

Tuesday, October 10, 2023

R Consortium Report

volcalc: Calculate Volatility of Chemical Compounds

Eric Scott
University of Arizona
ericscott@arizona.edu

Kristina Riemer
University of Arizona
kristinariemer@arizona.edu

Summary

This report includes progress updates on milestones 1 and 2 of our proposed work. The GitHub repository for the development version of `volcalc` can be found at: <https://github.com/Meredith-Lab/volcalc>.

We have completed milestone 1 to set up continuous integration, a package website, and automated test coverage tracking. Milestone 2 is near completion resulting in a fairly major re-factoring of the main functions in the package to allow `volcalc` to work not only with compounds in the KEGG database, but any compound that can be represented as a `.mol` file or a SMILES string.

Milestone 1: Setup phase

We've completed all of our goals for [milestone 1](#) including:

- Set up continuous integration with GitHub actions
- Track test coverage with codecov
- [Brainstorm package API/UI](#)

At the conclusion of milestone 1, we released [v1.0.0 of volcalc](#) on GitHub. We went on to release v1.0.1 and v1.0.2 with minor bug fixes in how calculations were done.

Currently the package is installable from GitHub and is also built on r-universe at <https://cct-datascience.r-universe.dev/volcalc>. The pkgdown website for the package is <https://meredith-lab.github.io/volcalc/>

Milestone 2: Refactoring

The main functions in `volcalc` underwent major re-factoring during milestone 2. Now, the main function `calc_vol()` is no longer tied to downloading compounds from the KEGG database and instead accepts a path to a `.mol` file or a SMILES string as input. Downloading `.mol` files from KEGG is still possible with the new `get_mol_kegg()` function. We also split out some of the code in `calc_vol()` into the separate function `simpol1()` to allow for easier extension to other methods in the future.

The full changelog can be viewed on the [release page](#).

There are currently two issues that we view as part of [milestone 2](#) still open at the time of writing this report:

- Allowing users to specify the environment (air or soil) for categorizing relative volatility into “low”, “medium”, and “high”
- Allowing users to specify the method of calculation as the original SIMPOL.1 method or the version modified by Meredith et al. This has an open [pull request](#) waiting for review.

After these last two elements are completed, we plan to make a v2.1.0 release. Then, we hope to distribute the package to a small group of chemoinformatics and metabolomics researchers to get feedback on the user interface, functionality, and documentation of the package.