Update of TTD: Therapeutic Target Database

Feng Zhu¹, BuCong Han¹,², Pankaj Kumar¹, XiangHui Liu¹, XiaoHua Ma¹, XiaoNa Wei¹,², Lu Huang¹,², YangFan Guo¹, LianYi Han¹, ChanJuan Zheng¹ and YuZong Chen¹,²,*

¹Bioinformatics and Drug Design Group, Center for Computational Science and Engineering, Department of Pharmacy and ²Computation 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)
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 col-
clected 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, Idexin, 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 informa-
tion 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).

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).

TARGET AND DRUG SIMILARITY SEARCHING

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):

\[
sim(i,j) = \frac{\sum_{d=1}^{l} x_{di}x_{dj}}{\sqrt{\sum_{d=1}^{l} (x_{di})^2} \times \sqrt{\sum_{d=1}^{l} (x_{dj})^2} - \sum_{d=1}^{l} x_{di}x_{dj}}
\]
Figure 1. Customized search page of TTD.

Figure 2. Target similarity search page of TTD.
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 \( \text{sim}(i,j) > 0.9 \), \( 0.85 < \text{sim}(i,j) < 0.9 \), \( 0.75 < \text{sim}(i,j) < 0.85 \), or \( \text{sim}(i,j) < 0.75 \), 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**


