SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules

Antoine Daina [1], **Olivier Michielin**[1,2,*] **and Vincent Zoete**[1,3,*]

[1]Molecular Modeling Group, SIB Swiss Institute of Bioinformatics, University of Lausanne, Quartier UNIL-Sorge, Bâtiment Amphipôle, CH-1015 Lausanne, Switzerland, [2]Department of Oncology, University Hospital of Lausanne, Ludwig Cancer Research – Lausanne Branch, CH-1011 Lausanne, Switzerland and [3]Department of Fundamental Oncology, University of Lausanne, Ludwig Cancer Research – Lausanne Branch, Route de la Corniche 9A, CH-1066 Epalinges, Switzerland

ABSTRACT

SwissTargetPrediction is a web tool, on-line since 2014, that aims to predict the most probable protein targets of small molecules. Predictions are based on the similarity principle, through reverse screening. Here, we describe the 2019 version, which represents a major update in terms of underlying data, back-end and web interface. The bioactivity data were updated, the model retrained and similarity thresholds redefined. In the new version, the predictions are performed by searching for similar molecules, in 2D and 3D, within a larger collection of 376 342 compounds known to be experimentally active on an extended set of 3068 macromolecular targets. An efficient back-end implementation allows to speed up the process that returns results for a druglike molecule on human proteins in 15–20 s. The refreshed web interface enhances user experience with new features for easy input and improved analysis. Interoperability capacity enables straightforward submission of any input or output molecule to other on-line computer-aided drug design tools, developed by the SIB Swiss Institute of Bioinformatics. High levels of predictive performance were maintained despite more extended biological and chemical spaces to be explored, e.g. achieving at least one correct human target in the top 15 predictions for >70% of external compounds. The new SwissTargetPrediction is available free of charge (www.swisstargetprediction.ch).

INTRODUCTION

In any endeavor to discover, develop or repurpose bioactive molecules, it has become key to identify the targeted proteins. Nowadays, efficient support can be provided by established bio-/chemo-informatics approaches to estimate the most probable targets of small molecules (1,2). These target prediction (also named target fishing) methods may be classified in one of the two traditional categories of computer-aided molecular design, i.e. making use of the three-dimensional structure of the protein (structure-based) or not (ligand-based) (3).

Ligand-based target prediction has proven highly performant and fast in predicting correct protein targets of compounds in drug discovery contexts (4,5). The quantification of similarity between compounds by different means enabled the validation of the intuitive 'molecular similarity hypothesis' which postulates common proteins targeted by similar molecules (6,7).

SwissTargetPrediction is a web-based tool, on-line since 2014, to perform ligand-based target prediction for any bioactive small molecule. The user-friendly graphical interface shields non-experts from methodological pitfalls and specialists from tedious technical efforts. This allows anyone to achieve reverse screening towards previously carefully prepared chemical libraries. Consequently, the website has been increasingly visited and the service increasingly used. In 2018, 20 726 unique visitors opened 44 641 sessions to submit 139 944 jobs. Compared to 2017, this represents an increase of 63%, 55% and 87%, respectively. As a result, at the time of writing the manuscript, we recorded total numbers reaching 62 890 unique visitors having opened 135 695 sessions to submit 357 808 jobs since 2014. Geographically, users are coming from 159 countries (top 10 is China, India, USA, Brazil, Russia, Egypt, Switzerland, Mexico, UK and Italy).

The unique engine behind SwissTargetPrediction, extensively detailed elsewhere (8), calculates the similarity between the user's query compounds and those compiled in curated, cleansed collections of known actives in well-defined experimental binding assays. The quantification of similarity is 2-fold. In both cases, it consists in computing a pair-wise comparison of 1D vectors describing molecular structures: the 2D measure uses the Tanimoto index between path-based binary fingerprints (FP2) (9), while the 3D measure is based on a Manhattan distance similarity quantity between Electroshape 5D (ES5D) (10,11) float vectors. The latter mines five descriptors for each atom of 20 previously generated conformations (Cartesian coordinates, partial charge and lipophilic contribution). For both 2D and 3D similarity measures, the principle is that two similar molecules are represented by analogous vectors, which exhibit a quantified similarity close to 1. The SwissTargetPrediction model was trained by fitting a multiple logistic regression on various size-related subsets of known actives in order to weight 2D and 3D similarity parameters in a so-called Combined-Score. A Combined-Score higher than 0.5 predicts that the molecules are likely to share a common protein target. In reverse screening, the Combined-Score allows to calculate for any query molecule, assumed as bioactive, the probability to target a given protein. As 2D and 3D description of molecules are complementary, this dual scoring ligand-based reverse screening showed high performance in predicting macromolecular targets in various test sets (8) (12,13).

The present report details the updates of SwissTargetPrediction released in 2019. The novelties primarily regard the dataset used to construct the collection of known actives. This set relies on the bioactivity data of ChEMBL version 23 (14,15) and thus describes more molecules and more targets than the old version (based on ChEMBL version 16). The predictive model was re-trained on this larger data set. Doing so, we focused the screening on similar molecules only, which contain the most useful information. The backend was also redesigned. The current implementation is more than 30% faster, although reverse screenings are performed toward 34% more molecules. Finally, although keeping a look-and-feel comparable to the trusted previous version, the new web interface includes numerous improvements. The reader is encouraged to employ the 2019 version of SwissTargetPrediction (www.swisstargetprediction.ch). However, the previous version will remain accessible for at least one year (old.swisstargetprediction.ch) to ensure the continuity of user's ongoing projects.

UPDATES

Technologies

The screening engine of SwissTargetPrediction was mainly coded in Python 2.7. Important improvements were implemented for chemical handling, which is now performed by JChem REST Web Services (www.chemaxon.com, version 18.29.0) and OpenBabel (www.openbabel.org, version 3.4.1) (9). A queuing system, based on Slurm (https://slurm.schedmd.com, version 17.11.2), was setup on the calculation server. This allows to increase both the process speed

and the user's comfort with a monitoring of the progress of the computation.

The website was written in HTML5, PHP (version 7.2.10) and JavaScript. Most significant novelties are (i) the full integration of JChem Web Services to Marvin JS sketcher (www.chemaxon.com, version 18.29.0) for providing multiple ways of inputting molecules; (ii) the use of the DataTables plug-in (version 1.10.18) for dynamic output tables and advanced export options; and (iii) the pie chart of target classes, which is now generated with JPgraph (https://jpgraph.net, version 4.2.6).

Data and model

The wealth of available bioactivity information regarding small molecules has greatly increased compared to the 2014 version of SwissTargetprediction. The same strict filtering criteria that were applied to ChEMBL16 (12) led to extract more high-quality interaction data from ChEMBL23 for more compounds, active on more proteins. Overall the dataset includes 376 342 unique compounds (increase of 34%), showing 580 496 binding activities (increase of 32%) on 3068 protein targets (increase of 19%). All numbers, broken down by species, are given in Table 1.

This significantly larger amount of data has a direct and positive impact on the screening since more active compounds can be found similar to the user's query molecule and more protein targets can be predicted. Moreover, the probability of sharing a common target as a function of 2D similarity (3D similarity, respectively) was calculated on the entire matrices of pair-wise comparisons of FP2 vectors (ES5D vectors, respectively; see Supplementary Figure S1 in the Supplementary Data). It clearly shows that the information meaningful for the model corresponds to Tanimoto indexes higher than 0.65 (Manhattan-based similarities higher than 0.85, respectively). Discarding information below these thresholds reduces noise in the model, which now focuses on highly similar active molecules only. In our experience, we found that the shorter lists of more similar known actives (in 2D and in 3D) selected among a larger pool of screenable molecules, returned by SwissTargetPrediction 2019 compared to the old version, are more descriptive and more informative for drug design/discovery applications.

Graphical interface (user experience)

The new web interface keeps a look-and-feel similar to the old version, yet with numerous additions and significant improvements meant for enhancing the user experience. As several processes are performed through JChem REST Web Services, the client's internet connection and browser must allow binding to web services. The user is now informed that the SwissTargetPrediction website is optimized for Google Chrome, Firefox and Safari, when trying to use another browser.

Input

The user's journey starts at the input page directly accessible without log in nor registration at URL

Table 1. Comparative volume of bioactivity data between versions 2019 and 2014 of SwissTargetPrediction (extracted from ChEMBL23 and ChEMBL16, respectively)

	SwissTargetPrediction 2019				SwissTargetPrediction 2014			
Species[a]	Human	Rat	Mouse	All	Human	Rat	Mouse	All
Number of targets	2092	535	441	3068	1768	469	342	2579
Number of active compounds	327 719	53 819	14 215	376 342	236 554	43 895	9540	280 381
Number of interactions	494 196	69 661	16 639	580 496	369 897	58 729	11 908	440 534

[a]Cow and horse data were withdrawn from version 2019, because not numerous enough and rarely accessed by users of the 2014 website.

Figure 1. SwissTargetPrediction input page. Major novelties and updated items are boxed in blue. These include the upper black banner that provides links to other in-house CADD online tools, including new ones. Links to the old version of SwissTargetPrediction, FAQ, help and download pages are provided in the header's menu. After having selected the species, the query molecule is inputted either as SMILES or by using the MarvinJS sketcher for opening, importing or modifying a molecular structure. A 'Clear' button allows deleting all inputs, and examples can be selected from a drop-down menu. The 'Predict targets' button becomes red and active when an input molecule is given.

www.swisstargetprediction.ch (Figure 1). The upper black banner provides links to an updated list of other CADD online tools developed by the SIB Molecular Modeling Group (16–21). Another way to access these web tools are the so-called interoperability icons appearing in the output pages (refer to the *Output* section). The header's menu gives access to the old version of SwissTargetPrediction (old.swisstargetprediction.ch) as well as to renewed FAQ, help and download pages (see the *Other pages* section). The latter was supplemented with the external test set on which the predictive power of the new model was evaluated (see the *Validation* section). After having selected the species of origin for the prediction among human, rat and mouse (cow and horse were removed because of the small amount of data and the scarce usage on the 2014 website), the query molecule can be inputted either as SMILES in the left-hand dedicated text box or through the MarvinJS molecular editor on the right, which enables convenient drawing of chemical structures. The new implementation of the sketcher includes features connected with JChem Web services for the import of the compound from a file (locally or through the network), or by name (chemical or common, if any). Note that the sketcher has now copy/paste capabilities and that it is synchronized with the SMILES text box. This provides an easy way to modify molecules. Examples can be selected from a drop-down menu. The 'Clear' button allows deleting the molecule from the text box and the sketcher.

The 'Predict targets' button becomes red and active upon providing a query molecule. By clicking on it, the computation process is initiated with chemical handling of the input structure through JChem Web services (version 18.29.0) and Openbabel (version 2.4.1). Kekulization, canonicalization and hydrogenation as at pH 7.4 are applied on the SMILES, which is then translated both into (i) FP2 binary vectors, and (ii) 20 conformers to finally generate ES5D vectors. A validation of the input molecule through JChem Structure Checker was implemented. This procedure returns an error message to the user when the query compound contains a bad valence, problems with aromaticity, broken bonds, or unknown atom/pseudoatom name. Then, the actual screenings (both 2D and 3D) are started to find similar molecules among the compounds known to be experimentally active on one or several of the 3068 proteins.

Output

Immediately after submission, the status of the computation appears, starting with the structural validation. This transient page is linked to the workload management system (Slurm version 17.11.2) to inform the user about his/her job being queued or in progress. The different steps of computation are transparent to the user and the overall process can be followed on a progression bar. Started calculations take about 15 to 20 seconds for a compound of the size of a druglike molecule. When the progression bar is completed, the predictions are displayed in a first output page. The major information presented in this result page (Figure 2) consists in the tabulated list of probable protein targets as predicted by the dual score ligand-based reverse screening of the query molecule towards the collection of known actives. By default, the table rows are ranked according to the probability of the corresponding protein (thoroughly described by clicking on the link to Genecard, UniProt or ChEMBL) to be an actual target of the query molecule. In version 2019, the targets consisting in protein complex are given in full name (linked to a unique ChEMBL ID), while their subunits/components are linked individually to Genecard and Uniprot. As mentioned earlier and described in details elsewhere (12), the probability values, depicted here as green bars, are calculated from the Combined-Scores of the most similar compounds to the query molecule (in 2D and 3D) known to be active on a given protein. Importantly, this value precisely illustrates the probability for a bioactive molecule to have a given protein as target, but *not* the probability of being bioactive. Moreover, targets tagged '(by homology)' are predictions based on similar molecules active on proteins showing sufficient level of homology. The method was extensively detailed in a dedicated publication (13). A typical case is related to orthology, where a target is predicted for the chosen species, yet based on the similarity of the query molecule with compounds active on the ortholog proteins. The last column shows the numbers of compounds known to be active on each listed target and highly similar to the query molecule. The thresholds are 0.65 for Tanimoto index on FP2 (2D) and 0.85 for the Manhattan-based similarity on ES5D (3D). These numbers are hypertext links to pages containing information about these compounds, which have yielded the prediction (know actives similar in 3D and 2D in Figure 3A and 3B, respectively). Enhanced capabilities compared to the old version is provided by the implementation of the DataTables plug-in (version 1.10.18). In particular, the number of predicted targets to display can be adjusted with page browsing to 15 (default), 25, 50 or all (maximum 100). Additionally, the table can be sorted according to any column by clicking on the header and a search box allows filtering the results. Moreover, advanced export options are provided to the users by clicking on dedicated icons. The table can be downloaded in various formats (PDF, CSV or Excel) or copied into the clipboard as well as directly printed.

A box on the top-right of the result page (Figure 2) provides a summary of the predicted target classes (column 5 of the table) as a pie chart produced with JPgraph (version 4.2.6). Percentages correspond to the top 15 proteins by default, but the user has the possibility to choose the top 25, 50 or all predicted targets with the buttons appearing on the left of the chart. A full-size picture can be opened in a new tab of the browser by clicking on the pie-chart, allowing for further save or printing.

A box on the top-left of the result page (Figure 2) recalls the chemical structure of the user's query molecule. Four icons appear inside this box and give the opportunity for straightforward submission of the molecule to other web tools, developed by the Molecular Modeling Group of the SIB Swiss Institute of Bioinformatics. These interoperability icons, together with link to ChEMBL Compound Report Card, are also appearing in all boxes relating the chemical structures of known actives (known actives output page, refer to Figure 3). The latter are ranked from the most to the less similar to the query molecules. At the time of
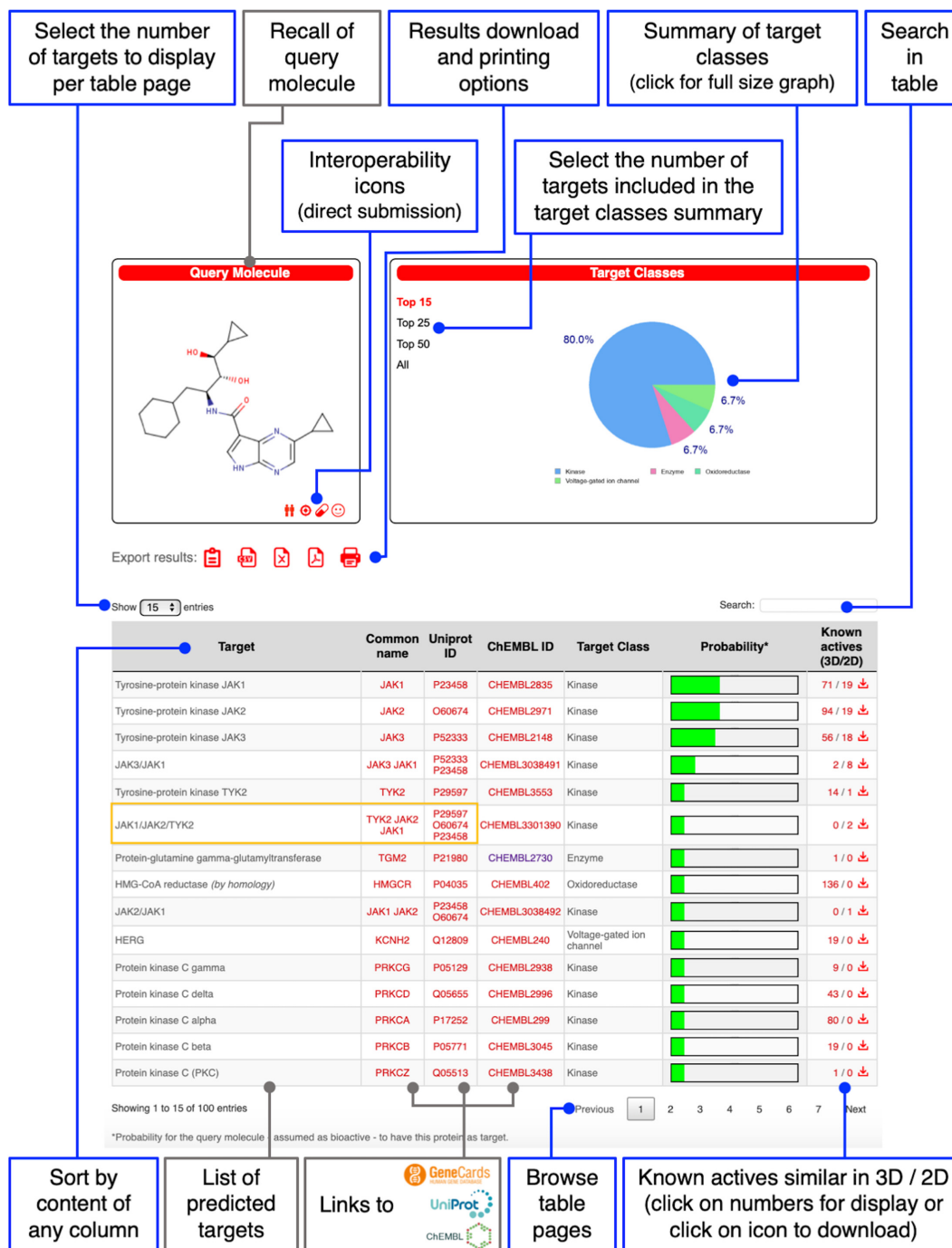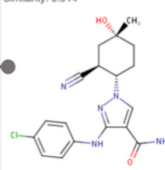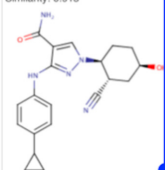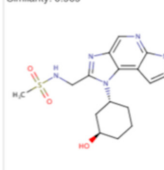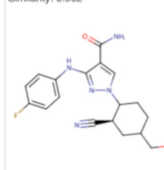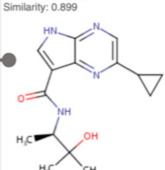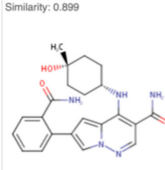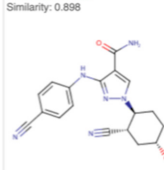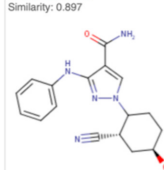
Figure 2. Prediction output page. Major novelties and updated items are boxed in blue. Upper-left: recall of the query molecule structure (here: a pyrrolopyrazine JAK inhibitor from patent US2014155376A1, not included in ChEMBL23) and interoperability icons to submit the compound to SwissADME, SwissSimilarity or SwissTargetPrediction for further analysis, or to get the query SMILES. Below, other icons allow downloading the results as PDF, CSV or Excel files, as well as printing or copying them to the clipboard. Upper-right: the summary of the predicted target classes is displayed as a pie chart, whose percentages are calculated using the top 15, 25, 50 or all predicted targets (maximum 100). The main table lists the predicted most probable targets for the query molecule. By default, proteins are ranked according to their probability of being the actual target (depicted has green bars), but the dynamic table can be sorted by any column by clicking on the corresponding header. By default, the number of displayed rows is 15, but the user can choose 25, 50 or all predicted targets (maximum 100). Also, a search box allows filtering the results. Links to GeneCards, UniProt and ChEMBL are provided. Protein complexes are now given in full name (as in ChEMBL, one example boxed in yellow); each subunit/component is linked individually to GeneCards and UniProt. The last column gives how many known actives on each listed target are highly similar to the query molecule based on either 2D or 3D similarity measures. These numbers are links to pages containing information about these similar molecules (refer to Figure 3). Clicking on the right-hand 'download' icon saves a ZIP archive on the user's computer including two files. One file contains all known actives on the relative target similar in 2D to the query molecule; the other file contains all known actives on the same target similar in 3D to the query molecule. Predictions obtained from molecules active on homologous (e.g. ortholog) proteins are tagged '(by homology)'.

Figure 3. Known actives output pages are tabulated lists of molecules experimentally defined as active on the protein of interest (here the tyrosine-protein kinase JAK1) that are highly similar to the query molecule (recalled on the top-left box, here: a pyrrolopyrazine JAK inhibitor from patent US2014155376A1, not included in ChEMBL23) according to 3D (panel **A**) or 2D (panel **B**) measures. These pages are obtained by clicking on the number of known actives in the previous prediction output page (see Figure 2). All molecules have similarity values to the query compound above the defined thresholds (0.65 Tanimoto index on FP2 for 2D, and 0.85 Manhattan-based measure on ES5D for 3D) and ranked from the most to less similar. A link to the ChEMBL compound report card is provided for each compound by clicking on the ChEMBL ID. The major novelty on the known actives page are the interoperability icons for straightforward submission of the known active compounds (or the query molecule) to the in-house webtools SwissSimilarity ('twins'), SwissTargetPrediction ('target'), SwissADME ('pill'). It is also possible to get the molecules SMILES. By clicking the 'Export table' icon, known actives can be downloaded as a CSV file containing ChEMBL ID, SMILES, similarity value to the query molecule (Tanimoto or Manhattan-based measures, for 2D or 3D respectively) and the source species of the homologous experimentally determined target (if this applies).

writing this manuscript, three computer-aided drug design tools are interoperable (both ways). The 'twins' icon submits the molecule to SwissSimilarity (21) for ligand-based direct screening, the 'target' for re-submitting the target prediction (on another species, for instance) and the 'pill' icon to estimate physicochemical, ADME or pharmacokinetic parameters by SwissADME (20). The planned extension of interoperability by including other in-house tools is an on-going work and therefore additional icons will be added in the different websites in the future. Also, the use of the molecule as potential input to other, possibly external, methods is facilitated by obtaining its SMILES through a click on the fourth icon. Finally, the list of known actives can be exported in CSV format by clicking on the right-hand icon, above the table, in the know actives output page (Figure 3). The file contains ChEMBL ID, SMILES, similarity value to the query molecule (i.e. Tanimoto index or Manhattan-based measure, for 2D or 3D respectively) and, in case, the source species of the experimentally-determined homologous target.

Other pages

Support for using website and analysing the predictions is provided by renewed FAQ and help pages accessible through the header menu. The latter contains a step-by-step description of input and output, with clickable screenshot linking to pre-filled submission form and pre-calculated results. Should users need more assistance, a contact form is available. For evaluation purpose, different datasets are given including the external test set employed for estimating the predictive power of the retrained model of SwissTargetPrediction 2019 (refer to the Validation section).

VALIDATION

The new SwissTargetPrediction model was validated by assessing its predictive capacity on an external test set of experimentally active compounds. 500 compounds meeting the following criteria were randomly picked up from ChEMBL24. They must (i) be annotated as direct binders, (ii) with an activity (K_i, K_D, IC_{50} or EC_{50}) < 1 nM, (iii) in an assay labelled with a confidence score >3, (iv) on at least two different human targets; (v) contain <80 heavy atoms; (vi) bind to single proteins or protein complexes (e.g. excluding targets corresponding to protein families) and (vii) not being recorded in ChEMBL23.

The 500 test molecules were submitted to the new reverse screening engine as implemented on the updated website to predict their most probable human protein targets. For 360 molecules (72%), at least one of the experimentally known targets can be found among the predicted top-15 (note that the table in the prediction output page has 15 rows by default).

The 2014 SwissTargetPrediction model, trained on ChEMBL16 data, was previously tested in a similar way on compounds taken from ChEMBL17. Likewise, the success rate at top-15 was 70% (12). However, interestingly, the success rate at top-15 dropped to 51% when applying the 2014 model on the recent 500-compound set from ChEMBL24, predicting correct targets for 253 molecules.

The massive growth of the bioactivity data increasingly included in the successive releases of ChEMBL has greatly extended both the biological and the chemical spaces (14,15) in regions that were covered only partially by the previous version of SwissTargetPrediction. In addition, the number of druggable proteins has increased. As extreme examples, the opioid growth factor receptor-like protein 1 and the CDK2/Cyclin A complex, which are known today to be targeted by three compounds of the ChEMBL23 test set, were absent from ChEMBL16 and therefore impossible to predict by the 2014 version of SwissTargetPrediction. Noticeably, all targets of the 500 compounds in the external test sets were possible to be predicted by the 2019 version of SwissTargetprediction (and similarly all targets of the previous ChEMBL17 test set were possible to be predicted by the 2014 version of SwissTargetPrediction). Together with the discovery of new bioactive chemotypes, this strongly supports the major update of the reverse screening engine described in the present article. As a result, the 2019 version showed remarkable robustness in predictive power, further demonstrated by the capacity to predict the correct target at rank 1 in 28% of the cases. Given that predictions were made among 2092 possible human targets, and that test compounds target 2.1 proteins on average, this success rate corresponds to a 280-fold enrichment compared to random picking.

CONCLUSION AND OUTLOOK

The 2019 version represents a major update of SwissTargetPrediction (www.swisstargetprediction.ch) with updated bioactivity data, a re-trained model, a more efficient back-end and important novelties in the web interface. Although the reverse screening engine yields predictions among more possible protein targets and is now based on molecular similarity with a larger collection of experimentally defined active compounds, the process is significantly faster (15–20 s for human proteins) and the high predictive performance level is maintained compared to the previous 2014 version (correct target listed in the top 15 for >70% of tested external compounds). Moreover, the user experience was enhanced for both input of molecules and analysis of the results. Noteworthy, the two-way interoperability capacity enables straightforward submission of any input or output molecule to other on-line tools, which are part of a core initiative of SIB Swiss Institute of Bioinformatics to provide free access to a collection of computer-aided drug design web-based methods: currently SwissTargetPrediction, SwissSimilarity (direct ligand-based screening) (21) and SwissADME (for physicochemical and pharmacokinetic parameters estimation) (20). Apart from regular updates, other in-house tools (such as SwissDock (16), SwissParam (17) or SwissBioisostere (19)) will be further connected in order to propose, ultimately, a fully integrated web-based drug design environment.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Byrne,R. and Schneider,G. (2019) In silico target prediction for small molecules. *Methods Mol. Biol.*, **1888**, 273–309.
2. Chaudhari,R., Tan,Z., Huang,B. and Zhang,S. (2017) Computational polypharmacology: a new paradigm for drug discovery. *Expert Opin. Drug Discov.*, **12**, 279–291.
3. Lavecchia,A. and Cerchia,C. (2016) In silico methods to address polypharmacology: current status, applications and future perspectives. *Drug Discov. Today*, **21**, 288–298.
4. Cereto-Massagué,A., Ojeda,M.J., Valls,C., Mulero,M., Pujadas,G. and Garcia-Vallvé,S. (2015) Tools for in silico target fishing. *Methods*, **71**, 98–103.
5. Ding,H., Takigawa,I., Mamitsuka,H. and Zhu,S. (2014) Similarity-based machine learning methods for predicting drug-target interactions: a brief review. *Brief. Bioinformatics*, **15**, 734–747.
6. Willett,P. and Winterman,V. (1986) A comparison of some measures for the determination of inter-molecular structural similarity measures of inter-molecular Structural Similarity. *Quant. Struct.-Act. Relat.*, **5**, 18–25.
7. Johnson,M., Lajiness,M. and Maggiora,G. (1989) Molecular similarity: a basis for designing drug screening programs. *Prog. Clin. Biol. Res.*, **291**, 167–171.
8. Gfeller,D., Michielin,O. and Zoete,V. (2013) Shaping the interaction landscape of bioactive molecules. *Bioinformatics*, **29**, 3073–3079.
9. O'Boyle,N.M., Banck,M., James,C.A., Morley,C., Vandermeersch,T. and Hutchison,G.R. (2011) OpenBabel: An open chemical toolbox. *J. Cheminform.*, **3**, 33.
10. Armstrong,M.S., Morris,G.M., Finn,P.W., Sharma,R., Moretti,L., Cooper,R.I. and Richards,W.G. (2010) ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. *J. Comput. Aided Mol. Des.*, **24**, 789–801.
11. Armstrong,M.S., Finn,P.W., Morris,G.M. and Richards,W.G. (2011) Improving the accuracy of ultrafast ligand-based screening: incorporating lipophilicity into ElectroShape as an extra dimension. *J. Comput. Aided Mol. Des.*, **25**, 785–790.
12. Gfeller,D., Grosdidier,A., Wirth,M., Daina,A., Michielin,O. and Zoete,V. (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res.*, **42**, W32–W38.
13. Gfeller,D. and Zoete,V. (2015) Protein homology reveals new targets for bioactive small molecules. *Bioinformatics*, **31**, 2721–2727.
14. Gaulton,A., Hersey,A., Nowotka,M., Bento,A.P., Chambers,J., Mendez,D., Mutowo,P., Atkinson,F., Bellis,L.J., Cibrián-Uhalte,E. *et al.* (2017) The ChEMBL database in 2017. *Nucleic Acids Res.*, **45**, D945–D954.
15. Mendez,D., Gaulton,A., Bento,A.P., Chambers,J., De Veij,M., Félix,E., Magariños,M.P., Mosquera,J.F., Mutowo,P., Nowotka,M. *et al.* (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.*, **47**, D930–D940.
16. Grosdidier,A., Zoete,V. and Michielin,O. (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.*, **39**, W270–W277.
17. Zoete,V., Cuendet,M.A., Grosdidier,A. and Michielin,O. (2011) SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.*, **32**, 2359–2368.
18. Gfeller,D., Michielin,O. and Zoete,V. (2013) SwissSidechain: a molecular and structural database of non-natural sidechains. *Nucleic Acids Res.*, **41**, D327–D332.
19. Wirth,M., Zoete,V., Michielin,O. and Sauer,W.H.B. (2013) SwissBioisostere: a database of molecular replacements for ligand design. *Nucleic Acids Res.*, **41**, D1137–D1143.
20. Daina,A., Michielin,O. and Zoete,V. (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.*, **7**, 42717.
21. Zoete,V., Daina,A., Bovigny,C. and Michielin,O. (2016) SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening. *J. Chem. Inf. Comput. Sci.*, **56**, 1399–1404.