Curating Target-Activity Information for NADI Compounds Based on CHEMBL Using Similarity Searching

Muhammad Jaziem bin Mohamed Javeed, Nurul Hashimah Ahamed Hassain Malim, Aini Athirah Rozali, Siti Zuraidah binti Mohamad Zobir, Habibah binti Abdul Wahab

nurulhashimah@usm.my, muhammadjaziem21@gmail.com

INTRODUCTION

A natural product is a chemical compound found in nature. It is produced by living organism. Most of the modern drug discovery are derived from natural products [1]. Besides, natural products shares the same advantage as traditional medicine when treating disease [2]. In Malaysia, natural products are extracted and stored in a database known as NADI.

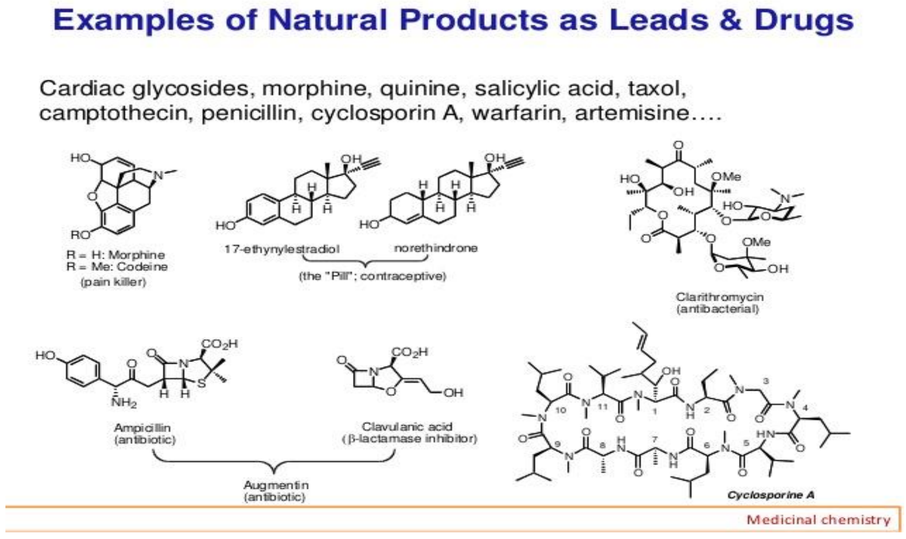

Figure 1: Samples of Drugs which are derived from Natural Products [3].

PROBLEM STATEMENT and MOTIVATION

Studies that utilizes NADI datasets have suggested the potential of NADI compounds being developed as drug-like molecules and later drugs. A study was conducted by using NADI dataset focusing on finding the compounds for Chagas Disease Treatment. To our knowledge, compounds in NADI are actually ligands that would show target-activity to specific targets. However, these compounds has not been linked to any target. Hence, we address this limitation through this study which aim to deduce ligand-target activity of NADI compounds in a process called activity class curation.

RESEARCH METHODOLOGY

Similarity searching which is the simplest method of virtual screening is being used to classify NADI compounds into their respective activity classes by comparing NADI with publicly-chemical database (CHEMBL).

Phase 1 – Data Pre-Processing

Figure 2: Process involved in data pre-processing.

- Canonicalization – generating unique SMILES for NADI compounds.
- Generating fingerprint - the canonicalized SMILES to Extended-connectivity fingerprints (ECFP4) using OpenCDK library.

Phase 2 – Similarity Searching

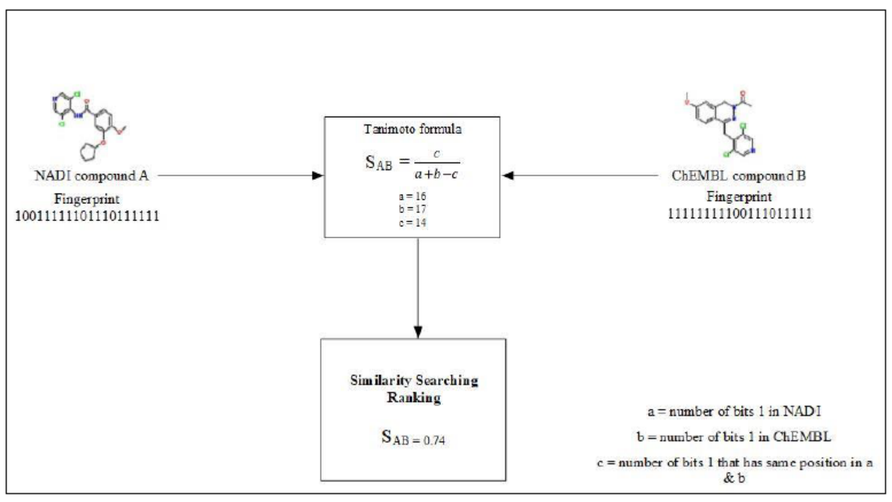

Figure 3: Process involved in similarity searching.

- Tanimoto coefficient – to measure the similarity
- 0.85-1.00 – cut-off range for CHEMBL compounds

Phase 3 – NADI Compound Activity Class Annotation

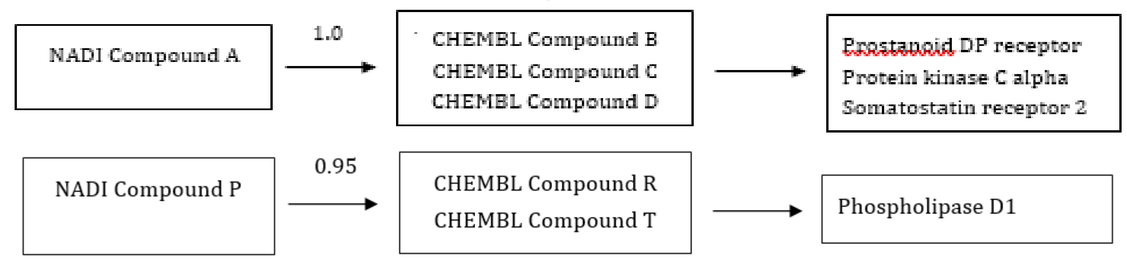

Figure 4: Process involved in target-activity curation

- Pattern matching – exact string matching between CHEMBL ID and its respective activity classes.
- Once a match is found the target activity is used as associate NADI compounds target-activity; else, the NADI compound will be labelled as undetermined.

RESULTS and DISCUSSION

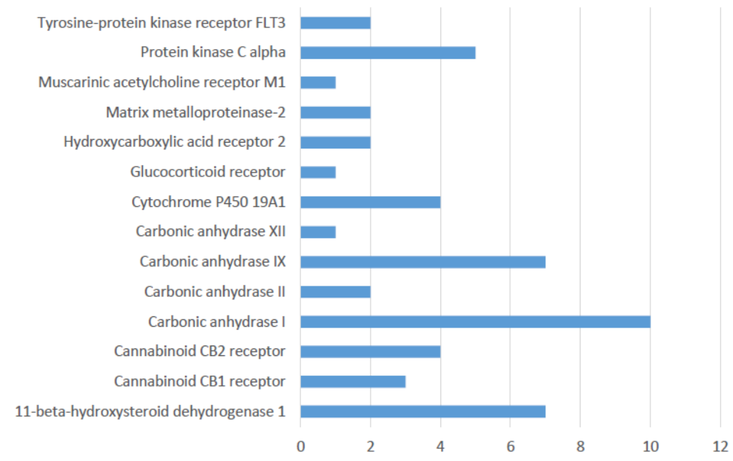

Figure 5: NADI compounds in their activity classes.

Out of 4042 compounds that originally exist in NADI database, only 3901 are able to be converted in canonical SMILES to be further studied by using similarity searching and pattern matching. Not all compounds in ChEMBL are already reported with the information of their activity classes. There are only 102 of target activity classes existed in ChEMBL and this leads to the undetermined activity class of NADI compounds even though there are similar compounds to it that exist in ChEMBL.

42 compounds out of 3901 compounds that exist in NADI that has been successfully assigned to at least one target activity information. Among those 42 compounds, 11 of them is identified interacting with two or more targets.

CONCLUSION and FUTURE WORK

Even though we are not able to classify all the NADI compounds into their respective activity classes, however, the method used in the experiment managed to predict the target for some of the NADI compounds. The main factor is the unreported activity classes for some of the CHEMBL compounds. Despite, some of the NADI compounds achieves a similarity searching score of 1.0, but they couldn't be classified to any of the activity class since their corresponding CHEMBL compounds do not have any target information.

In future, we would like to add more chemical database reference beside CHEMBL. One of them is CRC Dictionary of Natural Products. By implementing more reference, we hope we can identify all the NADI compounds targets accordingly.

REFERENCES

[1]W. Strohl, "The role of natural products in a modern drug discovery program", *Drug Discovery Today*, vol. 5, no. 2, pp. 39-41, 2000.

[2] P. Dewick, "Medicinal natural products: A biosynthetic approach Edited by Paul M. Dewick, Wiley, Chichester, 2002. 508 pp.; E34.95", *European Journal of Medicinal Chemistry*, vol. 36, no. 11-12, pp. 966-967, 2001.

[3]P. Wipf, "Medicinal Chemistry and Drug Discovery", *Slide Player*, 2018. [Online]. Available: https://slideplayer.com/slide/5733313/. [Accessed: 15- Sep- 2018].