# Chapter 50
# Analysis of Kinase Inhibitors and Druggability of Kinase-Targets Using Machine Learning Techniques

**S. Prasanthi**
*University of Hyderabad, India.*

**T. Sobha Rani**
*University of Hyderabad, India.*

**S.Durga Bhavani**
*University of Hyderabad, India.*

**Raju S. Bapi**
*University of Hyderabad, India.*

## ABSTRACT

*Vast majority of successful drugs or inhibitors achieve their activity by binding to, and modifying the activity of a protein leading to the concept of druggability. A target protein is druggable if it has the potential to bind the drug-like molecules. Hence kinase inhibitors need to be studied to understand the specificity of a kinase inhibitor in choosing a particular kinase target. In this paper we focus on human kinase drug target sequences since kinases are known to be potential drug targets. Also we do a preliminary analysis of kinase inhibitors in order to study the problem in the protein-ligand space in future. The identification of druggable kinases is treated as a classification problem in which druggable kinases are taken as positive data set and non-druggable kinases are chosen as negative data set. The classification problem is addressed using machine learning techniques like support vector machine (SVM) and decision tree (DT) and using sequence-specific features. One of the challenges of this classification problem is due to the unbalanced data with only 48 druggable kinases available against 509 non-drugggable kinases present at Uniprot. The accuracy of the decision tree classifier obtained is 57.65 which is not satisfactory. A two-tier architecture of decision trees is carefully designed such that recognition on the non-druggable dataset also gets improved. Thus the overall model is shown to achieve a final performance accuracy of 88.37. To the best of our knowledge, kinase druggability prediction using machine learning approaches has not been reported in literature.*

DOI: 10.4018/978-1-4666-3604-0.ch050

Copyright © 2013, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

## PATTERN DISCOVERY IN KINASES

Human genome contains about 518 protein kinase genes, which constitute about 2% of all human genes (Vulpetti & Bosotti, 2004). Protein kinases regulate almost all biochemical pathways. They play a critical role in signal transduction, physiological responses, and in the functioning of nervous and immune systems. They also control many other cellular processes like metabolism, transcription, cell cycle progression, cyto-skeletal rearrangement and cell movement, apoptosis, and differentiation (Bakheet & Doig, 2009).

Kinases are enzymes which help in phosphorylation of substrates facilitating the transfer of phosphate group from ATP. They may phosphorylate up to 30% of the proteome (Manning et al., 2002), (Manning, 2005). Since kinases participate in signal transduction pathways of cell cycle and cell differentiation they are known to be targets for diseases. Abnormal phosphorylation of the protein kinases is a cause of disease and hence needs to be inhibited by small drug-like molecules called kinase inhibitors. Some of the well-known inhibitors are Serine/Threonine kinase inhibitors and Tyrosine kinase inhibitors which are named on the basis of the amino acid whose phosphorylation is inhibited. Kinase inhibitors are developed in the treatment of diseases like cancers, inflammatory disorders, neurological disorders, diabetes mellitus, heart disease etc. Some of the available kinase inhibitor drugs are Imatinib, Nilotinib and Gefitinib.

In this study we present two perspectives of drug discovery: one from the view point of kinase target and the other from kinase inhibitor. Even though kinases are known to be targets for diseases, not all kinases are druggable. Hence it is important to distinguish druggable kinase targets from non-druggable kinases. Further kinase inhibitors need to be studied to understand the specificity of a kinase inhibitor in choosing a particular kinase target. The ultimate goal, in some sense, is to predict the matching between a target and its corresponding inhibitor(s) with the help of target and ligand properties individually and together with protein-ligand interaction features. In this paper we restrict ourselves to addressing the problem of druggability of kinases and conduct a feature analysis of kinase inhibitors. The problem of matching will be taken up in future. In the next section we present a study of significant properties of kinase inhibitors.

## BACKGROUND

Vieth et al., (2004) conduct a study of kinase targets and inhibitors in order to identify medicinally relevant kinase space. Using both sequence based information and the small molecule selectivity information, they presented the first dendogram of kinases based on small molecule data. This study concludes that the structural basis of kinase inhibitor selectivity will require knowledge of complexes of one ligand with multiple targets. Classification of kinase inhibitors with a bayesian model was studied by Xia et al., (2004). Using Bayesian statistics, a model for general and specific kinase inhibitors was proposed. They have considered serine/ threonine and tyrosine kinase inhibitors (Amgen compounds) from CORP data set. Kinase model was generated using properties like number of hydrogen bond donors, halogens, aromatic residues, value of AlogP and molecular weight. The general kinase model described was trained on tyrosine kinase inhibitors achieving prediction accuracy of 80%.

In order to initiate the study of kinase inhibitors we need both kinase target and inhibitor features that are available in various databases.

### Databases of Chemical Compounds

A drug molecule is required to satisfy the well-known properties known as Lipinski's rules (Lipinski et al., 1997). Drug Bank (http://www.drugbank.ca) is a popular data base housing FDA

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/analysis-kinase-inhibitors-druggability-kinase/76103

## Related Content

Cluster Analysis in R With Big Data Applications
Alicia Taylor Lamere (2024). *Research Anthology on Bioinformatics, Genomics, and Computational Biology (pp. 482-503).*
www.irma-international.org/chapter/cluster-analysis-big-data-applications/342540

Network Querying Techniques for PPI Network Comparison
Valeria Fiondaand Luigi Palopoli (2009). *Biological Data Mining in Protein Interaction Networks (pp. 312-334).*
www.irma-international.org/chapter/network-querying-techniques-ppi-network/5571

Clustering Genes Using Heterogeneous Data Sources
Erliang Zeng, Chengyong Yang, Tao Liand Giri Narasimhan (2010). *International Journal of Knowledge Discovery in Bioinformatics (pp. 12-28).*
www.irma-international.org/article/clustering-genes-using-heterogeneous-data/45163

Improving Prediction Accuracy via Subspace Modeling
Majid Masso (2012). *Computational Knowledge Discovery for Bioinformatics Research (pp. 33-48).*
www.irma-international.org/chapter/improving-prediction-accuracy-via-subspace/66703

Data Mining Approach for the Early Risk Assessment of Gestational Diabetes Mellitus
Saeed Rouhaniand Maryam MirSharif (2018). *International Journal of Knowledge Discovery in Bioinformatics (pp. 1-11).*
www.irma-international.org/article/data-mining-approach-for-the-early-risk-assessment-of-gestational-diabetes-mellitus/202360