# Combination of Molecular Simulation and Deep Learning for De Novo Drug Design

Drug discovery is a long and costly process that also oftentimes ends in failure, and its costs rise every year. Especially now as the world is still fighting the COVID-19 pandemic, medicinal chemistry could greatly benefit from the tools that would help to reduce costs of this process and shorten the path to the discovery of the new potent drugs. One of the ways to achieve that is to employ computational power to solve a number of drug discovery tasks on a computer. In particular, artificial intelligence has been successfully used for computer-aided drug design in many tasks, such as target identification, virtual screening, and molecular optimisation. However, solutions based on machine learning depend heavily on the data quality, which can be deceiving in the realm of pharmaceutics. Companies tend to report only successful compounds or do not publish their discoveries at all, and thus only small chunks of the vast chemical space can be explored by machine learning models. This is also the reason why these models are often criticised for either lacking novelty in the proposed molecules or creating unrealistic compounds that cannot be synthesised.

With the increasing amounts of chemical data emerges the need for methods that would screen the available libraries of chemical compounds in order to find molecules that interact well with the given target. To that end, molecular docking programs have been introduced, which make the tedious process of drug discovery more manageable. Docking software simulates the optimal pose of a molecule inside a binding pocket of the target. In particular, chemists use docking software to examine interactions between small molecules and proteins and make more insightful decisions when designing new drugs. Docking programs also facilitate the search of potentially active compounds in big libraries of chemical compounds, which is called a virtual screening. Recently, deep learning has been used to further accelerate virtual screening of enormous chemical datasets by approximating the output of a docking software using a faster neural network.

The main research objective of this project is to *explore the potential of molecular docking software for de novo drug discovery with the main focus on the quality of deep learning based methods.* More specifically, we propose a new high-level methodology for combining simulation with any generative model for *de novo* drug design or molecular optimisation. Several generative models will be implemented, and they will create new compounds that have improved molecular properties. To track these properties, predictive machine learning models will be also implemented. Then, we will investigate different methods of introducing the information from simulation into the models, e.g. training the models on the simulated data or optimising the models to generate molecules that achieve good results in simulation.

As the final effect of the project, a deep learning model that generates new compounds accordingly to the described methodology will be implemented. The implemented model should be useful in *de novo* drug design, i.e. creating novel drug candidates based on the biological target, but also in lead optimisation, which stands for the optimisation of existing compounds (lead compounds) to improve their physicochemical properties and identify drug candidates.