Bioinformatics, 2016, 1–3
doi: 10.1093/bioinformatics/btw055
Advance Access Publication Date: 28 January 2016
Applications Note

OXFORD

Databases and ontologies

ASDB: a resource for probing protein functions with small molecules

Zhihong Liu[1], Peng Ding[1], Xin Yan[1], Minghao Zheng[2], Huihao Zhou[1], Yuehua Xu[1], Yunfei Du[3], Qiong Gu[1] and Jun Xu[1,*]

[1]Research Center for Drug Discovery, School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China, [2]Department of Pharmacy, Cancer Hospital of Shantou University Medical College, Shantou 515031, China and [3]National University of Defense Technology, National Supercomupter Center in Guangzhou, Sun Yat-sen University, Guangzhou 510006, China

*To whom correspondence should be addressed.
Associate Editor: John Hancock

Received on 16 November 2015; revised on 29 December 2015; accepted on 23 January 2016

Abstract

Summary: Identifying chemical probes or seeking scaffolds for a specific biological target is important for protein function studies. Therefore, we create the Annotated Scaffold Database (ASDB), a computer-readable and systematic target-annotated scaffold database, to serve such needs. The scaffolds in ASDB were derived from public databases including ChEMBL, DrugBank and TCMSP, with a scaffold-based classification approach. Each scaffold was assigned with an InChIKey as its unique identifier, energy-minimized 3D conformations, and other calculated properties. A scaffold is also associated with drugs, natural products, drug targets and medical indications. The database can be retrieved through text or structure query tools. ASDB collects 333 601 scaffolds, which are associated with 4368 targets. The scaffolds consist of 3032 scaffolds derived from drugs and 5163 scaffolds derived from natural products. For given scaffolds, scaffold-target networks can be generated from the database to demonstrate the relations of scaffolds and targets.
Availability and implementation: ASDB is freely available at http://www.rcdd.org.cn/asdb/with the major web browsers.
Contact: junxu@biochemomes.com or xujun9@mail.sysu.edu.cn
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Medicinal chemists, phytochemists and chemical biologists are seeking novel chemical scaffolds or probes for a specific biological target (mainly a protein). Therefore, great attention has been paid in developing methods and resources to discover new chemical scaffolds or probes associated with biological targets. In the high-throughput synthesis and screening campaigns, privileged scaffolds were identified (Welsch *et al.*, 2010) to establish structure activity relations (SAR) for further property and activity optimizations. Small chemical scaffolds have been studied for many years (Schuffenhauer and Varin, 2011; Xu, 2002), and resulted in many scaffold-oriented techniques including scaffold hopping (Sun *et al.*, 2012), virtual screening (Ertl, 2014), and diversity-oriented synthesis (Galloway *et al.*, 2010). Bemis and Murcko are the pioneers in chemical scaffold concept (Bemis and Murcko, 1996). Xu (2002) used the scaffold concept to classify compounds and measure molecular diversity. Bajorath and colleagues studied scaffolds and activity cliffs relations (Hu and Bajorath, 2010), scaffolds and activities (Hu and Bajorath, 2011).

These scaffold related studies generated new needs for chemical scaffold resources. Shen and coworkers developed a scaffold database for scaffold hopping studies (Yan *et al.*, 2009). To systematically study current chemical scaffolds associated with biological targets, we created an annotated scaffold data base (ASDB). ASDB collects chemical scaffolds derived from the major databases for medicinal chemistry or chemical biology. The scaffolds have been

annotated with their known targets, drugs, natural product sources and medical indications. ASDB is also cross-linked with the ChEMBL (Gaulton *et al.*, 2012), DrugBank (Law *et al.*, 2014), TCMSP (Ru *et al.*, 2014) and UniProt (Bateman *et al.*, 2015) databases, and can be retrieved through structural query or text query tools. To demonstrate scaffold-target and scaffold-scaffold relations, ASDB provides a network viewing tool to present the database search hits.

2 Methods

2.1 Data and the protocol for deriving chemical scaffolds

Raw data were extracted from the ChEMBL_20 (Gaulton *et al.*, 2012) (bioactive compounds), DrugBank (Law *et al.*, 2014) (drugs) and TCMSP (Ru *et al.*, 2014) (natural products) databases. The raw data were processed with the following protocol: (i) removing counter ion moieties; (ii) filtering compounds in ChEMBL_20 for activity (IC_{50}, K_i, K_d) thresholds that were greater than 10 μM; (iii) deriving scaffolds with the scaffold classification approach; and finally, (iv) converting the scaffold data into InChI, InChIKey (scaffold identifier), SMILES, and MOL2 formats by means of the OpenBabel program (O'Boyle *et al.*, 2011). The protocol was written in the Golang language. The flowchart of the process is depicted in Figure 1.

2.2 Web-based database implementation

ASDB is a relational database implemented in MySQL technology and, can be accessed through a web browser using the browser/server framework, and the ChemDoodle web component (web.chemdoodle.com/) as a 2D molecule editor and 3D chemical structure viewer. Four types of queries (structure query, substructure query, 2D similarity and 3D similarity query) are provided for a user to access the ASDB. An online scaffold-target network can be dynamically generated and viewed in ASDB using D3.js (d3js.org/). Technologies for creating ASDB are summarized in Supplementary Table S1.

3 Results and discussion

3.1 ASDB statistics

Currently, ASDB collects 333 601 scaffolds, which are associated with 4368 targets and, contains 491 208 scaffold-target relations. 3032 scaffolds come from drugs; 5163 scaffolds come from natural products. 641 scaffolds were derived from natural products, and experimentally validated as bioactive scaffolds. 174 of the 641 scaffolds were associated with drugs. 4546 natural product associated scaffolds are not annotated with any bioactivity or biological target. These scaffolds are unexplored chemical diversity for chemical biology or medicinal chemistry studies (Supplementary Fig. S1A.). 38% of targets are associated with 0–5 scaffolds. 77% of targets are associated with 0–50 scaffolds. 93% scaffolds are associated with less than five targets (Supplementary Fig. S1C). This data indicates that most targets are biologically unexplored and most scaffolds are either selective or biologically unexplored. 23% targets are associated with 50–500+ scaffolds suggesting these targets have been significantly explored, and the related scaffolds can be privileged chemotypes (Xu *et al.*, 2013) (Supplementary Fig. S1B).

3.2 Web interface

ASDB can be accessed through web browsers. The 'Search' button at the menu bar provides two options for a user to access ASDB scaffolds. The first option is the text query, which can be used to search the target name, UniProt ID or ChEMBL ID. The second option is structure query, which allows a user to conduct full structure search, substructure searches, and 2D or 3D similarity searches. Users can draw a structural query with ChemDoodle, set hit criteria, and get combined search hits. There is a limit to avoid an undesirable amount of hits. The 'Service' button at the menu bar can display a scaffold-target network based upon your scaffold search results. The scaffold-target network is associated with a scaffold list or target list (UniProt IDs or CHEMBL IDs).

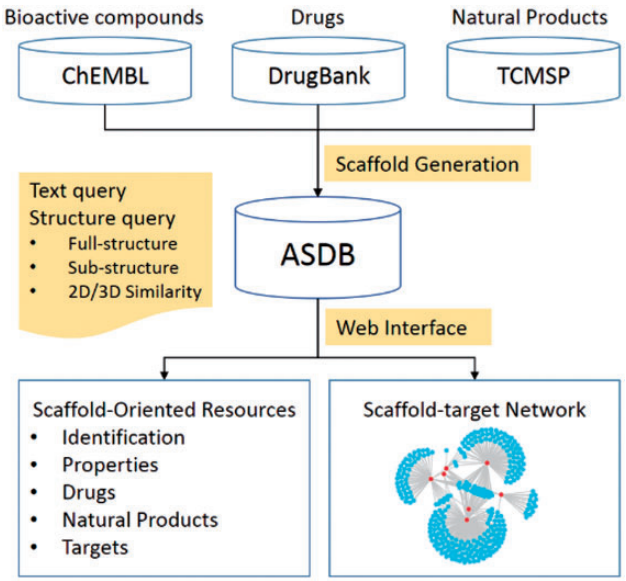

Fig. 1. The flowchart of ASDB data process

3.3 Case studies

Two case studies for target fishing and new scaffold exploring using ASDB are provided in Supplementary materials.

4 Conclusions

Chemical scaffold concept is extensively studied, although it is still controversial. ASDB provides a platform for scientists to explore novel scaffolds or chemical probes for pharmaceutical innovations and chemical biology studies. The scaffold concept is also related to fragment-based drug discovery (FBDD). Chemical scaffolds or fragments are associated with biological targets *per se*, and deserve to have a systemic resource for further explorations.

Funding

National Science Foundation of China (81173470); National High Technology Research and Development Program of China (863 Program, 2012AA020307).

Conflict of Interest: none declared.

References

Bateman, A. *et al.* (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.

Bemis,G.W. and Murcko,M.A. (1996) The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.*, **39**, 2887–2893.

Ertl,P. (2014) Intuitive ordering of scaffolds and scaffold similarity searching using scaffold keys. *J. Chem. Inf. Model.*, **54**, 1617–1622.

Galloway,W.R.J.D. *et al.* (2010) Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules. *Nat. Commun.*, **1**.

Gaulton,A. *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.*, **40**, D1100–D1107.

Hu,Y. and Bajorath,J. (2010) Molecular Scaffolds with High Propensity toForm Multi-Target Activity Cliffs. *J. Chem. Inf. Model.*, **50**, 500–510.

Hu,Y. and Bajorath,J. (2011) Combining horizontal and vertical substructure relationships in scaffold hierarchies for activity prediction. *J. Chem. Inf. Model.*, **51**, 248–257.

Law,V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–D1097.

O'Boyle,N.M. *et al.* (2011) Open Babel: An open chemical toolbox. *J. Cheminformatics*, **3**, 33–46.

Ru,J.L. *et al.* (2014) TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminformatics*, **6**.

Schuffenhauer,A. and Varin,T. (2011) Rule-Based Classification of Chemical Structures by Scaffold. *Mol. Inform.*, **30**, 646–664.

Sun,H. *et al.* (2012) Classification of scaffold-hopping approaches. *Drug Discov. Today*, **17**, 310–324.

Welsch,M.E. *et al.* (2010) Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.*, **14**, 347–361.

Xu,J. (2002) A new approach to finding natural chemical structure classes. *J. Med. Chem.*, **45**, 5311–5320.

Xu,J. *et al.* (2013) Chemomics and drug innovation. *Sci China Chem*, **56**, 71–85.

Yan,B.B. *et al.* (2009) ScafBank: a public comprehensive Scaffold database to support molecular hopping. *Acta Pharmacol. Sin.*, **30**, 251–258.