Blind prediction of cyclohexane-water distribution coefficients from the SAMPL5 challenge

Caitlin C. Bannan · Kalistyn H. Burley ·
Michael Chiu · Michael R. Shirts ·
Michael K. Gilson · David L. Mobley

Abstract In the recent SAMPL5 challenge, participants submitted predictions for cyclohexane/water distribution coefficients for a set of 53 small molecules. Distribution coefficients ($\log D$) replace the hydration free energies that were a central part of the past five SAMPL challenges. A wide variety of computational methods were represented by the 76 submissions from 18 participating groups. Here, we analyze submissions by a variety of error metrics and provide details for a number of reference calculations we performed. As in the SAMPL4 challenge, we assessed the ability of participants to evaluate not just their statistical uncertainty, but their model uncertainty – how well they can predict the magnitude of their model or force field error for specific predictions. Unfortunately, this remains an area where prediction and analysis need improvement. In SAMPL4 the top performing submissions achieved a root-mean-squared error (RMSE) around 1.5 kcal/mol. If we anticipate accuracy in $\log D$ predictions to be similar to the hydration free energy predictions in

C. C. Bannan
Department of Chemistry, University of California, Irvine

K. H. Burley
Department of Pharmaceutical Sciences, University of California, Irvine

M. Chiu
Qualcomm Institute, University of California, San Diego

M. R. Shirts
Chemical and Biological Engineering, University of Colorado, Boulder

M. K. Gilson
Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego

D. L. Mobley
Department of Chemistry and Department of Pharmaceutical Sciences, University of California, Irvine
147 Bison Modular, Irvine, CA 92697 Tel.: +949-824-6383
Fax: +949-824-2949
E-mail: dmobley@mobleylab.org

SAMPL4, the expected error here would be around 1.54 log units. Only a few submissions had an RMSE below 2.5 log units in their predicted $\log D$ values. However, distribution coefficients introduced complexities not present in past SAMPL challenges, including tautomer enumeration, that are likely to be important in predicting biomolecular properties of interest to drug discovery, therefore some decrease in accuracy would be expected. Overall, the SAMPL5 distribution coefficient challenge provided great insight into the importance of modeling a variety of physical effects. We believe these types of measurements will be a promising source of data for future blind challenges, especially in view of the relatively straightforward nature of the experiments and the level of insight provided.

Keywords SAMPL · distribution coefficient · blind challenge · free energy · alchemical · molecular simulation

1 Introduction

This year's Statistical Assessment of Modeling of Proteins and Ligand (SAMPL) challenge focuses on prediction of cyclohexane-water distribution coefficients. The inclusion of distribution coefficients replaces the previous focus on hydration free energies which was a fixture of the past five challenges (SAMPL0-4) [1–7]. Due to a lack of ongoing experimental work to generate new data, hydration free energies are no longer a practical property to include in blind challenges. It has become increasingly difficult to find unpublished or obscure hydration free energies and therefore impossible to design a challenge focusing on target compounds, functional groups or chemical classes. But this type of data is extremely valuable — the past SAMPL challenges have driven real improvements in a variety of methods for calculating hydration free energies [1] — so we sought to include a similar physical property in SAMPL5. The organizers of SAMPL5 settled on cyclohexane-water distribution coefficients, and thanks to a partnership with Genentech, this led to a series of measurements on drug-like compounds, discussed in detail in this issue [8]. The experimental measurements are also straightforward enough that future distribution coefficient challenges can be deliberately designed to focus on issues that merit attention to move the field forward.

Partition and distribution coefficients are important physical properties [9,10] which can provide a valuable opportunity for testing computational methods and molecular models. Distribution coefficients describe how all forms of a solute distributes itself across two immiscible solvents. In this case,

$$D = \frac{\sum_i [X_i]_{cyc}}{\sum_i [X_i]_{aq}} \tag{1}$$

where X_i represents a single protonation or tautomeric state of the solute in one of the solvents, and the sum runs over all protonation and tautomeric states [10]. Results are reported as the logarithm of this ratio ($\log D$). These

are more complicated than partition coefficients ($\log P$), which measure the concentration of the neutral solute in both solvents [9]. Specifically, if only one tautomer is relevant, $\log P$ is proportional to the transfer free energy, and can be calculated from solvation free energies [11–21]. In contrast, all relevant charged and neutral forms of the solute will need to be included to accurately calculate $\log D$, which can be estimated from a calculated $\log P$ and the relative populations of protonation states and tautomers in each solvent. Thus, accurate tautomer enumeration in both solvents may be an important part of predicting $\log D$, introducing new complexities to the SAMPL challenge which were avoided in previous hydration free energy challenges.

Here we give an overview of the analysis done for the SAMPL5 challenge, including the compounds considered, overall performance of submissions, and the metrics used for analysis. We also include details for a set of reference calculations we performed estimating $\log D$ as the cyclohexane/water partition coefficient as well as a series of follow-up studies focusing on the importance of tautomers in estimating $\log D$. Overall, we believe the outcome of the present SAMPL5 challenge highlights the potential benefits of this type of experimental data to improve computational methods, force fields, sampling algorithms, and treatment of protonation states and tautomers. Many of these issues will be highly relevant for nominally more challenging problems, such as predictions of protein-ligand binding affinities.

2 Challenge Logistics

SAMPL5 began on September 15, 2015 when the specifications and input files for the challenge were made available on the D3R website (http://drugdesigndata.org); these are also provided in the supporting information, made available on the University of California Dash (http://n2t.net/ark:/b7280/d1988w). The challenge deadline was February 2, 2016 and experimental results were provided to participants not long after. As in past SAMPL challenges, each group could submit multiple sets of predictions. There was also the option for participants' identities to remain anonymous, although their results and method descriptions would still be made available. A total of 76 prediction sets from 18 participants or participating groups were submitted and assigned a random 2 digit ID number, 01 to 76, that will be used throughout this paper. Predictions were analyzed and overview statistics, as well as individual analysis of each submission by various error metrics (as detailed below) were returned to each participant. The challenge culminated with discussions of participants experiences and results at the 1st D3R Workshop at the University of California, San Diego March 9-11, 2016.

For the prediction of distribution coefficients in SAMPL5, a total of 53 molecules were considered. They were assigned an identifier in the form SAMPL5_XXX and are pictured in table 2. The 53 molecules were divided into batches 0, 1, and 2 containing 13, 20, and 20 molecules respectively. We wanted each batch to have a similar dynamic range and for the molecules to increase in size across

Batch 0

| 003: 1.7 ± 0.2 | 015: 8.8 ± 1.4 | 017: 3.0 ± 0.3 | 020: 3.9 ± 0.4 | 037: 7.5 ± 1.3 | 045: 1.4 ± 0.2 | 055: 2.7 ± 0.2 |

| 058: 2.2 ± 0.3 | 059: 1.8 ± 0.2 | 061: 5.9 ± 1.0 | 068: 2.7 ± 0.3 | 070: 5.9 ± 0.8 | 080: 2.6 ± 0.2 |

Batch 1

| 004: 2.3 ± 0.3 | 005: 2.6 ± 0.3 | 007: 3.1 ± 0.4 | 010: 7.8 ± 1.6 | 011: 7.1 ± 1.3 | 021: 2.2 ± 0.2 | 026: 8.0 ± 1.7 |

| 027: 3.4 ± 0.3 | 042: 3.2 ± 0.4 | 044: 3.7 ± 0.4 | 046: 2.7 ± 0.4 | 047: 2.1 ± 0.3 | 048: 2.7 ± 0.3 | 056: 3.5 ± 0.3 |

| 060: 6.7 ± 1.6 | 063: 6.7 ± 1.0 | 071: 2.8 ± 0.3 | 072: 4.9 ± 0.7 | 081: 6.0 ± 0.8 | 090: 2.8 ± 0.3 |

Batch 2

| 002: 2.5 ± 0.2 | 006: 2.7 ± 0.4 | 013: 3.1 ± 0.3 | 019: 3.1 ± 0.4 | 024: 3.0 ± 0.4 | 033: 3.0 ± 0.3 | 049: 2.1 ± 0.2 |

| 050: 5.6 ± 0.4 | 065: 5.3 ± 0.5 | 067: 4.5 ± 0.6 | 069: 3.9 ± 0.5 | 074: 6.6 ± 0.4 | 075: 4.8 ± 0.6 | 082: 5.1 ± 0.6 |

| 083: 8.4 ± 0.7 | 084: 3.6 ± 0.5 | 085: 2.8 ± 0.3 | 086: 4.5 ± 0.6 | 088: 2.9 ± 0.4 | 092: 3.9 ± 0.4 |

Table 1 A complete list of compounds used in the SAMPL5 challenge, sorted by batch. The average unsigned error (AUE), reported in log units as SAMPL5 ID: AUE, was calculated with all submitted predictions for that compound. The 2D images were generated using OpenEye OEDepict toolkit [22] using the provided SMILES strings.

batches, so on average the smallest molecules are in batch 0 and the largest in batch 2. To ensure each batch had adequate dynamic range, the molecular weight and estimated octanol/water partition coefficient were computed for each compound. These partition coefficients were estimated with OpenEye's $\log P$ calculator. Molecules were then assigned to bins by estimated partition coefficient, and assigned to batches based on molecular weight. Specifically, the smallest molecules from each partition coefficient bin were added to batch 0, then batch 1, and the rest of the molecules comprise batch 2.

Participants could submit partial sets of predictions as long as they included full consecutive batches; that is, they could submit batch 0, batches 0 and 1, or batches 0, 1, and 2. The idea was that all participants should attempt predictions on the full set if at all possible, but grouping into batches would allow people with particularly demanding methods (such as polarizable force fields or methods requiring intensive quantum mechanics) to focus on smaller compounds and still be evaluated. Eight submissions from two participants included results for only batch 0, and an additional five submissions from two participants provided only batches 0 and 1. Here we focus on the results for the complete set of molecules (batches 0, 1, and 2). Separate analysis for data subsets is available in the supporting information on Dash

Participants were asked to report a cyclohexane/water distribution coefficient for each molecule. As discussed above, distribution coefficients are the ratio of concentrations for all forms of the solute in the cyclohexane and aqueous layers, at a specified pH. In this case, experiments were done with the water layer consisting of a buffered aqueous solution at pH 7.4. We also required participants to provide two estimates for uncertainty, a statistical uncertainty for their computational method and a model uncertainty that estimates agreement with experiment. The statistical uncertainty was intended to be the variation expected over repeated calculations of the same value. The model uncertainty, on the other hand, was intended to provide an estimate of how well the calculated value will agree with experiment. For example, in a recent study we computed cyclohexane/water partition coefficients using alchemical solvation free energy calculations in GROMACS where the statistical uncertainties were around 0.05 log units, but the root mean squared error was around 1.4 log units [23], so an appropriate estimated model uncertainty would have been 1.4 log units. A careful analysis of expected error could even yield model uncertainties which would vary based on the anticipated difficulty or complexity of a compound. Our interest in model uncertainties is in part based on the realization that an important part of creating predictive models is the ability to know when they will be unreliable or fail. Thus, analysis of model uncertainties is an important part of evaluating any model.

3 Analysis of Submission Performance

As in past SAMPL challenges, we considered a variety of error metrics in analyzing all predictions submitted to SAMPL5. Each error metric was calculated

for all submissions, by batch, and distributed to challenge participants before the workshop. Here we will focus primarily on six error metrics: the root-mean-squared error (RMSE), average unsigned error (AUE), average signed error (ASE), Pearson's R (R), Kendall's tau (tau), and the 'error slope' explained in depth below. We also calculated the maximum absolute error and the percent of predictions with the correct sign, but these are not included in the analysis here. However, these metrics were provided to challenge participants and are available in the supporting information on Dash. The uncertainty in each metric was calculated as the standard deviation over 1000 bootstrap trials, where each trial consists of creating a 'new' dataset by sampling pairs of (predicted, calculated) values from the original set, with replacement. As described previously, this bootstrapping technique also included variation in the experimental values based on their reported uncertainties [1].

As discussed above, an important factor influencing the utility of a predictive tool is the ability of the tool to not only provide predictions but to predict the accuracy of those predictions – that is, how well the calculated values are likely to agree with experiment – not just its statistical error. To assess this, as in SAMPL4 [1], a quantile-quantile plot (QQ Plot) was created for each prediction set [24]. QQ Plots compare the fraction of a normal distribution within a specified number of standard deviations to the distribution of errors (calculated minus experiment) that are within that number of model uncertainties. For example, consider the number of predictions within one standard deviation of the expected value; if the samples are drawn from a normal distribution, then 0.68 of the values ought to fall within one standard deviation, so the value on the x-axis is 0.68. The value on the y-axis will represent the fraction of predictions that are within one model uncertainty of the experimental value. If the model uncertainty is accurate, then this also ought to correspond to a value of 0.68. A linear regression analysis helps summarize these results. The 'error slope' is the slope of the line comparing the fraction of predictions within a specified range of experiment to the expected fraction from a normal distribution. An error slope of greater than one indicates that the calculated values are within uncertainty of experiment more often than expected, or in other words the model uncertainty was overestimated. In contrast, an error slope less than one suggests the model uncertainty was underestimated.

We also attempted to identify any individual molecules where most of the methods failed to accurately estimate the distribution coefficient. To accomplish this, we analyzed all predictions on a molecule-by-molecule basis via our usual set of error metrics. Here we will primarily focus on just average unsigned error for molecules, but all other error metrics were provided to participants and are available in supporting information on Dash.

4 Reference calculations

We calculated distribution coefficients through a few different methods as a reference. K.H.B., a graduate student in the Mobley group, performed a set

of blind calculations estimating the $\log D$ as a partition coefficient between cyclohexane and water calculated from solvation free energies. In addition, C.C.B. and D.L.M. performed post-challenge analysis of protonation and tautomeric states and used this to convert our calculated partition coefficients to distribution coefficients. We considered a null hypothesis where all molecules are assumed to distribute equally between cyclohexane and water. Many fast structure-based tools for octanol/water partition coefficients exist, and we used one of those to estimate partition coefficients, both with no correction for the fact that we are interested in cyclohexane, and with a small adjustment for this as discussed below.

4.1 Calculating partition coefficients from solvation free energies

We decided to estimate distribution coefficients via a $\log P$ calculation, by assuming only a single neutral tautomer of each solute, then calculating $\log P$ from a difference in solvation free energies. Before the challenge, each molecule was taken directly from the provided SMILES string. As demonstrated in the literature, [11–21] partition coefficients are directly proportional to the difference between the solvation free energy for the solute into each solvent. We use previously established and automated protocols [23] to calculate the solvation free energy of each molecule into water and cyclohexane. Then the calculated partition coefficient was reported as an estimate for $\log D$.

To calculate solvation free energies, we used automated tools created by the Mobley lab. Molecular dynamics simulations were performed in GROMACS [25–31] with the General AMBER Force Field (GAFF) [32] with AM1-BCC charges [33,34]. The TIP3P water model [35] was used for the aqueous phase. Topology and coordinate files for the solvated boxes with 1 solute molecule and 500 cyclohexane or 1000 water molecules were built using our Solvation Toolkit [23]. These files were then converted to AMBER, DESMOND, and LAMMPS formats and provided to SAMPL5 participants as reference calculations. The Solvation Toolkit takes advantage of many open source Python modules and is available at https://github.com/MobleyLab/SolvationToolkit. It converts SMILES strings or IUPAC names of any mixture of small organic compounds to parameterized molecules and builds topology and coordinate files for a variety of simulation packages. All molecular dynamics parameters are identical to previous studies [36,4,23]. The molecule is taken from the solvated box to a non-interacting gas phase in 20 lambda values. Solvation free energies are calculated with Alchemical Analysis tool [37] using the multistate Bennett acceptance ratio to extract the free energy difference between the beginning and end state. The partition coefficient was calculated as the difference between the cyclohexane solvation free energy and the hydration free energy. The statistical uncertainty was reported as the propagated uncertainty from the solvation free energy calculations. The model uncertainty was estimated to be the same for all molecules and reported as the root-mean-squared error from a recent study on calculating cyclohexane/water partition
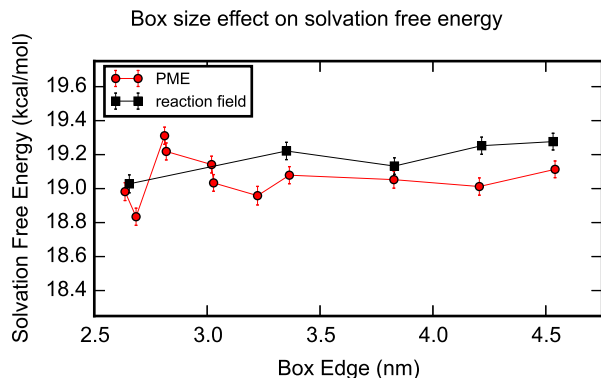
Fig. 1 Calculated solvation free energy for SAMPL5_024 is independent of box size for PME and reaction field coulomb interactions. Points are connected to help distinguish between the two data sets

coefficient, specifically 1.4 log units [23]. These reference calculations were assigned submission ID 39 and included in the error analysis performed on all submissions.

Simulation box size does not affect the calculated solvation free energy. Hydration free energies were previously shown to be independent of box size for box edges ranging from 2 to 9 nanometers, within calculated uncertainties [38]; however, here, because cyclohexane is much less polar, we had some concern that finite size effects could still be significant. To explore this, we performed some tests in which we varied the simulation box size. Because more polar solutes are more likely to have substantial long range interactions, we calculated the dipole moment of each SAMPL5 molecule using the position and charges on atoms in the mol2 files. SAMPL5_024 had the largest dipole moment so it was used as the solute for the box size investigation. The solvation free energy calculations were set up as described above, changing the number of cyclohexane molecules from 100 to 500. Our calculations above are performed with lattice-sum (PME) treatment of coulomb interactions. It was the primary focus of this check and we included duplicate calculations for the smaller box sizes where the initial coordinate file was the same, but new velocities were generated for the equilibrium phases. We also repeated the solvation free energy calculations with reaction field coulomb interactions assigning the dielectric coefficient for cyclohexane, 2.0243 [39]. For both types of simulation, the calculated solvation free energy fluctuated around an average of 19.1 kcal/mol with no trend that would suggest solvation free energy depends on box size (figure 1). There are many other explanations for fluctuations in calculated solvation free energies, including sampling, that might account for the 0.48 kcal/mol range in the PME simulations. Ultimately, we find that box edge lengths from 2.64 to 4.54 nm have no significant affect on the calculated solvation free energies. This suggests that in the future, smaller box sizes could be

used for computational efficiency. The input, results, molecular dynamics parameter and coordinate files, and tables of solvation free energies are available in the supporting information on Dash.

4.2 Consideration of tautomers after SAMPL5

To help understand how the results from our partition coefficient calculations could have been improved, we considered corrections for changes in the solutes' protonation or tautomeric states. Distribution coefficients differ from partition coefficients in that they include all forms of the solute in both solvents. A common way to convert between experimentally measured distribution coefficients and partition coefficients is with pKa values for the solute [40]. This is a simple correction using the Henderson-Hasselbalch equation:

$$pH = pK_a + \log \frac{[X]}{[HX]} \tag{2}$$

to relate the concentration of neutral species to the charged species at a given pH. This correction assumes the solute only has one other relevant protonation state and changes for acidic and basic molecules. Zwitterions and other neutral tautomers are not taken into account. The equation used to calculate a distribution coefficient ($\log D$) from a partition coefficient ($\log P$) for a basic solute (or X in equation 2) is below

$$\log D = \log P - log(1 + 10^{pK_a - pH}) \tag{3}$$

Alternatively for an acidic solute (or HX in equation 2) we would instead use:

$$\log D = \log P - log(1 + 10^{pH - pK_a}) \tag{4}$$

We use Schrödinger's Epik tool [41–43] to estimate pKa values for each molecule according to experimental conditions. We then estimated $\log D$ using the equations above, accounting for just one change in protonation state, meaning each solute was taken to be either acidic or basic. For acidic solutes, the smallest acidic pKa was used with equation 4, oppositely for basic solutes the largest basic pKa was used with equation 3 to estimate $\log D$ from $\log P$.

Using pKa values only accounts for one change in protonation, whereas a correct distribution coefficient should include all relevant tautomers and protonation states of the molecule in both solvents. To account for all other tautomer states, we used Schrödinger's LigPrep [44] to enumerate tautomers for each molecule in the aqueous solution. The results of the enumeration can include an energetic "state penalty" calculated with Epik which relates the population of that tautomer to all others. This state penalty can be converted into log units and used as a correction term to convert $\log P$ to $\log D$:

$$\log D = \log P + \frac{-E_{state\ penalty}}{k_B T \ln(10)} \tag{5}$$

where k_B is Boltzmann constant and T is temperature. LigPrep can only perform the tautomer enumeration with water or DMSO as a solvent, so we were unable to predict tautomers in cyclohexane. Therefore both of these corrections account for the protonation or tautomer states only in the aqueous layer and assume the tautomer remains fixed in cyclohexane as the one used in the initial simulation. In the results section below, the corrections performed with pK_a and the corrections made with the the calculated state penalty are referred to $\log D_{pK_a}$ and $\log D_{state\ penalty}$ respectively.

4.3 Estimating distribution coefficients with a fast, structural based partition coefficient calculator

Many structure-based tools exist for octanol/water partition coefficients; they are very fast and generally accurate. However, these tools are all trained on empirical data, meaning they are limited by the training data. We chose the Open-Eye tool OEXlogP [45, 46] as an example of such a tool. Two post-prediction sets were prepared with the OEXlogP tool. First, the predicted octanol/water partition coefficient was considered an estimate for cyclohexane/water distribution coefficient. In the second set, we used a linear regression to correct for the bias between the calculated XlogP values and a set of experimental cyclohexane/water partition coefficients [9]. For the rest of this paper we will refer to the octanol/water partition coefficient set as $XlogP_{oct}$ and the bias-corrected set as $XlogP_{corr}$.

4.4 Exploring the possibility of mixing solvents

Because no two solvents are perfectly immiscible, we wanted to explore the effect that a small amount of water present in cyclohexane would have on computed $\log D$ values for one of the more polar solutes. The experimental concentration of water in cyclohexane is 0.00047 mole fraction [47]. The presence of water in the cyclohexane phase has a possibility of affecting the transfer free energy, especially for solutes with many polar functional groups. Also, it has been suggested that particularly polar compounds can pull water with them from the aqueous layer into the organic solvent [9]; while this is a kinetic argument and should not apply to equilibrium thermodynamic properties like $\log D$, the point is well taken — some solutes have a particularly high affinity for water and may actually impact the amount of water present in cyclohexane when the solute is present at finite concentration. Thus, to investigate this, we took one of the most polar compounds – one for which we had particularly large errors relative to experiment – SAMPL5_074, and performed two additional sets of free energy calculations in cyclohexane. Both sets of simulations had the single solute in 150 cyclohexane molecules, but varied in water content – the first with seven water molecules present in cyclohexane, and the second with only a single water molecule. The input files for these

simulations were also created with Solvation Toolkit as described above and the same simulation protocol was followed. These simulations were conducted after the SAMPL submission deadline in order to help us understand the role of water in such cases.

The experiments were not done on completely pure water and cyclohexane – particularly, experimental distribution coefficients were measured with small amounts of dimethyl sulfoxide (DMSO) and acetonitrile present in solution. Therefore, we investigated the how acetonitrile and DMSO would distribute in simulations. The experimental concentrations reported for each solvent are approximately 50%, 50%, 1%, and 0.4% by volume for cyclohexane, water, DMSO, and acetonitrile respectively [8]. To explore how the DMSO and acetonitrile distribute between cyclohexane and water we performed a single simulation with 130 cyclohexane, 780 water, four DMSO, and two acetonitrile molecules mirroring those concentrations. Input topology and coordinate files were created with Solvation Toolkit. The system was minimized and equilibrated following our reference calculation procedure and then a 5 ns constant pressure and temperature production simulation was run. The trajectory from this simulation was visualized with VMD [48] and a movie is available in the supplementary information.

5 Results and Discussion

A broad range of methods were used for the 76 submissions predicting cyclohexane/water distribution coefficients for the SAMPL5 challenge. Many of these predictions used alchemical molecular dynamics simulations to estimate the solvation free energy in explicit solvent using several classes of force fields, including fixed-charge all-atom force fields [49–53], all-atom/coarse-grained hybrid force fields [54], and polarizable force fields [55]. One participant used Semi-Explicit Assembly, a type of implicit solvent solvation free energy method applied to one or more chosen solute conformations [56]. A variety of quantum mechanics (QM) methods were also used including QM/molecular mechanics (QM/MM) with explicit solvent [52,53], QM with non-Boltzmann Bennett free energy calculations [52,53], and QM energy calculations with a single optimized molecular geometry [52,57]. Two participants used variations on the reference interaction-site model (RISM), an integral equation approach, to predict solvation free energies [58,59]. One participant used QM calculations to derive parameters to tune an empirical model for activity coefficients and used these to estimate distribution coefficients [60]. A few submissions used empirically trained methods for calculating solvation free energies [61,62]. A particularly successful submission, which will be discussed again below, employed the Conductor-Like-Screening Model for Real Solvents (COSMO-RS) [63].

SAMPL5 is the first SAMPL challenge to include distribution coefficients, but we can estimate how well we expect submissions to do based on past SAMPL challenges which included hydration free energies. Distribution coeffi-

cients can be related to transfer free energy between solvents, which allows us to estimate an expected performance from root-mean-squared error (RMSE) in past hydration free energy calculations. In SAMPL4 [1], the average RMSE for the best half of submissions was about 1.5 kcal/mol which would correspond to a 1.54 log unit error in a distribution coefficient if both solvation free energies have comparable errors. Here, only five submissions had an RMSE less than 2.5 log units in SAMPL5. There are many reasons for this perceived change in accuracy, such as a more complex set of molecules, the use of cyclohexane as a solvent, and the complexity of estimating tautomer populations, discussed in depth below. Since this is the first challenge on predicting distribution coefficients, it is likely that participants had not yet developed good protocols to deal with many of these challenges, meaning that somewhat less accuracy ought to be expected. It took several challenges focused on hydration [1–7] before a range of methods could achieve the success noted in SAMPL4.

As discussed above, we calculated root-mean-squared error (RMSE), average unsigned error (AUE), average signed error (ASE), Pearson's R (R), Kendall's tau (tau), and the slope from the QQ plot (error slope) for each set of predictions. These are reported for all submissions (Table 5), but the rest of the analysis will focus only on submissions that reported results for batches 0, 1, and 2. For each submission, we also created a plot comparing the predicted and experimental values. Some example plots are provided (Fig. 2); these represent a typical submission, in that these submissions were in the middle of the pack by most error metrics. Comparison and QQ plots for every submission are available in the supporting information on Dash as well as error metric tables by batch rather than just for the full set.

To help visualize all of the error metrics, the data was compiled into a histogram where results are sorted from best to worst for that metric (closest to 1 for error slope for example). These metrics are split into measurements of deviation from experiment (Fig. 3) and correlation with experiment (Fig. 4) distinctions which helped in identifying high performing groups. This analysis included only submissions that included data for all molecules; the other submissions were indicated in Table 5 and generally fall in the middle of the pack on most metrics. In comparing methods by all of the error metrics, it is important to keep in mind the uncertainty in these error metrics. While figures 3 and 4 are ordered by method performance in some sense, the reality is that there are many submissions that are not significantly different from one another.

In the error slope analysis, the slopes are often substantially different from 1, indicating that participants generally provided poor estimates of model uncertainty. Only the top three submissions are within uncertainty of 1. Sebastian Diaz-Rodriguez *et. al* from Miami University used conservative estimates based on results in previous calculations for solubility and hydration free energy for submissions 53 and 60 [60]. Gerhard König *et. al* from Max-Planck-Institut für Kohlenforschung provided no explicit discussion of model uncertainty with submission 43 [53]. Only submission 40 significantly overestimated their model uncertainty. All other submissions have an error slope

ID	ASE	RMSE	AUE	tau	R	Err. slope
01[b]	2.3 ± 0.8	5.1 ± 0.5	4.3 ± 0.5	0.13 ± 0.13	0.20 ± 0.18	0.44 ± 0.09
02	-0.5 ± 0.3	2.3 ± 0.3	1.7 ± 0.2	0.48 ± 0.07	0.63 ± 0.07	0.69 ± 0.07
03[b]	-7.6 ± 3.4	21.3 ± 2.6	15.9 ± 2.4	0.52 ± 0.10	0.59 ± 0.12	-0.00 ± 0.00
04[a]	1.6 ± 0.5	2.5 ± 0.6	1.9 ± 0.4	0.77 ± 0.12	0.87 ± 0.05	0.77 ± 0.13
05	-8.2 ± 0.4	8.7 ± 0.4	8.2 ± 0.4	0.29 ± 0.08	0.39 ± 0.11	0.21 ± 0.04
06	1.8 ± 0.5	4.0 ± 0.3	3.4 ± 0.3	0.46 ± 0.09	0.61 ± 0.10	0.58 ± 0.07
07	0.5 ± 0.5	3.3 ± 0.5	2.5 ± 0.3	0.34 ± 0.08	0.51 ± 0.11	0.33 ± 0.07
08	-1.7 ± 0.4	3.5 ± 0.5	2.5 ± 0.4	0.58 ± 0.06	0.70 ± 0.06	0.60 ± 0.07
09	-5.5 ± 0.4	6.3 ± 0.5	5.5 ± 0.4	0.29 ± 0.08	0.40 ± 0.10	0.26 ± 0.05
10	0.3 ± 0.4	3.1 ± 0.3	2.6 ± 0.3	0.51 ± 0.07	0.69 ± 0.08	0.79 ± 0.07
11	-4.4 ± 1.7	13.3 ± 2.5	6.9 ± 1.6	0.45 ± 0.09	0.53 ± 0.09	0.39 ± 0.07
12	-5.5 ± 2.5	19.4 ± 1.9	15.0 ± 1.7	0.37 ± 0.09	0.39 ± 0.12	-0.00 ± 0.00
13[a]	-11.1 ± 5.0	21.0 ± 5.0	12.2 ± 4.8	0.56 ± 0.17	0.43 ± 0.22	0.59 ± 0.17
14	-0.7 ± 0.4	2.7 ± 0.4	2.0 ± 0.3	0.57 ± 0.06	0.72 ± 0.06	0.66 ± 0.08
15	-1.4 ± 0.4	3.3 ± 0.5	2.3 ± 0.3	0.57 ± 0.07	0.70 ± 0.06	0.61 ± 0.07
16	0.5 ± 0.3	2.1 ± 0.2	1.7 ± 0.2	0.73 ± 0.04	0.84 ± 0.03	0.46 ± 0.08
17	-4.2 ± 0.4	5.0 ± 0.5	4.2 ± 0.4	0.36 ± 0.08	0.51 ± 0.10	0.50 ± 0.06
18	-0.8 ± 0.4	2.7 ± 0.4	2.0 ± 0.3	0.47 ± 0.08	0.60 ± 0.07	0.62 ± 0.08
19	1.5 ± 0.3	2.7 ± 0.2	2.3 ± 0.2	0.54 ± 0.06	0.75 ± 0.06	0.83 ± 0.06
20	-2.3 ± 0.4	3.6 ± 0.5	2.7 ± 0.3	0.55 ± 0.07	0.70 ± 0.06	0.48 ± 0.07
21	-1.2 ± 0.4	3.4 ± 0.6	2.4 ± 0.3	0.44 ± 0.08	0.45 ± 0.16	0.58 ± 0.07
22	1.6 ± 0.5	3.9 ± 0.3	3.1 ± 0.3	0.29 ± 0.09	0.48 ± 0.11	0.68 ± 0.08
23	1.9 ± 0.5	4.0 ± 0.4	3.0 ± 0.4	0.42 ± 0.07	0.58 ± 0.07	0.78 ± 0.08
24[a]	2.3 ± 0.7	3.3 ± 0.8	2.5 ± 0.6	0.77 ± 0.13	0.88 ± 0.05	0.67 ± 0.15
25	0.0 ± 0.5	3.6 ± 0.4	2.9 ± 0.3	0.53 ± 0.07	0.70 ± 0.07	0.71 ± 0.07
26	2.3 ± 0.7	5.6 ± 0.4	4.6 ± 0.4	0.25 ± 0.08	0.37 ± 0.11	0.46 ± 0.07
27	-0.2 ± 0.4	2.6 ± 0.4	1.8 ± 0.2	0.49 ± 0.07	0.61 ± 0.08	0.66 ± 0.07
28	-2.3 ± 0.4	3.6 ± 0.5	2.7 ± 0.4	0.54 ± 0.07	0.69 ± 0.07	0.47 ± 0.07
29	-6.7 ± 0.4	7.2 ± 0.4	6.7 ± 0.4	0.33 ± 0.08	0.45 ± 0.11	0.28 ± 0.04
30	2.5 ± 0.5	4.3 ± 0.3	3.7 ± 0.3	0.39 ± 0.10	0.52 ± 0.12	0.53 ± 0.07
31	-1.0 ± 0.3	2.7 ± 0.3	2.0 ± 0.3	0.56 ± 0.07	0.72 ± 0.06	0.63 ± 0.08
32	2.5 ± 0.4	3.5 ± 0.2	3.1 ± 0.2	0.47 ± 0.07	0.64 ± 0.07	0.25 ± 0.06
33	-0.1 ± 0.5	3.4 ± 0.5	2.8 ± 0.3	0.53 ± 0.07	0.71 ± 0.07	0.73 ± 0.07
34	-1.3 ± 0.4	3.0 ± 0.4	2.2 ± 0.3	0.56 ± 0.06	0.69 ± 0.07	0.61 ± 0.07
35	0.5 ± 0.4	2.9 ± 0.3	2.2 ± 0.2	0.36 ± 0.08	0.54 ± 0.09	0.35 ± 0.07
36	1.1 ± 0.3	2.6 ± 0.2	2.1 ± 0.2	0.57 ± 0.07	0.75 ± 0.06	0.50 ± 0.07
37[a]	-7.1 ± 4.9	19.6 ± 4.1	13.9 ± 3.7	0.59 ± 0.16	0.41 ± 0.22	-0.00 ± 0.00
38	0.8 ± 0.4	3.3 ± 0.3	2.7 ± 0.3	0.41 ± 0.08	0.58 ± 0.08	0.78 ± 0.07
39	1.6 ± 0.3	2.6 ± 0.2	2.1 ± 0.2	0.49 ± 0.08	0.65 ± 0.10	0.63 ± 0.08
40	0.4 ± 0.3	2.6 ± 0.3	1.9 ± 0.2	0.48 ± 0.07	0.61 ± 0.08	1.16 ± 0.05
41	0.3 ± 0.4	3.2 ± 0.3	2.7 ± 0.3	0.53 ± 0.07	0.69 ± 0.07	0.77 ± 0.07
42	4.6 ± 0.4	5.3 ± 0.4	4.6 ± 0.3	0.50 ± 0.08	0.61 ± 0.11	0.15 ± 0.05
43	-0.7 ± 0.4	3.0 ± 0.4	2.3 ± 0.3	0.51 ± 0.08	0.67 ± 0.08	0.94 ± 0.07
44	-0.6 ± 0.3	2.4 ± 0.3	1.8 ± 0.2	0.47 ± 0.07	0.63 ± 0.07	0.70 ± 0.07
45	0.9 ± 0.5	3.6 ± 0.3	2.9 ± 0.3	0.38 ± 0.08	0.58 ± 0.10	0.71 ± 0.07
46	-8.3 ± 0.5	9.1 ± 0.6	8.3 ± 0.5	0.23 ± 0.08	0.31 ± 0.10	0.14 ± 0.03
47	-1.3 ± 0.4	3.3 ± 0.5	2.2 ± 0.3	0.58 ± 0.06	0.71 ± 0.06	0.62 ± 0.08
48	1.5 ± 0.4	3.0 ± 0.3	2.3 ± 0.3	0.38 ± 0.07	0.55 ± 0.08	0.42 ± 0.07
49	-1.1 ± 0.4	3.3 ± 0.4	2.6 ± 0.3	0.42 ± 0.07	0.58 ± 0.07	0.78 ± 0.07
50[b]	-7.1 ± 2.7	16.6 ± 3.2	9.2 ± 2.4	0.60 ± 0.09	0.66 ± 0.08	0.38 ± 0.10
51	1.7 ± 0.7	5.2 ± 0.4	4.3 ± 0.4	0.31 ± 0.08	0.46 ± 0.11	0.46 ± 0.08
52[a]	-3.5 ± 1.1	5.4 ± 0.6	4.8 ± 0.7	0.56 ± 0.14	0.59 ± 0.14	0.23 ± 0.10
53	0.5 ± 0.4	2.8 ± 0.3	2.2 ± 0.2	0.44 ± 0.09	0.58 ± 0.10	1.00 ± 0.06
54	-1.0 ± 0.3	2.7 ± 0.3	1.9 ± 0.2	0.56 ± 0.07	0.70 ± 0.06	0.65 ± 0.08
55[b]	-11.6 ± 3.3	22.3 ± 3.0	13.7 ± 3.1	0.59 ± 0.09	0.61 ± 0.11	0.38 ± 0.09
56	-1.1 ± 0.4	3.3 ± 0.5	2.2 ± 0.3	0.57 ± 0.06	0.71 ± 0.06	0.67 ± 0.08
57	-10.2 ± 2.4	20.2 ± 2.3	12.6 ± 2.2	0.43 ± 0.09	0.42 ± 0.12	0.38 ± 0.07
58	-2.9 ± 0.5	4.8 ± 0.5	3.8 ± 0.4	0.30 ± 0.09	0.44 ± 0.11	0.55 ± 0.08
59[a]	-4.2 ± 1.0	5.6 ± 0.6	5.2 ± 0.6	0.54 ± 0.15	0.55 ± 0.14	0.13 ± 0.07
60	0.2 ± 0.3	2.5 ± 0.4	1.9 ± 0.2	0.49 ± 0.08	0.60 ± 0.08	1.02 ± 0.06
61	-1.2 ± 0.5	3.4 ± 0.6	2.4 ± 0.3	0.44 ± 0.08	0.45 ± 0.16	0.53 ± 0.07
62	0.7 ± 0.5	3.5 ± 0.4	2.7 ± 0.3	0.27 ± 0.09	0.38 ± 0.12	0.73 ± 0.08
63	-4.5 ± 1.7	13.3 ± 2.5	6.9 ± 1.6	0.45 ± 0.09	0.52 ± 0.08	0.41 ± 0.07
64	1.3 ± 0.7	5.2 ± 0.4	4.4 ± 0.4	0.35 ± 0.08	0.51 ± 0.10	0.43 ± 0.07
65	-2.2 ± 0.5	4.4 ± 0.5	3.5 ± 0.4	0.24 ± 0.10	0.35 ± 0.12	0.61 ± 0.08
66	1.4 ± 0.7	5.4 ± 0.4	4.6 ± 0.4	0.34 ± 0.08	0.51 ± 0.10	0.41 ± 0.07
67[a]	-5.0 ± 3.1	11.9 ± 4.5	6.2 ± 2.9	0.59 ± 0.17	0.58 ± 0.13	0.56 ± 0.17
68	2.5 ± 0.4	3.6 ± 0.3	3.1 ± 0.2	0.47 ± 0.07	0.64 ± 0.07	0.25 ± 0.06
69[a]	-5.1 ± 2.9	11.9 ± 4.4	6.2 ± 2.8	0.59 ± 0.16	0.57 ± 0.12	0.59 ± 0.16
70[b]	-7.0 ± 2.6	16.5 ± 3.2	9.2 ± 2.4	0.60 ± 0.09	0.67 ± 0.08	0.36 ± 0.10
71	-10.7 ± 0.4	11.2 ± 0.5	10.7 ± 0.4	0.22 ± 0.08	0.29 ± 0.11	0.16 ± 0.03
72	-2.6 ± 0.5	4.2 ± 0.6	3.0 ± 0.4	0.56 ± 0.06	0.70 ± 0.06	0.45 ± 0.07
73	0.3 ± 0.3	2.4 ± 0.3	1.8 ± 0.2	0.48 ± 0.08	0.64 ± 0.08	0.50 ± 0.08
74	-2.7 ± 0.4	4.2 ± 0.5	3.0 ± 0.4	0.56 ± 0.07	0.70 ± 0.06	0.44 ± 0.07
75	4.1 ± 0.4	5.1 ± 0.3	4.4 ± 0.3	0.23 ± 0.09	0.34 ± 0.12	0.29 ± 0.06
76	1.7 ± 0.7	5.3 ± 0.4	4.3 ± 0.4	0.32 ± 0.08	0.47 ± 0.10	0.47 ± 0.08

Table 2 Error metrics were calculate for each set of predictions, including root-mean-squared error (RMSE), average unsigned error (AUE), average signed error (ASE), Kendall's tau (tau), and Pearson's R (R). Error slope refers to the slope of data in a QQ Plot. Indicated submissions included only batch 0[a] or batches 0 and 1[b].
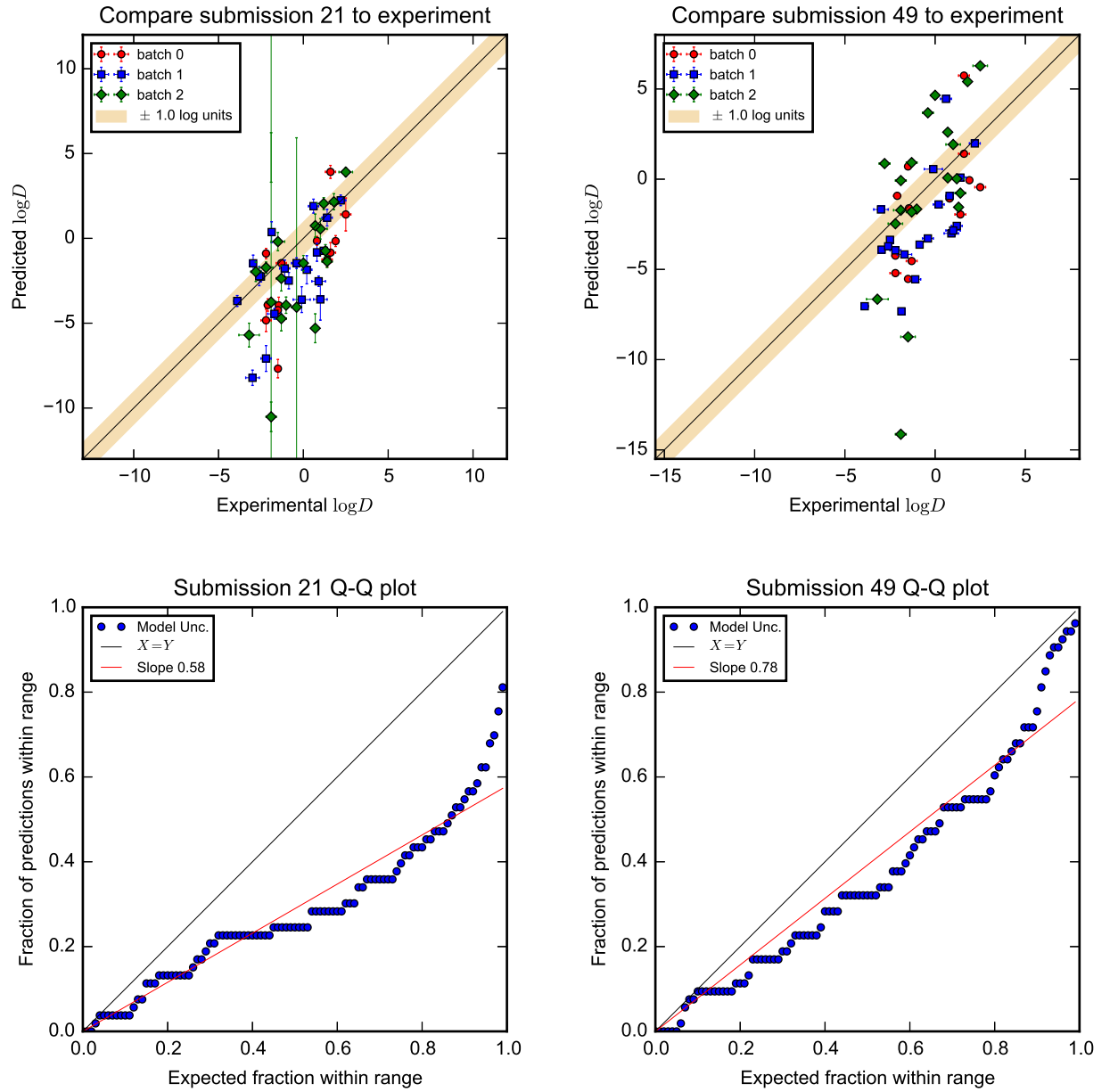
Fig. 2 Example plots created for each set of predictions. Submission 21 [53] and submission 49 [62] were chosen to try to represent average submissions, those that were in the middle by most error metrics. Comparison plots show how predicted distribution coefficients compared to experiment for both submissions. QQ Plots show how errors in the predictions were distributed compared to expectations given the model uncertainty.
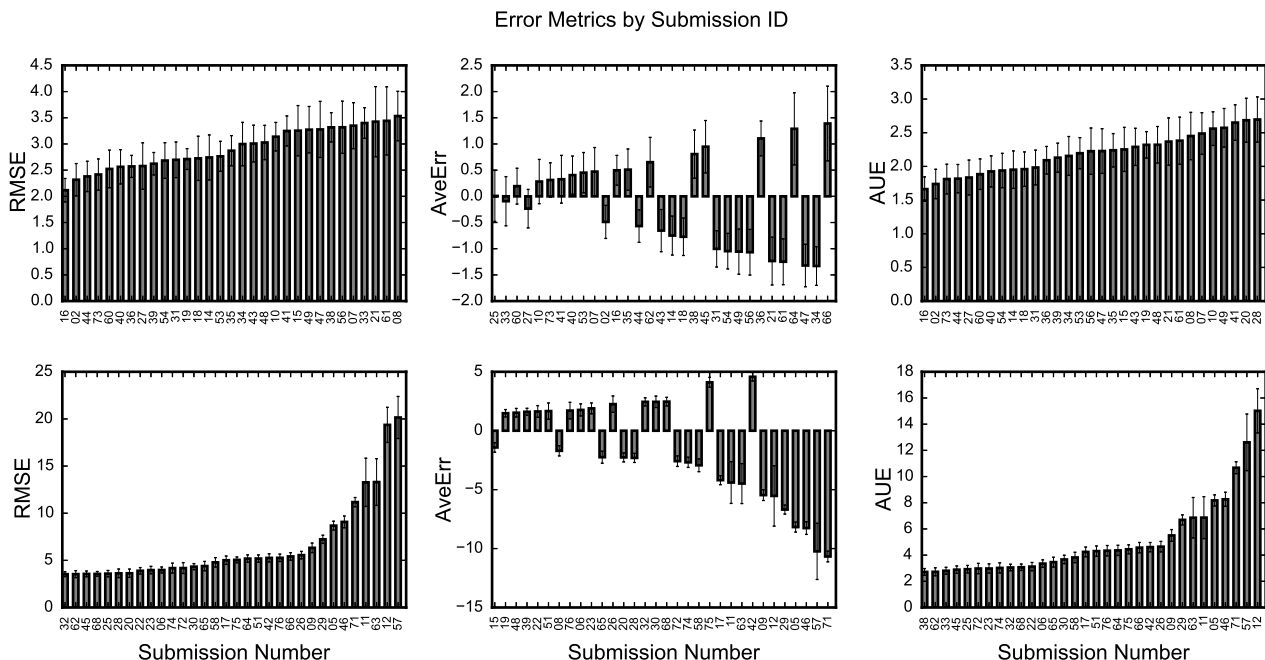
Fig. 3 Root-mean-squared error (RMSE), average error (AveErr), and average unsigned error (AUE) for every SAMPL5 submission covering all batches. The submissions on each plot are sorted from best to worst by that metric. Due to the number of submissions, data was split across the two panels, with a change in the y-axis scale.

below one, indicating a significant underestimation of the model uncertainty. This suggests that analysis and prediction of model uncertainty remains a key frontier for predictive molecular simulations, and further effort is needed in that area.

5.1 Top performing submissions

In order to determine which submissions performed the best, we group error metrics into two categories. The first category describes typical error relative to experiment, and includes metrics RMSE and AUE. The second category describes how well correlated the experimental values are with the experimental values, and includes Kendall τ and Pearson R. Unlike past SAMPL challenges, there does appear to be one submission which performs best by all of these metrics, submission 16, and for most metrics it is better by a statistically significant amount. Next, we considered the top ten submissions for each of these four metrics. There were only two submissions, other than 16, which performed in the top ten for at least three of these metrics, 14 and 36. Predictions from each of these submissions are compared to experiment in figure 5. For submission 16, Andreas Klamt *et. al.* from COSMOlogic

Error Metrics by Submission ID
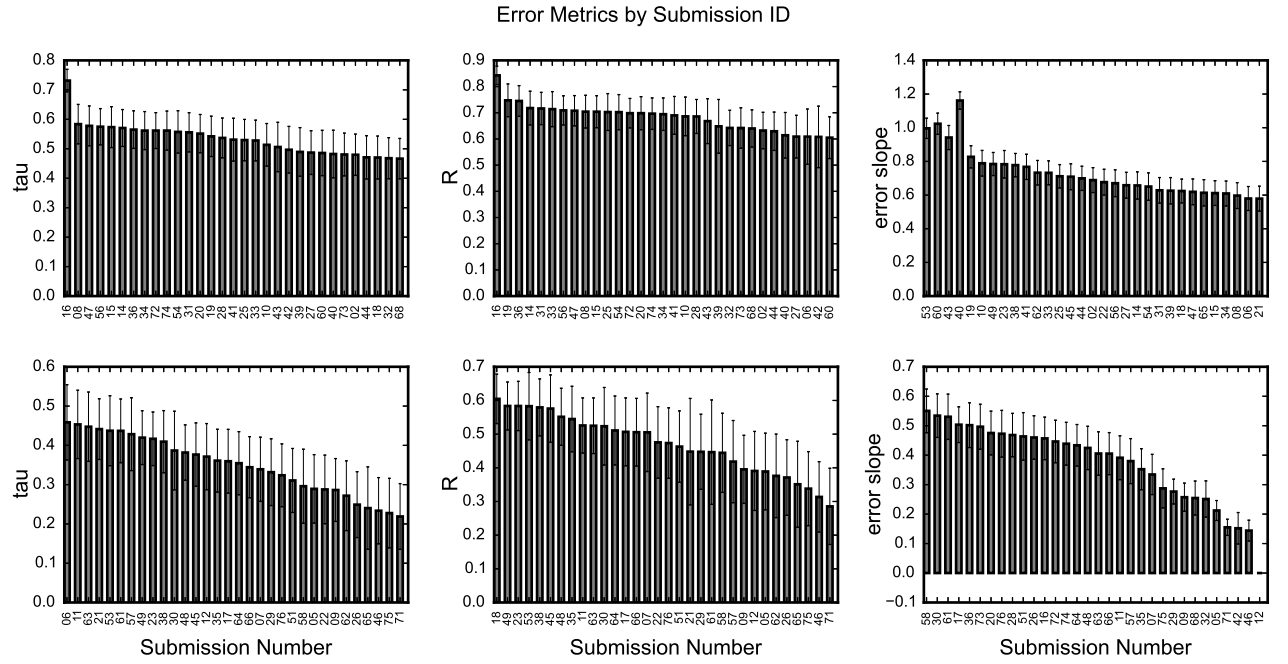
Fig. 4 Kendall's tau, Pearson's R, and the slope from a linear regression analysis on the QQ Plot ('error slope') for every SAMPL5 submission covering all batches. The submissions on each plot are sorted from best to worst by that metric. Due to the number of submissions, data was split across the two panels, with a change in the y-axis scale.
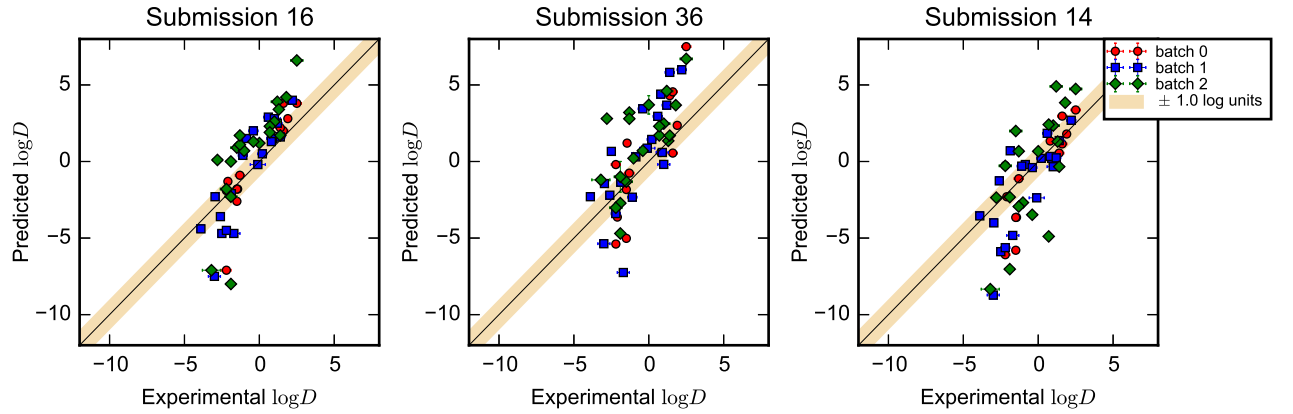
Fig. 5 These plots compare predicted and experimental distribution coefficients for the top performing submissions (16, 36, and 14).

Metric	Null	$XlogP_{oct}$	$XlogP_{corr}$
AveErr	0.5 ± 0.2	2.8 ± 0.2	1.1 ± 0.2
RMSE	1.8 ± 0.1	3.1 ± 0.1	1.6 ± 0.1
AUE	1.6 ± 0.1	2.8 ± 0.2	1.3 ± 0.1
tau	N/A	0.62 ± 0.04	0.62 ± 0.04
R	N/A	0.78 ± 0.04	0.78 ± 0.04

Table 3 Null hypothesis corresponds to $\log D = 0$ for all molecules. $XlogP_{oct}$ is a calculation octanol/water partition coefficient for each molecule and $XlogP_{corr}$ includes a linear regression correction.

used COSMO-RS to compute a partition coefficient for each solute from the difference in chemical potentials for the solute in each solvent [63]. To find distribution coefficients, calculations for the formation of different protonation states, zwitterions, and tautomers were performed in COSMO-RS for relevant molecules. For submission 14, Frank Pickard *et al.* from the National Institute of Health calculated solvation free energies from QM calculations with SMD implicit solvent in Gaussian. Absolute pK_a calculations were used to account for additional ionization states [52]. For submission 36, Sherin Shanaka Paranhewage *et al.* from Oklahoma State University estimated $\log D$ as a partition coefficient, calculated from the difference in alchemical solvation free energies where the solute was parameterized with the dielectrically corrected general AMBER force field, water was the dielectrically corrected H2O-DC model, and cyclohexane was a specially optimized united-atom model [49]. Further details for each of these submissions can be found in this issue so only a brief explanation of each method was provided here.

5.2 Comparisons to simple empirical models

One way of evaluating predictive models is to compare them to a null hypothesis, or default result of some kind. In the case of distribution coefficients, we chose a null hypothesis where we assume all solute molecules distribute equally between cyclohexane and water, corresponding to $\log D = 0$, as suggested by Christopher Fennell [64]. We performed all our standard error analyses discussed above (RMSE, AUE, and Ave. Err) on this simple model as a point of comparison (Table 5.2). The null hypothesis would have been the top submission for both RMSE and AUE. While this null hypothesis has no actual predictive power and could not be used to rank compounds, the fact that it performs better than any submission in terms of error statistics is a challenge for the other methods. These results may also provide commentary on the dataset, which contains a reasonably large percentage of $\log D$ values that are not that far off from zero (figure 6). Organizers had hoped to ensure equal coverage of all $\log D$ values within the assay range, but due to experimental time constraints this was not possible. It is possible the null model would look worse if the experimental results were more evenly dispersed across the entire dynamic range.
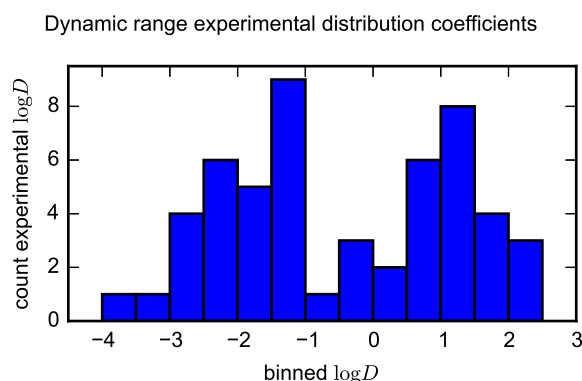
Dynamic range experimental distribution coefficients

Fig. 6 The experimental distribution coefficients of the SAMPL5 challenge have a relatively small dynamic range, with most falling within 2 log units of zero.

There are many structure-based and/or empirically trained prediction methods for octanol/water partition coefficients. To a first approximation, one might imagine that cyclohexane/water partition coefficients would follow similar trends to those in octanol/water. Therefore, we used OEXlogP from Openeye ($XlogP_{oct}$) to examine the possibility of estimating cyclohexane/water distribution coefficients with such a tool. Next, we compared $XlogP_{oct}$ results for a set of compounds with experimental cyclohexane/water partition coefficients [9]. A linear regression was used to correct the $XlogP_{oct}$ values with a slope of 0.7241 and a y-intercept of -1.0306 ($XlogP_{corr}$). $XlogP_{oct}$ would be in the top few submissions by tau and R, but ranked in the middle for all other metrics (Table 5.2). However, with a simple linear regression trained on experimental cyclohexane/water partition coefficients, $XlogP_{corr}$ has a better RMSE and AUE than any SAMPL5 submission. We do not wish to suggest that regression-trained tools are the best mechanism for predicting distribution coefficients; rather, this indicates the potential for cyclohexane/water distribution data to help drive improvements in our physical models, as clearly there are a range of physical effects here which are not yet well described by our models.

5.3 Results of reference calculations

We performed a set of blind reference calculations (submission 39) for the SAMPL5 challenge, calculating $\log P$ for the provided neutral tautomers of all solutes. Our protocol for these calculations was announced in advance, and parameter and coordinate files for the calculations were made available (as described above) in formats for a variety of simulation packages. Participants were encouraged to perform their own set of reference calculations for the full set, or at the very least several specified reference compounds, using these files. This would allow differences in performance to be traced back to methodological differences rather than force field differences. Unfortunately,
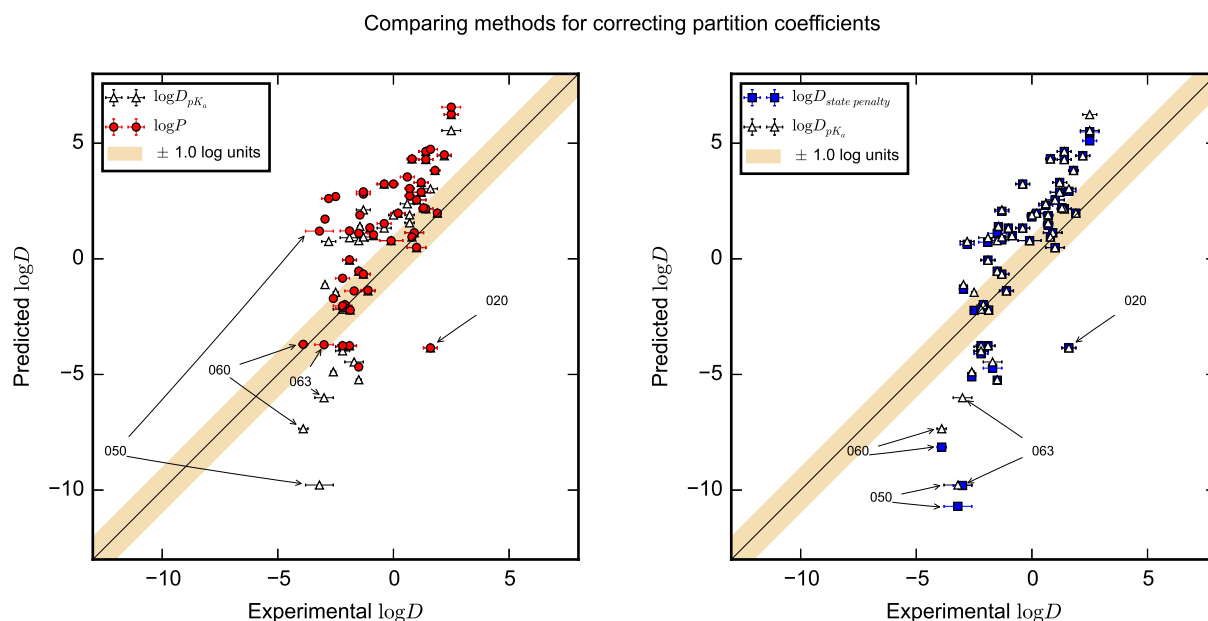
Comparing methods for correcting partition coefficients

Fig. 7 Plots showing our reference calculations compared to experiment. Shown here is the results for submission 39 in SAMPL5, with no tautomer correction ($\log P$), distribution coefficient corrected from calculated partition coefficient based on pKas ($\log D_{pK_a}$), and distribution coefficient corrected from calculated partition coefficient based on state penalties ($\log D_{state\ penalty}$).

no participants actually reported results of reference calculations, so this type of analysis has thus far been impossible. However, the results of our reference calculations are still helpful for understanding the challenges facing SAMPL5 participants.

For our reference calculations, solvation free energies were calculated using GROMACS with GAFF parameters and AM1-BCC charges. Our reference calculations yielded partition coefficients, determined from the difference in solvation free energies without correcting for variation in tautomers. These calculations were done blindly, and analyzed as submission number 39, which was in the top quarter of submissions by most error metrics (Table 5) although, there is a slight bias favoring concentrations in cyclohexane, evidenced by the average error (1.6 ± 0.3).

After the challenge we explored how including protonation and deprotonation would have affected the initial partition coefficient predictions. The first set of corrections involved calculating the pKa for each molecule using Schrödinger's Epik tool [41–43]. Next, $\log D$ was calculated using the pKa and partition coefficient determined in submission 39 using equations 3 and 4 for basic and acidic solutes, respectively. We assumed only one change in protonation state occurred so only one pK_a was used. This does not account for

Metric	$\log P$	$\log D_{pK_a}$	$\log D_{state\ penalty}$
AveErr	1.6 ± 0.3	0.7 ± 0.3	0.5 ± 0.4
RMSE	2.6 ± 0.2	2.4 ± 0.2	2.6 ± 0.3
AUE	2.1 ± 0.2	2.0 ± 0.2	2.1 ± 0.2
tau	0.49 ± 0.08	0.65 ± 0.07	0.65 ± 0.06
R	0.6 ± 0.1	0.78 ± 0.07	0.77 ± 0.06

Table 4 Shown are the results for error analysis on our reference calculations which estimated $\log D$ as a cyclohexane partition coefficient ($\log P$) and the correction to distributions coefficients by pK_a ($\log D_{pK_a}$) and by state penalty ($\log D_{state\ penalty}$). Included here are average error (AveErr), root-mean-squared error (RMSE), average unsigned error (AUE), Kendall's tau, and Pearson's R.

zwitterions or alternate neutral tautomers. This correction (labeled $\log D_{pK_a}$) showed a slight improvement by most error metrics (Table 5.3) including a decrease in the average error from 1.6 ± 0.3 to 0.7 ± 0.3 indicating less bias toward overly high concentration in cyclohexane.

For the next set of corrections, we used Schrödinger's Ligprep tool [44] to enumerate tautomers and calculate a state penalty or relative energy of each tautomer in an aqueous buffer at pH 7.4. The state penalty was used to correct the concentration in the aqueous layer, according to equation 5. This correction (labeled $\log D_{state\ penalty}$) results show improvements from the original partition coefficient coefficients for tau (0.49 ± 0.08 to 0.65 ± 0.06) and R (0.6 ± 0.1 to 0.77 ± 0.06), but no significant change in RMSE or AUE. It should be noted that no attempt was made to estimate uncertainties in the newly calculated $\log D_{pK_a}$ or $\log D_{state\ penalty}$ although both of these corrections would certainty have introduced uncertainty into the estimate. Future exploration in tautomer estimation will need to account for the uncertainty in those calculations.

Both of these correction methods only adjust the concentration in the aqueous layer; however, there may be tautomer affects that would change the concentration in cyclohexane as well. Outliers and molecules with particularly significant changes in $\log D$ were indicated by number in figure 7. SAMPL_050 for example, had an initial $\log P$ value of 1.20 ± 0.04 which was decreased significantly to -9.78 with the pKa correction and -10.70 with the state penalty correction compared to the experimental value -3.2 ± 0.6. SAMPL_060 and SAMPL_063 also changed by more than 3 log units due to these corrections. These state penalties allow us to account for other tautomers and protonation states *only* in the aqueous phase. Without tautomer enumeration in cyclohexane, we have to assume that the tautomer used for solvation free energy calculations, prior to correction, is the dominant state of the solute in the cyclohexane phase. If an alternate tautomer were relevant in water *and* cyclohexane, we could obtain dramatically incorrect values with this approach, since our state penalty will only recover alternate tautomer(s) in water, but not cyclohexane. This appears to be one of the reasons why these corrections seem to overshoot in the cases listed above. A better solution would compute

state penalties in both water and cyclohexane, but we do not currently have an adequate approach for doing so.

Even after correcting for tautomer enumeration in the aqueous phase, there is a slight bias for cyclohexane in the calculated distribution coefficients (table 5.3 and Fig. 7). There are a variety of factors that could be contributing to this slight bias. A recent study by the Mobley group calculated partition coefficients for small molecules and found a slight bias for alcohol compounds to over favor cyclohexane[23]. Possibly, this demonstrates a limitation in GAFF or atomistic force fields in general to accurately predict the behavior of solutes in polar and non-polar environments. That same study found that for large, flexible compounds insufficient conformational sampling can dramatically affect the calculated partition coefficients [23]. Given the number of large, flexible, and polyfunctional compounds in SAMPL5, more investigation into each of these effects along with improved tautomer enumeration and handling (especially for the cyclohexane phase) will be required to completely understand the slight bias seen here.

5.4 Examining individual molecules

With only 53 molecules, it is difficult to find any statistically significant trends in terms of functional groups which are well- or poorly-predicted in general; compared to past SAMPL challenges, this set of molecules is much more complex. They are on average larger, more flexible, and contain multiple functional groups per compound. For each molecule, we organized a data set of predicted distribution coefficients and compared them to the experimental values, calculating the average unsigned error for each (Table 2). There are only three molecules with an AUE less than 2.0 log units (SAMPL5_ 003, 045, and 059). While these three molecules are relatively small, there were no trends in AUE and molecular weight, which was a trend present in SAMPL4 hydration free energy results [1]. We tried grouping molecules by functional group, molecular mass, and estimated number of tautomers to see if size, presence or absence of particular functional groups, or number of tautomers played a role in how difficult each compound was in general. The only trend found in this process was that all five carboxylic acids (SAMPL5_ 010, 011, 015, 026, and 060) are in the worst ten molecules by AUE and RMSE. This could be due to poor treatment of the effects of protonation state changes. Among the bottom compounds, perhaps unsurprisingly, was SAMPL5_083 which is a large macrocycle and SAMPL5_050; both have many neutral tautomeric forms. Most submissions had significant errors in predicting SAMPL5_074, despite the fact that it is relatively small, rigid, and has no other significant tautomers. Below we will explore why some of these molecules may have had distribution coefficients which were particularly difficult to predict.

The provided SMILES strings may not be the most populated tautomeric form of the molecule. From our tautomer enumeration and discussions with other

	SAMPL5_050		SAMPL5_083	
	tautomer 1	tautomer 2	tautomer 1	tautomer 2
$\Delta G_{hydration}$	-11.45 ± 0.04	-21.50 ± 0.03	-33.98 ± 0.07	-32.68 ± 0.1
$\Delta G_{cyclohexane}$	-13.09 ± 0.04	-13.25 ± 0.04	-35.6 ± 0.1	-36.1 ± 0.2
$\log P_{cyc/wat}$	1.20 ± 0.04	-6.04 ± 0.03	1.21 ± 0.09	2.5 ± 0.2
state penalty correction	-11.902	-0.453	-0.488	-6.53
$\log D_{cyc/wat}$	-10.70 ± 0.04	-6.50 ± 0.03	0.72 ± 0.09	-4.0 ± 0.2
experimental $\log D$	-3.2 ± 0.6		-1.9 ± 0.4	

Table 5 Shown are the results for solvation free energy of two different tautomers of SAMPL5_ 050 and 083. Corrections from $\log P$ to $\log D$ account for tautomer populations in the aqueous phase. Energies are reported in *kcal/mol*.

SAMPL5 participants [65] it became clear that accurately estimating $\log D$ for molecules with many tautomers was difficult. For example, we compared population corrections we derived using Schrödinger's LigPrep [44] with corrections calculated by Pickard *et. al.* [66,52] and found significant differences between them. If we could perfectly calculate solvation free energies and tautomer populations in both solvents, the starting tautomer should not affect the final calculated distribution coefficient, but differing corrections – such as these – will yield different results. Additionally, whenever protonation state/tautomer populations are not estimated correctly or not included in *both solvents*, the initial choice of protonation state/tautomer is likely to affect computed $\log D$ values. Here, our initial solvation free energy calculations used provided SMILES strings without any consideration of other tautomers. To explore how this may have affected our $\log D$ calculation, we decided to repeat a few solvation free energy calculations with alternate tautomers. We used SAMPL5_050 and SAMPL5_083 as examples since both have other neutral tautomers that could be present in both the water and cyclohexane solutions. Also, most participants failed to accurately predict the correct $\log D$ for either solute. The new tautomer for SAMPL5_050 was found with Ligprep [44]; the one for SAMPL5_083 was found with COSMO-RS and was provided by Andreas Klamt [65,63]. For both SAMPL5_050 and SAMPL5_083 there were significant changes in their calculated solvation free energies and partition coefficients for the two different tautomers (Table 5.4). Distribution coefficients were calculated from the $\log P$ and state penalties calculated with Schrödinger's LigPrep tool [44]. In both cases the $\log D$ is still significantly different from the experimental values. Since both the calculated solvation free energies and the tautomer/protomer populations are needed to estimate the distribution coefficient, it is impossible for us to know which calculation introduces more error into our estimates.

[CCB1]: updated water in cyclohexane section

Solvents are not completely immiscible. Though the concentration of water in cyclohexane is very small, 0.00047 mole fraction [47], it may still affect how a solute is distributed across the two solvents. This will be particularly important for solutes with many polar groups; and may be one reason it was difficult to accurate estimate the $\log D$ for SAMPL5_074. We performed two new cal-

Number Molecules	0 water	1 water	7 water
$\Delta G_{hydration}$		-21.90 ± 0.04	
$\Delta G_{cyclohexane}$	-16.77 ± 0.04	-18.05 ± 0.04	-19.54 ± 0.04
$\log D_{cyc/wat}$	-3.76 ± 0.04	-2.82 ± 0.04	-1.73 ± 0.04
experimental $\log D$		-1.9 ± 0.3	

Table 6 Shown are results for the solvation free energy and distribution coefficient for SAMPL5_074 in cyclohexane with no water present, 1 water molecule, and 7 water molecules all with 150 cyclohexane molecules and 1 solute molecule. No changes were made for the aqueous phase so the $\Delta G_{hydration}$ is the same for all three calculations. Energy is reported in $kcal/mol$.

culations of solvation free energy of SAMPL5_074 into cyclohexane with water also in the solution. These simulations were set-up with Solvation Toolkit as described above. Both sets of simulations had a single solute molecule in 150 cyclohexanes, but one had seven water molecules and the other had a single water molecule. While neither of these results reach the low experimental concentration of water in cyclohexane, they do demonstrate the dramatic variation in solvation free energy and $\log D$ which can be caused by the presence of water in cyclohexane. With varying amounts of water in cyclohexane, the water dramatically impacts the computed solvation free energy for SAMPL5_074 into cyclohexane (Table 5.4) and thus calculated $\log D$ values. In this case particular case, the presence of water also dramatically improves the estimation for $\log D$, though as noted this is with far too high a water concentration in cyclohexane. We visualized trajectories from the production phase of our calculations and find that all water molecules stay adjacent to SAMPL5_074 for the full simulation, likely indicating a particularly high affinity for water. Addition of one or a few molecules of a second solvent (water) to cyclohexane might be expected to raise the uncertainty in calculated free energies significantly due to slow mixing of the second solvent in the simulation box, but we do not observe that here. It seems likely this is because the water molecules are so strongly attracted to the solute that they essentially stay bound throughout the simulations, so we do not observe slow mixing and the associated increased uncertainty. For a solute with so many polar functional groups, it is perhaps unsurprising that the water molecules in cyclohexane are drawn to the solute. In general, this suggests that the local concentration of water near highly polar solutes may be higher than the bulk concentration in cyclohexane, and this may potentially be important when considering simulation settings. It is important to note that the concentration of water in cyclohexane for these simulations are 0.0067 and 0.047 for 1 and 7 water molecules, both a gross overestimate of the amount of water in cyclohexane (by multiple orders of magnitude) as our focus here was to explore the sensitivity of $\log D$ to cyclohexane water content. However, our results are sufficient to show that the presence of water in the cyclohexane phase can dramatically affect the computed distribution coefficient, depending on the affinity of the solute for water. To accurately account for these effects, we will need long simulations at much larger box sizes to not only match the experimental water concentration, but

to determine how much water will localize around the solute at equilibrium. Also, longer simulations will be required to account for the water molecules moving from the solute and into the cyclohexane bulk.

[DLM2]: I don't actually understand what you're getting at in this last sentence.

Other buffer/solution components may affect distribution coefficients. The two phases for the distribution coefficients were cyclohexane and an aqueous buffer, however, DMSO and acetonitrile were used in the experiments [8] as well. While DMSO and acetonitrile were at very low concentrations, their presence in either solvent layer may affect how a solute distributes between phases. We created topology and coordinate files for a system with 780 water, 130 cyclohexane, 4 DMSO, and 2 acetonitrile molecules using SolvationToolkit, roughly matching the experimental concentrations. In the trajectory visualized with VMD, both DMSO and acetonitrile spend most of their time near the solvent interface, with very little movement into the bulk of the water or cyclohexane. In our view, this does not immediately suggest profound implications for calculated $\log D$ values, but a detailed understanding of the implications of this for both calculated and measured distribution coefficients will require further study.

6 Conclusions

Past SAMPL challenges often involved a broad range of methods for hydration free energies. Here, in our first SAMPL on cyclohexane/water distribution coefficients, we saw a similarly diverse set of methods. Predicting $\log D$ accurately for this set of molecules was rather difficult, exhibiting a number of the same challenges that will face accurate prediction of binding affinities. The best methods showed reasonable agreement with experiment with RMSEs around 2.5 log units and Kendall tau's around 0.6. However, considering that the null model and $XlogP_{corr}$ both would have topped the submissions list by RMSE and AUE, there is clear room for all of the methods employed in SAMPL5 to improve.

This SAMPL5 set was substantially more complex, flexible, and polyfunctional than typical molecules in SAMPL hydration challenge sets. Additionally, the most relevant protonation state and tautomer were not always clear for the compounds. Some compounds likely had multiple relevant protonation states and tautomers, and shifts in protonation/tautomeric state on transferring between phases. These issues are especially important given that the challenge focused on distribution coefficients rather than partition coefficients ($\log P$). Given these complexities, it is not surprising that we saw drop in performance relative to the accuracy that would have been expected for $\log D$ values if participants predicted $\log D$ based on solvation free energies in two solvents that could be computed as accurately as hydration free energies in previous SAMPL challenge (yielding an expected accuracy of about 1.5 log units). In

large part, this is probably because of the additional complexities of distribution coefficients such as the need to account for other protonation states and tautomers. Our solvation free energy calculations performed with a less dominant tautomer of SAMPL5_050 lead to a $\log D$ estimate of -10.70 ± 0.04 which is 7.5 log units below the experimental value. Thus, accurately accounting for tautomers appears to be a vital part of accurately calculating $\log D$, and better methods for treating protonation and especially tautomeric states in non-aqueous environments are needed.

As mentioned earlier, the molecules chosen for SAMPL5 are generally larger and more flexible than those in past SAMPL challenges. It follows that conformational sampling of the solutes might play a significant role in accurately predicting these distribution coefficients. In a previous study using the same protocol as the reference calculation here, the Mobley lab showed that changing the initial conformation of a large, flexible molecule dramatically changed its calculated partition coefficient due to insufficient conformational sampling [23]. In this issue, Tyler Luchko *et. al* spend time addressing the importance of conformational sampling for these molecules in the context of their submission results for the challenge [59].

We asked participants to estimate two forms of uncertainty, statistical uncertainty and model uncertainty, the latter of which should predict how well their calculation will agree with experiment. This latter uncertainty estimate is particularly key, as it would allow practitioners to predict how reliable their calculations are likely to be in applications. Here, we find that almost every participant dramatically underestimated their model uncertainty. The importance for the community to improve error estimation has been addressed in past SAMPL challenges [1], but clearly, much more work is still needed.

A common trend in SAMPL5 submissions was that the dynamic range in the predicted distribution coefficients was larger than the dynamic range observed experimentally – that is, generally, the smallest $\log D$ values were underestimated and the largest were overestimated. This issue of dynamic range is visible in figures 2, 5, and 7, but was evident in comparison plots for almost all submissions. For SAMPL5_074, the introduction of water to cyclohexane increased the $\log D$ estimate from -3.76 to -1.73; it is possible that accounting for water in cyclohexane would affect the apparent underestimation at the lower end of the $\log D$ scale. As discussed above, it is also possible that insufficient conformational sampling could account for some of this problem. For example, if a solute has a particular conformation that is more stable in a non-polar environment and that conformation was not found for the solute in cyclohexane due to insufficient conformational sampling, then the calculated distribution coefficient would over favor the aqueous environment. However, this will require further investigation. It is of course also possible that experimental issues could have compressed the measured dynamic range of compounds, but this too will require further investigation.

The results of this challenge strongly suggest predictions for solute partitioning will be extremely helpful for driving improvements to physical modeling needed in pharmaceutical research. The major challenges encountered

here are all very likely to occur when attempting to predict binding affinities or other biomolecular properties of interest to drug discovery. Specifically, accurately predicting the population of protonation and tautomeric states was a challenge, complicated by the fact that there is no simple way to predicted protonation states and tautomers to corresponding experimental values relevant to the conditions studied here. Most work so far on tautomer prediction has focused on tautomeric ratios in vacuum or in water, but tautomer populations are likely environment-dependent in ways that can dramatically affect computed physical properties. These same challenges are likely to apply when predicting biomolecular binding. An improved treatment of these effects within the context of SAMPL or similar challenges will drive advances in computational techniques also used to predict binding and related properties such as solubility.

Overall, distribution coefficients have been an extremely valuable part of this year's SAMPL5 challenge, and, since they can be measured in a relatively straightforward way, seem to be a promising potential source of future data for blind challenges. Additionally, this data highlights important issues, such as tautomer enumeration, that need better treatment in many of our models. The ability to create new, completely blind data sets make distribution coefficients a great option for future challenges.

6.1 Supporting Information

Supporting information is available free of charge online through the University of California Dash at http://n2t.net/ark:/b7280/d1988w It includes all of the files provided to SAMPL5 participants. That includes a table of SAMPL5_IDnumbers paired with SMILES, MOL2 files for each molecule, and the GROMACS, AMBER, DESMOND, and LAMMPS topology and coordinate files for each solute in water and cyclohexane. A separate directory is included with analysis associated with our reference calculations and post sample method development. This includes all python scripts, input files, output files, and results files required to repeat our simulations and calculations done with Schrödinger tools. We also provide the data files for all submission,

the scripts used for error analysis, plots for all submissions, and data for only batch 0 and batches 0 and 1. This is supported with detailed README files explaining the structure of the directories.

References

1. D.L. Mobley, K.L. Wymer, N.M. Lim, J.P. Guthrie, Journal of Computer-Aided Molecular Design **28**(3), 135 (2014)
2. M.T. Geballe, J.P. Guthrie, Journal of Computer-Aided Molecular Design **26**(5), 489 (2012)
3. M.T. Geballe, A.G. Skillman, A. Nicholls, J.P. Guthrie, P.J. Taylor, Journal of Computer-Aided Molecular Design **24**(4), 259 (2010)
4. P.V. Klimovich, D.L. Mobley, Journal of Computer-Aided Molecular Design **24**(4), 307 (2010)
5. D.L. Mobley, C.I. Bayly, M.D. Cooper, K.A. Dill, The Journal of Physical Chemistry B **113**(14), 4533 (2009)
6. D.L. Mobley, S. Liu, D.S. Cerutti, W.C. Swope, J.E. Rice, Journal of Computer-Aided Molecular Design **26**(5), 551 (2012)
7. A. Nicholls, D.L. Mobley, J.P. Guthrie, J.D. Chodera, C.I. Bayly, M.D. Cooper, V.S. Pande, Journal of Medicinal Chemistry **51**(4), 769 (2008)
8. A.S. Rustenburg, J. Dancer, B. Lin, J.A. Feng, D.F. Ortwine, D.L. Mobley, J.D. Chodera, Journal of Computer-Aided Molecular Design (2016)
9. A. Leo, C. Hansch, D. Elkins, Chemical Reviews **71**(6), 525 (1971)
10. R.J. Young, D.V.S. Green, C.N. Luscombe, A.P. Hill, Drug discovery today **16**(17-18), 822 (2011)
11. J.W. Essex, C.A. Reynolds, W.G. Richards, J. Am. Chem. Soc. **114**(10), 3634 (1992)
12. S.A. Best, K.M. Merz Jr, C.H. Reynolds, J. Phys. Chem. B **103**(4), 714 (1999)
13. J.E. Eksterowicz, J.L. Miller, P.A. Kollman, J. Phys. Chem. B **101**(50), 10971 (1997)
14. W.L. Jorgensen, Accounts of Chemical Research **22**, 187 (1989)
15. W.L. Jorgensen, J.M. Briggs, L. Contreras, Journal of Physical . . . **94**(4), 1683 (1990)
16. N.M. Garrido, A.J. Queimada, M. Jorge, E.A. Macedo, I.G. Economou, Journal of Chemical Theory and Computation **5**(9), 2436 (2009)
17. N.M. Garrido, M. Jorge, A.J. Queimada, J.R.B. Gomes, I.G. Economou, E.A. Macedo, Physical Chemistry Chemical Physics **13**(38), 17384 (2011)
18. N.M. Garrido, I.G. Economou, A.J. Queimada, M. Jorge, E.A. Macedo, AIChE J. **58**(6), 1929 (2012)
19. L. Yang, A. Ahmed, S.I. Sandler, Journal of Computational Chemistry **34**(4), 284 (2013)
20. J. Michel, M. Orsi, J.W. Essex, J. Phys. Chem. B **112**(3), 657 (2007)
21. S. Genheden, Journal of Chemical Theory and Computation **12**(1), 297 (2016)
22. I. OpenEye Scientific Software. Oechem. http://www.eyesopen.com (2010). URL www.eyesopen.com
23. C.C. Bannan, G. Calabró, D.Y. Kyu, D.L. Mobley, Journal of Chemical Theory and Computation **12**(8), 4015 (2016)
24. M.B. Wilk, R. Gnanadesikan, Biometrika **55**(1), 1 (1968)
25. H.J.C. Berendsen, D. Van Der Spoel, R. van Drunen, Comput. Phys. Commun. **91**(1-3), 43 (1995)
26. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theory Comput. **4**(3), 435 (2008)
27. E. Lindahl, B. Hess, D. van der Spoel, J. Mol. Model. **7**(8), 306 (2001)
28. D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, J. Comput. Chem. **26**(16), 1701 (2005)
29. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, Bioinformatics (Oxford, England) **29**(7), 845 (2013)
30. S. Páll, M.J. Abraham, C. Kutzner, B. Hess, E. Lindahl, in *Solving Software Challenges for Exascale*, vol. 8759 (Springer International Publishing, Stockholm, Sweden, 2014), pp. 3–27

31. M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, Software X **1-2**, 19 (2015)
32. J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Journal of Computational Chemistry **25**(9), 1157 (2004)
33. A. Jakalian, B.L. Bush, D.B. Jack, C.I. Bayly, Journal of Computational Chemistry **21**(2), 132 (2000)
34. A. Jakalian, D.B. Jack, C.I. Bayly, Journal of Computational Chemistry **23**(16), 1623 (2002)
35. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, The Journal of Chemical Physics **79**(2), 926 (1983)
36. S. Liu, S. Cao, K. Hoang, K.L. Young, A.S. Paluch, D.L. Mobley, Journal of Chemical Theory and Computation **12**(4), 1930 (2016)
37. P.V. Klimovich, M.R. Shirts, D.L. Mobley, Journal of Computer-Aided Molecular Design **29**(5), 397 (2015)
38. S. Parameswaran, D.L. Mobley, Journal of Computer-Aided Molecular Design **28**(8), 825 (2014)
39. D.R. Lide (ed.), *CRC handbook of chemistry and physics*, 76th edn. (CRC press, Boca Raton, 1996)
40. J. Sangster, J. Phys. Chem. Ref. Data **18**, 1111 (1989)
41. **Schrödinger Release 2014-4**: Epik, version 3.0, Schrödinger, LLC, New York, NY, 2014
42. J.C. Shelley, A. Cholleti, L.L. Frye, J.R. Greenwood, M.R. Timlin, M. Uchimaya, Journal of Computer-Aided Molecular Design **21**(12), 681 (2007)
43. J.R. Greenwood, D. Calkins, A.P. Sullivan, J.C. Shelley, Journal of Computer-Aided Molecular Design **24**(6-7), 591 (2010)
44. **Schrödinger Release 2014-4**: Ligprep, version 3.2, Schrödinger, LLC, New York, NY, 2014
45. R. Wang, Y. Fu, L. Lai, Journal of Chemical Information and Modeling **37**(3), 615 (1997)
46. R. Wang, Y. Gao, L. Lai, Perspectives in Drug Discovery and Design **19**(1), 47 (2000)
47. C. Black, G.G. Joris, H.S. Taylor, The Journal of Chemical Physics **16**(5), 538 (1948)
48. W. Humphrey, A. Dalke, K. Schulten, Journal of Molecular Graphics **14**(1), 33 (1996)
49. S.S. Paranahewage, C.S. Gierhart, C.J. Fennell, Journal of Computer-Aided Molecular Design (2016)
50. B. Iorga, I.M. Kenney, O. Beckstein, Journal of Computer-Aided Molecular Design (2016)
51. S. Bosisio, A.S.J.S. Mey, J. Michel, Journal of Computer-Aided Molecular Design (2016)
52. F. Pickard, G. König, F. Tofoleanu, J. Lee, A. Simmonett, Y. Shao, J. Ponder, B.R. Brooks, Journal of Computer-Aided Molecular Design (2016)
53. G. König, F.C. Pickard, J. Huang, A.C. Simmonett, F. Tofoleanu, J. Lee, P.O. Dral, F.N.U. Samarjeet, M. Jones, Y. Shao, W. Thiel, B.R. Brooks, Journal of Computer-Aided Molecular Design (2016)
54. S. Genheden, J. Essex, Journal of Computer-Aided Molecular Design (2016)
55. G. Kamath, I. Kurnikov, B. Fain, I. Leontyev, A. Illarionov, O. Butin, M. Olevanov, L. Pereyaslavets, Journal of Computer-Aided Molecular Design (2016)
56. E. Brini, S.S. Paranahewage, C.J. Fennell, K.A. Dill, Journal of Computer-Aided Molecular Design (2016)
57. M.R. Jones, B.R. Brooks, A.K. Wilson, Journal of Computer-Aided Molecular Design (2016)
58. N. Tielker, D. Tomazic, J. Heil, T. Kloss, S. Ehrhart, S. Güssregen, K.F. Schmidt, S. Kast, Journal of Computer-Aided Molecular Design (2016)
59. T. Luchko, N. Blinov, G.C. Limon, K.P. Joyce, A. Kovalenko, Journal of Computer-Aided Molecular Design (2016)
60. S. Diaz-Rodriguez, S.M. Bozada, J.R. Phifer, A.S. Paluch, Journal of Computer-Aided Molecular Design (2016)
61. H. Park, K.C. Chung, Journal of Computer-Aided Molecular Design (2016)
62. D. Santos-Martins, P.A. Fernandes, M.J.a. Ramos, Journal of Computer-Aided Molecular Design (2016)

63. A. Klamt, F. Eckert, J. Reinisch, K. Wichmann, Journal of Computer-Aided Molecular Design (2016)
64. C.J. Fennell. Personal Communication (2016)
65. A. Klamt. Personal Communication (2016)
66. F.C. Pickard IV. Personal Communication (2016)