BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 2 2013, pages 292–294
doi:10.1093/bioinformatics/bts651

Databases and ontologies

Advance Access publication November 18, 2012

PTID: an integrated web resource and computational tool for agrochemical discovery

Jiayu Gong[1], Xiaofeng Liu[2], Xianwen Cao[2], Yanyan Diao[2], Daqi Gao[1], Honglin Li[2,*] and Xuhong Qian[2,*]

[1]School of Information Science and Engineering and [2]State Key Laboratory of Bioreactor Engineering, Shanghai Key Laboratory of Chemical Biology, Shanghai Key Laboratory of New Drug Design, Institute of Pharmaceuticals and Pesticides, School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

Associate Editor: Alex Bateman

ABSTRACT

Summary: Although *in silico* drug discovery approaches are crucial for the development of pharmaceuticals, their potential advantages in agrochemical industry have not been realized. The challenge for computer-aided methods in agrochemical arena is a lack of sufficient information for both pesticides and their targets. Therefore, it is important to establish such knowledge repertoire that contains comprehensive pesticides' profiles, which include physicochemical properties, environmental fates, toxicities and mode of actions. Here, we present an integrated platform called Pesticide-Target interaction database (PTID), which comprises a total of 1347 pesticides with rich annotation of ecotoxicological and toxicological data as well as 13 738 interactions of pesticide-target and 4245 protein terms via text mining. Additionally, through the integration of ChemMapper, an in-house computational approach to polypharmacology, PTID can be used as a computational platform to identify pesticides targets and design novel agrochemical products.

Availability: http://lilab.ecust.edu.cn/ptid/.

Contact: hlli@ecust.edu.cn; xhqian@ecust.edu.cn

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

Received on August 13, 2012; revised on October 15, 2012; accepted on October 29, 2012

1 INTRODUCTION

Agrochemical products help us and our crops defend from the invasion of insects, weeds and fungi. Unfortunately, the resistance arising from the increasing selection pressure as well as the cross-resistance of pesticides acting on the same target limits the continuing use of almost all pesticides (Casida, 2009). On the other hand, environmental safety and toxicity are major factors to the failures in last stages of pesticides development. In the face of ever-increasing stringent demands of efficacy and environmental safety, the discovery of new agrochemicals is clearly imperative. Since the 1990's, *in vivo* and *in vitro* high throughput screen has been routinely used in agrochemical discovery (Tietjen *et al.*, 2005). However, both of these random screening methods require a large number of compounds and produce few hits. In the past decade, rational methods such as structure-based design were introduced into agrochemical industries, and the combination of rational and non-rational methods was realized to be more efficient than traditional high throughput screen (Walter, 2002). Recently *in silico* approaches, which are routine methods for developing new pharmaceuticals, have entered the lead discovery processes of agrochemical industry (Speck-Planche *et al.*, 2012a, b). These methods can discover new active ingredients with ideal properties at lower cost in much quicker time. Additionally, risk assessment and toxicity prediction are also considered.

However, both of rational *in vitro* and *in silico* approaches and analysis demand enormous information not only on the structure, physicochemical properties, toxicity and mode of action of existed pesticides but also the structural data of the active sites. Although these data are available via a diversity of sources including government departments, manufacturers and universities, the dispersed data source and a lack of search and analysis tools make accessing inconvenient. Moreover, insufficient knowledge of targets and three-dimensional structures also impedes the applications of computer aided methods.

Consequently, there is a severe need for a comprehensive database covering physicochemical properties, environmental fates and mode of action, with integration of high-level search system and analysis tools, for novel pesticide discovery and risk assessments. The purpose of Pesticide-Target interaction database (PTID) is to establish a web-accessible database providing association information of pesticides and corresponding potential targets via text mining.

2 DATA SETUP

The general pesticides data of PTID were extracted from US EPA Pesticide Programs and EU Pesticide database. These pesticides were grouped into 22 classifications such as insecticides, herbicides and fungicides based mainly on chemical structure and pesticide activity (Supplementary Table S1). In addition to general pesticides information, toxicological and environmental profiles, which came from US-EPA Pesticide project, EU Pesticide database, IPM and PPDB were also integrated. Subsequently, structural data were annotated through NCBI PubChem, and physicochemical properties for each pesticide were calculated (See Supplementary Material Table S2 for the details of data source).

*To whom correspondence should be addressed.

© The Author 2012. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

To derive potential pesticides–proteins relationships, which were hidden in a vast amount of dispersed literature, text mining method were introduced. Starting with 1347 pesticides and their synonyms, the text mining tool PolySearch was used for extracting the relevant text passages containing potential pesticide-target relationships from public abstracts listed in PubMed (Cheng *et al*., 2008). More than 13 000 interactions of pesticide-target were extracted, and 4245 protein terms were finally mapped onto the pesticides and classified accordingly. We assigned relevancy score as confidence score to reflect the level of significance and certainty of interactions. Details of relevancy score could refer to the Supplementary Material. In addition to literature-derived protein terms, sequence date and other annotation were searched from Entrez Gene and SwissProt/TrEMBL by querying the combination of protein terms and organism of pests.

3 USER INTERFACE AND SERVICE

3.1 Web interface

The data of PTID is accessible through a simple, user-friendly web interface. In addition to intelligent text search and advanced Boolean text search, PTID also offers a sequence alignment search and a chemical structure search.

The structure similarity search tool allows users to quickly and simply identify pesticide scaffolds by inputting query template. PTID provides three methods for similarity searching. Fingerprints molecular similarity and substructure searching implemented by Open Babel allow a quick identification of pesticide scaffolds by inputting query template, whereas 3D similarity method implemented by SHAFTS, which combines shape similarity and feature similarity and allows an identification of pesticides with similar physiochemical properties, may affect the same mode of action (Liu *et al*., 2011; Lu *et al*., 2011; O'Boyle *et al*., 2011).

Similarly, sequence alignment is particularly useful, as it can potentially allow users to quickly and simply identify pesticides from newly sequenced pathogens. Given the FASTA formatted sequence and chosen target organism, the sequence alignment search service performs a BLAST search on PTID target dictionary.

3.2 Local BLAST service

Sequence alignment is crucial for many biological analyses, including functional annotation, classifying protein and structure modeling. PTID provides local BLAST service in the detailed page of each sequence. Sequence index of non-redundant protein sequence, PDB and SwissProt were provided. The BLAST service allows users to specify database, sequence segment and other parameters. The output page, which is formatted by BLAST, includes a link to corresponding record of these protein databases.

3.3 ChemMapper service

Through the integration of ChemMapper, PTID provided the functions for novel pesticides discovery and potential mode of action prediction *via* polypharmacology effects computation (http://lilab.ecust.edu.cn/chemmapper/). ChemMapper service can be used to generate novel chemical scaffolds and lead discovery by the pattern of known agrochemicals in PTID through three-dimensional molecular similarity search in the large scale of chemical library. On the other hand, searching in a large repertoire of bioactive chemical database with target annotation can help users to retrieve new biological effects or elucidate their mechanism of action via predicting potential targets for those agrochemicals; moreover, potential toxicity can also be predicted by comparing with similar chemicals, which bind to anti-targets of Homo sapiens or mammalia. (See more details from Supplementary Material)

3.4 Pesticides-targets network visualization

Graphic network always provides intuitively view of interaction data. PTID presented current entry in context of an interactive graphic network implemented by customized Cytoscape web network visualization components of interaction partners (Lopes *et al*., 2010). The customized Cytoscape web plugin has a user-friendly graphical user interface and supplies various tools and layouts for network analysis. Users may filter nodes or edges according to confidence score or distance between other nodes and current entry. Using network analysis, it can be hypothesized that certain pesticides, which have high degrees, could show their toxicity and adverse environmental effect (Supplementary Figure S1).

4 CONCLUSION

The current version of PTID contains ~1347 pesticides with annotation of toxicity and environmental fates. More than 4245 potential target terms were derived by text mining tool; sequence data were also annotated for each target term. In addition to these data, several computational tools for target exploration and virtual screen were also integrated into PTID. Potential application of PTID includes identification of pesticides by structures or properties interest; prediction of potential targets; or assessment of toxicity and environmental effect. To our knowledge, PTID is the first attempt to establish a pesticide database, which is integrated with the knowledge of protein–protein interactions. It is expected that PTID will serve as a useful resource for the development of agrochemicals.

Funding: Fundamental Research Funds for the Central Universities, the National Natural Science Foundation of China (grants 21173076, 81102375, 81230076, 81230090 and 81222046), the Shanghai Committee of Science and Technology (grant 11DZ2260600), the 863 Hi-Tech Program of China (grant 2012AA020308) and National Basic Research Program of China (973 Program, 2010CB126100). H.L. is also sponsored by Program for New Century Excellent Talents in University (grant NCET-10-0378).

Conflict of Interest: none declared.

REFERENCES

Casida,J.E. (2009) Pest toxicology: the primary mechanisms of pesticide action. *Chem. Res. Toxicol*., **22**, 609–619.

Cheng,D. *et al*. (2008) PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. *Nucleic Acids Res*., **36**, W399–W405.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

Liu,X.F. *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.

Lopes,C.T. *et al.* (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.

Lu,W.Q. *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.

O'Boyle,N.M. *et al.* (2011) Open Babel: an open chemical toolbox. *J. Cheminform.*, **3**, 33.

Speck-Planche,A. *et al.* (2012a) Predicting multiple ecotoxicological profiles in agrochemical fungicides: a multi-species chemoinformatic approach. *Ecotoxicol. Environ. Saf.*, **80**, 308–313.

Speck-Planche,A. *et al.* (2012b) Fragment-based approach for the in silico discovery of multi-target insecticides. *Chemometr. Intell. Lab.*, **111**, 39–45.

Tietjen,K. *et al.* (2005) High throughput screening in agrochemical research. *Comb. Chem. High Throughput Screen*, **8**, 589–594.

Walter,M.W. (2002) Structure-based design of agrochemicals. *Nat. Prod. Rep.*, **19**, 278–291.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016