The Mu Opioid Receptor: Computational Tools for Ligand Efficiency & Discovery

Summary of Conducted Experiments:

Opioids have been long known for their properties as analgesics for the clinical treatment of severe pain, usually for providing positive reinforcements that alleviate the negative sensory and emotional aspects of postoperative and chronic pain. However, they present the drawbrack of drug dependence and analgesic tolerance (reduced pain-relieving effect). Previous studies on the class of opioid receptors, subtypes of the broader G protein-coupled receptor (GPCR) family, have led to the creation of new prescriptions which have the intended biological function with a minimized addiction potential through the use of partial agonists/antagonists or receptor-specific targeting.

However, the process of drug discovery is expensive, as well as time intensive, in a process often spanning over a decade. To expedite the process of searching for compounds that have the intended biological characteristics, machine learning techniques in drug discovery have been used to expedite the search for these desired "hit" compounds with the optimal protein-ligand interactions, using chemical attributes to train robust models. This study aims to focus on the application of a diverse set of Machine Learning Models and computational tools to explore the Mu opioid receptor and the binding affinities and efficacy of potential ligands to the receptor.

We first used a database of 10,000 commercial compounds from a ChEMBL Database of bioactive molecules, using their SMILES strings (an alphanumeric translation of a compound's three-dimensional chemical structure) to generate MACCS keys, a chemical fingerprint representing SMILES data as a large binary matrix for each compound. Data processing was used to filter out PAINS molecules, which are "false-positives" in the drug-discovery process that have no therapeutic potential. The Lipinski Rule of 5 was also applied to filter out compounds that did not have the intended drug-likeness (properties for oral and intestinal absorption). This included parameters for molecular mass, the number of hydrogen bond donors & acceptors, and the octanol-water partition coefficient value.

The remaining compounds were suitable to be used as data points for machine learning models. In addition to the generated chemical fingerprints, the ChEMBL file of compounds also contained the IC_{50} values of the compounds, which is a measure of the concentration of the compound required to inhibit a particular biological process by 50%. In this case, this value was a measure of the potency of each compound as an antagonist. This measured the performance of the compound in binding to the Mu opioid receptor and inhibiting the function of endogenous ligands (produced naturally in the body), associated with the body's pain response. For each compound, we took the negative logarithm of each IC_{50} value to obtain the pIC_{50} for better data processing.

Using the unrolled chemical fingerprint binary matrix (into a single vector) of each compound as input features, we trained the following models: Decision Tree, Random Forest, Gradient Boosting, Support Vector Machines, Neural Networks, K-Nearest Neighbors (KNN), Ridge Regression, Lasso Regression, and XGBoost (using randomly defined test, cross-validation, and training subsets of the data). We originally used a threshold on the pIC_{50} values of the compound to create a binary output variable for the machine learning models to predict. However, we later changed the models to predict a continuous value (pIC_{50} value) rather than confining it to a binary result. This change increased the model accuracies by over 30%. We evaluated results using ROC curves that took into account F1 scores (using precision

and recall model measurements), concluding the objective of predicting ligand efficiency (pIC_{50}) values based on their chemical structure (SMILES string).

We then discovered new biologically active ligands to the Mu opioid receptor using computational tools. Initially, we began by analyzing the crystal structure of an agonist (Protein Data Bank ID: BU72) interacting with the Mu receptor. We used the December 2020 version of the SMINA molecular docking tool to evaluate the binding affinity of this natural agonist to the Mu receptor. We then used a similarity search (with similarity threshold 75%) through PubChem to generate a list of 24 similar compounds (based on chemical substructures). We then used SMINA again to run a simulation docking each of the 24 compounds to the Mu receptor, finding the binding affinity values. Using the molecule with the highest binding affinity from this list of 24, we repeated the PubChem search to generate a new list of 100 similar compounds, again using SMINA to identify compounds with a higher affinity. Using the deep learning based molecular docking tool GNINA we obtained similar results. This branched similarity search identified a method for finding bioactive molecules that interacted with the Mu opioid receptor and exhibited drug-potential.

Our two-step research process utilized existing computational tools on the Mu opioid receptor to find new plausible “hit” compounds and to evaluate their potency as drug candidates. These findings can be furthered by machine learning models using more data, different activation functions and scoring functions, as well as a greater variety of input features than the MACCS chemical fingerprint. The similarity search in the second part of the study could also be run for more iterations to generate a more expansive list of results.