Notification No — 490
Date — 08/03/2021

Designing Inhibitor Targeting Autophagy Regulating Kinases for Cancer Using Machine Learning Techniques

EXTENDED ABSTRACT
of the PhD Thesis

Submitted to
Jamia Millia Islamia
for the award of the Degree of Doctor of Philosophy

Submitted by

Chetna Kumari

under the supervision of

Prof. Suraiya Jabin
(Supervisor)
Department of Computer Science
Jamia Millia Islamia
New Delhi

Dr. Naidu Subbarao
(Co-supervisor)
School of Computational & Integrative Sci.
Jawaharlal Nehru University
New Delhi

Dr. Muhammad Abulaish
(Co-supervisor)
Dept. of Computer Science
South Asian University
New Delhi

Department of Computer Science
Faculty of Natural Sciences
Jamia Millia Islamia
New Delhi

Chetna Kumari

September, 2020

Designing Inhibitor Targeting Autophagy Regulating Kinases for Cancer Using Machine Learning Techniques

Keywords: *Computer-Aided Drug Design, Autophagy, Cancer, Drug Targets, kinases, Inhibitor Design, Machine Learning, Molecular Docking, Molecular Dynamics Simulation*

Introduction: With a massive amount of 'multi-omics' data, such as genomics, proteomics, transcriptomics, and metabolomics, the field of drug design is now progressing towards a data-intensive science. The rapidly expanding biological databases have increased the complexity in handling, processing, and retrieving the relevant information. To deal with such complexities, biologists have started to join data-club. Consequently, computational approaches are integrated with the traditional drug design process, giving rise to a new field of study known as computer-aided drug design (CADD). The process of CADD starts with disease target identification, selection and validation, hit to lead discovery, lead optimization, and preclinical profiling called drug discovery and proceeds to clinical testing, regulatory affairs, and submission for launch to the market, called drug development. There are three different CADD approaches – structure-based approach, ligand-based approach, and systems-based approach. Structure-based approaches, such as protein structure prediction, molecular modeling, *de novo* drug design, virtual screening based on molecular docking, and molecular dynamics simulation are commonly used for drug target identification and virtual screening of compound library to find hit or lead molecule. Ligand-based approaches, such as molecular profiling, quantitative structure-activity relationship, and virtual screening based on pharmacophore, similarity, and machine learning are mainly applied for the identification and early validation of drug targets, multi-targeted drug discovery (*aka* polypharmacology), designing and screening compound libraries for prediction of drug-like com-

Chetna Kumari

pounds. Systems-based approaches, such as system pharmacology (*aka* network pharmacology) and proteochemometric modeling (i.e., bioactivity modeling based on the description of the ligands as well as proteins) are used for drug repurposing, while pathway analysis is used to study the complexity of interacting biomolecules such as gene, proteins, carbohydrates, and lipids to decode the molecular basis of life. Despite, the integration of computational approaches in the drug design process, the research and development (R&D) productivity in the pharmaceutical industry has not changed positively. Hence, a detailed analysis of R&D productivity in the pharmaceutical industry may help to improve the productivity of the drug design process.

In this thesis, the proposed methodology for inhibitor design against autophagy regulating kinases in cancer uses machine learning techniques and few other computational approaches. Autophagy is a type of programmed cell death that regulates homeostasis during adverse growing conditions in living cells. Despite the significant role of autophagy in the maintenance of normal physiology and its dysregulation in various diseases such as neurodegeneration and cancer, the pathway is still less explored for the development of drugs. Despite the advent of automation, miniaturization, high throughput techniques, and extensive applications of computational approaches in drug design, many of the identified drug targets are unaddressed. This is evidenced by the fact that out of approximately 20,000 protein-encoding genes, only 10,000 to 12,000 proteins are estimated to be ubiquitously expressed, while the currently approved drugs target only 618 proteins. This indicates that a large number of proteins are unidentified, and many of the identified proteins remain to be drugged. Before the selection of a drug target for inhibitor design, it is crucial to understand the mechanism of the dysregulated molecular pathway, the status of the selected drug target in a particular disease, and its capability to get modulated by the novel drug molecule.

Protein kinase is one of the prominent family of drug targets, constituting approximately 10% of human protein drug targets. Kinases are integral to the autophagy pathway, and many of the autophagy regulating protein kinases are already validated

as anticancer drug targets. Autophagy plays dual roles in tumor initiation and progression, hence therapeutically targeting autophagy in cancer is challenging but promising. To date, only a few FDA-approved drugs such as Rapamycin and its analogs that modulate the autophagy pathway are known to be used for clinical purposes. This study proposes a machine learning-based approach and few other computational approaches for predicting inhibitors against autophagy regulating kinases in cancer. The protein kinases selected in this study are – mammalian Target of Rapamycin (mTOR), c-Jun N-terminal Kinase 1 (JNK1), Unc-51 Like Kinase 1 (ULK1), and Death Associated Protein Kinase 1 (DAPK1).

Aim: "Designing Inhibitor Targeting Autophagy Regulating Kinases for Cancer Using Machine Learning Techniques".

Objectives: A summarized list of the objectives of this thesis is given below:

1. The first objective of this thesis is to understand regulation of autophagy by mammalian Target of Rapamycin (mTOR), and to predict small molecule inhibitors against it.

2. The second objective is to search c-Jun N-terminal kinase (JNK) signaling as a therapeutic target for cancer.

3. The third objective is to target UNC-51 Like Kinase 1 (ULK1), a mammalian counterpart of Atg1, for screening small molecule inhibitors of the autophagy pathway.

4. The fourth objective of this thesis is to evaluate Death-Associated Protein Kinase (DAPK) as a therapeutic target for designing small molecule inhibitor of autophagy.

   Finally, the integration and analysis of the inhibitors related to the targets mentioned above using machine learning techniques and virtual screening approaches.

Methodology: The 3D structures of selected kinases and related bioactivity datasets are deposited in the protein data bank (PDB) and Chemical database of the European Molecular Biology Laboratory (ChEMBL), respectively. Together, these may guide towards structure-based as well as ligand-based drug design against the selected drug targets. mTOR and JNK1 bioactivity datasets show sufficient compounds with known bioactivity values (IC50) that can be used for assigning labels to the compounds as *active* or *inactive*. Hence, we propose to develop predictive models using supervised machine learning approaches for predicting mTOR and JNK1 inhibitors. The computational intelligence in drug design has progressed towards artificial intelligence (AI)-related machine learning (ML) and deep learning (DL) techniques. Therefore, we have applied random forest variable importance measures and deep neural network-based autoencoders for molecular feature selection and dimension reduction. Moreover, both mTOR and JNK1 bioactivity datasets curated from ChEMBL show disproportionate distribution of *active* and *inactive* classes, and generally classifiers show poor performance over class-imbalance datasets. Hence, we have used different class-balancing methods such as undersampling, oversampling, and hybrid approaches to balance the skewed class-distribution of training datasets, and evaluated their performance in enhancing the efficacy of the predictive models. On the other hand, the bioactivity datasets of ULK1 and DAPK1 do not have sufficient compounds with known bioactivity values, that can be used to generate a labeled dataset. Hence, we have proposed the prediction of novel ULK1 Inhibitors by integrating structure-based molecular docking and unsupervised machine learning, whereas DAPK1 inhibitors are predicted by integrating structure-based approaches such as molecular docking and molecular dynamics simulation.

Results: We find that depending on the availability of the protein and bioactivity datasets, various computational approaches can be applied at the early stages of drug discovery to screen large compound library, and find novel chemical entities. The recent rise of AI, ML, and DL techniques in drug discovery may help in achieving the

success or failure in finding 'hit' or 'lead' molecule at a faster rate. Moreover, exploiting the autophagy pathway for therapeutic targeting in cancer is significant, as autophagy provides nutrients to the developing cancer cells in the advanced stage of tumorigenesis which is detrimental. Hence, designing inhibitors against autophagy regulating kinases in cancer seems to be beneficial, and may also help to overcome drug resistance in cancer. This thesis presents machine learning-based classification models developed over class-imbalanced as well as class-balanced mTOR and JNK1 bioactivity datasets to predict their inhibitors. The predictive models are validated using internal as well as external validation datasets. Moreover, a novel approach proposed to predict the ULK1 inhibitor is capable of identifying a few potent ULK1-inhibitor like compounds. A novel DAPK1 inhibitor is predicted using structure-based docking, consensus scoring, molecular interaction, and molecular dynamics simulation studies.

Consultation: We conclude that depending on the availability of the protein and bioactivity datasets, various computational approaches can be applied at initial stages of drug discovery to screen large compound library, and find novel chemical entities. The recent rise of AI, ML, and DL techniques in drug discovery may help in achieving the success or failure in finding 'hit' or 'lead' molecule at a faster rate. Moreover, exploiting the autophagy pathway for therapeutic targeting in cancer is significant, as autophagy provides nutrients to the developing cancer cells in the advanced stage of tumorigenesis which is detrimental. Hence, designing inhibitors against autophagy regulating kinases in cancer seems to be beneficial, and may also help to overcome drug resistance in cancer. However, the computationally predicted candidate drug molecules should be experimentally validated to get accepted for the clinical trials.

* * *