BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 14 2011, pages 2018–2020
doi:10.1093/bioinformatics/btr333

Data and text mining

Advance Access publication June 2, 2011

The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections

David Lagorce[1,*], Julien Maupetit[1,2], Jonathan Baell[3,4], Olivier Sperandio[1], Pierre Tufféry[1,2], Maria A. Miteva[1], Hervé Galons[5] and Bruno O. Villoutreix[1,2,*]

[1]MTi, [2]Ressource Parisienne en Bioinformatique Structurale (RPBS), Institut National de la Santé et de la Recherche Médicale (INSERM), UMR-S 973 – Paris Diderot University, 75205 Paris, Cedex 13, France, [3]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, [4]Department of Medical Biology, The University of Melbourne, Parkville, Victoria 3010, Australia and [5]UMR8601, Paris Descartes University, 75006 Paris, France

Associate Editor: John Quackenbush

ABSTRACT

Summary: The FAF-Drugs2 server is a web application that prepares chemical compound libraries prior to virtual screening or that assists hit selection/lead optimization before chemical synthesis or ordering. The FAF-Drugs2 web server is an enhanced version of the FAF-Drugs2 package that now includes Pan Assay Interference Compounds detection. This online toolkit has been designed through a user-centered approach with emphasis on user-friendliness. This is a unique online tool allowing to prepare large compound libraries with in house or user-defined filtering parameters.

Availability: The FAF-Drugs2 server is freely available at http://bioserv.rpbs.univ-paris-diderot.fr/FAF-Drugs/.

Contact: david.lagorce@inserm.fr; bruno.villoutreix@inserm.fr

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

Received on April 6, 2011; revised on May 16, 2011; accepted on May 28, 2011

1 INTRODUCTION

The most established method for the identification of hit compounds modulating the functions of a target is experimental high-throughput screening (HTS). Alternatively or in parallel, screening can be performed *in silico* (virtual ligand screening or VLS) in order to prioritize compounds for *in vitro* screening. The screening process involves measurement/prediction of activities of sometimes millions of chemical compounds *in vitro*, on cells or *in silico*. The first two, in particular, are time consuming and costly, and as such, the preparation of the compound collections is critical. Analyses of past failures have led to a much better understanding of crucial properties that distinguish any chemical from an interesting drug-candidate or a relevant chemical probe. Thus, the concept of screening high-quality compound collections in terms of improved ADMET (absorption, distribution, metabolism, excretion, toxicity) properties and containing a reduced number of 'nuisance compounds' is gaining momentum. Hence, we decided to develop a user-friendly online tool dedicated to the preparation of compound collections. In fact, such a service, freely available and able to handle several thousands of compounds with either predefined filtering parameters or user-tuned parameters, assisting decision making has not been reported to date. Our online *in silico* filtering engine is based on a significantly improved version of a previously reported stand-alone package named FAF-Drugs2, standing for Free ADMET Filtering-Drugs2 (Lagorce *et al.*, 2008), which was successfully employed for preparing compound datasets for different projects as in Reynes *et al.* (2010). While many free online services are available to compute molecular properties (Ertl *et al.*, 2007; Tetko, 2003; Walker *et al.*, 2010), FAF-Drugs2 is the first free web-based package capable of preparing compound libraries through physicochemical rules, functional groups and Pan Assay Interference Compounds (PAINS) detection (Baell *et al.*, 2010). In the present application note, we describe the FAF-Drugs2 server highlighting the new functionalities introduced as compared with the original stand-alone package.

2 THE FAF-DRUGS2 SERVER

2.1 Improvements of FAF-Drugs2 in the web server version

FAF-Drugs2 client version requires as input the compound collection in SD or SMILES file format and two parameter files, one containing among others physicochemical thresholds and another one listing the chemical substructures that have to be investigated. Various output files are generated, e.g. SDF filtered files or a tabulated file reporting all the computed values/descriptors. The main improvements on the server version as compared with the FAF-Drugs2 stand-alone tool include: (i) a step before the filtering process to prepare and clean the electronic input molecular data file, with removal of empty structures, salts, counterions, inorganics, mixtures, duplicates. Further, we apply a standardization procedure on eight common chemical groups using SMARTS search and the ChemAxon Standardizer utility (see Supplementary Material). Yet, in order to compute some descriptors the way they were originally described, a rule-based protonation protocol is internally performed with OpenBabel (O'Boyle *et al.*, 2008). (ii) Novel, optimized and user-defined filtering rules, with a major improvement of the logP calculation through the use of the XLOGP3 program enhanced with experimental logP values extracted from the PHYSPROP database (Syracuse Research Database). (iii) More than 200 SMARTS patterns used to search for undesirable moieties. (iv) Implementation of PAINS filters optimized by our research group. Such moieties

*To whom correspondence should be addressed.

have been described to appear as frequent hitters or promiscuous compounds in many HTS experiments. (v) The ESOL solubility estimation model (see Supplementary Material) and some others useful features are now available such as various oral bioavailability evaluations. (vi) Finally, an intuitive visualization of the results through any web browser is now implemented. More explanations about our filtering engine protocol are detailed in the Supplementary Material and the online userguide.

2.2 Implementation

The optimized FAF-Drugs2 engine has been embedded in the RPBS' Mobyle Portal (Neron *et al.*, 2009). Mobyle's features makes it the perfect solution for online toolkit implementation, because it offers: (i) a centralized workspace for the end-user (bookmarked results and parameters are stored on the server for further uses and backup); (ii) on-the-fly program results pipelining; and (iii) a user-friendly interface and a robust tool to interact with end users and control programs' execution on the server side (storage and resources quota, jobs tracking, etc.). The FAF-Drugs2 web server is organized as a toolkit composed of three services: (i) Bank-Formatter (e.g. converts SMILES input files to SD files format, see Supplementary Material); (ii) Filter-Editor (customizes user-defined filters); and (iii) the FAF-Drugs2 filtering service. The server has been validated and optimized: (i) on 10 000 molecules from the WEHI HTS library (see Supplementary Material); (ii) to be fully functional on popular operating systems and web browsers; and (iii) to run on our 800 nodes cluster (see Supplementary Material).

2.3 Interface features

SD files, as long as they do not exceed 50 000 molecules, can be submitted to the FAF-Drugs2 either by uploading input files on the server or by pasting data in a dedicated window. Instead of using the Filter-Editor service, users can also choose among various physicochemical elsewhere-published pre-tuned filters; among others REOS (Walters *et al.*, 2002) or ZINC (Irwin *et al.*, 2005). Furthermore, users can decide to activate the PAINS and functional group substructures detection. When the process is terminated, while filtering result files can be downloaded, various analyses can be done directly in several internet browser windows, like a statistical analysis of the main descriptor distributions (Fig. 1a), a pie chart summarizing the results for the substructures search (Fig. 1b), a table detailing the results for each compound: the filtering options, the compound's 2D depiction via the ChemAxon molconvert tookit (Fig. 1c), the compound projection (magenta) onto the first plan of the principal component analysis of the orally bioavailable (blue) DrugBank Small molecules (Fig. 1d) and a radar plot depicting some molecular properties important for oral absorption (Fig. 1e).

3 SUMMARY AND PERSPECTIVES

The FAF-Drugs2 web server is the first free web-based application devoted to compound library preparation and PAINS detection. The service can also be used to evaluate drug-like properties, the presence of an undesirable substructure and oral absorption of a small set of compounds before synthesis. This service should be of interest to both academic and private research groups.

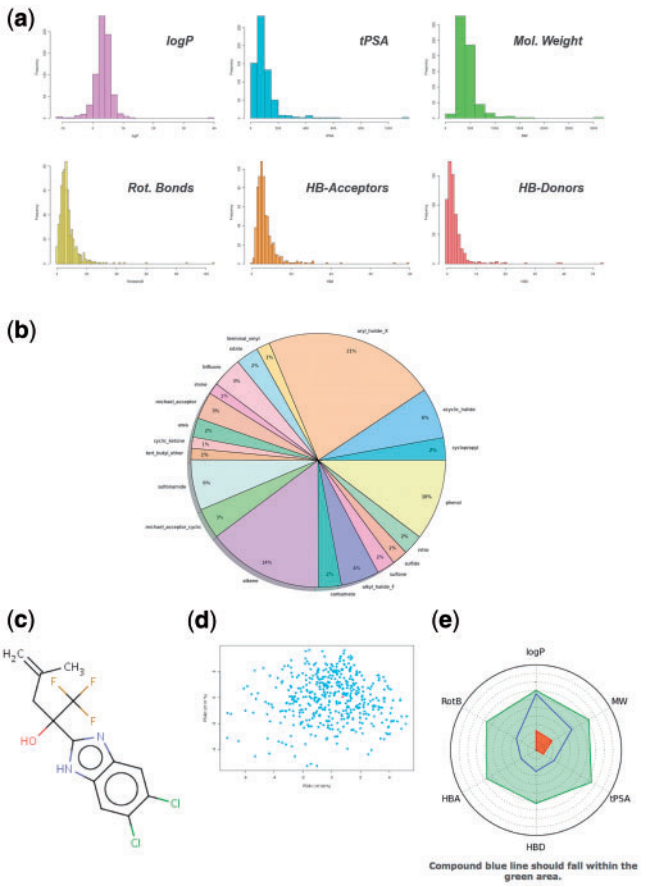

Fig. 1. Various screenshots from Internet browser windows that screenplay the FAF-Drugs2 results. (**a**) Main PhysChem descriptors analysis of random_bank_5000.sdf library; (**b**) Pie chart of the most found flagged-rejected chemical moieties in the random_bank_5000.sdf library; summary of the random_bank_5000_1_formatted_standardized_pH.sdf library: (**c**) compound 2D representation; (**d**) oral property space; (**e**) oral absorption estimation.

ACKNOWLEDGEMENTS

The authors thank INSERM and Paris Diderot University for supports and Dr G. Laconde and Mr. C. Gageat for helpful discussions. We are grateful to Chemaxon for providing a license allowing the use of several utilities included in their package.

Conflict of Interest: none declared.

REFERENCES

Baell,J.B. and Holloway,G.A. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.*, **53**, 2719–2740.

Ertl,P. and Jelfs,S. (2007) Designing drugs on the internet? Free web tools and services supporting medicinal chemistry. *Curr. Top. Med. Chem.*, **7**, 1491–1501.

Irwin,J.J. and Shoichet,B.K. (2005) ZINC–a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.*, **45**, 177–182.

Lagorce,D. *et al.* (2008) FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. *BMC Bioinformatics*, **9**, 396.

Neron,B. *et al.* (2009) Mobyle: a new full web bioinformatics framework. *Bioinformatics*, **25**, 3005–3011.

O'Boyle,N.M. *et al.* (2008) Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.*, **2**, 5.

Reynes,C. *et al.* (2010) Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods. *PLoS Comput. Biol.*, **6**, e1000695.

Tetko,I.V. (2003) The WWW as a tool to obtain molecular parameters. *Mini. Rev. Med. Chem.*, **3**, 809–820.

Walker,T. *et al.* (2010) Chembench: a cheminformatics workbench. *Bioinformatics*, **26**, 3000–3001.

Walters,W.P. and Murcko,M.A. (2002) Prediction of 'drug-likeness'. *Adv. Drug Deliv. Rev.*, **54**, 255–271.