Challenges **2014**, *5*, 334-337; doi:10.3390/challe5020334

OPEN ACCESS

challenges

ISSN 2078-1547

www.mdpi.com/journal/challenges

Communication

MyChEMBL: A Virtual Platform for Distributing Cheminformatics Tools and Open Data

Mark Davies [1], Michał Nowotka [1], George Papadatos [1], Francis Atkinson [1], Gerard J. P. van Westen [1], Nathan Dedman [1], Rodrigo Ochoa [2] and John P. Overington [1,*]

[1] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK; E-Mails: mdavies@ebi.ac.uk (M.D.); mnowotka@ebi.ac.uk (M.N.); georgep@ebi.ac.uk (G.P.); francis@ebi.ac.uk (F.A.); gerardvw@ebi.ac.uk (G.J.P.W.); ndedman@ebi.ac.uk (N.D.)

[2] Programa de Estudio y Control de Enfermedades Tropicales (PECET), Universidad de Antioquia, Medell ń 53-108, Colombia; E-Mail: rodrigo.ochoa@udea.edu.co

***** Author to whom correspondence should be addressed; E-Mail: jpo@ebi.ac.uk; Tel.: +44-(0)-1223-492-666; Fax: +44-(0)-1223-494-468.

External Editor: Luc Patiny

Received: 14 August 2014; in revised form: 22 September 2014 / Accepted: 24 September 2014 / Published: 29 September 2014

Abstract: MyChEMBL is an open virtual platform which provides a free, secure, standardised and easy to use chemoinformatics environment for bioactivity data mining, machine learning, application development, learning and teaching. The main technical features of myChEMBL along with its applications and future plans are discussed here.

Keywords: virtual machine; ChEMBL; open data; RDKit; web services

---

1. Introduction

MyChEMBL [1] is an open virtual platform that combines public domain bioactivity data with open source web, database and chemoinformatics technologies. MyChEMBL consists of a Linux (Ubuntu) Virtual Machine (VM), with key installed components including a PostgreSQL version of the ChEMBL database [2] and the latest RDKit chemoinformatics toolkit and chemistry cartridge [3].

The primary aim of the system is to remove the technical hurdles often associated with building and deploying chemoinformatic platforms, thus allowing both novice and expert users easy access to domain-specific data and tools. In addition to the ChEMBL database and RDKit libraries, myChEMBL VM also provides secure local access to the ChEMBL Web Services [4], interactive IPython notebook tutorials [5], the phpPgAdmin PostgreSQL schema browser [6] and example KNIME [7] workflows. Furthermore, these components are linked together by middleware developed in-house; the latter abstracts common tasks, such as interaction with the database and API, networking, etc. Access to all of these tools and services, along with additional documentation, is provided through the myChEMBL LaunchPad landing page.

2. Results and Discussion

Based on the technical features of myChEMBL described above, the platform has several applications and advantages:

- **No Costs**—myChEMBL uses exclusively free and open source tools and libraries, so it removes the expensive licensing costs often associated with similar applications.
- **Security**—myChEMBL runs locally behind a firewall, therefore the typical concerns regarding submission of sensitive data to web-based applications do not apply.
- **Application Development**—the source code is available for all myChEMBL applications, so developers can use this as a starting point for applications they wish to develop in the future.
- **Ease of use**—Due to the availability of interactive, web- and GUI-based tools, myChEMBL requires no prior programming experience or knowledge.
- **Learning**—myChEMBL provides a versatile platform for learning chemical data mining and cheminformatics in an intuitive and straightforward way. The combination of data with relevant pre-installed tools effectively lowers the 'activation barrier' and shifts the focus to hands-on programming and learning.
- **Training**—myChEMBL is a proven resource for training scientists on the use of essential tools in the field of chemoinformatics and computer-aided drug discovery.

3. Conclusions

In conclusion, the primary goal of the myChEMBL project (currently in its second release) has been to provide a truly open chemoinformatics platform, combining open data with open tools and tutorials. Although a fairly recent development, myChEMBL has already been adopted by both academic and industrial groups as a standardised chemoinformatics resource. Looking forward, we envisage broadening the scope of myChEMBL by integrating more open tools, such as Beaker [8], along with adding completely new functionality, such as a compound registration mechanism and bioactivity curation interface. The latter could be linked to an open electronic lab notebook (eLNB), thus offering a complete solution for reporting, storing and querying experimental data. Furthermore, it is hoped that the availability of a completely free, self-contained and extendable version of ChEMBL will catalyze further innovation and development in emerging economies and Open Science/Data projects in areas such as malaria and TB research [9]. Finally, due to the open philosophy of this project, we encourage

the community to provide feedback, new ideas, IPython notebooks or complete tools, in order to enhance and improve the current functionality.

Acknowledgments

The myChEMBL VM relies upon Open Source software packages. Please refer to the following link, where we attempt to acknowledge and reference all of the core software components and tools we have used to build myChEMBL: ftp://ftp.ebi.ac.uk/pub/databases/chembl/VM/myChEMBL/current/ACKNOWLEDGMENTS.

Funding: Strategic Award for Chemogenomics from the Wellcome Trust (WT086151/Z/08/Z); European Molecular Biology Laboratory.

Availability

The myChEMBL virtual machine is available to download from the following link: ftp://ftp.ebi.ac.uk/pub/databases/chembl/VM/myChEMBL/current.

The scripts and installation instructions used to build myChEMBL are available in the following GitHub repository: https://github.com/chembl/mychembl.

MyChEMBL is also available as a Vagrant development environment [10]. For more details: http://chembl.blogspot.co.uk/2014/06/how-to-install-mychembl-using-two.html.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Ochoa, R.; Davies, M.; Papadatos, G.; Atkinson, F.; Overington, J.P. MyChEMBL: A virtual machine implementation of open data and cheminformatics tools. *Bioinformatics*, **2014**, *30*, 298–300, doi:10.1093/bioinformatics/btt666.
2. Bento, A.P.; Gaulton, A.; Hersey, A.; Bellis, L.J.; Chambers, J.; Davies, M.; Krueger, F.A.; Light, Y.; Mak, L.; McGlinchey, S.; *et al*. The ChEMBL bioactivity database: An update. *Nucl. Acids Res. Database Issue.* **2014**, *42*, D1083–D1090, doi:10.1093/nar/gkt1031.
3. RDKit: Cheminformatics and Machine Learning Software. Available online: www.rdkit.org/ (accessed on 24 July 2014).
4. ChEMBL Web Services. Available online: www.ebi.ac.uk/chembl/ws (accessed on 4 August 2014).
5. The IPython Notebook. Available online: http://ipython.org/notebook.html (accessed on 4 August 2014).
6. phpPgAdmin. Available online: http://phppgadmin.sourceforge.net/doku.php (accessed on 4 August 2014).
7. Berthold, M.R.; Cebron, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer: Berlin, Germany, 2007; pp. 319–326.

8. ChEMBL Beaker. Available online: https://github.com/mnowotka/chembl_beaker (accessed on 4 August 2014).
9. Open Source Malaria. Available online: http://opensourcemalaria.org/ (accessed on 4 August 2014).
10. Vagrant. Available online: www.vagrantup.com/ (accessed on 4 August 2014).

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).