DaiCee: A database for anti-cancer compounds with targets and side effect profiles

Manikkam Rajalakshmi[*]**, Sugumaran Suveena, Periyasamy Vijayalakshmi**[a]**, Sabapathy Indu, Anita Roy, Antony Ludas**

DBT-BIF Centre, PG & Research Department of Biotechnology & Bioinformatics, Holy Cross College (Autonomous) (Affiliated to Bharathidasan University), Trichy, Tamilnadu, India; Dr. Rajalakshmi Manikkam E-mail: mdraji@gmail.com; rajalakshmi@hcctrichy.ac.in; Fax: +91 431-2713312; Phone: +91 99524 98098; ***Corresponding author

Declaration on official E-mail:
The corresponding author declares that official e-mail from their institution is not available for all authors

Declaration on Publication Ethics:
The authors state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Abstract:
Identification of the toxicity of compounds is more crucial before entering clinical trials. Awareness of physiochemical properties, possible targets and side effects has become a major public health issue to reduce risks. Experimental determination of analyzing the physiochemical properties of a drug, their interaction with specific receptors and identifying their side-effects remain challenging is time consuming and costly. We describe a manually compiled database named DaiCee database, which contains 2100 anticancer drugs with information on their physiochemical properties, targets of action and side effects. It includes both synthetic and herbal anti-cancer compounds. It allows the search for SMILES notation, Lipinski's and ADME/T properties, targets and side effect profiles of the drugs. This helps to identify drugs with effective anticancer properties, their toxic nature, drug-likeness for in-vitro and in-vivo experiments. It also used for comparative analysis and screening of effective anticancer drugs using available data for compounds in the database. The database will be updated regularly to provide the users with latest information. The database is available at the URL http://www.hccbif.org/usersearch.php

Keywords: Drug, properties, SMILES, ADME/T, target

Background:
Cancer is that the second leading explanation for death worldwide. As multiple targets are often involved in cancer disease, much attention has been paid to search drugs involving many targets [**1-3**]. Knowledge of protein targets of drugs is useful in solving various problems in drug discovery process [**4,5**]. Only a limited number of targets are identified as related to approved drugs so far [**4,6,7**]. The therapeutic goals of a variety of medicinal drugs are still unclear. Experiment for identifying target for a drug molecule is time-consuming, expensive, and limited in small-scale research, computational methods are needed to decrease time and costs for drug discovery [**8-11**]. Computational methods can provide supporting evidences to the drug target experiments and accelerate the drug discovery [**12**]. Therefore it is of interest to describe DaiCee, a Database for anti-cancer compounds with targets and side effect profiles for further evaluation using in vitro and in vivo assays. The flow chart of entire work was given in Figure 1.

Methodology:
Data collection:
The database includes 224 anticancer compounds. Its targets and the side effect profile were collected from scientific literatures, various databases, such as DIAB [**13**], Kegg [**14**], Supertarget [**15**], Pubchem [**16**], ChEMBL [**17**], Drug Bank [**18**] and SIDER [**19**]. The database was organized in the alphabetical order of the compounds that simplifies the process of finding the target, properties and side effect profiles. The dataset of the compound were prepared for property prediction.

Generating the properties:
Properties of each compound such as Lipinski's rule of five, ADMET were calculated using Accord Excel 6.1 version. Accord Excel did not provide exact molecular weight and LogP value, rather it gave,if molecular weight less than 500, the result as False, otherwise True. If logP value is less than 5, the result will be "False" otherwise "True". So in order to overcome this difficulty we used Discovery studio 2.1 version, with the help of which we calculated the value of molecular weight and LogP. Accelrys Discovery Studio (2.1) is a life science modeling and simulation suite of application focused on optimizing the drug discovery process. It makes easier to examine the properties of large and small molecules. Accord for Excel allows scientists to display chemical structures and reactions, perform chemical calculations, analyze R-groups, and query by substructure or similarity directly within Excel.

Database design:
The database is developed as a result of integration of all the collected and computed data into a single storehouse through PHP and HTML as front end and MySQL as back end.

Features of the new database:
The database includes compound name, SMILES, drug likeness properties, target and side effect profile as separate fields. Advanced search links were provided to access through compound name, SMILES, target name, molecular formula or molecular weight. This is open source and freely available.

Results & Discussion:
The compound details in the database can be accessed by text search by providing details such as compound name or Protein target details or compound formula or molecular weight of the compound or Smile notation of the compound (**Figure 2a**). Here, Compound name"Aspirin" has given for submission as an example. The results of Aspirin obtained from the database contains all compounds related to Aspirin in the database such as Aspirin aluminum, Aspirin calcium salt, Aspirin lysine salt and Aspirin sodium along with Aspirin are generated (**Figure 2b**). The main page for each entry provides the following information: Molecular formula, Lipinski's rule of five, ADME/T properties, Protein targetand side effect profile (**Figure 2c**).

Drug target discovery, which aims to rapidly and accurately identify drug targets with true potential, is a crucial step in the discovery process and also plays a vital role in new therapeutics. In pharmaceutics, drugs generally fail in the clinic for two reasons: they either do not work or are proved to be unsafe. Drug target validation and identification of undesirable effects are among the main challenges in developing new drugs. A database with collection of a set of information on drugs available for a particular disease on in general would help the researchers for the easy search and knowledge on the compounds, help to guide and speed up the laborious and costly experimental determination on screening or identification of the potential biological activity of the drug. Similar to Jayashree*et al.* 2010 & John *et al.,* 2012 [**20, 21**] the Daicee database is also enriched by user-friendly interface, important keywords are hyperlinked and pictorial representations. *DaiCee* database stands unique with information on both anticancer agents from different sources, not particular in its source. The ADME/T and Lipinski rule of five provided in the database was identified for those drugs with no data reported on it. The information on the side-effect profiles of the anticancer drugs of natural as well as synthetic has been found to be a distinctive parameter in *DaiCee*, which was lacking in other reported databases. It also contains some novel compounds having

anticancer effect. *Daicee* is a unique knowledge and analysis environment for small molecules. It provides calculated molecular descriptors, experimental assay results as target, and literature- based information side effects, allowing integration of both new and established information in a single, public resource.
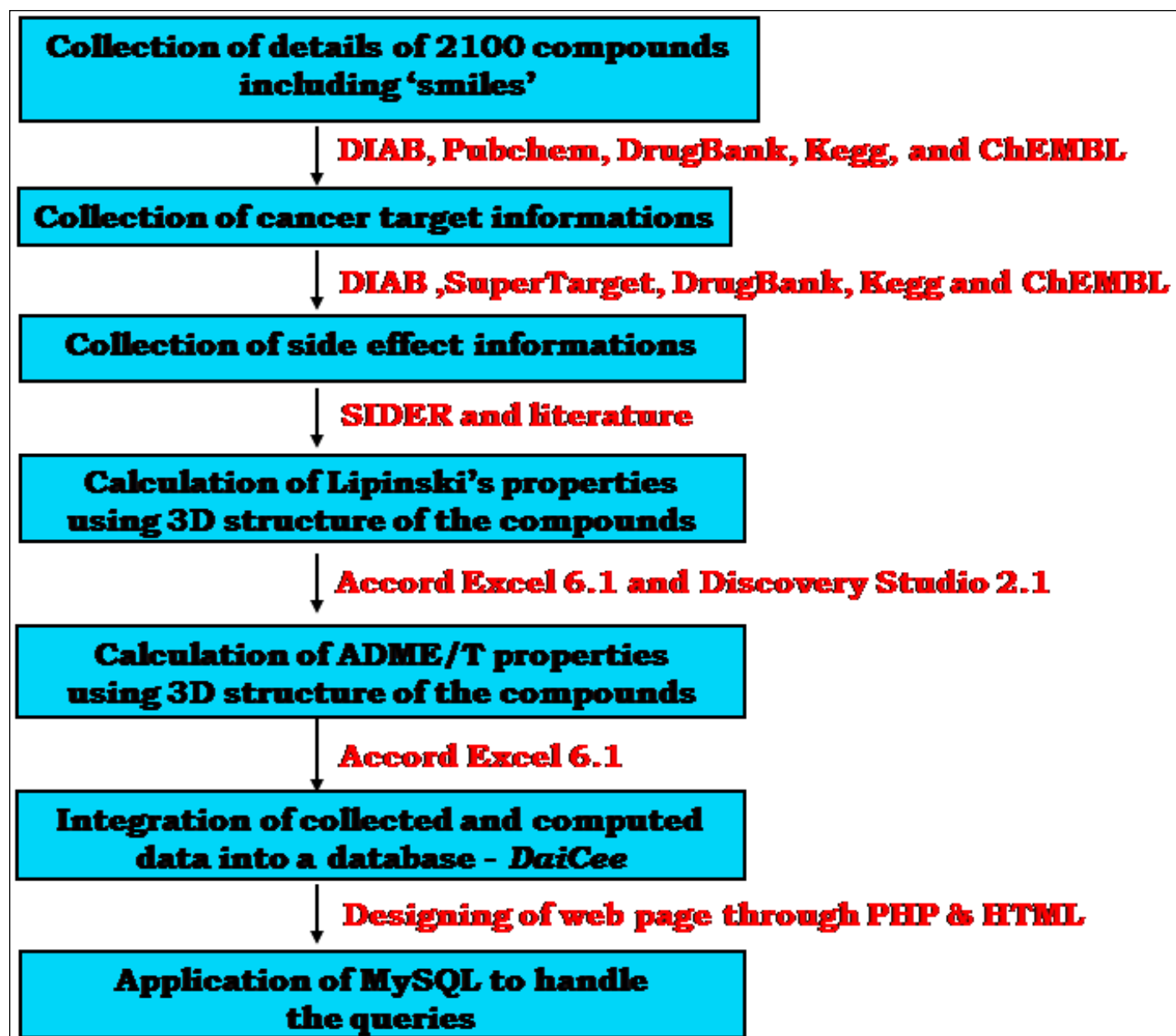

Figure 1: Flowchart for database creation

Figure 2: Web Interface for the database (a) Submission page of *Daicee* Database; (b) A sample of data retrieved for Aspirin from *Daicee* database; (c) ADME/T properties of the Aspirin retrieved from *Daicee* database

Conclusion:
We describe a manually compiled database named DaiCee database, which contains 2100 anticancer drugs with information on their physiochemical properties, targets of action and side effects. It includes both synthetic and herbal anti-cancer compounds. It allows the search for SMILES notation, Lipinski's and ADME/T properties, targets and side effect profiles of the drugs. This helps to identify drugs with effective anticancer properties, their toxic nature, and drug-likeness for in-vitro and in-vivo experiments.

References:
[1] Yang K *et al. Mol Syst Biol*. 2008 **4**:228. [PMID: 18985027]
[2] Zimmermann GR *et al. Drug Discov Today*. 2007 **12**:34. [PMID: 17198971]
[3] Frantz S, *Nature*. 2005 **437**:942 [PMID: 16222266]
[4] Drews J. *Science*. 2000 **287**:1960 [PMID: 10720314]
[5] Ohlstein EH *et al. Annu Rev Pharmacol Toxicol*. 2000 **40**:177 [PMID: 10836132]
[6] Overington JP *Nat Rev Drug Discov*. 2006 **5**:993 [PMID: 17139284]
[7] Landry Y, *Fundam Clin Pharmacol*. 2008 **22**:1. [PMID: 18251718]
[8] Wang YC *et al. Bioinformatics*. 2016 **32**:226. [PMID: 26415726]
[9] Haggarty SJ *et al. Chem Biol*. 2003 **10**:383. [PMID: 12770821]
[10] Kuruvilla FG *et al. Nature*. 2002 **416:**653 [PMID: 11948353]
[11] Yamanishi Y *et al. Bioinformatics*. 2008 **24**:i232. [PMID: 18586719]
[12] Pitchai D *et al. Bioinformation*. 2010 **5**:43. [PMID: 21346859]
[13] Ogata H *et al. Nucleic Acids Res*. 1999 **27**:29 [PMID: 9847135]
[14] Günther S *et al. Nucleic Acids Res*. 2008 **36l**:D919. [PMID: 17942422]
[15] Wang Y *et al. Nucleic Acids Res*. 2009 **37**:W623. [PMID: 19498078]
[16] Gaulton A *et al. Nucleic Acids Res*. 2012 **40**:D1100. [PMID: 21948594]
[17] Wishart DS *et al. Nucleic Acids Res*. 2006 **34**:D668 [PMID: 16381955]
[18] Kuhn M *et al. Mol Syst Biol*. 2010 **6**:343. [PMID: 20087340]
[19] Mohandass J *et al. Bioinformation*. 2010 **5**:1. [PMID: 21346869]
[20] Mary JA *et al. Bioinformation*. 2012 **8**:391. [PMID: 22570521]

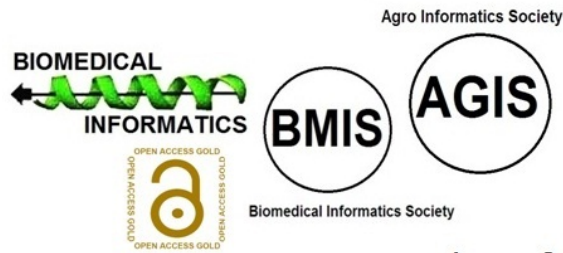Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

BIOINFORMATION
Discovery at the interface of physical and biological sciences

OPEN ACCESS GOLD

BIOMEDICAL INFORMATICS

Agro Informatics Society
AGIS
BMIS
Biomedical Informatics Society
OPEN ACCESS GOLD

since 2005

BIOINFORMATION
Discovery at the interface of physical and biological sciences

indexed in

PMC

Pub Med

EBSCO

INDEXED IN
EMERGING
SOURCES
CITATION
(Web of Science)
CLARIVATE ANALYTICS

WEB OF SCIENCE

Web of Science Group

Crossref

doi®

ResearchGate

R G

publons