Published online 20 November 2009 *Nucleic Acids Research, 2010, Vol. 38, Database issue* ***D787–D791***
doi:10.1093/nar/gkp1014

Update of TTD: Therapeutic Target Database

Feng Zhu[1], BuCong Han[1,2], Pankaj Kumar[1], XiangHui Liu[1], XiaoHua Ma[1], XiaoNa Wei[1,2], Lu Huang[1,2], YangFan Guo[1], LianYi Han[1], ChanJuan Zheng[1] and YuZong Chen[1,2,*]

[1]Bioinformatics and Drug Design Group, Center for Computational Science and Engineering, Department of Pharmacy and [2]Computation and Systems Biology, Singapore-MIT Alliance, National University of Singapore, Singapore, 117543

Received August 18, 2009; Revised October 16, 2009; Accepted October 19, 2009

ABSTRACT

Increasing numbers of proteins, nucleic acids and other molecular entities have been explored as therapeutic targets, hundreds of which are targets of approved and clinical trial drugs. Knowledge of these targets and corresponding drugs, particularly those in clinical uses and trials, is highly useful for facilitating drug discovery. Therapeutic Target Database (TTD) has been developed to provide information about therapeutic targets and corresponding drugs. In order to accommodate increasing demand for comprehensive knowledge about the primary targets of the approved, clinical trial and experimental drugs, numerous improvements and updates have been made to TTD. These updates include information about 348 successful, 292 clinical trial and 1254 research targets, 1514 approved, 1212 clinical trial and 2302 experimental drugs linked to their primary targets (3382 small molecule and 649 antisense drugs with available structure and sequence), new ways to access data by drug mode of action, recursive search of related targets or drugs, similarity target and drug searching, customized and whole data download, standardized target ID, and significant increase of data (1894 targets, 560 diseases and 5028 drugs compared with the 433 targets, 125 diseases and 809 drugs in the original release described in previous paper). This database can be accessed at http://bidd.nus.edu.sg/group/cjttd/TTD.asp.

INTRODUCTION

Pharmaceutical agents generally exert their therapeutic effects by binding to and subsequently modulating the activity of a particular protein, nucleic acid or other molecular (such as membrane) target (1,2). Target discovery efforts have led to the discovery of hundreds of successful targets (targeted by at least one approved drug) and >1000 research targets (targeted by experimental drugs only) (3–6). Rapid advances in genomic, proteomic, structural, functional and systems studies of the known targets and other disease proteins (7–13) enable the discovery of drugs, multi-target agents, combination therapies and new targets (3,5,7,14,15), analysis of on-target toxicity (16) and pharmacogenetic responses (17) and development of discovery tools (18–21).

To facilitate the access of information about therapeutic targets, publicly accessible databases such as Drugbank (22), Potential Drug Target Database (PDTD) (23) and our own Therapeutic Target Database (TTD) (24) have been developed. These databases complement each other to provide target and drug profiles. DrugBank is an excellent source for comprehensive drug data with information about drug actions and multiple targets (22). PDTD contains active-sites as well as functional information for potential targets with available 3D structures (23). TTD provides information about the primary targets of approved and experimental drugs (24).

While drugs typically modulate the activities of multiple proteins (25) and up to 14 000 drug-targeted-proteins have been reported (26), the reported number of primary targets directly related to the therapeutic actions of approved drugs is limited to 324 (6). Information about the primary targets of more comprehensive sets of approved, clinical trial and experimental drugs is highly useful for facilitating focused investigations and discovery efforts against the most relevant and proven targets (5,7,14,16,17,20). Therefore, we updated TTD by significantly expanding the target data to include 348 successful, 292 clinical trial and 1254 research targets, and added drug data for 1514 approved, 1212 clinical trial and 2302 experimental drugs linked to their primary targets (3382 small molecule and 649 antisense drugs with available structure and sequence, more structures will be added).

We collected a slightly higher number of successful targets than the reported number of 320 targets (6)

*To whom correspondence should be addressed. Tel: +65 6516 6877; Fax: +65 6774 6756; Email: csccyz@nus.edu.sg

because of the identification of protein subtypes as the targets of some approved drugs and the inclusion of multiple targets of approved multi-target drugs and non-protein/nucleic acid targets of anti-infectious drugs (e.g. bacterial cell wall and membrane components). Clinical trial drugs are based on reports since 2005 with the majority since 2008. Clinical trial phase is specified for every clinical trial drug. We also added new features for data access by drug mode of action, recursive search of related target and drug entries, similarity search of targets and drugs, customized and whole data download, and standardized target ID.

TARGET AND DRUG DATA COLLECTION AND ACCESS

Additional data about the approved, clinical trial and experimental drugs and their primary targets were collected from a comprehensive search of literatures, FDA Drugs@FDA webpage (http://www.accessdata.fda.gov) with information about FDA approved drugs, latest reports from 17 pharmaceutical companies that describe clinical trial and other pipeline drugs (Astrazeneca, Bayer, Boehringer Ingelheim, Genentech, GSK, Idenix, Incyte, ISIS, Merck, Novartis, Pfizer, Roche, Sanofi Aventis, Schering-Plough, Spectrum, Takeda, Teva). Literature search was conducted by searching Pubmed database using keyword combinations of 'therapeutic' and 'target', 'drug' and 'target', 'clinical trial' and 'drug', and 'clinical trial' and 'target', and by comprehensive search of such review journals as *Nature Reviews Drug Discovery*, *Trends of Pharmaceutical Science* and *Drug Discovery Today*. In particular, these searches identified 198 recent papers reporting approved and clinical trial drugs and their targets. As many of the experimental antisense drugs are described in US patents, we specifically searched US patent databases to identify 745 antisense drugs targeting 104 targets. Primary targets of 211 drugs and drug binding modes of 79 drugs are not specified in our collected documents. Further literature search was conducted to find the relevant information for these drugs. The criteria for identifying the primary target of a drug or targets of a multi-target drug is based on the developer or literature reported cell-based or *in vivo* evidence that links the target to the therapeutic effect of the drug. These searched documents are listed in the respective target or drug entry page of TTD and crosslink is provided to the respective PubMed abstract, US patent or developer web-page.

TTD data can be accessed by keyword or customized search. Customized search (Figure 1) fields include target name, drug name, disease indication, target biochemical class, target species, drug therapeutic class and drug mode of action. Further information about each target can be accessed via crosslink to UniProtKB\SwissProt, PDB, KEGG, OMID and Brenda database. Further drug information can be accessed via crosslink to PubChem, DrugBank, SuperDrug and ChEBI. Related target or drug entries can be recursively searched by clicking a disease or drug name. Similarity targets of an input protein sequence in FASTA format can be searched by using the BLAST sequence alignment tool (27). Similarity drugs of an input drug structure can be searched by using molecular descriptor based Tanimoto similarity searching method (28,29). Target and drug entries are assigned standardized TTD IDs for easy identification, analysis and linkage to other related databases. The whole TTD data, target sequences along with Swissprot and Entrez gene IDs, and drug structures can be downloaded via the download link. A separate downloadable file contains the list of TTD drug ID, drug name and the corresponding IDs in other cross-matching databases PubChem, DrugBank, SuperDrug and ChEBI. The corresponding HGNC name and Swissprot and Entrez gene ID of each target is provided in the target page. The SMILES and InCHI of each drug is provided in the drug page.

TARGET AND DRUG SIMILARITY SEARCHING

Target similarity searching (Figure 2) is based on the BLAST (27) algorithm to determine the similarity level between the sequence of an input protein and the sequence of each of the TTD target entries. The BLAST program was downloaded from NCBI website (http://www.ncbi.nlm.nih.gov/BLAST/download.shtml). The similarity targets are ranked by *E*-value and BLAST score (27). *E*-value has been reported to give reliable predictions of the homologous relationships (30) and *E*-value cutoff of 0.001 can be used to find 16% more structural relationships in the SCOP database than when using a standard sequence similarity with a 40% sequence-identity threshold (31). The majority of protein pairs that share 40–50% (or higher) sequence-identity differ by <1 Å RMS deviation (32,33), and a larger structural deviation probably alters drug-binding properties.

Drug similarity searching (Figure 3) is based on the Tanimoto similarity searching method (28). An input compound structure in MOL or SDF format is converted into a vector composed of molecular descriptors by using our MODEL software (34). Molecular descriptors are quantitative representations of structural and physicochemical features of molecules, which have been extensively used in deriving structure–activity relationships, quantitative structure–activity relationships and virtual screening tools for drug discovery (35,36). Based on the results of our earlier studies (29), a total of 98 1D and 2D descriptors were used as the components of the compound vector, which include 18 descriptors in the class of simple molecular properties, 3 descriptors in the class of chemical properties, 35 descriptors in the class of molecular connectivity and shape, and 42 descriptors in the class of electro-topological state. The vector of an input compound i is then compared with drug j in TTD by using the Tanimoto coefficient $sim(i,j)$ (28):

$$sim(i,j) = \frac{\sum_{d=1}^{l} x_{di}x_{dj}}{\sum_{d=1}^{l} (x_{di})^2 + \sum_{d=1}^{l} (x_{dj})^2 - \sum_{d=1}^{l} x_{di}x_{dj}}$$

Figure 1. Customized search page of TTD.

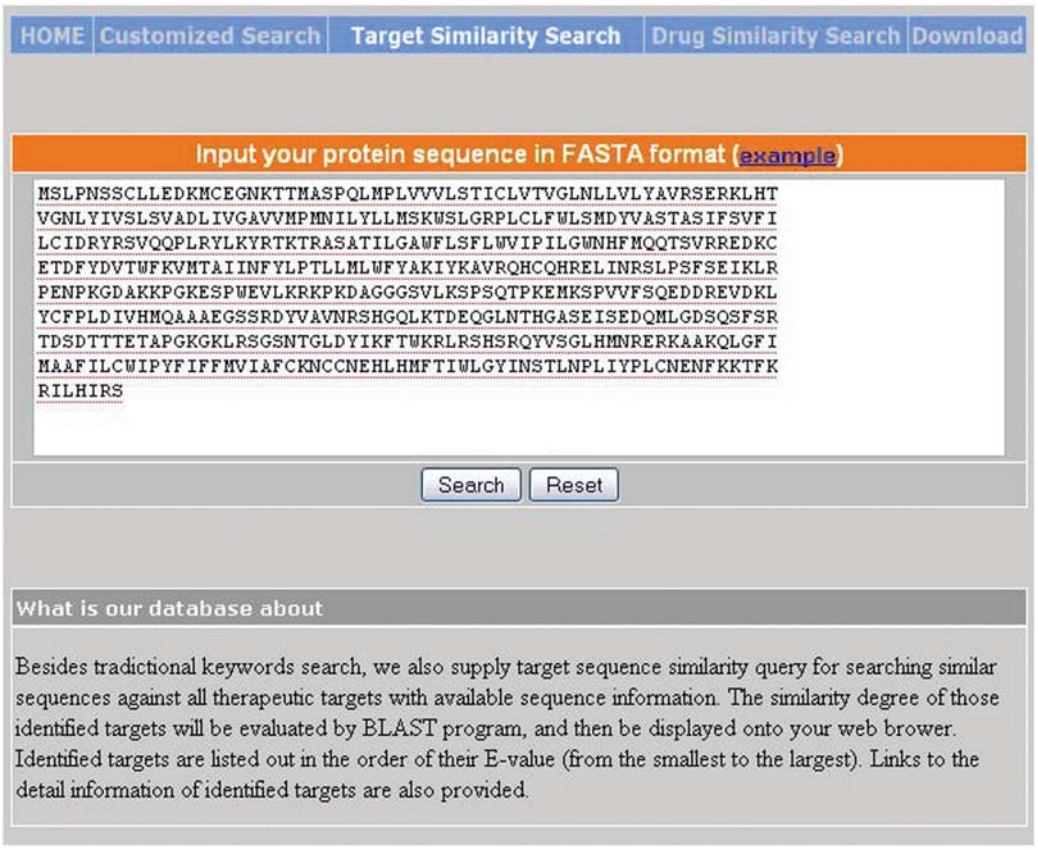

Figure 2. Target similarity search page of TTD.

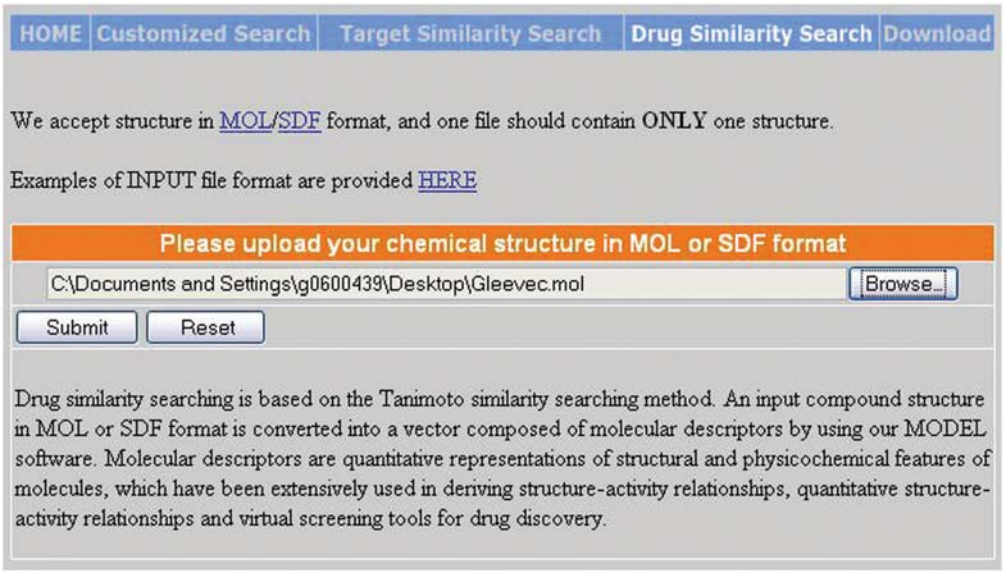

Figure 3. Drug similarity search page of TTDss.

where l is the number of molecular descriptors. Tanimoto coefficient of similarity compounds are typically in the range of 0.8–0.9 (37,38). Hence compound i is considered to be very similar, similar, moderately similar, or un-similar to drug j if $sim(i,j) > 0.9$, $0.85 < sim(i,j) < 0.9$, $0.75 < sim(i,j) < 0.85$, or $sim(i,j) < 0.75$, respectively.

REMARKS

The updated TTD is intended to be a more useful resource in complement to other related databases by providing comprehensive information about the primary targets and other drug data for the approved, clinical trial and experimental drugs. In addition to the continuous update of new target and drug information, efforts will be devoted to the incorporation of more features into TTD. Increasing amounts of data about the genomic, proteomic, structural, functional and systems profiles of therapeutic targets have been and are being generated (7–13). Apart from establishing crosslink to the emerging data sources, some of the profiles extracted or derived from the relevant data (3) may be further incorporated into TTD. Target data has been used for developing target discovery methods (18–20), some of these methods may be included in TTD in addition to the BLAST tool for similarity target searching. As in the case of PDTD (23), some of the virtual screening methods and datasets (35,36) may also be included in TTD for facilitating target oriented drug lead discovery.

FUNDING

Funding for open access charge: The Open Access charges for this article were partially waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Ohlstein,E.H., Ruffolo,R.R. Jr and Elliott,J.D. (2000) Drug discovery in the next millennium. *Annu. Rev. Pharmacol. Toxicol.*, **40**, 177–191.
2. Zambrowicz,B.P. and Sands,A.T. (2003) Knockouts model the 100 best-selling drugs–will they model the next 100? *Nat. Rev. Drug Discov.*, **2**, 38–51.
3. Zheng,C.J., Han,L.Y., Yap,C.W., Ji,Z.L., Cao,Z.W. and Chen,Y.Z. (2006) Therapeutic targets: progress of their exploration and investigation of their characteristics. *Pharmacol Rev.*, **58**, 259–279.
4. Golden,J.B. (2003) Prioritizing the human genome: knowledge management for drug discovery. *Curr. Opin. Drug Discov. Dev.*, **6**, 310–316.
5. Imming,P., Sinning,C. and Meyer,A. (2006) Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Discov.*, **5**, 821–834.
6. Overington,J.P., Al-Lazikani,B. and Hopkins,A.L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.*, **5**, 993–996.
7. Lindsay,M.A. (2003) Target discovery. *Nat. Rev. Drug Discov.*, **2**, 831–838.
8. Edwards,A. (2009) Large-scale structural biology of the human proteome. *Annu. Rev. Biochem.*, **78**, 541–568.
9. Lundstrom,K. (2006) Structural genomics: the ultimate approach for rational drug design. *Mol. Biotechnol.*, **34**, 205–212.
10. Kramer,R. and Cohen,D. (2004) Functional genomics to new drug targets. *Nat. Rev. Drug Discov.*, **3**, 965–972.
11. Dey,R., Khan,S. and Saha,B. (2007) A novel functional approach toward identifying definitive drug targets. *Curr. Med. Chem.*, **14**, 2380–2392.
12. Hopkins,A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.*, **4**, 682–690.
13. Giallourakis,C., Henson,C., Reich,M., Xie,X. and Mootha,V.K. (2005) Disease gene discovery through integrative genomics. *Annu. Rev. Genomics Hum. Genet.*, **6**, 381–406.
14. Zimmermann,G.R., Lehar,J. and Keith,C.T. (2007) Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov. Today*, **12**, 34–42.
15. Jia,J., Zhu,F., Ma,X., Cao,Z., Li,Y. and Chen,Y.Z. (2009) Mechanisms of drug combinations: interaction and network perspectives. *Nat. Rev. Drug Discov.*, **8**, 111–128.
16. Liebler,D.C. and Guengerich,F.P. (2005) Elucidating mechanisms of drug-induced toxicity. *Nat. Rev. Drug Discov.*, **4**, 410–420.
17. Eichelbaum,M., Ingelman-Sundberg,M. and Evans,W.E. (2006) Pharmacogenomics and individualized drug therapy. *Annu. Rev. Med.*, **57**, 119–137.

18. Barcellos,G.B., Pauli,I., Caceres,R.A., Timmers,L.F., Dias,R. and de Azevedo,W.F. Jr (2008) Molecular modeling as a tool for drug discovery. *Curr. Drug Targets*, **9**, 1084–1091.
19. Lee,G.M. and Craik,C.S. (2009) Trapping moving targets with small molecules. *Science*, **324**, 213–215.
20. Zhu,F., Han,L., Zheng,C., Xie,B., Tammi,M.T., Yang,S., Wei,Y. and Chen,Y. (2009) What are next generation innovative therapeutic targets? Clues from genetic, structural, physicochemical, and systems profiles of successful targets. *J. Pharmacol. Exp. Ther.*, **330**, 304–315.
21. Han,L.Y., Zheng,C.J., Xie,B., Jia,J., Ma,X.H., Zhu,F., Lin,H.H., Chen,X. and Chen,Y.Z. (2007) Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. *Drug Discov. Today*, **12**, 304–313.
22. Wishart,D.S., Knox,C., Guo,A.C., Cheng,D., Shrivastava,S., Tzur,D., Gautam,B. and Hassanali,M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.
23. Gao,Z., Li,H., Zhang,H., Liu,X., Kang,L., Luo,X., Zhu,W., Chen,K., Wang,X. and Jiang,H. (2008) PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics*, **9**, 104.
24. Chen,X., Ji,Z.L. and Chen,Y.Z. (2002) TTD: Therapeutic Target Database. *Nucleic Acids Res.*, **30**, 412–415.
25. Yildirim,M.A., Goh,K.I., Cusick,M.E., Barabasi,A.L. and Vidal,M. (2007) Drug-target network. *Nat. Biotechnol.*, **25**, 1119–1126.
26. Wishart,D.S., Knox,C., Guo,A.C., Shrivastava,S., Hassanali,M., Stothard,P., Chang,Z. and Woolsey,J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
27. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
28. Willett,P. (1998) Chemical similarity searching. *J. Chem. Inf. Comput. Sci.*, **38**, 983–996.
29. Ma,X.H., Wang,R., Yang,S.Y., Li,Z.R., Xue,Y., Wei,Y.C., Low,B.C. and Chen,Y.Z. (2008) Evaluation of virtual screening performance of support vector machines trained by sparsely distributed active compounds. *J. Chem. Inf. Model*, **48**, 1227–1237.
30. George,R.A. and Heringa,J. (2002) Protein domain identification and improved sequence similarity searching using PSI-BLAST. *Proteins*, **48**, 672–681.
31. Gerstein,M. (1998) Measurement of the effectiveness of transitive sequence comparison, through a third 'intermediate' sequence. *Bioinformatics*, **14**, 707–714.
32. Wood,T.C. and Pearson,W.R. (1999) Evolution of protein sequences and structures. *J. Mol. Biol.*, **291**, 977–995.
33. Koehl,P. and Levitt,M. (2002) Sequence variations within protein families are linearly related to structural variations. *J. Mol. Biol.*, **323**, 551–562.
34. Li,Z.R., Han,L.Y., Xue,Y., Yap,C.W., Li,H., Jiang,L. and Chen,Y.Z. (2007) MODEL-molecular descriptor lab: a web-based server for computing structural and physicochemical features of compounds. *Biotechnol. Bioeng.*, **97**, 389–396.
35. Yap,C.W., Li,H., Ji,Z.L. and Chen,Y.Z. (2007) Regression methods for developing QSAR and QSPR models to predict compounds of specific pharmacodynamic, pharmacokinetic and toxicological properties. *Mini Rev. Med. Chem.*, **7**, 1097–1107.
36. Li,H., Yap,C.W., Ung,C.Y., Xue,Y., Li,Z.R., Han,L.Y., Lin,H.H. and Chen,Y.Z. (2007) Machine learning approaches for predicting compounds that interact with therapeutic and ADMET related proteins. *J. Pharm. Sci.*, **96**, 2838–2860.
37. Bostrom,J., Hogner,A. and Schmitt,S. (2006) Do structurally similar ligands bind in a similar fashion? *J. Med. Chem.*, **49**, 6716–6725.
38. Huang,N., Shoichet,B.K. and Irwin,J.J. (2006) Benchmarking sets for molecular docking. *J. Med. Chem.*, **49**, 6789–6801.