should be highly motivated to develop state-of-the-art software and conduct research for open challenges in molecular modelling and computer-aided drug discovery.

Contact:

Please send your application with the keyword ‘PhD application’ as one single file including a CV, a motivation letter, copies of university transcripts (BSc and MSc, in Russian or English), and contact addresses of two academic references by email to p.popov@skoltech.ru.

Research Environment:

PhD project will be supervised by Petr Popov at Skolkovo Institute of Science and Technology in the Center for Computational and Data-Intensive Science and Engineering. PhD fellowship : 75000 rub.
Skolkovo Institute of Science and Technology is a private graduate research institute in Moscow, Russia. Established in 2011 in collaboration with MIT, Skoltech cultivates a new generation of researchers and entrepreneurs, promotes advanced scientific knowledge and fosters innovative technology to address critical issues facing Russia and the world. Skoltech applies the best Russian and international research and educational practices, with particular emphasis on entrepreneurship and innovation.
Skoltech Center for Computational and Data-Intensive Science and Engineering aims at the conduction of cross-cutting interdisciplinary research driven by modern applications in the fields of computational and data sciences. CDISE has accumulated several research groups with the expertise unique in Russia and competitive at the world level in their prospective areas.