ABSTRACT

PREDICTING OCTANOL/WATER PARTITION COEFFICIENTS USING MOLECULAR SIMULATION FOR THE SAMPL7 CHALLENGE: COMPARING THE USE OF NEAT AND WATER SATURATED 1-OCTANOL

by Spencer J. Sabatino

The need for more efficient drug design and development has become more prevalent in just the last few years, leading to the development of the SAMPL challenges to promote exploration of methods to compute physical properties key to drug development. Blind predictions of octanol/water partition coefficients at 298.15 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.

# PREDICTING OCTANOL/WATER PARTITION COEFFICIENTS USING MOLECULAR SIMULATION FOR THE SAMPL7 CHALLENGE: COMPARING THE USE OF NEAT AND WATER SATURATED 1-OCTANOL

A Thesis

Submitted to the

Faculty of Miami University

in partial fulfillment of

the requirements for the degree of

Master of Science

by

Spencer J. Sabatino

Miami University

Oxford, Ohio

2022

Advisor: Dr. Andrew Paluch

Reader: Dr. Jason Boock

Reader: Dr. Alim Dewan

This Thesis titled

# PREDICTING OCTANOL/WATER PARTITION COEFFICIENTS USING MOLECULAR SIMULATION FOR THE SAMPL7 CHALLENGE: COMPARING THE USE OF NEAT AND WATER SATURATED 1-OCTANOL

by

Spencer J. Sabatino

has been approved for publication by

The College of Engineering and Computing

and

Department of Chemical, Paper, and Biomedical Engineering

_____

Dr. Andrew Paluch

_____

Dr. Jason Boock

_____

Dr. Alim Dewan

# Table of Contents

Background …… 1

Theory …… 2

Partition Coefficient ……... 2

Distribution Coefficient ……..… 3

General Case ….....4

SAMPL7 …… 6

Introduction …… 6

Methodology …… 11

Computational Details …… 11

Molecular Dynamics …… 12

Free Energy Calculations …… 13

Results and Discussion …… 14

Quantitative Predictions ……... 14

Structural Analysis ……... 21

Simulation Time ……... 24

Conclusion ……... 25

Future Work: SAMPL8 …… 27

Introduction …….. 27

Methodology …… 28

pKa ……... 28

Distribution Coefficients (logD) ……. 29

uESE …… 29

References …… 30

Appendix I: Dimensionless Solvation Free Energies ……..…… 38

Appendix II: Additional Structural Analysis ……41

# List of Tables

Table 1: A summary of the octanol/water partition coefficients using water (TIP4P) and neat 1-octanol relative to the experimental values published by the SAMPL group. Each calculation is broken down into its respective LJ and electrostatic contributions. The numbers in the subscript represent the error in the first two decimal places……17

Table 2: A summary of the octanol/water partition coefficients using water (TIP4P) and water-saturated 1-octanol relative to the experimental values published by the SAMPL group. Each calculation is broken down into its respective LJ and electrostatic contributions. The numbers in the subscript represent the error in the first two decimal places……18

Table 3: A summary of the octanol/water partition coefficients using water (TIP4P-2005) and neat 1-octanol relative to the experimental values published by the SAMPL group. Each calculation is broken down into its respective LJ and electrostatic contributions. The numbers in the subscript represent the error in the first two decimal places…….19

Table 4: A summary of the octanol/water partition coefficients using water (TIP4P-2005) and water-saturated 1-octanol compared to the experimental values published by the SAMPL group. Each calculation is broken down into its respective LJ and electrostatic contributions. The numbers in the subscript represent the error in the first two decimal places……..20

Table 5: A summary of reference calculations performed by the SAMPL6 organizers comparing neat and water-saturated 1-octanol using the TIP3P water model and GAFF for 1-octanol. The numbers in the subscript represent the error in the first two decimal places……..21

## List of Figures

Figure 1: Chemical structure and name of the molecules from the SAMPL6 challenge…8

Figure 2: Chemical structure and name of the molecules from the SAMPL7 challenge..10

Figure 3: Parity plot of the predicted versus experimental octanol/water partition coefficients for the SAMPL7 challenge. The dashed green line (y=x) is drawn for reference……15

Figure 4: Plot of the difference in octanol/water partition coefficients predicted using both water-saturated 1-octanol and neat 1-octanol for each SAMPL7 molecule. Each prediction was broken down into its respective electrostatic and LJ contributions……...16

Figure 5: Local density of solvent O around key portions of SM36. The top portion is for neat and saturated 1-octanol, and the bottom pane is for water…….23

Figure 6: Local density of solvent O around key portions of SM33. The top portion is for neat and saturated 1-octanol, and the bottom pane is for water…….24

Figure 7: Local density of solvent O around key portions of SM33. The top portion is for neat and saturated 1-octanol, and the bottom pane is for water……25

Figure 8: Chemical structure and name of each molecule of interest in the SAMPL8 challenge……29

## Dedication

To my mother, Kelley, and my grandfather, John, for all their support they provided for me growing up.

## Acknowledgements

Computing support was provided by the Ohio Supercomputer Center [95]. We appreciate the National Institutes of Health for its support of the SAMPL Project via R01GM124270 to David L. Mobley (UC Irvine).

I would like to thank the entire faculty of the Department of Chemical, Paper, and Biomedical Engineering for all their work with helping build a reliable, engaged future generation of engineers and scientists over the past four years that I attended Miami University. I would like to specifically thank Dr. Boock, Dr. Paluch, and Dr. Dewan for all their hard work over the past four years as professors and academic advisors to help students like me succeed in chemical engineering. I would also like to thank Dr. Paluch for all his wisdom and motivation that he provided me while I was part of his research group from Fall 2019 to present day.

The work done in this paper for SAMPL7 can be found published in the Journal of Computer-Aided Molecular Simulation [1].

## Background and Introduction

The development of mathematical models for physical phenomena has always been one of the main goals of scientists and engineers. These models allow for quick, relatively low-cost estimations and predictions of a physical system such as material properties, system behavior, or thermodynamic data. Before the invention of computers and high-power processing, many models did not have a viable way to apply or solve for parameters in the model, so numerous assumptions would be made to simplify the model and governing equations down to a useable model, albeit a potentially less accurate one. With the rise of modern computing power over the last 75 years, we have seen growing access to more powerful computing software to simulate and evaluate models that previously were not useable due to the difficulty of the model's calculations. It becomes key that we begin to test common models and see how assumptions made to the model have an effect on the computed parameters and whether they are significant or uniform. These models can be applied in both academic research and in industry settings such as pharmaceuticals or aircraft design.

The pharmaceuticals industry is one of the largest and quickest-growing industries in the world, with 35 pharmaceutical companies alone amassing nearly $12 trillion revenue in a span of 18 years, with that number only expected to grow [2]. This unparalleled growth is only expected to significantly increase over time, especially given the COVID-19 pandemic and the demand increase for vaccines, anti-viral medication, and medical devices. This increasing demand for pharmaceuticals has led to a substantial increase in research projects and activities to develop new vaccines or drugs [3]. This increase in research activity leads to an increase in expenses on various chemicals and equipment, and more importantly, generation of a lot more chemical waste from the development process. This untreated chemical waste, from both production of the drugs and their development, has become of interest as of recent, as it has become one of the main contaminants found in wastewater treatment facilities in the world. This chemical waste has not only been classified as toxic to terrestrial life but has also been shown to be one of the possible factors for the rise of antibiotic-resistant microbes in the environment, providing another key risk to society [4-6].

With these risks in mind, it is ideal that pharmaceutical companies find a way to reduce these sources of waste to prevent any catastrophic consequences or fines from environmental regulation agencies throughout the world. We have seen the rise of different supply chain ideologies emerge over time, with many ideas for the pharmaceutical industry to become "zero waste" or to become a circular supply chain, where the waste produced from making pharmaceutical products can be used in other industries and thus be recycled and utilized elsewhere, promoting the "zero waste" goal proposed by many [7, 8]. Various green chemistry methods have been slowly being employed throughout the last 50 years to help minimize both the amount of waste and the toxicity of the waste being produced. Eldin et al notes the various tenets of green chemistry such as minimizing the use of toxic solvents for extractions, minimizing emissions, using less wasteful sampling techniques, and even reducing labor by relying on automated processes and computational work, albeit with the caveat that increased use in computational work generates more carbon dioxide to the environment [9]. Computational work for pharmaceutical drug research and design will be the focus of the remainder of this discussion.

Molecular modeling and design have become a significant field of study over the last 60 years, emerging from efforts from Hendrickson to prediction energies of hydrocarbons using the computational power at the time [10,11]. These computational techniques have been extensively utilized by companies like Eli Lilly and DuPont as early as the mid-1960's and has been advertised heavily to the pharmaceutical industry as a way for the industry to develop pharmaceutical products more easily for much cheaper. This rise in interest led to the founding of numerous prominent companies relating to molecular modeling and dynamics, such as Gaussian, which focuses mostly on electronic structure modeling, and Schrödinger and D.E. Shaw, with more of a focus on molecular dynamics and modeling software [10]. With this rise in prominence and importance, many innovators were able to design their own force fields and simulation methods using quantum chemistry or statistic mechanics to formulate their process, leading to the development of many different software packages.

With the development of numerous software packages, force fields, and other simulation techniques, a new challenge emerges: what software package or force field combination is the best for modeling a particular system of interest? While there are so many options for force fields and software packages, there are very few large testbeds available to the public for researchers to test, and comparative methods are few and far between since so many configurations of force fields and software packages exist. To combat this, in 2008, the molecular modeling community launched the Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) project designed to provide viable testbeds for researchers to practice novel computational methods on to predict various physical properties. The premise of the SAMPL challenges is to provide researchers with a novel set of pharmaceutical-like molecules to use in generating blind predictions of various physical properties of these molecules in various solvents. These projects typically have a timeline of about 1-2 months and allow the researchers involved in the contest to use any desired technique. Various techniques used in the competition include physical molecular modeling, physical quantum modeling, and empirical methods [12]. Physical molecular modeling involves the use of force fields to estimate a system's energy as a function of a molecule's position, essentially solving a system of equations of Newton's laws of motion for every molecule. Quantum molecular modeling uses a numerical solution to the Schrödinger equation to estimate physical properties, often employing numerous assumptions to provide a viable solution. Empirical methods are more varied but typically share common trends of being parameterized with large data sets to allow for accurate predictions [12]. In this document, I intend to describe my work in the SAMPL7 and SAMPL8 challenges to predict physical properties by using molecular modeling techniques and the methods employed.

# Theory

### Partition Coefficient (logP)

The partitioning of a neutral solute in a binary liquid phase has various applications in many industries and has a large significance in the pharmaceutical industry. In particular, octanol-water partition coefficients are of biggest interest to the pharmaceutical world due to their ability to quantify the hydrophilic and hydrophobic nature of the solute, such as a

pharmaceutical-grade drug or toxic compound, in the system [13, 14]. The 1-octanol simulates the hydrophobic portions of the body such as a lipid bilayer around living cells or the fat of an animal, and the water simulates the hydrophilic portions of the body such as the blood or other bodily fluids [15, 16]. During measurement of partition coefficients, it's crucial to keep the solute concentration relatively low so the solution can be assumed infinitely dilute, although this assumption is rarely accurate. [14, 15]. This infinitely dilute assumption minimizes solute interactions that may interfere with the equilibrium in the liquid-liquid phase while still modeling the solute-solvent interactions present in the system, simplifying our calculations. With the infinite dilution assumption, we can model partition coefficients in terms of solvation free energies [16]:

$$\log_{10} P_{1-0}^{\mathrm{I/II}}(T,P) = \log_{10} \frac{c_{1,I}}{c_{1,II}} = -\frac{\log_{10} e}{RT}\left[\Delta G_{1-0,\mathrm{I}}^{\mathrm{solv}}(T,P) - \Delta G_{1-0,\mathrm{II}}^{\mathrm{solv}}(T,P)\right] \quad (1)$$

where T and P are system temperature and pressure, R is the molar gas constant, $c_{1,I}$ is the concentration of the solute in liquid phase I, $c_{1,II}$ is the concentration of the solute in liquid phase II, $\Delta G_{1,\mathrm{I}}^{\mathrm{solv}}(T,P)$ is the solvation free energy of the solute in liquid phase I at a specific temperature and pressure, and $\Delta G_{1,\mathrm{II}}^{\mathrm{solv}}(T,P)$ is the solvation free energy of the solute in liquid phase II at a specific temperature and pressure. The difference in the solvation free energies above can also be rewritten as the transition free energy of the system. Being able to model these partition coefficients is crucial for drug design and has been the focus of the SAMPL physical property prediction challenges over the past several years in SAMPL challenges like SAMPL6 and SAMPL7 [12, 18]. In the case of octanol-water partition coefficients, phase I would be the octanol-phase and phase II would be the water-phase, and equation (1) from above can be rewritten specifically for water-octanol partition coefficients as the following:

$$\log_{10} P_{1}^{\mathrm{o/w}}(T,P) = \log_{10} \frac{c_{1,o}}{c_{1,w}} = -\frac{\log_{10} e}{RT}\left[\Delta G_{1,\mathrm{o}}^{\mathrm{solv}}(T,P) - \Delta G_{1,\mathrm{w}}^{\mathrm{solv}}(T,P)\right] \quad (2)$$

where the subscript "o" denotes the octanol-rich phase of the system, and the subscript "w" denotes the water-rich phase of the system. The sign of the octanol-water partition coefficient indicates a higher concentration of solute in one of the liquid phases, octanol if positive and water if negative, thus showing the solute's affinity for one of the liquid phases. The magnitude of the partition coefficient describes the strength of the affinity for the given phase, with a higher magnitude implying a much greater affinity for one of the phase relative to the other.

**Distribution Coefficient (logD)**

While partition coefficients are sufficient for neutral systems, they often neglect pH effects which would lead to the formation of nonneutral systems through ionization, deionization, or tautomeric effects in the aqueous phase of the system. When your solute is placed in the aqueous phase of a system, while we anticipate most of the solute molecules remain neutral, it's likely that some of the molecules will be gain or lose $H^+$ ions and become protonated or deprotonated in one or multiple spots in the solute molecule. This interaction could yield the formation of singly protonated species, singly deprotonated species, doubly protonated species, doubly deprotonated species, and possibly more depending on the nature of the system.

Likewise, some molecules could rearrange in aqueous solution and form a tautomer and rearrange between the normal and tautomeric structure depending on the equilibrium of the system. To account for this, the distribution coefficient is formulated to account for not just neutral species partitioned across both liquid phases, but all ionic and tautomeric species in both the water phase and the organic phase [20]. The distribution coefficient helps provide more insight into the hydrophilic and hydrophobic nature of the solute in question, as it considers any charged variations of the solute formed in the aqueous phase of the system. Likewise, this can more accurately model the solute's trajectory in the aqueous phase, blood in the case of a biological system since it can account for the acidic and basic compounds in blood potentially protonating or deprotonating the given solute. The general expression for the distribution coefficient is given below as the following:

$$\log_{10} D_1^{\mathrm{I/II}}(T,P) = \log_{10}\left(\frac{\sum_{i=0}^{N} c_{1,\,\mathrm{I}}}{\sum_{i=0}^{N} c_{1,\,\mathrm{II}}}\right) \quad (3)$$

where T is the temperature of the system, P is the pressure of the system, $c_{1,\,\mathrm{I}}$ is the total concentration of all species, neutral, ionic, and tautomeric, of the solute in solvent I at a certain pH, $c_{1,\,\mathrm{II}}$ denotes the total concentration of all species of solute in solvent II at a certain pH, and N denotes the number of ionic and tautomeric states present in each phase. Depending on the nature of the solute, some solutes many contain several different protonated and deprotonated species, so we must consider all possible species for all possible configurations to obtain the most accurate description of the system.

**General Case**

To obtain the general expression to compute $log_{10} D_1^{o/w}$ in the present study, we consider an example. Let's assume we have our neutral solute plus five additional microstates.

0. $H_2A$: our neutral reference molecule.
1. $HA^-$: deprotonated species with a formal charge of -1, caused by the loss of a hydrogen atom.
2. $H_3A^+$: protonated species with a formal charge of 1, caused by the donation of a hydrogen atom.
3. $AH_2$: a tautomer of our original species with a formal charge of 0.
4. $A^{2-}$: deprotonated species with a formal charge of -2, caused by the loss of two hydrogen atoms.
5. $H_4A^{2+}$: protonated species with a formal charge of 2, caused by the donation of two hydrogen atoms.

This leads to:

$$\log_{10} D_1^{\mathrm{o/w}}(T,P) = \log_{10}\left(\frac{c_{1-0,\mathrm{o}}}{c_{1-0,\mathrm{w}} + c_{1-2,\mathrm{w}} + c_{1-3,\mathrm{w}} + c_{1-4,\mathrm{w}} + c_{1-5,\mathrm{w}} + c_{1-6,\mathrm{w}}}\right) \quad (4)$$

We can readily compute $\log_{10} \mathrm{P}_1^{\mathrm{o/w}}$ from the transfer free energy of the neutral solute between the two phases. Let us therefore take the difference between $\log_{10} \mathrm{D}_1^{\mathrm{o/w}}$ and $\log_{10} \mathrm{P}_1^{\mathrm{o/w}}$ to obtain

an expression that may be used to ``correct" our predicted value of $\log_{10} \mathrm{P}_1^{\mathrm{o/w}}$. Taking the difference and simplifying we have:

$$= \log_{10}\left(\frac{c_{1-0,\mathrm{o}}}{c_{1-0,\mathrm{w}} + c_{1-1,\mathrm{w}} + c_{1-2,\mathrm{w}} + c_{1-3,\mathrm{w}} + c_{1-4,\mathrm{w}} + c_{1-5,\mathrm{w}}}\right) - \log_{10}\left(\frac{\mathrm{c}_{1-0,\mathrm{o}}}{\mathrm{c}_{1-0,\mathrm{w}}}\right)$$

$$= \log_{10}\left(\frac{c_{1-0,\mathrm{o}}}{c_{1-0,\mathrm{w}} + c_{1-1,\mathrm{w}} + c_{1-2,\mathrm{w}} + c_{1-3,\mathrm{w}} + c_{1-4,\mathrm{w}} + c_{1-5,\mathrm{w}}} \cdot \frac{c_{1-0,w}}{c_{1-0,o}}\right)$$

$$= \log_{10}\left(\frac{c_{1-0,w}}{c_{1-0,\mathrm{w}} + c_{1-1,\mathrm{w}} + c_{1-2,\mathrm{w}} + c_{1-3,\mathrm{w}} + c_{1-4,\mathrm{w}} + c_{1-5,\mathrm{w}}}\right)$$

$$= -\log_{10}\left(1 + \frac{c_{1-1,\mathrm{w}}}{c_{1-0,\mathrm{w}}} + \frac{c_{1-2,\mathrm{w}}}{c_{1-0,\mathrm{w}}} + \frac{c_{1-3,\mathrm{w}}}{c_{1-0,\mathrm{w}}} + \frac{c_{1-4,\mathrm{w}}}{c_{1-0,\mathrm{w}}} + \frac{c_{1-5,\mathrm{w}}}{c_{1-0,\mathrm{w}}}\right) \text{ (5)}$$

Next, we will work out expressions for the relative concentrations in terms of the free energies of reaction computed in the present study. Starting with state 1, we have:

$$\mathrm{H_{2A}} \rightleftharpoons \mathrm{H^+} + \mathrm{HA^-} \text{ (6)}$$

$$Ka_1 = \frac{[\mathrm{H^+}][\mathrm{HA^-}]}{[\mathrm{H_{2A}}]} = \frac{c_{\mathrm{H^+,w}} c_{1-1,\mathrm{w}}}{c_{1-0,\mathrm{w}}} \text{ (7)}$$

Where:

$$\mathrm{pH} = -\log_{10} c_{\mathrm{H^+,w}} \text{ (8)}$$

$$\ln c_{\mathrm{H^+,w}} = -\mathrm{pH}/\log_{10} e \text{ (9)}$$

And then:

$$-\frac{\Delta G_{1-0}^{\mathrm{rxn}}(T,P)}{RT} = \ln K\, a_1 = \ln c_{\mathrm{H^+,w}} + \ln \frac{c_{1-1,\mathrm{w}}}{c_{1-0,\mathrm{w}}}$$

$$-\frac{\Delta G_{1-0}^{\mathrm{rxn}}(T,P)}{RT} = -\frac{\mathrm{pH}}{\log_{10} e} + \ln \frac{c_{1-1,\mathrm{w}}}{c_{1-0,\mathrm{w}}}$$

$$-\frac{\Delta G_{1-0}^{\mathrm{rxn}}(T,P)}{RT} + \frac{\mathrm{pH}}{\log_{10} e} = \ln \frac{c_{1-1,\mathrm{w}}}{c_{1-0,\mathrm{w}}} \text{ (10)}$$

Where $\Delta G_{1-0}^{\mathrm{rxn}}(T,P)$ is the corresponding free energy of reaction computed in this study. This leads to the desired result:

$$\frac{c_{1-1,\mathrm{w}}}{c_{1-0,\mathrm{w}}} = \exp\left[-\frac{\Delta G^{\mathrm{rxn}}_{1-0}(T,P)}{RT} + \frac{\mathrm{pH}}{\log_{10} e}\right] \quad (11)$$

This procedure can be repeated for the other cases, yielding the equations below for the other 5 cases mentioned above:

$$\frac{c_{1-2,\mathrm{w}}}{c_{1-0,\mathrm{w}}} = \exp\left[-\frac{\Delta G^{\mathrm{rxn}}_{2-0}(T,P)}{RT} - \frac{\mathrm{pH}}{\log_{10} e}\right] \quad (12)$$

$$\frac{c_{1-3,\mathrm{w}}}{c_{1-0,\mathrm{w}}} = \exp\left[-\frac{\Delta G^{\mathrm{rxn}}_{3-0}(T,P)}{RT}\right] \quad (13)$$

$$\frac{c_{1-4,\mathrm{w}}}{c_{1-0,\mathrm{w}}} = \exp\left[-\frac{\Delta G^{\mathrm{rxn}}_{4-0}(T,P)}{RT} + 2\frac{\mathrm{pH}}{\log_{10} e}\right] \quad (14)$$

$$\frac{c_{1-5,\mathrm{w}}}{c_{1-0,\mathrm{w}}} = \exp\left[-\frac{\Delta G^{\mathrm{rxn}}_{5-0}(T,P)}{RT} - 2\frac{\mathrm{pH}}{\log_{10} e}\right] \quad (15)$$

Combining equations (11) – (15) into equation (5) and generalizing the expression to n states, we can derive our generalized distribution coefficient equation:

$$\log_{10} D_1^{\mathrm{o/w}}(T,P) = \frac{\log_{10} e}{RT}\left[\Delta G^{\mathrm{solv}}_{1-0,\mathrm{o}}(T,P) - \Delta G^{\mathrm{solv}}_{1-0,\mathrm{w}}(T,P)\right] - \log_{10}\left(1 + \sum_{i=1}^{N} \exp\left[-\frac{\Delta G^{\mathrm{rxn}}_{i-0}}{RT} - q_i \frac{\mathrm{pH}}{\log_{10} e}\right]\right) \quad (16)$$

# SAMPL7

# Introduction

The partitioning of neutral solutes between two liquid phases is important for a variety of industrial applications. Using equation (1) from above, we can compute the partition coefficient for a system for any two liquids assuming an infinitely dilute system. Likewise, equation (2) can be used to compute 1-octanol-water partition coefficients. When computing partition coefficients, one common assumption is that the liquid phases are pure and immiscible. For the case of 1-octanol and water, the mutual solubility of water in 1-octanol is $0.703 \times 10^{-4}$ mole fracs, so it is a reasonable assumption for the aqueous phase to be essentially pure [20]. However, we unfortunately cannot neglect the water saturation in the 1-octanol phase, so we can modify our equation (2) as the following:

$$\log_{10} P_1^{\mathrm{o^*/w}}(T,P) = \log_{10} \frac{c_{1,o*}}{c_{1,w}} = -\frac{\log_{10} e}{RT}\left[\Delta G^{\mathrm{solv}}_{1,\,\mathrm{o^*}}(T,P) - \Delta G^{\mathrm{solv}}_{1,w-}(T,P)\right] \quad (17)$$

where the new subscript, "o*" denotes the water-saturated 1-octanol phase as the organic phase, rather than the pure 1-octanol phase previously denoted.

Experimentally, the measurement of octanol/water partition coefficients can be laborious. Tse and Sandler [22] 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 [17, 23]. 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 by using equation (2). For the set of organic pollutants studied, they found that there was an appreciable difference between the limiting activity coefficient in pure water and pure 1-octanol, and in turn there was a noticeable 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 12 compounds used to train their model consisted of the following alkanes and chloro-alkanes: pentane, cyclohexane, hexane, heptane, $CCl_4$, $CHCl_3$, $CH_2Cl_2$, $C_2HCl_3$, 1,1,1-$Cl_3C_2H_3$, 1,1,2-$Cl_3C_2H_3$, 1,1-$Cl_2C_2H_4$, and 1,2-$Cl_2C_2H_4$. With reference $\log_{10} P_1^{o*/w}$ values ranging from 1.25 to 4.99, the average absolute error was 0.2 when using limiting activity coefficients in pure 1-octanol. For all the systems except 1,1,2-$Cl_3C_2H_3$, the reference $\log_{10} P_1^{o*/w}$ was less than that predicted using limiting activity coefficients in pure 1-octanol.

The liquid-liquid equilibrium of water and 1-octanol has been rigorously studied by Chen and Siepmann [24] 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 [24] 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 [25] and TraPPE-UA 1-octanol [26, 27] 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 [28], and SPC water with the modified GROMOS96 force field for 1-octanol resulted in an equilibrium concentration of 0.16 mole fracs [29].

Significant progress has been made in the use of molecular simulation (Monte Carlo and molecular dynamics) to predict octanol/water partition coefficients [7, 16, 18–27]. 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. (2). 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 [22] we know that $\Delta G_{1,o*}^{solv}$ 1,o* $\neq \Delta G_{1,o}^{solv}$. 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 system 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 to 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 0.05–0.10 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 [18].

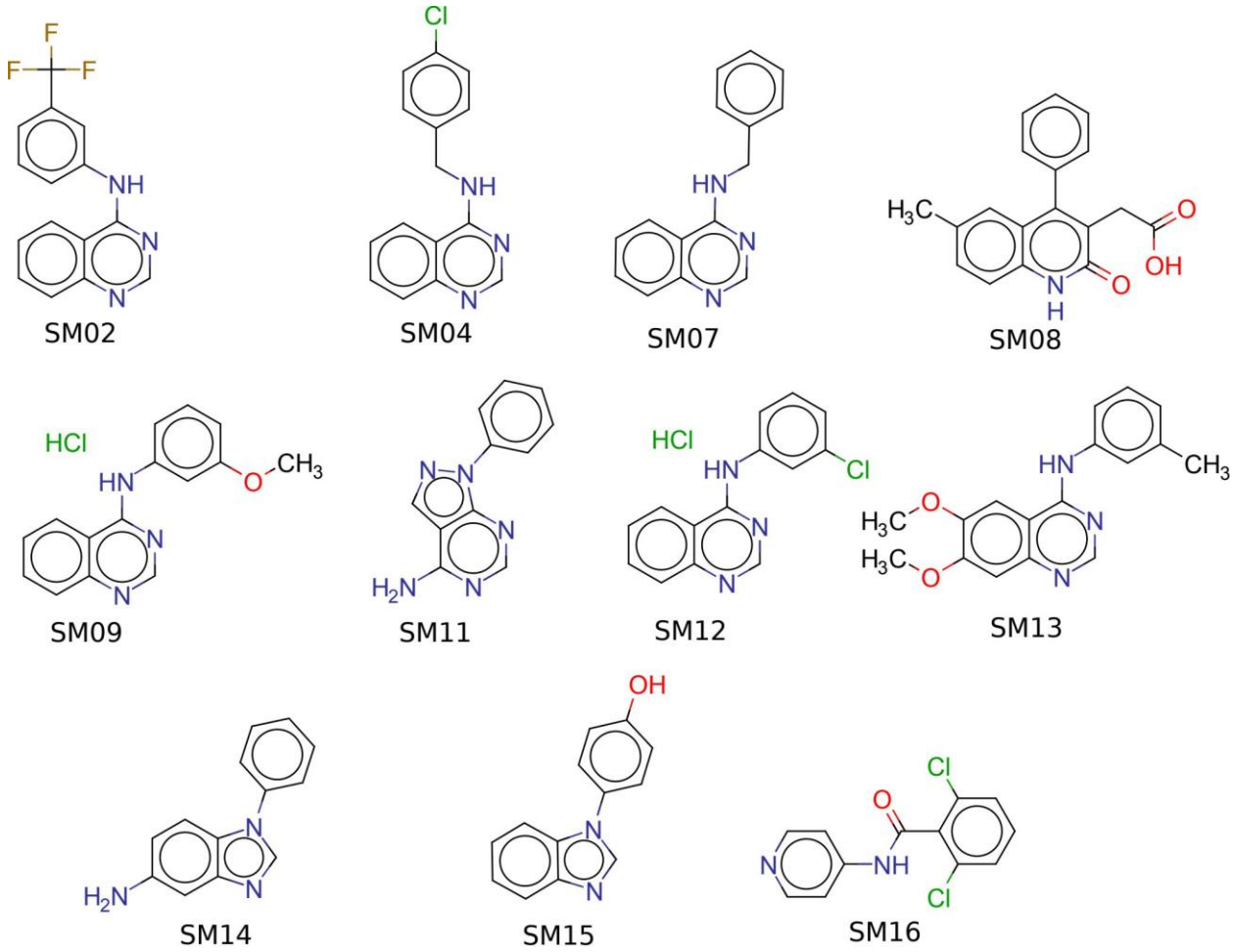


*Figure 1: Chemical structure and name of the molecules from the SAMPL6 challenge*

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 [12, 30]. 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 [24], 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 simulations 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. The predictions using pure 1-octanol corresponds to SAMPL7 entry "TFE MD neat oct (GAFF/TIP4P)" and the predictions using water-saturated 1-octanol corresponds to "TFE wet oct (GAFF/ TIP4P)" [12, 30]. After the close of the challenge, we additionally repeated the calculations with TIP4P-2005 to look at the effect of the water model [31]. 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, and the effect of water saturation is dependent on the chemical structure of the solute. The difference in RMSE of 0.39 log units is larger, albeit comparable to the experimental findings of Tse and Sandler [22].

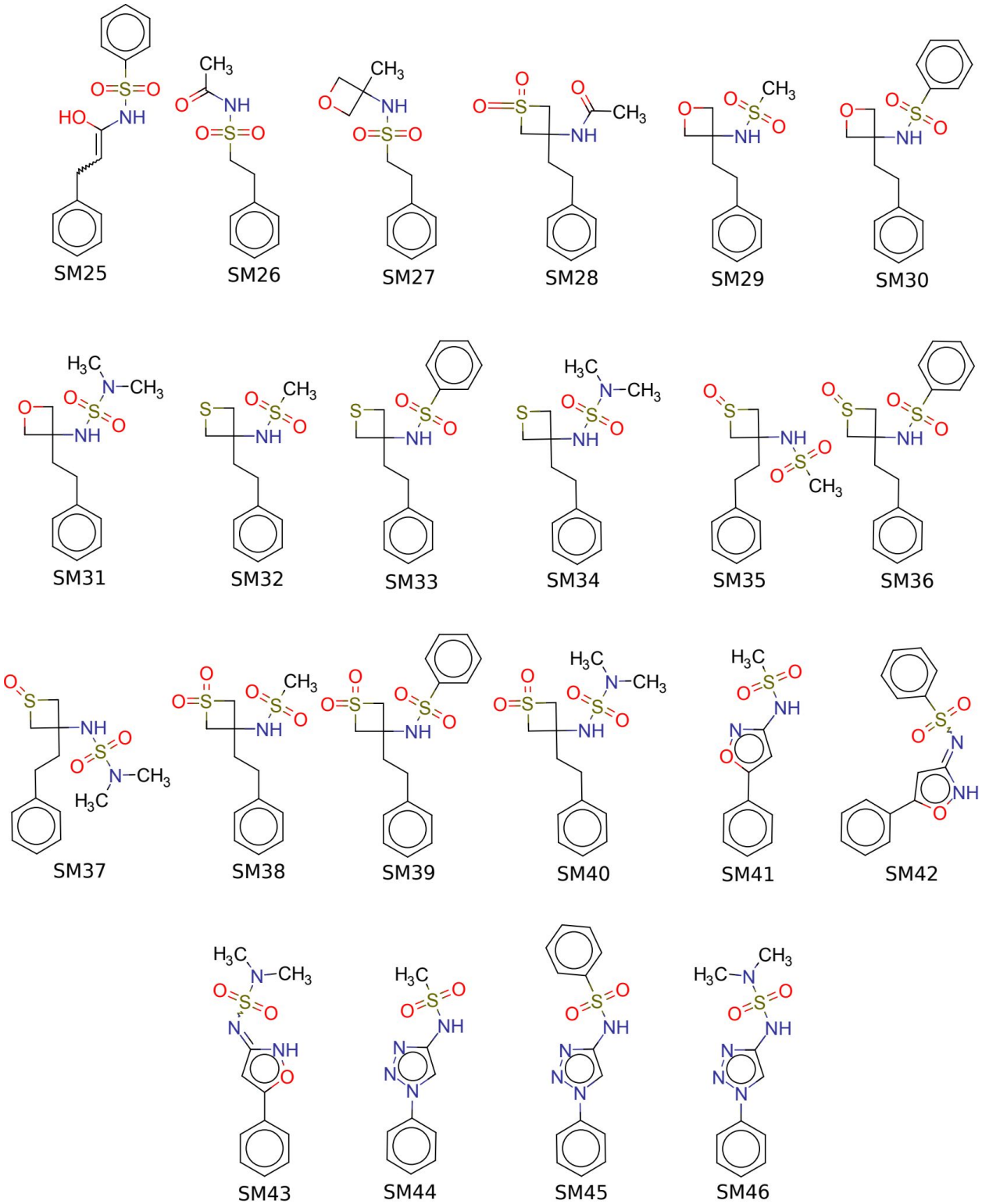


*Figure 2: Chemical structure and name of the molecules from the SAMPL7 challenge*

While our predictions here using neat 1-octanol are in better quantitative agreement with experiment, this should not be interpreted as a recommendation to ignore the effect of water

saturation. The goal of the present work is to study the effect of water saturation. We use the combination of TIP4P water and TraPPE-UA 1-octanol as it has been shown to accurately predict the mutual solubility of the binary system [24]. The ability to model the solutes using TraPPE is not straightforward, so we instead modeled the solutes using the General AMBER Force Field [32]. Future work is needed to study the effect of solvent and solute force fields.

# Methodology

### Computational Details:

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 [33, 34]

$$U_{\mathrm{nb}}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \qquad (21)$$

where $r_{ij}$ is the separation distance between sites i and j, $\epsilon_{ij}$ is the well-depth of the LJ potential, $\sigma_{ij}$ 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 [25] and 1-octanol was modeled with the United Atom Transferable Potentials for Phase Equilibria (TraPPE-UA) force field [26, 27]. The choice of solvent models was based on the work of Chen and Siepmann [24] 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 [35] for comparison after the conclusion of the SAMPL7 challenge.

The ability to model the solutes using TraPPE is not obvious. Therefore, the solutes were all modeled using the General AMBER Force Field version 2 (GAFF2) as implemented in the AMBER 20 simulation suite [32, 36, 37]. 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 started with the optimized structures found in our SAMPL7 entry "TFE-SMD-vacuum-opt" wherein we used electronic structure calculations to predict both the partition coefficients and pKa [12, 30]. In summary, the structures were obtained by first taking the Daylight SMILES [37, 38] provided by the challenge organizers [30] and generating an initial 3-D structure with Open Babel 2.3.2 [40, 41]. 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) [32] with Gasteiger partial charges [45] 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 [46, 47] using Gaussian 16 [45]. For this optimized structure we then performed 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) [46, 47] method in antechamber (within the AMBER 20 simulation suite). The use of RESP with HF/6-31G* follows the

recommendations of the original GAFF work [32]. All the GROMACS force field files used in the present study are provided in the Supporting Information accompanying the electronic version of this manuscript.

## 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 3000 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 [24]. 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 the simulations were carried out following the same procedure.

First, Packmol was used to generate initial structures [48, 49]. This was followed by 3000 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 [33, 34, 50, 51]. 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 [50–52]. This was followed by 4 ns of equilibration in an NPT ensemble at the same conditions using the stochastic velocity rescaling thermostat [46, 52–54] and the Parrinello-Rahman barostat [56]. 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 [57–60]. All bond lengths in 1-octanol, and bond lengths involving a hydrogen for the SAMPL7 molecules were constrained using P-LINCS [50, 61, 62]. Water was modeled as completely rigid using the SETTLE algorithm [63, 64]. The Verlet neighbor list was used [50] and LJ interactions were cut-of at 1.4 nm. Long-range analytic dispersion corrections were applied to the energy and pressure to accommodate the truncation [33, 34, 50, 51]. Lorentz-Berthelot mixing rules were used for unlike LJ sites [33]. The electrostatic terms were evaluated with the smooth particle-mesh-Ewald method (SPME) with tin-foil boundary conditions [50, 51, 65] 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 published in the Journal of Computer-Aided Molecular Design [1].

**Free Energy Calculations**

The free energy calculations were performed at 298.15 K and 1 bar following our previous work [66–69]. The solvation free energy, $\Delta G_1^{solv}$ , for the solute infinitely dilute in water, 1-octanol, and water-saturated 1-octanol was calculated using a multi-stage free energy perturbation method [70–74] with the multi-state Bennett's acceptance ratio method (MBAR) [75–78].

A "soft-core" potential was used to decouple the solute-solvent intermolecular LJ interactions. Stage (m) dependent decoupling parameters, $\lambda_m^{LJ}$ and $\lambda_m^{elec}$ controlled the LJ and electrostatic intermolecular interactions, respectively. The decoupling parameters varied from 0 to 1. When $\lambda_m^{LJ} = \lambda_m^{elec} = 1$, the solute is fully coupled to the system. When $\lambda_m^{LJ} = \lambda_m^{elec} = 0$, the solute is decoupled from the system. The "soft-core" potential had the form [79–82]

$$U_{\mathrm{LJ^{SC}}}(r_{ij};m) = 4\lambda_m^{\mathrm{LJ}}\varepsilon_{ij}\left\{\frac{\sigma_{ij}^{12}}{\left[\left(1-\lambda_m^{LJ}\right)\alpha_{LJ}\sigma_{ij}^{6}+r_{ij}^{6}\right]^{2}} - \frac{\sigma_{ij}^{6}}{\left[\left(-\lambda_m^{LJ}\alpha_{LJ}\sigma_{ij}^{6}+r_{ij}^{6}\right]}\right\} \qquad (22)$$

where $\alpha_{LJ}$ 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 $\lambda_m^{LJ}$ = 0 and 1), it additionally allows nearly decoupled molecules to overlap with a fnite energy (and hence finite probability). The electrostatic term in the intermolecular potential was decoupled linearly as

$$U_{\mathrm{elec}}(r_{ij};m) = \lambda_m^{\mathrm{elec}}\frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \quad (23)$$

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 [78] to determine $\Delta G_1^{\mathrm{solv}}$ . This analysis was performed using the Python implementation of MBAR (PyMBAR) and the GROMACS analysis script distributed with it [81]. The GROMACS analysis script has implemented an autocorrelation analysis so that only uncorrelated samples are used to determine $\Delta G_1^{solv}$ and the corresponding uncertainty [83–85].

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 = 1–10 the LJ interactions were increased from $\lambda_m^{LJ}$= 0.1–1.0 in 10 equal increments of 0.1. Electrostatic interactions were increased in a square root fashion following $\lambda_m^{elec}$ = {0.50,0.71,0.87,1.00} from m = 11–14 [86].

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 [50, 51, 87]. 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.

# Results and Discussion

## Quantitative predictions

The solvation free energy computed for each SAMPL7 molecule is summarized in tables S1, S2, and S3 of the Supporting Information of the original publication for water, neat 1-octanol, and water-saturated 1-octanol, respectively [1]. In all cases the (total) solvation free energy is broken down into its electrostatic (m = 10–14) and LJ (m = 0–10) 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 and Table 2, we additionally summarize the experimental values provided by the challenge organizers after the close of the challenge [30]. The predictions using TIP4P with pure 1-octanol corresponds to SAMPL7 entry "TFE MD neat oct (GAFF/TIP4P)" and the predictions with water-saturated 1-octanol corresponds to "TFE wet oct (GAFF/TIP4P)" [12, 30]. The predictions using the TIP4P/2005 water model were completed after the close of the challenge.

In Fig. 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 tend 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.

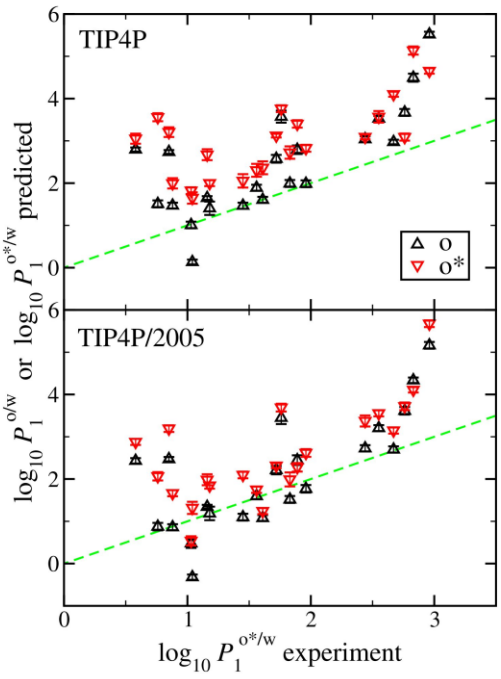


*Figure 3: Parity plot of the predicted vs experimental octanol/water partition coefficients for the SAMPL7 challenge. The dashed green line (y=x) is drawn for reference.*

In Fig. 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

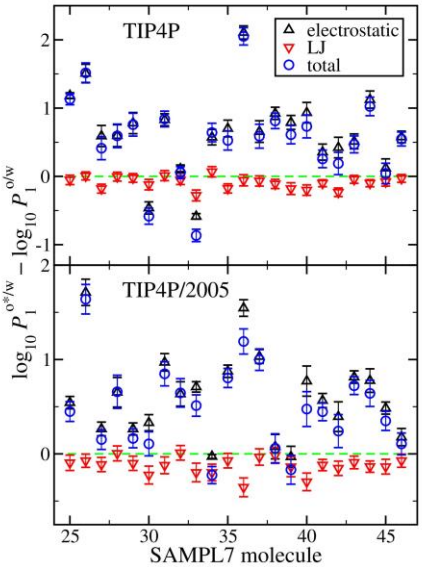


*Figure 4: Plot of the difference in octanol/water partition coefficients predicted using both water-saturated 1-octanol and neat 1-octanol for each SAMPL7 molecule. Each prediction was broken down into its respective electrostatic and LJ contributions.*

*Table 1: A summary of the octanol/water partition coefficients using TIP4P and neat 1-octanol relative to the experimental values published by the SAMPL group. Each calculation is broken down into its respective LJ and electrostatic contributions. The numbers in the subscript represent the error in the first two decimal places.*

| Molecule Number | Electrostatic | LJ | Total | Experimental Results | Difference from Experimental |
|---|---|---|---|---|---|
| 25 | $-7.67_{02}$ | $10.64_{05}$ | $2.97_{05}$ | $2.67_{01}$ | $0.30_{06}$ |
| 26 | $-8.53_{04}$ | $8.66_{04}$ | $0.13_{06}$ | $1.04_{01}$ | $-0.91_{06}$ |
| 27 | $-7.94_{05}$ | $9.83_{05}$ | $1.89_{07}$ | $1.56_{11}$ | $0.33_{13}$ |
| 28 | $-8.81_{15}$ | $10.22_{04}$ | $1.41_{16}$ | $1.18_{08}$ | $0.23_{18}$ |
| 29 | $-8.07_{05}$ | $9.67_{04}$ | $1.61_{07}$ | $1.61_{03}$ | $0.00_{08}$ |
| 30 | $-8.18_{06}$ | $11.86_{05}$ | $3.67_{08}$ | $2.76_{19}$ | $0.91_{21}$ |
| 31 | $-8.58_{06}$ | $10.57_{05}$ | $1.99_{08}$ | $1.96_{14}$ | $0.03_{16}$ |
| 32 | $-6.94_{05}$ | $9.98_{05}$ | $3.04_{07}$ | $2.44_{17}$ | $0.60_{18}$ |
| 33 | $-6.91_{03}$ | $12.43_{06}$ | $5.52_{06}$ | $2.96_{21}$ | $2.56_{22}$ |
| 34 | $-6.45_{08}$ | $10.95_{05}$ | $4.49_{10}$ | $2.83_{20}$ | $1.66_{22}$ |
| 35 | $-8.98_{05}$ | $10.45_{05}$ | $1.47_{07}$ | $0.88_{02}$ | $0.59_{07}$ |
| 36 | $-11.45_{06}$ | $12.96_{06}$ | $1.51_{08}$ | $0.76_{05}$ | $0.75_{10}$ |
| 37 | $-10.05_{05}$ | $11.51_{05}$ | $1.47_{07}$ | $1.45_{10}$ | $0.02_{12}$ |
| 38 | $-9.70_{06}$ | $10.71_{05}$ | $1.01_{08}$ | $1.03_{07}$ | $-0.02_{11}$ |
| 39 | $-10.29_{08}$ | $13.08_{06}$ | $2.79_{10}$ | $1.89_{13}$ | $0.90_{16}$ |
| 40 | $-9.95_{06}$ | $11.94_{06}$ | $1.99_{08}$ | $1.83_{05}$ | $0.16_{10}$ |
| 41 | $-6.35_{03}$ | $9.15_{04}$ | $2.80_{05}$ | $0.58_{02}$ | $2.22_{05}$ |
| 42 | $-8.13_{13}$ | $11.70_{04}$ | $3.57_{14}$ | $1.76_{03}$ | $1.81_{15}$ |
| 43 | $-7.67_{02}$ | $10.41_{04}$ | $2.74_{04}$ | $0.85_{01}$ | $1.89_{05}$ |
| 44 | $-7.54_{02}$ | $9.19_{04}$ | $1.65_{04}$ | $1.16_{03}$ | $0.49_{05}$ |
| 45 | $-8.00_{04}$ | $11.53_{05}$ | $3.53_{06}$ | $2.55_{04}$ | $0.98_{07}$ |
| 46 | $-7.76_{08}$ | $10.34_{04}$ | $2.58_{09}$ | $1.72_{01}$ | $0.86_{09}$ |

*Table 2: A summary of the octanol/water partition coefficients using TIP4P and water-saturated 1-octanol relative to the experimental values published by the SAMPL group. Each calculation is broken down into its respective LJ and electrostatic contributions. The numbers in the subscript represent the error in the first two decimal places*

| Molecule Number | Electrostatic | LJ | Total | Experimental Results | Difference from Experimental |
|---|---|---|---|---|---|
| 25 | $-6.48_{04}$ | $10.59_{05}$ | $4.10_{06}$ | $2.67_{01}$ | $1.43_{06}$ |
| 26 | $-7.02_{13}$ | $8.67_{04}$ | $1.65_{13}$ | $1.04_{01}$ | $0.61_{13}$ |
| 27 | $-7.35_{14}$ | $9.65_{05}$ | $2.30_{15}$ | $1.56_{11}$ | $0.74_{19}$ |
| 28 | $-8.22_{05}$ | $10.21_{05}$ | $2.00_{07}$ | $1.18_{08}$ | $0.82_{11}$ |
| 29 | $-7.29_{15}$ | $9.65_{05}$ | $2.36_{15}$ | $1.61_{03}$ | $0.75_{15}$ |
| 30 | $-8.65_{06}$ | $11.74_{05}$ | $3.09_{08}$ | $2.76_{19}$ | $0.33_{21}$ |
| 31 | $-7.75_{06}$ | $10.58_{05}$ | $2.83_{08}$ | $1.96_{14}$ | $0.87_{16}$ |
| 32 | $-6.83_{04}$ | $9.92_{05}$ | $3.09_{06}$ | $2.44_{17}$ | $0.65_{18}$ |
| 33 | $-7.50_{03}$ | $12.15_{06}$ | $4.65_{06}$ | $2.96_{21}$ | $1.69_{22}$ |
| 34 | $-5.88_{08}$ | $11.02_{05}$ | $5.14_{10}$ | $2.83_{20}$ | $2.31_{22}$ |
| 35 | $-8.28_{11}$ | $10.27_{05}$ | $2.00_{12}$ | $0.88_{02}$ | $1.12_{12}$ |
| 36 | $-9.34_{08}$ | $12.90_{06}$ | $3.56_{10}$ | $0.76_{05}$ | $2.80_{11}$ |
| 37 | $-9.39_{15}$ | $11.45_{05}$ | $2.05_{16}$ | $1.45_{10}$ | $0.60_{19}$ |
| 38 | $-8.77_{06}$ | $10.60_{05}$ | $1.82_{08}$ | $1.03_{07}$ | $0.79_{11}$ |
| 39 | $-9.50_{07}$ | $12.90_{06}$ | $3.40_{09}$ | $1.89_{13}$ | $1.51_{16}$ |
| 40 | $-9.02_{14}$ | $11.74_{05}$ | $2.72_{15}$ | $1.83_{05}$ | $0.89_{16}$ |
| 41 | $-6.00_{11}$ | $9.05_{04}$ | $3.05_{12}$ | $0.58_{02}$ | $2.47_{12}$ |
| 42 | $-7.71_{07}$ | $11.47_{04}$ | $3.76_{08}$ | $1.76_{03}$ | $2.00_{09}$ |
| 43 | $-7.16_{11}$ | $10.36_{04}$ | $3.21_{12}$ | $0.85_{01}$ | $2.36_{12}$ |
| 44 | $-6.42_{12}$ | $9.09_{04}$ | $2.67_{13}$ | $1.16_{03}$ | $1.51_{13}$ |
| 45 | $-7.88_{13}$ | $11.45_{04}$ | $3.57_{14}$ | $2.55_{04}$ | $1.02_{14}$ |
| 46 | $-7.18_{02}$ | $10.31_{04}$ | $3.12_{04}$ | $1.72_{01}$ | $1.40_{05}$ |

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 $\Delta G_{1,o*}^{solv}$ relative to $\Delta G_{1,o}^{solv}$, 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 in both cases is

0.62. 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 watersaturated 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

*Table 3: A summary of the octanol/water partition coefficients using TIP4P-2005 and neat 1-octanol relative to the experimental values published by the SAMPL group. Each calculation is broken down into its respective LJ and electrostatic contributions. The numbers in the subscript represent the error in the first two decimal places.*

| Molecule Number | Electrostatic | LJ | Total | Experimental Results | Difference from Experimental |
|---|---|---|---|---|---|
| 25 | $-7.73_{05}$ | $10.43_{06}$ | $2.70_{07}$ | $2.67_{01}$ | $0.03_{08}$ |
| 26 | $-8.68_{04}$ | $8.36_{05}$ | $-0.32_{06}$ | $1.04_{01}$ | $-1.36_{06}$ |
| 27 | $-7.86_{05}$ | $9.46_{05}$ | $1.60_{07}$ | $1.56_{11}$ | $0.04_{14}$ |
| 28 | $-8.67_{15}$ | $9.86_{05}$ | $1.19_{16}$ | $1.18_{08}$ | $0.01_{18}$ |
| 29 | $-8.23_{05}$ | $9.31_{05}$ | $1.08_{07}$ | $1.61_{03}$ | $-0.53_{08}$ |
| 30 | $-7.98_{06}$ | $11.59_{07}$ | $3.61_{09}$ | $2.76_{19}$ | $0.85_{21}$ |
| 31 | $-8.62_{07}$ | $10.39_{06}$ | $1.77_{09}$ | $1.96_{14}$ | $-0.19_{17}$ |
| 32 | $-6.92_{05}$ | $9.64_{05}$ | $2.73_{07}$ | $2.44_{17}$ | $0.29_{19}$ |
| 33 | $-7.05_{04}$ | $12.22_{07}$ | $5.16_{08}$ | $2.96_{21}$ | $2.20_{22}$ |
| 34 | $-6.51_{02}$ | $10.84_{06}$ | $4.34_{06}$ | $2.83_{20}$ | $1.51_{21}$ |
| 35 | $-9.23_{05}$ | $10.09_{05}$ | $0.86_{07}$ | $0.88_{02}$ | $-0.02_{07}$ |
| 36 | $-11.63_{06}$ | $12.50_{07}$ | $0.87_{09}$ | $0.76_{05}$ | $0.11_{10}$ |
| 37 | $-10.03_{05}$ | $11.13_{06}$ | $1.10_{08}$ | $1.45_{10}$ | $-0.35_{13}$ |
| 38 | $-9.69_{09}$ | $10.17_{06}$ | $0.48_{11}$ | $1.03_{07}$ | $-0.55_{13}$ |
| 39 | $-10.18_{09}$ | $12.62_{08}$ | $2.45_{11}$ | $1.89_{13}$ | $0.56_{17}$ |
| 40 | $-10.01_{05}$ | $11.52_{07}$ | $1.52_{08}$ | $1.83_{05}$ | $-0.31_{10}$ |
| 41 | $-6.25_{05}$ | $8.68_{04}$ | $2.43_{06}$ | $0.58_{02}$ | $1.85_{07}$ |
| 42 | $-7.79_{14}$ | $11.24_{05}$ | $3.45_{15}$ | $1.76_{03}$ | $1.69_{15}$ |
| 43 | $-7.44_{02}$ | $9.90_{05}$ | $2.47_{05}$ | $0.85_{01}$ | $1.62_{05}$ |
| 44 | $-7.32_{02}$ | $8.66_{04}$ | $1.34_{05}$ | $1.16_{03}$ | $0.18_{06}$ |
| 45 | $-7.81_{05}$ | $11.02_{05}$ | $3.21_{07}$ | $2.55_{04}$ | $0.66_{08}$ |
| 46 | $-7.57_{09}$ | $9.79_{05}$ | $2.21_{10}$ | $1.72_{01}$ | $0.49_{10}$ |

*Table 4: A summary of the octanol/water partition coefficients using TIP4P-2005 & water-saturated 1-octanol compared to the experimental values published by the SAMPL group. Each calculation is broken down into its respective LJ and electrostatic contributions. The numbers in the subscript represent the error in the first two decimal places.*

| Molecule Number | Electrostatic | LJ | Total | Experimental Results | Difference from Experimental |
|---|---|---|---|---|---|
| 25 | $-7.18_{05}$ | $10.33_{06}$ | $3.15_{07}$ | $2.67_{01}$ | $0.48_{07}$ |
| 26 | $-6.97_{13}$ | $8.28_{05}$ | $1.32_{14}$ | $1.04_{01}$ | $0.28_{14}$ |
| 27 | $-7.59_{05}$ | $9.34_{05}$ | $1.75_{07}$ | $1.56_{11}$ | $0.19_{13}$ |
| 28 | $-8.02_{04}$ | $9.86_{06}$ | $1.84_{07}$ | $1.18_{08}$ | $0.66_{11}$ |
| 29 | $-7.96_{05}$ | $9.21_{05}$ | $1.24_{07}$ | $1.61_{03}$ | $-0.37_{08}$ |
| 30 | $-7.65_{06}$ | $11.37_{07}$ | $3.72_{09}$ | $2.76_{19}$ | $0.96_{21}$ |
| 31 | $-7.65_{07}$ | $10.27_{06}$ | $2.62_{09}$ | $1.96_{14}$ | $0.66_{17}$ |
| 32 | $-6.28_{12}$ | $9.66_{06}$ | $3.37_{13}$ | $2.44_{17}$ | $0.93_{21}$ |
| 33 | $-6.35_{04}$ | $12.02_{07}$ | $5.67_{08}$ | $2.96_{21}$ | $2.71_{22}$ |
| 34 | $-6.54_{01}$ | $10.65_{06}$ | $4.11_{06}$ | $2.83_{20}$ | $1.28_{21}$ |
| 35 | $-8.35_{05}$ | $10.01_{06}$ | $1.66_{07}$ | $0.88_{02}$ | $0.78_{07}$ |
| 36 | $-10.08_{07}$ | $12.15_{07}$ | $2.07_{10}$ | $0.76_{05}$ | $1.31_{11}$ |
| 37 | $-9.00_{05}$ | $11.10_{06}$ | $2.10_{08}$ | $1.45_{10}$ | $0.65_{13}$ |
| 38 | $-9.61_{09}$ | $10.15_{06}$ | $0.54_{11}$ | $1.03_{07}$ | $-0.49_{13}$ |
| 39 | $-10.21_{07}$ | $12.49_{07}$ | $2.28_{11}$ | $1.89_{13}$ | $0.39_{16}$ |
| 40 | $-9.24_{15}$ | $11.23_{07}$ | $1.99_{17}$ | $1.83_{05}$ | $0.16_{17}$ |
| 41 | $-5.69_{06}$ | $8.56_{04}$ | $2.88_{07}$ | $0.58_{02}$ | $2.30_{07}$ |
| 42 | $-7.39_{07}$ | $11.08_{05}$ | $3.69_{09}$ | $1.76_{03}$ | $1.93_{10}$ |
| 43 | $-6.62_{06}$ | $9.81_{05}$ | $3.19_{08}$ | $0.85_{01}$ | $2.34_{08}$ |
| 44 | $-6.54_{12}$ | $8.52_{04}$ | $1.98_{13}$ | $1.16_{03}$ | $0.82_{13}$ |
| 45 | $-7.33_{05}$ | $10.88_{06}$ | $3.55_{07}$ | $2.55_{04}$ | $1.00_{08}$ |
| 46 | $-7.40_{03}$ | $9.71_{05}$ | $2.32_{06}$ | $1.72_{01}$ | $0.60_{06}$ |

coefficient 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 Fig. 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 to 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 0.05–0.10 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 [18].

Table 5 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 [18]. 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. While the difference of 0.22 log units is smaller than observed here for SAMPL7, interestingly is it very similar to the effect observed in the experimental study of Tse and Sandler [22]. Additionally, for 3 of the 11 solutes we find that the inclusion of water in the octanol-rich phase decreased 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.

*Table 5: A summary of reference calculations performed by the SAMPL6 organizers comparing neat and water-saturated 1-octanol using the TIP3P water model and GAFF for 1-octanol. The numbers in the subscript represent the error in the first two decimal places.*

| Molecule Number | $\log_{10} P_1^{o/w}$ | $\log_{10} P_1^{o^*/w}$ | Experimental $\log_{10} P_1^{o^*/w}$ |
|---|---|---|---|
| 02 | $5.54_{08}$ | $5.86_{18}$ | $4.09_{03}$ |
| 04 | $5.16_{06}$ | $5.18_{18}$ | $3.98_{03}$ |
| 07 | $4.21_{21}$ | $4.04_{28}$ | $3.21_{04}$ |
| 08 | $9.82_{08}$ | $9.37_{14}$ | $3.10_{03}$ |
| 09 | $4.51_{03}$ | $4.63_{07}$ | $3.03_{07}$ |
| 11 | $2.54_{07}$ | $2.94_{04}$ | $2.10_{04}$ |
| 12 | $5.59_{06}$ | $4.99_{04}$ | $3.83_{03}$ |
| 13 | $5.29_{12}$ | $5.33_{06}$ | $2.92_{04}$ |
| 14 | $2.42_{11}$ | $2.57_{12}$ | $1.95_{03}$ |
| 15 | $3.10_{08}$ | $3.31_{14}$ | $3.07_{03}$ |
| 16 | $3.88_{06}$ | $4.14_{29}$ | $2.62_{01}$ |

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 [24] 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. Consequently, 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 is also sensitive to the inclusion of water in the octanol-rich phase.

**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. (Fig. 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 the SAMPL7 molecules, and for SM33 this is the largest decrease of all 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 [88, 89]. 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 [33, 34]. 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 Fig. 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

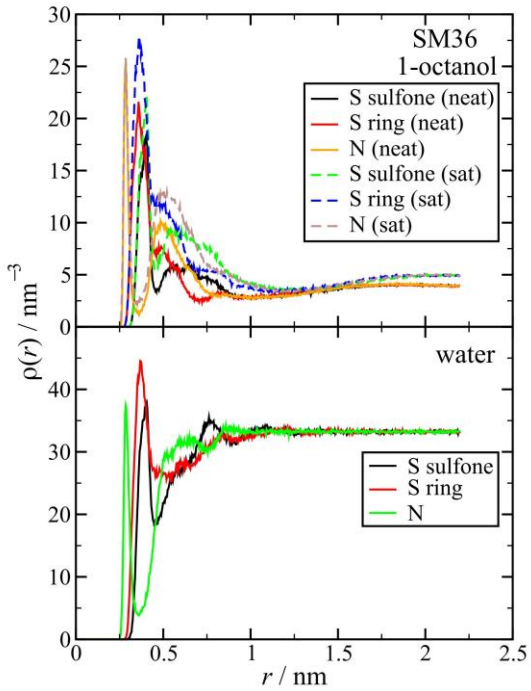


*Figure 5: Local density of solvent O around key portions of SM36. The top portion is for neat and saturated 1-octanol, and the bottom pane is for water.*

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 Fig. 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 like 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 found 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.

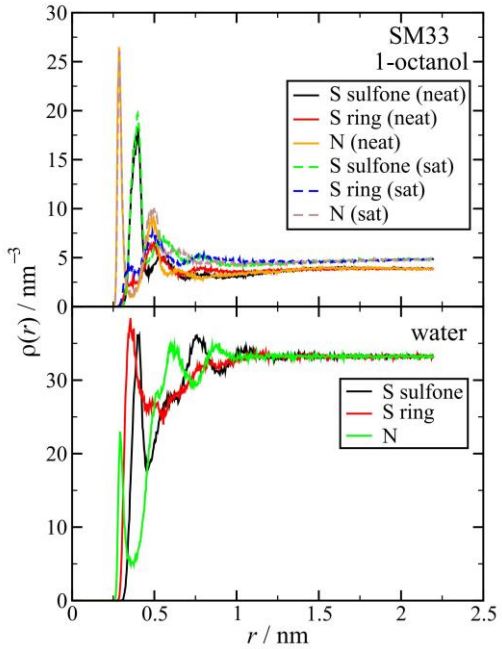


*Figure 6: Local density of solvent O around key portions of SM33. The top portion is for neat and saturated 1-octanol, and the bottom pane is for water.*

Compare this to the case of SM33 in Fig. 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 like 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 found 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 $\log_{10} P_1^{o*/w}$ 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 [89]. 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 [90]. 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 Fig. 7 we consider the case of SM39. For this case the ring sulfur is now in a sulfone group. The results are like 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 increases are 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.

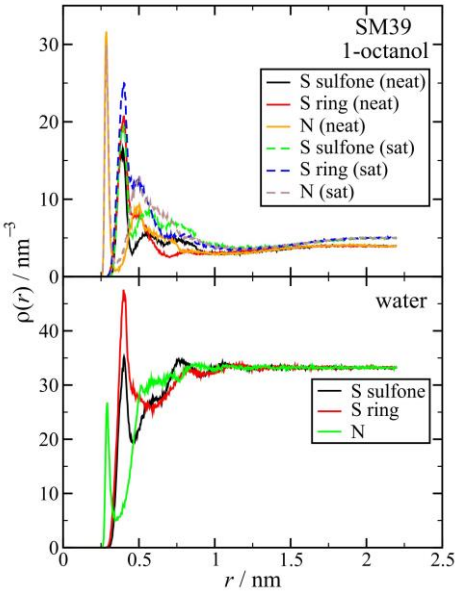


*Figure 7: : Local density of solvent O around key portions of SM39. The top portion is for neat and saturated 1-octanol, and the bottom pane is for water*

**Simulation Time**

For the solvation free energy calculations, for each stage we performed a 17.5 ns simulation where the first 1.5 ns was discarded from analysis as equilibration. This protocol is based on our

previous work using TraPPE-UA and the TIP4P water model [66–69]. However, in the context of SAMPL6 and SAMPL7, it has been shown that care must be taken when performing solvation free energy calculations in 1-octanol [91, 92]. As compared to water, the equilibration time of 1-octanol is much longer, requiring longer simulation times to ensure accurate results. Following their experience with SAMPL6, for SAMPL7 Beckstein and co-workers performed 50 ns simulations for each stage as compared to 5 ns simulations in SAMPL6 [91, 92], where 1-octanol was modeled using CHARMM, OPLS-AA, and GAFF force fields [32, 93–95].

A major difference of the present study is our use of a united atom model (TraPPE-UA) for 1-octanol, as compared to an all-atom model (CHARMM, OPLS-AA, and GAFF). We expect that the self-diffusion coefficient of a united atom model is larger than an analogous all atom model, which would lead to a shorter equilibration time [96]. From the long 100 ns NPT simulations for molecules SM30, SM33, SM34, SM36, SM37, and SM39 in pure water and pure 1-octanol, we computed the self-diffusion coefficient of TIP4P water and TraPPE-UA 1-octanol using the Einstein relation from the mean squared displacement computed using "gmx msd" in GROMACS 2020.2 [57–60]. We used the default options in "gmx msd" which determined the self-diffusion coefficient by linear regression of the mean squared displacement from 10 to 90 ns. Taking the average value computed from the six independent simulations and estimating the uncertainty as the corresponding standard deviation, for TIP4P water we compute a diffusion coefficient of $3.62 \pm 0.03 \times 10{-}5$ $cm^2/s$ at 298.15 K. This is in reasonable agreement with the reference published value of $3.73 \pm 0.02 \times 10{-}5$ $cm^2/s$ [96]. For TraPPE-UA 1-octanol we compute a diffusion coefficient of $0.171 \pm 0.009 \times 10{-}5$ $cm^2$ /s at 298.15 K. For comparison, the diffusion coefficient of 1-octanol modeled using OPLS-AA was recently computed to be $0.02 \pm 0.01 \times 10{-}5$ $cm^2$ /s at 298.15 K [98, 99]. We find that the self-diffusion coefficient of TraPPE-UA 1-octanol is an order of magnitude larger than OPLS-AA 1-octanol, which would translate to a shorter equilibration time using TraPPE-UA 1-octanol and the reason for the shorter simulation times used here. Nonetheless, based on the work on Beckstein and co-workers [91], this should be studied further in the future

# 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 [12], 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 [24] 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. Like the findings of the SAMPL6 challenge, we find that the predictions are sensitive to the choice of force fields [18]. 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 (Fig. 1), we expect hydrogen bonding to be more important

with the SAMPL7 solute molecules. As demonstrated by Chen and Siepmann [24], 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, this should not be interpreted as a recommendation to ignore the effect of water saturation. 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.

In the present study our focus was on the effect of water saturation on the prediction of octanol-water partition coefficients. For a comparison of the different prediction methodologies used by the SAMPL7 participants, please see the SAMPL7 overview article [12].

# Future Works: SAMPL8

# Introduction

The partitioning of a solute between two solvents has numerous applications in industry, particularly in the pharmaceutical industry, as shown above in the SAMPL7 challenge [12]. While partition coefficients are crucial to determining the dispersion of a neutral pharmaceutical in a human body, the use of partition coefficients is merely an assumption to simplify the actual behavior of the drugs in our body [101]. When a drug enters the aqueous phase within our body, such as our blood or other pH-dependent regions, there is the possibility for acid-base interactions between the aqueous phase and the drug, leading to the formation of charged or reacted species such as cations, anions, or tautomeric species [102]. Identifying the presence of these charged or reacted species can be key for future drug development, as protonation of certain drugs can lead to degradation of the drug in our bodies, rendering the compound useless for its intended need [103]. To account for this pH dependence and potential generation of charged/reacted species, we extend the definition of the partition coefficient in equation (1) to formulate the distribution coefficient.

As shown in equation (3), the distribution coefficient's form is identical to the partition coefficient's form, but it also considers the concentration of all charged and reactive species in both the aqueous phase and the organic phase. When charged/reacted species are generated in the aqueous phase, we see a diffusion of these new species into the organic phase, which is why we consider the concentration of charged/reacted species in the 1-octanol phase [104]. While partition coefficients are traditionally evaluated in a water/1-octanol system due to its ability to simulation partitioning of drugs in a body, distribution coefficients are often calculated with varying, more simple organic phases such as cyclohexane [105]. Solvents like cyclohexane or ethyl acetate are typically used since they are much more rigid and don't form complex heterogeneous micelle conformations, unlike 1-octanol, making the system much easier to model [106, 107].

While using bulkier, well-behaved solvents like cyclohexane is beneficial for facilitating computational estimates, distributions coefficients cannot be computed directly using physical molecular dynamics simulations, unlike partition coefficients from SAMPL7. Classical molecular dynamics cannot directly model formation and breaking of bonds. Several force fields have been developed to attempt to account for reactions and breaking bonds, but these are out of the scope of the discussion [108]. Due to this limitation, we cannot utilize the same methods employed in SAMPL7 to compute distribution coefficients and must rely on other methods such as electronic structure calculations. For the SAMPL8 challenge, we were asked to provide predictions on the macroscopic pKa values and distribution coefficients for every solvent and molecule as shown below in Figure (8). Without the option of using classical molecular modeling for this task, another computational method must be used to aid in predictions of pKa and logD values. Upon further inspection, we have found that electronic structure calculations are the most sufficient and widely available method to predicting these properties. This method of computing solvation free energy does not explicitly model reactions or collisions like molecular dynamics, but rather the main goal is to treat the system as a continuum of particles

with certain electrical properties and then using this continuum to approximate the solvation free energy of the solute in your bulk phase, which can be a better way to compute solvation free energies due to most ionic or tautomeric species being generated in solution rather than in the traditional gas phase [109].

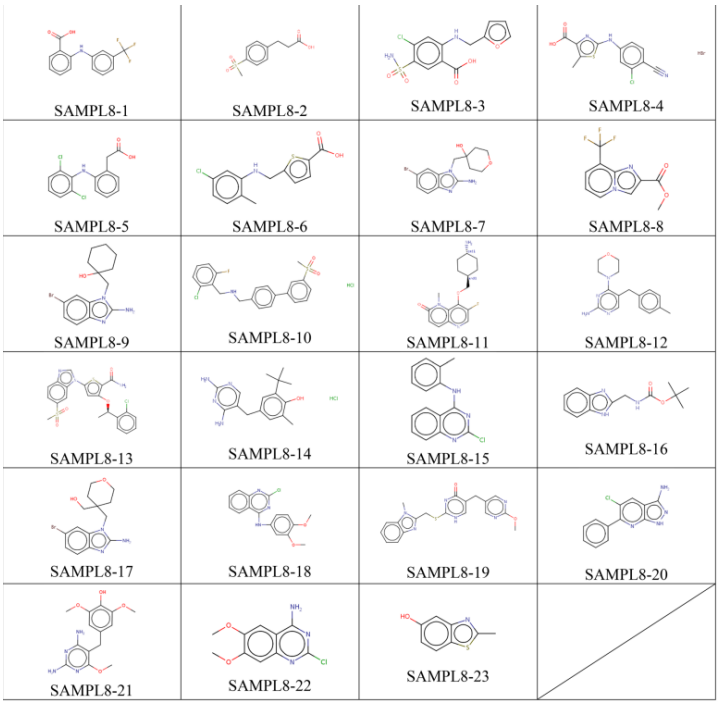


*Figure 8: Chemical structure and name of each molecule of interest in the SAMPL8 challenge*

# Methodology

## pKa

We computed pKa values by using electronic structure calculations through Gaussian. First, we obtained our data set of molecules and their respective microstates. We generated .mol2 files for each microstate and used these files throughout our work. We then optimized the geometry of each system in vacuum using the PM7 level of theory. PM7 is a type of modified neglect of diatomic differential overlap (MNDO) semiempirical model that approximates physical properties of a system based on quantum orbital theory combined with empirical data correlations to allow for better property prediction from already tabulated data [110, 111]. This method allowed us to calculate the partial charges for each structure and act as a check for determining the accuracy of the structures after optimization occurred. The structures were further optimized using B3LYP/def2-TZVP level of theory, a basis set available to model and describe the general behavior of your system and its electronic density function and wave function accordingly. While optimizing our structures, some of our molecules of interest were broken apart, possibly due to unfavorable initial configurations, so to complete the optimization, we froze some of the bonds in our systems to preserve the original molecule and not generate

new species even before the initial reaction. We evaluated the optimized structures using Avogadro to ensure the partial charge did not change due to structural changes because of the initial configurations [112, 113]. If the initial configuration has odd geometry, sometimes the Gaussian software can perceive that the structure changed, which can heavily affect our electronic structure calculations. Next, we computed single point energies to determine our CM5 partial charges and use Gaussian and SMD to predict our pKa values for various microstates followed by Boltzmann weighting to determine macroscopic pKa values.

## Distribution Coefficients (LogD)

Distribution coefficients were determined using uESE and the pKa values previously calculated.

### uESE

uESE is a newly developed, novel electronic structure calculation method derived from the COSMO solvation model. This model was developed to make a reliable, easy-to-use solvation model with an error below 1 kcal/mol that could be readily used on any neutral system [114-117]. This model was readily optimized around the Minnesota Solvation Database, containing solvent parameters and very accurate fitting to most of the compounds in the database. To use this model, all that is needed is the dielectric constant of your system, provided that is not already optimized in the system, and atomic partial charges of the molecules of interest using certain partial charge theories such as CM5 or ESP [118-121]. With the partial charges and dielectric constant, if using unoptimized systems, available, one can yield optimal estimates of hydration and solvation free energies for any neutral system [114-117].

We used uESE to run electronic structure calculations to compute solvation free energy values and partition coefficients for the list of provided solvents for each system. For some of our solvents, there was no data included in the original program, so we defined solvents based on their dielectric constant and other thermodynamic properties. For example, thermodynamic values for TBME were not readily found, so we parameterized the solvent using several different methods. For some cases, we used literature values we found in research papers, but one method we employed was estimating the behavior of TBME by parameterizing it with values from diethyl ether. Diethyl ether was used since its dielectric constant is very close to the dielectric constant of TBME, making our free energy calculations more precise. Next, we used equation (16) to relate the previously calculated solvation free energy values using uESE to compute distribution coefficient values for each solvent.

The work for SAMPL8 is forthcoming and will soon be published in the Journal of Computer Aided Molecular Design

# References


1. Sabatino, S.J., Paluch, A.S. Predicting octanol/water partition coefficients using molecular simulation for the SAMPL7 challenge: comparing the use of neat and water saturated 1-octanol. J Comput Aided Mol Des 35, 1009–1024 (2021). https://doi.org/10.1007/s10822-021-00415-4
2. Ledley, F. D.; McCoy, S. S.; Vaughan, G.; Cleary, E. G. Profitability of Large Pharmaceutical Companies Compared with Other Large Public Companies. *JAMA* **2020**, *323* (9), 834.
3. Ayati, N.; Saiyarsarai, P.; Nikfar, S. Short and Long Term Impacts of COVID-19 on the Pharmaceutical Sector. *DARU Journal of Pharmaceutical Sciences* **2020**, *28* (2), 799–805.
4. Akash, M. S.; Rehman, K.; Sabir, S.; Gul, J.; Hussain, I. Review Potential Risk Assessment of Pharmaceutical Waste: Critical Review and Analysis. *Pakistan Journal of Scientific & Industrial Research Series A: Physical Sciences* **2020**, *63* (3), 209–219.
5. Fatta-Kassinos, D., Kalavrouziotis, I.K., Koukoulakis, P.H., Vasquez, M.I. 2011. The risks associated with wastewater reuse and xenobiotics in the agroecological environment. The Science of the Total Environment, 409: 3555-3563.
6. Fatta-Kassinos, D., Meric, S., Nikolaou, A. 2011. Pharmaceutical residues in environmental waters and wastewater: current state of knowledge and future research. Analytical and Bioanalytical Chemistry, 399: 251-275.
7. Alshemari, A.; Breen, L.; Quinn, G.; Sivarajah, U. Can We Create a Circular Pharmaceutical Supply Chain (CPSC) to Reduce Medicines Waste? *Pharmacy* **2020**, *8* (4), 221.
8. Veleva, V.; Bodkin, G.; Todorova, S. The Need for Better Measurement and Employee Engagement to Advance a Circular Economy: Lessons from Biogen's "Zero Waste" Journey. *Journal of Cleaner Production* **2017**, *154*, 517–529.
9. Eldin, A. B.; Ismaiel, O. A.; Hassan, W. E.; Shalaby, A. A. Green Analytical Chemistry: Opportunities for Pharmaceutical Quality Control. *Journal of Analytical Chemistry* **2016**, *71* (9), 861–871.
10. Richon, A. B. An Early History of the Molecular Modeling Industry. *Drug Discovery Today* **2008**, *13* (15-16), 659–664.
11. Hendrickson, J. B. Molecular Geometry. I. Machine Computation of the Common Rings. *Journal of the American Chemical Society* **1961**, *83* (22), 4537–4547.
12. Bergazin, T. D.; Tielker, N.; Zhang, Y.; Mao, J.; Gunner, M. R.; Francisco, Karol; Ballatore, C.; Kast, S.; Mobley, D. Evaluation of Log P, PKA and Log D Predictions from the Sampl7 Blind Challenge. *Journal of Chemical & Engineering Data*. ***2021***.
13. Schoensee, C.; Bucheli, T. Experimental Determination of Octanol-Water Partition Coefficients of Selected Natural Toxins. **2019**.
14. OECD. 107: Partition Coefficient (n-octanol/water): Shake Flask Method. OECD Guidelines for the Testing of Chemicals, 1995
15. Sangster J (1989) Octanol-water partition coefficients of simple organic compounds. J Phys Chem Ref Data 18:1111–1227
16. Leo A, Elkins CHD (1971) Partition coefficients and their uses. Chem. Rev. 71:525–616

17. Ouimet JA, Paluch AS (2020) 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.
18. Işık M, Bergazin TD, Fox T, Rizzi A, Chodera JD, Mobley DL (2020) 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.
19. Bahr, M.; Nandkeolyar, A.; Kenna, J.; DaVia, L.; Nevins, N.; Isik, M.; Chodera, J.; Mobley, D. Automated High Throughput PKA and Distribution Coefficient Measurements of Pharmaceutical Compounds for the SAMPL8 Blind Prediction Challenge. **2021**.
20. Bannan CC, Calabro G, Kyu DY, Mobley DL (2016) Calculating partition coefficients of small molecules in octanol/water and cyclohexane/water. J Chem Theory Comput 12(8):4015–4024
21. Sørensen JM, Arlt W (eds) (1979) Liquid-liquid equilibrium data collection, part 1: binary systems. DECHEMA, Frankfurt a. M.
22. Tse G, Sandler SI (1994) Determination of infinite dilution activity coefficients and 1-octanol/water partition coefficients of volatile organic pollutants. J Chem Eng Data 39:354–357
23. Roese SN, Heintz JD, Uzat CB, Schmidt AJ, Margulis GV, Sabatino SJ, Paluch AS (2020) Assessment of the SM12, SM8, and SMD solvation models for predicting limiting activity coefficients at 298.15 K. Processes 8:623
24. Chen B, Siepmann JI (2006) Microscopic structure and solvation in dry and wet octanol. J Phys Chem B 110:3555–356
25. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
26. Martin MG, Siepmann JI (1998) Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. J Phys Chem B 102:2569–2577
27. Chen B, Potof JJ, Siepmann JI (2001) 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
28. Chen B, Siepmann JI (2000) Partitioning of alkane and alcohol solutes between water and (dry or wet) 1-octanol. J Am Chem Soc 122:6464–6467
29. MacCullum JL, Tieleman DP (2002) Structures of neat and hydrated 1-octanol from computer simulations. J Am Chem Soc 124:15085–15093
30. SAMPL7 logP Prediction Challenge. https://github.com/sampl challenges/SAMPL7/tree/master/physical_property. Accessed 11 March, 2021
31. Abascal JLF, Vega C (2005) A general purpose model for the condensed phase of water: TIP4P/2005. J Chem Phys 123:234505
32. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174
33. Leach AR (2001) Molecular modelling: principles and applications, 2nd edn. Pearson Education Limited, Harlow
34. Frenkel D, Smit B (2002) Understanding molecular simulation: from algorithms to applications, 2nd edn. Academic Press, San Diego, CA
35. Abascal JLF, Vega C (2005) A general purpose model for the condensed phase of water: TIP4P/2005. J Chem Phys 123:234505

36. 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., SchottVerdugo, 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)
37. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graphics Modell 25:247–260
38. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36
39. Daylight Chemical Information Systems, Inc. https://www.dayli ght.com/. Accessed 11 March, (2021)
40. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchinson GR (2011) Open babel: an open chemical toolbox. J Cheminf 3:33
41. Open Babel: The Open Source Chemistry Toolbox. http://openb abel.org/wiki/Main_Page. Accessed 26 June, (2019)
42. Gasteiger J, Marsili M (1978) A new model for calculating atomic charges in molecules. Tetrahedron Lett 34:3181–3184
43. Zhao Y, Truhlar DG (2008) 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
44. Cramer CJ (2002) Essentials of computational chemistry. Wiley, Chichester, West Sussex
45. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, 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, JA, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2019) Gaussian 16, Revision C01
46. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 97:10269–10280
47. Cieplak P, Cornell WD, Bayly C, Kollman PA (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. J Comput Chem 16:1357–1377

48. Martínez L, Andrade R, Birgin EG, Martínez JM (2009) Packmol: a package for building initial confgurations for molecular dynamics simulations. J Comput Chem 30(13):2157–2164
49. Packmol: Packing Optimization for Molecular Dynamics Simulations. http://www.ime.unicamp.br/~martinez/packmol/. Accessed 31 Mar (2021)
50. GROMACS development team: GROMACS Documentation: Release 2020.2 (2020). https://manual.gromacs.org/documentat ion/2020.2/manual-2020.2.pdf
51. Berendsen HJC (2007) Simulating the physical world: Hierarchial modeling from quantum mechanics to fuid dynamics. Cambridge University Press, New York
52. Berendsen HJC, Postma JPM, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
53. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity-rescaling. J Chem Phys 126:014101
54. Bussi G, Parrinello M (2008) Stochastic thermostats: comparison of local and global schemes. Compd Phys Commun 179:26–29
55. Bussi G, Zykova-Timan T, Parrinello M (2009) Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. J Chem Phys 130:74101
56. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 52:7182–7190
57. Hess B, Kutzner C, van der Spoel D, Lindal E (2008) GROMACS 4: Algorithms for highly efcient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447
58. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open-source molecular simulation toolkit. Bioinformatics 29:845–854
59. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25
60. GROMACS: Fast, flexible, free. http://www.gromacs.org/. Accessed 31 March (2021)
61. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472
62. Hess B (2008) P-LINCS: a parallel linear constraint solver for molecular simulation. J Chem Theory Comput 4:116–122
63. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341
64. Miyamoto S, Kollman PA (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithms for rigid water models. J Comput Chem 13:952–962
65. Deserno M, Holm C (1998) How to mesh up Ewald sums. i. a theoretical and numerical comparison of various particle mesh routines. J Chem Phys 109:7678–7693
66. Fuerst GB, Ley RT, Paluch AS (2015) 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

67. Ley RT, Fuerst GB, Redeker BN, Paluch AS (2016) 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
68. Long GE, Dhakal P, Redeker BN, Paluch AS (2019) 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
69. 2019 Solvation Free Energy Calculation Workshop at UFF. https://sites.google.com/a/miamioh.edu/2019-free-energy-workshop-atuf/. Accessed 31 March (2021)
70. Shing KS, Chung ST (1987) Computer simulation methods for the calculation of solubility in supercritical extraction systems. J Phys Chem 91:1674–1681
71. Kofke DA, Cummings PT (1997) Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation. Mol Phys 92:973–996
72. Shirts MR, Pitera JW, Swope WC, Pande VS (2003) 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
73. Kofke DA, Cummings PT (1998) Precision and accuracy of staged free-energy perturbation methods for computing the chemical potential by molecular simulation. Fluid Phase Equilib 150–151:41–49
74. Chipot C, Pohorille A (eds) (2007) Free energy calculations: theory and applications in chemistry and biology. Springer series in chemical physics, vol 86. Springer, New York
75. Bennett CH (1976) Efficient estimation of free energy differences from monte carlo data. J Comput Phys 22:245–268
76. Shirts MR, Bair E, Hooker G, Pande VS (2003) Equilibrium free energies from nonequilibrium measurements using maximum likelihood methods. Phys Rev Lett 91:140601
77. Lu N, Singh JK, Kofke DA (2003) Appropriate methods to combine forward and reverse free-energy perturbation averages. J Chem Phys 118:2977–2984
78. Shirts MR, Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. J Chem Phys 129:124105
79. Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF (1994) Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chem Phys Lett 222:529–539
80. Shirts MR, Pande VS (2005) Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. J Chem Phys 122:134508
81. Steinbrecher T, Mobley DL, Case DA (2007) Nonlinear scaling schemes for Lennard–Jones interactions in free energy calculations. J Chem Phys 127:214108
82. PyMBAR: Python implementation of the multistate bennett acceptance ratio (mbar). https://github.com/choderalab/pymbar. Accessed 31 March (2021)
83. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA (2007) Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. J Chem Theory Comput 3:26–41
84. Klimovich PV, Shirts MR, Mobley DL (2015) Guidelines for analysis of free energy calculations. J Comput-Aided Mol Des 29:397–411
85. Chodera JD (2016) A simple method for automated equilibration detection in molecular simulations. J Chem Theory Comput 12:1799–1805

86. 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 $\lambda_m^{elec}$= {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 MR, Mobley DL (2013) An introduction to best practices in free energy calculations. In Monticelli L, Salonen E (eds) Biomolecular simulations: methods and protocols, pp 271–311. Humana Press, Totowa, NJ
87. van Gunsteren WF, Berendsen HJC (1988) A leap-frog algorithm for stochastic dynamics. Mol Simul 1:173–185
88. Brehm M, Kirchner B (2011) TRAVIS—a free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories. J Chem Inf Model 51:2007–2023
89. Brehm M, Thomas M, Gehrke S, Kirchner B (2020) TRAVIS—a free analyzer for trajectories from molecular simulation. J Chem Phys 152:164105
90. Jeffrey GA (1997) An introduction to hydrogen bonding. Oxford University Press, New York
91. Fan S, Nedev H, Vijayan R, Iorga BI, Beckstein O (2021) Precise force-field-based calculations of octanol-water partition coefficients for the SAMPL7 molecules. J Comput-Aided Mol Des 35:853–870
92. Fan S, Iorga BI, Beckstein O (2020) Prediction of octanol-water partition coefficients for the SAMPL6-log P molecules using molecular dynamics simulations with OPLS-AA, AMBER and CHARMM force felds. J Comput-Aided Mol Des 34:543–560
93. MacKerell Jr., AD, Wiorkiewicz-Kuczera J, Karplus M (1995) An all-atom empirical energy function for the simulation of nucleic acids. J Am Chem Soc 117, 11946–11975
94. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31:671–690
95. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and Testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236
96. Kondratyuk N, Lenev D, Pisarev V (2020) Transport coefficients of model lubricants up to 400 MPa from molecular dynamics. J Chem Phys 152:191104
97. van der Spoel D, van Maaren PJ (2006) The origin of layer structure artifacts in simulations of liquid water. J Chem Theory Comput 2:1–11
98. Caleman C, van Maaren PJ, Hong M, Hub JS, Costa LT, van der Spoel D (2012) Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. J Chem Theory Comput 8:61–74
99. virtualchemistry.org. http://virtualchemistry.org/molecule.php? filename=1-octanol.sdf#refs. Accessed 11 March (2021)
100. Ohio Supercomputer Center: Ohio Supercomputer Center (1987). http://osc.edu/ark:/19495/f5s1ph73
101. Utsey K, Gastonguay MS, Russell S, Freling R, Riggs MM, Elmokadem A (2020) Quantification of the impact of partition coefficient prediction methods on physiologically based pharmacokinetic model output using a standardized tissue composition. Drug Metab Dispos 48(10):903–916

102. Bannan CC, Calabro G, Kyu DY, Mobley DL (2016) Calculating partition coefficients of small molecules in octanol/water and cyclohexane/water. J Chem Theory Comput 12(8):4015–4024
103. Fenton, O.S.; Katy, N.O.; Pillai, P.S.; Mitchell, M.J.; Langer, R. Advances in Biomaterials for Drug Delivery. Adv. Mater. 2018, 30, 1705328.
104. Hannesschlaeger, C.; Horner, A.; Pohl, P. Intrinsic Membrane Permeability to Small Molecules. *Chemical Reviews* **2019**, *119* (9), 5922–5953.
105. Arnott JA, Planey SL (2012) The influence of lipophilicity in drug discovery and design. Expert Opin Drug Discov 7(10):863–875
106. Linkov I, Ames MR, Crouch EA, Satterstrom FK (2005) Uncertainty in octanol-water partition coefficient: implications for risk assessment and remedial costs. Environ Sci Technol 39(18):6917–6922
107. Schönsee CD, Bucheli TD (2020) Experimental determination of octanol-water partition coefficients of selected natural toxins. J Chem Eng Data 65(4):1946–1953
108. Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; van Duin, A. C. The REAXFF Reactive Force-Field: Development, Applications and Future Directions. *npj Computational Materials* **2016**, *2* (1).
109. Voityuk, A. A.; Vyboishchikov, S. F. A Simple Cosmo-Based Method for Calculation of Hydration Energies of Neutral Molecules. *Physical Chemistry Chemical Physics* **2019**, *21* (34), 18706–18713.
110. Stewart, J. J. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *Journal of Molecular Modeling* **2012**, *19* (1), 1–32.
111. Dewar, M. J.; Thiel, W. Ground States of Molecules. 38. the MNDO Method. Approximations and Parameters. *Journal of the American Chemical Society* **1977**, *99* (15), 4899–4907.
112. Avogadro: an open-source molecular builder and visualization tool. Version 1.XX. http://avogadro.cc/
113. Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek and Geoffrey R Hutchison; "Avogadro: An advanced semantic chemical editor, visualization, and analysis platform" Journal of Cheminformatics 2012, 4:17.
114. S. F. Vyboishchikov, uESE program, Girona,2021
115. S. F. Vyboishchikov, A. A. Voityuk, J. Comput. Chem.,2021,42. DOI: 10.1002/jcc.26531
116. A. A. Voityuk, S. F. Vyboishchikov, Phys. Chem. Chem. Phys.2020,22, 14591–14598. DOI:10.1039/d0cp02667k
117. A. A. Voityuk, S. F. Vyboishchikov, Phys. Chem. Chem. Phys.,2019,21, 875–874. DOI:10.1039/c9cp03010g
118. S. R. Cox and D. E. Williams,J. Comput. Chem., 1981,2, 304.
119. U. C. Singh and P. A. Kollman,J. Comput. Chem., 1984,5,129–145, DOI: 10.1002/jcc.540050204
120. B. H. Besler, K. M. Merz Jr. and P. A. Kollman,J. Comput.Chem., 1990,11, 431–439, DOI: 10.1002/jcc.540110404.
121. A. V. Marenich, S. V. Jerome, C. J. Cramer and D. G. Truhlar,J. Chem. Theory Comput., 2012,8, 527–541.

# Appendix I: Dimensionless Solvation Free Energies

*Table S1: A summary of the dimensionless solvation free energies in water using both TIP4P and TIP4P 2005 water models. The total values are broken down into their respective LJ and electrostatic components. The subscripts correspond to the error in last two decimal places.*

| TIP4P | | | | TIP4P 2005 | | |
|---|---|---|---|---|---|---|
| Molecule Number | Electrostatic | LJ | Total | Electrostatic | LJ | Total |
| 25 | $-24.56_{04}$ | $3.15_{08}$ | $-21.41_{08}$ | $-24.71_{10}$ | $2.67_{10}$ | $-22.04_{15}$ |
| 26 | $-30.12_{08}$ | $2.42_{06}$ | $-27.70_{10}$ | $-30.47_{08}$ | $1.73_{09}$ | $-28.75_{12}$ |
| 27 | $-26.62_{12}$ | $3.22_{07}$ | $-23.40_{14}$ | $-26.44_{12}$ | $2.36_{09}$ | $-24.08_{15}$ |
| 28 | $-36.38_{08}$ | $2.41_{07}$ | $-33.97_{11}$ | $-36.05_{07}$ | $1.59_{10}$ | $-34.47_{12}$ |
| 29 | $-27.63_{12}$ | $2.93_{07}$ | $-24.70_{14}$ | $-28.00_{10}$ | $2.09_{09}$ | $-25.91_{14}$ |
| 30 | $-27.64_{14}$ | $3.74_{09}$ | $-23.90_{16}$ | $-27.16_{14}$ | $3.13_{12}$ | $-24.03_{18}$ |
| 31 | $-26.55_{14}$ | $3.59_{08}$ | $-22.96_{16}$ | $-26.65_{15}$ | $3.19_{11}$ | $-23.46_{19}$ |
| 32 | $-24.00_{08}$ | $2.31_{07}$ | $-21.69_{11}$ | $-23.95_{09}$ | $1.54_{10}$ | $-22.41_{13}$ |
| 33 | $-23.66_{07}$ | $3.38_{09}$ | $-20.28_{11}$ | $-23.98_{09}$ | $2.88_{13}$ | $-21.10_{16}$ |
| 34 | $-22.29_{18}$ | $3.05_{08}$ | $-19.23_{20}$ | $-22.41_{03}$ | $2.81_{01}$ | $-19.60_{11}$ |
| 35 | $-35.33_{11}$ | $2.31_{07}$ | $-33.02_{13}$ | $-35.90_{11}$ | $1.48_{09}$ | $-34.42_{14}$ |
| 36 | $-41.21_{13}$ | $3.23_{09}$ | $-37.98_{16}$ | $-41.63_{13}$ | $2.19_{13}$ | $-39.44_{18}$ |
| 37 | $-35.83_{11}$ | $2.88_{08}$ | $-32.95_{14}$ | $-35.80_{12}$ | $2.00_{11}$ | $-33.80_{16}$ |
| 38 | $-36.35_{14}$ | $1.69_{07}$ | $-34.67_{16}$ | $-36.33_{21}$ | $0.44_{11}$ | $-35.89_{24}$ |
| 39 | $-37.78_{14}$ | $2.60_{09}$ | $-35.19_{17}$ | $-37.53_{16}$ | $1.55_{13}$ | $-35.98_{21}$ |
| 40 | $-34.59_{14}$ | $2.36_{08}$ | $-32.24_{16}$ | $-34.73_{11}$ | $1.40_{12}$ | $-33.33_{16}$ |
| 41 | $-25.83_{06}$ | $1.23_{06}$ | $-24.60_{08}$ | $-25.60_{10}$ | $0.16_{07}$ | $-25.44_{12}$ |
| 42 | $-29.63_{10}$ | $1.35_{07}$ | $-28.28_{12}$ | $-28.84_{12}$ | $0.27_{10}$ | $-28.57_{15}$ |
| 43 | $-28.34_{03}$ | $1.50_{06}$ | $-26.85_{07}$ | $-27.80_{04}$ | $0.34_{08}$ | $-27.46_{09}$ |
| 44 | $-26.36_{04}$ | $0.22_{06}$ | $-26.14_{07}$ | $-25.83_{05}$ | $-1.02_{07}$ | $-26.84_{09}$ |
| 45 | $-28.72_{09}$ | $1.07_{07}$ | $-27.65_{12}$ | $-28.28_{10}$ | $-0.11_{10}$ | $-28.39_{14}$ |
| 46 | $-24.78_{04}$ | $1.01_{06}$ | $-23.76_{07}$ | $-24.35_{06}$ | $-0.25_{09}$ | $-24.60_{11}$ |

*Table S2: A summary of the dimensionless solvation free energies in pure 1-octanol.. The total values are broken down into their respective LJ and electrostatic components. The subscripts correspond to the error in last two decimal places.*

| $\Delta G_{1,o}^{solv}$ / RT | | | |
|---|---|---|---|
| Molecule Number | Electrostatic | LJ | Total |
| 25 | $-6.91_{04}$ | $-21.34_{08}$ | $-28.26_{09}$ |
| 26 | $-10.49_{03}$ | $-17.52_{07}$ | $-28.01_{08}$ |
| 27 | $-8.34_{02}$ | $-19.42_{08}$ | $-27.75_{08}$ |
| 28 | $-16.09_{34}$ | $-21.12_{08}$ | $-37.20_{35}$ |
| 29 | $-9.05_{02}$ | $-19.34_{08}$ | $-28.40_{08}$ |
| 30 | $-8.79_{02}$ | $-23.56_{09}$ | $-32.35_{10}$ |
| 31 | $-6.79_{02}$ | $-20.74_{09}$ | $-27.53_{09}$ |
| 32 | $-8.02_{08}$ | $-20.67_{08}$ | $-28.69_{11}$ |
| 33 | $-7.74_{01}$ | $-25.25_{10}$ | $-32.99_{10}$ |
| 34 | $-7.42_{02}$ | $-22.16_{09}$ | $-29.58_{10}$ |
| 35 | $-14.66_{02}$ | $-21.75_{08}$ | $-36.40_{09}$ |
| 36 | $-14.85_{02}$ | $-26.60_{10}$ | $-41.45_{10}$ |
| 37 | $-12.69_{04}$ | $-23.63_{09}$ | $-36.33_{10}$ |
| 38 | $-14.02_{02}$ | $-22.97_{08}$ | $-36.99_{08}$ |
| 39 | $-14.10_{12}$ | $-27.51_{11}$ | $-41.61_{16}$ |
| 40 | $-11.68_{02}$ | $-25.14_{10}$ | $-36.82_{10}$ |
| 41 | $-11.20_{05}$ | $-19.83_{06}$ | $-31.03_{08}$ |
| 42 | $-10.91_{30}$ | $-25.60_{08}$ | $-36.51_{31}$ |
| 43 | $-10.68_{02}$ | $-22.47_{07}$ | $-33.15_{07}$ |
| 44 | $-8.99_{02}$ | $-20.95_{07}$ | $-29.93_{07}$ |
| 45 | $-10.29_{03}$ | $-25.48_{08}$ | $-35.77_{08}$ |
| 46 | $-6.91_{19}$ | $-22.79_{07}$ | $-29.70_{20}$ |

*Table S3: A summary of the dimensionless solvation free energies in water-saturated 1-octanol using both TIP4P and TIP4P 2005 water models. The total values are broken down into their respective LJ and electrostatic components. The subscripts correspond to the error in last two decimal places.*

| | TIP4P | | | TIP4P 2005 | | |
|---|---|---|---|---|---|---|
| Molecule Number | Electrostatic | LJ | Total | Electrostatic | LJ | Total |
| 25 | $-9.64_{08}$ | $-21.22_{08}$ | $-30.86_{11}$ | $-8.16_{02}$ | $-21.12_{09}$ | $-29.29_{09}$ |
| 26 | $-13.97_{28}$ | $-17.53_{07}$ | $-31.50_{29}$ | $-14.43_{30}$ | $-17.35_{08}$ | $-31.78_{31}$ |
| 27 | $-9.70_{31}$ | $-19.00_{08}$ | $-28.70_{32}$ | $-8.95_{02}$ | $-19.16_{08}$ | $-28.10_{09}$ |
| 28 | $-17.46_{09}$ | $-21.11_{08}$ | $-38.57_{12}$ | $-17.59_{08}$ | $-21.13_{09}$ | $-38.72_{11}$ |
| 29 | $-10.85_{31}$ | $-19.29_{08}$ | $-30.14_{32}$ | $-9.66_{02}$ | $-19.11_{08}$ | $-28.77_{08}$ |
| 30 | $-7.73_{05}$ | $-23.28_{09}$ | $-31.01_{11}$ | $-9.55_{02}$ | $-23.04_{10}$ | $-32.60_{10}$ |
| 31 | $-8.70_{01}$ | $-20.77_{10}$ | $-29.47_{10}$ | $-9.03_{02}$ | $-20.46_{10}$ | $-29.49_{10}$ |
| 32 | $-8.27_{02}$ | $-20.54_{08}$ | $-28.82_{08}$ | $-9.48_{26}$ | $-20.69_{08}$ | $-30.18_{27}$ |
| 33 | $-6.39_{02}$ | $-24.61_{10}$ | $-31.00_{10}$ | $-9.37_{02}$ | $-24.79_{10}$ | $-34.17_{10}$ |
| 34 | $-8.74_{04}$ | $-22.32_{09}$ | $-31.06_{10}$ | $-7.36_{01}$ | $-21.71_{09}$ | $-29.07_{09}$ |
| 35 | $-16.27_{24}$ | $-21.34_{08}$ | $-37.61_{25}$ | $-16.67_{02}$ | $-21.58_{08}$ | $-38.25_{09}$ |
| 36 | $-19.71_{12}$ | $-26.47_{09}$ | $-46.17_{15}$ | $-18.41_{09}$ | $-25.79_{11}$ | $-44.20_{14}$ |
| 37 | $-14.20_{33}$ | $-23.48_{09}$ | $-37.68_{34}$ | $-15.07_{02}$ | $-23.56_{09}$ | $-38.62_{09}$ |
| 38 | $-16.15_{03}$ | $-22.71_{09}$ | $-38.86_{09}$ | $-14.20_{02}$ | $-22.93_{09}$ | $-37.13_{09}$ |
| 39 | $-15.92_{07}$ | $-27.10_{10}$ | $-43.02_{12}$ | $-14.03_{02}$ | $-27.20_{10}$ | $-41.23_{10}$ |
| 40 | $-13.83_{28}$ | $-24.68_{09}$ | $-38.51_{30}$ | $-13.46_{33}$ | $-24.46_{10}$ | $-37.92_{35}$ |
| 41 | $-12.02_{25}$ | $-19.60_{07}$ | $-31.62_{26}$ | $-12.51_{08}$ | $-19.56_{07}$ | $-32.06_{11}$ |
| 42 | $-11.88_{12}$ | $-25.06_{08}$ | $-36.94_{15}$ | $-11.82_{12}$ | $-25.25_{08}$ | $-37.06_{15}$ |
| 43 | $-11.86_{25}$ | $-22.37_{07}$ | $-34.23_{26}$ | $-12.56_{14}$ | $-22.26_{07}$ | $-34.81_{16}$ |
| 44 | $-11.58_{28}$ | $-20.72_{06}$ | $-32.29_{29}$ | $-10.77_{28}$ | $-20.64_{07}$ | $-31.41_{29}$ |
| 45 | $-10.57_{28}$ | $-25.30_{07}$ | $-35.87_{29}$ | $-11.41_{04}$ | $-25.17_{09}$ | $-36.57_{09}$ |
| 46 | $-8.24_{02}$ | $-22.72_{07}$ | $-30.96_{07}$ | $-7.32_{04}$ | $-22.62_{07}$ | $-29.94_{08}$ |

# Appendix II: Additional Structural Analysis

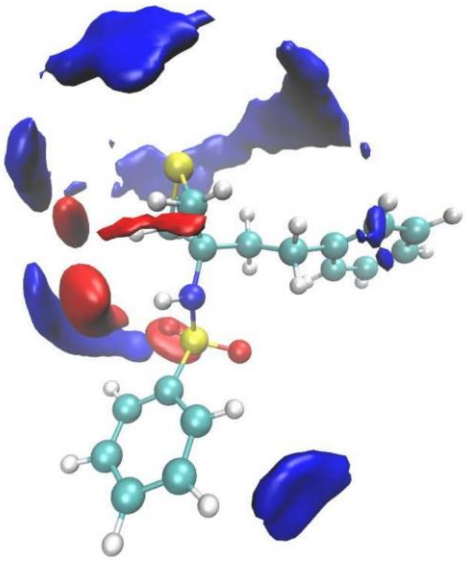

*Figure S1: Spatial distribution function (SDF) for the 1-octanol O (red) and methyl carbons (blue) around molecule SM33 in neat 1-octanol. The isosurfaces correspond to densities of 35 and 8 $nm^{-3}$ for O and the methyl carbons, respectively*

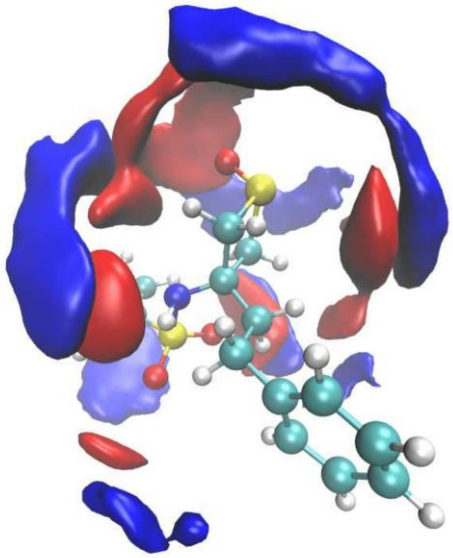

*Figure S2: Spatial distribution function (SDF) for the 1-octanol O (red) and methyl carbons (blue) around molecule SM36 in neat 1-octanol. The isosurfaces correspond to densities of 35 and 8 $nm^{-3}$ for O and the methyl carbons, respectively*

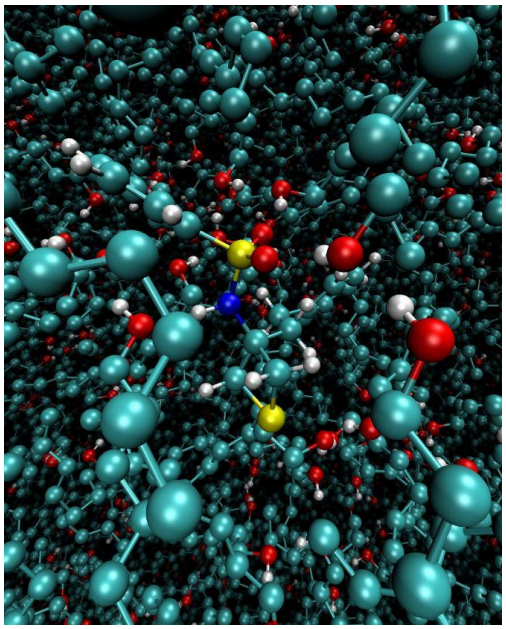

*Figure S3: A snapshot of the system SM33 in neat 1-octanol*

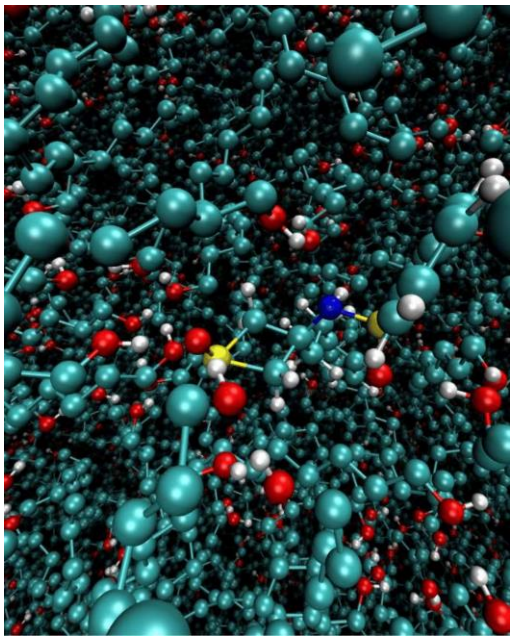

*Figure S4: A snapshot of the system SM36 in neat 1-octanol*