Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Jordi Mestres
Email	jmestres@imim.es
Institution	IMIM Hospital del Mar Institute of Medical Research
Website	http://syspharm.imim.cat/
Group	Systems Pharmacology

Project

Pharmacoinformatics & systems pharmacology

Project Title:

A knowledge-based approach to PROTACs design

Keywords:

PROTACs design

Summary:

PROteolysis TArgeting Chimeras (PROTACs) have emerged as a new revolutionary modality in drug discovery. PROTACs are heterobifunctional molecules comprising of a ligand targeting a protein of interest, a ligand targeting an E3 ligase and a connecting linker. The aim is, instead of inhibiting the target, to induce its proteasomal degradation [1,2]. In spite its wide exploitation in many therapeutic areas, there is still a lack of well-thought knowledge-based strategies to designing PROTACs. Accordingly, the main aim of this project will be to establish the knowledge basis to develop new approaches to PROTACs design.

References:

[1] M. Konstantinidou et al. PROTACs– a game-changing technology. Expert Opinion in Drug Discovery (2019) 14:1255. [2] Sun et al. PROTACs: great opportunities for academia and industry. Signal Transduction and Targeted Therapy (2019) 4:64.

Expected skills::

The ideal candidate should have good scripting/programming skills and a background on chemistry/biology/pharmacology/pharmacy.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Master project 2020-2021

Personal Information

Supervisor	Patrick Aloy
Email	patrick.aloy@irbbarcelona.org
Institution	Institute for Research in Biomedicine (IRB Barcelona)
Website	https://sbnb.irbbarcelona.org
Group	Structural Bioinformatics & Network Biology

Project

Pharmacoinformatics & systems pharmacology

Project Title:

Formatting Biological Big Data to Enable Systems Pharmacology

Keywords:

Systems pharmacology, complex diseases, biological and disease signatures

Summary:

The amount and complexity of the biological data generated in the last years, due to the popularization of high-throughput pipelines, is virtually flooding biomedical research. Indeed, the growth of biological databases is steeper than ever before, and the repertoire of possible read-outs spans all levels of biology. However, the nature of biological data is remarkably complex, and dealing with diversity, inconsistency and incompleteness, among other issues, demands heavy specialist processing, and prevents a widespread predictive approach to disease biology. Indeed, this deluge of data has not spurred the development of truly precision therapies, and the inherent limitations of the prevailing reductionist approaches have highlighted the need of moving away from the 'one disease, one target, one drug' paradigm and consider the complexity of human pathologies and physiological responses. The current project builds on the hypothesis that the disease-causing perturbations leave detectable traces at different - and variable -

levels of biological complexity (i.e. activation/inhibition of signaling pathways, transcriptional changes, etc) that capture both the direct effect of the perturbation and a global reaction of the system. Accordingly, the main aim of the project is to collect genuinely heterogeneous datasets, and offer a generic and intuitive means to bridge the gap between biological big data repositories and state-of-the-art machine-learning tools. Besides, we shall develop a generalized connectivity mapping, as a form of virtual phenotypic screening, to discover novel chemical or genetic modulators able to revert the specific signatures of disease and 'cancel out' the phenotypic traits of the disorder. The successful candidate shall be responsible for the implementation of a pipeline to collect and process biological big data, and to encapsulate it in the form of heterogeneous biological embeddings. Overall, we shall develop a novel strategy to integrate the deluge of biological data in a format that is readily suitable for modern machine learning. Additionally, he/she will develop a General Connectivity Mapping (GCMap) strategy to link biological and chemical signatures from the Chemical Checker (htts://chemicalchecker.org), so that the biological context of each small molecule can be incorporated as a descriptor. We shall then explore the added value of these biological descriptors to identify therapeutic opportunities to treat complex diseases.

References:

- Duran-Frigola M, et al. Formatting biological big data for modern machine learning in drug discovery. WIREs Comp Mol Sci (2018), e1408. - Duran-Frigola M, et al. Extending the small molecule similarity principle to all levels of biology. Nat Biotechnol (2020) In press. Available at bioRxiv.

Expected skills::

Highly motivated. Fluency in English. Good programming and scripting skills, with knowledge of Python and databases management (e.g. postgresSQL).

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Patrick Aloy
Email	patrick.aloy@irbbarcelona.org
Institution	Institute for Research in Biomedicine (IRB Barcelona)
Website	https://sbnb.irbbarcelona.org
Group	Structural Bioinformatics & Network Biology

Project

Pharmacoinformatics & systems pharmacology

Project Title:

Generative Models to create Precision Drugs

Keywords:

Systems pharmacology, generative networks, machine learning, precision drugs.

Summary:

Biological data is accumulating at an unprecedented rate, escalating the role of data-driven methods in computational drug discovery. The urge to couple biological data to cutting-edge machine learning has spurred developments in data integration and knowledge representation, especially in the form of heterogeneous, multiplex and semantically-rich biological networks. Today, thanks to the propitious rise in knowledge embedding techniques, these large and complex biological networks can be converted to a vector format that suits the majority of machine learning implementations. Indeed, we have generated biological embeddings (i.e. bioactivity signatures) that capture complex relationships between small molecules and other biological entities such as targets or diseases (Duran-Frigola et al. 2020 Nat Biotechnol). However, only a tiny fraction of the possible chemical space has been so far explored, meaning that most compounds able to modulate biological activities (i.e. drugs) are yet to be discovered. Accordingly, the main objective of this project is to couple our bioactivity signatures to inverse design algorithms to generate new chemical entities with a desired functionality. In particular, we aim at generating new chemical entities (NCEs) to modulate the activity of a specific set of targets, selected from a combination of perturbagen profiles, to revert the pathological state induced by Alzheimer´s disease (AD) and other complex disorders. All in all, the incorporation of machine learning methods to the drug discovery process will trigger the development of thousands of novel compounds, finally enabling precision medicine. The successful candidate shall be responsible for the implementation of ML-based Generative Models (i.e. cVAEs or GANs) to create new small molecules that fulfill the required polypharmacological properties to revert AD pathological signatures.

References:

- Duran-Frigola M, et al. Formatting biological big data for modern machine learning in drug discovery. WIREs Comp Mol Sci (2018), e1408. - Duran-Frigola M, et al. Extending the small molecule similarity principle to all levels of biology. Nat Biotechnol (2020) In press, available at bioRxiv.

Expected skills::

Highly motivated. Fluency in English. Excellent programming and scripting skills, with deep knowledge of Python. Previous experience on the use of machine learning and data science techniques (e.g. TensorFlow/AdaNet) and HPC environments will be an asset.

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Gerard Pujadas
Email	gerard.pujadas@gmail.com
Institution	Universitat Rovira i Virgili
Website	http://www.cheminformatics-nutrition.recerca.urv.cat/en/
Group	Cheminoformatics & Nutrition

Project

Pharmacoinformatics & systems pharmacology

Project Title:

Drug discovery and drug design for COVID-19 treatment

Keywords:

COVID-19, SARS-CoV-2, drug discovery, drug design

Summary:

The project aims to find new drugs that can be used for the treatment of COVID-19 and related patologies. In order to do that, the student will use structural information of therapeutic targets for COVID-19 treatment to define energy-based pharmacophores that allow him to mine in commercial databases of small molecules for finding those that can bind with high affinity to the target of interest. Then, the most promosing hits of this virtual screening will be used in hit-to-lead computational experiments to find more potents derivatives. The project will involve different drug-discovery or drug design technologies such protein-ligand docking, shape & electrostatic comparisons, FEP+. To achieve this project, the student will be trained in the use of the most common drug discovery & design suites (Schrödinger, OpenEye and Cresset). The result of his/her research project will be the several drugs for COVID-19 treatment.

References:

Understanding the variability of the S1' pocket to improve matrix metalloproteinase inhibitor selectivity profiles. Gimeno A, Beltrán-Debón R, Mulero M, Pujadas G, Garcia-Vallvé S. Drug Discov Today. 2020 Jan;25(1):38-57 Mining large databases to find new leads with low similarity to known actives: application to find new DPP-IV inhibitors. Ojeda-Montes MJ, Casanova-Martí À, Gimeno A, Tomás-Hernández S, Cereto-Massagué A, Wolber G, Beltrán-Debón R, Valls C, Mulero M, Pinent M, Pujadas G, Garcia-Vallvé S. Future Med Chem. 2019 Jun;11(12):1387-1401. The Light and Dark Sides of Virtual Screening: What Is There to Know? Gimeno A, Ojeda-Montes MJ, Tomás-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Mulero M, Pujadas G, Garcia-Vallvé S. Int J Mol Sci. 2019 Mar 19;20(6). Combined Ligand- and Receptor-Based Virtual Screening Methodology to Identify Structurally Diverse Protein Tyrosine Phosphatase 1B Inhibitors. Gimeno A, Ardid-Ruiz A, Ojeda-Montes MJ, Tomás-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Mulero M, Valls C, Aragonès G, Suárez M, Pujadas G, Garcia-Vallvé S. ChemMedChem. 2018 Sep 19;13(18):1939-1948 Activity and selectivity cliffs for DPP-IV inhibitors: Lessons we can learn from SAR studies and their application to virtual screening. Ojeda-Montes MJ, Gimeno A, Tomas-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Valls C, Mulero M, Pujadas G, Garcia-Vallvé S. Med Res Rev. 2018 Sep;38(6):1874-1915.

Expected skills::

Good skills with Python and shell scripting

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed
