Computational Approaches for Drug Target Identification

8

Pramod Katara

Abstract

It is assumed that due to the enormous investment in terms of time, money, human volunteers, and other resources, sometimes failure at the later stage mostly put pharmaceutical companies on the back foot. For the last two decades, pharmaceutical companies felt that the traditional drug designing process should be optimized to avoid huge financial loss and save time. Thus, despite its limitations, the use of computer-aided drug design (CADD) techniques in drug discovery and development process is successful. CADD approaches support almost all phases of the drug designing process, including drug target identification, lead identification, optimization of leads, and simulations. Drug target identification and characterization is a first and most essential step that begins with identifying the function of a possible molecular target (gene/protein) and its role in the disease. The availability of the huge amount of molecular data, i.e., big data, for human as well as pathogens with applications of knowledge-based data mining approaches can provide a list of probable drug targets which further can be validated through experiments can save time and cost of pharmaceutical companies and boost their research towards the development of new drugs. This chapter focuses on the computational approaches for drug target identification, which play a crucial role in the drug discovery and development process.

Keywords

Algorithm · Database · Drug target · Drug designing · Druggability · Biological network

P. Katara (✉)
Computational Omics Lab, Centre of Bioinformatics, University of Allahabad, Prayagraj, India

D. B. Singh (ed.), *Computer-Aided Drug Design*,
https://doi.org/10.1007/978-981-15-6815-2_8

8.1 Introduction

Drug designing deals with the discovery and development of therapeutic molecules for a drug target. The drug is a small molecule that has potential to modulate the function of drug targets, such as a protein and sometimes nucleic acid tool, i.e., regulatory RNAs (Dersch et al. 2017). Drug design involves the design of molecules that are complementary in shape to the chosen drug target and modulate in the desired manner (Zauhar et al. 2003). Nowadays various drug designing approaches are in practice, broadly they can be classified into two types: (1) traditional methods: traditional methods involve trial and error method of testing for chemicals on cultured or animals cell, and observe the outcome of treatments, and (2) rational drug design: this approach is based on the hypothesis that modulation of a specific biological target which will be considered as drug targets, may have therapeutic value. In this approach, a potential therapeutic target is identified and purified. The purified protein is used to develop a screening assay. In rational drug design, 3D structure of the drug target should be available. The small bioactive searched by screening libraries of a drug or bioactive compound. This can also be performed by the screening assay, which also known as chemical or wet screening assay.

Nowadays computational methods are also in practice to screen compounds virtually and are well known as virtual screening (McInnes 2007). After library screening, the molecules are subjected to biological screening to test toxicity and those who show positive screening enter into the clinical trials where they try on human volunteers/patients to check pharmacokinetics (ADMET) of the drug. In the case of the successful completion of the clinical trials, a molecule passes to the approval agency and then finally hits the market (Fig. 8.1). This whole drug designing process is very time consuming and expensive, and at any stage of the process, a lead molecule can fail. Failure of leads at a later stage is responsible for the loss of millions of dollars for pharmaceutical companies (Hughes et al. 2011).

To reduce the chance of later-failure and speed up the molecular screening process, computational approaches are in practice for the last one and a half decade. Nowadays, designing drug using computational approaches is well known as computer-aided drug designing (CADD). CADD involves various approaches such as QSAR, virtual screening, docking, etc. (Katsila et al. 2016). Computational approaches have speed up the process of drug discovery and have provided novel drug targets and lead structures (Katara 2013). The computational method can identify drug targets and leads against them, affinity and efficacy between them before clinical trials and saving enormous time and cost (Shekhar 2008; Katara 2017).

8.2 Drug Targets

The term drug target describes the native biomolecule in the human body whose function can be modulated by a drug molecule, which may have a therapeutic effect against the disease or some adverse effect. Mostly these drug targets are biological

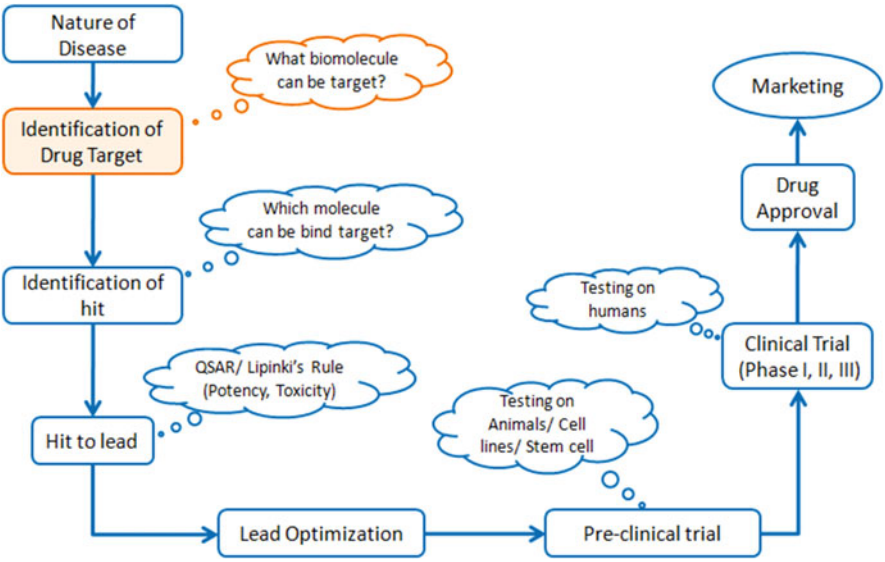

Fig. 8.1 The flow of drug designing process (Katara 2017)

Table 8.1 Details of frequently used drug target protein classes

S. No.	Target classes	Description
1	GPCRs	G protein-coupled receptors (GPCRs) play a central role in various signal transduction pathways responsible for cellular responses. Due to its indispensable role, GPCRs make up a large portion of the targets of approved drugs. Presently, more than hundreds of GPCRs are already in practice as targets of ~34% FDA approved drugs (Sriram and Insel 2018)
2	Ion channels	Ion channels play a very crucial role in controlling a very wide range of physiological processes in humans, and their dysfunction can lead to abnormalities, thus they are reported as one of the important drug targets (Kaczorowski et al. 2008)
3	Kinases	Kinase plays a pivotal role in the regulation of many cellular and biological processes. Abnormal kinase activity has been well reported to be linked with a variety of diseases and human cancers (Cohen 2002; Klaeger et al. 2017)
4	Proteases	Deficient or abnormal protease function is linked with many pathological conditions. An estimated 5–10% of all drugs under progress target the proteases (Docherty et al. 2003)

targets in nature. Various protein drug targets are currently utilized by available drugs, most of them belong to one of four major drug target protein classes (Table 8.1), in some cases, nucleic acids are also utilized by drugs as a target.

8.3 Drug Target Identification

After identifying the biological nature and origin of a disease, identification of potential drug targets is the first step in the discovery of a drug. Drug target identification follows the hypothesis that the most promising targets are tightly linked to the disease of interest, and have an established function in the underlying pathology, which can be observed with high frequency in the disease-associated population. By definition, it is not necessary for potential drug targets to be involved in the disease-causing process, or responsible for a disease, but they must be disease-modifying. Currently, various strategies are in practice for drug target identification, which is either based on experimental approaches or computational approaches.

Experimental approaches are mainly based on comparative genomics (expression profiling) and supplemented with the phenotype and genetic association analysis. Mostly, all experimental approaches provide reliable results, and theoretically, they should be the first choice methods for target identifications. Even though experimental approaches are more precise, they are suffering from some practical limitations, i.e., relatively high costs and intensive scientific labor required for experimental profiling of the full target space (>20,000 proteins, nucleic acid) of chemical compounds and they often end with few drug targets in hand. Due to all these limitations, mostly scientists and pharmaceutical companies utilize the computational methods for first-line research and then use the experimental approaches for further validation and other purposes.

8.4 Computational Approaches for Drug Target Identification

The development of bioinformatics has come up with various bioinformatics resources, including the database, algorithm, and software, which push the CADD in every aspect of the drug designing process (Table 8.2). One of the most important contributions is computational drug target identification, as discussed earlier that identification of the drug target is a very crucial and most decisive step of the drug designing process. In this regard, for the last one and half decades, various scientific studies carried out with the aim of drug target identification with the help of bioinformatics resources and proposed various approaches for drug target identifications. These approaches easily handle and deal with a huge amount of genomics, transcriptomics, and proteomics data, and also process it efficiently, and at the end provide potential drug targets in a short period at a low cost.

Currently, several computational approaches are available which utilized different molecular information, i.e., gene and genome sequence, molecular interaction information and protein 3D structure. Most of these approaches are interlinked. Still, based on their concept, they have broadly classified into two types: (1) homology-based approaches and (2) network-based approaches. The major features which are checked for drug target prediction are listed in Table 8.3 (Kim et al. 2017).

Table 8.2 Bioinformatics resources for drug target identification and CADD

S. No.	Database/ method	Description
1	DbMDR	It provides a collection of multidrug resistance genes and their orthologs, acting as potential drug targets (Gupta et al. 2011)
2	DEG	It contains all known essential genes from a different organism (Zhang et al. 2004)
3	DFVF	Collection of fungal virulence factors which collected>2000 pathogenic genes from a wide range of fungal sp. (Lu et al. 2012)
4	DrugBank	DrugBank is a richly annotated database, which provides detailed information about the drugs along with their target and drug action information (Wishart et al. 2008)
5	GEO	The database provides transcriptomics data (mainly array- and sequence-based) useful for functional genomics (Clough and Barrett 2016)
6	KEGG	KEGG offers information about the pathway, gene, and ligands in three different databases, i.e., Pathway, Gene, and Ligand (Kanehisa and Goto 2000)
7	MvirDB	Microbial protein toxins, virulence factors, and genes related to antibiotic resistance (Zhou et al. 2007)
8	PDTD	Database of potential proteins for in silico *drug* target identification (Gao et al. 2008)
9	TDR targets	Identification and prioritization of molecular targets for drug development (Magariños et al. 2012)
10	TTD	Publicly accessible cross-links database that provides inclusive information about known therapeutic targets with related information, i.e., pathway information and the corresponding drugs/ligands (Chen et al. 2002)
11	VFDB	Database contains virulence factors (VFs) of various medical significant bacterial pathogens (Chen et al. 2005)
12	Daspfind	Interactions between drugs and target proteins based on the similarities among them (Ba-Alawi et al. 2016)
13	iDTI-ESBoost	Evolutionary and structural feature-based model for identification of drug–target interactions (Rayhan et al. 2017)
14	NetCBP	Drug–target interaction prediction with the help of networks. It also predicts some new drugs without any known target interaction information (Chen and Zhang 2013)
15	SELF-BLM	It predicts drug–target interactions using a self-training support vector machine (SVM) based bipartite local model; SELF-BLM (Keum and Nam 2017)

8.5 Homology-Based Approaches

Homology-based approaches utilize sequence similarities among genes and proteins, further based on predicted homology, it takes the decision just like decision tree analysis. Mostly these methods consider the various level of homology test, which follows top-down direction. Each level of homology test scale down the data,

Table 8.3 Important features utilized in drug target identifications

S. No.	Features	Description
1	Essentiality of targets	To find out the indispensable nature of probable target for disease/pathogen
2	Gene ontology, biological process, involvement in pathways	To find out the biological process, pathways, and functional involvement of probable targets
3	Cellular localization	To find out the accessibility of probable target for a drug
4	Structural availability, druggability	To find out the binding pockets along with various physiochemical features involved in binding. It also helps to predict binding affinity and drug–target interaction mode
5	Gene expression patterns	Expression patterns play a significant role to check the availability of targets in given conditions. It also helps to predict the chance of adverse drug reaction, especially in the case of polypharmacological drugs

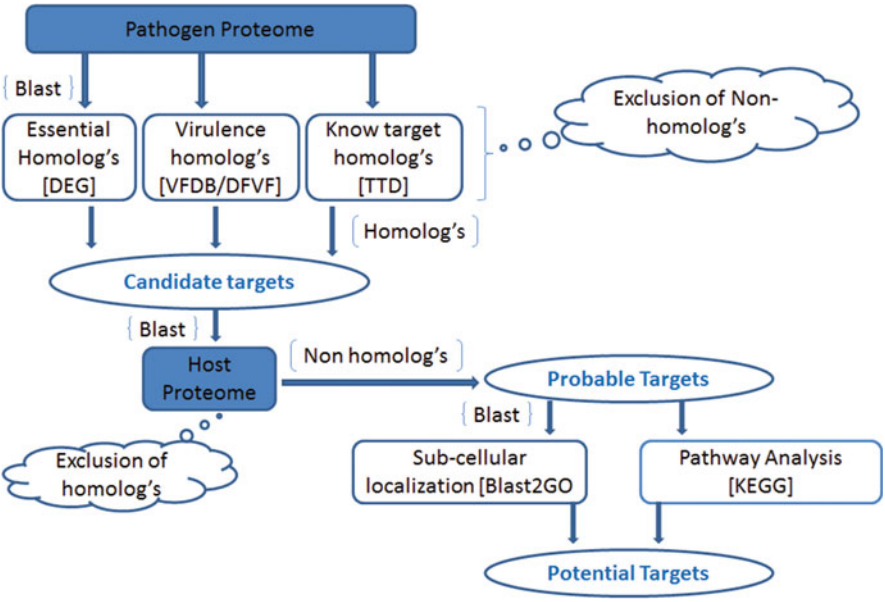

Fig. 8.2 Schematic diagram of the standard flowchart for drug target identification using homology-based approach

starting from complete genes or proteome, and step by step either eliminate those which fitted in "inappropriate" or select only those which fitted in "appropriate." Homology-based approaches always ended with countable potential drug targets (Fig. 8.2), and because of their scale down nature, these approaches are also known as subtractive (genomic or proteomic) approaches.

The term "inappropriate" and "appropriate" are conditional, and they are tested on various biological conditions that play a decisive role in target selection. The following are the major conditional tests that help to decide the further consideration of molecules for drug target identification.

8.5.1 Human Homologs

It is assumed that humans have various genes, and few of them are playing an indispensable biological role, considered as housekeeping genes. The use of human housekeeping genes or homologs of human housekeeping genes as a drug target can create lethal conditions and result in the death of human patients. To avoid such accidental use of the housekeeping gene as well as some important pathway-related gene as a drug target genes of the microbial pathogen are generally compared against the human, and those genes which show significant similarities with human housekeeping or crucial genes will be considered as "inappropriate" and mostly eliminate from rest of the process.

8.5.2 Human-Microbiome Homologs

The human body, especially, the gut has a lot of microbes that are already listed by the human microbiome project. Most of these microbes are involved in the biological process, which is beneficial for humans and thus considered beneficial microbes. Use of homologs from these beneficial microbes as a drug target can harm these bacteria, which can affect the related biological process in the human host, i.e., digestion, respiration process, etc., because of the above said reason, human-microbiome homologs are considered as "inappropriate" and eliminated from the further process.

8.5.3 Essentiality

Identification of drug targets against the microbial pathogen assumes that the essentiality of the target protein for pathogen-microbes is one of the advantageous and "appropriate" features. Without the function of essential proteins, microbial-pathogen will not able to survive. Various essential genes and proteins are identified by experimental approaches and enlisted in various databases. The database of essential genes (DEG) is one of the most active databases providing a collection of essential genes and protein sequences. Based on the above concept, those pathogenic genes/proteins which show homology with essential genes/proteins are considered as "appropriate" and include for the further process.

8.5.4 Virulence Factor Homologs

Those proteins whose role in virulence and pathogenicity is reported through the experiment are considered as virulence factors. Various such proteins are available, especially for microbes, and their molecular information is stored in various databases, i.e., virulence factor database (VFDB) and database of fungal virulence factors (DFVF). Genes/proteins of the pathogens that show homology with these virulence factors can be considered as "appropriate" and utilized as a potential drug target.

8.5.5 Drug Target Homologs

Information about known and explored drug/therapeutic targets is available, i.e., therapeutic target database (TTD). Homology mining with TTD is in practice, and those candidate molecules which show significant homology with these known targets are considered as "appropriate" and included for further exploration.

8.5.6 Cellular Location

The cellular location of the target protein is one of the very important features and plays a crucial role in target selection. In a homology-based approach, sequence-based gene ontology (GO) and annotation are in practice to look at the sub-cellular location along with the cellular component, biological process, and molecular function. Generally, those targets whose access is easy are preferable over others.

8.5.7 Role in the Biological Pathway

Biological pathways are responsible for the synthesis or metabolism of various bio-products. Few of these pathways are very important and unique, and they are solely responsible for their processes and products. The blockage of these pathways creates a scarcity of their products and finally reduces the chance of survival of the pathogen. Various pathway databases are available to conduct such checks. Current literature shows that the KEGG pathway is one of the richest and preferable pathway databases utilized for this purpose. Those pathways which are unique for pathogen are considered as appropriate pathways, and gene/proteins involved in them were considered for the further process. In contrarily those pathways which are also shared by human/host and their gene/proteins are "inappropriate" and excluded from further consideration.

It has been observed that homology-based approaches are very fast and almost cover the entire target space, and it only needs sequence information as input. Available reviews suggest that uses of homology-based approaches are very

common for microbial disease and generally restricted with them only. Their use for other types of infection or disease is not in common practice.

8.5.8 Case Study: Subtractive Approach for Drug Target Identification

The subtractive approach is one of the very famous approaches that have been utilized for target identification against various pathogens. In 2011 Katara et al. presented a subtractive approach exploiting the knowledge of global gene expression along with sequence comparisons to predict the potential drug targets in *Vibrio cholera*e, cholera causing bacterial pathogen, efficiently. Their analysis was based on the available knowledge of 155 experimentally proved virulence genes (seed information) (Fig. 8.3). For target identification, they utilized co-expression based gene mining and multilevel subtractive approach. At the end, they reported 36 gene products as a drug target, to check the reliability of the predicted targets they also performed gene ontology through Blast2GO. They observed these targets for their involvement in a crucial biological process and their cellular location. They found all these 36 gene products as reliable targets and conclude them as potential drug targets.

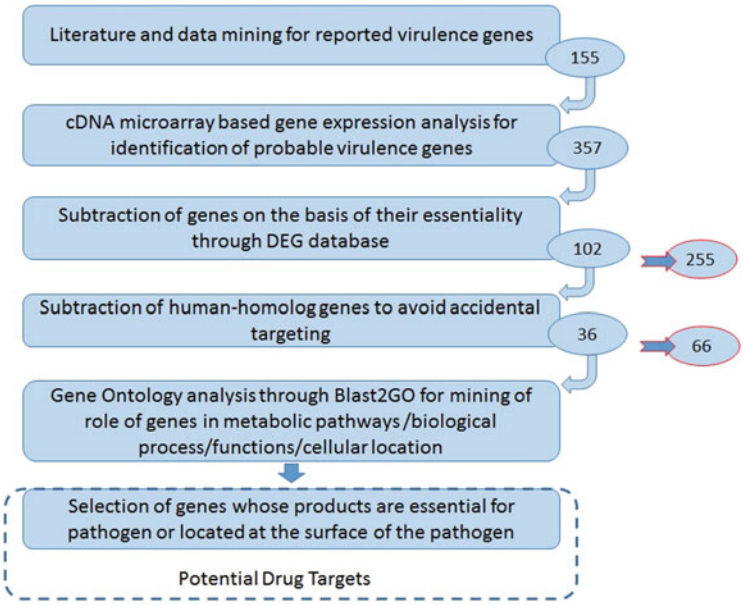

Fig. 8.3 Subtractive approach for drug target identification

8.6 Network-Based Approaches

It examines the effects of drugs in the context of molecular networks (i.e., protein–protein interactions, gene networks, transcriptional regulatory networks, metabolic networks, and biochemical reaction networks). In molecular network models, molecules refer as nodes, and each edge corresponds to an interaction between two molecules, based on the direction and importance of interaction between nodes, sometimes edges also mention the direction and weight (Fig. 8.4). Drug target identification through the network is based on the fact that networks have many important nodes that are vulnerable and can be targeted in many ways. Most of the time, these nodes are very crucial, and sometimes essential for the whole network structure, inhibition of such nodes can reduce their efficiency and damage of these nodes can shut down the complete network. Network inhibition process follows one of the following two models: (1) partial inhibitions: Partial knockout of the interactions of the target nodes, and (2) complete inhibition: all interactions around a given target node are eliminated.

In the drug designing process, these target nodes can be considered as potential drug targets. Various molecular networks (Table 8.4), including protein-interaction networks, regulatory, metabolic, and signaling networks individually or in integrated form can be subjected to a similar analysis (Imoto et al. 2007; Sridhar et al. 2008; Kotlyar et al. 2012; Shin et al. 2017).

8.6.1 Centrality Based Drug Target

Network centrality can be used as a potential tool for network-based target identification. Network centrality can prioritize proteins based on the network centrality measures (i.e., degree, closeness betweenness). It can be used to characterize the importance of proteins in the biological system.

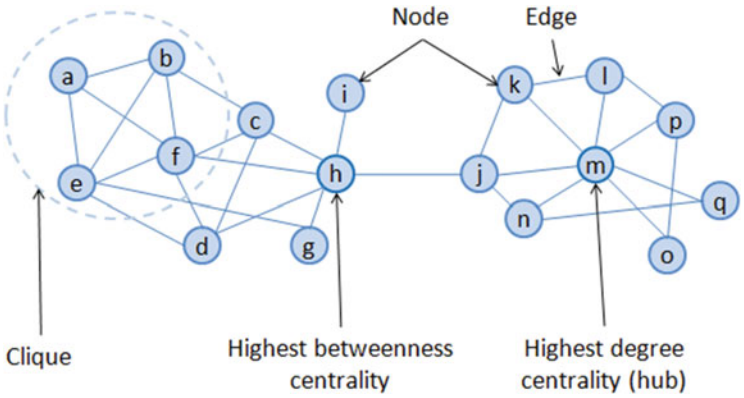

Fig. 8.4 Various components of a standard network

Table 8.4 Types of the biological network for drug target identification

S. No.	Network	Description
1	Protein–protein interactions (PPIs)	Here, proteins are nodes, and their interactions are edges. Proteins with high degrees of connectedness are likely to be more crucial than proteins with lesser degrees (Zheng et al. 2013; Shin et al. 2017; Verma et al. 2020)
2	Gene regulatory networks (GRN)	Transcription factors bind to multiple binding sites in a genome. As a result, all cells have complex networks between transcription factors (with respect to their target gene) that form a GRN (Imoto et al. 2007)
3	Gene co-expression network (GCN)	GCN is an undirected graph network that shows connectivity between co-expressed genes that supposed to be regulated by the same transcriptional regulatory system (Cheng et al. 2012; Yang et al. 2014)
4	Metabolic networks	The network of biochemical reactions is called metabolic network. Flux-balance analysis of these networks provides information about potential targets (Sridhar et al. 2008)
5	Cell signaling networks	Signaling networks represent connectivity between cellular signals typically, by combining PPIN, GRN, and metabolic networks (Behar et al. 2013)
6	Composite network	Composite cellular (transcriptional, signaling, PPI) networks identify the susceptible nodes which can act as a potential target (Pinto et al. 2014)

8.6.1.1 Hubs as Target

Real-world networks almost show a scale-free degree distribution, which means that in these networks, some nodes have a tremendous number of connections to other nodes (high degree), whereas most nodes have just a few. Here, nodes with a great number of connections than average called hubs. It assumes that the functionality of such scale-free networks heavily depends on these hubs, and if these hubs are selectively targeted, the information transfer through networks gets hindered and results in the collapse of the network (Pinto et al. 2014).

8.6.1.2 Betweenness Centrality Based Target

Hubs are the centers of local network topology, thus only provide the local picture of the network. Betweenness centrality is another approach that can be used to explain network centre, unlike, hub it provides central elements of the network in the global topology, thus, provide a global picture of network connections. Conceptually, betweenness is the number of times a node is in the shortest paths between two other nodes (Fig. 8.4), thus higher the betweenness means more importance of the node in quick network communication. Such higher betweenness centrality nodes can be utilized as a potential target against drugs (Melak and Gakkhar 2015).

8.6.1.3 Mesoscopic Centrality Based Target

Considering the advantage of both local and global centers of network topology for drug target identifications, the third class of centrality called mesoscopic centrality has also been reported. Mesoscopic centrality is neither fully based on local

information (such as hubs) nor global information (such as betweenness centrality) on network structure. It mainly considers long-range connections between high degree nodes, which make a profound effect on small-world networks.

8.6.1.4 Weight-Based Drug Target

Recently, the weighted-directed network is also reported for drug target identification studies (Wang et al. 2013). The weighted-directed network is closer to the real, cellular scenario, where PPIs are characterized by their affinity and dominance (link weight) as well as direction (e.g., in form of signaling), as mentioned in Fig. 8.5. It has been assumed that the deletion of the links with the highest weighted centralities is often more disturbing to network behavior than the removal of the most central links in the similar un-weighted network topology.

Utilization of the complex structural information of real-world networks to measure the centrality is not an easy task, and it requires more sophisticated methods to overcome these challenges. Bioinformatics provides various tools to support network construction, visualization, and network-based analysis, i.e., weight, centrality, interaction directions (Table 8.5).

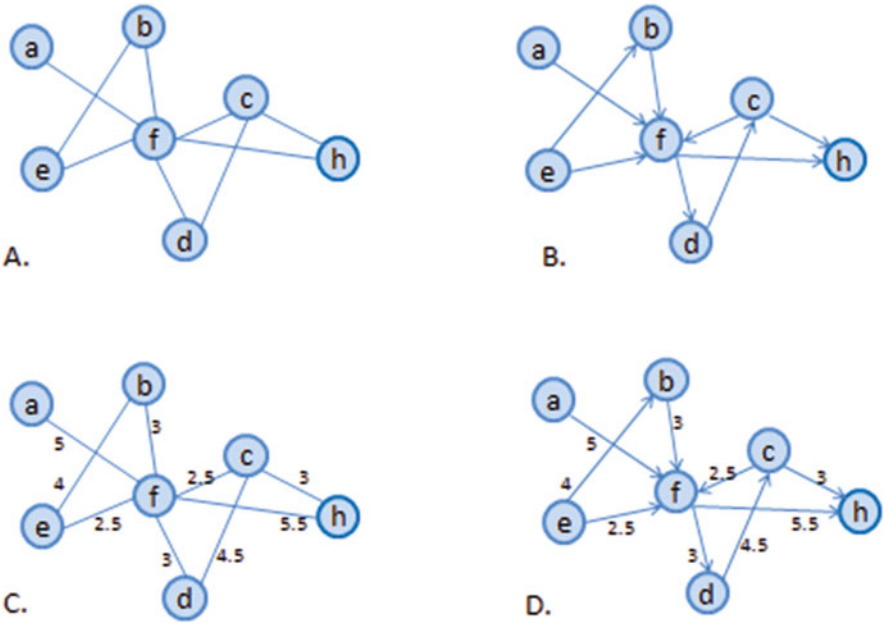

Fig. 8.5 Molecular network with a different type of connectivity between nodes (**a**) undirected (**b**) directed (**c**) weighted, and (**d**) weighted directed

Table 8.5 Tools supporting molecular network analysis for drug target identification

S. No.	Resource	Description
1	BioGRID	It is a repository of biological network information that can be visualized by Cytoscape (Oughtred et al. 2019)
2	BioMart	It contains data, software, and provides data services to facilitate scientific interactions and drug target discovery (Haider et al. 2009)
3	Connectivity map	It is a collection of genome-wide expression data from bioactive treated cultured human cells. It provides transcriptome based functional connections between drugs, genes, and diseases (Lamb et al. 2006)
4	MetaboAnalyst	It is an analysis tool for high-throughput metabolomics data, including data processing, biomarker discovery, and pathway analysis (Xia et al. 2015)
5	Netpredictor	Netpredictor is an R package that dealing with a unipartite or bipartite network. It can utilize to explore interactome and enrichment analysis for disease pathway and ontology (Seal and Wild 2018)

8.6.2 Limitations

Drug target identification through the biological network is an empirical approach, which relies on available information on molecular networks. However, numbers of molecular interaction databases are available, and most of them suffer from uncertainties, false-positive entries, and the average probability of particular interaction along with nomenclature as well as interpretation problems. However, to overcome these issues, recently, PPI databases are linked with protein structure data, which provides more reliable and validated interactions. At the same time, scientists also propose some alternative, i.e., use of the curated database and low-resolution network to surmount the above-mentioned problems (De-Alarcón et al. 2002).

8.7 Properties of an Ideal Drug Target

Identification of potential drug targets is not the last step. Nowadays, through various computational approaches, a huge number of probable targets are reported against different diseases and are available in databases and literature (Katara et al. 2011). It is not a good idea to recommend them directly for testing, its recommendation that first, we check them for an ideal property (Table 8.6), and then for druggability. Only those targets which fulfill most of them are considered as an ideal drug target and recommended them for further validation and testing (Gashaw et al. 2011).

Table 8.6 Important properties to assess the ideal drug targets

S. No.	Property	Detail
1	Disease-modifying	Target should be disease-modifying with proven function in disease pathophysiology
2	Disease specific modulation	Modulation of the target must be explicit to the targeted disease, should not affect standard physiology in normal or other disease conditions
3	Druggability assessment	Target druggability should be observable
4	Assay ability	Target should have favorable assay ability, specifically through high-throughput screening
5	Tissue-specific expression	Target expression should be tissue-specific, it should not affect unrelated tissue or organs

8.8 Druggability of Drug Target

In drug designing process, the potential of any target is defined by its druggability (affinity of the target to bind with drug-like molecules), thus the target must be druggable (Fauman et al. 2011). Biomolecules (i.e., protein, nucleic acid) with an activity that can be modulated by a drug are considered as a druggable target. These targets must have binding sites with typical structural and physicochemical properties that favor binding interaction with high affinity and specificity.

8.8.1 Importance of Druggability

Despite technological advancement in the drug designing process, most drug discovery projects fail because of the druggability problem. To avoid the failure of a drug discovery project, which is mostly very expensive, it is very important to understand the difficulties associated with a potential target. Druggability has become part of the target identification and validation process, more significantly in the case where targets do not belong to traditional classes (Finan et al. 2017).

8.9 Computational Methods for Druggability Assessment

To date, various targets are reported and documented through various methods, and few of them are already in practice (drugs are available against them), such targets are druggable. If no drug available for a target, then predict druggability is required. Various computational methods are available to evaluate the druggability of target protein, mainly rely on either sequence-based or 3D-structure based properties of proteins (Fauman et al. 2011).

8.9.1 Sequence-Based Methods

A protein is druggable if its other family members are known to be targeted by drugs. For such analysis, sequence alignment can be used to predict sequence similarity (homology) between probable target (query) proteins and database of known druggable targets (Finan et al. 2017). The sequence-based concept provides a significant approximation of druggability, but it suffers from the following limitations: (1) its predictions are limited to known drug target families, it does not attempt for those potential targets, which belong to the novel "un-drugged" protein family; and (2). It assumes that all members of the protein family are equally druggable, which is not true.

8.9.2 Structure-Based Methods

Structure-based methods rely on the availability of 3D structure information, thus only can apply to those proteins whose structures are available. Along with experimentally determined 3D structures, it also considers high-quality structure models through homology modeling. Several structure-based methods are available for the assessment of target druggability, irrespective of their different algorithms; all of them consist of the following three common components.

8.9.2.1 Identifying Cavities and Binding Pockets

Many computational methods and tools have been developed for binding pocket identification, which scans 3D surface and interior of the target protein for potential cavities (possess suitable properties for binding a ligand) that can act as binding pockets. These tools mainly tend to look for cavities with suitable size, shape, and composition to accommodate drug-like molecules.

Working of binding pockets detection methods depends on either energy-based or geometry-based detection algorithms (Nisius et al. 2012; Zheng et al. 2013). Energy-based detection predicts pockets by computing the interaction energy between atoms of protein and a probe molecule (Ghersi and Sanchez 2011). Geometry-based detection predicts the solvent accessible area that is embedded in the protein surface. Comparative studies suggest that both types of detection algorithms have good performance and advantages (Schmidtke et al. 2010). It has been observed that geometry-based detections are more suitable for large-scale pocket detection. Their inherent advantages, i.e., high speed and robustness against structural variations or missing atoms and residues in the input structures, provide the edge over an energy-based detection algorithm (Schmidtke et al. 2010). With the increasing availability of binding cavity information, recently, one new class of methods called information-based detection methods are developed. These methods utilize available cavity information from its neighbor and similar proteins whose binding cavities are known.

8.9.2.2 Druggability of Binding Pocket

This second step aims to calculate the physicochemical and geometric properties of the pocket to check whether these properties are complementary with the properties of drug-like molecules. Lipinski's rule of five (RO5) connects the physicochemical properties of a drug with its pharmacokinetic properties (Lipinski 2000). It is a well-known fact that the physicochemical properties of the druggable pocket should be the mirror image of the physicochemical properties of the drug-like molecule itself. This analogy gave the concept of a druggable pocket. Therefore, the complementary properties of the pockets reflect the Lipinski's rule of five of "drug-likeness" (H-bond donors >5, H-bond acceptors $= 10$, molecular weight > 500, and the Log P (CLog P) is >5).

The major features which define and affect the druggability of pockets are pocket descriptors. Characteristic features of a binding site play a very crucial role in druggability calculation, and the selection of those descriptors, which are crucial for binding drug-like molecules, needs to be described as accurate as possible. Observations suggest that none of the individual pocket descriptors is sufficient for druggability explanation, and a group of descriptors is required to describe and calculate pocket druggability. Both physiochemical and geometrical features play a crucial role as descriptors. Physiochemical descriptors and frequently used physiochemical pocket descriptors include size, shape, electrostatics, hydrogen bonding, hydrophobicity, polarity, amino acid composition, rigidity, and secondary structure (Halgren 2009; Krasowski et al. 2011). Geometrical descriptors: Along with physicochemical properties, geometrical properties, i.e., the shape and size of the binding pocket, play a crucial role in suitable interactions with a small molecule (Zheng et al. 2013). The following are the major geometrical features involved in pocket druggability measurement.

Position of the Atoms

It has been observed that the position of the atoms in pockets affects the contribution of an atom in interaction. Atoms located at the contact surface considerably give a major contribution in contact energy (hydrophobic interaction) than those who lie outside of the surface, i.e., within the bulk of the protein cavity.

Cavity Size

Large spherical cavities are more exposed to the solvent, thus not suitable for binding, especially with small drug molecules. Narrow (micro) cavity pockets are less exposed to the solvent and offer more van der Waals contact, thus they are more druggable. These micro-cavities are also defined as hot spots, which are characteristic of highly druggable targets.

8.9.2.3 Target Specificity Assessment

Drug target must be specific; structure similarity of drug target molecules with other unwanted molecules will create problems in the drug development process. Structural similarity of the binding sites could make the design of selective inhibitors difficult. During target selection, it is very important to assess the structural

landscape of the primary binding sites of the target to confirm the druggability. Sequence and structural alignment based computational methods are available to perform specificity assessment.

Sequence Alignment Based Assessment

It is based on the sequence information of binding sites of the target protein. It assumes that when the degree of conversation between the two sequences is sufficiently high, then identical amino acids in the sequence will likely correspond to identical binding site structure.

Structure Alignment Based Assessment

These methods are based on either structural superposition or pharmacophore features. Structural superposition generally utilizes a 3D grid force field around the binding sites, which can be calculated using various types of energy terms, i.e., electrostatic, hydrophobic, and hydrogen bonding. In the grid approach, the field potentials can be calculated for each suspicious protein and are used for comparing their binding sites. The structural similarity between a pair of proteins can be studied by correlation functions of the various molecular interaction fields (MIFs) of the two grids or by utilizing the Fourier transformation of correlation functions or related approaches. Another approach consists of identifying pharmacophore features that generally summarized with the help of surface chemical features (SCF), including hydrophobic centers, H-bond donors and acceptors, positive and negative charges, and aromatic centers, etc. This SCF based on the consideration can be determined on the whole protein surface or a chosen cavity. Binding sites with the highest SCF matches show the highest similarity with the query binding site. Various computational tools are already available, which provide the facilities to evaluate binding site similarities and assess the specificity (Table 8.7). Almost all tools rely on the available entries at the protein structural database.

8.9.3 Quantification of Druggability

Quantification of druggability could provide the best criteria for target selection, but till now, none of the standard explanation is available for this purpose. Each method has its measures for druggability, thus a druggability score of a specific target might vary. However, irrespective of an individual's weaknesses and strengths, all major druggability measures can classify targets into druggable, non-druggable, medium druggable, and difficult-druggable.

8.9.4 Major Concern

8.9.4.1 Size of Training Sets

Most of the druggability assessment methods are based on the machine learning algorithm, thus highly dependent on available training sets (ChEMBL, BindingDB,

Table 8.7 Bioinformatics resources for druggability detection and evaluation

S. No.	Tool/ algorithm	Description
1	CavityPlus	Protein cavity detection and functional analyses (Xu et al. 2018)
2	Dr. PIAS	A druggability assessment system. Along with druggability, it also provides functional annotation of interacting proteins (Sugaya and Furuya 2011)
3	DrugEBIlity	It evaluates the druggability of targets. The server can search with a sequence, PDB id, or uploaded structure (https://www.ebi.ac.uk/chembl/drugebility)
4	DrugPred	Structure-based druggability predictor that relies on the affinity between known drugs and their target proteins (Krasowski et al. 2011)
5	IsoCleft	Detection of local geometric and chemical similarities between potential binding cavities for small molecules (Kurbatova et al. 2013)
6	IsoMIF finder	Detection and comparison of binding site molecular interaction field (MIF) (Chartier et al. 2016)
7	MultiBind	Recognize the common spatial chemical binding patterns along with shared physicochemical binding site properties (Shulman-Peleg et al. 2008)
8	PockDrug-server	Pocket druggability with and without ligand proximity information. In both cases, it provides consistent druggability results using different pocket estimation methods (Hussein et al. 2015)
9	SiteAlign	Align, compare druggable ligand-binding sites, and to measure distances between druggable protein cavities (Schalon et al. 2008)
10	SiteMap's	Provide prediction of the target's binding sites with druggability. It also provides quantitative and graphical information about the target (Halgren 2009)

PubChem, etc.) used to train them. The size and quality of the available datasets in databases directly affect the reliability and scope of the assessment methods.

8.9.4.2 Binding Site Flexibility

The identification of the binding cavity in a rigid target is based on the assumption that the cavity already exists. There are some proteins whose binding pockets do not exist in their native structure, and their active pockets behave like inducible allosteric sites, which only revealed after protein conformational changes. In such a case, it is very difficult to assess the binding pockets, and this situation is considered as a binding site "flexibility problem." The presence of multiple X-ray conformers for a specific target can help us to handle binding site flexibility. Multiple conformers allow us to assess the relative variability of certain residues within the binding site pockets. Based on such relative variability information, it is possible to assess the plasticity of the binding site.

8.10 Target-Based Drug Discovery

As discussed, drug targets are the most crucial element of the drug designing process, and selection of the targets decides the fate of the drug designing process that it will succeed or get fail at a later stage. For several decades, pharmaceutical companies are successfully using well established one drug-one target approach for drug designing purposes. By realizing the scenario, the central dogma of the drug designing process has now shifted from one drug-one target to one drug-multi-target concept and considers multiple targets for a single drug.

8.10.1 Multi-Target Drug Designing

Computational approaches specifically those, which are based on system biology concepts are very crucial in the identification of multi-targets, thus play a major role in the success of the multi-target-based drug designing (Vasaikar et al. 2016). Multi-target-based drug designing approach is, to some extent, similar to single target-based drug designing, but it initiated with the set of targets multi-targets (Fig. 8.6). The following are the main steps of multi-target drug designing.

8.10.1.1 Identification of a Set of Targets "Multi-Targets"

This is the most crucial step which decides the fate of the whole following process. System biology-based molecular networks are in practice to identify multi-targets.

8.10.1.2 Generation of Multi-Target Pharmacophore

Computational methods are available to design multi-target (structure) based pharmacophore, which utilizes combinatorial algorithms (Kumar et al. 2018; Ramsay et al. 2018). The most common steps in multi-target pharmacophore generation include (1) interaction profiling (MIFs) of all targets, (2) identification of common MIFs/features, and (3) multi-target specific and selective ensembles development.

8.10.1.3 Virtual Screening

Pharmacophore generation is followed by virtual screening of chemical libraries to find suitable compounds against multi-target pharmacophore.

8.10.1.4 Generation or Selection of Multi-Target Compound

Multi-target compounds are generated through the integration of pharmacophore of above-selected molecules (already known drugs or drug candidates).

8.10.1.5 Evaluation and Optimization of Multi-Target Specific Compound

Evaluation and optimization process mainly includes multi-target specific interaction assay (to avoid off-targeting), QSAR, and degree of modulation. Though multi-target drugs seem promising and designing of these compounds is not a

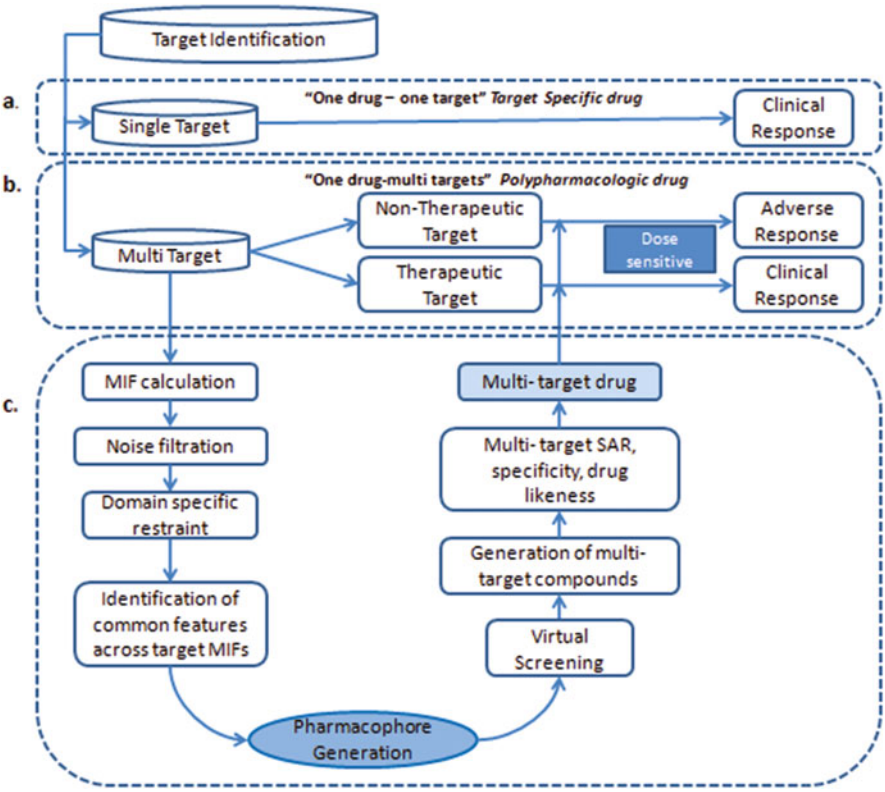

Fig. 8.6 Target-based drug designing (**a**) single target-based drug designing, (**b**) multi-target-based drug designing and (**c**) major steps involve in multi-target-based drug designing

straightforward task. It needs to deal with various crucial issues, i.e., right target-sets selection, balanced activity towards them, and excluding activity at off-target(s), while at the same time retaining drug-like properties (Hopkins 2008; Bottegoni et al. 2012). Available experimental methods are not enough to handle these issues, thus the feasibility of multi-target drugs profoundly depends on computational approaches and resources. Various databases are also there, i.e., DrugBank, STITC H, BindingDB ZINC, PubChem, KEGG DRUG, which provide required information about molecular pathways, 3D structure, chemical reactions, side effects, and known drug targets, thus help in the success of poly-pharmacologic drugs.

8.11 Summary

Now day's computational biology becomes an indispensable tool for almost every aspect of biology and related fields, and drug designing is not an exception. CADD is now a mature field, and its success influenced by its first and pivotal step that is the

identification of drug targets. This chapter provides an overview of various computational approaches available for drug target identification. It also discusses various bioinformatics resources, i.e., database, methods, and software, which can be handy for drug target identification purposes.

Competing Interest The author declares that there are no competing interests.

References

Ba-Alawi W, Soufan O, Essack M, Kalnis P, Bajic VB (2016) DASPfind: new efficient method to predict drug-target interactions. J Cheminf 8:15

Behar M, Barken D, Werner SL, Hoffmann A (2013) The dynamics of signaling as a pharmacological target. Cell 155(2):448–461

Bottegoni G, Favia AD, Recanatini M, Cavalli A (2012) The role of fragment-based and computational methods in polypharmacology. Drug Discov Today 17(1–2):23–34

Chartier M, Adriansen E, Najmanovich R (2016) IsoMIF Finder: online detection of binding site molecular interaction field similarities. Bioinformatics 32(4):621–623

Chen H, Zhang Z (2013) A semi-supervised method for drug-target interaction prediction with consistency in networks. PLoS One 8(5):e62975

Chen X, Ji ZL, Chen YZ (2002) TTD: therapeutic target database. Nucleic Acids Res 30 (1):412–415

Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q (2005) VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:D325–D328

Cheng F, Liu C, Jiang J et al (2012) Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol 8(5):e1002503

Clough E, Barrett T (2016) The gene expression omnibus database. Methods Mol Biol 1418:93–110

Cohen P (2002) Protein kinases—the major drug targets of the twenty-first century? Nat Rev Drug Discov 1(4):309–315

De-Alarcón PA, Pascual-Montano A, Gupta A, Carazo JM (2002) Modeling shape and topology of low-resolution density maps of biological macromolecules. Biophys J 83(2):619–632

Dersch P, Khan MA, Mühlen S, Görke B (2017) Roles of regulatory RNAs for antibiotic resistance in bacteria and their potential value as novel drug targets. Front Microbiol 8:803

Docherty AJ, Crabbe T, O'Connell JP, Groom CR (2003) Proteases as drug targets. Biochem Soc Symp 70:147–161

Fauman EB, Rai BK, Huang ES (2011) Structure-based druggability assessment--identifying suitable targets for small molecule therapeutics. Curr Opin Chem Biol 15(4):463–468

Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, Galver L, Kelley R, Karlsson A, Santos R, Overington JP, Hingorani AD, Casas JP (2017) The druggable genome and support for target identification and validation in drug development. Sci Transl Med 9(383):eaag1166

Gao Z, Li H, Zhang H, Liu X, Kang L, Luo X, Zhu W, Chen K, Wang X, Jiang H (2008) PDTD: a web-accessible protein database for drug target identification. BMC Bioinf 9:104

Gashaw I, Ellinghaus P, Sommer A, Asadullah K (2011) What makes a good drug target? Drug Discov Today 16(23–24):1037–1043

Ghersi D, Sanchez R (2011) Beyond structural genomics: computational approaches for the identification of ligand binding sites in protein structures. J Struct Funct Genom 12(2):109–117

Gupta S, Mishra M, Sen N, Parihar R, Dwivedi GR, Khan F, Sharma A (2011) DbMDR: a relational database for multidrug resistance genes as potential drug targets. Chem Biol Drug Des 78 (4):734–738

Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A (2009) BioMart Central Portal—unified access to biological data. Nucleic Acids Res 37:W23–W27

Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model 49(2):377–389
Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4 (11):682–690
Hughes JP, Rees S, Kalindjian SB, Philpott KL (2011) Principles of early drug discovery. Br J Pharmacol 162(6):1239–1249
Hussein HA, Borrel A, Geneix C, Petitjean M, Regad L, Camproux AC (2015) PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Res 3(W1):W436–W442
Imoto S, Tamada Y, Savoie CJ, Miyano S (2007) Analysis of gene networks for drug target discovery and validation. Methods Mol Biol 360:33–56
Kaczorowski GJ, McManus OB, Priest BT, Garcia ML (2008) Ion channels as drug targets: the next GPCRs. J Gen Physiol 131(5):399–405
Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30
Katara P (2013) Role of bioinformatics and pharmacogenomics in drug discovery and development process. Netw Model Anal Health Inform Bioinf 2(4):225–230
Katara P (2017) Stem cell: a key to solving the drug screening enigma. In: Verma V, Singh MP, Kumar M (eds) Stem cells from culture dish to clinic. Nova Science, New York, pp 257–268
Katara P, Grover A, Kuntal H, Sharma V (2011) In silico prediction of drug targets in Vibrio cholerae. Protoplasma 248(4):799–804
Katsila T, Spyroulias GA, Patrinos GP, Matsoukas MT (2016) Computational approaches in target identification and drug discovery. Comput Struct Biotechnol J 14:177–184
Keum J, Nam H (2017) SELF-BLM: prediction of drug-target interactions via self-training SVM. PLoS One 12(2):e0171839
Kim B, Jo J, Han J, Park C, Lee H (2017) In silico re-identification of properties of drug target proteins. BMC Bioinf 18(Suppl 7):248
Klaeger S, Heinzlmeir S, Wilhelm M et al (2017) The target landscape of clinical kinase drugs. Science 358(6367):eaan4368
Kotlyar M, Fortney K, Jurisica I (2012) Network-based characterization of drug-regulated genes, drug targets, and toxicity. Methods 57(4):499–507
Krasowski A, Muthas D, Sarkar A, Schmitt S, Brenk R (2011) DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set. J Chem Inf Model 51(11):2829–2842
Kumar P, Kaalia R, Srinivasan A, Ghosh I (2018) Multiple target-based pharmacophore design from active site structures. SAR QSAR Environ Res 29(1):1–19
Kurbatova N, Chartier M, Zylber MI, Najmanovich R (2013) IsoCleft Finder—a web-based tool for the detection and analysis of protein binding-site geometric and chemical similarities. F1000Res 2:117
Lamb J, Crawford ED, Peck D et al (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795):1929–1935
Lipinski CA (2000) Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods 44:235–249
Lu T, Yao B, Zhang C (2012) DFVF: database of fungal virulence factors. Database 2012:bas032
Magariños MP, Carmona SJ, Crowther GJ et al (2012) TDR targets: a chemogenomics resource for neglected diseases. Nucleic Acids Res 40:D1118–D1127
McInnes C (2007) Virtual screening strategies in drug discovery. Curr Opin Chem Biol 11 (5):494–502
Melak T, Gakkhar S (2015) Comparative genome and network centrality analysis to identify drug targets of Mycobacterium tuberculosis H37Rv. Biomed Res Int 2015:1. https://doi.org/10.1155/2015/212061
Nisius B, Sha F, Gohlke H (2012) Structure-based computational analysis of protein binding sites for function and druggability prediction. J Biotechnol 159(3):123–134

Oughtred R, Stark C, Breitkreutz BJ et al (2019) The BioGRID interaction database: 2019 update. Nucleic Acids Res 47(D1):D529–D541

Pinto JP, Machado RS, Xavier JM, Futschik ME (2014) Targeting molecular networks for drug research. Front Genet 5:160

Ramsay RR, Popovic-Nikolic MR, Nikolic K, Uliassi E, Bolognesi ML (2018) A perspective on multi-target drug discovery and design for complex diseases. Clin Transl Med 7(1):3

Rayhan F, Ahmed S, Shatabda S et al (2017) iDTI-ESBoost: identification of drug target interaction using evolutionary and structural features with boosting. Sci Rep 7(1):17731

Schalon C, Surgand JS, Kellenberger E, Rognan D (2008) A simple and fuzzy method to align and compare druggable ligand-binding sites. Proteins 71(4):1755–1778

Schmidtke P, Le Guilloux V, Maupetit J, Tufféry P (2010) Fpocket: online tools for protein ensemble pocket detection and tracking. Nucleic Acids Res 38:W582–W589

Seal A, Wild DJ (2018) Netpredictor: R and shiny package to perform drug-target network analysis and prediction of missing links. BMC Bioinf 19(1):265

Shekhar C (2008) In silico pharmacology: computer-aided methods could transform drug development. Chem Biol 15(5):413–414

Shin WH, Christoffer CW, Kihara D (2017) In silico structure-based approaches to discover protein-protein interaction-targeting drugs. Methods 131:22–32

Shulman-Peleg A, Shatsky M, Nussinov R, Wolfson HJ (2008) MultiBind and MAPPIS: webservers for multiple alignment of protein 3D-binding sites and their interactions. Nucleic Acids Res 36:W260–W264

Sridhar P, Song B, Kahveci T, Ranka S (2008) Mining metabolic networks for optimal drug targets. Pac Symp Biocomput 13:291–302

Sriram K, Insel PA (2018) G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs? Mol Pharmacol 93(4):251–258

Sugaya N, Furuya T (2011) Dr. PIAS: an integrative system for assessing the druggability of protein-protein interactions. BMC Bioinf 12:50

Vasaikar S, Bhatia P, Bhatia PG, Chu Yaiw K (2016) Complementary approaches to existing target based drug discovery for identifying novel drug targets. Biomedicines 4(4):E27

Verma Y, Yadav A, Katara P (2020) Mining of cancer core-genes and their protein interactome using expression profiling based PPI network approach. Gene Rep 18:100583

Wang W, Yang S, Li J (2013) Drug target predictions based on heterogeneous graph inference. Biocomputing 2013:53–64

Wishart DS, Knox C, Guo AC et al (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36:D901–D906

Xia J, Sinelnikov IV, Han B, Wishart DS (2015) MetaboAnalyst 3.0—making metabolomics more meaningful. Nucleic Acids Res 43(W1):W251–W257

Xu Y, Wang S, Hu Q et al (2018) CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res 46(W1):W374–W379

Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H (2014) Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat Commun 5:3231

Zauhar RJ, Moyna G, Tian L, Li Z, Welsh WJ (2003) Shape signatures: a new approach to computer-aided ligand- and receptor-based drug design. J Med Chem 46(26):5674–5690

Zhang R, Ou HY, Zhang CT (2004) DEG: a database of essential genes. Nucleic Acids Res 32: D271–D272

Zheng X, Gan L, Wang E, Wang J (2013) Pocket-based drug design: exploring pocket space. AAPS J 15(1):228–241

Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T (2007) MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res 35:D391–D394