Title
Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water

Journal
Journal of Chemical Theory and Computation, 12(8)

ISSN
1549-9618

Authors
Bannan, Caitlin C
Calabró, Gaetano
Kyu, Daisy Y
et al.

Publication Date
2016-08-09

DOI
10.1021/acs.jctc.6b00449

Calculating partition coefficients of small molecules in octanol/water and cyclohexane/water

Caitlin C. Bannan,[†] Gaetano Calabró,[‡] Daisy Y. Kyu,[‡] and David L. Mobley[*,‡,†]

[†]*Department of Chemistry, University of California, Irvine*

[‡]*Department of Pharmaceutical Sciences, University of California, Irvine*

E-mail: dmobley@mobleylab.org

Phone: 949-824-6383

Abstract

Partition coefficients describe how a solute is distributed between two immiscible solvents. They are used in drug design as a measure of a solute's hydrophobicity and a proxy for its membrane permeability. We calculate partition coefficients from transfer free energies using molecular dynamics simulations in explicit solvent. Setup is done by our new Solvation Toolkit which automates the process of creating input files for any combination of solutes and solvents for many popular molecular dynamics software packages. We calculate partition coefficients between octanol/water and cyclohexane/water with the Generalized AMBER Force Field (GAFF) and the Dielectric Corrected GAFF (GAFF-DC). With similar methods in the past we found a root-mean-squared error (RMSE) of 6.3 kJ/mol in hydration free energies which would correspond to an error of around 1.6 log units in partition coefficients if solvation free energies in both solvents were estimated with comparable accuracy. Here we find an overall

RMSE of about 1.2 log units with both force fields. Results from GAFF and GAFF-DC seem to exhibit systematic biases in opposite directions with GAFF and GAFF-DC for calculated cyclohexane/water partition coefficients.

Introduction

Partition coefficients provide a way to test the accuracy of atomistic force fields in various solvent environments. They describe the ratio of concentrations of a neutral solute molecule in a system with two immiscible solvents:

$$P = \frac{[solute]_{Organic}}{[solute]_{Aqueous}} \tag{1}$$

where *solute* refers to the neutral solute in both solvents and are typically reported as the logarithm of this concentration ratio ($\log P$).[1–3] This differs from a distribution coefficient or "apparent partition coefficient" which includes all ionized and unionized forms of the solute.[4] $\log P$ is proportional to the transfer free energy between the two solvents and can be related to the solvation free energies. Solvation free energies have been used to benchmark[5–14] and inform changes[15–17] to atomistic force fields, including the GROMOS 53A5 force field, which was parameterized in part using solvation free enthalpies in cyclohexane.[18] However, solvation free energies can be difficult to measure experimentally[10] compared to partition coefficients, which are relatively easy to measure and are measured routinely.[19] The access to experimental $\log P$ values and their straightforward relationship to the solvation free energy makes partition coefficients an excellent property to test and improve the accuracy of atomistic force fields in different solvent environments.

Due to the popularity of these values in the pharmaceutical industry, many tools already exist to predict partition coefficients between octanol and water ($\log P_{oct}$). Partition coefficients are used in the pharmaceutical industry to estimate how a drug may transfer between different biological environments[4,20] and are regularly used to predict a molecule's

hydrophobicity.[21] Additionally, many quantitative structure activity relationship (QSAR) methods use calculated $\log P_{oct}$ as an input parameter.[19] Methods for estimating partitioning are often based on additivity principles; for example, in 1964, Fujita et al. predicted $\log P_{oct}$ by determining the change in the partition coefficient when a functional group was added to a benzene, then using these functional group contributions to predict partition coefficients for new compounds.[22] Similar methods are still used today, where partition coefficients are calculated from contributions determined by molecular fragments[2,23–29] or atom types, characterized by element and bonding order.[30–36] In recent years, machine learning techniques have been developed using physical properties of the solute as parameters to predict $\log P_{oct}$.[19,37–47] However, very little work has been done to extend these methods to other organic solvents, so there are only a few examples of empirically trained methods for predicting cyclohexane/water partition coefficients.[48–51] These methods vary in accuracy and efficiency, but are all trained on experimental data, meaning that they have a limited domain of applicability. To some extent, more physical methods would be preferable, as these could be general enough to cover any organic solute-solvent combination, even combinations well outside the range of training data. One example of a relatively more physical approach is COSMO-RS, which predicts solvation and partitioning with a quantum mechanics-based approach combined with a variety of empirically-derived correction factors.[5,52,53]

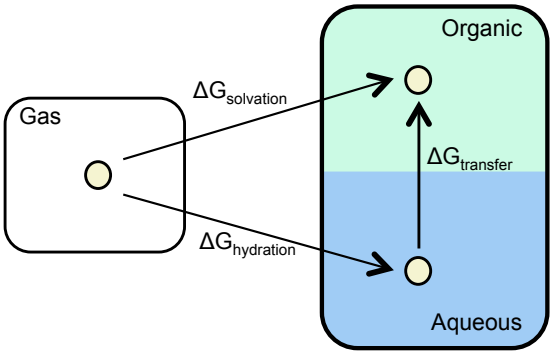

Figure 1: Alchemical thermodynamic cycle used to calculate transfer free energies from solvation free energies. Here we calculate $\Delta G_{Transfer}$ from the difference in $\Delta G_{solvation}$ into cyclohexane or octanol and $\Delta G_{hydration}$. $\log P$ is directly proportional to $\Delta G_{Transfer}$.

Here we take a physical approach to calculating partition coefficients from solvation free energies in octanol/water ($\log P_{oct}$) and cyclohexane/water ($\log P_{cyc}$) using alchemical methods to calculate the transfer free energy. Alchemical free energy calculations take a system through computationally accessible thermodynamic pathways to connect the target end states[54,55] – in this case, taking the solute out of one solvent and into the other. As discussed elsewhere, the logarithm of the partition coefficient is proportional to the transfer free energy between solvents,[56] which means that we can use standard free energy techniques to compute partition coefficients via an appropriate combination of solvation free energy calculations. Specifically, we calculate the transfer free energy as the difference between the hydration free energy and the solvation free energy into the organic phase (Figure 1). Similar methods for calculating hydration free energies have yielded fairly accurate results.,[6,11,13,14] as did a recent effort for predicting relative solubilities in a variety of organic solvents this approach.[57]

A variety of previous studies have calculated partition coefficients from solvation free energies. Early attempts to calculate $\log P$ based molecular dynamics simulations used free energy perturbation, changing the identity of the solute in both solvents to obtain octanol/water[56,58] and chloroform/water[59–61] relative partition coefficients (comparing partition of two different solutes). A number of attempts have been made to calculate $\log P$ from absolute solvation free energies with all atom force fields.[16,62–64] For example, two recent studies used a hydrid atomic level/coarse grained force field with the General AMBER Force Field (GAFF)[65] for solute parameters.,[66,67] the same force field we apply here. Our present study is the first attempt at $\log P$ calculations with the Dielectric Corrected General AMBER Force Field (GAFF-DC)[15] for such calculations, and we are not aware of any prior applications of alchemical techniques to the calculation of water-cyclohexane partition coefficients other than the relative work using free energy perturbation noted above. Another new feature of this work is that we provide an automated work flow for solvation free energies and therefore partition coefficient calculations, in that our new Solvation Toolkit was used to create all

input files and is available on GitHub at https://github.com/MobleyLab/SolvationToolkit, and analysis of the solvation free energies was fully automated with the Alchemical Analysis tool.[68] To facilitate additional applications of this technique, all setup and analysis scripts are also available in the Additional information, which is provided free on charge online at http://n2t.net/ark:/b7280/d15k5m

Computational Theory and Methods

Automated setup for arbitrary mixtures using Solvation Toolkit

Molecular modeling in different software packages requires a variety of input file types, terminal applications, and work flows. In an effort to automate our simulations, we have created Solvation Toolkit to generalize workflows across different software packages. It aims to deliver a simple tool to set mixtures of arbitrary combinations of solutes and solvents for use in popular molecular dynamics software packages such as GROMACS[69–75] and AMBER This piece of software relies on other software packages including OpenMolTools,[76,77] OpenEye's Python toolkits,[78–88] AmberTools,[89–94] Packmol,[95] ParmEd,[96] and MDTraj.[97]

The program can be logically divided into three main sections related to different tasks: converting simple input to molecular structures, generating force field parameters, and building a solvated box. These are followed by output to the proper file formats. The toolkit begins by requiring the IUPAC name or the SMILES string and the desired number of molecules for all compounds in the system. As recommended by OpenEye,[98] the OEChem toolkit is used to convert these into a molecular structure and generate up to 800 conformations of the molecule. Next, the Quacpac toolkit is used to automatically select the best conformation and to assign symmetric AM1-BCC charges.[88,99] OpenEye's toolkits were used rather than Antechamber as OpenEye's AM1-BCC implementation is maintained by Christopher Bayly, one of the original authors of AM1-BCC, and has several features beyond Antechamber's AM1-BCC, including better handling of multi-conformer molecules, proper

symmetrization of charges which ought to be symmetric, and others.[98] Using AMBER modules, GAFF parameters are assigned to the charged molecule with Antechamber and LEaP is used to produce a monomer topology file. Then the monomers for each compound in the mixture are assembled using Packmol to pack the different types of molecules into a box region defined by geometric constraints to keep atoms from different molecules at a safe pairwise distance.[95] Finally, LEaP is used to generate the final AMBER topology and coordinate files of the solvated mixture. In an optional extra step, ParmEd is used to convert the AMBER files into GROMACS files.

The SolvationToolkit package is released as open source to the scientific community. It can be download via github at github.com/mobleylab/SolvationToolkit and example files are provided to illustrate the main capabilities of the software package.

Theory for Computing Partition Coefficients

Partition coefficients can be estimated from solvation free energies derived from molecular dynamics simulations. We estimate $\log P$ directly from the transfer free energy of the solute moving from the aqueous to organic layer. The transfer free energy can be calculated as the difference in the solvation free energies obtained via the thermodynamic pathway presented in Figure 1. Therefore $\log P$ is directly proportional to a difference in the solvation free energies:

$$\log P = \frac{-\Delta G_{transfer}}{R\,T\ln(10)} = \frac{\Delta G_{hydration} - \Delta G_{solvation}}{R\,T\ln(10)} \tag{2}$$

Here R is the molar Boltzmann constant and T is the temperature, 298.15 K in this case. We will calculate the solvation free energy for each solute in all three solvents. The equation above can then be used to calculate $\log P_{oct}$ and $\log P_{cyc}$.

A few assumptions were made in order to simplify this estimation of the transfer free energy and partition coefficient. First, the two solvents were taken to be completely immiscible. Depending on the solubility of the organic solvent in water, this assumption may not

always hold. In theory, very polar solutes could carry water molecules with them across the solvent interface.[1] Separately calculating the solvation free energy in each solvent does not allow for this possibility, and – because we use pure solvents here – assumes immiscibility. Additionally, during solvation free energy calculations, the solutions were also taken to be at infinite dilution, meaning there will be one solute molecule in a bulk solution of solvent. When experimental measurements were taken at sufficiently low concentrations, calculated solvation free energies have been shown to have good agreement.[57,100] If experimental measurements of partition coefficients were made at concentrations deviating substantially from infinite dilution (i.e. where dimerization or oligomerization plays a role) this would be a potential source of error.

Selecting a Data Set

Experimental data was collected from the literature for both $\log P_{oct}$ (36 molecules) and $\log P_{cyc}$ (41 molecules). There are many sources for experimental data for $\log P_{oct}$ due to its use in the pharmaceutical industry. Albert Leo et al.'s collection of partition coefficients remains one of the larger collections of partition coefficients with a diverse set of organic solvents.[1] Initially, we assumed finding experimental octanol/water partition coefficients would be easy, as they are so regularly measured, therefore 41 solutes with cyclohexane partition coefficients provided were chosen from this collection. We made sure a diverse set of functional groups and a wide range of $\log P_{cyc}$ were represented in this set. In the literature, the solutes are labeled with their common names, which do not always include proper specification for stereochemistry. The chosen molecules only include those with unambiguous stereochemistry. The set of 41 molecules were chosen based on the available $\log P_{cyc}$ data, but ultimately we were only able to find experimental $\log P_{oct}$ for 36 of these compounds, so the octanol set studied here is a subset of the cyclohexane set. For $\log P_{oct}$ values not provided in Leo et al.,[1] other sources of experimental data were found.[21,101] The final set of molecules included a range of $\log P_{cyc}$ from -3.32 to 3.42 and $\log P_{oct}$ from -0.82 to 5.01.

Creating Input Files

In order to setup the simulations, OpenEye's OEChem and Omega[79] Python toolkits were used to convert the molecule names provided by Leo et al. into SMILES strings and IUPAC names. The SMILES strings for the solutes and solvents were used to build coordinate and topology files for use in GROMACS via the Solvation Toolkit described above. Recent studies have shown that calculated hydration free energies are independent of box size over a range of typical simulation box sizes (2-8 nm box edge).[102] Solvation free energy in cyclohexane is also independent of box size (2-4.5 nm box edge), as shown by members of the Mobley group in a recent study.[103,104] Here, the number of solvent molecules were chosen such that the box edge was around 3nm, sufficiently large for the GROMACS parameters specified below (100 octanol or 150 cyclohexane solvent molecules). As discussed above, the simulations were taken at infinite dilution, therefore only 1 solute molecule was added to each box.

SolvationToolkit uses OpenMolTools[76] Amber module to build monomer files. Originally, the OpenMolTools Amber module only represented water as flexible GAFF water, whereas we wanted to use TIP3P.[105] The monomer topology file for the solute was built according to the protocol in Solvation Toolkit described above. The final solvated mixture was built with OpenMolTools Gromacs module to make cubic boxes of the solute in water with at least 1.2nm between the solute and the nearest box edge. OpenMolTools version 0.7.0 forward uses TIP3P water in the Amber module so water input files can now be built directly with the SolvationToolkit.

Simulation Protocols with GROMACS

Generally protocols were taken from previous work with relative solubility calculations[57] and updated to work with GROMACS 5.0.6.[69–72] The alchemical solvation was broken into 20 lambda states. In the first 5 lambda states the electrostatic interactions between the solute and solvent were turned off. Then the Van der Waals interactions (modeled by Lennard Jones potentials) were switched off in the last 15 lambda states. Specifically, Lennard-Jones

interactions were scaled using λ = [0.0, 0.0, 0.0, 0.0, 0.0, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.0] and electrostatic interactions were scaled using λ = [0.0, 0.25, 0.50, 0.75, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]. Each state was minimized using GROMACS steepest decent algorithm and then equilibrated for a total of 150ps. The equilibration was broken into three steps: (1) 50ps constant volume, (2) 50ps constant pressure with the Berendsen barostat,[106] and (3) 50ps constant pressure with the Parrinello-Rahman barostat.[107] These were followed by a 5ns production phase at each λ, still using the Parrinello-Rahman barostat. The initial 100ps of the production stage was also removed to give the system extra time to reach equilibrium.

The new Hydroxynator used to setup calculations in GAFF-DC

We also recomputed solvation free energies using the dielectric corrected AMBER force field (GAFF-DC) parameters proposed by Fennell et al.[15] GAFF-DC implements changes to the Lennard-Jones parameters of hydroxyl group oxygens and scales the charges for all atoms in the hydroxyl group. These changes improve the accuracy of hydration free energy calculations[15] and to some extent relative solubilities.[57] Partition coefficients calculated from GAFF-DC solvation free energies provide a way to monitor how the accuracy of calculations with GAFF-DC compares with GAFF. The Hydroxynator tool changes the parameters of a topology file from GAFF to GAFF-DC.[15] It was applied to the topology files for only alcohols in water and cyclohexane and all solutes in octanol, as the solvent is an alcohol in that case. Since GAFF-DC only changes parameters around hydroxyl groups, applying Hydroxynator to the topology files for non-alcohols in water or cyclohexane would result in no change to the file. Following the protocol above, simulations were run in GROMACS with the GAFF-DC files. Results from these simulations will be labeled "GAFF-DC".

The initial implementation of Hydroxynator[15] could only handle topology files with one molecule containing a hydroxyl group. Up until now this was sufficient as GAFF-DC had only been tested on systems where the solute was the alcohol. However, as our octanol sys-

tems contain multiple molecules with hydroxyl groups, we rewrote Hydroxynator, relying on ParmEd (version 2.0.4) to read in a GAFF topology file and parse through each molecule to identify hydroxyl groups and adjust the parameters for those molecules. Like the original, this tool is open source and can be downloaded from https://github.com/MobleyLab/Hydroxynator.[108]

Analysis of Simulations and Results

The partition coefficients for both solvents ($\log P_{oct}$ and $\log P_{cyc}$) were calculated from the calculated solvation free energies. The free energy difference between each lambda value and the solvation free energy was calculated using the Multistate Bennett Acceptance Ratio (MBAR)[109] through the Alchemical Analysis tool.[68] As demonstrated above, the partition coefficients were calculated from the transfer free energy (eq. 2). A variety of error metrics were calculated in order to compare the calculated $\log P$ to experiment, including root-mean-squared error (RMSE), average signed error (ASE), Pearson's correlation coefficient (R), and the percent of calculated $\log P$ with the correct sign. Following established methods,[9] each metric was calculated for 1,000 bootstrap trials and the uncertainty was reported as the standard deviation from these results. In order to compare our results with an empirically trained method, we estimated $\log P_{oct}$ with the OpenEye OEXLogP tool.[31,78] The error metric analysis was repeated to compare the OEXLogP values with experiment. To examine the statistical difference between GAFF and GAFF-DC, a t-test was performed to compare GAFF and GAFF-DC values using methods available with the SciPy[110] Python module. Density convergence plots were all created for all GAFF simulations showing the cumulative average versus time. Uncertainty for each measurement was given by $\sigma = \frac{std}{\sqrt{\frac{N}{g}}}$ where std is the standard deviation in the density measurements, N is the number of measurements and g is the statistical inefficiency.[109]

Results and Discussion

We estimated cyclohexane/water and octanol/water partition coefficients from solvation free energies as described above (eq. 2) for a set of small molecules, with both standard GAFF parameters and the GAFF-DC parameter set. Our results are organized into four sets based on the organic solvent (cyclohexane or octanol) and force field (GAFF or GAFF-DC). Complete tables of our solvation free energies and partition coefficients, both calculated and experimental, are available in the supporting information.

While experimental values were reported without uncertainties, experimental measurements do involve uncertainty, raising questions about the typical uncertainty in the reported values. To assist with this, we found there were 17 molecules from our set with multiple experimental values reported,[1] and we used the difference in these measurements to estimate an average experimental uncertainty of about 0.3 log units.[a] In reality, some measurements are likely susceptible to higher errors than others, but without experimental uncertainty estimates or repeated measurements for all compounds, we have no way of assigning different uncertainties to different measurements. Therefore, the same uncertainty is used for all results analysis.

In general, we saw rather good agreement between the experimental and calculated partition coefficients (Figure 2). Considering all partition coefficients measured, we found a root-mean-square error (RMSE) of 1.2 ± 0.2 in GAFF and 1.2 ± 0.1 in GAFF-DC , a Pearson's correlation coefficient (R) of 0.8 ± 0.1 in both force fields and an average signed error of 0.5 ± 0.1 in GAFF and 0.2 ± 0.1 in GAFF-DC. A common metric for evaluating the accuracy of partition coefficients in octanol is to check that the sign of the $\log P$ for the calculated

[a]From the data provided in Leo et al. 13 of the compounds had at least 2 $\log P_{cyc}$ measurements: butanol (950), ethyl benzalcyanoacetate (4312), p-ethylphenol (2938), benzaldehyde (2133), 2-4-dimethylphenol (2914), o-methylphenol (2336), aniline (1711), p-methylphenol (2348), p-toluidine (2422), p-iodophenol (1425), 2-5-dimethylphenol (2917), m-methylphenol (2323), salicylic acid (2185). 8 of the compounds had at least 2 $\log P_{oct}$ measurements: o-toluidine (2411), butylamine (1014), p-methylphenol (2348), diethylamine (1028), p-toluidine (2422), o-nitrophenol (1453), m-methylphenol (2323), salicylic acid (2185)

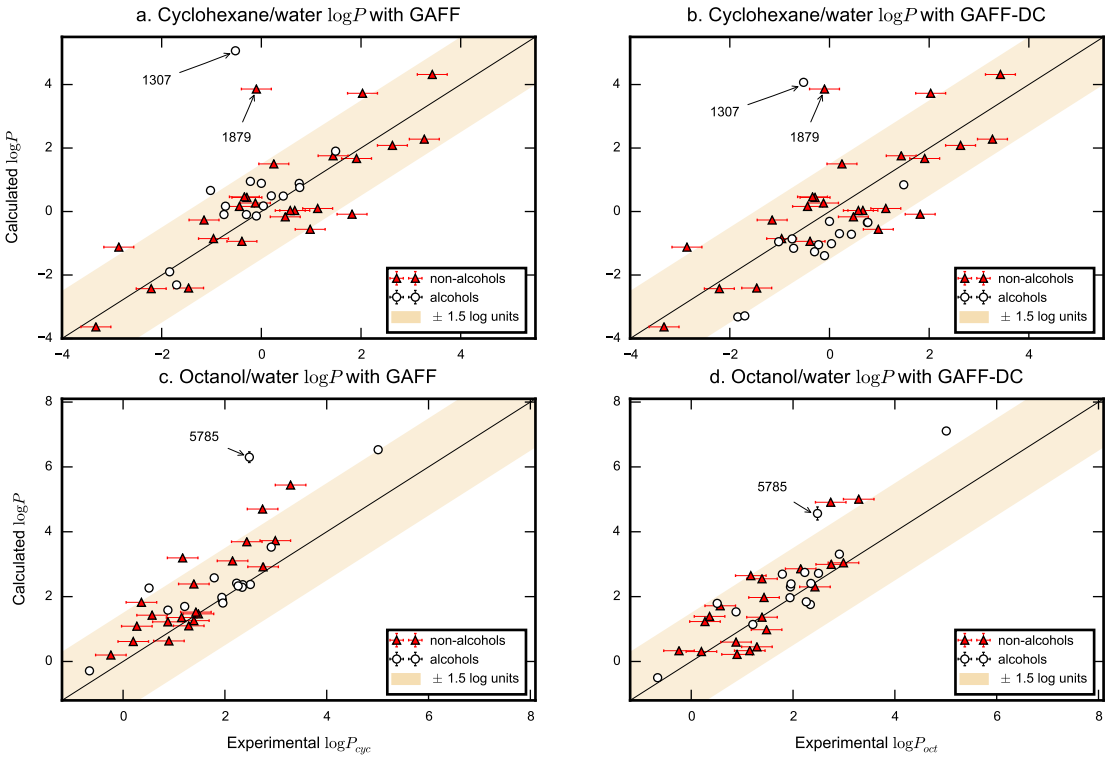

Figure 2: Comparison of the calculated $\log P$ to the experimental $\log P$. Data is divided into four sets, covering partition from water into both organic solvents (cyclohexane and octanol) for both force fields (GAFF and GAFF-DC). Alcohols are indicated separately since GAFF-DC affect the parameters for hydroxyl groups. The shaded region indicates where experimental and calculated $\log P$ agree within 1.5 log units, the predicted error given the accuracy of hydration free energy calculations in the FreeSolv Database.[111] Outliers greater than 3 log units are labeled by number.

and experimental data is the same.[67] The sign determines if the solute prefers the organic or aqueous layer. For all measured $\log P$, $82 \pm 4\%$ in GAFF and $79 \pm 5\%$ in GAFF-DC agreed by sign with the experimental data. These error metrics were computed for $\log P_{oct}$ and $\log P_{cyc}$ separately and for alcohol solute molecules as a subset (Table 1).

There were a few clear outliers in both solvents. We will focus on those with larger than 3 log unit difference between calculated and experimental values (Figure 2). In octanol, the calculated $\log P_{oct}$ for erythromycin (5785) was overestimated by 3.8 log units. The outliers in cyclohexane were pentachlorophenol (1307) by 5.6 log units and nitrocyclohexane (1879) by 3.9 log units. Past studies of hydration free energies have shown that atom-centered charges can be inadequate to describe the electrostatic potential around poly-

Table 1: Shown are the Pearson correlation coefficient (R), the root-mean-squared error (RMSE), average signed error (ASE), and percent correct sign for the comparison of calculated and experimental partition coefficients for both organic solvents and the selected force fields. Analysis is presented for the entire set of solutes (All) and separately for only hydroxyl containing solutes (Alcohols).

Solvent	**ForceField**	**Subset**	**R**	**RMSE**	**ASE**	**Correct Sign**
Cyclohexane	GAFF	All	0.7 ± 0.1	1.4 ± 0.3	0.4 ± 0.2	$68 \pm 7\%$
Cyclohexane	GAFF-DC	All	0.7 ± 0.1	1.3 ± 0.2	-0.1 ± 0.2	$63 \pm 7\%$
Octanol	GAFF	All	0.83 ± 0.07	1.1 ± 0.2	0.7 ± 0.1	$97 \pm 3\%$
Octanol	GAFF-DC	All	0.86 ± 0.06	0.9 ± 0.1	0.5 ± 0.1	$97 \pm 3\%$
Cyclohexane	GAFF	Alcohols	0.5 ± 0.3	1.5 ± 0.6	0.7 ± 0.3	$70 \pm 10\%$
Cyclohexane	GAFF-DC	Alcohols	0.5 ± 0.3	1.5 ± 0.4	-0.5 ± 0.4	$60 \pm 10\%$
Octanol	GAFF	Alcohols	0.8 ± 0.2	1.2 ± 0.4	0.6 ± 0.3	$100 \pm 1\%$
Octanol	GAFF-DC	Alcohols	0.9 ± 0.1	0.9 ± 0.2	0.5 ± 0.2	$100 \pm 1\%$

chlorinated compounds, likely explaining difficulty reproducing the partition coefficients for pentachlorophenol.[11] Erythromycin is a large molecule with many rotatable bonds and rings. It seems likely that our errors for erythromycin could be due to problems with conformational sampling, though another possible source of errors could be conformation-dependence of partial charges. To check the latter, we took many snapshots of erythromycin from our simulations and calculated partial charges for each conformation separately, then looked at the average and standard deviation in these charge sets for each atom. The standard deviation of all charge sets was orders of magnitude smaller than the average, thus our data does not support charge variation as the explanation for the error here. Instead, we suspect that the problem is conformational sampling, and that in order to accurately sample all of the possible configurations, longer simulation times or enhanced sampling methods may be needed. We decided to explore the possibility of sampling issues with erythromycin further, these results are discussed in depth below.

$\log P_{oct}$ values were also calculated with OpenEye's OEXLogP tool. These results are included in the Supporting Information tables and the scripts to used to calculate them are available with the additional information online. The OEXLogP results in an R value of 0.83 ± 0.09, RMSE of 0.6 ± 0.1, ASE of -0.1 ± 0.1, and correct sign of $94 \pm 4\%$. Clearly, OEXLogP more accurately estimates $\log P_{oct}$ than the values calculated with either force

field. This is perhaps not surprising given the wealth of experimental octanol partitioning data that can be used for training empirical methods like this one. However, there are a few examples where the $\log P_{oct}$ calculated by our alchemical solvation free energy method results were closer to experiment than OEXLogP. Most notably, salicyclic acid (2185) and o-nitrophenol (1453) where the OEXLogP under estimates $\log P_{oct}$ by 1.5 and 2.1 log units respectively. But given the small size of our present test set, we cannot draw meaningful predictions about whether and when alchemical methods will be more accurate in general than OEXLogP or other empirically trained methods.

Comparing results with GAFF and GAFF-DC to previous work with both force fields

Our results are about as accurate in GAFF as past work would predict, but significantly less accurate with GAFF-DC. The simulation parameters used here to calculate solvation free energies with GAFF are the same used by Mobley and collaborators for the calculations reported in the FreeSolv database.[111] FreeSolv includes the calculated and experimental hydration free energies for 643 neutral solutes. For the calculations reported in FreeSolv, there is an RMSE of 6.3 kJ/mol for the whole database. Assuming the solvation free energy calculations in each solvent (i.e. water and cyclohexane) would have a similar error, this would translate to an expected error of 1.6 log units in $\log P$. Fennell et al. also used these GAFF and GAFF-DC parameters when testing the development of GAFF-DC.[15] They found a similar RMSE for hydration free energy calculations with GAFF, but significantly better performance in GAFF-DC with an RMSE of only 1.9 kJ/mol, which would lead to an expected error of 0.47 log units for $\log P$. When we consider all our $\log P$ calculations in both solvents the RMSE is 1.2 log units for both force fields. Our calculated $\log P$ results are within uncertainty of what would be expected given the accuracy of hydration free energies in FreeSolv. Despite significant improvement in past hydration free energy calculations,[15] we see no significant improvement in $\log P$ values calculated with GAFF-DC compared to

GAFF (Table 1).

For solvation free energies in cyclohexane and water, only alcohols are affected by the change from GAFF to GAFF-DC, so they should be addressed as a subset of the data. Our results show a significant bias in the calculated $\log P_{cyc}$. Partition coefficients calculated with GAFF show alcohols prefer cyclohexane to water more strongly than in experiment as demonstrated by the average signed error (0.7 ± 0.3) and Figure 2a. The calculated $\log P_{cyc}$ in GAFF-DC show alcohols to prefer water over cyclohexane with an average signed error of -0.5 ± 0.3 (Figure 2b). Thus, GAFF-DC shifts the systematic error for alcohol $\log P_{cyc}$ from having one sign to having the other, as shown in Figure 2. GAFF-DC was parameterized to improve simulations in pure methanol, leading to more polarized hydroxyl groups.[15] It follows that the less polarized alcohols in GAFF might over-favor cyclohexane, but when more polarized would move to favoring water.

In octanol, however, there is no significant change in the accuracy of the calculated partition coefficients with GAFF compared to GAFF-DC (Figures 2c,d). The calculated $\log P_{oct}$ indicates all solutes prefer octanol over water more strongly than in experiment for both force fields as indicated by the average signed error (GAFF: 0.7 ± 0.1 and GAFF-DC: 0.5 ± 0.1). When only considering alcohol solutes there is still a preference for octanol over water with no significant change in the average signed error (GAFF: 0.6 ± 0.3 and GAFF-DC: 0.5 ± 0.2). The RMSE also decreases slightly for GAFF-DC with all solutes and alcohols, but the change is still within uncertainty (Table 1). T-tests comparing the $\log P_{oct}$ values calculated with GAFF and GAFF-DC resulted in $t = 2.4$ and $p = 0.022$ indicating the two sets of calculated values are significantly different, even though there is no statistically significant change in the overall error metrics. This significant difference appears to be consistent with a slight overall improvement in calculated values with GAFF-DC

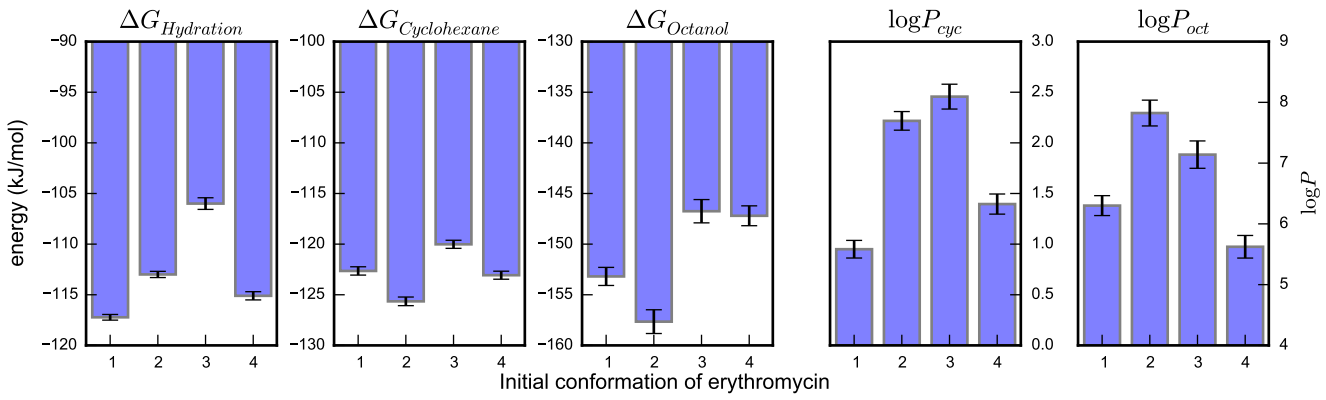

Figure 3: Comparison of calculated solvation free energies and partition coefficients for 4 different initial conformations of erythromycin (5785). If the conformational space was thoroughly sampled, changes in initial conformation would have no effect on the solvation free energy. Here we see significant changes in all three free energies, dramatically affecting calculated partition coefficients.

Sampling in largest solute molecule

Our solutes varied in size and flexibility from only 2 heavy atoms and no rotatable bonds (methanol, 159) to 51 heavy atoms and 7 rotatable bonds (erythromycin, 5785). As mentioned above, erythromycin was also an obvious outlier in octanol. For these reasons, we chose erythromycin as a test case to check how the initial solute conformation might affect solvation free energy and therefore $\log P$. Each new conformation is placed in an independent box of solvent molecules. If the conformational space of the solute is being thoroughly sampled and solvent sampling is adequate, then the solvation free energy should not be affected by a change in the initial conformation. We simulated erythromycin in each of the solvents three additional times each with a different initial conformation. Variations in solvation free energies of about 4 kJ/mol in each solvent led to variations of up to 1.6 log units in $\log P_{cyc}$ and 2.2 log units in $\log P_{oct}$ (Figure 3). These results suggest a significant dependence on the initial conformation of the erythromycin, which in turn points to issues in sampling the conformational space of the solute. Longer time scales or different sampling techniques will need to be applied for solutes as large and flexible as erythromycin.

(a) Density convergence of octanol with o-nitrophenol (1453) as solute

(b) Density convergence of octanol with 2,5-dimethylphenol (2917) as solute

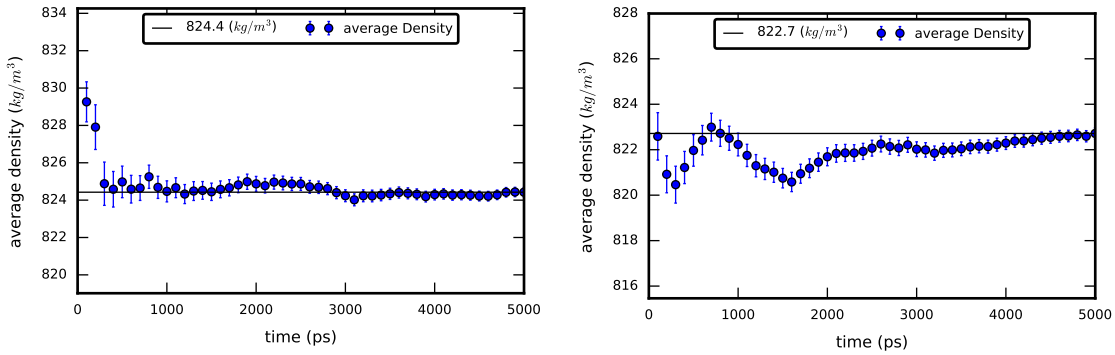

Figure 4: A cumulative average of the density during simulations of the solute in octanol. The overall average is indicated with a solid black line.

Examining convergence of solvation free energies in octanol

Octanol has a tendency to form clusters where the polar hydroxyl groups group together and separate from the non-polar carbon chains. These clusters have been suggested as a source of slow equilibration in simulations.[58,112] Convergence of physical properties can help indicate that a system has sufficiently equilibrated. To observe how well the simulations in octanol converged, we plotted average density versus simulation time (Figure 4). These plots were made with data from the production phase, meaning the system has already spent considerable time equilibrating. In a system that has converged, the average density should be within uncertainty of a constant value. In the case of o-nitrophenol (1453), the cumulative average density is within uncertainty of the overall average within the first 300 ps the density appears to be converging (Figure 4a). However in the case for 2,5-dimethylphenol (2917), the system does not get within uncertainty of the overall average until 4400 ps meaning the density may not have completely converged in the 5000 ps production phase (Figure 4b). In the solvation free energy calculations, we remove the first 100 ps of the production phase for additional equilibration, but it is included here. If the initial part of the production phase had not reached equilibrium, we might expect removing data from the beginning of the

simulation would decrease the amount of time it takes the density to apparently converge. To test if this was the case, we considered density convergence plots with the first 100, 300, and 1000 ps removed from the beginning of the data sets for 1453 and 2917. The plots from this analysis are available in the additional information, but there was no significant improvement in the apparent time to convergence when any amount of data was removed from the 2917 simulation.

These two systems are meant to represent the best and worst examples of simulations in octanol; plots for all other solutes are available in the additional information. We do not necessarily expect that the solute identity for these relatively small solutes dramatically impacts the convergence of the density of the entire system. Indeed, the density of the simulation with erythromycin, which has been shown to have sampling problems, converged more quickly than simulations with smaller solutes. Since each system is set up independently, the difference in convergence of the density for these different systems is likely the result of changes in initial conditions. If that is the case, it clearly indicates that octanol rearrangement is quite slow. As a follow-up, we performed three additional simulations for (4-bromo-benzal)-acetylacetone (4299) with larger box sizes (400 octanol molecules and 1 solute molecule) and fully interacting solvent ($\lambda = 0$). In one of these cases, the average density did not converge within the 5 ns production time, evidence that the slow rearrangement of octanol does not appear to be dependent on box size. The data and resulting convergence plots from these simulations are available in the additional information. The fact that the density converges quite slowly raises concerns about how well these simulations are converged. Thus, despite fairly accurate results for $\log P_{oct}$ there are still issues to be resolved to guarantee sufficient sampling when octanol is used as the solvent.

Conclusions

We did not initially expect accurate results for transfer energies into cyclohexane. Beauchamp et al. recently reported results for a benchmarking study for GAFF based on calculations of dielectric constants in pure solvents. They argued that GAFF would not accurately predict transfer free energies into cyclohexane because of an inaccurate representation of the solvent's dielectric constant.[77] They suggested transfer free energies from aqueous solutions into cyclohexane would have an error around -3.8 kJ/mol. Since partition coefficients are directly proportional to the transfer free energy, this translates to an error of +0.7 log units in $\log P_{cyc}$. While there is a slight bias for alcohols in cyclohexane, there is no obviously trend in the $\log P_{cyc}$ for all solutes (Figure 2a). While we do observe an overall average signed error in that same direction (0.4 ± 0.2), its magnitude suggests less bias than was anticipated for transfer free energies in cyclohexane. One possible reason for the lack of bias is Beauchamp et al. based their conclusions on calculations of the dielectric constant. This limits the free energy estimation to contributions from the electrostatic interactions. It is possible that Van der Waals interactions may play a role in counteracting a potential electrostatic bias introduced by the solvent dielectric constant. Given the diverse sizes and functional groups represented in this small set of molecules, we find GAFF's performance on cyclohexane partition coefficients to be surprisingly good.

As discussed in the methods section, a number of assumptions were made in order to perform these simulations. Some of these assumptions will need to be addressed in the future if partition coefficient calculations are going to be pursued. For one, we treated the solvents as completely immiscible. This approximation will be worse as the organic solvent becomes more polar and more miscible with water. For more polar organic solvents, water is more likely to permeate the organic/aqueous interface and may impact the accuracy of the calculated partition coefficient. Calculating solvation free energies into wet octanol, instead of the dry used here, may improve calculated $\log P_{oct}$ values. Additionally, there is some evidence to suggest polar solutes carry water molecules with them into the organic layer.[1]

The assumption of complete immiscibility does not allow for this possibility. One way to examine if a polar solute is carrying water across the interface would be to perform a set of simulations near the interface between the aqueous and organic layers. Umbrella sampling at various distances from the interface could then be used to monitor if the solute is pulling water into the organic layer with it.

In any study evaluating the accuracy of a computational model or method, the reliability of the experimental data must be considered. Partition coefficients are no different and a number of specific concerns have been well documented for experimental $\log P$ data. For example, a recent survey found that sources of partition coefficient measurements cited in some databases are not actually the original measurement of the reported value.[113,114] The miscibility of water and octanol can also pose problems with experimental measurements. Above we discussed how the clustering of water around polar solutes has been proposed as a concern. Especially in shake-flask techniques, octanol transferring into the aqueous phase can form small clusters around a non-polar solute to over-favor the aqueous phase.[113,115] With these concerns, a more non-polar solvent may be a better option for using partition coefficients to evaluate force fields. When comparing any computed physical property with experiment, experimental data must always be carefully curated to ensure high quality.

We propose that partition coefficients, given the ease and frequency with which they are measured experimentally, provide a new way to benchmark the accuracy of atomistic force fields. The results for $\log P_{cyc}$ in particular show continued biases for alcohol solutes, indicating these calculations may be helpful for further force field improvements. While our set contains relatively few solutes containing hydroxyl groups, there is still a clear bias for these results with both force fields. The calculated $\log P_{cyc}$ for alcohols overestimates the solute concentration in cyclohexane with GAFF and in water with GAFF-DC. GAFF-DC increases the polarization of the hydroxyl group. It was parameterized to correct the dielectric constant of pure liquid methanol. When initially proposed, Fennell et al. showed GAFF-DC significantly improves calculated hydration free energies for solutes with hydroxyl

groups.[15] However, a recent study calculated relative solubilities and did not see a significant change in the accuracy of results in GAFF-DC compared to GAFF.[57] We would expect GAFF-DC to improve calculated properties in environments where they would be naturally more polarized, such as water or other polar solvents. Cyclohexane would not be expected to increase the polarization in the hydroxyl group, so in our view GAFF-DC could actually decrease the accuracy of calculated properties in such an environment, though we do not observe a significant change in overall accuracy here. However, we do see that GAFF-DC results in a systematic error (averaged signed error) for alcohol partition coefficients between cyclohexane and water which has the opposite sign of that with GAFF and differs by more than one log unit. Essentially, it appears likely that because of the lack of polarization of cyclohexane, GAFF-DC results in over polarized solutes in cyclohexane. We do not see this same effect in octanol, possibly because GAFF-DC also polarizes octanol. We find that GAFF-DC outperforms standard GAFF for hydration free energies, but shows no significant improvement in overall accuracy for partition coefficients. This may indicate that we are encountering the limits of fixed charge force fields. Fixed charge force fields do not allow for changes in polarization as the environment around the solute molecule changes. A similar issue was encountered in GROMOS force field development, where it was found that no single charge set could adequately capture both hydration free enthalpies and the thermodynamics of pure organic liquids.[18] Overcoming this limitation will be a key step in improving the accuracy of such simulations.

The first goal of this project was to create an automated protocol for calculating partition coefficients from solvation free energies using GROMACS. By introducing Solvation Toolkit, automating this setup was successful. We plan to use these protocols to extend this small database of partition coefficients in the near future.

Additional Files Available Online

A database with all information, scripts, and files required to repeat the simulations and data analysis presented here is made available free of charge online at http://n2t.net/ark:/b7280/d15k5m . This includes GROMACS molecular dynamics parameter files, topology files, and coordinate files. It contains all results files: GROMACS files required to run Alchemical Analysis and the corresponding results for each solute in each solvent with both force fields, density data used to create convergence plots, convergence plots for each solute in each solvent, and GROMACS energy output files. It also has a directory of python scripts used to create, run, and store data for all simulations and electronically readable files with data on every molecule. There are multiple README documents explaining how the content is organized. Alchemical Analysis, Hydroxynator, and Solvation Toolkit are all available open source and are maintained on GitHub at http://github.com/MobleyLab.

Acknowledgement

All authors appreciate the financial support from the National Science Foundation (CHE 1352608) and the National Institutes of Health (1R01GM108889-01) and computing support from the UCI GreenPlanet cluster, supported in part by NSF Grant CHE-0840513. We thank Michael Shirts (University of Colorado), John Chodera (Memorial Sloan Kettering Cancer Center (MSKCC)), and Christopher Fennell (Oklahoma State) for their input. We particularly appreciate Kyle Beauchamp (currently Counsyl, previously MSKCC) for his work on preparing pure solvent simulations with OpenMolTools, which provided a starting point for our work on SolvationToolkit. We also thank present and past members of the Mobley lab for their support and sharing of knowledge, particularly Guilherme Matos and Shuai Liu (currently University of California, San Diego).

Supporting Information Available

One pdf file with the tables considered too long for the main text are available in the Supporting Information. Two tables with complete lists of calculated solvation free energies for all solutes in water, cyclohexane, and octanol with determined with both force fields (GAFF and GAFF-DC). $\log P_{cyc}$ and $\log P_{oct}$ tables with experimental and calculated values using GAFF and GAFF-DC. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525–616.

(2) Leo, A. J. *Chem. Rev.* **1993**, *93*, 1281–1306.

(3) Nernst, W. *Z. Phys. Chem.* **1891**, *8*, 110–139.

(4) Young, R. J.; Green, D. V. S.; Luscombe, C. N.; Hill, A. P. *Drug Discovery Today* **2011**, *16*, 822–830.

(5) Zhang, J.; Tuguldur, B.; van der Spoel, D. *J. Chem. Inf. Model.* **2015**, *55*, 1192–1201.

(6) Shivakumar, D.; Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919–930.

(7) Guthrie, J. P. *J. Phys. Chem. B* **2009**, *113*, 4501–4507.

(8) Villa, A.; Mark, A. E. *J. Comput. Chem.* **2002**, *23*, 548–553.

(9) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 135–150.

(10) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008**, *51*, 769–779.

(11) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 551–562.

(12) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. *J. Chem. Theory Comput.* **2012**, *8*, 2553–2558.

(13) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616–17626.

(14) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740–5761.

(15) Fennell, C. J.; Wymer, K. L.; Mobley, D. L. *J. Phys. Chem. B* **2014**, *118*, 6438–6446.

(16) Garrido, N. M.; Jorge, M.; Queimada, A. J.; Gomes, J. R. B.; Economou, I. G.; Macedo, E. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17384–17394.

(17) Knight, J. L.; Yesselman, J. D.; Brooks III, C. L. *J. Comput. Chem.* **2013**, *34*, 893–903.

(18) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(19) Liao, Q.; Yao, J.; Yuan, S. *Mol. Diversity* **2006**, *10*, 301–309.

(20) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug. Delivery Rev.* **2012**, *64*, 4–17.

(21) Sangster, J. *J. Phys. Chem. Ref. Data* **1989**, *18*, 1111–1227.

(22) Fujita, T.; Iwasa, J.; Hansch, C. *J. Am. Chem. Soc.* **1964**, *86*, 5175–5180.

(23) Klopman, G.; Wang, S. *J. Comput. Chem.* **1991**, *12*, 1025–1032.

(24) Hopfinger, A. J.; Esposito, E. X.; Llinas, A.; Glen, R. C.; Goodman, J. M. *J. Chem. Inf. Model.* **2009**, *49*, 1–5.

(25) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 2878–2888.

(26) Leo, A. *J. Chem. Soc., Perk. Trans. 2* **1983**, 825–838.

(27) Iwase, K.; Komatsu, K.; Hirono, S.; Nakagawa, S.; Moriguchi, I. *Chem. Pharm. Bull.* **1985**, *33*, 2114–2121.

(28) Rekker, R. F.; Mannhold, R. *Calculation of drug lipophilicity: the hydrophobic fragmental constant approach*; Wiley-VCH: Weinheim, 1992.

(29) Suzuki, T.; Kudo, Y. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 155–198.

(30) Wang, R.; Fu, Y.; Lai, L. *J. Chem. Inf. Model.* **1997**, *37*, 615–621.

(31) Wang, R.; Gao, Y.; Lai, L. *Perspect. Drug Discovery Des.* **2000**, *19*, 47–66.

(32) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. *Chem. Pharm. Bull.* **1992**, *40*, 127–130.

(33) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Phys. Chem. A* **1998**, *102*, 3762–3772.

(34) Klopman, G.; Iroff, L. D. *J. Comput. Chem.* **1981**, *2*, 157–160.

(35) Bodor, N.; Gabanyi, Z.; Wong, C. K. *J. Am. Chem. Soc.* **1989**, *111*, 3783–3786.

(36) Sun, H. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 748–757.

(37) Hilal, S. H.; Karickhoff, S. W.; Carreira, L. A. *QSAR Comb. Sci.* **2004**, *23*, 709–720.

(38) Xing, L.; Glen, R. C. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 796–805.

(39) Bodor, N.; Huang, M.-J. *J. Pharm. Sci.* **1992**, *81*, 272–281.

(40) Sasaki, Y.; Kubodera, H.; Matuszaki, T.; Umeyama, H. *J. Pharmacobio-Dyn.* **1991**, *14*, 207–214.

(41) Molnár, L.; Keserű, G. M.; Papp, Á.; Gulyás, Z.; Darvas, F. *Bioorg. Med. chem. Lett.* **2004**, *14*, 851–853.

(42) Beck, B.; Breindl, A.; Clark, T. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1046–1051.

(43) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. *J. Chem. Inf. Model.* **2002**, *34*, 752–781.

(44) Devillers, J.; Domine, D.; Guillon, C. *Eur. J. Med. Chem.* **1998**, *33*, 659–664.

(45) Chuman, H.; Mori, A.; Tanaka, H.; Yamagami, C.; Fujita, T. *J. Pharm. Sci.* **2004**, *93*, 2681–2697.

(46) Schnackenberg, L. K.; Beger, R. D. *J. Chem. Inf. Model.* **2005**, *45*, 360–365.

(47) In, Y.; Chai, H. H.; No, K. T. *J. Chem. Inf. Model.* **2005**, *45*, 254–263.

(48) Platts, J. A.; Abraham, M. H.; Burina, D.; Hersey, A. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 71–80.

(49) Khadikar, P. V.; Mandloi, D.; Bajaj, A. V.; Joshi, S. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 419–422.

(50) Zerara, M.; Brickmann, J.; Kretschmer, R.; Exner, T. E. *J. Comput.-Aided Mol. Des.* **2008**, *23*, 105–111.

(51) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. *J. Pharm. Sci.* **1994**, *83*, 1085–1100.

(52) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.

(53) Zhang, J.; Tuguldur, B.; van der Spoel, D. *J. Chem. Inf. Model.* **2016**, *56*, 819–820.

(54) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2010**, *31*, 1569–1582.

(55) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.

(56) Essex, J. W.; Reynolds, C. A.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 3634–3639.

(57) Liu, S.; Cao, S.; Hoang, K.; Young, K. L.; Paluch, A. S.; Mobley, D. L. *J. Chem. Theory Comput.* **2016**, *12*, 1930–1941.

(58) Best, S. A.; Merz Jr, K. M.; Reynolds, C. H. *J. Phys. Chem. B* **1999**, *103*, 714–726.

(59) Eksterowicz, J. E.; Miller, J. L.; Kollman, P. A. *J. Phys. Chem. B* **1997**, *101*, 10971–10975.

(60) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 187–189.

(61) Jorgensen, W. L.; Briggs, J. M.; Contreras, L. *J. Phys. Chem.* **1990**, *94*, 1683–1686.

(62) Garrido, N. M.; Queimada, A. J.; Jorge, M.; Macedo, E. A.; Economou, I. G. *J. Chem. Theory Comput.* **2009**, *5*, 2436–2446.

(63) Garrido, N. M.; Economou, I. G.; Queimada, A. J.; Jorge, M.; Macedo, E. A. *AIChE J.* **2012**, *58*, 1929–1938.

(64) Yang, L.; Ahmed, A.; Sandler, S. I. *J. Comput. Chem.* **2013**, *34*, 284–293.

(65) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(66) Michel, J.; Orsi, M.; Essex, J. W. *J. Phys. Chem. B* **2007**, *112*, 657–660.

(67) Genheden, S. *J. Chem. Theory Comput.* **2016**, *12*, 297–304.

(68) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 397–411.

(69) Berendsen, H. J. C.; Van Der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(70) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(71) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.

(72) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(73) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics (Oxford, England)* **2013**, *29*, 845–854.

(74) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. *Solving Software Challenges for Exascale*; Springer International Publishing: Stockholm, Sweden, 2014; Vol. 8759; pp 3–27.

(75) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1-2*, 19–25.

(76) Beauchamp, K. A.; Rustenburg, A. S.; Rizzi, A.; Behr, J. M.; Matos, G.; Wang, L.-P.; McGibbon, R. T.; Mobley, D. L.; Chodera, J. D. OpenMolTools. https://github.com/choderalab/openmoltools, (accessed September 30, 2015).

(77) Beauchamp, K. A.; Behr, J. M.; Rustenburg, A. S.; Bayly, C. I.; Kroenlein, K.; Chodera, J. D. *J. Phys. Chem. B* **2015**, *119*, 12912–12920.

(78) OEChem, version 1.7.4, OpeneEye Scientific Software, Inc. Santa Fe, NM, USA. http://www.eyesopen.com, 2010. (accessed June 16, 2015).

(79) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.

(80) Hawkins, P. C. D.; Nicholls, A. *J. Chem. Inf. Model.* **2012**, *52*, 2919–2936.

(81) Perola, E.; Charifson, P. S. *J. Med. Chem.* **2004**, *47*, 2499–2510.

(82) Kristam, R.; Gillet, V. J.; Lewis, R. A.; Thorner, D. *J. Chem. Inf. Model.* **2005**, *45*, 461–476.

(83) Stahl, M.; Mauser, H. *J. Chem. Inf. Model.* **2005**, *45*, 542–548.

(84) Marcou, G.; Rognan, D. *J. Chem. Inf. Model.* **2006**, *47*, 195–207.

(85) Cannon, E. O. *J. Chem. Inf. Model.* **2012**, *52*, 1124–1131.

(86) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357–1377.

(87) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.

(88) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(89) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER15*; University of California: San Francisco, 2015.

(90) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *WIREs Comput. Mol. Sci.* **2012**, *3*, 198–210.

(91) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(92) Cheatham III, T. E.; Case, D. A. *Biopolymers* **2013**, *99*, 969–977.

(93) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(94) Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003**, *66*, 27–85.

(95) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(96) Swails, J.; Hernandez, C.; Mobley, D. L.; Nguyen, H.; Wang, L.-P.; Janowski, P. Parmed. https://github.com/ParmEd/ParmEd, (accessed October 9, 2015).

(97) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. *Biophys. J.* **2015**, *109*, 1528–1532.

(98) Generating Canonical AM1-BCC Charges, 2015. OpenEye Scientific. https://docs.eyesopen.com/toolkits/cookbook/python/modeling/am1-bcc.html, (accessed June 6, 2016).

(99) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.

(100) Klimovich, P. V.; Mobley, D. L. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 307–316.

(101) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR.*; ACS Professional Reference Book; American Chemical Society: Washington, DC, 1995; Vol. 2.

(102) Parameswaran, S.; Mobley, D. L. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 825–829.

(103) Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L. **submitted for publication**,

(104) Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L. Analysis Scripts and Data for SAMPL5 Distribution Coefficients Submission Analysis and Reference Calculations, UC Irvine Collection. doi:10.7280/d1988w.

(105) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(106) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(107) Parrinello, M.; Rahman, A. *J. App. Phys.* **1981**, *52*, 7182–7190.

(108) Fennell, C. J. Personal Communication, 2015.

(109) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105–1–124105–10.

(110) Van Der Walt, S.; Colbert, S. C.; Varoquaux, G. *Comput. Sci. Eng.* **2011**, *13*, 22–30.

(111) Mobley, D. L.; Guthrie, J. P. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711–720.

(112) DeBolt, S. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5316–5340.

(113) Renner, R. *Environ. Sci. Technol.* **2002**, *36*, 410A–413A.

(114) Pontolillo, J.; Eganhouse, R. P. *The search for reliable aqueous solubility (Sw) and octanol-water partition coefficient (Kow) data for hydrophobic organic compounds: DDT and DDE as a case study*; US Department of the Interior, US Geological Survey: Reston, Virginia, 2001.

(115) De Bruijn, J.; Busser, F.; Seinen, W.; Hermens, J. *Environ. Toxicol. Chem.* **1989**, *8*, 499–512.

Graphical TOC Entry