

SAMPL5 Distribution Coefficient Instructions, v2

Updated October 27, 2015

Due date

Your predictions must be uploaded on the D3R SAMPL5 web-page by February 1, 2016. The experimental results will be released immediately after the challenge closes.

Anonymous versus public participation

When you signed up for this challenge, you were given the option of participating anonymously. Anonymous participation means that we may report on your predictions and methods, but not your identity. Public participations means we may also say who you are. Please note that, although we will work to protect the identity of anonymous participants, we cannot make any guarantees.

You may use the D3R website to change your anonymous/public status until the challenge has closed. However, after the challenge has closed, you may not change your anonymous/public status.

SAMPL5 workshop March 9-12

Participants are invited to share and discuss their results, as well as the D3R and SAMPL projects more broadly, at the first D3R and SAMPL workshop, which is scheduled for March 9-13, 2016, at UC San Diego, La Jolla, CA. Note that the workshop immediately precedes the ACS National Meeting in San Diego, whose theme is Computers in Chemistry.

Molecular systems

Water-cyclohexane *distribution* coefficients (LogD) at were measured for a variety of small molecules at pH 7.4. We initially provided a set of 95 compounds, but this was reduced to 53 compounds as the experimental data was curated for quality. More experimental details are given in the SAMPL5_MoleculesSimulations document.

Computational methods

You may use any method(s) you like to generate your predictions; e.g., MD with implicit or explicit solvent; quantum methods; docking; etc.

Files provided

All small molecules are provided in mol2 and SDfile formats in the MoleculeFiles directory. These were generated from the original isomeric SMILES strings, which

are provided (together with compound IDs of the form "SAMPL5_XXX") in the SMILES_by_ID.txt file in the same directory. Provided .mol2 and .sdf files were generated using the OpenEye toolkits, with OpenEye Omega for conformer generation, and AM1-BCC partial charges are provided.

There is no guarantee or representation that the protonation and tautomer states provided are optimal. It is also possible that the protonation state or tautomer of a molecule could change on transfer between environments.

The organizers plan to carry out a complete *reference* set of distribution coefficient calculations (typically for a single protonation state/tautomer of each molecule) by calculating solvation free energies in water and cyclohexane. These will be done in GROMACS, using GAFF/AM1-BCC force field parameters and TIP3P water. In case you are interested in using the same force field for your calculations, Kalli Burley (Mobley Lab, UC Irvine) has generated and provided a full set of GROMACS simulation files, and Dr. Michael Shirts (U Colorado), working with Ms. Burley, David Mobley (UC Irvine), and Christoph Klein (Vanderbilt U.) have generously converted the original GROMACS-format files into matched files for LAMPPS, AMBER and Desmond, and have checked for consistent energies across these software packages. We are also looking into the possibility of providing CHARMM-format files, but they are not available in the initial download. The following initial setups will be provided in the download package (though only GROMACS files are available in the initial download, the package will be updated shortly with the other formats):

- AMBER: .rst and .prmtop
- LAMPPS: .input and .lmp
- GROMACS: .gro and .top
- Desmond: .cms

Details of the energy comparison will shortly be provided in the files SAMPL5Energies.xls and .csv. In brief, for LAMPPS, GROMACS and Desmond, the energies for the starting conformation agree to within a relative error of approximately 10^{-6} , given matched nonbonded and other parameters. For AMBER, similar agreement is obtained for most terms, but there are discrepancies, with relative error approximately 10^{-5} , for energies that include an electrostatic contribution. These discrepancies appear to result from a deviation, relative to the other three programs, at the fifth significant digit of Coulomb's constant converting q_1q_2/r to energy in kcal/mol. We hope to fully clarify this situation soon.

An additional discussion of the purpose of the reference calculations is provided in the accompanying document -- in brief, these are designed primarily to allow easier separation of methodological differences from force field differences.

Standardization runs

If you are using MD with explicit solvent to compute distribution coefficients, we ask that you would at least apply your method to the standard setups (see above) for

the following two cases, so that we will all have a common point of comparison across packages and methods:

- SAMPL5_059
- SAMPL5_037

In these cases, regardless of how your planned approach in general, we would ask that you would calculate the distribution coefficients via solvation free energy calculations into the pure solvents (as in the provided input files) for a point of direct comparison with our standard calculations. Hopefully, we will all get the same results to within numerical precision. If not, we can work together to reconcile the discrepancies.

In our calculations, we will use a protocol very similar to that in Reference ¹ for calculating hydration free energies, and essentially the same protocol but in cyclohexane for solvation in cyclohexane (including collecting data for 5 ns at each lambda state). We will then estimate the distribution coefficient based on these calculated solvation free energies. Sample GROMACS .mdp files can be provided upon request.

Uploading your predictions

D3R is currently outfitting the SAMPL5 page with the ability to accept your uploaded predictions, and we are also preparing a template file to be used in submitting predictions. As soon as these are ready, you may upload your predictions. If you want to upload more than one set of predictions, generated by different methods, each set must be uploaded as a separate file. Please use the template provided, as the predictions will be parsed and analyzed with automated scripts.

A complete set of predictions constitutes predicted distribution coefficients for the full final set of SAMPL5 compounds (which may be reduced slightly from the initial set posted, in which case you will be notified), with predicted numerical uncertainties. Incomplete submissions - such as for a subset of compounds - will also be accepted, but will not necessarily be evaluated together with the rest of the submissions.

Note that distribution coefficients should be given as a ratio of concentrations (as will be the experimental values) rather than as a ratio of mole fractions.

Pending items, error reports, questions

¹ Mobley, D. L., Liu, S., Cerutti, D. S., Swope, W. C., & Rice, J. E. (2012). Alchemical prediction of hydration free energies for SAMPL. *Journal of Computer-Aided Molecular Design*, 26(5), 551–562. <http://doi.org/10.1007/s10822-011-9528-8>

We will email you during the challenge about the following pending items:

- Changes to the set as we finalize analysis of our experimental data
- Workshop details

Please feel free to contact us if you notice any errors in the information provided or have questions about SAMPL5: samplchallenge@gmail.com