Brian Wiley

Final Project

AS.410.712.81.SU21 Advanced Practical Computer Concepts for Bioinformatics

**Protein blast to drug leads**

For this project I decided to choose a topic that would relate to my future endeavors in pursuing a PhD in computational chemistry or computer aided drug design (CADD) start next year. Proteins are the “workhorses” of all multi and single cellular organisms. They encompass and influence all biological processes. As such “the vast majority of drugs used in medicine are targeted to proteins” (Patrick 2013). In short, this web application allows for the user to insert a peptide either obtained from some biological assay with instruments such as a protein sequencer or a mass spectrometer as well as already existing curated proteins from SwissProt to the NCBI blastp program using the SwissProt database. From the hits the user can select accession IDs and pull all the associated DrugBank IDs.

A key aspect in the drug design process is finding drug leads for a target in which can be slightly altered many times by adding various substituents and functional groups (Figure 1). This allows for testing compounds in assays with the targets to improve the absorption, distribution, metabolism, and excretion (ADME), the toxicity levels including increasing the target specificity, or the efficacy by increasing the molecule affinity for the target. Another interesting part that of drug design that has resurfaced (pun to be intended soon) is not building and creating new drug compounds but discovering new targets for existing drugs. I once heard talk on cancer therapy, I forget if it was this one specifically, [New treatment horizon : Chimeric Antigen Receptor CAR T Cell Therapy](https://www.youtube.com/watch?v=gmjLSeAXAmw), but the presenter mentioned pharmaceutical research was proving new targets for “rediscovering” the use of aspirin (acetylsalicylic acid), a drug that is over 100 years old.

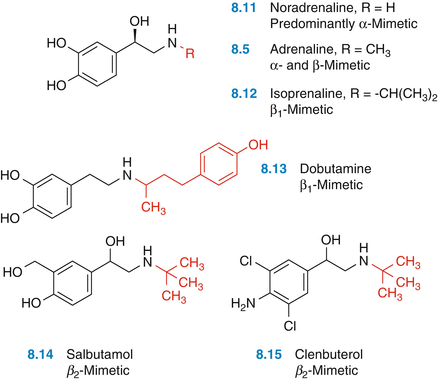


Figure 1. Adding substituents to norepinephrine to create new drug molecules that may be more specific to certain alpha and beta adrenergic receptors. *Figure taken from Optimization of Lead Structures* by *Gerhard Klebe*.

Just as compounds can be used as building blocks, I implemented an application that can also be built upon. **Specifically, this programs only works off of protein sequences, blasting against a single database, SwissProt, and only pulls from drugs from one drug database, DrugBank as drug leads**. There can be building blocks that include other macromolecules (nucleic acids, lipids, and carbohydrates), blasting more protein databases, other molecule type blast databases, or implementing non-blast searching such as structure homology, motif searching, fold recognition (RNA or proteins), or comparative modelling, as well as including many other drug databases such as ZINC, PubChem, and ChEMBL/ChEBI.

I hope this program will be able to influence other ideas as well for searching proteins in which are functionally and structurally similar to existing proteins in which there are successful and efficacious therapies for. Some of the challenges I endured in my development included are as follows. Determining the scope was a challenge. Despite being able to summarize the program in one sentence (above in bold), I initially wanted to search many chemical databases such as those indicated above as well and also included some functionality of the REACTOME Pathway Browser which include Chemical Compounds in their pathways, for example these 12

References:

1. Patrick, G.L. An Introduction to Medicinal Chemistry. 5th Edition. 2013
2. NCBI blastp.
3. SwissProt
4. Drugbank