

# 36 Web Resources For Target Hunting In Drug Discovery

Nov. 22, 2017 by Alfred Ajami , Andrii Buvailo

Choosing the right biological target or a combination of targets is a fundamental task for any successful drug discovery project. All the subsequent efforts -- be it a small molecule hit identification, lead optimization, pharmacokinetic studies, or a clinical trial -- will just be as effective, at the end of the day, as was the initial decision to choose one target or another.

No wonder, why target identification and validation play central roles in the overall drug discovery process. The number of promising drug targets increases every year thanks to new scientific advances. Understanding the definition and range of drug targets is facilitated by web resources covering molecular mechanisms, modes of action and signaling network interactions, especially when the evidence results from the linkage of target with disease and specific experimental factor ontologies, as illustrated in the accompanying diagram on the "ETDO" architecture.

Below is a curated list of 36 online open access resources, with their accompanying explanatory publications (mostly free), all crafted around ETDO variations. They should prove useful for both educational and research purposes:

## 1. BindingDB

A searchable database of experimentally measured binding affinities, focusing chiefly on the interactions of proteins considered to be drug-targets with small, drug-like molecules. BindingDB contains 1,419,347 binding data, for 7,000 protein targets and 635,301 small molecules.

## 2. BioGRID

The Biological General Repository for Interaction Datasets is an open access database on protein, genetic and chemical interactions for humans and all major model organism species and humans. To facilitate network-based approaches for drug discovery, BioGRID now incorporates 27,501 chemical-protein interactions for human drug targets.

## 3. canSAR

A cancer research and drug discovery knowledgebase that contains chemical and pharmacological data for over one million, bioactive, small molecule drugs and compounds corresponding to ~8 million pharmacological bioactivities as well as over 10 million calculated chemical properties intended to enable target selection and validation in drug discovery.

#### 4. CARLSBAD

A confederated database of more than 1 million bioactivity values for over 400,000 compounds associated with 3,500 protein targets. A one-click user query can determine potential leads for a target, associated off-targets, and druggable targets in associated disease pathways.

#### 5. ChEMBL

A large-scale database of bioactive drug-like small molecules, it contains 2-D structures, calculated properties (e.g. logP, Molecular Weight, Lipinski Parameters, etc.) and abstracted bioactivities (e.g. binding constants, pharmacology and ADMET information). Target data can be searched via keyword, protein sequence search (BLAST), or by navigating the target classification hierarchy. Research reference.

#### 6. CSNAP

Chemical Similarity Network Analysis Pull-down is a computational approach for compound target identification. Query and reference compounds are populated on the network connectivity map and a graph-based neighbor counting method is applied to rank the consensus targets. The CSNAP approach facilitates high-throughput target discovery and off-target prediction.

#### 7. DGIdb 3.0

The Drug-Gene interaction database consolidates, organizes and presents drug-gene relationships and gene druggability information from papers, databases and web resources. It encapsulates multiple gene categories (e.g. kinases, G-protein coupled receptors) that are expected to be good drug targets.

#### 8. DINIES

Drug-target Interaction Network Inference Engine built on Supervised analysis enables prediction of potential interactions between drug molecules and target proteins, based on drug data and omics-scale protein data.

## 9. DisGeNET

A platform integrating information on human disease-associated genes and variants. It can be used for investigation of the molecular underpinnings of specific human diseases, analysis of the properties of disease genes, generation of hypothesis on drug therapeutic action and drug adverse effects, validation of computationally predicted disease genes, and retrieval of druggable targets and biological pathways for a disease of interest.

## 10. DrugBank 3.0

A combination of detailed drug data with comprehensive drug target information, including searchable details on the nomenclature, ontology, chemistry, structure, function, action, pharmacology, pharmacokinetics, metabolism and pharmaceutical properties. Content covers 10,510 drug entries, 871 approved biotech (protein/peptide) drugs, 105 nutraceuticals and over 5,028 experimental drugs.

## 11. DrugCentral

Online drug compendium integrating structure, bioactivity, regulatory, pharmacologic actions and indications for active pharmaceutical ingredients. At the molecular level, DrugCentral bridges drug-target interactions with pharmacological action and indications to provide mechanistic understanding and relate protein targets to human disease and symptoms.

## 12. DTO

Drug Target Ontology offers searchable database access to classify and integrate drug discovery data based on formalized and standardized classifications and annotations of druggable protein targets. DTO integrates phylogenecity, function, target development level, disease association, tissue expression, chemical ligand and substrate characteristics, and target-family specific characteristics.

## 13. ECOdrug

A research platform tool connecting drugs and conservation of their targets across species. It harmonizes ortholog predictions from multiple sources via a simple user interface underpinning critical applications for a wide range of studies in pharmacology, ecotoxicology and comparative evolutionary biology.

## 14. HitPick

A web server that facilitates the analysis of chemical screenings by identifying hits and predicting their molecular targets, with a focus on analysis and interpretation of chemical phenotypic screens. The target prediction functionality can also be used in a stand-alone fashion.

## 15. iDrug-Target

A package of web-services for predicting drug-target interactions between drug compounds and target proteins in cellular networking via a benchmark dataset optimization approach. It contains four predictors: iDrug-GPCR, iDrug-ChI, iDrug-Ezy, and iDrug-NR, specialized for GPCRs (G protein-coupled receptors), ion channels, enzymes, and NR (nuclear receptors), respectively.

## 16. idTarget

A web server that can predict possible binding targets of a small chemical molecule via a divide-and-conquer docking approach. It is intended also to reproduce known off-targets of drugs or drug-like compounds. Unlike previous approaches that screen against a specific class of targets or a limited number of targets, idTarget addresses nearly all protein structures deposited in the Protein Data Bank (PDB).

## 17. IUPHAR/BPS

An interactive guide to pharmacology aggregating drug target information from Ensembl, UniProt, PubChem, ChEMBL and DrugBank, as well as curated citations from PubMed. Coverage includes the key properties and selective ligands and tool compounds available for each target family. Queries can then be expanded into the pharmacological, physiological, structural, genetic and pathophysiological properties of each target.

## 18. MANTRA

Mode of Action by NeTwoRk Analysis is a computational tool for the evaluation of the target mode of action (MoA) of novel drug structures and the identification of known and approved candidates for “drug repositioning”. Structural queries are automatically integrated into a network of compounds where the topology reveals similarities and differences in MoA compared to reference compounds in the known pharmacopoeia.

## 19. Open Targets

A platform for therapeutic target identification and validation, providing either a target-centric workflow to identify diseases that may be associated with a specific target, or a disease-centric workflow to identify targets that may be associated with a specific disease. Coverage includes genetic associations, somatic mutations, known drugs, gene expression, affected pathways, literature mining and animal models.

## 20. PASS

Prediction of Activity Spectra for Substances is an online tool for evaluating the general biological potential of an organic drug-like molecule. PASS provides simultaneous predictions of many types of biological activity, including molecular and cellular targets, with readouts modeled after the Anatomical Therapeutic Chemical classification system (ATC). In this context, see also SuperPred.

## 21. PharmMapper

A web server for potential drug target identification with a target database of over 53,000 receptor-based pharmacophore models created with ~23,000 protein poses predicted to qualify as druggable binding sites. The expanded target data repository in Version 7 now covers 450 indications and 4,800 molecular functions.

## 22. Pharos

A multimodal web interface and target search portal at the front end of the underlying databases in the Illuminating the Druggable Genome initiative. It also integrates with the functionality of the DrugCentral and DTO resources (also cited in this compendium). Pharos now serves as an entry point to targets in the entire human proteome. Its design is intended to shed light on the dark corners thereby expanding what is considered druggable.

## 23. PhID

A network pharmacology portal of targets, diseases, genes, side-effects, pathways, and drugs designed to provide visualization of complex relationships and specifically to predict drug-target interactions. The underlying databases cover approximately 330,000 chemical structures, including the known pharmacopoeia; 24,000 targets; 8,500 diseases; 43,000 genes; 4,500 side effects; and 867 pathways.

## 24. PPB

The Polypharmacology Browser is a multi-fingerprint target prediction tool using ChEMBL bioactivity data. It allows users to find out if a newly identified bioactive molecule, or any compound, is closely related to

molecules with documented bioactivity and therefore likely to interact with the corresponding biological target.

## 25. PROMISCUOUS

A database and visualization tool for network-based drug-repositioning designed to deliver complex relations among drugs, their respective targets and side-effects of the drugs. The web portal has been designed with a novel interface that offers a "natural" way of exploring the network: database entities (drugs, targets and side effects) are represented as nodes in a network with edges, which represent the relations between them. Coverage of the 25,000 drug plus drug-like molecules collection includes annotations on drug-protein and protein-protein relationships.

## 26. SEA

The Similarity Ensemble Approach categorizes proteins by comparing the set-wise chemical similarity among their ligands. It can be used to rapidly search large compound databases and to build cross-target similarity maps. A collection of ~65,000 ligands annotated for drug targets are incorporated into the query algorithms. These relate targets according to ligand chemistry in order to reveal unexpected relationships that may be assayed using the ligands themselves.

## 27. SLAP

Semantic Link Association Prediction uncovers "missing links" from a wide range of databases, including compound-gene, drug-drug, protein-protein, drug-side effects, to create a complex network of compound-target interactions for which there is no experimental data but which are statistically probable.

## 28. SPiDER

Computer assisted drug design that identifies the macromolecular targets of chemical entities using self-organizing maps, consensus scoring, and statistical analysis to successfully identify targets for both known drugs and computer-generated molecular scaffolds.

## 29. SuperDRUG2

Knowledge-base of approved and marketed drugs covering 4,587 active pharmaceutical ingredients, including small molecules and biological products. The database is intended as a one-stop resource providing data on: chemical structures, regulatory details, indications, drug targets, side-effects, physicochemical properties, pharmacokinetics and drug-drug interactions.

### 30. SuperPred

A prediction web server for ATC codes and target prediction of compounds. Predicting ATC codes or targets of small molecules, and thus gaining information about the compounds, assists in the drug development process for new compounds. The web server's ATC prediction as well as target prediction is based on an internal pipeline consisting of 2D fragment and 3D similarity search results. In this context, also see PASS.

### 31. SuperTarget

A web-based, searchable data warehouse integrating drug-related information associated with medical indications, adverse drug effects, drug metabolism, pathways and Gene Ontology (GO) terms for target proteins. The current database version contains more than 6,000 target proteins, which are annotated with more than 330,000 relationships to 196,000 compounds (including approved drugs).

### 32. SwissTargetPrediction

A web server for target prediction of bioactive small molecules using drawn structures or SMILES notation for input. By analyzing a combination of 2D and 3D similarity measures, the underlying software compares the query molecule to a library of 280,000 compounds active on more than 2,000 targets in 5 different organisms. Mapping predictions by homology within and between different species is enabled for close paralogs and orthologs.

### 33. TargetMine

An integrative data analysis platform for gene set analysis and knowledge discovery in a data warehouse framework. The current version has been optimized for target prioritization in response to queries, while also surveying the wider biological and chemical data space, relevant to drug discovery and development. Search functions permit users to identify chemicals and drugs that interact with a given set of genes/proteins and integrate them into gene-disease associations and regulatory networks.

### 34. TargetNet

A web service for predicting potential drug-target interaction profiling via multi-target SAR models. Naïve Bayes models together with various molecular fingerprints are employed to construct prediction models against 623 human proteins with highly validated druggable features.

### 35. TTD

---

Therapeutic Target Database is a resource for facilitating bench-to-clinic research of targeted therapeutics. Current coverage for searches based on similarities to the input query includes interconnected information on 3,100 target and 34,000 drugs and drug-like compounds, consisting in part of 2,500 approved drugs and 18,900 investigational agents.

### 36. WITHDRAWN

A resource for withdrawn and discontinued drugs to be used as heuristic templates for future predictions. The searchable database comprises 578 withdrawn or discontinued drugs, their structures, important physico-chemical properties, protein targets and relevant signaling pathways. A special feature of the database calls out the drugs withdrawn due to adverse reactions and toxic effects, so that similarity searches can uncover possible liabilities in query molecules.

**Please, leave your comments below, or suggest other useful resources, which did not make it to this list.**