

ChemMapper: a versatile web server for exploring pharmacology and chemical structure association based on molecular 3D similarity method

Jiayu Gong^{1,†}, Chaoqian Cai^{1,†}, Xiaofeng Liu^{2,*}, Xin Ku³, Hualiang Jiang², Daqi Gao¹ and Honglin Li^{1,2,*}

¹School of Information Science and Engineering, ²Shanghai Key Laboratory of New Drug Design, State Key Laboratory of Bioreactor Engineering, School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China and ³Department of Proteomics and Bioanalytics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Freising 85354, Germany

Associate Editor: Jonathan Wren

ABSTRACT

Summary: ChemMapper is an online platform to predict polypharmacology effect and mode of action for small molecules based on 3D similarity computation. ChemMapper collects >350 000 chemical structures with bioactivities and associated target annotations (as well as >3 000 000 non-annotated compounds for virtual screening). Taking the user-provided chemical structure as the query, the top most similar compounds in terms of 3D similarity are returned with associated pharmacology annotations. ChemMapper is designed to provide versatile services in a variety of chemogenomics, drug repurposing, polypharmacology, novel bioactive compounds identification and scaffold hopping studies.

Availability: <http://ililab.ecust.edu.cn/chemmapper/>.

Contact: xliu@ecust.edu.cn or hlli@ecust.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on February 20, 2013; revised on April 21, 2013; accepted on May 7, 2013

1 INTRODUCTION

Recent advances in systems biology and chemical biology exhibit great challenges to the current drug discovery paradigm that drugs selectively bind with one or two targets and show that most drugs interact with multiple targets, namely, polypharmacology (Keiser *et al.*, 2009; Paolini *et al.*, 2006). On the other hand, identification of new hit compounds and optimization of lead compounds targeting specific proteins or exhibiting desired pharmacological effects are still playing a pivotal role in conventional drug discovery projects in the form of virtual screening and scaffold hopping. In this context, there is an urgent request for developing new strategies to acquire knowledge of the complete pharmacology profile and bridge the chemical and pharmacological spaces to improve the success rates of current drug discovery research, including reducing the side effects and

increasing the regulatory effects on the complete networks (Keiser *et al.*, 2009).

One of the emerging scenarios to address this question is the application of molecular similarity searching on basis of the well-known similar property principle, namely, similar structures may have similar bioactivity and the same potential drug targets, which is expected to relate the pharmacology space by chemical structural similarity information (Johnson *et al.*, 1990). According to this principle, small molecule similarity had been used to identify off-targets for a long period (Boehm *et al.*, 2008). The 2D similarity methods have been widely used in various online tools, like Similarity Ensemble Approach (SEA) (Keiser *et al.*, 2007), SuperPred (Dunkel *et al.*, 2008) and ChemProt (Kim Kjaerulff *et al.*, 2013), which relate protein pharmacology and disease networks by fingerprint-based ligand similarity, and FtreeWeb (Rarey and Dixon, 1998), which adopts graph theory to encode molecular 2D descriptions to perform virtual screening against large chemical databases. Also, numerous online servers using 3D similarity methods are also emerging, like Superimposé (Bauer *et al.*, 2008), which provides two different 3D superimposition algorithms and three different databanks for screening, and wwLigCSRre (Sperandio *et al.*, 2009), which is built on top of CSR algorithm that searches for the maximal common substructure between two sets of unordered coordinates to screen focused chemical libraries. Recently, a direct comparison of 2D and 3D methods for on- and off-target prediction reveals that the benefit of 3D over 2D was obvious for prediction of polypharmacology and drug pairs that shared high 3D similarity but low 2D similarity (i.e. a novel scaffold) were shown to be much more likely to exhibit pharmacologically relevant differences in terms of specific protein target modulation (Yera *et al.*, 2011).

Herein, we present a versatile web-based tool ChemMapper for exploring target pharmacology and chemical relationships against any given small molecules via SHAFTS, a fast 3D similarity method in which the 3D similarity calculation is driven by the hybrid information of molecular shape and chemotype features (Liu *et al.*, 2011; Lu *et al.*, 2011). ChemMapper assembles a large repertoire of bioactive chemical database annotated with target information and multiple screening databases from different catalogs of the chemical vendors. It facilitates identification

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

of potential targets that may play a pivotal role in biological response to the drugs and other active chemicals, as well as the discovery of new hit compounds with similar pharmacology profiles but novel scaffolds. Moreover, ChemMapper can also be useful in the application of drug repurposing, as well as investigation of potential side effects related to the drugs and other active chemicals.

2 METHOD

2.1 Data set-up

The polypharmacology information for the bioactive compounds and biological annotations of the corresponding targets was gathered from various public databases, including ChEMBL (ChEMBL 14), DrugBank (version 3.0), BindingDB (December 2012), KEGG (May 2011) and PDB (November 2010). The chemical–protein interaction annotations were categorized as follows: protein name, UniProt access ID, species, biological functions, Gene Ontology (GO) annotations (molecular function and biological process involved) and activity data. Other non-annotated chemical structures are also incorporated for virtual screening from public databases like ZINC. The final database consists of nearly 350 000 molecular entries annotated for >20 000 proteins and >3 million compounds for virtual screening. Multiple 3D conformers were pre-generated for each compound to facilitate 3D similarity calculation (Supplementary Table S1).

2.2 Similarity metrics

ChemMapper uses SHAFTS as the 3D similarity calculation method. SHAFTS adopts a hybrid similarity metric combined with molecular shape and colored (labeled) chemical groups annotated by pharmacophore features, which is designed to integrate the strength of pharmacophore matching and volumetric overlay approaches. Previous studies proved SHAFTS is more efficient and accurate in ligand-based virtual screening than the common 2D fingerprint-based similarity and other 3D similarity methods (Liu *et al.*, 2011; Supplementary Material).

3 WEB INTERFACE

ChemMapper accepts a chemical structure (sketched online or uploaded in multiple chemical structure file formats) as the query and provides two types of online calculation services: Target Navigator and Hit Explorer. The input 2D structure will be automatically converted into single 3D conformer. In the Target Navigator mode, the users are allowed to search either bioactive compound collections with corresponding target information (ChEMBL, DrugBank and BindingDB) or chemical substrates from the KEGG enzyme database. The results table can be filtered, re-ordered and grouped in terms of various rules (Supplementary Fig. S1a). A list of potential protein targets is returned ranked by the inference of chemical–protein association network whose edges weighed by the bioactivity of the similar compounds to the query (Supplementary Fig. S1b). The biological annotations for each target, including name, species, function and involved pathway, are displayed as well (Supplementary Fig. S1c). In the Hit Explorer mode, only the most similar compounds from various public chemical repositories to the query are returned, which presumably exhibit similar bioactivity or function. Alternatively, the users are able to upload their

customized mini-database (in sdf or mol2 format) for virtual screening by 3D similarity calculation. In both mode, the overlay poses between the query and each target compound can be visualized. All the information in the result table can be downloaded in Comma Separated Values (csv) format as well. Two test cases (polypharmacology effect of loratadine and scaffold-hopping for EGFR inhibitors) relevant to the two modes are described in detail in the Supplementary Materials.

Comparing with 2D fingerprint-based similarity calculation, 3D similarity calculation is slower: a typical query to the ChEMBL (339 624 entities) in ChemMapper takes 12–24 h. A monitoring page is provided for bookmarking to redirect to the result page as the job finishes.

4 CONCLUSION

The goal of ChemMapper is to provide an online versatile framework for fast exploring the target pharmacology and chemical relationship via molecular 3D similarity methods. ChemMapper links bioactivity and target protein annotation data associated to small molecules to chemical 3D similarities, which can lead to novel perspectives and new applications to the old drugs and suggest new associations between the chemical and pharmacological spaces.

Funding: Fundamental Research Funds for the Central Universities, the National Natural Science Foundation of China (21173076, 81102375, 81230090, 81222046 and 81230076), the Specialized Research Fund for the Doctoral Program of Higher Education of China (grant 20110074120009), the Special Fund for Major State Basic Research Project (2009CB918501), the Shanghai Committee of Science and Technology (11DZ2260600 and 12401900801), the National S&T Major Project of China (2011ZX09307-002-03) and the 863 Hi-Tech Program of China (2012AA020308). H.L. is also sponsored by Program for New Century Excellent Talents in University (NCET-10-0378).

Conflict of Interest: none declared.

REFERENCES

- Bauer, R.A. *et al.* (2008) Superimpose: a 3D structural superposition server. *Nucleic Acids Res.*, **36**, W47–W54.
- Boehm, M. *et al.* (2008) Similarity searching and scaffold hopping in synthetically accessible combinatorial chemistry spaces. *J. Med. Chem.*, **51**, 2468–2480.
- Dunkel, M. *et al.* (2008) SuperPred: drug classification and target prediction. *Nucleic Acids Res.*, **36**, W55–W59.
- Johnson, M.A. *et al.* (1990) *Concepts and Applications of Molecular Similarity*. Wiley, New York, NY.
- Keiser, M.J. *et al.* (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.*, **25**, 197–206.
- Keiser, M.J. *et al.* (2009) Predicting new molecular targets for known drugs. *Nature*, **462**, 175–181.
- Kim Kjaerulf, S. *et al.* (2013) ChemProt-2.0: visual navigation in a disease chemical biology database. *Nucleic Acids Res.*, **41**, D464–D469.
- Liu, X. *et al.* (2011) SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. *J. Chem. Inf. Model.*, **51**, 2372–2385.

- Lu, W. *et al.* (2011) SHAFTS: a hybrid approach for 3D molecular similarity calculation. 2. Prospective case study in the discovery of diverse p90 ribosomal S6 protein kinase 2 inhibitors to suppress cell migration. *J. Med. Chem.*, **54**, 3564–3574.
- Paolini, G.V. *et al.* (2006) Global mapping of pharmacological space. *Nat. Biotechnol.*, **24**, 805–815.
- Rarey, M. and Dixon, J.S. (1998) Feature trees: a new molecular similarity measure based on tree matching. *J. Comput. Aided Mol. Des.*, **12**, 471–490.
- Sperandio, O. *et al.* (2009) wwLigCSRre: a 3D ligand-based server for hit identification and optimization. *Nucleic Acids Res.*, **37**, W504–W509.
- Yera, E.R. *et al.* (2011) Chemical structural novelty: on-targets and off-targets. *J. Med. Chem.*, **54**, 6771–6785.