Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents

Ying Zhou1,†, Yintao Zhang2,†, Xichen Lian2, Fengcheng Li2, Chaoxin Wang3, Feng Zhu2,4,*, Yunqing Qiu1,* and Yuzong Chen5,6,*

1State Key Laboratory for Diagnosis and Treatment of Infectious Disease, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang Provincial Key Laboratory for Drug Clinical Research and Evaluation, The First Affiliated Hospital, Zhejiang University, 79 QingChun Road, Hangzhou, Zhejiang 310000, China, 2College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China, 3Department of Computer Science, Kansas State University, Manhattan 66506, USA, 4Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 330110, China, 5State Key Laboratory of Chemical Oncogenomics, Key Laboratory of Chemical Biology, The Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China and 6Qian Xuesen Collaborative Research Center of Astrochemistry and Space Life Sciences, Institute of Drug Discovery Technology, Ningbo University, Ningbo 315211, China

Received September 09, 2021; Revised September 29, 2021; Editorial Decision September 30, 2021; Accepted October 04, 2021

ABSTRACT
Drug discovery relies on the knowledge of not only drugs and targets, but also the comparative agents and targets. These include poor binders and non-binders for developing discovery tools, prodrugs for improved therapeutics, co-targets of therapeutic targets for multi-target strategies and off-target investigations, and the collective structure-activity and drug-likeness landscapes of enhanced drug feature. However, such valuable data are inadequately covered by the available databases. In this study, a major update of the Therapeutic Target Database, previously featured in NAR, was therefore introduced. This update includes (a) 34 861 poor binders and 12 683 non-binders of 1308 targets; (b) 534 prodrug-drug pairs for 121 targets; (c) 1127 co-targets of 672 targets regulated by 642 approved and 624 clinical trial drugs; (d) the collective structure-activity landscapes of 427 262 active agents of 1565 targets; (e) the profiles of drug-like properties of 33 598 agents of 1102 targets. Moreover, a variety of additional data and function are provided, which include the cross-links to the target structure in PDB and AlphaFold, 159 and 1658 newly emerged targets and drugs, and the advanced search function for multi-entry target sequences or drug structures. The database is accessible without login requirement at: https://idrblab.org/ttd/.

INTRODUCTION
Drug discovery is promoted by not only the knowledge of drugs (1) and their therapeutic targets (2–4), but
also the comparative data with respect to other bioactive agents and other targets. Such comparative data include the knowledge of poor binders or non-binders of individual target that are useful for developing drug discovery tool of enhanced performance (5–7); the information of prodrugs that facilitates drug design by improving pharmacokinetic/pharmacodynamic features (8); the co-targets of therapeutic targets that facilitate the investigations of multi-target strategies (9), off-target (10,11) & undesired effect (9); the collective structure-activity landscapes of drugs against individual target that reveal important pharmaceutical features such as activity cliffs (12); and the drugs’ profiles of their drug-like properties that provide drug-likeness landscapes of the explored bioactive chemical space for therapeutic targets (13). Particularly, there is a rapid trend of the discovery of Artificial Intelligence (AI) tools for the drug discovery (14,15), including the AI tools for identifying bioactive compounds, and the construction of such tools requires data of poor binders and non-binders of a specific target (16). In the meantime, the existing prodrug data may inspire new ideas to avoid the drug development challenges that limit formulation option or result in undesired biopharmaceutical/pharmacokinetic performance (8). Thus, such comparative data above are urgently needed by researchers in drug discovery community. Moreover, the data of target’s 3D structure are the key information for drug discovery (5). Apart from the increasing number of experimentally-resolved target crystal structures (17), advanced AI technologies (e.g. AlphaFold) have enabled the prediction of target’s crystal structures of high-confidence (18,19), which requires the target-related databases, especially TTD, to include such valuable data.

While the established databases provide the comprehensive information of both drugs and targets (20–24), there is an inadequate coverage of the comparative data for the targeted agents and high-confidence 3D structures of human targets. To provide such valuable data, several major updates of Therapeutic Target Database (https://idrblab.org/ ttd/) were thus introduced in this study. The first was the inclusion of > 34,800 poor binders (the target activity within the range of 50–200 μM) and > 12,600 non-binders (target activity > 200 μM) for 383 and 309 successful targets (STs), 392 and 275 clinical trial targets (CTs), 137 and 91 preclinical or patented targets (PTs), and 331 and 195 research targets (RTs) respectively. Second, we added > 500 prodrug-drug pairs for 91 STs, 30 CTs. Third, we provided > 1,100 co-targets of 423 STs and 249 CTs. These STs and CTs are targeted by 642 approved, and 624 clinical trial drugs, respectively. Fourth, we provided the 2D collective structure-activity landscapes (containing > 427,200 bioactive agents) for 444 STs, 469 CTs, 163 PTs and 489 RTs. Fifth, the drugs’ profiles of drug-like property of > 33,500 agents of 435 STs, 356 CTs, 125 PTs and 186 RTs were also shown. Meanwhile, additional structural data were updated, which included the cross-links to 930 experimentally-resolved PDB structures and 1,824 AlphaFold-generated structures; and 159 and 1,658 newly emerged targets and drugs were also collected. Table 1 gave the statistics of targets and drugs among different database versions, and Table 2 summarized the new features and their corresponding statistics updated to the latest database. Moreover, the schema, search engine, and adopted ontology of this database were also provided in the TTD website.

**POOR BINDERS AND NON-BINDERS OF THERAPEUTIC TARGETS**

Molecular docking is a widely-used structure-based drug discovery method (17), which employs scoring functions for scoring the binding of molecules to a target site (25). Poor binders and non-binders are useful decoy molecules for the development of the scoring functions (6). AI methods have also been extensively explored to develop bioactive molecule and pharmaceutical property screening tools, which have been primarily trained by actives (e.g. binders) and non-actives (e.g. poor binders, non-binders) (26–28). Particularly, the molecules of <10 μM activity were typically considered as inhibitors or actives (29), while those of 50–200 μM activity were reported as poor inhibitors (30,31). Meanwhile, the molecules of >200 μM activity were regarded to be inactive of little effect (32,33). In other words, it is essential to have a conveniently-accessible resource for poor binders and non-binders of the therapeutic targets. Thus, the molecules with experimentally measured activities against each TTD target were first collected by reviewing PubMed literatures (34) using keyword combinations between target names/synonyms and ‘inhibitor’, ‘antagonist’, ‘agonist’, ‘activity’, ‘binding’, ‘affinity’, ‘IC50’, ‘Ki’, etc. Second, these PubMed literatures were manually checked to discover those containing the molecule with experimentally measured quantitative activity against any target of interest. Third, based on these collected activity values, the poor binders and non-binders were tentatively defined as of 50–200 μM (30,31) and >200 μM (32,33) activity, respectively. Using the above criteria, a total of 34,861 poor binders and 12,683 non-binders were collected for 393 and 309 STs, 392 and 275 CTs, 137 and 91 PTs, 331 and 195 RTs, respectively.

**PRODRUGS**

Good therapeutic drugs possess not only potent activities but also desirable pharmacokinetic and toxicological properties (35). In some cases, the drug leads may possess potent activity but poor pharmacokinetic property, which could be overcome using the prodrug strategy (8). Prodrugs are molecules modified from the parent drugs, with little or no activity but also desirable pharmacokinetic and toxicological properties (e.g. poor binders, non-binders) (26–28). Particularly, the molecules of <10 μM activity were typically considered as inhibitors or actives (29), while those of 50–200 μM activity were reported as poor inhibitors (30,31). Meanwhile, the molecules of >200 μM activity were regarded to be inactive of little effect (32,33). In other words, it is essential to have a conveniently-accessible resource for poor binders and non-binders of the therapeutic targets. Thus, the molecules with experimentally measured activities against each TTD target were first collected by reviewing PubMed literatures (34) using keyword combinations between target names/synonyms and ‘inhibitor’, ‘antagonist’, ‘agonist’, ‘activity’, ‘binding’, ‘affinity’, ‘IC50’, ‘Ki’, etc. Second, these PubMed literatures were manually checked to discover those containing the molecule with experimentally measured quantitative activity against any target of interest. Third, based on these collected activity values, the poor binders and non-binders were tentatively defined as of 50–200 μM (30,31) and >200 μM (32,33) activity, respectively. Using the above criteria, a total of 34,861 poor binders and 12,683 non-binders were collected for 393 and 309 STs, 392 and 275 CTs, 137 and 91 PTs, 331 and 195 RTs, respectively.
Table 1. Accumulation of drugs and their corresponding targets in the latest and previous versions of TTD database

| TTD statistics for targets and drugs |  |  |  |  |  |  |
|--------------------------------------|---|---|---|---|---|---|
| All targets                          | 2022 | 2020 | 2018 | 2016 | 2014 | 2012 | 2010 |
| Successful targets                   | 3578 | 3419 | 3101 | 2589 | 2360 | 2025 | 1894 |
| Clinical targets                     | 498  | 461  | 445  | 397  | 388  | 364  | 348  |
| Preclinical/patented targets         | 1542 | 1191 | 1121 | 723  | 461  | 286  | 292  |
| Research targets                     | 185  | 155  | 0    | 0    | 0    | 0    | 0    |
| All drugs                            | 38760| 37102| 34019| 31614| 20667| 17816| 5028 |
| Approved drugs                       | 2797 | 2649 | 2544 | 2071 | 2003 | 1540 | 1514 |
| Clinical trial drugs                 | 10831| 9465 | 8103 | 7291 | 3147 | 1423 | 1212 |
| Preclinical/patented drugs           | 5009 | 4845 | 18923| 17803| 14856| 14853| 2302 |

Table 2. New features and their corresponding statistics added to the 2022 TTD. These new features included structure-based activity landscape of targets, profile of drug-like properties of studied targets, prodrugs together with their parent drug and target, co-targets modulated by approved or clinical trial drugs, and the poor binders and non-binders of targets

| Structure-based activity landscape of studied targets
| No. of targets with chemical structure based activity landscape | No. of drug structures |
|---------------------------------------------------------------|------------------------|
| Successful Clinical trial                                     | 444                    |
| Preclinical/patented Research                                 | 163                    |
| Research                                                      | 489                    |
| 427 262                                                       |

| Drug-like properties of studied targets
| No. of targets with drug property profile                     | No. of drugs |
|---------------------------------------------------------------|--------------|
| Successful Clinical trial                                     | 435          |
| Preclinical/patented Research                                 | 125          |
| Research                                                      | 186          |
| 33 598                                                       |

| Prodrugs together with their parent drug and target
| No. of prodrugs                                               | No. of targets for prodrugs |
|---------------------------------------------------------------|----------------------------|
| Approved Clinical trial                                      | 91                        |
| Preclinical/patented Experimental                             | 30                        |
| Research                                                     | 1                         |

| Co-targets modulated by approved/clinical trial drugs
| No. of targets with co-targets                                | No. of drugs modulating co-targets |
|---------------------------------------------------------------|-----------------------------------|
| Successful Clinical trial                                     | 423                  |
| Preclinical/patented Research                                 | 249                  |
| Approved Clinical trial                                       | 642                  |
| Clinical trial                                               | 624                  |
| 1127                                                          |

| Poor binders and non-binders of studied targets
| No. of targets with poor binder(s)                            | No. of poor binders interacting with TTD targets |
|---------------------------------------------------------------|-----------------------------------------------|
| Successful Clinical trial                                     | 383                        |
| Preclinical/patented Research                                 | 137                        |
| Research                                                      | 331                        |
| 34 861                                                        |
| No. of targets with non-binder(s)                             | 309                        |
| Clinical trial                                               | 275                        |
| Preclinical/patented Research                                 | 91                         |
| Research                                                      | 195                        |
| 12 683                                                        |

the structures of the prodrug and its parent drug were drawn using ChemDraw based on the structures reported in each corresponding literature. As shown in Figure 1, both the detailed data and structures of prodrugs were explicitly described in the TTD prodrug page. All in all, a total of 534 prodrug-drug pairs of 91 STs and 30 CTs were collected to this update of TTD.

CO-TARGETS OF THERAPEUTIC TARGETS

Many drugs are known to interact with more macromolecular targets than their intended primary therapeutic target. In particular, a multi-target drug produces its therapeutic effect by modulating multiple targets (9). Some clinical trial drugs have been found to produce their therapeutic effects via interacting with off-targets, i.e., a macromolecular target other than their originally intended primary target (10). On the one hand, such beneficial effects of off-target have been explored for drug repurposing against complex diseases (36–39); on the other hand, off-target activity may in some instances lead to undesirable effect (40). Based on multiple targets of drugs, one can define the co-targets of a therapeutic target as the additional targets of all drugs targeting the therapeutic target. In other words, these co-targets represent both the targets co-modulated by a multi-target drug (5) and the off-target of a drug (11). Thus, those co-targets of a therapeutic target were first collected by reviewing PubMed literatures (34) by combining the target name with the keywords 'multi-target', 'off-target', 'multiple targets', 'poly-pharmacology', 'co-targets', 'co-targeting', etc. Second, all these literatures were manually checked to discover those having the information of co-targets, and the
drugs of clinical importance (approved or clinical trials) that co-regulating a therapeutic target and its co-targets were also identified from literatures, company reports, and other official resources providing drug-target information. Third, detailed data of each co-target were collected to TTD and cross-linked to other reputable databases (e.g. UniProt (41) and NCBI Gene (34)). As a result, 1127 co-targets of 423 STs and 249 CTs co-modulated by 642 approved and 624 clinical trial drugs were identified and collected for this update.

### COLLECTIVE STRUCTURE-ACTIVITY LANDSCAPES OF INDIVIDUAL TARGET

In the design of drugs against individual target, the molecular structure of the hit against a target (first molecule found to bind to the target) should be modified to optimize target binding activity (42,43). Those modified molecules, particularly the structural derivatives of a hit, largely follow certain structure-activity relationship (44), and can also lead to the dramatical activity variations, namely activity cliff
Such structure-activity relationships can be further evaluated by the collective structure-activity landscape of all known binders of studied targets. As described in Figure 2, all known binders of a target were clustered based on their structural similarities, each binder was represented by a colored bar with its height proportional to the level of target binding activity (−log IC50, −log Ki, etc.) and color indicating each binder’s clinical status (orange, yellow, blue and grey denote approved, clinical, discontinued and investigational drugs, respectively). The clustering of all binders of target was constructed using the sequential steps as follows. First, the molecular fingerprints of all binders were computed using R package ChemmineR (47). Second, the Tanimoto coefficient-based similarities among binders were computed by ChemmineR (47). Third, the complete linkage hierarchical clustering based on Euclidean distance (48) was adopted to cluster all target binders. Finally, a 2D graph was generated using the Data-Driven Documents (49), which was displayed on TTD webpage. In this update, the chemical structure-based activity landscapes of 444 STs, 469 CTs, 163 PTs and 186 RTs were provided. Figure 2 presents the 2D graph of such landscape for carbonic anhydrase VI (TTD Target ID: T06569).

Such collective structure-activity landscape of individual target is, to the best of our knowledge, unique in the following aspects. First, each landscape in TTD is dedicated to all drugs and other binders of individual therapeutic target. Such target-specific landscape provides the overview of the structural similarity among all target-specific binders, which could help the readers to gain a quick understanding of all available binding scaffolds of a studied target. Second, such landscape gives the activities of all drugs and binders for a target along with their structural characteristics, which is useful for describing QSARs and activity cliffs. Third, this provided landscape includes the valuable information of each drug’s clinical status, which demonstrated a unique perspective illustrating the relationships between drug structures and clinical development stages. Therefore, such collective structure-activity landscape of individual therapeutic target provided in TTD was of great merit for modern drug discovery.

COLLECTIVE PROFILES OF DRUG-LIKE PROPERTIES OF INDIVIDUAL TARGET

The potential of a bioactive molecule to become a drug is partly judged by the evaluation of its drug-like properties (13,50). The drug-likeness rules such as the Lipinski’s rule of five have been developed and widely used for evaluating the drug development potential of bioactive molecules (50–53). Such rules exploit drug’s distinguished physicochemical property, including molecular weight and the number of hydrogen bond donors, as the basis for drug-likeness evaluations (54). The value of these drug-like properties may vary from the drugs of one target to those of another. Therefore, target-specific profiles of drug-like property may be useful for facilitating the analysis of the landscape of drug-like property for targeted therapeutics (55). As illustrated in Figure 3, the 2D profiles of the target-specific drug-like properties for those targets in TTD were provided. Particularly, all known drugs of a target were clustered based on multiple (the top plot in Figure 3) or single (six plots at the bottom of Figure 3) drug-like properties, which was displayed using the hierarchical clustering map, heatmap and bar plot. The bar color indicates the highest clinical status of the corresponding drugs (approved, clinical trial, etc.). Users can move the mouse over the bar to find the basic information (status, PubChem CID, property, etc.) of specific drugs, and the detailed information of each drug can be also found by clicking that drug. Within each graph, the known drugs of a target were clustered according to their similarities in drug-like properties, which was constructed by a process similar to that described in previous section. Each drug was represented by a vertical line with the amplitude proportional to the values of drug-like property. All in all, the profiles of 6 drug-like properties (such as molecular weight, octanol/water partition coefficient, hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count & topological polar surface area) for 435 STs, 356 CTs, 125 PTs and 186 RTs were shown. Figure 3 presents the 2D profile of drug-like property for HIV integrase (TTD Target ID: T39087).

ENRICHED STRUCTURAL DATA AND ADVANCED SEARCH FUNCTION

The structures of macromolecules are important for drug discovery (56) and protein engineering or design (57). With the availability of target’s 3D structures, one can employ the structure-based drug discovery methods (such as molecular docking (56,58), 3D QSAR (59,60), structure-based pharmacophore (61) and molecular dynamics simulation (62)) to identify the binders of specific target (63). The number of experimental 3D structural entries of macromolecules have increased to >180 000 (17). These nonetheless only represent a minority of known protein sequences, with 35% proteins in human proteome having structure(s) in Protein Data Bank (18). Recent progress of AI technique like AlphaFold have enabled high-confidence prediction of protein 3D structures for most human proteins (18). AlphaFold employs a deep learning architecture to predict the 3D structure of a protein from its sequence (18). Thus, the AlphaFold-generated 3D structures could greatly expand the range of targets covered by structure-based drug discovery methods (64). To have a convenient access of the structures for each TTD target, the crosslinks to PDB (providing experimentally-resolved crystal structure) and AlphaFold (describing the predicted 3D structure) were reviewed and provided in TTD, which helped to link 2754 targets to their structure data.

Sequence similarity searching is the search of proteins with similar sequences to a known target, which is useful for identifying potential targets (65) and tracing protein evolution (66). It is based on the hypothesis that proteins of similar sequences have similar functions (67). Drug similarity searching is the search of small molecules with similar structures as that of a known drug, which is useful for finding molecules with similar activities or drug-like properties (68). TTD and other databases (41,69) have already provided target similarity and drug similarity searching facilities. Nonetheless, during practical applications, multiple proteins or chemical libraries are frequently searched and analyzed for the potential target and bioactive molecule. In other words, there is a need for the facilities that can support
Figure 2. A typical plot in TTD showing the chemical structure-based activity landscape for a target. All known drugs of a target are clustered based on their structural similarity. Moreover, the binding activity (e.g. $-\log IC_{50}$, $-\log Ki$) for each drug against the target is represented by bar chart. The color of the bar indicates the highest clinical status of the corresponding drug (approved, clinical trial, etc.). Users can move the mouse over the bar to get the basic information (status, PubChem CID, activity, etc.) of each drug. The detailed drug data can be found by clicking that particular drug.
Figure 3. A typical plot in TTD providing the information of the drug-like property-based profile for a target. All known drugs of a target are clustered based on multiple (the top plot) or single (six plots at the bottom) drug-like properties, which is displayed using the hierarchical clustering map, heatmap and bar plot. The bar color indicates the highest clinical status of the corresponding drugs (approved, clinical trial, etc.). The user can move the mouse over the bar to find the basic data (status, PubChem CID, property, etc.) of that drug. The detailed drug data can be found by clicking that drug.
multi-entry target and drug similarity searching. Therefore, a multi-entry target similarity searching and a multi-entry drug similarity searching facility was introduced, where the users can upload a file of multiple protein sequences or multiple molecular structures for finding TTD targets or drugs that are similar in sequence or structure. Particularly, the target similarity searching is based on the BLAST algorithm. Input with one protein sequence or a batch upload of multiple sequences for similarity search is now available in the latest version of TTD. The identified targets are ranked according to the BLAST outcomes. Moreover, the drug similarity searching is based on Tamimoto coefficients. The compound structure is first converted to PubChem Fingerprin by PaDEL-descriptors (70), and the similarity between input compound and TTD drugs was then calculated.

CONCLUDING REMARKS

With the rapid advances in modern drug discovery (71–75), there is an explosion of publications on revealing the mechanism underlying both disease and therapeutic (76–78), which in turn lead to the accumulation of huge amount of data for drug discovery. The expanded coverage of these data in TTD and other established databases collectively provide the enriched resources for drug discovery and the development of drug identification tool. The enriched data further enhance the ability to analyze and explore these derived data. Drug discovery efforts have benefited from this cycle of technology advancements, expanded knowledge and data, enhanced capabilities for the exploration of these derived data, and the advancements to the next round of the cycle. TTD and other established databases (79–81) will continue to update the new pharmaceutical data and play enhanced facilitating roles in current drug discovery efforts.

FUNDING

Scientific Research Grant of Ningbo University [215-432000282]; Ningbo Top Talent Project [215-432094250]; Zhejiang Provincial Science and Technology Department [2020C03046]; National Natural Science Foundation of China [81971982, 81872798, U1909208]; Natural Science Foundation of Zhejiang Province [LR21H300001]; Leading Talent of the ‘Ten Thousand Plan’ – National High-Level Talents Special Support Plan of China; Fundamental Research Fund for Central Universities [2018QNA7023]; ‘Double Top-Class’ University Project [181201*194232101]; Key R&D Program of Zhejiang Province [2020C03010]; Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare; Alibaba Cloud; Information Technology Center of Zhejiang University. Funding for open access charge: National Natural Science Foundation of China [81872798].

Conflict of interest statement. None declared.

REFERENCES

1. Shih, H.P., Zhang, X. and Aronov, A.M. (2018) Drug discovery effectiveness from the standpoint of therapeutic mechanisms and indications. Nat. Rev. Drug Discov., 17, 19–33.
2. Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I. et al. (2017) A comprehensive map of molecular drug targets. Nat. Rev. Drug Discov., 16, 19–34.

Downloaded from https://academic.oup.com/nar/article/50/D1/D1398/6413598 by National Science & Technology Library Root Admin user on 05 January 2022
DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res., 46, D1074–D1082.

21. Wang,Y., Zhang,S., Li,F., Zhou,Y., Zhang,Y., Wang,Z., Zhang,R., Zhu,J., Ren,Y., Tan,Y. et al. (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. Nucleic Acids Res., 48, D1031–D1041.

22. Avram,S., Bologa,C.G., Holmes,J., Rock,G., Wilson,T.B., Nguyen,D.T., Curpan,R., Halip,L., Bora,A., Yang,J.J. et al. (2021) DrugCentral 2021 supports drug discovery and repurposing. Nucleic Acids Res., 49, D1160–D1169.

23. Armstrong,J.F., Facenda,E., Harding,S.D., Pawson,A.J., Southan,C., Sharrman,J., Campo,B., Cavanagh,D.R., Alexander,S.P., Gelpi,J.L. et al. (2021) The IUPHAR/BPS Guide to Pharmacology in 2020: extending immunopharmacology content and introducing the UPHAR/MMV guide to malaria pharmacology. Nucleic Acids Res., 48, D1006–D1021.

24. Yang,Q., Wang,Y., Zhang,Y., Li,F., Xia,W., Zhou,Y., Qiu,Y., Li,H. and Zhu,F. (2020) NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. Nucleic Acids Res., 48, W436–W448.

25. Warren,G.L., Andrews,C.W., Capelli,A.M., Clarke,B., LaLonde,J., Bender,A. and Cortes-Ciriano,I. (2021) Artificial intelligence in drug discovery: the potential of deep learning for drug discovery and development. Curr. Opin. Drug Discov. Dev., 6, 68–80.

26. Han,L.Y., Ma,X.H., Lin,H.H., Jia,J., Zhu,F., Xue,Y., Li,Z.R., Xue,Y., Li,Z.R., Yap,C.W., Sun,L.Z., Chen,X. and Chen,Y.Z. (2004) Inhibition of glycinamide ribonucleotide formyltransferase and other folate enzymes by homofolate polyglutamates in human lymphoma and murine leukemia cell extracts. Cancer Res., 49, 158–163.

27. Bader,A. and Cortes-Ciriano,I. (2021) Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. Drug Discov. Today, 26, 1040–1052.

28. Thorndike,J., Gaumont,Y., Kisliuk,R.L., Sirotnak,F.M., Lambert,M.H., Lindvall,M., Nevins,N., Semus,S.F., Senger,S. et al. (2021) ChemmineR: a compound mining framework for R. J.Chem.Inf.Comput.Sci., 51, 1733–1734.

29. Seng, S., Gao, J., and Duan, X. (2021) Chemoinformatics: a tool for drug discovery. Curr. Top. Med.Chem., 21, 150–163.

30. Lipinski,C.A. (2004) Lead- and drug-like compounds: the rule-of-five requirements for glycogen synthase kinase 3 inhibitors. J. Med. Chem., 48, 7103–7112.

31. Fiscon,G., Conte,F., Farina,L. and Paci,P. (2021) SAveRUNNER: a software package for the prediction of protein-ligand complexes. Eur. J. Med.Chem., 207, 6177–6196.

32. Lipinski,C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. J. Med.Chem., 44, 1038–1040.

33. Lipinski,C.A. (2004) Lead- and drug-like compounds: the rule-of-five requirements for glycogen synthase kinase 3 inhibitors. J. Med. Chem., 48, 7103–7112.

34. Friesner,R.A., Murphy,R.B., Repasky,M.P., Frye,L.L., Halgren,E.A., Kuntz,B.I., Goetz,F.S., Atwood,J.W., Esber,D.T. et al. (2020) Glide: rationalization based on X-ray structures. J. Med.Chem., 49, 7979–7989.

35. Fiscon,G. and Paci,P. (2021) SAveRUNNER: an R-based tool for drug repurposing. BMC Bioinformatics, 22, 150.

36. Fiscon,G., Conte,F., Farina,L. and Paci,P. (2021) SAveRUNNER: a network-based algorithm for drug repurposing and its application to COVID-19. PLoS Comput. Biol., 17, e1008680.

37. Kumar,S., Jang,C., Subedi,L., Kim,S.Y. and Kim,M.H. (2020) Repurposing of FDA approved ring systems through bi-directional target-ring system dual screening. Sci. Rep., 10, 21133.

38. Tang,J., Fu,J., Wang,Y., Luo,Y., Yang,Q., Li,B., Tu,G., Hong,J., Cui,X., Chen,Y. et al. (2019) Simultaneous improvement in the precision, accuracy, and robustness of label-free proteome quantification by optimizing data manipulation chains. Mol. Cell. Proteomics, 18, 1683–1699.

39. Loukine,E., Keiser,M.J., Whitebread,S., Mikhailov,D., Hamon,J., Jenkins,J.L., Lavan,P., Weber,E., Doak,A.K., Cote,S. et al. (2012) Large-scale prediction and testing of drug activity on side-effect targets. Nucleic Acids Res., 40, D480–D489.

40. UniProt. (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res., 49, D480–D489.

41. Wu,T., Nagle,A., Kuhn,K., Garaging,K., Borboa,R., Francek,C., Chen,Z., Plouffe,D., Goh,A., Akamshinarianaya,S.B. et al. (2011) Imidazolopiperazines: hit to lead optimization of new antimalarial agents. J. Med. Chem., 54, 5116–5130.

42. Teli,M.K., Kumar,S., Yadav,D.K. and Kim,M.H. (2021) In silico identification of prolyl hydroxylase inhibitor by per-residue energy decomposition-based pharmacophore approach. J. Cell. Biochem., 122, 1098–1112.

43. Martinez,A., Alonso,M., Castello,A., Dorransoro,L., Gelpi,J.L., Luque,F.J., Perez,C. and Moreno,F.J. (2005) SAR and 3D-QSAR studies on thiazolidinodione derivatives: exploration of structural requirements for glycogen synthase kinase 3 inhibitors. J. Med. Chem., 48, 1167–1174.

44. Thorndike,J., Gaumont,Y., Kisliuk,R.L., Sirotnak,F.M., Lambert,M.H., Lindvall,M., Nevins,N., Semus,S.F., Senger,S. et al. (2021) ChemmineR: a compound mining framework for R. Bioinformatics, 24, 1733–1734.

45. Leeson,P.D. and Sprangthorpe,B. (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. Nat. Rev. Drug Discov., 6, 881–890.

46. Lipinski,C.A., Lombardo,F., Dominy,B.W. and Feeney,P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev., 46, 3–26.

47. Lipinski,C.A. (2004) Lead- and drug-like compounds: the rules-of-five revolution. Drug Discov. Today Technol., 1, 337–341.

48. Leeson,P.D., Bento,A.P., Gaulton,A., Hersey,A., Manners,E.J., Radoux,C.J. and Leach,A.R. (2021) Target-based evaluation of drug-like properties and ligand efficiencies. J. Med. Chem., 64, 7310–7320.

49. Gorgulla,C., Boesezemmyani,A., Wang,Z.F., Fischer,P.D., Coote,P.W., Padmanabha Das,K.M., Malets,Y.S., Radchenko,D.S., Moroz,Y.S., Scott,D.A. et al. (2020) An open-source drug discovery platform enables ultra-large virtual screens. Nature, 580, 663–668.

50. Taulaje,R., Venkat,A., Huang,L.C., Zhou,Z., Yeung,W., Rasheed,K.M., Li,S., Edison,A.S., Moremen,K.W. and Kannan,N. (2020) Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. Elife, 9, e54532.

51. Friesner,R.A., Murphy,R.B., Repasky,M.P., Frye,L.L., Greenwood,J.R., Halgren,T.A., Sanschagrin,P.C. and Mainz,D.T. (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enforcement for protein-ligand complexes. J. Med.Chem., 49, 6177–6196.

52. Verma,J., Khedkar,V.M. and Coutinho,E.C. (2010) 3D-QSAR in drug design—a review. Curr. Top. Med. Chem., 10, 95–115.
