

Predicting octanol/water partition coefficients using molecular simulation for the SAMPL7 challenge: comparing the use of neat and water saturated 1-octanol

Spencer J. Sabatino · Andrew S. Paluch

Submitted to ChemRxiv: May 3, 2021

Abstract Blind predictions of octanol/water partition coefficients at 298 K for 22 drug-like compounds were made for the SAMPL7 challenge. The octanol/water partition coefficients were predicted using solvation free energies computed using molecular dynamics simulations, wherein we considered the use of both pure and water-saturated 1-octanol to model the octanol-rich phase. Water and 1-octanol were modeled using TIP4P and TrAPPE-UA, respectively, which have been shown to well reproduce the experimental mutual solubility, and the solutes were modeled using GAFF. After the close of the SAMPL7 challenge, we additionally made predictions using TIP4P/2005 water. We found that the predictions were sensitive to the choice of water force field. However, the effect of water in the octanol-rich phase was found to be even more significant and non-negligible. The effect of inclusion of water was additionally sensitive to the chemical structure of the solute.

Keywords log P · partition coefficient · solvation free energy · SAMPL7

1 Introduction

The partitioning of a neutral solute between two liquid phases is important for a range of industrial, environmental, and biological processes. The resulting partition coefficient is of great fundamental significance as it can both readily be measured and can be related to the underlying intermolecular interactions. During its measurement,

Spencer J. Sabatino
Department of Chemical, Paper, and Biomedical Engineering
Miami University
Oxford, Ohio 45056, USA

Andrew S. Paluch
Department of Chemical, Paper, and Biomedical Engineering
Miami University
Oxford, Ohio 45056, USA
Tel.: (513) 529-0784
E-mail: PaluchAS@MiamiOH.edu

it is desired to keep the solute concentration dilute (less than 0.01 M) in each phase so that the solute may be assumed infinitely dilute [1,2]. In doing so, the partition coefficient of a solute (1) between two liquid phases at equilibrium (*I* and *II*, $P_1^{I/II}$) may be computed as [3]:

$$\begin{aligned} \log_{10} P_1^{I/II}(T,P) &= \log_{10} \frac{c_{1,I}}{c_{1,II}} = -\frac{\log_{10} e}{RT} \left[\Delta G_{1,I}^{\text{solv}}(T,P) - \Delta G_{1,II}^{\text{solv}}(T,P) \right] \\ &= -\frac{\log_{10} e}{RT} \Delta G_{1,II \rightarrow I}^{\text{tran}} \quad (1) \end{aligned}$$

where T and P correspond to the temperature and pressure, R is the molar gas constant, c_1^I and c_1^{II} are the molar or mass concentration (moles/volume or mass/volume) of the solute in phase *I* and *II*, respectively, $\Delta G_{1,I}^{\text{solv}}$ and $\Delta G_{1,II}^{\text{solv}}$ are the solvation free energy in phase *I* and *II*, respectively, and $\Delta G_{1,II \rightarrow I}^{\text{tran}}$ is the transfer free energy from phase *II* to *I*. In assuming the solute is infinitely dilute, the partition coefficient is a measure of the relative solute-solvent interactions, while minimizing any effect the solute may have on the liquid-liquid equilibrium of the two phases. For example, the partitioning of a solute between 1-octanol and water can be used to characterize the lipophilic/hydrophilic balance of a solute, and is an important parameter for biological systems [4,2,5]. For this reason, the prediction of octanol/water partition coefficients remains an important component of the SAMPL (Statistic Analysis of the Modeling of Proteins and Ligands) physical property prediction challenges [6].

For the case of octanol/water partition coefficients, phase *I* corresponds to the octanol-rich phase and the phase *II* corresponds to the water-rich phase. The mutual solubility of water in 1-octanol is 0.207 mole fracs, while the mutual solubility of 1-octanol in water is 0.703×10^{-4} mole fracs [7]. It is therefore reasonable to take the water-rich phase (phase *II*) to be pure water. On the other hand, the water in the octanol-rich phase (phase *I*) is not negligible. From eq. 1 this leads to:

$$\begin{aligned} \log_{10} P_1^{o^*/w}(T,P) &= \log_{10} \frac{c_{1,o^*}}{c_{1,w}} = -\frac{\log_{10} e}{RT} \left[\Delta G_{1,o^*}^{\text{solv}}(T,P) - \Delta G_{1,w}^{\text{solv}}(T,P) \right] \\ &= -\frac{\log_{10} e}{RT} \Delta G_{1,w \rightarrow o^*}^{\text{tran}} \quad (2) \end{aligned}$$

where the superscript “w” is used to indicate the phase taken to be pure water, and “o*” is used to indicate water-saturated octanol. A positive value of $\log_{10} P_1^{o^*/w}$ is indicative of a preference of the solute for the octanol-rich phase, while a negative value is indicative of a preference for water. The greater the absolute value, the greater the preference for the respective phase.

Experimentally, the measurement of octanol/water partition coefficients can be laborious. Tse and Sandler [8] therefore investigated the ability to determine octanol/water partition coefficients indirectly by instead measuring values of the limiting activity coefficient of the solute in each phase; the log limiting activity coefficient

and solvation free energy are directly related [3,9]. Furthermore, just as it was desirable to avoid the direct measurement of $P_1^{o*/w}$, it was desirable to avoid measurements with water-saturated octanol and instead use only pure (neat) solvents:

$$\log_{10} P_1^{o/w}(T,P) = -\frac{\log_{10} e}{RT} \left[\Delta G_{1,o}^{\text{solv}}(T,P) - \Delta G_{1,w}^{\text{solv}}(T,P) \right] \quad (3)$$

where “o” is used to indicate pure 1-octanol. For the set of organic pollutants studied, they found that there was a large difference between the limiting activity coefficient in pure water and pure 1-octanol, and in turn there was a large difference in the limiting activity coefficient in water-saturated 1-octanol and pure 1-octanol. While this led to a difference in the computed partition coefficient using neat and water-saturated 1-octanol, they did find that the partition coefficients were linearly correlated.

The liquid-liquid equilibrium of water and 1-octanol has been rigorously studied by Chen and Siepmann [10] using Monte Carlo simulations with advanced sampling techniques. In both pure and water-saturated 1-octanol, microscale heterogeneities exist consisting of polar and non-polar domains. In water-saturated 1-octanol, it was shown that considerably more large hydrogen bond aggregates (i.e., polar domains) exist, where a hydrogen bond aggregate is defined as a collection of molecules where every molecule shares at least one hydrogen bond with another molecule belonging to the same aggregate. For water-saturated 1-octanol, approximately 50% of the water and 1-octanol molecules belong to clusters with aggregation numbers greater than 25; this is approximately double that in pure 1-octanol. Additionally, Chen and Siepmann [10] rigorously computed the mutual solubility of water and 1-octanol and found that the results were highly sensitive to the force fields used. They found that their combination of TIP4P water [11] and TraPPE-UA 1-octanol [12,13] resulted in an equilibrium concentration of 0.21 mole fracs of water in the octanol rich-phase in good agreement with experiment. In earlier studies, it was found that a combination of TIP4P water and OPLS-UA 1-octanol resulted in an equilibrium concentration of 0.09 mole fracs [14], and SPC water with a the modified GROMOS96 force field for 1-octanol resulted in an equilibrium concentration of 0.16 mole fracs [15].

Significant progress has been made in the use of molecular simulation (Monte Carlo and molecular dynamics) to predict octanol/water partition coefficients [16–25, 6, 14]. In these efforts, the partition coefficient may be predicted with knowledge only of the structure of the solute. Moreover, these methods simultaneously may be used to understand the underlying molecular-level details. In general, the octanol/water partition coefficient is computed as the difference in solvation free energy in pure water and 1-octanol via eq. 3. While this has resulted in many accurate predictions, it physically is not in agreement with the experimental measurements. Specifically, the experimental measurements involve water-saturated 1-octanol. And following Tse and Sandler [8] we know that $\Delta G_{1,o*}^{\text{solv}} \neq \Delta G_{1,o}^{\text{solv}}$. Previous work has acknowledged this and attempted to model water-saturated 1-octanol using the experimental mutual solubility. However, it has been found that the mutual solubility predicted with common molecular models differ from experiment. If a simulation were performed at experimental conditions which corresponds to a mutual solubility greater than that predicted by the model, it would correspond to a metastable system. While the sys-

tem sizes and timescales are relatively small so that phase separation would not be observed, it would nonetheless correspond to a metastable system.

The most recent SAMPL6 challenge involved the 11 molecules in fig. 1 that resemble fragments of small molecule protein kinase inhibitors. The challenge organizers encourage participants to consider the effect of water-saturation on the predicted partition coefficients. The experimental $\log_{10} P_1^{o*/w}$ values were all positive, indicating a preference for the octanol-rich phase, with values ranging from 1.94–4.09. In general, it was found that the use of water-saturated 1-octanol instead of pure 1-octanol only slightly lowered the root mean squared error (RMSE) by 0.05 to 0.10 log units as compared to experiment. Methodological differences and the choice of force field were found to have a greater impact on the prediction accuracy than the composition of the 1-octanol phase [6].

For the SAMPL7 challenge participants were asked to make blind predictions of $\log_{10} P_1^{o*/w}$ for the 22 molecules in fig. 2 [26]. We submitted two sets of predictions which used solvation free energies computed using molecular dynamics simulations. In the first set, we treated the octanol-rich phase as pure 1-octanol, and in the second set of calculations we used water-saturated 1-octanol. Following the work of Chen and Siepmann [10], we used a combination of TIP4P water and TraPPE-UA 1-octanol, with their simulation determined mutual solubility of water in 1-octanol of 0.21 mole fracs. In this way we could be assured the simulation were performed at the equilibrium conditions of the model. Our goal was to understand the effect of water-saturation. Based on the structures of the molecules in SAMPL7 as compared to SAMPL6, we hypothesize that hydrogen bonding plays a more important role in their solvation. Knowing that water-saturated 1-octanol has larger hydrogen bond aggregates as compared to pure 1-octanol, we suspect the effect will be more pronounced in SAMPL7 as compared to SAMPL6. As compared to the experimental data provided at the close of the challenge, our predictions using pure 1-octanol had a RMSE of 1.08 and ranked 1/10 in the “Physical (MM)” category, while our predictions using water-saturated 1-octanol had a RMSE of 1.47 and ranked 6/10 in the Physical (MM) category. After the close of the challenge, we additionally repeated the calculations with TIP4P/2005 to look at the effect of the water model [27]. Overall, we agree that the accuracy of the predictions is sensitive to the choice of force field. However, we additionally find that the inclusion of water is not insignificant. While our predictions here using neat 1-octanol are in better quantitative agreement with experiment, we believe this is a fortuitous result. Moreover, we find that the effect of water saturation is dependent on the chemical structure of the solute.

2 Computational Details

2.1 Force Fields

Interactions were modeled using a “class I” potential energy function where all non-bonded intermolecular interactions (U_{nb}) were accounted for using a combined Lennard-Jones (LJ) plus fixed point charge model of the form [28, 29]

$$U_{\text{nb}}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (4)$$

where r_{ij} is the separation distance between sites i and j , ϵ_{ij} is the well-depth of the LJ potential, σ_{ij} is the distance at which the LJ potential is zero, and q_i and q_j are the partial charges of sites i and j , respectively.

Water was modeled with TIP4P [11] and 1-octanol was modeled with the United Atom Transferable Potentials for Phase Equilibria (TraPPE-UA) force field [12, 13]. The choice of solvent models was based on the work of Chen and Siepmann [10] who demonstrated the sensitivity of the mutual solubility of water and 1-octanol on the solvent models, and the good performance of TIP4P with TraPPE-UA. An additional set of calculations was performed with the TIP4P/2005 water model [27] for comparison after the conclusion of the SAMPL7 challenge.

The solutes were all modeled using the General AMBER Force Field version 2 (GAFF2) as implemented in the AMBER 20 simulation suite [30–32]. Parameters were generated using antechamber and converted from AMBER to GROMACS format using ParmEd. To generate partial charges for each of the SAMPL7 solute molecules, we first took the Daylight SMILES [33, 34] provided by the challenge organizers [26] and generated an initial 3-D structure with Open Babel 2.3.2 [35, 36]. Next, using Open Babel we performed a systematic conformation search to identify the lowest energy conformer followed by geometry optimization, all using the General Amber Force Field (GAFF) [31] with Gasteiger partial charges [37] as implemented in Open Babel. We then took the final structure for each compound from Open Babel and performed a geometry optimization in vacuum at the M06-2X/6-31+G** level of theory/basis set [38, 39] using Gaussian 16 [40]. This was followed by a single point energy calculation at the HF/6-31G* level of theory/basis with Gaussian 16 to generate an electrostatic potential from which partial charges were obtained using the restrained electrostatic potential (RESP) [41, 42] method in antechamber (within the AMBER 20 simulation suite). The use of RESP with HF/6-31G* is based on the recommendations of the original GAFF work [31].

All of the GROMACS force field files used in the present study are provided in the Supporting Information accompanying the electronic version of this manuscript.

2.2 Molecular Dynamics

Simulations were performed for each SAMPL7 solute molecule in water, 1-octanol, and water-saturated 1-octanol. In all cases we had a single solute molecule infinitely dilute in solution. The number of solvent molecules was chosen to obtain a cubic box with an edge length of approximately 4.5 nm at 298.15 K and 1 bar. The simulations in water and 1-octanol consisted of 3,000 and 350 molecules, respectively. With water-saturated octanol, we had 340 1-octanol molecules and 90 water molecules, resulting in a water mole fraction of 0.21. This was the equilibrium concentration of water in 1-octanol found for TIP4P water and TraPPE-UA 1-octanol by Chen and Siepmann [10]. After the close of the SAMPL7 challenge, calculations were additionally performed with the TIP4P/2005 water model for comparison to look at the

sensitivity of the water model. All of the simulations were carried out following the same procedure.

First, Packmol was used to generate initial structures [43,44]. This was followed by 3,000 steepest descent minimization steps to remove any bad contacts that might have resulted from the packing. The next two steps involved dynamics with the equations of motion integrated using the Verlet leap-frog algorithm [45,46,29,28]. The system was first equilibrated in an *NPT* ensemble at 298.15 K and 1 bar for 1 ns using the Berendsen thermostat and barostat [45–47]. This was followed by 4 ns of equilibration in an *NPT* ensemble at the same conditions using the stochastic velocity rescaling thermostat [45,48–50] and the Parrinello-Rahman barostat [51]. The final structure from this series of simulations was then used as the initial structure for our free energy calculations, as will be described momentarily. For systems involving SAMPL7 molecules SM30, SM33, SM34, SM36, SM37, and SM39, the final *NPT* equilibration was continued for an additional 100 ns which was used for subsequent structural analysis.

For all of the molecular dynamics simulations in this study, the simulations were performed using GROMACS 2020.2 [52–55]. All bond lengths in 1-octanol, and bond lengths involving a hydrogen for the SAMPL7 molecules were constrained using P-LINCS [45,56,57]. Water was modeled as completely rigid using the SETTLE algorithm [58,59]. The Verlet neighbor list was used [45] and LJ interactions were cut-off at 1.4 nm. Long-range analytic dispersion corrections were applied to the energy and pressure to accommodate the truncation [45,46,28,29]. Lorentz-Berthelot mixing rules were used for unlike LJ sites [28]. The electrostatic terms were evaluated with the smooth particle-mesh-Ewald method (SPME) with tin-foil boundary conditions [45,46,60] with real space interactions truncated at 1.4 nm. The SPME B-spline was order 4, the Fourier spacing was 0.12 nm, and the relative tolerance between long and short-range energies was 10^{-8} . The equations of motion were integrated with a timestep of 2 fs, the time constant for the thermostat was 1 ps and the time constant for the barostat was 4 ps.

Sample GROMACS input files are provided in the Supporting Information accompanying the electronic version of this manuscript.

2.2.1 Free Energy Calculations

The free energy calculations were performed at 298.15 K and 1 bar following our previous work [61–64]. The solvation free energy, ΔG_1^{solv} , for the solute infinitely dilute in water, 1-octanol, and water-saturated 1-octanol was calculated using a multi-stage free energy perturbation method [65–69] with the multi-state Bennett’s acceptance ratio method (MBAR) [70–73].

A “soft-core” potential was used to decouple the solute-solvent intermolecular LJ interactions. Stage (m) dependent decoupling parameters, λ_m^{LJ} and λ_m^{elec} controlled the LJ and electrostatic intermolecular interactions, respectively. The decoupling parameters varied from 0 to 1. When $\lambda_m^{\text{LJ}} = \lambda_m^{\text{elec}} = 1$, the solute is fully coupled to the system. When $\lambda_m^{\text{LJ}} = \lambda_m^{\text{elec}} = 0$, the solute is decoupled from the system. The “soft-core” potential had the form [74–76]

$$U_{\text{LJ}}^{\text{sc}}(r_{ij}; m) = 4\lambda_m^{\text{LJ}} \varepsilon_{ij} \left\{ \frac{\sigma_{ij}^{12}}{\left[(1 - \lambda_m^{\text{LJ}}) \alpha_{\text{LJ}} \sigma_{ij}^6 + r_{ij}^6 \right]^2} - \frac{\sigma_{ij}^6}{\left[(1 - \lambda_m^{\text{LJ}}) \alpha_{\text{LJ}} \sigma_{ij}^6 + r_{ij}^6 \right]} \right\} \quad (5)$$

where α_{LJ} is a constant, which had a value of 1/2. The advantage of using a “soft-core” potential to decouple the LJ interactions is that while it yields the correct limiting value of the potential (when $\lambda_m^{\text{LJ}} = 0$ and 1), it additionally allows nearly decoupled molecules to overlap with a finite energy (and hence finite probability). The electrostatic term in the intermolecular potential was decoupled linearly as

$$U_{\text{elec}}(r_{ij}; m) = \lambda_m^{\text{elec}} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (6)$$

At each stage m , an independent MD simulation was performed. The simulation time for each stage m was 17.5 ns, where the first 1.5 ns was discarded from analysis as equilibration. The change in the Hamiltonian with the current configuration between stage m and the other stages is computed every 0.20 ps. This is saved for subsequent post-simulation analysis with MBAR [73] to determine $\Delta G_1^{\text{sol}}v$. This analysis was performed using the Python implementation of MBAR (PyMBAR) and the GROMACS analysis script distributed with it [77–79].

A total of 15 different stages were used for the free energy calculations where $m = 0$ corresponds to a non-interacting (ideal gas) state and $m = 14$ is a fully interacting system. From $m = 1$ to 10 the LJ interactions were increased from $\lambda_m^{\text{LJ}} = 0.1$ to 1.0 in 10 equal increments of 0.1. Electrostatic interactions were increased in a square root fashion following $\lambda_m^{\text{elec}} = \{0.50, 0.71, 0.87, 1.00\}$ from $m = 11$ to 14 [80].

The simulation parameters for the free energy calculations were the same as the last step of equilibration except the equations of motion were integrated with the GROMACS “stochastic dynamics” integrator, corresponding to stochastic or velocity Langevin dynamics integrated with the leap-frog algorithm [45, 81, 46]. The time constant for the stochastic (or Langevin) thermostat was 1.0 ps. This change is necessary as a local thermostat is required to correctly control the temperature of a decoupled and weakly coupled solute molecule.

Sample GROMACS input files are provided in the Supporting Information accompanying the electronic version of this manuscript.

3 Results and discussion

3.1 Quantitative Predictions

The solvation free energy computed for each SAMPL7 molecule is summarized in tables S1, S2, and S3 of the Supporting Information for water, neat 1-octanol, and water-saturated 1-octanol, respectively. In all cases the (total) solvation free energy is broken down into its electrostatic ($m = 10$ to 14) and LJ (LJ, $m = 0$ to 10) contributions. The LJ contribution is taken as the change in free energy in going

from the non-interacting ideal gas state ($m = 0$) to fully interacting LJ interactions ($m = 10$). The electrostatic contribution is taken as the change in free energy of going from the state with fully interacting LJ interactions but no electrostatic interactions ($m = 10$) to fully interacting LJ and electrostatic interactions ($m = 14$). In tables 1 and 2 we summarize the computed octanol/water partition coefficient using both neat 1-octanol ($\log_{10} P_1^{o/w}$) and water-saturated 1-octanol ($\log_{10} P_1^{o*/w}$), using the TIP4P and TIP4P/2005 water models, respectively. The predictions are additionally broken down into their electrostatic and LJ contributions by using the respective contribution of the solvation free energy. In table 1 we additionally summarize the experimental values provided by the challenge organizers after the close of the challenge [26].

In figure 3 we provide a parity plot of the predicted versus experimental octanol/water partition coefficient. First, we find that for all cases the experimental data has $\log_{10} P_1^{o*/w} > 0$, ranging from 0.58 to 2.96. The positive values are indicative of a preference of the solute for the octanol-rich phase over water. Considering the use of TIP4P water, in all cases we predict the correct sign of the octanol/water partition coefficient in agreement with experiment. In general, we have a tendency to predict octanol/water partition coefficients that are too large, thereby over-predicting the affinity of the solute for the octanol-rich phase. Comparing the use of neat 1-octanol and water-saturated 1-octanol, the inclusion of water in general further increases the value of the octanol/water partition coefficient, thereby increasing the affinity of the solute for the octanol-rich phase. Only for the case of SM30 and SM33 does the inclusion of water decrease the value of the octanol/water partition coefficient. The average difference and the average absolute difference between $\log_{10} P_1^{o*/w}$ and $\log_{10} P_1^{o/w}$ is 0.56 and 0.69, respectively.

In figure 4 we plot the difference in $\log_{10} P_1^{o*/w}$ and $\log_{10} P_1^{o/w}$, along with the difference in its electrostatic and LJ contributions, for each SAMPL7 molecule. Considering again the use of TIP4P water, we first find that the difference in the LJ contribution is relatively small. The average difference and average absolute difference for the LJ contribution is -0.09 and 0.10 , respectively. The difference for SM26, SM31 and SM34 were positive while all others were negative. The difference in $\log_{10} P_1^{o*/w}$ and $\log_{10} P_1^{o/w}$ is dominated by the electrostatic contribution, for which the average difference and average absolute difference is 0.65 and 0.75 , respectively. The increase in the octanol/water partition coefficient corresponds to an increase in the affinity of the solute for the octanol-rich phase. This stems from a general decrease in $\Delta G_{1,o}^{\text{solv}}$ relative to $\Delta G_{1,o}^{\text{solv}}$, with the change dominated by electrostatic interactions.

The results using TIP4P/2005 are similar. However, we do predict $\log_{10} P_1^{o*/w} < 0$ for SM26. The average difference and the average absolute difference between $\log_{10} P_1^{o*/w}$ and $\log_{10} P_1^{o/w}$ is 0.49 and 0.53 , respectively, where again the inclusion of water in general increases the value of the octanol/water partition coefficient. The change is likewise dominated by electrostatic interactions for which the average difference and average absolute difference is 0.62 and 0.62 , respectively. The LJ contribution is smaller, for which the average difference and average absolute difference is -0.12 and 0.13 , respectively.

Next, let us compare our predicted octanol/water partition coefficients to the provided experimental data. For the competition, we made submissions using neat 1-octanol and water-saturated 1-octanol using TIP4P water. At the close of the challenge the challenge organizers analyzed the results and reported for our predictions using neat 1-octanol a RMSE of 1.08 which ranked 1/10 in the “Physical (MM)” category, while our predictions using water-saturated 1-octanol had a RMSE of 1.47 and ranked 6/10 in the Physical (MM) category. Computing independently here we find an average absolute error (AAE) and RMSE of 0.83 and 1.11 for neat 1-octanol, and an AAE and RMSE of 1.30 and 1.47 for water-saturated 1-octanol. Considering the average uncertainty in our predicted octanol/water partition coefficients is 0.08 and 0.10 using neat 1-octanol and water-saturated 1-octanol, respectively, the difference in AAE and RMSE of 0.48 and 0.37 is significant. Interestingly, while the use of water-saturated 1-octanol is a better representation of the physical system, we find that the predictions using neat 1-octanol are in better quantitative agreement with the experimental data. This is a fortuitous result.

The set of predictions using TIP4P/2005 were conducted after the close of the challenge to look at the effect of the water model. Comparing our predicted octanol/water partition coefficients to the provided experimental data, for neat 1-octanol we obtain an AAE and RMSE of 0.70 and 0.97, respectively. For water-saturated 1-octanol we obtain an AAE and RMSE of 0.97 and 1.20, respectively. We again find that our results using neat 1-octanol are in better quantitative agreement with the experimental data. Additionally, the computed error using TIP4P/2005 is less than using TIP4P. Using neat 1-octanol, the AAE and RMSE decreased by 0.13 and 0.14, respectively when using TIP4P/2005. Likewise, using water-saturated 1-octanol the AAE and RMSE decreased by 0.34 and 0.27, respectively, when using TIP4P/2005. The choice of force field is important and can be tuned to improve the accuracy of the predictions. However, here we find that the effect of the inclusion of water in the octanol-rich phase makes an even larger difference than the choice of water model.

For the SAMPL6 challenge which involved the 11 molecules in figure 1, challenge organizers encourage participants to consider the effect of water-saturation on the predicted partition coefficients. The experimental $\log_{10} P_1^{o*/w}$ values were all positive, indicating a preference for the octanol-rich phase, with values ranging from 1.94–4.09. This is larger than the SAMPL7 range of 0.58–2.96, indicating a greater preference of the SAMPL6 molecules for the octanol-rich phase. In general, it was found that the use of water-saturated 1-octanol instead of pure 1-octanol only slightly lowered the RMSE by 0.05 to 0.10 log units as compared to experiment. Methodological differences and the choice of force field were found to have a greater impact on the prediction accuracy than the composition of the 1-octanol phase [6]. Table 3 provides a set of reference predictions provided by the challenge organizers comparing the use of neat 1-octanol (*REF07*) and water-saturated 1-octanol (*REF02*) using the TIP3P water model and GAFF for 1-octanol and the solutes, along with the experimental values [6]. For this set of reference predictions, the average difference and the average absolute difference between $\log_{10} P_1^{o*/w}$ and $\log_{10} P_1^{o/w}$ is 0.03 and 0.25, respectively. This is noticeably smaller than observed here for SAMPL7. Additionally, for 3 of the 11 solutes we find that the inclusion of water in the octanol-rich phase de-

creased the predicted octanol/water partition coefficient, indicating a decreased affinity for the octanol-rich phase. For SM07, SM08, and SM12 the decrease is 0.17, 0.45 and 0.60, respectively.

Based on the structures of the molecules in SAMPL7 as compared to SAMPL6, we suspect that hydrogen bonding plays a more important role in the solvation of the SAMPL7 molecules. From the work of Chen and Siepmann [10] we know that water-saturated 1-octanol has larger hydrogen bond aggregates as compared to neat 1-octanol, which likely results in the larger effect of water in SAMPL7 as compared to SAMPL6. As a consequence, we find that effect of inclusion of water is dependent on the solute. While we agree that methodological differences and the choice of force field are very important, the result are also sensitive to the inclusion of water in the octanol-rich phase.

3.2 Structural Analysis

For the systems involving SAMPL7 molecules SM30, SM33, SM34, SM36, SM37, and SM39, additional 100 ns *NPT* simulations were performed to generated sufficient statistics for structural analysis to better understand the effect of inclusion of water in the octanol-rich phase; we only considered TIP4P water which was used in our SAMPL7 entries. Here we focus on SM33, SM36 and SM39 to highlight the role of water. Structurally, the three molecules differ only in the sulfur group in the four membered ring. For SM33, SM36, and SM39 we have a sulfide, sulfoxide, and sulfone group, respectively. (See figure 2). The difference between $\log_{10} P_1^{O^*/w}$ and $\log_{10} P_1^{O/w}$ is -0.87 , 2.05 , and 0.61 , for SM33, SM36, and SM39, respectively. For the case of SM36 this is the largest increase of all of the SAMPL7 molecules, and for SM33 this is the largest decrease of all of the SAMPL7 molecules. And interestingly SM39 falls in between close to the observed average change for the SAMPL7 set. The structural analyses were all performed using TRAVIS [82,83].

Figures 5, 6, and 7 plot the local density of the solvent O around the ring S, (non-ring) sulfone S, and N of the solute. These interactions were chosen to capture the effect of hydrogen bonding; the solvent O was chosen as there is only one O per solvent molecule. It is common to consider the radial distribution function, $g(r)$. Note that the local density and radial distribution function are related, with the radial distribution function equal to the local density normalized by the bulk density [28,29]. Comparing neat and water-saturated 1-octanol, the solvent bulk density of O increases by approximately 20%, and the normalization of the radial distribution function can hide the local change, motivating the use of local density here. When considering the case of water-saturated 1-octanol, we do not differentiate between the solvent O from water and 1-octanol.

Consider first the case of SM36 in figure 5. In water, the local density in the first solvation shell around the solute sulfone S, sulfoxide (ring) S, and amine N are all greater than bulk, with the largest density around the sulfoxide S. As a solvent, dimethyl sulfoxide (DMSO) is well known for its strong hydrogen bond accepting ability, so this result is not surprising. Consider next the case of neat 1-octanol. The local density in the first solvation shell around the solute sulfone S, sulfoxide (ring)

S, and amine N are again all greater than bulk. Here the largest density is around N, followed by the sulfoxide S and then the sulfone S. With the addition of water, in water-saturated 1-octanol the change in local density around N is insignificant, however, we notice a large change in local density around the sulfoxide and sulfone S, with the sulfoxide having the largest increase. We also point out the larger width of the peak around the sulfoxide S, which is indicative of the larger population of solvent O. We therefore find that the presence of water increases the local density of solvent O around the solute sulfone and sulfoxide S.

Compare this to the case of SM33 in figure 6. In water, the local density in the first solvation shell is largest for the sulfide (ring) S, closely followed by the sulfone S, and then the amine N is noticeably lower. In all cases the local density is lower than we found for SM36. Considering next the case of neat 1-octanol, we again find that the local density in the first solvation shell around the solute amine N and sulfone S are all greater than bulk, with the largest density around N, followed by sulfone S. The values are similar to those observed in neat 1-octanol for SM36. However, this is not the case for the sulfide (ring) S. The local density in the first solvation shell is very small and less than bulk. Moreover, we find that with the addition of water, in water-saturated 1-octanol, the change in local density in the first solvation shell is insignificant, and noticeably the local density around the sulfide (ring) S remains very small.

In summary, for the case of SM36 where we observe the largest increase between $\log_{10} P_1^{O*/w}$ and $\log_{10} P_1^{o/w}$, we observe a large increase in the local density of solvent O in the first solvation shell around the sulfoxide (ring) S and sulfone S in going from neat to water-saturated 1-octanol. On the other hand, for the case of SM33 where we observe the largest decrease between $\log_{10} P_1^{O*/w}$ and $\log_{10} P_1^{o/w}$, we observe that there is very little change in the local density of solvent O in the first solvation shell around the sulfide (ring) S and sulfone S in going from neat to water-saturated 1-octanol, and moreover that the local density around the sulfide (ring) S is very small. In figures S1 and S2 of the Supporting Information we provide spatial distribution functions (SDFs) for SM33 and SM36, respectively, in neat 1-octanol. We see the absence of the 1-octanol O near the (ring) sulfide S in SM33. In figures S3 and S4 of the Supporting Information we provide a snapshot of the system for SM33 and SM36, respectively, in neat 1-octanol. The snapshot likewise demonstrates the absence of the 1-octanol O near the (ring) sulfide S in SM33.

Interestingly, in both SM33 and SM36 the difference in local density in the first solvation shell around the amine N in neat 1-octanol and water-saturated 1-octanol is insignificant. We find that in neat 1-octanol, in both cases, the local density in the first solvation shell is larger than the ring S and sulfone S. The secondary amine N-H is the only solute hydrogen bond donating site, and the secondary amine N-H is known to be a moderate hydrogen bond donor and acceptor [84]. Possibly in neat 1-octanol the secondary amine is already "saturated" with hydrogen bonds, such that the local density is not impacted by the addition of water. For the case of SM36, the (ring) sulfoxide S=O is a moderate hydrogen bond acceptor [84]. But as compared to the secondary amine, it is more sterically free, and can better accommodate hydrogen

bonding with the solvent. We see this in the large halo of 1-octanol O around the sulfoxide in the SDF in figure S2 of the Supporting Information.

Last, in figure 7 we consider the case of SM39. For this case the ring sulfur is now in a sulfone group. The results are similar to SM36 in that in water-saturated 1-octanol the local density around both sulfone groups increases as compared to neat 1-octanol. However, the relative increase is less than that observed for SM36. While the value of $\log_{10} P_1^{O*/w}$ is greater than $\log_{10} P_1^{O/w}$ for both SM36 and SM39, the increase is greater for SM36 than SM39. Comparing the partial charges used by the force field for the ring sulfoxide group in SM36 and ring sulfone group in SM39, the O and S partial charges are -0.54 and $+0.29$, respectively for sulfoxide, and -0.59 and 1.06 , respectively, for sulfone. While the O carries a similar partial charge in both cases, the local O=S dipole will be larger in the sulfoxide group than each of the O=S dipoles in sulfone. As a result, the sulfoxide group is a stronger hydrogen bond acceptor.

4 Conclusion

In the present study we made blind predictions of the octanol/water partition coefficient for the 22 molecules in fig. 2 for the SAMPL7 challenge, wherein we treated the octanol-rich phase as pure 1-octanol and water-saturated 1-octanol. We used a combination of TIP4P water and TraPPE-UA 1-octanol which were shown by Chen and Siepmann [10] to well reproduce the experimental mutual solubility. As compared to the experimental data provided at the close of the challenge, our predictions using pure 1-octanol had a RMSE of 1.08 and ranked 1/10 in the "Physical (MM)" category, while our predictions using water-saturated 1-octanol had a RMSE of 1.47 and ranked 6/10 in the Physical (MM) category. After the close of the challenge, we additionally repeated the calculations with TIP4P/2005 to look at the effect of the water model. With TIP4P/2005, for neat 1-octanol we obtained a RMSE of 0.97 and for water-saturated 1-octanol we obtained a RMSE of 1.20. Similar to the findings of the SAMPL6 challenge, we find that the predictions are sensitive to the choice of force fields [6]. However, here we find that the effect of water in the octanol-rich phase is not negligible. Additionally, we find that the effect of inclusion of water is dependent on the chemical structure of the solute. As compared to the solute molecules in the SAMPL6 challenge (see fig. 1), we expect hydrogen bonding to be more important with the SAMPL7 solute molecules. As demonstrated by Chen and Siepmann [10], as compared to pure 1-octanol, in water-saturated 1-octanol considerably more large hydrogen bond aggregates exist. Looking at the local solvent density around the solute molecules we were able to see the effect.

While our predictions here using neat 1-octanol are in better quantitative agreement with experiment, we believe this is a fortuitous result. The use of water-saturated 1-octanol is a better representation of the physical system of interest. In future SAMPL challenges we will consider the sensitivity of the employed solute force field. Additional studies are also needed to examine the predicted mutual solubility of other common force fields for water and 1-octanol, to allow for the consideration of additional solvent models.

Acknowledgements Computing support was provided by the Ohio Supercomputer Center [85]. We appreciate the National Institutes of Health for its support of the SAMPL project via R01GM124270 to David L. Mobley (UC Irvine).

Conflict of interest

The authors declare that they have no conflict of interest.

Table 1 Summary of the computed octanol/water partition coefficient using the TIP4P water model. Predictions are made using both neat 1-octanol ($\log_{10} P_1^{o/w}$) and water-saturated 1-octanol ($\log_{10} P_1^{os/w}$). The predictions are additionally broken down into their electrostatic and Lennard-Jones contributions by using the respective contribution of the solvation free energy. In the last column we provide the experimental values provided by the challenge organizers after the close of the challenge. The subscript corresponds to the error in the last two decimal places.

SAMPL7 molecule	TIP4P						$\log_{10} P_1^{os/w}$ experiment
	$\log_{10} P_1^{o/w}$			$\log_{10} P_1^{os/w}$			
	electrostatic	LJ	total	electrostatic	LJ	total	
SM25	-7.67 ₀₂	10.64 ₀₅	2.97 ₀₅	-6.48 ₀₄	10.59 ₀₅	4.10 ₀₆	2.67 ₀₁
SM26	-8.53 ₀₄	8.66 ₀₄	0.13 ₀₆	-7.02 ₁₃	8.67 ₀₄	1.65 ₁₃	1.04 ₀₁
SM27	-7.94 ₀₅	9.83 ₀₅	1.89 ₀₇	-7.35 ₁₄	9.65 ₀₅	2.30 ₁₅	1.56 ₁₁
SM28	-8.81 ₁₅	10.22 ₀₄	1.41 ₁₆	-8.22 ₀₅	10.21 ₀₅	2.00 ₀₇	1.18 ₀₈
SM29	-8.07 ₀₅	9.67 ₀₄	1.61 ₀₇	-7.29 ₁₅	9.65 ₀₅	2.36 ₁₅	1.61 ₀₃
SM30	-8.18 ₀₆	11.86 ₀₅	3.67 ₀₈	-8.65 ₀₆	11.74 ₀₅	3.09 ₀₈	2.76 ₁₉
SM31	-8.58 ₀₆	10.57 ₀₅	1.99 ₀₈	-7.75 ₀₆	10.58 ₀₅	2.83 ₀₈	1.96 ₁₄
SM32	-6.94 ₀₅	9.98 ₀₅	3.04 ₀₇	-6.83 ₀₄	9.92 ₀₅	3.09 ₀₆	2.44 ₁₇
SM33	-6.91 ₀₃	12.43 ₀₆	5.52 ₀₆	-7.50 ₀₃	12.15 ₀₆	4.65 ₀₆	2.96 ₂₁
SM34	-6.45 ₀₈	10.95 ₀₅	4.49 ₁₀	-5.88 ₀₈	11.02 ₀₅	5.14 ₁₀	2.83 ₂₀
SM35	-8.98 ₀₅	10.45 ₀₅	1.47 ₀₇	-8.28 ₁₁	10.27 ₀₅	2.00 ₁₂	0.88 ₀₂
SM36	-11.45 ₀₆	12.96 ₀₆	1.51 ₀₈	-9.34 ₀₈	12.90 ₀₆	3.56 ₁₀	0.76 ₀₅
SM37	-10.05 ₀₅	11.51 ₀₅	1.47 ₀₇	-9.39 ₁₅	11.45 ₀₅	2.05 ₁₆	1.45 ₁₀
SM38	-9.70 ₀₆	10.71 ₀₅	1.01 ₀₈	-8.77 ₀₆	10.60 ₀₅	1.82 ₀₈	1.03 ₀₇
SM39	-10.29 ₀₈	13.08 ₀₆	2.79 ₁₀	-9.50 ₀₇	12.90 ₀₆	3.40 ₀₉	1.89 ₁₃
SM40	-9.95 ₀₆	11.94 ₀₆	1.99 ₀₈	-9.02 ₁₄	11.74 ₀₅	2.72 ₁₅	1.83 ₀₅
SM41	-6.35 ₀₃	9.15 ₀₄	2.80 ₀₅	-6.00 ₁₁	9.05 ₀₄	3.05 ₁₂	0.58 ₀₂
SM42	-8.13 ₁₃	11.70 ₀₄	3.57 ₁₄	-7.71 ₀₇	11.47 ₀₄	3.76 ₀₈	1.76 ₀₃
SM43	-7.67 ₀₂	10.41 ₀₄	2.74 ₀₄	-7.16 ₁₁	10.36 ₀₄	3.21 ₁₂	0.85 ₀₁
SM44	-7.54 ₀₂	9.19 ₀₄	1.65 ₀₄	-6.42 ₁₂	9.09 ₀₄	2.67 ₁₃	1.16 ₀₃
SM45	-8.00 ₀₄	11.53 ₀₅	3.53 ₀₆	-7.88 ₁₃	11.45 ₀₄	3.57 ₁₄	2.55 ₀₄
SM46	-7.76 ₀₈	10.34 ₀₄	2.58 ₀₉	-7.18 ₀₂	10.31 ₀₄	3.12 ₀₄	1.72 ₀₁

Table 2 Summary of the computed octanol/water partition coefficient using the TIP4P/2005 water model. Predictions are made using both neat 1-octanol ($\log_{10} P_1^{o/w}$) and water-saturated 1-octanol ($\log_{10} P_1^{o*/w}$). The predictions are additionally broken down into their electrostatic and Lennard-Jones contributions by using the respective contribution of the solvation free energy. The subscript corresponds to the error in the last two decimal places.

SAMPL7 molecule	TIP4P/2005					
	$\log_{10} P_1^{o/w}$			$\log_{10} P_1^{o*/w}$		
	electrostatic	LJ	total	electrostatic	LJ	total
SM25	-7.73 ₀₅	10.43 ₀₆	2.70 ₀₇	-7.18 ₀₅	10.33 ₀₆	3.15 ₀₇
SM26	-8.68 ₀₄	8.36 ₀₅	-0.32 ₀₆	-6.97 ₁₃	8.28 ₀₅	1.32 ₁₄
SM27	-7.86 ₀₅	9.46 ₀₅	1.60 ₀₇	-7.59 ₀₅	9.34 ₀₅	1.75 ₀₇
SM28	-8.67 ₁₅	9.86 ₀₅	1.19 ₁₆	-8.02 ₀₄	9.86 ₀₆	1.84 ₀₇
SM29	-8.23 ₀₅	9.31 ₀₅	1.08 ₀₇	-7.96 ₀₅	9.21 ₀₅	1.24 ₀₇
SM30	-7.98 ₀₆	11.59 ₀₇	3.61 ₀₉	-7.65 ₀₆	11.37 ₀₇	3.72 ₀₉
SM31	-8.62 ₀₇	10.39 ₀₆	1.77 ₀₉	-7.65 ₀₇	10.27 ₀₆	2.62 ₀₉
SM32	-6.92 ₀₅	9.64 ₀₅	2.73 ₀₇	-6.28 ₁₂	9.66 ₀₆	3.37 ₁₃
SM33	-7.05 ₀₄	12.22 ₀₇	5.16 ₀₈	-6.35 ₀₄	12.02 ₀₇	5.67 ₀₈
SM34	-6.51 ₀₂	10.84 ₀₆	4.34 ₀₆	-6.54 ₀₁	10.65 ₀₆	4.11 ₀₆
SM35	-9.23 ₀₅	10.09 ₀₅	0.86 ₀₇	-8.35 ₀₅	10.01 ₀₅	1.66 ₀₇
SM36	-11.63 ₀₆	12.50 ₀₇	0.87 ₀₉	-10.08 ₀₇	12.15 ₀₇	2.07 ₁₀
SM37	-10.03 ₀₅	11.13 ₀₆	1.10 ₀₈	-9.00 ₀₅	11.10 ₀₆	2.10 ₀₈
SM38	-9.69 ₀₉	10.17 ₀₆	0.48 ₁₁	-9.61 ₀₉	10.15 ₀₆	0.54 ₁₁
SM39	-10.18 ₀₉	12.62 ₀₈	2.45 ₁₁	-10.21 ₀₇	12.49 ₀₇	2.28 ₁₀
SM40	-10.01 ₀₅	11.52 ₀₇	1.52 ₀₈	-9.24 ₁₅	11.23 ₀₇	1.99 ₁₇
SM41	-6.25 ₀₅	8.68 ₀₄	2.43 ₀₆	-5.69 ₀₆	8.56 ₀₄	2.88 ₀₇
SM42	-7.79 ₁₄	11.24 ₀₅	3.45 ₁₅	-7.39 ₀₇	11.08 ₀₅	3.69 ₀₉
SM43	-7.44 ₀₂	9.90 ₀₅	2.47 ₀₅	-6.62 ₀₆	9.81 ₀₅	3.19 ₀₈
SM44	-7.32 ₀₂	8.66 ₀₄	1.34 ₀₅	-6.54 ₁₂	8.52 ₀₄	1.98 ₁₃
SM45	-7.81 ₀₅	11.02 ₀₅	3.21 ₀₇	-7.33 ₀₅	10.88 ₀₆	3.55 ₀₇
SM46	-7.57 ₀₉	9.79 ₀₅	2.21 ₁₀	-7.40 ₀₃	9.71 ₀₅	2.32 ₀₆

Table 3 A summary of the set of reference calculations performed by the SAMPL6 challenge organizers comparing the use of neat 1-octanol (*REF07*, $\log_{10} P_1^{o/w}$) and water-saturated 1-octanol (*REF02*, $\log_{10} P_1^{o*/w}$) using the TIP3P water model and GAFF for 1-octanol and the solutes, along with the experimental values [6]. The subscript corresponds to the error in the last two decimal places.

SAMPL6 molecule	TIP3P/GAFF		experiment
	$\log_{10} P_1^{o/w}$	$\log_{10} P_1^{o*/w}$	$\log_{10} P_1^{o*/w}$
SM02	5.54 ₀₈	5.86 ₁₈	4.09 ₀₃
SM04	5.16 ₀₆	5.18 ₁₈	3.98 ₀₃
SM07	4.21 ₂₁	4.04 ₂₈	3.21 ₀₄
SM08	9.82 ₀₈	9.37 ₁₄	3.10 ₀₃
SM09	4.51 ₀₃	4.63 ₀₇	3.03 ₀₇
SM11	2.54 ₀₇	2.94 ₀₄	2.10 ₀₄
SM12	5.59 ₀₆	4.99 ₀₄	3.83 ₀₃
SM13	5.29 ₁₂	5.33 ₀₆	2.92 ₀₄
SM14	2.42 ₁₁	2.57 ₁₂	1.95 ₀₃
SM15	3.10 ₀₈	3.31 ₁₄	3.07 ₀₃
SM16	3.88 ₀₆	4.14 ₂₉	2.62 ₀₁

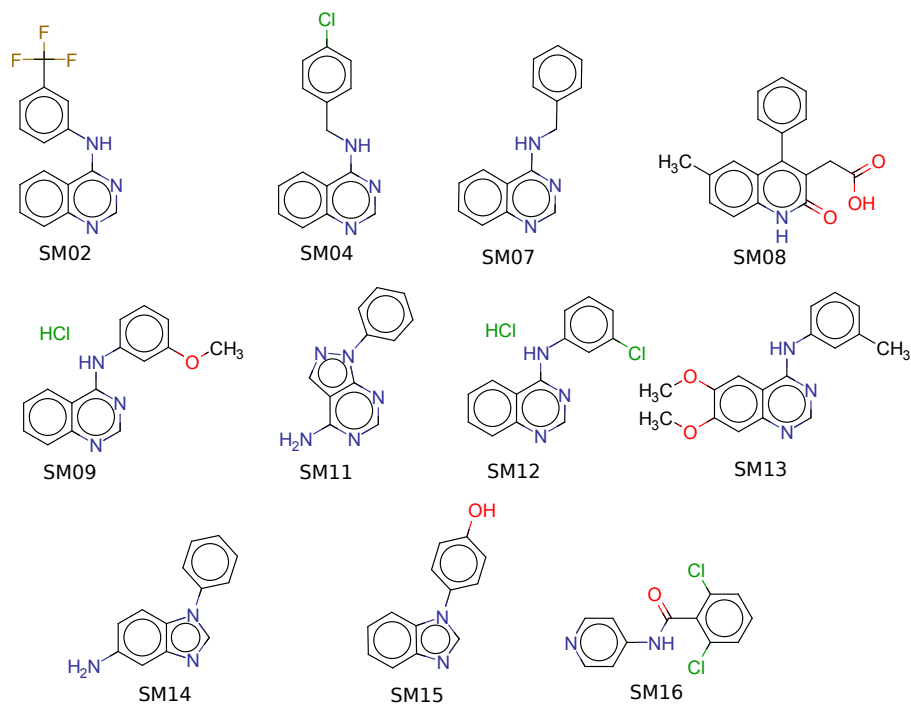


Fig. 1 Chemical structure and name of the molecules from the SAMPL6 challenge.

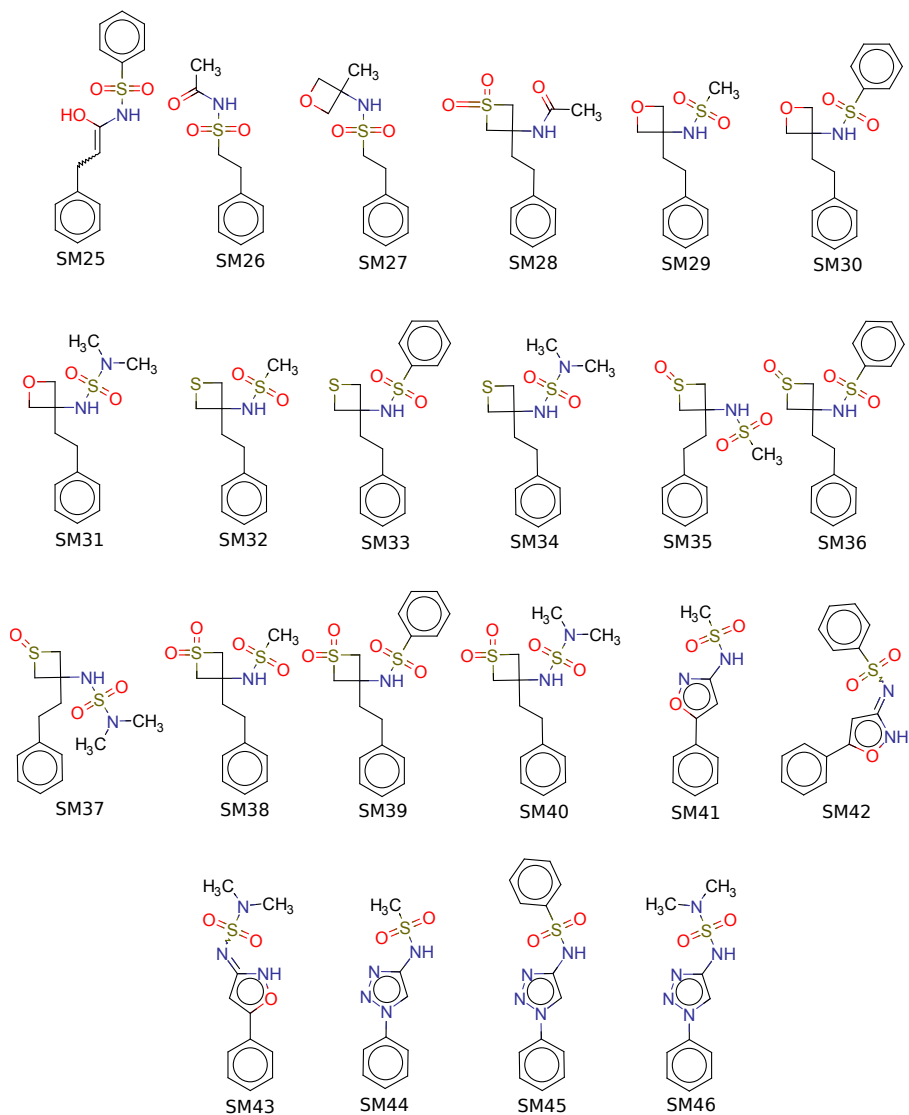


Fig. 2 Chemical structure and name of the molecules from the SAMPL7 challenge.

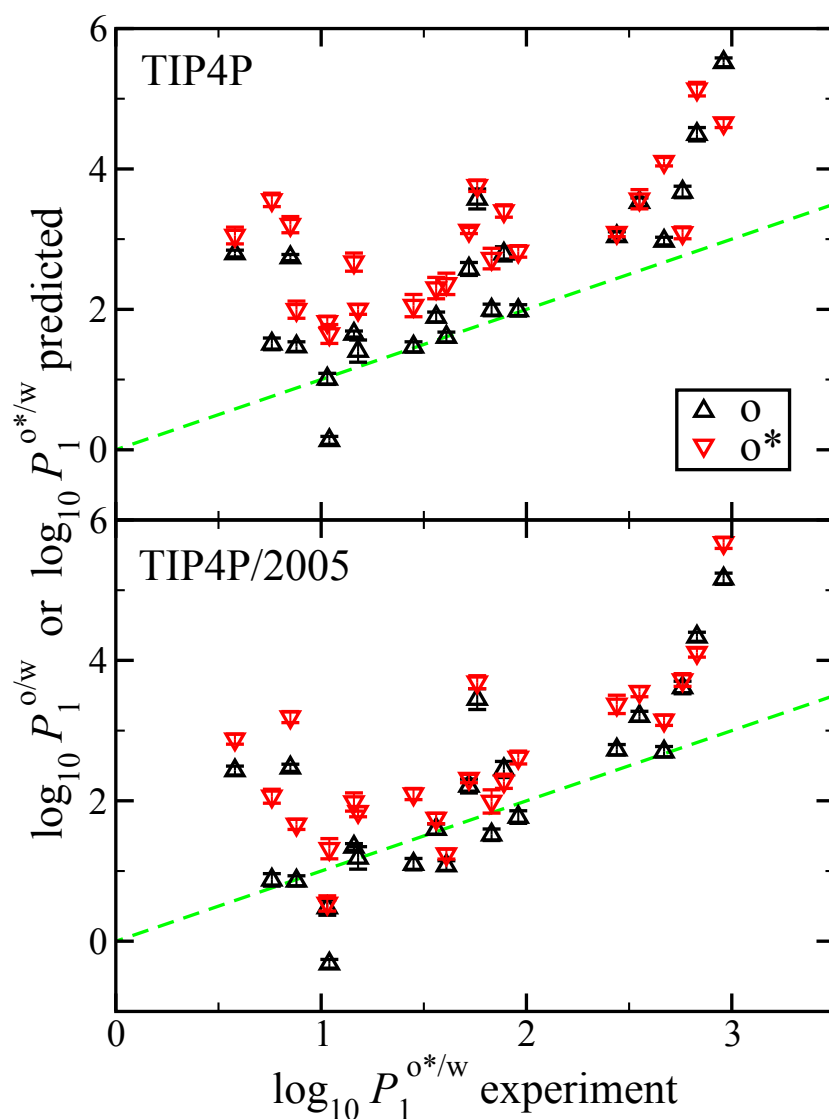


Fig. 3 Parity plot of the predicted versus experimental octanol/water partition coefficient for the SAMPL7 challenge. Predictions are made using both neat 1-octanol ($\log_{10} P_1^{o/w}$) and water-saturated 1-octanol ($\log_{10} P_1^{o*/w}$), and using both the TIP4P and TIP4P/2005 water models. The green dashed line ($y = x$) is drawn for reference.

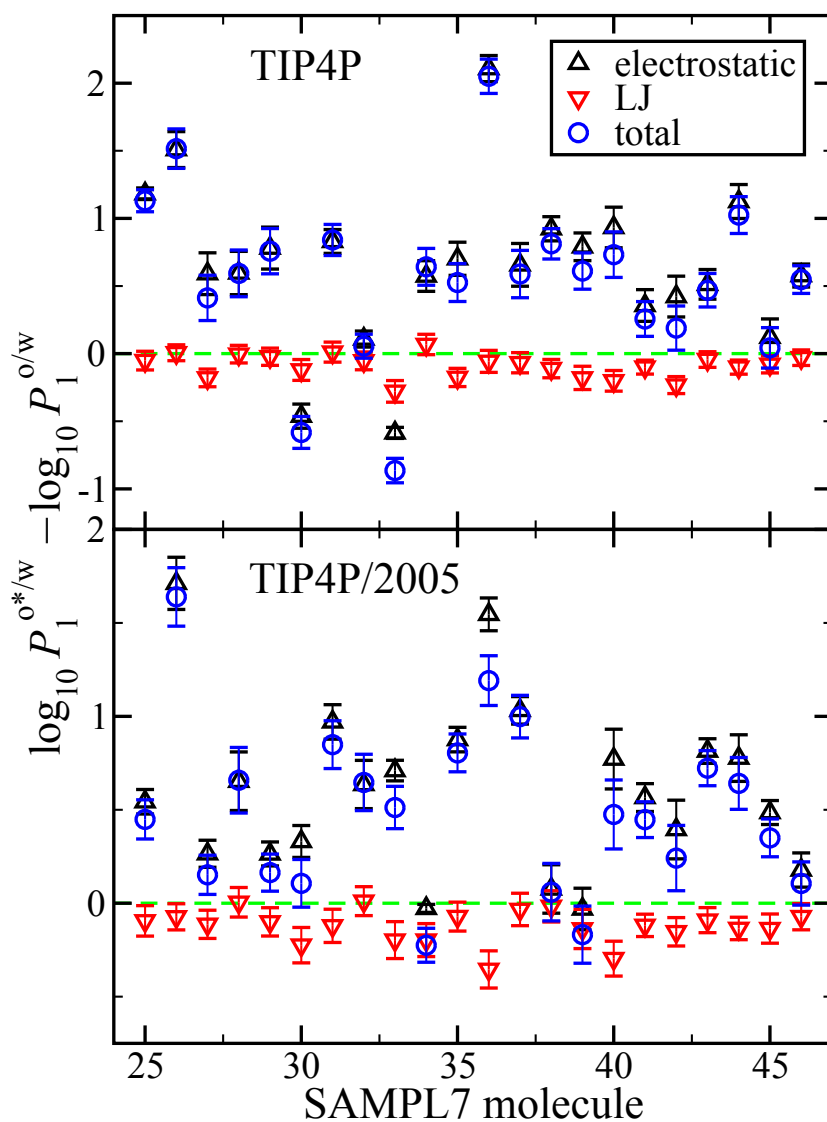


Fig. 4 Plot of the difference in the octanol/water partition coefficient predicted using water-saturated 1-octanol ($\log_{10} P_1^{o*/w}$) and neat 1-octanol ($\log_{10} P_1^{o/w}$) for each SAMPL7 molecule, using both the TIP4P and TIP4P/2005 water models. The predictions are additionally broken down into their electrostatic and Lennard-Jones contributions by using the respective contribution of the solvation free energy (see tables ?? and 2). The green dashed line is a horizontal line ($y = 0$) drawn for reference.

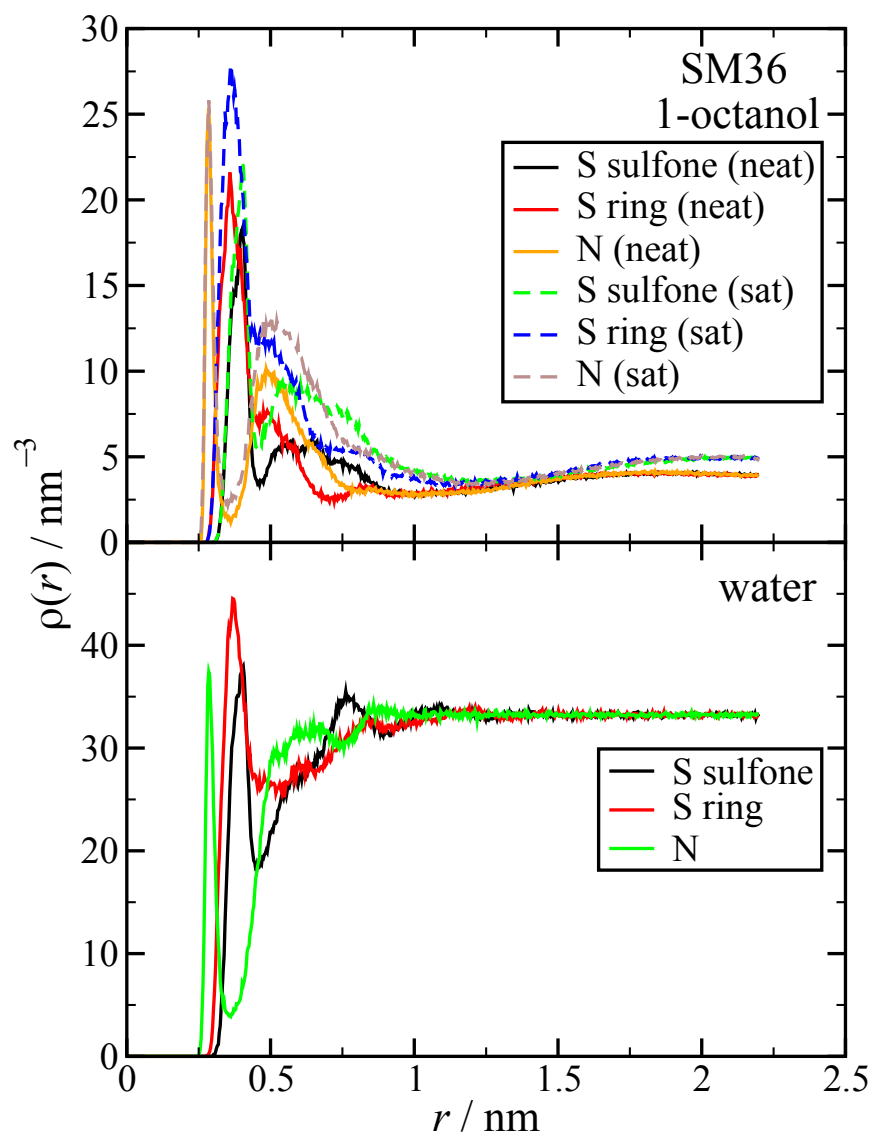


Fig. 5 Local density, $\rho(r)$, of solvent O around the S in the sulfone group, the (sulfoxide) S in the ring, and the (secondary amine) N in molecule SM36. The top pane is for neat and water-saturated (sat) 1-octanol, and the bottom pane is for water.

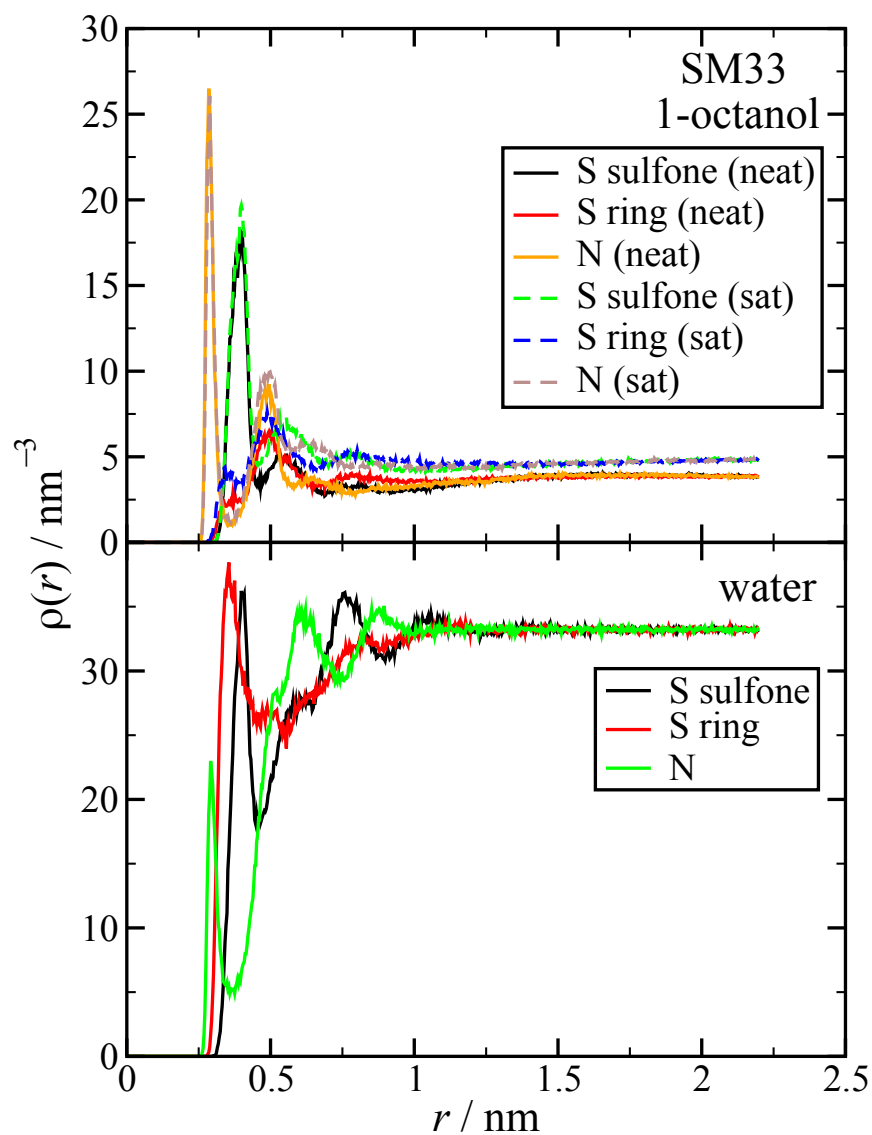


Fig. 6 Local density, $\rho(r)$, of solvent O around the S in the sulfone group, the (sulfide) S in the ring, and the (secondary amine) N in molecule SM33. The top pane is for neat and water-saturated (sat) 1-octanol, and the bottom pane is for water.

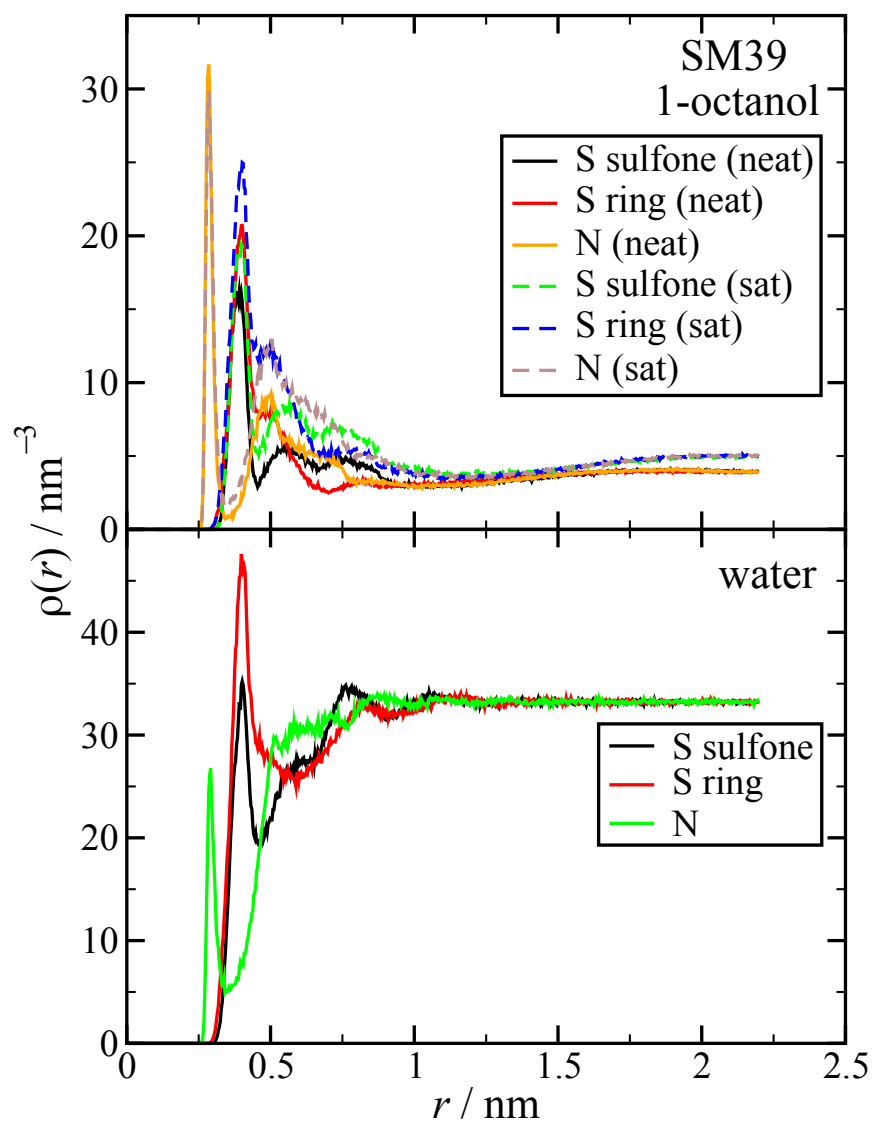


Fig. 7 Local density, $\rho(r)$, of solvent O around the S in the (non-ring) sulfone group, the (sulfone) S in the ring, and the (secondary amine) N in molecule SM36. The top pane is for neat and water-saturated (sat) 1-octanol, and the bottom pane is for water.

References

1. OECD: Test No. 107: Partition Coefficient (n-octanol/water): Shake Flask Method (1995). DOI <https://doi.org/https://doi.org/10.1787/9789264069626-en>. URL <https://www.oecd-ilibrary.org/content/publication/9789264069626-en>
2. Sangster, J.: Octanol-Water Partition Coefficients of Simple Organic Compounds. *J. Phys. Chem. Ref. Data* **18**, 1111–1227 (1989)
3. Ouimet, J.A., Paluch, A.S.: Predicting octanol/water partition coefficients for the SAMPL6 challenge using the SM12, SM8, and SMD solvation models. *J. Comput.-Aided Mol. Des.* **34**, 575–588 (2020)
4. Leo, A., Elkins, C.H.D.: Partition Coefficients and their Uses. *Chem. Rev.* **71**, 525–616 (1971)
5. Sangster, J.: Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry. John Wiley and Sons, Chichester, New York (1997)
6. Işık, M., Bergazin, T.D., Fox, T., Rizzi, A., Chodera, J.D., Mobley, D.L.: Assessing the accuracy of octanol-water partition coefficient predictions in the SAMPL6 Part II log *P* Challenge. *J. Comput.-Aided Mol. Des.* **34**, 335–370 (2020)
7. Sørensen, J.M., Arlt, W. (eds.): Liquid-Liquid Equilibrium Data Collection, Part 1: Binary Systems. DECHEMA, Frankfurt am Main, Germany (1979)
8. Tse, G., Sandler, S.I.: Determination of Infinite Dilution Activity Coefficients and 1-Octanol/Water Partition Coefficients of Volatile Organic Pollutants. *J. Chem. Eng. Data* **39**, 354–357 (1994)
9. Roese, S.N., Heintz, J.D., Uzat, C.B., Schmidt, A.J., Margulis, G.V., Sabatino, S.J., Paluch, A.S.: Assessment of the SM12, SM8, and SMD Solvation Models for Predicting Limiting Activity Coefficients at 298.15 K. *Processes* **8**, 623 (2020)
10. Chen, B., Siepmann, J.I.: Microscopic Structure and Solvation in Dry and Wet Octanol. *J. Phys. Chem. B* **110**, 3555–3563 (2006)
11. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L.: Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926–935 (1983)
12. Martin, M.G., Siepmann, J.I.: Transferable Potentials for Phase Equilibria. 1. United-Atom Description of *n*-Alkanes. *J. Phys. Chem. B* **102**, 2569–2577 (1998)
13. Chen, B., Potoff, J.J., Siepmann, J.I.: Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Tertiary Alcohols. *J. Phys. Chem. B* **105**, 3093–3104 (2001)
14. Chen, B., Siepmann, J.I.: Partitioning of Alkane and Alcohol Solutes between Water and (Dry or Wet) 1-Octanol. *J. Am. Chem. Soc.* **122**, 6464–6467 (2000)
15. MacCullum, J.L., Tieleman, D.P.: Structures of Neat and Hydrated 1-Octanol from Computer Simulations. *J. Am. Chem. Soc.* **124**, 15085–15093 (2002)
16. Bhatnagar, N., Kamath, G., Chelst, I., Potoff, J.J.: Direct calculation of 1-octanol-water partition coefficients from adaptive biasing force molecular dynamics simulations. *J. Chem. Phys.* **137**, 014502 (2012)
17. Kamath, N.B.G., Potoff, J.J.: Prediction of 1-octanol-water and air-water partition coefficients for nitro-aromatic compounds from molecular dynamics simulations. *Phys. Chem. Chem. Phys.* **15**, 6467–6474 (2013)
18. Zhang, H., Jiang, Y., Cui, Z., Yin, C.: Force Field Benchmark of Amino Acids. 2. Partition Coefficients between Water and Organic Solvents. *J. Chem. Inf. Model.* **58**, 1669–1681 (2018)
19. Bannan, C.C., Calabró, G., Kyu, D.Y., Mobley, D.L.: Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water. *J. Chem. Theory Comput.* **12**, 4015–4024 (2016)
20. Ogata, K., Hatakeyama, M., Nakamura, S.: Effect of Atomic Charges on Octanol-Water Partition Coefficient Using Alchemical Free Energy Calculation. *Molecules* **23**, 425 (2018)
21. Garrido, N.M., Queimada, A.J., Jorge, M., Macedo, E.A., Economou, I.G.: 1-Octanol/Water Partition Coefficients of *n*-Alkanes from Molecular Simulation of Absolute Solvation Free Energies. *J. Chem. Theory Comput.* **5**, 2436–2446 (2009)
22. Garrido, N.M., Economou, I.G., Queimada, A.J., Jorge, M., Macedo, E.A.: Prediction of the *n*-Hexane/Water and 1-Octanol/Water Partition Coefficients for Environmentally Relevant Compounds using Molecular Simulation. *AIChE J.* **58**, 1929–1938 (2012)
23. Yang, L., Ahmed, A., Sandler, S.I.: Comparison of Two Simulation Methods to Compute Solvation Free Energies and Partition Coefficients. *J. Comput. Chem.* **34**, 284–293 (2013)
24. Lyubartsev, A.P., Jacobsson, S.P., Sundholm, G., Laaksonen, A.: Solubility of Organic Compounds in Water/Octanol Systems. A Expanded Ensemble Molecular Dynamics Simulation Study of log *P* Parameters. *J. Phys. Chem. B* **105**, 7775–7782 (2001)

25. Fan, S., Iorga, B.I., Beckstein, O.: Prediction of octanol-water partition coefficients for the SAMPL6- $\log P$ molecules using molecular dynamics simulations with OPLS-AA, AMBER and CHARMM force fields. *J. Comput.-Aided Mol. Des.* **34**, 543–560 (2020)
26. SAMPL7 $\log P$ Prediction Challenge. URL https://github.com/samplchallenges/SAMPL7/tree/master/physical_property. (accessed March 11, 2021)
27. Abascal, J.L.F., Vega, C.: A general purpose model for the condensed phase of water: TIP4P/2005. *J. Chem. Phys.* **123**, 234505 (2005)
28. Leach, A.R.: *Molecular Modelling: Principles and Applications*, 2nd edn. Pearson Education Limited, Harlow, England (2001)
29. Frenkel, D., Smit, B.: *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd edn. Academic Press, San Diego, CA (2002)
30. Case, D.A., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham, III, T.E., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Giambasu, G., Gilson, M.K., Gohlke, H., Goetz, A.W., Harris, R., Izadi, S., Izmailov, S.A., Kasavajhala, K., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T.S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Man, V., Merz, K.M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., Onufriev, A., Pan, F., Pantano, S., Qi, R., Roe, D.R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C., R.Skrynnikov, N., Smith, J., Swails, J., Walker, R.C., Wang, J., Wilson, L., Wolf, R.M., Wu, X., Xiong, Y., Xue, Y., York, D.M., Kollman, P.: AMBER 20 (2020)
31. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A.: Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **25**, 1157–1174 (2004)
32. Wang, J., Wang, W., Kollman, P.A., Case, D.A.: Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **25**, 247–260 (2006)
33. Weininger, D.: SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988)
34. Daylight Chemical Information Systems, Inc. URL <https://www.daylight.com/>. (accessed March 11, 2021)
35. O’Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchinson, G.R.: Open Babel: An open chemical toolbox. *J. Cheminf.* **3**, 33 (2011)
36. Open Babel: The Open Source Chemistry Toolbox. URL http://openbabel.org/wiki/Main_Page. (accessed June 26, 2019)
37. Gasteiger, J., Marsili, M.: A New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.* **34**, 3181–3184 (1978)
38. Zhao, Y., Truhlar, D.G.: The M06 theory of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Account* **120**, 215–241 (2008)
39. Cramer, C.J.: *Essentials of Computational Chemistry*. John Wiley & Sons Ltd, Chichester, West Sussex, England (2002)
40. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A.V., Bloino, J., Janesko, B.G., Gomperts, R., Mennucci, B., Hratchian, H.P., Ortiz, J.V., Izmaylov, A.F., Sonnenberg, J.L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V.G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, Jr., J.A., Peralta, J.E., Ogliaro, F., Bearpark, M.J., Heyd, J.J., Brothers, E.N., Kudin, K.N., Staroverov, V.N., Keith, T.A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A.P., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Millam, J.M., Klene, M., Adamo, C., Cammi, R., Ochterski, J.W., Martin, R.L., Morokuma, K., Farkas, O., Foresman, J.B., Fox, D.J.: Gaussian 16, Revision C.01 (2019)
41. Bayly, C.I., Cieplak, P., Cornell, W.D., Kollman, P.A.: A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993)
42. Cieplak, P., Cornell, W.D., Bayly, C., Kollman, P.A.: Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **16**, 1357–1377 (1995)
43. Martínez, L., Andrade, R., Birgin, E.G., Martínez, J.M.: Packmol: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**(13), 2157–2164 (2009)

44. Packmol: Packing Optimization for Molecular Dynamics Simulations. URL <http://www.ime.unicamp.br/martinez/packmol/>. (accessed Mar 31, 2021)
45. GROMACS development team: *GROMACS Documentation: Release 2020.2* (2020). URL <https://manual.gromacs.org/documentation/2020.2/manual-2020.2.pdf>
46. Berendsen, H.J.C.: *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*. Cambridge University Press, New York, NY (2007)
47. Berendsen, H.J.C., Postma, J.P.M., DiNola, A., Haak, J.R.: Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984)
48. Bussi, G., Donadio, D., Parrinello, M.: Canonical sampling through velocity-rescaling. *J. Chem. Phys.* **126**, 014101 (2007)
49. Bussi, G., Parrinello, M.: Stochastic thermostats: comparison of local and global schemes. *Comp. Phys. Comm.* **179**, 26–29 (2008)
50. Bussi, G., Zykova-Timan, T., Parrinello, M.: Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **130**, 074101 (2009)
51. Parrinello, M., Rahman, A.: Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981)
52. Hess, B., Kutzner, C., van der Spoel, D., Lindal, E.: GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008)
53. 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.: GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013)
54. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E.: GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015)
55. GROMACS: Fast, flexible, free. URL <http://www.gromacs.org/>. (accessed Mar 31, 2021)
56. Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M.: LINCS: A Linear Constraint Solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472 (1997)
57. Hess, B.: P-LINCS: A Parallel Linear Constraint Solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008)
58. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C.: Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977)
59. Miyamoto, S., Kollman, P.A.: SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comp. Chem.* **13**, 952–962 (1992)
60. Deserno, M., Holm, C.: How to mesh up Ewald sums. i. a theoretical and numerical comparison of various particle mesh routines. *J. Chem. Phys.* **109**, 7678–7693 (1998)
61. Fuerst, G.B., Ley, R.T., Paluch, A.S.: Calculating the Fugacity of Pure, Low Volatile Liquids via Molecular Simulation with Application to Acetanilide, Acetaminophen, and Phenacetin. *Ind. Eng. Chem. Res.* **54**, 9027–9037 (2015)
62. Ley, R.T., Fuerst, G.B., Redeker, B.N., Paluch, A.S.: Developing a Predictive Form of MOSCED for Nonelectrolyte Solids Using Molecular Simulation: Application to Acetanilide, Acetaminophen, and Phenacetin. *Ind. Eng. Chem. Res.* **55**, 5415–5430 (2016)
63. Long, G.E., Dhakal, P., Redeker, B.N., Paluch, A.S.: Using limiting activity coefficients to efficiently evaluate the ability of fixed-charge force fields to model miscible water plus cosolvent mixtures. *Mol. Simul.* **45**, 322–335 (2019)
64. 2019 Solvation Free Energy Calculation Workshop at UFF. URL <https://sites.google.com/a/miamioh.edu/2019-free-energy-workshop-at-uff/>. (accessed Mar 31, 2021)
65. Shing, K.S., Chung, S.T.: Computer simulation methods for the calculation of solubility in supercritical extraction systems. *J. Phys. Chem.* **91**, 1674–1681 (1987)
66. Kofke, D.A., Cummings, P.T.: Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation. *Mol. Phys.* **92**, 973–996 (1997)
67. Shirts, M.R., Pitera, J.W., Swope, W.C., Pande, V.S.: Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **119**, 5740–5761 (2003)
68. Kofke, D.A., Cummings, P.T.: Precision and accuracy of staged free-energy perturbation methods for computing the chemical potential by molecular simulation. *Fluid Phase Equilib.* **150-151**, 41–49 (1998)

69. Chipot, C., Pohorille, A. (eds.): Free Energy Calculations: Theory and Applications in Chemistry and Biology. *Springer Series in Chemical Physics*, vol. 86. Springer, New York, NY (2007)
70. Bennett, C.H.: Efficient estimation of free energy differences from monte carlo data. *J. Comp. Phys.* **22**, 245–268 (1976)
71. Shirts, M.R., Bair, E., Hooker, G., Pande, V.S.: Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.* **91**, 140601 (2003)
72. Lu, N., Singh, J.K., Kofke, D.A.: Appropriate methods to combine forward and reverse free-energy perturbation averages. *J. Chem. Phys.* **118**, 2977–2984 (2003)
73. Shirts, M.R., Chodera, J.D.: Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008)
74. Beutler, T.C., Mark, A.E., van Schaik, R.C., Gerber, P.R., van Gunsteren, W.F.: Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **222**, 529–539 (1994)
75. Shirts, M.R., Pande, V.S.: Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **122**, 134508 (2005)
76. Steinbrecher, T., Mobley, D.L., Case, D.A.: Nonlinear scaling schemes for lennard-jones interactions in free energy calculations. *J. Chem. Phys.* **127**, 214108 (2007)
77. PyMBAR: Python implementation of the multistate bennett acceptance ratio (mbar). URL <https://github.com/choderalab/pymbar>. (accessed Mar 31, 2021)
78. Chodera, J.D., Swope, W.C., Pitera, J.W., Seok, C., Dill, K.A.: Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput.* **3**, 26–41 (2007)
79. Klimovich, P.V., Shirts, M.R., Mobley, D.L.: Guidelines for analysis of free energy calculations. *J. Comput.-Aided Mol. Des.* **29**, 397–411 (2015)
80. In summary, from $m = 0$ to 14, we have $\lambda_m^{LJ} = \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.0, 1.0, 1.0, 1.0\}$ and $\lambda_m^{\text{elec}} = \{0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.50, 0.71, 0.87, 1.00\}$. A detailed discussion of best practices when coupling/decoupling intermolecular interactions during free energy calculations is provided in the recent review by Shirts and Mobley [86].
81. van Gunsteren, W.F., Berendsen, H.J.C.: A leap-frog algorithm for stochastic dynamics. *Mol. Sim.* **1**, 173–185 (1988)
82. Brehm, M., Kirchner, B.: TRAVIS - A Free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories. *J. Chem. Inf. Model.* **51**, 2007–2023 (2011)
83. Brehm, M., Thomas, M., Gehrke, S., Kirchner, B.: TRAVIS - A free analyzer for trajectories from molecular simulation. *J. Chem. Phys.* **152**, 164105 (2020)
84. Jeffrey, G.A.: An Introduction to Hydrogen Bonding. Oxford University Press, New York, NY (1997)
85. Ohio Supercomputer Center: Ohio Supercomputer Center (1987). URL <http://osc.edu/ark:/19495/f5s1ph73>
86. Shirts, M.R., Mobley, D.L.: Biomolecular Simulations: Methods and Protocols, chap. An Introduction to Best Practices in Free Energy Calculations, pp. 271–311. Humana Press, Totowa, NJ (2013)