Tox2020: Toxicity Prediction of pre clinical trial drug using Physicochemical Properties and Machine Learning Approaches

File No : ECR/2015/000150 (Ver-1)

Submitted By : Prashant Singh Rana

Submission Date : 29-Sep-2015

Proposal

File No : ECR/2015/000150 (Ver-1)

Proposal Details

Project Title :	Tox2020: Toxicity Prediction of pre clinical trial drug using Physicochemical Properties and Machine Learning Approaches		
Scheme :	Early Career Research Award		
Broad Area :	Life Sciences		
Duration In Month :	24	**Total Cost (in Rs.) :**	*51,14,000*
Name of Principal Investigator :	Prashant Singh Rana	**Email ID :**	psrana@gmail.com
Date of Birth :	15-Mar-1982	**Contact No :**	919313889932
Category :	GEN	**Gender :**	Male
Designation :	Assistant Professor	**Is differently abled :**	No
Department :	Computer Science & Engineering Department	**PI Address :**	Computer science & engineering department, Thapar university, Patiala
PI Institute :	Thapar Institute of Engineering & Technology	**State :**	PUNJAB
PI Institute Address :	Adarsh Nagar, Prem Nagar, Patiala, Pb, Patiala, Punjab-147004	**Pin Code :**	147004
Institute Joining Date :	15 July, 2015		

CO PI Details :

Name & Designation	Date of Birth	Mobile No.	Email	Institute Details
Deepak Garg (Associate Professor)	05-Aug-1975	919815599654	deep108@yahoo.com	Thapar Institute of Engineering & Technology, Adarsh Nagar, Prem Nagar, Patiala, PB, PUNJAB-147004
Vijay Kumar (Assistant Professor)	10-Jun-1982	917073169619	vijaykumarchahar@gmail.com	Thapar Institute of Engineering & Technology, Adarsh Nagar, Prem Nagar, Patiala, PB, PUNJAB-147004
Avinash Mishra (Post doc research Fellow)	14-Mar-1984	919311618895	avinash.mishra@novoinformatics.com	Indian Institute of Technology Delhi , Hauz Khas, New Delhi, DELHI-110016

Project Summary : Most people are exposed to many different chemicals during the course of their lifetimes through

sources including food, household cleaning products and medicines. In some cases, these chemicals can be toxic. In fact, more than 30 percent of promising pharmaceuticals have failed in human clinical trials because they are determined to be toxic despite promising of pre-clinical studies in animal models (Nat Rev Drug Discov. 2004;3(8):711–715). Creating new methods (Toxicology and Applied Pharmacology. 2010; 245(2):153-159) for assessing chemical toxicity has the potential to improve how scientists evaluate environmental chemicals and develop new medicines. There are more than 40 million (ZINC, PubChem) commercially-available compounds for virtual screening. The goal is to quickly and efficiently test whether certain chemical compounds have the potential to disrupt processes in the human body that may lead to adverse health effects. Physicochemical properties of chemical compounds always guide to determine its activity (active or inactive); therefore it has been rigorously used to classify as drug and non-drug. In this work we try to explore the machine learning methods (such as random forest, support vector machine, neural network, linear model, M5P, cubist, foba and decision stump etc) with physicochemical properties to predict the toxicity of chemical compound. There are more than 8000 molecular descriptors (physicochemical properties) are identified for chemical compounds that will help in train and test the predictive model. To calculate the molecular descriptors several open source and commercial tools are available. The approach are classified into eight phases. In the first phase 10,000 compound libraries will be collected from ZINC and PubChem databases. Data cleansing and filtering will take place in phase 2. Physicochemical parameters for 10,000 collected compounds will be calculated in phase 3, Feature selection will be performed in phase 4. In phase 5, Machine learning methods will be train and test and evaluated in phase 6. K-fold cross validation would be used to measure the robustness of the best predictive method an it will be performed in phase7. Finally, a stand-alone and/or web based application will be developed for the research community. The over all objectives are: (i) To develop better toxicity assessment features, methods and algorithms along with better data structures to handle the big data. (ii) To create new methods for assessing chemical toxicity which will have the potential to improve the procedure followed by scientists to evaluate environmental chemicals and develop new medicines. (3) To develop a framework to quickly and efficiently test certain chemical compounds for their probable chances to disrupt processes in the human body. (4) To develop a stand-alone and/or web-based application(s) that helps the researchers and research community to predict the toxicity of the newly discovered chemical compound.

Keywords : Toxicity Prediction, Machine Learning, Feature Selection, stress response pathways, nuclear receptor signaling pathways

Objective :

- To develop better toxicity assessment features, methods and algorithms along with better data structures to handle the big data.
- To create new methods for assessing chemical toxicity which will have the potential to improve the procedure followed by scientists to evaluate environmental chemicals and develop new medicines.
- To develop a framework to quickly and efficiently test certain chemical compounds for their probable chances to disrupt processes in the human body.
- To develop a stand-alone and/or web-based application(s) that helps the researchers and research community to predict the toxicity of the newly discovered chemical compound.

Expected Output and Outcome of the proposal :

- The outcome of Tox2020 will be available in the form of stand-alone and/or web-based application that help the researchers and research community to predict the activity of the newly discovered chemical compound.
- Tox2020 can also be used as a decision support system that helps to improve the activity a drug molecule.