INDO GLOBAL JOURNAL OF PHARMACEUTICAL SCIENCES
ISSN 2249- 1023

Bioinformatics Tools for Pharmaceutical Drug Product Development

Johra Khan [1,2], Rajeev K. Singla [3,4*]

[1] *Department of Medical Laboratory Sciences, College of Applied Medical Sciences, Majmaah University, 11952, Al Majmaah, Saudi Arabia. E-mail address: j.khan@gmail.com*

[2] *Health and Basic Sciences Research Center, Majmaah University, Al Majmaah 11952, Saudi Arabia*

[3] *Institutes for Systems Genetics, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Xinchuan Road 2222, Chengdu, Sichuan, China*

[4] *School of Pharmaceutical Sciences, Lovely Professional University, Phagwara, Punjab-144411, India*

Address for Correspondence: Rajeev K. Singla, rajeevsingla26@gmail.com

Received:
28.01.2022
Accepted:
03.09.2022
Published:
24.12.2022
Keywords
Drug Discovery, Bioinformatics, Cheminformatics , Drug Development.

ABSTRACT: Drug discovery and production is a long and expensive process which starts with target identification followed by validation of targets to lead optimization, taking years to develop a drug which sometime false to reach marked resulting in loss of time, effort, and huge amount of money. Bioinformatics tools are becoming more and more important in drug product development. Repurposing large amount of data needs to be exploited and generated from genomics, epigenetics, cistromic, proteomics, transcriptomics, ribosomal profiling, and genomic based studies of drug targets. Bioinformatics analysis and data mining are effective tools to explore big series of biological and biomedical data, however the advance tools are often found difficult to understand making their use limited to difficult to access by the researchers working in drug discovery. In this review we focused on systematically presenting the different tools used for drug target identification and product development. The tools are broadly classified according to disease based computational tools, gene based tools, and web based tools and ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) study for drug repurposing. The focus was on the basic principle of these tools functioning, uses and limitations in drug target identification, validation, data analysis, comparison with other similar tools in target analysis. © 2022 Caproslaxy Media. All rights reserved.

INTRODUCTION

Drug discovery for any disease is always an important process not only for human welfare but also for pharmaceutical industries which invest nearly 1.8 US dollars and a time of more than a decade to bring a new drug to market [1]. From 2019 with starting of this pandemic (COVID-19) the need for fast and efficient drug discovery increased [2]. One of the methods in the drug discovery pipeline is to reduce the time required to gather more information from basic science and research [3]. The translational drug discovery method is an effective approach not only in new drug discovery but also allows research and treatment to be patient-specific. Bioinformatics is an interdisciplinary science that uses molecular data for drug discovery. In bioinformatics molecular data of patients, animals, different disease models, cell lines, and controls are compared to connect symptoms of disease with epigenetic modifications, mutations, other changes [4]. Bioinformatics helps in the identification of drug targets that can function in resorting to cellular activities or in removing malfunctioning cells. It also helps in providing information of possible drug candidates to target or design therapeutic approaches against a particular disease and can also help in evaluating effect of environment on health of different human beings with potential drug resistance.

Drug repurposing is a very recent technique of re-evaluating old drugs and drug compounds in the pharmaceutical industry for therapeutical potential towards other diseases [5]. Repurposing of old drugs helps in faster drug development with a great impact on personalized medicine. Bioinformatics with repurposing significantly reduce research and development time and financial burden on pharma industries [6]. In this chapter, we are discussing different bioinformatics tools for drug product development in pharmaceutics that can

Cite this article as: Khan, J.; Singla, R.K. Bioinformatics Tools for Pharmaceutical Drug Product Development. Indo Global J. Pharm. Sci., 2022; 12: 281-294. **DOI**: http://doi.org/10.35652/IGJPS.2022.12037

make the process of drug discovery and repurposing faster and cheaper.

TARGET BASED BIOINFORMATICS TOOLS

To discover a potential drug, it is necessary to identify whether the target is druggable or not. Target-based analysis helps in reducing the risk of project failure and saving time and money investments [7]. Identification of drug-target includes understanding the molecular data including gene sequence analysis, protein interaction, and metabolic pathway analysis [8]. Target identification also needs combining data from genomics, proteomics, transcriptomics, and metabolic aspects of a disease. Computational target analysis is a most rigorous exercise that includes a study of the human genome and associated annotations, algorithms for gene sequence analysis, protein structure prediction, and proteomics analysis [9].

1. Gene to Target Method

In the gene to target method initial step is to select a common group of drug targets followed by designing a computational method to discover new members of this group to forecast their function based on the available information and knowledge of the target group [10]. The first step of target identification is a screening of the gene sequence database. The Discovery of new members of a target gene class is an important part of drug discovery and understanding the molecular basis of a disease condition. The two important strategies in early target prediction are; a) Genome data mining, and b) Expressed sequence tags [11].

a) *Genome Data mining:* Human genome sequence data mining is used to detect new protein-coding genes that can become the new targets. G-protein coupled receptors (GPCRs) are one of the most important targets of protein classes for drug discovery, studied using primary database search tools such as BLAST (Basic Local Alignment Search Tool) or PRINTS (Protein Fingerprints) [12]. Some studies reported BLAST and PRINT as significant tools for GPCRs identification whereas some studies found them insufficient as GPCRs is a very divergent family of proteins with strikingly small similar sequences shared between the groups. Researchers focused on other *in silico* methods to overcome the limitations of BLAST and PRINT by incorporating other features like trans-membrane topology, amino acid configurations, and physiological properties of these GPCRs [13]. *ab-initio* is another technique for gene sequence prediction and is helpful in the discovery of new GPCR targets [14].

b) *Expressed sequence tags* (ESTs): A large number of expressed sequence tags helps in the collection of resources for gene identification, their characteristics, and tissue-specific gene expression [15]. The most specific function of the ESTs database is to classify new gene expression levels. Researchers like Wittenberger et al. (2001) used EST database search method to find new GPCRs family with 14 new ESTs, five of which GPR84, 86, 87, 90, and 91 were experimentally validated as promising candidates for new putative GPCRs [16]. Out of these five GPR86 was reported to be the center of many pathophysiologies and immune system diseases. A similar investigation was also conducted by Marvanova et al. (2002), who used ESTs as an initial point for map brain expression and as a potential drug target [17].

Understanding gene function is also one of the essential requirements for drug target identification [18]. *In silico* method of bioinformatics is used to explain the gene function but still finding protein function is the most challenging issue in this bioinformatics era. In different organisms most studied like *Escherichia coli* and *Plasmodium falciparum* also have 30 to 60 % of their functions of all identified genes are still unknown [19]. The biggest limitation in studying gene function is the absence of fully assayed signal-specific metabolic events and expected changes in protein phosphorylation and gene expression [20].

2. Disease based approaches

The target identification for drug designing needs basic knowledge of the etiology of disease and its related biological processes and control systems [21]. The disease-based approach focuses on a particular disease or therapeutic category of the disease. Different pharmaceutical companies focus on different diseases for target identification. To discover gene expression profiles in disease to find drug target microarray technology is used as it can identify novel molecular targets and related therapeutically biochemical pathways [22]. Microarray technology also helps in understanding disease regulatory networks, their biological processes, and related cellular pathway which leads to the identification of potential drug targets. The next step is to reduce the drug target gene which appears centrally related to the disease etiology [23]. To identify the possible drug targets for Alzheimer's disease, Cellzome Ltd used the microarray technique to develop a series of small molecule γ- secretase modulators [24].

COMPUTATIONAL DRUG REPURPOSING TOOLS

Computational drug discovery includes drug repurposing in which a large number of servers some of which are available online free and some are paid are used for drug-target interaction studies [25]. Some of the steps (**Figure 1**) used are:

a) Ligand fingerprint encoding- The basic principle in using ligand cantered calculations for fingerprint encoding is useful due to their structural resemblance and comparability in biological functions and properties [26]. To detect unknown leads, previous knowledge of the same type of compounds and their target binding is required. The data available of publicly accessible

Cite this article as: Khan, J.; Singla, R.K. Bioinformatics Tools for Pharmaceutical Drug Product Development. Indo Global J. Pharm. Sci., 2022; 12: 281-294. **DOI**: http://doi.org/10.35652/IGJPS.2022.12037

compounds are huge in comparison to studied and screened data based on target – agnostic ligand-similarity-based strategies [27].

b) ChemMapper- ChemMapper is a 3D similarity algorithm also known as SHAFTS (SHApe-FeaTure Similarity) to find the polypharmacological appearance of a target. It uses a triple hashing method for fast alignment of molecular confirmation by using shape and chemical structure type for assessing alignment [28]. ChemMapper assimilates data related to target annotation from different sources like KEGG (Kyoto Encyclopedia of Genes and Genomes), ChEMBL, BindingDB, and Protein data bank [29]. To validate data from SHAFTS standard virtual screening data set are used and also to identify targets [30].

c) ChemProt- It provides a heat map that connects bioactive compounds with proteins that have a database of more than 7 million stored connections collected from annotating compounds and diseased proteins [31]. ChemProt is one of the biggest confederated database sets of proteins, diseases, and interactions collected from different sources [32]. Concerning drug reposting, it offers a similarity ensemble approach (SEA), reimplementation, and Quantitative Structure-Activity Relationships (QSAR) and it also provides an ensemble-based estimation of the probable target to query molecules. The query molecules can be associated with a drug set and a related map provides a technique to direct the known connections as per the combined database [33]. To assess new interactions similarity of fingerprints is used to produce a set of like drugs. Simplified Molecular Input Line Entry System (SMILES) shows the prediction can be input first then followed by protein selection from the available list and finally downloaded as positive or negative prediction results but till now no validation has been found about any prediction [34].

d) Molecular Docking method – It is a method to forecast the intermolecular complex structure between two known molecules and it also helps to find the best suitable ligand orientation which can make a complex with overall least energy [35]. These 3D positions of connected ligands can be seen by using different visualizing tools available such as pymol and RasMol [36]. They help in interpreting the best fit ligand and the result will be given in the form of a score based on docking algorithms made because of different possible combinations of the structure. In molecular docking different macromolecules like lipids, proteins, and nucleic acid are important to predict affinity between these molecules for finding a suitable drug candidate. X-ray crystallography and NMR (nuclear magnetic resonance) are used to find the structure of the macromolecules [37]. This stimulation method analyses the time-dependent behavior of macromolecules and gives information about polypeptide-based protein structure [38].

During proteomics study a different aspect of drug interaction like modifications in protein abundance, their relation partner's network, and explains cellular processes. The use of bioinformatics tools and proteomics can make analysis faster and easier. Validation of target with technology helps in drug product development [39]. The main hurdle in drug target selection is finding potential drug targets. Another important issue with drug product development is the selection of drug targets from a limited pool of potential drug targets [40]. Due to incorrect target identification, most of the drugs fail to cross the early pre-clinical stage. Support vector machine is a new technique to predict drug-target based on protein sequence properties in place of homology annotation and 3D structures. This method is 84% accurate in tenfold validation and differentiating significant drug targets from non-drug targets [41].

The tools used for target validation are

a) Gene logic- It is a foremost integrated genomics company that provides laboratory information management system solutions, and reference genomic database [42]. The database consists of gene expression data of tumors and normal cells. This target list is then screened for their functional target validation. This database also provides expression patterns of gene expression and their level of expression in different disease tissues. Researchers can use this database to study the expression pattern of different proteins which can play important role in drug discovery. Another known technology is Ribonucleic interference Intradigm, which is based on RNA interference technology (RNAi) where siRNA oligos use gene inhibitors for damaging homologs mRNA with high efficiency and specificity [43]. In disease conditions like cancer, autoimmune disease, and inflammation Intradigm focuses on the angiogenesis pathway. siRNA delivery gives high-value information in regards to studying the role of proteins or gene-related to disease processes, numerous genes of the similar pathway, and connection of different pathways to disease. Information of siRNA is also critical for drug target discovery and significant therapeutic siRNA production [44].

b) Immusol – It provides an inducible RNAi vector that can be introduced to cell culture resulting in RNAi expression. It can be used for target validation as its recently launched technology for fast and effective *in vivo* target validation [45]. The study of the inducible vector was done on xenograft tumor mouse model in which aptamers nascacells work together with aptamers that binds with the active site were small molecules of drug binds to deactivate functional

Cite this article as: Khan, J.; Singla, R.K. Bioinformatics Tools for Pharmaceutical Drug Product Development. Indo Global J. Pharm. Sci., 2022; 12: 281-294. **DOI**: http://doi.org/10.35652/IGJPS.2022.12037

epitope present on a protein without affecting other structures [46]. These aptamers mimic the small drug molecule and help in differentiating between many post-translational modifications by deactivating stable proteins using physiological turnover rate [47].

c) Lead Identification- A process that starts with compound library screening. Focusing on compounds that are related to target proteins and modulating their activities [48]. It is a complex process of drug target identification in which the confirmed hits are further optimized to be better drug candidate identification. Lead optimization is done by modification of complex structures of the compounds. Computational scrutinizing helps in performing the complex task of lead optimization in less time with more accuracy [49]. After successful lead optimization, the next big hurdle is predicting drug toxicity.

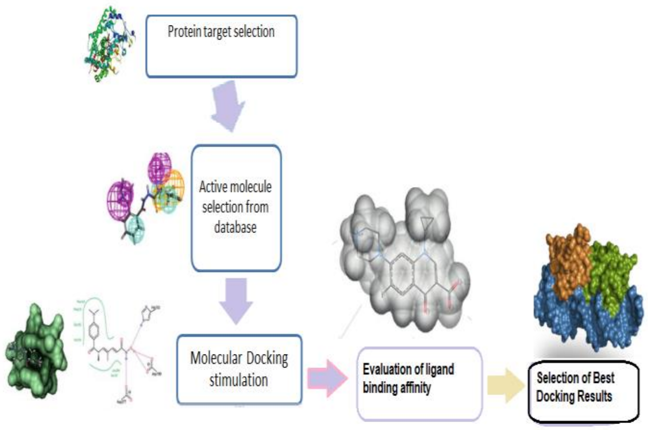
Protein target selection
Active molecule selection from database
Molecular Docking stimulation
Evaluation of ligand binding affinity
Selection of Best Docking Results

Figure 1: Flow chart of Bioinformatics tools for drug discovery.

Some of the drugs optimizing software database are:

i) Comprehensive medicinal chemistry database – This database gives important information in regards to biochemical processes of drug target class consisting of pKa, and Log P data of about 8,000 molecules [50].

ii) Drug bank- A database that provides comprehensive information of various drag targets linked with their pharmacological and chemical constituents. It also provides information on gene sequence, structure, and related pathways. This drug bank provided many drug-target data for some of the rare diseases and help in drug product development [51].

iii) PharmaGKB – A computational method or tool which can predict the reaction of a drug in comparison to variations in the human genome. It is a big pharmacogenomics database that includes information related to doses, gene-drug association, and correlation of genotype and phenotype [52].

iv) Quantitative structural activity relationship database- It is a technique used to predict biochemical properties and their activities in the human body which are not tested but have structural similarities with other drug compounds [53].

1. Web-based tools

The web-based integrating system connected with the biological network helps unscrupulously mark pharmacological properties of drugs to repurpose them. These interaction databases save various information related to drug-related entities including targets. In the web-based approach, ligand-target connections are multi-dimensional. They are also known as network-based polypharmacology and algorithm systems developed. Some of these networks are:

a) Balestra Web: It is used to predict interfaces of leads, drug targets based on AL (active learning), and collaborated filtering techniques. Operators can only discover repurposing chances by using drug-drug and target–target likeness in different tabs [54]. The drugs and drug targets approved from drug banks forms nodes of the bipartite graph. The interaction present on the edge of the bipartite graph represents known interactions that are used to study LV (latent variable vectors) expressing each drug protein. Latent variable vectors dot products of drug-target pair are the power of interfaces that can be forecasted using this method [55]. An AL method is adopted based on PMF (probabilistic matrix factor) to calculate the statistical weight of all drug targets which are associated with approved drugs. BalestraWeb depends on interaction profiles to predict any drug target without depending on their chemical or structural similarities [56]. PMF is an extensively validated method, which is validated by using 4 different classes of targets and five-fold validation. It is also reported that validation with another 2 drug target network topology-based learning algorithms shows the power of 3 algorithms differ in all four target classes [57].

b) Chemical Similarity Network Analysis Pull-down (CSNAP) – It uses an arrangement of CSN and chemical agreement to make a chemo-type based sub-network to forecast targets in different drug classes. ChEMBL, PubChem, and many similar bioactivity databases are used to recover compounds and indirectly contain bioactivity facts to match them with target hits [58]. Target annotations and query compounds are collected into CSNs and the priority of the consensus statistic is target prediction by using the frequency of target and other neighbor query compounds. CNSAP is used in cluster compounds to separate small networks which represent a specific chemotype [58]. In CNSAP the compounds are shown by cluster nodes and similarity of target by edges whereas the ligands are represented by FP2 fingerprints which are compared by Tanimoto coefficient and trials of Z-score similarity [59].

Ligands with structural diversity can also become part of a similar sub-network due to connection metric which is based on chemotypes. The neighbor compounds are ranked using S-score and the significance of every compound protein group is calculated by H-score. This method is used as a benchmark for the SEA method. It also has a high analytic ability in the case of pre-annotated target protein in 6 different classes and is specifically known for drug promiscuity. DUD (Discovery of Useful Decoys) are used to collect diverse data sets and CSNAP helps in explaining target compounds using high throughput chemical screens because they have characteristically nonspecific binding patterns [60]. A study of 212 mitotic compounds was performed using CSNAP to identify compounds with known structure in comparison to new mitotic drug targets, also to produce novel microtubules. CSNAP is considered the largest and most effective web-based tool for multivariate chemical screen profiling [61].

c) DASPfind – It uses drug-drug, drug–target, and target – target based three different subgraphs to discover new drug-target connections [62]. These connections are recovered using BRENDA (Braunschweig Enzyme Database), Drug bank, KEGG, and SuperTarget to create a new heterogeneously interconnections network that can rank new associations. The similarity is the weight of an edge between 2 drugs which is measured by using SIMCOMP (SIMilar COMPound), whereas in protein-protein interaction is calculated by using the Smith and Waterman algorithm [63]. The weight graph in DASPfind is made of nodes as drugs and proteins. DASPfind depends on a simple path to discover new connections and the score is produced by penalizing a longer path [64]. The result of these calculations is validated by HGBI (Heterogeneous Graph Based Inference) data sets by using established data sets of approved drugs from DrugBanks. The best function of DASPfind was observed with a subjective test that use 'top 1^{+}' candidates [65]. The predictive strength of any tool can only be confirmed based on a database or old literature search. Using these databases a researcher can only hypothesize new drug targets but they cannot be validated.

d) Domain Tuned Web (DT- Web) **-** It is a tool that covers approval based on bipartite network projection by combining old drug-drug, drug–target, and target-target interactions to a diverse network [66]. Then these web-based edges are connected to DT- Hybrid algorithm. This tool takes input as three matrices including a drug-drug similarity matrix using SIMCOMP (SIMilar COMPound) creating a drug similarity matrix. The similarity score of target-target protein interaction depends on the sequence similarity of these proteins. These target similarities can also be obtained using BLAST, Smith-Waterman, and the validated drug-target interactions and adjacency matrix can be obtained by Drug Bank. Every drug target interaction is made of 3 different matrices, each of these interaction networks between drug-target is mapped by using Entrez identifier and the annotations with (GO) Gene Ontology terms [67]. In ontology directed acyclic graph node distance is used to compute similarity for each pair of GO. The P-value of each drug is used to mark the interaction between the targets predicted and validated. DT- Web can predict the combinations of drugs with optimal target connotation benchmarks just by gene sequence data input. The evaluations of DT-Web are based on tenfold cross-validation followed by 30 repetitions and its performance is calculated by using precision and recollection enhancement and average AUC (Area under the ROC Curve) for 20 top predictions. Studies till now confirm the improved function of DT-Web in comparison to NBI and Hybrid [68].

e) nAnnolyze – It offers a web-based edge to network-based relative docking method known as Annolyze. In this method, only protein structures with the solved 3D structure are used [69]. In this network four important components used are; PDB (Protein Data Banks) components that apply pharmacological effects on crystallized proteins, human structural proteins from ModBase, DrugBank compounds, and LigBase based protein binding sites. nAnnolyze uses a bipartite network structural connections and similarities.

In this method, a subnetwork of ligands is made by using PDB ligands with drug-likeness above a particular threshold. Random Forest Classifier (RFC) derived from similarity is used to reduce the subnetwork to a k-core network for avoiding redundancy [70]. The protein subnetwork is made by applying targets that can bind ligands together above a drug-likeness threshold. ProBis is a network-based tool that is used to link structural similarities of binding sites with the same filtering as ligand sub-network is connected. Human structural proteomics results from purifying structures of protein by using ModBase and ProBis, also two subnetworks with known ligand-target interactions are merged using PDB. A large part of DrugBank compounds is connected by RFC for calculating similarity and edge adding to best similar compounds to ligand sub-network. Different studies on eAnnolyze use positive benchmarks made of drug-protein sets annotated between PDB and FDA (Food and Drug Administration) permitted drugs [71].

f) Promiscuous – It is the first web-based public network that can be used for repurposing. This network is made of proteins, drugs, side-effects as nodes, drug side effects, and interactions of drugs- targets, drug-drug, and protein-protein working as edges [72]. This web-based network collects data from different public databases like Uniport, SIDER (Side Effect Resource), PDB, and SuperDrug. PROMISCUOUS helps in predicting a drug target by transitive mapping but does not provide a rank or prioritize the target prediction of any kind. It provides an explorative network output [73]. Most researches showed it as the best tool for drug identification by using Memantine which is used for dementia but also repurposed for treating Parkinson. Memantine is the same as Amantadine which is used as an anti-Parkinson drug, both share NMDA glutamate. Semantic link association prediction (SLAP), forecasts the association between a drug and its target by using database incorporation and statistical modeling. SLAP function depends on path patterns that are pre-defined association paradigms comparing between nodes and edges [74]. Nodes and edges are part of a sematic network made by using protein-protein, and drug-drug similarities with drug-target interactions from Chem2Bio2RDF and ontology semantic annotations [75]. The original drug target sets built by the above connection network were recovered from DrugBank. To find the shortest path between 2 nodes of length less than 3, the Heap-based Dijkstra algorithm is used. The target predicted is ranked as P-value and related association score, which is the sum of total validation of 2 nodes and their z-scores [76].

The edge can receive inputs as drug-pair, drug-to output predicted targets, and drugs with the same biological activity and only protein and connected links. Authentication of forecast drug–target relations through SLAP was done by MATADOR. They show better presentation in comparison to other similar link forecasting methods by calculating AUROCs. These results can be compared with SEA for drug-target forecasting with CMap for predicting drug associations [75].

g) Search Tool for Interacting chemicals (STITCH) – It is the most modified version of the search tool that focuses on providing substantially broad maps of drug-target associations with the most refined filters and imaging [77]. In these years in which a huge number of the database are connected with the author or server to provide details. It provides common lines that incorporate data resources of different drug target connections starting from high throughput experiments to physically curated databases and to many analytical algorithms. Additionally, this STITCH also applied automated text mining algorithms that can forecast interactions based on the co-existence of data in different web databases like NIH RePORTER, PubMed, and MEDLINE. Each version brings in grades of selectivity and adds different resources like; users of version 5 can filter out connections created on tissue specificity [78]. Every important set of information is recorded separately and combined with statistics from text mining.

In STITCH, the confidence-based scores show the level of significance and confidence of a connection [79]. The inputs in STITCH are accepted in the form of names of chemical compounds, genes, structures of proteins, and chemical compounds as the query. The edge thickness of drug-target connections is measured according to binding affinity which shows all known Ki values. STITCH is a well-firmed source with all updates which provides the user with many published studies from different groups that directly uses result from this site [80]. Binding site parameterization is important for STITCH as these are the region present in protein structure that binds with non-bonded interaction. The binding region also has many conserved regions that can be used for identifying new protein structures and related fold families. These methods of target hunting based on binding site resemblance mapping algorithms are revised for better search results.

h) ProBis - Protein Binding site (ProBis) uses native binding sites likeness as the basic index to discover the targets matching with the query. It practices the maximum clique algorithm under the same nomenclature for physiochemical and structural properties of components and backbone of amino acids to compare two different protein binding sites [81]. In response to a quire related to protein, ProBis gives results in the form of similar binding sites, nucleic acid particles, forecasted ligands, and small molecular binding patterns. This database works as a repository for a huge number of non-redundant binding sites and related PDB structures that are updated weekly [82]. Users of this database can choose pre-calculated data to receive an immediate result. The only limitation for ProBis is it can only accept protein as a query and does not accept drugs as input [83].

i) Pocket Similarity Search using Multiple Sketches (PoSSuM) – It is a web-based search tool that is based on an algorithm that can search the complete PDB database for all similar bindings [84]. A ligand-binding region is considered significant if the result is in the form of a probe cluster with more than 200 probes. A ligand-binding site is a set of amino acids near a non-polymer molecule known as a putative binding site. PoSSuM accepts three different types of inputs which are; a) ligand-binding site, b) protein structure and c) ligand. PoSSuM, when searched with a PDB protein structure, finds all similar ligand binding sites, similarly, with ligand binding query the result will be in the form of a similar site to the input [85].

The query in PoSSuM can be studied by using ligand binding sites or putative binding sites as inputs or both can also be used. The output of the query will be in the form of one million similar binding sites. Based on geometric and physicochemical properties, the binding sites are programed as feature vectors and related sites are inserted using SketchSort, which is a fast search algorithm. A ligand can also be used as a query and the result is in the form of binding site sets which are like pockets with ligands known as bind. Measures of likeness are specified using P-value and cosine similarity [84]. The dissimilarities in ligand binding sites are represented by root mean square deviation. Similarities of all the pairs can be applied to more than three million ligand binding sites and around twenty-four million associated sites with 6 residues that configure the PoSSuM social database. PoSSuM has a limitation as its result validation is yet not obtained [55].

In this chapter, we have included most of the database for drug target identification-related servers but these are not all, a few other databases that are used to predict drug-target association are DRugome, PROteome, and DISeasome also known as DR. PRODIS which uses the FINDSITEcomb algorithm to find similarity-based targets and it also depends on assumptions that depend on evolutionarily related proteins having similar functions with binding capacity to similar types of ligands. Another is Drug E-bank that uses resemblance based features, hybrids, and descriptors with joint learning method to find the drug targets, whereas Self Organizing map based Prediction of Drug Equivalence Relationship (SPiDER) forecast the targets by using Self Organizing Maps (SOM), estimates of pharmacophore descriptors, and physiochemical properties of drug compounds.

2. Disease linked Drugs

To develop drug products, it is important to study the relation of a drug to a disease condition and the tools to study annotations dependent disease connotations. Disease-based methods are developed during the absence of pharmacology of drug or present but not considerable. Computational approaches which use drug-disease connections are researched and reviewed many times and the 2 most common types based on web access are; MeSHDD, and MEDLINE.

a) MeSHDD – MeSH based drug-drug similarity and repositioning (MeSHDD) is a cluster-based on drug-drug similarities resulting in connection which are based on disease cantered MeSH as found in MEDLINE [86]. The input is the drug name that is searched for similar matches consists of approved drugs from DrugBank. MeSH co-existed drug similarity is calculated using Bonferroni corrections and hypergeometric P-value. In this method, the drug-drug resemblance is measured by calculating the bitwise distance achieved from converting the P-values into binary symbols [87].

These methods consist of clustered drugs based on group-wise distance and mean values of bootstrap clustering methods and Jaccard index used to match clustering of different k values. The cluster of disease is measured by comparing data from TTD. MeSHDD was used to validate the discovery of Metformin against cystic fibrosis [86].

b) RE-fine drugs – This method is based on the integration of drug-gene-disease data in a transformative method to produce drug-disease connections, forecasting new suggestions for current drugs [88]. In this web-based tool, the disease is used as a query with output in the form of a list of drugs possible to be used as drug treatment. RE-fine drug tool classifies forecasted drug-disease pair as known or repurposed connections when present in drug bank. These drugs are strongly maintained if related drug data is present in NIH clinical trial registry or literature but if not found in any of these registers then considered as novel. Daclizumab is a renal disorder drug that was repurposed for asthma by using the RE-fine drug tool [55]. Drug-induced gene expression is used to compare mRNA expression in research based on cell lines to predict the drug-disease expression before it should be used for therapeutic use. Gene expression works as a disease signature that can help in describing the effect of a drug on the human system. Gene expression not only helps in understanding drug mechanisms and new biomarkers of a disease, but its signature expressions also help in identifying resemblance with other drug compounds based on likeness with their specific positive or negative expression profile in comparison to specific disease conditions and results in discovery of other repurposing drug candidates. Drug repurposing can be done by comparing disease-specific expressions signatures and biomarkers, and related pathways for inducing drug manifestation signatures that seek drugs having opposite effects on the disease condition and are effective to study drug-disease relation for repurposing and identification of drug targets [89].

c) Connective map (CMap) – It is reinforced by a cellular response database of different chemical biomarkers and their normal controls. CMap helps in providing mRNA expression data from different DNA microarrays based on research that is working on recording different gene expressions in different disease conditions to create a database with similar and reverse signature expressions [90]. The connections in these expressions are measured using Kolmogorov- Smirnov statistical test so, in the situation of reposting, CMap can classify both antagonists and agonists. The research on the CMap tool consists of different classes like HDAC inhibitors, phenothiazines produced by CMap, and estrogen which are produced or changed during different disease conditions using Different Gene Expression (DGE) data for validating the results of drug repurposing [91].

Different researchers identified various reverse drug-disease signatures, as in the case of obesity and Alzheimers induced by diet. Moreover, they delivered identification of chemical compounds from a diet that can reverse the drug resistance in a disease condition as reported in the case of acute lymphoblastic leukemia and Obesity. From the time of development of the CMap tool it has had a big impact on research on therapeutic drug-related to different diseases, also it opened a new line of research and inquiry in the field of drug reposting, target and lead discovery, MoA explanation, system biology, and biological consideration. It provides a very effective and direct method of research investigation in the therapeutic potential of drugs, also its CMap dependent approach has been widely searched by different groups of researchers in the field of drug product discovery and repurposing of old drugs for different [92].

d) Differentially Expressed Gene Signatures- inhibitors (DeSigN) - CMap and DeSigN both function on the same basic principle that is the disease signatures in response of drug mechanism associated with gene signature based on IC50 data. DeSigN is made by using GDSC [93]. CMap and DeSigN both use gene expression profiles and DeSigN uses baseline gene expression profiles that can be tested by using 4 GEO studies and the collective score of these studies with drug response was found consistent with published research on GEO studies [94].

e) Go-Predict- It is a tool used to integrate data from different public data sources including signaling pathway databases and drug–target-related information with cancer genomic data to use this information for purposes of significant drugs effective on gene expression. Go-Predict uses gene expression as input and in response, the output is in the form of similar drug targets. The databases used as a reference in Go-Predict are Gene Ontology, TCGA, DrugBank, and KEGGDrug [95].

Go-Predict measures gene rank linked to its effect on the regulation of different pathways. In this database, the gene-drug set is arranged based on particular GO processes to validate the drug ranks of all the genes regulated by that pathway. The researchers and authors have produced novel drug-DGE linkages which are also reported in different literature to validate drug ranks [96].

f) L1000CDS – It is a web-based interface that connects and uses the CD (Characteristic Direction) signatures data of LINCS-L100 to forecast new signals. CMap and many other databases used diluted Z-score [97]. The multivariate method and CD are more complex to identify DGE. CDS measures the angle between the input gene signature and LINCS-1000 data to make a list of possible aspirant molecules that can reverse and sometimes mimic the query gene and its expression. The researchers working on L1000CDS have forecasted candidates for disease signature from the GEO database. Moreover, they also anticipated a drug known as Kenpaullone to be effective against the Ebola virus and offer many related studies and investigations to support their claim [98].

g) Mode of Action by NeTwoRK (MANTRA 2.0) – Different molecular targets for drug signature can be identified using MANTRA 2.0 using gene expression profile before and after uploading drug perturbation that gets fixed into a cooperative learning environment [99]. A network made of the visual database with a new node helps the user find a nearby neighbor to discover new hints. They input a prototype ranked list (PRL) for a drug to compare between two different PRLs by using GSEA (Gene Set Ensemble Approach) method. It provides a collective, investigative, environment and opportunity for the users to distribute their data in other databases and to different users [100].

h) NFFinder – Even though there are many databases to compare and validate studies related to drug-gene interaction, NFFinder uses the MARQ technique to associate the signatures of gene expression [101]. In this method, two sets of up and down-regulated genes are successively compared with GEO, DrugMatrix, and CMap data. They follow a two-step validation method, first is the TCA (Trichostatin A) method, which is found effective in destroying MPNST (Malignant Peripheral Nerve Sheath Tumors), and the next is recovered TCA as a target hit during the gene expression profile study of a known tumor cells suppressor cocktail (PD901/JQ1) besides MPNST cells which are used as query [102].

i) Prediction of Drugs having Opposite effects on Disease genes (PDOD) – As many database and web-based tools uses gene expression as signatures, PDOD focuses on effect-direction and effect- type using KEGG, drug target information from GEO and DrugBank for communication-based data to find any possible drug is available that can pay off for differentially measured disease genes [103]. Limma was used to discover the differentially expressed genes with a role they established to estimate drug-disease score based on parameterizing the relation between drugs – drug relation. The studies on this database cannot predict drugs products for all diseases as the prediction depends on the availability of data in different databases. As many researchers concluded these tools are successful only against selective disease classes [104].

3. Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) studies

ADMET studies are based on absorption, distribution, metabolism, excretion, and toxicity components of a drug product development. The entire failed drug targets are found to fulfil the ADMET criteria. Drug candidate discovery needs ADMET profiling and its early profiling helps in reducing the risk of study attrition. Different medium in vitro ADMET screening method was developed to contribute to data analysis at an early stage of identification of drug targets. It is an expensive method in comparison to other tools especially when the numbers of compounds to be screened are in large numbers [105, 106]. ADMET is more important as it reduces animal testing, which is a priority of research. Various *in silico* tools are developed for facilitating fast and economical means of ADMET profiling (**Figure 2**). By focusing on experimental properties of ADMET different QSAR (quantitative structure-activity/property relationship) models were generated which can forecast different ADMET properties for new chemical components [107]. Various other methods used ADMET-based predictions for evaluating drug similarities of a compound, whereas other models are used as part of profitable software sets based on exclusive datasets. A significant need for open basis software was found even though much soft wear is commercially available [108].

ADMET Lab is one of the famous services which offer fifty-three different predictions that are measured by using a multi-task and operational graph network structural data. This technique can make modified fingerprints using general characters of a specific assignment [109]. SwissADME is a modified form of ADMET used to assess the pharmacokinetics and drug similarity of a small molecule or compound [110]. These forecasts are based on the arrangement of fragmental techniques and on machine learning-dependent binary classification techniques to consider other ADMET-related properties.

Cite this article as: Khan, J.; Singla, R.K. Bioinformatics Tools for Pharmaceutical Drug Product Development. Indo Global J. Pharm. Sci., 2022; 12: 281-294. **DOI**: http://doi.org/10.35652/IGJPS.2022.12037

ADMETSar is a model-based application used for drug discovery and environmental risk assessments are made using Morgan fingerprints and MACCS. ProTox uses toxicity models based on the chemical likeness between compounds that are already identified as toxic effects and toxic fragments. Morgan and other similar models for mutagenicity, hepatotoxicity, mutagenicity, and carcinogenicity depend on fingerprints [111]. A protracted connective fingerprint makes the basis for the forecast of fifteen ADMET characteristics in the vNN server in which models are qualified by adaptable to nearest neighbor technique. pkCSM is another tool related to ADMET that uses graph-based gene or protein signature to grow prognostic models of dominant ADMET characteristics. Some other software tools like CarcinoPred-EL, CapsCarcino, and MDCKPred emphasize only one property that is forecasting of permeability constant and carcinogenic complexes. All these models are based on molecular depiction with a basis of different physiochemical molecular descriptions like fingerprints, 2D or 3D, and graph signatures. All these models are very famous and found as an alternative for QSPR studies, which gave computational tools easy and user-friendly for prediction and repurposing easily [112].

During the last decade, a large number of fingerprinting including feature circular and path range were described and developed but ADMET-based fingerprints studies are very selective and limited. Some studies used this method to calculate the efficacy of 20 different types of fingerprints going from structural to extended connectivity and different path-centered encoding like; depth-first search, local path environment, and shortest path types. Some research used fingerprint-centered regression models to calculate fifty ADMET and related endpoints by data collected from different literature sources, which is one of the most inclusive compilations studied to date. The endpoint results of most of the predictions were found analogous with other classy descriptor preparations. The drawback of ADMET studies is constant poor fingerprint results in comparison to other studies using PubChem, ECFP, and MACCS, where encoding was found to give better results for most of the drug product-related properties. These software and other related tools are created in a downloadable pack under the license of GNU.

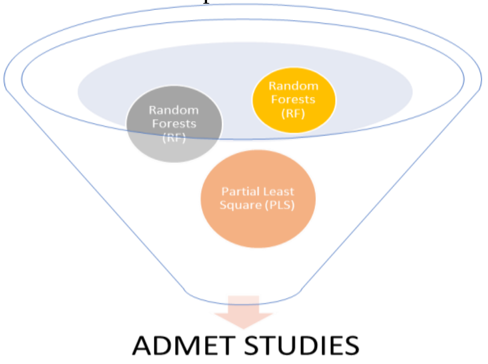

Figure 2: Different tools of ADMET studies for drug repurposing method.

DATA MODELLING

Data modeling is one of the important parts of studying the classification and regression of drug candidates. Random Forest Algorithm (RFA) was selected to build the model. In algorithm bagging and feature, randomness is used to build multiple decision trees and to connect them [113]. The training of these models was done using the ranger library in arithmetic computing environment R. The average number of forecast values of trees used for computing in modeling was fixed at 500. To each endpoint, the data was divided into different sets of 80 and 20% test sets. For cross-validation, a fivefold method was used to classify the finest performing model, and to avoid any choice bias researchers repeated tests were randomly repeated 3 times and an average result is considered to understand the variability. The *Y*-randomization method can be conducted to calculate the robustness of the ultimate model. To solve the problems related to the unequal spreading of samples in different classes and the data extension of the minority can be carried out by using the synthetic minority over SMOTE (sampling minority oversampling technique) [114].

In the regression model, their performances were measured by using squared regression coefficient (R_2R_2) for correlating values of experiment and forecast. Root mean square error (RMSE) and mean absolute error (MAE) are used in classification models and metrics which are sensitive to class inequality. Each model consists of a fixed applicability domain (AD) in the limits of which its forecast can be trusted. In regression model prediction intervals can be measured using the quantile regression prediction method [115]. In this technique, the smallest forecast interval shows their highest stability predictions. In classification techniques, confidence and credibility are two important values that are related to forecasted labels based on CPF (conformal prediction framework). Confidence in modeling provides a degree of likeness of forecasting with which the comparison of all other classifications measures provides a sign of a good training method and sets for classifying the highest *p*-value of one or more probable organizations of true labels.

CONCLUSION

Bioinformatics and web-based tools can facilitate drug discovery and drug product development. Drug product development depends on predicting drug-target connections and validations. These can be done by using algorithms, computer-aided drug designing (CADD), and computational chemistry. Bioinformatics tools help in repurposing drugs to reduce the time, money, and effort needed to develop new drug products. These tools of bioinformatics also help in big data including transcriptomics, gene sequence data, and proteomics. Bioinformatics tools need more improvement for the analysis of high throughput pangenomic, protomics, metabolomics and metagenomimc data. The bioinformatics effective tools are required for better genomic assembly and annotation with high accuracy, to improve quality of sequenced genomics without gaps, sub-genomic, polypeptide species, and genomes of single cells [116]. Pharmacogenomics and bioinformatics are still in a developmental phase and the tools for drug target prediction have many limitations and hurdles but they show huge potential to help in drug product development even to be patient-specific in near future.

ACKNOWLEDGEMENT

Authors would like to express a special thanks to Majmaah University for providing facilities and resources to carry out an extensive literature survey on the presented topic.

AUTHOR'S CONTRIBUTION

Manuscript was prepared, revised and submitted by both the authors.

ETHICS STATEMENT

The authors have taken all the necessary permissions as per ethical guidelines wherever applicable. The authors will be responsible for all the technical content mentioned in the manuscript. Journal and Publisher will not be responsible for any copyright infringement and plagiarism issue.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY

All the key information is already available in the manuscript, still, authors are ready to share the raw data, if the proper channel for the inquiry will be followed which will be routed through journal and affiliation authorities.

FUNDING SOURCE

No external funding source has been disclosed.

REFERENCES

1. Power, A., A.C. Berger, and G.S. Ginsburg, Genomics-enabled drug repositioning and repurposing: insights from an IOM Roundtable activity. Jama, 2014. **311**(20): p. 2063-2064.
2. Scannell, J.W., et al., Diagnosing the decline in pharmaceutical R&D efficiency. Nature reviews Drug discovery, 2012. **11**(3): p. 191-200.
3. Padhy, B. and Y. Gupta, Drug repositioning: re-investigating existing drugs for new therapeutic indications. Journal of postgraduate medicine, 2011. **57**(2): p. 153.
4. Buchan, N.S., et al., The role of translational bioinformatics in drug discovery. Drug discovery today, 2011. **16**(9-10): p. 426-434.
5. Van Driel, M.A. and H.G. Brunner, Bioinformatics methods for identifying candidate disease genes. Human genomics, 2006. **2**(6): p. 1-4.
6. Josset, L., et al., Gene expression signature-based screening identifies new broadly effective influenza a antivirals. PloS one, 2010. **5**(10): p. e13169.
7. Shameer, K., et al., Systematic analyses of drugs and disease indications in RepurposeDB reveal pharmacological, biological and epidemiological factors influencing drug repositioning. Briefings in bioinformatics, 2018. **19**(4): p. 656-678.
8. Li, J., et al., A survey of current trends in computational drug repositioning. Briefings in bioinformatics, 2016. **17**(1): p. 2-12.
9. Dai, Y.-F. and X.-M. Zhao, A survey on the computational approaches to identify drug targets in the postgenomic era. BioMed research international, 2015. **2015**.
10. March-Vila, E., et al., On the integration of in silico drug design methods for drug repurposing. Frontiers in pharmacology, 2017. **8**: p. 298.
11. Corsello, S.M., et al., The Drug Repurposing Hub: a next-generation drug library and information resource. Nature medicine, 2017. **23**(4): p. 405-408.
12. Wang, X., et al., PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. Nucleic acids research, 2017. **45**(W1): p. W356-W360.
13. Li, C., et al., Applications of three-dimensional printing in surgery. Surgical innovation, 2017. **24**(1): p. 82-88.
14. Jacobo, O.M., et al., Three-dimensional printing modeling: application in maxillofacial and hand fractures and resident training. European Journal of Plastic Surgery, 2018. **41**(2): p. 137-146.
15. Papanikolaw, J., Bioinformatics emerges as key technology for developing new drugs. Chemical Market Reporter; May 24, 1999; 255, 21; ABI/INFORM Global pg, 1999. **22**.
16. Wittenberger, T., H.C. Schaller, and S. Hellebrand, An expressed sequence tag (EST) data mining strategy succeeding in the discovery of new G-protein coupled receptors. Journal of molecular biology, 2001. **307**(3): p. 799-813.
17. Marvanova, M., et al., Synexpression analysis of ESTs in the rat brain reveals distinct patterns and potential drug targets. Molecular brain research, 2002. **104**(2): p. 176-183.
18. Nayal, M. and B. Honig, On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. Proteins: Structure, Function, and Bioinformatics, 2006. **63**(4): p. 892-906.
19. Mehlin, C., et al., Heterologous expression of proteins from Plasmodium falciparum: results from 1000 genes. Molecular and biochemical parasitology, 2006. **148**(2): p. 144-160.
20. Dong, Y., et al., Anopheles gambiae immune responses to human and rodent Plasmodium parasite species. PLoS pathogens, 2006. **2**(6): p. e52.

21. Dudley, J.T., T. Deshpande, and A.J. Butte, Exploiting drug–disease relationships for computational drug repositioning. Briefings in bioinformatics, 2011. **12**(4): p. 303-311.
22. Veber, D.F., et al., Molecular properties that influence the oral bioavailability of drug candidates. Journal of medicinal chemistry, 2002. **45**(12): p. 2615-2623.
23. Spangenberg, T., et al., The open access malaria box: a drug discovery catalyst for neglected diseases. PloS one, 2013. **8**(6): p. e62906.
24. Garofalo, A.W., Patents targeting γ-secretase inhibition and modulation for the treatment of Alzheimer's disease: 2004–2008. Expert Opinion on Therapeutic Patents, 2008. **18**(7): p. 693-703.
25. Ou-Yang, S.-s., et al., Computational drug discovery. Acta Pharmacologica Sinica, 2012. **33**(9): p. 1131-1140.
26. Sliwoski, G., et al., Computational methods in drug discovery. Pharmacological reviews, 2014. **66**(1): p. 334-395.
27. Schaduangrat, N., et al., Towards reproducible computational drug discovery. Journal of cheminformatics, 2020. **12**(1): p. 1-30.
28. Gong, J., et al., ChemMapper: a versatile web server for exploring pharmacology and chemical structure association based on molecular 3D similarity method. Bioinformatics, 2013. **29**(14): p. 1827-1829.
29. Liu, X., H. Jiang, and H. Li, SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. Journal of chemical information and modeling, 2011. **51**(9): p. 2372-2385.
30. Lu, W., et al., SHAFTS: a hybrid approach for 3D molecular similarity calculation. 2. Prospective case study in the discovery of diverse p90 ribosomal S6 protein kinase 2 inhibitors to suppress cell migration. Journal of medicinal chemistry, 2011. **54**(10): p. 3564-3574.
31. Liu, X., et al., PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. Nucleic acids research, 2010. **38**(suppl_2): p. W609-W614.
32. Taboureau, O., et al., ChemProt: a disease chemical biology database. Nucleic acids research, 2010. **39**(suppl_1): p. D367-D372.
33. Kim Kjærulff, S., et al., ChemProt-2.0: visual navigation in a disease chemical biology database. Nucleic acids research, 2012. **41**(D1): p. D464-D469.
34. Kringelum, J., et al., ChemProt-3.0: a global chemical biology diseases mapping. Database, 2016. **2016**.
35. Rose, P.W., et al., The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic acids research, 2016: p. gkw1000.
36. Rose, P.W., et al., The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic acids research, 2015. **43**(D1): p. D345-D356.
37. Pagadala, N.S., K. Syed, and J. Tuszynski, Software for molecular docking: a review. Biophysical reviews, 2017. **9**(2): p. 91-102.
38. Hernández-Santoyo, A., et al., Protein-protein and protein-ligand docking. Protein engineering-technology and application, 2013: p. 63-81.
39. Zheng, H., et al., X-ray crystallography over the past decade for novel drug discovery–where are we heading next? Expert opinion on drug discovery, 2015. **10**(9): p. 975-989.
40. Eweas, A.F., I.A. Maghrabi, and A.I. Namarneh, Advances in molecular modeling and docking as a tool for modern drug discovery. Der Pharma Chemica, 2014. **6**(6): p. 211-228.
41. del Carmen Fernández-Alonso, M., et al., Protein-carbohydrate interactions studied by NMR: from molecular recognition to drug design. Current Protein and Peptide Science, 2012. **13**(8): p. 816-830.
42. Lipfert, J. and S. Doniach, Small-angle X-ray scattering from RNA, proteins, and protein complexes. Annu. Rev. Biophys. Biomol. Struct., 2007. **36**: p. 307-327.
43. Schiffelers, R.M., et al., Effects of treatment with small interfering RNA on joint inflammation in mice with collagen-induced arthritis. Arthritis & Rheumatism, 2005. **52**(4): p. 1314-1318.
44. Xu, J., Strategic Research Institute--first international siRNA conference. Prospect for new therapeutics and commercial opportunities for pharma and biotech. 24-25 March 2003, LaJolla, CA, USA. IDrugs: the investigational drugs journal, 2003. **6**(5): p. 449-450.
45. Xie, F.Y., M.C. Woodle, and P.Y. Lu, Harnessing in vivo siRNA delivery for drug discovery and therapeutic development. Drug discovery today, 2006. **11**(1-2): p. 67-73.
46. Iorns, E., et al., A new mouse model for the study of human breast cancer metastasis. PloS one, 2012. **7**(10): p. e47995.
47. Gondi, C.S. and J.S. Rao, Concepts in in vivo siRNA delivery for cancer therapy. Journal of cellular physiology, 2009. **220**(2): p. 285-291.
48. Gamo, F.-J., et al., Thousands of chemical starting points for antimalarial lead identification. Nature, 2010. **465**(7296): p. 305-310.
49. Singh, J., et al., Application of genetic algorithms to combinatorial synthesis: A computational approach to lead identification and lead optimization. Journal of the American Chemical Society, 1996. **118**(7): p. 1669-1676.
50. Nicola, G., T. Liu, and M.K. Gilson, Public domain databases for medicinal chemistry. Journal of medicinal chemistry, 2012. **55**(16): p. 6987-7002.

Cite this article as: Khan, J.; Singla, R.K. Bioinformatics Tools for Pharmaceutical Drug Product Development. Indo Global J. Pharm. Sci., 2022; 12: 281-294. **DOI**: http://doi.org/10.35652/IGJPS.2022.12037

51. Wishart, D.S., et al., DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research, 2008. **36**(suppl_1): p. D901-D906.

52. Thorn, C.F., T.E. Klein, and R.B. Altman, PharmGKB: the pharmacogenomics knowledge base, in Pharmacogenomics. 2013, Springer. p. 311-320.

53. Zhao, Y., et al., Toxicity of ionic liquids: database and prediction via quantitative structure–activity relationship method. Journal of hazardous materials, 2014. **278**: p. 320-329.

54. Cobanoglu, M.C., et al., BalestraWeb: efficient online evaluation of drug–target interactions. Bioinformatics, 2015. **31**(1): p. 131-133.

55. Sam, E. and P. Athri, Web-based drug repurposing tools: a survey. Briefings in bioinformatics, 2019. **20**(1): p. 299-316.

56. Malik, S.I., et al. Mathematical Modeling and Docking of Medicinal Plants and Synthetic drugs to determine their effects on Abnormal Expression of Cholinesterase and Acetyl Cholinesterase Proteins in Alzheimer. in International Work-Conference on Bioinformatics and Biomedical Engineering. 2019. Springer.

57. Munir, A., et al., In silico repositioning of alendronate and cytarabine drugs to cure mutations of FPPS, HAP, PTPRS, PTPRE, PTN4, GGPPS gene and mutant DNA, DPOLB, TOP2a, DPOLA, DNMT, RNA, TYSY, RIR genes. International Journal Bioautomation, 2016. **20**(3): p. 317.

58. Lo, Y.-C., et al., 3D chemical similarity networks for structure-based target prediction and scaffold hopping. ACS chemical biology, 2016. **11**(8): p. 2244-2253.

59. Cherkasov, A., et al., Mapping the protein interaction network in methicillin-resistant Staphylococcus aureus. Journal of proteome research, 2011. **10**(3): p. 1139-1150.

60. Lo, Y.-C., et al., Machine learning in chemoinformatics and drug discovery. Drug discovery today, 2018. **23**(8): p. 1538-1546.

61. Rzuczek, S.G., M.R. Southern, and M.D. Disney, Studying a drug-like, RNA-focused small molecule library identifies compounds that inhibit RNA toxicity in myotonic dystrophy. ACS chemical biology, 2015. **10**(12): p. 2706-2715.

62. Ba-Alawi, W., et al., DASPfind: new efficient method to predict drug–target interactions. Journal of cheminformatics, 2016. **8**(1): p. 1-9.

63. Zaki, N., et al., Protein-protein interaction based on pairwise similarity. BMC bioinformatics, 2009. **10**(1): p. 1-12.

64. Xu, Q., E.W. Xiang, and Q. Yang. Protein-protein interaction prediction via collective matrix factorization. in 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2010. IEEE.

65. Skrabanek, L., et al., Computational prediction of protein–protein interactions. Molecular biotechnology, 2008. **38**(1): p. 1-17.

66. Alaimo, S., et al., DT-Web: a web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference. BMC systems biology, 2015. **9**(3): p. 1-11.

67. Alaimo, S., et al., Drug–target interaction prediction through domain-tuned network-based inference. Bioinformatics, 2013. **29**(16): p. 2004-2008.

68. Alaimo, S., R. Giugno, and A. Pulvirenti, Recommendation techniques for drug–target interaction prediction and drug repositioning, in Data Mining Techniques for the Life Sciences. 2016, Springer. p. 441-462.

69. Martínez-Jiménez, F. and M.A. Marti-Renom, Ligand-target prediction by structural network biology using nAnnoLyze. PLOS computational biology, 2015. **11**(3): p. e1004157.

70. Azar, A.T., et al., A random forest classifier for lymph diseases. Computer methods and programs in biomedicine, 2014. **113**(2): p. 465-473.

71. Chaudhary, A., S. Kolhe, and R. Kamal, An improved random forest classifier for multi-class classification. Information Processing in Agriculture, 2016. **3**(4): p. 215-222.

72. Burkhardt, H.A., et al. Predicting adverse drug-drug interactions with neural embedding of semantic predications. in AMIA Annual Symposium Proceedings. 2019. American Medical Informatics Association.

73. Von Eichborn, J., et al., PROMISCUOUS: a database for network-based drug-repositioning. Nucleic acids research, 2010. **39**(suppl_1): p. D1060-D1066.

74. Fu, G., et al., Predicting drug target interactions using meta-path-based semantic network analysis. BMC bioinformatics, 2016. **17**(1): p. 1-10.

75. Chen, B., Y. Ding, and D.J. Wild, Assessing drug target association using semantic linked data. PLoS computational biology, 2012. **8**(7): p. e1002574.

76. Cheng, T., et al., Large-scale prediction of drug-target interaction: a data-centric review. The AAPS journal, 2017. **19**(5): p. 1264-1275.

77. Kuhn, M., et al., STITCH: interaction networks of chemicals and proteins. Nucleic acids research, 2007. **36**(suppl_1): p. D684-D688.

78. Kuhn, M., et al., STITCH 4: integration of protein–chemical interactions with user data. Nucleic acids research, 2014. **42**(D1): p. D401-D407.

79. Gao, Y.-F., et al., Predicting metabolic pathways of small molecules and enzymes based on interaction information of chemicals and proteins. 2012.

80. Szklarczyk, D., et al., STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. Nucleic acids research, 2016. **44**(D1): p. D380-D384.

Cite this article as: Khan, J.; Singla, R.K. Bioinformatics Tools for Pharmaceutical Drug Product Development. Indo Global J. Pharm. Sci., 2022; 12: 281-294. **DOI**: http://doi.org/10.35652/IGJPS.2022.12037

81. Konc, J. and D. Janežič, ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. Bioinformatics, 2010. **26**(9): p. 1160-1168.

82. Konc, J., et al., ProBiS-CHARMMing: web interface for prediction and optimization of ligands in protein binding sites. 2015, ACS Publications.

83. Jukič, M., et al., ProBiS H2O MD approach for identification of conserved water sites in protein structures for drug design. ACS medicinal chemistry letters, 2020. **11**(5): p. 877-882.

84. Ito, J.-I., et al., PoSSuM: a database of similar protein–ligand binding and putative pockets. Nucleic acids research, 2012. **40**(D1): p. D541-D548.

85. Ito, J.-i., et al., PoSSuM v. 2.0: data update and a new function for investigating ligand analogs and target proteins of small-molecule drugs. Nucleic acids research, 2015. **43**(D1): p. D392-D398.

86. Brown, A.S. and C.J. Patel, MeSHDD: literature-based drug-drug similarity for drug repositioning. Journal of the American Medical Informatics Association, 2017. **24**(3): p. 614-618.

87. Zeng, X., et al., Measure clinical drug–drug similarity using electronic medical records. International journal of medical informatics, 2019. **124**: p. 97-103.

88. Moosavinasab, S., et al., 'RE: fine drugs': an interactive dashboard to access drug repurposing opportunities. Database, 2016. **2016**.

89. Shukla, R., et al., Signature-based approaches for informed drug repurposing: Targeting CNS disorders. Neuropsychopharmacology, 2021. **46**(1): p. 116-130.

90. Musa, A., et al., A review of connectivity map and computational approaches in pharmacogenomics. Briefings in bioinformatics, 2018. **19**(3): p. 506-523.

91. Cheng, J., et al., Systematic evaluation of connectivity map for disease indications. Genome medicine, 2014. **6**(12): p. 1-8.

92. Qu, X.A. and D.K. Rajpal, Applications of Connectivity Map in drug discovery and development. Drug discovery today, 2012. **17**(23-24): p. 1289-1298.

93. Lee, B.K.B., et al., DeSigN: connecting gene expression with therapeutics for drug repurposing and development. BMC genomics, 2017. **18**(1): p. 1-11.

94. Jones, J., et al., Gene signatures of progression and metastasis in renal cell cancer. Clinical cancer research, 2005. **11**(16): p. 5730-5739.

95. Dabra, R., T.R. Singh, and R.M. Yennamalli, In Silico Screening of Putative Drug Mplecules to Target MSI Pathway for Colorectal Cancer and HNPCC. 2017.

96. Louhimo, R., Biomedical Data Integration in Cancer Genomics. 2015.

97. Duan, Q., et al., L1000CDS 2: LINCS L1000 characteristic direction signatures search engine. NPJ systems biology and applications, 2016. **2**(1): p. 1-12.

98. Clark, N.R., et al., The characteristic direction: a geometrical approach to identify differentially expressed genes. BMC bioinformatics, 2014. **15**(1): p. 1-16.

99. Carrella, D., et al., Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis. Bioinformatics, 2014. **30**(12): p. 1787-1788.

100. Iorio, F., et al., Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 2010. **107**(33): p. 14621-14626.

101. Setoain, J., et al., NFFinder: an online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning. Nucleic acids research, 2015. **43**(W1): p. W193-W199.

102. Vazquez, M., et al., MARQ: an online tool to mine GEO for experiments with similar or opposite gene expression signatures. Nucleic acids research, 2010. **38**(suppl_2): p. W228-W232.

103. Yu, H., et al. Prediction of drugs having opposite effects on disease genes in a directed network. in BMC systems biology. 2016. Springer.

104. Fang, M., et al., Drug perturbation gene set enrichment analysis (dpGSEA): a new transcriptomic drug screening approach. BMC bioinformatics, 2021. **22**(1): p. 1-14.

105. Davis, A.M. and R.J. Riley, Predictive ADMET studies, the challenges and the opportunities. Current opinion in chemical biology, 2004. **8**(4): p. 378-386.

106. Rudrapal, M., et al., Repurposing of phytomedicine-derived bioactive compounds with promising anti-SARS-CoV-2 potential: Molecular docking, MD simulation and drug-likeness/ADMET studies. Saudi journal of biological sciences, 2021.

107. Tareq Hassan Khan, M., Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. Current drug metabolism, 2010. **11**(4): p. 285-295.

108. Khan, M.T. and I. Sylte, Predictive QSAR modeling for the successful predictions of the ADMET properties of candidate drug molecules. Current drug discovery technologies, 2007. **4**(3): p. 141-149.

109. Yang, H., et al., admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. Bioinformatics, 2019. **35**(6): p. 1067-1069.

110. Daina, A., O. Michielin, and V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific reports, 2017. **7**(1): p. 1-13.

Cite this article as: Khan, J.; Singla, R.K. Bioinformatics Tools for Pharmaceutical Drug Product Development. Indo Global J. Pharm. Sci., 2022; 12: 281-294. **DOI**: http://doi.org/10.35652/IGJPS.2022.12037

111. Mahanthesh, M., et al., Swiss ADME prediction of phytochemicals present in Butea monosperma (Lam.) Taub. J. Pharmacogn. Phytochem, 2020. **9**: p. 1799-1809.

112. Mishra, S. and R. Dahima, In vitro ADME studies of TUG-891, a GPR-120 inhibitor using SWISS ADME predictor. Journal of Drug Delivery and Therapeutics, 2019. **9**(2-s): p. 366-369.

113. Kumar, M.S., et al. Credit card fraud detection using random forest algorithm. in 2019 3rd International Conference on Computing and Communications Technologies (ICCCT). 2019. IEEE.

114. Chawla, N.V., et al., SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 2002. **16**: p. 321-357.

115. Chicco, D., M.J. Warrens, and G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 2021. **7**: p. e623.

116. Chavda, Vivek P., et al. Advanced computational methodologies used in the discovery of new natural anticancer compounds. Frontiers in Pharmacology, 2021. 12.

Indo Global Journal of Pharmaceutical Sciences(ISSN 2249 1023; CODEN- IGJPAI; NLM ID: 101610675) indexed and abstracted in CrossRef (DOI Enabling), CNKI, EMBASE (Elsevier), National Library of Medicine (NLM) Catalog (NCBI), ResearchGate, Publons (Clarivate Analytics), CAS (ACS), Index Copernicus, Google Scholar and many more. For further details, visit http://iglobaljournal.com

Cite this article as: Khan, J.; Singla, R.K. Bioinformatics Tools for Pharmaceutical Drug Product Development. Indo Global J. Pharm. Sci., 2022; 12: 281-294. **DOI**: http://doi.org/10.35652/IGJPS.2022.12037