ChemmineOB

Kevin Horan & Thomas Girke

October 22, 2013

Contents

1 Introduction

ChemmineOB provides an R interface to a subset of cheminformatics functionalities implemented by the OpelBabel C++ project (O'Boyle et al., 2008, 2011). OpenBabel is an open source cheminformatics toolbox that includes utilities for structure format interconversions, descriptor calculations, compound similarity searching and more. *ChemineOB* aims to make a subset of these utilities available from within R. For non-developers, *ChemineOB* is primarily intended to be used from *ChemmineR* (Cao et al., 2008; Backman et al., 2011; Wang et al., 2013) as an add-on package rather than used directly.

2 Installation

To use the *ChemmineOB* package on Linux or Mac, OpenBabel 2.3.0 or greater needs to be installed on a system. On Linux systems, the header files are also required to compile. The windows distribution will include its own version of OpenBabel. The OpenBabel site (`http://openbabel.org/wiki/Get_Open_Babel`) provides excellent instructions for installing the OpenBabel software on Mac or Linux systems. The *ChemmineR* and *ChemmineOB* packages can be installed from within R with the `biocLite` install script.

If the installation fails on Linux, you may need to manually set the locations of the open babel libraries and header files. This is best done through configure flags. For example, at the command prompt do:

$ R CMD INSTALL –configure-args='–with-openbabel-include=... –with-openbabel-lib=...' ChemmineOB

where the '...' are replaced by the relevant paths. See the README file for more details.

```
> source("http://bioconductor.org/biocLite.R")
> biocLite(c("ChemmineR", "ChemmineOB"))
> library("ChemmineR")
> library("ChemmineOB")
```

3 User Manual in ChemmineR Vignette

Detailed instructions for using *ChemmineOB* are provided in the vignette of the *ChemmineR* package instead of this document. The main reason for consolidating the documentation in one central document rather than distributing it across several vignettes is that it helps minimizing duplications and inconsistencies. It also is the more suitable format for providing a task-oriented description of functionalities for users. To obtain an overview of the OpenBabel utilities supported by *ChemmineOB*, we recommend consulting the *OpenBabel Functions* section of the *ChemmineR* vignette. To open the *ChemmineR* vignette from R, one can use the following command.

```
> vignette("ChemmineR")
```

4 Version Information

```
> toLatex(sessionInfo())
```

- R version 3.0.2 (2013-09-25), `i386-w64-mingw32`
- Locale: `LC_COLLATE=C`, `LC_CTYPE=English_United States.1252`, `LC_MONETARY=English_United States.1252`, `LC_NUMERIC=C`, `LC_TIME=English_United States.1252`
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: ChemmineOB 1.0.1
- Loaded via a namespace (and not attached): BiocStyle 1.0.0, tools 3.0.2, zlibbioc 1.8.0

5 Funding

This software was developed with funding from the National Science Foundation: ABI-0957099, 2010-0520325 and IGERT-0504249.

6 References

T W Backman, Y Cao, and T Girke. ChemMine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Res*, 39(Web Server issue):486–491, Jul 2011. doi: 10.1093/nar/gkr320. URL `http://www.hubmed.org/display.cgi?uids=21576229`.

Y Cao, A Charisi, L C Cheng, T Jiang, and T Girke. ChemmineR: a compound mining framework for R. *Bioinformatics*, 24(15):1733–1734, Aug 2008. doi: 10.1093/bioinformatics/btn307. URL `http://www.hubmed.org/display.cgi?uids=18596077`.

N M O'Boyle, C Morley, and G R Hutchison. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem Cent J*, 2(1):5–5, Mar 2008. doi: 10.1186/1752-153X-2-5. URL `http://www.hubmed.org/display.cgi?uids=18328109`.

N M O'Boyle, M Banck, C A James, C Morley, T Vandermeersch, and G R Hutchison. Open Babel: An open chemical toolbox. *J Cheminform*, 3:33–33, 2011. doi: 10.1186/1758-2946-3-33. URL `http://www.hubmed.org/display.cgi?uids=21982300`.

Y Wang, T W Backman, K Horan, and T Girke. fmcsR: Mismatch Tolerant Maximum Common Substructure Searching in R. *Bioinformatics*, Aug 2013. doi: 10.1093/bioinformatics/btt475. URL `http://www.hubmed.org/display.cgi?uids=23962615`.