Application of computational methods for the design of innovative molecules with multitarget profile

Corso di Dottorato di ricerca in 'RESEARCH METHODS IN SCIENCE AND TECHNOLOGY'
Settore scientifico disciplinare (SSD): Area 03, CHIM/08 CHIMICA FARMACEUTICA
Settori ERC: PE5_18, LS7_4

1. State-of-the-art and general description of the project

To date, the treatment of multifactorial diseases is often performed by single-target molecule or drug cocktail administration. Both these strategies may fail at obtaining meaningful therapeutic outcomes or may lead to several issues such as drug-drug interaction or less predictable pharmacokinetics. For these reasons, there is growing interest in developing molecules capable of binding multiple targets involved in multifactorial diseases, and these ligands have come to be referred to as 'MultiTarget-Directed Ligands' (MTDLs). MTDLs are defined as compounds that are rationally designed to modulate multiple targets of relevance to a disease, with the overall goal of enhancing efficacy and/or improving safety.(1) However, there are some negative aspects to be taken into consideration. MTDLs are usually believed to possess a poor pharmacokinetic profile mostly caused by their high molecular weight (MW) and LogP value. Moreover, their design is a complicated process that is deeply affected by the similarity of the chemical structures of known ligands for the selected targets and the structural similarity of the binding pockets.
The most widely used approach for designing MTDLs is the 'knowledge-based approach' which involves starting with a molecule active at one target of interest and engineering other activities into it. Thus, MTDLs designed by this strategy are obtained by the combination of different pharmacophoric frameworks from several drugs that are responsible for the activity on different targets. While fairly effective, this strategy tends to yield molecules with limited chemical novelty.
With the aim of retrieving chemically novel scaffolds with multitarget profile at selected targets, this project can be divided into four steps: i) through the employment of virtual screening (VS), check for the pre-existing presence of molecules capable of exerting an activity on all the proteins of interest among their known binders; ii) with the insights obtained from this approach, apply a generative machine learning

algorithm to design new drug-like compounds with the sought activity profile; iii) synthesize some selected molecules generated by machine learning and test whether these are active on the selected targets; iv) attempt to improve the affinity of selected hits while maintaining a MTDL profile.

2. Aim & Objective

Aim

- Identify a chemically novel series with activity at multiple previously selected targets

Objective

- Verify the existence of molecules active on all targets of interest among known binders using computational methods such as virtual screening
- Apply generative machine learning methods to design compounds active on targets of interest and identification of hit compounds
- Optimization of hit compounds affinity to proteins of interest

3. Experimental Methods

To achieve the goals described in Section 2, I propose a combination of chemoinformatic tools and resources.

For the first part of this project, publicly available data on known ligands from ChEMBL, a manually curated chemical database of bioactive molecules with drug-like properties, (https://www.ebi.ac.uk/chembl/), will be required.
This step will be accomplished using scripts written in Python, a general-purpose, free, object-oriented programming language known for its power and flexibility. One of its strengths is the presence of extensive standard libraries (packages) that expand the capabilities of the core language, making it possible to employ Python on multiple fields.

The second part of this project involves the use of computational techniques such as virtual screening and the application of generative models. The former is a computational technique used to prioritise from a large library of compounds those that likely bind to a specific target, usually an enzyme or receptor.[(2)] This technique is normally employed in the early steps of the drug discovery process and its main purpose is to retrieve hit compounds that may be improved in subsequent stages.

In this project, VS will be used to check for molecules with a multitarget profile among the known binders of the selected proteins, and to understand the molecular basis of their activity. Then, these results will be used to refine the molecule generation process afforded by generative models.

Generative models are deep learning algorithms that, once trained with a set of samples, manage to understand the underlying probability distribution and output a new sample from that same probability distribution. This approach can be applied to generate samples of molecules where the samples consist of molecular SMILES (a mono-dimensional molecular annotation).

4. Bibliography

1. Morphy R, Rankovic Z. Designed multiple ligands. An emerging drug discovery paradigm. Vol. 48, Journal of Medicinal Chemistry. 2005. p. 6523–43.
2. Gimeno A, Ojeda-Montes MJ, Tomás-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Mulero M, et al. The light and dark sides of virtual screening: What is there to know? Vol. 20, International Journal of Molecular Sciences. 2019.

5. Research Description

The project can be organized over a three-year period:

1°) Query ChEMBL by means of purposely written Python scripts to interact with the database's API, to automatically retrieve information about the known binders of the chosen proteins. Through the employment of the virtual screening technique, check for the presence of already existing molecules with a multitarget profile among the known binders of the selected proteins and identify the chemical features essential for interaction with receptors.

2°) Use the knowledge obtained to fine-tune the molecule generation process. Apply the refined generative model to design new drug-like compounds with desired activities.

3°) Through the use of virtual screening, verify if machine learning-generated molecules are active on the selected targets, identify those with a better binding score,

and attempt to improve their pharmacokinetic and pharmacodynamic characteristics in subsequent hit-to-lead and lead-optimization stages.