



Structured Data Matrixes of Active Molecules for Drug-Repurposing

Sivakami Dhulap*, Anita Mandhare and Veena Deshpande

Department of Pharmacology, CSIR - Indian Institute of Chemical Technology, Maharashtra, India

Short Communication

The traditional approach to drug discovery involves *de-novo* identification and the validation of new molecular entities (NME) responsible for treating a disorder in humans or animals. However this process is time-consuming and costly. Despite of huge investment in this traditional drug discovery and development process the number of new drugs introduced into the clinical trial and drug approval phase has not increased significantly. According to some estimates it costs about 2.5 billion dollar [1] and around 10-15 years of time to discover, develop and validate a new drug which can be launched in the market. Intervention of computers at some plausible steps is imperative to bring down the cost and time required in the drug discovery process. The progress of chemistry in terms of the development of numerous synthetic chemical active molecules, isolation and characterization of actives from natural products and the possibilities opened up by experimental pharmacology have lead to single molecules that are rigorously tested, optimized, toxicologically cleared, clinically proven and elucidation of their mechanism of action. A new climate was created by methods evolved on the basis of modern physiology, concerned predominantly with effects that were measurable under experimental conditions. As a result it became possible to elucidate the mode of action of many synthetic as well as natural product based actives which has brought major advances in the investigation and use of these molecules. This new approach, which allowed the subject to become open to scientific investigation, has generated considerable data on the mode of action and efficacy of individual molecules in treating various diseases. The advancement of Information and Communication technologies has given rise to the disclosure of large amount of such data which is widely dispersed in number of scientific journals and patents and are not available via a single source of information. A structured compilation of all such information relating to the chemical as well as pharmacological aspects of the active molecules which are in public domain can hence be of immense use for the discovery of newer drugs or new indications. In light of this background, this communication firstly provides the information about the development of such structured data matrixes containing the chemical as well as pharmacological data for the active molecules. Secondly we also illustrate the application of computational tools to such structured information for the discovery of newer hits useful for drug discovery. A search focused on collating the data on the phytoconstituents and molecules in public domain from scientific literature as well as from patent documents was carried out. The search method consisted of using data mining and text mining. The search for the phytoconstituents was mainly focused to collect the information regarding 500 Indian medicinal plants known in the traditional Indian medicinal system. On the other hand the information with respect to the therapeutically active molecules in public domain was mainly collected using various scientific and expired patent documents using a number of publicly accessible and subscribed chemistry and Pharma oriented databases. The information regarding the phytoconstituents as well as the therapeutically active molecules in public domain was initially screened for their safety issues with respect to the major toxicity endpoints. The molecules were searched in various paid and subscribed toxicity databases to list out the existing toxicity data if any. The potentially toxic molecules were screened out based on the toxicity data retrieved based on different endpoints namely acute toxicity, chronic toxicity, skin sensitization, genotoxicity, carcinogenicity, reproductive/developmental toxicity, neurotoxicity and metabolism [2]. The toxicity data was documented in the form of structured data matrixes taking into account the regulatory requirement. The molecules which were potentially free of any major toxicity reports were further subjected to in-silico prediction using a proprietary knowledge based software. Finally only those molecules which were predicted to be less toxic were selected and subjected for further analysis and inclusion into the dataset. For the shortlisted phytoconstituents and the therapeutically active molecules, the bio-evaluation information along with their physical, chemical and drug like properties from the scientific as well as patent documents were enlisted in the form of data matrix.

OPEN ACCESS

*Correspondence:

Sivakami Dhulap, Department of Pharmacology, CSIR - Indian Institute of Chemical Technology, Maharashtra, India,
E-mail: sivakami@urdip.res.in

Received Date: 21 Mar 2017

Accepted Date: 24 Apr 2017

Published Date: 26 Apr 2017

Citation:

Dhulap S, Mandhare A, Deshpande V. Structured Data Matrixes of Active Molecules for Drug-Repurposing. *Ann Pharmacol Pharm.* 2017; 2(5): 1050.

Copyright © 2017 Dhulap S. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For each of the active molecule included in the matrix the general information such as name, synonymous, molecular formula, molecular weight, physical constants, drug like properties, solubility data, pKa value, Log P value, H-bond donor values, H-bond receptor values, biological activity and binding data for the receptor sites were indexed [3]. The data matrix can be searched graphically using exact, similar or substructure search options, including by activity or by giving values for a combination of parameters. More generic searches may be performed using a variety of keyword options. This component of the data matrix thus integrates together all the above information about molecules that have been reported to be active constituents and exhibiting particular biological activity. The information with respect to the chemical structures was included in the form of simplified molecular-input line-entry system (SMILES) [4] as well as the 2D chemical structure. The 2D structures were then converted into 3D structures using proprietary software tools. The 3D structures were then stored in the form of Structure Data Files (SDF) files. This resulted into a virtual library comprising of approximately 10400+ drug like molecules and devoid of any major toxicity endpoints. The computational tools such as cheminformatics and bioinformatics tools were then applied to this virtual library in order to screen out the potential hits for a target of interest. As a next step the recent drug discovery methods of using virtual screening and molecular docking studies in the drug discovery process was applied to virtually screen out the potential hits against the targets of interest. For example the E-pharmacophore based structural screening [4] was applied to screen the virtual library to enlist the hits with a potential to regulate the function of cyclooxygenase-2 inhibitors (COX-2) [5-7] during inflammation and a four-point pharmacophore model for screening hits to regulate the function of P-glycoprotein (P-gp) in Alzheimer's disease [8]. The hits obtained as hits from the virtual screening were subjected to molecular docking studies using proprietary software tools. This permitted to provide an overview on the feasible interactions between ligand molecule and a protein receptor. The examples of hits as retrieved from the databases for COX-2 for inflammation and P-gP for Alzheimer's disease are incorporated herewith from our publications. The application of the molecular modelling approaches for screening the virtual library under consideration lead to the discovery of alkaloids from the stems of *Fissistigma oldhamii* and triterpenoids present in the resinous exudates of *Commiphora myrrha* as potential hits for regulating the over expression of COX-2 enzyme during inflammation [6]. Further the studies related to the screening of hits to regulate the function of P-glycoprotein protein in Alzheimer's disease revealed the potential

role of alkaloid having the dihydrobenzofuran indol-3-yl scaffold in preventing the onset of the Alzheimer's disease [7]. These results thus illustrate that the application of computation tools to the know data can lead to hits which find can be useful in the development of potential leads. The present study thus illustrates that collating the chemical and pharmacological information from scientific as well as expired patent documents using data mining techniques and integrating into a value added data matrix has successfully lead to the development of searchable data matrixes comprising the chemical and pharmacological information in a indexed and structured manner. This value added information can be used generate knowledge in terms of hits or leads in the drug discovery process. As the most of the preliminary data required in a traditional drug discovery process is available through such indexed information, the time and cost involved in the selection and optimization of the hit is reduced considerably. Further the application of cheminformatics and bioinformatics tools and software to such data, adds value to such information. Such an integrated data in a computer readable form can be searched and manipulated in a desirable way for the design and development of novel molecules and formulations. Modern software can use such data at the back end to provide faster 'leads' for new drug molecules and weed out non-active candidates early in the process.

References

1. Rick Mullin. Cost to Develop New Pharmaceutical Drug Now Exceeds \$2.5B. *J Comput Chem*. 2014.
2. Archana P, Anushree V, Veena D, Raj H. Evaluation of Potential Toxicity of Bioactives of *Anagallis arvensis*. *A Toxic Plant International*. *J Pharmacy Res*. 2016;8(3):163-72.
3. Weininger D, SMILES a chemical language and information system. Introduction to methodology and encoding rules. *J Chem Inf Model*. 1988;28(1):31-6.
4. Salam NK, Nuti R, Sherman W. Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. *J Chem Inf Model*. 2009;49:2356-68.
5. Turini ME, DuBois RN. Cyclooxygenase-2: a therapeutic target. *Annu Rev Med*. 2002;53:35-57.
6. Abhijeet Dhulap, Sivakami, Dhulap, R R Hirwani. *J Acc Chem Res*. 2013;3(4):6-10.
7. Pravin Shinde, Nikhil V. *Central Nervous System Agents in Medicinal Chemistry*. 2016;16:000-000.