The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data

Mahendra Awale and Jean-Louis Reymond*

Abstract

Background: Several web-based tools have been reported recently which predict the possible targets of a small molecule by similarity to compounds of known bioactivity using molecular fingerprints (fps), however predictions in each case rely on similarities computed from only one or two fps. Considering that structural similarity and therefore the predicted targets strongly depend on the method used for comparison, it would be highly desirable to predict targets using a broader set of fps simultaneously.

Results: Herein, we present the polypharmacology browser (PPB), a web-based platform which predicts possible targets for small molecules by searching for nearest neighbors using ten different fps describing composition, substructures, molecular shape and pharmacophores. PPB searches through 4613 groups of at least 10 same target annotated bioactive molecules from ChEMBL and returns a list of predicted targets ranked by consensus voting scheme and p value. A validation study across 670 drugs with up to 20 targets showed that combining the predictions from all 10 fps gives the best results, with on average 50% of the known targets of a drug being correctly predicted with a hit rate of 25%. Furthermore, when profiling a new inhibitor of the calcium channel TRPV6 against 24 targets taken from a safety screen panel, we observed inhibition in 5 out of 5 targets predicted by PPB and in 7 out of 18 targets not predicted by PPB. The rate of correct (5/12) and incorrect (0/12) predictions for this compound by PPB was comparable to that of other web-based prediction tools.

Conclusion: PPB offers a versatile platform for target prediction based on multi-fingerprint comparisons, and is freely accessible at www.gdb.unibe.ch as a valuable support for drug discovery.

Keywords: Target prediction, Polypharmacology, Drug–target interactions, Molecular fingerprints

Background
The vast majority of small molecule drugs interact with multiple targets, a general phenomenon known as polypharmacology and a key parameter to be addressed in the course of drug development [1–3]. Many computational tools have been developed that exploit databases containing detailed structural information on the activity of small molecule drugs [4–8] and their protein targets [9] to predict the polypharmacology of any hit compound or drug candidate [10–37]. Several of these tools are accessible as target prediction websites (Table 1). Each of these websites returns a list of predicted targets based on similarity calculations using molecular fingerprints (fps) or on docking scores.

Ligand-based methods using fp comparisons are particularly versatile because they are applicable to any biological activity, i.e. the target may be a protein but also a cell line or a whole organism. Although some of these ligand based tools offer a selection of different fps, none of them permits the simultaneous use of more than one fp. Considering the fact that molecular similarity and therefore the predicted targets strongly depend on which fp is used for comparison, it would be highly desirable to predict targets using multiple fps simultaneously,
provided that each fp would perform well individually in virtual screening and target prediction test cases. Herein we present the polypharmacology browser (PPB), a multi-fingerprint browser for target prediction which addresses this issue by performing target predictions searches using six different fps and four fused molecular fingerprints (Ffps) (Table 2). Similarities are measured using the city-block distance because this similarity measure is rapidly computed and therefore well-suited for web-based similarity searches in large databases [38–45]. PPB searches through 2.7 M ligand-target interactions extracted from ChEMBL 21 and generates a list of predicted targets, each linked to the lists of known actives used for the prediction. PPB validation is presented for 670 drugs of known polypharmacology as well as in a predictive application of off-targets for a recently reported inhibitor of transient receptor potential vanilloid 6 (TRPV6) [46]. PPB is freely accessible at www.gdb.unibe.ch and works on computers, tablets and phones.

**Methods**

**Dataset**

We analyzed ChEMBL 21 and constructed the target database containing 4613 groups of at least 10 bioactive molecules with documented activity against the same biological target. Briefly, target database was constructed as follows: initially all targets along with their ligands were retrieved from the ChEMBL version 21.

| Table 1 Publicly accessible web-based target prediction tools |
| Website | Similarity method | Database | Ref. |
|---------|------------------|----------|------|
| www.gdb.unibe.ch | 10 different fingerprints | ChEMBL 21 | This work |
| http://www.pharmaexpert.ru/passonline/tarfsdock/ | Multilevel neighbourhoods of atoms (MNA) descriptors | WDI and ACD | [10] |
| http://sea.bkslab.org | Docking | PDTD | [12] |
| http://59.78.96.61/pharmmapper/ | ECfp4 | CHEMBL 16, WOMBAT, MDOR and StarLite | [14] |
| https://cpi.bio-x.cn/drar/ | Receptor-based pharmacophore models | TargetBank, DrugBank, BindingDB, PDTD | [17] |
| www.cbigand.org/TargetHunter | Docking | PDB, DrugBank | [18] |
| http://ilab.eust.edu.cn/chemmapper/ | ECfp6, ECfp4 and Openbabel FP2 | ChEMBL 11 and PubChem bioassay | [22] |
| http://mips.helmholtz-muenchen.de/proj/hitpick | Circular fingerprint FCFP | STITCH | [24] |
| http://modlab-cadd.ethz.ch/software/spider/ | CATS and MOE physiochemical descriptors | COBRA | [26] |
| www.swisstargetprediction.ch | Openbabel FP2 and Electroshape descriptors | ChEMBL 16 | [28] |
| http://prediction.charite.de/index.php | ECfp4 | ChEMBL, SuperTarget and BindingDB | [29] |
| www.dddc.ac.cn/tarpred | ECfp4 | BindingDB | [33] |
| http://potential.cbs.dtu.dk/ChemProt | Sfp | ChEMBL 14, BindingDB, DrugBank, PharmGKB, PubChem bioassay, WOMBAT, IUPHAR, CTD and STITCH | [37] |

| Table 2 Molecular fingerprints used for target prediction with PPB |
| Name | Description | Ref. |
|-------|-------------|------|
| APfp | 21-D atom-pair fingerprint, perceives molecular shape | [43] |
| Xfp | 55-D atom category extended atom-pair fingerprint, perceives pharmacophores | [43] |
| MQN | 42-D Molecular Quantum Numbers, scalar fingerprint counting atoms, bonds, polarity and ring features, perceives constitution, topology and molecular shape | [38] |
| SMIfp | 34-D scalar fingerprint counting occurrence of characters in SMILES, perceives rings, aromaticity, and polarity | [42] |
| Sfp | 1024-D binary daylight type substructure fingerprint, perceives detailed substructures | [49] |
| ECfp4 | 1024-D binary circular extended connectivity fingerprint, perceives detailed substructures and pharmacophores | [50] |
| Ffp1 | Fusion fingerprint, Xfp + SMIfp + Sfp | This work |
| Ffp2 | Fusion fingerprint, Xfp + MQN + SMIfp | This work |
| Ffp3 | Fusion fingerprint, Xfp + SMIfp + Sfp + ECfp4 | This work |
| Ffp4 | Fusion fingerprint, Xfp + MQN + SMIfp + Sfp + ECfp4 | This work |
For each target we retained the compounds having IC_{50}, EC_{50}, GI_{50}, K_{i}, K_{D}, or potency value of \( \leq 10 \) \( \mu \)M or percent inhibition of \( \geq 50\% \). All molecules were processed as non-stereo SMILES and ionized at pH 7.4 using an in-house developed Java program utilizing the JChem chemistry library from ChemAxon Pvt. Ltd. Afterwards, duplicate molecules were removed in the context of each target. Finally, targets with at least 10 bioactive compounds were retained in database. In total, these 4613 targets represent 871,673 unique bioactive compounds and 2.7 M ligand-target interactions. Of these targets, 60% are single protein type, 55% are human targets, and 45% have less than 50 bioactive compounds (Fig. 1a–c).

### Fingerprints

For each of the 871,673 selected ChEMBL compounds we computed each of the six fingerprints described in Table 2. In short, APfp, Xfp, MQN and SMIfp were calculated by using an in house developed Java program utilizing various calculator plugins from JChem chemistry library, in particular TopologyAnalyserPlugin to determine shortest topological path for atom-pair, HBDAPlugins to determine hydrogen bond donor and acceptor atoms, and MajorMicrospeciesPlugin to adjust the ionization state of molecules. The detailed procedure for the generation of these fingerprints can be found in the respective publication [38, 42, 43]. For Sfp, a daylight type 1024-bit hash fingerprint was computed using the ChemicalFingerprint class of the JChem library. The source codes for computation of fingerprints are freely available for download at www.gdb.unibe.ch.

### Fused fingerprints (Data fusion)

To generate additional molecular fingerprint descriptions we further investigated data fusion between different combinations of these fingerprints [47]. Since we aimed at using the city-block distance (Eq. 1) as similarity measure, we scaled each fingerprint by analyzing the distance distribution of 50 M random pairs of compounds in each fingerprint space, and scaled values to adjust the most frequently occurring distance in each fp to the value for Xfp (Fig. 1d/e). We then performed enrichment studies of ligands against decoys in the directory of useful decoys (DUD) [48] and evaluated the average performance of 57 different combinations of the scaled fingerprints in terms of area under the curve (AUC) and enrichment factor at 1% screening in the receiver operator characteristic (ROC) curves (data not shown). We selected the four fusion fingerprints Ffp1–Ffp4 (Table 2) due to their good performance in this enrichment study (Fig. 1f/g), and computed the corresponding Ffp1–4 values for the 871,673 ChEMBL compounds.

\[
CBD_{A,B} = \sum_{j=1}^{K} |A_j - B_j|
\]

### p value calculation

Each target prediction for a given query molecule is based on the city-block distance between the query molecule and the closest member of a group of compounds associated with this target. A \( p \) value can be computed for each prediction as the degree of randomness of the observed city block distance [51] and therefore the probability that the corresponding query–target association occurs at random. To compute the \( p \) value, we generated a random distance distribution for each of the 4613 targets in each of the ten fingerprint spaces by computing distances between the ChEMBL compounds associated with the target and randomly selected molecules from the ZINC database [52], taken as representative molecules of screening compounds for which target predictions might be carried out (Fig. 1h). For each target up to 1 M random pairs were considered. We then fitted each of the 46,130 distance distributions using a negative binomial distribution function, and generated the corresponding cumulative density functions giving the \( p \) value as a function of the city block distance (Fig. 1i). The choice of a negative binomial distribution function was based on the discrete nature of the city-block distance. As the \( p \) value calculation is specific for each target protein and fingerprint space it can only be used in this context, and should not be used to compare molecules from different targets or fingerprint spaces. The curve fitting was carried out using the R statistical package version 3.2.5 using the “fitdistrplus” library with default maximum likelihood method for parameters estimation.

### Validation set

The validation set containing 670 drugs and their targets annotation was created as follows: (a) compounds labelled as approved drug or drug in clinical trial were extracted from ChEMBL database, (b) for each drug, target list was constructed by comparing SMILES string of drug to SMILES of bioactive compounds of targets used in the PPB. When the SMILES string was matched, corresponding PPB target was added to known target list for the drug and (c) retained the drugs with \( \leq 20 \) targets in the list.

### The PPB web-interface

Given a query molecule, PPB computes each of the 10 fps for this compound, sorts the 871,673 ChEMBL compounds by city-block distance to the query using each fp
independently, and collects a predefined number of compound associated targets in each case. The $p$ value for each target is calculated from the nearest neighbor found for that target. Finally, PPB merges the different target lists.

The graphical user interface of PPB starts with an initial page wherein the user can input the structure of a query molecule using the JavaScript based JSME molecular editor (http://peter-ertl.com/jsme/) [53]. The structure can be drawn or copy-pasted in SMILES or sdf file format. An option is available to extract the query molecule from the Protein Data Bank (PDB) using the PDB id of the protein–ligand complex of interest (the PDB ligand data was downloaded from http://ligand-expo.rcsb.org/ website in March 2016 and stored on our web server, and will be updated once a year). The option “No. of targets” allows the user to input the number of targets to be returned by each fp. By default this parameter is set to 20. Following the entry of a query molecule, the target search can be initiated by clicking on the “Submit” button. Typically, execution of a search takes less than 1 min.

The target prediction results are presented as list of targets annotated with a probability bar for each fp (Fig. 2a). The $p$ value is shown by a green bar of decreasing length up to $p = 0.01$, with values above 0.01 written as number in

![Fig. 1](image.png)

**Fig. 1** Overview of the data used for constructing PPB. Distribution of a target type as defined in ChEMBL and b source of targets. c Distribution of targets as per number of associated bioactive compounds. d Histogram of city block distances (log scale) calculated for 50 million random pairs of compounds from ChEMBL 21 using six molecular fingerprints (APfp, MQN, SMIfp, Sfp and ECfp4) fingerprints were scaled with respect to Xfp to adjust to the value of the most frequent distance. Scaling factors are shown in parentheses. e Enrichment of 40 set of DUD actives from corresponding decoys set by six different fingerprints (APfp, Xfp, MQN, SMIfp, Sfp and ECfp4) and four similarity fusion methods (Fp1–4). City block distance was used as sorting function. Data is represented as average of f Area under ROC curve and g Enrichment factor at 1% of screen database for 40 targets from DUD. h Example of $p$ value calculation. h Observed (red) and fitted (black) random distance distributions for the muscarinic acetylcholine receptor M1 (CHRM1, CHEMBL216) in MQN fingerprint space. City block distances were calculated for 1788 ligands of CHRM1 with respect to random compounds from ZINC database. Negative binomial distribution was used for curve fitting. i Cumulative density plot indicating area under fitted curve in h.
the structures of these compounds labelled with finger-
target (furthest right column) opens a new tab displaying
remaining 29% had 10–20 associated targets (Fig. 3).
these drugs 71% had less than 10 associated targets and the
removed from our target compound database. Among
get interactions). Prior to evaluation these 670 drugs were
labelled as approved drug or drug in clinical trial and anno-
gated to provide in each window.
prints, city-block distance and ChEMBL compound id
(target name, and by the number of selected compounds by
2 different fps. For each drug, we calculated the fraction
of known targets in the results list, the total number of
predicted targets and the hit rate (ratio of known targets
found to predicted targets).
In terms of the fraction of correctly predicted known
targets (Fig. 4a) the fusion fingerprints Ffp1, Ffp3 and
Ffp4 showed the highest average value, followed by the
binary fingerprints Sfp and ECfp4, the pharmacophore
fingerprint Xfp, and finally MQN, SMIfp and APfp. To
evaluate if the performance of these fingerprints signifi-
cantly differ from one another or not, the student t tests
(at confidence interval of 0.95) were performed for all the
possible pairs of fingerprints (Additional file 1: Tables S1
and S2). Although most of the individual fingerprint pairs
show significant differences in performance, no signifi-
cant differences were found between performance of Sfp,
ECfp4, Ffp1 and Ffp2. Overall the performance trend fol-
lowed the complexity of each fingerprint and highlights
that detailed structural encoding tends to increase pre-
diction performance.
These single fp approaches were outperformed by using
the combination of all 10 results lists (combined method),
however at the expense of a relatively low hit rate result-
ing from checking a larger number of targets for each
drug as compared to individual fps (Fig. 4b/c). The com-
bined method also showed the highest success rate for
finding at least one known target among the 5 top pre-
dicted targets for each fp (Fig. 4d). For all methods except
SMIfp the predicted target list with p value ≤0.01 showed
a significantly greater chance of success compared to the
target lists with p value >0.01. The importance of a low p
value for target prediction was particularly striking with
the fused fingerprints Ffp1–4 and the combined method,
for which more than 75% of all correct predictions origi-
nated from predicted targets with p value ≤0.01.
Although several of the fps were not statistically dif-
f erent in terms of performance, the pairwise overlap
between predicted targets by each of the 10 different fps
showed that on average less than 45% of targets were com-
mon between any two fps (Fig. 4f). Furthermore each fp
returned the highest percentages of unique targets which are
not found by any other fp, further highlighting the utility
of each fp. Interestingly, APfp and Xfp, which perceive
the shape and pharmacophore patterns in molecules,
returned the highest percentages of unique targets (52
and 31% respectively). For fused fingerprints Ffp1–4 the
percentages of unique targets were relatively low (3–10%)
due to the considerable overlap among themselves.
A similar analysis performed by categorizing targets
according to their p values (Fig. 4g/h) showed that the

### Results and discussion

The performance of PPB was evaluated by challenging
its ability to recall the known targets of 670 compounds
labelled as approved drug or drug in clinical trial and anno-
tated with up to 20 targets in ChEMBL (4794 drug–tar-
get interactions). Prior to evaluation these 670 drugs were
removed from our target compound database. Among
these drugs 71% had less than 10 associated targets and the
remaining 29% had 10–20 associated targets (Fig. 3).

The 670 different searches were performed using PPB
with the default search settings. The predicted targets
were analysed for each fp considering (a) all targets, (b)
targets with p value ≤0.01 and (c) targets with p value

![](image.png)

**Fig. 3** Distribution of 670 ChEMBL compounds labelled as drug in
clinical trial or as approved drug used in validation study according to
number of associated targets.
pairwise overlap between high confidence (low \( p \) value) targets of different fingerprints were significantly higher (on average 47% overlap) as compared to low confidence targets (high \( p \) value, on average 24% overlap). This can be explained by the fact that at high \( p \) values the structural similarity between a query and its nearest neighbour compound associated with the target becomes less obvious and difficult to capture (Fig. 4e).

**Prediction of off-targets of a new TRPV6 inhibitor**

We recently reported the identification of CIS22a (Fig. 5a) as the first potent and selective inhibitor of
TRPV6, a transmembrane calcium channel overexpressed in breast and prostate cancer [46]. We tested the polypharmacology of this inhibitor for 24 out of 44 targets present in the “safety screen” panel of Cerep Pvt. Ltd (Fig. 5b/c). To assemble this target list we first inspected the PPB results list and chose five targets selected by multiple fps considering in each case the target from human and rat origin as the same target. These were the adrenergic α1A receptor (ADRA1A), the dopamine receptor subtypes D1 (DRD1), D2 (DRD2), and D4 (DRD4), the 5-hydroxytryptamine receptor 1A (HTR1A). We then added further subtypes of these five targets as well as all ion channels present in the safety screen panel, resulting in a list of 17 GPCRs and seven ions channels.

In-vitro profiling showed that CIS22a bound significantly (>50% inhibition at 10 µM) to 12 targets of the 24 selected targets (Fig. 5b). Five of these targets (ADRA1A, DRD1, DRD2, DRD4 and 5HTR1A) were proposed by multiple fps in the PPB. Only two of the fps (Ffp1 and Ffp3) and the combined method were able to predict all of the five targets, illustrating the usefulness of similarity fusion methods and combined data analysis. The analysis of bioactive compounds which linked these five targets to CIS22a showed that the linking compounds...
were shared by different fps and closely related targets (1–8 in Fig. 5d). Interestingly, the linking compounds for three targets (DRD2, DRD4 and HTR1A) suggested by Xfp were not shared by any other fps, probably because of their relatively low substructure similarity to the query compound. On the other hand Xfp predicted an activity on ADRA2A, which was not confirmed experimentally (only 31% inhibition at 10 μM).

For comparison we successfully ran target predictions for CIS22a using six of the fourteen target prediction web-based tools listed in Table 1. Results comparable to PPB were obtained with SwissTarget, SuperPred, Target-Hunter, ChemMapper and ChEMBLPred. On the other hand, SEA only returned a single, correct target, and PharmMapper did not predict any of the tested targets.

Conclusion

The PPB web tool features a unique, intuitive and exhaustive search platform for target prediction. The comparative view of target list from various fingerprint spaces provides a simple yet efficient way for selection of targets by consensus voting. PPB provides a valuable support to drug discovery projects.

Additional file

Additional file 1. Figure S1: Recovery statistics of targets of 670 drugs by various fingerprints and combined method used in PPB Tables S1 and S2: Student t-test significance values for pair wise comparison of target prediction performance of all the fingerprints used in PPB.

Authors’ contributions

MA designed and realized the project and wrote the paper. J-LR co-designed and supervised the project and wrote the paper. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The website, source code and all the data used for development and validation of PPB is freely available for download at the PPB home page.

Funding

This work was supported financially by the University of Berne, the Swiss National Science Foundation and the NCCR TransCure.

Received: 29 August 2016 Accepted: 10 February 2017 Published online: 21 February 2017

References

1. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? Nat Rev Drug Discov 5(12):993–996
2. Anighoro A, Bajorath J, Rastelli G (2014) Polypharmacology: challenges and opportunities in drug discovery. J Med Chem 57(19):7874–7887
3. Lavecchia A, Cerchia C (2016) In silico methods to address polypharmacology: current status, applications and future perspectives. Drug Discov Today 21(2):288–298
4. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang JW (2012) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 34(suppl 1):D668–D672
5. Olah M, Rad R, Ostopovici L, Bora A, Hadaruga N, Hadaruga D, Moldovan R, Fialus A, Mractc M, Oprea TI (2008) WOMBAT and WOMBAT-PK: bioactivity databases for lead and drug discovery, chemical biology: from small molecules to systems biology and drug design. Wiley-VCH Verlag GmbH, Weinheim, pp: 760–786
6. Wang Y, Xiao J, Suzeck TO, Zhang J, Wang J, Bryant SH (2009) Pubchem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res 37(Web Server issue):W623–W633
7. Gaulton A, Bells LJ, Bento AF, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Malichovách D, Al-Lazikani B et al (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40(D1):D1100–D1107
8. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong JS (2015) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res 44(D1):D1045–D1053
9. Rose PW, Prié A, Bi C, Bluhm WF, Christie CH, Dutta S, Green RK, Goodsell DS, Westbrook JD, Woot J et al (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res 43(D1):D345–D356
10. Lagunin A, Stepanchikova A, Filimonov D, Porokov V (2000) PASS: prediction of activity spectra for biologically active substances. Bioinformatics 16(8):747–748
11. Jenkins JL, Bender A, Davies JW (2006) In silico target fishing: predicting biological targets from chemical structure. Drug Discov Today Technol 3(4):413–421
12. Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K, Luo X, Zhu W, Chen K, Shen J et al (2006) TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Res 34(suppl 2):W219–W224
13. Nidhi G, Glick M, Davies JW, Jenkins JL (2006) Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. J Chem Inf Model 46(3):1124–1133
14. Keiser MJ, Roth BL, Armbuster BN, Ermsberger P, Irwin JJ, Shoichet BK (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25(2):197–206
15. Nigsch F, Bender A, Jenkins JL, Mitchell JBO (2008) Ligand–target prediction using winnow and naive Bayesian algorithms and the implications of overall performance statistics. J Chem Inf Model 48(12):2313–2325
16. Wale N, Karypis G (2009) Target fishing for chemical compounds using target–ligand activity data and ranking based methods. J Chem Inf Model 49(10):2190–2201
17. Liu X, Ouyang S, Yu B, Liu Y, Huang K, Gong J, Zheng S, Li Z, Li H, Jiang H (2010) PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. Nucleic Acids Res 38(suppl 2):W609–W614
18. Luo H, Chen J, Shi L, Mikallov M, Zhu H, Wang K, He L, Yang L (2011) DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical–protein interactome. Nucleic Acids Res 39(suppl 2):W492–W498
19. Koutsoukas A, Simms B, Kirchmayr J, Bond PJ, Whitmore AV, Zimmer S, Young MP, Jenkins JL, Glick M, Lan RC et al (2011) From in silico target prediction to multi-target drug design: current databases, methods and applications. J Proteome 74(12):2554–2574
20. Abdulhameed MDM, Chaudhury S, Singh N, Sun H, Wallqvist A, Tawa GJ (2012) Exploring polypharmacology using a ROCs-based target fishing approach. J Chem Inf Model 52(2):492–505
21. Mavridis F, Bender A, Jenkins JL, Mitchell JBO (2008) Ligand–target prediction to multi-target drug design: current databases, methods and applications. J Chem Inf Model 48(12):2313–2325
22. Wang L, Ma C, Wipf P, Liu H, Su W, Xie X-Q (2013) TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. AAPS J 15(2):395–406
23. Gong J, Cai C, Liu X, Xu K, Jiang H, Gao D, Li H (2013) ChEMapper: a versatile web server for exploring pharmacology and chemical structure association based on molecular 3D similarity method. Bioinformatics 29(14):1827–1829
24. Liu X, Vogt I, Haque T, Campillos M (2013) HitPick: a web server for hit identification and target prediction of chemical screenings. Bioinformatics 29(15):1910–1912
25. Peragovics Á, Simon Z, Tombor L, Jelinek B, Hári P, Czobor P, Málnási-Csizmadia A (2013) Virtual affinity fingerprints for target fishing: a new application of drug profile matching. J Chem Inf Model 53(1):103–113
26. Reker D, Rodrigues T, Schneider P, Schneider G (2014) Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proc Natl Acad Sci 111(11):4067–4072
27. Clark AM, Sarker M, Ekins S (2014) New target prediction and visualization tools incorporating open source molecular fingerprints for TB Mobile 2.0. J Cheminform 6(1):1–17
28. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 42(W1):W32–W38
29. Nickel J, Gohlke B-O, Erehman J, Banerjee P, Rong WW, Goede A, Dunkel M, Preissner R (2014) SuperPred: update on drug classification and target prediction. Nucleic Acids Res 42(W1):W26–W31
30. Cereto-Massagué A, Ojeda MJ, Valls C, Mulero M, Pujadas G, Garcia-Vallve S (2015) Tools for in silico target fishing. Methods 71:98–103
31. Lusci A, Fooshee D, Browning M, Swamidass J, Baldi P (2015) Accurate and efficient target prediction using a potency-sensitive influence-relevance voter. J Cheminform 7(1):1–13
32. Liu X, Gao Y, Peng J, Xu Y, Wang Y, Zhou N, Xiong J, Luo X, Jiang H, Zheng M (2015) TarPred: a web application for predicting therapeutic and side effect targets of chemical compounds. Bioinformatics 31(12):2049–2051
33. Afzal AM, Mussa HY, Turner RE, Bender A, Glen RC (2015) A multi-label approach to target prediction taking ligand promiscuity into account. J Cheminform 7(1):1–14
34. Cao R, Wang Y (2016) Predicting molecular targets for small-molecule drugs with a ligand-based interaction fingerprint approach. ChemMedChem 11(12):1352–1361
35. Willett P (2013) Fusing similarity rankings in ligand-based virtual screening. Comput Struct Biotechnol J 5(6):1–6
36. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49(23):6789–6801
37. Hagadone TR (1992) Molecular substructure similarity searching: efficient retrieval in two-dimensional structure databases. J Chem Inf Comput Sci 32(5):515–521
38. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50(5):742–754
39. Baldissera E, Nastri R (2010) When is chemical similarity significant? The statistical distribution of chemical similarity scores and its extreme values. J Chem Inf Model 50(7):1205–1222
40. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 52(7):1757–1768
41. Bienfait B, Ertl P (2013) JSME: a free molecule editor in JavaScript. J Cheminform 5(1):1–6