R Consortium Report

volcalc: Calculate Volatility of Chemical Compounds

Kristina Riemer Eric R. Scott Laura Meredith
S. Marshall Ledford

2024-06-04

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 avilable here: https://meredith-lab.github.io/volcalc/. The code for `volcalc` has been archived on Zenodo (Riemer, Scott, and Meredith 2023) and a methods paper has been published in *Frontiers in Microbiology* (Meredith et al. 2023).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 in the DESCRIPTION file of a Bioconductor dependency, an issue that is unfortunately out of our hands. As of today, this issue appears to have been fixed, but now checks on RHub are failing because of issues with the `RCurl` package. 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.

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 parsing errors from Open Babel are encountered, these error messages are displayed on the R console, but cannot be captured (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, the `InChI` value is *always* missing on Windows, so our validation code doesn't work on Windows. 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, and in talks. 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 that we have implemented in `volcalc`, one could specify the temperature at which to make these calculations (Pankow and Asher 2008). In `volcalc`, users currently can only estimate volatility at a fixed temperature of 20°C. Implementing a variable temperature would not be trivial, but would add a great deal of flexibility.

References

Meredith, Laura K., S. Marshall Ledford, Kristina Riemer, Parker Geffre, Kelsey Graves, Linnea K. Honeker, David LeBauer, Malak M. Tfaily, and Jordan Krechmer. 2023. "Automating Methods for Estimating Metabolite Volatility." *Frontiers in Microbiology* 14 (December). https://doi.org/10.3389/fmicb.2023.1267234.

Pankow, J F, and W E Asher. 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, Kristina, Eric R. Scott, and Laura K. Meredith. 2023. *Volcalc: Calculate Volatility of Chemical Compounds*. Zenodo. https://doi.org/10.5281/ZENODO.8015154.