Friday, June 7, 2024

R Consortium Report

volcalc: Calculate Volatility of Chemical Compounds

Kristina Riemer
Communications & Cyber Technologies, University of Arizona

Eric R. Scott
Communications & Cyber Technologies, University of Arizona

Laura Meredith
School of Natural Resources and the Environment, University of Arizona

S. Marshall Ledford
School of Natural Resources and the Environment, University of Arizona

Summary

This is the final report on the ISC-funded work on the `volcalc` package. The GitHub repository for the current version of `volcacl` can be found at: https://github.com/Meredith-Lab/volcalc. We have completed the proposed milestones 3 and 4 as well as a modified version of our proposed dissemination plan.

Milestone 3

`volcalc` now includes two vignettes—one demonstrating downloading .mol files from KEGG, and another showing general usage for calculating estimated volatility. A `pkgdown` website including these vignettes is available here: https://meredith-lab.github.io/volcalc/. The code for `volcalc` has been archived on Zenodo (Riemer et al., 2023) and a methods paper has been published in *Frontiers in Microbiology* (Meredith et al., 2023). We've submitted `volcalc` to CRAN and it was rejected due to issues that are unfortunately out of our hands—the "License" field in the DESCRIPTION file of a Bioconductor dependency was not

CRAN compliant. As of today, this issue appears to have been fixed, but now automated checks on Windows are failing because of issues with the `RCurl` package and on macOS because of issues building `ChemmineOB`. When these problems in dependencies are fixed, we may attempt another CRAN submission. In the meantime, `volcalc` can be installed from GitHub or r-universe.

Milestone 4

We added functionality to supply chemical representations as SMILES strings in addition to paths to .mol files. We have not (yet) added a vignette about integrating `webchem` and `volcalc` because few `webchem` functions provide results with SMILES strings. `ChemmineOB` is able to translate from more commonly returned formats such as InChI, however that feature is not available on Windows, so we are questioning the value of adding this feature to `volcalc`.

Other improvements

We also added some validation abilities to `volcalc` that are important due to the way `ChemmineOB` and `ChemmineR` interact with the command line program Open Babel. When Open Babel encounters errors parsing a .mol file or SMILES string, error messages are displayed on the R console, but cannot be captured and acted on (see discussion in issue #56). The workaround we implemented was to include code to look for the "symptoms" of parsing errors—namely a missing value for `InChI` returned by `ChemmineR::propOB()`. Unfortunately, this strategy doesn't work on Windows because InChI generation is not available on that OS. We implemented this as a `validate` argument to `calc_vol()` and `get_fx_groups()` which defaults to `TRUE` and prints an warning when set to `TRUE` and run on Windows.

Dissemination

In addition to the *Frontiers in Microbiology* publication, we have shared `volcalc` on social media (Mastodon), in blog posts, in the rOpenSci Slack group, and in talks. The package authors have also reached out personally to researchers who may be interested in these calculations.

Future opportunities

Besides continuing to try to get `volcalc` on CRAN, we have a few open issues that we see as potential improvements to `volcalc` that could be part of future work. Because of various difficulties encountered with `ChemmineOB` and `ChemmineR`, we are considering dropping these Bioconductor dependencies and implementing interactions with Open Babel directly in `volcalc`. This would potentially lead to a more stable and self-contained package, however it would be a substantial amount of work. Similarly, we may consider dropping the `KEGGREST` dependency as it is only used for one function to access the KEGG API, and we've already created other functions in `volcalc` that access different KEGG API endpoints.

In the original SIMPOL.1 method, equations are provided to calculate group contribution coefficients at any temperature (Pankow & Asher, 2008). As it is implemented currently in `volcalc`, however, volatility is only estimated at a fixed temperature of 20°C. Implementing a variable temperature would not be trivial, but would add a great deal of flexibility for estimating relative volatility at a range of environmental conditions.

We have opened GitHub issues for all of these future improvements and will consider pursuing additional funding for another round of `volcalc` development as the user base grows.

References

References

Meredith, L. K., Ledford, S. M., Riemer, K., Geffre, P., Graves, K., Honeker, L. K., LeBauer, D., Tfaily, M. M., & Krechmer, J. (2023). Automating methods for estimating metabolite volatility. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1267234

Pankow, J. F., & Asher, W. E. (2008). SIMPOL.1: A simple group contribution method for predicting vapor pressures and enthalpies of vaporization of multifunctional organic compounds. *Atmos. Chem. Phys.* https://doi.org/10.5194/acp-8-2773-2008

Riemer, K., Scott, E. R., & Meredith, L. K. (2023, December 15). *volcalc: Calculate volatility of chemical compounds*. Zenodo. https://doi.org/10.5281/ZENODO.8015154